Intraday liquidity risk estimation using transaction data: an extreme value theory approach

Sofie van den Hoogen

August, 2017

A THESIS SUBMITTED TO THE

DELFT INSTITUTE OF APPLIED MATHEMATICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

IN

APPLIED MATHEMATICS

SUPERVISORS AND REVIEWERS

J. J. CAI (DELFT UNIVERSITY OF TECHNOLOGY)

G. JONGBLOED (DELFT UNIVERSITY OF TECHNOLOGY)

D. KUROWICKA (DELFT UNIVERSITY OF TECHNOLOGY)

E. NIKOLAKAKIS (ABN AMRO BANK N.V.)

Abstract

Intraday liquidity risk is a subject that applies to all banks, and arises whenever there is a timing mismatch between incoming and outgoing payments within a business day. In case such a mismatch occurs, the bank is exposed to the risk that it is unable to meet its payment obligations at the time expected. A liquidity buffer could help to mitigate this risk.

This thesis presents a framework for intraday liquidity risk management within ABN AMRO Bank, while taking different priorities of transactions into account. We examine the use of extreme value theory (EVT) and propose two metrics to capture the risk: the univariate and multivariate risk metric. The univariate risk metric represents the size of the liquidity buffer for each priority group separately and provides granular view. Making use of a Monte Carlo simulation algorithm in combination with univariate EVT, we are able to estimate the size of the liquidity buffer for a specified time interval within a business day. We forecast the buffer size 30 days out-of-sample and test the violations against the conditional coverage (CC) hypothesis. Satisfactory results are obtained for the groups with high and moderate priority when the highest confidence levels are considered: $\alpha = 0.1$ and 0.05. For the group with low priority, the risk metric performs well for the lowest confidence levels: $\alpha = 0.025$ and 0.01. The multivariate risk metrics aggregates the size of the liquidity buffer, while taking the diversification of the priority groups into account. We define a failure set and investigate the use of multivariate EVT.

Contents

List of Abbreviations

ACD	Autoregressive Conditional Duration
BM	Block Maxima
CC	Conditional Coverage
CLT	Central Limit Theorem
EVT	Extreme Value Theory
DNB	Dutch Central Bank
DQ	Dynamic Quantile
EBA	European Bank Authority
GEV	Generalized Extreme Value
GP	Generalized Pareto
IID	Independent and Identically Distributed
ILAAP	Internal Liquidity Adequacy Assessment Process
IND	Independence
IRM	Intraday Risk Measure
IVaR	Intraday Value at Risk
LM	Lagrange Multiplier
LCRM	Liquidity and Capital Risk Management
ML	Maximum Likelihood
POF	Proportion Of Failures
POT	Peaks Over Threshold
QML	Quasi Maximum Likelihood
UC	Unconditional Coverage
VaR	Value at Risk

List of Figures

List of Tables

1

Introduction

1.1 Intraday liquidity risk

Regulators are concerned with the protection of the financial system against catastrophic events. In recent years, a central issue in risk management has been to determine capital and liquidity requirements for financial institutions to meet catastrophic risk (Bali, 2007, [Bal07]).

One specific type of risk concerns intraday liquidity risk, which is a subject that applies to all banks (Neijs et al., 2015, [NW15]). Intraday liquidity refers to the resources that can be accessed by the bank during a business day, usually to enable banks to make payments in real time. Intraday liquidity risk arises whenever there is a mismatch in timing between incoming and outgoing payments during a business day. Whenever such a timing mismatch occurs, the bank is exposed to the risk that it is unable to meet its payment obligations at the time expected (Ball et al., 2011, [BDM11]). This may impact the liquidity position of the bank and, in extreme situations due to the high interdependency between different payment systems, that of other parties.

In order to moderate these risks and to strengthen the resilience of banks in times of extreme financial distress, a number of regulatory measures have been introduced. The 2008 - 2009 financial crisis has driven the introduction of the Basel III regulatory framework. Basel III focuses on strengthening capital and liquidity regulations in the international banking sector (Mohanta, 2014, [Moh14]). In particular, Basel III requires banks to implement monitoring tools for intraday liquidity management. The exact requirement concerning intraday liquidity management (i.e. BCBS 248: principal 8) is stated below.

"A bank should actively manage its intraday liquidity positions and risks to meet payment and settlement obligations on a timely basis under both normal and stressed conditions and thus contribute to the smooth functioning of payment and settlement systems."

In addition to Basel III, the Dutch Central Bank (DNB) introduced the Internal Liquidity Adequacy Assessment Process (ILAAP) in the Netherlands in June 2011. Dutch banks are expected, when reporting the ILAAP to DNB, to give a description together with an internal assessment on the way in which liquidity risk is managed within the organization. Regarding this assessment, it is important for banks to address how the required liquidity buffer is calculated. Hence, in this thesis the size of the liquidity buffer will be the risk metric of interest.

1.2 Research question

ABN AMRO Bank distinguishes different types of priorities among their payments, both for incoming and outgoing payments. A payment is considered to be high priority, moderate priority or low priority. A payment with a high priority has to be paid immediately, while a low priority payments could be postponed. Hence, it is desired to take these three different priorities into account while estimating the size of the liquidity buffer of ABN AMRO Bank. This thesis aims to answer the following research question:

"How should the intraday liquidity buffer size of ABN AMRO Bank, when a distinction is made between different priority groups, be determined?

(1) *Determine the liquidity buffer size for each group separately within a univariate framework.*

(2) *Determine the aggregated liquidity buffer size within a multivariate framework.*

In this thesis, item (1) will be referred to as the univariate risk metric and item (2) will be referred to as the multivariate risk metric. The univariate risk metric provides a granular analysis of intraday liquidity risk for each priority group separately and allows for monitoring. The multivariate risk metric specifies the aggregated level of risk and takes the diversification of these groups into account. Hence, by developing these risk metrics we develop a practice to determine the intraday liquidity buffer size in both normal and stressed financial conditions. This thesis proposes an extreme value theory (EVT) application to obtain these risk metrics.

1.3 Outline

Figure 1.1 illustrates the outline of this thesis and highlights the links between the different chapters. Three different layers are distinguished. The first layer is the research framework, in which the research question and the data set is presented. Chapter 2 introduces the ABN AMRO Bank transaction data set, which is considered to be ultra high frequency.

The second layer introduces the theoretical fundamentals on which our application will rely. Chapter 3 presents results on both univariate and multivariate EVT. The results are rather theoretical and especially recommended for people who are not familiar with EVT. Subsequent chapters will be build upon the results.

The third layer consists of Chapter 4, 5 and 6 and presents the risk metrics. Chapter 4 describes the class of Autoregressive Conditional Duration (ACD) models, which are used to model the durations between high frequency transactions. Chapter 5 relies on the results of univariate EVT of Chapter 3 and the ACD model of Chapter 4. This chapter introduces a Monte Carlo simulation algorithm, where the time steps are defined by the ACD model of Chapter 4. The algorithm enables us to estimate our univariate risk metric: the size of the liquidity buffer size of each group separately. Chapter 6 relies heavily on the results of multivariate EVT of Chapter 3, and presents the multivariate risk metric: the size of aggregated liquidity buffer size.

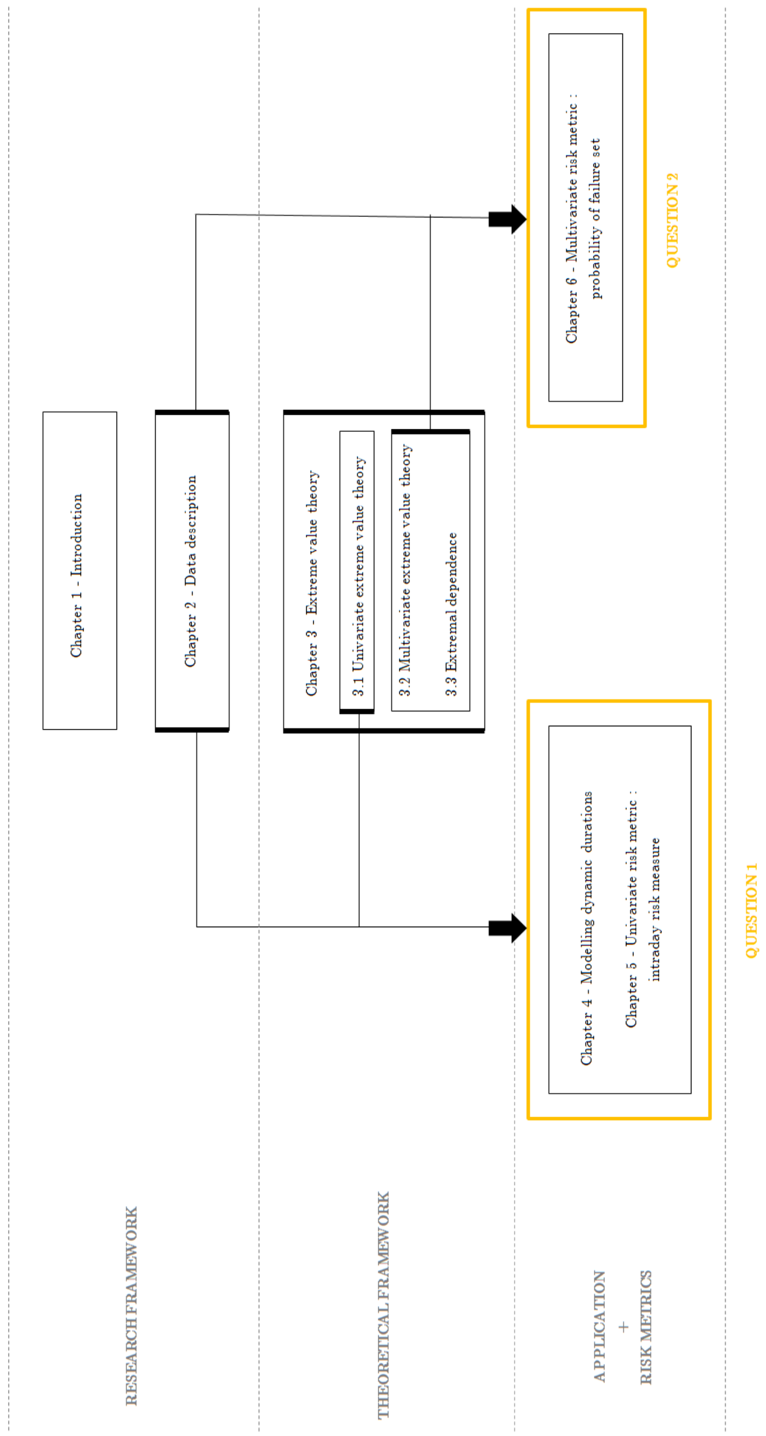

Figure 1.1: Graphic representation of the thesis outline.

2

Data description

2.1 Introduction

Author's note: this section is confidential.

2.2 Description of transaction data set

Author's note: this section is confidential.

2.2.1 Notation of transaction data set

Author's note: this section is confidential.

2.3 Description of priority transaction data set

Author's note: this section is confidential.

2.3.1 Notation regarding segmentation

Author's note: this section is confidential.

3

Extreme value theory

3.1 Introduction

The study of exceptional risks has become an important subject in probability and statistical research, as these type of risks can have major impact. Extreme value theory (EVT) is considered as a classical mathematical framework to evaluate these exceptional risks (Cai et al., 2013, [CFM13]). EVT provides an asymptotic distribution for extreme data, analogous to the way that the Central Limit Theorem (CLT) suggests the normal distribution as an approximate model for the sample means of a large enough sample (Northrop et al., 2011, [NJ11]).

This chapter is rather theoretical and serves as an introduction on EVT. Results on both univariate and multivariate EVT are presented. These results will be applied in the subsequent chapters. The univariate EVT will be used in Chapter 5, which enables us to answer the first research question. The multivariate EVT will play a key role in Chapter 6, which addresses the second research question.

3.2 Univariate extreme value theory

Univariate EVT is concerned with finding the limit distribution for extreme observations of a sequence of independent and identically distributed random variables. Two fundamental approaches can be distinguished in the way these extreme observations are identified: the Block Maxima (BM) approach and the Peaks Over Threshold (POT) approach. We start by making an intuitive distinction between these two methods.

The BM approach divides the observation period into equally sized and non overlapping blocks in order to obtain the maximum realization in each block. The limiting distribution arising from these maxima is the generalized extreme value (GEV) distribution. The POT approach focuses on the realizations exceeding a given (high) threshold and searches for a distribution that is able to capture the limit behavior of these exceedances. It turns out that the limit behavior of these exceedances can be captured by the generalized pareto (GP) distribution. Figure 3.1 gives a graphic representation of both methods for the realizations $x_1, x_2, \ldots, x_{16}$ of an arbitrary process. The vertical lines in the left figure represent the blocks, where the crosses form the maxima. The horizontal line in the right figure defines the threshold, where the crosses represent the corresponding exceedances of this threshold. The BM approach identifies the realizations $x_2,\ x_5,\ x_{12},\ x_{14}$ as extreme events, while the POT approach considers the realizations $x_2,\ x_3,\ x_5,\ x_7,\ x_8,\ x_{10},\ x_{11},\ x_{12},\ x_{14},\ x_{16}$ to be extreme.

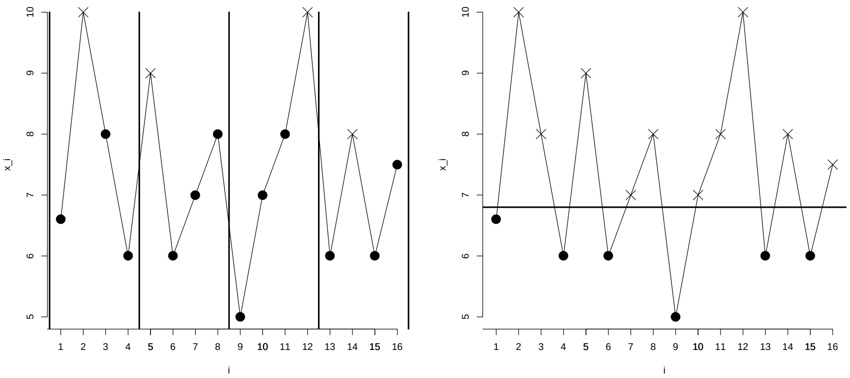

Figure 3.1: Graphic representation of the BM approach (left) and the POT approach (right) to define extreme events for an arbitrary sequence of realizations.

It should be noted that the BM approach could miss some of the high observations and might retain some lower observations. Hence, the BM approach is considered to be wasteful in case other approaches are present for the identification of extremes observations (Coles, 2011, [Col11]). Hence, the POT approach is preferred in this thesis.

3.2.1 Peaks Over Threshold approach

This section elaborates the general mathematical framework of the POT approach. Section 3.2.1.1 introduces the GEV distribution, the extreme value index and the max-domain of attraction. Definition 3.1, Remark 3.2 and Theorem 3.1 form the highlights of this section. These highlights enables us to the link the extreme value index with the GP distribution. The GP distribution is presented in Section 3.2.1.2.

3.2.1.1 GEV distribution and max-domain of attraction

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables with common cumulative distribution function (cdf) F. As extremes intuitively occur near the upper end of the support of F, define the right endpoint of F by $x_F = \sup\left\{x \in \mathbb{R} \ : \ F(x) < 1\right\}$ which may be infinite. We define the partial maxima $M_n = \max(X_1, X_2, \ldots, X_n)$. It can be shown that, under the assumption of independent and identically distributed random variables, the distribution of these partial maxima can be expressed by

$$P(M_n \leq x) = (F(x))^n \ . \tag{3.1}$$

It should be noticed that $P(M_n \leq x) = (F(x))^n \to 0$ as $n \to \infty$ for all $x < x_F$ and $P(M_n \leq x) = (F(x))^n = 1$ as $n \to \infty$ for all $x \geq x_F$. This implies that the limit distribution of the partial maxima converges to a degenerate distribution. Thus, in order to find a non degenerate limit distribution for the partial maxima, a proper transformation is necessary. Suppose there exist two sequences of real numbers $a_n > 0$ and $b_n \in \mathbb{R}$ $(n = 1, 2, \ldots)$ and a non degenerate distribution function G such that

$$\lim_{n\to\infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F^n(a_n x + b_n) = G(x) \tag{3.2}$$

for every continuity point x of G. Definition 3.1 shows all distribution functions G that can occur as a limit, which belong to the class of generalized extreme value (GEV) distributions. Therefore, we call G an extreme value distribution. This implies that the distribution G, i.e. the limit distribution of the properly transformed maxima, is completely characterized by a one parameter family. Note that via the relation $\min(X_1, X_2, \ldots, X_n) = -\max(-X_1, -X_2, \ldots, -X_n)$ the same results can be applied for minima too.

Definition 3.1 (Class of extreme value distributions)
The class of extreme value distributions is given by

$$G_\gamma(x) = \begin{cases} \exp\left(-(1+\gamma x)_+^{-\frac{1}{\gamma}}\right) & \gamma \neq 0 \\ \exp\left(-\exp\left(-x\right)\right) & \gamma = 0 \end{cases} \ , \tag{3.3}$$

where $\gamma \in \mathbb{R}$ is referred to as the extreme value index. The notation $(\cdot)_+ = \max(\cdot, 0)$ is used.

Remark 3.1 (Block Maxima approach)
The class of extreme value distributions, denoted by Equation (3.3), are the possible limiting distributions of the partial maxima. This approach, in order to obtain the limit distribution for the tail observations of a sample, is known as the BM approach.

However, it should be noted that the convergence of the properly transformed partial maxima to an extreme value distribution G_γ does not hold in general. Therefore, we are interested in conditions on the cdf F that ensure that there indeed exist two sequences of real numbers $a_n > 0$ and $b_n \in \mathbb{R}$ $(n = 1, 2, \ldots)$ such that properly transformed partial maxima converge to an extreme value distribution G_γ. In order to state these conditions, first the concept of max-domain of attractions is introduced.

Definition 3.2 (Max-domain of attraction)
The cdf F is said to be in the max-domain of attraction of an extreme value distribution G_γ if there exist two sequences of real numbers $a_n > 0$ and $b_n \in \mathbb{R}$ $(n = 1, 2, \ldots)$ such that

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(x) \tag{3.4}$$

for each continuity point x of G. We write $F \in MDA(G_\gamma)$.

Alternatively, by taking logarithms of both sides of Equation (3.4), the concept of max-domain of attraction can also be described by Equation (3.5). This is denoted by Collorary 3.1.

Corollary 3.1
Alternatively, we could write: for each continuity point x for which $0 < G_\gamma(x) < 1$ we have

$$\lim_{n\to\infty} n\left(1 - F(a_n x + b_n)\right) = -\log G_\gamma(x)\ . \tag{3.5}$$

Proof 3.1

We refer to Appendix E.1 for the proof.

□

Remark 3.2
The extreme value index $\gamma \in \mathbb{R}$ characterizes the heaviness of the tail of the distribution.

Based on the sign of the extreme value index γ, the class of generalized extreme value distribution can be divided into three categories.

- **$\gamma > 0$**

  The case $\gamma > 0$ represents the Fréchet max-domain of attraction, and translates into Equation (3.6). We write $F \in MDA(G_{\gamma>0})$. The distribution has a heavy right tail with an infinite right endpoint, i.e. $x_F = \infty$. Moments of order greater than $\frac{1}{\gamma}$ do not exist. Examples of distributions in the Fréchet domain of attraction are Cauchy, Student and Pareto distributions.

$$G_{\gamma>0}(x) = \begin{cases} 0 & x \leq 0 \\ \exp\left(-x^{-\frac{1}{\gamma}}\right) & x > 0 \end{cases} \tag{3.6}$$

- **$\gamma = 0$**

  The case $\gamma = 0$ represents the Gumbel max-domain of attraction, and translates into Equation (3.7). We write $F \in MDA(G_{\gamma=0})$. The distribution has a light right tail with an infinite or finite right endpoint. All moments exist. Examples of distributions in the Gumbel domain of attractions are the normal and gamma distribution.

$$G_{\gamma=0}(x) = \exp\left(-\exp\left(-x\right)\right) \tag{3.7}$$

- **$\gamma < 0$**

  The case $\gamma < 0$ represents the Weibull max-domain of attraction, and translates into Equation (3.8). We write $F \in MDA(G_{\gamma<0})$. The distribution has a finite right endpoint, i.e. $x_F < \infty$. All moments exist. Examples of distributions in the Weibull domain of attraction are the uniform and beta distribution.

$$G_{\gamma<0}(x) = \begin{cases} 1 & x \geq 0 \\ \exp\left(-\left(-x\right)^{-\frac{1}{\gamma}}\right) & x < 0 \end{cases} \tag{3.8}$$

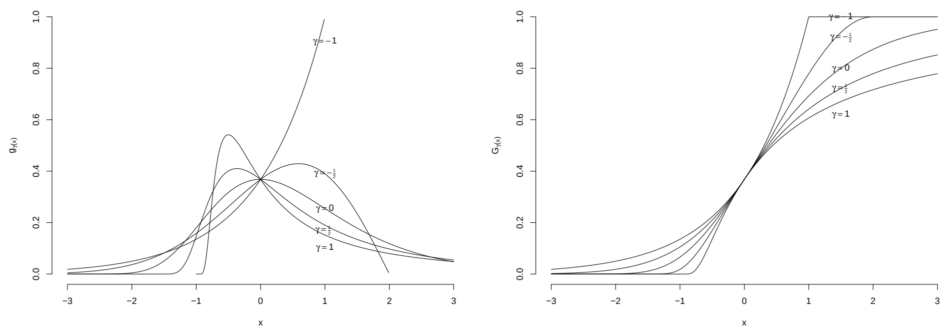

Figure 3.2: Probability density function (left) and cumulative distribution function (right) of the GEV distribution for $\gamma = -1,\ -\frac{1}{2}$ (with upper end points $x = 1,\ 2$), $\gamma = 0,\ \frac{1}{2}$ and $\gamma = 1$ (with lower end point $x = -1$).

Theorem 3.1, which reformulates the definition of max-domain of attraction, puts conditions on the cdf F that ensure that there exist two sequences of real numbers $a_n > 0$ and $b_n \in \mathbb{R}$ $(n = 1, 2, \ldots)$ such that properly transformed partial maxima converge to an extreme value distribution G_γ. As this theorem includes the concept of tail quantile functions, first the tail quantile function is introduced. Definition 3.3 introduces the concept of tail quantile functions.

Definition 3.3 (Tail quantile function)
Let F be a cdf with left continuous inverse $F^{\leftarrow}(x) = \inf\left\{t \ : \ F(t) \geq x\right\}$. Then the tail quantile function U is given by

$$U(t) = F^{\leftarrow}\left(1 - \frac{1}{t}\right) \ . \tag{3.9}$$

Theorem 3.1
For the extreme value index $\gamma \in \mathbb{R}$ the following statements are equivalent

1. *$F \in MDA(G_\gamma)$.*

2. *There exist real constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that for all x*

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(x) \ . \tag{3.10}$$

3. *There exists a positive function a such that*

$$\lim_{t\to\infty} t\left(1 - F\left(a(t)x + U(t)\right)\right) = (1 + \gamma x)_+^{-\frac{1}{\gamma}} \ , \tag{3.11}$$

for all x. The notation $(\cdot)_+ = \max(\cdot, 0)$ is used.

4. *There exists a positive function a such that for $x > 0$*

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma} \ , \tag{3.12}$$

where for $\gamma = 0$ the right-hand side is interpreted as $\log x$.

5. There exists a positive function u such that

$$\lim_{t \uparrow x_F} \frac{1 - F(t + xu(t))}{1 - F(t)} = (1 + \gamma x)_+^{-\frac{1}{\gamma}} \ , \tag{3.13}$$

for all x. The notation $(\cdot)_+ = \max(\cdot, 0)$ is used.

Moreover, Equation (3.10) holds with $a_n := a(n)$ and $b_n := U(n)$. Also, Equation (3.14) holds with $u(t) = a\left(\frac{1}{1-F(t)}\right)$.

Proof 3.2

We refer to (de Haan et al., 2006, [HF06]) for the proof.

□

Example 3.1

Assume we have a sequence $X_1, X_2, \ldots, X_n$ of independent and identically distributed random variables that follow the standard exponential distribution, i.e. $X_i \overset{IID}{\sim} Exp(1)$. This implies that we have cdf $F(x) = 1 - e^{-x}$ for $x > 0$. It is easy to verify that we get $a_n = 1$, $b_n = \log n$ and $\gamma = 0$. Thus, $F \in MDA(G_{\gamma=0})$ and by choosing a_n and b_n accordingly, $M_n - \log n$ converges to the Gumbel distribution.

Example 3.2

Assume we have a sequence $X_1, X_2, \ldots, X_n$ of independent and identically distributed random variables that follow the uniform distribution on the interval $(0, 2)$, i.e. $X_i \overset{IID}{\sim} U(0, 2)$. This implies that we have cdf $F(x) = \frac{x}{2}$ for $0 \leq x \leq 2$. It is easy to verify we get $a_n = \frac{2}{n}$, $b_n = 2 - \frac{2}{n}$ and $\gamma = -1$. Thus, $F \in MDA(G_{\gamma<0})$ and by choosing a_n and b_n accordingly, $n\left(\frac{M_n}{2} - 1\right) + 1$ converges to the Weibull distribution.

3.2.1.2 GP distribution

Theorem 3.1 forms the basis of the POT approach, as it provides the bridge between the POT approach and the introduced framework of the BM approach. Item (5) of Theorem 3.1 enables us to establish a link between the BM approach and the POT approach. We can reformulate item (5) of Theorem 3.1 as follows

$$
\begin{aligned}
\lim_{t \uparrow x_F} \frac{1 - F(t + xu(t))}{1 - F(t)} &= \lim_{t \uparrow x_F} \frac{1 - P(X \leq t + xu(t))}{1 - P(X \leq t)} \\
&= \lim_{t \uparrow x_F} \frac{P(X > t + xu(t))}{P(X > t)} \\
&= \lim_{t \uparrow x_F} \frac{P(X > t + xu(t) \ \cap \ X > t)}{P(X > t)} \quad \forall\, x \geq 0 \\
&= \lim_{t \uparrow x_F} P\left(X > t + xu(t) \;\middle|\; X > t\right) \quad \forall\, x \geq 0 \\
&= \lim_{t \uparrow x_F} P\left(\frac{X - t}{u(t)} > x \;\middle|\; X > t\right) \quad \forall\, x \geq 0 \, . \qquad (3.14)
\end{aligned}
$$

Let X be a random variable with cdf $F \in MDA(G_\gamma)$. Let t be a threshold large enough. Define the sequence $\frac{X_1 - t}{u(t)}, \ \frac{X_2 - t}{u(t)}, \ \ldots, \ \frac{X_n - t}{u(t)}$ of independent threshold exceedances scaled by an appropriate factor. Then the limit distribution, conditional on X exceeding t, is given by Equation (3.15) as $t \uparrow x_F$. This class of distribution functions is referred to as the class of generalized pareto (GP) distributions. The formal definition of the class of GP distributions is stated below.

Definition 3.4 (Class of generalized pareto distributions)
The class of generalized pareto distributions is given by

$$
G_v(x) = \lim_{t \uparrow x_F} P\left(\frac{X - t}{u(t)} \leq x \;\middle|\; X > t\right) = \begin{cases} 1 - (1 + \gamma x)_+^{-\frac{1}{\gamma}} & \gamma \neq 0 \\ 1 - \exp(-x) & \gamma = 0 \end{cases} , \qquad (3.15)
$$

where $\gamma \in \mathbb{R}$ *is referred to as the extreme value index. The notation* $(\cdot)_+ = \max(\cdot, 0)$ *is used.*

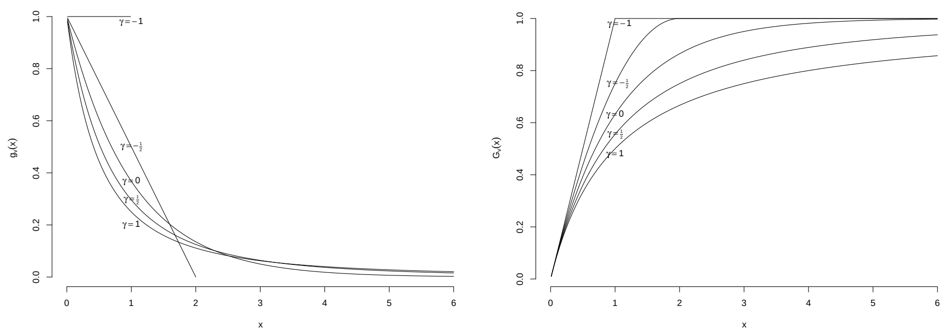

Figure 3.3: Probability density function (left) and cumulative distribution function (right) of the GP distribution for $\gamma = -1, \ -\frac{1}{2}$ (with upper end points $x = 1, \ 2$), $\gamma = 0, \ \frac{1}{2}, \ 1$.

It should be noticed that the parameter γ of the GP distribution is uniquely determined by the associated GEV distribution. To be more specific, the parameter γ stated in Definition 3.5 is equal to that of the associated GEV distribution (Coles, 2011, [Col11]).

Remark 3.3 (Peaks Over Threshold approach)
The class of generalized pareto distributions, denoted by Equation (3.15), are the possible limiting distributions of the exceedances over a large enough threshold. This approach, in order to obtain the limit distribution of the tail observations of a sample, is known as the POT approach. In this thesis, the POT approach is preferred over the BM approach.

Remember that in EVT one is interested in the tail behavior and not the whole distribution. In case we have a sequence $X_1, X_2, \ldots, X_n$ of independent and identically distributed random variables, only the largest $k < n$ observations are included in the estimation process. Let $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ denote the order statistics. Choose the $(n-k)$-th order statistic $X_{n-k,n}$ as threshold, i.e. $t = X_{n-k,n}$. Recall Theorem 3.1 with relation $u(t) = a\left(\frac{1}{1-F(t)}\right)$. Hence,

$$\begin{aligned} u(X_{n-k,n}) &= a\left(\frac{1}{1-F(X_{n-k,n})}\right) \\ &\approx a\left(\frac{1}{1-F_n(X_{n-k,n})}\right) \\ &= a\left(\frac{1}{1-\frac{n-k}{n}}\right) \\ &= a\left(\frac{n}{k}\right) . \end{aligned} \tag{3.16}$$

Remark 3.4 (Moment estimators)
Many estimators exist for $a\left(\frac{n}{k}\right)$, $b\left(\frac{n}{k}\right) = U\left(\frac{n}{k}\right)$ and γ, and each estimator has its own advantages and disadvantages. The most common estimators are the Hill estimator for $\gamma > 0$, the ML estimator for $\gamma > -1$, the probability weighted moment estimator for $\gamma < 1$ and the moment estimator for $\gamma \in \mathbb{R}$. For an overview of these estimators, the reader is referred to Chapter 3 and 4 in (De Haan et al., 2006, [HF06]). In this thesis, we will use the moment estimators given by

$$M_n^{(j)} = \frac{1}{k}\sum_{i=1}^{k} (\log X_{n,n-i+1} - \log X_{n,n-k})^j \tag{3.17}$$

$$\widehat{\gamma} = M_n^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{\left(M_n^{(1)}\right)^2}{M_n^{(2)}}\right)^{-1} \tag{3.18}$$

$$\widehat{a}\left(\frac{n}{k}\right) = X_{n-k,n} M_n^{(1)} \left(1 - \widehat{\gamma} + M_n^{(1)}\right) \tag{3.19}$$

$$\widehat{b}\left(\frac{n}{k}\right) = X_{n-k,n} \ , \tag{3.20}$$

where $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ *denotes the ordered sample. Moreover, we have* $j = 1,\ 2$ *and provide that* $k = k(n) \to \infty,\ \frac{k}{n} \to 0$ *as* $n \to \infty$.

The question that remains is how to choose an appropriate value for k, as a trade off exists between bias and variance. Choosing large values for k (i.e. high threshold) will result in a high variance, as we will end up with few exceedances to estimate the parameters of the GP distribution. Choosing small values for k (i.e. low threshold) will result in bias, as the asymptotic basis will be violated (Coles, 2011, [Col11]). Two methods exist. The first method is to generate the mean residual life plot and look where this plot is linear in the threshold $t = X_{n-k,n}$. The second method is to take a range of values for k, estimate the parameters of the GP distribution and look for stability of parameter estimates. For more information about these methods one is referred to respectively section 4.3.1 and 4.3.4 in (Coles, 2011, [Col11]).

Example 3.3 (Application to ABN AMRO Bank transaction data set)
Author's note: this example is confidential.

Remark 3.5
Author's note: this remark is confidential.

3.3 Multivariate extreme value theory

In this section, the focus will be on multivariate extremes. We restrict our attention to the bivariate case. In this manner the main concepts can be highlighted without becoming unnecessary complex due to notation issues that arise while considering higher dimensions.

In contrast to the univariate case, the identification of extreme values is not straightforward and therefore many possibilities exist. It turns out that a very naive approach considering pointwise maxima leads to a rich enough theory (Coles, 2011, [Col11]), and therefore we will follow this approach. By following this approach we will rely on the results presented in the previous sections of this chapter. Our main interest will be the limiting behavior of these pointwise maxima.

Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be a sequence of independent and identically distributed random vectors with common cdf F. Let F_1 and F_2 be the marginal distribution functions of X and Y, respectively. Define the pointwise maxima $M_{x,n} = \max(X_1, \ldots, X_n)$ and $M_{y,n} = \max(Y_1, \ldots, Y_n)$. Then $\mathbf{M_n} = (M_{x,n}, M_{y,n})$ is called the vector of pointwise maxima. It should be noticed that the index i for which the variable X_i attains its maximum value need not be necessarily the same as that of variable Y_i. Therefore, the vector $\mathbf{M_n}$ is not necessarily observed in the original sequence.

Similarly to the univariate framework, in order to avoid the limiting distribution of $\mathbf{M_n}$ to be degenerate, we look for sequences of real numbers $a_n,\ c_n > 0$ and $b_n,\ d_n \in \mathbb{R}$ $(n = 1, 2, \ldots)$ and a bivariate distribution function G with non degenerate marginals such that

$$\lim_{n\to\infty} P\left(\frac{M_{x,n} - b_n}{a_n} \leq x, \frac{M_{y,n} - d_n}{c_n} \leq y\right) = \lim_{n\to\infty} F^n(a_n x + b_n, c_n y + d_n) = G(x, y) \tag{3.21}$$

for all continuity points (x, y) of G. Any distribution function G that can occur as a limit is called a multivariate extreme value (MEV) distribution. In contrast to the univariate framework, the class of MEV distributions cannot be represented by a finite dimensional parametric family. However, a specific characterization of the MEV distribution can be found. Note that we have the following restrictions on the distribution function G due to the convergence of the marginal distributions of the pointwise maxima $M_{x,n}$ and $M_{y,n}$. Here, the function G_{γ_1} and G_{γ_2} refers to the univariate GEV distribution as denoted in Equation (3.3).

$$\lim_{n\to\infty} P\left(\frac{M_{x,n} - b_n}{a_n} \leq x\right) = G(x, \infty) = G_{\gamma_1}(x) \tag{3.22}$$

$$\lim_{n\to\infty} P\left(\frac{M_{y,n} - d_n}{c_n} \leq y\right) = G(\infty, y) = G_{\gamma_2}(y) \tag{3.23}$$

It would be desirable to separate the modelling of the MEV distribution into inference on the marginal distributions and inference on the dependence structure. A way to accomplish this is to assume specific margins for the distribution function G. To be more specific, one aims to find an appropriate transformation for the sequence $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ such that the transformed tails (i.e. transformed pointwise maxima) follow an univariate extreme value distribution (De Haan et al., 1998, [HR98]). A common and useful choice is to set the transformed tails to follow the standard Fréchet distribution. This is where Theorem 3.3, which is stated below, can be of use.

Theorem 3.2

Suppose Equations (3.21), (3.22) and (3.23) hold. Then

$$\lim_{n\to\infty} F^n(U_1(nx), U_2(ny)) = G_0(x,y) \ , \tag{3.24}$$

where

$$G_0(x,y) = G\left(\frac{x^{\gamma_1}-1}{\gamma_1}, \frac{y^{\gamma_2}-1}{\gamma_2}\right) \ . \tag{3.25}$$

Proof 3.3

We refer to Appendix E.2 for the proof.

□

Corollary 3.2

Alternatively, we could write: for each continuity point (x,y) *for which* $0 < G_0(x,y) < 1$ *we have*

$$\lim_{n\to\infty} n\left(1 - F(U_1(nx), U_2(ny)\right) = -\log G_0(x,y) \ . \tag{3.26}$$

Proof 3.4

We refer to Appendix E.2 for the proof, which is almost identical to the univariate case.

□

One important observation is that Equation (3.24) can be reformulated into the form of Equation (3.21). The result is shown by Equation (3.27), and the corresponding derivation can be found in Appendix E.3. Consequently, Equations (3.28) and (3.29) follow. Here, $\tilde{M}_{x,n}$ and $\tilde{M}_{y,n}$ refer respectively to $\max\left(\tilde{X}_1, \ldots, \tilde{X}_n\right) = \max\left(\frac{1}{1-F_1(X_1)}, \ldots, \frac{1}{1-F_1(X_n)}\right)$ and $\max\left(\tilde{Y}_1, \ldots, \tilde{Y}_n\right) = \max\left(\frac{1}{1-F_2(Y_1)}, \ldots, \frac{1}{1-F_2(Y_n)}\right)$. The reformulation implies that the marginal distribution of these maxima after normalizing with constants $a_n = c_n = n$ and $b_n = d_n = 0$ is fixed as $n \to \infty$. The normalized maxima follow an univariate extreme value distribution in the limit, i.e. the standard Fréchet distribution.

$$\lim_{n\to\infty} F^n(U_1(nx), U_2(ny)) = \lim_{n\to\infty} P\left(\frac{\tilde{M}_{x,n}}{n} \leq x, \frac{\tilde{M}_{y,n}}{n} \leq y\right) = G_0(x,y) \tag{3.27}$$

$$\lim_{n\to\infty} P\left(\frac{\tilde{M}_{x,n}}{n} \leq x\right) = G_0(x,\infty) = \exp\left(-\frac{1}{x}\right) \tag{3.28}$$

$$\lim_{n\to\infty} P\left(\frac{\tilde{M}_{y,n}}{n} \leq y\right) = G_0(\infty, y) = \exp\left(-\frac{1}{y}\right) \tag{3.29}$$

Now we are able to formulate the class of limiting distributions in Equation (3.21), which is stated below in Theorem 3.4. This theorem implies that the bivariate extreme value distribution G from Equation (3.21) is completely characterized by the extreme value indices γ_1 and γ_2 of the marginal distributions and the exponent measure that captures the dependence structure. Remark that also other characterizations exist, for instance by including the spectral measure. For more information about these type of characterizations the reader is referred to Theorem 6.1.14 in (de Haan et al., 2006, [HF06]).

Theorem 3.3

For any extreme value distribution distribution G from Equation (3.21) with (3.22) and (3.23) there exist a distribution function H on $[0,1]$ with mean $\frac{1}{2}$, i.e.

$$\int_0^1 w\mathrm{d}H(w) = \frac{1}{2} \ , \tag{3.30}$$

such that for $x, y > 0$

$$G_0(x,y) = G\left(\frac{x^{\gamma_1}-1}{\gamma_1}, \frac{y^{\gamma_2}-1}{\gamma_2}\right) = \exp\left(-V(x,y)\right) \ , \tag{3.31}$$

where the function $V(x,y)$ is defined by

$$V(x,y) = 2\int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) \mathrm{d}H(w) \ . \tag{3.32}$$

Conversely, any finite measure represented by distribution function H gives rise to a limit distribution G in Equation (3.21) via Equation (3.31).

Remark 3.6 (Exponent measure)

The function $V(x,y)$, defined by Equation (3.32), determines the exponent measure ν in the following way:

$$\nu(A_{x,y}) := V(x,y) \ , \tag{3.33}$$

with

$$A_{x,y} := \left\{(s,t) \in \mathbb{R}_+^2 \ : \ s > x \cup t > y\right\} \ . \tag{3.34}$$

Moreover, we have

$$\nu(A_{sx,sy}) = \nu(sA_{x,y}) = \frac{1}{s}\nu(A_{x,y}) \ . \tag{3.35}$$

Remark 3.7 (Preview)

The exponent measure will play a key role in Chapter 6.

In the univariate framework, we introduced the concept of max-domain of attraction. This concept enabled us to state conditions on the cdf F to ensure that there exist two sequences $a_n > 0$ and $b_n \in \mathbb{R}$ $(n = 1, 2, \ldots)$ such that the transformed partial maxima converge to a extreme value distribution G_γ. In the bivariate framework we can introduce a similar concept, which puts conditions on both the marginal distributions and the extremal dependence.

Theorem 3.4 (Max-domain of attraction)

The cdf F, with continuous marginal cdfs F_1 and F_2, is said to be in the max-domain of attraction of a bivariate extreme value distribution G with marginal distributions G_{γ_1} and G_{γ_2} if and only if

1. *$F_i \in MDA(G_{\gamma_i})$ for $i = 1,\ 2$.*

2. *There exsists a positive function $s(x, y)$ such that for $x,\ y > 0$*

$$\lim_{n\to\infty} \frac{1 - F(U_1(nx), U_2(ny))}{1 - F(U_1(n), U_2(n))} = s(x, y) \ . \tag{3.36}$$

We write $F \in MDA(G)$.

3.4 Extremal dependence

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sequence of independent and identically distributed random vectors. A common interest in multivariate EVT is estimating the probability of an extreme event, i.e. a probability of type

$$P\left(X_i > u, Y_i > v\right) \tag{3.37}$$

for some values u and v. In the environmental context one could think of the variables X_i and Y_i representing wave heights and still water levels of which the exceedance of both at some prespecified levels u and v would result in a dangerous situation that should be avoided, see for instance the application in (de Haan et al., 1998, [HR98]). In the financial context one could think of the variables X_i and Y_i representing the losses of two stock indices, see for instance the application in (Poon et al., 2004, [PRT04]).

In multivariate extreme value theory, one important issue is to indicate whether the pointwise maxima are dependent or not in the limit (Hüsler, 2009, [Hus09]) (De Haan et al., 1998, [HR98]). In most literature, the distinction between asymptotic dependence and asymptotic independence is not made. Often asymptotic dependence of the variables is assumed in case independence is rejected. However, in case the pointwise maxima are asymptotically independent, the procedures for estimating a probability of type (3.37) are not applicable anymore (Draisma et al., 2004, [DDFH04]). In this thesis, we want to avoid this situation and make a clear distinctions between these two types of dependence structures in the limit.

We start this section by explaining the concept of asymptotic dependence and defining a measure for this type of dependence. Hereafter, another type of dependence, known as residual dependence is discussed.

3.4.1 Asymptotic dependence

In case two random variables are asymptotic independent, the limit distribution G of the pointwise maxima, corresponds to the joint distribution function of two independent random variables. However, this does not necessarily imply that the two variables are independent. The formal definition of asymptoc independence in a bivariate framework, is denoted by Definition 3.5.

Definition 3.5 (Asymptotic dependence)
Let the cdf F be in the max-domain of attraction of a bivariate extreme value distribution G with marginal distributions G_{γ_1} and G_{γ_2}. Then

1. *If $G(x, y) = G_{\gamma_1}(x)G_{\gamma_2}(y)$, then X and Y are said to be asymptotic independent.*
2. *Otherwise, then X and Y are said to be asymptotic dependent.*

Note that Equation (3.26) can be translated into Equations (3.38) and (3.39). In case the variables X and Y are said to be asymptotically independent, using Definition 3.6, we can see that the right-hand side of Equation (3.39) is identically zero (i.e. $\log G_0(x, y) = \log (G_0(x)G_0(y)) = \log G_0(x) + \log G_0(y)$). Consequently, Equation (3.39) cannot help us by estimating a probability of type (3.37).

$$\lim_{n\to\infty} nP\left(\frac{1}{1-F_1(X)} > nx \cup \frac{1}{1-F_2(Y)} > ny\right) = -\log G_0(x,y) \tag{3.38}$$

$$\lim_{n\to\infty} nP\left(\frac{1}{1-F_1(X)} > nx, \frac{1}{1-F_2(Y)} > ny\right) = \log G_0(x,y) - \log G_0(x) - \log G_0(y) \tag{3.39}$$

This implicates that it is important to preliminary check whether the variables X and Y are asymptotic independent. We start by defining a measure of asymptotic dependence in a bivariate framework, which is given by Equation (3.40). In case $\chi = 0$ the components X and Y are said to be asymptotically independent. In case $\chi > 0$ the component X and Y are said to be asymptotically dependent, and the measure χ gives the strength of asymptotic dependence.

$$\chi = \lim_{t\to\infty} P\left(X > U_1(t) \;\middle|\; Y > U_2(t)\right) \tag{3.40}$$

Proposition 3.1 (Asymptotic independence)
If X and Y are asymptotic independent, then

$$\chi = 0 \ . \tag{3.41}$$

Proof 3.5 (Asymptotic independence)

We refer to Appendix E.4 for the proof.

□

Based on the bivariate sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ an estimator for the measure χ can be derived. Let $X_{1,n} \le X_{2,n} \le \ldots \le X_{n,n}$ and $Y_{1,n} \le Y_{2,n} \le \ldots \le Y_{n,n}$ be the ordered samples. Then the estimator $\hat{\chi}$ is defined by

$$\hat{\chi} = \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}_{\{X_i > X_{n-k,n},\ Y_i > Y_{n-k.n}\}} \ . \tag{3.42}$$

It should be noticed that the measure χ is only defined for the bivariate framework so far. However, this measure is extended by de Haan et al. (2006) for a d-dimensional framework. The corresponding theorem is stated below. The theorem shows that pairwise asymptotic dependence results in joint asymptotic independence.

Theorem 3.5
Let $F : \mathbb{R}^d \to \mathbb{R}_+$ be a cdf. Suppose that its marginal distribution functions $F_i : \mathbb{R} \to \mathbb{R}_+$ satisfy

$$\lim_{n\to\infty} F_i^n \left(a_n^{(i)} x + b_n^{(i)}\right) = \exp\left(-\left(1+\gamma_i x\right)^{-\frac{1}{\gamma_i}}\right) \tag{3.43}$$

for all x for which $1+\gamma_i x > 0$ and where $a_n^{(i)} > 0$ and $b_n^{(i)} \in \mathbb{R}$ are sequences of real numbers, $i = 1, 2, \ldots, d$. Let $(X_1, X_2, \ldots, X_d)$ be a random vector with cdf F. If

$$\lim_{t\to\infty} P\left(X_i > U_i(t) \;\middle|\; X_j > U_j(t)\right) = 0 \tag{3.44}$$

for all $i \leq i < j \leq d$, then

$$\lim_{n\to\infty} F^n \left(a_n^{(1)} x_1 + b_n^{(1)}, \ldots, a_n^{(d)} x_d + b_n^{(d)}\right) = \exp\left(-\sum_{i=1}^{d} \left(1+\gamma_i x_i\right)^{-\frac{1}{\gamma_i}}\right) \tag{3.45}$$

for $1+\gamma_i x_i > 0$, $i = 1, 2, \ldots, d$. Hence the components of $(X_1, X_2, \ldots, X_d)$ are asymptotically independent.

Proof 3.6

We refer to (de Haan et al., 2006, [HF06]) for the proof.

□

Example 3.4 (Application to ABN AMRO Bank transaction data set)
Author's note: this example is confidential.

3.4.2 Residual dependence

In order to overcome the problem of estimating a probability of type (3.37) in case of asymptotic independence, Ledford et al. (1998) introduced an additional parameter regarding the extremal dependence in a bivariate setting. This additional parameter is known as the residual dependence index, or the coefficient of tail dependence.

Consider a cdf F that is said to be in the max-domain of attraction of a bivariate extreme value distribution G. This implies that the marginal cdfs F_1 and F_2 are said to be in the max-domain of attraction of some univariate extreme value distributions G_{γ_1} and G_{γ_2}. The univariate extreme value distributions G_{γ_1} and G_{γ_2} are given by Equation (3.2). Secondly, Ledford et al. assumed that in case of asymptotic independence an additional function S exists such that

$$\lim_{t\to\infty} \frac{P\left(1-F_1(X)<\frac{x}{t}, 1-F_2(Y)<\frac{y}{t}\right)}{P\left(1-F_1(X)<\frac{1}{t}, 1-F_2(Y)<\frac{1}{t}\right)} := S(x,y) \tag{3.46}$$

for $x, y > 0$. The function S is homogeneous of order $\frac{1}{\eta}$ for some $\eta \in (0,1]$, i.e. $S(ax, ay) = a^{\frac{1}{\eta}} S(x,y)$. The index η is called the residual dependence index of the variables (X,Y). The value of the residual dependence index η characterizes the extremal dependence of the variables (X,Y). In case $\eta = 1$, we have asymptotic dependence which indicates that the extremes of X and Y tend to occur simultaneously. In case $\eta < 1$, we have asymptotic independence.

Remark 3.8
In case of $\eta < 1$, we can distinguish three different categories.

- $\boldsymbol{\eta \in \left(\frac{1}{2}, 1\right)}$
  The case $\eta \in \left(\frac{1}{2}, 1\right)$ represents a positive association between the variables (X,Y). The extremes of X and Y tend to occur simultaneously more often than those when X and Y are independent. Mathematically, this can be translated into $P(X > x, Y > y) \gg P(X > x)P(Y > y)$.

- $\boldsymbol{\eta = \frac{1}{2}}$
  The case $\eta \in \left(\frac{1}{2}, 1\right)$ represents near extremal independence. The extremes of X and Y behave as if X and Y are independent.

- $\boldsymbol{\eta \in \left(0, \frac{1}{2}\right)}$
  The case $\eta \in \left(0, \frac{1}{2}\right)$ represents a negative association between the variables (X,Y). The extremes of X and Y tend to occur simultaneously less often than those when X and Y are independent. Mathematically, this can be translated into $P(X > x, Y > y) \ll P(X > x)P(Y > y)$.

Using Equation (3.46) and the fact that the function S is homogeneous of order $\frac{1}{\eta}$, we can derive

$$S\left(\frac{1}{x}, \frac{1}{x}\right) = x^{-\frac{1}{\eta}} S(1,1) = x^{-\frac{1}{\eta}} \ . \tag{3.47}$$

Additionally, introduce $T = \min\left(\frac{1}{1-F_1(X)}, \frac{1}{1-F_2(Y)}\right)$. Now, rewriting Equation (3.46) using the newly introduced variable T results in

$$S\left(\frac{1}{x}, \frac{1}{x}\right) = \lim_{t\to\infty} \frac{P\left(\min\left(\frac{1}{1-F_1(X)}, \frac{1}{1-F_2(Y)}\right) > tx\right)}{P\left(\min\left(\frac{1}{1-F_1(X)}, \frac{1}{1-F_2(Y)}\right) > t\right)} = \lim_{t\to\infty} \frac{P(T > tx)}{P(T > t)} \ . \tag{3.48}$$

Hence, combining Equation (3.47) and (3.48) gives

$$\lim_{t\to\infty} \frac{P(T > tx)}{P(T > t)} = x^{-\frac{1}{\eta}} \ . \tag{3.49}$$

Referring to Theorem 1.2.1 in (de Haan et al., 2006, [HF06]), this implies that the variable T is in the max-domain of attraction of the extreme value distribution G_η with extreme value index $\eta > 0$. Going back to the univariate framework, this implies that the variable T has a distribution with a heavy right tail with an infinite right endpoint. Therefore, we can use the same estimators we used for the univariate framework.

Based on the bivariate sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ an estimator for the residual dependence index η can now be derived following the procedure in (Draisma et al., 2004, [DDFH04]). Let $T_{1,n} \leq T_{2,n} \leq \ldots \leq T_{n,n}$ be the ordered sample of $T_i = \min\left(\frac{1}{1-F_1(X_i)}, \frac{1}{1-F_2(Y_i)}\right)$ for $i = 1, 2, \ldots, n$. The estimator $\hat{\eta}$ is given by Equation (3.50). In case the marginal distribution F_1 and F_2 are unknown, T_i is replaced by $T_i^{(n)} = \min\left(\frac{n+1}{n+1-R_i^X}, \frac{n+1}{n+1-R_i^Y}\right)$. Then the estimator $\hat{\eta}$ is defined by Equation (3.51).

$$\hat{\eta} = \frac{1}{k} \sum_{i=1}^{k} \log\left(\frac{T_{n,n-i+1}}{T_{n,n-k}}\right) \tag{3.50}$$

$$\hat{\eta} = \frac{1}{k} \sum_{i=1}^{k} \log\left(\frac{T^{(n)}_{n,n-i+1}}{T^{(n)}_{n,n-k}}\right) \tag{3.51}$$

Example 3.5 (Application to ABN AMRO Bank transaction data set)
Author's note: this example is confidential.

4

Modelling dynamic durations

4.1 Introduction

This chapter introduces the class of Autoregressive Conditional Duration (ACD) models. The ACD model will form the main building block of Chapter 5, in which a methodology is proposed for the estimation of the univariate risk metric.

The chapter is organized as follows. First the need of the ACD model is elaborated. Section 4.4 presents the ACD model framework and the different specifications that are available: the standard ACD model and the Log ACD models. Section 4.5 and 4.6 adress model estimation and tests available to check the correctness of the estimated model. In Section 4.7 a model selection methodology is proposed. The aim of this methodology is to quantify the quality of each ACD model and to provide a straightforward guideline for selecting an appropriate model. We end this chapter by applying this methodology to the ABN AMRO Bank transaction data set of Chapter 2, of which the results will be used in Chapter 5.

4.2 Ultra high frequency data

Over the last years, improvements in computer technology, data recording and data storage have made it possible to obtain high frequency time series. However, whether time series are classified as high frequency depends on the domain of study. In finance, time series are classified as high frequency in case the observations are taken at a time scale finer than once per day (Yan et al., 2003, [YZ03]). Driving this frequency to the ultimate limit results in real-time measurements, which are in the financial context also known as transaction-by-transaction data, tick-by-tick data or ultra high frequency data. Hence, the ABN AMRO Bank transaction data set as discussed in Chapter 2, can be classified as ultra high frequency.

The accessibility of this kind of high frequency time series has challenged the development of new methods in statistical modelling, as ultra high frequency financial time series exhibit distinct characteristics compared to the classical financial time series that we are usually dealing with (such as daily stock prices, monthly interest rates or quarterly inflation rates). Most ultra high frequency time series are positive-valued, positively autocorrelated, strongly persistent and exhibit strong intraday periodicities (Hautsch, 2012, [Hau12]). Moreover, as illustrated by Figure 4.1, ultra high frequency time series have the main characteristic of being irregularly spaced over time with a random number of observations per time interval. This implies that the duration between consecutive events is not constant, but can be interpreted as a random variable.

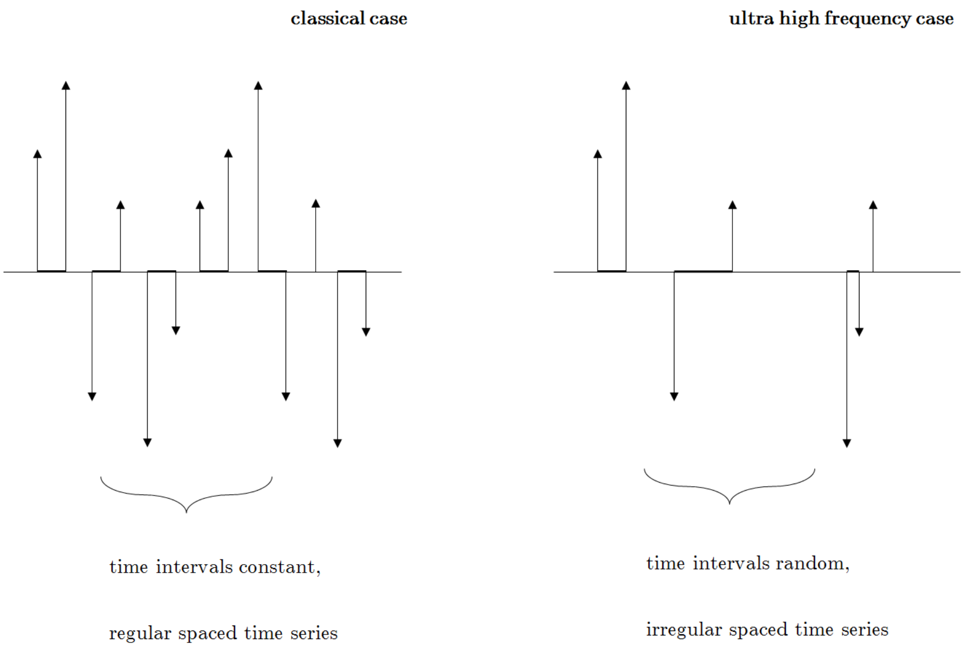

Figure 4.1: Graphic representation of the characteristics of ultra high frequency time series.

As classical time series analysis makes use of discrete time series methods with fixed time intervals, these methods are not applicable anymore in the case of ultra high frequency time series. A solution could be to divide the observations into fixed time intervals prior to analysis. However, it can be argued that the time interval between consecutive events might contain valuable information. Dividing observations into short time intervals for periods of relatively low activity could result in heteroskedasticity, due to the presence of intervals that contain no additional information. In the same manner, dividing observations into long time intervals for periods of high activity could result in the loss of information contained in rapidly arriving data (Engle et al., 1998, [ER98]). Thus, to avoid the aforementioned problems we must be able to model the irregularity.

A solution to account for the irregularity of the timing between events is to consider the event times as random variables that follow a point process, as first introduced by Engle. This introduction was the starting point for the rapid development of research in the area of ultra high frequency financial econometrics (Hautsch, 2012, [Hau12]). The next section introduces the concept of point processes.

4.3 Point processes

One of the characteristics of ultra high frequency time series is that the events do not occur at regularly spaced time intervals. One way of representing ultra high frequency time series is by means of point processes. A point process is a stochastic process that generates a random collection of points on the time axis (Bauwens et al., 2001, [BG01]). Point processes are common in domains of study such as neuroscience and queuing theory, but have gained great interest in finance over the last few years (Pacurar, 2006, [Pac08]).

Now, let t be a physical (calendar) time and let t_i be the time of occurrence of the i-th event such that $0 \leq t_1 \leq t_2 \leq \ \ldots$. Then, the sequence of event times $\Big\{t_i : i = 1, 2, \ \ldots\Big\}$ is called a point process on $[0, \infty)$. Additionally, in case $t_i < t_{i+1}$ for all i, the possibility is excluded that events occur simultaneously. The sequence of event times $\Big\{t_i : i = 1, 2, \ \ldots\Big\}$ is then referred to as a simple point process. In this thesis we will only focus on simple point processes.

Often, additional information is associated with the event times. In case the events are financial transactions, the information could for instance be the amount or type of the transaction. Each type of additional information is known as a mark. Let $\Big\{y_i : i = 1, 2, \ \ldots\Big\}$ be the sequence of marks associated with the events times $\Big\{t_i \ : \ i = 1, 2, \ \ldots,\Big\}$ where $y_i \in \Big\{1, 2, \ldots, K\Big\}$ denotes the type of the i-th event. Then, the sequences of event times and corresponding marks $\Big\{(t_i, y_i) \ : \ i = 1, 2 \ \ldots\Big\}$ is referred to as a (simple) marked point process or a (simple) K-dimensional point process. The next sections will be dedicated to models for simple point processes without marks.

Define the counting function $N(t) = \sum_{i \geq 1} \mathbb{1}_{\{t_i \leq t\}}$ which denotes the number of events that have occurred by time t. This counting function is a step function that is continuous from the left with limits from the right, also known as a cadlag function. Now, the simple point process $\Big\{t_i : i = 1, 2, \ \ldots\Big\}$ and the (simple) marked point process $\Big\{(t_i, y_i) \ : \ i = 1, 2 \ \ldots\Big\}$

can be translated in terms of the counting function, i.e. $\Big\{t_i : i = 1, \ldots, N(T)\Big\}$ and $\Big\{(t_i, y_i) : i = 1, \ldots, N(T)\Big\}$ such that $t_1 < \ldots < t_{N(T)}$.

A point process is considered to evolve without after-effects if for any $t > t_0$ the realization of the point process in the interval $[t, \infty)$ does not depend on the realization of the point process in the previous interval $[t_0, t)$. Alternatively, a point process evolves with after-effects in case there is a non-zero dependence between subsequent intervals. A point process is called conditionally orderly at time $t \geq t_0$ if for a sufficiently short time interval and conditional on any event P defined by the realization of the point process on $[t_0, t)$, the probability of two or more events occurring is infinitesimal relative to the probability of one event. It should be noticed that in the remainder of this thesis only conditionally orderly point processes with after-effects will be considered.

A natural way of formulating a conditionally orderly point process is by means of the conditional density function, conditional survivor function or the conditional intensity function of the event times. The conditional intensity function describes the expected event arrival rate conditional on the filtration of event arrival times. The exact definition of the conditional intensity function is stated below. By using one of these functions the complete conditionally orderly point process is specified (Engle et al., 2004, [ER04]).

Definition 4.1 (Conditional intensity function)
Let $N(t) = \sum_{i \geq 1} \mathbb{1}_{\{t_i \leq t\}}$ *be the counting function for the sequence of event times* $\Big\{t_i : i = 1, 2, \ldots,\Big\}$*. Then the conditional intensity function* λ *is given by*

$$\lambda\Big(t \;\Big|\; N(t), t_1, \ldots, t_{N(t)}\Big) = \lim_{\Delta t \to 0} \frac{P\Big(N(t + \Delta t) > N(t) \;\Big|\; N(t), t_1, \ldots, t_{N(t)}\Big)}{\Delta t} \,. \tag{4.1}$$

Many different representations for the conditional intensity function exist in the literature, and we can distinguish four different classes. The first three classes are denoted by

$$\lambda\Big(t \;\Big|\; N(t), t_1, \ldots, t_{N(t)}\Big) = \omega + \sum_{i=1}^{N(t)} \pi\left(t - t_i\right) \,, \tag{4.2}$$

$$\lambda\Big(t \;\Big|\; N(t), t_1, \ldots, t_{N(t)}\Big) = \omega + \sum_{i=1}^{N(t)} \pi_i\left(t_{N(t)+1-i} - t_{N(t)-i}\right) \,, \tag{4.3}$$

$$\lambda\Big(t \;\Big|\; z_1, \ldots, z_{N(t)}\Big) = \lambda_0(t) \exp\left(\beta' z_{N(t)}\right) \,. \tag{4.4}$$

The fourth class is known as the accelerated failure time model and is denoted by Equation (4.5). Let z_i be a vector of explanatory variables associated with event time t_i. Assume H_i is a random time-to-event variable which can be expressed in terms of the vector z_i via the relation $\log H_i = z_i\beta + W_i$. Here it is assumed that the errors W_i are IID random variables which are independent of β. Then by taking the exponent on both sides of the equation we obtain $H_i = \exp(z_i\beta)\exp(W_i)$ where $\exp(W_i) > 0$. Let $\lambda_0(t)$ be the hazard function of $\exp(W_i)$. Then the conditional intensity function can be expressed by

$$\lambda\left(t \;\middle|\; z_1,\ \ldots,\ z_{N(t)}\right) = \lambda_0\left(t\exp\left(-\beta' z_{N(t)}\right)\right)\exp\left(-\beta' z_{N(t)}\right) \ . \tag{4.5}$$

It will turn out that the Autoregressive Conditional Duration (ACD) model, which is introduced in Section 4.3, belongs to this specific class.

Remark 4.1
The class denoted by Equation (4.4) is known as the proportional hazard model. While the explanatory variables act multiplicatively on the time in the accelerated failure time model, the explanatory variables act multiplicatively on the hazard function in the proportional hazard model.

4.4 ACD models

Engle et al. (1998) introduced a model for a sequence of time events that arrive at irregular time intervals, which obtained in particular popularity in modelling the durations of financial transactions such as trade durations and price durations. These type of models are called Autoregressive Conditional Duration (ACD) models. The ACD model is a non-linear model that captures both the clustering effect often seen in high frequency financial durations and the time dependence between these durations. First the general ACD framework will be introduced, where after the standard ACD model, Log ACD model and some other extensions will be discussed.

Let x_i be the i-th duration, which is defined as the time interval between the events occurring at successive event times t_i and t_{i-1}. This is represented by

$$x_i = t_i - t_{i-1} \ . \tag{4.6}$$

In this thesis, as we focus on the transaction process of two specific accounts within ABN AMRO Bank, a duration x_i is the time elapsed in seconds between two consecutive events. It should be noted that simultaneously recorded transactions are considered to be one event. For more information about the transaction process, the reader is referred to Chapter 2.

Let $\mathcal{F}_{i-1}$ denote the information set associated with the sequence of durations up to time t_{i-1}. Let ψ_i be the expectation of the duration x_i conditional on this information set, which is given by

$$E\left(x_i \;\middle|\; \mathcal{F}_{i-1}\right) = \psi_i\left(\mathcal{F}_{i-1}, \boldsymbol{\theta}\right) := \psi_i \ . \tag{4.7}$$

Often, ψ_i is also referred to as the conditional mean function. Furthermore, the ACD model is based on the assumption that the durations are decomposed into the product of their conditional expectation and an error term. Let ϵ_i be the error. This assumption can be translated into

$$x_i = \psi_i \epsilon_i \ . \tag{4.8}$$

The errors, also referred to as standardized durations in the literature, are assumed to have values that are non negative and to be independent and identically distributed (IID) with a probability density f_ϵ. Because ψ_i is defined as the conditional expectation of x_i, ϵ_i should have an expectation of 1. Therefore, the probability density f_ϵ should be chosen such that $E(\epsilon_i) = 1$.

In order to derive a general expression for the conditional intensity function, let S_ϵ be the survivor function of the errors ϵ_i. Then $\lambda_0(t) = \dfrac{f_\epsilon(t)}{S_\epsilon(t)}$ is the hazard function of the errors ϵ_i. Now the ACD model can be represented in terms of the conditional intensity function, which is denoted by

$$\lambda\left(t \;\middle|\; N(t), t_1,\ \ldots,\ t_{N(t)}\right) = \lambda_0\left(\frac{t - t_{N(t)-1}}{\psi_{N(t)}}\right)\frac{1}{\psi_{N(t)}} \ . \tag{4.9}$$

This representation indeed demonstrates that the ACD model belongs to the class of accelerated failure time models.

The ACD model possesses flexibility because various specifications for both the conditional mean function ψ_i and the probability density f_ϵ of the errors ϵ_i can be considered. The choice of the probability density f_ϵ affects the conditional intensity function. For instance, if we assume that the errors are distributed according to the exponential distribution or the Weibull distribution with unit mean, this results in the following conditional intensity functions. The conditional intensity function for exponential distributed errors turns out to be constant, while the conditional intensity function for Weibull distributed errors could be either increasing (i.e. $\gamma > 1$) or decreasing (i.e. $\gamma < 1$).

$$\epsilon_i \overset{IID}{\sim} Exp(\lambda) \implies \lambda\left(t \,\middle|\, N(t), t_1, \ldots, t_{N(t)}\right) = \frac{1}{\psi_{N(t)}} \ . \tag{4.10}$$

$$\epsilon_i \overset{IID}{\sim} Weibull(\lambda, \gamma) \implies \lambda\left(t \,\middle|\, N(t), t_1, \ldots, t_{N(t)}\right) = \gamma\left(t - t_{N(t)-1}\right)^{\gamma-1}\left(\frac{\Gamma\left(1+\frac{1}{\gamma}\right)}{\psi_{N(t)}}\right)^{\gamma} \ . \tag{4.11}$$

4.4.1 Diurnal adjustment

Financial duration processes are subject to strong intraday periodicities, or time-of-day seasonality (Hautsch, 2012, [Hau12]). For instance, it is known that the frequency of transactions is higher near the open and the close of the market, and the frequency f transactions is lower during lunch breaks. The consequence is that these events result in durations that exhibit a deterministic intraday seasonality. Therefore, Engle et al. (1998) assumed that the duration process can be decomposed into a stochastic component and a deterministic component, whereby the deterministic component can be formulated as a multiplicative function. Such a decomposition is given by

$$x_i = \phi(t_i)\tilde{x}_i \ , \tag{4.12}$$

where $\tilde{x}_i$ is the stochastic component and is assumed to follow an ACD model, while $\phi(t_i)$ is the deterministic component which is only dependent on the time of the day. This deterministic component is also known as the diurnal factor. Now, using the decomposition in Equation (4.12), one can translate the conditional mean function ψ_i in Equation (4.7) into the following equation

$$E\left(x_i \,\middle|\, \mathcal{F}_{i-1}\right) = \phi(t_i)E\left(\tilde{x}_i \,\middle|\, \mathcal{F}_{i-1}\right) = \phi(t_i)\tilde{\psi}_i\left(\mathcal{F}_{i-1}, \boldsymbol{\theta}\right) := \phi_i\tilde{\psi}_i \ . \tag{4.13}$$

The deterministic component $\phi(t_i)$ can be specified in different manners. A common approach, used extensively in the literature, is to specify the deterministic component by using spline functions. The main idea is to average the 'raw' durations x_i over q-minute intervals, whereafter spline functions are used to smooth the deterministic component and moreover to extrapolate for any time along the day (Engle, 2000, [Eng00]) (Engle et al., 2004, [ER04]) (Bauwens et al., 2004, [BGGV04]) (Dionne et al., 2009, [DDP09]). For instance, Engle (2000) averages the 'raw' durations over 60-minute intervals in combination with a cubic spline, Engle et al. (2004) average over 60-minute intervals in combination with a linear spline, and Bauwens et al. (2004) and Dionne et al. (2009) average over 30-minute intervals in combination with a cubic spline.

Let $\tau_0, \ldots, \tau_Q$ be the time points splitting the business day into $(Q+1)$ q-minute time intervals. Then the deterministic component can be denoted by Equation (4.14), where M represents the order of the spline function. The most common choices in literature are the linear ($M = 1$) and cubic ($M = 3$) spline function, which are denoted by respectively Equation (4.15) and (4.16). It should be noticed that the constant c_j is identified by setting the mean of the predicted deterministic factor equal to the observed sample mean.

$$\phi_i = \sum_{j=1}^{Q} \mathbb{1}_{\tau_j \le t_i < \tau_{j+1}} \left(c_j + \sum_{m=1}^{M} d_{m,j} (t_i - \tau_j)^m \right) \tag{4.14}$$

$$\implies \phi_i = \sum_{j=1}^{Q} \mathbb{1}_{\tau_j \le t_i < \tau_{j+1}} \left(c_j + d_{1,j}(t_i - \tau_j) \right) \quad \text{for } M = 1 \tag{4.15}$$

$$\implies \phi_i = \sum_{j=1}^{Q} \mathbb{1}_{\tau_j \le t_i < \tau_{j+1}} \left(c_j + d_{1,j}(t_i - \tau_j) + d_{2,j}(t_i - \tau_j)^2 + d_{3,j}(t_i - \tau_j)^3 \right) \quad \text{for } M = 3 \tag{4.16}$$

It should be noticed that the parameters of both the deterministic and stochastic component can be estimated jointly. Also a two-step procedure can be applied by standardizing the 'raw' durations by the deterministic component prior to further model estimation. The ACD model will then be estimated for the standardized durations $\tilde{x}_i$ instead of the 'raw' durations x_i. The first stage will entail the estimation of the parameters of the deterministic component, while the second stage will involve the estimation of the parameters of the conditional mean function of the standardized durations. We will adopt this two-step procedure in this thesis.

4.4.2 Standard ACD model

As mentioned previously in this chapter, the ACD model has a lot of flexibility because different specifications are available for both the conditional mean function ψ_i and the probability density function f_ϵ of the errors ϵ_i. By combining these specifications for both the conditional mean and the probability density a wide range of alternative parameterizations for the ACD model can be obtained. The most general specification of the conditional mean function ψ_i is based on a linear parametrization, given by

$$\psi_i = \omega + \alpha x_{i-1} + \beta \psi_{i-1} \ , \tag{4.17}$$

where the sufficient conditions $\omega > 0$, $\alpha \ge 0$ and $\beta \ge 0$ are imposed in order to ensure positivity of ψ_i. This parametrization is also known as the basic ACD$(1,1)$ model. The basic ACD$(1,1)$ model can be extended to higher orders resulting in the basic ACD(p,q) model, of which the formal definition is denoted by Definition 4.2.

Definition 4.2 (ACD(p,q) model)
The time series $\{x_i \ : \ i = 1, \ \ldots, \ n\}$ is ACD(p,q) if it satisfies the following conditions

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} \ , \tag{4.18}$$

$$x_i = \psi_i \epsilon_i \ , \tag{4.19}$$

where it is assumed that the errors ϵ_i are IID non-negative random variables such that $E(\epsilon_i) = 1$. Let $\boldsymbol{\theta} = (\omega, \ \alpha_1, \ \ldots, \ \alpha_p, \beta_1, \ \ldots, \ \beta_q)' \in \mathbb{R}^k$. The conditions $\omega > 0$, $\alpha_j \geq 0$ for $j = 1, \ldots, p$ and $\beta_j \geq 0$ for $j = 1, \ldots, q$ are needed to ensure positive conditional durations for all possible realizations. The condition $\sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j < 1$ is required to ensure covariance stationarity.

Remark 4.2 (Diurnal adjustment)
In practice it is known that financial durations exhibit intraday seasonality (see Section 4.3.1). Therefore, in order to avoid distortions, this seasonality must be removed from the durations prior to model estimation. This is done by standardizing the raw durations by a deterministic component. Let ϕ_i be the deterministic component. Then we can rewrite the ACD(p,q) model as

$$\tilde{\psi}_i = \omega + \sum_{j=1}^{p} \alpha_j \tilde{x}_{i-j} + \sum_{j=1}^{q} \beta_j \tilde{\psi}_{i-j} \ , \tag{4.20}$$

$$x_i = \phi_i \tilde{\psi}_i \epsilon_i \ , \tag{4.21}$$

where $\tilde{x}_i = \dfrac{x_i}{\phi_i}$ is called the standardized duration.

Now, by assuming a specific probability density f_ϵ for the error term, a complete specification of the ACD model is obtained. A common choice for the probability density f_ϵ is the exponential distribution, as first introduced in (Engle et al., 1998, [ER98]). The resulting specification yields the exponential ACD (EACD) model. Because the exponential distribution is quite restrictive due to a constant intensity function, also other densities are considered in the literature. Possible candidates are the Weibull distribution and the generalized gamma distribution with their parameters chosen such that their means equal unity. These models are known as the Weibull ACD (WACD) and generalized gamma ACD (GACD), respectively.

It should be noticed that the ACD(p, q) model is closely related to the GARCH(p, q) model, as the conditional expected durations in the ACD(p, q) model are modelled in a similair way as the conditional variances in the GARCH(p, q) model. Rewriting Equations (4.18) and (4.19) by taking $\epsilon_i = z_i^2,\ x_i = y_i^2$, and $\psi_i = \sigma_i^2$ yields a general GARCH(p, q) model formulation. As a result, the ACD(p, q) model allows for empirical features such as clustering of events, i.e. small (large) durations are most likely followed by other small (large) durations in a similar manner as the GARCH(p, q) model allows for volatility clustering. Therefore, the ACD(p, q) model is considered as the counterpart of the GARCH(p, q) model for duration processes (Bauwens et al., 2009, [BH09]) (Pacurar, 2008, [Pac08]). Many results and specifications from the GARCH literature can be carried over to the ACD literature.

Another interesting property of the ACD(p, q) model is that it can be formulated in terms of an ARMA($\max(p, q), q$) model for the durations x_i. Let $\eta_i = x_i - \psi_i$, which is a martingale difference by definition. Then the ARMA formulation in Equation (4.22) can be derived. Although Equation (4.22) is more interesting from a theoretical perspective, in practice Equations (4.18) and (4.19) are easier to work with.

Remark 4.3 (ACD(p, q) model as ARMA($\max(p, q), q$) model)
The ACD(p, q) model, as stated in Definition 4.2, can be formulated in terms of an ARMA($\max(p, q), q$) model for the durations x_i, i.e.

$$x_i = \omega + \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) x_{i-j} - \sum_{j=1}^{q} \beta_j \eta_{i-j} + \eta_i \ , \tag{4.22}$$

where $\alpha_j = 0 \ \forall \ j > p$, $\beta_j = 0 \ \forall \ j > q$ and η_i are non-Gaussian innovations.

Derivation 4.1 (ACD(p, q) model as ARMA($\max(p, q), q$) model)

We refer to Appendix E.5 for the derivation.

□

Remark 4.4 (Expectation of ACD(p, q) model)
The conditional expectation of x_i is equal to ψ by definition 4.20. The unconditional expectation of x_i is given by

$$E(x_i) = \frac{\omega}{1 - \sum_{j=1}^{p} \alpha_j - \sum_{j=1}^{q} \beta_j} \ . \tag{4.23}$$

Derivation 4.2 (Expectation of ACD(p, q) model)

We refer to Appendix E.6 for the derivation.

□

4.4.3 Log ACD model

As described in the previous section, the standard ACD model can be quite restrictive as several conditions on the parameters of the conditional mean function are needed in order to ensure that the sequence of durations $\left\{x_i : i = 1, 2, \ \ldots\right\}$ has non-negative conditional durations. Bauwens et al. (2000) introduced a more flexible model, in which the conditional mean function takes a logarithmic form, namely

$$\ln \psi_i = \omega + \sum_{j=1}^{p} \alpha_j g(\epsilon_{i-j}) + \sum_{j=1}^{q} \beta_j \ln \psi_{i-j} \ , \tag{4.24}$$

where $g(\epsilon_{i-j})$ is given by $\ln \epsilon_{i-j}$ or ϵ_{i-j}. The resulting parameterizations are known as respectively the Log ACD$_1(p, q)$ model and the Log ACD$_2(p, q)$ model. In contrast to the standard ACD model, both Log ACD models do not impose any conditions on the parameters of the conditional mean function in order to ensure the non-negativity of conditional durations. Consequently, this enables us to introduce exogenous explanatory variables into the model. The formal definitions are denoted by Definition 4.3 and 4.4.

Definition 4.3 (Log ACD$_1(p, q)$ model)
The time series $\{x_i \ : \ i = 1, \ \ldots, \ n\}$ is Log ACD(p, q) if it satisfies the following conditions

$$\ln \psi_i = \omega + \sum_{j=1}^{p} \alpha_j \ln \epsilon_{i-j} + \sum_{j=1}^{q} \beta_j \ln \psi_{i-j} \ , \tag{4.25}$$

$$x_i = \psi_i \epsilon_i \ . \tag{4.26}$$

Or equivalently, when letting $\ln \psi_i = \kappa_i$,

$$\kappa_i = \bar{\omega} + \sum_{j=1}^{p} \alpha_j \ln \epsilon_{i-j} + \sum_{j=1}^{q} \beta_j \kappa_{i-j} \ , \tag{4.27}$$

$$x_i = e^{\kappa_i} \epsilon_i \ , \tag{4.28}$$

where it is assumed that the errors ϵ_i *are IID non-negative random variables such that* $E(\epsilon_i) = 1$. *Let* $\boldsymbol{\theta} = (\omega,\ \alpha_1,\ \ldots,\ \alpha_p, \beta_1,\ \ldots,\ \beta_q)' \in \mathbb{R}^k$. *The condition* $\sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) < 1$ *is required to ensure covariance stationarity.*

Definition 4.4 (Log ACD$_2(p,q)$ model)
The time series $\{x_i \ : \ i = 1,\ \ldots,\ n\}$ *is Log ACD*(p,q) *if it satisfies the following conditions*

$$\ln \psi_i = \omega + \sum_{j=1}^{p} \alpha_j \epsilon_{i-j} + \sum_{j=1}^{q} \beta_j \ln \psi_{i-j} \ , \tag{4.29}$$

$$x_i = \psi_i \epsilon_i \ . \tag{4.30}$$

Or equivalently, when letting $\ln \psi_i = \kappa_i$,

$$\kappa_i = \omega + \sum_{j=1}^{p} \alpha_j \epsilon_{i-j} + \sum_{j=1}^{q} \beta_j \kappa_{i-j} \ , \tag{4.31}$$

$$x_i = e^{\kappa_i} \epsilon_i \ , \tag{4.32}$$

where it is assumed that the errors ϵ_i *are IID non-negative random variables such that* $E(\epsilon_i) = 1$. *Let* $\boldsymbol{\theta} = (\omega,\ \alpha_1,\ \ldots,\ \alpha_p, \beta_1,\ \ldots,\ \beta_q)' \in \mathbb{R}^k$. *The condition* $\sum_{j=1}^{q} \beta_j < 1$ *is required to ensure covariance stationarity.*

Remark 4.5 (Diurnal adjustment)
Because financial durations exhibit intraday seasonality (see Section 4.3.1), this seasonality must be removed from the durations prior to model estimation. This is done by standardizing the raw durations by a deterministic component. Let ϕ_i *be the deterministic component. Then we can rewrite the Log ACD*$_1(p,q)$ *model as*

$$\ln \tilde{\psi}_i = \omega + \sum_{j=1}^{p} \alpha_j \ln \epsilon_{i-j} + \sum_{j=1}^{q} \beta_j \ln \tilde{\psi}_{i-j} \ , \tag{4.33}$$

$$x_i = \phi_i \tilde{\psi}_i \epsilon_i \ , \tag{4.34}$$

and the Log ACD$_2(p,q)$ *model as*

$$\ln \tilde{\psi}_i = \omega + \sum_{j=1}^{p} \alpha_j \epsilon_{i-j} + \sum_{j=1}^{q} \beta_j \ln \tilde{\psi}_{i-j} \ , \tag{4.35}$$

$$x_i = \phi_i \tilde{\psi}_i \epsilon_i \ , \tag{4.36}$$

where $\tilde{x}_i = \dfrac{x_i}{\phi_i}$ *is called the standardized duration.*

In the same view as the ACD(p, q) model can be seen as the counterpart of the GARCH(p, q) model for duration processes, both Log ACD(p, q) models are considered as the counterpart of the EGARCH(p, q) model for duration processes (Bauwens et al., 2009, [BH09]). Also, the Log ACD$_1(p, q)$ model has the property that it can be formulated in terms of an ARMA model for $\ln x_i$.

Remark 4.6 (Log ACD$_1(p, q)$ model as ARMA$(\max(p, q), \max(p, q))$ model)
The Log ACD$_1(p, q)$ model, as stated in Definition 2.2, can be formulated in terms of an ARMA$(\max(p, q), \max(p, q))$ model for $\ln x_i$, i.e.

$$\ln x_i = \tilde{\omega} + \sum_{j=1}^{\max(p,q)} \delta_j \ln x_{i-j} + \sum_{j=1}^{\max(p,q)} \theta_j \nu_{i-j} + \nu_i \ , \tag{4.37}$$

where

$$\tilde{\omega} = \bar{\omega} + \sum_{j=1}^{\max(p,q)} \theta_j E(\ln \epsilon_i) + E(\ln \epsilon_i) \ , \tag{4.38}$$

$$\nu_i = \ln \epsilon_i - E(\ln \epsilon_i) \ . \tag{4.39}$$

Derivation 4.3 (Log ACD$_1(p, q)$ model as ARMA$(\max(p, q), \max(p, q))$ model)

We refer to Appendix E.7 for the proof.

□

4.5 Parameter estimation

Given the sequence of durations $\{x_i \ : \ i = 1, \ \ldots, \ N(nT)\}$, the aim is to estimate the unknown parameter set $\boldsymbol{\theta} = (\omega, \ \alpha_1, \ \ldots, \ \alpha_p, \beta_1, \ \ldots, \ \beta_q)' \in \mathbb{R}^k$ corresponding to the ACD(p, q) model as defined in Definition 4.2. The most common method is to estimate the parameter set by using the maximum likelihood (ML) method. Before using the ML method, the joint density function of the sequence of durations $\{x_i \ : \ i = 1, \ \ldots, \ N(nT)\}$ conditioned on the parameter set $\boldsymbol{\theta}$ is needed. Using the fact that the joint density function can be written as the product of the conditional density, the following result is retrieved

$$\begin{aligned} f\left(\boldsymbol{x_m} \,\middle|\, \boldsymbol{\theta}\right) &= f\left(\boldsymbol{x}_{\max(\boldsymbol{p},\boldsymbol{q})} \,\middle|\, \boldsymbol{\theta}\right) f\left(x_{\max(p,q)+1} \,\middle|\, \boldsymbol{x}_{\max(\boldsymbol{p},\boldsymbol{q})}, \boldsymbol{\theta}\right) \cdots f\left(x_m \,\middle|\, \boldsymbol{x_{m-1}}, \boldsymbol{\theta}\right) \\ &= f\left(\boldsymbol{x}_{\max(\boldsymbol{p},\boldsymbol{q})} \,\middle|\, \boldsymbol{\theta}\right) \prod_{i=\max(p,q)+1}^{m} f\left(x_i \,\middle|\, \boldsymbol{x_{i-1}}, \boldsymbol{\theta}\right) \ , \end{aligned} \tag{4.40}$$

where $m = N(nT)$ is used for better readability. Mind the subtle difference in notation, as x_i refers to a single duration defined by Equation (4.6) and $\boldsymbol{x_i}$ refers to the vector of durations $(x_1, \ \ldots, \ x_i)' \in \mathbb{R}^i$.

It should be noticed that the joint density function of the sequence $\{x_i \ : \ i = 1, \ \ldots, \ \max(p,q)\}$, conditioned on the parameter set $\boldsymbol{\theta}$, can be ignored when applying the ML method. This is due to the fact that this joint density function is complicated but the impact on the likelihood (or log likelihood) diminishes as the sample size increases (Tsay, 2001, [Tsa01]). Thus, we obtain the following likelihood (or log likelihood) function

$$\ell\left(\boldsymbol{\theta} \,\middle|\, \boldsymbol{x_m}\right) = \prod_{i=\max(p,q)+1}^{m} f\left(x_i \,\middle|\, \boldsymbol{x_{i-1}}, \boldsymbol{\theta}\right) \ , \tag{4.41}$$

$$\mathcal{L}\left(\boldsymbol{\theta} \,\middle|\, \boldsymbol{x_m}\right) = \log\left(\prod_{i=\max(p,q)+1}^{m} f\left(x_i \,\middle|\, \boldsymbol{x_{i-1}}, \boldsymbol{\theta}\right)\right) = \sum_{i=\max(p,q)+1}^{m} \log f\left(x_i \,\middle|\, \boldsymbol{x_{i-1}}, \boldsymbol{\theta}\right) \ . \tag{4.42}$$

Let f_ϵ be the probability density of the errors ϵ_i. In case the true distribution of ϵ_i is known, the estimation procedure is straightforward. The parameters of the ACD(p,q) model are estimated by maximizing the likelihood (or log likelihood) function with respect to the parameter set $\boldsymbol{\theta}$. In case the distribution of ϵ_i is unknown, which is usually the case in practice, some distribution has to be chosen for ϵ_i. It should be noted that any distribution defined on positive support can be specified for f_ϵ (Pacurar, 2008, [Pac08]). Some choices for the density f_ϵ found in literature are the Weibull, generalized gamma and Burr, resulting in respectively the WACD model, GACD model and Burr-ACD model. A more natural choice for the density f_ϵ is the exponential, as the exponential distribution is considered to be the central distribution for stochastic processes defined on positive support (Hautsch, 2012, [Hau12]).

Remark 4.7

An advantage of choosing the exponential distribution for probability density f_ϵ is that it provides quasi-maximum likelihood (QML) estimators for the parameter set $\boldsymbol{\theta}$ that are consistent. Hence, the QML method based on the exponential distribution will produce consistent estimators regardless of the true distribution of ϵ_i. In case the exponential distribution is assumed, the conditional density is given by

$$f\left(x_i \;\middle|\; \boldsymbol{x_{i-1}}, \boldsymbol{\theta}\right) = \frac{1}{\psi_i} \exp\left(-\frac{x_i}{\psi_i}\right) \ , \tag{4.43}$$

resulting in the quasi log likelihood

$$\mathcal{L}\left(\boldsymbol{\theta} \;\middle|\; \boldsymbol{x_m}\right) = -\sum_{i=\max(p,q)+1}^{m} \left(\log \psi_i + \frac{x_i}{\psi_i}\right) \ . \tag{4.44}$$

Note that this result only holds when conditional mean function, as defined in Equation (4.18), is correctly specified. The result was first proven by Gourieroux et al. (1984) for independent observations. Engle (2000) extended the result for observations that exhibit dependence.

In this thesis, we will use the ML method to estimate the unknown parameter set $\boldsymbol{\theta}$. We derive the log likelihood, as defined by Equation (4.42), in accordence with the chosen distribution for the probability density f_ϵ. Hereafter, both the score function and the Hessian function are derived. The score function is defined as the first order derivative of the likelihood function (or log likelihood function), which we denote by

$$s\left(\boldsymbol{\theta} \;\middle|\; \boldsymbol{x_m}\right) = \frac{\partial \mathcal{L}\left(\boldsymbol{\theta} \;\middle|\; \boldsymbol{x_n}\right)}{\partial \boldsymbol{\theta}} \ . \tag{4.45}$$

The Hessian function is defined as the second order derivative of the likelihood function (or log likelihood function), represented by

$$H\left(\boldsymbol{\theta} \;\middle|\; \boldsymbol{x_m}\right) = \frac{\partial^2 \mathcal{L}\left(\boldsymbol{\theta} \;\middle|\; \boldsymbol{x_n}\right)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \ . \tag{4.46}$$

The log likelihood function is maximized by setting the score function equal to zero and solving it for the parameter set $\boldsymbol{\theta}$. This gives us our estimate $\widehat{\boldsymbol{\theta}}$. In case the Hessian function is negative definite for this estimate $\widehat{\boldsymbol{\theta}}$, the log likelihood is indeed maximized.

4.6 Testing

After fitting the ACD model using the estimation techniques presented in the previous section, it is crucial to check the correctness of the estimated model such that we can treat the obtained parameters with confidence. While there is a lot of literature on the different model specifications of the ACD model, the evaluation of these model specifications has not yet received much attention (Li et al., 2003, [LY03]) (Fernandes et al., 2005, [FG05]) (Meitz et al., 2006, [MT06]) (Pacurar, 2008, [Pac08]).

In this section the available tests concerning the adequacy of the estimated ACD model are elaborated. The flexibility of the ACD model arises from the fact that different choices can be made for both the functional form of the conditional mean function and the distributional form of the probability density function of the error term. In this thesis a distinction will be made between tests on the specification of the functional form of the conditional mean function and tests on the specification of the error term. The latter one can again be divided into two categories: tests on the independence assumption and tests on the distributional assumption. We will start by discussing the tests on the specification of the functional form of the conditional mean function by using Lagrange Multiplier (LM) tests. Hereafter, tests on the specification of the error term are elaborated. Firstly, tests regarding the distributional assumption of the error term are discussed which includes an introduction to probability integral transforms in combination with the Kolmogorov-Smirnov test. Secondly, tests regarding the independence assumption of the error term are described. These tests are more common in empirical studies and include portmanteau tests, such as the classical Ljung-Box test and the relatively new Li-Yu test. The testing framework is summarized by Figure 4.2.

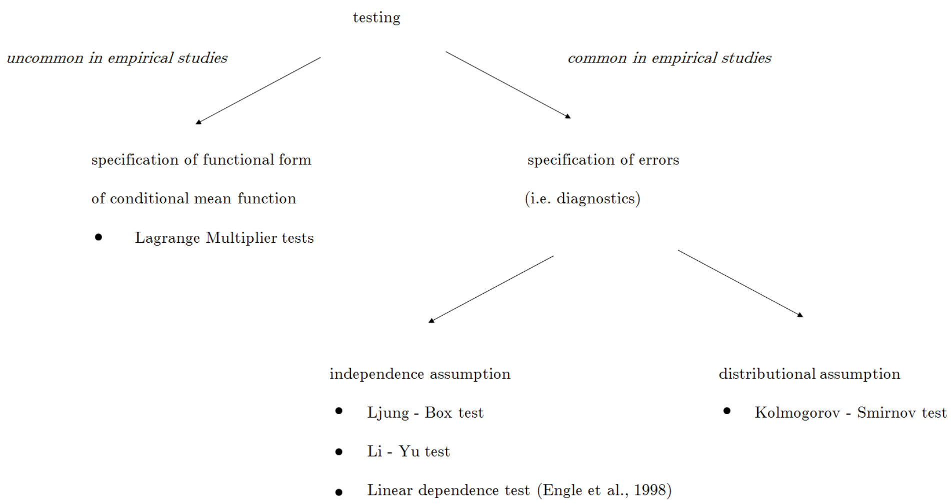

Figure 4.2: Graphic representation of the testing framework for ACD models.

4.6.1 Specification of conditional mean function

Remember that in case the exponential distribution is chosen for the error term, the QML method will produce consistent estimators regardless of the true distribution of the error term. However, this result only holds in case the conditional mean function is correctly specified. Hence, in case the conditional mean function is misspecified, we might run into the risk that the estimated parameter set does not converge in probability to the true parámete set as the sample size increases. Therefore, testing this type of misspecification is an essential part of assessing the adequacy of the estimated model in case the QML method is used.

Meitz et al. (2006) introduced, for the class of exponential ACD models, a Lagrange Multiplier (LM) test in order to detect misspecification of the functional form of the conditional mean function. Let $\psi_i = \psi_i\left(\mathcal{F}_{i-1}, \boldsymbol{\theta}_1\right)$ refer to the conditional mean function as defined in Equation (4.7) and let $\boldsymbol{\theta}_1$ be the corresponding parameter set. Two different types of misspecification are distinguished: additive misspecification and multiplicative misspecification. The conditional mean function is either additively misspecified,

$$x_i = \left(\psi_i + \xi_i\right)\epsilon_i \ , \tag{4.47}$$

or multiplicatively misspecified,

$$x_i = \psi_i \xi_i \epsilon_i \ , \tag{4.48}$$

where $\xi_i = \xi_i\left(\mathcal{F}_{i-1}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right)$ denotes an $\mathcal{F}_{i-1}$ measurable function that depends on an additional parameter set $\boldsymbol{\theta}_2$. Equation (4.47) and (4.48) will form the basis of the LM tests for additive and multiplicative misspecification of the functional form of the conditional mean function. Both tests can be found Appendix B.1. As the tests are rather general and only defined for the class of exponential ACD models, we will not apply these tests in this thesis. For more information about these LM tests the reader is referred to (Meitz et al., 2006, [MT06]).

4.6.2 Specification of error term

Next to testing the specification of the functional form of the conditional mean function, the obtained residuals after estimation could be investigated in order to check the correctness of the estimated ACD model. Going back to the definition of the ACD model, the obtained residuals are defined by

$$\hat{\epsilon_i} = \frac{x_i}{\hat{\psi}_i} \ . \tag{4.49}$$

The properties of these obtained residuals facilitate a way to assess the correctness of the estimated model. To be more specific, in case the estimated model is adequate the obtained residuals $\hat{\epsilon}_i$ should be IID non-negative random variables with a distribution that corresponds with the specified distribution of the ACD errors and a mean equal to 1. In this thesis, a distinction will be made between testing the distributional assumption of the error term and testing the independence assumption.

4.6.2.1 Distributional assumption

Bauwens et al. (2004) proposed a test to verify the correctness of the distributional form of the conditional probability density, which is based on using probability integral transforms. Let $\{y_i \; : \; i = 1, \; \ldots, \; n\}$ be the sequence of the realizations of the process of interest. Consider the sequence of (unknown) true densities $\{p^0 \left(y_i \;\middle|\; \boldsymbol{y_{i-1}} \right) \; : \; i = 1, \; \ldots, \; n\}$. Furthermore, consider the sequence of estimated 1-step-ahead density forecasts $\{p \left(y_i \;\middle|\; \boldsymbol{y_{i-1}} \right) \; : \; i = 1, \; \ldots, \; n\}$, i.e. the sequence of densities defined for the next observation of the variable y_i. These densities are connected with each other by the probability integral transform $z_{i \,|\, i-1}$, which is defined by

$$z_{i \,|\, i-1} = \int_{-\infty}^{y_i} p \left(s \;\middle|\; \boldsymbol{y_{i-1}} \right) \mathrm{d}s = P \left(y_i \;\middle|\; \boldsymbol{y_{i-1}} \right) \; . \tag{4.50}$$

Diebold et al. (1998) proved that the sequence of probability integral transforms will be IID random variables distributed according to the standard uniform distribution. The exact proposition is stated below.

Proposition 4.1 (Distribution probability integral transforms)
Suppose $\{y_i \; : \; i = 1, \; \ldots, \; n\}$ is generated from $\{p^0 \left(y_i \;\middle|\; \boldsymbol{y_{i-1}} \right) \; : \; i = 1, \; \ldots, \; n\}$. If a sequence of density forecasts $\{p \left(y_i \;\middle|\; \boldsymbol{y_{i-1}} \right) \; : \; i = 1, \; \ldots, \; n\}$ coincides with $\{p^0 \left(y_i \;\middle|\; \boldsymbol{y_{i-1}} \right) \; : \; i = 1, \; \ldots, \; n\}$, then under the usual condition of a nonzero Jacobian with continuous partial derivatives, the sequence of probability transforms of $\{y_i \; : \; i = 1, \; \ldots, \; n\}$ with respect to $\{p \left(y_i \;\middle|\; \boldsymbol{y_{i-1}} \right) \; : \; i = 1, \; \ldots, \; n\}$ is IID $U(0,1)$. That is,

$$z_{i \,|\, i-1} = \int_{-\infty}^{y_i} p \left(s \;\middle|\; \boldsymbol{y_{i-1}} \right) \mathrm{d}s = P \left(y_i \;\middle|\; \boldsymbol{y_{i-1}} \right) \overset{IID}{\sim} U(0,1) \tag{4.51}$$

Proof 4.1 (Distribution probability integral transforms)

We refer to (Diebold et al., 1998, [DGT98]) for the proof.

□

Applying these results to the probability integral transforms within the ACD model framework, will give us information on the (mis)specification of the distributional form of the estimated ACD model. Let $\{x_i \ : \ i = 1, \ \ldots, \ m = N(nT)\}$ be the sequence of durations. Let $f\left(x_i \;\middle|\; \boldsymbol{x_{i-1}}\right)$ be the density forecast of the true conditional density $f^0\left(x_i \;\middle|\; \boldsymbol{x_{i-1}}\right)$ for the i-th duration. If we for instance assume that the errors are distributed according to the exponential distribution or Weibull distribution, Equation (4.50) translates into respectively Equation (4.52) and (4.53).

$$\epsilon_i \overset{IID}{\sim} Exp(\lambda) \implies \hat{z}_{i \,\mid\, i-1} = 1 - e^{-\frac{x_i}{\hat{\psi}_i}} \tag{4.52}$$

$$\epsilon_i \overset{IID}{\sim} Weibull(\lambda, \gamma) \implies \hat{z}_{i \,\mid\, i-1} = 1 - e^{-\left(\frac{x_i \Gamma\left(1+\frac{1}{\gamma}\right)}{\hat{\psi}_i}\right)^{\gamma}} \tag{4.53}$$

The (one-sided) Kolmogorov-Smirnov test can now be used to the test obtained probability integral transforms against the standard uniform distribution. The Kolmogorov-Smirnov test is used to test whether a sample comes from a certain distribution, and compares the empirical cumulative distribution of the sample against a hypothetical cumulative distribution. The test is stated below. Note that in our case the sequence $Y_1, \ \ldots, \ Y_n$ corresponds with the obtained sequence of probability integral transforms and the hypothetical cumulative distribution G_0 corresponds with the standard uniform distribution.

Theorem 4.1 (Kolmogorov-Smirnov test)
Let $Y_1, \ \ldots, \ Y_n$ be IID random variables with empirical cumlative distribution function $\hat{G}_n$. Let G_0 be a specified hypothetical continuous distribution function. Furthermore, assume that under the null hypothesis $H_0 : \hat{G}_n = G_0$. Then, under H_0, the Kolmogorov-Smirnov test statistic

$$D_n = \sup_{y \in \mathbb{R}} \left| \hat{G}_n(y) - G_0(y) \right| \tag{4.54}$$

satisfies

$$\lim_{n \longrightarrow \infty} P\left(\sqrt{n} D_n \leq y\right) = L(y) = 1 - 2 \sum_{i=1}^{\infty} (-1)^{i-1} e^{-i^2 y^2} \ . \tag{4.55}$$

H_0 is rejected at significance level $(1-\alpha)\%$ in case $\sqrt{n} D_n > L^{-1}(1-\alpha)$.

Proof 4.2 (Kolmogorov-Smirnov test)

We refer to Section 3 in (Feller, 1948, [Fel48]) for the proof.

□

The critical values of the Kolmogorov-Smirnov test statistic are tabulated for small values of n, see for instance (Smirnov, 1948, [Smi48]). The critical values of the Kolmogorov-Smirnov test statistic for large enough values of n can be approximated by using the asymptotic distribution above, denoted by Equation (4.55). One important property of the Kolmogorov-Smirnov test statistic is that it is distribution-free, which means that under the null hypothesis the distribution of the test statistic does not depend on the hypothetical distribution function G_0. This property is stated below and demonstrates that it suffices to study the case when the hypothetical cumulative distribution function is the uniform distribution on $[0, 1]$.

Remark 4.8 (Kolmogorov-Smirnov test statistic distribution-free)
Under H_0, the distribution of D_n does not depend on the continuous underlying function G_0.

Proof 4.3 (Kolmogorov-Smirnov test statistic distribution-free)

We refer to Appendix E.8 for the proof.

□

Rejection of the null hypothesis can be due to both the misspecification of the distributional form of the conditional probability density and the misspecification of the functional form of the conditional mean function. Plotting a histogram of the sequence of probability integral transforms can reveal more information about the reasons for misspecification. By plotting a histogram, departures from the standard uniform distribution can be easily detected. For instance, a humped shaped histogram suggests that the forecasts do not take into account the tails of the true distribution. An U shaped histogram suggests that the forecasts under- or overestimate too often.

4.6.2.2 Independence assumption

By testing the presence of autocorrelation in the obtained residuals, the independence assumption is investigated. The most common test found in the literature for testing the presence of autocorrelation in the residuals after fitting an ACD(p, q) model are portmanteau tests such as the classical Ljung-Box test. Ljung et al. (1978) examined the residuals after estimating an appropriate ARMA(p, q) model, proposed a test statistic based on these residuals and derived the corresponding asymptotic distribution. This asymptotic distribution turns out to be a χ^2-distribution. The exact theorem concerning the Ljung-Box test is stated below.

Theorem 4.2 (Ljung-Box test)
Consider the ARMA(p, q) model, and denote the residuals with $\hat{\epsilon}_i$ and the k-th lag sample autocorrelations are given by

$$\hat{\rho}_k = \frac{\sum_{i=k+1}^{n} (\hat{\epsilon}_i - \bar{\hat{\epsilon}})(\hat{\epsilon}_{i-k} - \bar{\hat{\epsilon}})}{\sum_{i=1}^{n} (\hat{\epsilon}_i - \bar{\hat{\epsilon}})^2} \qquad k = 1, 2, \ldots \tag{4.56}$$

Let $s = p + q$ represent the number of parameters to be estimated (remark $s = \max(p, q)$ in the case of an ACD(p, q) model). Furthermore, assume that under the null hypothesis H_0 : $\rho_1 = \ \ldots = \rho_k = 0$. Then, under H_0, the Ljung-Box test statistic

$$Q_{LB}(m) = n(n+2) \sum_{k=1}^{m} \frac{\hat{\rho}_k^2}{n-k} \overset{a}{\sim} \chi^2_{m-s} \tag{4.57}$$

H_0 is rejected at significance level $(1-\alpha)\%$ in case $Q_{LB}(m) > \chi^2_{m-s}(1-\alpha)$.

Proof 4.4 (Ljung-Box test)

We refer to (Ljung et al., 1978, [LB78]) for the proof.

□

Although the Ljung-Box test is the most applied (and only) test found in empirical studies in order to check for autocorrelation in the residuals after estimating ACD models, the test results are only valid when applied to the residuals after estimating ARMA models. For instance, when the Ljung-Box test is applied to the residuals after estimating GARCH model the test results are invalid as the test statistic does not follow the asymptotic χ^2-distribution under the null hypothesis (Li et al., 1994, [LM94]). Therefore, Pacurar (2008) stresses that the Ljung-Box results regarding the asymptotic distribution of the test statistic cannot be carried over to ACD models. Nevertheless, the Ljung-Box test can be used to give an approximation about the presence of autocorrelation in the residuals.

Li et al. (2003) investigated the the asymptotic distribution of the residuals after estimating ACD models, which results in a goodness of fit test for this kind of model. The latter test will be named the Li-Yu test throughout the remainder of this thesis, and is preferred over the Ljung-Box test. However, it should be noted that the test is restricted to the ACD$(p, 0)$ model.

Theorem 4.3 (Li-Yu test)
Consider the ACD$(p, 0)$ model such that $\epsilon_i \sim Exp(1)$, and denote the residuals with $\hat{\epsilon}_i$ and the k-th lag sample autocorrelations are given by

$$\hat{\rho}_k = \frac{\sum_{i=k+1}^{n} (\hat{\epsilon}_i - 1)(\hat{\epsilon}_{i-k} - 1)}{n} \qquad k = 1, 2, \ldots \tag{4.58}$$

Assume that under the null hypothesis $H_0 : \rho_1 = \ldots = \rho_k = 0$. Then, under H_0, the Li-Yu test statistic

$$Q_{LY}(m) = n\hat{\boldsymbol{\rho}}' \left(\mathbf{I}_m - \mathbf{X}\mathbf{G}^{-1}\mathbf{X}'\right)^{-1} \hat{\boldsymbol{\rho}} \overset{a}{\sim} \chi^2_m \tag{4.59}$$

where

$$\hat{\boldsymbol{\rho}} = (\hat{\rho}_1, \; \ldots, \; \hat{\rho}_k)' \tag{4.60}$$

$$\mathbf{I}_m \; m \times m \text{ identity matrix} \tag{4.61}$$

$$\mathbf{X} = \begin{pmatrix} \frac{1}{n}\sum_{i=2}^{n} \frac{x_i}{\psi_i^2}(e_{i-1} - 1) & \frac{1}{n}\sum_{i=2}^{n} \frac{x_i x_{i-1}}{\psi_i^2}(e_{i-1} - 1) \\ \vdots & \vdots \\ \frac{1}{n}\sum_{i=2}^{n} \frac{x_i}{\psi_i^2}(e_{i-k} - 1) & \frac{1}{n}\sum_{i=2}^{n} \frac{x_i x_{i-1}}{\psi_i^2}(e_{i-k} - 1) \end{pmatrix} \tag{4.62}$$

$$\mathbf{G} = -E\left(\frac{1}{n} H\left(\boldsymbol{\theta} \;\middle|\; \boldsymbol{x_n}\right)\right) \tag{4.63}$$

H_0 is rejected at significance level $(1 - \alpha)\%$ in case $Q_{LY}(m) > \chi^2_m(1 - \alpha)$.

Proof 4.5 (Li-Yu test)

We refer to (Li et al., 2003, [LY03]) for the proof.

□

The conditional mean function, as given by Equation (4.17), is assumed to be a linear parametrization of the past durations and the past conditional durations both belonging to the set $\mathcal{F}_{i-1}$. Engle et al. (1998) proposed an additional test regarding the independence assumption of the error term, which can also verify the correctness of this linear parametrization. The main idea is to divide the durations into appropriate sized bins ranging from 0 to ∞. Call these bins $B_1, B_2, \ldots, B_k$. Then regress the residuals $\hat{\epsilon}_i$ on a constant and the indicators of the size of the previous duration, i.e.

$$\hat{\epsilon}_i = \alpha_0 + \alpha_1 \mathbb{1}_{x_{i-1} \in B_1} + \alpha_2 \mathbb{1}_{x_{i-1} \in B_2} + \ \ldots \ + \alpha_k \mathbb{1}_{x_{i-1} \in B_k} = \alpha_0 + \sum_{i=1}^{k} \alpha_i \mathbb{1}_{x_{i-1} \in B_i} \quad (4.64)$$

Under the null hypothesis of the residuals being independently distributed, no predictability should be implied from this regression and the coefficients $\alpha_0,\ \alpha_1, \ldots, \alpha_k$ should not be significantly different from zero. In case the coefficients $\alpha_0,\ \alpha_1, \ldots, \alpha_k$ are significantly different from zero, this null hypothesis is rejected. The coefficients that are significantly different from zero can be examined in order to identify the sources of misspecification of the functional form of the conditional mean function.

4.7 Model selection

Section 4.3 introduced the ACD model. By combining different specifications for the conditional mean function and the probability density function of the error term, a wide variety of ACD models is available for application. While ACD model applications appear often in literature (see for instance (Dionne et al., 2009, [DDP09]), (Engle, 2000, [Eng00]), (Engle et al., 1998, [ER98])), the methodology behind the model selection remains rather vague. In this section, we address this gap by introducing a model selection methodology. The aim of this methodology is to quantify the quality of each ACD model and to provide a straightforward and easily adjustable framework for selecting an appropriate model.

4.7.1 Scoring function

The central element of the model selection methodology is the so called scoring function, which will be used to quantify the quality of each estimated ACD model. Let m denote the number of candidate ACD models. These m models form the input of the scoring function. The output is a vector $\mathbf{Y} \in \mathbb{R}^m$, where Y_i represents the score of the i-th ACD model.

A distinction is made between two different types of scores: a performance score and a specifcation score. The performance score focuses on the performance capability of the estimated ACD model. This capability is measured in terms of the Mean Squared Error (MSE) and Mean Squared Prediction Error (MSPE), which will be discussed in the next section. The specification score indicates the appropriateness of the chosen specification of the ACD model (i.e. specification of the conditional mean function and the error term). Here, the tests discussed in Section 4.5 will be used. This scoring strategy is summarized by Figure 4.3. Each bullet point reflects an element that contributes to the score Y_i. The scoring function calculates Y_i, by summing the scores

$$[\,] \times \frac{p}{m} \;, \tag{4.65}$$

for each element. Here, $p \in [1, m]$ and $[\,]$ reflects the weight of the element. $p = m$ if the i-th ACD model performs best, $p = m-1$ if the i-th ACD model performs second best, etcetera.

The linear dependence test and the Kolmogorov-Smirnov test, as discussed in Section 4.5, form an exception as these tests can be either passed of failed. Therefore, the scoring function calculates Y_i, by adding up the scores

$$[\,] \times p \;, \tag{4.66}$$

for these elements. Here, $p \in \{0, 1\}$ and $[\,]$ reflects again the weight of the element. $p = 1$ if the i-th ACD model passes the test and 0 otherwise. As the asympotic distribution of the Ljung-Box test statistic does not hold for ACD models and only gives an approximation about the presence of autocorrelation in the residuals, it is decided to treat this test not as a pass of fail type of test. The score will be determined by Equation (4.65). A high value of the Ljung-Box test statistic reflects the presence of autocorrelation in the residuals. Hence, the ACD model with the highest test statistic is considered to be the worst performing (i.e. $p = 1$) and the ACD model with the lowest test statistic is considered to be the best performing (i.e. $p = m$).

Hence, the ACD model for which $Y_i = \max_{1 \leq i \leq m} Y_i$ will be selected in this thesis.

Remark 4.9 (Weights)
Table 4.1 presents the weights that are assigned to each element of the scoring function. Hence, we have 2 elements that address the performance score (total weight of [30]*), and 6 elements that reflect the specification score (total weight of* [70]*). The maximum score possible equals 100 and the lowest score possible equals 6. Mind that the weights can be adjusted to the preference of the modeller.*

element	weight
score with respect to performance	
Mean Squared Error (MSE)	[15]
Mean Squared Prediction Error (MSPE)	[15]
score with respect to specification	
Akaike Information Criterion (AIC)	[10]
Bayesian Information Criterion (BIC)	[10]
Ljung - Box test statistic residuals	[10]
Ljung - Box test statistic residuals^2	[10]
Linear dependence test (Engle et al., 1998)	[10]
Kolmogorov - Smirnov test	[20]

Table 4.1: Weights that are assigned to each element of the scoring function.

4.7.1.1 Performance measures

As stated in the previous section, the performance capability of the ACD model is measured in terms of the Mean Squared Error (MSE) and Mean Squared Prediction Error (MSPE). Given the sequence of durations, the unknown parameter set $\boldsymbol{\theta} = (\omega,\ \alpha_1,\ \ldots,\ \alpha_p, \beta_1,\ \ldots,\ \beta_q)' \in \mathbb{R}^k$ is estimated corresponding to the chosen ACD model (for instance ACD(p, q), Log ACD$_1(p, q)$, Log ACD$_2(p, q)$) In this thesis the ML method is used, and $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^k$ denotes the parameter set that maximizes the (log) likelihood function.

In our application the original dataset $\{x_i\ :\ i = 1,\ \ldots,\ m = N(nT)\}$ is split into two subsets: an estimation set and a validation set. Let $\{x_i\ :\ i = 1,\ \ldots,\ k\}$ denote the estimation set and $\{x_i\ :\ i = k+1,\ \ldots,\ k+l = N(nT)\}$ denote the validation set. The unknown parameter set $\boldsymbol{\theta}$ is estimated based on the estimation set. The estimated parameter set $\widehat{\boldsymbol{\theta}}$ is used to obtain the sequence $\{\widehat{x}_i\ :\ i = 1,\ \ldots,\ m = N(nT)\}$. The MSE, as stated in Definition 4.5, measures the mean squared difference between this sequence and the actual durations for the estimation set. The MSPE, as stated by Definition 4.6, measures the mean squared difference between this sequence and the actual durations of the validation set. Hence, the MSPE allows us to compare the predictive power of a set of candidate ACD models.

Definition 4.5 (Mean Squared Error)
The Mean Squared Error (MSE) is given by

$$MSE = \frac{1}{k} \sum_{i=1}^{k} (\hat{x}_i - x_i)^2 \ . \tag{4.67}$$

Given a set of candidate ACD models, the preferred model is the one with the minimum MSE.

Definition 4.6 (Mean Squared Prediction Error)
The Mean Squared Prediction Error (MSPE) is given by,

$$MSPE = \frac{1}{N(nT) - k} \sum_{i=k+1}^{N(nT)} (\widehat{x}_i - x_i)^2 \ . \tag{4.68}$$

Given a set of candidate ACD models, the preferred model is the one with the minimum MSPE.

4.7.1.2 Specification measures

To determine the specification score, the tests of Section 4.5 will be used. For an overview of the Kolmogorov-Smirnov test to test the distributional assumption of the error term, the reader is reffered to Section 4.5.2.1. For an overview of the Ljung-Box test and the linear dependence test (Engle et al., 1998) to test the independence assumption of the error term, the reader is referred to Section 4.5.2.2.

However, as mentioned in Section 4.5.1, the available tests concerning the specification of the conditional mean function are only defined for the class of exponential ACD models. Hence, two different criteria are used to assess the specification of the conditional mean function: the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The AIC is an estimate of the Kullback-Leibler distance, and takes into account the value maximized log likelihood function in combination with the number of parameters. A penalty is imposed for increasing the number of parameters. Definition 4.7 denotes the AIC. The BIC is closely related to the AIC, but the penalty imposed for increasing the number of parameter is stronger. Definition 4.8 gives the definition of the BIC.

Definition 4.7 (Akaike Information Criterion)
Let $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^k$ be he parameter set that maximizes the (log) likelihood function $\mathcal{L}$ (see Section 4.3). Then, the Akaike Information Criterion (AIC) is defined by

$$AIC = 2k - 2\mathcal{L}\left(\widehat{\boldsymbol{\theta}} \,\middle|\, \boldsymbol{x_m}\right) \ . \tag{4.69}$$

Given a set of candidate ACD models, the preferred model is the one with the minimum AIC value.

Definition 4.8 (Bayesian Information Criterion)
Let $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^k$ *be he parameter set that maximizes the (log) likelihood function* $\mathcal{L}$ *(see Section 4.3). Then, the Bayesian Information Criterion (BIC) is defined by*

$$BIC = \log(n)k - 2\mathcal{L}\left(\widehat{\boldsymbol{\theta}} \;\middle|\; \boldsymbol{x_m}\right) \ . \tag{4.70}$$

Given a set of candidate ACD models, the preferred model is the one with the minimum BIC value.

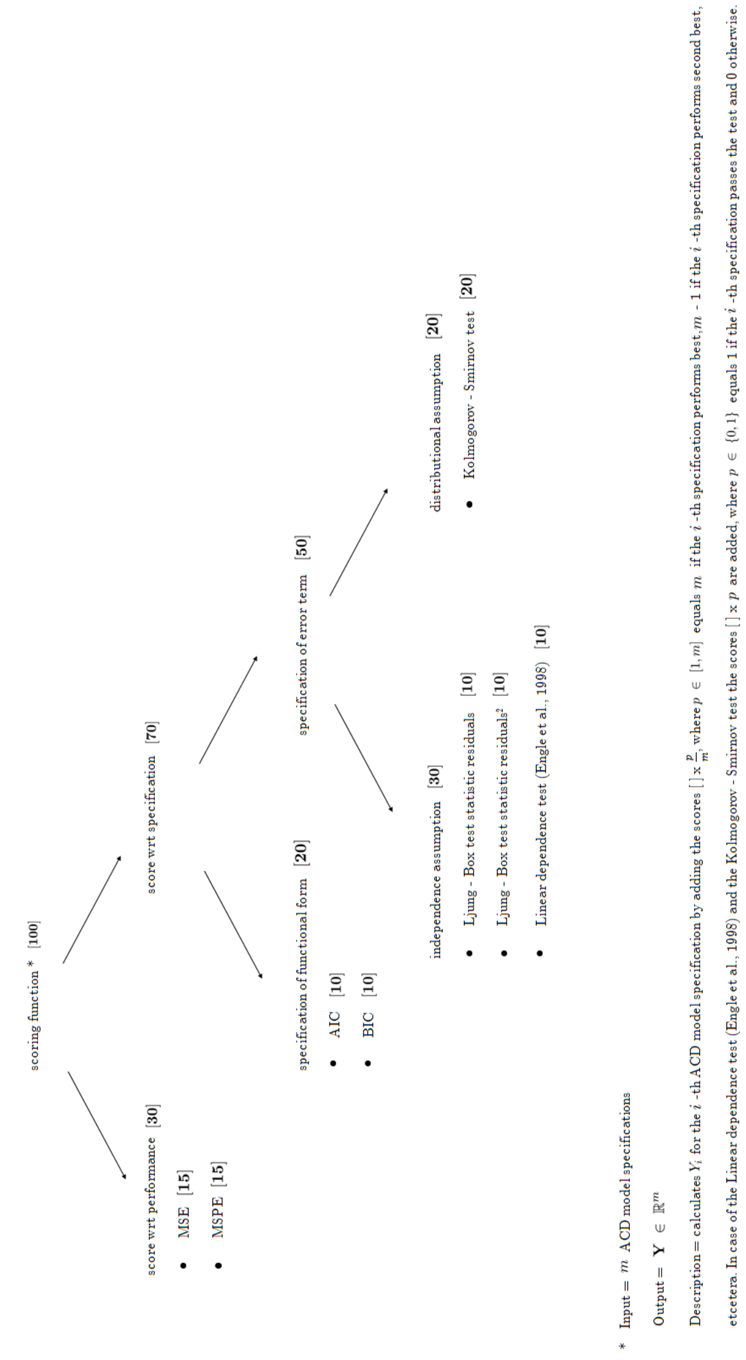

Figure 4.3: Graphic representation of the scoring function scheme.

4.8 Results from application

Author's note: this section is confidential.

5

Univariate risk metric : intraday risk measure

5.1 Introduction

Chapter 4 introduced the concept of point processes, and proposed the class of ACD models to capture the behavior of the durations of simple point processes. In this chapter, the associated marks to these durations will be studied. We focus on the net positions associated with the durations.

Despite the rise of econometric models for irregular spaced data, very few contributions link high frequency data to risk management (Dionne et al., 2009, [DDP09]). Dionne et al. (2009) examined the use of ultra high frequency data for market risk management with an application to stock returns. They introduced an intraday risk measure by using Monte Carlo simulations. The general idea is to use the ACD model to define the time steps of the simulations, and to use the new UHF-GARCH model (which will be elaborated in the coming sections) to generate the corresponding tick-by-tick returns that are rescaled into fixed time intervals.

This chapter deals with the first research question, and aims to estimate the size of the liquidity buffer for each group separately. We follow the approach of Dionne et al. (2009), and tailor the intraday risk measure further for our purpose by using univariate EVT. Let IRM_i be the intraday risk measure at time t_i. Then, the aim is to find a risk measure that satisfies the following three conditions.

- IRM_i can be expressed as a function of variables known at time t_{i-1} and an unknown parameter set.
- IRM_i can be estimated by an algorithm.

- IRM_i can be tested to asses the quality of the estimator.

We will start this chapter by introducing the definition of the high frequency risk measure. Hereafter, the corresponding algorithm including Monte Carlo simulations will be elaborated. We end this chapter by presenting a testing framework for this risk measure.

5.2 Intraday risk measure

In this thesis, we focus on the transaction process of two specific accounts within ABN AMRO Bank. In this chapter, the transaction process will be viewed as a marked point process in order to account for the irregularity of durations between consecutive transactions. For the exact specification of the transaction process, the reader is referred to Chapter 2. For an introduction to the concept of point processes, the reader is referred to Chapter 4. The durations form the points and the net positions form the associated marks. Then, the complete transaction process can be viewed as $\left\{(x_i, NP_i) \ : \ i = 1, \ldots, N(nT)\right\}$ where n denotes the number of business days and T denotes the number of seconds within a business day. In this manner, the duration x_i allows us to keep trace of the time step of the net position NP_i.

In order to define our intraday risk measure, we first have to introduce the rescaled net positions. By defining rescaled net positions we follow the approach elaborated in (Dionne et al., 2009, [DDP09]), where rescaled returns are introduced for the purpose of market risk management. Let $\left\{y_l \ : \ l \in \mathbb{Z}\right\}$ be the sequence of net positions rescaled at fixed intervals of size T_{fixed}. Consider the realization $\left\{y_l \ : \ l = 1, \ldots, n^{'}\right\}$ with y_l obtained at the event times $t_l^{'}$ such that $t_l^{'} - t_{l-1}^{'} = T_{fixed}$. We define the rescaled net position corresponding to the l-th interval by

$$y_l = \sum_{i=\tau(l-1)}^{\tau(l)-1} NP_i \ , \tag{5.1}$$

with the restrictions $\sum_{i=\tau(0)}^{\tau(l)-1} x_i \leq lT_{fixed}$ and $\sum_{i=\tau(0)}^{\tau(l)} x_i > lT_{fixed}$ where $\tau(0) = 1$. This implies that $\tau(l)$ for $l = 1, \ldots, n^{'}$ is chosen such that the sum of the durations exceeds lT_{fixed} for the first time. A visual representation of these restrictions is given by Figure 5.1. Thus, the rescaled net positions are obtained by summing up the tick-by-tick net positions such that the sum of durations does not exceed T_{fixed}. The rescaled net positions y_l are also referred to as the T_{fixed}-period net positions of the l-th fixed interval.

Remark 5.1

The rescaled net position y_l of the l-th interval is equivalent to the net cumulative position at *the end of the l-th interval. Consequently, the rescaled net position y_l describes the liquidity need of ABN AMRO Bank at the end of the l-th interval. In case $y_l < 0$, ABN AMRO Bank needs access to liquidity to fund this balance. This remark will play a key role in the remainder of this chapter.*

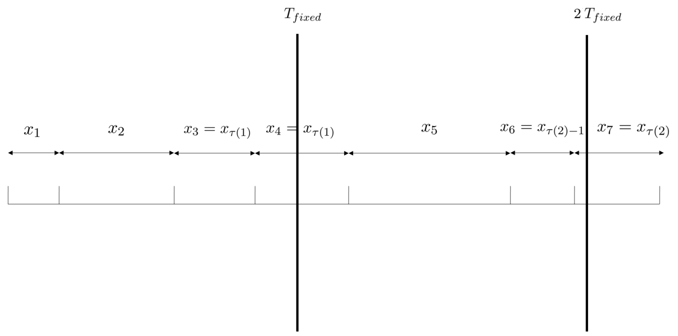

Figure 5.1: Graphic representation of the restrictions $\sum_{i=\tau(0)}^{\tau(l)-1} x_i \leq lT_{fixed}$ and $\sum_{i=\tau(0)}^{\tau(l)} x_i > lT_{fixed}$ for obtaining the T_{fixed}-period net positions.

Remark 5.1 implies that the rescaled net position y_l is not the variable of interest, as it only reflects the liquidity need of the bank at the end of the l-th interval. More interesting would be to observe the liquidity need of the bank within the l-th interval. Hence, the largest negative rescaled net position occurring within the l-th interval will be the building block of our intraday risk measure. We define the largest negative rescaled net position occurring within the l-th fixed interval by

$$z_l = \max\left\{ - \sum_{i=\tau(l-1)}^{j} NP_i \;:\; j = \tau(l-1), \ldots, \tau(l) - 1 \right\}, \tag{5.2}$$

for which still the restrictions $\sum_{i=\tau(0)}^{\tau(l)-1} x_i \leq lT_{fixed}$ and $\sum_{i=\tau(0)}^{\tau(l)} x_i > lT_{fixed}$ hold.

Definition 5.1 formally defines the intraday risk measure (IRM). The intraday risk measure $IRM_l(\alpha)$ corresponds to the conditional quantile of the largest negative rescaled net position z_l. $IRM_l(\alpha)$ represents the maximum liquidity need not to be succeeded within the l-th interval for a given confidence level $100(1-\alpha)\%$. In this thesis, the size of the liquidity buffer associated with the l-th interval is set to be equal to $IRM_l(\alpha)$. By estimating this conditional quantile for each interval l, we are able to obtain the size of the liquidity buffer for each interval.

Definition 5.1 (Intraday risk measure)
The intraday risk measure (IRM), for shortfall probability α, $IRM_l(\alpha)$ is defined by

$$P\left(z_l > IRM_l(\alpha) \;\middle|\; \mathcal{I}_{l-1}\right) = \alpha \ , \tag{5.3}$$

where the information set $\mathcal{I}_{l-1}$ contains the information until moment $\tau(l-1)$.

Remark 5.2
By defining the risk measure for the largest negative rescaled net positions instead of the rescaled net positions, we differ from the approach elaborated in (Dionne et al., 2009, [DDP09]). Dionne et al. (2009) defines the risk measure for rescaled returns. In the context of tick-by-tick stock returns, Definition 5.1 is referred to as Intraday Value at Risk (IVaR) which is denoted by $IVaR_l(\alpha)$. For more information about IVaR in the context of tick-by-tick stock returns, the reader is referred to (Dionne et al., 2009, [DDP09]).

5.3 Methodology

The question remains how the intraday risk measure $IRM_l(\alpha)$, as given by Definition 5.1, can be estimated. In this section this question will be elaborated. Again, the approach discussed in (Dionne et al., 2009, [DDP09]) will be our guide. The general idea is to use Monte Carlo simulations to obtain the estimator $\widehat{IRM}_l(\alpha)$. The ACD model is used to define the time steps of the simulations, and the new UHF-GARCH model is used to generate the corresponding net positions rescaled at fixed time intervals of size T_{fixed}. We start by introducing the UHF-GARCH model.

5.3.1 UHF-GARCH model

Engle (2000) proposed adapting the general GARCH model for the application to tick-by-tick data. Let NP_i be the net position at event time t_i. Let $\mathcal{G}_{i-1}$ denote the information set including both the information set $\mathcal{F}_{i-1}$ and the current duration x_i. Engle (2000) defined two conditional variances: the conditional variance per transaction, h_i and the conditional variance per unit time, σ_i^2. Let $h_i = V\left(NP_i \;\middle|\; \mathcal{G}_{i-1}\right)$ and $\sigma_i^2 = V\left(\frac{NP_i}{\sqrt{x_i}} \;\middle|\; \mathcal{G}_{i-1}\right)$. This implies the linear relationship $h_i = \sigma_i^2 x_i$. In case it is assumed that $E\left(NP_i \;\middle|\; \mathcal{G}_{i-1}\right) = 0$, then the conditional variance per unit time is modeled by a GARCH(r, s) process given by

$$\sigma_i^2 = \dot{\omega} + \sum_{j=1}^{r} \dot{\alpha}_j \left(\frac{NP_{i-j}}{\sqrt{x_{i-j}}}\right)^2 + \sum_{j=1}^{s} \dot{\beta}_j \sigma_{i-j}^2 \ , \tag{5.4}$$

which is known as the UHF-GARCH model. The formal definition is denoted in Definition 5.2. The parameters $\dot{\omega}$, $\dot{\alpha}_j$ and $\dot{\beta}_j$ of the UHF-GARCH model differ from the parameters of the ACD model. Dots are placed above the parameters of the UHF-GARCH model in order to differentiate.

Definition 5.2 (UHF-GARCH(r,s) model)
Under the assumptions $\sigma_i^2 = V\left(\frac{NP_i}{\sqrt{x_i}} \,\Big|\, \mathcal{G}_{i-1}\right)$ *and* $h_i = V\left(NP_i \,\Big|\, \mathcal{G}_{i-1}\right)$*, the conditional variance per unit time follows a GARCH(r,s) model:*

$$\sigma_i^2 = \dot{\omega} + \sum_{j=1}^{r} \dot{\alpha}_j \left(\frac{NP_{i-j}}{\sqrt{x_{i-j}}}\right)^2 + \sum_{j=1}^{s} \dot{\beta}_j \sigma_{i-j}^2 \ , \tag{5.5}$$

$$NP_i = \sqrt{x_i}\sigma_i\nu_i \ . \tag{5.6}$$

Or equivalently,

$$h_i = \dot{\omega} x_i + \sum_{j=1}^{r} \left(\dot{\alpha}_j \frac{x_i}{x_{i-j}}\right) NP_{i-j}^2 + \sum_{j=1}^{s} \left(\dot{\beta}_j \frac{x_i}{x_{i-j}}\right) h_{i-j} \ , \tag{5.7}$$

$$NP_i = \sqrt{h_i}\nu_i \ , \tag{5.8}$$

where it is assumed that the errors ν_i *are IID random variables with mean 0 and variance 1.*

Remark 5.3
The UHF-GARCH(r,s) model of Definition 5.2 can be easily extended by changing the representation of the conditional variance per unit time. Below three potential different representations are described, of which the third one has proven to be the most successful (Engle, 2000, [Eng00]).

1. $\sigma_i^2 = \dot{\omega} + \dot{\alpha}\left(\frac{NP_{i-1}}{\sqrt{x_{i-1}}}\right)^2 + \dot{\beta}\sigma_{i-1}^2 \ .$ (5.9)
2. $\sigma_i^2 = \dot{\omega} + \dot{\alpha}\left(\frac{NP_{i-1}}{\sqrt{x_{i-1}}}\right)^2 + \dot{\beta}\sigma_{i-1}^2 + \gamma\frac{1}{x_i} \ .$ (5.10)
3. $\sigma_i^2 = \dot{\omega} + \dot{\alpha}\left(\frac{NP_{i-1}}{\sqrt{x_{i-1}}}\right)^2 + \dot{\beta}\sigma_{i-1}^2 + \gamma_1\frac{1}{x_i} + \gamma_2\frac{x_i}{\psi_i} + \gamma_3\frac{1}{\psi_i} \ .$ (5.11)

The assumption $E\left(NP_i \,\Big|\, \mathcal{G}_{i-1}\right) = 0$ is too restrictive in case micro structure effects are present. Hence, wee denote $E\left(NP_i \,\Big|\, \mathcal{G}_{i-1}\right) = \mu_i$. To remove these micro structure effects an ARMA(p,q) model is proposed to model the net positions. This is represented by

$$NP_i = c + \sum_{j=1}^{p} \xi_j NP_{i-j} + \sum_{j=1}^{q} \theta_j \epsilon_{i-j} + \epsilon_i \ , \tag{5.12}$$

in which ϵ_i are referred to as the demeaned net positions. The demeaned net positions are extracted, and we denote $\epsilon_i = NP_{c,i}$ from now on. Note that we can also write $NP_i = \mu_i + NP_{c,i}$. Hereafter, the UHF-GARCH(r,s) model is applied to the demeaned net positions per unit time,

$$\sigma_i^2 = \dot{\omega} + \sum_{j=1}^{r} \dot{\alpha}_j \left(\frac{NP_{c,i-j}}{\sqrt{x_{i-j}}} \right)^2 + \sum_{j=1}^{s} \dot{\beta}_j \sigma_{i-j}^2 \ , \tag{5.13}$$

$$NP_{c,i} = \sqrt{x_i} \sigma_i \nu_i \ . \tag{5.14}$$

This model is considered to be an extension of the UHF-GARCH(r,s) model, and is referred to as the UHF-ARMA(p,q)-GARCH(r,s) model. The estimation of the UHF-GARCH model is done in two stages. The first stage covers the estimation of the ACD model to capture the durations, assuming the error follows a specific distribution with positive support and expectation equal to one. The second stage covers the estimation of the GARCH model, augmented with the duration as represented by Definition 5.2.

Remark 5.4 (Diurnally adjustments)
It should be noticed that in real world applications, not only the durations but also the net positions could exhibit intraday seasonality (see Chapter 4). Therefore, the deterministic component must be removed from the durations and net positions prior to model estimation. This is done by standardizing the raw durations and the raw net positions. Let $\phi_{x,i}$ *and* $\phi_{NP^2,i}$ *be respectively the deterministic components of the durations and squared net positions. This gives the following result*

$$x_i = \tilde{x}_i \phi_i^x \ , \tag{5.15}$$

$$NP_i = \sqrt{\phi_i^{NP^2}} \left(\mu_i + \tilde{NP}_{c,i} \right) \ . \tag{5.16}$$

5.3.2 Dionne-Duchesne-Pacurar algorithm

Having the ACD model available for the durations and the UHF-GARCH model present to capture the behavior of the net positions, Monte Carlo simulations can be used to obtain the estimator $\widehat{IRM_l(\alpha)}$. In this subsection this Monte Carlo approach and the resulting algorithm to obtain this estimator are elaborated.

Let the last event time of the in-sample period be t_n, and let t_{n+1} be the first event time of the out-of-sample period. Then n is called the forecasting origin. Given that the ACD model and the UHF-GARCH model are estimated for both the diurnally adjusted durations and diurnally adjusted and demeaned net positions, one-step and i-step ahead forecasts can be derived analytically given the information set $\mathcal{F}_n$. The information set $\mathcal{F}_n$ contains all information of the in-sample period. In case the durations follow a Log ACD$_2(p,q)$ model the forecasts are given by

$$\widehat{\kappa}_{n+1} = \widehat{\omega} + \sum_{j=1}^{p} \widehat{\alpha}_j + \sum_{j=1}^{q} \widehat{\beta}_j \ln \tilde{x}_{n+1-j} \ , \tag{5.17}$$

$$\Longrightarrow \widehat{\kappa}_{n+i} = \widehat{\omega} + \sum_{j=1}^{p} \widehat{\alpha}_j + \sum_{j=1}^{i-1} \widehat{\beta}_j \widehat{\kappa}_{n+i-j} + \sum_{j=i}^{q} \widehat{\beta}_j \ln \tilde{x}_{n+i-j} \text{ for } q > i,\ i > 1 \ , \tag{5.18}$$

and in case the net positions follow an UHF-GARCH(r,s) model we derive

$$\widehat{\sigma}^2_{n+1} = \widehat{\dot{\omega}} + \sum_{j=1}^{r} \widehat{\dot{\alpha}}_j \left(\frac{\tilde{NP}_{c,n+1-j}}{\sqrt{x_{n+1-j}}} \right)^2 + \sum_{j=1}^{s} \widehat{\dot{\beta}}_j \sigma^2_{n+1-j} \ , \tag{5.19}$$

$$\Longrightarrow \widehat{\sigma}^2_{n+i} = \widehat{\dot{\omega}} + \sum_{j=1}^{i-1} \left(\widehat{\dot{\alpha}}_j + \widehat{\dot{\beta}}_j \right) \widehat{\sigma}^2_{n+i-j} + \sum_{j=i}^{\max(r,s)} \left(\widehat{\dot{\alpha}}_j + \widehat{\dot{\beta}}_j \right) \sigma^2_{n+i-j} \text{ for } \max(r,s) > i,\ i > 1 \ . \tag{5.20}$$

For the derivations of the one-step and i-step ahead forecasts the reader is referred to Appendices E.9 and E.10. In case another model is preferred for the durations, the one-step and i-step ahead forecasts can be derived in an analogous manner. Having the one-step and i-step forecasts in place, we are able to describe the Monte Carlo algorithm discussed in (Dionne et al., 2009, [DDP09]). The algorithm is described by Algorithm 5.1, and referred to as the Dionne-Duchesne-Pacurar algorithm.

Algorithm 5.1 (Dionne-Duchesne-Pacurar algorithm)

Estimate the parameters of the Log ACD$_2$-GARCH model using diurnally adjusted durations $\left\{ \tilde{x}_i \ : \ i = 1, \ \ldots, \ n \right\}$ *and diurnally adjusted and demeaned net positions* $\left\{ \tilde{NP}_{c,i} \ : \ i = 1, \ \ldots, \ n \right\}$.

for $p = 1$ **to** P **do**

$i = 1$

for $l = 1$ **to** L **do**

(1) Estimate the 1*-step-head forecast*

$$\widehat{x}^{p}_{n+1} = \widehat{\tilde{x}}^{p}_{n+1}\phi^{x}_{n+1} \ , \tag{5.21}$$

where $\widehat{\tilde{x}}^{p}_{n+1} = \exp(\widehat{\kappa}^{p}_{n+1})\epsilon^{p}_{1}$ *and* $\widehat{\kappa}^{p}_{n+1}$ *is obtained by Equation (5.17).*

(2) Estimate the 1*-step-head forecast*

$$\widehat{NP}^{p}_{n+1} = \sqrt{\phi^{NP^2}_{n+1}}\left(\widehat{\mu}^{p}_{n+1} + \widehat{\tilde{N}P}^{p}_{c,n+1}\right) \ , \tag{5.22}$$

where $\widehat{\tilde{N}P}^{p}_{c,n+1} = \sqrt{\widehat{x}^{p}_{n+1}}\widehat{\sigma}^{p}_{n+1}\nu^{p}_{1}$ *and* $\widehat{\sigma}^{p}_{n+1}$ *is obtained by Equation (5.19).*

while $\sum_{i=\tau(0)}^{i} \widehat{x}^{p}_{n+i} \leq lT_{fixed}$ **do**

(i) Estimate the i*-step-head forecast*

$$\widehat{x}^{p}_{n+i} = \widehat{\tilde{x}}^{p}_{n+i}\phi^{x}_{n+i} \ , \tag{5.23}$$

where $\widehat{\tilde{x}}^{p}_{n+i} = \exp(\widehat{\kappa}^{p}_{n+i})\epsilon^{p}_{i}$ *and* $\widehat{\kappa}^{p}_{n+i}$ *is obtained by Equation (5.18).*

(ii) Estimate the i*-step-head forecast*

$$\widehat{NP}^{p}_{n+i} = \sqrt{\phi^{NP^2}_{n+i}}\left(\widehat{\mu}^{p}_{n+i} + \widehat{\tilde{N}P}^{p}_{c,n+i}\right) \ , \tag{5.24}$$

where $\widehat{\tilde{N}P}^{p}_{c,n+i} = \sqrt{\widehat{x}^{p}_{n+i}}\widehat{\sigma}^{p}_{n+i}\nu^{p}_{i}$ *and* $\widehat{\sigma}^{p}_{n+i}$ *is obtained by Equation (5.20).*

(iii) $i = i + 1$

end while

(3) $\tau(l) = i$

(4) Derive the largest negative rescaled net position

$$z_{p,l} = \max\Big\{ - \sum_{i=\tau(l-1)}^{j} \widehat{NP}^{p}_{n+i} \ : \ j = \tau(l-1), \ldots, \tau(l) - 1 \Big\} \ . \tag{5.25}$$

Note that $z_{p,l}$ corresponds to the largest negative rescaled net position of the l-th fixed interval of length T_{fixed} of path p. Hence, $z_{1,1}$ corresponds to the largest negative rescaled net position of the 1-th fixed interval of path 1, $z_{1,2}$ corresponds to the largest negative rescaled net position of the 2-nd fixed interval of path 1, etcetera.

end for

end for

for $l = 1$ **to** L **do**

(1) Estimate, based on the sample $\Big\{z_{p,l} := z_p \ : \ p = 1, \ldots, P\Big\}$, the moment estimators $\widehat{\gamma}$, $\widehat{a}\left(\frac{P}{k}\right)$ and $b\left(\frac{P}{k}\right)$ as stated in Remark 3.4. These moment estimators are given by

$$M_P^{(j)} = \frac{1}{k} \sum_{i=1}^{k} (\log z_{P,P-i+1} - \log z_{P,P-k})^j \tag{5.26}$$

$$\widehat{\gamma} = M_P^{(1)} + 1 - \frac{1}{2} \left(1 - \frac{\left(M_P^{(1)}\right)^2}{M_P^{(2)}} \right)^{-1} \tag{5.27}$$

$$\widehat{a}\left(\frac{P}{k}\right) = z_{P-k,P} M_P^{(1)} \left(1 - \widehat{\gamma} + M_P^{(1)}\right) \tag{5.28}$$

$$\widehat{b}\left(\frac{P}{k}\right) = z_{P-k,P} \ , \tag{5.29}$$

where and $z_{1,P} \leq z_{2,P} \leq \ldots \leq z_{P,P}$ denotes the ordered sample. Moreover, we have $j = 1, \ 2$ and provide that $k = k(P) \to \infty$, $\frac{k}{P} \to 0$ as $P \to \infty$.

(2) Estimate the out-of-sample forecast $\widehat{IRM}_l(\alpha)$ of the l-th fixed interval, given by

$$\widehat{IRM}_l(\alpha) = \widehat{b}\left(\frac{P}{k}\right) + \frac{\widehat{a}\left(\frac{P}{k}\right)}{\widehat{\gamma}} \left(\left(\frac{P\alpha}{k}\right)^{-\widehat{\gamma}} - 1 \right) \ . \tag{5.30}$$

end for

Remark 5.5

It is assumed $\left\{z_{p,l}\ :\ p = 1, \ldots, P\right\}$ *represents a sequence of IID random variables. Hence, the algorithm allows for the application of univariate EVT. The out-of-sample forecast* $\widehat{IRM}_l(\alpha)$ *is derived by combining Theorem 3.1 and Definition 5.1. For the derivation of the forecast, the reader is referred to Appendix E.11. For an extensive overview of univariate EVT, the reader is referred to Chapter 3. For each fixed interval* l*, the parameters* γ *and the normalizing constants* $a\left(\frac{P}{k}\right)$ *and* $b\left(\frac{P}{k}\right)$ *are estimated based on this sequence* $\left\{z_{p,l}\ :\ p = 1, \ldots, P\right\}$*. The visual representation of Algorithm 5.1 is given by Figure 5.2.*

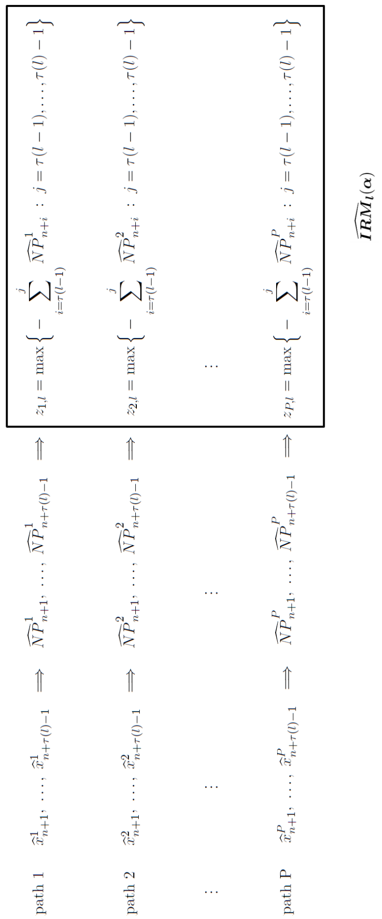

Figure 5.2: Graphic representation of the Dionne-Duchesne-Pacurar algorithm.

5.4 Backtesting procedure

Evaluating the accuracy of the $IRM_l(\alpha)$ forecasts is the next step, which is also referred to as backtesting. An advantage of rescaling the net positions into fixed intervals is that it allows for the evaluation of the accuracy in a traditional way. In this section two traditional evaluation methodologies for measuring $IRM_l(\alpha)$ accuracy are elaborated.

In general, the accuracy of the forecast of an economic variable is assessed by comparing the ex ante forecasted value with the ex post realization (Banulescu et al., 2016, [BCHT16]). However, in case of $IRM_l(\alpha)$ forecasts the ex ante forecasted value cannot be compared with the ex post realization because the true quantile of the distribution is not observable. Therefore, the accuracy of $IRM_l(\alpha)$ forecasts cannot be based on this comparison.

Currently, in the literature two different approaches can be distinguished in backtesting procedures that aim to assess the accuracy of the forecast of an economic variable. The first approach is based on the concept of violation as introduced by Christoffersen (1998). An event is considered to be a violation in case the ex post realization is more negative than the ex ante forecasted value. The main idea of this approach is to test the process of violations against two hypotheses: unconditional coverage (UC) hypothesis and independence (IND) hypothesis. The second approach is based on the concept of probability integral transforms as introduced by Diebold et al. (1998). In this section we will skip the latter approach and focus on the first one.

5.4.1 Framework and testable hypotheses

The starting point of the backtesting procedure is to define an indicator variable. Let $H_l(\alpha)$ be the indicator variable associated with the ex post realization of an $IRM_l(\alpha)$ violation. Hence, the indicator variable $H_l(\alpha)$ is denoted by

$$H_l(\alpha) = \begin{cases} 1 \text{ for } z_l > IRM_l(\alpha) \\ 0 \text{ for } z_l \leq IRM_l(\alpha) \end{cases} . \tag{5.31}$$

This indicator variable is also known as the so called hit variable, and can be more compactly denoted as $H_l(\alpha) = \mathbb{1}_{z_l > IRM_l(\alpha)}$. In case $H_l(\alpha) = 1$, the ex post realization of the largest negative rescaled net position corresponding to the l-th fixed interval is larger than the estimated $IRM_l(\alpha)$. In this case the estimated $IRM_l(\alpha)$ is considered to be violated. On the contrary, in case $H_l(\alpha) = 0$, the ex post realization of the largest negative rescaled net position corresponding to the l-th fixed does not violate the estimated $IRM_l(\alpha)$.

Following the methodology in (Christoffersen, 1998, [Chr98]), the $IRM_l(\alpha)$ forecasts are assumed to be accurate if and only if the associated indicator variable $H_l(\alpha)$ satisfies both the Unconditional Coverage (UC) hypothesis and the Independence (IND) hypothesis. These hypotheses are formally denoted by Definition 5.3 and Definition 5.4. The UC hypothesis states that the probability that an ex post realization of the largest negative rescaled net position z_l exceeding the ex ante $IRM_l(\alpha)$ forecast should be equal to the coverage rate α. The IND hypothesis states that violations for the same coverage rate α observed at two different dates must be independently distributed. Observe that in case $P(H_l(\alpha) = 1) > \alpha$, $IRM_l(\alpha)$ under-estimates the true level of risk. In case $P(H_l(\alpha) = 1) < \alpha$, $IRM_l(\alpha)$ over-estimates the true level risk.

Definition 5.3 (Unconditional Coverage (UC) hypothesis)
Let $H_l(\alpha) = \mathbb{1}_{z_l > IRM_l(\alpha)}$ be a violation indicator variable or hit variable, and consider the event $H_l(\alpha) = 1$ to be a violation. Then the unconditional probability of a violation must be equal to the coverage rate α, i.e.

$$P(H_l(\alpha) = 1) = E(H_l(\alpha)) = \alpha \ . \tag{5.32}$$

Definition 5.4 (Independence (IND) hypothesis)
Let $H_l(\alpha) = \mathbb{1}_{z_l > IRM_l(\alpha)}$ be a violation indicator variable or hit variable, and consider the event $H_l(\alpha) = 1$ to be a violation. Then violations observed at two different dates must be independently distributed, i.e.

$$P\left(H_l(\alpha) = 1 \;\middle|\; \mathcal{F}_{l-1}\right) = P(H_l(\alpha) = 1) \ . \tag{5.33}$$

When both hypothesis are simultaneously valid, the forecasts are said to have a correct conditional coverage. Combining both hypothesis results in the conditional coverage (CC) hypothesis, formally denoted by Definition 5.5. In case the CC hypothesis is satisfied, the demeaned violation indicator variable is said to be a martingale difference sequence. That means, the demeaned violations form a martingale difference sequence with respect to the information set $\mathcal{I}_{l-1}$. This result is demonstrated by

$$\begin{aligned}
E\left(H_l(\alpha) - \alpha \;\middle|\; \mathcal{I}_{l-1}\right) &= E\left(H_l(\alpha) \;\middle|\; \mathcal{I}_{l-1}\right) - \alpha \\
&= P\left(H_l(\alpha) = 1 \;\middle|\; \mathcal{I}_{l-1}\right) - \alpha \\
&= P(H_l(\alpha) = 1) - \alpha \\
&= 0 \ .
\end{aligned}$$

Definition 5.5 (Conditional Coverage (CC) hypothesis)
Let $H_l(\alpha) = \mathbb{1}_{z_l > IRM_l(\alpha)}$ be a violation indicator variable or hit variable, and consider the event $H_l(\alpha) = 1$ to be a violation. Then the conditional probability of a violation must be equal to the coverage rate α, i.e.

$$P\Big(H_l(\alpha) = 1 \;\Big|\; \mathcal{F}_{l-1}\Big) = E\Big(H_l(\alpha) \;\Big|\; \mathcal{F}_{l-1}\Big) = \alpha \ . \tag{5.34}$$

It should be noted that the CC hypothesis can be translated into a distributional assumption. Under the CC hypothesis, each violation indicator variable $H_l(\alpha)$ takes the value 1 with probability α and the value 0 with probability $1 - \alpha$. Therefore, the sequence of violation indicator variables or violation sequence $\{H_l(\alpha)\}$ is assumed to be a random sample from the Bernoulli distribution with a success probability equal to α. This implies that the violation sequence of an accurate $IRM_l(\alpha)$ forecast agrees with a sequence of IID Bernoulli distributed variables. In the same manner, the sequence of the sum of hit variables or the sequence of the number of violations $\Big\{\sum_{l=1}^{L} H_l(\alpha)\Big\}$ is assumed to be a random sample from the binomial distribution with success probability equal to α. This is summarized by Equation (5.35) and (5.36).

$$\text{CC hypothesis} \implies H_l(\alpha) \overset{IID}{\sim} Ber(\alpha) \ . \tag{5.35}$$

$$\text{CC hypothesis} \implies \sum_{l=1}^{L} H_l(\alpha) \overset{IID}{\sim} Bin(L, \alpha) \ . \tag{5.36}$$

5.4.2 Testing unconditional coverage hypothesis

In this section the testing of the UC hypothesis, as stated in Definition 5.3, will be elaborated. Christoffersen (1998) proposed a likelihood ratio test in order to verify the UC hypothesis. This likelihood test models the violation sequence as a sequence of IID random variables distributed according to the Bernoulli distribution with unknown success probability $\pi_1 \in [0, 1]$. This can be translated into

$$H_l(\alpha) \overset{IID}{\sim} Ber(\pi_1) \ . \tag{5.37}$$

The likelihood function for the violation sequence in Equation (5.37) is then given by

$$\ell\Big(\pi_1 \;\Big|\; H_1(\alpha),\ \dots,\ H_L(\alpha)\Big) = \prod_{l=1}^{L} f\Big(H_l(\alpha) \;\Big|\; \pi_1\Big) = \pi_1^{\sum_{l=1}^{L} H_l(\alpha)} (1 - \pi_1)^{L - \sum_{l=1}^{L} H_l(\alpha)} \ . \tag{5.38}$$

Remember that L denotes the number of fixed intervals. Now taking the first derivative of this likelihood function with respect to α and setting this derivative equal to zero, one maximizes the likelihood which results in the maximum likelihood estimator. This maximum likelihood estimator is $\hat{\pi}_1 = \frac{\sum_{l=1}^{L} H_l(\alpha)}{L}$. Theorem 5.1 formally defines the test.

Theorem 5.1 (Christoffersen test for unconditional coverage)
Let $H_l(\alpha) = \mathbb{1}_{z_l > IRM_l(\alpha)}$ be a violation indicator variable or hit variable. Define $H = \sum_{l=1}^{L} H_l(\alpha)$ as the sum of indicator variables, or the number of violations. Assume that the sequence of indicator variables $\{H_l(\alpha) \ : \ l = 1, \ \ldots, L\}$ is independent over time. Furthermore, assume that under the null hypothesis $H_0 \ : \pi_1 = \alpha$. Then, under H_0, the likelihood is given by

$$\ell\left(\pi_1 \;\middle|\; H_1(\alpha), \ \ldots, \ H_L(\alpha)\right) = \alpha^H (1-\alpha)^{L-H} \ , \tag{5.39}$$

which results in the likelihood ratio test statistic

$$LR_{UC} = 2\left(\ln \ell\left(\hat{\pi}_1 \;\middle|\; H_1(\alpha), \ \ldots, \ H_L(\alpha)\right) - \ln \ell\left(\pi_1 \;\middle|\; H_1(\alpha), \ \ldots, \ H_L(\alpha)\right)\right) \overset{a}{\sim} \chi_1^2 \ , \tag{5.40}$$

where

$$\hat{\pi}_1 = \frac{H}{L} \ . \tag{5.41}$$

H_0 is rejected at significance level $(1-\beta)\%$ in case $LR > \chi_1^2(1-\beta)$.

Proof 5.1 (Christoffersen test for unconditional coverage)

We refer to (Christoffersen, 1998, [Chr98]) for the short derivation of the asymptotic distribution of the test statistic.

□

This test is often referred to as the Proportion Of Failures (POF) test. The test only gives information about whether the number of violations agrees with the coverage rate. The number of ones of the violation sequence is taken into account, however the order of the ones of the violation sequence is not taken into account. Therefore, it could be that the number of violations are clustered together in a time-dependent fashion. This will be discussed in the next section.

5.4.3 Testing independence hypothesis

Christoffersen (1998) introduced a likelihood ratio test in order to verify the IND hypothesis. This likelihood test models the violation sequence as a first order Markov chain. The test is based on a Markov chain model which has two states: violation or no violation. This implies that the first order Markov chain can be written as

$$H_l(\alpha) \;\Big|\; H_{l-1}(\alpha) \overset{IID}{\sim} Ber(p_l(\theta)) \ , \tag{5.42}$$

with success probability

$$p_l(\boldsymbol{\theta}) = H_{l-1}(\alpha)\pi_{11} + (1 - H_{l-1}(\alpha))\pi_{01} \ , \tag{5.43}$$

where $\pi_{ij} = P\left(H_l(\alpha) = j \;\Big|\; H_{l-1}(\alpha) = i\right)$ such that $\boldsymbol{\theta} = (\pi_{11}, \pi_{01})' \in [0,1]^2$. Equivalently, this can also be denoted in terms of the transition probability matrix. Let $\boldsymbol{\Pi}$ be the transition probability matrix, denoted by

$$\boldsymbol{\Pi} = \begin{pmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{pmatrix} \ . \tag{5.44}$$

The probability of a violation occurring in the current interval (i.e. $H_l(\alpha) = 1$) given that no violation occurred in the previous interval (i.e. $H_{l-1}(\alpha) = 0$) is given by π_{01}. The probability of a violation occurring in the current interval (i.e. $H_l(\alpha) = 1$) given that a violation also occurred in the previous interval (i.e. $H_{l-1}(\alpha) = 1$) is given by π_{11}. Analogously, the probability of no violation occurring in the current interval conditional on the fact that no violation occurred in the previous interval is given by $1 - \pi_{01}$. The probability of no violation occurring in the current interval conditional on the fact that a violation occurred in the previous interval is given by $1 - \pi_{11}$.

The aim of the likelihood ratio test is to examine whether or not the likelihood of the violation sequence corresponding to the $IRM_l(\alpha)$ forecast depends on whether a violation occurred on the previous interval. In case the $IRM_l(\alpha)$ forecast reflects the amount of risk accurately, the chance of violating the $IRM_l(\alpha)$ forecast in the current interval should not dependent on whether or not the $IRM_l(\alpha)$ was violated in the previous interval. The likelihood for the violation sequence is then given by

$$\ell\left(\boldsymbol{\Pi} \;\Big|\; H_1(\alpha), \ \ldots, \ H_L(\alpha)\right) = p\left(H_1(\alpha), \ \ldots, \ H_L(\alpha) \;\Big|\; \boldsymbol{\Pi}\right) = (1-\pi_{01})^{L_{00}} \, \pi_{01}^{L_{01}} \, (1-\pi_{11})^{L_{10}} \, \pi_{11}^{L_{11}} \ . \tag{5.45}$$

Now taking the first derivatives of this likelihood function (or log likelihood function to simplify the maximization process) with respect to π_{01} and π_{11} and setting these derivatives equal to zero, the maximum likelihood estimators found. Let L_{ij} denote the number of observations of the violation sequence where a j follows an i. Then the maximum likelihood estimators result in respectively and $\hat{\pi}_{01} = \dfrac{L_{01}}{L_{00} + L_{01}}$ and $\hat{\pi}_{11} = \dfrac{L_{11}}{L_{10} + L_{11}}$. The formal test is given by Theorem 5.2.

Theorem 5.2 (Christoffersen test for independence)
Let $H_l(\alpha) = \mathbb{1}_{y_l < -RSIVaR_l(\alpha)}$ be a violation indicator variable or hit variable. Assume that the sequence of indicator variables $\{H_l(\alpha) \; : \; l = 1, \; \ldots, L\}$ is dependent over time and can be represented as a first order Markov chain with two states and transition probability matrix

$$\mathbf{\Pi} = \begin{pmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{pmatrix} , \tag{5.46}$$

where $\pi_{ij} = P\left(H_l(\alpha) = j \;\middle|\; H_{l-1}(\alpha) = i\right)$. Furthermore, assume that under the null hypothesis $H_0 \; : \; \pi_{01} = \pi_{11} = \pi$. Then, under H_0, the likelihood is given by

$$\ell\left(\mathbf{\Pi} \;\middle|\; H_1(\alpha), \; \ldots, \; H_L(\alpha)\right) = (1-\pi)^{L_{00}} \, \pi^{L_{01}} \, (1-\pi)^{L_{10}} \, \pi^{L_{11}} \; , \tag{5.47}$$

which results in the likelihood ratio test statistic

$$LR_{IND} = 2\left(\ln \ell\left(\hat{\mathbf{\Pi}} \;\middle|\; H_1(\alpha), \; \ldots, \; H_L(\alpha)\right) - \ln \ell\left(\mathbf{\Pi} \;\middle|\; H_1(\alpha), \; \ldots, \; H_L(\alpha)\right)\right) \overset{a}{\sim} \chi_1^2 \; , \tag{5.48}$$

where

$$\hat{\mathbf{\Pi}} = \begin{pmatrix} 1-\hat{\pi}_{01} & \hat{\pi}_{01} \\ 1-\hat{\pi}_{11} & \hat{\pi}_{11} \end{pmatrix} , \tag{5.49}$$

$$\hat{\pi}_{01} = \frac{L_{01}}{L_{00} + L_{01}} \; , \tag{5.50}$$

$$\hat{\pi}_{11} = \frac{L_{11}}{L_{10} + L_{11}} \; . \tag{5.51}$$

H_0 is rejected at significance level $(1-\beta)\%$ in case $LR > \chi_1^2(1-\beta)$.

Proof 5.2 (Christoffersen test for independence)

We refer to (Christoffersen, 1998, [Chr98]) for the short derivation of the asymptotic distribution of the test statistic.

□

5.4.4 Testing conditional coverage hypothesis

As stated before, it is also possible to test the UC hypothesis and the IND hypothesis jointly. That means, testing the CC hypothesis at once. In this section two different tests are proposed. We start with the Christoffersen test which is based on a Markov chain model with two states, that is very similar to the test proposed for testing the IND hypothesis. We conclude this chapter with the dynamic quantile (DQ) test, which is based on a linear autoregressive regression.

The test proposed by Christoffersen (1998) to test the UC hypothesis and the IND hypothesis jointly is stated in Theorem 5.3. It should be remarked that this test is rather limited. First, the test is limited to modeling the violation sequence as a first order Markov chain, and therefore is not able to capture dependencies higher than order one. Moreover, the test does not allow to include any other explanatory variables.

Theorem 5.3 (Christoffersen test for conditional coverage)
Let $H_l(\alpha) = \mathbb{1}_{z_l > IRM_l(\alpha)}$ be a violation indicator variable or hit variable. Assume that the sequence of indicator variables $\{H_l(\alpha)\ :\ l = 1,\ \dots, L\}$ is dependent over time and can be represented as a first order Markov chain with two states and transition probability matrix

$$\mathbf{\Pi} = \begin{pmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{pmatrix} , \tag{5.52}$$

where $\pi_{ij} = P\left(H_l(\alpha) = j \;\middle|\; H_{l-1}(\alpha) = i\right)$. Furthermore, assume that under the null hypothesis $H_0\ :\ \pi_{01} = \pi_{11} = \alpha$. Then, under H_0, the likelihood is given by

$$\ell\left(\mathbf{\Pi} \;\middle|\; H_1(\alpha),\ \dots,\ H_L(\alpha)\right) = (1-\alpha)^{L_{00}}\, \alpha^{L_{01}}\, (1-\alpha)^{L_{10}}\, \alpha^{L_{11}} , \tag{5.53}$$

which results in the likelihood ratio test statistic

$$LR_{CC} = 2\left(\ln \ell\left(\hat{\mathbf{\Pi}} \;\middle|\; H_1(\alpha),\ \dots,\ H_L(\alpha)\right) - \ln \ell\left(\mathbf{\Pi} \;\middle|\; H_1(\alpha),\ \dots,\ H_L(\alpha)\right)\right) \overset{a}{\sim} \chi^2_2 , \tag{5.54}$$

where

$$\hat{\mathbf{\Pi}} = \begin{pmatrix} 1 - \hat{\pi}_{01} & \hat{\pi}_{01} \\ 1 - \hat{\pi}_{11} & \hat{\pi}_{11} \end{pmatrix} , \tag{5.55}$$

$$\hat{\pi}_{01} = \frac{L_{01}}{L_{00} + L_{01}} , \tag{5.56}$$

$$\hat{\pi}_{11} = \frac{L_{11}}{L_{10} + L_{11}} . \tag{5.57}$$

H_0 is rejected at significance level $(1-\beta)\%$ in case $LR > \chi^2_2(1-\beta)$.

Proof 5.3 (Christoffersen test for conditional coverage)

We refer to Appendix in (Christoffersen, 1998, [Chr98]) for the proof.

□

Engle et al. (2004) proposed a linear regression based test in order to overcome these two downsides. In order to apply this test, first a sequence of hit variables $Hit_l(\alpha)$ has to be defined. The hit variables $Hit_l(\alpha)$ are defined by

$$Hit_l(\alpha) = H_l(\alpha) - \alpha \ . \tag{5.58}$$

From this equation it is clear that the hit variable takes value $1 - \alpha$ in case the estimated $IRM_l(\alpha)$ is exceeded by the ex post realization of the largest negative rescaled net position corresponding to the l-th fixed interval. On the other hand, the hit variable takes value α in case the estimated $IRM_l(\alpha)$ is not exceeded.

Second a linear regression of the variable $Hit_l(\alpha)$ is applied on the past variables of $Hit_l(\alpha)$ and any other explanatory variables seem to be suited for the application. This linear regression takes the following form

$$Hit_l(\alpha) = \delta + \sum_{k=1}^{K} \beta_k Hit_{l-k}(\alpha) + \sum_{k=1}^{K} \gamma_k g\left(Hit_{l-k}(\alpha), Hit_{l-k-1}(\alpha), \ldots, z_{l-k}, z_{l-k-1}, \ldots\right) + \epsilon_l \ , \tag{5.59}$$

$$\Longrightarrow \mathbf{Hit}(\alpha) = \mathbf{X}\boldsymbol{\Psi} + \boldsymbol{\epsilon} \ . \tag{5.60}$$

Here, it is assumed that the errors ϵ_l are IID random variables. The vector $\boldsymbol{\Psi} = (\delta, \beta_1, \ldots, \beta_K, \gamma_1, \ldots, \gamma_K)' \in \mathbb{R}^{2K+1}$ contains the regression coefficients and the matrix $\mathbf{X} \in \mathbb{R}^{P\times(2K+1)}$ contains the corresponding explanatory variables. Moreover, the function g takes the past Hit_{l-k} variables and z_{l-k} variables as input. The z_{l-k} variables can be any informative variable from the information set Ω_{l-1}. In this manner, testing the CC hypothesis boils down to testing the null hypothesis of the coefficients δ, β_l and γ_l being equal to zero. Note that under the null hypothesis of joint nullity of the coefficients we have

$$E\left(Hit_l(\alpha) \;\middle|\; \mathcal{F}_{l-1}\right) = E\left(\epsilon_l \;\middle|\; \mathcal{F}_{l-1}\right) = 0 \ . \tag{5.61}$$

Then using the definition of the variable $Hit_l(\alpha)$ leads to Equation (5.34), which corresponds with the CC hypothesis. This is shown below. Hence, testing the jointly nullity of the coefficients of the regression corresponds with testing the CC hypothesis. Here, $\beta_l = 0$ and $\gamma_l = 0$ covers the IND hypothesis while $\delta = 0$ covers the UC hypothesis. The test is known as the dynamic quantile test and is given by Theorem 5.4.

$$E\left(Hit_l(\alpha) \,\middle|\, \mathcal{F}_{l-1}\right) = 0 \;, \tag{5.62}$$

$$\Longrightarrow E\left(H_l(\alpha) - \alpha \,\middle|\, \mathcal{F}_{l-1}\right) = 0 \;, \tag{5.63}$$

$$\Longleftrightarrow E\left(H_l(\alpha) \,\middle|\, \mathcal{F}_{l-1}\right) - \alpha = 0 \;, \tag{5.64}$$

$$\Longleftrightarrow E\left(H_l(\alpha) \,\middle|\, \mathcal{F}_{l-1}\right) = \alpha \;, \tag{5.65}$$

$$\Longrightarrow P\left(H_l(\alpha) = 1 \,\middle|\, \mathcal{F}_{l-1}\right) = \alpha \;. \tag{5.66}$$

Theorem 5.4 (Dynamic quantile test)

Let $Hit_l(\alpha) = H_l(\alpha) - \alpha$ be the demeaned process on α associated to $H_l(\alpha)$. Consider the linear regression in Equation (5.60). $\boldsymbol{\Psi} = (\delta, \beta_1, \ldots, \beta_K, \gamma_1, \ldots, \gamma_K)' \in \mathbb{R}^{2K+1}$ represents the vector containing the coefficients included in the regression. $\mathbf{X} \in \mathbb{R}^{P\times(2K+1)}$ denotes the matrix containing the corresponding explanatory variables. Furthermore, assume that under the null hypothesis $H_0 \;:\; \delta = \beta_l = \gamma_l = 0 \;\forall\; l = 1, \;\ldots, K$. Then, under H_0, this results in the Wald statistic

$$DQ_{CC} = \frac{\hat{\boldsymbol{\Psi}}'_{OLS}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\Psi}}_{OLS}}{\alpha(1-\alpha)} \overset{a}{\sim} \chi^2_{2K+1} \;, \tag{5.67}$$

where

$$\hat{\boldsymbol{\Psi}}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Hit} \;. \tag{5.68}$$

H_0 is rejected at significance level $(1-\beta)\%$ in case $LR > \chi^2_{2K+1}(1-\beta)$.

Proof 5.4 (Dynamic quantile test)

We refer to Appendix B in (Engle et al., 2004, [EM04]) for the proof.

□

The question that remains is how the explanatory variables should be chosen. Clearly, this depends on the specific model one is dealing with. In the literature, often the lagged hit variables in combination with the forecast are chosen as explanatory variables of the linear regression. For instance, Engle et al. (2004) use a constant, the first four lagged hit variables and the current VaR forecast, while Dionne et al. (2009) use the first five lagged hit variables and the current IVaR forecast. Then, the regression as denoted in Equation (5.59) translates into respectively Equation (5.69) and (5.70). Note that Engle et al. (2004) used the dynamic quantile test in order to assess the accuracy of their $VaR_t(\alpha)$ forecast in the context of daily returns where t denotes day t (i.e. low frequency). Dionne et al. (2009) used the dynamic quantile test in order to examine the correctness of their $IVaR_l(\alpha)$ forecast in the context of tick-by-tick returns where l denotes interval l during a business day (i.e. high frequency). Key point is that the fact whether we are using low frequency or high frequency returns does not change the way the dynamic quantile test is applied.

$$Hit_t(\alpha) = \mathbb{1}_{r_t < -VaR_t(\alpha)} \Longrightarrow Hit_t(\alpha) = \delta + \sum_{k=1}^{4} \beta_k Hit_{t-k}(\alpha) + \gamma_1 VaR_t(\alpha) \ . \tag{5.69}$$

$$Hit_l(\alpha) = \mathbb{1}_{y_l < -IVaR_l(\alpha)} \Longrightarrow Hit_l(\alpha) = \delta + \sum_{k=1}^{5} \beta_k Hit_{l-k}(\alpha) + \gamma_1 IVaR_l(\alpha) \ . \tag{5.70}$$

Remark 5.6
It should be noticed that throughout this thesis it is recommended to test the UC hyothesis and IND hypothesis separately, instead of testing the CC hypothesis at once. This gives us more information in case the $IRM_l(\alpha)$ estimator turns out to be inaccurate. Only testing the CC hypothesis gives us indeed whether the estimator is accurate or inaccurate, but does not give us the cause. Testing the UC hypothesis and the IND hypothesis separately gives us the reason in case the estimator turns out to be inaccurate: inaccurate coverage, clustered violations or even both.

5.5 Results from application

Author's note: this section is confidential.

5.5.1 Model estimation for durations and net positions

Author's note: this section is confidential.

5.5.1.1 Model estimation durations

Author's note: this section is confidential.

5.5.1.2 Model estimation net positions

Author's note: this section is confidential.

5.5.2 Backtesting results

Author's note: this section is confidential.

5.5.2.1 Backtesting results for group 1

Author's note: this section is confidential.

5.5.2.2 Backtesting results for group 2

Author's note: this section is confidential.

5.5.2.3 Backtesting results for group 3

Author's note: this section is confidential.

6

Multivariate risk metric : probability of failure set

6.1 Introduction

In Chapter 5, we examined the use of ultra high frequency data for liquidity risk management in a univariate framework. The intraday risk measure was introduced and estimated by a Monte Carlo simulation algorithm. The ACD model, as introduced in Chapter 4, was used to define the time steps of the simulation algorithm. The (UHF-)GARCH model was used to generate the corresponding net positions at each time step, rescaled to an appropriate time interval. Hence, this enabled us to estimate the size of the liquidity buffer for a specific time interval within a business day. However, this size was derived for each group separately.

In this chapter the second research question is addressed, and we aim to estimate the size of the liquidity buffer for all groups together. However, extensions to a multivariate framework have proven to be very complex. This complexity arises from the nature of ultra high frequency data, which are by definition not aligned in time (Rengifo et al., 2004, [RH04]). In this chapter, we will make use of the statistical tools from multivariate EVT instead. For an introduction to multivariate EVT, the reader is referred to Chapter 3. In this chapter, we will rely heavily on the results developed in Chapter 3.

The focus of this chapter are the largest daily negative net cumulative positions of the ABN AMRO Bank transaction data set. The specifics of the ABN AMRO Bank transaction data set can be found in Chapter 2. The largest daily negative net cumulative positions for each group d have been recorded during $n = 824$ business days, and are denoted by

$$DM_{k,d} = \max_{N_d((k-1)T)+1 \leq i \leq N_d(kT)} (-NCP_{i,d}) \tag{6.1}$$

for $k = 1, \ldots, n$ and $d = 1, 2, 3$. Note that $DM_{k,d}$ reflects the maximum liquidity need of ABN AMRO Bank within business day k based on group d. We obtain 3 sequences $\Big\{DM_{k,1} \ : \ k = 1, \ldots, n\Big\}$, $\Big\{DM_{k,2} \ : \ k = 1, \ldots, n\Big\}$ and $\Big\{DM_{k,3} \ : \ k = 1, \ldots, n\Big\}$ representing the largest daily negative net cumulative positions of group 1, group 2 and group 3, respectively. It is assumed that $(DM_{1,1}, DM_{1,2}, DM_{1,3})$, $(DM_{2,1}, DM_{2,2}, DM_{2,3})$,. . ., $(DM_{n,1}, DM_{n,2}, DM_{n,3})$ represents a sequence of independent and identically distributed random vectors.

Now, the approach elaborated in (De Haan et al., 1998, [HR98]) (De Haan et al., 2006, [HF06]) is followed closely. Define

$$C := \Big\{ (DM_1, DM_2, DM_3) \ : \ DM_1 + DM_2 + DM_3 > \ell \Big\} , \tag{6.2}$$

which we refer to as the failure set. Additionally, let

$$p_n := P((DM_1, DM_2, DM_3) \in C) \tag{6.3}$$

denote the probability of the failure set. We are interested in those values for ℓ, such that the probability p_n is low (i.e. one could think of $p_n = 0.05$, 0.01, etcetera). In case this probability is low, ℓ reflects the size of the liquidity buffer. By estimating ℓ, we are able to derive the aggregated size of the liquidity buffer of ABN AMRO Bank.

The chapter is organized as follows. In Section 6.2, the mathematical framework is introduced. The probability of the failure set will be derived in Section 6.3. In Section 6.4, an estimator for this probability is explicated. In both sections we start with a general bivariate framework and extend the results to the trivariate framework of the ABN AMRO Bank transaction data set, as defined by Equation (6.3). Section 6.4 presents the results of the application to the ABN AMRO transaction data set. Here, the probability of the failure set is estimated for different values of ℓ.

6.2 Framework

Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be a sequence of independent and identically distributed random vectors with common cdf F. Let F_1 and F_2 denote the marginal distribution functions of X and Y, respectively. Define the block maxima $M_{x,n} = \max(X_1, \ldots, X_n)$ and $M_{y,n} = \max(Y_1, \ldots, Y_n)$. Suppose that there exist sequences of real numbers $a_n,\ c_n > 0$ and $b_n,\ d_n \in \mathbb{R}$ $(n = 1, 2, \ldots)$ and a bivariate distribution function G with non degenerate marginals such that

$$\lim_{n\to\infty} P\left(\frac{M_{x,n} - b_n}{a_n} \leq x, \frac{M_{y,n} - d_n}{c_n} \leq y\right) = \lim_{n\to\infty} F^n(a_n x + b_n, c_n y + d_n) = G(x, y) \quad (6.4)$$

for all continuity points (x, y) of G.

Chapter 3 tells us that we have restrictions on the distribution function G due to the convergence of the marginal distributions of the block maxima $M_{x,n}$ and $M_{y,n}$. These restrictions are denoted by

$$\lim_{n\to\infty} P\left(\frac{M_{x,n} - b_n}{a_n} \leq x\right) = G(x, \infty) = G_{\gamma_1}(x)\ , \quad (6.5)$$

$$\lim_{n\to\infty} P\left(\frac{M_{y,n} - d_n}{c_n} \leq y\right) = G(\infty, y) = G_{\gamma_2}(y)\ . \quad (6.6)$$

Here, the functions G_{γ_1} and G_{γ_2} refer to the univariate GEV distribution as denoted in Equation (3.3). Hence, γ_1 and γ_2 denote the extreme values indices that characterize the heaviness of the tail of the distribution. Based on this bivariate framework, the probability of failure and its estimator will be derived.

6.3 Derivation probability of failure set

The aim of this section is to derive the probability of a failure set for a general bivariate setting, as introduced in Section 6.2. In accordance with Equations (6.2) and (6.3), define the probability of the failure set as

$$p_n = P((X, Y) \in C_n)\ , \quad (6.7)$$

where

$$C_n = \Big\{(x, y)\ :\ x + y > \ell_n\Big\} \quad (6.8)$$

denotes the failure set. Introduce Lemma 6.1, which involves the exponent measure ν as introduced in Remark 3.6. Lemma 6.1 forms the core of our derivation. The main idea is to transform the random variables X and Y in such a way that the probability of the failure set p_n can be approximated by making use of Equation (6.9).

Lemma 6.1

Suppose Equations (6.4), (6.5) and (6.6) hold. Then

$$\lim_{n\to\infty} \frac{n}{k} P\left(\left(\left(1+\gamma_1 \frac{X-U_1\left(\frac{n}{k}\right)}{a_1\left(\frac{n}{k}\right)}\right)^{\frac{1}{\gamma_1}}, \left(1+\gamma_2 \frac{Y-U_2\left(\frac{n}{k}\right)}{a_2\left(\frac{n}{k}\right)}\right)^{\frac{1}{\gamma_2}}\right) \in A_{x,y}\right) = \nu(A_{x,y}) \ , \tag{6.9}$$

where k is an intermediate sequence such that $k = k(n) \to \infty$, $\frac{k}{n} \to 0$ as $n \to \infty$.

Proof 6.1

We refer to Appendix E.12 for the proof.

□

Define the functions f_1 and f_2 such that

$$f_1(x) = \left(1+\gamma_1 \frac{x-b_1\left(\frac{n}{k}\right)}{a_1\left(\frac{n}{k}\right)}\right)^{\frac{1}{\gamma_1}} \ , \tag{6.10}$$

$$f_2(y) = \left(1+\gamma_2 \frac{y-b_2\left(\frac{n}{k}\right)}{a_2\left(\frac{n}{k}\right)}\right)^{\frac{1}{\gamma_2}} \ . \tag{6.11}$$

Mind that the functions f_1 and f_2 do not represent the marginal probability density functions of X and Y, respectively. Using the functions f_1 and f_2 to transform the random variables X and Y, the following relation is derived

$$\begin{aligned}
&(X,Y) \ \in \ C_n \\
\Longrightarrow\ &(X,Y) \ \in \ \Big\{(x,y) \ : \ x+y > \ell_n\Big\} \\
\Longrightarrow\ &(f_1(X), f_2(Y)) \ \in \ \Big\{(f_1(x), f_2(y)) \ : \ x+y > \ell_n\Big\} \\
\Longrightarrow\ &(f_1(X), f_2(Y)) \ \in \ \Big\{(x,y) \ : \ f_1^{\leftarrow}(x) + f_2^{\leftarrow}(y) > \ell_n\Big\} \\
\Longleftrightarrow\ &(f_1(X), f_2(Y)) \ \in \ \Big\{(x,y) \ : \ \frac{a_1\left(\frac{n}{k}\right)}{\gamma_1}(x^{\gamma_1}-1) + b_1\left(\frac{n}{k}\right) + \frac{a_2\left(\frac{n}{k}\right)}{\gamma_2}(y^{\gamma_2}-1) + b_2\left(\frac{n}{k}\right) > \ell_n\Big\} \ .
\end{aligned}$$

Now, define the set

$$Q_n = \left\{(x,y)\ :\ \frac{a_1\left(\frac{n}{k}\right)}{\gamma_1}\left(x^{\gamma_1}-1\right)+b_1\left(\frac{n}{k}\right)+\frac{a_2\left(\frac{n}{k}\right)}{\gamma_2}\left(y^{\gamma_2}-1\right)+b_2\left(\frac{n}{k}\right)>\ell_n\right\}. \quad (6.12)$$

Combining Lemma 6.1 with the above derived relation, an approximation is found for the probability of the failure set p_n. This probability is expressed in terms of the transformed random variables X and Y, i.e. $f_1(X)$ and $f_2(Y)$. Hence, the probability p_n is approximated by

$$\begin{aligned}
p_n &= P\left((X,Y)\ \in\ C_n\right)\\
&= P\left((f_1(X),f_2(Y))\ \in\ Q_n\right)\\
&= P\left(\left(\left(1+\gamma_1\frac{X-b_1\left(\frac{n}{k}\right)}{a_1\left(\frac{n}{k}\right)}\right)^{\frac{1}{\gamma_1}},\left(1+\gamma_2\frac{Y-b_2\left(\frac{n}{k}\right)}{a_2\left(\frac{n}{k}\right)}\right)^{\frac{1}{\gamma_2}}\right)\ \in\ Q_n\right)\\
&\approx \frac{k}{n}\nu(Q_n)\\
&= \frac{k}{n}\nu\left(c_n\frac{Q_n}{c_n}\right)\\
&= \frac{k}{nc_n}\nu\left(\frac{Q_n}{c_n}\right)\\
&= \frac{k}{nc_n}\nu\left(S\right)\ , \qquad (6.13)
\end{aligned}$$

where

$$S = \left\{(x,y)\ :\ \frac{a_1\left(\frac{n}{k}\right)}{\gamma_1}\left((c_nx)^{\gamma_1}-1\right)+b_1\left(\frac{n}{k}\right)+\frac{a_2\left(\frac{n}{k}\right)}{\gamma_2}\left((c_ny)^{\gamma_2}-1\right)+b_2\left(\frac{n}{k}\right)>\ell_n\right\}, \quad (6.14)$$

and c_n denotes a positive sequence.

Remark 6.1

The question remains how to choose an appropriate positive sequence c_n. Two methods are distinguished in (De Haan et al., 2006, [HF06]). The first method considers c_n to be known, and the value of c_n is chosen by the statistician. The second method assumes c_n is unknown, and the estimation of c_n is incorporated in the problem itself. The latter method will be utilized in this chapter. It is assumed that there exists some boundary point of the failure set C_n, (u_n,v_n), such that

$$C_n \subset D_n = \left\{(x,y)\ :\ x \geq u_n \cup y \geq v_n\right\} \quad (6.15)$$

holds for all n. Define for some $r > 0$

$$c_n = \frac{\sqrt{q_n^2 + r_n^2}}{r} \ , \tag{6.16}$$

where

$$q_n = f_1(u_n) = \left(1 + \gamma_1 \frac{u_n - b_1\left(\frac{n}{k}\right)}{a_1\left(\frac{n}{k}\right)}\right)^{\frac{1}{\gamma_1}} \ , \tag{6.17}$$

$$r_n = f_2(v_n) = \left(1 + \gamma_2 \frac{v_n - b_2\left(\frac{n}{k}\right)}{a_2\left(\frac{n}{k}\right)}\right)^{\frac{1}{\gamma_2}} \ . \tag{6.18}$$

Remark 6.2

Note that the point $\left(u_n = \frac{\ell_n}{2}, v_n = \frac{\ell_n}{2}\right)$ could be an appropriate choice, translating Equation (6.15) into

$$C_n \subset D_n = \left\{(x, y) \ : \ x \geq \frac{\ell_n}{2} \cup y \geq \frac{\ell_n}{2}\right\} \ . \tag{6.19}$$

Figure 6.1 illustrates the choice $u_n = v_n = \frac{\ell_n}{2}$, as denoted by Equation (6.19). The black area represents the set C_n, while the yellow area demonstrates the set D_n. The figure confirms that D_n is indeed an subset of C_n.

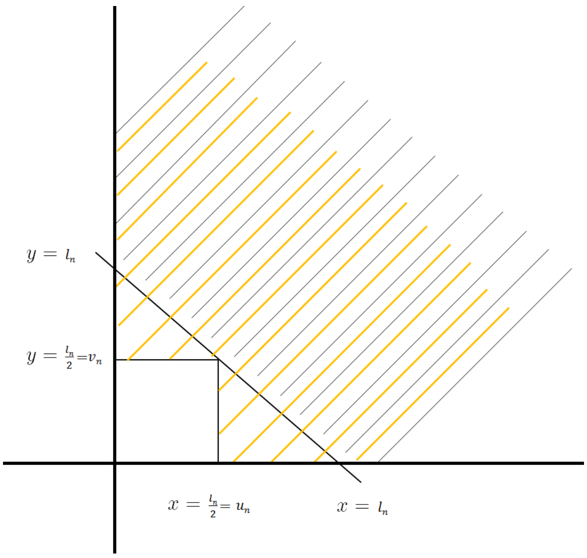

Figure 6.1: Graphic representation of the set $C_n = \left\{(x, y) \ : \ x + y > \ell_n\right\}$ and $D_n = \left\{(x, y) \ : \ x \geq \frac{\ell_n}{2} \cup y \geq \frac{\ell_n}{2}\right\}$.

Remark 6.3 (Extension to trivariate framework)
It is straightforward to extend the probability of the failure set to the trivariate framework, as it is completely analogous to the bivariate framework (De Haan et al., 2006, [HF06]). In the trivariate framework, the probability of interest is defined by

$$p_n = P((X, Y, Z) \in C_n) \ , \tag{6.20}$$

where

$$C_n = \Big\{(x, y, z) \ : \ x + y + z > \ell_n\Big\} \tag{6.21}$$

denotes the failure set. Then, the probability of interest can be approximated by Equation (6.13) such that

$$\begin{aligned} S =& \Big\{(x,y,z) \ : \ \frac{a_1\left(\frac{n}{k}\right)}{\gamma_1}\left((c_n x)^{\gamma_1} - 1\right) + b_1\left(\frac{n}{k}\right) + \frac{a_2\left(\frac{n}{k}\right)}{\gamma_2}\left((c_n y)^{\gamma_2} - 1\right) + b_2\left(\frac{n}{k}\right) + \\ & \frac{a_3\left(\frac{n}{k}\right)}{\gamma_3}\left((c_n z)^{\gamma_3} - 1\right) + b_3\left(\frac{n}{k}\right) > \ell_n\Big\} \ . \end{aligned} \tag{6.22}$$

Moreover, the positive sequence c_n is, for some $r > 0$, defined by

$$c_n = \frac{\sqrt{q_n^2 + r_n^2 + s_n^2}}{r} \ , \tag{6.23}$$

where

$$q_n = f_1(u_n) = \left(1 + \gamma_1 \frac{u_n - b_1\left(\frac{n}{k}\right)}{a_1\left(\frac{n}{k}\right)}\right)^{\frac{1}{\gamma_1}} \ , \tag{6.24}$$

$$r_n = f_2(v_n) = \left(1 + \gamma_2 \frac{v_n - b_2\left(\frac{n}{k}\right)}{a_2\left(\frac{n}{k}\right)}\right)^{\frac{1}{\gamma_2}} \ , \tag{6.25}$$

$$s_n = f_3(w_n) = \left(1 + \gamma_3 \frac{w_n - b_3\left(\frac{n}{k}\right)}{a_3\left(\frac{n}{k}\right)}\right)^{\frac{1}{\gamma_3}} \ , \tag{6.26}$$

such that

$$C_n \subset D_n = \Big\{(x, y, z) \ : \ x \geq u_n \cup y \geq v_n \cup z \geq w_n\Big\} \ . \tag{6.27}$$

Hence, an appropriate choice could be $u_n = v_n = w_n = \dfrac{\ell_n}{3}$.

6.4 Estimation probability of failure set

In Section 6.3 an expression for the probability p_n, as defined by Equation (6.7), was derived. In this section we continue with this expression and aim to find an estimator $\widehat{p}_n$, that is based on the independent and identically distributed random vectors $(X_1, Y_1), (X_2, Y_2), \ldots,$ (X_n, Y_n). Denote this estimator by

$$\widehat{p}_n = \frac{k}{n\widehat{c}_n} \widehat{\nu}\left(\widehat{S}\right) \ . \tag{6.28}$$

Once more, Lemma 6.1 is of great importance and plays a leading role. Lemma 6.1 tells us that the exponent measure of the set S, as defined by Equation (6.14), can be approximated by

$$\nu(S) \approx \frac{n}{k} P\left(\left(\left(1+\gamma_1 \frac{X - U_1\left(\frac{n}{k}\right)}{a_1\left(\frac{n}{k}\right)}\right)^{\frac{1}{\gamma_1}}, \left(1+\gamma_2 \frac{Y - U_2\left(\frac{n}{k}\right)}{a_2\left(\frac{n}{k}\right)}\right)^{\frac{1}{\gamma_2}}\right) \in S\right) \ . \tag{6.29}$$

Hence, by replacing the probability by its empirical counterpart and using the definition of the set S, we obtain

$$\begin{aligned}
\widehat{\nu}(S) &= \frac{n}{k}\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\left(\left(1+\gamma_1 \frac{X_i - U_1\left(\frac{n}{k}\right)}{a_1\left(\frac{n}{k}\right)}\right)^{\frac{1}{\gamma_1}}, \left(1+\gamma_2 \frac{Y_i - U_2\left(\frac{n}{k}\right)}{a_2\left(\frac{n}{k}\right)}\right)^{\frac{1}{\gamma_2}}\right) \in S} \\
&= \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}_{\left(\left(1+\gamma_1 \frac{X_i - U_1\left(\frac{n}{k}\right)}{a_1\left(\frac{n}{k}\right)}\right)^{\frac{1}{\gamma_1}}, \left(1+\gamma_2 \frac{Y_i - U_2\left(\frac{n}{k}\right)}{a_2\left(\frac{n}{k}\right)}\right)^{\frac{1}{\gamma_2}}\right) \in S} \\
&= \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}_{(f_1(X_i), f_2(Y_i)) \in S} \\
&= \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}_{\frac{a_1\left(\frac{n}{k}\right)}{\gamma_1}((c_n f_1(X_i))^{\gamma_1}-1)+b_1\left(\frac{n}{k}\right)+\frac{a_2\left(\frac{n}{k}\right)}{\gamma_2}((c_n f_2(Y_i))^{\gamma_2}-1)+b_2\left(\frac{n}{k}\right)>\ell_n} \\
&= \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}_{\frac{a_1\left(\frac{n}{k}\right)}{\gamma_1}\left((c_n^{\gamma_1}\left(1+\gamma_1 \frac{X_i - b_1\left(\frac{n}{k}\right)}{a_1\left(\frac{n}{k}\right)}\right)-1\right)+b_1\left(\frac{n}{k}\right)+\frac{a_2\left(\frac{n}{k}\right)}{\gamma_2}\left(c_n^{\gamma_2}\left(1+\gamma_2 \frac{Y_i - b_2\left(\frac{n}{k}\right)}{a_2\left(\frac{n}{k}\right)}\right)-1\right)+b_2\left(\frac{n}{k}\right)>\ell_n} \\
&= \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}_{\frac{a_1\left(\frac{n}{k}\right)}{\gamma_1}c_n^{\gamma_1}+c_n^{\gamma_1}\left(X_i - b_1\left(\frac{n}{k}\right)\right)-\frac{a_1\left(\frac{n}{k}\right)}{\gamma_1}+b_1\left(\frac{n}{k}\right)+\frac{a_2\left(\frac{n}{k}\right)}{\gamma_2}c_n^{\gamma_2}+c_n^{\gamma_2}\left(Y_i - b_2\left(\frac{n}{k}\right)\right)-\frac{a_2\left(\frac{n}{k}\right)}{\gamma_2}+b_2\left(\frac{n}{k}\right)>\ell_n} \ .
\end{aligned} \tag{6.30}$$

Now, combining Equations (6.28) and (6.30) and substituting γ_1, γ_2, $a_1\left(\frac{n}{k}\right)$, $a_2\left(\frac{n}{k}\right)$, $b_1\left(\frac{n}{k}\right)$ and $b_2\left(\frac{n}{k}\right)$ by its estimators $\widehat{\gamma}_1$, $\widehat{\gamma}_2$, $\widehat{a_1}\left(\frac{n}{k}\right)$, $\widehat{a}_2\left(\frac{n}{k}\right)$, $\widehat{b}_1\left(\frac{n}{k}\right)$ and $\widehat{b}_2\left(\frac{n}{k}\right)$, yields the final result. Consequently, the estimator $\widehat{p}_n$ is given by

$$\widehat{p}_n = \frac{1}{n\widehat{c}_n}\sum_{i=1}^{n}\mathbb{1}_{\frac{\widehat{a}_1\left(\frac{n}{k}\right)}{\widehat{\gamma}_1}\widehat{c}_n^{\widehat{\gamma}_1}+\widehat{c}_n^{\widehat{\gamma}_1}\left(X_i-\widehat{b}_1\left(\frac{n}{k}\right)\right)-\frac{\widehat{a}_1\left(\frac{n}{k}\right)}{\widehat{\gamma}_1}+\widehat{b}_1\left(\frac{n}{k}\right)+\frac{\widehat{a}_2\left(\frac{n}{k}\right)}{\widehat{\gamma}_2}\widehat{c}_n^{\widehat{\gamma}_2}+\widehat{c}_n^{\widehat{\gamma}_2}\left(Y_i-\widehat{b}_2\left(\frac{n}{k}\right)\right)-\frac{\widehat{a}_2\left(\frac{n}{k}\right)}{\widehat{\gamma}_2}+\widehat{b}_2\left(\frac{n}{k}\right)>\ell_n}. \tag{6.31}$$

Remember that various estimators exist for γ_1, γ_2, $a_1\left(\frac{n}{k}\right)$, $a_2\left(\frac{n}{k}\right)$, $b_1\left(\frac{n}{k}\right)$ and $b_2\left(\frac{n}{k}\right)$. In this thesis, the moment estimators of Remark 3.4 are used. Note that for $M_{n,2}^{(j)}$, $\widehat{\gamma}_2$, $\widehat{a}_2\left(\frac{n}{k}\right)$ and $\widehat{b}_2\left(\frac{n}{k}\right)$, X is replaced by Y in Equations (3.17), (3.18), (3.19) and (3.20).

Remark 6.4
Note that Lemma 6.1 is applied two times in this chapter. In Section 6.3 the lemma is applied to derive an expression for the probability of the failure set, p_n. In this section, the lemma is applied again to find an estimator for the probability of the failure set, $\widehat{p_n}$.

Remark 6.5 (Extension to trivariate framework)
In this thesis, the transaction process within ABN AMRO Bank is divided into three different groups. Hence, we are interested in an estimator for the probability of the failure set in a trivariate context. Fortunately, as discussed in Remark 6.3, this estimator can be extended to the trivariate context. The estimator is given by

$$\widehat{p}_n = \frac{1}{n\widehat{c}_n}\sum_{i=1}^{n}\mathbb{1}_{\frac{\widehat{a}_1\left(\frac{n}{k}\right)}{\widehat{\gamma}_1}\widehat{c}_n^{\widehat{\gamma}_1}+\widehat{c}_n^{\widehat{\gamma}_1}\left(X_i-\widehat{b}_1\left(\frac{n}{k}\right)\right)-\frac{\widehat{a}_1\left(\frac{n}{k}\right)}{\widehat{\gamma}_1}+\widehat{b}_1\left(\frac{n}{k}\right)+\frac{\widehat{a}_2\left(\frac{n}{k}\right)}{\widehat{\gamma}_2}\widehat{c}_n^{\widehat{\gamma}_2}+\widehat{c}_n^{\widehat{\gamma}_2}\left(Y_i-\widehat{b}_2\left(\frac{n}{k}\right)\right)-\frac{\widehat{a}_2\left(\frac{n}{k}\right)}{\widehat{\gamma}_2}+\widehat{b}_2\left(\frac{n}{k}\right)+\frac{\widehat{a}_3\left(\frac{n}{k}\right)}{\widehat{\gamma}_3}\widehat{c}_n^{\widehat{\gamma}_3}+\widehat{c}_n^{\widehat{\gamma}_3}\left(Z_i-\widehat{b}_3\left(\frac{n}{k}\right)\right)-\frac{\widehat{a}_3\left(\frac{n}{k}\right)}{\widehat{\gamma}_3}+\widehat{b}_3\left(\frac{n}{k}\right)>\ell_n} \tag{6.32}$$

6.5 Results from application

This section applies the estimator of the probability of the failure set, as defined by Equation (6.32), to the three dimensional data set of ABN AMRO Bank. The ABN AMRO Bank data set consists of $n = 824$ independent and identically distributed observations of the largest daily negative net cumulative positions of three different groups, which are denoted by

$$DM_{k,d} = \max_{N_d((k-1)T)+1\leq i\leq N_d(kT)}\left(-NCP_{i,d}\right) \tag{6.33}$$

for $k = 1, \ldots, n$ and $d = 1, 2, 3$.

The aim is to estimate the multivariate risk metric: the aggregated liquidity buffer size. As described in Section 6.1 of this chapter, this implies that those values of ℓ are of interest, such that the probability

$$p_n = P\left((DM_1, DM_2, DM_3) \in C\right) \ , \tag{6.34}$$

of the failure set

$$C = \Big\{(DM_1, DM_2, DM_3) \ : \ DM_1 + DM_2 + DM_3 > \ell\Big\} \ , \tag{6.35}$$

is low.

Both the probability p_n and the value ℓ are unknown. In this chapter we proceed as follows. Different values of ℓ will be chosen. For each chosen value of ℓ, the corresponding probability p_n will be estimated. In this manner, we obtain a table that contains values of ℓ with corresponding probability p_n. In this thesis, we aim to find values of ℓ such that the probability $p_n = 0.10,\ 0.05,\ 0.025,\ 0.01$. These probabilities are common practice in risk management. This procedure is summarized by Figure 6.2.

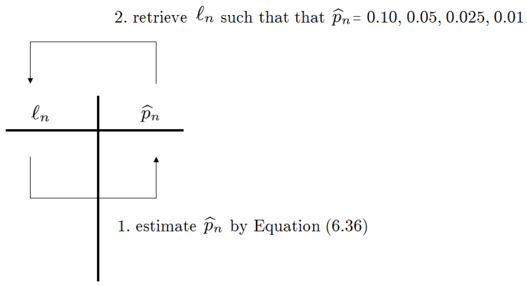

Figure 6.2: Graphic representation of estimation procedure.

6.5.1 Estimation results of marginals

Author's note: this section is confidential.

6.5.2 Estimation results of probability of failure set

Author's note: this section is confidential.

7

Conclusion and recommendations

The aim of this thesis was to propose a framework for intraday liquidity risk management within ABN AMRO Bank, while taking different priorities of transactions into account. For this purpose, two risk metrics were introduced. The univariate risk metric that presented the liquidity buffer size for each priority group separately, and the multivariate risk metric that aggregated the liquidity buffer size while taking the diversification of the priority groups into account. The univariate risk metric provided a granular view, while the multivariate risk metric specified the aggregated level of risk.

An extreme value theory application was proposed to obtain these risk metrics. Hence, the theoretical fundamentals of EVT were elaborated in Chapter 3. Both results from univariate and multivariate EVT were derived and discussed extensively.

Chapter 5 examined the use of ultra high frequency data in combination with results from univariate EVT to obtain the univariate risk metric. The intraday risk measure was introduced. By estimating the intraday risk measure, we were able to obtain the size of the liquidity buffer for a specified time interval within a business day. A Monte Carlo simulation algorithm was proposed to obtain an estimator for this intraday liquidity risk measure. The ACD model, as introduced in Chapter 4, was used to define the time steps of this simulation algorithm.

The performance of the intraday risk measure forecasts was evaluated for each group separately. Based on the frequency of transactions, different interval lengths were considered. For the group with high priority payments, the business day was split up in 3, 4 and 5 intervals. The resulting interval lengths were 13200, 9900 and 7920 seconds. For the group with moderate and low priority payments, the business day was split up in 5, 10 and 15 intervals. Hence, the resulting interval lengths were equal to 7920, 3960 and 2640 seconds.

We forecasted the intraday risk measure 30 days out-of-sample. The main findings are summarized below.

- For the group with high priority payments, the intraday risk measure performed well out-of-sample in case an interval length of 13200 seconds was considered. In case smaller interval lengths were considered, the forecasts only performed well for high values of α: $\alpha = 0.1$ and 0.05. For lower values of α the hypothesis of unconditional coverage was rejected, due to an inaccurate coverage.

- For the group consisting of payments with moderate priority, the intraday risk measure performed satisfactory out-of-sample if an interval length of 7920 seconds was selected. If smaller interval lengths were selected, the forecasts only performed satisfactory for the highest level of α: $\alpha = 0.1$ Again, an inaccurate coverage seemed to be the cause.

- For the group with low priority payments, the forecasts performed adequate in case the interval length of 7920 seconds was chosen. In case smaller interval were chosen, the forecasts only performed well for low choices of α: $\alpha = 0.025$ and 0.01. For higher values of α both the hypotheses of unconditional coverage and independence were rejected. The forecasts had an inaccurate coverage and clustered violations.

Chapter 6 investigated the use of multivariate EVT to obtain the multivariate risk metric.

Author's note: this alinea is confidential.

7.1 Future developments

The following items could be taken into account when considering future developments.

- The diurnal component of the raw durations and net positions is estimated using step functions, in order to avoid negative results. This is a rather simple solution to the problem at hand. Other options for the diurnal component could be considered.

- The Monte Carlo simulation algorithm is very time consuming. The implementation of the algorithm could be improved.

- The aggregated buffer size is estimated based on historical data only. One could investigate the use of the univariate framework to enable future predictions of the aggregated buffer size. In case it is assumed that the different priority groups are completely independent, the estimated models for the durations and net positions could be used to obtain future predictions of the aggregated buffer size. In case this assumption is not made, the dependence between the different priority groups serves as a starting point for further research.

A

Additional figures and tables Chapter 2

A.1 Mapping of heterogenous payment priority groups

Author's note: this section is confidential.

B

Additional figures and tables Chapter 4

B.1 LM tests for specification of conditional mean function

B.1.1 LM test for additive misspecification

Theorem B.1 (LM test for additive misspecification)
Consider the ACD(p, q) model with $\epsilon_i \overset{IID}{\sim} Exp(\lambda)$ *and the duration is additive misspecified, i.e. instead of (4.8) we specify the duration as (4.47). Furthermore, assume that the standard regularity conditions (Engle, 2000, [Eng00]) apply, and that under the null hypothesis* H_0 *the function* ξ_i *satisfies* $\xi_i\left(\mathcal{F}_{i-1}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right) = 0$. *Then, under* H_0, *Equation (4.47) reduces to Equation (4.8) and the LM test statistic*

$$LM = \left(\sum_{i=1}^{n} \hat{c}_i \hat{\mathbf{b}}_{\mathbf{i}}'\right) \times \left(\sum_{i=1}^{n} \hat{\mathbf{b}}_{\mathbf{i}} \hat{\mathbf{b}}_{\mathbf{i}}' - \left(\sum_{i=1}^{n} \hat{\mathbf{b}}_{\mathbf{i}} \hat{\mathbf{a}}_{\mathbf{i}}'\right) \left(\sum_{i=1}^{n} \hat{\mathbf{a}}_{\mathbf{i}} \hat{\mathbf{a}}_{\mathbf{i}}'\right)^{-1} \left(\sum_{i=1}^{n} \hat{\mathbf{a}}_{\mathbf{i}} \hat{\mathbf{b}}_{\mathbf{i}}'\right)\right)^{-1} \times \left(\sum_{i=1}^{n} \hat{c}_i \hat{\mathbf{b}}_{\mathbf{i}}\right) \overset{a}{\sim} \chi^2_{\dim \boldsymbol{\theta_2}}$$

where

$$\mathbf{a}_{\mathbf{i}}(\boldsymbol{\theta_1}) = \frac{1}{\psi_i(\boldsymbol{\theta_1})} \frac{\partial \psi_i(\boldsymbol{\theta_1})}{\partial \boldsymbol{\theta_1}}$$

$$\mathbf{b}_{\mathbf{i}}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) = \frac{1}{\psi_i(\boldsymbol{\theta_1})} \frac{\partial \xi_i(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})}{\partial \boldsymbol{\theta_2}}$$

$$c_i(\boldsymbol{\theta_1}) = \frac{x_i}{\psi_i(\boldsymbol{\theta_1})} - 1$$

H_0 *is rejected at significance level* $(1-\alpha)\%$ *in case* $LM > \chi^2_{\dim \boldsymbol{\theta_2}}(1-\alpha)$.

Proof B.1 (LM test for additive misspecification)

We refer to Appendix A in (Meitz et al., 2006, [MT06]) for the proof.

□

B.1.2 LM test for multiplicative misspecification

Theorem B.2 (LM test for multiplicative misspecification)
Consider the ACD(p, q) model with $\epsilon_i \overset{IID}{\sim} Exp(\lambda)$ the duration is multiplicative misspecified, i.e. instead of (4.8) we specify the durations as (4.48). Furthermore, assume that the standard regularity conditions (Engle, 2000, [Eng00]) apply, and that under the null hypothesis H_0 the function ξ_i satisfies $\xi_i(\mathcal{F}_{i-1}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = 1$. Then, under H_0, Equation (4.48) reduces to Equation (4.8) and the LM test statistic

$$LM = \left(\sum_{i=1}^{n} \hat{\psi}_i \hat{c}_i \hat{\mathbf{b}}_\mathbf{i}'\right) \times \left(\sum_{i=1}^{n} \hat{\psi}_i^2 \hat{\mathbf{b}}_\mathbf{i} \hat{\mathbf{b}}_\mathbf{i}' - \left(\sum_{i=1}^{n} \hat{\psi}_i \hat{\mathbf{b}}_\mathbf{i} \hat{\mathbf{a}}_\mathbf{i}'\right) \left(\sum_{i=1}^{n} \hat{\mathbf{a}}_\mathbf{i} \hat{\mathbf{a}}_\mathbf{i}'\right)^{-1} \left(\sum_{i=1}^{n} \hat{\psi}_i \hat{\mathbf{a}}_\mathbf{i} \hat{\mathbf{b}}_\mathbf{i}'\right)\right)^{-1} \times \left(\sum_{i=1}^{n} \hat{\psi}_i \hat{c}_i \hat{\mathbf{b}}_\mathbf{i}\right) \overset{a}{\sim} \chi^2_{\dim \boldsymbol{\theta_2}}$$

where

$$\begin{aligned}
\mathbf{a}_\mathbf{i}(\boldsymbol{\theta_1}) &= \frac{1}{\psi_i(\boldsymbol{\theta_1})} \frac{\partial \psi_i(\boldsymbol{\theta_1})}{\partial \boldsymbol{\theta_1}} \\
\mathbf{b}_\mathbf{i}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) &= \frac{1}{\psi_i(\boldsymbol{\theta_1})} \frac{\partial \xi_i(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})}{\partial \boldsymbol{\theta_2}} \\
c_i(\boldsymbol{\theta_1}) &= \frac{x_i}{\psi_i(\boldsymbol{\theta_1})} - 1
\end{aligned}$$

H_0 is rejected at significance level $(1-\alpha)\%$ in case $LM > \chi^2_{\dim \boldsymbol{\theta_2}}(1-\alpha)$.

Proof B.2 (LM test for multiplicative misspecification)

We refer to Appendix A in (Meitz et al., 2006, [MT06]) for the proof, which is almost identical.

□

B.2 Descriptive statistics of (diurnally adjusted) durations

Author's note: this section is confidential.

B.2.1 Descriptive statistics of (diurnally adjusted) durations for group 1

Author's note: this section is confidential.

B.2.2 Descriptive statistics of (diurnally adjusted) durations for group 2

Author's note: this section is confidential.

B.2.3 Descriptive statistics of (diurnally adjusted) durations for group 3

Author's note: this section is confidential.

B.3 Estimation results of diurnal components

Author's note: this section is confidential.

B.3.1 Estimation results of diurnal components for group 1

Author's note: this section is confidential.

B.3.2 Estimation results of diurnal components for group 2

Author's note: this section is confidential.

B.3.3 Estimation results of diurnal components for group 3

Author's note: this section is confidential.

B.4 Estimation results of ACD models

B.4.1 Estimation results of ACD models for group 1

	Log $GACD_1$(1,2)		
	estimate	**SE**	**$p-$value**
ω	-0.11012	0.02504	0.00001
α_1	-0.01124	0.00326	0.00057
α_2	-	-	-
β_1	-1.14223	0.21770	0.00000
β_2	-0.40705	0.18863	0.03095
γ	0.13207	0.00405	0.00000
κ	3.74017	0.09990	0.00000
$\sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) < 1$	-1.56052		
log likelihood	-9356.55		
AIC	18725.10		
BIC	18769.13		
MSE	0.8691		
$Q_{LB,1}(15)$	58		
$Q_{LB,2}(15)$	42		
linear dependence test R^2 (Engle et al. (1998))	0.0322		
mean	1.0399		
D_n	0.0433		

Table B.1: Estimation results of fitting a Log $GACD_1(1,2)$ model to the diurnally adjusted durations of group 1 (i.e. estimation sample).

	Log $GACD_2$(2,1)			Log $GACD_2$(2,2)		
	estimate	**SE**	**$p-$value**	**estimate**	**SE**	**$p-$value**
ω	-0.01520	0.00740	0.04009	-0.01697	0.00000	0.00000
α_1	-0.04888	0.00664	0.00000	-0.04492	0.00502	0.00000
α_2	0.04579	0.00676	0.00000	0.03962	0.00523	0.00000
β_1	0.55084	0.01430	0.00000	0.19523	0.08366	0.01964
β_2	-	-	-	0.29300	0.08555	0.00062
γ	0.13200	0.00411	0.00000	0.13251	0.00355	0.00000
κ	3.74764	0.09710	0.00000	3.73192	0.07514	0.00000
$\sum_{j=1}^{q} \beta_j < 1$	0.55084			0.48823		
log likelihood	-9339.48			-9335.32		
AIC	18690.96			18684.65		
BIC	18734.98			18736.01		
MSE	0.8664			0.8647		
$Q_{LB,1}(15)$	65			42		
$Q_{LB,2}(15)$	39			25		
linear dependence test R^2 (Engle et al. (1998))	0.0285			0.0283		
mean	1.0406			1.0427		
D_n	0.0459			0.0466		

Table B.2: Estimation results of fitting a Log $GACD_2(2,1)$ and Log $GACD_2(2,2)$ model to the diurnally adjusted durations of group 1 (i.e. estimation sample).

B.4.2 Estimation results of ACD models for group 2

	GACD(2,1)		
	estimate	**SE**	**p−value**
ω	0.00980	0.00026	0.00000
α_1	0.00202	0.00099	0.04043
α_2	0.10093	0.00151	0.00000
β_1	0.89432	0.00108	0.00000
β_2	-	-	-
γ	5.09275	0.13341	0.00000
κ	0.32594	0.00456	0.00000
$\sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j < 1$	0.99728		
log likelihood	-299308.38		
AIC	598628.76		
BIC	598693.42		
MSE	2.1234		
$Q_{LB,1}(15)$	2933		
$Q_{LB,2}(15)$	122		
linear dependence test R^2 (Engle et al. (1998))	0.0086		
mean	0.9782		
D_n	0.0683		

Table B.3: Estimation results of fitting a GACD$(2, 1)$ model to the diurnally adjusted durations of group 2 (i.e. estimation sample).

	Log $GACD_2$(1,2)			Log $GACD_2$(2,1)		
	estimate	**SE**	**p−value**	**estimate**	**SE**	**p−value**
ω	-0.02353	0.00035	0.00000	-0.06674	0.00059	0.00000
α_1	0.02361	0.00036	0.00000	0.01785	0.00084	0.00000
α_2	-	-	-	0.04921	0.00094	0.00000
β_1	1.67772	0.00517	0.00000	0.97618	0.00037	0.00000
β_2	-0.68668	0.00509	0.00000	-	-	-
γ	4.72563	0.12405	0.00000	4.63026	0.12751	0.00000
κ	0.33865	0.00470	0.00000	0.34192	0.00501	0.00000
$\sum_{j=1}^{q} \beta_j < 1$	0.99104			0.97618		
log likelihood	-299905.82			-300363.05		
AIC	599823.63			600738.10		
BIC	599888.29			600802.76		
MSE	488.9518			1337.5664		
$Q_{LB,1}(15)$	135			294		
$Q_{LB,2}(15)$	19			24		
linear dependence test R^2 (Engle et al. (1998))	0.0105			0.0098		
mean	0.9722			0.9730		
D_n	0.0775			0.0776		

Table B.4: Estimation results of fitting a Log $GACD_2(1,2)$ and Log $GACD_2(2,1)$ model to the diurnally adjusted durations of group 2 (i.e. estimation sample).

B.4.3 Estimation results of ACD models for group 3

	Log $\mathbf{EACD}_2(1,2)$		
	estimate	**SE**	**$p-$value**
ω	-0.12197	0.00111	0.00000
α_1	0.11948	0.00109	0.00000
α_2	-	-	-
β_1	0.86649	0.00977	0.00000
β_2	0.09812	0.00959	0.00000
$\sum_{j=1}^{q} \beta_j < 1$	0.96462		
log likelihood	-467925.21		
AIC	935858.41		
BIC	935902.93		
MSE	1.5158		
$Q_{LB,1}(15)$	226		
$Q_{LB,2}(15)$	171		
linear dependence test R^2 (Engle et al. (1998))	0.0041		
mean	1.0000		
D_n	0.0558		
excess dispersion test (Engle et al. (1998))	73.17		

Table B.5: Estimation results of fitting a Log $\mathrm{EACD}_2(1,2)$ model to the diurnally adjusted durations of group 3 (i.e. estimation sample).

	Log $\text{WACD}_2(1,2)$		
	estimate	**SE**	***p*−value**
ω	-0.12252	0.00128	0.00000
α_1	0.11999	0.00126	0.00000
α_2	-	-	-
β_1	0.86454	0.01102	0.00000
β_2	0.10071	0.01081	0.00000
γ	0.94425	0.00103	0.00000
$\sum_{j=1}^{q} \beta_j < 1$	0.96525		
log likelihood	-466502.12		
AIC	933014.24		
BIC	933069.89		
MSE	1.5357		
$Q_{LB,1}(15)$	243		
$Q_{LB,2}(15)$	168		
linear dependence test R^2 (Engle et al. (1998))	0.0042		
mean	1.0006		
D_n	0.0445		
excess dispersion test (Engle et al. (1998))	23.22		

Table B.6: Estimation results of fitting a Log $\text{WACD}_2(1,2)$ model to the diurnally adjusted durations of group 3 (i.e. estimation sample).

	Log GACD$_2(1,1)$		
	estimate	**SE**	**$p-$value**
ω	-0.11585	0.00067	0.00000
α_1	0.11438	0.00068	0.00000
α_2	-	-	-
β_1	0.96796	0.00056	0.00000
β_2	-	-	-
γ	2.28858	0.02760	0.00000
κ	0.58495	0.00409	0.00000
$\sum_{j=1}^{q} \beta_j < 1$	0.96796		
log likelihood	-463770.85		
AIC	927551.70		
BIC	927607.35		
MSE	1.4904		
$Q_{LB,1}(15)$	399		
$Q_{LB,2}(15)$	158		
linear dependence test R^2 (Engle et al. (1998))	0.0039		
mean	0.9947		
D_n	0.0450		

Table B.7: Estimation results of fitting a Log GACD$_2(1,1)$ model to the diurnally adjusted durations of group 3 (i.e. estimation sample).

C

Additional figures and tables Chapter 5

C.1 Descriptive statistics (diurnally adjusted) net positions

Author's note: this section is confidential.

C.1.1 Descriptive statistics of (diurnally adjusted) net positions for group 1

Author's note: this section is confidential.

C.1.2 Descriptive statistics of (diurnally adjusted) net positions for group 2

Author's note: this section is confidential.

C.1.3 Descriptive statistics of (diurnally adjusted) net positions for group 3

Author's note: this section is confidential.

C.2 AICc values

C.2.1 AICc values for group 1

	p−order			
q−order	**0**	**1**	**2**	**3**
0	31799	31801	31803	31794
1	31801	31803	31791	31782
2	31803	31790	31775	31775
3	31794	31781	31774	31776

Table C.1: AICc values corresponding to the ARMA(p, q) models estimated on the diurnally adjusted net positions of group 1 (i.e. estimation sample). The preferred ARMA(p, q) model, is the one with minimum AICc value: ARMA(2,3).

C.2.2 AICc values for group 2

	p−order					
q−order	**0**	**1**	**2**	**3**	**4**	**5**
0	1239247	1001716	1001565	1001416	1001387	1001389
1	1137520	1001559	1001397	1001385	1001389	1001388

Table C.2: AICc values corresponding to the ARIMA(p, q) models estimated on the diurnally adjusted net positions of group 2 (i.e. estimation sample). The preferred ARIMA(p, q) model, is the one with minimum AICc value: ARIMA(1,1,3).

C.2.3 AICc values for group 3

	p−order					
q−order	**0**	**1**	**2**	**3**	**4**	**5**
0	1428478	1428080	1427921	1427894	1427890	1427890
1	1428066	1427885	1427886	1427862	1427888	1427890
2	1427912	1427886	1427884	1427861	1427861	1427858
3	1427892	1427861	1427861	1427863	1427862	1427850
4	1427889	1427887	1427887	1427863	1427881	1427850
5	1427890	1427889	1427859	1427850	1427850	1427850

Table C.3: AICc values corresponding to the ARMA(p, q) models estimated on the diurnally adjusted net positions of group 3 (i.e. estimation sample). The preferred ARMA(p, q) model, is the one with minimum AICc value: ARMA(3,5), ARMA(4,5), ARMA(5,3), ARMA(5,4) and ARMA(5,5). ARMA(3,5) is selected, due to the number of parameters and significant coefficients.

C.3 Estimation results of AR(I)MA(-EGARCH) models

C.3.1 Estimation results of ARMA model for group 1

	ARMA(2,3)		
	estimate	**SE**	**p−value**
c	0.04880	0.01014	0.00000
ξ_1	1.43614	0.12707	0.00000
ξ_2	-0.68611	0.10320	0.00000
θ_1	-1.43738	0.12672	0.00000
θ_2	0.68097	0.10261	0.00000
θ_3	-0.01584	0.00135	0.00000

Table C.4: Estimation results of fitting an ARMA(2,3) model to the diurnally adjusted net positions of group 1 (i.e. estimation sample).

C.3.2 Estimation results of ARIMA-EGARCH model for group 2

	ARIMA(1,1,3)		
	estimate	**SE**	**p−value**
ξ_1	0.33028	0.02214	0.00000
θ_1	-0.54379	0.11007	0.00000
θ_2	0.32428	0.02160	0.00000
θ_3	-0.01472	0.00063	0.00000

Table C.5: Estimation results of fitting an ARIMA(1,1,3) model to the diurnally adjusted net positions of group 2 (i.e. estimation sample).

	EGARCH(1,1)		
	estimate	**SE**	**p−value**
ω	-0.10522	0.00027	0.00000
α_1	-0.12215	0.00040	0.00000
β_1	0.73030	0.00014	0.00000

Table C.6: Estimation results of fitting an EGARCH(1,1) model to the residuals, obtained after fitting an ARIMA(1,1,3) model to the diurnally adjusted net positions of group 2 (i.e. estimation sample).

C.3.3 Estimation results of ARMA model for group 3

	ARMA(3,5)		
	estimate	**SE**	**$p-$value**
c	0.00066	0.00033	0.04909
ξ_1	0.57176	0.11008	0.00000
ξ_2	-0.44535	0.12527	0.00038
ξ_3	0.69576	0.08185	0.00000
θ_1	-0.54379	0.11007	0.00000
θ_2	0.44746	0.12329	0.00028
θ_3	-0.68544	0.07998	0.00000
θ_4	-0.01183	0.00150	0.00000
θ_5	-0.01119	0.00165	0.00000

Table C.7: Estimation results of fitting an ARMA(3,5) model to the diurnally adjusted net positions of group 1 (i.e. estimation sample).

C.4 Backtesting plots

Author's note: this section is confidential.

C.4.1 Backtesting plots for group 1

Author's note: this section is confidential.

C.4.2 Backtesting plots for group 2

Author's note: this section is confidential.

C.4.3 Backtesting plots for group 3

Author's note: this section is confidential.

D

Additional figures and tables Chapter 6

D.1 Estimation results moment estimators

D.1.1 Estimation results moment estimators for group 1

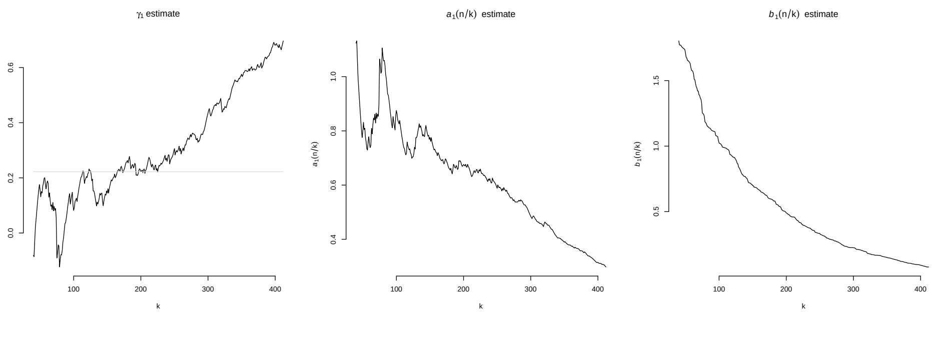

Figure D.1: Estimates of γ_1, $a_1\left(\frac{n}{k}\right)$ and $b_1\left(\frac{n}{k}\right)$ for various values of k (i.e. $10 \leq k \leq 400$), obtained by using the moment estimators given by Equations (3.18), (3.19) and (3.20). The estimates are based on the largest daily negative net cumulative positions of the ABN AMRO Bank transaction data set, see Equation (6.1) for $d = 1$.

D.1.2 Estimation results moment estimators for group 2

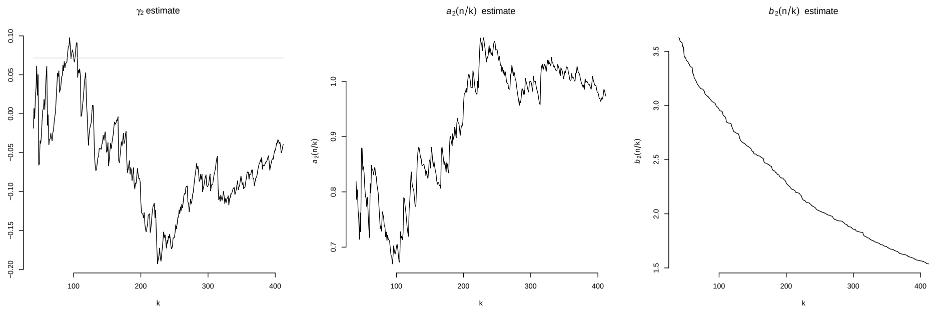

Figure D.2: Estimates of γ_2, $a_2\left(\frac{n}{k}\right)$ and $b_2\left(\frac{n}{k}\right)$ for various values of k (i.e. $10 \leq k \leq 400$), obtained by using the moment estimators given by Equations (3.18), (3.19) and (3.20). The estimates are based on the largest daily negative net cumulative positions of the ABN AMRO Bank transaction data set, see Equation (6.1) for $d = 2$.

D.1.3 Estimation results moment estimators for group 3

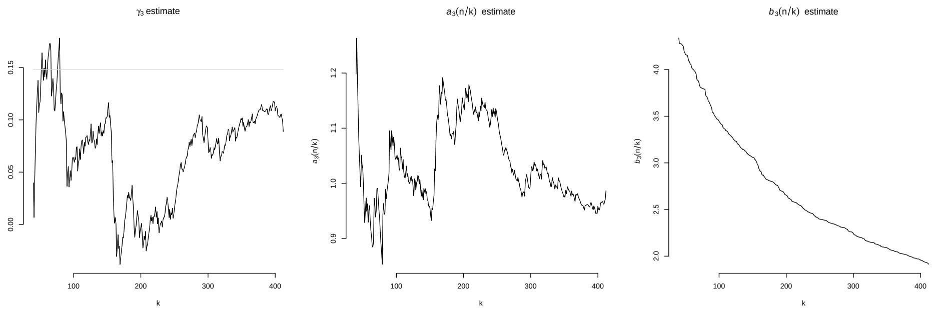

Figure D.3: Estimates of γ_3, $a_3\left(\frac{n}{k}\right)$ and $b_3\left(\frac{n}{k}\right)$ for various values of k (i.e. $10 \leq k \leq 400$), obtained by using the moment estimators given by Equations (3.18), (3.19) and (3.20). The estimates are based on the largest daily negative net cumulative positions of the ABN AMRO Bank transaction data set, see Equation (6.1) for $d = 3$.

D.2 Estimation results probability of failure set

Author's note: this section is confidential.

D.2.1 Estimation results probability of failure set for $\ell_n = 4.0$

Author's note: this section is confidential.

D.2.2 Estimation results probability of failure set for $\ell_n = 4.5$

Author's note: this section is confidential.

D.2.3 Estimation results probability of failure set for $\ell_n = 5.0$

Author's note: this section is confidential.

D.2.4 Estimation results probability of failure set for $\ell_n = 5.5$

Author's note: this section is confidential.

D.2.5 Estimation results probability of failure set for $\ell_n = 6.0$

Author's note: this section is confidential.

D.2.6 Estimation results probability of failure set for $\ell_n = 6.5$

Author's note: this section is confidential.

D.2.7 Estimation results probability of failure set for $\ell_n = 7.0$

Author's note: this section is confidential.

D.2.8 Estimation results probability of failure set for $\ell_n = 7.5$

Author's note: this section is confidential.

D.2.9 Estimation results probability of failure set for $\ell_n = 8.0$

Author's note: this section is confidential.

D.2.10 Estimation results probability of failure set for $\ell_n = 8.5$

Author's note: this section is confidential.

E

Proofs and derivations

E.1 Proof Corollary 3.1

From Equation (3.4) we get by taking logarithms left and right

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(x)$$
$$\iff \lim_{n\to\infty} \log F^n(a_n x + b_n) = \log G_\gamma(x)$$
$$\iff \lim_{n\to\infty} n \log F(a_n x + b_n) = \log G_\gamma(x) \ .$$

This implies that $\lim_{n\to\infty} F(a_n x + b_n) = 1$. Now, use the following inequality

$$t < -\log(1-t) < \frac{t}{1-t} \text{ for } 0 < t < 1$$
$$\iff 1 < \frac{-\log(1-t)}{t} < \frac{1}{1-t} \text{ for } 0 < t < 1$$
$$\implies 1 < \frac{-\log F(a_n x + b_n)}{1 - F(a_n x + b_n)} < \frac{1}{F(a_n x + b_n)} \ .$$

Hence

$$\lim_{n\to\infty} \frac{-\log F(a_n x + b_n)}{1 - F(a_n x + b_n)} = 1 \ .$$

Thus

$$\begin{aligned}
&\lim_{n\to\infty} n \log F(a_n x + b_n) = \log G_\gamma(x) \\
\Longrightarrow\ &\lim_{n\to\infty} -n(1 - F(a_n x + b_n)) = \log G_\gamma(x) \\
\Longleftrightarrow\ &\lim_{n\to\infty} n(1 - F(a_n x + b_n)) = -\log G_\gamma(x)\ .
\end{aligned}$$

□

E.2 Proof Theorem 3.2

From the assumption that Equations (3.22) and (3.23) hold, follows

$$\lim_{n\to\infty} F^n(a_n x + b_n, \infty) = G(x, \infty) = G_{\gamma_1}(x)$$

$$\lim_{n\to\infty} F^n(\infty, c_n x + d_n) = G(\infty, y) = G_{\gamma_2}(y) \ .$$

Making use of item (4) of Theorem 3.1, it follows that there exist positive functions $a_1,\ a_2$ such that for $x,\ y > 0$

$$\lim_{t\to\infty} \frac{U_1(tx) - U_1(t)}{a_1(t)} = \frac{x^{\gamma_1} - 1}{\gamma_1} \ ,$$

$$\lim_{t\to\infty} \frac{U_2(ty) - U_2(y)}{a_2(t)} = \frac{y^{\gamma_2} - 1}{\gamma_2} \ .$$

Now, let

$$g_t(x) = \frac{U_1(tx) - U_1(t)}{a_1(t)} \implies U_1(tx) = a_1(t) g_t(x) + U_1(t) \ ,$$

$$h_t(x) = \frac{U_2(ty) - U_2(t)}{a_2(t)} \implies U_2(ty) = a_2(t) h_t(y) + U_2(t) \ .$$

Hence, by the continuity of G and the monotonicity of F, we conclude

$$\begin{aligned}\lim_{n\to\infty} F^n(U_1(nx), U_2(ny)) &= \lim_{n\to\infty} F^n\left(a_1(n) g_n(x) + U_1(n), a_2(n) h_n(y) + U_2(n)\right) \\ &= G\left(\frac{x^{\gamma_1} - 1}{\gamma_1}, \frac{y^{\gamma_2} - 1}{\gamma_2}\right) \ .\end{aligned}$$

□

E.3 Reformulation Theorem 3.3

$$\begin{aligned}
\lim_{n\to\infty} F^n\left(U_1(nx), U_2(ny)\right) &= \lim_{n\to\infty} F\left(U_1(nx), U_2(ny)\right)^n \\
&= \lim_{n\to\infty} P\left(X \le U_1(nx), Y \le U_2(ny)\right)^n \\
&= \lim_{n\to\infty} P\left(X \le F_1^{\leftarrow}\left(1-\frac{1}{nx}\right), Y \le F_2^{\leftarrow}\left(1-\frac{1}{ny}\right)\right)^n \\
&= \lim_{n\to\infty} P\left(F_1(X) \le 1-\frac{1}{nx}, F_2(Y) \le 1-\frac{1}{ny}\right)^n \\
&= \lim_{n\to\infty} P\left(\frac{1}{1-F_1(X)} \le nx, \frac{1}{1-F_2(Y)} \le ny\right)^n \\
&= \lim_{n\to\infty} P\left(\frac{1}{1-F_1(X_1)} \le nx, \frac{1}{1-F_2(Y_1)} \le ny\right) \dots P\left(\frac{1}{1-F_1(X_n)} \le nx, \frac{1}{1-F_2(Y_n)} \le ny\right) \\
&= \lim_{n\to\infty} P\left(\frac{1}{1-F_1(X_1)} \le nx, \frac{1}{1-F_2(Y_1)} \le ny, \dots, \frac{1}{1-F_1(X_n)} \le nx, \frac{1}{1-F_2(Y_n)} \le ny\right) \\
&= \lim_{n\to\infty} P\left(\tilde{X}_1 \le nx, \tilde{Y}_1 \le ny, \dots, \tilde{X}_n \le nx, \tilde{Y}_n \le ny\right) \\
&= \lim_{n\to\infty} P\left(\max\left(\tilde{X}_1, \dots, \tilde{X}_n\right) \le nx, \max\left(\tilde{Y}_1, \dots, \tilde{Y}_n\right) \le ny\right) \\
&\overset{(*)}{=} \lim_{n\to\infty} P\left(\tilde{M}_{x,n} \le nx, \tilde{M}_{y,n} \le ny\right) \\
&= \lim_{n\to\infty} P\left(\frac{\tilde{M}_{x,n}}{n} \le x, \frac{\tilde{M}_{y,n}}{n} \le y\right) \ .
\end{aligned}$$

where

$$(*) \quad \tilde{M}_{x,n} = \max\left(\tilde{X}_1, \dots, \tilde{X}_n\right) \text{ with } \tilde{X}_i = \frac{1}{1-F_1(X_i)} \text{ for } i = 1, \dots, n \ ,$$

$$\tilde{M}_{y,n} = \max\left(\tilde{Y}_1, \dots, \tilde{Y}_n\right) \text{ with } \tilde{Y}_i = \frac{1}{1-F_2(Y_i)} \text{ for } i = 1, \dots, n \ .$$

E.4 Proof Proposition 3.1

$$
\begin{aligned}
\chi &= \lim_{t\to\infty} P\left(X > U_1(t) \;\middle|\; Y > U_2(t)\right) \\
&= \lim_{t\to\infty} \frac{P\left(X > U_1(t), Y > U_2(t)\right)}{P\left(Y > U_2(t)\right)} \\
&= \lim_{t\to\infty} \frac{P\left(X > U_1(t)\right) + P\left(Y > U_2(t)\right) - P\left(X > U_1(t) \;\cup\; Y > U_2(t)\right)}{P\left(Y > U_2(t)\right)} \\
&= \lim_{t\to\infty} \frac{1 - P\left(X \leq U_1(t)\right) + 1 - P\left(Y \leq U_2(t)\right) - P\left(X > U_1(t) \;\cup\; Y > U_2(t)\right)}{1 - P\left(Y \leq U_2(t)\right)} \\
&= \lim_{t\to\infty} \frac{2 - P\left(X \leq U_1(t)\right) - P\left(Y \leq U_2(t)\right) - P\left(X > U_1(t) \;\cup\; Y > U_2(t)\right)}{1 - P\left(Y \leq U_2(t)\right)} \\
&= \lim_{t\to\infty} \frac{2 - P\left(X \leq F_1^{\leftarrow}\left(1 - \frac{1}{t}\right)\right) - P\left(Y \leq F_2^{\leftarrow}\left(1 - \frac{1}{t}\right)\right) - P\left(X > U_1(t) \;\cup\; Y > U_2(t)\right)}{1 - P\left(Y \leq F_2^{\leftarrow}\left(1 - \frac{1}{t}\right)\right)} \\
&= \lim_{t\to\infty} \frac{2 - F_1\left(F_1^{\leftarrow}\left(1 - \frac{1}{t}\right)\right) - F_2\left(F_2^{\leftarrow}\left(1 - \frac{1}{t}\right)\right) - P\left(X > U_1(t) \;\cup\; Y > U_2(t)\right)}{1 - F_2\left(F_2^{\leftarrow}\left(1 - \frac{1}{t}\right)\right)} \\
&= \lim_{t\to\infty} \frac{2 - \left(1 - \frac{1}{t}\right) - \left(1 - \frac{1}{t}\right) - P\left(X > U_1(t) \;\cup\; Y > U_2(t)\right)}{1 - \left(1 - \frac{1}{t}\right)} \\
&= \lim_{t\to\infty} \frac{2 - 1 + \frac{1}{t} - 1 + \frac{1}{t} - P\left(X > U_1(t) \;\cup\; Y > U_2(t)\right)}{1 - 1 + \frac{1}{t}} \\
&= \lim_{t\to\infty} \frac{\frac{2}{t} - P\left(X > U_1(t) \;\cup\; Y > U_2(t)\right)}{\frac{1}{t}} \\
&= \lim_{t\to\infty} t\left(\frac{2}{t} - P\left(X > U_1(t) \;\cup\; Y > U_2(t)\right)\right) \\
&= \lim_{t\to\infty} 2 - tP\left(X > U_1(t) \;\cup\; Y > U_2(t)\right) \\
&= \lim_{t\to\infty} 2 - t\left(1 - P\left(X \leq U_1(t) \;\cup\; Y \leq U_2(t)\right)\right) \\
&= \lim_{t\to\infty} 2 - t\left(1 - F\left(U_1(t), U_2(t)\right)\right) \\
&\overset{(*)}{=} 2 + \log G_0(1,1) \\
&\overset{(**)}{=} 2 - \log e^{-2} \\
&= 2 - 2 \\
&= 0\ ,
\end{aligned}
$$

where

(∗) Use Corollary 3.2 .

$$
\begin{aligned}
(**) \quad G_0(x,y) &= G\left(\frac{x^{\gamma_1}-1}{\gamma_1}, \frac{y^{\gamma_2}-1}{\gamma_2}\right) \\
&= G\left(\frac{x^{\gamma_1}-1}{\gamma_1}\right) G\left(\frac{y^{\gamma_2}-1}{\gamma_2}\right) \\
&= e^{-\left(+\gamma_1\left(\frac{x^{\gamma_1}-1}{\gamma_1}\right)\right)^{-\frac{1}{\gamma_1}}} e^{-\left(+\gamma_2\left(\frac{y^{\gamma_2}-1}{\gamma_2}\right)\right)^{-\frac{1}{\gamma_2}}} \\
&= e^{-(1-x^{\gamma_1}-1)^{-\frac{1}{\gamma_1}}} e^{-(1-y^{\gamma_2}-1)^{-\frac{1}{\gamma_2}}} \\
&= e^{-(x^{\gamma_1})^{-\frac{1}{\gamma_1}}} e^{-(y^{\gamma_2})^{-\frac{1}{\gamma_2}}} \\
&= e^{-x^{-1}} e^{-y^{-1}} \; .
\end{aligned}
$$

□

E.5 Derivation Remark 4.2

$$
\begin{aligned}
& \psi_i = \omega + \sum_{j=1}^{p} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} \\
\Longleftrightarrow\ & \psi_i + x_i = \omega + \sum_{j=1}^{p} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} + x_i \\
\Longleftrightarrow\ & x_i = \omega + \sum_{j=1}^{p} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} + x_i - \psi_i \\
\Longleftrightarrow\ & x_i = \omega + \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) x_{i-j} - \sum_{j=1}^{\max(p,q)} \beta_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} + x_i - \psi_i \\
\overset{(*)}{\Longleftrightarrow}\ & x_i = \omega + \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) x_{i-j} - \sum_{j=1}^{q} \beta_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} + x_i - \psi_i \\
\Longleftrightarrow\ & x_i = \omega + \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) x_{i-j} - \sum_{j=1}^{q} (\beta_j x_{i-j} - \beta_j \psi_{i-j}) + x_i - \psi_i \\
\Longleftrightarrow\ & x_i = \omega + \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) x_{i-j} - \sum_{j=1}^{q} \beta_j \left(x_{i-j} - \psi_{i-j} \right) + x_i - \psi_i \\
\overset{(**)}{\Longrightarrow}\ & x_i = \omega + \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) x_{i-j} - \sum_{j=1}^{q} \beta_j \eta_{i-j} + \eta_i \ ,
\end{aligned}
$$

where

$$
\begin{aligned}
&(*) \quad \beta_j = 0 \ \forall \ j > q \ . \\
&(**) \ \eta_i = x_i - \psi_i \ .
\end{aligned}
$$

□

E.6 Derivation Remark 4.3

Assume x_i is covariance stationary and let $E(x_i) = \mu$, then we can derive the following

$$
\begin{aligned}
&x_i = \omega + \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) x_{i-j} - \sum_{j=1}^{q} \beta_j \eta_{i-j} + \eta_i \\
\iff\ & x_i - \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) x_{i-j} = \omega - \sum_{j=1}^{q} \beta_j \eta_{i-j} + \eta_i \\
\implies\ & E\left(x_i - \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) x_{i-j} \right) = E\left(\omega - \sum_{j=1}^{q} \beta_j \eta_{i-j} + \eta_i \right) \\
\implies\ & E\left(x_i\right) - \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) E\left(x_{i-j}\right) = \omega - \sum_{j=1}^{q} \beta_j E\left(\eta_{i-j}\right) + E\left(\eta_i\right) \\
\implies\ & \mu - \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) \mu = \omega - \sum_{j=1}^{q} \beta_j E\left(\eta_{i-j}\right) + E\left(\eta_i\right) \\
\implies\ & \mu - \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) \mu = \omega - \sum_{j=1}^{q} \beta_j E\left(x_{i-j} - \psi_{i-j}\right) + E\left(x_i - \psi_i\right) \\
\implies\ & \mu - \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) \mu = \omega - \sum_{j=1}^{q} \beta_j E\left(E\left(x_{i-j} - \psi_{i-j} \;\middle|\; \mathcal{F}_{i-1} \right)\right) + E\left(E\left(x_i - \psi_i \;\middle|\; \mathcal{F}_{i-1} \right)\right) \\
\implies\ & \mu - \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) \mu = \omega - \sum_{j=1}^{q} \beta_j E\left(\psi_{i-j} - \psi_{i-j}\right) + E\left(\psi_i - \psi_i\right) \\
\iff\ & \mu - \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) \mu = \omega \\
\iff\ & \mu \left(1 - \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) \right) = \omega \\
\iff\ & \mu = \frac{\omega}{1 - \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j)} \\
\iff\ & \mu = \frac{\omega}{1 - \sum_{j=1}^{\max(p,q)} \alpha_j - \sum_{j=1}^{\max(p,q)} \beta_j} \\
\overset{(*)}{\iff}\ & \mu = \frac{\omega}{1 - \sum_{j=1}^{p} \alpha_j - \sum_{j=1}^{q} \beta_j} ,
\end{aligned}
$$

where

$$(*)\quad \alpha_j = 0 \ \forall\ j > p,\ \beta_j = 0 \ \forall\ j > q\ .$$

□

E.7 Derivation Remark 4.4

First it should be noted that

$$\begin{aligned}
& x_i = e^{\kappa_i}\epsilon_i \\
\iff\ & \ln x_i = \ln e^{\kappa_i}\epsilon_i \\
\iff\ & \ln x_i = \ln e^{\kappa_i} + \ln \epsilon_i \\
\iff\ & \ln x_i = \kappa_i + \ln \epsilon_i \\
\iff\ & \ln x_i - \kappa_i = \ln \epsilon_i \ .
\end{aligned}$$

Then we can derive

$$\begin{aligned}
& \kappa_i = \bar{\omega} + \sum_{j=1}^{p} \alpha_j \ln \epsilon_{i-j} + \sum_{j=1}^{q} \beta_j \kappa_{i-j} \\
\iff\ & \kappa_i = \bar{\omega} + \sum_{j=1}^{p} \alpha_j (\ln x_{i-j} - \kappa_i) + \sum_{j=1}^{q} \beta_j \kappa_{i-j} \\
\iff\ & \kappa_i = \bar{\omega} + \sum_{j=1}^{p} \alpha_j \ln x_{i-j} + \sum_{j=1}^{\max(p,q)} (\beta_j - \alpha_j)\kappa_{i-j} \\
\overset{(*)}{\implies}\ & \kappa_i = \bar{\omega} + \sum_{j=1}^{p} \alpha_j \ln x_{i-j} + \sum_{j=1}^{\max(p,q)} \gamma_j \kappa_{i-j} \\
\iff\ & \kappa_i + \ln x_i = \bar{\omega} + \sum_{j=1}^{p} \alpha_j \ln x_{i-j} + \sum_{j=1}^{\max(p,q)} \gamma_j \kappa_{i-j} + \ln x_i \\
\iff\ & \ln x_i = \bar{\omega} + \sum_{j=1}^{p} \alpha_j \ln x_{i-j} + \sum_{j=1}^{\max(p,q)} \gamma_j \kappa_{i-j} + \ln x_i - \kappa_i \\
\iff\ & \ln x_i = \bar{\omega} + \sum_{j=1}^{\max(p,q)} (\alpha_j + \gamma_j) \ln x_{i-j} - \sum_{j=1}^{\max(p,q)} \gamma_j \ln x_{i-j} + \sum_{j=1}^{\max(p,q)} \gamma_j \kappa_{i-j} + \ln x_i - \kappa_i \\
\iff\ & \ln x_i = \bar{\omega} + \sum_{j=1}^{\max(p,q)} (\alpha_j + \gamma_j) \ln x_{i-j} - \sum_{j=1}^{\max(p,q)} (\gamma_j \ln x_{i-j} + \gamma_j \kappa_{i-j}) + \ln x_i - \kappa_i \\
\iff\ & \ln x_i = \bar{\omega} + \sum_{j=1}^{\max(p,q)} (\alpha_j + \gamma_j) \ln x_{i-j} - \sum_{j=1}^{\max(p,q)} \gamma_j (\ln x_{i-j} + \kappa_{i-j}) + \ln x_i - \kappa_i \\
\overset{(**)}{\implies}\ & \ln x_i = \bar{\omega} + \sum_{j=1}^{\max(p,q)} \delta_j \ln x_{i-j} + \sum_{j=1}^{\max(p,q)} \theta_j \ln \epsilon_{i-j} + \ln \epsilon_i \\
\iff\ & \ln x_i = \bar{\omega} + \sum_{j=1}^{\max(p,q)} \delta_j \ln x_{i-j} + \sum_{j=1}^{\max(p,q)} \theta_j (\ln \epsilon_{i-j} - E(\ln \epsilon_{i-j})) + \sum_{j=1}^{\max(p,q)} \theta_j E(\ln \epsilon_{i-j}) + \ln \epsilon_i - E(\ln \epsilon_i) + E(\ln \epsilon_i) \\
\iff\ & \ln x_i = \bar{\omega} + \sum_{j=1}^{\max(p,q)} \theta_j E(\ln \epsilon_{i-j}) + E(\ln \epsilon_i) + \sum_{j=1}^{\max(p,q)} \delta_j \ln x_{i-j} + \sum_{j=1}^{\max(p,q)} \theta_j (\ln \epsilon_{i-j} - E(\ln \epsilon_{i-j})) + \ln \epsilon_i - E(\ln \epsilon_i) \\
\overset{(***)}{\implies}\ & \ln x_i = \tilde{\omega} + \sum_{j=1}^{\max(p,q)} \delta_j \ln x_{i-j} + \sum_{j=1}^{\max(p,q)} \theta_j \nu_{i-j} + \nu_i \ ,
\end{aligned}$$

where

$$(*) \quad \gamma_j = \beta_j - \alpha_j \ .$$

$$(**) \quad \delta_j = \alpha_j + \gamma_j, \ \theta_j = -\gamma_j \ .$$

$$(***) \ \tilde{\omega} = \bar{\omega} + \sum_{j=1}^{\max(p,q)} \theta_j E(\ln \epsilon_i) + E(\ln \epsilon_i), \ \nu_i = \ln \epsilon_i - E(\ln \epsilon_i) \ .$$

□

E.8 Proof Remark 4.5

Define the inverse of G_0 by $G_0^{-1}(z) = \min\{y \ : G_0(y) \leq z\}$.

Using the substitution $G_0^{-1}(z) = y$ yields

$$
\begin{aligned}
D_n &= \sup_{y\in\mathbb{R}} \left|\hat{G}_n(y) - G_0(y)\right| \\
&= \sup_{z\in[0,1]} \left|\hat{G}_n(G_0^{-1}(z)) - G_0(G_0^{-1}(z))\right| \\
&= \sup_{z\in[0,1]} \left|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{Y_i \leq G_0^{-1}(z)} - z\right| \\
&= \sup_{z\in[0,1]} \left|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{G_0(Y_i) \leq z} - z\right| .
\end{aligned}
$$

Additionally, note that for $z \in [0, 1]$

$$
\begin{aligned}
P(G_0(Y_i) \leq z) &= P(Y_i \leq G_0^{-1}(z)) \\
&= G_0(G_0^{-1}(z)) \\
&= z \ .
\end{aligned}
$$

Hence, the sequence of $G_0(Y_i) \overset{IID}{\sim} U(0, 1)$ for $i \leq n$.

Using the substitution $G_0(Y_i) = U_i$ for $i \leq n$ yields

$$
D_n = \sup_{z\in[0,1]} \left|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{U_i \leq z} - z\right| ,
$$

which concludes the proof.

□

E.9 Derivation one-step and i-step ahead forecasts of durations

$$\begin{aligned}
\widehat{\kappa}_{n+1} &= E\left(\kappa_{n+1} \;\middle|\; \mathcal{F}_n\right) \\
&= E\left(\omega + \sum_{j=1}^{p} \alpha_j \epsilon_{n+1-j} + \sum_{j=1}^{q} \beta_j \kappa_{n+1-j} \middle| \; \mathcal{F}_n\right) \\
&= \omega + \sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j \ln \tilde{x}_{n+1-j}
\end{aligned}$$

$$\begin{aligned}
\widehat{\kappa}_{n+2} &= E\left(\kappa_{n+2} \;\middle|\; \mathcal{F}_n\right) \\
&= E\left(\omega + \sum_{j=1}^{p} \alpha_j \epsilon_{n+2-j} + \sum_{j=1}^{q} \beta_j \kappa_{n+2-j} \middle| \; \mathcal{F}_n\right) \\
&= \omega + \sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j E\left(\phi_{n+2-j} \middle| \; \mathcal{F}_n\right) \\
&= \omega + \sum_{j=1}^{p} \alpha_j + \beta_1 E\left(\kappa_{n+1} \middle| \; \mathcal{F}_n\right) + \sum_{j=2}^{q} \beta_j E\left(\kappa_{n+2-j} \middle| \; \mathcal{F}_n\right) \text{ for } q > 2,\ i > 1 \\
&= \omega + \sum_{j=1}^{p} \alpha_j + \beta_1 \widehat{\kappa}_{n+1} + \sum_{j=2}^{q} \beta_j \ln \tilde{x}_{n+2-j} \text{ for } q > 2
\end{aligned}$$

$$\begin{aligned}
\widehat{\kappa}_{n+i} &= E\left(\kappa_{n+i} \;\middle|\; \mathcal{F}_n\right) \\
&= E\left(\omega + \sum_{j=1}^{p} \alpha_j \epsilon_{n+i-j} + \sum_{j=1}^{q} \beta_j \kappa_{n+i-j} \middle| \; \mathcal{F}_n\right) \\
&= \omega + \sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j E\left(\kappa_{n+i-j} \middle| \; \mathcal{F}_n\right) \\
&= \omega + \sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{i-1} \beta_j E\left(\kappa_{n+i-j} \middle| \; \mathcal{F}_n\right) + \sum_{j=i}^{q} \beta_j E\left(\kappa_{n+i-j} \middle| \; \mathcal{F}_n\right) \text{ for } q > i \\
&= \omega + \sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{i-1} \beta_j \widehat{\kappa}_{n+i-j} + \sum_{j=i}^{q} \beta_j \ln \tilde{x}_{n+i-j} \text{ for } q > i,\ i > 1
\end{aligned}$$

E.10 Derivation one-step and i-step ahead forecasts of net positions

$$\begin{aligned}
\widehat{\sigma}^2_{n+1} &= E\left(\sigma^2_{n+1} \;\middle|\; \mathcal{F}_n\right) \\
&= E\left(\dot{\omega} + \sum_{j=1}^{r} \dot{\alpha}_j \left(\frac{\tilde{NP}_{c,n+1-j}}{\sqrt{x_{n+1-j}}}\right)^2 + \sum_{j=1}^{s} \dot{\beta}_j \sigma^2_{n+1-j} \middle| \mathcal{F}_n\right) \\
&= \dot{\omega} + \sum_{j=1}^{r} \dot{\alpha}_j \left(\frac{\tilde{NP}_{c,n+1-j}}{\sqrt{x_{n+1-j}}}\right)^2 + \sum_{j=1}^{s} \dot{\beta}_j \sigma^2_{n+1-j}
\end{aligned}$$

$$\begin{aligned}
\widehat{\sigma}^2_{n+2} &= E\left(\sigma^2_{n+2} \;\middle|\; \mathcal{F}_n\right) \\
&= E\left(\dot{\omega} + \sum_{j=1}^{r} \dot{\alpha}_j \left(\frac{\tilde{NP}_{c,n+2-j}}{\sqrt{x_{n+2-j}}}\right)^2 + \sum_{j=1}^{s} \dot{\beta}_j \sigma^2_{n+2j} \;\middle|\; \mathcal{F}_n\right) \\
&= E\left(\dot{\omega} + \sum_{j=1}^{r} \dot{\alpha}_j \left(\frac{\sqrt{x_{n+2-j}}\sigma_{n+2-j}\nu_{n+2-j}}{\sqrt{x_{n+2-j}}}\right)^2 + \sum_{j=1}^{s} \dot{\beta}_j \sigma^2_{n+2-j} \;\middle|\; \mathcal{F}_n\right) \\
&= E\left(\dot{\omega} + \sum_{j=1}^{r} \dot{\alpha}_j \sigma^2_{n+2-j}\nu^2_{n+2-j} + \sum_{j=1}^{s} \dot{\beta}_j \sigma^2_{n+2-j} \;\middle|\; \mathcal{F}_n\right) \\
&= E\left(\dot{\omega} + \sum_{j=1}^{\max(r,s)} \left(\dot{\alpha}_j \nu^2_{n+2-j} + \dot{\beta}_j\right) \sigma^2_{n+2-j} \;\middle|\; \mathcal{F}_n\right) \\
&= \dot{\omega} + \sum_{j=1}^{\max(r,s)} \left(\dot{\alpha}_j + \dot{\beta}_j\right) E\left(\sigma^2_{n+2-j} \;\middle|\; \mathcal{F}_n\right) \\
&= \dot{\omega} + \left(\dot{\alpha}_1 + \dot{\beta}_1\right) E\left(\sigma^2_{n+1} \;\middle|\; \mathcal{F}_n\right) + \sum_{j=2}^{\max(r,s)} \left(\dot{\alpha}_j + \dot{\beta}_j\right) E\left(\sigma^2_{n+2-j} \;\middle|\; \mathcal{F}_n\right) \text{ for } \max(r,s) > 2 \\
&= \dot{\omega} + \left(\dot{\alpha}_1 + \dot{\beta}_1\right) \widehat{\sigma}^2_{n+1} + \sum_{j=2}^{\max(r,s)} \left(\dot{\alpha}_j + \dot{\beta}_j\right) E\left(\sigma^2_{n+2-j} \;\middle|\; \mathcal{F}_n\right) \text{ for } \max(r,s) > 2
\end{aligned}$$

$$\begin{aligned}
\widehat{\sigma}^2_{n+i} &= E\left(\sigma^2_{n+i} \;\middle|\; \mathcal{F}_n\right) \\
&= E\left(\dot{\omega} + \sum_{j=1}^{r} \dot{\alpha}_j \left(\frac{\tilde{NP}_{c,n+i-j}}{\sqrt{x_{n+i-j}}}\right)^2 + \sum_{j=1}^{s} \dot{\beta}_j \sigma^2_{n+i-j} \;\middle|\; \mathcal{F}_n\right) \\
&= E\left(\dot{\omega} + \sum_{j=1}^{r} \dot{\alpha}_j \left(\frac{\sqrt{x_{n+i-j}}\sigma_{n+i-j}\nu_{n+i-j}}{\sqrt{x_{n+i-j}}}\right)^2 + \sum_{j=1}^{s} \dot{\beta}_j \sigma^2_{n+i-j} \;\middle|\; \mathcal{F}_n\right) \\
&= E\left(\dot{\omega} + \sum_{j=1}^{r} \dot{\alpha}_j \sigma^2_{n+i-j}\nu^2_{n+i-j} + \sum_{j=1}^{s} \dot{\beta}_j \sigma^2_{n+i-j} \;\middle|\; \mathcal{F}_n\right) \\
&= E\left(\dot{\omega} + \sum_{j=1}^{\max(r,s)} \left(\dot{\alpha}_j \nu^2_{n+i-j} + \dot{\beta}_j\right) \sigma^2_{n+i-j} \;\middle|\; \mathcal{F}_n\right) \\
&= \dot{\omega} + \sum_{j=1}^{\max(r,s)} \left(\dot{\alpha}_j + \dot{\beta}_j\right) E\left(\sigma^2_{n+i-j} \;\middle|\; \mathcal{F}_n\right)
\end{aligned}$$

$$= \dot{\omega} + \sum_{j=1}^{i-1} \left(\dot{\alpha}_j + \dot{\beta}_j\right) E\left(\sigma^2_{n+i-j} \;\middle|\; \mathcal{F}_n\right) + \sum_{j=i}^{\max(r,s)} \left(\dot{\alpha}_j + \dot{\beta}_j\right) E\left(\sigma^2_{n+i-j} \;\middle|\; \mathcal{F}_n\right) \text{ for } \max(r,s) > i,\ i > 1$$

$$= \dot{\omega} + \sum_{j=1}^{i-1} \left(\dot{\alpha}_j + \dot{\beta}_j\right) \widehat{\sigma}^2_{n+i-j} + \sum_{j=i}^{\max(r,s)} \left(\dot{\alpha}_j + \dot{\beta}_j\right) \sigma^2_{n+i-j} \text{ for } \max(r,s) > i,\ i > 1$$

E.11 Derivation Remark 5.5

From item (5) of Theorem 3.1 we deduce

$$\begin{aligned}
&\lim_{t\uparrow x_F} \frac{1-F(t+xu(t))}{1-F(t)} = (1+\gamma x)_+^{-\frac{1}{\gamma}} \\
\iff &\lim_{t\uparrow x_F} \frac{1-P(X\le t+xu(t))}{1-F(t)} = (1+\gamma x)_+^{-\frac{1}{\gamma}} \\
\iff &\lim_{t\uparrow x_F} \frac{1-P\left(\frac{X-t}{u(t)}\le x\right)}{1-F(t)} = (1+\gamma x)_+^{-\frac{1}{\gamma}} \\
\iff &\lim_{t\uparrow x_F} \frac{P\left(\frac{X-t}{u(t)}> x\right)}{1-F(t)} = (1+\gamma x)_+^{-\frac{1}{\gamma}} \\
\iff &\lim_{t\uparrow x_F} P\left(\frac{X-t}{u(t)}> x\right) = (1-F(t))\,(1+\gamma x)_+^{-\frac{1}{\gamma}} \quad .
\end{aligned}$$

Now, using Definition 5.1, we obtain

$$\begin{aligned}
&P\left(z_l > IRM_l\left(\alpha\right) \;\middle|\; \mathcal{I}_{l-1}\right) = \alpha \\
\iff &P\left(\frac{z_l-t}{u(t)} > \frac{IRM_l\left(\alpha\right)-t}{u(t)} \;\middle|\; \mathcal{I}_{l-1}\right) = \alpha \\
\implies &(1-F(t))\left(1+\gamma\frac{IRM_l(\alpha)-t}{u(t)}\right)_+^{-\frac{1}{\gamma}} \approx \alpha \\
\iff &\left(1+\gamma\frac{IRM_l(\alpha)-t}{u(t)}\right)_+^{-\frac{1}{\gamma}} \approx \frac{\alpha}{(1-F(t))} \\
\iff &1+\gamma\frac{IRM_l(\alpha)-t}{u(t)} \approx \left(\frac{\alpha}{(1-F(t))}\right)^{-\gamma} \\
\iff &\gamma\frac{IRM_l(\alpha)-t}{u(t)} \approx \left(\frac{\alpha}{(1-F(t))}\right)^{-\gamma} - 1 \\
\iff &IRM_l(\alpha)-t \approx \frac{u(t)}{\gamma}\left(\left(\frac{\alpha}{(1-F(t))}\right)^{-\gamma} - 1\right) \\
\iff &IRM_l(\alpha) \approx t + \frac{u(t)}{\gamma}\left(\left(\frac{\alpha}{(1-F(t))}\right)^{-\gamma} - 1\right) \quad .
\end{aligned}$$

Now, choosing $t = \widehat{b}\left(\frac{P}{k}\right) =$ and $u(t) = a\left(\frac{1}{1-F(t)}\right) = \widehat{a}\left(\frac{P}{k}\right)$ gives the result

$$\widehat{IRM}_l(\alpha) = \widehat{b}\left(\frac{P}{k}\right) + \frac{\widehat{a}\left(\frac{P}{k}\right)}{\widehat{\gamma}}\left(\left(\frac{P\alpha}{k}\right)^{-\widehat{\gamma}} - 1\right) \ ,$$

where $\widehat{\gamma}$, $\widehat{a}\left(\frac{P}{k}\right)$ and $\widehat{b}\left(\frac{P}{k}\right)$ are the moment estimators of Remark 3.4, where n is replaced by P and X is replaced by z_l (see Equations (5.27), (5.28) and (5.29)).

$\square$

E.12 Proof Lemma 6.1

From the assumption that Equations (6.4), (6.5) and (6.6) hold, Corollary 3.2 tells us that for each continuity point (x, y) for which $0 < G_0(x, y) < 1$ we have

$$\lim_{t\to\infty} t\,(1 - F(U_1(tx), U_2(ty)) = -\log G_0(x, y)\ . \tag{E.1}$$

Additionally, define for $x > 0$

$$g_t(x) = \frac{U_1(tx) - U_1(t)}{a_1(t)}\ ,$$

$$h_t(x) = \frac{U_2(ty) - U_2(t)}{a_2(t)}\ ,$$

where a_1 and a_2 are positive functions. Now, making use of item (4) of Theorem 3.1, deduce

$$\begin{aligned}
\lim_{t\to\infty} t(1 - F(U_1(tx), U_2(ty)) &= \lim_{t\to\infty} t(1 - P(X \le U_1(tx), Y \le U_2(ty)) \\
&= \lim_{t\to\infty} t\,(1 - P\,(X \le a_1(t)g_t(x) + U_1(t), Y \le a_2(t)h_t(y) + U_2(t))) \\
&= \lim_{t\to\infty} tP\,(X > a_1(t)g_t(x) + U_1(t) \cup Y > a_2(t)h_t(y) + U_2(t)) \\
&= \lim_{t\to\infty} tP\left(X > a_1(t)\frac{x^{\gamma_1} - 1}{\gamma_1} + U_1(t) \cup Y > a_2(t)\frac{y^{\gamma_2} - 1}{\gamma_2} + U_2(t)\right) \\
&= \lim_{t\to\infty} tP\left(\frac{X - U_1(t)}{a_1(t)} > \frac{x^{\gamma_1} - 1}{\gamma_1} \cup \frac{Y - U_2(t)}{a_2(t)} > \frac{y^{\gamma_2} - 1}{\gamma_2}\right) \\
&= \lim_{t\to\infty} tP\left(\left(1 + \gamma_1\frac{X - U_1(t)}{a_1(t)}\right)^{\frac{1}{\gamma_1}} > x \cup \left(1 + \gamma_2\frac{Y - U_2(t)}{a_2(t)}\right)^{\frac{1}{\gamma_2}} > y\right)\ .
\end{aligned} \tag{E.2}$$

Hence, using Equations (E.1) and (E.2) in combination with Theorem 3.3, results in

$$\lim_{t\to\infty} tP\left(\left(1 + \gamma_1\frac{X - U_1(t)}{a_1(t)}\right)^{\frac{1}{\gamma_1}} > x \cup \left(1 + \gamma_2\frac{Y - U_2(t)}{a_2(t)}\right)^{\frac{1}{\gamma_2}} > y\right) = V(x, y)\ ,$$

where the function $V(x, y)$ is defined by Equation (3.32). Remark 3.6 tells us that the function $V(x, y)$ determines the exponent measure. Hence, a more general expression is obtained, i.e.

$$\lim_{t\to\infty} tP\left(\left(\left(1+\gamma_1\frac{X-U_1(t)}{a_1(t)}\right)^{\frac{1}{\gamma_1}}, \left(1+\gamma_2\frac{Y-U_2(t)}{a_2(t)}\right)^{\frac{1}{\gamma_2}}\right)\in A_{x,y}\right) = \nu(A_{x,y})\ ,$$

where

$$A_{x,y} := \left\{(s,t)\in\mathbb{R}^2_+\ :\ s>x\cup t>y\right\}\ .$$

Now, let $k=k(n)\to\infty$, $\frac{k}{n}\to 0$ as $n\to\infty$. Replacing t by $\frac{n}{k}$ completes the proof

$$\lim_{n\to\infty}\frac{n}{k}P\left(\left(\left(1+\gamma_1\frac{X-U_1\left(\frac{n}{k}\right)}{a_1\left(\frac{n}{k}\right)}\right)^{\frac{1}{\gamma_1}}, \left(1+\gamma_2\frac{Y-U_2\left(\frac{n}{k}\right)}{a_2\left(\frac{n}{k}\right)}\right)^{\frac{1}{\gamma_2}}\right)\in A_{x,y}\right) = \nu(A_{x,y})\ .$$

□

Bibliography

[Bal07] T. Bali. “A Generalized Extreme Value Approach to Financial Risk Measurement”. In: *Journal of Money, Credit and Banking* (2007). Vol. 39, No. 7, pp. 1613 - 1649.

[Ban13] Banking Committee on Banking Supervision. “Monitoring tools for intraday liquidity management”. In: - (2013). Bank for International Settlements, pp.

[BCHT16] D. Banulescu, G. Colletaz, C. Hurlin, and S. Tokpavi. “Forecasting High-Frequency Risk Measures”. In: *Journal of Forecasting* (2016). No. 35, pp. 224 - 249.

[BDM11] A. Ball, E. Denbee, and M. Manning. “Intraday liquidity: risk and regulation”. In: *Financial Stability Paper* (2011). No. 11, pp.

[BG00] L. Bauwens and P. Giot. “The Logaritmic ACD Model: An Application to the Bid-Ask Quote Process of Three NYSE Stocks”. In: *Annales D’Économie et de Statisque* (2000). No. 60, pp. 117 - 149.

[BG01] L. Bauwens and P. Giot. *Econometric Modelling of Stock Market Intraday Activity*. Vol. 38, pp. 67. Springer-Science+Business Media, B.V., 2001.

[BGG03] L. Bauwens, F. Galli, and P. Giot. “The moments of Log-ACD models”. In: - (2003). pp.

[BGGV04] L. Bauwens, P. Giot, J. Grammig, and D. Veredas. “A comparison of financial durations models via density forecasts”. In: *Journal of Forecasting* (2004). Vol. 20, pp. 589 - 609.

[BH09] L. Bauwens and N. Hautsch. *Handbook of Financial Time Series*. pp. 953 - 979. Springer Berlin Heidelberg, 2009. Chap. Modelling Financial High Frequency Data Using Point Processes.

[Bol86] T. Bollerslev. “Generalized Autoregressive Conditional Heteroskedasticity”. In: *Journal of Econometrics* (1986). Vol. 31, pp. 307 - 327.

[Bys04] H. Byström. “Managing extreme risks in tranquil and volatile markets using conditional extreme value theory”. In: *International Review of Financial Analysis* (2004). No. 13, pp. 133 - 152.

[Cam05] S. Campbell. “A Review Of Backtesting and Backtesting Procedures”. In: *Finance and Economics Discussion Series* (2005). pp.

[CFM13] J. J. Cai, A. Fougères, and C. Mercaier. “Environmental data: multivariate Extreme Value Theory in practice”. In: *Journal de la Société Française de Statistique* (2013). Vol. 154, No. 2, pp. 178 - 199.

[Chr10] P. Christoffersen. “Encyclopedia of Quantitative Finance”. In: *Backtesting* (2010). pp.

[Chr98] P. Christoffersen. “Evaluating Interval Forecasts”. In: *International Economic Review* (1998). Vol. 39, No. 4, pp. 841 - 862.

[Col11] S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. pp. 74 - 91. Springer, 2011.

[DDFH04] G. Draisma, H. Drees, A. Ferreira, and L. de Haan. “Bivariate tail estimation: dependence in asymptotic independence”. In: *Bernouli* (2004). Vol. 10, No. 2, pp. 251 - 280.

[DDP09] G. Dionne, P. Duchesne, and M. Pacurar. “Intraday Value at Risk (IVaR) Using Tick-by-Tick Data with Application to the Toronto Stock Excange”. In: *Journal of Empirical Finance* (2009). Vol. 16, No. 5, pp. 777 - 792.

[DGT98] F. Diebold, T. Gunther, and A. Tay. “Evaluating Density Forecasts with Applications to Financial Risk Management”. In: *International Economic Review* (1998). Vol. 39, No. 7, pp. 863 - 883.

[EM04] R. Engle and S. Manganelli. “CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles”. In: *Journal of Business and Economic Statistics* (2004). Vol. 22, No. 4, pp. 367 - 381.

[Eng00] R. Engle. “The Econometrics of Ultra-High-Frequency Data”. In: *Econometrica* (2000). Vol. 68, No. 1, pp. 1 - 22.

[ER04] R. Engle and J. Russell. “Analysis of High Frequency Financial Data”. In: - (2004). New York University, University of California, University of Chicago, Graduate School of Business, pp.

[ER98] R. Engle and J. Russell. “Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data”. In: *Econometrica* (1998). Vol. 66, No. 5, pp. 1127 - 1162.

[Fel48] W. Feller. “On the Kolmogorov-Smirnov Limit Theorems for Empirical Distributions”. In: *The Annals of Mathematical Statistics* (1948). Vol. 19, pp. 177 - 189.

[FG05] M. Fernandes and J. Grammig. “Nonparametric specification tests for conditional duration models”. In: *Journal of Econometrics* (2005). Vol. 127, No. 1, pp. 35 - 68.

[GJ98] E. Ghysels and J. Jasiak. “GARCH for irregularly spaced financial data: the ACD-GARCH model”. In: *Studies in Nonlinear Dynamics and Econometrics* (1998). Vol. 2, No. 4, pp. 133 - 149.

[GK06] M. Gilli and E. Këllezi. “An Application of Extreme Value Theory for Measuring Financial Risk”. In: *Computational Economics* (2006). Vol. 27, pp. 1 - 23.

[GMT84] C. Gourieroux, A. Monfort, and A. Trognon. “Pseudo Maximum Likelihood Methods: Theory”. In: *Econometrica* (1984). Vol. 52, No. 3, pp. 681 - 700.

[Hau12] N. Hautsch. *Econometrics of Financial High-Frequency Data*. pp. 99 - 142. Springer, 2012.

[HF06] L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. pp. 15 - 16. Springer, 2006.

[HR98] L. de Haan and J. de Ronde. “Sea and Wind: Multivariate Extremes at Work”. In: *Extremes* (1998). Vol. 1, No. 1, pp. 7 - 45.

[Hus09] J. Hüsler. “Extreme value analysis in biometrics”. In: *Biometrical Journal* (2009). Vol. 51, No. 2, pp. 252 - 272.

[LB78] G. Ljung and G. Box. “On a measure of lack of fit in time series models”. In: *Biometrika* (1978). Vol. 65, pp. 297 - 303.

[LM94] W. Li and T. Mak. “On the squared residual autocorrelations in non-linear time series with conditional heteroscedasticity”. In: *Journal of Time Series Analysis* (1994). Vol. 15, No. 6, pp. 627 - 636.

[LT98] A. Ledford and J. Tawn. “Concomitant Tail Behaviour for Extremes”. In: *Advances in Applied Probability* (1998). Vol. 30, No. 1, pp. 197 - 215.

[LY03] W. Li and P. Yu. “On the residual autocorrelation of the autoregressive conditional duration model”. In: *Economic Letters* (2003). Vol. 79, No. 6, pp. 169 - 175.

[Moh14] A. Mohanta. “Impact of Basel III Liquidity Requirements on the Payments Industry: Liquidity management strategy for banks providing payment services”. In: - (2014). Capgemini, pp.

[MRW06] N. Meddahi, E. Renault, and B. Werker. “GARCH and irregularly spaced data”. In: *Economics Letters* (2006). Vol. 90, No. 2, pp. 200 - 204.

[MT06] M. Meitz and T. Teräsvirta. “Evaluating Models of Autoregressive Conditional Duration”. In: *Journal of Business and Economic Statistics* (2006). Vol. 24, No. 1, pp. 104 - 124.

[NJ11] P. Northrop and P. Jonathan. “Threshold modelling of spatially-dependent non-stationary extremes with application to hurricane-induced wave heights”. In: *Environmetrics* (2011). Vol. 22, No. 7, pp. 799 - 809.

[NW15] T. Neijs and M. Wycisk. “Robust Intraday Liquidity Management”. In: - (2015). Zanders Treasury and Finance Solutions, pp. 3.

[Pac08] M. Pacurar. “Autoregressive Conditional Duration (ACD) Models in Finance: A Survey of the Theoretical and Empirical Literature”. In: *Journal of Economic Surveys* (2008). Vol. 22, No. 4, pp. 711 - 751.

[PRT04] S. Poon, M. Rockinger, and J. Tawn. “Extreme Value Dependence in Financial Markets: Diagnostics, Models and Financial Implications”. In: *The Review of Financial Studies* (2004). Vol. 17, No. 2, pp. 581 - 610.

[RH04] E. Rengifo and A. Heinen. “Comovements in Trading activity: A Multivariate Autoregressive Model of Time Series Count Data Using Copulas”. In: *Econometric Society 2004 Far Eastern Meetings* (2004). No. 755, pp.

[Smi48] N. Smirnov. "Table for estimating the goodness of fit of empirical distributions". In: *The Annals of Mathematical Statistics* (1948). Vol. 19, No. 2, pp. 279 - 281.

[Tsa01] R. Tsay. *Palgrave Handbook of Econometrics: Volume 2 Applied Econometrics.* pp. 1004 - 1024. Palgrave Macmillan UK, 2001. Chap. Autoregressive Conditional Duration Models.

[WCJZ10] Z. Wang, X. Chen, Y. Jin, and Y. Zhou. "Estimating risk of foreign exchange portfolio: Using VaR and CVaR based on GARCH-EVT-Copula model". In: *Physica A: Statistical Mechanics and its Applications* (2010). Vol. 389, No. 21, pp. 4981 - 4928.

[Woo91] J. Wooldridge. "Specification testing and quasi-maximum-likelihood estimation". In: *Journal of Econometrics* (1991). No. 48, pp. 29 - 55.

[YZ03] B. Yan and E. Zivot. "Analysis of High-Frequency Financial Data with S-Plus". In: - (2003). University of Washington, Department of Economics, pp.