Time-Changed Brownian Motions and Copulas

Estimation, Inference, and Applications

Toke Christian Zinn

Gruppe 1.211b
Mathematics-Economics

February 1st to June 3rd

This page is intentionally left blank.

School of Engineering and Science
Mathematics-Economics
Aalborg University
http://www.ses.aau.dk

AALBORG UNIVERSITY
STUDENT REPORT

Title:
Time-Changed Brownian Motions and Copulas
Estimation, Inference, and Applications

Theme:
Continuous-time Financial Modelling

Project Period:
Spring Semester 2020

Project Group:
Group 1.211b

Participant(s):
Toke Christian Zinn

Supervisor(s):
Orimar Arregui Sauri

Page Numbers: 104

Date of Completion:
2020-28-05

Abstract:

We propose an estimator for the conditional copula of specific models which can be represented time-changed Brownian motions. We show that under suitable restrictions, the estimator is consistent and we provide a limit theorem in a very restricted case. Through numerical analysis, we argue that the result generalises to a larger class of models, and the asymptotic results for the restricted case are somewhat preserved. We then propose a method for forecasting the distribution of a portfolio and show its validity in a controlled simulation study. While the simulation study is unrealistic, we propose ways to augment the method to allow for much greater flexibility, which may render the method more useful in real applications.

Preface

This Master's Thesis in written in the Spring Semester 2020 by Toke Christian Zinn, a student in Mathematics and Economics at the Department of Mathematical Sciences, Aalborg University.

The document is typeset in LaTeX. Computations, modeling and the majority of figures are performed using the `R` language, [29]. Additional figures are produced in in the Open Source image editor Inkscape. In addition the base packages provided by the Core `R` team the following packages were used:

- R Studio, as an *Integrated Development Environment* for `R`, [32].
- `ggplot2`, for figures and data visualisation, [42].
- `DBI`, for interfacing the produced database for simulations studies. The database is available upon request, [30].
- `tikzDevice`, for exporting `ggplot2` figures to a Tikz figure in LaTeX, [36].

The primary functionalities are developed by the author and the code is available at **github.com/TokeZinn**.

The author would like to acknowledge the project advisor, Assistant Professor Orimar Sauri, for his inputs, patience, and continued interest in the project. In particular, the author thanks him for his invaluable inputs regarding Theorem 3.2, which would not have been proved otherwise.

Aalborg University, 2020-06-03

Toke Christian Zinn
<tzinn15@student.aau.dk>

"Young man, in mathematics you don't understand things. You just get used to them."

– John von Neumann, [[44], footnote page 208]

Contents

1

Introduction

The general theory of stochastic processes has applications almost everywhere; while specific observations can be considered as *independent* or sharing a weak dependence, it is more likely to be the exception rather than the rule.

Temporal dependence creates an immediate problem, especially from an inferential or statistical standpoint which is best explained by an example: Suppose we are an investor seeking to model the behaviour of multiple assets. We choose a selection of seemingly unrelated assets, that is we choose not to include stocks from companies within the same market. While it is unrealistic, one may argue that over short periods, different assets are independent, but one can hardly argue that there is temporal independence. Here lies the problem. Namely, each path can be considered as the single realisation of an abstract random variable whose image is an appropriate function space. Hence, when we wish to make inference or predictions, we have to include the temporal dependence and previous observations.

Initially, one may welcome this problem as a blessing; clearly, more information is better, and from a financial standpoint, one might even expect that additional information can help one speculate. However, one will soon be faced with the computational curse of highly multivariate data. To make matters worse, the rate at which financial data is arriving is increasing, and hence one must decide on an action at an increasingly faster pace.

To make the computational burden even worse, many models today rely on an excessive number of parameters. Models such as Recurrent Neural Networks are gaining momentum in the use of finance, see, e.g., [17], based on their success in modelling difficult behaviour in other fields. These models typically rely on an extremely high dimensional parameters space. While the computational burden of evaluation is typically limited the optimisation of these models present a problem in a high-frequency setting.

As a rule of thumb, economists have defined so-called *stylised facts* to which economic models should adhere. Several models have been implemented to accommodate the so-called stylised facts. Some models rely on discrete time steps and others continuous-time.

Only modelling the market at distinct times presents challenges. If one supposes regularly spaced data with increasing frequency, then the natural extension to time-series are continuous-time

stochastic processes. These models are robust to any frequency of data which is advantageous. Furthermore, under weak assumptions, time-series may be identified in the so-called *Skorokhod space* or space of cádlág functions. We will explore the Skorokhod Space and associated topologies later in the thesis.

Philosophically, one could argue for time-series in finance by arguing that data is only recorded once a bid or offer is placed. Alternatively, one may use the so-called *tick*-data arriving at regular frequencies. However, this does not pose a problem for stochastic processes either. We may define so-called *canonical projections* which are a collection of mappings that map a function to its evaluation at specific points, hence typically an element of $\mathbb{R}^n$ for some natural number n. The result is that one will end up with a multivariate distribution.

Copulas are a tool for modelling multivariate distributions. In particular, a multivariate distribution function may be decomposed into its marginal distributions and a copula. Abe Sklar, who initially claimed and later proved this decomposition, and his co-author Ben Schweizer writes "The name was chosen to express the fact that a copula embodies the manner in which a joint distribution function is coupled to its one-dimensional margins" in their book *Probabilistic Metric Spaces* [[35], Section 6.2].

Copulas were initially a tool for solving problems related to the probabilistic metric spaces. However, the immediate advantage of copulas was that since univariate distributions were well studied, then the statistician could apply his or her knowledge of univariate distributions to multivariate distributions via copulas. Furthermore, they allowed for the marginals to be chosen independently, which allowed for more realistically modelling.

In most classical statistics, the general measure of dependence is correlation. Typical linear correlation measures the linear dependence between model, but at times linear dependence is not sufficient. An example of non-linear dependence is found in returns. Returns typically exhibit little to no linear dependence, but the squared or absolute returns exhibit linear dependence. Copulas are an alternative as to the linear dependence modelling, as they allow for highly non-linear specifications.

While this indicates that copulas may solve many issues, they have also been met with criticism. Gaussian copulas received criticism after the financial crisis in 2008. Some claimed that it was partially to blame for the crisis since the Gaussian copula model was widely used but made unrealistic assumptions. In particular, the model of Li, [19], received heavy criticism. Studies were made to address the validity of the claim. David M. Zimmer found that the asymptotic independence of extreme events present in the Gaussian copula was unrealistic and presented alternatives [43]. Samuel Watts argued that it was not the particular copula, but rather the use of copulas in the specific context which was flawed [41]. Furthermore, Thomas Mikosch in his papers *Copulas: Tales and Facts* and *Copulas: Tales and Facts - Rejoinder* addresses several issues with the use of copulas, some of which relate to temporal dependence [21],[22].

However, Schmitz showed that the temporal dependence of several stochastic processes could be represented via copulas, [34]. He argued that some proofs of original and central theorems in stochastic processes could be expressed in terms of the copulas and argues that these proofs are simpler. He also derived the copulas of several classical stochastic processes such as the Brownian motion, the Ornstein-Uhlenbeck Process, and the Brownian Bridge with the possibility of extensions to continuous local martingales and semimartingales.

Of course, in order to apply the results from Schmitz, we need an estimator for these copulas. In order for the estimator to be useful, we must assert its consistency. We propose a method using realised variance for estimating the conditional copula of time-changed Brownian motion and investigates its asymptotic properties. Our method will rely on non-parametrical estimates of the quadratic variation, and so they are computationally favourable. We show the rate of convergence for such an estimator and limiting distribution in a simple case. We show that numerical experiments indicate that the type of convergence may be stronger than what we prove.

We also propose a semi-parametric forecasting procedure making use the aforementioned copula to forecast the distribution of a semimartingale. The forecasting shows how one may incorporate copulas temporally. We show how the forecasting procedure fares in a controlled environment where we simulate data from the Heston model, [11].

The thesis is divided into several chapters. In Chapter 2, we define the preliminary theory to prove our results.

In Chapter 3, we provide proof of the asymptotic properties of our proposed estimator. We show that it is consistent and provide a limit theorem in a very restricted setting. Using numerical analysis, we argue that the result may be enhanced.

In Chapter 4, we turn our focus to financial applications. Here we show a distributional forecasting procedure for portfolios, and apply the conditional coverage test from Christoffersen, [8], to test the validity.

In Chapter 5, we summarise the thesis. We discuss the results obtained, their applications, and future research. We then conclude our findings.

Problem statement: How can we use realised variance to estimate the conditional copula of a time-changed Brownian motion? How can temporal copulas be used in the assessment of portfolio risk?

2

Preliminary Theory

2.1 Notation

Notation	
(E, d_E)	Metric or Polish space
$\mathcal{B}(E)$	The Borel σ-algebra on (E, d_E)
$(\Omega, \mathcal{F}, \mathbb{P})$	Probability space
$\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,\infty)}, \mathbb{P}\right)$	Filtered probability space
$\xrightarrow{\mathbb{P}}$	Convergence in probability
$\xrightarrow{a.s}$	Almost sure convergence
$\overset{u.c.p.}{\Rightarrow}$	Uniform convergence on compacts in probability
$\xrightarrow{\mathcal{L}_S}$	Stable convergence in law
$\overset{\mathcal{L}_S}{\Rightarrow}$	Stable convergence in law for the Skorokhod topology
t, s	Elements in $[0, \infty)$
$a \vee b$	$\max(a, b), \quad a, b \in \mathbb{R}$.
$a \wedge b$	$\min(a, b), \quad a, b \in \mathbb{R}$.
$C(\mathcal{T}; E)$	Space of continuous functions from $\mathcal{T}$ to E.
$\mathbb{D}(\mathcal{T}; E)$	Space of cádlág functions from $\mathcal{T}$ to E.
$C^d(\mathcal{T}, E)$	Space of d-times continuously differentiable functions from $\mathcal{T}$ to $\mathcal{E}$.

2.2 Copulas

This section is based on [34], [7], and [24].

In classical multivariate probability, a typical approach is to assume the joint distribution of two random variables is some specific distribution. That is, we assume $(X, Y) \sim H$ where H is some multivariate distribution. However, in assuming a joint distribution one typically assumes the marginal distributions as well. A classic example is the multivariate Gaussian distribution which is widely used.

Copulas provide a framework for overcoming the implicitly assumed marginals and allowing a more flexible and modular approach to modelling multivariate distributions.

2.2.1 Construction

A large part of the construction relies on the *Probability Integral Transform*. Suppose that we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X : \Omega \to \mathbb{R}$. The law of X is the probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined by

$$\mathcal{L}_X(B) = \mathbb{P}(X \in B), \ B \in \mathcal{B}(\mathbb{R}).$$

Since X is a $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable function, we have that $\mathcal{L}_X$ is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We can then define the *cumulative distribution function*

$$F_X(x) = \mathcal{L}_X((-\infty, x]), \quad x \in \mathbb{R}.$$

Furthermore, the function F_X uniquely characterises $\mathcal{L}_X$.

Proposition 2.1 (Probability Integral Transform)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X : \Omega \to \mathbb{R}$ a random variable. Then, if the cumulative distribution function F_X is continuous and strictly increasing we have*

$$F_X(X) \sim \text{Unif}([0, 1])$$

Proof. It suffices to show that $Y = F_X(X)$ has a uniform law.

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(F_X(X) \leq y) \\ &= \mathbb{P}(X \leq F_X^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) \\ &= y, \end{aligned}$$

where the invertibility follows the fact that F_X is continuous and strictly increasing. ■

If F_X is not continuous and strictly increasing, then the proposition cannot be recovered in full. However, it does hold in a "weaker" sense. Before we can elaborate, we need the following definition.

Definition 2.1 (Quantile Function)
Let $F : \mathbb{R} \to [0, 1]$ be a cumulative distribution function. Define,

$$F^{-1}(p) = \inf\{x \in \mathbb{R} \mid F(x) \geq p\}, \quad p \in [0, 1].$$

The function F^{-1} is called the generalised inverse or quantile function.

We have the following proposition

Proposition 2.2. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \to \mathbb{R}$ be a random variable with cumulative distribution function $F_X : \mathbb{R} \to [0, 1]$. Let $U \sim \text{Unif}([0, 1])$, then*

$$F_X^{-1}(U) \sim F_X$$

Proof. Let $Y = F_X^{-1}(U)$ then

$$\mathbb{P}(Y \leq y) = \mathbb{P}(F_X^{-1}(U) \leq y) = \mathbb{P}(U \leq F_X(y)) = F_X(y)$$

■

Proposition 2.2 tells us that while $F_X(X)$ does not necessarily follow a uniform law we have $F_X^{-1}(U)$ follows F_X. The implication is also convenient for sampling from a known distribution function F_X.

Now, suppose $(X, Y) \sim H$, where H is some bi-variate cumulative distribution function. Assume for simplicity that the marginal distributions $F_X(x) = H(x, \infty)$ and $F_Y(y) = H(\infty, y)$ are continuous and strictly increasing.

Suppose we apply the marginal distributions to X and Y. By Proposition 2.1 we have that $U = F_X(X) \sim \text{Unif}([0, 1])$ and similarly $V = F_Y(Y) \sim \text{Unif}([0, 1])$. However, what distribution does the pair (U, V) follow? It seems intuitive that what remains is only the dependence between X and Y. Such a construct is called a *copula*.

Definition 2.2 (Copula)
Let $C : [0, 1]^d \to [0, 1]$ be a function. If C a distribution function on $[0, 1]^d$ with uniform marginals, then C is said to be a copula. Furthermore, we define

$$\mathcal{C}^d = \{C : [0, 1]^d \to [0, 1] \mid C \text{ is a copula}\}.$$

In order to ease notation significantly we introduce the following partial relation on $\mathbb{R}^d$.

Definition 2.3 (Partial Relation on $\mathbb{R}^d$)
Let $d \in \mathbb{N}$. For $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ and $y = (y_1, y_2, \ldots, y_d) \in \mathbb{R}^d$ we define the following notation

$$x \leq y \text{ if } x_i \leq y_i, \ \forall i \in \{1, 2, \ldots, d\}.$$

Furthermore, for $x, y \in \mathbb{R}^d$ such that $x \leq y$ we define

$$(x, y] = \bigtimes_{i=1}^{d} (x_i, y_i].$$
$$[x, y] = \bigtimes_{i=1}^{d} [x_i, y_i].$$

We have the following characterisation of a copula.

Proposition 2.3. *Let $C : [0,1]^d \to [0,1]$ be a function. Then C is a copula if, and only if, it satisfies*

1 *$C(1, \ldots, 1, x, 1 \ldots, 1) = x$ for every $x \in [0,1]$.*

2 *$C(x) = 0$, where $x = (x_1, \ldots x_d)$ if there is at least one $i \in \{1, 2, \ldots, d\}$ such that $x_i = 0$.*

and finally, for all $a, b \in [0,1]^d$ with $a \leq b$ we have

$$\sum_{x \in \bigtimes_{i=1}^{d} \{a_i, b_i\}} (-1)^{|\{i \in \{1,2,\ldots,d\} | x_i = a_i\}|} C(x) \geq 0,$$

where $|\{i \in \{1, 2, \ldots, d\} \mid x_i = a_i\}|$ is the cardinality of the set $\{i \in \{1, 2, \ldots, d\} \mid x_i = a_i\}$.

Remark: Note that $x \in \bigtimes_{i=1}^{d} \{a_i, b_i\}$ simply means that x iterates through all possible configurations of boundary points of the set $(a, b]$. That is, x_i is either a_i or b_i.

We will not prove Proposition 2.3. However, note that the first requirement corresponds to C having uniform marginals. The second requirement is the so-called grounded property. It corresponds to the event

$$\mathbb{P}(U \leq u, V \leq 0) = C(u, 0) = 0, \quad (U, V) \sim C,$$

which is intuitive since copulas are continuous distribution functions - we will later justify this claim. Finally, the third condition guarantees that C "behaves" like a distribution; let μ be a measure and let $A \subset B$ belong to the underlying σ-algebra, then we have

$$\mu(A) \leq \mu(B).$$

Let $I = (a, b]$ and $J = (a, c]$ where $a, b, c \in [0,1]^d$ with $a < b \leq c$. Since C is the distribution of a random variable with uniform marginals it induces a unique probability measure, μ, on $([0,1]^d, \mathcal{B}([0,1]^d)$ and we have

$$\mu(I) = C(b) - C(a) \leq C(c) - C(a) = \mu(J).$$

Conversely, given a probability measure μ on $([0,1]^d, \mathcal{B}([0,1]^d))$ and letting $u \in \{u \in [0,1]^d \mid \exists! i \in \{1,2,\ldots,d\} : u_i \neq 1\}$, where $\exists!$ is the unique existence quantifier, i.e. u is of the form $(1,1,\ldots,u_i,\ldots,1,1)$, if

$$\mu((0,u]) = u_i,$$

then μ induces a copula by defining

$$C(u) = \mu((0,u]),\ \forall u \in [0,1]^d.$$

A well known example is the Lebesgue measure, λ_d, on $([0,1]^d, \mathcal{B}([0,1]^d))$, since

$$\lambda_d((0,u]) = \prod_{i=1}^{d} \lambda_1((0,u_i]) = \prod_{i=1}^{d} u_i,$$

where λ_1 is the Lebesgue measure on $([0,1], \mathcal{B}([0,1]))$. λ_d corresponds to the so-called independence copula, since it is just the product of the uniform marginals.

In order to further ease notation, we will use the following notation. If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $X : \Omega \to \mathbb{R}^d$ is a random variable, then the marginals follow $X_i \sim F_{X_i}$ for some distribution F_{X_i} for all $i \in \{1,2,\ldots,d\}$. We then define

$$F_X(x) = (F_{X_1}(x_1), F_{X_2}(x_2), \ldots, F_{X_d}(x_d)), \tag{2.1}$$

and we say that F_X is the *marginal distribution function* of X. We say that F_X is continuous if all the marginals F_{X_i} are continuous.

We claimed that copulas were the remaining dependence between random variables after applying each of their marginal distributions. The following theorem justifies that claim.

Theorem 2.1 (Sklar's Theorem)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X : \Omega \to \mathbb{R}^d$ be a random variable. Let $X \sim H$ where H is some distribution function with marginal distribution function F_X, in the sense of* (2.1)*, then there exists a copula C such that*

$$H(x) = C(F_X(x)),\ \forall x \in \mathbb{R}^d.$$

Furthermore, if F_X is continuous the copula is unique, otherwise it is uniquely defined of the range of F_X. Conversely, if C is a d-dimensional copula and $F_i, i \in \{1,2,\ldots,d\}$ are univariate distribution functions then defining $F(x) = (F_1(x_1), F_2(x_2), \ldots, F_d(x_d))$ we have that $C(F(x))$ defines a distribution function on $\mathbb{R}^d$.

The original theorem was presented in [37] and an excellent proof can be found in [10].

If the marginals are continuous then Theorem 2.1 follows immediately from the fact that F_{X_i} are invertible on their image and we may then invert F_X entry wise. Hence, for any distribution take

$$C(u) = H(F_X^{-1}(u)),\ u \in [0,1]^d. \tag{2.2}$$

Many of the earliest studied copulas arose from (2.2). For instance, the Gaussian copula. Let $\Phi_\Sigma(x)$ denote the standardized multivariate Gaussian distribution with correlation matrix Σ. Then clearly the marginals follow the standard univariate Gaussian distribution, Φ. The Gaussian copula is then defined by

$$C_\Phi((u_1, u_2, \ldots, u_d); \Sigma) = \Phi_\Sigma((\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_d)), \quad u \in [0,1]^d. \tag{2.3}$$

By Theorem 2.1 we immediately have that we may construct a distribution with arbitrary marginals but the dependence of a Gaussian distribution.

Now, in finance, we often wish to include as much information as possible. Hence, when some information is revealed to us we wish to condition our random variable to this information.

In order to rigorously define these conditional laws we first define conditional expectations and by extension conditional probabilities. However, for sufficiently "nice" image spaces conditional probabilities admit a slightly more interpretable definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, (S, Σ) a measurable space and $X : \Omega \to S$ a random variable. If S is sufficiently nice, then there exist a function $\mu_X : S \times \mathcal{F} \to [0,1]$ such that for each $x \in S$ it is a probability measure on $(\Omega, \mathcal{F})$ and for each $F \in \mathcal{F}$ it is a $\Sigma/\mathcal{B}([0,1])$ measurable map. Furthermore, it satisfies

$$\mathbb{P}(F \cap X^{-1}(B)) = \int_F \mu_X(x, A) d\mathcal{L}_X(x), \quad \forall F \in \mathcal{F}.$$

Typically, we denote $\mu_X(x, A) = \mathbb{P}(A \mid X = x)$. Furthermore, we get that for each $x \in S$ we can define the conditional probability distribution function of another random variable $Y : \Omega \to \mathbb{R}$ as $F_{Y|X}(y \mid x) = \mathbb{P}(Y \leq y \mid X = x)$.

We now wish to relate conditional distributions to copulas. In a similar fashion to the original Sklar's Theorem we initially define for a random variable $X_i : \Omega \to \mathbb{R}$ with $i \in \{1, 2, \ldots, d\}$ and a random variable W

$$F_{X_i|W}(x_i \mid w) = \mathbb{P}(X_i \leq x_i \mid W = w),$$

where x_i is in the image of X_i and similarly w is in the image of W. For $X = (X_1, X_2, \ldots, X_d)$ we define

$$F_{X|W}(x \mid w) = (F_{X_1|W}(x_1 \mid w), F_{X_2|W}(x_2 \mid w), \ldots, F_{X_d|W}(x \mid w))$$

We say that $F_{X|W}$ is continuous in x if each $F_{X_i|W}$ is continuous in x_i.

Theorem 2.2 (Sklar's Theoreom for Conditional Probabilities)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ we a probability space and $X = (X_1, X_2, \ldots, X_d)$ and W be a random variables on Ω and assume the appropriate regular conditional probabilities exist. Let $H_{X|W}(x \mid w) = \mathbb{P}(X \leq x \mid W = w)$. If $F_{X|W}$ is continuous in x for every w, then for each w there exists a unique copula $C_{X|W}(\cdot \mid w)$ such that*

$$H_{X|W}(x \mid w) = C_{X|W}(F_{X|W}(x \mid w) \mid w).$$

for every x in the domain of $F_{X|W}(\cdot \mid w)$ and every w in the image of W. Conversely, if $F_{X|W}$ is a marginal distribution function for each w in the image of W and $C_{X|W}$ is copula for each w the image of W, then $x \mapsto C_{X|W}(F_{X|W}(x \mid w) \mid w)$ defines a distribution function for each w in the image of W.

We refer to [26] for the proof. The general idea here is that copulas allow us to capture the multivariate dependence, even conditionally. It follows fairly naturally that

$$C_X(u) = \int C_{X|W}(u \mid w) d\mathcal{L}_W(w), \quad u \in [0,1]^d$$

where the domain of integration is the image of W. Conditional copulas may also be considered

as "random" copulas. Of course, by our definition, a copula cannot be random. A conditional copula is still a copula, or rather a family of copulas for each w in the domain of the random variable W.

2.2.2 Properties

In this subsection we prove some properties for the space $\mathcal{C}^d$. First note that the space $\mathcal{C}^d \subset L^p([0,1]^d)$ for $p \geq 1$, including $p = \infty$, since copulas are bounded by 1. However, they satisfy even stronger properties.

Proposition 2.4. *Let $C \in \mathcal{C}^d$, then C is Lipschitz continuous.*

See [[34], Corollary 2.8] for a proof.

The next results are related to the partial derivatives of a copula. To ease notation we will use the multi-index notation. Let $\alpha \in \mathbb{N}^d$, and let f be a function. We then define

$$D_\alpha f = \frac{\partial^\alpha f}{\partial^{\alpha_1} x_1 \partial^{\alpha_2} x_2 \dots \partial^{\alpha_d} x_d}.$$

D_α may be considered either as a purely mathematical expression defined only when the expression makes sense or in terms of an operator. We merely need it to ease notation. We provide a brief example: let $\alpha = (1,0,0)$ and $f(x,y,z) = x^2 + y^2 + z^2$, then $D_\alpha f = 2x$. Furthermore, for functions defined in distinct variables we define

$$D_x = \frac{\partial}{\partial x},$$

for instance if $C(u,v)$ is a bi-variate copula, then

$$D_u C(u,v) = \frac{\partial}{\partial u} C(u,v).$$

We now consider the significance of the partial derivatives. Suppose $(U,V) \sim C$ for some copula $C \in \mathcal{C}^2$. Then,

$$\mathbb{P}(U \leq u, V \leq v) = C(u,v), \quad u,v \in [0,1].$$

Suppose we want the conditional probability of U given V, then

$$\begin{aligned} \mathbb{P}(U \leq u \mid V = v) &= \lim_{\varepsilon \to 0} \mathbb{P}(U \leq u, v \leq V \leq v + \varepsilon) && (2.4) \\ &= \lim_{\varepsilon \to 0} \frac{C(u, v+\varepsilon) - C(u,v)}{\varepsilon} && (2.5) \\ &= D_v C(u,v), && (2.6) \end{aligned}$$

and by Rademacher's theorem we know that (2.6) exists for almost all $(u,v) \in (0,1)^2$. The implication of (2.6) is the following theorem.

Theorem 2.3. *Let $(X_1, X_2, \dots, X_n)$ where $n \in \mathbb{N}$ be continuous random variables with marginal distribution functions $F_i, i \in \{1,2,\dots,n\}$. Let C be the copula of $(X_1, X_2, \dots, X_n)$, and let K be the copula of $(X_1, X_2, \dots X_k)$ where $k < N$. Then, if the marginals are continuously differentiable and the copula K is absolutely differentiable and C is continuously differentiable with respect to the k first entries, i.e. $\alpha = \sum_{i=1}^k e_i$, where $(e_i)_{i=1}^k$ is the standard basis of $\mathbb{R}^k$, and*

$$D_\alpha C(u)$$

is continuous. Then,

$$\frac{D_\alpha C(F_1(X_k), \ldots, F_k(X_k), F_{k+1}(x_{k+1}), \ldots F_n(x_n)}{D_\alpha K(F_1(X_1), \ldots, F_k(X_k) \prod_{i=1}^k F_i'(X_i)}$$

is a version of

$$\mathbb{P}(X_{k+1} \leq x_{k+1}, \ldots, X_n \leq x_n \mid X_1, \ldots, X_k).$$

Corollary 2.1. *Let $(X, Y) \sim H$ where H is a continuous and differentiable bi-variate cumulative distribution function with continuously differentiable marginals. Let $C(u, v)$ be the copula of (X, Y), then*

$$D_{(1,0)} C(F_X(X), F_Y(y))$$

is a version of

$$\mathbb{P}(Y \leq y \mid X).$$

for all y in the domain of F_Y.

We will not prove Theorem 2.3 or Corollary 2.1 but refer to [[34], pp. 21-22 Corollary 2.28] for a proof. The analogous statement holds if one differentiates with respect to the second argument.

Corollary 2.1 has direct applications in sampling since it allow us to generate a sample of X by first sampling U and applying Proposition 2.2. Then, we can obtain the conditional law using Corollary 2.1 and again to sample V such that $(U, V) \sim C$ and finally applying Proposition 2.2 to obtain Y from V. Similarly, we may use Theorem 2.3 to sample from high dimensional distributions. It may also help us identify the copula; if we know the conditional distribution of Y given X, and the marginals are continuous and strictly increasing, then we can recover the copula by integration.

2.3 The Skorokhod Space

This section is based on [14], [13], [4].

We will first define the Skorokhod Space. We will not recover the theory in full, but refer to the sources mentioned above, specifically [14] and [4] for extensive treatment.

We will later propose an estimator for the conditional copula of certain time-changed Brownian motions at two distinct times. We want our estimator to not only converge point-wise for each distinct pair of times but to converge uniformly over all pairs of time. In order to do so, we need to develop tools for asserting uniform convergence in this sense.

We first define a Polish space.

Definition 2.4 (Polish Space)
Let (X, τ) be a topological space. If (X, τ) is separable and completely metrizable, we say that (X, τ) is a Polish space.

Remark: For an introduction of the topological implications of a Polish Space we refer to Appendix A.1.

Polish spaces are, in a sense, a class of well-behaved topological spaces. Namely, we can find a complete metric which allows us to describe convergence. Furthermore, since Polish spaces are separable, then we avoid many measurability issues since measures are countably additive and exhibit algebraic properties under countably many set operations. A lot of fundamental theorems of probability typically require a Polish structure in the image of random variables.

We will now present some function spaces, which are Polish. Our reason for doing so is that we are interested in the convergence of an entire process and not only point-wise convergence. In order to determine convergence, we will need a metric. Now, while our limit may be a continuous function our approximations are typically not, therefore the space of continuous functions, $C([0,\infty),\mathbb{R})$, is not sufficient.

Definition 2.5 (Cádlág and Skorokhod Space)
Let $\mathcal{T}$ be of the form $[0,T]$ for some $T<\infty$ or $[0,\infty)$ and (E,d_E) a metric space. A function $f:\mathcal{T}\to E$ is said to be cádlág if $\lim_{s\uparrow t} f(s)=f(t-)$ exists for every $t\in\mathcal{T}$ and

$$\lim_{s\downarrow t} f(s)=f(t),\quad t\in\mathcal{T}.$$

Furthermore, the space of cádlág function from $\mathcal{T}$ to E denoted $\mathbb{D}(\mathcal{T};E)=\{f:\mathcal{T}\to E\mid f \text{ is cádlág}\}$ is called the Skorokhod Space.

We can think of cádlág functions as an extension of continuous functions which "jumps", hence why cádlág processes appear naturally in mathematical finance; sometimes the price-path of an asset will jump as a result of unexpected news. This argument is only partially valid, as one may define processes which consist only of jumps.

Having introduced the Skorokhod space, it remains to introduce the appropriate metric for which it is Polish. At first glance, the metric may seem quite counter-intuitive. Therefore, we initially consider the space of continuous functions.

2.3.1 The Space of Continuous Functions

In classical and functional analysis the space of continuous functions is well-known. Formally, let (E,d_E) and $(E',d_{E'})$ be metric spaces and $K\subseteq E'$ be compact. Now, define

$$C(K;E)=\{f:K\to E\mid f \text{ is continuous}\}.$$

We may endow $C(K;E)$ with a the metric

$$d_\infty(f,g)=\sup_{t\in K} d_E(f(t),g(t)).$$

We immediately have the following theorem.

Theorem 2.4. *Let (E,d_E) and $(E',d_{E'})$ be Polish metric spaces and $K\subset E'$ compact. Define*

$$C(K;E)=\{f:K\to E\mid f \textit{ is continuous}\},$$

and

$$d_\infty(f,g) = \sup_{t\in K} d_E(f(t),g(t)).$$

then $(C(K,E), d_\infty)$ *is a Polish metric space.*

Remark: Weaker assumptions can be placed on the domain. However, for our use Theorem 2.4 is sufficient. Namely, the domain K does not have to belong to any space, it just has to be a compact metrizable space and E has to be Polish.

Theorem 2.4 is well-known and it tells us that $C([0,T];\mathbb{R})$ is Polish. However, many processes are defined on the entire real line, or the positive real half-space, i.e. $C([0,\infty);\mathbb{R})$.

It turns out that if one naively tries to extend the metric of uniform convergence, d_∞, then it fails to be a metric. Since continuous functions on $[0,\infty)$ do not need to be bounded, we have that the supremum metric is not a real metric; it may take infinite values which violates the fact that a metric is a mapping to $[0,\infty)$. One can recover a metric on $C([0,\infty),\mathbb{R})$ from the metric above, by taking an increasing sequence of compact sets and weighting them appropriately. We cover this construction in Appendix A.2.

Many approximations are naturally cádlág, especially in the case of stochastic processes where we may not know a priori if the process is continuous or not.

Example 2.1. Consider the identity function

$$\begin{aligned} \mathrm{Id} : [0,1] &\to [0,1] \\ t &\mapsto t. \end{aligned}$$

We now construct an approximation of the identity using indicators. We define

$$f_n(t) = \mathbb{1}_{\{1\}}(t) + \sum_{i=1}^{n-1} \frac{i}{n} \mathbb{1}_{\left[\frac{i}{n},\frac{(i+1)}{n}\right)}(t).$$

Clearly we have that $\sup_{t\in[0,1]} |t - f_n(t)| = \frac{1}{n}$, hence f_n converges uniformly to Id which is visualised in Figure 2.1.

□

The next question is, does $\mathbb{D}([0,T];\mathbb{R})$ or $\mathbb{D}([0,\infty);\mathbb{R})$ endowed with the topology of (local) uniform convergence preserves the Polish structure? Sadly, the answer is no. Specifically, $\mathbb{D}([0,\infty),\mathbb{R})$ is completely metrizable under the topology of local uniform convergence, but it lacks separability. However, Skorokhod, after whom the space is named, showed in his famous paper from 1954 that $\mathbb{D}([0,\infty),\mathbb{R})$ indeed is a Polish space with a topology induced by a metric which resembles the metric of uniform convergence [38].

We conclude with two brief example where we visualize the open balls in $\mathbb{D}([0,T];\mathbb{R})$ or $C([0,T];\mathbb{R})$ under the topology of uniform convergence and highlight its issues in $\mathbb{D}([0,T];\mathbb{R})$.

Example 2.2. Let $\varepsilon > 0$ and consider the ball

$$\overline{B_\varepsilon(\mathrm{Id})} = \{f \in \mathbb{D}([0,T];\mathbb{R}) \mid d(f,\mathrm{Id}) \leq \varepsilon)\} \tag{2.7}$$

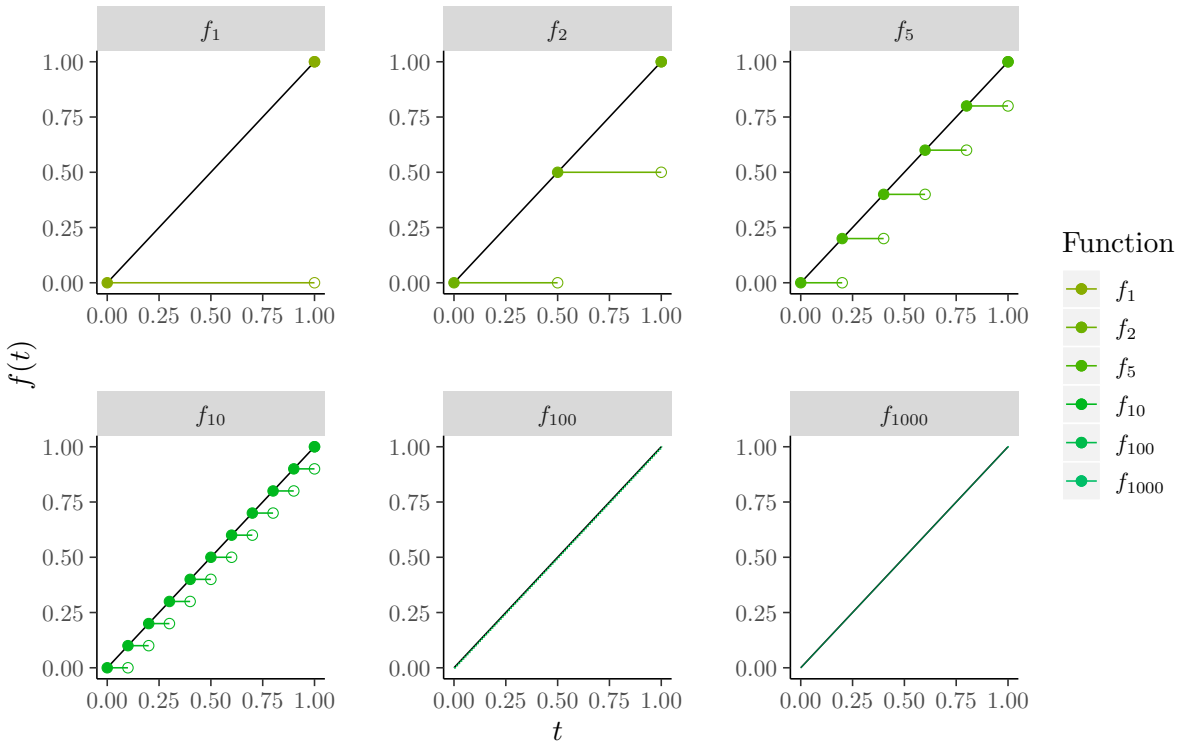

Figure 2.1: Approximation of Identity using Indicators

where Id is the identity function. Let $\varepsilon = \frac{1}{5}$ and consider the functions

$$f(t) = t + \frac{1}{6}\sin(6t)$$

$$g(t) = \mathbb{1}_{\{1\}}(t) + \sum_{i=1}^{4} \frac{i}{5}\mathbb{1}_{\left[\frac{i}{5},\frac{(i+1)}{5}\right)}(t).$$

Clearly, both $f, g \in \overline{B_\varepsilon(\mathrm{Id})}$; see Figure 2.2.

□

Example 2.2 shows that spheres in the topology of uniform convergence form a band of width ε around their centre. For continuous functions, this space is connected. However, when the function is cádlág the area becomes disconnected for small ε; consider Example 2.3.

Example 2.3. Consider the function

$$g(t) = \left(\frac{4}{5}\mathbb{1}_{[\frac{4}{5},1]}(t)\right) + \sum_{i=1}^{3} \frac{i}{5}\mathbb{1}_{\left[\frac{i}{5},\frac{(i+1)}{5}\right)}(t).$$

and recall the notation $g(t-) = \lim_{s\uparrow t} g(s)$ and define

$$D_g = \{t \in [0,1] \mid g(t) - g(t-) \neq 0\}.$$

Suppose we wish to approximate g using a continuous function. We know that

$$D_g = \left\{\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}\right\}$$

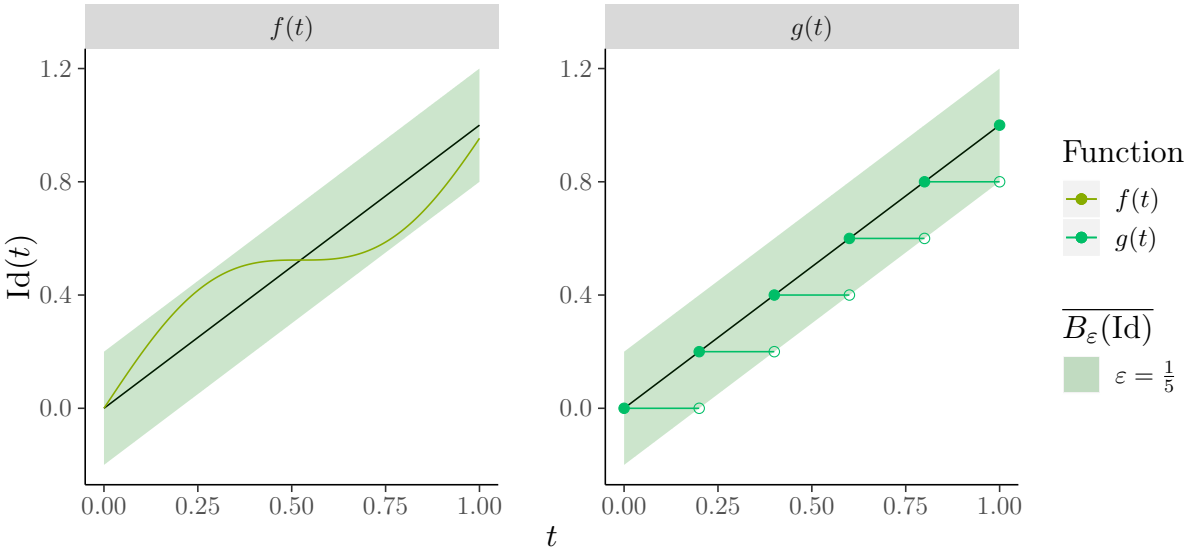

Figure 2.2: Element of $C([0,1];\mathbb{R})$ and $\mathbb{D}([0,1];\mathbb{R})$ in $\overline{B_\varepsilon(\mathrm{Id})}$.

Now, let $\varepsilon > 0$ be given. Note that while $\varepsilon > \frac{1}{2}\sup_{s\in[0,1]}|g(s) - g(s-)|$ then we can still approximate g using continuous functions in the sense that there exists a continuous function such that $d_\infty(f,g) < \varepsilon$. However, when $\varepsilon \leq \frac{1}{2}\sup_{s\in[0,1]}|g(s)-g(s-)|$ then any function $f \in B_\varepsilon(g)$ must satisfy

$$D_g \subseteq D_f. \tag{2.8}$$

We visualise in Figure 2.3, where (2.8) becomes clear.

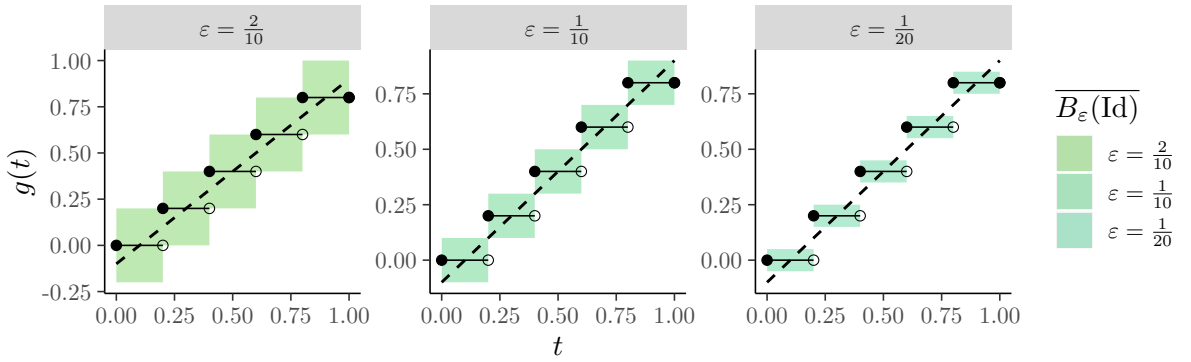

Figure 2.3: Approximation of $g(t)$ using a straight line. It succeeds for $\varepsilon \in (\frac{1}{10}, \infty)$ but fails for $\varepsilon \in (0, \frac{1}{10}]$.

□

We saw in Example 2.3 that for sufficiently small epsilon our approximation will have to share at least the discontinuities of the approximant. It is a very strong requirement that the discontinuities are preserved in approximation, especially if we are to approximate a process and we do not observe it for all times; the exact moment the process jumps may be indeterminable.

2.3.2 The Space of Cádlág Functions

We saw in Examples 2.2 and 2.3 that continuous functions may be approximated by cádlág functions.

In the topology of uniform convergence, two function are in some way near each other if we can make small changes in the image space such that the graphs of the functions coincide. For cádlág functions, we need to extend this to the domain as well; consider the function $f_n(t) = \mathbb{1}_{[1+\frac{1}{n},\infty)}(t)$ and the function $f(t) = \mathbb{1}_{[1,\infty)}(t)$. For large n the two function seem *close*, however in the topology of uniform convergence the f_n does not approach f, since only the image is considered.

To deal with this issue we take a function $\lambda : [0,\infty) \to [0,\infty)$ which is strictly increasing and invertible, i.e. injective and surjective, then we can think of λ as a change of time. We denote this space of functions $\Lambda([0,\infty))$. We present a small technical lemma.

Lemma 2.1. *Let $\mathcal{T} \in \{[0,T] \mid 0 < T < \infty\} \cup \{[0,\infty)\}$ Let $\Lambda(\mathcal{T})$ be defined by*

$$\Lambda(\mathcal{T}) = \{\lambda : \mathcal{T} \to \mathcal{T} \mid \lambda(s) < \lambda(t),\ s < t \in \mathcal{T}, \lambda \text{ is invertible } \}.$$

The following are true for $\lambda \in \Lambda(\mathcal{T})$.

(1) λ is continuous with a continuous inverse $\lambda^{-1} \in \Lambda(\mathcal{T})$.

(2) $\lambda_2 \circ \lambda_1 \in \Lambda(\mathcal{T})$ for $\lambda_1, \lambda_2 \in \Lambda(\mathcal{T})$.

(3) $d_\infty(\lambda, Id) = d_\infty(\lambda^{-1}, Id)$ for $\mathcal{T} = [0,T]$ with $T < \infty$.

(4) $d_\infty(f \circ \lambda, g) = d_\infty(f, g \circ \lambda^{-1})$ for $f, g \in \mathbb{D}([0,T];\mathbb{R})$.

Proof.

Claim (1): We know by surjectivity that $\lambda(\mathcal{T}) = \{\lambda(t) \mid t \in \mathcal{T}\} = \mathcal{T}$. Assume for contradiction that λ is not continuous, then there exists $t \in \mathcal{T}$ and an $\varepsilon > 0$ such that no matter what $\delta > 0$ there is an $s \in \mathcal{T}$ such that $|\lambda(s) - \lambda(t)| \geq \varepsilon$, but since λ is strictly increasing and surjective we arrive at a contradiction. Hence λ is continuous. Note that by assumption λ^{-1} is well-defined. First, note that λ^{-1} is strictly increasing since

$$s < t \iff \lambda(s) < \lambda(t),$$

applying λ^{-1} to both sides yields the desired result. Since λ^{-1} is strictly increasing and clearly invertible - by the original λ - we conclude $\lambda^{-1} \in \Lambda(\mathcal{T})$. It follows by the arguments above that λ^{-1} is also continuous.

Claim (2): Follows directly from

$$s < t \iff \lambda_1(s) < \lambda_1(t)$$
$$s < t \iff \lambda_2(s) < \lambda_2(s).$$

Resulting in

$$s < t \iff \lambda_1(s) < \lambda_1(t) \iff \lambda_2(\lambda_1(s)) < \lambda_2(\lambda_1(t)).$$

Now clearly by injectivity and surjectivity of both λ_1 and λ_2 we have the map $\lambda_2 \circ \lambda_1$ is invertible and strictly increasing. The result is that $\lambda_2 \circ \lambda_1 \in \Lambda(\mathcal{T})$.

Claim (3): We must show that

$$\sup_{t\in[0,T]} |\lambda(t) - t| = \sup_{t\in[0,T]} |\lambda^{-1}(t) - t|.$$

Note that $\lambda(0) = \lambda^{-1}(0) = 0$ and similarly $\lambda(T) = \lambda^{-1}(T) = T$. By surjectivity we have that

$$\forall t' \in [0,T] \exists t \in [0,T] : t' = \lambda(t), \tag{2.9}$$
$$\forall t'' \in [0,T] \exists t \in [0,T] : t'' = \lambda^{-1}(t). \tag{2.10}$$

Furthermore, by Claim 1, we have that λ and λ^{-1} are continuous. Hence the mappings

$$t \mapsto |\lambda(t) - t|$$
$$t \mapsto |\lambda^{-1}(t) - t|$$

are continuous. Now, by continuity we know that

$$\exists t' \in [0,T] : |\lambda(t') - t'| = \sup_{t\in[0,T]} |\lambda(t) - t| \tag{2.11}$$
$$\exists t'' \in [0,T] : |\lambda^{-1}(t'') - t''| = \sup_{t\in[0,T]} |\lambda^{-1}(t) - t|. \tag{2.12}$$

Now let t' be such that (2.11) holds. By (2.10) we now have

$$\begin{aligned}\sup_{t\in[0,T]} |\lambda(t) - t| &= |\lambda(t') - t'| \\ &= |\lambda(\lambda^{-1}(t'')) - \lambda^{-1}(t'')| \\ &= |\lambda^{-1}(t'') - t''| \\ &\leq \sup_{t\in[0,T]} |\lambda^{-1}(t) - t|\end{aligned}$$

Similar arguments using (2.12) and (2.9) show that

$$\sup_{t\in[0,T]} |\lambda(t) - t| \geq \sup_{t\in[0,T]} |\lambda^{-1}(t) - t|.$$

It follows that $\sup_{t\in[0,T]} |\lambda^{-1}(t) - t| = \sup_{t\in[0,T]} |\lambda(t) - t|$.

Claim (4): We must show that

$$\sup_{t\in[0,T]} |f(\lambda(t)) - g(t)| = \sup_{t\in[0,T]} |f(t) - g(\lambda^{-1}(t))|, \quad f, g \in \mathbb{D}([0,T];\mathbb{R}).$$

First, define

$$h(t) = f(\lambda(t)) - g(t), \quad t \in [0,T]$$

Now, since the domain of h is $[0,T]$ we have that $h(\lambda^{-1}(t))$ is well-defined for each t. Similarly to the argument in Claim 2 we have that

$$\forall t' \in [0,T] \exists t \in [0,T] : h(t') = h(\lambda^{-1}(t)).$$

Hence,

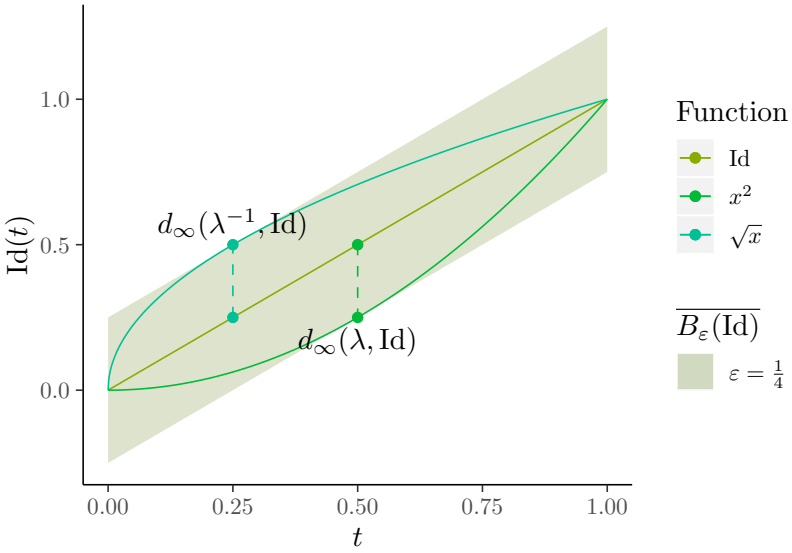

Figure 2.4: Visualising the shared value of the supremum between time-changes in $\Lambda([0,1])$.

$$\sup_{t\in[0,T]} |h(t)| = \sup_{t\in[0,T]} |h(\lambda^{-1}(t))|,$$

but we have

$$|h(\lambda^{-1}(t))| = |f(t) - g(\lambda(t))|,$$

which proves the claim. ■

The claims in Lemma 2.1 feel quite natural since they correspond to shifting the surface in an invertible way. The arguments can be extended visually to some extend. We provide a brief example.

Example 2.4. Consider $\Lambda([0,1])$ and the functions

$$\lambda(x) = x^2$$
$$\lambda^{-1}(x) = \sqrt{x}.$$

The property (3) of Lemma 2.1 is visualised in Figure 2.4. The intuition here is that λ and its inverse are both equally far away from the identity.

Next, we consider the following functions

$$f(x) = \sin(12x) + x$$
$$g(x) = \mathbb{1}_{[1/2,1]}(x).$$

We visualise property (4) of Lemma 2.1 in Figure 2.5. The intuition of property (4) is that we may either time change f to fit g or time change g to fit f with the inverse time change.

□

Using the space $\Lambda([0,\infty))$ we can make small deformations in time which will essentially allow us to align the discontinuities so that the metric of uniform convergence is "usable" again. Recall Example 2.3 where we encountered that the metric of uniform convergence would

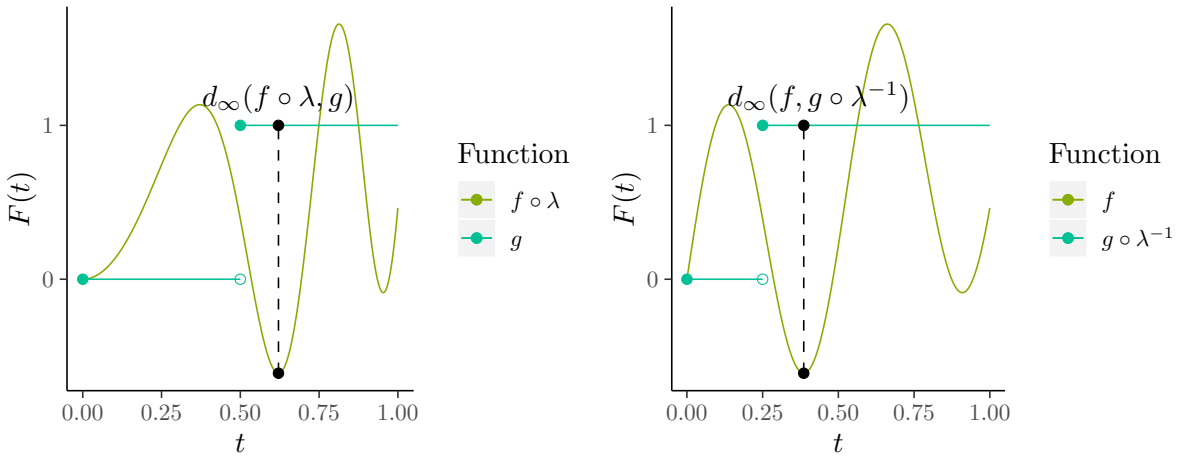

Figure 2.5: Visualising $d_\infty(f \circ \lambda, g) = d_\infty(f, g \circ \lambda^{-1})$.

eventually require our approximation to share at least the discontinuities of the function we are approximating. By deforming time slightly, we can align the discontinuities even if they do not happen at the same time. However, we are not interested in deforming time in the limit. Therefore, we wish to find a sequence of time changes such that

$$f_n \circ \lambda_n \to f$$

uniformly in t and

$$\lambda_n \to \mathrm{Id}$$

uniformly in t. Now, initially we consider the space $\mathbb{D}([0,T]; E)$. We have the following lemma:

Lemma 2.2. *Let (E, d_E) be a Polish space. The space $\mathbb{D}([0,T]; E)$ can be endowed with a metric*

$$d_S(f,g) = \inf_{\lambda \in \Lambda([0,T])} \{d_\infty(\lambda, Id) \vee d_\infty(f, g \circ \lambda)\}.$$

Proof. Clearly, $d_S \geq 0$ and $f = g$ clearly implies $d_S(f,g) = 0$. To show the converse, let for that $f, g \in \mathbb{D}([0,T], \mathbb{R})$ and $d_S(f,g) = 0$. In this case note that we must then have $d_\infty(\lambda, \mathrm{Id}) = 0$ and $d_\infty(f, g \circ \lambda) = d_\infty(f,g) = 0$. By the norm properties of d_∞ we have that $f = g$.

To see why $d_S(x,y) = d_S(y,x)$ first note that λ is invertible with $\lambda^{-1} \in \Lambda([0,T])$ by Lemma 2.1. Then take $d_\infty(f, g \circ \lambda) = d_\infty(f \circ \lambda^{-1}, g)$, intuitively this happens since we can either time change g to fit f or time change f to fit g with the reverse time change. Clearly $d_\infty(f \circ \lambda, g) = d_\infty(f, g \circ \lambda^{-1}) = d_\infty(g \circ \lambda^{-1}, f)$ and similarly $d_\infty(\lambda, \mathrm{Id}) = d_\infty(\mathrm{Id}, \lambda^{-1}) = d_\infty(\lambda^{-1}, \mathrm{Id})$. It follows that $d_S(f,g) = d_S(g,f)$.

We remain to show the triangle inequality, i.e.

$$d_S(f,g) \leq d_S(f,h) + d_S(h,g).$$

To this end, observe that if $\lambda_1 \in \Lambda([0,T])$ and $\lambda_2 \in \Lambda([0,T])$, then $\lambda_2 \circ \lambda_1 \in \Lambda([0,T])$.

Furthermore, we have

$$d_\infty(\lambda_2 \circ \lambda_1, \mathrm{Id}) \leq d_\infty(\lambda_1, \mathrm{Id}) + d_\infty(\lambda_2, \mathrm{Id}),$$

since

$$\begin{aligned} d_\infty(\lambda_2 \circ \lambda_1, \mathrm{Id}) &= d_\infty(\lambda_2, \lambda_1^{-1}) \\ &\leq d_\infty(\lambda_2, \mathrm{Id}) + d_\infty(\mathrm{Id}, \lambda_1^{-1}) \\ &= d_\infty(\lambda_2, \mathrm{Id}) + d_\infty(\lambda_1, \mathrm{Id}) \end{aligned}$$

and similarly we have

$$d_\infty(x, y \circ \lambda_2 \circ \lambda_1) \leq d_\infty(x, z \circ \lambda_1) + d_\infty(z, y \circ \lambda_2),$$

from which the triangle inequality follows. ■

Let us interpret the metric d_S. First, let x and y be cádlág functions. Then,

$$d_S(x, y) = \inf_{\lambda \in \Lambda([0,T]} \{d_\infty(\lambda, \mathrm{Id}) \vee d_\infty(x, y \circ \lambda)\}.$$

The term

$$d_\infty(\lambda, \mathrm{Id}), \tag{2.13}$$

describes the deformation of time. The second term

$$d_\infty(x, y \circ \lambda) \tag{2.14}$$

describes the largest difference between our functions under a time change. Finally, we choose the time-change that makes the maximum of Equations 2.13 and 2.14 as small as possible.

One can show that d_S induces a separable topological space, but it is not complete. So why introduce it? Well, it turns out that $\mathbb{D}([0,T];\mathbb{R})$ may be endowed with a metric which is equivalent to the metric d_S for which $\mathbb{D}([0,T];\mathbb{R})$ is a Polish metric space.

The reader may ask, why not just complete d_S? A typical approach to simply "add" the Cauchy sequences that do not converge in $(\mathbb{D}([0,T];R), d_S)$ and obtain an extension $(\overline{\mathbb{D}([0,T];\mathbb{R})}, \overline{d_S})$, similar to how $C([0,T];\mathbb{R})$ can be completed using the L^p metrics to obtain the L^p spaces. The difference here is that, we are interested in preserving cádlág properties of the elements of $\mathbb{D}([0,T];R)$ and such an extension may not preserve the property.

We now return to the question at hand. Consider

$$\|\lambda\|^\circ = \sup_{0 \leq s < t \leq T} \left| \log\left(\frac{\lambda(t) - \lambda(s)}{t - s} \right) \right|,$$

that is, if the slope of λ is close to 1 everywhere, then $\|\lambda\|^\circ = 0$. The point here is, that $\|\cdot\|^\circ$ characterises deformations in time which are not too excessive and where the secant lines do not "explode". Now, $\|\lambda\|^\circ$ may be infinite. However, when we consider

$$d^\circ(x, y) = \inf_{\lambda \in \Lambda([0,T])} \{\|\lambda\|^\circ \vee d(x, y \circ \lambda)\}$$

we force finiteness since we are taking the infimum. Suppose $\lambda \in \Lambda([0,T])$ has $\|\lambda\|^\circ = \infty$, then clearly we have $d_\infty(x, y) < \|\lambda\|^\circ = \infty$ and so, in this case, we see that choosing the identity function then yields

$$d^\circ(x, y) \leq \|\mathrm{Id}\|^\circ \vee d_\infty(x, y) = d_\infty(x, y) < \|\lambda\|^\circ \vee d_\infty(x, y \circ \lambda) = \infty.$$

We now have the following relationship between d° and d_S.

Proposition 2.5. *The space $\mathbb{D}([0,T],\mathbb{R})$ can be endowed with the metric*

$$d^\circ(f,g) = \inf_{\lambda\in\Lambda([0,T])}\{\|\lambda\|^\circ \vee d(x, y\circ\lambda)\}$$

and we have $(\mathbb{D}([0,T],\mathbb{R}), d_S)$ is topologically equivalent to $(\mathbb{D}([0,T],\mathbb{R}), d^\circ)$.

Remark: We denote the topology induced by either d_S or d° as the *Skorokhod topology.* When we consider convergence in the Skorokhod topology we refer to the metric d°.

We will not prove Proposition 2.5 but refer to ([4] pp. 125-127 and Theorem 2).

Theorem 2.5. *The space $\mathbb{D}([0,T];\mathbb{R})$ is a Polish space under the Skorokhod topology. Specifically, $\mathbb{D}([0,T];d^\circ)$ is a Polish metric space.*

We refer the reader to [[4], p. 138 Theorem 12.2] for a proof. Similar to the space $C([0,\infty);\mathbb{R})$ we cover the space $\mathbb{D}([0,\infty);\mathbb{R})$ in Appendix A.2.

The Skorokhod topology is in general not very nice to work with; finding the infimum over $\Lambda([0,\infty)$ is no simple task. It is therefore convenient to characterise convergence. We consider the following proposition.

Proposition 2.6. *A sequence $(x_n)_{n\in\mathbb{N}} \subset \mathbb{D}([0,T],\mathbb{R})$ converges to $x \in \mathbb{D}([0,T];\mathbb{R})$ in the Skorokhod topology if, and only if, there exists a sequence $(\lambda_n)_{n\in\mathbb{N}} \subset \Lambda([0,T])$ such that*

$$\lim_{n\to\infty} d_\infty(\lambda_n, Id) = 0 \text{ and } \lim_{n\to\infty} d_\infty(x_n\circ\lambda_n, x) = 0.$$

Where d_∞ is the metric of topology of uniform convergence.

See [[14], pp.328-330] for a proof. The Skorokhod feels more natural when considered in light of Proposition 2.6, which is also why we chose to introduce it in this way. A simple but consequential property is now presented in Corollary 2.2.

Corollary 2.2. *Let $(x_n)_{n\in\mathbb{N}} \subset \mathbb{D}([0,\infty);\mathbb{R})$ be a sequence, and suppose $x_n \to x$ in the topology of uniform convergence. Then $x_n \to x$ in the Skorokhod topology.*

Proof. Choose $\lambda_n = \mathrm{Id}, \quad \forall n \in \mathbb{N}$ and apply Proposition 2.6. ■

Hence, we have established a link between the uniform topology and the Skorokhod topology. The converse of Corollary 2.2 is not true in general. Consider Example 2.5.

Example 2.5. We consider the space $\mathbb{D}([0,2];\mathbb{R})$. We wish to approximate $\mathbb{1}_{[1,2]}$ by $\mathbb{1}_{\left[1+\frac{1}{2^n},2\right]}$. We first consider the topology of uniform convergence. For each $n \in \mathbb{N}$ we have

$$\sup_{t\in[0,2]}\left|\mathbb{1}_{[1,2]}(t) - \mathbb{1}_{\left[1+\frac{1}{2^n},2\right]}(t)\right| = 1,$$

since $\mathbb{1}_{[1,2]}(1) = 1$ while $\mathbb{1}_{\left[1+\frac{1}{2^n},2\right]}(1) = 0$. It follows that

$$\lim_{n\to\infty}\sup_{t\in[0,2]}\left|\mathbb{1}_{[1,2]}(t) - \mathbb{1}_{\left[1+\frac{1}{2^n},2\right]}(t)\right| = \lim_{n\to\infty} 1 = 1.$$

Hence, the sequence does not converge uniformly. However, it seems counter-intuitive that this does not converge as the graphs become seemingly indistinguishable as seen in Figure 2.6.

However, if we can find a time change λ_n for each n such that $\lambda_n \to \text{Id}$ uniformly in t and $\mathbb{1}_{[1+1/n,2]}(\lambda_n(t))$ goes to $\mathbb{1}_{[1,2]}(t)$ uniformly in t then we have convergence in the Skorokhod topology,

To this end define

$$\lambda_n(t) = \begin{cases} \left(1 + \frac{1}{2^n}\right) t & t \in [0,1] \\ \left(2 - \left(1 + \frac{1}{2^n}\right)\right)(t-1) + \left(1 + \frac{1}{2^n}\right) & t \in (1,2] \end{cases}.$$

Clearly λ_n is injective, surjective and strictly increasing. Notice that $\sup_{t\in[0,2]} |\lambda_n(t) - t| = \frac{1}{2^n}$, so clearly λ_n converges uniformly to the identity in Figure 2.7.

Furthermore, we have that $\mathbb{1}_{[1+1/n,2]}(\lambda_n(t)) = \mathbb{1}_{[1,2]}(t)$ so clearly uniform convergence is obtained. Hence, we have convergence in the Skorokhod topology.

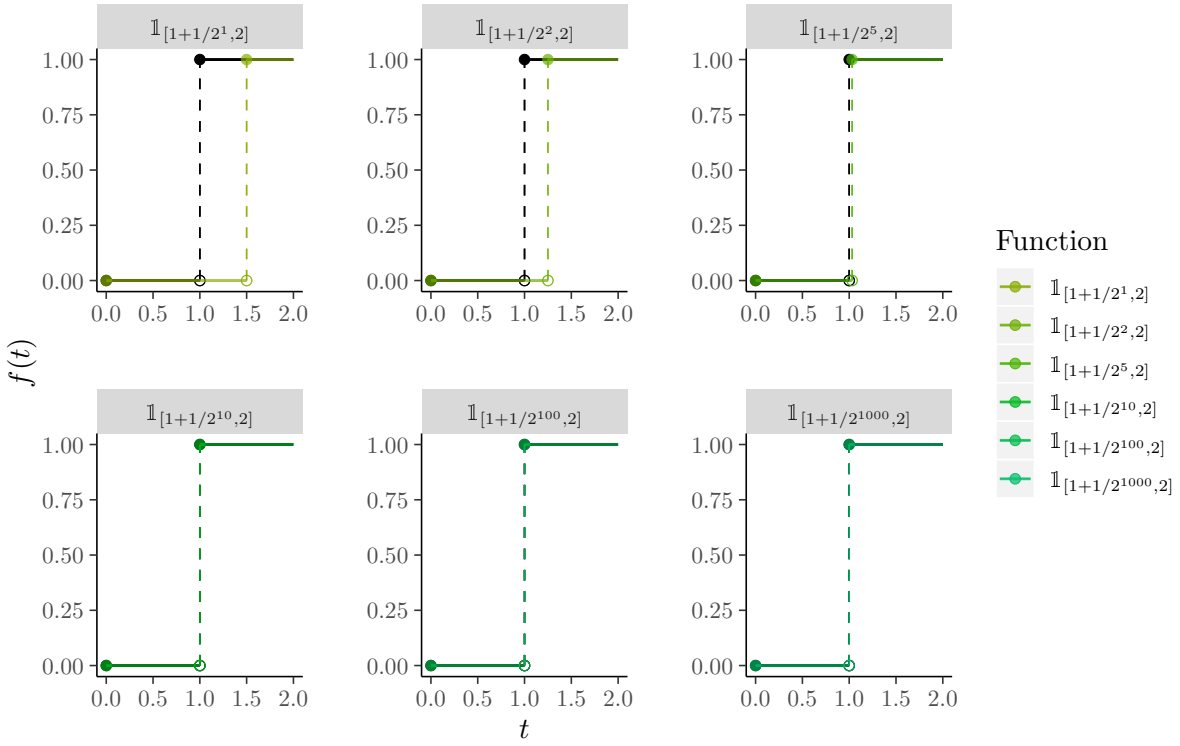

Figure 2.6: Convergence of Indicator Functions

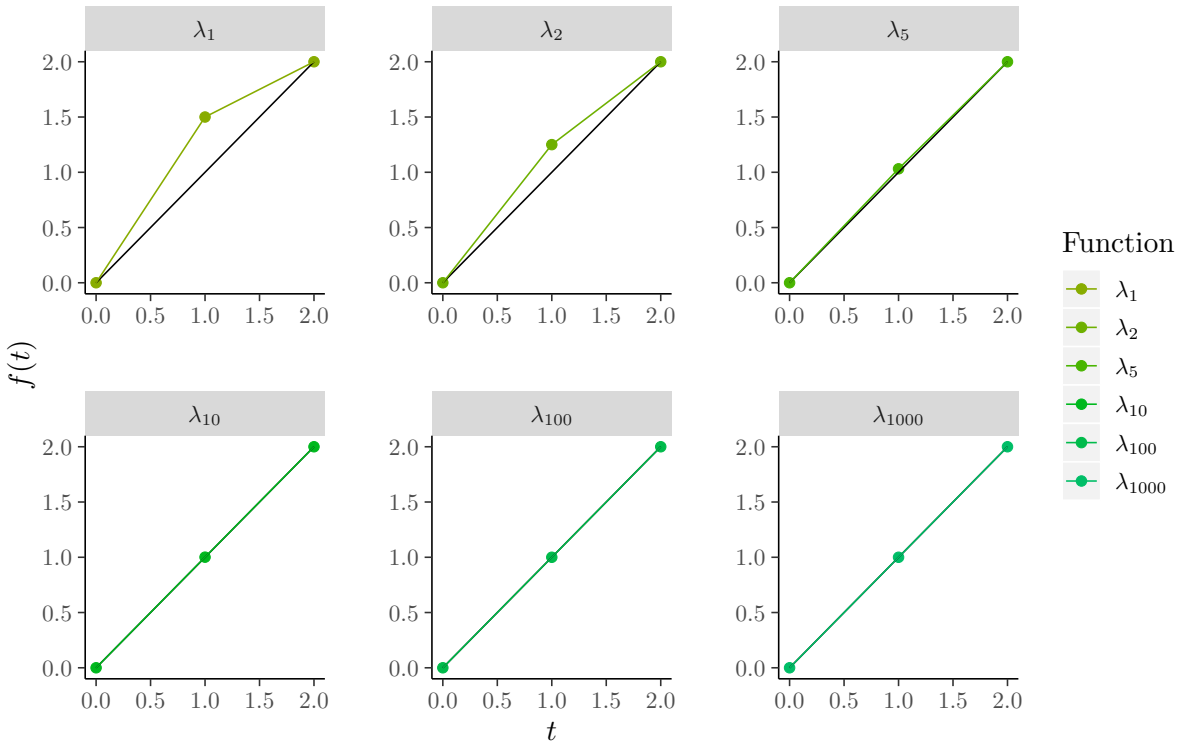

Figure 2.7: Convergence of Time Changes

□

So far it seems that the Skorokhod topology is superior to the uniform topology in the sense that any convergent sequence in the uniform topology is also convergent in the Skorokhod topology while allowing for convergence of additional functions. However, it does come at a cost; while the uniform topology induces a non-separable Banach space and in particular a topological vector space, the Skorokhod topology does not induce a topological vector space. We provide an example.

Example 2.6. Let $(x_n)_{n\in\mathbb{N}}, (y_n)_{n\in\mathbb{N}}$ be sequences in $\mathbb{D}([0,2];\mathbb{R})$ defined by

$$x_n(t) = \mathbb{1}_{[1+1/2^n,2]}(t)$$
$$y_n(t) = \mathbb{1}_{[1-1/2^n,2]}(t)$$

Clearly, $y_n, x_n \to \mathbb{1}_{[1,2]}$ in the Skorokhod topology. However, consider $x_n + y_n$. By virtue of Proposition 2.6 we need to find $(\lambda_n)_{n\in\mathbb{N}} \subset \Lambda([0,2])$ such that $\lambda_n \to \mathrm{Id}$ uniformly in t and

$$(x_n + y_n)(\lambda_n(t)) \to 2 \cdot \mathbb{1}_{[1,2]}(t)$$

uniformly in time. We claim that no such sequence can exist. Note that unless $(x_n+y_n)(\lambda_n(t)) = 2\mathbb{1}_{[1,2]}(t)$ then $x_n + y_n$ cannot approach $2\mathbb{1}_{[1,2]}$ since

$$\sup_{t\in[0,2]} |(x_n + y_n)(\lambda_n(t)) - 2\mathbb{1}_{[1,2]}(t)| < 1 \implies x_n(\lambda_n(t)) = y_n(\lambda_n(t)) = \mathbb{1}_{[1,2]}(t), \quad \forall t \in [0,2]$$

However, this means $\mathbb{1}_{[1+1/2^n,2]}(\lambda_n(t)) = \mathbb{1}_{[1-1/2^n,2]}(\lambda_n(t)) = \mathbb{1}_{[1,2]}(t)$ for all t. The result is that we must have $\lambda_n(1) = 1 - 1/2^n$ and $\lambda_n(1) = 1 + 1/2^n$. Clearly, we then have λ_n is not a function

and hence cannot belong to $\Lambda([0,2])$. Even if we made a cádlág λ_n with $\lim_{s\uparrow 1}\lambda_n(s) = 1 - 1/2^n$ and $\lambda_n(1) = 1 + 1/2^n$ then it would have to be discontinuous in $t = 1$, hence it cannot be surjective and it would not belong to $\Lambda([0,2])$. □

2.3.3 Multivariate Skorokhod Spaces

This subsection is based on [18] and [25].

We have introduced arguably the most intuitive Skorokhod space, namely $\mathbb{D}([0,T];\mathbb{R})$. Now, the definition of continuity extends naturally to multivariate spaces; we may replace absolute values with norms or metrics. However, for cádlág functions, the extension requires slightly more care. In particular, one may think of cádlág function, whose domain is some interval $\mathbb{R}$, as continuous from above with limits from below. When we try to extend to a multivariate setting, there is no clear definition of continuity from "above" and limits from "below". Of course, the product space $\bigtimes_{i=1}^{d}\mathbb{D}([0,T];\mathbb{R})$ is a multivariate Skorokhod space, and since it is a countable product of Polish spaces it is also a Polish space. However, the product space is not sufficient for our purposes.

In order to construct an extension which preserves the aforementioned structure, we must first define what "above" and "below" means in a multivariate setting.

Definition 2.6 (Quadrant)
Let $\mathcal{R} = \{<, \geq\}$ where $<$ and $\geq$ are the typical relations on $\mathbb{R}$. Let $d \in \mathbb{N}$ and $R = (R_1, R_2, \ldots, R_d) \in \mathcal{R}^d = \bigtimes_{i=1}^{d}\mathcal{R}$. We define the R quadrant as the map

$$\begin{aligned} &Q_R : \mathbb{R}^d \to \mathcal{P}(\mathbb{R}^d) \\ &t \mapsto \{s \in \mathbb{R}^d \mid s_i R_i t_i, i \in \{1,2,\ldots,d\}\}, \end{aligned}$$

where $\mathcal{P}(\mathbb{R}^d)$ denotes the power-set of $\mathbb{R}^d$.

Remark: It is worthwhile noting that Q_R by construction maps to Borel-sets of $\mathbb{R}^d$. Furthermore, it is easy to verify that $\mathcal{R}^d$ defines a set of partial orders on $\mathbb{R}^d$.

We provide a brief example.

Example 2.7. Let $R = (<, \geq)$, and let $t \in \mathbb{R}^2$. Then

$$Q_R(t) = \{s \in \mathbb{R}^2 \mid s_1 < t_1, s_2 \geq t_2\}.$$

We visualise in Figure 2.8

□

We are now ready to define the multivariate Skorokhod spaces.

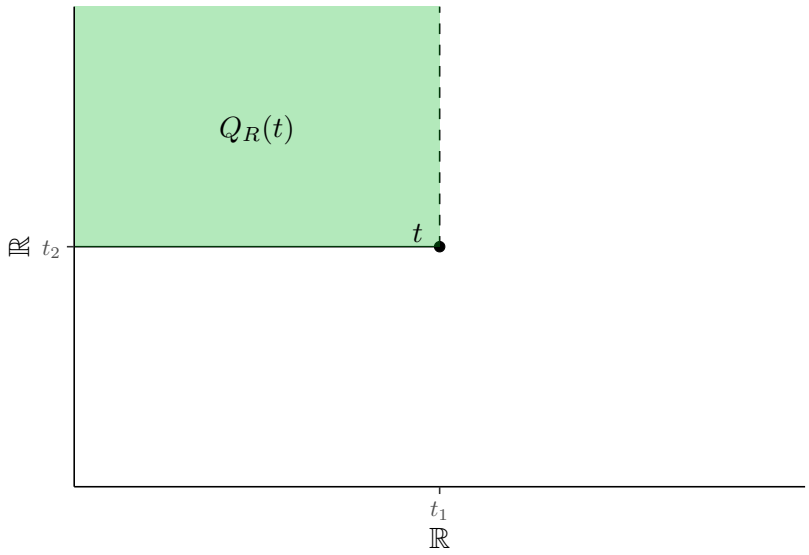

Figure 2.8: The quadrant $Q_R(t)$.

Definition 2.7 (Multivariate Cádlág and Skorokhod Space)
Let (E, d) be a metric space. Let $d \in \mathbb{N}$ and $\mathcal{T} \subseteq \mathbb{R}^d$ be of the form $\bigtimes_{i=1}^d [0, T_i]$ with $T_i < \infty$ for all $i \in \{1, 2, \ldots, d\}$ or $\mathcal{T} = \bigtimes_{i=1}^d [0, \infty)$. A function $f : \mathcal{T} \to E$ is said to be cádlág if

$$\lim_{\substack{s \to t \\ s \in Q_R(t)}} f(s) \text{ exists}, \quad \forall R \in \mathcal{R}^d,$$

and

$$\lim_{\substack{s \to t \\ s \in Q_R(t)}} f(s) = f(t), \quad R = (\geq, \geq, \ldots, \geq).$$

We denote the space of cádlág function $\mathbb{D}(\mathcal{T}; E)$ and call it *the Skorokhod space* over $\mathcal{T}$ mapping to E.

Remark: To ease notation we will write $\mathbb{D}([0, T]^d, E)$ when $\mathcal{T}$ is of the form $\bigtimes_{i=1}^d [0, T]$. We will also primarily assume multivariate Skorokhod spaces of this type.

The quadrant $Q_R(t)$ with $R = (\geq, \geq, \ldots, \geq) \in \mathcal{R}^d$ can be thought of as the points above the point $t \in \mathbb{R}^d$ in every dimension, and every other quadrant defines the points below t. Relating this to the univariate definition it preserves the property of "limits from below" and "continuous from above".

Let (E, d_E) be a Polish space. We are now interested in endowing $\mathbb{D}([0, T]^d; E)$ with a metric. First, we note that $\mathbb{D}([0, T]^d; E)$ may be endowed with the supremum metric

$$d_\infty(x, y) = \sup_{t \in [0,T]^d} d_E(x(t), y(t)), \quad x, y \in \mathbb{D}([0, T]^d; E).$$

In a very similar fashion, it may also be endowed with the Skorokhod metric. In order to define the Skorokhod metric we need to extend $\|\lambda\|^\circ$ to the multivariate case. In order to do so, we make use of the product space $\bigtimes_{i=1}^d \Lambda([0, T]) = \Lambda([0, T])^d$. For $\lambda \in \Lambda([0, T])^d$ we define

$$\|\lambda\|^\circ = \sup_{\substack{0\le s<t<\infty \\ s,t\in[0,\infty)^d}} \left|\left|\left(\log\left(\frac{\lambda_1(t_1)-\lambda_1(s_1)}{t_1-s_1}\right), \log\left(\frac{\lambda_2(t_2)-\lambda_2(s_2)}{t_2-s_2}\right), \ldots, \log\left(\frac{\lambda_d(t_d)-\lambda_d(s_d)}{t_d-s_d}\right)\right)\right|\right|,$$

with the norm being the Euclidean norm on $\mathbb{R}^d$. We may then define the Skorokhod metric on $\mathbb{D}([0,T]^d; E)$ as

$$d^\circ(x,y) = \inf_{\lambda\in\Lambda([0,T])^d}\{\|\lambda\|^\circ \vee d_\infty(x, y\circ\lambda)\}, \quad x,y\in\mathbb{D}([0,T]^d;E),$$

where $y\circ\lambda(t) = y(\lambda_1(t_1), \lambda_2(t_2), \ldots, \lambda_d(t_d))$ for $t\in[0,T]^d$. We claim without proof that the space $(\mathbb{D}([0,T]^d;\mathbb{R}^m), d_\infty)$ is complete and $(\mathbb{D}([0,T]^d;\mathbb{R}^m), d^\circ)$ is a Polish metric space for $d,m\in\mathbb{N}$. We will not prove this claim but refer to [18] or [25].

We also claimed that that the product space $\bigtimes_{i=1}^d \mathbb{D}([0,T];\mathbb{R}) = \mathbb{D}([0,T];\mathbb{R})^d$ could be considered as multivariate Skorokhod spaces. There are two ways to think of this product space; one where the evaluation is considered to at the same $t\in[0,T]$, i.e. $z(t) = (z_1(t), z_2(t), \ldots, z_d(t))$. The other way of considering the product space is

$$z(t) = (z_1(t_1), z_2(t_2), \ldots, z_d(t_n)), \quad t\in[0,T]^n, z\in\mathbb{D}([0,T];\mathbb{R})^d. \tag{2.15}$$

In a sense we may think of (2.15) as a "field" rather than a path. We will only consider elements of $\mathbb{D}([0,T];\mathbb{R})^d$ as described in (2.15). When we consider the representation (2.15) then the natural metrics are defined to correspond to the product space, i.e.

$$d_\infty^m(x,y) = \sum_{i=1}^m d_\infty(x_i,y_i), \quad x,y\in\mathbb{D}([0,T];\mathbb{R})^m,$$
$$d_m^\circ(x,y) = \sum_{i=1}^m d^\circ(x_i,y_i), \quad x,y\in\mathbb{D}([0,T];\mathbb{R})^m.$$

for $m\in\mathbb{N}$. It should be no surprise that the space $\mathbb{D}([0,T];\mathbb{R})^m$ and $\mathbb{D}([0,T]^m;\mathbb{R}^m)$ are related for $m\in\mathbb{N}$. To see why, let $x\in\mathbb{D}([0,T];\mathbb{R})^m$, we now claim $x\in\mathbb{D}([0,T]^m;\mathbb{R}^m)$. We must show that

$$\lim_{\substack{s\to t \\ s\in Q_R(t)}} x(s) \text{ exists}, \quad \forall R\in\mathcal{R}^m,$$

and

$$\lim_{\substack{s\to t \\ s\in Q_R(t)}} x(s) = x(t), \quad R = (\ge,\ge,\ldots,\ge).$$

Note that R decomposes into $(R_1, R_2, \ldots, R_m)$, with $R_i\in\mathcal{R} = \{<,\ge\}$ for $i\in\{1,2,\ldots,m\}$. Clearly,

$$\lim_{\substack{s\to t \\ s\in Q_R(t)}} x(s) \text{ exists} \iff \lim_{\substack{s_i\to t_i \\ s_i\in Q_{R_i}(t_i)}} x_i(s_i) \text{ exists}, \quad \forall i\in\{1,2,\ldots m\}.$$

But since $R_i\in\{<,\ge\}$ we have the two options

$$\lim_{\substack{s_i\to t_i \\ s_i\in Q_{R_i}(t_i)}} x_i(s_i) = \begin{cases} \lim_{s_i\uparrow t_i} x_i(s_i) & R_i = < \\ \lim_{s_i\downarrow t_i} x_i(s_i) & R_i = \ge, \end{cases}$$

and since $x_i \in \mathbb{D}([0,T];\mathbb{R})$ both limits clearly exist. Exactly the same arguments show that

$$\lim_{\substack{s\to t \\ s\in Q_R(t)}} x(s) = x(t), \quad R = (\geq, \geq, \ldots, \geq),$$

and hence $x \in \mathbb{D}([0,T]^d;\mathbb{R}^d)$. With this result we can show the more general case.

Lemma 2.3. *Let $m, n \in \mathbb{N}$ and $g \in C(\mathbb{R}^n;\mathbb{R}^m)$ and let $x = (x_1, x_2, \ldots, x_n) \in \mathbb{D}([0,T];\mathbb{R})^n$. Then,*

$$z = g \circ x \in \mathbb{D}([0,T]^n;\mathbb{R}^m).$$

Proof. The proof follows directly from the continuity of g since

$$\lim_{\substack{s\to t \\ s\in Q_R(t)}} g(x(s)) = g\left(\lim_{\substack{s\to t \\ s\in Q_R(t)}} x(s)\right), \quad \forall R \in \mathcal{R}^n$$

■

In light of Lemma 2.3 it seems natural to *extend* g from a function on $\mathbb{R}^n$ to a function on the product space. That is, g defines a map

$$\begin{aligned} \Psi_g : \mathbb{D}([0,T];\mathbb{R})^n &\to \mathbb{D}([0,T]^n;\mathbb{R}^m), \\ x &\mapsto g \circ x. \end{aligned}$$

Lemma 2.3 justifies the image of Ψ_g when g is continuous. The next natural question is then whether Ψ_g extends to a continuous function on on $\bigtimes_{i=1}^d \mathbb{D}([0,T];\mathbb{R})$ endowed with d_∞^m or d_m° to $\mathbb{D}([0,T]^m;\mathbb{R}^n)$ endowed with either d_∞ or d°.

2.3.4 Continuous Extensions

A version similar to the theorems presented here is found in [[39], Appendix B, Theorem B.2.5] without proof.

Often we are not only interested in the behaviour of our approximation, but also how it acts under certain functions. Continuous functions are a class of functions that preserve limits which is justified in Proposition A.1. Our interest in doing so is to recover convergence for a large class of processes. Suppose we have

$$x_n \to x, \quad (x_n)_{n\in\mathbb{N}} \subset \mathbb{D}([0,T];\mathbb{R}), x \in \mathbb{D}([0,T];\mathbb{R}).$$

We realise that proving everything case-by-case is cumbersome and recognise that a large class of functions are created from compositions. Hence, we are interested in knowing if

$$x_n \to x \implies g \circ x_n \to g \circ x.$$

We know that continuous functions are a class of functions which have exactly this property. However, the space $C(\mathbb{D}([0,T];\mathbb{R});\mathbb{D}([0,T];\mathbb{R}))$ is not as well studied as the space $C(\mathbb{R};\mathbb{R})$. In this section we show that under suitable conditions we may identify $C(\mathbb{R};\mathbb{R})$ in $C(\mathbb{D}([0,T];\mathbb{R}),\mathbb{D}([0,T];\mathbb{R}))$ by extending continuous functions on $\mathbb{R}$ to a continuous function on $\mathbb{D}([0,T];\mathbb{R})$. We first require the following results

It is well known that every cádlág function is bounded on compact sets. Note that this result has previously been used implicitly; a metric is a mapping that must be finite for each pair in the underlying space and therefore d_∞ could not be a metric if cádlág function were not bounded on compact sets, since $d_\infty(x,0) = \sup_{t\in[0,T]} |x(t)| < \infty$.

Theorem 2.6 (Heine-Cantor Theorem)**.** *Let $f \in C(E', E)$ with (E, d_E) and $(E', d_{E'})$ metric spaces, and let $K \subseteq E'$ be compact, then f is uniformly continuous on K.*

We now provide the first extension. While this extension is not particularly useful for our problem it shows a particular strategy to prove more general cases.

In order to emphasize the metric, or equivalently the topology, we write

$$\Psi : (X, d_X) \to (Y, d_Y),$$

to imply that Ψ is a continuous function from the metric space (X, d_X) to the metric space (Y, d_Y).

Lemma 2.4. *Let $g \in C(\mathbb{R};\mathbb{R})$. Define the function*

$$\begin{aligned} \Psi_g : (\mathbb{D}([0,T];\mathbb{R}), d_\infty) &\to (\mathbb{D}([0,T];\mathbb{R}), d_\infty). \\ x &\mapsto g \circ x \end{aligned}$$

The function Ψ_g is continuous in the topology of uniform convergence.

Proof. We must show, that

$$\forall x \in \mathbb{D}([0,T];\mathbb{R}) \forall \varepsilon > 0 \exists \delta > 0 \forall y \in \mathbb{D}([0,T];\mathbb{R}) : d_\infty(x,y) < \delta \implies d_\infty(\Psi_g(x), \Psi_g(y)) < \varepsilon.$$

The proof is in two parts: First, we construct a compact set for each $x \in \mathbb{D}([0,T];\mathbb{R})$ where the restriction of g is uniformly continuous. Second, we extend g and show that it is indeed continuous for each $x \in \mathbb{D}([0,T];\mathbb{R})$.

(i) Construction of a compact set

Let $x \in \mathbb{D}([0,T];\mathbb{R})$ be arbitrary and $\varepsilon > 0$. Since cádlág functions are bounded on compact intervals, we may choose a constant which only depends on x such that

$$\sup_{t\in[0,T]} |x(t)| < C_x < \infty, \quad \forall x \in \mathbb{D}([0,T];\mathbb{R}).$$

We then define the following sphere

$$\overline{B_{C_x}(x)} = \{y \in \mathbb{D}([0,T];\mathbb{R}) \mid d_\infty(x,y) \le C_x\}.$$

For $y \in \overline{B_{C_x}(x)}$ we have

$$\left| \sup_{t\in[0,T]} |y(t)| - \sup_{t\in[0,T]} |x(t)| \right| \le \sup_{t\in[0,T]} |y(t) - x(t)| = d_\infty(y,x) \le C_x$$

by the reverse triangle inequality, which implies

$$\sup_{t\in[0,T]}|y(t)| \leq C_x + \sup_{t\in[0,T]}|x(t)|$$
$$< 2C_x.$$

The immediate implication is that

$$y(t) \in [-2C_x, 2C_x], \quad \forall t \in [0,T].$$

Furthermore, $x(t) \in [-C_x, C_x] \subseteq [-2C_x, 2C_x]$ for all $t \in [0,T]$. Therefore, define

$$K_x = [-2C_x, 2C_x].$$

Clearly K_x is compact.

(ii) Asserting continuity

Now, let $g\restriction_{K_x}$ denote the restriction of g to K_x. By Theorem 2.6 we have that $g\restriction_{K_x}$ is uniformly continuous. We omit writing the restriction as we will henceforth only consider the restriction. Therefore,

$$\forall\varepsilon > 0 \exists\delta_0 > 0 : |t-s| < \delta_0 \implies |g(t) - g(s)| < \varepsilon, \quad t,s \in K_x \tag{2.16}$$

Now, we have to show that

$$\forall x \in \mathbb{D}([0,T];\mathbb{R}) \forall\varepsilon > 0 \exists\delta > 0 \forall y \in \mathbb{D}([0,T];\mathbb{R}) : d_\infty(x,y) < \delta \implies d_\infty(\Psi_g(x), \Psi_g(y)) < \varepsilon.$$

We claim that $\delta = \min(\delta_0, C_x)$ is sufficient. Note that

$$d_\infty(x,y) = \sup_{t\in[0,T]}|x(t) - y(t)| \geq |x(t) - y(t)|, \quad \forall t \in [0,T]. \tag{2.17}$$

Now, by the uniform continuity of g on K_x we have

$$|x(t) - y(t)| < \delta_0 \implies |g(x(t)) - g(y(t))| < \varepsilon.$$

Taking $d(x,y) < \delta = \min(\delta_0, C_x)$, then the continuity of g and (2.17) implies

$$|g(x(t)) - g(y(t))| < \varepsilon, \quad \forall t \in [0,T].$$

Consequently, we have

$$d_\infty(\Psi_g(f), \Psi_g(h)) = \sup_{t\in[0,T]}|g(f(t)) - g(h(t))| < \varepsilon.$$

■

While subtle one of the applications of Lemma 2.4 is that convergence is preserved under continuous maps, and that this will hold even for approximations in $\mathbb{D}([0,T];\mathbb{R})$.

Before we continue we first present a technical lemma which will simplify the proofs for the Skorokhod topology significantly.

Lemma 2.5. *Let $x, y \in \mathbb{D}([0,T];\mathbb{R})$ be such that*

$$d^\circ(x,y) < \delta,$$

then there exists a $\lambda \in \Lambda([0,T])$ such that

$$(\|\lambda\|^\circ \vee d_\infty(x, y \circ \lambda)) < \delta.$$

Proof. Suppose for contradiction that

$$d^\circ(x, y) < \delta,$$

but

$$\nexists \lambda \in \Lambda([0, T]) : (\|\lambda\|^\circ \vee d_\infty(x, y \circ \lambda)) < \delta.$$

We then have

$$(\|\lambda\|^\circ \vee d_\infty(x, y \circ \lambda)) \geq \delta, \quad \forall \lambda \in \Lambda([0, T])$$

and hence certainly for the infimum, but then $\delta > d^\circ(x, y) \geq \delta$, which is a contradiction. ■

We also require the following Lemma.

Lemma 2.6. *Let $x, y \in \mathbb{D}([0, T]; \mathbb{R})$, then*

$$d^\circ(x, y) \leq d_\infty(x, y).$$

Proof. We know that $\mathrm{Id} \in \Lambda([0, T])$. Now, it follows immediately that $\|\mathrm{Id}\|^\circ = 0$. Hence,

$$\begin{aligned} d^\circ(x, y) &= \inf_{\lambda \in \Lambda([0,T])} \{\|\lambda\|^\circ \vee d_\infty(x, y \circ \lambda)\} \\ &\leq \|\mathrm{Id}\|^\circ \vee d_\infty(x, y) = d_\infty(x, y). \end{aligned}$$

■

We can now present the analog of Lemma 2.4 for the Skorokhod topology.

Lemma 2.7. *Let $g \in C(\mathbb{R}; \mathbb{R})$ be a continuous function. Define the function*

$$\begin{aligned} \Psi_g : (\mathbb{D}([0, T]; \mathbb{R}), d^\circ) &\to (\mathbb{D}([0, T]; \mathbb{R}), d^\circ). \\ x &\mapsto g \circ x \end{aligned}$$

The function Ψ_g is continuous in the Skorokhod topology.

Proof. We have to show

$$\forall x \in \mathbb{D}([0, T]; \mathbb{R}) \forall \varepsilon > 0 \exists \delta > 0 \forall y \in \mathbb{D}([0, T]; \mathbb{R}) : d^\circ(x, y) < \delta \implies d^\circ(\Psi_g(x), \Psi_g(y)) < \varepsilon.$$

Let $\varepsilon > 0$ and $x \in \mathbb{D}([0, T]; \mathbb{R})$ be arbitrary. Analogously to the proof of Lemma 2.4 we need to construct a compact set K_x on which g is uniformly continuous. By Lemma 2.6 we may choose the same compact set. The uniform continuity of g on K_x implies

$$\forall \varepsilon > 0 \exists \delta_0 > 0 : |t - s| < \delta_0 \implies |g(t) - g(s)| < \varepsilon. \quad t, s \in K_x. \tag{2.18}$$

Now, whenever

$$\delta > d_\infty(x, y \circ \lambda) = \sup_{t \in [0,T]} |x(t) - y(\lambda(t))| \geq |x(t) - y(\lambda(t))|, \quad \forall t \in [0, T]. \tag{2.19}$$

By (2.18) we have

$$|x(t) - y(\lambda(t))| < \delta_0 \implies |g(x(t)) - g(y(\lambda(t)))| < \varepsilon.$$

By (2.19) we have this is true for all $t \in [0,T]$ and then certainly also true for the supremum. Hence,

$$d_\infty(x, y \circ \lambda) < \delta_0 \implies d_\infty(g \circ x, g \circ y \circ \lambda) < \varepsilon. \tag{2.20}$$

Now, by Lemma 2.5 we have that when $d(x,y)^\circ < \min(\delta_0, C_x)$ there exists a λ^* such that $(\|\lambda^*\| \vee d_\infty(x, y \circ \lambda^*)) < \min(\delta_0, C_x)$. For the same λ^* we now have that

$$(\|\lambda^*\|^\circ \vee d_\infty(x, y \circ \lambda^*)) < \min(\delta_0, C_x) \implies d_\infty(g \circ x, g \circ y \circ \lambda^*) < \varepsilon.$$

Now, choosing $\delta < \min(\varepsilon, \delta_0, C_x)$ we have

$$(\|\lambda^*\|^\circ \vee d_\infty(x, y \circ \lambda^*)) < \delta \implies (\|\lambda^*\|^\circ \vee d_\infty(g \circ x, g \circ \lambda^*)) < \varepsilon \implies d^\circ(\Psi_g(x), \Psi_g(y)) < \varepsilon,$$

since certainly the infimum over $\lambda \in \Lambda([0,T])$ must be smaller than the value of the metric from λ^*. Hence, Ψ_g is continuous in the Skorokhod topology. ■

Suppose we now wish to define continuous maps depending on several paths. This will be detrimental to the main result.

Lemma 2.8. *Let $g \in C(\mathbb{R}^n; \mathbb{R}^m)$. Define the function*

$$\Psi_g : \left(\bigtimes_{i=1}^{n} \mathbb{D}([0,T]; \mathbb{R}), d_\infty^n \right) \to (\mathbb{D}([0,T]^n; \mathbb{R}^m), d_\infty)$$
$$(x_1, x_2, \dots, x_n) = x \mapsto g(x_1(t_1), x_2(t_2), \dots x_n(t_n)), \quad \forall (t_1, t_2, \dots t_n) \in [0,T]^n.$$

Then function Ψ_g is continuous in between the topologies generated by d_∞^n and d_∞.

Proof. The proof is completely analogous to Lemma 2.4, replacing absolute values with norms and using d_∞^n in place of d_∞. ■

2.4 Convergence of Stochastic Processes

This section is based on [4], [5], [13], and [20]

The previous theory has mainly been extending a continuous function in a way such that it "remains continuous". Our reason for doing so was to preserve convergence. The following will justify our reason for doing so.

Recall, that convergence in probability for random variables $(X_n)_{n \in \mathbb{N}}$ to a random variable X on a metric space (E, d_E) is defined by $\lim_{n \to \infty} \mathbb{P}(d_E(X_n, X) > \varepsilon) = 0$.

Suppose $X : \Omega \to E$, where (E,d) is a metric space and suppose we have $g : E \to E'$ where g is a continuous map. We are now interested in the convergence

$$g(X_n) \to g(X).$$

Luckily, convergence is preserved, by the following theorem.

Theorem 2.7 (Continuous Mapping Theorem)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let (E, d_E) and $(E', d_{E'})$ be metric spaces. Suppose $X : \Omega \to E$ is a random variable, and $(X_n)_{n\in\mathbb{N}}$ is a sequence of random variables such that $X_n : \Omega \to E$. Let $g \in C(E, E')$. Then, if*

1 $X_n \xrightarrow{a.s.} X$ we have $g(X_n) \xrightarrow{a.s.} g(X)$.

2 $X_n \xrightarrow{\mathbb{P}} X$ we have $g(X_n) \xrightarrow{\mathbb{P}} g(X)$.

Proof. First, by Proposition A.1 we have and for almost all ω

$$\lim_{n\to\infty} g(X_n(\omega)) = g(X(\omega)). \tag{2.21}$$

By definition of almost sure convergence

$$\mathbb{P}(\{\omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1,$$

but then by (2.21) we have

$$\mathbb{P}(\{\omega \mid \lim_{n\to\infty} g(X_n(\omega)) = g(X(\omega))\}) = \mathbb{P}(\{\omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1.$$

For convergence in probability fix $\varepsilon > 0$, note that by definition of continuity we have

$$\forall x \forall \varepsilon > 0 \exists \delta_0 > 0 \forall y : d_E(x, y) < \delta \implies d_{E'}(g(x), g(y)) < \varepsilon.$$

and by convergence in probability we have

$$\lim_{n\to\infty} \mathbb{P}(\{\omega \mid d_E(X_n(\omega), X(\omega)) \geq \delta\}) = 0, \quad \forall \delta > 0. \tag{2.22}$$

Now, notice that

$$d_{E'}(g(X_n(\omega)), g(X(\omega))) \geq \varepsilon \iff d_E(X_n(\omega), X(\omega)) \geq \delta_0.$$

It follows immediately by (2.22) that $g(X_n) \xrightarrow{\mathbb{P}} g(X)$.

■

We will primarily be interested in two specific types of convergence in probability. We, therefore, state a definition to establish notation.

Definition 2.8 (Convergence in Probability, Skorokhod and Uniform on Compacts)
Let $(X_n)_{n\in\mathbb{N}}$ such that $X_n : \Omega \to \mathbb{D}([0,T];\mathbb{R})$ and let $X : \Omega \to \mathbb{D}([0,T];\mathbb{R})$. We say that $(X_n)_{n\in\mathbb{N}}$ converges in the Skorokhod metric or topology in probability if

$$\lim_{n\to\infty} \mathbb{P}\left(d^{\circ}(X_n, X) \geq \varepsilon\right) = 0,$$

and we write

$$X_n \overset{Sk}{\Rightarrow} X.$$

Similarly, we say that $(X_n)_{n\in\mathbb{N}}$ convergences in the local uniform metric or topology in probability if

$$\lim_{n\to\infty} \mathbb{P}\left(d_{\infty}(X_n, X) \geq \varepsilon\right) = 0,$$

and we write

$$X_n \overset{u.c.p.}{\Rightarrow} X.$$

Remark: If we make use of $\mathbb{D}([0,\infty);\mathbb{R})$ we will use the same notation when there is no confusion.

Now, a stochastic process may be considered as a jointly $\Omega\times[0,T]$ measurable mapping. However, in many cases, there will be a process indistinguishable from the original stochastic process, which is cádlág. Therefore, it is not too restrictive to consider certain stochastic processes as measurable functions to the Skorokhod space.

It is also worth noticing that we have introduced several Skorokhod spaces, and we will sometimes go from one type to another. We will make use of the results presented in our section on continuous extensions, namely that if $(X_n, Y_n)_{n\in\mathbb{N}} \subset \mathbb{D}([0,T];\mathbb{R})^2$ which converges to (X, Y) in the product space and g is a continuous function from $\mathbb{R}^2$ to, e.g., $\mathbb{R}$, then we will still say $\Psi_g(X_n, Y_n) \overset{u.c.p.}{\Rightarrow} \Psi_g(X, Y)$. We justify this notation since the convergence will still be uniformly on compact sets in probability.

We now present a weaker type of convergence. First, recall the definition of convergence in law. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, (E, d_E) a Polish space with Borel σ-field $\mathcal{B}(E)$, and let $((\Omega_n, \mathcal{F}_n, \mathbb{P}_n))_{n\in\mathbb{N}}$ be a sequence of probability spaces. Let $Z_n : \Omega_n \to E$ be a sequence of random variables. $(Z_n)_{n\in\mathbb{N}}$ is said to converge in law if there exists a probability measure μ on $(E, \mathcal{B}(E))$ such that

$$\mathbb{E}_n[f(Z_n)] \to \int_E f(x)d\mu(x), \quad \forall f \text{ continuous and bounded on } E.$$

We can think of the limit as a random variable Z with $\mathcal{L}_Z = \mu$, hence the name. In particular, we can choose $Z : E \to E$ being the identity on $(E, \mathcal{B}(E), \mu)$ and we write

$$Z_n \overset{\mathcal{L}}{\to} Z.$$

However, for our purposes, we need a type of convergence related to convergence in law. First, we will define what we mean by an extension of our probability space. We also define some stronger properties needed for later theorems.

Let $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,\infty)}, \mathbb{P}\right)$ be filtered probability space and $(\Omega', \mathcal{F}')$ another measurable space. Let

$\mathbb{Q}$ be a transition probability from $(\Omega, \mathcal{F})$ to $(\Omega', \mathcal{F}')$, i.e. $\mathbb{Q}(\omega, \cdot)$ is a probability measure on $(\Omega', \mathcal{F}')$ for $\omega \in \Omega$ and $\mathbb{Q}(\cdot, F')$ is a measurable function on $(\Omega, \mathcal{F})$ for $F' \in \mathcal{F}'$, in particular recall the definition of a conditional probability which makes use of transition probabilities. We can now extend our probability space with Ω' by defining

$$\begin{aligned}
\widetilde{\Omega} &= \Omega \times \Omega' \\
\widetilde{\mathcal{F}} &= \mathcal{F} \otimes \mathcal{F}' = \sigma(\{F \times F' \mid F \in \mathcal{F}, F' \in \mathcal{F}'\}) \\
\widetilde{\mathbb{P}} &= \mathbb{P}(d\omega)\mathbb{Q}(\omega, d\omega'),
\end{aligned}$$

where the latter is to be understood in terms of the integrals. Similarly, we may need to extend our filtration. To this end we call a filtration $(\widetilde{\mathcal{F}}_t)_{t\in[0,\infty)}$ on $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ an extension of $(\mathcal{F}_t)_{t\in[0,\infty)}$ if $\mathcal{F}_t \subseteq \widetilde{\mathcal{F}}_t$ for every t where the subset property is to be understood by the inclusion $\mathcal{F}_t \otimes \{\emptyset, \Omega'\}$. We then say $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, (\widetilde{\mathcal{F}}_t)_{t\in[0,\infty)}, \widetilde{\mathbb{P}})$ is a filtered extension of $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,\infty)}, \mathbb{P}\right)$. The filtered extension is said to be *very good* if the mapping

$$\omega \mapsto \int_{\Omega'} \mathbb{1}_A(\omega, \omega')\mathbb{Q}(\omega, d\omega')$$

is measurable for every ω and $A \in \widetilde{\mathcal{F}}_t$ given any $t \in [0, \infty)$. In particular, very good filtered extensions preserve the semimartingale property which we will later define.

Definition 2.9 (Stable Convergence in Law)
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and (E, d_E) a Polish space with Borel σ-algebra $\mathcal{B}(E)$. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of E valued random variables. We say that $(X_n)_{n\in\mathbb{N}}$ converges stably in law if there exist a probability measure μ on $(\Omega \times E, \mathcal{F} \otimes \mathcal{B}(E))$ such that

$$\mathbb{E}[Y f(Z_n)] = \int_{\Omega\times E} Y(\omega) f(x) d\mu(\omega, x),$$

for every Y bounded and f continuous and bounded.

It may again be more interpretable to realise the limit as a random variable. In order to realise the limit, we first suitable extend the space and σ algebra,

$$\begin{aligned}
\widetilde{\Omega} &= \Omega \times E \\
\widetilde{\mathcal{F}} &= \mathcal{F} \otimes \mathcal{B}(E).
\end{aligned}$$

Now, we may endow $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$ with the probability measure μ from Definition 2.9. Now, rather than the identity we realise the limit Z as the random variable $Z(\omega, x) = x$. We then have

$$\mathbb{E}[Y f(Z_n)] \to \widetilde{\mathbb{E}}[Y f(Z))],$$

for every bounded random variable Y and continuous and bounded f. We denote stable convergence in law by

$$Z_n \xrightarrow{\mathcal{L}_S} Z.$$

If we have explicit stable convergence in law for the Skorokhod topology we write

$$Z_n \xRightarrow{\mathcal{L}_S} Z.$$

We now present two crucial lemmas.

Lemma 2.9. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and (E, d_E) a Polish space endowed with its Borel σ-algebra, $\mathcal{B}(E)$. Let $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ be sequences of E valued random variables such that*

$$X_n \xrightarrow{\mathbb{P}} X, \quad Y_n \xrightarrow{\mathcal{L}_S} Y,$$

with Y defined on a very good extension of $(\Omega, \mathcal{F}, \mathbb{P})$. Then,

$$(X_n, Y_n) \xrightarrow{\mathcal{L}_S} (X, Y).$$

See [[20], appendix B] for a proof. Now, we have the following version of the Continuous Mapping theorem.

Lemma 2.10. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, (E, d_E) a Polish space and $(Z_n)_{n\in\mathbb{N}}$ a sequence of random variables, such that*

$$Z_n \xrightarrow{\mathcal{L}_S} Z,$$

with Z defined on an extension $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$. Let $g : E \to E'$ be a continuous function on a Polish space (E', d'_E). Then,

$$g(Z_n) \xrightarrow{\mathcal{L}_S} g(Z)$$

Proof. Note, that by definition

$$\mathbb{E}[Y f(Z_n)] \to \widetilde{\mathbb{E}}[Y f(Z)], \quad \forall f \text{ continuous and bounded and } Y \text{ bounded.}$$

However, $f \circ g$ is continuous and bounded whenever f is continuous and bounded on $(E', d_{E'})$ and g is continuous so we have

$$\mathbb{E}[Y f(g(Z_n))] \to \widetilde{E}[Y f(g(Z))], \quad \forall f \text{ continuous and bounded and } Y \text{ bounded.}$$

■

2.4.1 Semimartingales and Quadratic Variation

This section is based on [14], [13], and [4].

We first define the class of semimartingales.

Definition 2.10 (Semimartingale)
Let $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,\infty)}, \mathbb{P}\right)$ be a filtered probability space. A process X is said to be a semimartingales if

$$X = X_0 + M + A,$$

where M is a local martingale with $M_0 = 0$ almost surely, A is a process of locally bounded variation with $A_0 = 0$ almost surely.

Remark: Note that throughout this section we consider not necessarily finite intervals, i.e. $[0, \infty)$. However, all results still apply to the finite intervals $[0, T]$.

Our primary focus has been considering processes as $\mathbb{D}([0,\infty);\mathbb{R})$ valued random variables. However, in this section and going forward we will need the random variables X_t for distinct $t \in [0,\infty)$. For $X : \Omega \to \mathbb{D}([0,\infty);\mathbb{R})$ valued random variables we may define the so-called canonical projections to make sense of X_t. Let $\tau = (t_1, t_2, \ldots, t_n) \subset [0,\infty)$ then π_τ is defined as the the map $\pi_\tau : \mathbb{D}([0,T];\mathbb{R}) \to \mathbb{R}^n$ mapping X to $(X_{t_1}, X_{t_2}, \ldots X_{t_n})$. We justify the measurability in Appendix A.3. In order to ease notation we define

$$X_t = \pi_t X, \quad t \in [0,\infty).$$

Furthermore, we define

$$X_{t-} = \lim_{s \to t} X_s,$$

and the process X_- by the mapping $t \mapsto X_{t-}$. In addition this allows us to define the process ΔX by the mapping $t \mapsto X_t - X_{t-}$. Note that if $X : \Omega \to \mathbb{D}([0,\infty);\mathbb{R})$ then the processes X_- and ΔX are well defined. We will now need to recover methods of integration with respect to the processes that are semimartingales. We will not develop this theory in full, and for a more explicit treatment we refer the reader to [14].

Let H be bounded process such that

$$H \in \begin{cases} \{Y \cdot \mathbb{1}_{(a,b]} \mid a < b \in [0,\infty), Y \text{ bounded and } \mathcal{F}_a\text{-measurable. } \} \\ \{Y \cdot \mathbb{1}_{\{0\}} \mid Y \text{ bounded and } \mathcal{F}_0\text{-measurable. } \} \end{cases} \tag{2.23}$$

Here, the process is "simply" a scaled indicator by some random variable Y. Now, intuitively the integral of processes in (2.23) is defined by the process

$$H \bullet X_t = \begin{cases} Y(X_{b\wedge t} - X_{a\wedge t}) & H \in \{Y \cdot \mathbb{1}_{(a,b]} \mid a < b \in [0,\infty), Y \text{ bounded and } \mathcal{F}_a\text{-measurable.}\} \\ 0 & H \in \{Y \cdot \mathbb{1}_{\{0\}} \mid Y \text{ bounded and } \mathcal{F}_0\text{-measurable. } \} \end{cases} \tag{2.24}$$

Now, in this case simply define our integral to mean $H \bullet X$, i.e.

$$\int_0^t H_s dX_s = H \bullet X_t.$$

Using this simple integral we may extend it to a larger class of integrands.

Theorem 2.8. *Let X be a semimartingale. The map $H \mapsto H \bullet X$ defined for*

$$H \in \begin{cases} \{Y \cdot \mathbb{1}_{(a,b]} \mid a < b \in [0,\infty), Y \text{ bounded and } \mathcal{F}_a\text{-measurable. } \} \\ \{Y \cdot \mathbb{1}_{\{0\}} \mid Y \text{ bounded and } \mathcal{F}_0\text{-measurable. } \}. \end{cases} \tag{2.25}$$

Can be extended to H in the space of locally bounded and predictable processes with the following properties:

1. *$H \bullet X$ is $\mathbb{D}([0,\infty);\mathbb{R})$ valued and adapted.*
2. *$H \mapsto H \bullet X$ is linear, up to indistinguishability.*
3. *If $(H_n)_{n\in\mathbb{N}}$ is a sequence of predictable process with pointwise limit H, and if $|H_n(t)| \leq K(t)$ for all t where K is a locally bounded predictable process, then $H_n \bullet X \to H \bullet X$ weakly.*

If we consider finite intervals in 3 then $H_n \bullet X \overset{u.c.p.}{\Rightarrow} H \bullet X$. *Furthermore, this extension is unique up to indistinguishability.*

See [[14], pp. 46-47, Theorem 4.31] for a proof. The integral futhermore satisfies the following. In particular, we recognise that 3 in Theorem 2.8 is a "Dominated Convergence Theorem" for the stochastic integral.

We have not strictly defined what it means for the integral to converge weakly. However, a definition is given in the appendix.

Proposition 2.7. *Let* X *be a semimartingale and* H *a locally bounded predictable process, then the following statements hold up to indistinguishability.*

1. $X \mapsto H \bullet X$ *is linear.*
2. $H \bullet X$ *is a semimartingale.*
3. *If* X *is a local martingale, then* $H \bullet X$ *is a local martingale.*

We are now ready to define the quadratic variation

Definition 2.11 (Quadratic Covariation)
The quadratic co-variation of two semimartingales X and Y is given by

$$[X, Y] = XY - X_0 Y_0 - X_- \bullet Y - Y_- \bullet X,$$

when $Y = X$ we say the quadratic variation and denote it $[X] = [X, X]$.

Proposition 2.8. *Let* X *and* Y *be semimartingales. Then,*

1. $[X, Y]_0 = 0$.
2. $[X, Y] = [X - X_0, Y - Y_0]$.
3. $[X, Y] = \frac{1}{4}\left([X + Y] - [X - Y]\right)$

At the moment, all processes are still considered as $\mathbb{D}([0, T]; \mathbb{R})$ valued random variables. We are now interested in whether finite dimensional convergence can imply convergence in the Skorokhod space. We firs present some definitions.

Definition 2.12 (Adapted Subdivisions)
Let $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,\infty)}, \mathbb{P}\right)$ be a filtered probability space. Let $\tau = (t_n)_{n \in \mathbb{N}_0}$ be a sequence of stopping times satisfying $t_0 = 0$ and $t_n < t_{n+1}$ and tends to $t_n \to \infty$ as $n \to \infty$ almost surely, then τ is called an adapted subdivision.

We are now ready to define an approximation of our stochastic integral.

Definition 2.13 (Riemann Approximation)
Let $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,\infty)}, \mathbb{P}\right)$ be a filtered probability space and τ be an adapted subdivision. Recall that $\tau = (t_0, t_1, t_2, \dots) = (t_n)_{n\in\mathbb{N}_0}$. Let H be a predictable process of locally bounded variation and X be a semimartingale. We define the Riemann-approximant of $H \bullet X$ as the process

$$\tau(H \bullet X)_t = \sum_{n\in\mathbb{N}_0} H_{t_n}(X_{t_{n+1}\wedge t} - X_{t_n\wedge t}).$$

Our goal is to prove that when τ becomes a finer partition, then the above converges to the true integral $H \bullet X$. To this end, we need to define what it means for a partition to become finer. Now, let $(\tau_m)_{m\in\mathbb{N}_0} = ((t_{m,n})_{n\in\mathbb{N}_0})_{m\in\mathbb{N}_0}$ be a sequence of adapted subdivisions satisfying

$$\lim_{m\to\infty} \sup_{n\in\mathbb{N}_0} |t_{m,n+1} \wedge t - t_{m,n} \wedge t| = 0, \quad \forall t \in [0,\infty), \tag{2.26}$$

we have that the covering becomes finer. Note that the times $t_{m,n}$ need not be regularly spaced, although we will assume them to be for our result. In this event we have that the times are not random.

Proposition 2.9. *Let $(\tau_m)_{m\in\mathbb{N}_0}$ be a sequence of adapted subdivisions satisfying* (2.26)*. Let X be a semimartingale and H a process of locally bounded variation. Then*

$$\tau_m(H \bullet X) \overset{u.c.p.}{\Rightarrow} H \bullet X.$$

Proof. $(\tau_m)_{m\in\mathbb{N}_0}$ be a sequence satisfying (2.26) and recall $\tau_m = (t_{m,0}, t_{m,1}, \dots)$. Define

$$H^m(t) = \sum_{n\in\mathbb{N}} H_{t_{m,n}} \mathbb{1}_{(t_{m,n}, t_{m,n+1}]}(t).$$

H^m is clearly predictable with point-wise limit H. Define $K(t) = \sup_{s\in[0,t]} |H(s)|$. We now have $|H^m(t)| \leq K(t)$ for all $t \in [0,\infty)$. Now,

$$\tau_m(H \bullet X)_t = \sum_{n\in\mathbb{N}_0} H_{t_{m,n}}(X_{t_{m,n+1}\wedge t} - X_{t_{m,n}\wedge t})$$

Similarly, note that H^m is defined on disjoint intervals. Applying (2.24) to each interval yields

$$H^n \bullet X = \sum_{n\in\mathbb{N}_0} H_{t_{m,n}}(X_{t_{m,n+1}\wedge t} - X_{t_{m,n}\wedge t}).$$

The result is that $\tau_m(H \bullet X) = H^m \bullet X$. By Theorem 2.8 the result follows. ■

Hence, our integral can be viewed as the limit of a Riemann-approximation. Now, the same methodology can be applies for the quadratic variation.

Lemma 2.11. *Let X be a semimartingales let $(\tau_m)_{m\in\mathbb{N}_0}$ be a sequence of adapted subdivisions satisfying* (2.26)*. Define*

$$S_{\tau_m}(X)_t = \sum_{n\in\mathbb{N}_0} (X_{t_{m,n+1}\wedge t} - X_{t_{m,n}\wedge t})^2,$$

then

$$S_{\tau_m}(X) \overset{u.c.p.}{\Rightarrow} [X].$$

Proof. First, consider

$$\begin{aligned}(x-y)^2 &= x^2 + y^2 - 2xy \\ &= x^2 - y^2 - 2y(x-y).\end{aligned}$$

By definition, we may rewrite $S_{\tau_m}(X)$ as

$$S_{\tau_m}(X) = \sum_{m\in\mathbb{N}_0} X^2_{t_{m,n+1}\wedge t} - X^2_{t_{m,n}\wedge t} - 2(X_{t_{m,n}\wedge t})(X_{t_{m,n+1}\wedge t} - X_{t_{m,n}\wedge t}).$$

Note that the first two terms are telescoping. Hence,

$$S_{\tau_m}(X) = X_t^2 - X_0^2 - 2\left(\sum_{m\in\mathbb{N}_0} (X_{t_{m,n}\wedge t})(X_{t_{m,n+1}\wedge t} - X_{t_{m,n}\wedge t})\right).$$

Note, by Definition 2.13 that

$$\tau_m(X_- \bullet X) = \sum_{m\in\mathbb{N}_0} (X_{t_{m,n}\wedge t})(X_{t_{m,n+1}\wedge t} - X_{t_{m,n}\wedge t})$$

Hence,

$$S_{\tau_m}(X) = X^2 - X_0^2 - 2\tau_m(X_- \bullet X).$$

Now, by Proposition 2.9 we have $\tau_m(X_- \bullet X) \overset{u.c.p.}{\Rightarrow} X_- \bullet X$. Now, by definition of the quadratic variation we now have

$$S_{\tau_m}(X) \overset{u.c.p.}{\Rightarrow} X^2 + X_0^2 - 2X_- \bullet X = [X].$$

■

We now present some practical applications of the previous theory.

Definition 2.14 (Realised Variance and Integral Approximation)
Let $\tau = (t_0, t_1, \dots, t_n)$ be an equidistant partition of $[0,T]$, i.e. $t_i = \Delta_n i$ where $\Delta_n = T/n$ and $i \in \{0, 1, \dots n\}$. We define the realised variance of a process X as

$$\widehat{[X]}^n_t = \sum_{i=1}^{[t/\Delta_n]} (X_{i\Delta_n} - X_{(i-1)\Delta_n})^2,$$

where $[t/\Delta_n]$ is the integer part of t/Δ_n. Furthermore, let H be predictable process of locally bounded variation. We define

$$\widehat{H \bullet X}^n_t = \sum_{i=1}^{[t/\Delta_n]} H_{(i-1)\Delta_n}(X_{i\Delta_n} - X_{(i-1)\Delta_n}).$$

Consider with slight alterations the approximation S_τ defined in Lemma 2.11 and it's relation to $\widehat{[X]}$. The difference is subtle, but suppose we consider $t^* \in [0,T]$, then for a semimartingale X

$$S_\tau(X)_{t^*} = (X_{t^*} - X_{([t^*/\Delta_n])\Delta_n})^2 + \sum_{i=1}^{[t^*/\Delta_n]} (X_{i\Delta_n} - X_{(i-1)})^2 = (X_{t^*} - X_{([t^*/\Delta_n])\Delta_n})^2 + \widehat{[X]}^n_{t^*},$$

now, it should be no surprise that the term $(X_{t^*} - X_{([t^*/\Delta_n])\Delta_n})^2$ becomes negligible as our partition becomes finer. A similar result applies to $\widehat{H \bullet X}^n$. Indeed, we consider the following corollary.

Corollary 2.3. *Let X be a semimartingale on $[0,T]$ and H a process of locally bounded variation, then*

$$\widehat{[X]}^n \overset{u.c.p}{\Rightarrow} [X], \quad \text{as } n \to \infty.$$
$$\widehat{H \bullet X}^n \overset{u.c.p}{\Rightarrow} H \bullet X, \quad \text{as } n \to \infty.$$

We refer to [14] and [28] for explicit proofs but note that it follows essentially from Proposition 2.9 and Lemma 2.11.

Corollary 2.3 states that convergence of our estimator is ensured, but it fails to tell us at which rate. We will make some restrictions to the structure of our processes here. Essentially, we now require the quadratic variation to be absolutely continuous with respect to the Lebesgue measure. Therefore, we introduce the so-called *Itô semimartingales*. Typically, Itô semimartingales are allowed to jump, but we restrict ourselves to continuous processes.

Definition 2.15 (Continuous Itô Semimartingale)
Let $(\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,\infty)}, \mathbb{P}\right)$ be a filtered probability space, and let X be a semimartingale. X is said to be a continuous Itô semimartingale if it admits the representation

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s,$$

where W is a Brownian Motion, σ is predictable and cádlág and b is a process satisfying $\int_0^t |b_s| ds < \infty$ for every $t > 0$.

Now, the following limit theorem is due to Jacod and Protter, [13].

Theorem 2.9. *Let $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,\infty)}, \mathbb{P}\right)$ be a filtered probability space satisfying the usual conditions. Let X be a continuous Itô semimartingale, i.e. of the form*

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s,$$

satisfying

$$\int_0^t (|b_s|^2 + |\sigma_s|^4) ds < \infty, \quad \forall t > 0.$$

Consider an equidistant partitioning of $[0,T]$ with distance Δ_n.

$$\frac{1}{\sqrt{\Delta_n}}(\widehat{[X]}^n - [X]) = Z^n \overset{\mathcal{L}_S}{\Rightarrow} Z,$$

where Z is a process defined on a very good filtered extension of $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,\infty)}, \mathbb{P}\right)$. Furthermore, Z is $\mathcal{F}$ conditionally centered Gaussian with independent increments and finite second moment given by

$$\mathbb{E}\left[Z_t^2 \mid \mathcal{F}\right] = 2\int_0^t \sigma_s^4 ds.$$

See [[13], p. 162 Theorem 5.42]. The central limit theorem here tells us that we can expect a rate of convergence proportional to $\sqrt{n}$ when the convergence happens on an equidistant mesh. Now, the above can be paraphrased as

$$Z_t^n \overset{\mathcal{L}_S}{\Rightarrow} Z_t = \sqrt{2}\int_0^t \sigma_s^2 dW_s', \tag{2.27}$$

where W' is a Brownian motion on the very good filtered extension from the theorem. Jacod and Protter states that W' is independent of $\mathcal{F}$. It implies that the second moment in Theorem 2.9 is exactly the quadratic variation of the process Z. We now need an estimator of

$$\int_0^t \sigma_s^4 ds,$$

Furthermore, the estimator must converge in probability on the Skorokhod space, since this will immediately imply the joint convergence of Z^n and the estimator. Now, Jacod and Protter show the following convergence.

$$\frac{1}{\Delta_n}\sum_{i=1}^{[t/\Delta_n]} (|X_{t_i} - X_{t_{i-1}}|)^4 \overset{u.c.p.}{\Rightarrow} 3\int_0^t |\sigma_s|^4 ds, \tag{2.28}$$

under the added assumptions that

$$\int_0^t |b_s|^{\frac{4}{6}} ds < \infty, \quad \int_0^t |\sigma_s|^4 ds < \infty.$$

In fact both Lemma 2.11 and (2.28) are special cases of the so called power variation convergence, which for continuous processes can be found in [[13], p. 70, Theorem 3.4.1]. We can now present the central limit theorem.

Theorem 2.10. *Let $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,\infty)}, \mathbb{P}\right)$ be a filtered probability space satisfying the usual conditions. Let X be a continuous Itô semimartingale, i.e. of the form*

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s,$$

satisfying

$$\int_0^t (|b_s|^2 + |\sigma_s|^4) ds < \infty, \quad \forall t > 0.$$

Consider an equidistant partitioning of $[0,T]$ with $\Delta_n = T/n$. Then

$$\frac{\sqrt{3}\left(\sum_{i=1}^{[t/\Delta_n]}(X_{t_i} - X_{t_{i-1}})^2 - [X]_t\right)}{\sqrt{2}\sqrt{\sum_{i=1}^{[t/\Delta_n]}(|X_{t_i} - X_{t_{i-1}}|)^4}} \overset{\mathcal{L}_S}{\Rightarrow} N,$$

where N is defined on a very good filtered extension in restriction to $\{[X]_t > 0\}$. Furthermore, N is $\mathcal{F}$ conditionally standard Gaussian.

Proof. By Theorem 2.9 we know that

$$\frac{1}{\sqrt{\Delta_n}}(\widehat{[X]}^n - [X]) = Z^n \overset{\mathcal{L}_S}{\Rightarrow} Z,$$

where Z is defined by (2.27). In particular,

$$Z_t^n \xrightarrow{\mathcal{L}_S} Z_t, \quad t > 0.$$

Furthermore, by (2.28) we have

$$\frac{1}{\Delta_n} \sum_{i=1}^{[t/\Delta_n]} (|X_{t_i} - X_{t_{i-1}}|)^4 \overset{u.c.p.}{\Rightarrow} 3 \int_0^t |\sigma_s|^4 ds,$$

and in particular it implies

$$G_t^n = \frac{1}{\Delta_n} \sum_{i=1}^{[t/\Delta_n]} (|X_{t_i} - X_{t_{i-1}}|)^4 \xrightarrow{\mathbb{P}} 3 \int_0^t |\sigma_s|^4 ds = G_t, \quad t > 0.$$

By Lemma 2.9 we have (G_t^n, Z_t^n) converges stably in law to (G_t, Z_t). By Lemma 2.10 we have

$$\frac{Z_t^n}{\sqrt{G_t^n}} \xrightarrow{\mathcal{L}_S} \frac{Z_t}{\sqrt{3 \int_0^t \sigma_s^4 ds}},$$

when restricted to the set $\{\int_0^t |\sigma|_s^4 ds > 0\}$ which coincides with $\{[X]_t > 0\}$. Now, by Theorem 2.9 Z_t is an $\mathcal{F}$ conditional centered Gaussian with variance

$$2 \int_0^t \sigma_s^4 ds,$$

and so

$$\frac{Z_t}{\sqrt{3 \int_0^t \sigma_s^4 ds}}$$

is $\mathcal{F}$ conditionally centered Gaussian with variance 2/3. It follows that

$$\frac{\sqrt{3}\left(\sum_{i=1}^{[t/\Delta_n]} (X_{t_i} - X_{t_{i-1}})^2 - [X]_t\right)}{\sqrt{2}\sqrt{\sum_{i=1}^{[t/\Delta_n]} (|X_{t_i} - X_{t_{i-1}}|)^4}} \xrightarrow{\mathcal{L}_S} \frac{\sqrt{3} Z_t}{\sqrt{2}\sqrt{3 \int_0^t \sigma_s^4 ds}},$$

which is $\mathcal{F}$ conditionally centered Gaussian with variance 1. ■

Remark: Jacod and Protter show that the restrictions on the processes b and σ can be reduced to b being locally bounded and σ cádlág. The explicit statement can be found in [[13], Theorem 5.6.1] where the assumption is that X is a continuous Itô semimartingale satisfying the above. We simply state some additional assumptions to make the proof rely only on theory we have presented.

2.4.2 Processes Embedded in a Brownian Motion

This subsection is based on [28], [23], and [40].

In the previous section we introduces semimartingales. Semimartingales are frequently used in finance due to their flexibility and the fact that they provide a *good* basis for integration.

In this section our goal is to motivate the use of a time-changed Brownian motion. We show the Dubins-Schwarz theorem and associated proof, which essentially tells us that any continuous local martingale is a time-changed Brownian motion. We also refer to a theorem from Monroe,

which states that any semimartingale can be represented as a time-changed Brownian motion. However, while the first proof is constructive the latter is not.

We say that a process can be embedded into a Brownian motion if we can apply a time-change which transforms the process to a Brownian motion.

In order to show the Dubins-Schwarz theorem we require the following lemmas. We include some results in Appendix A.4 to keep the focus of the section on the Brownian embedding.

Lemma 2.12. *Let M be a local martingale. Then M is a martingale with $\mathbb{E}\left[M_t^2\right] < \infty$ for all $t \in [0, \infty)$ if, and only if, $\mathbb{E}\left[[M]_t\right] < \infty$ for all $t \in [0, \infty)$. We then have $\mathbb{E}\left[M_t^2\right] = \mathbb{E}\left[[M]_t\right]$*

Lemma 2.12 allows us to identify true martingales and in a sense justifies the expression quadratic variation.

Lemma 2.13. *Let M be a local square integrable martingale, then $[M]$ is the unique process such that with locally bounded variation satisfying*

$$M^2 - [M],$$

is a local martingale.

Lemma 2.13 tells us that the quadratic variation is unique, but also tells us that $M^2 - [M]$ is a local martingale.

Lemma 2.14. *Let M be a continuous local martingale, then $[M]$ is continuous a.s., and if $I \subset [0, \infty)$ is an interval over which M is constant, then $[M]$ is constant a.s. on I. Conversely, if I is an interval over which $[M]$ is constant, then M is constant on I.*

Lemma 2.14 tells us, that if our process becomes constant on some interval, then the quadratic variation will also become constant. That is, if the process is deterministic over some interval, then the quadratic variation does not change on this interval.

Theorem 2.11 (Lévy's Characterisation)**.** *Let M be a local-martingale with $M_0 = 0$, then the following are equivalent*

1. *M is a standard Brownian motion.*
2. *M has continuous sample paths and $M_t^2 - t$ is a local martingale.*
3. *$[M]_t = t$*

Theorem 2.11 is known as *Lévy's Characterisation of a Brownian Motion.* In light of Lemmas 2.12, 2.13, 2.14 it should be no surprise. However, the theorem is extremely significant, as it allows us to identify a Brownian motion.

Theorem 2.12 (Dubins-Schwarz Theorem)**.** *Let M be a continuous local martingale with $M_0 = 0$ and whose quadratic variation satisfies $[M]_t \to \infty$ as $t \to \infty$ almost surely. Let*

$$T_s = \inf\{t > 0 : [M]_t > s\},$$

and define $\mathcal{G}_s = \mathcal{F}_{T_s}$ *and* $B_s = M_{T_s}$ *then* B *is a standard Brownian motion with respect to* $(\mathcal{G})_{t\in[0,\infty)}$. *Moreover,* $([M]_t)_{t\in[0,\infty)}$ *are stopping times for* $(\mathcal{G}_t)_{t\in[0,\infty)}$ *and*

$$M_t = B_{[M]_t}, \quad a.s., \ \forall t \in [0,\infty).$$

Proof. Note that $(T_t)_{t\in[0,\infty)}$ are stopping times, and T_t is finite almost surely by hypothesis, since $\lim_{t\to\infty}[M]_t = \infty$ almost surely. Hence $\mathcal{G}_t$ is a well-defined σ-algebra. Furthermore note that $\{\omega \mid [M]_t(\omega) \leq s\}$ is equal to $\{\omega \mid T_s(\omega) \geq t\}$ by construction, and so $([M]_t)_{t\in[0,\infty)}$ are stopping times for the filtration $(\mathcal{G}_t)_{t\in[0,\infty)}$.

By Lemma 2.12 we have $\mathbb{E}\left[M_{T_s}^2\right] = \mathbb{E}\left[[M]_{T_s}\right] = s < \infty$, since $[M]_{T_s} = s$ by the continuity, which follows from Lemma 2.14. Furthermore,

$$\mathbb{E}\left[B_t - B_s \mid \mathcal{G}_s\right] = \mathbb{E}\left[M_{T_t} - M_{T_s} \mid \mathcal{F}_{T_s}\right] = 0,$$

which follows from Theorem A.4 and since we have $[M]_t \to \infty$ as $t \to \infty$ we obtain that B is a local martingale, since clearly $T_s \leq T_t$ almost surely and $T_t \to \infty$ as $t \to \infty$.

$$\begin{aligned}
\mathbb{E}\left[B_t^2 - B_s^2 \mid \mathcal{G}_s\right] &= \mathbb{E}\left[(B_t - B_s)^2 \mid \mathcal{G}_s\right] \\
&= \mathbb{E}\left[(M_{T_t} - M_{T_s})^2 \mid \mathcal{F}_{T_s}\right] \\
&= \mathbb{E}\left[[M]_{T_t} - [M]_{T_s} \mid \mathcal{F}_{T_s}\right] \\
&= t - s
\end{aligned}$$

It follows that $\mathbb{E}\left[B_s^2\right] - s$ is a martingale, and so $[B]_s = s$ and so by Theorem 2.11 we obtain that B is a standard Brownian motion with respect to $(\mathcal{G}_t)_{t\in[0,\infty)}$; the sample path continuity follows by the converse of Lemma 2.14.

We must now show that $M_t = B_{[M]_t}$. Note that $B_t = M_{T_t}$ and by the continuity of $[M]$ it follows that $B_{[M]_t} = B_{T_{[M]_t}}$ a.s. Now, $(T_t)_{t\in[0,\infty)} \geq t$ for all t with $T_t = t$ if, and only if, t is a point of increase, i.e. a point t such that all $\varepsilon > 0$ we have $T_{t+\varepsilon} > T_t$. But now note that if $T_{[M]_t} > t$ implies $t \mapsto [M]_t$ is constant on $(t, T_{[M]_t})$ and therefore M is constant on $(t, T_{[M]_t})$ by Lemma 2.14 and it follows that $M_{T_{[M]_t}} = M_t$ a.s. which then implies $B_{[M]_t} = M_t$ a.s. ■

In general, the requirement that $\lim_{t\to\infty}[M]_t = \infty$ is not very restrictive. Revuz and Yor show that in the case $\lim_{t\to\infty}[M] = M_\infty < \infty$ we can extend our probability space in a certain way such that any continuous local martingale is a time-changed Brownian motion on an extension, see [[31], Theorem (1.7)].

A much stronger result which we will not prove is given by Monroe in [23].

Theorem 2.13. *Let* X *be a semimartingale then* X *is equivalent to time-changed Brownian motion on a suitably extended probability space. That is, there exists a process of stopping times* $(T_t)_{t\in[0,\infty)}$ *on on the extended space such that*

$$X_t = W_{T_t}$$

The proof is beyond the scope of this thesis and we refer to [23] for a proof.

Rather than a proof, consider the following thought experiment for the finite dimensional distribution of a semimartingale X defined on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,\infty)}, \mathbb{P}\right)$.

Let $\tau = (t_1, t_2, \ldots, t_n) \subset [0, \infty)$ with $t_i < t_{i+1}$ for i in $\{1, 2, \ldots, n-1\}$ and $X_\tau = (X_{t_1}, X_{t_2}, \ldots, X_{t_n})$. Let W be a Brownian motion on a separate probability filtered space, $\left(\Omega^*, \mathcal{F}^*, (\mathcal{F}_t^*)_{t\in[0,\infty)}, \mathbb{Q}\right)$ independent of the underlying probability space of X. Now, consider the hitting times

$$T_n(\omega^*, \omega) = \inf\{t > 0 \mid W_t(\omega^*) \in [X_{t_n}(\omega), \infty)\},$$

on the extension $\overline{\Omega} = \Omega^* \times \Omega$. While this interpretation is in no way rigorous, one can imagine that *simply waiting for the Brownian motion to hit the right value* is a valid strategy and Theorem 2.13 justifies the claim to some extent.

A final remark to this section is that Monroe showed the statement conversely, i.e. X is a semimartingale only if there exists an extension for which is is a time-changed Brownian motion. The application is yet much less obvious. Suppose we are given a sample $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ we could perform the numerical work an simulate paths of a Brownian motion to estimate the laws of $(T_{t_i})_{i=1}^N$. However, if the values are excessive, then the likelihood that any algorithm would ever complete a reasonable approximation is very low.

In general, one can now define semimartingales and martingales by the converse of Theorem 2.12 and Theorem 2.13. In the literature there are primarily two types of time-changes which are studied. The first being the so-called *subordinators*. They are Lévy processes, i.e. stationary processes starting at 0 with independent increments and which are continuous in probability, with the added requirement that they are almost surely increasing, see, e.g., [1].

The second type, which we will focus on, are processes $\mathcal{T}$ which are either continuous and strictly increasing with $\mathcal{T}_0 = 0$ or strictly increasing, and absolutely continuous with respect to the Lebesgue measure. To motivate the use of these models, consider the following. We take a Brownian motion W and an independent process ν which is cádlág and non-negative almost surely. Then we define a process

$$X_t = \int_0^t \nu_s dW_s, \quad t \geq 0.$$

The quadratic variation of this process is given by

$$[X]_t = \int_0^t \nu_s^2 ds, \quad t \geq 0.$$

Now, define $\mathcal{T}_t = [X]_t$ and note that $\mathcal{T}$ is absolutely continuous with respect to the Lebesgue measure. Furthermore, by Dubins-Schwarz we have

$$X_t = W'_{\mathcal{T}_t}, \quad t \geq 0,$$

where W' is a Brownian motion. It should be no surprise that W' and $\mathcal{T}_t$ are independent in this case. We will not prove this, but refer to [[40], Corollary 2]. It follows that X is a continuous local martingale by the Dubins-Schwarz theorem. Given the previous results, we may estimate $\mathcal{T}_t$ using the realised variance.

2.4.3 Copulas and Finite Dimensional Distributions

Suppose we wish to analyse a stochastic process $X = (X_t)_{t\in\mathcal{T}}$. Many applications do not require complete knowledge of X but rather only the distribution of X in certain times and the dependence.

$$(X_{t_1}, X_{t_2}, \ldots, X_{t_n}). \tag{2.29}$$

There are several ways to go about analysing (2.29). We previously presented the so-called *Skorokhod space* $\mathbb{D}(\mathcal{T};\mathbb{R})$ for which we can make sense of (2.29) in terms of a canonical projection onto $\mathbb{R}^n$ and since the Skorokhod space is infinite dimensional while (2.29) is finite dimensional we refer to the distribution of (2.29) as a *finite dimensional distribution* of X.

It is no surprise that for many stochastic processes the most valuable information comes from deriving the distribution of

$$(X_s, X_t),\ s < t.$$

We previously showed, a large class of stochastic processes may be embedded in the Brownian motion under a time change. That is, let W be a standard Brownian motion, then many processes may be described by

$$X_t = W_{\mathcal{T}_t}.$$

for some appropriate time-change $\mathcal{T}$. Therefore, we wish to find the copula between the same Brownian motion at two distinct times.

Theorem 2.14. *Let W be a Brownian Motion and $s < t$, then the copula of (W_s, W_t) is given by*

$$C^W_{s,t}(u,v) = \int_0^u \Phi\left(\frac{\sqrt{t}\Phi^{-1}(v) - \sqrt{s}\Phi^{-1}(w)}{\sqrt{t-s}}\right) dw$$

Proof. Note that by the independent increments and stationarity we have

$$\mathbb{P}(W_t \le x \mid W_s = y) = \mathbb{P}(W_t - W_s \le x - y).$$

By virtue of Theorem 2.1 we have

$$\begin{aligned} C^W_{s,t}\left(\Phi\left(\frac{y}{\sqrt{s}}\right), \Phi\left(\frac{x}{\sqrt{t}}\right)\right) &= \mathbb{P}(W_t \le x, W_s \le y) \\ &= \int_{-\infty}^{y} \mathbb{P}(X_t \le x \mid X_s = z) d\Phi\left(\frac{z}{\sqrt{s}}\right) \\ &= \int_{-\infty}^{y} \Phi\left(\frac{x-z}{\sqrt{t-s}}\right) d\Phi\left(\frac{z}{\sqrt{s}}\right) \end{aligned}$$

Using the change of variables $v = \Phi\left(\frac{x}{\sqrt{t}}\right)$ and $u = \Phi\left(\frac{y}{\sqrt{s}}\right)$ we obtain

$$C^W_{s,t}(u,v) = \int_0^u \Phi\left(\frac{\sqrt{t}\Phi^{-1}(v) - \sqrt{s}\Phi^{-1}(w)}{\sqrt{t-s}}\right) dw$$

■

We would like to remark that $C^W_{s,t}$ is indeed a copula; we only consider the marginals the other

requirements of Proposition 2.3 are clear. Clearly,

$$C^W_{s,t}(0,v) = C^W_{s,t}(u,0) = 0.$$

Furthermore, for the identities $C^W_{s,t}(u,1) = u$ and $C_{s,t}(1,v) = v$ follows essentially from the proof of Theorem 2.14.

We will now abuse the notation $[X]$ slightly. That is, we will take a process $[X]$ which is continuous and strictly increasing with $[X]_0 = 0$ almost surely and define a process X by time-changing an independent Brownian motion using $[X]$. Hence $[X]$ will coincide with the quadratic variation and justify the abuse of notation. While the argumentation becomes rather circular, we feel that the notation is suitable for interpreting the presented results and relating the result to the Dubins-Schwarz theorem.

Corollary 2.4. *Let X be a continuous local martingale of the form*

$$X_t = W_{[X]_t}, \quad t \geq 0.$$

where W is a Brownian motion and $[X]$ is a strictly increasing, continuous process independent of W with $[X]_0 = 0$. Let $s < t$. Then the conditional copula of (X_s, X_t) given $([X]_s, [X]_t)$ is given by

$$C^X_{s,t}(u,v \mid [X]_s, [X]_t) = \int_0^u \Phi\left(\frac{\sqrt{[X]_t}\Phi^{-1}(v) - \sqrt{[X]_s}\Phi^{-1}(w)}{\sqrt{[X]_t - [X]_s}}\right) dw.$$

Proof. By Theorem 2.12 we have that $X_t = W_{[X]_t}$ for some Brownian motion. Hence, by Theorem 2.14 and independence between $[X]$ and W we have

$$C^X_{s,t}(u,v \mid [X]_s, [X]_t) = C^W_{[X]_s,[X]_t}(u,v).$$

■

Remark: In particular the copula will allow us to model

$$\mathbb{P}(X_t \leq x, X_s \leq y \mid [X]_t = b, [X]_s = a) = C^X_{s,t}\left(\Phi\left(\frac{x}{b}\right), \Phi\left(\frac{y}{a}\right) \mid [X]_s = a, [X]_t = b\right)$$

In the event that $[X]$ is non-random, for instance in the case of the Brownian motion, then $C^X_{s,t}$ is the unconditional copula of (X_s, X_t).

Lemma 2.15. *Under the assumptions of Corollary 2.4 the copula of (X_s, X_t) given $([X]_s, [X]_t)$ admits the following limit as $s \to t$ or $t \to s$:*

$$\lim_{|t-s|\to 0} C^X_{s,t}(u,v \mid [X]_t, [X]_s) = \min(u,v).$$

Proof. By assumption $[X]$ is continuous and strictly increasing, hence

$$\lim_{|s-t|\to 0} \frac{1}{\sqrt{[X]_t - [X]_s}} = \infty.$$

Furthermore, we have that

$$\lim_{s\to t} \sqrt{[X]_t}\Phi^{-1}(v) - \sqrt{[X]_s}\Phi^{-1}(w) = \sqrt{[X]_t}(\Phi^{-1}(v) - \Phi^{-1}(w))$$

and similarly

$$\lim_{t\to s}\sqrt{[X]_t}\Phi^{-1}(v)-\sqrt{[X]_s}\Phi^{-1}(w)=\sqrt{[X]_s}(\Phi^{-1}(v)-\Phi^{-1}(w)).$$

The result is that

$$\lim_{|t-s|\to 0}\frac{\sqrt{[X]_t}\Phi^{-1}(v)-\sqrt{[X]_s}\Phi^{-1}(w)}{\sqrt{[X]_t-[X]_s}}=\begin{cases}\infty & v>w\\ -\infty & v<w\\ 0 & v=w\end{cases}$$

Now,

$$\lim_{|t-s|\to 0}\Phi\left(\frac{\sqrt{[X]_t}\Phi^{-1}(v)-\sqrt{[X]_s}\Phi^{-1}(w)}{\sqrt{[X]_t-[X]_s}}\right)=\begin{cases}1 & v>w\\ 0 & v<w\\ 0.5 & v=w\end{cases}.$$

Note that $\{(w,v)\mid w=v\}$ has Lebesgue measure 0, and hence

$$\lim_{|t-s|\to 0}\Phi\left(\frac{\sqrt{[X]_t}\Phi^{-1}(v)-\sqrt{[X]_s}\Phi^{-1}(w)}{\sqrt{[X]_t-[X]_s}}\right)=\mathbb{1}_{[w,1]}(v),\ \text{almost everywhere.}$$

Clearly, the $\mathbb{1}_{[}w,1](v)$ is bounded and we may apply the Dominated Convergence Theorem to obtain.

$$\lim_{|t-s|\to 0}\int_0^u\Phi\left(\frac{\sqrt{[X]_t}\Phi^{-1}(v)-\sqrt{[X]_s}\Phi^{-1}(w)}{\sqrt{[X]_t-[X]_s}}\right)dw=\int_0^u\mathbb{1}_{[w,1]}(v)dw=\min(u,v).$$

■

Lemma 2.15 is quite intuitive. As $|t-s|\to 0$ we are essentially asking what is the probability that X_t is both less than or equal to $\sqrt{[X]_t}\Phi^{-1}(u)$ and $\sqrt{[X]_t}\Phi^{-1}(v)$ which should correspond to the minimum of the two. It may also be beneficial to define the copula $C^X_{s,t}(u,v)=\min(u,v)$ for $s=t$.

Now, another interesting observations is the following

Lemma 2.16. *Under the assumptions of Corollary 2.4 the copula of (X_0,X_t) given $[X]_t$ is given by*

$$C^X_{0,t}(u,v)=u\cdot v,\quad t>0,\ \textit{almost surely.}$$

Proof. Note that almost surely we have $[X]_0=0$, hence

$$C^X_{0,t}(u,v)=\int_0^u\Phi\left(\frac{\sqrt{[X]_t}\Phi^{-1}(v)}{\sqrt{[X]_t}}\right)dw=\int_0^u vdw=u\cdot v.$$

■

We now present a generalisation to a larger class of processes.

Theorem 2.15. *Let X be a semimartingale of the form*

$$X_t=X_0+A_t+W_{[X]_t},\quad t\geq 0,$$

where A_t is a process of locally bounded variation, $[X]$ is a continuous, strictly increasing process with $[X]_0=0$ and W a Brownian motion. Assume that X_0, A and $[X]$ are independent of W. Then, the conditional copula of (X_s,X_t) given $G_{s,t}=(A_s,A_t,[X]_s,[X]_t,X_0)$ is

$$C^X_{s,t}(u,v \mid G_{s,t}) = \int_0^u \Phi\left(\frac{\sqrt{[X]_t}\Phi^{-1}(v) - \sqrt{[X]_s}\Phi^{-1}(w)}{\sqrt{[X]_t - [X]_s}}\right) dw.$$

Proof. Note that

$$\mathbb{P}(X_t \leq x \mid X_s = y, G_{t,s}) = \Phi\left(\frac{x - y - (A_t - A_s)}{\sqrt{[X]_t - [X]_s}}\right),$$

Similar to the martingale case we now have

$$D_u C^X_{s,t}(F_s(y), F_t(x) \mid G_{s,t}) = \mathbb{P}(X_t \leq x \mid X_s = y, G_{t,s}) = \Phi\left(\frac{x - y - (A_t - A_s)}{\sqrt{[X]_t - [X]_s}}\right),$$

with

$$F_t(x) = \mathbb{P}(X_t \leq x \mid G_{t,s}) = \mathbb{P}(X_t \leq x \mid X_0, A_t, [X]_t) = \Phi\left(\frac{x - X_0 - A_t}{\sqrt{[X]_t}}\right),$$

and similarly for F_s. Now,

$$C^X_{s,t}(F_s(y), F_t(x)) = \int_{-\infty}^y \Phi\left(\frac{x - z - (A_t - A_s)}{\sqrt{[X]_t - [X]_s}}\right) dF_s(z).$$

Applying the change of variables $u = F_s(y)$ and $v = F_t(x)$ yields

$$C^X_{s,t}(u,v) = \int_0^u \Phi\left(\frac{\sqrt{[X]_t}\Phi^{-1}(v) + A_t + X_0 - \sqrt{[X]_s}\Phi^{-1}(w) - A_s - X_0 - (A_t - A_s)}{\sqrt{[X]_t - [X]_s}}\right) dw \tag{2.30}$$

$$= \int_0^u \Phi\left(\frac{\sqrt{[X]_t}\Phi^{-1}(v) - \sqrt{[X]_s}\Phi^{-1}(w)}{\sqrt{[X]_t - [X]_s}}\right) dw. \tag{2.31}$$

■

Remark: While it may seem surprising that the copula is the "same" as in the martingale case note that all we're doing is conditioning until we are left with essentially the Brownian part of our process. Also note that the assumptions here may very well be relaxed, but it will suffice for our purposes.

We now have an immediate application of these copulas, namely simulation. We need to be able to simulate from the copula $C^W_{s,t}(u,v)$, but note that

$$\mathbb{P}(V \leq v \mid U = u) = D_u C^W_{s,t}(u,v) = \Phi\left(\frac{\sqrt{t}\Phi^{-1}(v) - \sqrt{s}\Phi^{-1}(u)}{\sqrt{t-s}}\right).$$

Now, let $w \in (0,1)$, then

$$w = \Phi\left(\frac{\sqrt{t}\Phi^{-1}(v) - \sqrt{s}\Phi^{-1}(u)}{\sqrt{t-s}}\right) \iff v = \Phi\left(\frac{\sqrt{t-s}\Phi^{-1}(w) + \sqrt{s}\Phi^{-1}(u)}{\sqrt{t}}\right).$$

We define

$$Q^W_{s,t}(u,w) = \Phi\left(\frac{\sqrt{t-s}\Phi^{-1}(w) + \sqrt{s}\Phi^{-1}(u)}{\sqrt{t}}\right).$$

The asymptotic behavior of $w \in \{0,1\}$ can be recovered by extending Φ onto $\mathbb{R} \cup \{\infty, -\infty\}$ with $\Phi(\infty) = 1$. It then follows that $w = 1$ if, and only if $v = 1$, and similarly $w = 0$ if, and only if,

$v = 0$ - since $D_u C^W_{s,t}$ is only defined for $u \in (0,1)$. Note that the inverse is defined even in the point $t = s$.

To simulate from $C^W_{s,t}$ first we simulate $\widetilde{U}, U \sim \text{Unif}([0,1])$. By Proposition 2.2 we can now take

$$V = Q^W_{s,t}(U, \widetilde{U}),$$

then $(U, V) \sim C^W_{s,t}$. Now, let X be a process of the form

$$X_t = W_{T_t}, \quad t \geq 0,$$

where T is an almost surely increasing process and W is a Brownian motion independent of T. Then, we may define

$$Q^X_{s,t}(u, w \mid T_s, T_t) = \Phi\left(\frac{\sqrt{T_t - T_s}\Phi^{-1}(w) + \sqrt{T_s}\Phi^{-1}(u)}{\sqrt{T_t}}\right). \tag{2.32}$$

Of course, this applies to a continuous local martingale when the quadratic variation is independent of the Brownian motion. The same holds true for semimartingales of the form presented in Theorem 2.15. The only change here is that the marginals needed for inverting the pair $(U, V) \sim C^X_{s,t}(\cdot \mid T_s, T_t)$ will change. We now present a sampling algorithm.

Algorithm 1 Sampling Algorithm

Sampling a path from a time changed Brownian motion on an equidistant mesh over $[0, T]$. Let $0 = t_0$, $T = t_n$ and $t_i = i(T/n)$ for i in $1, 2, \ldots, n$. We assume an initial sample of an appropriate independent time-change $(T_{t_i})_{i=1}^n$ with $T_{t_0} = 0$ and initial value X_{t_0}. Furthermore, we assume the desired marginals $(F_{t_i})_{i=1}^n$ is known.

$U_0 = 1$
for $i \in \{1, 2, \ldots, n\}$ **do**
  if $T_{t_i} = T_{t_{i-1}}$ **then**
    $U_i = U_{i-1}$
    $X_{t_i} = X_{t_{i-1}}$
  else
    $W_i \sim \text{Unif}([0,1])$
    $U_i = Q_{s,t}(U_i, W_i \mid T_i, T_{i-1})$
    $X_{t_i} = F^{-1}_{t_i}(U_i)$
  end if
end for
return $(X_{t_0}, X_{t_1}, X_{t_2}, \ldots, X_{t_n})$

Note that the mesh need not be equidistant, in fact any mesh will suffice.

3

Results

In this chapter we present an estimator of the copula of time-changed Brownian motion. In particular, we will assume that the process is related to the Dubins-Schwarz theorem by being a continuous local martingale represented by a time-changed Brownian motion, with the additional assumption that the time-change is independent of the Brownian motion.

Afterwards, we present a limit theorem in related, but rather restricted, setting. Here, we will assume that the time-change is deterministic and absolutely continuous with respect to the Lebesgue measure.

We will then analyse the results provided numerically to investigate whether the limits may be enhanced.

We propose the natural estimator defined by

$$\widehat{C}^{X_n}_{s,t}(u,v) = \begin{cases} \min(u,v) & \widehat{[X]}^n_s = \widehat{[X]}^n_t \\ \int_0^u \Phi\left(\frac{\sqrt{\widehat{[X]}^n_t}\Phi^{-1}(v) - \sqrt{\widehat{[X]}^n_s}\Phi^{-1}(w)}{\sqrt{\widehat{[X]}^n_t - \widehat{[X]}^n_s}} \right) dw & \widehat{[X]}^n_s < \widehat{[X]}^n_t \,, \end{cases} \tag{3.1}$$

3.1 Limit Theorems

In this section we provide the relevant limit theorems for the estimator (3.1). We first consider a process defined by Brownian motion time-changed by a continuous and strictly increasing process independent of the Brownian motion. We first assert the consistency of our estimator.

Theorem 3.1. *Let X be a continuous local martingale defined by*

$$X_t = W_{[X]_t}, \quad t \geq 0, \tag{3.2}$$

where W is a Brownian motion and $[X]$ is a continuous, strictly increasing process independent of W with $[X]_0 = 0$. Let $\widehat{[X]}^n$ denote the realised variance of X on an equidistant mesh over $[0,T]$. Then we have

$$\lim_{n\to\infty} \sup_{\substack{t\in[0,T]\\ s\in[0,t]}} |\widehat{C}^{X_n}_{s,t}(u,v) - C^X_{s,t}(u,v)| = 0, \qquad \textit{in probability,}$$

for every $(u,v) \in [0,1]^2$*, where* $\widehat{C}^{X_n}_{s,t}$ *is defined by* (3.1)*.*

Proof. Note, that by Proposition 2.3 and the fact that both $C^X_{s,t}(u,v)$ and $\widehat{C}^{X_n}_{s,t}(u,v)$ are copulas in u, v we have that the boundary convergence is ensured, i.e.

$$C^X_{s,t}(u,v) = \widehat{C}^{X_n}_{s,t}(u,v), \quad \forall (u,v) \in \partial[0,1]^2,$$

with $\partial[0,1]^2$ being $[0,1]\times\{0\} \cup [0,1]\times\{1\} \cup \{0\}\times[0,1] \cup \{1\}\times[0,1]$. Hence, it suffices to show for $(u,v) \in (0,1)^2$. Hence, fix $(u,v) \in (0,1)^2$. Now, define

$$g(s,t) = \begin{cases} \min(u,v) & s \geq t \\ \int_0^u \Phi\left(\frac{\sqrt{t}\Phi^{-1}(v) - \sqrt{s}\Phi^{-1}(w)}{\sqrt{t-s}}\right) dw & s < t \end{cases}$$

Note that the function is continuous on $[0,T]\times[0,T]$; in the case $s > t$ it is trivial. For $s < t$ note that it is a composition of continuous function and $s = t$ follows from Lemma 2.15. By Lemma 2.8 it extends to a mapping $\Psi_g : \mathbb{D}([0,T];\mathbb{R})^2 \to \mathbb{D}([0,T]^2;\mathbb{R})$ which is continuous between $\mathbb{D}([0,T];\mathbb{R})^2$ endowed with the product supremum metric and $\mathbb{D}([0,T]^2;\mathbb{R})$ endowed with the supremum metric. By, Lemma 2.11 we have

$$\widehat{[X]}^n \overset{u.c.p.}{\Rightarrow} [X].$$

By Theorem 2.7 and the continuity of Ψ_g we have

$$\Psi_g(\widehat{[X]}^n, \widehat{[X]}^n) \overset{u.c.p}{\Rightarrow} \Psi_g([X],[X]), \qquad \text{in } \mathbb{D}([0,T]^2;\mathbb{R}),$$

but $\Psi_g([X],[X])(s,t) = C^X_{s,t}(u,v)$ for $s \leq t$, and the result immediately follows. ■

Theorem 3.1 tells us that we may consistently estimate the conditional copula of (X_s, X_t) given $([X]_s, [X]_t)$ by using the realised variance.

We now wish to provide a limit theorem in a very restricted case. In order to do so we consider a process X which has the representation

$$X = W_{T_t}, \quad t \geq 0, \tag{3.3}$$

Where T_t is a strictly increasing and deterministic process. Furthermore, we assume that T is absolutely continuous and that there is σ bounded such that

$$T_t = \int_0^t \sigma_r^2 dr, \quad t \geq 0.$$

We remark that these assumptions are very restrictive. We will later discuss how certain assumptions may be relaxed.

Theorem 3.2. *Let* X *be a process which admits the representation*

$$X_t = W_{T_t}, \quad t \geq 0,$$

and T_t *be a deterministic, strictly increasing and absolutely continuous function satisfying*

$$T_t = \int_0^t \sigma_r^2 dr, \quad t \geq 0$$

for some bounded σ*. Let* $f \in C^1(O;\mathbb{R})$ *where* $O \subseteq [0,\infty)^2$ *is open and non-empty. Fix* $s < t \in [0,\infty)$*. Let*

$$Y = (T_s, T_t), \quad Y_n = (\widehat{[X]}^n_s, \widehat{[X]}^n_t),$$

and assume $Y \in O$*. Then,*

$$\sqrt{n}(f(Y_n) - f(Y)) \xrightarrow{\mathcal{L}_S} \sqrt{V_{s,t}} Z, \quad Z,$$

Z *is defined on a very good filtered extension of* $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,\infty)}, \mathbb{P}\right)$ *and*

$$V_{s,t} = 2\nabla f(Y) \begin{bmatrix} \int_0^s \sigma_r^4 dr & \int_s^t \sigma_r^4 dr \\ \int_s^t \sigma_r^4 dr & \int_0^t \sigma_r^4 dr \end{bmatrix} \nabla f(Y)^\top,$$

where ∇f *is the gradient of* f *and* $\nabla f(Y)^\top$ *is the transpose of* $\nabla f(Y)$*.*

Proof. To avoid confusion, all vectors are considered as "row vectors", therefore if $x \in \mathbb{R}^2$ is a vector then $xx^\top \in \mathbb{R}$ and $x^\top x \in \mathbb{R}^{2\times 2}$, where $x^\top$ denotes the transpose of x.

First, by the Taylor's Theorem, or the Mean-Value Theorem, we have

$$f(Y_n) - f(Y) = \left(\int_0^1 \nabla f(Y + h(Y_n - Y)) dh \right) (Y_n - Y)^\top.$$

By Theorem 2.9 we have

$$\frac{1}{\sqrt{\Delta_n}} (\widehat{[X]}^n - T) \xrightarrow{\mathcal{L}_S} Z,$$

since $[X] = T$. The limit Z is defined on a very good filtered extension of $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,\infty)}, \mathbb{P}\right)$ and is $\mathcal{F}$ conditionally centered Gaussian with second moment

$$\mathbb{E}\left[Z_t^2 \mid \mathcal{F}\right] = 2\int_0^t \sigma_r^4 dr.$$

In particular, this yields

$$\frac{1}{\sqrt{\Delta_n}} (Y_n - Y) \xrightarrow{\mathcal{L}_S} (Z_s, Z_t),$$

where (Z_s, Z_t) is an $\mathcal{F}$ conditional centered Gaussian with second moment

$$\mathbb{E}\left[(Z_s, Z_t)^\top (Z_s, Z_t) \mid \mathcal{F}\right] = 2Q_{s,t},$$

where

$$Q_{s,t} = \begin{bmatrix} \int_0^s \sigma_r^4 dr & \int_s^t \sigma_r^4 dr \\ \int_s^t \sigma_r^4 dr & \int_0^t \sigma_r^4 dr \end{bmatrix}.$$

Now, it suffices to show

$$\int_0^1 \nabla f(Y - h(Y_n - Y)) dh \xrightarrow{\mathbb{P}} \nabla f(Y),$$

since the continuous mapping theorem for stable convergence in law then yields the desired result. Since σ is bounded by assumption, the paths of

$$T_t = \int_0^t \sigma_s^2 ds$$

are Lipschitz continuous. In particular, it follows from [[9], Theorem 4.5][1] that for $0 < s < t$ we have

$$Y_n \overset{a.s.}{\to} Y.$$

Now, since we have by assumption that $Y \in O$ we have can find a radius $r > 0$ such that open ball $B_r(Y) \subseteq O$. Now, let $\Omega_{s,t} = \{\omega \in \Omega \mid Y_n(\omega) \to Y\}$. Clearly, this set has probability one. Now, for every ε such that $r > \varepsilon > 0$, there is an n_0, which depends on ω, such that

$$\|Y_n(\omega) - Y\| < \varepsilon, \quad \forall n \geq n_0(\omega), \omega \in \Omega_{s,t}.$$

By the convexity of a sphere this immediately implies

$$Y + h(Y_n(\omega) - Y) \in B_\varepsilon(Y), \quad \forall h \in [0,1], n \geq n_0(\omega), \omega \in \Omega_{s,t},$$

and in particular we can find a compact set $K_\omega \subseteq O$ such that $Y + h(Y_n(\omega) - Y) \in K_\omega$. By the continuity of ∇f on O, and by extension K_ω, we get that $\nabla f(Y + h(Y_n(\omega) - Y)$ is almost surely bounded for ever $h \in [0,1]$, $n \geq n_0(\omega)$, and $\omega \in \Omega_{s,t}$. The Dominated Convergence theorem then yields

$$\int_0^1 \nabla f(Y + h(Y_n(\omega) - Y))dh \to \nabla f(Y), \quad \forall \omega \in \Omega_{s,t}.$$

Since $\mathbb{P}(\Omega_{s,t}) = 1$ we have convergence in probability. By Lemma 2.9 we have the joint stable convergence and by Lemma 2.10 we have

$$\frac{1}{\sqrt{\Delta_n}}(f(Y_n) - f(Y)) \overset{\mathcal{L}_S}{\to} \nabla f(Y)(Z_s, Z_t)^\top.$$

Now, conditional on $\mathcal{F}$ we have that $(Z_s, Z_t)^\top$ was centered Gaussian. We now have that the second moment of $\nabla f(Y)(Z_s, Z_t)^\top$ is given by

$$2\nabla f(Y) Q_{s,t} \nabla f(Y)^\top$$

■

Note that the theorem is severely restricted. We will later discuss how some of the restrictions may be lifted. In order to apply the result to our estimator we will make use of the function

$$g(s,t) = \int_0^u \Phi\left(\frac{\sqrt{t}\Phi^{-1}(v) - \sqrt{s}\Phi^{-1}(w)}{\sqrt{t-s}}\right) dw, \tag{3.4}$$

which was also applied in Theorem 3.1. Note that g is differentiable on the open set

$$O = \{(s,t) \mid t \in 0 < s < t\}. \tag{3.5}$$

with gradient

[1] The theorem is quite different from the rest of the thesis, and we choose to omit it. Essentially, due to the Lipschitz paths they induce a partition with of order $o(1/\log(n))$, on which we have almost sure convergence for the quadratic variation of a Brownian motion.

$$\nabla g(s,t)^\top = \begin{bmatrix} \dfrac{\partial g(s,t)}{\partial s} \\ \\ \dfrac{\partial g(s,t)}{\partial t} \end{bmatrix} \tag{3.6}$$

$$= \begin{bmatrix} \int_0^u \varphi\left(\frac{\sqrt{t}\Phi^{-1}(v)-\sqrt{s}\Phi^{-1}(w)}{\sqrt{t-s}}\right)\left(\frac{\Phi^{-1}(v)}{2\sqrt{t(t-s)}} - \frac{\sqrt{t}\Phi^{-1}(v)-\sqrt{s}\Phi^{-1}(w)}{2\sqrt{t-s}^3}\right) dw \\ \\ \int_0^u \varphi\left(\frac{\sqrt{t}\Phi^{-1}(v)-\sqrt{s}\Phi^{-1}(w)}{\sqrt{t-s}}\right)\left(\frac{\sqrt{t}\Phi^{-1}(v)\sqrt{s}\Phi^{-1}(w)}{2\sqrt{t-s}^3} - \frac{\Phi^{-1}(w)}{2\sqrt{s(t-s)}}\right) dw \end{bmatrix}. \tag{3.7}$$

Where φ is the density function of a standard Gaussian. Note that (3.7) specifies the transpose of ∇g such that it is a "row vector" in concordance with the remark in the proof of Theorem 3.2. Now, with the gradient we have the immediate corollary.

Corollary 3.1. *Under the assumptions from Theorem 3.2 we have for fixed $(u,v) \in (0,1)^2$ and $(T_s, T_t) \in O$, with O defined by (3.5), then*

$$\frac{1}{\sqrt{\Delta_n}}(C_{s,t}^{X_n}(u,v) - C_{s,t}^{X}(u,v)) \xrightarrow{\mathcal{L}_S} \sqrt{V_{s,t}}Z$$

where Z is defined on a very good filtered extension and is $\mathcal{F}$ conditionally standard Gaussian. $V_{s,t}$ is given by

$$V_{s,t} = 2\nabla g([X]_s, [X]_t) Q_{s,t} \nabla g([X]_s, [X]_t)^\top,$$

where ∇g is given by (3.7) for the fixed (u,v) and $Q_{s,t}$ is as in Theorem 3.2.

We would like to remark that there are some obvious caveats here; ∇g for fixed $(u,v) \in (0,1)^2$ has some cases where it may explode. Of course, our restriction to the open set $\{(s,t) \mid 0 < s < t < T\}$ alleviates these, but for all practical purposes it is still an issue. In particular the case $s \approx 0$ or $s \approx t$ causes immediate problems, as these will cause the gradient to explode. Likewise, if (u,v) tend towards $\partial[0,1]^2$ the gradient may explode.

Now, we may very well estimate $V_{s,t}$ to provide a standardised limit theorem. In order to ease notation we define

$$\widehat{Q}_t = \frac{1}{3\Delta_n} \sum_{i=1}^{[t/\Delta_n]} (|X_{t_i} - X_{t_{i-1}}|)^4,$$

recall that under certain restrictions it converges uniformly on compacts in probability to

$$\widehat{Q}_t \overset{u.c.p.}{\Rightarrow} \int_0^t \sigma_r^4 dr. \tag{3.8}$$

By assumption σ is bounded in our case, and so we meet the assumptions for the convergence (3.8).

Lemma 3.1. *Under the assumptions of Theorem 3.2, we define*

$$\widehat{V}_{s,t} = 2\nabla f(Y_n) \widehat{Q}_{s,t} \nabla f(Y_n)^\top,$$

where

$$\widehat{Q}_{s,t} = \begin{bmatrix} \widehat{Q}_s & \widehat{Q}_t - \widehat{Q}_s \\ \widehat{Q}_t - \widehat{Q}_s & \widehat{Q}_t \end{bmatrix},$$

with $\widehat{Q}_t$ *defined by* (3.8)*. Then*

$$\frac{1}{\sqrt{\Delta_n \widehat{V}_{s,t}}}(f(Y_n) - f(Y)) \overset{\mathcal{L}_S}{\to} Z,$$

where Z *is defined on a very good filtered extension and is* $\mathcal{F}$ *conditionally standard Gaussian.*

Proof. The proof follows directly from the fact that $\widehat{V}_{s,t} \overset{\mathbb{P}}{\to} V_{s,t}$. Applying Lemma 2.9 and Lemma 2.10 yields the desired. ■

The immediate result is that we may estimate the variance of our copula estimate. Note that for certain steps we may have $\widehat{[X]}^n_s = \widehat{[X]}^n_t$ in which ∇g is not well defined, which is of course quite restrictive for practical purposes. Furthermore, numerical stability is again a concern. We would like to remark that the the boundaries $s = 0$ ans $s = t$ are quite uninteresting from a modelling perspective; in the case $s = 0$ we showed that it corresponds to the independence copula, and so one would, or rather should, not use the copula estimate in this case. Similarly, when $s = t$ we showed that the copula reduces to the minimum of (u, v). Again, from the modelling perspective, this situation is rather uninteresting.

For fixed $(u, v) \in (0, 1)^2$ we define our variance estimator as

$$\widehat{V}_{s,t} = 2\nabla g(\widehat{[X]}^n_s, \widehat{[X]}^n_t)\widehat{Q_{s,t}}\nabla g(\widehat{[X]}^n_s, \widehat{[X]}^n_t)^\top. \tag{3.9}$$

In summary, the result is that

$$\frac{1}{\sqrt{\Delta_n \widehat{V}_{s,t}}}(C^{X_n}_{s,t}(u, v) - C^X_{s,t}(u, v)) \overset{\mathcal{L}_S}{\to} N,$$

for N on a suitable extension and fixed (u, v). Furthermore, N is $\mathcal{F}$ conditionally standard Gaussian.

3.2 Numerical Analysis

We now apply numerical analysis to investigate whether the result can be generalised to stochastic volatility. We are interested in both the rate of convergence and whether Corollary 3.1 remains valid. Furthermore, we wish to estimate the rate of convergence for an enhanced problem, i.e.

$$\begin{array}{ll} \sup & |\widehat{C}^{X_n}_{s,t}(u, v) - C^X_{s,t}(u, v)| \\ \text{s.t.} & (u, v) \in [0, 1]^2 \\ & t \in [0, 1] \\ & s \in [0, t] \end{array}. \tag{3.10}$$

In this case our assumption is that we do not have to fix $(u, v) \in [0, 1]^2$ as is the case for Theorem 3.1, but rather that the convergence is also uniform over $(u, v) \in [0, 1]^2$. For practical purposes we implement the following problem, which is equivalent to (3.10).

$$\begin{array}{ll} \sup & |\widehat{C}^{X_n}_{s,s+\Delta}(u,v) - C^X_{s,s+\Delta}(u,v)| \\ \text{s.t.} & (u,v) \in [0,1]^2 \\ & \Delta \in [0,1] \\ & s \in [0, 1-\Delta] \end{array} . \tag{3.11}$$

In order to ease notation we define $\rho(\widehat{C}^{X_n}, C^X)$ to mean 3.11. Now, to estimate the rate of convergence we employ the following:

$$\mathbb{E}[\rho(\widehat{C}^{X_n}, C^X)] \leq Kn^{-\gamma}$$

which implies

$$\log(\mathbb{E}[\rho(\widehat{C}^{X_n}, C^X)]) \leq \log(K) - \gamma \log(n). \tag{3.12}$$

We sample a process from the so-called Cox-Ingersoll-Ross, abbreviated CIR, model, which follows the stochastic differential equation

$$dV_t = \kappa(\theta - V_t)dt + \sigma\sqrt{V_t}dW_t, \quad V_0 = v_0,$$

where κ is the rate of mean reversion, θ is the long-run mean and σ is the volatility of the CIR process. The mean reversion is present due to the fact that if $V_t > \theta$, then the drift term becomes negative, assuming $\kappa > 0$, hence forcing the process closer to θ. In order for the process to stay positive, it must satisfy the so-called Feller condition. That is,

$$2\kappa\theta > \sigma^2,$$

which ensures that the process is positive. We simulate a path from the process and define the process X as

$$X_t = \int_0^t \sqrt{V_t}dW'_t,$$

where W is some Brownian motion independent of V. We then have

$$[X]_t = \int_0^t V_t dt.$$

We consider equidistant discretisations of the process X over the interval $[0,1]$ with 10^n points for $n \in \{100, 1000, 10000, 100000\}$. We sample 1000 paths for each disretisation scheme. We simulate paths under the parameters described in Table 3.1.

Parameter	κ	θ	σ	V_0
Value	0.5	1.5	1	1.5

Table 3.1: Parameters of the CIR model.

We initially consider the rate of convergence and compute the quantity $\rho(\widehat{C}^{X_n}, C^X)$. We calculate a density estimate of the results and plot them alongside the mean, mode, and $1/\sqrt{n}$ to visually inspect the distribution of $\rho(\widehat{C}^{X_n}, C^X)$.

From Figure 3.1 we see that the mean is skewed towards the lower tail, this is caused by values much closer to 0 than depicted in Figure 3.1. Furthermore, we see that as the number of points per path increase, the distribution of $\rho(\widehat{C}^{X_n}, C^X)$ will shift towards 0, with a decrease in variance;

the scales on Figure 3.1 are both logarithmic, so the entire scale is decreased as the number of points per path increases.

Next, we estimate the regression coefficients using (3.12). The results are presented in Table 3.2. Furthermore, we apply a bootstrapping procedure to obtain the confidence quantiles of our estimate.

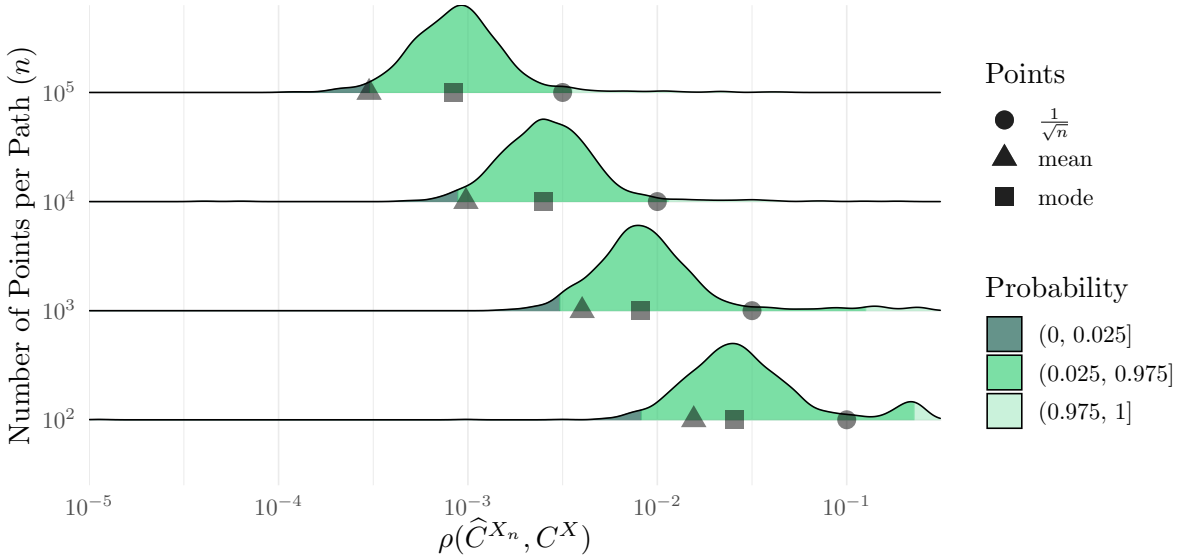

Figure 3.1: Density estimate of $\rho(C^{X_n}, C^X)$.

Estimate of both K and γ						
		Confidence Quantiles				
Variable	Estimate	0%	25%	50%	75%	100%
γ	0.5279	0.5708	0.5356	0.5283	0.5209	0.4813
K	0.4965	0.3629	0.4696	0.4967	0.5258	0.6816
Estimate of γ with $K = 0$						
		Confidence Quantiles				
Variable	Estimate	0%	25%	50%	75%	100%
γ	0.6067	0.6206	0.6096	0.6071	0.6043	0.5903
Estimate of K with $\gamma = 1/2$						
		Confidence Quantiles				
Variable	Estimate	0%	25%	50%	75%	100%
K	0.3966	0.3525	0.3875	0.3956	0.4042	0.4513

Table 3.2: Regression parameters for the order of convergence.

From Table 3.2 we see that based on the bootstrapping procedure we cannot see that the estimate for γ is roughly $1/2$, implying that the order of convergence could be $1/2$. Hence, it may indicate that even in the case of stochastic volatility the rate of convergence is proportional to $1/\sqrt{\Delta_n}$. The visual inspection in Figure 3.2 seems to agree with the estimation. Therefore, it is not

unlikely that the order of convergence is roughly 1/2 implying a rate of convergence proportional to $1/\sqrt{\Delta_n}$.

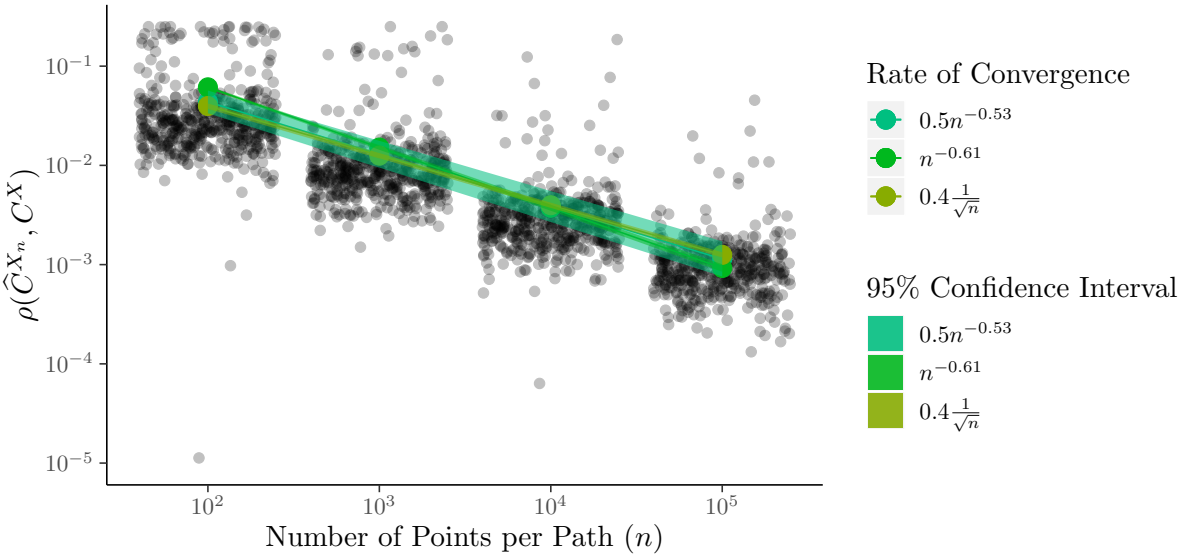

Figure 3.2: Depiction of the rate of convergence.

We now wish to verify Corollary 3.1 for stochastic volatility. In order to investigate, we calculate the variance estimator presented in (3.9) and provide a quantile-quantile plot against the theoretical quantiles of a standard Gaussian distribution.

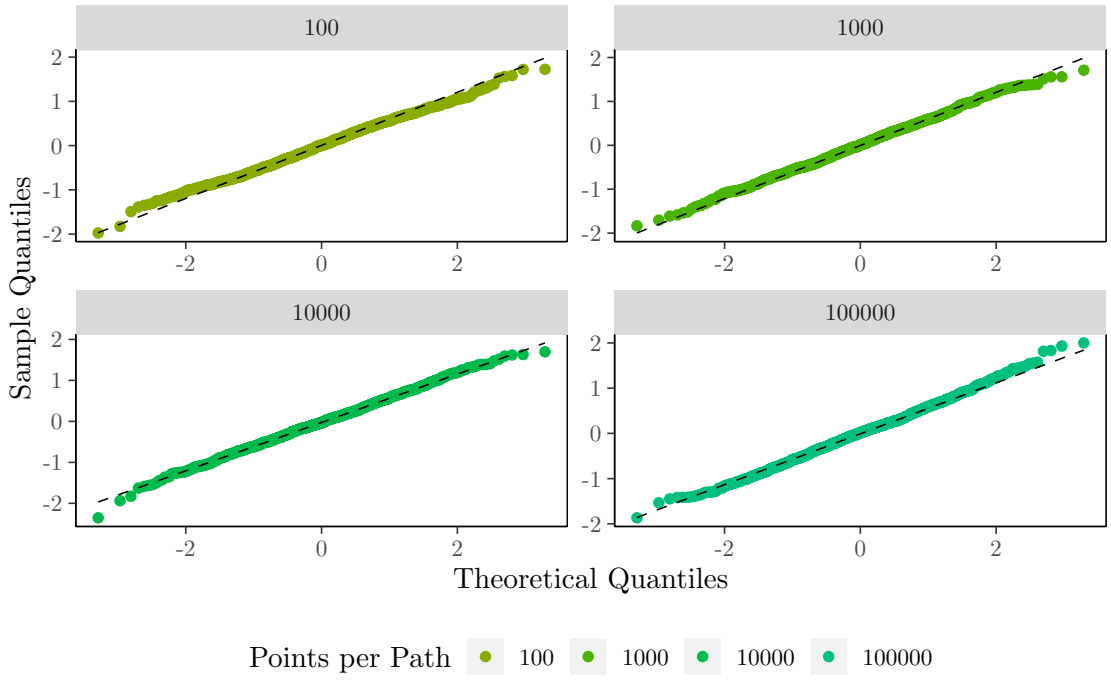

Figure 3.3: Quantile-Quantile Plot of Standardised Errors with $s = 0.3$, $t = 0.7$, $u = 0.4$, $v = 0.6$.

From Figure 3.3 we see that the distribution indeed does seem like it is Gaussian. We now

consider an example where $|t-s|$ tends towards 0. Figure 3.4 does not seem to indicate a Gaussian limit. As discussed, the limit is very sensitive to the boundaries $s \approx t$ and $s \approx 0$. In both cases inspection of the gradient may yields a partial answer, with the gradient in question being (3.7). Note that in (3.7) we have denominator terms such as $\sqrt{t-s}^3$ and $\sqrt{s(t-s)}$. These terms seem in particular troublesome, since they will tend to 0 quite fast in the cases $s \approx t$ and $s \approx 0$. From the figures 3.3 and Figure 3.4, we conclude that while the limit *may* be Gaussian, then the number of points required per path for the Gaussian distribution to become valid depends very much on which (s,t) and (u,v) one wishes to investigate. To some extend this was expected.

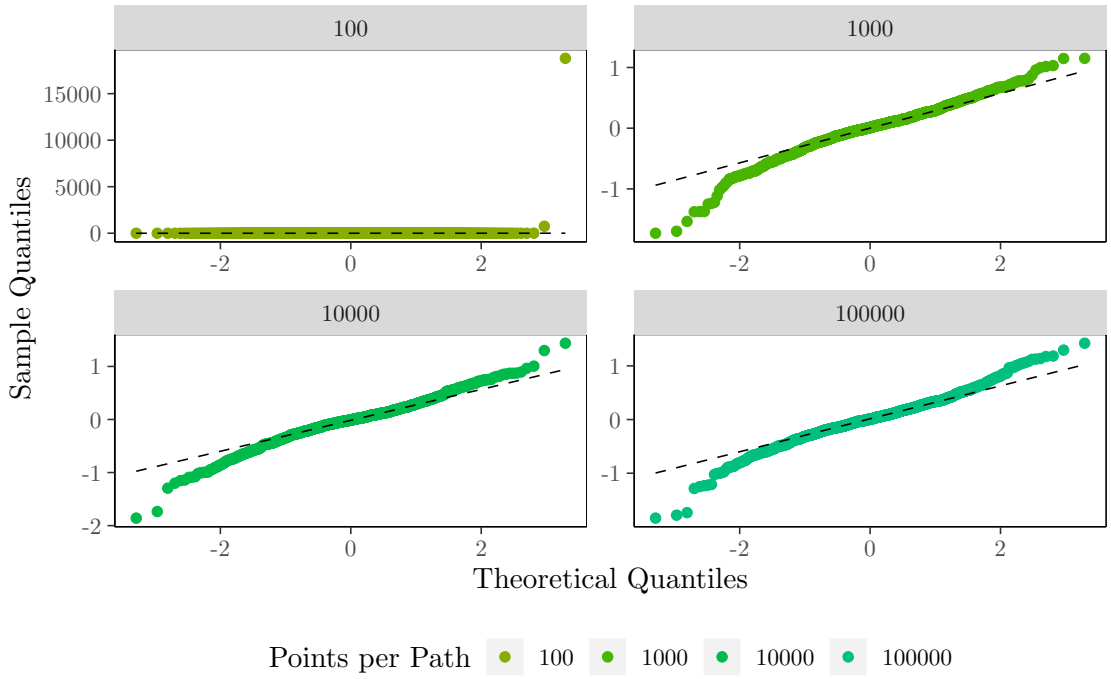

Figure 3.4: Quantile-Quantile Plot of Standardised Errors with $s = 0.45$, $t = 0.55$, $u = 0.4$, $v = 0.6$.

4

Financial Theory

This section is based on [13] and [33].

We start with the definition of a financial market.

Definition 4.1 (Financial Market)
Let $T \in (0, \infty)$. Let $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,T]}, \mathbb{P}\right)$ be a filtered probability space satisfying the usual conditions. Let $\mathcal{SM}$ be the space of semimartingales on the probability space.

Numéraire: Let $A \in \mathcal{SM}$. A is said to be a *numéraire* if

$$\mathbb{P}(A_t > 0, \forall t \in [0,T]) = 1.$$

Financial Market: Let $P \subset \mathcal{SM}$ be a finite collection of of $n+1$ semimartingales; n assets $S^i, i \in \{1, 2, \ldots, n\}$ and a unique numéraire A. A finite horizon financial market is the pair

$$\mathscr{M} = \left(\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,T]}, \mathbb{P}\right), P\right).$$

Intuitively, the numéraire is the risk-free assets which could either be the interest rate from a bank or government bonds.

Now, there are several ways to consider a portfolio mathematically. One may think of them as weights, and for non-temporal treatments such as Markowitz portfolio theory, this works well. In a sense, we wish to generalise this thought. However, we now need our "weights" to be processes themselves.

Let us consider a single asset S assumed to be semimartingale. Suppose we wish to buy the asset at time t_1 and sell it at time t_2. Our portfolio would then satisfy

$$\mathbb{1}_{(t_1,t_2]}(t) S_t, \quad \forall t \in [0,T].$$

If we wish to know our accumulated profit or loss, then we can do this via integration, in this particular instance we have a simple integrand and so

$$\int_0^t \mathbb{1}_{(t_1,t_2]}(s) dS_s = \mathbb{1}_{(t_1,t_2]} \bullet S_t = S_{t_2} - S_{t_1}.$$

Such a strategy is called a *buy-and-hold* strategy. Now, more sophisticated strategies can be approximated by buy-and-hold strategies which may be random. Here stopping times may naturally occur. For instance, a so-called *stop-loss* strategy may be implemented by the use of the stopping time

$$\tau(\omega) = \inf\{t \in [0,T] \mid S_t(\omega) \leq K\}, \quad K \in \mathbb{R}$$

Here K is some level at which the investor does not want to hold the asset. In general, we will assume that the portfolio or integrands are S integrable and that they are predictable. The predictable requirement must intuitively be satisfied since else the investor would be able to follow spontaneous movements in the market. In other words, the investor has to know what he is going to do at time t prior to time t.

Through the previous sections, we also know that simple integrands form a basis for stochastic integration. Hence, it would seem that the theory aligns perfectly with the mathematical theory. However, recall that we extended our definition beyond simple integrands to processes which are locally bounded and predictable in Theorem 2.8. If we allow these arguably more sophisticated class of strategies, then some problems arise.

One problem is the so-called *doubling* strategies. Consider a game where we bet on a fair coin showing heads. If we bet $\$X$ and the coin shows head, we receive $\$2X$, and if it shows tails, we get \$0. Suppose we start with betting \$1 and if we win then we get \$2 and hence \$1 profit. If we lose, then we bet \$2 and our total investment amounts to \$3, if we then win then we get \$4 and a \$1 profit, else we double our investment again. Since heads will eventually show up given that the probability is strictly greater than 0, then this strategy will, with probability 1, yield \$1 profit. The strategy requires infinite capital, which is unrealistic. This particular gambling example is called a martingale. The problem is solved mathematically by defining so-called *admissible strategies*.

Definition 4.2 (Admissible Strategies)
Let $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,T]}, \mathbb{P}\right)$ be a probability space and S a semimartingale. We say that a predictable process H is admissible or an admissible strategy for S if

$$H \bullet S \text{ is well defined.}$$

and there exists a constant $M > 0$

$$H \bullet S_t \geq -M, \quad \forall t \in [0,T]$$

The constant M may be thought of as a margin of leverage. In finance many brokers or banks utilise *margin-calls* exactly as in Definition 4.2.

Margin calls are typically used in the following context. Suppose an investor finances his investment partially with equity and partially with a loan. He will need to make payments on his loan since it is an independent financial contract. Now suppose he buys into an asset S with half financed by equity and half financed by a loan. Assume that the investment drops in value that is $S_{t_1} > S_{t_2}$ where $t_1 < t_2$ - this is similar to a losing streak of tails in the doubling strategy. From the investors perspective, this means that his equity has dropped in value, but his loan remains the same. If the investment drops too much in value the bank, or broker, will

call for extra margin to even the level between equity and loan. Otherwise, the bank will carry all the risk.

While there are several other reasons, we will use margin calls to motivate why an investor is interested in not only forecasting his expected return but is interested in forecasting the entire distribution of the asset. Suppose an asset S has value S_0 at the date of purchase, and that the investor has financed this purchased with leverage, either from a bank or short-selling other assets and that the investor, therefore, is subject to a margin M. In mathematics, we can define our set of admissible portfolios to satisfy $H \bullet S_t \geq -M$. However, in practice, this is an actual problem for the investor.

Suppose for simplicity, that the investor decides on a buy-and-hold strategy buying at $t_0 = 0$ and selling at time T, for some $T > 0$. Hence, in order to avoid the margin call, he must know whether or not the asset will drop sufficiently to incur the margin call. Of course, with assets being inherently risky and with the potential for loss of investment, it is better to phrase the question in terms of probability. In this particular instance, suppose that the margin call is incurred if the asset drops below some price S^*. In order to assert his or her risk, the investor will have to know for each $t \in [0, T]$ the probability that $S_t \leq S^*$. However, as time progresses, these probabilities change, and so arises the need for conditional distributional forecasts.

Single assets are one thing, but the same applies to entire portfolios; if our portfolio is financed partially with equity and partially with a loan, then the investor faces the same problem. Suppose we are given a market $\mathscr{M}$ with a collection of $d+1$ assets, with one asset being the numéraire, A, and the remaining d are risky assets, S^i for $i \in \{1, 2, \ldots, d\}$. Then, a strategy may be considered as a $d+1$ dimensional process of predictable processes $\theta = (\phi, \pi^1, \pi^2, \ldots, \pi^d)$ such that ϕ is integrable with respect to the numéraire and π_i is integrable with respect to asset $i \in \{1, 2, \ldots, d\}$. We may then define a portfolio as

$$V_t^\theta = \phi_t A_t + \sum_{i=1}^{d} \pi_t^i S_t^i, \quad t > 0.$$

Typically, we require the portfolio to be self-financed. It must then satisfy

$$V_t^\theta = \theta \bullet A_t + \sum_{i=1}^{d} \pi^i \bullet S_t^i, \quad t > 0.$$

In this scenario we have that V_t^θ is itself a semimartingale, and for a given θ is may be expressed as

$$V_t^\theta = V_0^\theta + Z_t + M_t, \quad t > 0.$$

V_0^θ can be thought of as the initial investment, Z_t a process of locally bounded variation and M a martingale. For large portfolios there is a numerical issue in assessing the risk of our portfolio. By making some structural assumptions, we may use the copula derived before to forecast the distribution of our portfolio. Namely, we will assume it is of the form

$$V_t^\theta = V_0^\theta + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s.$$

For simplicity, we assume the processes μ and σ are continuous and independent of W. We remark that these assumptions are unlikely to be true. However, one may argue that if the

portfolio is sufficiently well diversified, then the portfolio should be less influenced by jumps in each asset. Hence, the continuity of the portfolio process may be more valid, than in the individual asset case.

Being able to make good forecasts of ones portfolio solves multiple problems. If one has used external capital such as a loan to finance investments, then a distributional forecast may help prevent incurring a margin call. Furthermore, if the computational burden is feasible then it may be used to reduce the complexity of portfolio selection. Of course, by omitting the intricacies of the dependence between assets a lot of valuable information is lost, but from a portfolio selection standpoint these should be incorporated once we treat the portfolio as its own entity.

We would like to remark that at the moment we are not looking for the minimal assumptions necessary for the procedure to work and throughout the following procedure we assume that the processes μ and σ satisfy the criterion of the applied theorems.

4.1 Forecasting Procedure

We propose a heuristic forecasting procedure of the distribution of a portfolio. The forecasting procedure is adaptable, and we will later discuss how to augment the procedure to accommodate the shortcomings of the proposed method. The method proposed here also showcases how copulas may be applied to the temporal treatment. We will benchmark the forecasting procedure on the Heston model, [11]. While it is unrealistic that an entire portfolio would follow the dynamics of the Heston model we use the model to showcase how one can proceed in using the copula of a time-changed Brownian motion. We will later discuss whether or not is justified to apply this heuristic method. We make the further structural assumptions on our portfolio:

$$V_t^\theta = V_0^\theta + \int_0^t \mu V_r^\theta dr + \int_0^t \sigma_r V_r^\theta dW_r,$$

with σ some continuous stochastic process independent of W and $\mu \in \mathbb{R}$ a constant. For simplicity, we fix our strategy θ. By application of Ito's lemma we obtain

$$X_t = \ln(V_t^\theta) = X_0 + \int_0^t \left(\mu - \frac{\sigma_r^2}{2}\right) dr + \int_0^t \sigma_r dW_r. \tag{4.1}$$

Now, the quadratic variation is given by

$$[X]_t = \int_0^t \sigma_r^2 dr,$$

Now, let $M_t = \int_0^t \sigma_r dW_r$. We can estimate the conditional copula M. However, this highlights a glaring problem which we have so far ignored; in order to model the copula of (M_s, M_t) we need to know $[M]_t$, on which our particular estimator $[M]_t^n$ requires knowledge of M_t. Hence, our problem becomes rather circular.

However, due to the structural assumptions on X, we can employ the following heuristic forecasting method for the quadratic variation

$$\begin{aligned}[M]_t &= [X]_t \\ &= \int_0^t \sigma_r^2 dr\end{aligned}$$

$$= \int_0^s \sigma_r^2 dr + \int_s^t \sigma_r^2 dr$$
$$\approx [X]_s + (t-s) \cdot \sigma_s^2,$$

which of course converges as $s \to t$. However, we have now traded one problem for another; we now need an estimator of σ_s. Luckily, we may apply the following theorem.

Theorem 4.1. *Let X be an Itô semimartingale of the form*

$$X_t = X_0 + \int_0^t \mu_r dr + \int_0^t \sigma_r dW_r,$$

where μ is locally bounded and σ is continuous. Let $(t_i)_{i=0}^n$ be an equidistant partitioning of $[0,T]$ with $|t_{i+1} - t_i| = \Delta_n$ and $t_0 = 0$ satisfying $\Delta_n \to 0$ and define $\Delta_n^i X = X_{i\Delta_n} - X_{(i-1)\Delta_n}$ and $(k_n)_{n\in\mathbb{N}}$ be a sequence such that $k_n \to \infty$ and $k_n\Delta_n \to 0$. Define the realised spot volatility, abbreviated RSV,

$$(\widehat{\sigma}_t^n)^2 = \frac{1}{\Delta_n k_n} \sum_{i=0}^{k_n-1} (\Delta_n^{i+m-k_n} X)^2, \quad t \in [t_m, t_{m+1}).$$

Then

$$(\widehat{\sigma}_t^n)^2 \xrightarrow{\mathbb{P}} \sigma_t^2, \quad t \in [0,T].$$

An enhanced statement and proof can be found in [[13], pp. 255-256], in particular when σ is cádlág this estimator converges to σ_{t-} rather than σ_t. However, we assumed that σ is continuous and it follows that $\sigma_{t-} = \sigma_t$. From the perspective of our forecasting procedure this has little relevance.

Now, we define our forecast as

$$\widehat{[X]}_{t,h}^n = \underbrace{\widehat{[X]}_t^n}_{\text{RV}} + \underbrace{h}_{\text{Time Step}} \cdot \underbrace{(\widehat{\sigma}_t^n)^2}_{RSV} \tag{4.2}$$

n particular, σ can be viewed in terms of the Fundamental Theorem of Calculus as the "derivative" of $[X]_t$, which is also reflected in the construction of the estimator. For instance,

$$\frac{1}{\Delta_n}(\Delta_n^m X)^2$$

could be considered as an estimator for the derivative. However, due to the fact that our data is contaminated by noise we need to "smooth" the estimator, so that it is not fitting to noise, this is accomplished by averaging over the previous k_n observations yielding

$$\frac{1}{\Delta_n k_n} \sum_{i=0}^{k_n-1} (\Delta_n^{i+m-k_n} X)^2,$$

which is exactly our estimator. Hence, we extrapolate by assuming that the quadratic variation will grow at the same rate as it currently is, this assumption is of course a stretch, but in short terms it may be applicable. In order to investigate the if this method yields anything we will compare it to the simplest forecast, compare it to a "perfect forecast", where we are given $\widehat{[X]}_{t+h}$.

Since X is a semimartingale we know that it be definition decomposes to

$$X = X_0 + M + A,$$

where X_0 is the initial value, M the martingale part and A the process of locally bounded variation. Again, in our case each process is continuous by assumption. We have accounted for M by means of the quadratic variation and conditional copula, but we still need to account for A. However, note that by (4.1) A decomposes into

$$A_t = \int_0^t \mu ds - \frac{1}{2}\int_0^t \sigma_s^2 ds = \mu t - \frac{1}{2}\int_0^t \sigma_s^2 ds$$

Then we have the following

$$X_t - X_0 + \frac{1}{2}\int_0^t \sigma_s^2 ds = \mu t + \int_0^t \sigma_r dW_r.$$

Now, estimation of μ has not been our primary concern. It is beyond the scope of this thesis to solve every aspect of the problem. Therefore, we now rely on heuristics to sketch the idea. Now, heuristically for small h we have

$$X_{t+h} - X_t \approx \mu \cdot h - \frac{1}{2}\sigma_t^2 \cdot h + \sigma_t(W_{t+h} - W_t), \tag{4.3}$$

which implies

$$X_{t+h} \approx X_t + \mu \cdot h - \frac{1}{2}\sigma_t^2 \cdot h + \sigma_t\sqrt{h}N, \quad N \sim \mathcal{N}(0,1), \tag{4.4}$$

so in a very heuristic manner we have X_{t+h} is approximately Gaussian given information of the other processes up until time t. This practice is not uncommon, and many pseudo-maximum likelihood estimators rely on inference made on the discretisation of a process, see e.g. [12] or [6]. However, in our case we acknowledge that we are not applying best practices, as we are using estimates of σ in the place of true σ. We would like to remark that there are better procedures, see e.g. the comments made in [2].

Equation (4.4), then yields

$$\mathbb{E}\left[X_{t+h} - X_t + \frac{1}{2}\sigma_t^2 h\right] \approx \mu h,$$

and so

$$\frac{\mathbb{E}\left[X_{t+h} - X_t + \frac{1}{2}\sigma_t^2 h\right]}{h} \approx \mu,$$

by replacing σ^2 with its estimator yields the estimator for μ

$$\widehat{\mu} = \frac{1}{\Delta_n n}\left(\sum_{i=1}^{n} X_{t_i} - X_{t_{i-1}} + \frac{1}{2}\Delta_n(\widehat{\sigma}^n)^2_{t_{i-1}}\right).$$

However, the sum is telescoping and we obtain

$$\widehat{\mu} = \frac{1}{\Delta_n n}\left(X_{t_n} - X_0 + \sum_{i=1}^{n}\frac{1}{2}\Delta_n(\widehat{\sigma}^n)^2_{t_{i-1}}\right).$$

However, we would like to only use the information up until time t, so we define

$$\widehat{\mu}^n_t = \frac{1}{\Delta_n[t/\Delta_n]}\left(X_{t_{[t/\Delta_n]}} - X_0 + \sum_{i=1}^{[t/\Delta_n]}\frac{1}{2}\Delta_n(\widehat{\sigma}^n)^2_{t_{i-1}}\right).$$

Now, we can transform X to the unit interval $[0,1]$ by the transformation

$$\widehat{U}_t = \Phi\left(\frac{X_t - X_0 - \widehat{\mu}_t^n t + \frac{1}{2}\widehat{[X]}_t^n}{\sqrt{\widehat{[X]}_t^n}}\right). \tag{4.5}$$

We have that $\widehat{U}_s$ and $\widehat{U}_t$ *approximately* follows the copula $C_{s,t}^M$, which we can approximate by $C_{s,t}^{M_n}$.

To forecast we use the estimator of $[M]_{t+h}$ defined in (4.2). Hence, we may find our $\alpha \in (0,1)$ percentile forecast by taking

$$\widehat{U}_{t+h}^{\alpha} = Q_{t,t+h}^M(\widehat{U}_t, \alpha \mid \widehat{[M]}_t, \widehat{[M]}_{t+h}). \tag{4.6}$$

with $Q_{s,t}^M$ defined by (2.32). To bring it back to scale we apply the transformation and obtain the α quantile we apply the forecasted marginal quantile function

$$\widehat{X}_{t+h}^{\alpha} = \sqrt{\widehat{[M]}_{t,h}}\Phi^{-1}(\widehat{U}_{t+h}^{\alpha}) + \widehat{\mu}_t^n \cdot (t+h) + \frac{1}{2}\widehat{[M]}_{t,h}^n + X_0. \tag{4.7}$$

We can take it further and apply the exponential to obtain an estimate of the α-quantile of V_{t+h}^{θ}. In order to validate this approach we need a testing procedure.

4.2 Conditional Coverage Test

This section is based on [8].

We remark, that we only briefly go through the test as a testing procedure has not been our primary focus. Therefore, we deem it beyond the scope of the thesis to derive the procedure and refer to [8] for a detailed derivation.

The setup of the testing procedure is as follows, let $p \in (0,1)$ denote the desired coverage probability. Let $(\Omega, \mathcal{F}, \mathcal{F}_{[0,\infty)}, \mathbb{P})$ be a filtered probability space. We observe a a process at discrete times $(y_{t_i})_{i=1}^N$ and produce the conditional interval forecasts $([L_{t_i|t_{i-1}}(p), U_{t_i|t_{i-1}}(p)])_{i=1}^N$, where $L_{t_i|t_{i-1}}(p)$ represents the conditional forecast of the lower bound of the interval given some filtration $\mathcal{G}_{t_{i-1}} \subseteq \mathcal{F}_{t_{i-1}}$ with $\mathcal{G}_0 \subseteq \mathcal{F}_0$ the trivial σ-algebra augmented by $\mathbb{P}$-null sets and possibly initial value of the process y. Now, let

$$I_{t_i} = \begin{cases} 1 & y_i \in [L_{t_i|t_{i-1}}(p), U_{t_i|t_{i-1}}(p)] \\ 0 & \text{else.} \end{cases}$$

We say that the interval forecast $([L_{t_i|t_{i-1}}(p), U_{t_i|t_{i-1}}(p)])_{i=1}^N$ is efficient with respect to a filtration $(\mathcal{G}_t)_{t\in[0,\infty)}$ if $\mathbb{E}\left[I_{t_i} \mid \mathcal{G}_{t_{i-1}}\right] = p$ for every $i \in \{1,2,\ldots,N\}$. We have the following lemma from [8].

Lemma 4.1. *Testing $\mathbb{E}\left[I_{t_i} \mid I_{t_{i-1}}, \ldots, I_{t_1}\right] = p$ for every $i \in \{1,2,\ldots,N\}$ is equivalent to testing if the sequence $(I_{t_i})_{i=1}^N$ is identically and independently distributed Bernoulli with parameter p.*

See [[8], Lemma 2] for a proof. Hence, from the assumption that our forecasts are efficient we obtain the likelihood

$$L(p \mid (I_i)_{i=1}^N) = (1-p)^{n_0} p^{n_1},$$

where p is the coverage parameter, n_0 is the number of times I_{t_i} is equal to zero and similarly n_1 is the number of times I_{t_i} is equal to 1. Now, similarly,

$$L(\pi \mid (I_{t_i})_{i=1}^N) = (1-\pi)^{n_0} \pi^{n_1}.$$

The test for unconditional coverage is then the likelihood ratio test

$$\begin{aligned}
\Lambda_{UC} &= -2\log\left(\frac{L(p \mid (I_{t_i})_{i=1}^N)}{L(\widehat{\pi} \mid (I_{t_i})_{i=1}^N)}\right) \\
&= -2\log\left(\frac{(1-p)^{n_0} p^{n_1}}{(1-\widehat{\pi})^{n_0}\widehat{\pi}^{n_1}}\right) \\
&= -2n_0\log(1-p) - 2n_1\log(p) + 2n_0\log((1-\widehat{\pi})) + 2n_1\log(\widehat{\pi}).
\end{aligned}$$

which asymptotically follows a $\chi^2(1)$ distribution as $N \to \infty$. The maximum likelihood estimator, $\widehat{\pi}$, of π is given by

$$\widehat{\pi} = \frac{n_1}{n_0 + n_1}. \tag{4.8}$$

Furthermore, we can test for independence between forecasts using a similar approach; we assume the forecasts form a Markov-chain. The general Markov chain transition matrix is given by

$$\Pi = \begin{bmatrix} \pi_{0,0} & \pi_{0,1} \\ \pi_{1,0} & \pi_{1,1} \end{bmatrix},$$

Where $\pi_{i,j}$ is the probability of the interval forecast transitioning from state $i \in \{0,1\}$ to state $j \in \{0,1\}$. Π has the likelihood

$$L(\Pi \mid (I_i)_{i=1}^N) = (1-\pi_{0,1})^{n_{0,0}} \pi_{0,1}^{n_{0,1}} (1-\pi_{1,1})^{n_{1,0}} \pi_{1,1}^{n_{1,1}}.$$

with $n_{i,j}$ being the number of times we go from state j to state i. Maximum likelihood yields the estimator

$$\widehat{\Pi} = \begin{bmatrix} \dfrac{n_{0,0}}{n_{0,1}+n_{0,0}} & \dfrac{n_{0,1}}{n_{0,1}+n_{0,0}} \\ \dfrac{n_{1,0}}{n_{1,0}+n_{1,1}} & \dfrac{n_{1,1}}{n_{1,0}+n_{1,1}} \end{bmatrix}.$$

Under the independence assumption, the Markov chain has the transition matrix

$$\Pi_\perp = \begin{bmatrix} 1-\pi_\perp & \pi_\perp \\ 1-\pi_\perp & \pi_\perp \end{bmatrix},$$

Now, the likelihood ratio test is similarly given by

$$L(\pi_\perp \mid (I_i)_{i=1}^N)) = (1-\pi_\perp)^{(n_{0,0}+n_{1,0})} \pi_\perp^{(n_{0,1}+n_{1,1})},$$

where the maximum likelihood estimate of $\pi_\perp$ is given by $\widehat{\pi}_\perp = (n_{0,1}+n_{1,1})/(n_{0,0}+n_{0,1}+n_{1,0}+n_{1,1})$. Now, the likelihood ratio for independence is given by

$$\begin{aligned}
\Lambda_\perp &= -2\log\left(\frac{(1-\widehat{\pi}_\perp)^{(n_{0,0}+n_{1,0})}\widehat{\pi}_\perp^{(n_{0,1}+n_{1,1})}}{(1-\widehat{\pi}_{0,1})^{n_{0,0}}\widehat{\pi}_{0,1}^{n_{0,1}}(1-\widehat{\pi}_{1,1})^{n_{1,0}}\widehat{\pi}_{1,1}^{n_{1,1}}}\right) \\
&= -2(n_{0,0}+n_{1,0})\log(1-\widehat{\pi}_\perp) - 2(n_{0,1}+n_{1,1})\log(\widehat{\pi}_\perp)
\end{aligned}$$

$$+ 2n_{0,0}\log(1-\widehat{\pi}_{0,1}) + 2n_{0,1}\log(\widehat{\pi}_{0,1}) + 2n_{1,0}\log(1-\widehat{\pi}_{1,1}) + 2n_{1,1}\log(\widehat{\pi}_{1,1}).$$

The latter is preferable when N is large for numerical stability, since q^N goes to 0 very quickly when $q \in [0,1)$. We again have $\Lambda_\perp$ is asymptotically $\chi^2(1)$ distributed. Finally, we can present the conditional coverage.

$$\Lambda_{CC} = -2\log\left(\frac{L(p \mid (I_i)_{i=1}^N)}{L(\widehat{\Pi} \mid (I_i)_{i=1}^N)}\right),$$

which asymptotically follows a $\chi^2(2)$ distribution. Expansion yields

$$\begin{aligned}\Lambda_{CC} &= -2n_0\log(1-p) - 2n_1\log(p) \\ &\quad + 2n_{0,0}\log(1-\widehat{\pi}_{0,1}) + 2n_{0,1}\log(\widehat{\pi}_{0,1}) + 2n_{1,0}\log(1-\widehat{\pi}_{1,1}) + 2n_{1,1}\log(\widehat{\pi}_{1,1}).\end{aligned}$$

Now, Christoffersen notes that if one conditions on the first observation, then

$$\Lambda_{CC} = \Lambda_{UC} + \Lambda_\perp.$$

We end with the remark, that all proofs can be found in [8].

The testing procedure here only describes the relationship between forecasts of a single step, in our case it implies that we can only test forecasts with a fixed step length. However, as we are forecasting multiple steps ahead, which implies that we cannot, and would not, expect intra-step independence. Therefore, in order to validate the apply the following procedure.

Suppose we wish to forecast $I_{t+h|t}$; in our case $h = k\Delta_n$ for some $k \in \mathbb{N}$, on a mesh over $[0,T]$ with $\Delta_n = T/n$. We then group the times modulo k, i.e. the first group would be $t_0, t_k, t_{2k}\ldots$, the second group would be t_1, t_{k+1}, t_{2k+1}.

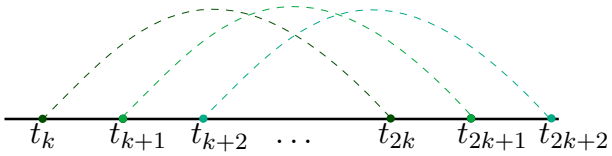

Figure 4.1: Visualisation of the Grouped Times Modulo k.

The result is a total of k groups where the step length between each observation is exactly $h = k\Delta_n$, and hence the conditional coverage test may be applied in each individual instance. In order to report the accuracy, we report the total number of times the test could not be rejected per group. For $k = 1$, it reduces to the conditional coverage test, but for $k > 1$ we have will have multiple groups per simulation. In our simulations study, we will simply report the number of times the test could not be rejected over all groups and simulations.

4.3 Simulation Study

Before conducting the simulation study we go through the procedure step-by-step, which will highlight some of the issues with this method. We will address some of these issues in our discussion at a later stage.

Example 4.1. We simulate a path from the Heston model, i.e.

$$dV_t = \mu V_t dt + \sqrt{\sigma_t^2} V_t dW_t^1$$
$$d\sigma_t^2 = \kappa(\theta - \sigma_t^2)dt + \xi\sqrt{\sigma_t^2} dW_t^2,$$

where $\mu \in \mathbb{R}$, $\xi > 0$, $\theta > 0$ and W_1 and W_2 are standard Brownian motions with quadratic co-variation $[W_1, W_2]_t = \rho t$, i.e. the Brownian Motions are correlated with correlation ρ.

Here, V represents our portfolio and σ^2 its volatility. The portfolio drifts with rate μ. The model for the volatility is the Cox-Ingersoll-Ross model, which we also used in the assessment of our proposed estimator. Here, ξ is the volatility of the process σ^2, θ represents the long-run mean of σ^2, and κ is the speed of mean reversion.

Using Euler discretisation, we simulate the path of an asset under the Heston-model with the following parameters.

Parameter	κ	θ	ξ	μ	ρ	V_0	σ_0^2
Value	2	0.025	0.02	0.1	0	20	0.025

Table 4.1: Simulation Parameters

We simulate the asset over $[0, 1]$ with $N = 5760$ observations. In particular, if $[0, 1]$ represents a trading day of 8 hours, then and $(t_i)_{i=1}^N$ is an equidistant covering of $[0, 1]$ then t_i corresponds to every fifth second on a given trading day.

First, we simulate trajectories of the our portfolio. The paths of the portfolio value V, the log value $X = \log(V)$, and volatility σ^2 are visualised in Figure 4.2.

Next, we calculate the quadratic variation and compare it with the integrated volatility process. We see that there is visually no difference between the realised variance and the true quadratic variation, as depicted in Figure 4.3. We proceed to calculate the estimator for σ^2.

From Figure 4.4 we see that the estimator for σ^2 varies greatly with the choice of k_n. Here k_n is written in parenthesis next to the estimator to emphasize what k_n was chosen, i.e. $\widehat{\sigma}_t^2(64)$ implies $k_n = 64$. Recall that k_n can be considered as a smoothing factor; the higher k_n the more secant lines we average over. Furthermore, the dashed line represents the time at which there has been k_n observations. Prior to the dashed line σ is estimated on all available data.

For the mean we relied heavily on heuristics and it is quite noticeable from Figure 4.5 that the estimator does not enjoy the same properties as the previous estimators. We provide an 80% confidence interval based on bootstrapping. We do not bootstrap new confidence intervals for every single observation but rely on linear interpolation yield accurate confidence bands. It does seem that the true value of μ is, mostly, contained in the 80% confidence interval, but it is also clear that the estimator does not seem to be very robust. However, it seems that the estimator of σ does not influence the estimation very much, which is no surprise given that it emphasizes only the first and last observation specifically.

Finally, we visually evaluate the forecasting of the quadratic variation by extrapolation of the current estimated spot volatility.

From Figure 4.6 we see that the forecasting procedure works quite well in this case. The subscript indicates how many steps ahead the forecast is done, i.e. $\widehat{[X]}_{t,50}^n$ corresponds to forecasting 50

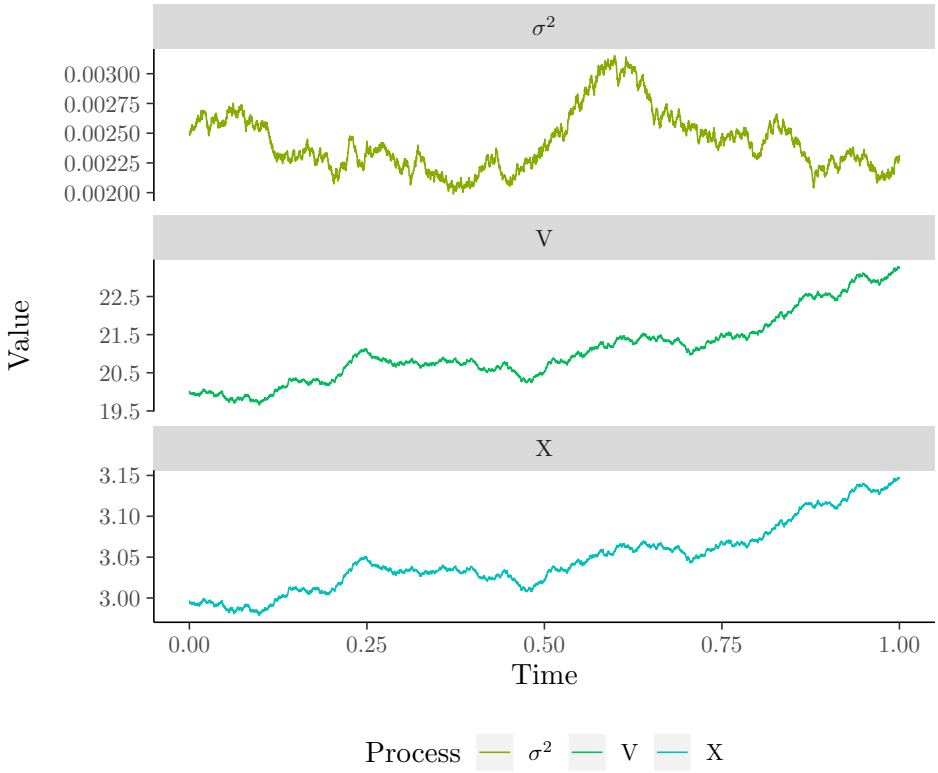

Figure 4.2: Sample Paths in the Heston Model.

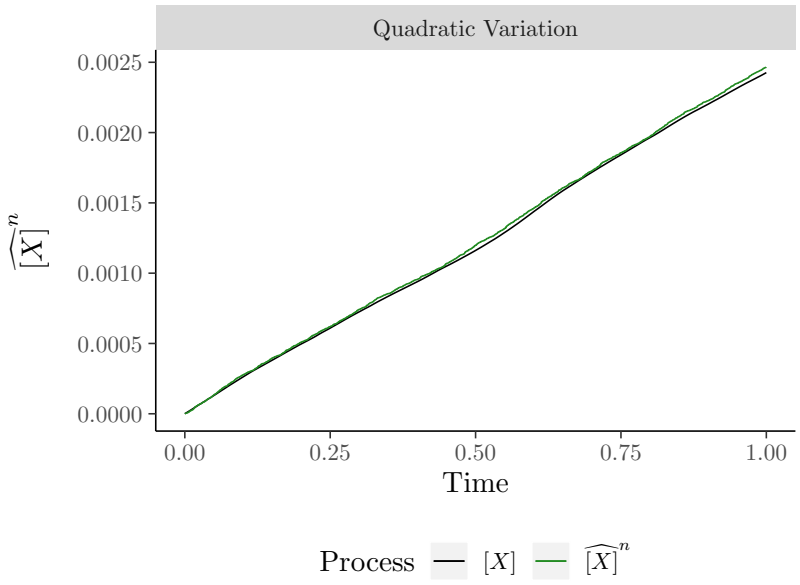

Figure 4.3: Comparison of the Integrated Volatility, or Quadratic Variation, with the Estimated Realised Volatility.

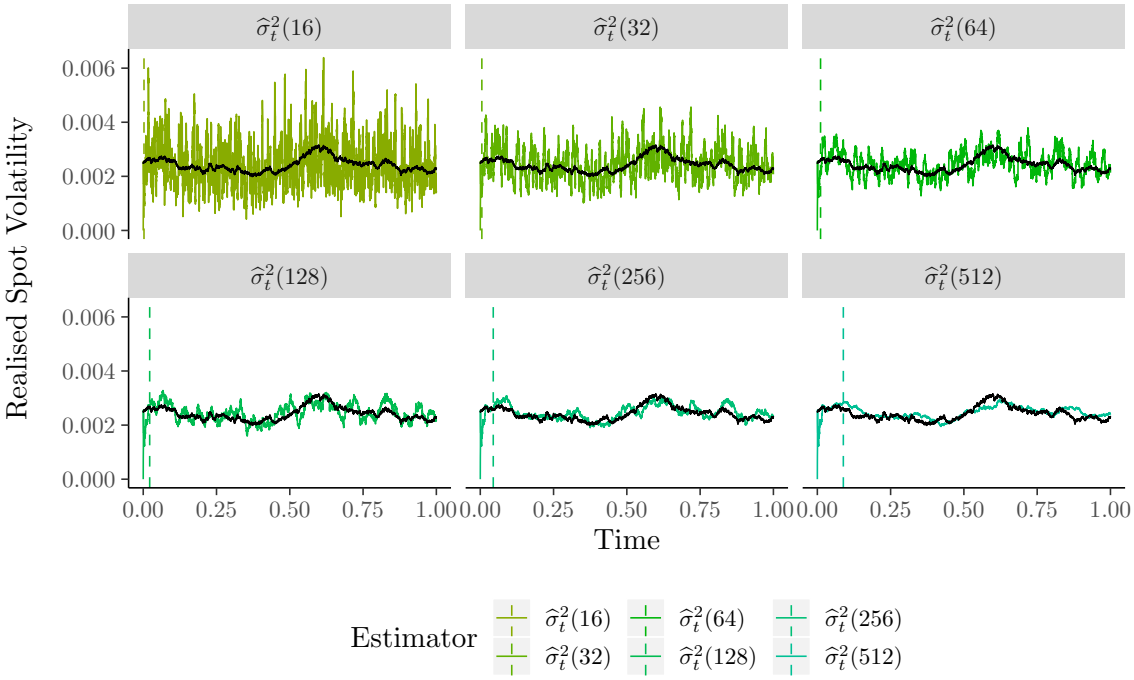

Figure 4.4: $\widehat{\sigma}_t^2(k_n)$, with n fixed.

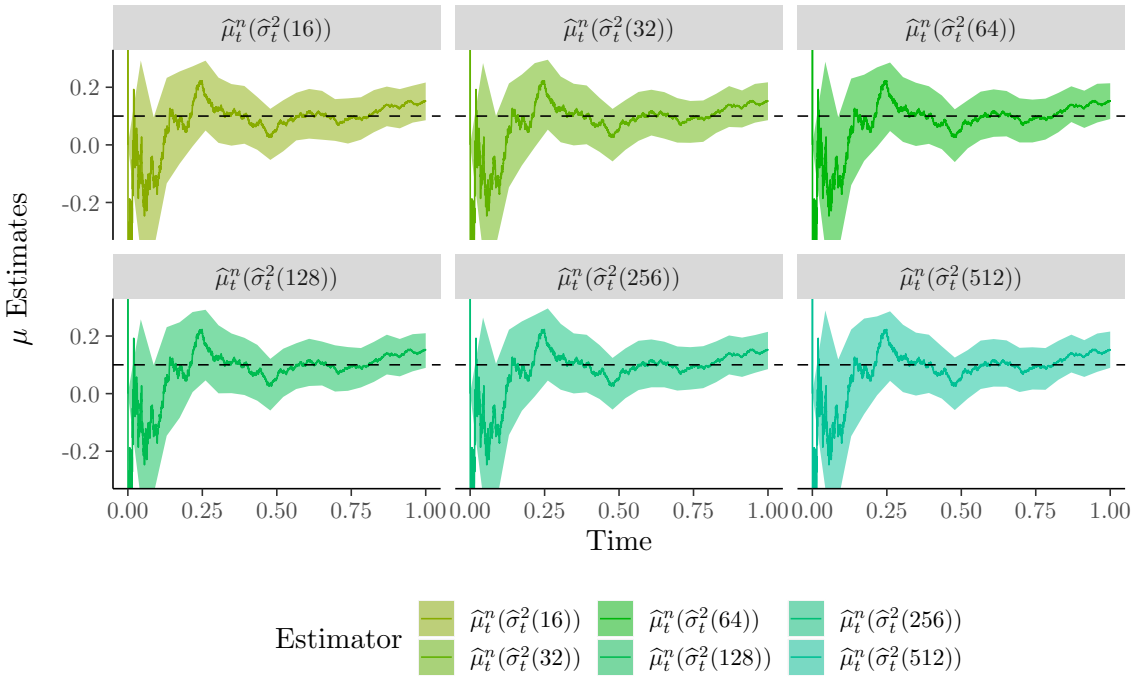

Figure 4.5: The estimate of μ and the corresponding σ used in the estimation. The shaded area corresponds to $[0.1, 0.9]$ confidence intervals based on bootstrapping.

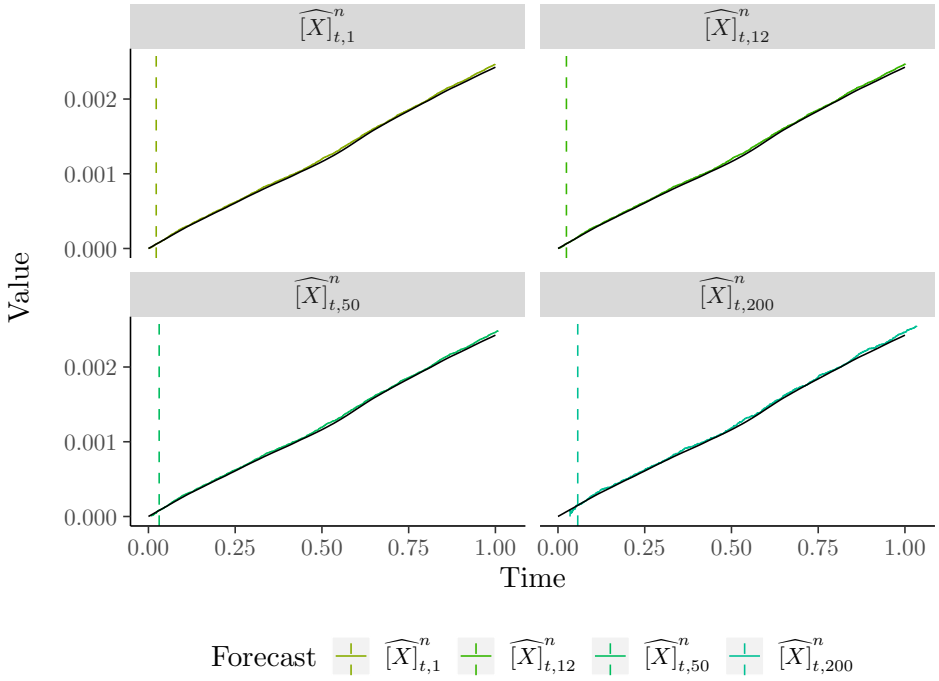

Figure 4.6: Forecasting the Quadratic Variation using the estimator $\widehat{\sigma}_t(256)$.

steps ahead. The dashed line in the figure represents the number of steps ahead with the addition of 256, the number of points reserved to estimate $\widehat{\sigma}_t^2(256)$. Hence, one may consider the dashed line as an indicator for when the actual forecasting starts.

At this stage we would like to remind the reader that the volatility process σ^2, and volatility of the volatility ξ, are quite low, so while it may seem from Figure 4.6 that the estimator is working extremely well we guess that increasing the volatility may lead to larger variations.

We now transform our data to the $[0, 1]$ scale. We compare three scalings; one with the estimated μ and quadratic variation. One with the true μ and estimated quadratic variation, and finally a "perfect" transformation using the true μ and quadratic variation, or integrated volatility. We choose $k_n = 256$ for the estimation of σ^2 and subsequently μ. Using the estimates we approximate the "percentile process" U by

$$\widehat{U}_t = \Phi\left(\frac{X_t - X_0 - \widehat{\mu}_t^n t + \frac{1}{2}\widehat{[X]}_t^n}{\sqrt{\widehat{[X]}_t^n}}\right)$$

It is quite noticeable that the adaptive estimation of μ is causing specification; the percentile process denoted "Estimated" in Figure 4.7 corresponds to the estimator in (4.5), whereas "Realised Variance" uses the true drift but replaces the quadratic variation with the estimated realised variance. The "True" percentile process corresponds to the integrated volatility and true drift μ. For the estimated one, we see that the process becomes, more or less, constant and equal to 0.5. To elaborate, consider the estimator of μ

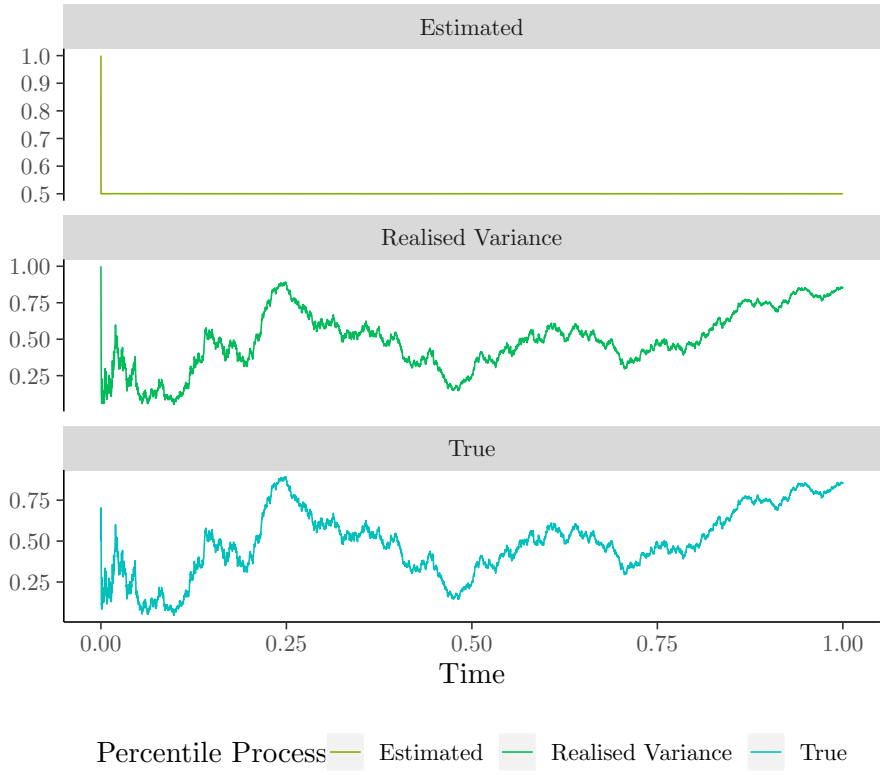

Figure 4.7: Percentile Processes.

$$\widehat{\mu}_t^n = \frac{1}{\Delta_n[t/\Delta_n]} \left(X_{t_{[t/\Delta_n]}} - X_0 + \sum_{i=1}^{[t/\Delta_n]} \frac{1}{2} \Delta_n (\widehat{\sigma}^n)^2_{t_{i-1}} \right).$$

and note that

$$\sum_{i=1}^{[t/\Delta_n]} \frac{1}{2} \Delta_n (\widehat{\sigma}^n)^2_{t_{i-1}} \approx \frac{1}{2} \widehat{[X]}_t^n$$

Now, this implies

$$\mu_t^n \cdot t \approx X_t - X_0 + \frac{1}{2} \widehat{[X]}_t^n,$$

since $t \cdot 1/(\Delta_n[t/\Delta_n]) \approx 1$ and so

$$\begin{aligned}
\widehat{U}_t &= \Phi \left(\frac{X_t - X_0 - \widehat{\mu}_t^n t + \frac{1}{2} \widehat{[X]}_t^n}{\sqrt{\widehat{[X]}_t^n}} \right) \\
&\approx \Phi \left(\frac{X_t - X_0 - X_t + X_0 - \frac{1}{2} \widehat{[X]}_t^n + \frac{1}{2} \widehat{[X]}_t^n}{\sqrt{\widehat{[X]}_t^n}} \right) \\
&= \Phi \left(\frac{0}{\sqrt{\widehat{[X]}_t^n}} \right) = 0.5.
\end{aligned}$$

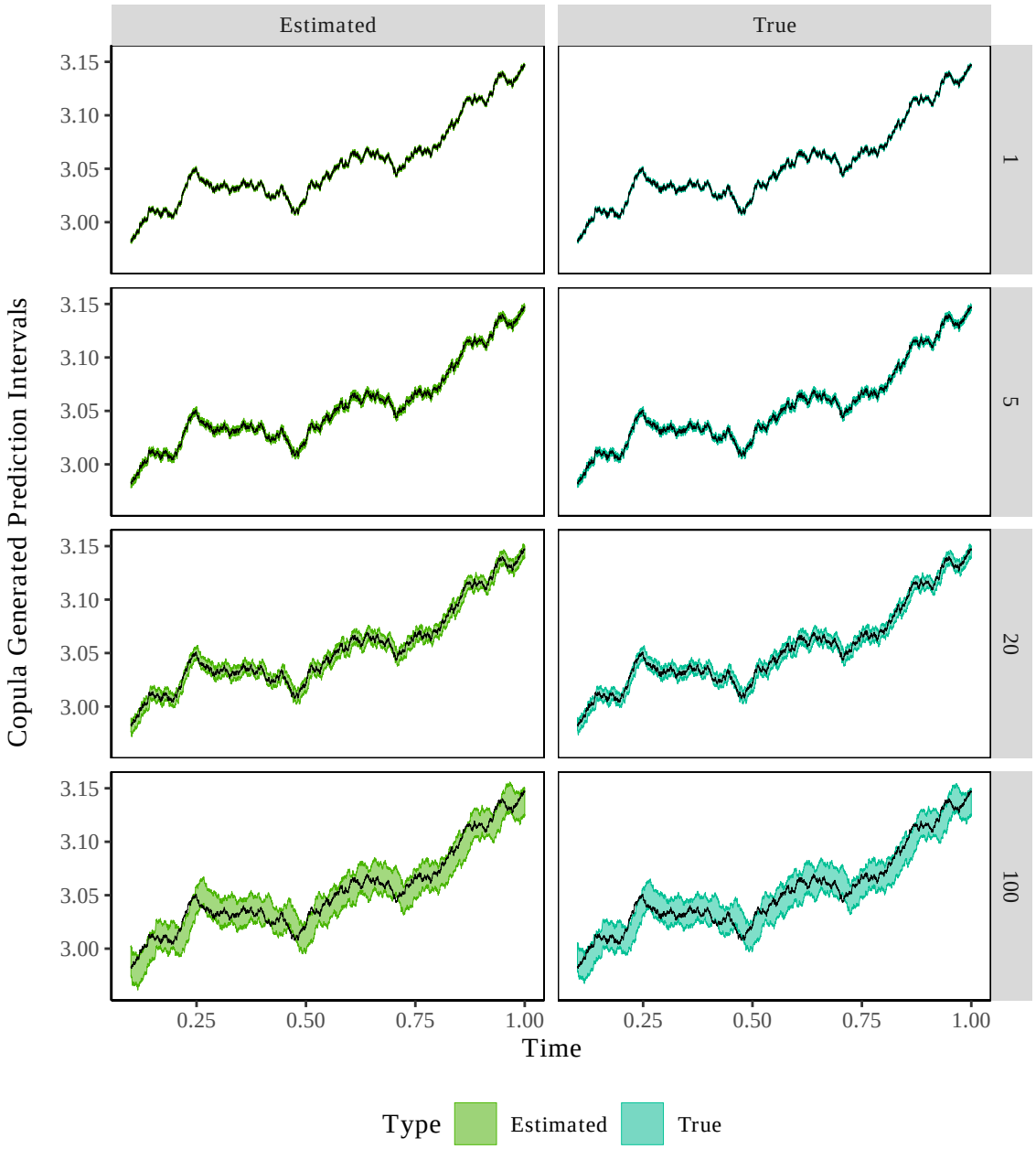

Figure 4.8: Forecasts for the 1, 5, 20, and 100 steps ahead.

We will later address how one can use this knowledge to try an prevent the issue, but it also verifies why the estimated percentile process Figure 4.7 is constantly 0.5.

Applying the forecasting routing specified by equations (4.6) and (4.7) with $\alpha = 0.975$ and $\alpha = 0.025$, corresponding to the 95% centered interval forecast, we obtain the obtain the upper and lower paths of our forecast, depicted in Figure 4.8.

Visual inspection shows practically no difference. We now proceed to conduct the conditional coverage test. We report the direct result of the conditional coverage test for 1 step ahead forecasts and report the total number of times we could not reject the null over the grouped

observations for the remaining step sizes.

Estimated Model			
Test	Statistic	p-value	Reject Null
Unconditional Coverage	0.0627	0.802	No
Independence	0.0564	0.812	No
Conditional Coverage	0.119	0.942	No
Model with true parameters.			
Test	Statistic	p-value	Reject Null
Unconditional Coverage	0.0964	0.756	No
Independence	0.283	0.595	No
Conditional Coverage	0.380	0.827	No

Table 4.2: Conditional Coverage Test for 1-step ahead interval forecasts.

From Table 4.2 we see that we cannot reject that the forecasts provide conditional coverage, which is what we would expect. Note that based on the value of the statistics the estimated model seems to outperform the true parameters. However, one should be cautious of this conclusion, since it may be that it is not providing as tight bounds as the model under the true parameters. Furthermore, in both cases we reject none of the hypotheses.

For the multi step models we report the summary statistics outlined in Section 4.2; we partition the path into groups modulo the number of steps and perform the conditional coverage test in each group. In order to interpret the results, we establish some notation.

$$
\begin{aligned}
n_{\mathrm{p}} &= \text{Number of times null hypothesis was not rejected.} \\
n_{\mathrm{c}} &= \text{Number of times test was computable.} \\
r_{\mathrm{p}} &= \frac{n_{\mathrm{p}}}{n_{\mathrm{c}}}. \\
n_{\mathrm{total}} &= \text{Steps} \cdot \text{Simulations}. \\
r_{\mathrm{c}} &= \frac{n_{\mathrm{c}}}{n_{\mathrm{total}}}
\end{aligned}
$$

We would like to elaborate why the test is not always computable, hence the need for n_{c} and r_{c}. Given a simulation of length 5760 we apply the grouping procedure described in Section 4.2. When we apply the grouping procedure with, say, step-size 100, we are essentially getting 100 groups of roughly 57 observations. In this case it becomes quite likely that for a given simulation the interval forecasts of one, or more, groups are never breached. Since the forecast is not breached, the maximum likelihood estimator of unconditional coverage for the group becomes $\widehat{\pi} = 1$. Now, in the likelihood ratio test the entity $\log(1 - \widehat{\pi})$ is to be computed. Since $\widehat{\pi} = 1$ we have $\log(1 - \widehat{\pi})$ is undefined, hence we cannot compute it in this case. The same case applies when $\widehat{\pi} = 0$ although inspection of the results revealed this to not be a problem. Hence, r_{p} is the ratio of passed tests out of all the computable tests and r_{c} is the ratio between computable tests and total tests.

In this example we are only considering a single simulation so n_{total} is simply the number of steps, but we establish the notation for later use.

Estimated Model				
Steps		Unconditional Coverage	Independent Coverage	Conditional Coverage
5	n_{p}	5	5	5
$(r_{\mathrm{c}} = 1)$	r_{p}	1	1	1
20	n_{p}	7	8	7
$(r_{\mathrm{c}} = 0.4)$	r_{p}	0.875	1	0.875
100	n_{p}	10	52	24
$(r_{\mathrm{c}} = 0.52)$	r_{p}	0.192	1	0.462
Model with true parameters.				
5	n_{p}	5	5	5
$(r_{\mathrm{c}} = 1)$	r_{p}	1	1	1
20	n_{p}	6	8	8
$(r_{\mathrm{c}} = 0.4)$	r_{p}	0.75	1	1
100	n_{p}	13	60	32
$(r_{\mathrm{c}} = 0.6)$	r_{p}	0.217	1	0.533

Table 4.3: Summary of the grouped Conditional Coverage tests.

From Table 4.3 we see that for 5 steps ahead the null-hypothesis cannot be rejected in neither the estimated model or the model under the true parameters, as indicated by both $n_{\mathrm{p}} = 5$ and r_{p} being 1. Furthermore, all tests were computable since $r_{\mathrm{c}} = 1$. Now, for 20 and a 100 steps ahead we see that r_{c} drops quite dramatically, to the extent where it is invalidating the test. Based on the few tests that were computable for 20 steps ahead, we see that quite often we cannot reject the null hypothesis. However, in the case of 100 steps ahead, the null is rejected frequently in both the estimated and the true model.

The indication is that we may rely on the presented methodology for short-term forecasting of the distribution, which could present a short-term forecasting for the value at risk. Furthermore, by inspection of Figure 4.8, we see that the prediction intervals are fairly tight around the process. Of course, there is much room for improvement and we address how one may proceed to augment the procedure in our discussion later. □

Having presented the methodology we now conduct a large scale study; we cannot rely on this single instance of the test to show us the behaviour in general. We showcase one simulation study under the same parameters used in the previous example, i.e. Table 4.1. We simulate 1000 paths, and we obtain the following results for the conditional coverage test. We report the total number of times we could not reject the null-hypothesis in each of the components of the conditional coverage test.

For the estimator of σ^2 we choose $k_n = 256$, based on visual inspection of Figure 4.4. Note that one could implement a procedure to adapt k_n. We will later see that the choice of k_n may have a significant impact on the forecasts. However, in order to limit the scope of the thesis we will simply use $k_n = 256$.

From Table 4.4 we see that all steps are comparable based on the ratio r_{p}. Note that the closer r_{p} is to 1, the more we could not reject the null hypothesis in the computable instances. In all steps r_{p} remains *comparable*. However, there is a caveat. We see the the ratio between the

number of computable test and total number of tests, r_c, drops dramatically. For 1 and 5 steps ahead we are fairly confident in our result, but at 20 steps ahead only roughly 50% of the tests were computable which lowers the confidence in the result, even if the result indicates that the forecasts were good. For 100 steps ahead only roughly 20% of the tests were computable for the estimated and 12% for the model with true mu and quadratic variation which is invalidating.

Estimated Model				
Steps		Unconditional Coverage	Independent Coverage	Conditional Coverage
1	n_p	966	931	944
($r_c = 1$)	r_p	0.966	0.931	0.944
5	n_p	4538	4617	4610
($r_c = 0.956$)	r_p	0.949	0.965	0.964
20	n_p	10626	10779	10590
($r_c = 0.561$)	r_p	0.947	0.961	0.944
100	n_p	17187	18547	17990
($r_c = 0.201$)	r_p	0.856	0.924	0.896
Model with true parameters				
1	n_p	947	950	9474
($r_c = 1$)	r_p	0.947	0.95	0.947
5	n_p	4477	4584	4557
($r_c = 0.939$)	r_p	0.954	0.976	0.971
20	n_p	9501	9566	9460
($r_c = 0.494$)	r_p	0.961	0.967	0.957
100	n_p	10987	10839	11258
($r_c = 0.122$)	r_p	0.899	0.886	0.921

Table 4.4: Summary of the Conditional Coverage test.

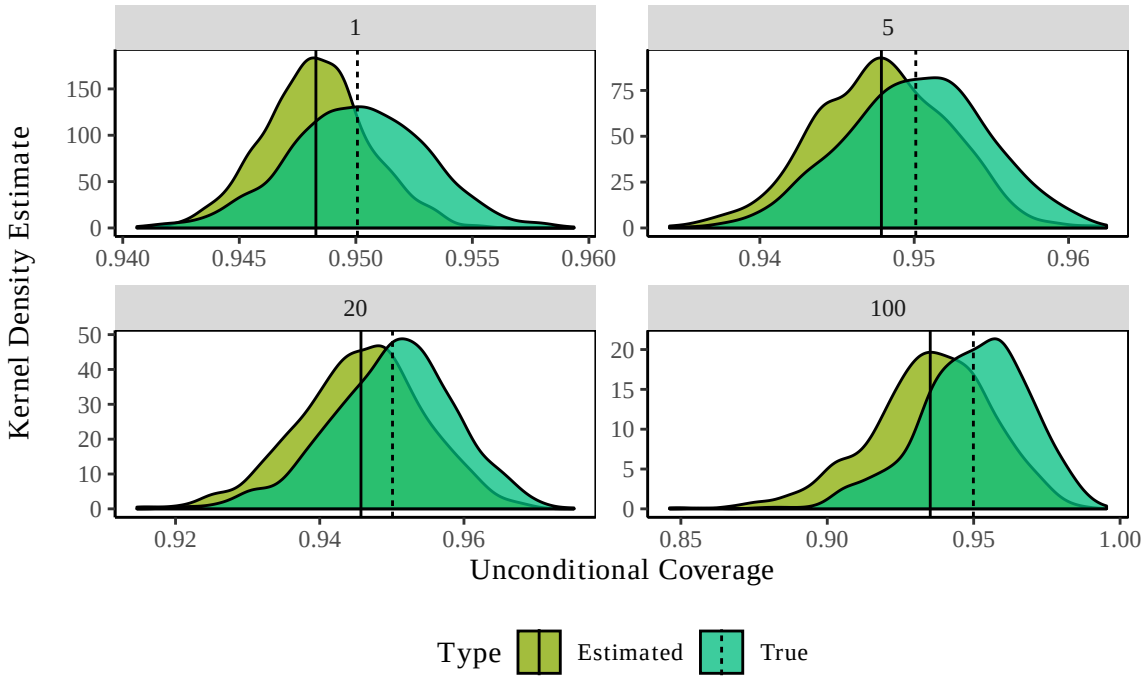

Figure 4.9: Kernel density estimates of the unconditional coverage rate. The dashed line represents the mean of the model with true parameters, the solid line the mean from the estimated model.

Based purely on value of r_{p} in Table 4.4 we see that the simply heuristic forecasting procedure produces *fairly* accurate results. However, the value of r_{p} does not tell us to which extend our forecasts were accurate. In order to see the difference between the forecasts generated by the procedure versus the true parameters we calculate the unconditional coverage for each group, i.e. $\widehat{\pi}$ of equation (4.8), by simply dividing the number of times the process was within the forecasted interval by the total number of forecasts. We then make a density estimate for the estimated and true model respectively and plot them together.

From Figure 4.9 we see that the distribution of unconditional coverage between the estimated model and the model with the true parameters are quite comparable. We see that the mean of the model with true parameters has an average unconditional coverage of almost exactly 95%, as indicated by the dashed line, where the estimated model is slightly lower in all instances. Recall that we were trying to produce an 95% centered forecast so these results are quite positive. It is also worth noting that the estimated model is slightly more leptokurtic for the step-sizes 1 and 5, an hence has more tail density. In the latter instances, 20 and 100 steps, it seems to be shifted slightly to the left.

We now conduct a similar experiment. However, we now use the following parameters:

Again, we simulate 1000 paths over an equidistant mesh $[0, 1]$ with 5670 observations. In this scenario the volatility is more persistent as seen from Table 4.5; κ has been reduced. Furthermore, the volatility process σ has become more volatile and μ has been reduced.

Parameter	κ	θ	ξ	μ	ρ	V_0	σ_0^2
Value	0.5	0.36	0.25	0.05	0	20	0.25

Table 4.5: Simulation Parameters for the Second Simulation

From Table 4.6 we see quite a different picture; the short term forecasts are noticeably worse for the estimated model indicated by r_{p} for all models being quite low. This indicates, that the short term forecasts did not provide an accurate cover of the centered 95% interval. Furthermore, the accuracy of the estimated model seems to increase as we forecast further ahead.

At first glance, this should seem counter intuitive. Clearly, when κ is reduced, the variance process becomes more persistent, and since our forecast of the quadratic variation relies on extrapolation of the current variance we would expect it to perform better. However, ξ is increased and so the variance of the process is also increased. Now, we have again chosen to estimate $(\widehat{\sigma}_t^2)^n$ using $k_n = 256$ points. One possible explanation to the behaviour is, that k_n is too large, and so we are smoothing the variance process excessively in this case. This would also explain why the long term forecasts are better; if the estimator of σ^2 is smoothed too much, we expect, roughly speaking, that $(\widehat{\sigma}_t^2)^n \to \theta$, i.e. we model the long term mean of the volatility rather than the spot volatility. Hence, on average our long term forecasts become more reasonable. In a sense, this motivates choosing k_n fairly small for short term forecasting and k_n large for long term forecasting.

Estimated Model				
Steps		Unconditional Coverage	Independent Coverage	Conditional Coverage
1	n_{p}	341	467	303
($r_{\mathrm{c}} = 1$)	r_{p}	0.341	0.467	0.303
5	n_{p}	3631	4044	3460
($r_{\mathrm{c}} = 0.989$)	r_{p}	0.734	0.818	0.700
20	n_{p}	12847	13463	12363
($r_{\mathrm{c}} = 0.743$)	r_{p}	0.865	0.906	0.832
100	n_{p}	23314	27692	25074
($r_{\mathrm{c}} = 0.302$)	r_{p}	0.771	0.916	0.829
Model with true parameters				
1	n_{p}	911	948	926
($r_{\mathrm{c}} = 0.997$)	r_{p}	0.914	0.951	0.929
5	n_{p}	4028	4160	4073
($r_{\mathrm{c}} = 0.854$)	r_{p}	0.944	0.975	0.954
20	n_{p}	7938	8068	7991
($r_{\mathrm{c}} = 0.423$)	r_{p}	0.939	0.955	0.946
100	n_{p}	9352	9924	9750
($r_{\mathrm{c}} = 0.109$)	r_{p}	0.858	0.910	0.894

Table 4.6: Summary of the Conditional Coverage test.

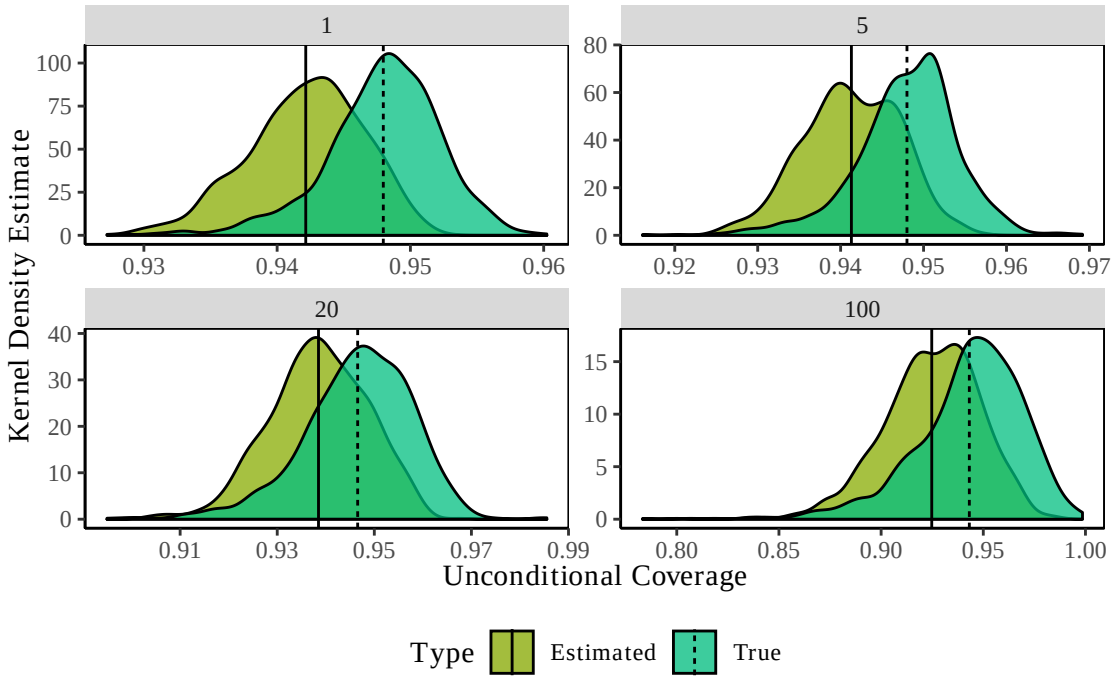

Figure 4.10: Kernel density estimates of the unconditional coverage rate. The dashed line represents the mean of the model with true parameters, the solid line the mean from the estimated model.

Another thing to notice from Table 4.6 is that even in the case of 1 step ahead there were tests that failed which implies that the forecast was not breached in at any time during these simulations. However, the procedure is subject to some numerical issues which we later address, and so it is not unlikely that these three paths were subject to numerical issues. While it only happens in 3 simulated paths we would have liked to investigate why it happens given additional time.

Again, as r_c decays towards 0 as the step-size increases the test is invalidated at large steps ahead. Therefore, in order to investigate the unconditional coverage we again plot the density estimates.

Figure 4.10 shows us that the unconditional coverage was overall too low for the estimated model. However, from the plot we can see that the forecasts were not completely unreasonable, although they did not pass the conditional coverage test. We see that the average unconditional coverage is roughly 94% for the estimated model in most scenarios. We also see that there is quite a large difference between the estimated densities of the estimated model and model under the true parameters, especially for the 1 step ahead forecasts.

5

Discussion and Conclusion

5.1 Discussion of the Results

We show that it is indeed possible to estimate the conditional copula of a continuous local martingale, under the additional assumption that its quadratic variation is independent of the Brownian motion it time-changes. Furthermore, we show a limit theorem, allowing us to quantify to which degree of certainty the result can be trusted although it is in a very restricted setting.

The consistency theorem states that as long as the quadratic variation is independent of the Brownian motion it time changes, then the consistency is uniform over the pairs (s,t) in the space $\{(s,t) \mid 0 \leq s \leq t \leq T\}$. Our numerical investigation revealed that it may be possible to enhance the result to uniform consistency over $(s,t,u,v) \in \{(s,t) \mid 0 \leq s \leq t \leq T\} \times [0,1]^2$.

In regards to the limit theorem we proceed with slightly more scepticism. We would expect the result to remain valid. In our case we made some restriction in regard to the boundedness of the process σ. Jacod and Protter show a so-called *localisation procedure* which may help us alleviate this restriction to, say, a requirement that σ is cádlág. The strategy of the current proof cannot be extended to stochastic volatility as the result from Dudley, [9], may not be extended. The numerical analysis and prior arguments reveal that when (s,t,u,v) are chosen suitably we *may* obtain a standard Gaussian limit. However, due to complications on the boundary of the set

$$\partial\{(s,t) \mid 0 \leq s \leq t \leq T\} \times \partial[0,1]^2,$$

the applications of the limit theorem may be very limited. Perhaps, the most important result in this direction is that the rate of convergence seems to be preserved, which seems to be the case even on the stronger uniformity over (s,t,u,v), but one should be sceptical of even this result if the limit here is non-Gaussian, or at least not determinable, since there may very-well be a non-trivial transformation depending on n which yields the correct limit. It also immediately implies we may not expect a "functional" limit theorem for the copula in the same sense that our consistency theorem was essentially functional convergence.

From a practical perspective we find it unlikely to cause issues, and in this case the boundary is also uninteresting. In a sense, it is only natural that the boundary induces pathological behaviour

since it, typically, reduces the problem to a univariate distribution.

It is clear that the estimation relies on being able to consistently estimate the quadratic variation. We have for now omitted jumps. However, quadratic variation can be estimated even when jumps are present, see e.g. [13]. Furthermore, the quadratic variation can be "disentangled"; as jumps are present they will contribute to the quadratic variation, but there are estimators which allow us to estimate only the continuous part, in this case we may be able to estimate the conditional copula for the continuous part of the model, see e.g. [3]. In general, a semimartingale admits the decomposition

$$X_t = X_0 + A_t + X_t^c + M_t, \quad t \geq 0,$$

where $A_0 = X_0^c = M_0 = 0$, A is a process of locally bounded variation, X^c is a continuous local martingale and M is a "purely discontinuous" martingale, see [13], which roughly relates to all movements of M being jumps. In this case, one may still estimate the copula of X_t^c under suitable restrictions.

For the result obtained there are some clear caveats. Noticeably, there are some quite restrictive assumptions which must be made. Classical models, such as the Heston model, have already alleviated themselves from being independent of the "driving" Brownian motion, such that they may model the leverage effect, which the framework we show here is not capable of.

In regards to our simulation study we find that the heuristic forecasting procedure works *quite* well. There are caveats to the procedure, some which may have an immediate solution and others which require slightly more care.

The procedure is subject to some numerical instability, as a lot of numerical integral and inversions are required to implement the procedure. Perhaps one may increase the stability by implementing certain asymptotic properties on the boundary directly.

We argue that the procedure could serve as a benchmark for a lot of forecasting procedures. Now, in our specific case we could have bypassed the copula completely. Hence making it redundant. The use of the copula may be more applicable in a so-called *marginal matching procedure* which we discuss in the upcoming section.

Through our simulation study we find, that the forecasting procedure works *somewhat* when tested on the Heston model. One test was overly positive, while the other one showed that the forecasts may not be as accurate. We would like to remark that given additional time, it would be interesting to forecast more than just the 95% centered forecast in an attempt to evaluate it over various forecasting ranges.

5.2 Future Research

While the applicability of the result at this stage is questionable there are some questions which arose while working on the results presented here.

We first address the marginal matching procedure. Suppose we are given data from some process X. Under assumptions that X is a semimartingale we may estimate $[X]$. Note that X by definition decomposes to $X_t = X_0 + A_t + M_t$. If we furthermore assume it is continuous and M

is represented by a time-changed Brownian motion independent of the quadratic variation, then we can consistently estimate the copula. Now, at time t we may then estimate

$$\widehat{U}_t = \Phi\left(\frac{X_t - X_0 - \widehat{A}_t}{\sqrt{\widehat{[X]}_t^n}}\right),$$

where $\widehat{A}$ is an estimator of the process of locally bounded variation. Then, when we make our forecast, say, $\widehat{U}_{t,h}$ for some $h = k\Delta_n$, nothing dictates that we invert using the Gaussian distribution. Hence, we may match the marginals to fit the data in a more flexible way.

In order elaborate, consider the following situation: We are tasked with providing a 95% centered interval forecast. We find that our upper bounds match empirically, that is we obtain a coverage rate for the upper bound which is approximately 97.5% and based on some testing procedure we cannot reject the upper bound. However, the lower bound seems to be slightly off; we find more than 2.5% are below the lower bound, say 5%. That is, our lower tail has more density than the model suggest. We could then choose marginal distributions which allow for this type of behaviour.

Another situation is in the framework of our forecasting procedure, we saw that our μ estimator made the "percentile process" constant. We could start by assuming $\mu = 0$, provide forecasts and based on evaluation over several quantiles attempt to shift μ such that all the quantiles match based on a routine similar to, e.g., bisection. It is unclear whether the outcome would be the same, but certainly interesting.

The marginal matching procedure may be compared to calibration of market models using option prices, but here we calibrate to certain quantile we wish to match.

We would also like to paint a more grand picture of what we are trying to accomplish. Our initial motivation for this thesis was in part to see if stochastic processes could be considered in a more graphical manner by using copulas. We elaborate with an example: suppose for simplicity that M is a continuous local martingale with deterministic quadratic variation. Then, the dependence of (M_s, M_t) can decomposed to the graph in Figure 5.1, where $H_{s,t}$ is the joint distribution of (M_s, M_t), F_s is the marginal at time s, F_t the marginal at time t and $C^M_{s,t}$ the copula between (M_s, M_t).

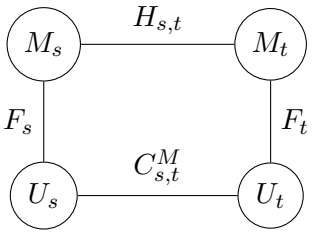

Figure 5.1: Dependence Graph of a Martingale

Our hope is to continue research in this direction and to extend the graph to accommodate more information such as volatility, jumps, and drift. In essence, we have so far established the estimation of the edge between (U_s, U_t), perhaps only partially if one considers stochastic volatility.

The applications would be that stochastic processes may be more representable in terms of functions; from a mathematical point of view one could argue that stochastic processes are to be considered as the measure they induce and that there is little to no use of other representations because they will only ever be restrictions of the more general case. From a practical point of view it is hard, if not impossible, to implement such measures. Functions over Euclidean domains may on the other hand be implemented through different numerical procedures.

Our hope was that via copulas the laws of certain quantities could be assumed or predetermined to allow for a more functional representation of stochastic processes with the addition of the flexibility which is associated with copulas. Time will tell whether this research will be fruitful and we would like to remark that after the work performed in this thesis we are perhaps more sceptical of positive results than prior to writing the thesis. However, the work in this thesis is an attempt at making the, relatively, new copula method compatible with the classical and rigorous theory. Perhaps a better approach is to start anew and see if one can eventually tie the ends together.

Of course, additional work may be conducted on the current model in an attempt to recover a more general limit theorem. However, in light of our findings we find it unlikely to be of much practical use.

In the thesis we also worked the Heston model. In particular, the Heston model is able to capture so-called *volatility clustering* and *the leverage effect*. The volatility clustering can be described by the mean reversion; if the volatility is high it will decay to its long run mean, which is controlled by the parameter κ. Similarly, if $\rho < 0$, then it also models the leverage effect, since higher volatility is then correlated with lower returns or price of the asset. In our very restricted setting we were not able to model the leverage effect, since independence with the driving Brownian motion is crucial and so $\rho = 0$ is the only theoretically valid parameter for the copula model.

However, $\rho < 0$ still implicitly assumes this dependence is linear, which is quite restrictive. Now, the problem can be solved using copulas. For instance, rather than two correlated Brownian motions we can construct two copula dependent Brownian motions.

Let $((U_i, V_i))_{i=1}^n \sim C$ be independent samples for some copula and take $X_i = \Phi^{-1}(U_i)$ and $Y_i = \Phi^{-1}(V_i)$. Now, marginally X and Y are unconditionally Gaussian, but they are not jointly Gaussian unless C is a Gaussian copula. By Donsker's Theorem, see Theorem A.5, we have the processes

$$t \mapsto \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} X_i \Rightarrow W_t^X \tag{5.1}$$

$$t \mapsto \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} Y_i \Rightarrow W_t^Y, \tag{5.2}$$

where the convergence is so-called weak convergence of probability measures and W^X and W^Y are Brownian motions on $\mathbb{D}([0,1];\mathbb{R})$. Hence, marginally the process is a Brownian motion, but unless C is the Gaussian copula, then (W^X, W^Y) is not a bi-variate Brownian motion. This may allow for more flexible modelling of the leverage effect, and one can indeed define a natural extension of the Heston model using the copula dependent Brownian motions in (5.1) and (5.2).

Note that since the marginal processes are Brownian motions, then nothing restricts us from

integrating with respect to either, hence copula dependent systems of stochastic differential equations are well defined.

The aforementioned procedures could contrast the current trends of inducing additional stochasticity to our models. For instance, in practice one might observe that two assets at times appear to be positively correlated, at other negatively correlated, and other times completely independent. From a theoretical point of view, it may be very natural to then consider the correlation, ρ, between two assets to be driven by some stochastic process, such that the dependence between two assets is random, and while it seems very intuitive to do so, with the current state of parameter estimation of such models become increasingly unreliable. One can draw parallels to a system of linear equations, where the system is only solvable if the system is not singular. In a sense, if we only observe, say, 2 processes, but we are trying to filter a total of 5 processes, then it is quite natural that the procedures are less robust. Much like solving a system of equations there might not be a single solution or any satisfactory solution at all. Here we see copula models possibly providing natural intermediate; rather than depending on linear correlation, we may augment our copula or marginals to accommodate such behaviour.

We would like to remark that the theoretical justifications for applying the aforementioned procedures are scarce. Perhaps the best justification of applying the aforementioned procedures is to claim that it may yield interpretable results and acknowledging that one is, to some extent, abandoning a rigorous statistical framework in favor of a somewhat heuristic framework.

We argue that this should be done with caution. Very flexible models can be derived from copulas, in particular any type of behaviour should be able to be modeled. However, the copula should be chosen with some clear goal in mind, and it is up to the individual to incorporate them responsibly. For instance, one could imagine that one re-calibrates a model every m'th period testing every available combination within reason of typical copulas and marginals to see which fit is better based on some criteria, and without regard for what is being modeled this can lead to heavy misuse.

5.3 Conclusion

We show that one may consistently estimate the copula of a time-changed Brownian motion using realised variance under suitable restrictions. We find that the estimator is uniformly consistent over the temporal domain and argue that the uniform consistency can be extended to the spatial domain as well.

We show a limit theorem in a very restricted case and via numerical analysis we argue that the result may be extended even to stochastic volatility. However, the strategy applied in the proof provided here will likely be of little use.

We argue that the rate of convergence of our estimator is likely to be proportional to $\sqrt{n}$. We claim this rate of convergence based partially on our limit theorem and partially on our numerical example. Further testing must be conducted in order to verify whether or not this result is indeed valid.

We propose a forecasting procedure using the derived copula. While the current state of the

procedure does not require the use of copulas we propose a way to augment to procedure to provide a more flexible modelling approach. We remark that one should be cautious in doing so and that applying such procedure may require one to abandon a rigorous statistical framework.

The forecasts provides reasonable coverage based on the conditional coverage test from Christoffersen, [8]. Enough so that we feel it warrants further research into developing the procedure.

We conclude that copulas can be made compatible with temporal modelling, but at this stage there is no clear advantage in doing so. Further research may uncover results which provide a clear benefit in applying copulas, but at the current state we are not convinced that any theoretically valid reason exists. However, we are optimistic and hope that the results provided here may yield new insights into how to proceed in using copulas for temporal and spatial modelling and we will conduct further research into this topic.

Appendix A

Appendix

A.1 Metric Spaces and Topology

This section is based on [16] and [15].

A fundamental part of mathematical analysis is the study of limits. Suppose we have a non-empty set S and a sequence $(s_n)_{n\in\mathbb{N}} \subseteq S$ and we are interested in whether or not the sequence approaches a limit s as n goes to infinity and equally important does the limit belong to S or does the limit belong to some other set Σ for which $S \subset \Sigma$.

Typically in doing so, we will require a metric. Metrics are notions of distance that must fulfil certain axioms, which we will define later. The goal of this section is to define the properties that make up a Polish space.

We are often interested in maps that preserve analytical structures as they typically enlighten the underlying space. Continuous functions are maps that preserve *open* sets. However, before we can transform open sets we will need to define what it means to be an open set.

Definition A.1 (Topology)
Let X be a set and τ a family of subsets of X satisfying

1 $X, \emptyset \in \tau$.

2 $A, B \in \tau$ implies $A \cup B \in \tau$.

3 $\{A_n\}_{n=1}^{N} \subset \tau$ implies $\bigcap_{n=1}^{N} A_n \in \tau$ for any $N \in \mathbb{N}$.

Then τ is a topology, and we say the pair (X, τ) is a topological space. Furthermore, the elements of the topology are called open with respect to the topology τ.

A topology is a collection of subsets of the original space which allows us to specify what we mean by openness. We require some consistency axioms namely closure under intersections, countable and finite unions, and inclusion of the original space and the empty set.

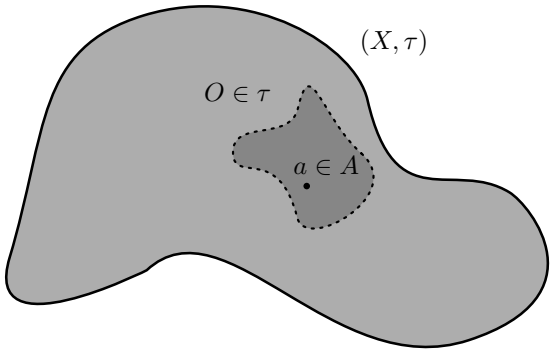

Figure A.1: Visualisation of a Dense Set. For every $O \in \tau$ such that $O \neq \emptyset$ we must be able to find at least one $a \in A \cap O$.

Definition A.2 (Dense and Separable)
Let (X, τ) be a topological space and let $A \subset X$. We say that A is dense in X if

$$A \cap O \neq \emptyset, \quad \forall O \in \tau \backslash \{\emptyset\}.$$

Furthermore, (X, τ) is separable if there exists $A \subseteq X$ where A is countable and dense.

When a set A is dense in in (X, τ) it satisfies that for every O in our topology we can find an a that both belongs to A and O. We visualise a single instance of the inclusion $a \in O \cap A$ in Figure A.1. When A is countable and dense then we must be able to enumerate every point in A and have it be dense in X. Intuitively we may think of A as a "mesh" that covers X with respect to the topology.

A simple example of a separable space is $\mathbb{R}$; we have $\mathbb{Q}$ is dense in $\mathbb{R}$ and $\mathbb{Q}$ is countable. In this simple example, we know that we can *approximate* elements of $\mathbb{R}$ with elements of $\mathbb{Q}$, and the definition of separability can be interpreted as being able to approximate a topological space in a countable way.

Now, a topology is in some sense the minimal analytic structure a space can have; it allows us to define continuity. However, oftentimes we have space endowed with a slightly stronger analytic structure. Note that in Definition A.2 we have no way to determine whether or not the element of $O \in \tau$ and $a \in A$ are close, merely that they belong to the same open set. Now, a metric will allow us to determine the distance.

Definition A.3 (Metric Space)
Let X be a non-empty set. A function $d : X \times X \to [0, \infty)$ is said to be a metric, if for all $x, y, z \in X$ it satisfies

1 $d(x, y) = 0$ if, and only if, $x = y$.

2 $d(x, z) \leq d(x, y) + d(y, z)$.

3 $d(x, y) = d(y, x)$.

Let $\varepsilon > 0$. The *open ball* with radius ε is defined as

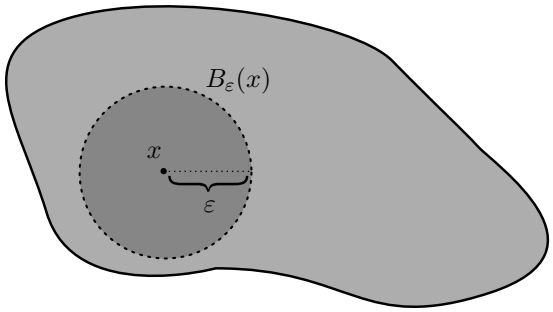

Figure A.2: Visualisation of Open Sets Induced by a Metric

$$B_\varepsilon(x) = \{y \in X \mid d(y,x) < \varepsilon\}, \quad x \in X. \tag{A.1}$$

Similarly, we may define the *closed ball* with radius ε as

$$\overline{B_\varepsilon(x)} = \{y \in X \mid d(y,x) \leq \varepsilon\}, \quad x \in X. \tag{A.2}$$

Every metric space is a topological space. We will not prove this fact as it is well known; open balls in a metric space provide a basis for the topology. We visualise this in Figure A.2.

Definition A.4 (Cauchy Sequence)
Let (X,d) be a metric space and let $(x_n)_{n\in\mathbb{N}}$ be a sequence. $(x_n)_{n\in\mathbb{N}}$ is said to be a Cauchy, or fundamental, sequence if for every $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that

$$m, n \geq N \implies d(x_n, x_m) < \varepsilon.$$

The metric space (X,d) is said to be complete if every Cauchy sequence converges in X.

Remark: Cauchy sequences are a fundamental part of mathematical analysis. In general, one may think of Cauchy sequences as sequences that always have a "limit" but that the limit does not necessarily belong to the original metric space (X,d). Indeed, if one encounters an incomplete metric space then it may be completed by constructing equivalence classes of Cauchy sequences in (X,d) and then identifying (X,d) in the resulting space using constant sequences which are certainly Cauchy. This approach can be used to construct $\mathbb{R}$ from $\mathbb{Q}$ or the $L^p(\Omega)$ spaces from the space of continuous functions $C(\Omega)$.

Topological and metric spaces serve as a basis for continuous functions. We now present a well-known proposition as it will play a detrimental role in the majority of theorems.

Proposition A.1. *Let (E, d_E) and $(E', d_{E'})$ be metric spaces and $f : E \to E'$ a continuous function. Let $(x_n)_{n\in\mathbb{N}}$ be a convergent sequence in E with limit x. Then*

$$\lim_{n\to\infty} f(x_n) = f(x).$$

Proof. We have to prove, that for any $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that

$$n \geq N \implies d(f(x), f(x_n))_{E'} < \varepsilon.$$

By the continuity of f we know that

$$\forall \varepsilon > 0 \exists \delta > 0 : d(x,y)_E < \delta \implies d(f(x), f(y))_{E'} < \varepsilon. \tag{A.3}$$

Since $(x_n)_{n\in\mathbb{N}}$ is convergent, we know that

$$\forall\delta > 0 \exists N \in \mathbb{N} : n \geq N \implies d(x, x_n)_E < \delta. \tag{A.4}$$

Using (A.3) we get that for $\varepsilon > 0$ there exists a $\delta = \delta(\varepsilon) > 0$, for the same δ we deduce by (A.4) that there exists $N \in \mathbb{N}$ such that $n \geq N \implies d(x, x_n)_E \leq \delta$. It now follows that

$$n \geq N \implies d(x, x_n)_E < \delta \implies d(f(x), f(x_n))_{E'} < \varepsilon.$$

■

Remark: Proposition A.1 justifies that when f is continuous we may push the limit inside the function, when we know $(x_n)_{n\in\mathbb{N}}$ is convergent.

Definition A.5 (Completely Metrizable Space)
Let (X, τ) be a topological space. (X, τ) is said to be completely metrizable if there exists at least one complete metric d, such that (X, d) is a complete metric space, where d induces τ.

A metric space is typically easier to work with than a topological space, since we may characterize the elements in the topology using the metric. Hence, a completely metrizable space is a sufficiently "nice" topological space, in the sense that we may find at least one complete metric which induces (X, τ).

A.2 The Space of Continuous Functions Over the Postive Reals

In order to construct a metric on $C([0, \infty); \mathbb{R})$ from the supremum metric we will have to bound our metric in a convenient way. Therefore, we will need the following definition.

Definition A.6 (Equivalent Metrics)
Let X be a non-empty space and d, ρ be metrics on X. Then the spaces (X, d) and (X, ρ) are topologically equivalent if, and only if, they generate the same topology. We write $(X, d) \cong (X, \rho)$.

Intuitively given a metric d we can induce a topology. However, suppose we only have balls of radius less than 1. Intuitively, we could build balls with radius $r > 1$ by joining balls of radius less than one. We justify this claim below.

Lemma A.1. *Let (X, d) be a metric space. Define*

$$d_\wedge(x, y) = 1 \wedge d(x, y), \quad x, y \in X.$$

Then $(X, d_\wedge) \cong (X, d)$.

Remark: Recall that for real numbers x and y we have $x \wedge y = \min\{x, y\}$.

Proof. We argue that it is trivial to see that $d_\wedge$ is indeed a metric if d is a metric.

We must show that an open set in (X, d) is open in $(X, d_\wedge)$. Let O be open in (X, d). By definition we know that

$$\forall x \in O \exists r > 0 : B_r(x) \subseteq O.$$

Now, clearly for $r \geq 1$ we have that $B_{r'}(x) \subseteq B_r(x)$ for all $r' \in (0, r]$. It follows that O is open in $(X, d_\wedge)$. Conversely, let O be open in $(X, d_\wedge)$. Note that in $(X, d_\wedge)$ we may characterise open balls by

$$\forall x \in O \exists r \in (0, 1) : B_r(x) \subseteq O,$$

but clearly for $r \in (0, 1)$ we have that the metrics agree, so for the same r we have that O is open in (X, d). Hence, $(X, d) \cong (X, d_\wedge)$. ■

We now have that $C([0, T]; \mathbb{R})$ may be either endowed with the metrics $d_\infty(x, y)$ or $1 \wedge d_\infty(x, y)$. The reason why we want $1 \wedge d_\infty(x, y)$ is that it's bounded by 1, hence we can appropriately weight compact intervals to define a metric on $C([0, \infty); \mathbb{R})$. That is, we take an increasing sequence of compact sets that eventually cover the entire domain and appropriately weigh them.

Lemma A.2. *Let (E, d_E) be a Polish space, then $C([0, \infty); E)$ admits a metric defined by*

$$d_\infty(f, g) = \sum_{n=1}^{\infty} \frac{1}{2^n} \left(1 \wedge \sup_{t \in [0,n]} d(f(t), g(t))_E \right).$$

Proof. First, note that $d_\infty(f, g) \geq 0$ for all $f, g \in C([0, \infty); E)$. Next, we have that

$$d_\infty(f, g) = 0 \iff f = g.$$

Clearly, the implication that $f = g \implies d_\infty(f, g) = 0$ is trivial. We wish to show that $d_\infty(f, g) = 0$ implies $f = g$. Note that if $d_\infty(f, g) = 0$, then each term is the sum must be identically 0, and it follows that $f = g$.

Now, clearly $d_\infty(f, g) = d_\infty(g, f)$ by extension of $1 \wedge \sup_{t \in [0,N]} d_E(f(t), g(t))$. It remains to show the triangle inequality. Consider

$$\begin{aligned}
d_\infty(f, g) &= \sum_{n=1}^{\infty} \frac{1}{2^n} \left(1 \wedge \sup_{t \in [0,N]} d_E(f(t), g(t)) \right) \\
&\leq \sum_{n=1}^{\infty} \frac{1}{2^n} \left(\left(1 \wedge \sup_{t \in [0,N]} d_E(f(t), h(t)) \right) + \left(1 \wedge \sup_{t \in [0,N]} d_E(h(t), g(t)) \right) \right) \\
&= \sum_{n=1}^{\infty} \frac{1}{2^n} \left(1 \wedge \sup_{t \in [0,N]} d_E(f(t), h(t)) \right) + \sum_{n=1}^{\infty} \frac{1}{2^n} \left(1 \wedge \sup_{t \in [0,N]} d_E(h(t), g(t)) \right) \\
&= d_\infty(f, h) + d_\infty(h, g).
\end{aligned}$$

Where the inequality follows from the fact that $1 \wedge \sup_{t \in [0,N]} d(f(t), g(t))_E$ is a metric and we use the associated triangle inequality. ■

Remark: We will use d_∞ for both the finite version and the infinite version stated above and it should be clear from context if we refer to $(C([0, T]; \mathbb{R}), d_\infty)$ or $(C([0, \infty); \mathbb{R}), d_\infty)$.

We denote the topology induced by d_∞ in Lemma A.2 as *the topology of local uniform convergence.* Now we have the following extension of Theorem 2.4.

Theorem A.1. *Let E be a Polish space. Then $(C([0,\infty),E),d_\infty)$ is a Polish metric space.*

We will not prove Theorem A.1. However, the essence is that we can use the local compactness of $[0,\infty)$ to recover Theorem 2.4 for each $[0,N]$, $N\in\mathbb{N}$. Note that in a very similar fashion we can endow $\mathbb{D}([0,\infty);\mathbb{R})$ with the topology of local uniform convergence.

Now, relatively analogous to the case $C([0,\infty),\mathbb{R})$ we have the following extension for the Skorokhod topology.

Theorem A.2. *The space $\mathbb{D}([0,\infty);\mathbb{R})$ can be endowed with a the metric*

$$d^\circ(f,g)=\sum_{n=1}^{\infty}\frac{1}{2^n}\left(1\wedge d_n^\circ(f^n,g^n)\right),$$

where

$$d_n^\circ(f,g)=\inf_{\lambda\in\Lambda([0,n])}\{\|\lambda\|^\circ\vee d_\infty(f,g\circ\lambda)\},$$

and

$$f^n(t)=f(t)\mathbb{1}_{[0,n-1)}(t)+f(t)(n-t)\mathbb{1}_{[n-1,n)}(t).$$

Furthermore, the space $(\mathbb{D}([0,\infty),\mathbb{R}),d^\circ)$ is a Polish metric space.

Remark: Similar to d_∞ we will use d° for both the Skorokhod metric in on finite and infinite time domains.

We refer the reader to [4] for a full proof.

A.3 Measurability of Canonical Projections

This section is based on [4] and [27].

While the entire law of our process is certainly of interest it is more often convenient to consider the process in some specified times. Very often in probability we are not so fortunate that we can observe the entire path, but rather only observe the path at some specified times. In order to preserve measurability we will need the so-called *canonical projections*.

Definition A.7 (Finite Sequences and Canonical Projections)
Let $\mathcal{T}$ be of the form $[0,T]$ with $T<\infty$ or $[0,\infty)$. We define the set

$$\mathscr{T}(\mathcal{T})=\{(t_n)_{n=1}^N\subset\mathcal{T}\mid N\in\mathbb{N},t_i<t_{i+1},i\in\{1,2,\dots,N-1\}\}\tag{A.5}$$

as the set of finite sequences. We define the natural, or canonical, projection onto $\tau=(t_1,t_2,\dots,t_{|\tau|})\in\mathscr{T}(\mathcal{T})$ as

$$\begin{aligned}\pi_\tau:\mathbb{D}([0,\infty);\mathbb{R})&\to\mathbb{R}^{|\tau|},\\ x&\mapsto(x(t_1),x(t_2),\dots,x(t_{|\tau|})).\end{aligned}$$

where $|\tau|$ is the cardinality, or length, of τ.

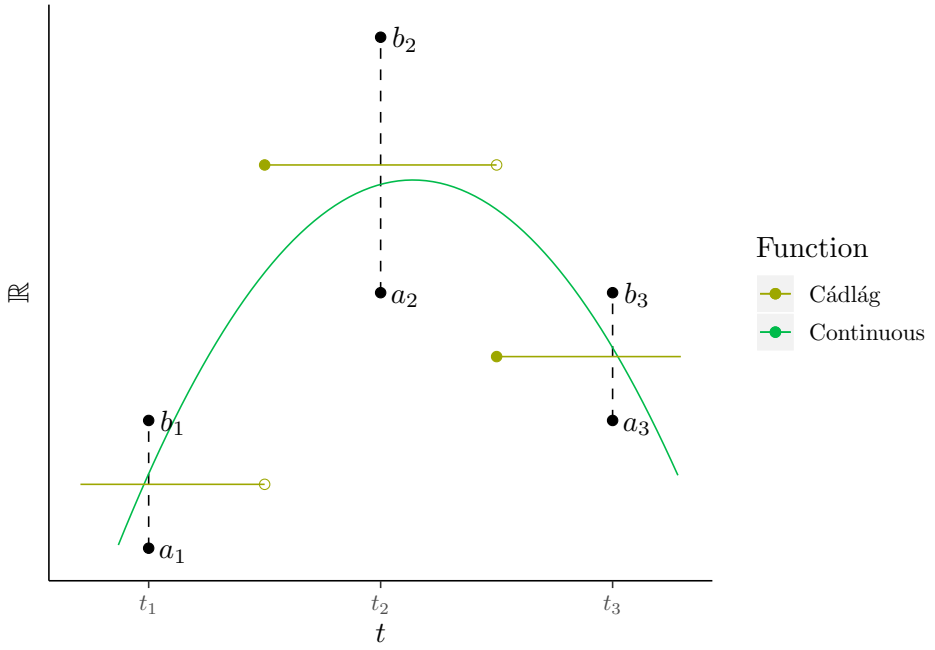

Figure A.3: Function surfaces in $\mathbb{D}([0,\infty);\mathbb{R})$ passing through $[a_i, b_i]$ at time t_i, where $i \in \{1,2,3\}$

Remark: The special case π_τ where $\tau = t$ for some $t \in \mathcal{T}$ is called the evaluation mapping.

We wish to describe the law of X evaluated at times $(t_1, t_2, \ldots, t_n) = \tau \in \mathscr{T}([0,\infty))$. Furthermore, as time progresses some values of the path will be realised which may alter the distribution for the remainder of the path. We will need to introduce some formal tools for describing this behaviour.

Definition A.8 (Finite Dimensional Distributions)
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathcal{T}$ be of the form $[0,T]$ with $T < \infty$ or $[0,\infty)$ and $X : \Omega \to \mathbb{D}(\mathcal{T};\mathbb{R})$ be a stochastic process. We define the finite distributions of X as

$$\mathcal{L}(X_\tau)(B) = \mathbb{P} \circ X^{-1} \circ \pi_\tau^{-1}(B) = \mathbb{P}(\pi_\tau X \in B), \quad \tau \in \mathscr{T}(\mathcal{T}), B \in \mathcal{B}(R^{|\tau|}),$$

where $|\tau|$ is the cardinality, or length, of the sequence τ.

Let us interpret the push-forward measure in Definition A.8. For simplicity we consider $\tau = (t_1, t_2, t_3)$ and $B = [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$ with $a_i \leq b_i$ for all $i \in \{1,2,3\}$. Now, we may think of $\pi_\tau^{-1}(B) = \pi^{-1}_{(t_1,t_2,t_3)}([a_1, b_1] \times [a_2, b_2] \times [a_3, b_3])$ as asking which function surfaces in $\mathbb{D}([0,\infty);\mathbb{R})$ passes through $[a_1, b_1]$ at time t_1 and through $[a_2, b_2]$ at time t_2 and finally $[a_3, b_3]$ at time t_3. We visualise in Figure A.3, where we see both a continuous and cádlág function.

Given the preimage $\pi_\tau^{-1}(B)$ is a Borel set in $\mathcal{B}(\mathbb{D}([0,\infty);\mathbb{R}))$ induced by either the topology of local uniform convergence or the Skorokhod topology we then use the push-forward measure $\mathbb{P} \circ X^{-1}$ to ask what is the probability of "sampling" a cádlág process that passes through $[a_i, b_i]$ at time t_i.

We have yet to justify that π_τ is measurable. This is important because otherwise $\mathbb{P}(\pi_\tau X \in B)$ is not necessarily well defined. We will now justify their measurability.

Proposition A.2. *Let $\mathcal{T}$ be of the form $[0,T]$ for $T<\infty$ or $[0,\infty)$ and let $\tau \in \mathscr{T}(\mathcal{T})$. Then π_τ is continuous in topology of uniform convergence.*

Proof. We must show that whenever for all $\varepsilon > 0$ and every $x \in \mathbb{D}([0,T];\mathbb{R})$ exists a $\delta > 0$ such that

$$d_\infty(x,y) < \delta \implies \|\pi_\tau x - \pi_\tau y\| < \varepsilon. \tag{A.6}$$

Where $\|\cdot\|$ is the norm on $\mathbb{R}^{|\tau|}$. Note that $\pi_\tau x \in \mathbb{R}^{|\tau|}$. To this end, note that

$$\delta > d_\infty(x,y) = \sup_{t\in[0,T]} |x(t)-y(t)| \geq |x(t)-y(t)|, \quad \forall s \in [0,T]. \tag{A.7}$$

And note that $\|\pi_\tau f - \pi_\tau g\|$ can be written as

$$\|\pi_\tau f - \pi_\tau g\| = \sqrt{\sum_{i=1}^{|\tau|}(f(t_i)-g(t_i))^2}$$

Now, choosing $\delta = \varepsilon/\sqrt{|\tau|+1}$ and using (A.7) yields

$$\begin{aligned}\sqrt{\sum_{i=1}^{|\tau|}(f(t_i)-g(t_i))^2} &< \sqrt{\sum_{i=1}^{|\tau|}(\delta)^2} \\ &= \sqrt{\sum_{i=1}^{|\tau|}\frac{\varepsilon^2}{|\tau|+1}} \\ &= \sqrt{\frac{|\tau|}{|\tau|+1}}\varepsilon < \varepsilon.\end{aligned}$$

■

This justifies that π_τ is measurable on the topology of uniform convergence. For the Skorokhod topology π_τ is continuous only if τ are points at which the function $x \in \mathbb{D}([0,T];\mathbb{R})$ is continuous, as shown in Proposition A.2. However, we can use the result above to recover measurability in the Skorokhod topology using a result shown in [27].

Theorem A.3. *Let $l : \mathbb{D}([0,T];\mathbb{R}) \to \mathbb{R}$ be a continuous linear functional on the topology of uniform convergence, then it is measurable with respect to the Borel σ-algebra generated by the Skorokhod topology.*

Then, for each $t \in [0,T]$ we have π_t is continuous and hence measurable by Proposition A.2. Clearly, π_t is linear, so π_t is measurable with respect to the Skorokhod topology. It follows that π_τ is measurable, as it may be decomposed as π_{t_i} for each $t_i \in \tau$ with $\tau \in \mathscr{T}([0,T])$.

For certain points we may recover continuity even in the Skorokhod topology.

Proposition A.3. *Consider the space $\mathbb{D}([0,T];\mathbb{R})$ where $T<\infty$. The projections π_0 and π_T are continuous, and for $t_0 \in (0,T)$. Then π_{t_0} is continuous in the Skorokhod topology only on the set*

$$\mathscr{C}(t) = \{x \in \mathbb{D}([0,T];\mathbb{R}) \mid x \text{ is continuous at } t_0\}.$$

Proof. We first use the characterisation of continuity in the Skorokhod topology and show it for the metric

$$d_S(x,y) = \inf_{\lambda\in\Lambda([0,T])}\{d_\infty(\lambda,\mathrm{Id}) \vee d_\infty(x, y\circ\lambda)\}.$$

We must show that for every $x \in \mathscr{C}(t)$ and $\varepsilon > 0$ exists a $\delta > 0$ such that

$$d_S(x,y) < \delta \implies |\pi_t x - \pi_t y| < \varepsilon$$

Let $\varepsilon > 0$ be given. First, note that

$$d_S(x,y) \geq \max(|x(0)-y(0)|, |x(1)-y(1)|) = \max(|\pi_0 x - \pi_0 y|, |\pi_T x - \pi_T y|),$$

since $\lambda(0) = 0$ and $\lambda(T) = T$, $\forall\lambda \in \Lambda([0,T])$. It follows that π_0 and π_T are continuous in the Skorokhod topology by choosing $\delta = \varepsilon$.

Now, fix $t_0 \in (0,T)$. We will show that π_{t_0} continuous at x if, and only if, $x \in \mathscr{C}(t_0)$. First, note that $x \in \mathscr{C}(t_0)$ implies that

$$\exists\delta_0 > 0 : |t - t_0| < \delta_0 \implies |x(t) - x(t_0)| < \frac{\varepsilon}{2}.$$

It follows that if $\delta_0 > d_\infty(\lambda, \mathrm{Id})$ we have

$$|\lambda(t_0) - t_0| < \delta_0 \implies |x(\lambda(t_0)) - x(t_0)| < \frac{\varepsilon}{2}.$$

Now,

$$\begin{aligned}|\pi_{t_0} y - \pi_{t_0} x| = |y(t_0) - x(t_0)| &= |y(t_0) - x(\lambda(t_0)) + x(\lambda(t_0)) - x(t_0)| \\ &\leq |y(t_0) - x(\lambda(t_0))| + |x(\lambda(t_0)) - x(t_0)|. \end{aligned} \tag{A.8}$$

Now, let $\delta = \min(\delta_0, \varepsilon/2)$, $x \in \mathscr{C}(t)$, $y \in \mathbb{D}([0,T];\mathbb{R})$, and $\lambda \in \Lambda([0,T])$ such that

$$\delta > (d_\infty(\lambda,\mathrm{Id}) \vee d_\infty(y, x\circ\lambda)) \geq d_S(y,x). \tag{A.9}$$

Then, clearly

$$\delta > d_\infty(\lambda,\mathrm{Id}) \geq |\lambda(t) - t|, \quad \forall t \in T.$$

and hence also specifically in $t = t_0$ which implies

$$\delta > d_\infty(\lambda,\mathrm{Id}) \implies \delta > |\lambda(t_0) - t_0| \implies |x(\lambda(t_0)) - x(t_0)| < \frac{\varepsilon}{2}.$$

Furthermore,

$$\delta > d_\infty(y, x\circ\lambda) \geq |y(t) - x(\lambda(t))|, \quad \forall t \in T,$$

which also holds at $s = t$. Hence,

$$|y(t_0) - x(\lambda(t_0))| < \delta \leq \frac{\varepsilon}{2}.$$

By (A.8) we now have

$$|\pi_{t_0} y - \pi_t x| \leq |y(t_0) - x(\lambda(t_0))| + |x(\lambda(t_0)) - x(t_0)| < \varepsilon.$$

Taking the infimum of $\lambda \in \Lambda([0,T])$ will by (A.9) yield the same. Hence $x \in \mathscr{C}(t_0)$ implies π_{t_0} continuous at x.

We must now show that if $x \notin \mathscr{C}(t_0)$ yields π_{t_0} not continuous. By Proposition A.1 it suffices to show that if $x_n \to x$ in the Skorokhod topology, but $\pi_{t_0} x_n \not\to \pi_{t_0} x$ then π_{t_0} is not continuous at x. To this end define

$$x_n(t) = \begin{cases} x(0) & t \in [0, \frac{1}{2^n}] \\ x(t - \frac{1}{2^n}) & t \in (\frac{1}{2^n}, T] \end{cases}.$$

Clearly, $x_n \to x$ in the Skorokhod topology, by similar arguments as in Example 2.5. Since x is cádlág and discontinuous at t_0 we have $x(t_0-) = \lim_{t \uparrow t_0} x(t) \neq x(t_0)$. It now follows that

$$\lim_{n\to\infty} \pi_{t_0} x_n = \lim_{n\to\infty} x_n(t_0) = \lim_{n\to\infty} x\left(t_0 - \frac{1}{2^n}\right) = \lim_{t\uparrow t_0} x(t) \neq x(t_0) = \pi_t x.$$

■

A small corollary follows.

Corollary A.1. *Consider the space $\mathbb{D}([0,T];\mathbb{R})$ where $T < \infty$. Let $\tau \in \mathscr{T}((0,T))$, then π_τ is continuous only on the set*

$$\mathscr{C}(\tau) = \bigcap_{t\in\tau} \mathscr{C}(t)$$

Remark: Beware that we are taking $\tau \in \mathscr{T}((0,T))$ rather than $\mathscr{T}([0,T])$ but the definition is analogous. That is,

$$\mathscr{T}(0,T) = \{(t_n)_{n=1}^N \subset (0,T) \mid N \in \mathbb{N}, t_i < t_{i+1}, i \in \{1,2,\ldots,N-1\}\}.$$

A.4 Martingales and Stopping Times

This section is based on [28].

In this section we present some definition and results for Martingales.

Definition A.9 (Martingale)
Let $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\infty)}, \mathbb{P}\right)$ be a filtered probability space. A stochastic process X is said to be a martingale if $\mathbb{E}[|X_t|] < \infty$ and

$$\mathbb{E}[X_t \mid \mathcal{F}_s] = X_s$$

Remark: Martingales rely heavily on the probability measure and filtration; a martingale on a space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,\infty)}, \mathbb{P}\right)$ may not be a martingale if we endow Ω with a different σ-algebra, filtration, and probability measure.

Definition A.10 (Stopping Time)
Let $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\infty)}, \mathbb{P}\right)$ be a filtered probability space. $\tau : \Omega \to [0,\infty)$ is said to be a stopping time if

$$\{\tau \leq t\} \overset{def}{=} \{\omega \mid \tau(\omega) \leq t\} \in \mathcal{F}_t.$$

The term *stopping-time* can be slightly misleading. In themselves they are not stopping anything. However, if X is a process then

$$X_{\tau \wedge t} = X_{\min\{\tau, t\}}.$$

is a process, which is constant - hence stopped - after time τ. Some interesting problems can be turned into a stopping-time problem.

Lemma A.3. *Let X be real valued adapted cádlág stochastic process and K be a closed subset of $\mathbb{R}$. Then the random variable*

$$T(\omega) = \inf\{t > 0 \mid X_t(\omega) \in K \text{ or } X_{t-}(\omega) \in K\}$$

is a stopping time.

Proof. We must show

$$\{T \leq t\} \in \mathcal{F}_t$$

We first define

$$A_n = \left\{ x \;\middle|\; \inf_{k \in K} |x - k| < \frac{1}{n} \right\}.$$

A_n is an open set with $K \subset A_n$. The set $[0, t)$ is uncountable, but consider $\mathbb{Q} \cap [0, t)$ which is clearly dense and countable. Hence,

$$\bigcup_{s \in \mathbb{Q} \cap [0,t)} \{\omega \mid X_s(\omega) \in A_n\}$$

clearly belongs to $\mathcal{F}_t$ since X is measurable and adapted and A_n is open for each n. By extension

$$\bigcap_{n \in \mathbb{N}} \bigcup_{s \in \mathbb{Q} \cap [0,t)} \{\omega \mid X_s(\omega) \in A_n\}$$

is a countable intersection and belongs to $\mathcal{F}_t$. Clearly,

$$\{\omega \mid X_t \in K \text{ or } X_{t-} \in K\}$$

is measurable, and so it follows that

$$\{\omega \mid X_t \in K \text{ or } X_{t-} \in K\} \cup \left(\bigcap_{n \in \mathbb{N}} \bigcup_{s \in \mathbb{Q} \cap [0,t)} \{\omega \mid X_s(\omega) \in A_n\} \right) \in \mathcal{F}_t.$$

But we have

$$\{\omega \mid T(\omega) \leq t\} = \{\omega \mid X_t \in K \text{ or } X_{t-} \in K\} \cup \left(\bigcap_{n \in \mathbb{N}} \bigcup_{s \in \mathbb{Q} \cap [0,t)} \{\omega \mid X_s(\omega) \in A_n\} \right).$$

■

The stopping time T in Lemma A.3 is called the *first hitting time* of the set K. The result can be extended into hitting any Borel set, but we will not cover this result.

From a financial perspective the first hitting time is naturally of interest; suppose we wish to hold portfolio of assets until the portfolio has reached some return. Naturally, there is a 1-to-1 correspondence between the return and the value of the portfolio, i.e. we obtain the desired

return r^* when the value of our portfolio reaches S^* and so the hitting time of the set $[S^*, \infty)$ is of interest. Deriving the law of T is of particular interest as it yields knowledge of the uncertainty as well.

Now, similarly to how we may be interested in the history of a stochastic process prior to a time t, we may similarly be interested in the history of a stochastic process up until time T, where T is a stopping time.

Definition A.11 (Stopping Time Sigma Algebra)
Let $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,\infty)}, \mathbb{P}\right)$ be a filtered probability space and let T be a stopping time. We define the σ-algebra $\mathcal{F}_T$ as

$$\mathcal{F}_T = \{B \in \mathcal{F} \mid B \cap \{T \leq t\} \in \mathcal{F}_t, t \in [0, \infty)\}.$$

We have the following Lemma

Lemma A.4. *let $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,\infty)}, \mathbb{P}\right)$ and $T : \Omega \to [0, \infty)$ be a stopping time. Then $\mathcal{F}_T$ is the coarsest, or smallest, σ-algebra containing all cádlág processes which are measurable at time T.*

Proof. Let $\mathcal{G} = \sigma(\{\omega \mid X_{T(\omega)}(\omega) \in B, B \in \mathcal{B}(\mathbb{R}), X \text{ cádlág and adapted }\})$. Let $F \in \mathcal{F}_T$. Note that $\mathbb{1}_F \mathbb{1}_{[T,\infty)}(t)$ is a cádlág process with $X_T = \mathbb{1}_F$. Hence, $F \in \mathcal{G}$ and it follows that $\mathcal{F}_T \subseteq \mathcal{G}$.

Now, let X be an adapted cádlág process. Consider $X : \Omega \times [0, \infty) \to \mathbb{R}$; it must be measurable since it is adapted. Now, fix $t \in [0, \infty)$ and define

$$\begin{aligned} &\varphi : \{\omega \mid T(\omega) \leq t\} \to [0, \infty) \times \Omega, \\ &\omega \mapsto (T(\omega), \omega). \end{aligned}$$

Note that both X and φ are measurable measurable, so the mapping $X \circ \varphi$ is a measurable mapping from $(\{\omega \mid T(\omega) \leq t\}, \mathcal{F}_t \cap \{\omega \mid T(\omega) \leq t\})$ to $\mathcal{B}(\mathbb{R})$, i.e.

$$\{\omega \mid X(T(\omega), \omega) \in B, B \in \mathcal{B}(\mathbb{R})\} \cap \{\omega \mid T(\omega) \leq t\} \in \mathcal{F}_t,$$

but this implies $X_T \in \mathcal{F}_T$ from which it follows that $\mathcal{G} \subseteq \mathcal{F}_T$. ■

Now, the goal of the previous Lemma, to construct a filtration induced by stopping times. The proceeding Corollary is therefore central to justifying

Corollary A.2. *Let S and T be stopping times and let*

$$S \leq T, \quad a.s.$$

Then

$$\mathcal{F}_S \subseteq \mathcal{F}_T.$$

Now, the following definition is a technical one.

Definition A.12 (Closure of a Martingale)
Let X be a martingale on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0,\infty)}, \mathbb{P}\right)$. X is said to be closed by a random variable Y if $\mathbb{E}[|Y|] < \infty$ and

$$X_t = \mathbb{E}[Y \mid \mathcal{F}_t], \quad \forall t \in [0, \infty).$$

We can now present the so-called *Optional Sampling Theorem.*

Theorem A.4 (Optional Sampling Theorem)**.** *Let X be a right continuous martingale with, which is closed by a random variable Y. Let S and T be stopping times such that $S \leq T$ a.s., then X_S and X_T are integrable and*

$$X_S = \mathbb{E}[X_T \mid \mathcal{F}_S], \quad a.s.$$

while at the surface the Theorem seems obvious; once S and T are realised by some ω then they correspond to some $s, t \in [0, \infty)$ with $s \leq t$, and for this s and t the martingale property hold. However, the theorem is incredibly valuable when we have to prove certain properties of martingales, as we will later see.

Definition A.13 (Local Martingale)
Let $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{[0\infty}, \mathbb{P}\right)$ be a probability space and let $X : \Omega \to \mathbb{D}([0, \infty); \mathbb{R})$ be a stochastic process. If there exists a sequence of stopping times $(\tau_n)_{n\in\mathbb{N}}$ such that

$$\mathbb{P}(\tau_n \leq \tau_{n+1}) = 1, \quad n \in \mathbb{N}$$
$$\mathbb{P}(\lim_{n\to\infty} \tau_n = \infty) = 1,$$

such that the process

$$X_{t\wedge\tau_n}$$

is a martingale for every n, then X is said to be a local martingale.

Clearly, every martingale is a local martingale, but the converse is not true.

A.5 Weak Convergence and Donsker's Theorem

This subsection is based on [4].

In this subsection we present the notion of weak convergence of probability measures and Donsker's theorem, included only for self-refrencing.

Definition A.14 (Weak Convergence of Probability Measures)
Let (E, d_E) be a Polish metric space with Borel σ-algebra $\mathcal{B}(E)$. A sequence of probability measures $(\mathbb{P}_n)_{n\in\mathbb{N}}$ is said to converge weakly to a probability measure $\mathbb{P}$ if

$$\lim_{n\to\infty} \int_E f d\mathbb{P}_n = \int_E f d\mathbb{P}, \quad \forall f \text{ continuous, real valued and bounded.}$$

Similarly to stable convergence in law, it may be ideal to think of the limit as a random variable on an appropriate probability space and $\mathbb{P}_n$ as the law of a sequence of random variables, i.e. $\mathbb{P}_n = \mathcal{L}_{Z_n}$ for some sequence of random variables Z_n with limit Z such that $\mathcal{L}_Z = \mathbb{P}$. We then write

$$Z_n \Rightarrow Z,$$

to mean that Z_n converges weakly to Z. This motivates the so-called *Functional Central Limit Theorem* also known as Donsker's Theorem.

Theorem A.5 (Donsker's Theorem)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $(\xi_m)_{m\in\mathbb{N}}$ be a sequence of independent random variables with mean 0 and variance σ^2. Then, define S by the mapping*

$$t \mapsto \sum_{i=1}^{[nt]} \xi_i,$$

then

$$\frac{1}{\sigma\sqrt{n}} S \Rightarrow W,$$

on $(\mathbb{D}([0,1];\mathbb{R}), \mathcal{B}(\mathbb{D}([0,1];\mathbb{R})))$.

For a proof see [[4], p. 146 Theorem 14.1]

Bibliography

[1] Barndorff-Nielsen, Ole E. "Normal inverse Gaussian distributions and stochastic volatility modelling". In: *Scandinavian Journal of statistics* 24.1 (1997), pp. 1–13.

[2] Barndorff-Nielsen, Ole E and Shephard, Neil. "Econometric analysis of realized volatility and its use in estimating stochastic volatility models". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64.2 (2002), pp. 253–280.

[3] Barndorff-Nielsen, Ole E and Shephard, Neil. "Econometrics of testing for jumps in financial economics using bipower variation". In: *Journal of financial Econometrics* 4.1 (2006), pp. 1–30.

[4] Billingsley, Patrick. *Convergence of probability measures*. John Wiley & Sons, 2013.

[5] Billingsley, Patrick. *Probability and measure*. John Wiley & Sons, 2008.

[6] Brouste, Alexandre et al. "The yuima project: A computational framework for simulation and inference of stochastic differential equations". In: American statistical association. 2014.

[7] Cherubini, Umberto et al. *Dynamic copula methods in finance*. John Wiley & Sons, 2011.

[8] Christoffersen, Peter F. "Evaluating interval forecasts". In: *International economic review* (1998), pp. 841–862.

[9] Dudley, RM et al. "Sample Functions of the Gaussian Process". In: *The Annals of Probability* 1.1 (1973), pp. 66–103.

[10] Durante, Fabrizio, Fernández-Sánchez, Juan, and Sempi, Carlo. "Sklar's theorem obtained via regularization techniques". In: *Nonlinear Analysis: Theory, Methods & Applications* 75.2 (2012), pp. 769–774.

[11] Heston, Steven L. "A closed-form solution for options with stochastic volatility with applications to bond and currency options". In: *The review of financial studies* 6.2 (1993), pp. 327–343.

[12] Iacus, Stefano M. *Simulation and inference for stochastic differential equations: with R examples*. Springer Science & Business Media, 2009.

[13] Jacod, Jean and Protter, Philip. *Discretization of processes*. Vol. 67. Springer Science & Business Media, 2011.

[14] Jacod, Jean and Shiryaev, Albert. *Limit theorems for stochastic processes*. Vol. 288. Springer Science & Business Media, 2013.

[15] Kechris, Alexander. *Classical descriptive set theory*. Vol. 156. Springer Science & Business Media, 2012.

[16] Kerr, David and Li, Hanfeng. "Ergodic theory". In: *Springer Monographs in Mathematics. Springer, Cham* (2016).

[17] Kim, Sangyeon and Kang, Myungjoo. "Financial series prediction using Attention LSTM". In: *arXiv preprint arXiv:1902.10877* (2019).

[18] Lagodowski, Zbigniew A and RYCHLIK, Zdzislaw. "Weak Convergence of Probability Measures on the Function Space Dd [0, oo". In: (1986).

[19] Li, David X. "On default correlation: A copula function approach". In: *The Journal of Fixed Income* 9.4 (2000), pp. 43–54.

[20] Martin, Ole. *High-Frequency Statistics with Asynchronous and Irregular Data*. Springer, 2019.

[21] Mikosch, Thomas. *Copulas: tales and facts*. University of Copenhagen, laboratory of Actuarial Mathematics, 2005.

[22] Mikosch, Thomas. "Copulas: Tales and facts–rejoinder". In: *Extremes* 9.1 (2006), pp. 55–62.

[23] Monroe, Itrel. "Processes that can be embedded in Brownian motion". In: *The Annals of Probability* (1978), pp. 42–56.

[24] Nelsen, Roger B. "Properties and applications of copulas: A brief survey". In: *Proceedings of the first brazilian conference on statistical modeling in insurance and finance*. Citeseer. 2003, pp. 10–28.

[25] Neuhaus, Georg et al. "On weak convergence of stochastic processes with multidimensional time parameter". In: *The Annals of Mathematical Statistics* 42.4 (1971), pp. 1285–1295.

[26] Patton, Andrew John. "Applications of copula theory in financial econometrics." In: (2003).

[27] Pestman, Wiebe R et al. "Measurability of linear operators in the Skorokhod topology". In: *Bulletin of the Belgian Mathematical Society-Simon Stevin* 2.4 (1995), pp. 381–388.

[28] Protter, Philip. "Stochastic Integration and Differential Equation". In: *Stochastic Modeling and Applied Probability* 21 (2004).

[29] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2019. URL: `https://www.R-project.org/`.

[30] R-SIG-DB, Wickham, Hadley, and and, Kirill Müller. *DBI: Database interface for R*. 2016. URL: `https://ggplot2.tidyverse.org`.

[31] Revuz, Daniel and Yor, Marc. *Continuous martingales and Brownian motion*. Vol. 293. Springer Science & Business Media, 2013.

[32] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc. Boston, MA, 2015. URL: `http://www.rstudio.com/`.

[33] Sauri, Orimar A. *Measure Theory and Stochastic Processes: Stochastic Analysis and its Applications in Finance*. 2018.

[34] Schmitz, Volker. "Copulas and stochastic processes". PhD thesis. Bibliothek der RWTH Aachen, 2003.

[35] Schweizer, Berthold and Sklar, Abe. *Probabilistic metric spaces*. Courier Corporation, 2011.

[36] Sharpsteen, Charlie and Bracken, Cameron. "tikzDevice: R Graphics Output in LaTeX Format". In: *R package version 0.7. 0. URL http://CRAN. R-project. org/package= tikzDevice* (2013).

[37] Sklar, Abe, SKLAR, A, and Sklar, CA. "Fonctions de reprtition an dimensions et leursmarges". In: (1959).

[38] Skorokhod, Anatoly V. "Limit theorems for stochastic processes". In: *Theory of Probability & Its Applications* 1.3 (1956), pp. 261–290.

[39] Steland, Ansgar. *Financial statistics and mathematical finance: methods, models and applications*. John Wiley & Sons, 2012.

[40] Veraart, Almut ED and Winkel, Matthias. "Time change". In: *Encyclopedia of quantitative finance* (2010).

[41] Watts, Samuel. *The Gaussian Copula and the Financial Crisis: A Recipe for Disaster or Cooking the Books?* 2016.

[42] Wickham, Hadley. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN: 978-3-319-24277-4. URL: `https://ggplot2.tidyverse.org`.

[43] Zimmer, David M. "The role of copulas in the housing crisis". In: *Review of Economics and Statistics* 94.2 (2012), pp. 607–620.

[44] Zukav, Gary. *The dancing Wu Li masters: An overview of the new physics*. Random House, 2012.