UNIVERSITY
OF OSLO

Dennis Schroers

New Topics in Nonlinear Functional Data Analysis

Thesis submitted for the degree of Philosophiae Doctor

Department of Mathematics
Faculty of Mathematics and Natural Sciences

University of Oslo

2022

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 2545

ISSN 1501-7710

Cover: Hanne Baadsgaard Utigard.
Print production: Graphics Center, University of Oslo.

Acknowledgements

First, I want to thank my supervisors Fred Espen Benth and Giulia Di Nunno for their guidance and patience as well as mathematical insights that I have benefited from over the past three years. Their enthusiasm for mathematics is contagious and inspiring. I also want to thank Almut Veraart for the plenty of mathematical discussions and for giving helpful advice and valuable insights into various aspects of academic life. I am grateful to all my great colleagues and friends from UiO for making the office a stimulating and familiar place to be. In particular, I would like to thank Anton Yurchenko-Tytarenko, with whom I shared an office throughout all three Ph.D. years, had countless mathematical discussions, and solved many problems, not just mathematical ones. Finally, I want to thank my family for supporting me unconditionally until this point. I especially thank Siri for her endless optimism and all her love.

Funding for this position within the project STORM: Stochastics for Time-Space Risk Models, from the Research Council of Norway (Project number: 274410) is gratefully acknowledged.

Dennis Schroers
Oslo, August 2022

List of Papers

Paper I

Benth, F.E., Di Nunno, G. and Schroers, D. "Copula Measures and Sklar's Theorem in Arbitrary Dimensions". To appear in *Scandinavian Journal of Statistics.* (2021), DOI: 10.1111/sjos.12559.

Paper II

Benth, F.E., Di Nunno, G. and Schroers, D. "A Topological Proof of Sklar's Theorem in Arbitrary Dimensions". In: *Dependence Modeling.* Vol. 10, no. 1 (2022), pp. 22–28. DOI: 10.1515/demo-2022-0103.

Paper III

Benth, F.E., Schroers, D., and Veraart, A.E.D. "A Weak Law of Large Numbers for Realised Covariation in a Hilbert Space Setting". In: *Stochastic Processes and their Applications.* Vol. 145, (2022), pp. 241–268. DOI: 10.1016/j.spa.2021.12.011.

Paper IV

Benth, F.E., Schroers, D., and Veraart, A.E.D. "A Feasible Central Limit Theorem for Realised Covariation of SPDEs in the Context of Functional Data". *Submitted for publication.*

Contents

Chapter 1

Introduction

This dissertation introduces two well-known concepts from multivariate statistics to the field of functional data analysis: *power variations* and *copulas*. Although these techniques are applied in different situations, both can be classified as nonlinear methods.

Functional data analysis comprises methods for the statistical analysis of data that can be considered as discretely or fully observed random curves, surfaces, or related smooth objects. Numerous applications can be found in areas such as finance, economics, meteorology, biomechanics, psychology or neurophysiology (cf. [89]), where functional data analysis has proven to be an invaluable approach to reveal smooth features from inherently infinite-dimensional data. In that regard, the realm of functional data differs fundamentally from the field of multivariate statistics and entails many new mathematical challenges.

In the simplest situation, we assume to observe i.i.d. copies $X_1, ..., X_n$ of a second-order random variable X taking values in a separable Hilbert space H such that $\mathbb{E}[\|X\|_H^2] < \infty$. The emblematic problem is to make inferences on the mean $m := \mathbb{E}[X]$ and the covariance operator $C := \mathbb{E}[X^{\otimes 2}]$, where $h^{\otimes 2} = \langle h, \cdot \rangle_H h$ denotes the usual tensor product. Often, $X = (X_t)_{t \in [0,T]}$ is a real-valued stochastic process such that $\mathbb{E}[X_t^2] < \infty$ for all $t \in [0, T]$ and is considered as an element in the Lebesgue space $L^2([0, T])$ of (equivalence classes of) square-integrable functions $f : [0, T] \to \mathbb{R}$ for some $T > 0$. In this case, the covariance operator can be expressed by an integral operator

$$Cf(t) = \int_0^T c(t, s) f(s) ds \quad t \in [0, T],$$

where $c(t, s) = cov(X(t), X(s))$ is the covariance kernel of X. Finding ways to estimate C (or equivalently, the kernel c) is of the utmost importance as this allows for the well-known functional principal component analysis, arguably the cornerstone of the theory and a key method to make infinite-dimensional data tractable. The idea is to make use of the fact that the nuclearity of C yields an eigendecomposition of the form

$$C = \sum_{j=1}^{\infty} \lambda_j e_j^{\otimes 2},$$

where $\lambda_j \geq 0$ are the positive eigenvalues of C in decreasing order and $(e_j)_{j \in \mathbb{N}}$ is the corresponding orthonormal system of eigenvectors. If c is continuous, this decomposition can also be expressed on the level of the covariance kernel for $s, t \in [0, T]$ by

$$c(s, t) = \sum_{j=1}^{\infty} \lambda_j e_j(s) e_j(t)$$

where the convergence holds uniformly in $[0,T]^2$. This is commonly known as Mercer's Lemma, cf. [26, Lemma 1.3]. The derivation of these eigenelements yields an optimal linear approximation of the infinite-dimensional random variable X by virtue of the fundamental Karhunen-Loève expansion (cf. [26, Theorem 1.5]). More precisely, the family $(Z_n)_{n\in\mathbb{N}} := (\langle X, e_n\rangle)_{n\in\mathbb{N}}$ of real-valued random variables is by definition pairwise uncorrelated with mean zero, variance $\mathbb{E}\left[Z_n^2\right] = \lambda_n$, and for all $t \in [0,T]$ we have with respect to the $L^2(\Omega)$-distance for large N that

$$X_t^N - m_t := \sum_{i=1}^{N} Z_n e_n(t) \approx \sum_{i=1}^{\infty} Z_n e_n(t) = X_t - m_t.$$

The convergence of the random series on the right hand-side holds uniformly on $[0,T]$ with respect to the $L^2(\Omega)$-norm. This allows us to approximate the infinite-dimensional random variable X with just a finite number of random sources $Z_1, ..., Z_N$ for $N \in \mathbb{N}$ by the best N-dimensional linear approximation in terms of the mean squared error. That is, for any orthonormal system $(f_n)_{n\in\mathbb{N}}$ and $\bar{X}_t^N = \sum_{n=1}^{N} \langle X, f_n\rangle f_n(t)$ we have

$$\mathbb{E}\left[\|X - X^N\|^2_{L^2([0,T])}\right] \leq \mathbb{E}\left[\|X - \bar{X}^N\|^2_{L^2([0,T])}\right].$$

It is clear that the success of this procedure hinges on the quality of the estimator for the covariance C. In the simple i.i.d. setting with fully observed curves, it is not hard to prove that the empirical mean $\hat{m}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$ and empirical Covariance $\hat{C}_n := \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{m}_n)^{\otimes 2}$ form consistent estimators of m and C respectively. The empirical covariance has finite rank n and has at most n positive eigenvalues $\hat{\lambda}_{1,n}, ..., \hat{\lambda}_{n,n}$, which we can derive from $\hat{C}_n$ together with the corresponding eigenfunctions $\hat{e}_{1,n}, ..., \hat{e}_{n,n}$. These values and functions are consistent estimators of the first n eigenvalues $\lambda_1, ..., \lambda_n$ and eigenvectors $e_1, ..., e_n$ of C (cf. Lemma 4.2 and Theorem 4.4 in [26]).

In that regard, Many authors consider the works of Karhunen [76] and Grenander [59] as a potential starting point of the field. Since then, the literature on the statistical theory for functional data has evolved considerably and includes in addition to functional principal component analysis (cf. [45], [92], [94], [61]) techniques such as functional linear regression (see [29],[28], [60]), the analysis of functional time series (cf. [26], [86]) and corresponding methods for sparsely sampled functional data (cf. [98], [99], [84], [62]). Multiple textbooks on the topic such as [90], [67], [49] or [66] have appeared as well. These techniques are remarkable in the sense that they can extract major information from infinite-dimensional objects parsimoniously. However, the majority of available methods is inherently linear and assumes simple or weak forms of dependence between the functional data. Indeed, the functional principal component analysis described in the basic case above accounts for linear dependence patterns (covariance) and relies on the often unrealistic i.i.d. assumption on the random curves. According to the survey [96], the predominant focus on linear models might be due to "the complexity of functional data analysis, which blends stochastic process theory, functional analysis, smoothing and multivariate techniques" [96, p.24]. At the same time, they argue that as "more and more functional data are being generated, it has emerged that many such data have inherent nonlinear features that make linear methods less effective" [96, p.4].

As an example in which the success of a statistical method for functional data might be prone to neglecting nonlinear patterns, we can consider the evolution of forward curves in infinite-dimensional term structure models. Term structure models relate the time to maturity x of a financial derivative $f_t(x)$ (e.g. forward interest rates) at time t to their empirical and theoretical characteristics and naturally appear in numerous financial contexts. The curve $x \mapsto f_t(x)$ is usually called the forward curve and can be thought of taking values in a function space or a space of equivalence classes of functions such as $L^2([0,T])$. Apart from the classical fixed-income literature (e.g. [53], [31]), forward curves play an important role in markets for variance swaps ([27]), stock markets ([93], [75], [30]) as well as new markets, like the modern intraday or forward markets for energy or weather derivatives (e.g. [13], [14]), where price variation is crucially driven by changes in weather and renewable power production. Based on data from contracts with different maturities that were traded on a particular day, forward curves are usually smoothed either by a parametric method like the classical Nelson-Siegel model or a nonparametric method, such as splines (see [55] for an overview). Recently, kernel-ridge regression was found to perform very well for this purpose in [51]. In that way, one obtains a functional time series of daily observations, which, at least in the nonparametric case, can readily be analysed by methods from functional data analysis. Indeed, an important question is how many factors are essentially driving the evolution of forward curves, and often a principal component analysis is conducted on this time series in order to obtain an answer to this (cf. [31, Section 1.7], [55, Section 3.4]). However, various aspects make an ad hoc usage of (functional) principal component analysis questionable in that regard.

One example is the sensitivity to heavy tails and the difficulty to describe complex patterns of tail dependence, which may be ignored by the decorrelation procedure induced by the Karhunen-Loève expansion. Like in many financial contexts there is, however, evidence for exactly these features along the maturities on interest rate forward curves (see e.g. [74]). Arguably, complex tail dependencies and varying tail patterns of the distributions of points along random curves can barely be covered by just a few components of a linear decomposition. As [25, p.1] points out, the "chance of having outliers or other types of imperfections in the data increases both with the number of observations and their dimension", such issues are particularly delicate in functional data analysis. This has motivated research on robust versions of the functional principal component analysis such as [82], [57], [5], [78], [25] or [101].

A valuable tool which is tailor-made to meet these challenges but has obtained little attention in the context of functional data are copulas. In a nutshell, the theory of copulas allows one to decouple the dependence pattern of a multivariate distribution from its marginal distributions and in that way describes multivariate statistical dependence in a fairly general sense. More precisely, a copula is a multivariate cumulative distribution function $C : \mathbb{R}^d \to [0,1]$ with uniform marginal distributions $C_i(u) = u$ for $u \in [0,1]$ and for any multivariate distribution function $F : \mathbb{R}^d \to [0,1]$ with univariate marginals $F_1, ..., F_d$, we can find a copula C such that

$$C(F_1(x_1), ..., F_d(x_d)) = F(x_1, ..., x_d) \quad \forall x_1, ..., x_d \in \mathbb{R}. \tag{1.1}$$

Vice versa, for any copula and any collection of marginals (1.1) returns a multivariate distribution function with univariate marginals $F_1, ..., F_d$. This is the assertion of Sklar's theorem (cf. [85, Theorem 2.10.9]). By this means, it lays the path for various parametric, semiparametric, or entirely nonparametric statistical methods. Nevertheless, to the best of the author's knowledge, the only paper that considered copulas in a functional analytic context is [63]. There the authors introduce copulas as a countable family of consistent multivariate copulas, which are used to model the dependence structure between basis coefficients of random variables in Hilbert spaces. Equivalently, the authors mention that the copula can be expressed as a probability measure C on the product space $\mathbb{R}^{\mathbb{N}}$, where the finite-dimensional distributions corresponding to this measure resemble the distributions for the basis components. This probabilistic view is also the one that we will take throughout the first half of this work. Importantly, the authors of [63] observe that the second part of Sklar's theorem which allows constructing a probability distribution from a copula and a family of marginals just returns a cylindrical probability measure. This is crucial for the applicability of the method and, arguably, the bottleneck for applying copula theory in function spaces. We will refer to this as the *construction problem* in our work.

In view of potential statistical applications, it might also be useful to go beyond the framework of *basis copulas* in Hilbert spaces from [63]. Indeed, a further problem which reflects the nonlinear nature of copulas is their inconsistency under linear transformations. In contrast to the covariance, copulas are not independent under the change of the orthonormal basis that is used to expand the infinite-dimensional random variable. Indeed, two $L^2[0,T]$-valued random variables X_1 and X_2 could have the same underlying copula in the sense of [63] for one basis (e.g. Fourier), but not for another basis (e.g., Haar wavelets). This is not an infinite-dimensional phenomenon and can happen already for basis changes in $\mathbb{R}^2$ (see Example I.3.7 in Paper I). This is bad news, as we usually do not have a priori a natural candidate for such a basis at hand (such as the eigenvector basis of the covariance) and we mostly sample data on a discrete grid of points on the curve, rather than in the form of basis coefficients. This makes any attempt to find a model-agnostic way for dimension reduction in the spirit of the Karhunen-Loève expansion difficult on the level of such basis copulas. It is also not very satisfactory from the point of view of modelling. A striking feature of copulas in finite dimensions is that they allow modelling marginals and dependence structure of a distribution separately. However, in many cases, we could be interested in modelling the dependence between all multivariate evaluations $X(t_1), ..., X(t_d)$ of a stochastic process X. In the context of forward curves we might want to model the tail dependence of contracts with various maturities. In this case, copulas which model the dependence between basis coefficients cannot always be an adequate tool, since they entail information on the dependence structure of these multivariate distributions only in combination with the marginals of the basis coefficients. This is a strong motivation to continue the research started in [63] and to generalise their framework to allow for a more flexible notion of marginals. We develop this in the first part of this dissertation, in which we also show different solutions to the construction problem in various cases and derive a general theory of copulas in infinite-dimensional settings, which is suitable for applications in functional data analysis.

Another pitfall regarding nonlinear structures in the data may lie in their temporal dependence, as X_i and X_{i-1} for $i = 1, ..., n$ could be sampled closely after each other, and their dependence might not be adequately expressed by a time-invariant (cross-)covariance pattern, but, for instance, show patterns of tail dependence. There are various attempts to model serial dependence in functional time series, such as [26], [64] or [86]. Recent research in functional data analysis evolved around the generalisation of ARCH and GARCH models (see [65],[4],[32], [77] or [91]), which account for heteroskedasticity. These models aim to analyse discrete functional time series, which, for instance, arise as (daily) partitions of a real-valued continuous-time stochastic process and eventually make an assumption of stationarity. They are, moreover, not tailor-made for conducting statistical analyses of the characteristics of infinite-dimensional stochastic processes in continuous time. For instance, we could have a particular interest in the spatio-temporal regularity of such processes which can just be accessed by the proximity of the data in time and space (where "space" means time-to-maturity in the context of forward curves). A reason for this is that entirely nonparametric methods for spatio-temporal data analyses quickly become infeasible. Some research in this direction can nevertheless be found for independent samples of spatio-temporal data such as [83]. Keeping in mind daily observations of forward curves, it is questionable at which time points we should split the data set to obtain a time series of spatio-temporal observations, which lets this heavy machine appear rather brute in this context.

Fortunately, for various spatio-temporal stochastic processes, including forward curves, the underlying theory (e.g. in finance, physics or biology) allows us to sharpen the setting drastically and, in that way, enables us to analyse the data efficiently and adequately. Quite often, it is reasonable to assume the data $X_0 = Y_0, X_1 = Y_{\Delta_n}, X_2 = Y_{2\Delta_n} ..., X_n = Y_T$, $\Delta_n = \frac{1}{n}$, to be discrete observations of a solution $(Y_t)_{t\in[0,1]}$ of a stochastic partial differential equation of the form

$$dY_t = (\mathcal{A}Y_t + \alpha_t)\, dt + \sigma_t dW_t, \quad t \in [0, 1]. \tag{1.2}$$

Here, $\mathcal{A}$ (often a differential operator) is the generator of a strongly continuous semigroup $(\mathcal{S}(t))_{t\geq 0}$ on H and W is a cylindrical Wiener process potentially on another separable Hilbert space U. Moreover, $\alpha = (\alpha_t)_{t\in[0,1]}$ is an almost surely Bochner-integrable adapted stochastic process with values in H and $\sigma = (\sigma_t)_{t\in[0,1]}$ is a Hilbert-Schmidt operator-valued process that is stochastically integrable with respect to W (cf. for instance Chapter 2.5 in [81] for the definition of the stochastic integral in this context). A fairly weak and hence general concept of a solution to such an equation is the mild solution of the form

$$X_t = \mathcal{S}(t)Y_0 + \int_0^t \mathcal{S}(t-s)\alpha_s ds + \int_0^t \mathcal{S}(t-s)\sigma_s dW_s, \quad t \in [0,1], \tag{1.3}$$

(cf. one of the standard textbooks [43], [88] or [56]). In [42] processes of this form were coined *mild Itô processes*. Under the assumption of the absence of arbitrage opportunities (and disregarding the possibility of jumps) forward rates could also be modelled as a Volterra process of the form (1.3) and the corresponding stochastic partial differential equation is known as the *Heath-Jarrow-Morton-Musiela equation* (cf. [53] or [14]). For instance, in [53] the Hilbert space H is a space of absolutely continuous functions $f : \mathbb{R}_+ \to \mathbb{R}$ (forward curves), $\mathcal{A} = \frac{d}{dx}$ is just the derivative operator and the semigroup $(\mathcal{S}(t))_{t\in[0,T]}$ is the semigroup of left-shifts, such that $\mathcal{S}(t)f(x) = f(x+t)$ for $x, t \geq 0$ and $f \in H$. Furthermore, due to arbitrage arguments, under the risk-neutral probability measure, the drift is determined by the volatility σ (cf. [53, Lemma 4.3.3]). Hence, under the risk-neutral measure, the Hilbert-Schmidt operator-valued volatility process characterises the entire evolution of forward curves and is hence pivotal for further analysis of the term structure. This alone makes the volatility process more useful than the process $(C_t)_{t\in[0,T]}$ of covariance operators.

The volatility might, moreover, be advantageous to approach the elemental task of dimension reduction in this context. Let us assume that we have daily observations of the forward curves $f_{i\Delta_n}, i = 1, ..., n$, where time could be measured in years (i.e. $n = 365$). As already mentioned before, the usual procedure is then to take finitely many points on these curves and to conduct a principal component analysis based on this time series of multivariate data to derive the major modes of variation in this model. A straightforward generalisation of this would be to treat these curves as functions and conduct a functional principal component analysis just as described before. Apart from nonlinear tail dependencies within the points of the curves, there are further problems with this approach. First, at least some structural assumptions must be made about the underlying forward curve process, such as stationarity or independence after differencing (see [31, Section 1.7]). This would already be a massive limitation on the form of the volatility from a probabilistic point of view. Moreover, even if we assume that we have estimated the covariance C of f_1, say, and from that derived the first d eigenelements $e_1, ..., e_d$ and $\lambda_1, ..., \lambda_d$, the projection onto the subspace $\langle e_1, ..., e_d \rangle$ would most likely not yield a theoretically viable model in the sense that it still is the solution to an equation of the form (1.2). However, this is important for the absence of arbitrage in the context of forward markets. Besides additional regularity conditions on the coefficients of the model, we would necessarily have $e_1, ..., e_d \in D(\mathcal{A})$, which already excludes simple functions such as $e_i = \mathbb{I}_{[0,1]}$ in the case $\mathcal{A} = \frac{d}{dx}$, $H = L^2(\mathbb{R})$. In view of the investigation of the spatio-temporal regularity of the process $(Y_t)_{t\in[0,1]}$ it is important to notice that this forces $(Y_t)_{t\in[0,1]}$ to be a semimartingale, as it must necessarily be a strong solution to (1.2) (cf. [50, Theorem 1]). In the special case of forward curves, such restrictions become even more intricate and lie at the core of the work on consistency problems for the Heath-Jarrow-Morton-Musiela equation (cf. [53], [54], [24], [23], [52] [95]). So at least theoretically, basing the dimension reduction on the covariance operator yields the potential problem that the estimated eigenvectors exclude the existence of a corresponding consistent finite-dimensional model. However, we do not necessarily have to pursue the goal of reducing the dimension of the state space of the forward curves, but rather to make out the number of random drivers and their precise form. This is exactly the information that is encoded by the volatility term structure.

So in comparison to the role of the covariance in the basic i.i.d. setting of functional data analysis, in this framework the role of the central second-order object is taken by the volatility σ, or rather the integrated volatility

$$\int_0^t \Sigma_s ds := \int_0^t \sigma_s \sigma_s^* ds \quad t \in [0,1],$$

where σ_s^* is the adjoint of σ_s in H. In fact, this process can tell us exactly the number of random drivers that are effectively needed to describe the term structure evolution. The reduction of the dimensionality of the noise, by exchanging the volatility in (1.2) by a projected volatility $\sigma_s^d := \sum_{i=1}^d \langle \sigma_s(\cdot), e_i \rangle e_i$ for the leading eigenvectors $e_1, ..., e_d$ of the integrated volatility at $t = 1$, say, avoids any consistency problem, while still providing a reasonable approximation of the process. More importantly, we might obtain valuable insights into the term structure of volatility itself by conducting statistical inference on the integrated volatility. To this end, we will develop a general asymptotic theory for power variations in this infinite-dimensional context in the second part of this dissertation.

If $\mathcal{A} = 0$, the semigroup is constant and corresponds simply to the identity operator I_H on H. In that case, the theory of stochastic processes from finite dimensions suggests to base estimation of the integrated volatility on the infinite-dimensional realised quadratic covariation

$$RV_1^n := \sum_{i=1}^n (Y_{i\Delta_n} - Y_{(i-1)\Delta_n})^{\otimes 2}. \tag{1.4}$$

Indeed, in finite dimensions the theory of volatility estimation based on the realised quadratic covariation or related measures of volatility is rich and we provide a glimpse of it later in this chapter. Important contributions are among many others [70], [71], [7] [100], [10] or [73]. We also refer to the textbooks [72] and [1]. In infinite-dimensional settings like ours, the situation changes significantly. If the semigroup is not uniformly continuous, which is the case for the forward curve evolution, the realised covariation may fail to estimate the integrated volatility. The reason is that in that case, Y is in general not an H-valued semimartingale, which is a purely infinite-dimensional issue. Therefore, the sequence of increments $\Delta_i^n Y := Y_{i\Delta_n} - Y_{(i-1)\Delta_n}$ no longer forms an array of martingale differences. To overcome the lack of semimartingality, the research on volatility estimation for Volterra processes, such as [38], [9], [8], [39], [37] and [58, 87] and the growing literature on volatility estimation for pointwise evalutations of the solution to the second-order stochastic partial differential equation such as [21] or [34] use particular weighting schemes, which account for the lower regularity of the corresponding process. Such an approach does not seem feasible for analysing volatility in our infinite-dimensional operator-setting. The reason is that already with regard to the weak operator topology we would run into problems, as for some functionals $\langle \cdot, h\rangle$, the process $\langle Y_t, h\rangle$ is indeed a semimartingale and the associated (one-dimensional) realised variation $\sum_{i=1}^n \langle \Delta_i^n Y, h\rangle_H^2 = \langle RV^n h, h\rangle_H$ a consistent estimator for $\int_0^t \langle \Sigma_s h, h\rangle ds$. However, for other functionals g, $\langle RV^n g, g\rangle_H$ can diverge. Luckily, since we are dealing with functional data which are assumed to be smoothed curves, we can adjust these functional observations in a way which recovers the martingale difference property. Instead of the increments $\Delta_i^n Y$, we will analyse the variation induced by the *semigroup adjusted increments*

$$\tilde{\Delta}_i^n Y := Y_{i\Delta_n} - \mathcal{S}(\Delta_n) Y_{(i-1)\Delta_n}.$$

For the example of forward curves, this adjustment comes in quite handy, as the semigroup $\mathcal{S}$ is just the left-shift semigroup, such that $\mathcal{S}(\Delta_n)f(x) = f(x+\Delta_n)$. In the second part of this work, we therefore develop an asymptotic theory for the estimation of the integrated volatility via the *semigroup adjusted realised covariation* (SARCV)

$$SARCV_1^n := \sum_{i=1}^n \tilde{\Delta}_i^n Y^{\otimes 2}. \tag{1.5}$$

In fact, we develop a more general theory based on semigroup adjusted increments, since a crucial role in establishing a feasible asymptotic distribution theory for $SARCV$ is taken by *semigroup adjusted realised multipower variations* $(SAMPV)$ given by

$$SAMPV_1^n(m_1, ..., m_k) := \sum_{i=1}^n \bigotimes_{j=1}^k \tilde{\Delta}_{i+j-1}^n Y^{\otimes m_j}. \tag{1.6}$$

We refer to Paper IV for a technical definition of the general tensor-power notation.

We structure this dissertation into two parts. After this introduction, we recall the relevant definitions and results for copulas and power variations in finite dimensions and give an overview on the motivation and challenges for their generalisation to infinite dimensions. After that we provide a summary of the four articles I, II, III and IV, which are provided in the subsequent part of this dissertation. The first two of these articles (I and II) are about copula theory and the last two articles (III and IV) are concerning power variations.

1.1 Copulas

In this section we will introduce the notion of copulas in the classical finite-dimensional context and briefly discuss the motivation and challenges for its generalisation to infinite dimensions.

The theory of copulas is a widely used instrument in the toolbox of multivariate statistics. It allows to decompose the distribution of a multivariate random vector $X = (X_1, ..., X_d)$ with values in $\mathbb{R}^d$, $d \in \mathbb{N}$, into the distributions of its marginal one-dimensional components $X_1, ..., X_d$ and its dependence structure, in terms of a copula. Technically speaking, a copula is the cumulative distribution function C of a multivariate random vector $(U_1, ..., U_d)$ with uniform marginal distributions $U_1, ..., U_d \sim Unif(0, 1)$. That is, for all $u_j \in [0, 1]$ and $j = 1, ..., d$ we have

$$C_j(u_j) := C(1, ..., 1, u_j, 1, ..., 1) = u_j.$$

Equivalently, each copula C can be uniquely identified with a probability measure μ_C, which corresponds to the cumulative distribution function C. The capability of copulas to describe statistical dependence of a multivariate distribution is due to the fundamental Theorem of Sklar ([85, Theorem 2.10.9]):

Theorem 1.1.1 (Sklar's Theorem in finite dimensions)**.** *Let $d \in \mathbb{N}$ and $F : \mathbb{R}^d \to [0, 1]$ be a d-dimensional cumulative distribution function on $\mathbb{R}^d$ with marginal one-dimensional cumulative distribution functions F_j for each $j = 1, ..., d$. Then there exists a copula C on $\mathbb{R}^d$, such that for all $x_1, ..., x_d \in \mathbb{R}$ it is*

$$F(x_1, ..., x_d) = C\left(F_1(x_1), ..., F_d(x_d)\right). \tag{1.7}$$

If the marginals F_j are continuous for each $j = 1, ..., d$, C is unique. If in contrast C is a copula in $\mathbb{R}^d$ and F_j are one-dimensional cumulative distribution functions for each $j = 1, ..., d$, then F defined by (1.7) *is a cumulative distribution function on $\mathbb{R}^d$ with marginals F_j for each $j = 1, ..., d$.*

For a univariate cumulative distribution function F let $F^{[-1]}$ denote the quantile function given by

$$F^{[-1]}(u) := \inf\left\{x \in (-\infty, \infty) : F(x) \geq u\right\}.$$

If for a d-dimensional cumulative distribution function F the marginals $F_1, ..., F_d$ are continuous, then $F_i^{[-1]}(F(x)) = x$ for $x \in \mathbb{R}$ and hence the unique copula induced by Sklar's theorem is given by

$$C(u_1, ..., u_d) = F(F_1^{[-1]}(u_1), ..., F_d^{[-1]}(u_d)) \quad u_1, ..., u_d \in [0, 1]. \tag{1.8}$$

An important example of copulas, which are constructed by the inversion formula (1.8) are copulas underlying a multivariate Gaussian distribution $Y = (Y_1, ..., Y_d) \sim N(\mu, \Sigma)$, where μ and Σ are the mean and the covariance matrix of Y. As copulas are invariant under strictly increasing transformations (cf. [85, Theorem 2.4.3]), we can always set $\mu = 0$ and $\Sigma = P$, where P is the correlation matrix of Y. Gaussian copulas are widely used and distributions $X = (X_1, ..., X_d)$, which have an underlying Gaussian copula, were coined *nonparanormal distributions* in [80]. The appeal of these specific copulas is that one can profit from the simplicity and flexibility of the dependence structure of multivariate Gaussian distributions while still having the possibility to capture features such as heavy tails via the marginals. This is also promising if we have in mind generalisations to functional data analysis. In fact, in order to estimate the Gaussian copula underlying a nonparanormal distribution, we just have to derive its latent correlation matrix P, which is suitable for the generalisation to the level of covariance operators in Hilbert spaces.

Let us describe how one can derive P in the case that $d = 2$. This can be done in a very sound way, namely by estimating Kendall's tau correlation coefficient τ. Let $X = (X_1, X_d)$ be a nonparanormal distribution with underlying Gaussian copula C, induced by a bivariate Gaussian random vector (Y_1, Y_2) with correlation matrix P. The copula C has exactly one parameter $P_{1,2}$ that has to be derived from observations of X. Although C is latent, this can be done in a theoretically sound way by appealing to the concept of rank correlation, in particular, Kendall's tau. For two real-valued random variables Z_1, Z_2 Kendalls tau rank correlation coefficient is defined as

$$\tau_{Z_1,Z_2} := \mathbb{E}\left[sign(Z_1 - \tilde{Z}_1)sign(Z_2 - \tilde{Z}_2)\right],$$

where $(\tilde{Z}_1, \tilde{Z}_2)$ is an independent copy of (Z_1, Z_2). If we now assume that we have $2n$ independent copies $X^1, ..., X^{2n}$ of the nonparanormal random variable X, τ_{X_1,X_2} can be consistently estimated by its empirical counterpart

$$\hat{\tau}_{X_1,X_2} = \frac{2}{n}\sum_{k=1}^{n} sign(X_1^{2k} - X_1^{2k-1})sign(X_2^{2k} - X_2^{2k-1}).$$

One can then exploit the relation

$$P_{1,2} = \sin(\frac{\pi}{2}\tau_{X_1,X_2}). \tag{1.9}$$

(cf. [48, Theorem 3.1]) and estimate $P_{1,2}$ consistently by the corresponding plug-in estimator.

Like Kendall's tau, many important statistical concepts that are used to describe the distribution of a multivariate distribution are exclusive properties of the underlying copula of a multivariate distribution. One example are the coefficients $\lambda_u^{X_1,X_2}$ and $\lambda_l^{X_1,X_2}$ of upper and lower tail dependence of a bivariate random variable (X_1, X_2) defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and having continuous cumulative marginal distribution functions F_{X_1} and F_{X_2}. These coefficients are defined by

$$\lambda_u^{X_1,X_2} := \lim_{q\uparrow 1} \mathbb{P}\left[X_1 > F_{X_1}^{[-1]}(q)|X_2 > F_{X_2}^{[-1]}(q)\right]$$

and

$$\lambda_l^{X_1,X_2} := \lim_{q\downarrow 0} \mathbb{P}\left[X_1 \leq F_{X_1}^{[-1]}(q)|X_2 \leq F_{X_2}^{[-1]}(q)\right]$$

provided that the limits exist in $[0, 1]$. If $\lambda_u^{X_1,X_2} > 0$ one says that X_1 and X_2 are upper tail dependent and if $\lambda_l^{X_1,X_2} > 0$ one says that X_1 and X_2 are lower tail dependent. The upper tail dependence is the limit for $q \to \infty$ of the probability that X_1 is in the q-Quantile of the distribution of X_1 conditional on the event that X_2 is in the q-quantile of its distribution. If this limit is positive, extremely large values of X_1 (in size of its distribution) are more likely to entail extremely large values of X_2 (in size of its distribution) and vice versa. The analogous reasoning holds for lower tail dependence. In terms of the (unique) underlying copula C of X_1 and X_2 we have

$$\lambda_u^{X_1,X_2} := \lim_{q\uparrow 1} \frac{1-C(q,q)}{1-q}$$

and

$$\lambda_l^{X_1,X_2} := \lim_{q\downarrow 0} \frac{C(q,q)}{q}$$

(cf. [85, Theorem 5.4.2]). Tail dependence can be found empirically in many financial contexts. This led to criticism of the usage of Gaussian copulas, since they necessarily induce a tail coefficient of 0, i.e. tail independence. A very flexible class of copulas, which is able to account for tail dependence and at the same time does not lose the appealing feature that multivariate dependence patterns are largely captured by a covariance matrix are elliptical copulas. A popular example of this class are t-copulas, i.e. copulas underlying a multivariate t-distribution. Recall that a multivariate centred t-distribution with $\nu > 0$ degrees of freedom and strictly positive definite scatter matrix Σ is given by the density

$$f_{t_{\Sigma,\nu}}(x_1,...,x_d) = \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\pi\nu)^d det(\Sigma)}} \left(1 + \frac{(x_1,...,x_d)\Sigma^{-1}(x_1,...,x_d)'}{\nu}\right)^{-\frac{\nu+d}{2}}.$$

Equivalently, one can introduce this as a multivariate normal variance mixture, i.e. a random vector $Y = (Y_1,...,Y_d)$ is t-distributed with parameters ν and Σ, if there is a standard multivariate normal distribution $Z = (Z_1,...,Z_d)$ such that

$$Y = SZ$$

where S is a univariate positive random variable, independent of Z, such that $\frac{\nu}{S^2} \sim \chi^2_\nu$ is chi-squared distributed with $\nu > 0$ degrees of freedom. Let us again consider the bivariate case $X = (X_1, X_2)$, where X_1 and X_2 have continuous cumulative distribution functions and an underlying t-copula with ν degrees of freedom and scatter matrix Σ. Again, due to invariance under monotone transformations, we can assume $\Sigma_{1,1} = \Sigma_{2,2} = 1$ and the only parameters left to derive are $\Sigma_{i,j}$ and $\nu > 0$. Luckily, relation 1.9 still holds for the coefficient $\Sigma_{1,2}$ (instead of $P_{1,2}$) and yields a straightforward way to estimate the scatter matrix. Moreover, it is well known that the multivariate t-distribution yields tail dependence between the coefficients. Precisely, we have

$$\lambda^{X_1,X_2} := \lambda_u^{X_1,X_2} = \lambda_l^{X_1,X_2} = 2F_{t_{\nu+1}}\left(-\sqrt{\nu+1}\frac{1-\Sigma_{1,2}}{1+\Sigma_{1,2}}\right)$$

where F_{t_ν} is the distribution function of the univariate t-distribution with ν degrees of freedom (cf. [46, Proposition 1]). So X_1 and X_2 can have arbitrarily low or high tail dependence, as λ^{X_1,X_2} approach 0 for $\nu \to \infty$ and becomes 1 for $\Sigma_{1,2} = 1$. The parameter ν might be estimated by a maximum likelihood procedure, while holding Σ, which could be derived by Kendall's tau, fixed (see [46, Section 4]).

Copulas for Functional Data: Motivation and Challenges

Copulas in infinite dimensions appeared in various contexts in the literature. Most notably, they are used to model the temporal dependence of Markov processes in [44], [79], [69], [68], [20] and [33]. In fact, copulas yield an elegant alternative characterisation of the Markov property (cf. [85, Theorem 6.4.3]). Infinite-dimensional Gaussian copulas as copulas of stochastic processes appeared in a Machine-Learning context in [97]. In stark contrast to the setting of [63] and the framework in which we operate in this dissertation, these earlier works do not directly link copulas to a topological structure of the space, in which the corresponding stochastic processes takes its values. In that regard, they are not tailor-made for usage in functional data analysis and leave the natural concept of copulas still absent from its toolbox. The attempt to change this goes along with some mathematical challenges but might open up a few valuable statistical methods as well.

Let us subsume the main motivational points for establishing the theory.

- **Nonlinear dependence patterns**: As remarked in the introduction, the classical functional principal component analysis takes into account linear dependence patterns and largely ignores more subtle aspects such as tail dependence. Copulas are capable of describing the dependence structure of multivariate random variables regardless of the existence of moments and various copula models, such as the t-copula, can model tail-dependence very targeted.

- **Control over local distributional properties**: The functional principal component analysis enables us to approximate random curves just by a few univariate random sources. However, one arguably loses control over local properties of the law of the random curves, such as the distribution of a stochastic process evaluated at a certain point. As Sklar's theorem allows us to model marginals and dependence structure separately, copulas naturally overcome this issue and allow us to include statistical information such as heavy tails freely along the random curve.

- **Probabilistic insights**: The question, if a cylindrical random variable is an actual random variable in a function space is fundamental in probability theory. Approaching this from the point of view of copula theory makes it possible to find answers to this question by digesting the interplay of marginals and the dependence structure of stochastic processes.

Let us repeat as well the major challenges that one faces regarding copula theory in function spaces:

- **Notion of marginals**: Yet, it is unclear what a proper notion of marginals should be. For instance in the Banach space $C[0,1]$ of continuous functions $f : [0,1] \to \mathbb{R}$, it would be an immediate generalisation of the framework in [63] to define marginals as the distributions of basis coefficients for some Schauder basis (see Paper I for this concept). On the other hand, an equally appropriate choice could be to treat all the evaluations $X(t)$ for $t \in [0,1]$ as marginals of a random curve X with values in $C[0,1]$. We need to provide a flexible notion of marginals that unifies all concepts of infinite-dimensional copulas and is adaptable to any practical situation.

- **Construction problem**: In general, the copula construction induced by the second part of Sklar's theorem results in a cylindrical probability measure in function spaces. For example, we show later that the path copula corresponding to a Brownian motion cannot induce a probability measure on the Banach space $C[0,1]$. Constructing random curves in the spirit of Sklar's theorem in such situations requires therefore practical criteria to decide whether a respective construction induces a probability measure in the function space we targeted.

- **Estimation and Approximation**: The appeal of functional data analysis stems partially from the fact, that we can make inferences on the law of random curves by pooling information from neighbouring points. Hence, in addition to the probabilistic approximation of copulas and marginals of a random curve, also analytic approximations have to be taken into account. For instance, we later want to generalise the nonparanormal distributions from multivariate statistics to "nonparanormal processes", i.e. we like to approximate a random variable with values in $L^1[0,1]$, say, and assume an underlying Gaussian copula (i.e. the copula corresponding to a square-integrable Gaussian process). We might then be able to approximate the marginals of such a nonparanormal process and also the covariance operator of the latent Gaussian process. Is there a metric that allows us to measure the convergence of this approximation procedure by measuring the convergence of the marginals and the latent covariance separately? In fact, copulas share a natural link to the so-called Wasserstein distance between probability measures (cf. for instance [2]). We will spend some time in Paper I to exploit this connection in order to obtain consistency and approximation results for copula constructions.

1.2 Power Variations

Throughout this section, let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ be a filtered probability space, on which the stochastic processes in this section are defined, if not stated differently. Fix $T > 0$. We assume to have observations $Y_{i\Delta_n}$, $i = 0, 1, ..., \lfloor T/n \rfloor$, $\Delta_n = \frac{1}{n}$, of a

d-dimensional continuous Itô-semimartingale

$$Y_t = Y_0 + \int_0^t \alpha_s ds + \int_0^t \sigma_s dW_s,$$

where $\alpha = (\alpha_t)_{t\in[0,T]}$ is an adapted $\mathbb{R}^d$ valued, almost surely integrable stochastic process, $W = (W_t)_{t\in[0,T]}$ is an $\mathbb{R}^{d'}$-valued standard Brownian motion for some $d' \in \mathbb{N}$ and $\sigma = (\sigma_t)_{t\in[0,T]}$ is a predictable $\mathbb{R}^{d\times d'}$-valued, almost surely square-integrable stochastic process.

The goal is then to make inferences on the integrated volatility

$$\int_0^t \Sigma_s ds := \int_0^t \sigma_s \sigma_s^\top ds \quad t \in [0,T],$$

where $\sigma^\top$ is the tranposed of the matrix σ_s. Interest in inferential methods for integrated volatility comes primarily from financial econometrics. One reason is, for instance, that if we assume α and σ to be independent of the driving Wiener process, we find that the process $(Y_t)_{t\in[0,T]}$ is conditionally Gaussian and the integrated volatility corresponds to the conditional covariance of the increments, i.e.

$$Y_{t_2} - Y_{t_1} | \alpha, \sigma \sim N(\int_{t_1}^{t_2} \alpha_s ds, \int_{t_1}^{t_2} \sigma_s \sigma_s^\top ds).$$

This fact was exploited, for instance in [3] in order to integrate high-frequency data into the process of volatility forecasting.

Due to its convenient form and since it is consistent with the notation in infinite-dimensional Hilbert spaces that are investigated in the papers III and IV, we will use for two vectors $x = (x_1, ..., x_d)^\top$ and $y = (y_1, ..., y_d)^\top$ the tensor notation

$$x \otimes y := xy^\top = (x_i y_j)_{i,j=1,\dots,d} \in \mathbb{R}^{d\times d}.$$

This can also be identified with the linear operator $x \otimes y(v) = \langle x, v\rangle y$. If $x = y$ we write $x^{\otimes 2} := x \otimes x$. The tensor notation easily carries over to more general Hilbert spaces, H and G (for instance the space of $d \times d$ matrices) for which the tensor product is defined in the same way as a linear operator from H into G by

$$h \otimes g := \langle h, \cdot\rangle_H g. \tag{1.10}$$

In that way, it is possible to define for instance the tensor $(x \otimes y) \otimes (v \otimes w)$, which can be identified as a linear operator from $\mathbb{R}^{d\times d}$ into itself or as an element in $\mathbb{R}^{d\times d\times d\times d}$.

The general theory of stochastic calculus tells us that the sum of squared returns or the realised variation

$$RV_t^n := \sum_{i=1}^{\lfloor t/n \rfloor} (Y_{i\Delta_n} - Y_{(i-1)\Delta_n})^{\otimes 2} \tag{1.11}$$

converges to $\int_0^t \Sigma_s ds$ uniformly on compacts in probability (u.c.p.), that is, for all $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}[\sup_{t\in[0,T]} \|RV_t^n - \int_0^t \Sigma_s ds\|_d > \epsilon] = 0.$$

This convergence can be considered to take place in the Skorohod space $D([0,T],\mathbb{R}^{d\times d})$ of right-continuous functions with left limits (càdlàg) equipped with the sup-norm. This norm makes $D([0,T],\mathbb{R}^{d\times d})$ a Banach space, but it is not separable. We therefore recall, that we can equip $D([0,T],E)$, where for the moment E is an abstract Polish (i.e. seperable and completely metrisable) space, with another metric d, that makes it separable and complete. The precise definition of this metric is slightly technical and not important in this dissertation. We will henceforth call the topology induced by this metric the Skorohod topology. It is however important to record that the Polish structure of the Skorohod space under this metric is paid with the price of some unusual properties. For example, for two sequences of processes $X_n, Y_n \in \mathcal{D}([0,T],E)$ such that $d(X_n,X) \to 0$ and $d(Y_n,Y) \to 0$ for some $X,Y \in \mathcal{D}([0,T],E)$ we do not necessarily have $d(X_n+Y_n, X+Y) \to 0$. Moreover, point evaluations are not continuous, that is, we can have $d(X_n,X) \to 0$, but not $X_n(t) \to X(t)$ for some $t \in [0,T]$. We refer to [22] for a detailed account on the Skorohod space.

To introduce an asymptotic distribution theory for the estimation of the integrated volatility, the concept of stable convergence in law is important. A sequence of random variables $(X_n)_{n\in\mathbb{N}}$ defined on $(\Omega,\mathcal{F},\mathbb{P})$ and with values in a Polish space E converges stably in law to a random variable X defined on an extension $(\tilde{\Omega},\tilde{\mathcal{F}},\tilde{\mathbb{P}})$ of $(\Omega,\mathcal{F},\mathbb{P})$ with values in E, if for all bounded continuous $f: E \to \mathbb{R}$ and all bounded random variables Y on $(\Omega,\mathcal{F})$ we have $\mathbb{E}[Yf(X_n)] \to \tilde{\mathbb{E}}[Yf(X)]$ as $n \to \infty$, where $\tilde{\mathbb{E}}$ denotes the expectation with respect to $\tilde{\mathbb{P}}$.

Under the condition that almost surely

$$\int_0^T \|\alpha_s\|^2 + \|\sigma_s\|^4 ds < \infty,$$

one also has a functional stable central limit theorem (cf. [72, Theorem 5.4.2]). That is, the the convergence

$$\left(\Delta_n^{-\frac{1}{2}} \left(RV_t^n - \int_0^t \Sigma_s ds\right)\right)_{t\in[0,T]} \to (N(0,\Gamma_t))_{t\in[0,T]}$$

holds stably in law with respect to the Skorohod topology. Here $(N(0,\Gamma_t))_{t\in[0,T]}$ is a continuous mixed Gaussian process defined on an extension $(\tilde{\Omega},\tilde{\mathcal{F}},\tilde{\mathcal{F}}_t,\tilde{\mathbb{P}})$ of $(\Omega,\mathcal{F},\mathcal{F}_t,\mathbb{P})$ and with values in $\mathbb{R}^{d\times d}$ and with random covariance Γ_t, which can be precisely described and is determined by the volatility σ. We refer to [6] for a precise formula of this covariance (or to the more general case in paper IV, which makes use of the tensor notation).

In comparison to the common convergence in distribution, the stronger stable convergence of processes is more pleasant in this setting, in view of the unknown covariance Γ_t, which must be estimated from the data as well, since otherwise the central limit theorem is not feasible. Among other things, stable convergence guarantees that if we have a consistent estimator $\hat{\Gamma}_t^n$ such that $\hat{\Gamma}_t^n \to \Gamma_t$ u.c.p., we obtain the joint convergence

$$\left(\Delta_n^{-\frac{1}{2}}(RV_t^n - \int_0^t \Sigma_s ds), \hat{\Gamma}_t^n\right)_{t\in[0,T]} \to \left(\mathcal{N}(0,\Gamma_t),\Gamma_t), \hat{\Gamma}_t^n\right)_{t\in[0,T]},$$

stably as a process, leading to feasible central limit theorems. In matrix notation, for $i, j \in \{1, ..., d\}$ and all $t \in [0, T]$, this means that in distribution and conditional on $\{(\Gamma_t)_{i,j} \neq 0\} \in \mathcal{F}$,

$$\frac{\Delta_n^{-\frac{1}{2}}\left((RV_t^n)_{i,j} - \int_0^t (\Sigma_s)_{i,j} ds\right)}{(\hat{\Gamma}_t^n)_{i,j}} \to \mathcal{N}(0,1),$$

which allows one to form confidence intervals and conduct hypothesis tests.

In fact, the unknown asymptotic variance Γ_t can be conveniently estimated from the data by appealing to multipower variations. For $m_1, ..., m_k, k \in \mathbb{N}$ we define the multipower variation $MPV(m_1, ..., m_k)$ as

$$\begin{aligned} &MPV_t^n(m_1, ..., m_k) \\ := &\sum_{i=1}^{\lfloor t/n \rfloor - k+1} (Y_{i\Delta_n} - Y_{(i-1)\Delta_n})^{\otimes m_1} \otimes ... \otimes (Y_{(i+k-1)\Delta_n} - Y_{(i+k-2)\Delta_n})^{\otimes m_k} \end{aligned}$$

If $k = 1$, these are simply called power variations. We give the following law of large numbers for multipower variations, which are special cases of Theorems 3.4.1 and 8.4.1 in [72].

Theorem 1.2.1. *For a postive semidefinite symmetric $d \times d$-matrix Σ, we define the matrix $\rho_\Sigma(m)$, as the m'th tensor moment of an $\mathbb{R}^d$-valued random variable $U \sim \mathcal{N}(0, \Sigma)$, i.e., $\rho_\Sigma(m) = \mathbb{E}[U^{\otimes m}]$.*

(i) *(power variations) Let $m \geq 2$ be a natural number. If*

$$\mathbb{P}\left[\int_0^T \|\alpha_s\|^{\frac{2m}{2+m}} ds + \int_0^T \|\sigma_s\|^m ds < \infty\right] = 1, \tag{1.12}$$

the following law of large numbers holds:

$$\Delta_n^{1-\frac{m}{2}} MPV^n(m) \xrightarrow{u.c.p.} \left(\int_0^t \rho_{\Sigma_s}(m) ds\right)_{t\in[0,T]}.$$

(ii) *(multipower variations) Let α be locally bounded, σ is a cádlág process and $m, m_1, m_2, \ldots, m_k$ be natural numbers such that $m_1 + ... + m_k = m$. Then*

$$\Delta_n^{1-\frac{m}{2}} MPV_t^n(m_1, ..., m_k) \xrightarrow{u.c.p.} \int_0^t \bigotimes_{j=1}^k \rho_{\Sigma_s}(m_j) ds, \tag{1.13}$$

Note that $\rho_\Sigma(m) = 0$, if m is odd.

By this law of large numbers, one can argue that a consistent estimator for the asymptotic variance Γ_t is given by

$$\hat{\Gamma}_t^n = MPV_t^n(4) - MPV_t^n(2,2)$$

(see for instance [6] or paper IV for a precise derivation of this).

We might also want to have confidence in this estimator, or integrated functions of the volatility estimated via this law of large numbers, which is why we also recall the corresponding central limit theorem, which is an immediate implication of the central limit theorem 11.2.1 in [72].

Theorem 1.2.2. *Let $m = m_1 + ... + m_k$ and m_i even. For a positive semidefinite symmetric $\Sigma \in R^{d\times d}$, we define a sequence $(U_i)_{i\in\mathbb{N}}$ of independent random variables $U_i \sim \mathcal{N}(0,\Sigma)$ and the σ-fields $\mathcal{G} = \sigma(U_1, ..., U_{k-1})$ and $\mathcal{G}' = \sigma(U_1, ..., U_k)$. We define*

$$\Gamma_t(m_1, ..., m_k) := \int_0^t R_{\Sigma_s}(m_1, ..., m_k) ds,$$

where

$$\begin{aligned} &R_\Sigma(m_1, ..., m_k) \\ &= \mathbb{E}\left[\left(\sum_{j=0}^{k-1} \mathbb{E}\left[\bigotimes_{i=1}^k U_{k-j+(i-1)}^{\otimes m_i} | \mathcal{G}'\right] - \mathbb{E}\left[\bigotimes_{i=1}^k U_{k-j+(i-1)}^{\otimes m_i} | \mathcal{G}\right]\right)^{\otimes 2}\right] \\ &= \mathbb{E}\left[\left(\sum_{j=0}^{k-1} \mathbb{E}\left[\bigotimes_{i=1}^k U_{k-j+(i-1)}^{\otimes m_i} | \mathcal{G}'\right]\right)^{\otimes 2} - \left(\sum_{j=0}^{k-1} \mathbb{E}\left[\bigotimes_{i=1}^k U_{k-j+(i-1)}^{\otimes m_i} | \mathcal{G}\right]\right)^{\otimes 2}\right] \end{aligned}$$

Let α be locally bounded and and σ itself an continuous Itô-process on $\mathbb{R}^{d\times d}$ of the form

$$\sigma_t = \sigma_0 + \int_0^t \tilde{\alpha}_s ds + \int_0^t \tilde{\sigma}_s d\tilde{W}_s,$$

where $\tilde{\alpha}$ is progressively measurable and $\tilde{\sigma}$ is cádlág and stochastically integrable with respect to the Wiener process $\tilde{W}$. Then the convergence

$$\begin{aligned} &\Delta_n^{\frac{1-m}{2}} \left(MPV_t^n(m_1, ..., m_k) - \int_0^t \bigotimes_{j=1}^k \rho_{\Sigma_s}(m_j) ds \right)_{t\in[0,T]} \\ &\to \left(\mathcal{N}\left((0, \Gamma_t(m_1, ..., m_k)\right)\right)_{t\in[0,T]}. \end{aligned}$$

holds stably in law with respect to the Skorohod topology as $n \to \infty$, where the limiting process on the right is conditionally on $\mathcal{F}$ a continuous centered Gaussian process with independent increments defined on an extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathcal{F}}_t, \tilde{\mathbb{P}})$ of $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$.

Power Variations for Functional Data: Motivation and Challenges

Although a lot of data and processes, ranging from term structure models in mathematical finance to physical applications, stem from infinite-dimensional evolution equations, the theory on infill-asymptotics in the sense of the asymptotics for power variations as discussed above seems to be sparse and often specific. In general, “statistical inference for SPDEs is in its developing stage with many fundamental problems still open” as witnessed in the survey paper [35, p.1]. Most articles deal with second-order parabolic SPDEs and exploit the spectral decompositions of the corresponding differential operator. Some works such as [21] and [34] manage to derive an asymptotic theory, which is fully applicable on the basis of observations of a spatio-temporal process at a fixed finite number of spatial points. Our goals are somewhat complementary to these works in the sense that we want to make inferences on smooth features underlying the spatio-temporal data and therefore necessarily assume to be in the luckier situation of (at least locally) densely observed curves. We list our major aims and motivational points here to put our developments in perspective:

- **Dimension Reduction:** As already argued before, the integrated volatility is an adequate object from which to reveal the effective number of random drivers needed to describe the stochastic evolution of stochastic partial differential equations. Extracting the eigenelements of the integrated volatility can yield valuable insights into the term structure of volatility itself, which is key for pricing bonds or energy derivatives or capturing the risk induced from the variation of interest rates.

- **Robustness:** The estimation of volatility does not impose any moment assumption on the process or the volatility method. This is very appealing in view of the active research in robust estimation methods for functional data analysis.

- **Shape information:** Our method naturally takes into account information about slope (or curvature if well defined) indicated by the data. One can see this intuitively, as the semigroup adjusted increments are dependent on the differential operator governing the equation. Taking shape information into account when modelling forward curve evolutions was proposed, for example in [36], where a second-order stochastic partial differential equation for the evolution of forward interest rates was suggested. Unlike the setting of [36], we explicitly want to include the no-arbitrage framework of the Heath-Jarrow-Morton-Musiela equation into our statistical considerations. Functional data analysis is the perfect statistical tool to develop such methods.

(iv) **Hilbertian Volatility Models:** Our method connects high-frequency estimation of volatility to the recent literature on infinite-dimensional volatility models. In these models, the volatility is itself a mild Itô process in a Hilbert space and they are originally designed to describe nontrivial patterns in the time-varying volatility term structure, mostly in connection to the energy market (see [15], [19],[18], [40], [41]). So far, there is no approach to estimate the parameters (which are in general operators) of such models. Our contribution in this work has the potential to make such models feasible and hence pave their way to applicability.

There are some characteristic differences between the finite-dimensional theory and the theory established for high-frequency estimation of stochastic partial differential equations so far:

- **No Semimartingality:** If Y is a mild Itô process of the form (1.3), it is in general not a semimartingale, and the standard quadratic variation estimator could diverge. This is essentially dependent on the regularity of the semigroup on the range of the volatility and leads us to the task of developing an asymptotic theory for the semigroup adjusted realised covariation as defined earlier in (1.5). This turns out to be a consistent estimator of the integrated volatility and can, thus, be seen as a legitimate generalisation of the quadratic variation from finite dimensions. However, its convergence rate hinges again on the interplay of the semigroup and the volatility, and the asymptotic theory that we build up here has to take into account cases of potentially low regularity.

- **Asymptotic Behaviour of Realised Covariation:** Since we want to establish a general theory for estimation of the integrated volatility operator in an operator norm, we cannot hope to find a weighting scheme as in [21] or [34] for the realised variation, which is independent of the volatility. The problem is that the realised variation might diverge already in the weak operator topology, while it does converge when it is projected onto very regular functionals. Nevertheless, the convergence of the realised covariation, if it takes place, is valuable, since we do not have to rely on (locally) dense data in space. We should therefore also consider cases, in which the quadratic variation can be used as an estimator for the integrated volatility.

- **Discrete Approximations** Full observations of curves are in most cases not realistic to hope for. This is arguably an intricate problem, since we are forced to apply the semigroup on partially observed curves in order to reconstruct the semigroup adjusted increments. We should therefore discuss on how to estimate infinite-dimensional operators in fully discrete frameworks. Luckily, this can be done quite handy for forward rates and the Heath-Jarrow-Morton-Musiela equation, as the operators belonging to the semigroup are just left-shifts.

1.3 Overview of Articles

This section provides an overview of the subsequent articles.

1.3.1 Copula Measures and Sklar's Theorem in Arbitrary Dimensions

This work corresponds to the research article [12], in which we develop a unified framework for copulas in infinite dimensions. In the most general sense, for an arbitrary index set I, we define copulas on the product space $\mathbb{R}^I$ as a probability measure C (a copula measure) on $\otimes_{i\in I}\mathcal{B}(\mathbb{R})$, such that its marginals $C_i = C\circ\pi_i^{-1}$, where $\pi_i(f) = f(i)$ is the canonical projection onto the ith component of elements $f \in \mathbb{R}^I$, are uniformly distributed on $[0,1]$ for each $i \in I$. In the product space setting, we can prove Sklar's theorem in the same generality as it can be done in finite dimensions.

We will introduce the general notion of marginals in a topological vector space V over $\mathbb{R}$ equipped with a σ-algebra $\mathcal{V}$ as the pushforward measures $\mu_m = \mu \circ m^{-1}$ for elements m of a measurable and linearly independent subset M of the algebraic dual $Hom(V,\mathbb{R})$, which separates the points of V. We can then define a copula corresponding to a probability measure μ on V as a copula on the product space $\mathbb{R}^M$ via the embedding

$$V \ni v \mapsto (m\,(v))_{m\in M} \in \mathbb{R}^M. \tag{1.14}$$

In that way, Sklar's theorem for product spaces always guarantees that there is a (not necessarily unique) copula underlying a probability distribution on V. However, as the embedding (1.14) is almost never a surjection from V into $\mathbb{R}^M$, we cannot guarantee that any choice for a copula C on $\mathbb{R}^M$ and marginals $(\mu_m)_{m\in M}$ correspond to a probability measure on V. This is the construction problem. In the case that $V = H$ is a reproducing kernel Hilbert space of functions $f : [0,1] \to \mathbb{R}$, say, such that the evaluation functionals $\delta_x f = f(x)$ for $x \in [0,1]$ are continuous, both the framework of [63] such that $M = \{\langle e_n, \cdot\rangle : n \in \mathbb{N}\}$ for an orthonormal basis $(e_n)_{n\in\mathbb{N}}$ of H as well as the choice $M = \{\delta_x : x \in [0,1]\}$ can be reasonable.

We will provide practical solutions to the construction problem for a few function spaces (and marginal-specifications), such as the Lebesgue spaces $L^p([0,T])$, the space of continuous functions $C[0,T]$ or sequence spaces l^p. Afterwards we elaborate approximation methods for copula models, mainly designed for the spaces $L^p([0,T])$ and l^p. This is done by measuring the distance of probability measures in the Wassertstein metric

$$d(\nu^1,\nu^2) =: \mathbb{W}_p(\nu^1,\nu^2) := \inf_{\rho <_{\nu^1}^{\nu^2}} \left(\int_{E\times E} \|x-y\|_E^p \rho(dxdy)\right)^{\frac{1}{p}},$$

where $\rho <_{\nu^1}^{\nu^2}$ indicates that ρ is a coupling of ν_1 and ν_2, i.e. ρ is a probability measure on $E \times E$ that has marginal distributions ν^1 and ν^2. We will outline a few useful connections of copulas and the Wasserstein distance, and then exploit these to derive a robustness inequality. We will use this inequality to derive practical approximation and estimation results for some copula models, such as elliptical copulas and a generalisation of the nonparanormal distribution to the functional data framework.

1.3.2 A Topological Proof of Sklar's Theorem in Arbitrary Dimensions

This short work corresponds to the research article [11] and shows the compactness of copula measures as measures in product spaces and outlines some implications of this fortunate circumstance. Among other things, we give an alternative way of proving Sklar's theorem on product spaces $\mathbb{R}^I$ with arbitrary index set I by generalising the proof for Sklar's theorem in finite dimensions from [47]. The basic steps of this are as follows: First, we show the set of copula measures is compact with respect to the topology of convergence of the finite-dimensional distributions. Then we prove the second part of Sklar's theorem (that every copula measure can be merged with any family of marginals to a probability measure). After this we show that the operation of merging a copula measure with marginals is a continuous mapping and use the compactness of the set of copulas to conclude that this map has closed image. The second part of Sklar's theorem follows by arguing that this image is also dense in the space of probability measures.

1.3.3 A Weak Law of Large Numbers for Realised Covariation in a Hilbert Space Setting

This work corresponds to the research article [17] and introduces a new estimator for the integrated volatility operator in the setting of mild Itô processes of the form (1.3). Namely, we prove uniform convergence in probability of the semigroup adjusted realised covariation SARCV, defined in (1.5), to the integrated volatility. Under additional regularity conditions, we will describe the exact speed of the convergence of this estimator, which turns out to be different to the $\mathcal{O}(\sqrt{\Delta_n})$-rate that is guaranteed in finite dimensions under mild conditions. Here $\Delta_n = 1/n$ is the distance between the data points in time. Rather, the rate is determined by the continuity property of the semigroup $(\mathcal{S}(t))_{t\geq 0}$ on the range of the volatility, and becomes $\mathcal{O}(\sqrt{\Delta_n} + b_n(T))$ under the assumption $\mathbb{E}\left[\sup_{s\in[0,T]} \|\sigma_s\|^4_{L_{\text{HS}}(U,H)}\right] < \infty$, where

$$b_n(T) := \left(\int_0^T \sup_{x\in[0,T]} \mathbb{E}\left[\|(I-\mathcal{S}(x))\sigma_s\|^2_{\text{op}}\right] ds\right)^{\frac{1}{2}}.$$

It is important to note, that in this article, we assumed the driving Wiener process to be a Q-Wiener process. At least in distribution there is equivalence between these two concepts, due to the martingale representation theorems (cf. [56, Section 2.2.5]). Thus, the results in this article also hold for the framework we introduced in the introduction.

We will elaborate on the magnitude of $b_n(T)$ for various examples, such as uniformly continuous semigroups and the semigroup of left-shifts in the context of forward curves. We also consider applications to some stochastic volatility models in Hilbert spaces.

1.3.4 A Feasible Central Limit Theorem for Realised Covariation of SPDEs in the Context of Functional Data

This work corresponds to the research article [16] and comprises the generalisation of the asymptotic theory in Section 1.2 for mild Itô processes of the form (1.3).

The most important contribution in this work is the proof of a feasible central limit theorem for the SARCV as introduced in (1.5). To this end, we will first establish the central limit theorem

$$\left(\Delta_n^{-\frac{1}{2}}\left(SARCV_t^n - \int_0^t \Sigma_s ds\right)\right)_{t\in[0,1]} \stackrel{\mathcal{L}-s}{\Longrightarrow} (\mathcal{N}(0,\Gamma_t))_{t\in[0,1]}\,, \tag{1.15}$$

where $\stackrel{\mathcal{L}-s}{\Longrightarrow}$ stands for the stable convergence in law as a process in the Skorokhod space $\mathcal{D}([0,T],\mathcal{H})$. $(\mathcal{N}(0,\Gamma_t))_{t\in[0,1]}$ is a multivariate continuous mixed Gaussian process defined on an extension of the initial probability space and with values in $\mathcal{H}$, the space of Hilbert-Schmidt operators on H, and with a covariance operator Γ_t, called the *asymptotic variance*.

To this end, it will be important to reconsider the findings in [17], which suggest that the rate of convergence of the $SARCV$ can in general be slower than the $\mathcal{O}(\sqrt{\Delta_n})$ speed that we obtain for semimartingales and the realised covariation in finite dimensions. We formulate sharp regularity conditions on the continuity of the semigroup on the range of the volatility, which account for this problem and allow us to prove (1.15). Putting much stronger regularity conditions on the semigroup and the volatility in this regard, we also show in which situations the standard realised covariation (1.4) yields a consistent and asymptotically normal estimator.

The asymptotic variance Γ_t is unknown, which makes the central limit theorem (1.15) infeasible for applications. To obtain a feasible central limit theorem, we introduce an estimator for the asymptotic variance based on the SAMPV as defined in (1.6). More precisely, we derive the following limit theorem, which holds uniformly on compacts in probability:

$$\left(\Delta_n^{-1}\left(SAMPV_t^n(4) - SAMPV_t^n(2,2)\right)\right)_{t\in[0,1]} \to (\Gamma_t)_{t\in[0,1]} \quad \text{as } n\to\infty. \tag{1.16}$$

Rather than just proving (1.16), we provide general laws of large numbers and central limit theorems for the SAMPV.

We will also outline how our results can be applied in fully discrete settings if the semigroup is either the identity or the semigroup of left-shifts, which is the important case for the analysis of forward curves.

References

[1] Aït-Sahalia, Y. and Jacod, J. *High-Frequency Financial Econometrics.* Princeton, New Jersey: Princeton University Press, 2014.

[2] Alfonsi, A. and Jourdain, B. “A remark on the optimal transport between two probability measures sharing the same copula”. In: *Statist. Probab. Lett.* Vol. 84 (2014), pp. 131–134.

[3] Andersen, T. et al. “Modeling and Forecasting Realized Volatility”. In: *Econometrica* vol. 71, no. 2 (2003), pp. 579–625.

[4] Aue, A., Horváth, L., and F. Pellatt, D. “Functional generalized autoregressive conditional heteroskedasticity”. In: *J. Time Ser. Anal.* Vol. 38, no. 1 (2017), pp. 3–21.

[5] Bali, J. et al. “Robust functional principal components: a projection-pursuit approach”. In: *Ann. Statist.* Vol. 39, no. 6 (2011), pp. 2852–2882.

[6] Barndorff-Nielsen, O. E. and Shephard, N. “Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics”. In: *Econometrica* vol. 72, no. 3 (2004), pp. 885–925.

[7] Barndorff-Nielsen, O. E. and Shephard, N. “Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models”. In: *J. R. Statist. Soc. B* vol. 64, no. 2 (2002), pp. 253–280.

[8] Barndorff-Nielsen, O. E., Corcuera, J., and Podolskij, M. “Limit theorems for functionals of higher order differences of Brownian semistationary processes”. In: *Prokhorov and Contemporary Probability.* Ed. by Shiryaev, A. E., Varadhan, S. R. S., and Presman, E. Vol. 33. Springer Proceedings in Mathematics and Statistics. 2013, pp. 69–96.

[9] Barndorff-Nielsen, O. E., Corcuera, J., and Podolskij, M. “Multipower variation for Brownian semistationary processes”. In: *Bernoulli* vol. 17, no. 4 (2011), pp. 1159–1194.

[10] Barndorff-Nielsen, O. E. et al. “Designing realized kernels to measure the ex post variation of equity prices in the presence of noise”. In: *Econometrica* vol. 76, no. 6 (2008), pp. 1481–1536.

[11] Benth, F., Di Nunno, G., and Schroers, D. “A topological proof of Sklar’s theorem in arbitrary dimensions”. In: *Depend. Model.* Vol. 10, no. 1 (2022), pp. 22–28.

[12] Benth, F., Di Nunno, G., and Schroers, D. “Copula Measures and Sklar’s Theorem in Arbitrary Dimensions”. In: *Scand. J. Statist. (forthcoming)* (2021).

[13] Benth, F., Šaltytė Benth, J., and Koekebakker, S. *Stochastic Modeling of Electricity and Related Markets.* World Scientific, 2008.

[14] Benth, F. E. and Krühner, P. “Representation of infinite-dimensional forward price models in commodity markets”. In: *Commun. Math. Statist.* Vol. 2, no. 1 (2014), pp. 47–106.

[15] Benth, F. E., Rüdiger, B., and Süss, A. “Ornstein–Uhlenbeck processes in Hilbert space with non-Gaussian stochastic volatility”. In: *Stoch. Proc. Applic.* Vol. 128, no. 2 (2018), pp. 461–486.

[16] Benth, F. E., Schroers, D., and Veraart, A. E. D. “A Feasible Central Limit Theorem for Realised Covariation of SPDEs in the Context of Functional Data”. In: *Available at ArXiv:2205.03927* (2022).

[17] Benth, F. E., Schroers, D., and Veraart, A. E. D. “A weak law of large numbers for realised covariation in a Hilbert space setting”. In: *Stoch. Proc. Applic.* Vol. 145 (2022), pp. 241–268.

[18] Benth, F. E. and Sgarra, C. “A Barndorff-Nielsen and Shephard model with leverage in Hilbert space for commodity forward markets.” In: *Available at SSRN 3835053* (2021).

[19] Benth, F. E. and Simonsen, I. C. “The Heston stochastic volatility model in Hilbert space”. In: *Stoch. Analysis Applic.* Vol. 36, no. 4 (2018), pp. 733–750.

[20] Bibbona, E., Sacerdote, L., and Torre, E. “A copula-based method to build diffusion models with prescribed marginal and serial dependence”. In: *Methodol. Comput. Appl. Probab.* Vol. 18, no. 3 (2016), pp. 765–783.

[21] Bibinger, M. and Trabs, M. “Volatility estimation for stochastic PDEs using high-frequency observations”. In: *Stoch. Proc. Applic.* Vol. 130, no. 5 (2020), pp. 3005–3052.

[22] Billingsley, P. *Convergence of Probability Measures.* 2nd ed. New York: John Wiley and Sons, 1999.

[23] Björk, T. and Landén, C. “On the construction of finite dimensional realizations for nonlinear forward rate models”. In: *Finance Stoch.* Vol. 6, no. 3 (2002), pp. 303–331.

[24] Björk, T. and Svensson, L. “On the existence of finite-dimensional realizations for nonlinear forward rate models”. In: *Math. Finance* vol. 11, no. 2 (2001), pp. 205–243.

[25] Boente, G. and Salibian-Barrera, M. “S-estimators for functional principal component analysis”. In: *J. Amer. Statist. Assoc.* Vol. 110, no. 511 (2015), pp. 1100–1111.

[26] Bosq, D. *Linear Processes in Function Space: Theory and Applications.* Vol. 149. Springer, 2012.

[27] Buehler, H. “Consistent Variance Curve Models”. In: *Finance Stoch.* Vol. 10, no. 2 (2006), pp. 178–203.

[28] Cai, T. and Hall, P. “Prediction in functional linear regression”. In: *Ann. Statist.* Vol. 34, no. 5 (2006), pp. 2159–2179.

[29] Cardot, H., Ferraty, F., and Sarda, P. “Functional linear model”. In: *Statist. Probab. Lett.* Vol. 45, no. 1 (1999), pp. 11–22.

[30] Carmona, R. and Nadtochiy, S. “Local volatility dynamic models”. In: *Finance Stoch.* Vol. 13, no. 1 (2009), pp. 1–48.

[31] Carmona, R. and Tehranchi, M. *Interest Rate Models: an Infinite Dimensional Stochastic Analysis Perspective.* Springer Finance. Springer Berlin Heidelberg, 2007.

[32] Cerovecki, C. et al. "Functional GARCH models: the quasi-likelihood approach and its applications". In: *J. Econometrics* vol. 209, no. 2 (2019), pp. 353–375.

[33] Cherubini, U. et al. *Dynamic Copula Methods in Finance*. John Wiley & Sons, Ltd, 2012.

[34] Chong, C. "High-frequency analysis of parabolic stochastic PDEs". In: *Ann. Statist.* Vol. 48, no. 2 (Apr. 2020), pp. 1143–1167.

[35] Cialenco, I. "Statistical Inference for SPDEs: an overview". In: *Statist. Inf. Stoch. Proc.* Vol. 20, no. 2 (Dec. 2018), pp. 309–329.

[36] Cont, R. "Modeling Term Structure Dynamics: An Infinite Dimensional Approach". In: *Int. J. Theor. Appl. Finance* vol. 08, no. 03 (2005), pp. 357–380.

[37] Corcuera, J., Nualart, D., and Podolskij, M. "Asymptotics of weighted random sums". In: *Commun. Appl. Indust. Math.* Vol. 6, no. 1 (2014), e–486, 11.

[38] Corcuera, J. M., Nualart, D., and Woerner, J. H. "Power variation of some integral fractional processes". In: *Bernoulli* vol. 12, no. 4 (Aug. 2006), pp. 713–735.

[39] Corcuera, J. M. et al. "Asymptotic theory for Brownian semi-stationary processes with application to turbulence". In: *Stoch. Proc. Applic.* Vol. 123, no. 7 (2013), pp. 2552–2574.

[40] Cox, S., Karbach, S., and Khedher, A. "Affine pure-jump processes on positive Hilbert-Schmidt operators". In: *Available at ArXiv:2012.10406* (2020).

[41] Cox, S., Karbach, S., and Khedher, A. "An infinite-dimensional affine stochastic volatility model". In: *Math. Finance* (2021).

[42] Da Prato, G., Jentzen, A., and Röckner, M. "A mild Itô formula for SPDEs". In: *Trans. Amer. Math. Soc.* Vol. 372, no. 6 (2019), pp. 3755–3807.

[43] Da Prato, G. and Zabczyk, J. *Stochastic Equations in Infinite Dimensions*. Second. Vol. 152. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, 2014.

[44] Darsow, W. F., Nguyen, B., and Olsen, E. T. "Copulas and Markov processes". In: *Illinois J. Math.* Vol. 36, no. 4 (Dec. 1992), pp. 600–642.

[45] Dauxois, J., Pousse, A., and Romain, Y. "Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference". In: *J. Multivariate Anal.* Vol. 12, no. 1 (1982), pp. 136–154.

[46] Demarta, S. and McNeil, A. "The t copula and related copulas". In: *Internat. Statist. Rev.* Vol. 73, no. 1 (2005), pp. 111–129.

[47] Durante, F., Fernández-Sánchez, J., and Sempi, C. “A topological proof of Sklar’s theorem”. In: *Appl. Math. Lett.* Vol. 26, no. 9 (2013), pp. 945–948.

[48] Fang, H., Fang, K., and Kotz, S. “The Meta-elliptical Distributions with Given Marginals”. In: *J. Multivariate Anal.* Vol. 82, no. 1 (2002), pp. 1–16.

[49] Ferraty, F. and Vieu, P. *Nonparametric Functional Data Analysis: Theory and Practice.* New York: Springer, 2006.

[50] Filipović, D. “Invariant manifolds for weak solutions to stochastic equations”. In: *Probab. Theory Related Fields* vol. 118, no. 3 (2000), pp. 323–341.

[51] Filipović, D., Pelger, M., and Ye, Y. *Stripping the Discount Curve - a Robust Machine Learning Approach.* Swiss Finance Institute Research Paper Series 22-24. Swiss Finance Institute, 2022.

[52] Filipović, D. and Teichmann, J. “On the geometry of the term structure of interest rates”. In: vol. 460. 2041. Stochastic analysis with applications to mathematical finance. 2004, pp. 129–167.

[53] Filipović, D. “Consistency Problems for HJM Interest Rate Models”. In: vol. 1760. Lecture Notes in Mathematics. Berlin: Springer, 2001.

[54] Filipović, D. “Exponential-polynomial families and the term structure of interest rates”. In: *Bernoulli* vol. 6, no. 6 (2000), pp. 1081–1107.

[55] Filipović, D. *Term-Structure Models: A Graduate Course.* Springer Finance. Springer Berlin Heidelberg, 2009.

[56] Gawarecki, L. and Mandrekar, V. *Stochastic Differential Equations in Infinite Dimensions: with Applications to Stochastic Partial Differential Equations.* Probability and Its Applications. Springer Berlin Heidelberg, 2010.

[57] Gervini, D. “Robust functional estimation using the median and spherical principal components”. In: *Biometrika* vol. 95, no. 3 (2008), pp. 587–600.

[58] Granelli, A. and Veraart, A. E. D. “A central limit theorem for the realised covariation of a bivariate Brownian semistationary process”. In: *Bernoulli* vol. 25, no. 3 (Aug. 2019), pp. 2245–2278.

[59] Grenander, U. “Stochastic processes and statistical inference”. In: *Ark. Mat.* Vol. 1 (1950), pp. 195–277.

[60] Hall, P. and Horowitz, J. “Methodology and convergence rates for functional linear regression”. In: *Ann. Statist.* Vol. 35, no. 1 (2007), pp. 70–91.

[61] Hall, P. and Hosseini-Nasab, M. “On properties of functional principal components analysis”. In: *J. R. Stat. Soc. Ser. B Stat. Methodol.* Vol. 68, no. 1 (2006), pp. 109–126.

[62] Hall, P., Müller, H.-G., and Wang, J.-L. “Properties of principal component methods for functional and longitudinal data analysis”. In: *Ann. Statist.* Vol. 34, no. 3 (2006), pp. 1493–1517.

[63] Hausenblas, E. and Riedle, M. “Copulas in Hilbert spaces”. In: *Stochastics* vol. 89, no. 1 (2017), pp. 222–239.

[64] Hörmann, S. and Kokoszka, P. “Weakly dependent functional data”. In: *Ann. Statist.* Vol. 38, no. 3 (2010), pp. 1845–1884.

[65] Hörmann, S., Horváth, L., and Reeder, R. “A FUNCTIONAL VERSION OF THE ARCH MODEL”. In: *Econometric Theory* vol. 29, no. 2 (2013), pp. 267–288.

[66] Horváth, L. and Kokoszka, P. *Inference for Functional Data with Applications*. Springer Series in Statistics. Springer, 2012.

[67] Hsing, T. and Eubank, R. *Theoretical foundations of functional data analysis, with an introduction to linear operators.* Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2015.

[68] Ibragimov, R. “Copula-based characterizations for higher order Markov processes”. In: *Econom. Theory* vol. 25, no. 3 (2009), pp. 819–846.

[69] Ibragimov, R. and Lentzas, G. “Copulas and long memory”. In: *Probab. Surv.* Vol. 14 (2017), pp. 289–327.

[70] Jacod, J. and Protter, P. “Asymptotic error distributions for the Euler method for stochastic differential equations”. In: *Ann. Probab.* Vol. 26, no. 1 (1998), pp. 267–307.

[71] Jacod, J. “Asymptotic properties of realized power variations and related functionals of semimartingales”. In: *Stoch. Proc. Applic.* Vol. 118, no. 4 (2008), pp. 517–559.

[72] Jacod, J. and Protter, P. *Discretization of Processes.* Vol. 67. Stochastic Modelling and Applied Probability. Springer, Heidelberg, 2012.

[73] Jacod, J. et al. “Microstructure noise in the continuous case: the pre-averaging approach”. In: *Stoch. Proc. Applic.* Vol. 119, no. 7 (2009), pp. 2249–2276.

[74] Junker, M., Szimayer, A., and Wagner, N. “Nonlinear term structure dependence: Copula functions, empirics, and risk implications”. In: *J. Bank. Financ.* Vol. 30, no. 4 (2006), pp. 1171–1199.

[75] Kallsen, J. and Krühner, P. “On a Heath-Jarrow-Morton approach for stock options”. In: *Finance Stoch.* Vol. 19, no. 3 (2015), pp. 583–615.

[76] Karhunen, K. “Zur Spektraltheorie stochastischer Prozesse”. In: *Ann. Acad. Sci. Fennicae Ser. A. I. Math.-Phys.* (1946), p. 7.

[77] Kühnert, S. “Functional ARCH and GARCH models: A Yule-Walker approach”. In: *Electron. J. Stat.* Vol. 14, no. 2 (2020), pp. 4321–4360.

[78] Kraus, D. and Panaretos, V. “Dispersion operators and resistant second-order functional data analysis”. In: *Biometrika* vol. 99, no. 4 (2012), pp. 813–832.

[79] Lagerås, A. N. “Copulas for Markovian dependence”. In: *Bernoulli* vol. 16, no. 2 (2010), pp. 331–342.

[80] Liu, H., Lafferty, J., and Wasserman, L. “The Nonparanormal: Semiparametric Estimation of High Dimensional Undirected Graphs”. In: *J. Mach. Learn. Res.* Vol. 10, no. 80 (2009), pp. 2295–2328.

[81] Liu, W. and Röckner, M. *Stochastic Partial Differential Equations: An Introduction*. Universitext. Springer International Publishing, 2015.

[82] Locantore, N. et al. “Robust principal component analysis for functional data”. In: *Test* vol. 8, no. 1 (1999). With discussion and a rejoinder by the authors, pp. 1–73.

[83] Masak, T. and Panaretos, V. “Random Surface Covariance Estimation by Shifted Partial Tracing”. In: *J. Amer. Statist. Assoc.* (2022).

[84] Müller, H.-G., Sen, R., and Stadtmüller, U. “Functional data analysis for volatility”. In: *J. Econometrics* vol. 165, no. 2 (2011), pp. 233–245.

[85] Nelsen, R. B. *An Introduction to Copulas*. 2. Springer Series in Statistics. Springer, New York, 2006.

[86] Panaretos, V. M. and Tavakoli, S. “Fourier analysis of stationary time series in function space”. In: *Ann. Statist.* Vol. 41, no. 2 (Apr. 2013), pp. 568–603.

[87] Passeggeri, R. and Veraart, A. E. D. “Limit theorems for multivariate Brownian semistationary processes and feasible results”. In: *Adv. Appl. Probab.* Vol. 51 (2019), pp. 667–716.

[88] Peszat, S. and Zabczyk, J. *Stochastic Partial Differential Equations with Lévy Noise: An Evolution Equation Approach*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2007.

[89] Ramsay, J. O. and Silverman, B. W. *Applied functional data analysis*. Springer Series in Statistics. Methods and case studies. Springer-Verlag, New York, 2002.

[90] Ramsay, J. and Silverman, B. W. *Functional Data Analysis*. Second. Springer Series in Statistics. Springer, 2005.

[91] Rice, G., Wirjanto, T., and Zhao, Y. “Tests for conditional heteroscedasticity of functional data”. In: *J. Time Series Anal.* Vol. 41, no. 6 (2020), pp. 733–758.

[92] Rice, J. and Silverman, B. “Estimating the mean and covariance structure nonparametrically when the data are curves”. In: *J. Roy. Statist. Soc. Ser. B* vol. 53, no. 1 (1991), pp. 233–243.

[93] Schweizer, M. and Wissel, J. “Arbitrage-free market models for option prices: the multi-strike case”. In: *Finance Stoch.* Vol. 12, no. 4 (2008), pp. 469–505.

[94] Silverman, B. “Smoothed functional principal components analysis by choice of norm”. In: *Ann. Statist.* Vol. 24, no. 1 (1996), pp. 1–24.

[95] Tappe, S. “An alternative approach on the existence of affine realizations for HJM term structure models”. In: *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* Vol. 466, no. 2122 (2010), pp. 3033–3060.

[96] Wang, J.-L., Chiou, J.-M., and Müller, H.-G. “Functional data analysis”. In: *Annu. Rev. Stat. Appl* vol. 3 (2016), pp. 257–295.

[97] Wilson, A. G. and Ghahramani, Z. “Copula processes”. In: *Adv. Neural Inf. Process. Syst.* 2010, pp. 2460–2468.

[98] Yao, F., Müller, H.-G., and Wang, J.-L. “Functional data analysis for sparse longitudinal data”. In: *J. Amer. Statist. Assoc.* Vol. 100, no. 470 (2005), pp. 577–590.

[99] Yao, F., Müller, H.-G., and Wang, J.-L. “Functional linear regression analysis for longitudinal data”. In: *Ann. Statist.* Vol. 33, no. 6 (2005), pp. 2873–2903.

[100] Zhang, L., Mykland, P., and Aït-Sahalia, Y. “A tale of two time scales: Determining integrated volatility with noisy high-frequency data”. In: *J. Amer. Statist. Assoc.* Vol. 100, no. 472 (2005), pp. 1394–1411.

[101] Zhong, R. et al. “Robust functional principal component analysis for non-Gaussian longitudinal data”. In: *J. Multivariate Anal.* Vol. 189 (2022), Paper No. 104864, 14.

Papers

Paper I

Copula Measures and Sklar's Theorem in Arbitrary Dimensions

Fred Espen Benth, Giulia Di Nunno, Dennis Schroers

To appear in *Scandinavian Journal of Statistics*, October 2021, DOI: 10.1111/sjos.12559.

Abstract

Although copulas are used and defined for various infinite-dimensional objects (e.g. Gaussian processes and Markov processes), there is no prevalent notion of a copula that unifies these concepts. We propose a unified functional analytic framework, show how Sklar's theorem can be applied in certain examples of Banach spaces and provide a semiparametric estimation procedure for second order stochastic processes with underlying Gaussian copula.

Contents

I.1 Introduction

The investigation of linear and nonlinear dependence structures between the elements in an arbitrary family of random objects is inherent in many problems, ranging from modelling dependence within Markov processes (see e.g. [15],[28], [23]), Gaussian processes, and general processes with continuous marginals (see e.g. [45]), to the modelling of dependence between semimartingale processes (see e.g. [25],[5]) or the components of a random measure (see e.g. [38]). One of the most powerful tools, which captures the whole structure of statistical dependence for a finite collection of real-valued random variables, are copulas. The theory for copulas is rather well-developed for the finite-dimensional case (see [34] for an introduction to the topic). In this paper, we develop a general theory of copulas in infinite-dimensional vector spaces and point out its capability to approximate and estimate the laws of certain second-order stochastic processes in a semiparametric (metaelliptical) framework.

Copulas are cumulative distribution functions with uniform marginals, which can generally be interpreted as the dependence structure separated from the laws of the marginals by virtue of Sklar's theorem. The first part of this theorem states that each multivariate Borel distribution possesses an underlying copula (representing its dependence structure), whereas the second part enables us to merge any copula with a freely chosen family of one-dimensional marginals to a multivariate Borel measure (with the copula measure as the specified dependence structure). It is straightforward to extend this result to laws on infinite-dimensional product spaces by replacing the notion with copulas as cumulative distribution functions by *copula measures* as we point out in Section I.2.

Nevertheless, as we also set an eye towards applications, that is, numerical approximations or (functional) data analysis, it is relevant to have sufficient knowledge about various properties like integrability or regularity of the corresponding stochastic processes. Thus it is inevitable to consider the framework for measures in (topological) function spaces.

Unfortunately, the advent of copulas in function spaces is subject to some nontrivial difficulties compared to the finite-dimensional setting: first of all, it is not immediately clear what marginals are in an infinite-dimensional vector space. If X is a random variable in a Hilbert space H, projections onto an orthonormal basis $(\langle X, e_n\rangle_H)_{n\in\mathbb{N}}$ are reasonable candidates. This case was treated in [20]. Nevertheless, if, in addition, the space considered is a reproducing kernel Hilbert space of functions, say over $[0,1]$, an equally natural option for marginals are function evaluations $(X(t))_{t\in[0,1]}$. This motivates the introduction of a flexible framework, which is needed but does not yet exist. We, therefore, propose a general concept of marginals for measurable vector spaces.

Another critical point in the infinite-dimensional setting is that even if we fix a certain notion for marginals and then construct a measure with some given dependence structure (i.e. a copula) and marginals via Sklar's theorem, this measure is not necessarily a Borel measure on the desired function space, but may rather be just a cylindrical premeasure. Proving whether a cylindrical premeasure corresponds to a proper probability measure is a considerably difficult task. This technical complication was already encountered in [20] and we will refer to it as the *construction problem*, in the context of applying the second part of Sklar's theorem. In applications, we further wish to be flexible in the choice of copulas and marginals and hence want to avoid to be overly restrictive by stating complicated conditions on the mutual behaviour of dependence structure and one-dimensional distributions. Otherwise useful criteria, for instance, those based on compactness arguments (e.g. Theorem 6.2 in [10]), may turn out to be cumbersome to translate into feasible criteria to overcome the construction problem. Nevertheless, there are several important situations in which one can find a satisfying framework to work with. One part of our work describes respective constructions, namely in the space of continuous functions $C(T)$, in Hölder spaces, in the Lebesgue spaces $L^p(T)$, and in the sequence spaces l^p. For the latter two cases, simple moment criteria on the marginals prove to be sufficient (and sometimes even necessary), which makes them attractive in practice.

The latter is good news, as these are presumably one of the most appealing cases for applications. Even better, in some important situations, it is possible to approximate infinite dimensional copula models in $L^p(T)$ for T compact and to measure the distance of this approximation feasibly. We can conveniently bound the $L^p(T)$ distance of two random variables from above by the Wasserstein distance of their one-dimensional marginals and the $L^q(T)$-distance of the corresponding underlying copula processes (i.e. processes that have the corresponding copulas as their laws) for any $q \geq 1$, under suitable smoothness and tail assumptions on the marginals of one of the variables. Namely, for all $q \geq 1$, one can find ρ and K such that

$$\|X - Y\|^p_{L^p(\Omega \times T)} \leq \|\mathbb{W}_p(F_{X.}, F_{Y.})\|_{L^p(T)} + K\|U^X - U^Y\|^\rho_{L^q(\Omega \times T)}, \tag{I.1}$$

where U^X, U^Y are the underlying copula processes and F_{X_t} and F_{Y_t} for $t \in T$ are the marginal cumulative distribution functions of the stochastic processes X and Y. Moreover, we demonstrate how to apply this bound in order to approximate (heavy-tailed and tail-dependent) stochastic processes with underlying elliptical copula and regularly varying marginals. We finally show how to estimate the laws of *nonparanormal stochastic processes* (wording in line with [29]) in the presence of functional data. We derive convergence rates for a law of large numbers in Wasserstein space by inequality (I.1) for these processes, which are assumed to have an underlying Gaussian copula but no parametric restriction on the marginals.

The paper is organised as follows. We describe the basic framework of copulas in product spaces and prove Sklar's theorem in Section I.2. Section I.3 is devoted to copula constructions in function spaces, where in Subsection I.3.1 we introduce a general framework for marginals in measurable vector spaces and describe the abstract construction problem. Subsection I.3.2 presents criteria to overcome the latter in various function spaces. Finally, Section I.4 provides distance estimates for the copula construction, where we recall the connection of copulas to Wasserstein spaces in Subsection I.4.1 and derive an estimate of the $L^p(T)$ distance of two processes in terms of the difference of the underlying copula and the one-dimensional Wasserstein distances of their marginals in Subsection I.4.2. Section I.4.3 demonstrates the applicability of this inequality to the estimation of nonparanormal stochastic processes. We give our conclusions and outlook in Section I.5.

Notation

For any measure μ on a measurable space $(B, \mathcal{B})$ and a measurable function $f : (B, \mathcal{B}) \to (A, \mathcal{A})$ into another measurable space $(A, \mathcal{A})$ we denote by $f_*\mu$ the pushforward measure with respect to f given by $f_*\mu(S) := \mu(f^{-1}(S))$ for all $S \in \mathcal{A}$. If $B = \mathbb{R}^I$, where I is an arbitrary index set, and $\mathcal{B} = \otimes_{i\in I}\mathcal{B}(\mathbb{R})$, we use the shorter notations $\pi_{J*}\mu =: \mu_J$ for a subset $J \subseteq I$ and $\pi_{\{i\}*}\mu =: \mu_i$ for an element $i \in I$, where π_J denotes the projection on $\mathbb{R}^J$. If $J \subset I$ is finite, we denote the corresponding finite-dimensional cumulative distribution functions by F_{μ_J} or F_{μ_i} respectively. We will frequently refer to the one-dimensional distributions $\mu_i, i \in I$ and equivalently $F_{\mu_i}, i \in I$ as marginals of the measure μ. Throughout the paper all random variables are considered on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and we write $L^0(\Omega, \mathcal{F}; A, \mathcal{A}) =: L^0(\Omega; A)$ for the measurable functions $f : (\Omega, \mathcal{F}) \to (A, \mathcal{A})$, i.e., A-valued random variables.

I.2 Copulas and Sklar's Theorem in Infinite Dimensions

Following the natural interpretation of copulas as measures in finite dimensions (see Appendix I.A for a short treatment of copulas in finite dimensions), we suggest defining the concept in the same line also in infinite dimensions:

Definition I.2.1. A copula measure (or simply copula) on $\mathbb{R}^I$ is a probability measure C on $\otimes_{i\in I}\mathcal{B}(\mathbb{R})$, such that its marginals C_i are uniformly distributed on $[0, 1]$.

For finite-dimensional index sets I the notions of measures and cumulative distribution functions have a one-to-one correspondence, which is the reason why in this case a copula measure C can be uniquely identified with the copula F_C in the classical sense of Definition I.A.2. For the same reason the finite-dimensional distributions C_J of an infinite-dimensional copula measure C, correspond uniquely to the copula F_{C_J} in the familiar sense of copulas in finite dimensions.

We also introduce the important notion of *copula processes.*

Definition I.2.2. We call a random variable $U \in L^0(\Omega; \mathbb{R}^I)$ with uniform marginals on $[0, 1]$ a *copula process.* That is, the law of a copula process is a copula measure.

Since for each copula, we can find a probability space, and a copula process with law C on it, the notion of copulas has a one-to-one correspondence with the one of copula processes.

As in finite dimensions, the most important result for the use of copulas is Sklar's theorem:

Theorem I.2.3 (Sklar's theorem)**.** *Let I be an index set and μ be a probability measure on $\otimes_{i\in I}\mathcal{B}(\mathbb{R})$ with marginal one-dimensional distributions $\mu_i, i \in I$.*

There exists a copula measure C, such that for each finite subset $J \subseteq I$, we have

$$F_{C_J}\left(\left(F_{\mu_j}(x_j)\right)_{j\in J}\right) = F_{\mu_J}\left((x_j)_{j\in J}\right) \tag{I.2}$$

for all $(x_j)_{j\in J} \in \mathbb{R}^J$. Moreover, C is unique if F_{μ_i} is continuous for each $i \in I$. Vice versa, let C be a copula measure on $\mathbb{R}^I$ and let $(\mu_i)_{i\in I}$ be a collection of (one-dimensional) Borel probability measures over $\mathbb{R}$. Then there exists a unique probability measure μ on $\otimes_{i\in I}\mathcal{B}(\mathbb{R})$, such that (I.2) *holds.*

In the following proof and the rest of the paper, we often use for a one-dimensional Borel measure μ_i on $\mathbb{R}$ the notation $F_{\mu_i}^{[-1]}$ for the quantile functions

$$F_{\mu_i}^{[-1]}(u) := \inf\left\{x \in (-\infty,\infty) : F_{\mu_i}(x) \geq u\right\}, \tag{I.3}$$

Proof. To prove the first part, let $(X_i)_{i\in I}$ be a random vector having μ as its law. Let U be a standard uniformly distributed real-valued random variable on the same probability space, such that U is independent of $(X_i)_{i\in I}$. For a one-dimensional distribution function we denote its left-limit by $F_{\mu_i}(x-) := \lim_{y\uparrow x} F_{\mu_i}(y)$. Define the distributional transform process $(U_i)_{i\in I}$ by

$$U_i := F_{\mu_i}(X_i-) + U\left(F_{\mu_i}(X_i) - F_{\mu_i}(X_i-)\right)$$

and C to be the law of $(U_i)_{i\in I}$. Since each U_i is uniformly distributed on $[0,1]$ and the finite-dimensional laws C_J fulfill (I.2) by Theorem I.A.4, C is the copula measure we looked for. Observe that in case of continuous marginals all finite-dimensional marginals of C must be uniquely determined by the unique copulas of the finite-dimensional laws of μ induced by Sklar's theorem in finite dimensions.

To prove the other direction of Sklar's theorem, observe that, since $F_{\mu_i}^{[-1]} : [0,1] \to \mathbb{R}$ is measurable for every $i \in I$ we have that $(F_{\mu_i}^{[-1]})_{i\in I}$ is a measurable map from the product space $([0,1]^I, \otimes_{i\in I}\mathcal{B}([0,1]), C)$ to $(\mathbb{R}^I, \otimes_{i\in I}\mathcal{B}(\mathbb{R}))$. The measure μ on $\otimes_{i\in I}\mathcal{B}(\mathbb{R})$ given by the corresponding pushforward measure

$$\mu := ((F_{\mu_i}^{[-1]})_{i\in I})_* C \tag{I.4}$$

has the desired properties. To see this, we just have to verify that μ has the finite-dimensional distributions induced by (I.2). Observe that, for all $i \in I$, by the monotonicity of the cumulative distribution functions we have that, for all $x \in (-\infty,\infty)$,

$$\left\{u \in [0,1] : F_{\mu_i}^{[-1]}(u) \leq x\right\} \supseteq \left\{u \in [0,1] : u < F_{\mu_i}(x)\right\} = [0, F_{\mu_i}(x))$$

and

$$\left\{u \in [0,1] : F_{\mu_i}^{[-1]}(u) \leq x\right\} \subseteq \left\{u \in [0,1] : u \leq F_{\mu_i}(x)\right\} = [0, F_{\mu_i}(x)].$$

Thus, for $J \subset I$ finite, we have for all $(x_j)_{j\in J} \in \mathbb{R}^J$ that

$$\left([0, F_{\mu_j}(x_j)]\right)_{j\in J} \setminus \left(\left\{u \in [0,1] : F_{\mu_j}^{[-1]}(u) \leq x_j\right\}\right)_{j\in J} \subseteq \left(\{F_{\mu_j}(x_j)\}\right)_{j\in J}$$

is a C_J null set. Therefore we obtain

$$
\begin{aligned}
C_J\left(\left(\left(F_{\mu_j}^{[-1]}\right)^{-1}(-\infty, x_j]\right)_{j\in J}\right) &= C_J\left(\left(\left\{u \in [0,1] : F_{\mu_j}^{[-1]}(u) \le x_j\right\}\right)_{j\in J}\right)\\
&= C_J\left(\left([0, F_{\mu_j}(x_j)]\right)_{j\in J}\right)\\
&= F_{C_J}\left(F_{\mu_j}\left((x_j)\right)_{j\in J}\right).
\end{aligned}
$$

This concludes the proof. ■

Remark I.2.4. If I is a finite set, Theorem I.2.3 coincides with Sklar's theorem I.A.3 in finite dimensions by identifying the copula measure uniquely with its corresponding cumulative distribution function.

Remark I.2.5. From the proof above it follows that for a copula measure C on $\mathbb{R}^I$ and a collection of marginals $(\mu_i)_{i\in I}$, the pushforward measure in (I.4) represents a probability measure μ on $\mathbb{R}^I$ having this underlying copula and marginals.

The following examples review some existing concepts of copulas, which can be embedded into our framework:

Example I.2.6. (Complete dependence and independence copulas) The complete dependence copula measure on $\mathbb{R}^I$ is the law corresponding to the consistent family of finite-dimensional cumulative distribution functions given by $M_J((u_j)_{j\in J}) = \min_{j\in J} u_j$. Observe, that its finite-dimensional distribution functions are Fréchet-Hoeffding upper bounds for the corresponding finite-dimensional copulas, that is, for all $J \subset I$ finite and any copula C on $\mathbb{R}^I$ we have

$$
F_{C_J}\left((u_j)_{j\in J}\right) \le M_J\left((u_j)_{j\in J}\right) \quad \forall (u_j)_{j\in J} \in [0,1]^J.
$$

The independence copula measure on $\mathbb{R}^I$ is the law of the consistent family of finite-dimensional cumulative distribution functions given by $\Pi_J((u_j)_{j\in J}) = \Pi_{j\in J} u_j$.

Example I.2.7. (Inversion method and Gaussian copulas) Given a law μ with continuous marginals $F_{\mu_i}, i \in I$, the underlying copula measure C induced by Sklar's theorem I.2.3 is given by its finite-dimensional cumulative distribution functions for each finite $J \subseteq I$ via

$$
F_{C_J}\left((u_j)_{j\in J}\right) := F_{\mu_J}\left(\left(F_{\mu_j}^{[-1]}(u_j)\right)_{j\in J}\right) \quad \forall (u_j)_{j\in J} \in [0,1]^J. \tag{I.5}
$$

This method is known as the inversion method (see e.g. [34]). In this way, we can derive, for instance, the copula measures that are underlying a Gaussian process (that is, each μ_J is Gaussian), which are called Gaussian copulas. Infinite-dimensional Gaussian copulas were applied already for example in [45] in a machine learning context.

Example I.2.8. (Archimedean copulas): Let I be an infinite dimensional index set. A continuous function $\psi : [0,\infty) \to [0,1]$ is called an Archimedean generator, if $\psi(0) = 1$ and $\lim_{x\to\infty} \psi(x) = 0$ and ψ is strictly decreasing on $[0, \inf\{x \geq 0 : \psi(x) = 0\})$. In this case ψ is said to generate an Archmimedean copula measure C, if

$$F_{C_J}\left((u_j)_{j\in J}\right) := \psi\left(\sum_{j\in J} \psi^{-1}(u_j)\right)$$

for each finite $J \subset I$, with the convention $\psi^{-1}(0) = \inf\{x \geq 0 : \psi(x) = 0\}$ (cf. chapter 2 in [32]). In the case of a finite index set I, say with cardinality $2 \leq d < \infty$, by [32, Theorem 2.2] we know that ψ generates an Archimedean copula if and only if it is d-monotone, that is, it has derivatives of up to order d on $(0,\infty)$ and $(-1)^k f^{(k)}(x) \geq 0$ for all $k = 0, ..., d-2$ and $x \in (0,\infty)$ and $(-1)^{d-2} f^{(d-2)}$ is nonincreasing and convex on $(0,\infty)$. This is translated to the infinite dimensional case by replacing condition of d-monotonicity by the condition that ψ is completely monotone, i.e. ψ has derivatives of all orders on $(0,\infty)$ and $(-1)^k f^{(k)}(x) \geq 0$ for all $k \in \mathbb{N}$ and $x \in (0,\infty)$. In fact, [32, Proposition 2.4] yields that if I has infinite cardinality, ψ generates an Archimedean copula measure, if and only if it is completely monotone. A prototypical example is the Clayton copula measure, with the generator $\psi_\theta(x) := \max((1+\theta x)^{-\frac{1}{\theta}}, 0)$ for a parameter $\theta \in \mathbb{R} \setminus \{0\}$ (as a limit from both sides, the case $\theta = 0$ also has an interpretation as the independence copula). While in finite dimensions ψ_θ generates an Archimedean copula for some $\theta < 0$, this possibility is completely ruled out in infinite dimensions, i.e. if I is infinite, ψ_θ generates an Archimedean copula if and only if θ is nonnegative (cf. Example 3 in [32]). It is also worth noting that by definition, these probability measures are exchangeable and it was shown in [13] that Archimedean copulas in infinite dimensions can be related to Dirichlet distributions.

In addition, our framework accommodates also Markov copulas, introduced in [15] and developed also, e.g., in [28], [24], [23], and [4], copulas for time series introduced in [12] and copulas in Hilbert spaces from [20].

I.3 Copulas in Topological Vector Spaces

In this section, we formulate a unified setting for the notion of copulas in the framework of vector spaces.

I.3.1 Marginals in Vector Spaces

Let V be a vector space over $\mathbb{R}$ and $\mathcal{V}$ a σ-algebra over this space. Recall that the algebraic dual of V is defined as the vector space

$$Hom(V, \mathbb{R}) := \{\varphi : V \to \mathbb{R} : \varphi \text{ is linear}\}.$$

Definition I.3.1 (M-Marginals)**.** Let X be a random variable on V. Let M be a linearly independent subspace of measurable functions in $Hom(V,\mathbb{R})$ that separates the points of V (i.e. for all $v, w \in V$ there is an $m \in M$ such that $m(v) \neq m(w)$). Then we call the random variables $(m(X) : m \in M)$ the M-marginals of X.

Observe that by the definition above, we can embed the vector space framework into the framework of product spaces, by the embedding

$$V \ni v \mapsto (m(v))_{m\in M} \in \mathbb{R}^M, \tag{I.6}$$

which is necessary for the application of our copula theory.

Some choices of M which are of practical importance are given in the sequel.

Example I.3.2 (Marginals in finite dimensions)**.** In the finite-dimensional case, that is $V = \mathbb{R}^d$ for some $d \in \mathbb{N}$, M is necessarily of the form

$$M = \{\langle e_1, \cdot\rangle, ..., \langle e_d, \cdot\rangle\} \tag{I.7}$$

for a basis $e_1, ..., e_d$ of $\mathbb{R}^d$ and where $\langle\cdot,\cdot\rangle$ denotes an inner product on $\mathbb{R}^d$. In terms of finite-dimensional copula theory, the natural choice is the standard basis $e_1 = (1, 0...., 0)$, ... , $e_d = (0, ..., 0, 1)$.

Example I.3.3 (Product space)**.** It is possible to embed the product space setting from Section I.2 into the framework of measurable vector spaces: the product space $V = \mathbb{R}^I$ for some index set I becomes a measurable vector space, if we equip it with the product σ-algebra $\otimes_{i\in I}\mathcal{B}(\mathbb{R})$. The projections (or evaluation functionals) $\pi_j((v_i)_{i\in I}) := v_j$ for $j \in I$ are measurable (even continuous) by definition, linearly independent and separate the points. Thus, we can take

$$M = \{\pi_i : i \in I\}. \tag{I.8}$$

Observe that we can do this with every space of functions, in which the evaluations are linearly independent. For instance, we can take the space of p-integrable functions over a subset T of $\mathbb{R}^d$, $d \in \mathbb{N}$

$$\begin{aligned}\mathcal{L}^p(T) &:= \mathcal{L}^p(T, \mathcal{A}, \mu; \mathbb{R}) \\ &= \left\{ f : T \to \mathbb{R} : f \text{ is measurable and } \|f\|_{\mathcal{L}^p(T)} := (\int_T f(t)^p \mu(dt))^{\frac{1}{p}} < \infty \right\}\end{aligned} \tag{I.9}$$

for some natural number p and a measure space $(T, \mathcal{A}, \mu)$. Observe that in this setting we work in a space of functions, rather than of equivalence classes. The reason is that point evaluations are not well defined in Banach spaces of equivalence classes. This serves also as motivation for Subsection I.3.2.1 where we construct copulas under these circumstances.

Example I.3.4 (Path marginals for Banach spaces of functions)**.** Let V be a separable Banach space of real-valued functions on a set T such that the evaluation functionals $\delta_t f := f(t)$ (or projections in terms of product spaces) are continuous and $\mathcal{V} = \mathcal{B}(V)$ is the Borel σ-algebra with respect to the corresponding norm topology. In most of these settings, the subset

$$M = \{\delta_t : t \in T\} \tag{I.10}$$

of evaluations is linearly independent and, due to continuity, it consists of measurable functionals. Important examples in this framework are the continuous functions $V = C(T)$ and $V = B_K$, where B_K is a reproducing kernel Banach or Hilbert space in the sense of [46] or [3].

Example I.3.5. (Basis marginals) If V is a Banach space that possesses a Schauder basis (cf. Definition I.3.19) we can take

$$M = \{f_n : n \in \mathbb{N}\}\,. \tag{I.11}$$

where $(f_n)_{n\in\mathbb{N}}$ is the sequence of coefficient functionals of the Schauder basis. Examples of Banach spaces that possess such a basis are $C([0,1])$, $L^p([0,1])$, the sequence spaces l^p and, as a special case, all separable Hilbert spaces with orthonormal bases as Schauder bases. Note that in the latter case we are effectively in the setting of *consistent copulas* from [20].

Remark I.3.6 (Marginals for nonlinear subspaces). Note, that if we are just interested in defining random variables on particular subsets of a vector space V, the set M must not necessarily be separating for all elements of V. One example is the construction of random probability measures, as a certain subset of random variables in the Banach space of signed measures on the real line. In this case, it suffices to take

$$M = \{F_\cdot(t) : F_\mu(t) = \mu(-\infty, t], t \in \mathbb{R}\}\,, \tag{I.12}$$

that is, we identify a random probability measure with the corresponding random cumulative distribution function.

We will refer to the choice of marginals in Examples I.3.3 and I.3.4 as *path marginals* and the corresponding copulas in this framework as *path copulas*. In contrast, the corresponding constructions in Example I.3.5 will be referred to as *basis marginals* and *basis copulas*.

Already in finite dimensions, due to different basis specifications, there is not just one choice for M. Unfortunately, copulas are not invariant under change of the notion of marginals, as shown by the following example:

Example I.3.7. Suppose (X_1, X_2) and (Y_1, Y_2) are two bivariate real random variables on the same probability space, such $C_{X_1,X_2}(u,v) = C_{Y_1,Y_2}(u,v) = uv$ is the independence copula. Assume, moreover, that $X_1 \sim N(0,1)$ and $X_2, Y_1, Y_2 \sim U(0,1)$. By Proposition 3.4.1 in [12] we know that

$$C_{X_2,X_1+X_2}(u,v) = \int_0^u \frac{d}{d_{x_1}} C_{X_1,X_2}\left(w, F_{X_1}\left(F^{[-1]}_{X_1+X_2}(v) - F^{[-1]}_{X_2}(w)\right)\right) dw.$$

One can easily show that $F_{X_1+X_2}(\frac{1}{2}) = \frac{1}{2}$ and $F_{Y_1+Y_2}(1) = \frac{1}{2}$. Thus, choosing $v = \frac{1}{2}$ and $u = \frac{1}{2}$ and since the independence copula is simply the product of the one-dimensional uniform distributions, the distribution functions of Y_2 and X_2 are the identity on $[0,1]$ and $F_{X_1}(\frac{1}{2}) < \frac{7}{10}$, we have

$$C_{X_2,X_1+X_2}(\frac{1}{2},\frac{1}{2}) = \int_0^{\frac{1}{2}} F_{X_1}(\frac{1}{2}-w)dw < \frac{7}{20} < \frac{3}{8} = \frac{1}{2}F^{[-1]}_{Y_1+Y_2}(\frac{1}{2}) - \frac{(\frac{1}{2})^2}{2}$$
$$= C_{Y_2,Y_1+Y_2}(\frac{1}{2},\frac{1}{2}).$$

Therefore (X_1, X_2) and (Y_1, Y_2) do not share the same copula with respect to $\{(0,1),(1,1)\}$-marginals, but with respect to $\{(1,0),(0,1)\}$-marginals they do.

If we want to construct a measure on a vector space V by virtue of the second part of Sklar's theorem the naive procedure reads now as follows:

Construction 1.

(i) Choose some set M which satisfies the conditions of Definition I.3.1.

(ii) Choose a copula C on $\mathbb{R}^M$ (or a copula process $(U_m)_{m\in M}$) and one-dimensional distributions $(\mu_m)_{m\in M}$ and merge them with Sklar's theorem to a law μ (or a process) on $\otimes_{m\in M}\mathcal{B}(\mathbb{R})$.

(iii) (Construction Problem) Check if μ can be identified with a measure on $\mathcal{V}$ via the embedding (I.6).

As anticipated in the introduction, the third point will not necessarily carry an affirmative answer. The choice of marginals and dependence structure in (ii) must be based on criteria that guarantee a solution to (iii), which is hereafter referred to as the *construction problem* for copulas in function spaces.

Consider now the following framework (which covers all mentioned examples): V is a topological vector space, $\mathcal{V} = \mathcal{B}(V)$ the corresponding Borel σ-algebra and M a subset of the dual that satisfies the conditions of Definition I.3.1. In addition, assume that each $m \in M$ is continuous, that is,

$$M \subset V^*,$$

where V^* denotes the topological dual of V, given by

$$V^* := \{v^* : V \to \mathbb{R} : v^* \text{ is linear and continuous}\}.$$

Then Construction 1 induced by Sklar's theorem effectively culminates in the construction of a cylindrical premeasure on that vector space (see for instance [10] or [40] for a treatment of cylindrical measure theory). In the case that V is even a separable Banach space and $M \subset V^*$, we however have the following useful criterion for our setting:

Lemma I.3.8. *Let V be a separable Banach space. Assume that M is a fundamental set with respect to the weak*-topology, that is, its linear span is dense. If the probability measure defined in Construction 1 is the law of a process $X := (X_m)_{m \in M}$, such that X is almost surely in the range of the embedding* (I.6)*, then it is the image of a Borel measurable random variable $\tilde{X}$ in V under this embedding.*

Proof. If $(X_m)_{m\in M}$ is almost surely in the range of the embedding, there is an $\tilde{\Omega} \subseteq \Omega$ with full measure and a random variable $\tilde{X}$ such that $m(\tilde{X}(\omega)) = X_m(\omega)$ for all $\omega \in \tilde{\Omega}$. Since M is a fundamental set, we have that for all $v^* \in V^*$ there is a sequence $(\sum_{i=1}^{N_n} \lambda_i^n m_i^n)_{n\in\mathbb{N}}$ in $lin(M)$ such that $\sum_{i=1}^{N_n} \lambda_i^n m_i^n \to v^*$ with respect to the weak*-topology. Thus

$$v^*(\tilde{X}) = \lim_{n\to\infty} \sum_{i=1}^{N_n} \lambda_i^n m_i^n(\tilde{X}) \qquad \text{a.s.}$$

is measurable, since linear combinations and limits of measurable functions are measurable. We conclude that $\tilde{X}$ is a weakly measurable random variable on a separable Banach space and hence, by the Pettis theorem [36, Theorem 1.1] strongly measurable, that is, measurable with respect to the Borel σ-algebra. ■

Remark I.3.9. Due to the existence of Hamel bases on V^* and the Hahn-Banach theorem (cf. Corollary 5.80 in [2]), the existence of a set M that satisfy the conditions in Definition I.3.1 is always guaranteed in locally convex Hausdorff spaces.

We will concentrate in the next sections on special cases of path- and basis constructions, which are adequate to solve the construction problem (for instance by Lemma I.3.8), which is why they are foremost of practical importance.

I.3.2 Solutions to the Construction Problem

I.3.2.1 Path Copulas for p-Integrable Stochastic Processes

We describe in this section how the copula construction induced by Sklar's theorem I.2.3 works for the function space $\mathcal{L}^p(T, \mathcal{B}(T), \mu; \mathbb{R}) =: \mathcal{L}^p(T)$ for $p \in \mathbb{N}$, a measurable set $T \subset \mathbb{R}^d$ with $d \in \mathbb{N}$ and a σ-finite measure μ. As mentioned in Example I.8, we take $M = \{\delta_t : t \in T\}$-marginals, that is, we identify a function $f \in \mathcal{L}^p(T)$ by all its function values $(f(t))_{t\in T}$. Moreover, we denote by $[f]$ the corresponding equivalence class of almost everywhere coinciding functions with f, which forms an element in the Banach space of equivalence classes $L^p(T)$.

For a stochastic process $X = (X_t)_{t\in T}$ we say that it is measurable, if the mapping $(t, \omega) \mapsto X_t(\omega)$ is $\mathcal{B}(T) \otimes \mathcal{A}/\mathcal{B}(\mathbb{R})$-measurable.

Lemma I.3.10. *Let $X = (X_t)_{t\in T}$ be a measurable stochastic process.*

(a) *Assume $X = (X_t)_{t\in T}$ is a stochastic process with sample path's in $\mathcal{L}^p(T)$ almost surely. Then there is a Borel measurable random variable Y in*

$\mathcal{L}^p(T)$ (with respect to the pseudometric induced by $\|\cdot\|_{\mathcal{L}^p(T)}$), such that $Y = X$ almost surely.

(b) *Let X be measurable. Then there is a Borel measurable random variable Y in $\mathcal{L}^p(T)$, such that $Y = X$ almost surely and $Y \in L^p(\Omega; L^p(T))$ if and only if*

$$\int_T \mathbb{E}\left[|X_t|^p\right] \mu(dt) < \infty.$$

Proof. We will verify that $[X]$ is a Borel measurable random variable on the Banach space $L^p(T)$. In that case we have that since $\mathcal{O} \in \mathcal{B}(\mathcal{L}^p(T))$ is open if and only if $[\mathcal{O}] \in \mathcal{B}(L^p(T))$ is open, thus X is $\mathcal{A}/\mathcal{B}(\mathcal{L}^p(T))$-measurable if and only if $[X]$ is $\mathcal{A}/\mathcal{B}(L^p(T))$-measurable. By Pettis theorem [36, Theorem 1.1] we have that $[X]$ is measurable, if and only if $\int_T X(t)y(t)dt$ is measurable for all $y \in L^q(T)$ with $q = \frac{p}{p-1}$ if $p \geq 2$ and for all $y \in L^\infty(T)$ if $p = 1$. Due to measurability of the process X, these integrals are indeed measurable. This shows (a).

To show (b), observe that by Fubini's theorem we have

$$\mathbb{E}\left[\int_T |X_t|^p \mu(dt)\right] = \int_T \mathbb{E}\left[|X_t|^p\right] \mu(dt)$$

whenever one of the terms in this equation is finite. Using (a), this shows the assertion. ∎

Lemma I.3.10 yields the following simple construction of random variables X such that $[X] \in L^p(\Omega; L^p(T))$:

Construction 2.

(i) Specify a measurable copula process $U = (U_t)_{t\in T}$.

(ii) define marginals $(F_t)_{t\in T}$, with corresponding pth moments $(m_t^p)_{t\in T}$, such that $(t, x) \mapsto F_t(x)$ is jointly measurable and

$$\int_T m_t^p \mu(dt) < \infty. \tag{I.13}$$

(iii) construct the new process X with underlying copula process U and marginals $(F_t)_{t\in T}$ via Sklar's theorem I.2.3, that is,

$$X_t = F_t^{[-1]}(U_t) \quad \forall t \in T.$$

This process has values in $\mathcal{L}^p(T)$ almost surely and by Lemma (I.3.10), $[X]$ is therefore an element in $L^p(\Omega; L^p(T))$.

Notice that the interpretability of the underlying path copula of X is complicated if transferred to the equivalence class $[X]$. Indeed path copulas specify dependence between point evaluations of the random function, which are not well defined anymore for equivalence classes. If one wants to specify dependence between equivalence classes, one should approach this by considering the notion of basis marginals, as described in Subsection I.3.2.3.

From a measure theoretical point of view, Banach spaces in which evaluation functionals are well defined and continuous are favourable and we will discuss this in the sense of spaces of continuous functions in the next subsection.

I.3.2.2 Path Copulas for Continuous Processes

Let T be a topological vector space. We want to establish a 'Sklar-like' theorem in the space of real continuous functions $C(T) := C(T;\mathbb{R})$. If T is a compact metrizable Hausdorff space, we equip it with the norm $\|f\|_\infty := \sup_{t\in T}|f(t)|$, making $C(T)$ a separable Banach space (cf. [2, Theorem 9.14]).

Recall that a process $X = (X_t)_{t\in T}$ with marginals $(F_t)_{t\in T}$ is continuous in distribution, if for all $t \in T$

$$\lim_{s\to t} F_s(x_0) = F_t(x_0) \quad \text{for all continuity points } x_0 \text{ of } F_t. \tag{I.14}$$

If we assume that all the marginals $F_t, t \in T$ are continuous (in x), (I.14) simplifies to the condition that

$$(t,x) \mapsto F_t(x) \quad \text{is continuous in both variables separately}$$

In the latter case we have even joint continuity:

Lemma I.3.11. *Assume that the marginals $F_t, t \in T$ of an almost surely continuous process $X = (X_t)_{t\in T}$ are continuous. Then*

$$(t,x) \mapsto F_t(x) \quad \textit{is jointly continuous.} \tag{I.15}$$

If the marginals are strictly increasing between the points $F_t^{[-1]}(0+)$ and $F_t^{[-1]}(1)$ in x we have that

$$(t,x) \mapsto F_t^{[-1]}(x) \quad \textit{is jointly continuous.} \tag{I.16}$$

Proof. Due to Lemma 21.2 from [43] we have that $t \mapsto F_t^{[-1]}$ is pointwise continuous. The proof follows analogously to the arguments of the proof of Proposition 1 in [27]:

Let $(t_0, x_0) \in T \times \mathbb{R}$. Then for all $\epsilon > 0$ there are $\delta > 0$ and neighbourhoods U_1, U_2 of t_0 in T such that

$$|F_{t_0}(x) - F_{t_0}(x_0)| \le \frac{\epsilon}{2} \quad \text{if } |x - x_0| \le \delta \tag{I.17}$$

$$|F_t(x_0+\delta) - F_{t_0}(x_0+\delta)| \le \frac{\epsilon}{2} \quad \text{if } t \in U_1 \tag{I.18}$$

$$|F_t(x_0-\delta) - F_{t_0}(x_0-\delta)| \le \frac{\epsilon}{2} \quad \text{if } t \in U_2. \tag{I.19}$$

Define $U_0 := U_1 \cap U_2$ and let $(t, x) \in T \times \mathbb{R}$ such that $t \in U_0$ and $|x - x_0| < \eta$. Then due to monotonicity in x we have

$$\begin{aligned} F_t(x) - F_{t_0}(x_0) \leq & F_t(x+\delta) - F_{t_0}(x_0) \\ = & F_t(x_0+\delta) - F_{t_0}(x_0+\delta) + F_{t_0}(x_0+\delta) - F_{t_0}(x_0) \end{aligned} \tag{I.20}$$

and

$$\begin{aligned} F_t(x) - F_{t_0}(x_0) \geq & F_t(x-\delta) - F_{t_0}(x_0) \\ = & F_t(x_0-\delta) - F_{t_0}(x_0-\delta) + F_{t_0}(x_0-\delta) - F_{t_0}(x_0). \end{aligned} \tag{I.21}$$

Combining (I.17) with (I.20) and (I.21) we obtain

$$|F_t(x) - F_{t_0}(x_0)| \leq \epsilon \tag{I.22}$$

and thus, joint continuity in (t_0, x_0). ■

Since processes with continuous sample paths are continuous in distribution, (I.14) (resp. (I.15)) forms a necessary condition on the marginals.

Theorem I.3.12. *Let $X = (X_t)_{t\in T}$ be a stochastic process with sample paths that belong almost surely to $C(T)$ and such that it has continuous marginals F_t for all $t \in T$. Then $U = (U_t)_{t\in T}$ defined by*

$$t \mapsto U_t := F_t(X_t) \tag{I.23}$$

is a copula process for X and almost surely continuous on T. Vice versa, if U is a copula process that is almost surely continuous on T and $F_t, t \in T$ are strictly increasing marginals between the points $F_t^{[-1]}(0+) := \lim_{x\downarrow 0} F_t^{[-1]}(x)$ and $F_t^{[-1]}(1)$, which are continuous in distribution, then $Y = (Y_t)_{t\in T}$ defined by

$$Y_t = F_t^{[-1]}(U_t) \tag{I.24}$$

is a random variable, which is almost surely in $C(T)$ with marginals F_t and underlying copula process U. Moreover, if T is a compact metrisable Hausdorff space, Y is measurable with respect to the Borel σ-algebra on the separable Banach space $C(T)$.

Proof. Observe, that since we are in the case of continuous marginals, the process $(U_t)_{t\in T}$ defined by $U_t = F_t(X_t)$ for all $t \in T$ is a copula process underlying X. Its continuity follows by Lemma I.3.11 and the continuity of $s \mapsto X_s$.

To show the second part, observe that $(t, x) \mapsto F_t^{[-1]}(x)$ is continuous in x, since the marginals $F_t, t \in T$ are strictly increasing between $F_t^{[-1]}(0+)$ and $F_t^{[-1]}(1)$. Hence Y is a random variable with values in $C(T)$ almost surely by Lemma I.3.11 and the continuity of $s \mapsto U_s$.

Its Borel measurability in the case that T is a compact metrisable Hausdorff space follows from Lemma I.3.8, as the point evaluations form a fundamental set with respect to the *weak**-topology on the dual of $C(T)$. ■

Remark I.3.13. In the framework of stochastic processes, the initial value X_0 is often chosen to be deterministic. Therefore it has neither a continuous nor strictly increasing distribution in the initial value. Possibly, for some processes, we still manage to define a continuous underlying copula (if the copula process has a limit from above in 0 which is uniformly distributed), but since this might be hard to check in general, it is reasonable to start the process a little bit later than in the origin. In fact, in some cases (like for the Brownian motion in Example I.3.14 below), it is possible to prove the nonexistence of a continuous version of the copula process on the whole real line.

Example I.3.14. For some $t_0 > 0$ let $X_t = B_t$ for $t \in [t_0, \infty)$ be a standard Brownian motion with sample paths in $C(\mathbb{R}_+)$.

$$U_t = F_t(B(t)) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{B(t)} e^{-\frac{x^2}{2t}}\, dx.$$

We note that the copula of a Brownian motion was investigated for instance in [41] in the framework of Markov copulas.

To see that we cannot find a continuous copula process for the Brownian motion starting in 0, we can argue as follows: As F_1 is a homeomorphism, we have that any version of $F_t(B_t) = F_1(B_t/\sqrt{t})$ is again of the form $F_t(\tilde{B}_t)$ for a continuous version $(\tilde{B}_t)_{t\geq 0}$ of the Brownian motion $(B_t)_{t\geq 0}$. So in order to find a continuous modification of the process $(F_t(B_t))_{t\geq 0}$, the limit $\lim_{t\to 0} F_t(B_t)$ must exist almost surely and be uniformly distributed. But we have that a Brownian motion is almost surely not 1/2-Hölder continuous in 0. In fact, for all $K > 0$ we have by applying Blumenthal's 0-1-law (cf. Example 21.16 in [26])

$$\mathbb{P}\left[\inf\{t > 0 : \frac{B_t}{\sqrt{t}} \geq K\} = 0\right] = 1$$

and therefore also

$$\mathbb{P}\left[\inf\{t > 0 : \frac{B_t}{\sqrt{t}} \leq -K\} = 0\right] = 1$$

Thus,

$$\limsup_{t\to 0} \frac{B_t}{\sqrt{t}} = \infty, \quad \liminf_{t\to 0} \frac{B_t}{\sqrt{t}} = -\infty.$$

But this implies that almost surely

$$\limsup_{t\to 0} F_t(B_t) = \limsup_{t\to 0} F_1(\frac{B_t}{\sqrt{t}}) = 1, \quad \liminf_{t\to 0} F_t(B_t) = \liminf_{t\to 0} F_1(\frac{B_t}{\sqrt{t}}) = 0,$$

which means that $(F_t(B_t))_{t\geq 0}$ cannot be continuous in 0.

We now move our attention to the regularity of paths induced by the copula construction. Therefore let now $T = I_1 \times ... \times I_d$ be an interval in $\mathbb{R}^d$ for some $d \in \mathbb{N}$. Recall that for a constant $\gamma > 0$ a function $f : T \to \mathbb{R}$ is called locally

γ-Hölder continuous, if for each $t \in T$ there is a neighbourhood $N(t)$ of t in T and a constant $K_t > 0$, such that for all $s, r \in N(t)$ we have

$$|f(s) - f(r)| \leq K_t |s - r|^\gamma.$$

For a nonnegative integer k, $\gamma \in (0, 1]$ and $m \in \mathbb{N}$, we introduce the Hölder spaces $C^{k,\gamma}(T; \mathbb{R}^m)$ to be the space of functions $f : T \to \mathbb{R}^m$ which are continuously differentiable up to order k and the kth derivative is locally γ-Hölder continuous.

Recall the following fact about locally Hölder continuous functions:

Lemma I.3.15. *Let $I_1, ..., I_m$ be intervals and $f = (f_1, ..., f_m) \in C^{k,\gamma}(T; \mathbb{R}^m)$ and $g \in C^{l,\eta}(I_1 \times ... \times I_m; \mathbb{R})$ such that $f(T) \subset I_1 \times ... \times I_m$. Then*

$$g \circ f \in \begin{cases} C^{0,\gamma\eta}(T; \mathbb{R}) & k = l = 0 \\ C^{k,\gamma}(T; \mathbb{R}) & k > l \\ C^{l,\eta}(T; \mathbb{R}) & l > k \\ C^{k,\min(\gamma,\eta)}(T; \mathbb{R}) & l = k \geq 1. \end{cases}$$

Proof. This is a special case of Theorem 4.3 in [30]. ■

As a consequence of Lemma I.3.15 we obtain the following immediately:

Corollary I.3.16. *Let $X \in C^{k,\gamma}(T; \mathbb{R})$ almost surely such that $(t, x) \mapsto F_{X_t}(x) \in C^{l,\delta}(T \times \mathbb{R}; \mathbb{R})$. Let U denote the associated copula process given by $U_t = F_t(X_t), t \in T$. Then almost surely*

$$U \in \begin{cases} C^{0,\gamma\eta}(T; \mathbb{R}) & k = l = 0 \\ C^{k,\gamma}(T; \mathbb{R}) & k > l \\ C^{l,\eta}(T; \mathbb{R}) & l > k \\ C^{k,\min(\gamma,\eta)}(T; \mathbb{R}) & l = k \geq 1. \end{cases}$$

For a copula process $U \in C^{k,\gamma}(T; \mathbb{R})$ almost surely and marginal cumulative distribution functions $(G_t)_{t \in T}$, such that $(t, u) \mapsto G_t^{[-1]}(u) \in C^{l,\delta}(T \times (0, 1); \mathbb{R})$, Y denotes the process given by $Y_t = G_t^{[-1]}(U_t)$. Then we have almost surely that

$$Y \in \begin{cases} C^{0,\gamma\eta}(T; \mathbb{R}) & k = l = 0 \\ C^{k,\gamma}(T; \mathbb{R}) & k > l \\ C^{l,\eta}(T; \mathbb{R}) & l > k \\ C^{k,\min(\gamma,\eta)}(T; \mathbb{R}) & l = k \geq 1. \end{cases}$$

By virtue of the previous Corollary I.3.16 we can determine the regularity of a copula process underlying a fractional Brownian motion:

Example I.3.17. Assume that $(U_t)_{t \in (t_0, \infty)}$ is a copula process underlying a fractional Brownian motion $(B_t^H)_{t \in [t_0, \infty)}$ for some $t_0 > 0$ with Hurst parameter $H \in (0, 1)$, that is, a centered Gaussian process with covariance function

$$\mathbb{E}[B_t^H B_s^H] = \frac{1}{2} \left(t^{2H} + s^{2H} + |t - s|^{2H} \right).$$

The process $(U_t)_{t\in(t_0,\infty)}$ has locally H-Hölder continuous paths. To see this, we just have to verify the local H-Hölder continuity of $(t,x)\mapsto \Phi_t^H(x)$ as stated in the Corollary I.3.16, where we denoted by Φ_t^H the cumulative distribution functions of B_t^H. We can estimate for $s,t\in[t_0,\infty)$ (with the constant $c=1/\sqrt{2\pi}$)

$$\begin{aligned}|\Phi_t^H(y)-\Phi_s^H(y)| = \left|\Phi_1^H\left(\frac{y}{t^H}\right)-\Phi_1^H\left(\frac{y}{s^H}\right)\right| = &\left|\int_{\min\left(\frac{y}{t^H},\frac{y}{s^H}\right)}^{\max\left(\frac{y}{t^H},\frac{y}{s^H}\right)} \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}dz\right|\\ \le& c|y||t^{-H}-s^{-H}|\\ \le& |y|\frac{c}{\min(t^H,s^H)^2}|t-s|^H\\ \le& \frac{c|y|}{t_0^{2H}}|t-s|^H.\end{aligned}$$

Analogously, for $x,y\in\mathbb{R}$ we get

$$\begin{aligned}|\Phi_t^H(x)-\Phi_t^H(y)| = \left|\Phi_1^H\left(\frac{x}{t^H}\right)-\Phi_1^H\left(\frac{y}{t^H}\right)\right| = &\left|\int_{\min\left(\frac{y}{t^H},\frac{x}{t^H}\right)}^{\max\left(\frac{y}{t^H},\frac{x}{t^H}\right)} \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}dz\right|\\ \le& c|x-y||t^{-H}|\\ \le& c|x-y||t_0^{-H}|.\end{aligned}$$

By the triangle inequality, we obtain the joint Hölder continuity

$$\begin{aligned}|\Phi_t^H(x)-\Phi_s^H(y)| &= |\Phi_t^H(x)-\Phi_t^H(y)|+|\Phi_t^H(y)-\Phi_s^H(y)|\\ &\le c\max\left(|t_0^{-H}|,\frac{|y|}{t_0^{2H}}\right)\left(|x-y|+|t-s|^H\right).\end{aligned}$$

Example I.3.18 (Exponential Marginals and fBM copula)**.** Several modelling situations (e.g., when modelling stochastic volatility, interest rates, etc. in financial mathematics) necessitate positive stochastic processes. It is simple to see that copula constructions might lead to good interpretable and alternative methods to model such processes since we are free to put any continuous family of marginals onto a Gaussian process (this was for example suggested in [45]).

As a simple example, take exponential marginals of the form

$$G_t(x) := \mathbb{I}_{x>0}\left(1-e^{-\frac{x}{t^H}}\right), \quad t\in[t_0,\infty), x\in\mathbb{R}$$

for some $t_0>0$, a (Hurst-)parameter $H=(0,1)$ corresponding to the copula process $(U_t)_{t\in T}$ of a fractional Brownian motion B^H (we take the parameter $\frac{1}{t^H}$ for the marginals to keep the same variance as the underlying fractional Brownian motion). By the smoothness of

$$G_t^{-1}(y) = -\log(1-y)t^h$$

we obtain that the transformed fractional Brownian motion

$$Y_t := G_t^{-1}\left(\Phi_t^H(B_t^H)\right) := -\log\left(1 - \int_{-\infty}^{B_t^H} \frac{e^{-\frac{z^2}{2t^{2H}}}}{\sqrt{2\pi}t^H} dz\right) t^H$$
$$= -\log\left(\int_{B_t^H}^{\infty} \frac{e^{-\frac{z^2}{2t^{2H}}}}{\sqrt{2\pi}t^H} dz\right) t^H$$

has underlying Gaussian copula U, is γ-Hölder continuous for all $\gamma < H$ and has exponential marginals (with parameters $\frac{1}{t^H}$).

In [19] it is argued empirically for lognormal marginals with a fractional Brownian motion copula for the stochastic volatility of asset prices. Our example shows that one can easily modify the marginals (to exponential, say, as in our example here), or other positively supported distributions, in so-called rough volatility models of asset prices. Moreover, the flexibility in the copula framework allows also to go beyond the specific dependency yielded by the copula induced by fractional Brownian motion.

I.3.2.3 Construction on Schauder Bases

In this section, we will characterise copula-constructed processes for random variables in Banach spaces with a Schauder basis. This includes $L^p([0,1])$-spaces (with the Haar wavelets as Schauder basis), $C([0,1])$ (with the original Schauder basis), l^p-sequence spaces, and therefore in particular, all separable Hilbert spaces (with an orthonormal basis as Schauder basis). For a detailed account of the theory of bases in Banach spaces, we refer to [21].

Definition I.3.19. A sequence $(e_n)_{n\in\mathbb{N}} \subseteq V$ of linearly independent vectors is called a basis of a locally convex Hausdorff space V, if for all $v \in V$ there is a unique sequence $a_n(v)$, such that

$$v = \sum_{n\in\mathbb{N}} a_n(v) e_n,$$

where the series converges with respect to the locally convex topology on V. A basis of V^* is called *weak**-basis of V^*, if it is a basis with respect to the *weak**-topology. If V is a Banach space and $v \mapsto a_n(v)$ is continuous with respect to the norm topology for all $n \in \mathbb{N}$, we call $(e_n)_{n\in\mathbb{N}} \subseteq V$ a Schauder basis.

The continuity of the function $(a_n)_{n\in\mathbb{N}}$ is automatically satisfied if V is a separable Banach space (see Theorem 3.1. in [42]). Note, that every Banach space that possesses a basis is separable. However, for a separable Banach space, the existence of a basis cannot be guaranteed, due to the counterexample by Enflo in [16]. For a Banach space with Schauder basis we can verify, that the corresponding coefficient functions are always contained in the topological dual:

Lemma I.3.20. *Let V be a Banach space with Schauder basis $(e_n)_{n\in\mathbb{N}}$ and coefficient functions $(a_n)_{n\in\mathbb{N}}$. Then $\{a_n : n \in \mathbb{N}\} \subset V^*$ and for $m, n \in \mathbb{N}$ we have*

$$a_n(e_m) = \begin{cases} 0 & m \neq n \\ 1 & m = n. \end{cases}$$

Proof. Linearity of the coefficients is clear due to uniqueness of the representation. Moreover for the same reason, $a_n(e_n) = 1$ and $a_m(e_n) = 0$ gives a valid series representation of e_n for all $n \in \mathbb{N}$ and by uniqueness of this, the assertion follows. ∎

That the coefficient functionals are linearly independent and separate the points is a consequence of the following theorem:

Theorem I.3.21. *Let V be a Banach space. A sequence $(a_n)_{n\in\mathbb{N}}$ is a weak* Schauder basis of V^* if and only if there exists a Schauder basis $(e_n)_{n\in\mathbb{N}}$ of V that has $(a_n)_{n\in\mathbb{N}}$ as its coefficient functionals. The coefficient functionals for the basis $(a_n)_{n\in\mathbb{N}}$ are then given by the bidual elements $(\iota_{e_n})_{n\in\mathbb{N}}$, where $\iota_v(v^*) = v^*(v)$.*

Proof. See Theorem 14.1. in [42]. ∎

Observe that for a Banach space with Schauder basis, the corresponding set

$$M = \{a_n : n \in \mathbb{N}\}$$

of coefficient, marginals satisfy all the conditions of Definition I.3.1. Embedding (I.6) reads now

$$v \mapsto (a_n(v))_{n\in\mathbb{N}}. \tag{I.25}$$

Theorem I.3.22. *Let V be a Banach space with Schauder basis $(e_n)_{n\in\mathbb{N}}$ and coefficient functions $(a_n)_{n\in\mathbb{N}}$. The following are equivalent*

(i) *A sequence $(a_n)_{n\in\mathbb{N}}$ is in the range of the embedding* (I.25)*;*

(ii) *$\sum_{n=1}^{\infty} a_n e_n \in V$ is a convergent series in the norm topology;*

(iii) *$\sup_{N\in\mathbb{N}} \|\sum_{n=1}^{N} a_n e_n\| < \infty$.*

Proof. See Theorem 4.13 in [21]. ∎

Thus, for the checkup of the Construction 1 we have the following:

Corollary I.3.23. *Let V be a Banach space with Schauder basis $(e_n)_{n\in\mathbb{N}}$ and $(X_n)_{n\in\mathbb{N}}$ be a stochastic process. Then the following are equivalent:*

(i) *$X = \sum_{n=1}^{\infty} X_n e_n$ is a Borel measurable random variable in V;*

(ii) *$\sup_{N\in\mathbb{N}} \|\sum_{n=1}^{N} X_n e_n\| < \infty \quad \mathbb{P}-$ almost surely.*

Proof. This follows directly from Theorem I.3.22 and Lemma I.3.8. ∎

Let us now describe how we can construct a Banach space probability measure with predescribed dependence structure and marginals for the basis components:

Construction 3.

(i) V a Banach space with Schauder basis $(e_n)_{n\in\mathbb{N}} \subseteq V$;

(ii) Choose a copula measure C on $\mathbb{R}^{\mathbb{N}}$ (which models the dependency between basis elements) and marginals $(\mu_n)_{n\in\mathbb{N}}$. Merge them to a law of a random sequence $(X_n)_{n\in\mathbb{N}}$ taking values in $\mathbb{R}^{\mathbb{N}}$ via Sklar's theorem I.2.3.

(iii) Define $X := \sum_{n\in\mathbb{N}} X_n e_n$.

(iv) Check if this sum converges in V almost surely (corresponding to Corollary I.3.23).

For the verification of (iv) we obtain conditions on the moments of the marginals.

Corollary I.3.24. *Let X be given as in Construction 3(iii). Then $X \in L^1(\Omega; V, \mathcal{B}(V))$ if the marginals have finite first moment and*

$$\sum_{n=1}^{\infty} \mathbb{E}[|X_n|] < \infty. \tag{I.26}$$

Proof. This follows immediately by using Corollary I.3.23 and the triangular inequality. ∎

In the case of sequence spaces, we obtain even sufficient and necessary conditions to construct laws with finite moments of a certain order. Denote

$$l^p := \left\{ (x_n)_{n\in\mathbb{N}} \subset \mathbb{R}^{\mathbb{N}} : \|(x_n)_{n\in\mathbb{N}}\|_p := \left(\sum_{n=1}^{\infty} |x_n|^p \right)^{\frac{1}{p}} < \infty \right\}$$

for some $p \in [1, \infty)$.

Corollary I.3.25. *Let $V = l^p$ and X be given as in Construction 3(iii). Then $X \in L^p(\Omega; l^p, \mathcal{B}(l^p))$ if and only if the marginals have finite pth moment and*

$$\sum_{n=1}^{\infty} \mathbb{E}[|X_n|^p] < \infty. \tag{I.27}$$

Proof. The standard basis $(\delta_n)_{n\in\mathbb{N}}$ is the sequence which has components equal to zero everywhere, except on the n'th entry, where it is 1. This defines a Schauder basis on l^p with coefficient functionals δ_i^* given by $\delta_i^*((x_n)_{n\in\mathbb{N}}) = x_i$, since

$$(x_n)_{n\in\mathbb{N}} = \sum_{i=1}^{\infty} x_i \delta_i = \sum_{i=1}^{\infty} \delta_i^*((x_n)_{n\in\mathbb{N}}) \delta_i.$$

Thus,

$$\sup_{N\in\mathbb{N}} \left\| \sum_{n=1}^{N} X_n \delta_n \right\|_p^p = \sup_{N\in\mathbb{N}} \sum_{n=1}^{N} |X_n|^p = \sum_{n=1}^{\infty} |X_n|^p.$$

This implies

$$\mathbb{E}[\|X\|_p^p] = \sum_{n=1}^{\infty} \mathbb{E}[|X_n|^p] < \infty.$$

The assertion follows. ∎

Remark I.3.26. Observe that Corollary I.3.25 generalises Corollary 4.3 in [20], where the case of separable Hilbert spaces, that is, $p = 2$, was considered and the notion of a Schauder basis is reduced to the concept of orthonormal bases.

The results derived above just impose conditions on the marginals, which makes them useful from a practical viewpoint. Still, the concept of copulas for random variables in the space $L^p(\Omega; l^p)$, or equivalent for laws in the Wasserstein space $\mathcal{W}_p(l^p)$ (see (I.28) below) is characterized completely by Corollary I.3.25. We will obtain another characterization of copulas as underlying solutions to certain restricted optimization problems in these Wasserstein spaces in the next section.

I.4 Approximation and Estimation

The previous section suggested that copula theory is well suited for the spaces $\mathcal{L}^p(T) := \mathcal{L}^p(T, \mathcal{B}(T), \mu; \mathbb{R})$ for a finite Borel measure μ and $T \subset \mathbb{R}^d$ a compact interval and the sequence spaces l^p, due to simple moment criteria to overcome the construction problem. In this section, we will provide distance estimates of random variables in these spaces in terms of their copula and marginals separately.

Hereafter we shorten the notation as follows: For a random variable X with values in E (where E equals l^p or $\mathcal{L}^p(T)$ respectively) we denote the operators

$$F_X(x)_n := F_{X_n}(x_n), \;\; n \in \mathbb{N} \quad (F_X(f)(t) := F_{X_t}(f(t)), \;\; t \in T \;\; \text{resp.})$$

and

$$F_X^{[-1]}(x)_n := F_{X_n}^{[-1]}(x_n), \;\; n \in \mathbb{N} \quad (F_X^{[-1]}(f)(t) := F_{X_t}^{[-1]}(f(t)), \;\; t \in T \;\; \text{resp.})$$

for all $x \in l^p$ (and $f \in \mathcal{L}^p(T)$ respectively). Moreover, we use the notation U^X for the underlying copula process of X. We will for convenience switch between the spaces $\mathcal{L}^p(T)$ and $L^p(T)$ whenever there is no confusion. If we say that an $[X] \in L^p(T)$ has underlying copula process $U \in \mathcal{L}^p(T)$, we mean that there is a representative $X \in \mathcal{L}^p(T)$ of the corresponding element, that has this path copula. We will also drop equivalence class notation from time to time, to ease the writing (especially, when we work with Wasserstein spaces in the next section) and just refer to the representative X, no matter if we mean the equivalence class or the actual stochastic process.

I.4.1 Copulas and Wasserstein Spaces

In this subsection, we characterise copulas for measures in Wasserstein spaces. For two laws ν^1 and ν^2 on E we write $\rho <_{\nu^1}^{\nu^2}$ for a law ρ on $E \times E$ that has marginal distributions ν^1 and ν^2, that is, ρ is a coupling of ν^1 and ν^2. Recall that the p-Wasserstein space over a separable Banach space E is a complete separable metric space (see e.g. [44]) given by

$$\mathcal{W}_p(E) := \left\{ \nu : \nu \text{ is a Borel law on } E, \int_E \|x\|_E^p \nu(dx) < \infty \right\} \tag{I.28}$$

equipped with the metric (in the case that we interpret $E = L^p(T)$ instead of $E = \mathcal{L}^p(T)$)

$$d(\nu^1, \nu^2) =: \mathbb{W}_p(\nu^1, \nu^2) := \inf_{\rho <_{\nu^1}^{\nu^2}} \left(\int_{E \times E} \|x - y\|_E^p \rho(dxdy) \right)^{\frac{1}{p}}.$$

If there are two random variables $X \sim \nu^1$ and $Y \sim \nu^2$, we also say that (X, Y) is a coupling and write $\mathbb{W}_p(\nu^1, \nu^2) = \mathbb{W}_p(X, Y)$. If $E = \mathbb{R}$, we have the following closed form of the Wasserstein distance (see e.g. Theorem 3.1.2 in [37]):

$$\mathbb{W}_p^p(X, Y) = \int_{[0,1]} |F_X^{[-1]}(u) - F_Y^{[-1]}(u)|^p du. \tag{I.29}$$

The next theorem is an immediate implication of the results in [37]. As the argument was not written down in infinite dimensions and the notion of copulas was not used directly, we provide a proof in Appendix I.B for convenience.

Theorem I.4.1. *Let X, Y be random variables in l^p (in $\mathcal{L}^p(T)$ respectively) for some $p \in \mathbb{N}$. Then the following are equivalent:*

(i) *X and Y share the same underlying basis copula (path copula respectively) C;*

(ii) *$(F_X^{[-1]}(U), F_Y^{[-1]}(U))$ is an optimal coupling of X and Y, where $U \sim C$;*

(iii) *The Wasserstein distance between X and Y is given by*

$$\mathbb{W}_p^p(X, Y) = \sum_{n \in \mathbb{N}} \mathbb{W}_p^p(X_n, Y_n)$$

$$(\textit{respectively } \mathbb{W}_p^p(X, Y) = \int_T \mathbb{W}_p^p(X_t, Y_t)\mu(dt)).$$

In particular, if one of the above holds we have

$$\sum_{n \in \mathbb{N}} \mathbb{W}_p^p(X_n, Y_n) = \|F_X^{[-1]}(U) - F_Y^{[-1]}(U)\|_{L^p(\Omega; l^p)}^p$$

$$(\textit{respectively } \int_T \mathbb{W}_p^p(X_t, Y_t)\mu(dt) = \|F_X^{[-1]}(U) - F_Y^{[-1]}(U)\|_{L^p(\Omega; L^p(T))}^p).$$

Remark I.4.2. Observe that the implications $(ii) \Rightarrow (i)$ and $(iii) \Rightarrow (i)$ in Theorem I.4.1 must be interpreted in the sense that there is always a representative of the equivalence classes that possesses the same path copula.

Remark I.4.3. The assertion of Theorem I.4.1 does not hold for the q-Wasserstein distance over l^p ($\mathcal{L}^p(T)$ respectively) if $q \neq p$, as it was shown in [1] for the finite-dimensional case.

Remark I.4.4. Theorem I.4.1 is useful because the one-dimensional Wasserstein distance has a closed form given by (I.29). This expression can oftentimes be estimated rather well from above (see for instance chapter 4.7 in [35] for a discussion of the convergence of empirical measures).

Remark I.4.5. The copula construction effectively solves the following optimization problem for $E = l^p$ or $E = \mathcal{L}^p(T)$ respectively:

$$(P) = \begin{cases} \min_{\nu \in \mathcal{W}_p(E)} \mathbb{W}_p(\nu^0, \nu) \\ \text{s.t.} \quad \nu \text{ has marginals } (\nu_n)_{n\in\mathbb{N}} \text{ (respectively } (\nu_t)_{t\in T}) \end{cases}$$

for any family of marginals $(\nu_n)_{n\in\mathbb{N}}$ (respectively $(\nu_t)_{t\in T}$), and the optimal value is given by $\nu = ((F_{\nu_n}^{[-1]})_{n\in\mathbb{N}})_* C$ (respectively $\nu = ((F_{\nu_t}^{[-1]})_{t\in T})_* C$) for the underlying copula measure C of ν^0.

Moreover, Theorem I.4.1 implies the following.

Corollary I.4.6. *Let X, Y be stochastic processes with values in $\mathcal{L}^2(T)$. If $(e_n)_{n\in\mathbb{N}}$ is an orthonormal basis in $L^2(T)$, then the following are equivalent:*

(a) $\int_T \mathbb{W}_2^2(X_t, Y_t)dt = \sum_{n=1}^{\infty} \mathbb{W}_2^2(\langle X, e_n\rangle_{L^2(T)}, \langle Y, e_n\rangle_{L^2(T)})$

(b) *X and Y have the same basis copula if and only if X and Y have the same path copula.*

I.4.2 A Robustness Inequality in $\mathcal{L}^p(T)$

In order to derive a distance estimate between random variables in $\mathcal{L}^p(T)$ for a finite Borel measure μ and $T \subset \mathbb{R}^d$ compact, based on the copula and the marginals separately, we impose a smoothness and tail-assumption on marginals of the distribution function.

Assumption 1. For all t, the marginals F_t are continuously differentiable and strictly increasing on $(F_t^{[-1]}(0+), F_t^{[-1]}(1))$. Moreover, assume that for the corresponding densities f_t there is a measurable

$$g : T \times \mathbb{R} \to \mathbb{R},\ g(t, x) =: g_t(x)$$

such that each g_t is ultimately monotone (see, e.g. [6]), that is, restricted on $(m_t + x_0^t, \infty)$ and restricted on $(-\infty, m_t - x_0^t)$ it is monotone for some $x_0^t \in \mathbb{R}_+$, $m_t \in \mathbb{R}$, with g_t bounded away from 0 on $[m_t - x_0^t, m_t + x_0^t]$ by some $\lambda > 0$

(independent of t) and $f_t(x) \geq g_t(x) > 0$ for all $x \in (F_{Y_t}^{[-1]}(0), F_{Y_t}^{[-1]}(1))$ and for all $t \in T$, and there is an $0 < \beta \leq 1$ such that

$$\int_T \int_{\mathbb{R}} \frac{f_t(x)}{g_t^{\beta}(x)} dx dt < \infty. \tag{I.30}$$

Remark I.4.7. If a density function $(x, t) \mapsto f_t(x)$ is continuous and ultimately monotone, that is, there are $x_0^t > 0$, $m_t \in \mathbb{R}$ such that f_t is monotone on $[m_t - x_0^t, m_t + x_0^t]^c$, the best candidate for the choice of g in Assumption 1 is f_t itself.

Observe that Assumption 1 is satisfied for Gaussian marginals:

Example I.4.8. Assume that $Y = W$ is a zero mean continuous Gaussian process. Clearly, the densities f_t of W_t are ultimately monotone and and we can choose $x_0^t = 0$ and $g_t = f_t$. Then Condition I.30 holds, as for all $\beta \in (0, 1)$

$$\begin{aligned}\int_T \int_{\mathbb{R}} f_t^{1-\beta}(x) dx dt = \int_T \frac{\mathbb{E}[e^{\beta \frac{W_t^2}{2\sigma_t^2}}]}{(\sqrt{2\pi}\sigma_t)^{-\beta}} dt =& (\sqrt{2\pi})^{\beta} \int_T \sigma_t^{\beta} dt \mathbb{E}[e^{\beta \frac{Z^2}{2}}] \\ =& \frac{(\sqrt{2\pi})^{\beta}}{\sqrt{1-\beta}} \int_T \sigma_t^{\beta} dt,\end{aligned}$$

where Z is a standard normally distributed random variable. Thus, since $t \mapsto \sigma_t^{\beta}$ is continuous, it is integrable over T and Assumption 1 holds.

Another example, for which Assumption 1 holds, is the following class of heavy-tailed marginals.

Example I.4.9. (Regularly varying marginals) A measurable function $h : [z_0, \infty) \to (0, \infty)$ for some $z_0 > 0$ is regularly varying with tail index $\alpha \in \mathbb{R}$, if

$$\lim_{x \to \infty} \frac{h(tx)}{h(x)} = t^{\alpha} \quad \forall t > 0$$

We write $h \in \mathcal{R}(\alpha)$. If $\alpha = 0$, h is called slowly varying. A one-dimensional law given by its cumulative distribution function F is said to have regularly varying tails, if the survival function $\bar{F} := 1 - F$ is regularly varying and for convenience, supported on some corresponding interval $[z_0, \infty) \subset (0, \infty)$.

Let Y be a càdlàg stochastic process, such that its marginals $(F_t)_{t \in T}$ are continuously differentiable, strictly increasing and supported on $[z_0, \infty)$ for some $z_0 > 0$ and regularly varying with tail index $-\alpha_t$ for $\alpha_t > 0$ where we assume that $t \mapsto \alpha_t$ is continuous. Moreover, assume that the densities f_t are ultimately monotone on $[z_0, y_0^t]^c$ for some $y_0^t \geq z_0$, $f_t(z_0) > 0$ and jointly continuous in x and t. This enables us to use the monotone density theorem (cf. Theorem 1.7.2 in [6]) to conclude that $f_t \in \mathcal{R}(-(\alpha_t + 1))$. Hence, there are slowly varying functions $l_t : [z_0, \infty) \to (0, \infty)$ such that

$$f_t(x) = x^{-(1+\alpha_t)} l_t(x).$$

For convenience, let us assume that $(t,x) \mapsto l_t(x)$ is bounded. By choosing $\beta < \min_{t\in T} \frac{\alpha_t}{1+\alpha_t}$ we obtain

$$l_t^{1-\beta}(x)x^{-(1-\beta)(1+\alpha_t)} \in \mathcal{R}(-(1-\beta)(1+\alpha_t))$$

such that $-(1-\beta)(1+\alpha_t) < -1$ and by Karamata's theorem (cf. Proposition 1.5.10 in [6]) we can find $x_0^t > y_0^t$ and some $\delta > 0$ such that

$$\int_{x_0^t}^{\infty} \frac{f_t(x)}{f_t^{\beta}(x)} dx = \int_{x_0^t}^{\infty} f_t^{1-\beta}(x)dx = \int_{x_0^t}^{\infty} l_t^{1-\beta}(x)x^{-(1-\beta)(1+\alpha_t)}dx$$
$$\le (1+\delta)l_t^{1-\beta}(x_0^t)(x_0^t)^{-(1-\beta)(1+\alpha_t)+1} < \infty.$$

Assume moreover that we can choose $t \mapsto x_0^t$ to be continuous (this is possible for instance if all l_t's are supported on a compact domain) and hence, since $f_t(x) > 0$ for all $x \in [z_0, \infty)$ we have for

$$\lambda := \min_{t\in T} \min_{x\in[z_0,x_0^t]} f_t(x) > 0$$

that each f_t is bounded away from 0 on $[z_0, x_0^t]$ by this $\lambda > 0$. Moreover, it holds

$$\int_T \int_{x_0^t}^{\infty} \frac{f_t(x)}{f_t^{\beta}(x)} dxdt \le \int_T (1+\delta)l_t^{1-\beta}(x_0^t)(x_0^t)^{-(1-\beta)(1+\alpha_t)+1}dt < \infty$$

by the continuity of $t \mapsto x_0^t$. Thus, (I.30) in Assumption 1 is valid with

$$g_t(x) := \begin{cases} 0 & x < z_0 \\ \lambda & z_0 \le x < x_0^t \\ f_t(x) & x \ge x_0^t. \end{cases} \tag{I.31}$$

The next Theorem gives an idea about the robustness of the copula construction. The proof can be found in Appendix I.B.

Theorem I.4.10. *Let $X = (F_{X_t}^{[-1]}(U_t^X))_{t\in T}$ and $Y = (F_{Y_t}^{[-1]}(U_t^Y))_{t\in T}$ be càdlàg stochastic processes, such that $[X] \in L^p(\Omega\times T)$ for some $p \ge 1$, $[Y] \in L^{p+\epsilon}(\Omega\times T)$ for some $\epsilon > 0$ and let the marginals F_Y of Y satisfy Assumption 1. Then for all $q \ge 1$ there are constants $K := K(\beta, q, p, \epsilon, F_Y)$ and $\rho := \rho(\beta, q, p, \epsilon)$ such that*

$$\|X - Y\|_{L^p(\Omega\times T)} \le \|\mathbb{W}_p(F_{X_\cdot}, F_{Y_\cdot})\|_{L^p(T)} + K\|U^X - U^Y\|_{L^q(\Omega\times T)}^{\rho} \tag{I.32}$$

and

$$\mathbb{W}_p(X,Y) \le \|\mathbb{W}_p(F_{X_\cdot}, F_{Y_\cdot})\|_{L^p(T)} + K\mathbb{W}_q(U^X, U^Y)^{\rho}. \tag{I.33}$$

The constants are given explicitly by

$$\rho := \frac{\epsilon q\beta}{p(p+\epsilon)(q+\beta) - pq\beta} \tag{I.34}$$

and

$$K := \left(\lambda^{-\beta}\int_T \mathbb{I}_{(0,\infty)}(x_0^t)dt + 2\|g_t^{-\beta}(Y_t)\|_{L^1(\Omega\times T)}\right)^{\frac{\rho}{\beta}} (2\|Y\|_{L^{p+\epsilon}(\Omega\times T)})^{(1-\rho)}. \tag{I.35}$$

Remark I.4.11. Although the marginals of Y must fulfil Assumption 1, the marginals of X can be chosen more freely and neither have to be absolutely continuous nor must satisfy a tail condition. For instance, this allows approximating a smooth law of Y with discrete marginal measures (e.g. empirical measures).

The following Theorem is useful if the copula processes stem from other processes, like Gaussian or elliptical copulas. Its proof can be found in Appendix I.B.

Theorem I.4.12. *Let $F_{\tilde{Y}}, F_{\tilde{X}}$ be marginals with finite qth moment for $q \geq 1$ and define $\tilde{X} := F_{\tilde{X}}^{[-1]}(U^X)$ and $\tilde{Y} := F_{\tilde{Y}}^{[-1]}(U^Y)$. Assume $F_{\tilde{Y}_t}$ is absolutely continuous, strictly increasing, and the corresponding density function is bounded, that is,*

$$\|f_{\tilde{Y}}\|_\infty := \sup_{t\in T, x\in\mathbb{R}} |f_{\tilde{Y}_t}(x)| < \infty.$$

Then

$$\|U^X - U^Y\|_{L^q(T\times\Omega)} \leq \|f_{\tilde{Y}}\|_\infty \left(\|\tilde{X} - \tilde{Y}\|_{L^q(T\times\Omega)} + \|\mathbb{W}_q^q(F_{\tilde{X}_\cdot}, F_{\tilde{Y}_\cdot})\|_{L^1(T)}^{\frac{1}{q}}\right)$$

and

$$\mathbb{W}_q(U^X, U^Y) \leq \|f_{\tilde{Y}}\|_\infty \left(\mathbb{W}_q(\tilde{X}, \tilde{Y}) + \|\mathbb{W}_q^q(F_{\tilde{X}_\cdot}, F_{\tilde{Y}_\cdot})\|_{L^1(T)}^{\frac{1}{q}}\right).$$

In particular,

$$\|U^X - U^Y\|_{L^q(T\times\Omega)} \leq 2\|f_{\tilde{Y}}\|_\infty \|\tilde{X} - \tilde{Y}\|_{L^q(T\times\Omega)}$$

and

$$\mathbb{W}_q(U^X, U^Y) \leq 2\|f_{\tilde{Y}}\|_\infty \mathbb{W}_q(\tilde{X}, \tilde{Y}).$$

Example I.4.13. Assume that U^Y is an elliptical copula corresponding to an elliptical random variable $\tilde{Y}$ in $L^2(T)$, that is,

$$\tilde{Y} = SV \tag{I.36}$$

for some positive, real-valued random variable S with finite second moment and a Gaussian process $V \sim \mathcal{N}(0, C)$, independent of S (see [7] for the exact description and the relation to finite-dimensional elliptical distributions). First, observe that without loss of generality we can assume $V_t \sim \mathcal{N}(0,1)$ since the process $(\frac{\tilde{Y}_t}{\mathbb{E}[|\tilde{Y}_t|^2]})_{t\in T}$ has by Lemma I.B.1 the same copula as $\tilde{Y}$. If S has a finite inverse moment, then $\tilde{Y}_t$ has for each t a bounded density, since by the formula for the density of two independent products, we get

$$f_{\tilde{Y}_t}(z) = \int_{-\infty}^{\infty} f_S(x) f_{V_t}(\frac{z}{x}) \frac{1}{|x|} dx \leq \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \frac{f_S(x)}{x} dx = \frac{\mathbb{E}[S^{-1}]}{\sqrt{2\pi}}.$$

Hence, for any other copula process $U^{\tilde{X}}$ corresponding to another process $\tilde{X}$ in $L^2(\Omega \times T)$, we have by Theorem I.4.12 that

$$\|U^{\tilde{Y}} - U^{\tilde{X}}\|_{L^2(\Omega\times T)} \leq \sqrt{\frac{2}{\pi}} \mathbb{E}[S^{-1}] \|\tilde{Y} - \tilde{X}\|_{L^2(\Omega\times T)}.$$

Copulas may be suitable to capture tail behaviour in functional data. This can be seen by the following example, combining the last ones.

Example I.4.14 (Approximating Pareto marginals on an elliptical copula). Assume that a process Y given by $Y_t := F_t^{[-1]}(U_t)$ where U is a copula process corresponding to an elliptical process $\tilde{Y}$ given by (I.36) and the marginals F_Y are regularly varying as in Example I.4.9. More specifically we can take F_Y to follow Pareto marginals, that is,

$$l_t(x) = \begin{cases} \alpha_t x_{\min}^{\alpha_t} & x \geq x_{min} \\ 0 & x \leq x_{min} \end{cases}$$

for some constant $x_{min} > 0$, such that f_t are the densities of a Pareto distribution $Par(x_{min}, \alpha_t)$, where $t \mapsto \alpha_t$ is assumed to be continuous. Assume now $\alpha_t > 2+\gamma$ for some $\gamma > 0$. Then Y takes values in $\mathcal{L}^2(T)$.

Consider a situation in which we can approximate the marginal function F_Y by another marginal function F_n (for example by empirical cumulative distribution functions). The underlying elliptical process is in the $L^2(T)$-norm best approximated over all processes with n dimensional spectral decomposition by the projection

$$\tilde{Y}^n = \sum_{i=1}^{n} Z_i e_i$$

of the first n principal components in the corresponding Karhunen–Loève expansion

$$\tilde{Y} = \sum_{i=1}^{\infty} Z_i e_i.$$

Here $(e_i)_{i\in\mathbb{N}}$ is an orthonormal basis of eigenvectors of the covariance operator of $\tilde{Y}$, where the corresponding eigenvalues $(\lambda_i)_{i\in\mathbb{N}}$ are ordered decreasingly (see for instance Theorem (1.5) in [9] for a proof and [7] more optimality properties of the principal components for elliptical processes). Then

$$\|\tilde{Y}_n - \tilde{Y}\|^2_{L^2(\Omega\times T)} = \sum_{i=n+1}^{\infty} \lambda_i. \quad \text{(I.37)}$$

Assume that U^n is the path copula process underlying $\tilde{Y}^n$. Let us investigate how well $Y^n = F_n^{[-1]}(U^n)$ approximates Y. We can choose $p = 1$, $\epsilon = 1$, $q = 2$, $\beta = \frac{2}{3}$ and hence $\rho = \frac{1}{3}$, $m_t = x_{min}$ (using the notation of Theorem I.4.10) and $x_0^t = 0$. Thus, by (I.32)

$$\mathbb{E}[\|Y - Y^n\|_{L^1(T)}] \leq \|\mathbb{W}_1(F_{Y.}, F_{n.})\|_{L^1(T)} + K\|U^Y - U^n\|^{\frac{1}{3}}_{L^2(\Omega\times T)}. \tag{I.38}$$

Since in this case

$$\mathbb{E}\left[f_t^{-\frac{2}{3}}(Y_t)\right] = \int_{x_{min}}^{\infty} \left(x^{-(\alpha_t+1)}\alpha_t x_{\min}^{\alpha_t}\right)^{\frac{1}{3}} dx = \frac{3\alpha_t^{\frac{1}{3}}}{\alpha_t - 2} x_{min}^{\frac{2}{3}} \leq \frac{3}{\gamma}\alpha_t^{\frac{1}{3}} x_{\min}^{\frac{2}{3}}$$

and

$$\|Y\|^{p+\epsilon}_{L^{p+\epsilon}(\Omega\times T)} = \int_T \mathbb{E}[Y_t^2]dt = \int_T \frac{\alpha_t x_{min}^2}{\alpha_t - 2} dt \leq \frac{x_{min}^2}{\gamma}\int_T \alpha_t dt$$

we get by (I.35)

$$K = \left(\frac{6}{\gamma} x_{\min}^{\frac{2}{3}} \int_T \alpha_t^{\frac{1}{3}} dt\right)^{\frac{1}{2}} \left(\frac{2}{\gamma} x_{min}^2 \int_T \alpha_t dt\right)^{\frac{2}{3}}. \tag{I.39}$$

Thus, using also Example I.4.13 and combining (I.37), (I.38) and (I.39) we obtain

$$\begin{aligned} &\|Y - Y^n\|_{L^1(T\times\Omega)} \\ \leq &\|\mathbb{W}_1(F_{Y.}, F_{n.})\|_{L^1(T)} + K\left(\sqrt{\frac{2}{\pi}}\mathbb{E}[S^{-1}]\right)^{\frac{1}{3}} \left(\sum_{i=n+1}^{\infty} \lambda_i\right)^{\frac{1}{6}} \end{aligned}$$

and therefore, the convergence rate is $\frac{1}{3}$ of the convergence rate of the principal components and the rate of convergence induced by the Wasserstein distance of the marginals, which depends on the respective approximation technique for the marginals.

I.4.3 Semiparametric Estimation of Nonparanormal Processes

We show how a consistent estimator for the copula model can be derived under the assumption that the unknown law of a stochastic process $X = (X_t)_{t\in T}$ with values in $\mathcal{L}^{2+\epsilon}(T)$ for some $\epsilon > 0$ has an underlying Gaussian copula process U^X. No parametric model on the marginals is assumed (in finite dimensions, this was termed *nonparanormal* in [29]). Although U^X corresponds to a Gaussian process $Y \sim \mathcal{N}(0, C)$ for some covariance operator C, the latter is not unique, since we can still impose different Gaussian marginals without changing the copula. Therefore, we fix in this section the unique choice C such that $Y_t \sim \mathcal{N}(0, 1)$. Assume throughout that we have n i.i.d copies

$$X_1, ..., X_n$$

of X.

I.4.3.1 Estimation of the Underlying Rank Correlation

We first estimate the underlying covariance such that

$$C = \sum_{j=1}^{\infty} \lambda_j e_j^{\otimes 2},$$

where $(\lambda_j)_{j\in\mathbb{N}}$ denote the eigenvalues of C in decreasing order and $(e_j)_{j\in\mathbb{N}}$ the corresponding eigenvectors. We follow the well known approach from finite dimensions by inferring on the rank correlation structure. Define Kendalls τ rank correlation function by

$$\tau(s,t) = \mathbb{E}[sign(X(s) - \tilde{X}(s))sign(X(t) - \tilde{X}(t))]$$

for an independent copy $\tilde{X}$ of X. In particular τ is the covariance kernel of the centered process

$$sign(X - \tilde{X}) = (sign(X(t) - \tilde{X}(t)))_{t\in T}$$

in $\mathcal{L}^p(T)$. As $sign(X_2 - X_1), sign(X_4 - X_3), ..., sign(X_n - X_{n-1})$ are i.i.d. samples of $sign(X - \tilde{X})$ (assuming w.l.o.g. n to be even), we obtain that the empirical covariance function

$$\hat{\tau}_n(s,t) = \frac{2}{n}\sum_{i=1}^{n/2} sign(X_{2i}(s) - X_{2i-1}(s))sign(X_{2i}(t) - X_{2i-1}(t)) \qquad \text{(I.40)}$$

is a consistent and asymptotically normal estimator of τ (see e.g. [22]).

The covariance function c of the underlying Gaussian process corresponds one-to-one to Kendalls τ (see for instance Theorem 3.1 in [17]) via the relation

$$c(s,t) = \sin(\frac{\pi}{2}\tau(s,t)). \qquad \text{(I.41)}$$

We define the plug-in estimator of the covariance kernel

$$\hat{c}_n(s,t) = \sin(\frac{\pi}{2}\hat{\tau}_n(s,t)). \qquad \text{(I.42)}$$

Observe that for any kernel c the integral operator

$$T_c f(t) := \int_T c(t,s)f(s)ds \quad f \in L^2(T),\, t \in T$$

is of special interest, since we have $T_c = C$ for the corresponding covariance operator. Let $(\tilde{\lambda}_{jn})_{n\in\mathbb{N}}$ and $(e_{jn})_{n\in\mathbb{N}}$ be the spectral decomposition of $T_{\hat{c}_n}$. The operator $T_{\hat{c}_n}$ might not be positive definite, which is why we define $\lambda_{jn} = \min(0, \tilde{\lambda}_{jn})$ for $j, n \in \mathbb{N}$ and set

$$\hat{C}_n := \sum_{j=1}^{\infty} \lambda_{jn} e_{jn}^{\otimes 2}.$$

The following law of large numbers is proved in Appendix I.B.

Theorem I.4.15. *We have*

$$\mathbb{E}[\|C - \hat{C}_n\|_{HS}^2] \leq \frac{\pi^2 \mu(T)^2}{n}$$

and almost surely

$$\limsup_{n\to\infty} \sqrt{\frac{\frac{n}{2}}{\log \frac{n}{2}}} \|C - \hat{C}_n\|_{HS} \leq 2\pi.$$

Remark I.4.16. Observe that the moment condition $X \in \mathcal{L}^{p+\epsilon}(T)$ is not needed for Theorem I.4.15 to be valid.

I.4.3.2 Consistent Estimation of Nonparanormal Processes

In the result above, we showed that $\hat{C}_n$ approximates C in Hilbert-Schmidt norm. We now derive convergence rates of the whole estimated copula model in the Wasserstein-distance. Therefore, we introduce

Assumption 2. The eigenspaces of C are one dimensional.

Define the empirical eigenvectors $(e_{jn})_{j\in\mathbb{N}}$ and in decreasing order $(\lambda_{jn})_{j\in\mathbb{N}}$ as the eigenelements of the estimator, that is, $\hat{C}_n = \sum_{j=1}^{\infty} \lambda_{jn} e_{jn}^{\otimes 2}$. Under Assumption 2 we can introduce the auxiliary operators

$$\hat{Q}_n = \sum_{j=1}^{\infty} \frac{1}{j^2 b_j} e_{jn}^{\otimes 2}, \quad Q = \sum_{j=1}^{\infty} \frac{1}{j^2 b_j} e_j^{\otimes 2}$$

where for $j \geq 2$ we set $b_j > \min((\lambda_j - \lambda_{j-1})^{-1}, (\lambda_{j+1} - \lambda_j)^{-1})$ and $b_1 > (\lambda_2 - \lambda_1)^{-1}$. Observe that $\hat{Q}_n$ and Q are positive semidefinite symmetric trace-class operators, which commute with $\hat{C}_n$ and C respectively. Therefore also $\hat{Q}_n^2 \hat{C}_n = \hat{Q}_n \hat{C}_n \hat{Q}_n$ and $Q^2 C = QCQ$ are positive semidefinite symmetric trace-class operators. Set $Z_j^n := \langle e_{jn}, F^{[-1]}_{\mathcal{N}(0,\hat{Q}_n^2 \hat{C}_n)}(F_X(X)) \rangle_{L^2(T)}$ and

$$\hat{Y}^n := \sum_{j=1}^{\infty} Z_j^n e_{jn} \sim \mathcal{N}(0, \hat{Q}_n^2 \hat{C}_n)$$

and let F^n be an arbitrary approximation of the marginals F_X of X (just satisfying the second moment condition of Lemma I.3.10). Now we can define a semiparametric estimator for (the law of) X by the (law of the) process $\hat{X}^n$ given by

$$\hat{X}_t^n := (F_t^n)^{[-1]}(F_{\mathcal{N}(0,\hat{Q}_n^2 \hat{C}_n)}(\hat{Y}_t^n)). \tag{I.43}$$

Remark I.4.17. If one wants to write down the estimator in practice, we must impose upper bounds on the differences of the neighbouring eigenvalues, that is, we assume to know suitable b_j's.

The proof of the following Theorem can be found in Appendix I.B.

Theorem I.4.18. *Let $X = (X_t)_{t\in T} \in \mathcal{L}^{2+\epsilon}(\Omega \times T)$ be a stochastic process with underlying Gaussian copula process U^X. We assume that Assumption 1 holds for the marginals of X and 2 holds for the underlying Gaussian copula. Then there are constants $0 < L, \kappa < \infty$ such that*

$$\mathbb{E}[\mathbb{W}_1(X, \hat{X}^n)] \le \|\mathbb{E}[\mathbb{W}_1(F_{X_\cdot}, F_\cdot^n)]\|_{L^1(T)} + L\mathbb{E}[\|C - \hat{C}_n\|_{op}^{\kappa}] \tag{I.44}$$

where

$$\kappa := \frac{\epsilon\beta}{(1+\epsilon)(2+\beta) - 2\beta}$$

and

$$L := K(\sup_{s\in T} \frac{1}{\sqrt{q(s,s)}} \frac{\pi\sqrt{\pi}}{3}(2\|C\|_{op}^{\frac{1}{2}} + \frac{1}{\sqrt{2}}))^{\rho}$$

for K and ρ as given in Theorem I.4.10 with $p = 1$, $q = 2$ and q is the covariance kernel corresponding to the covariance operator QC. In particular, by virtue of Theorem I.4.15 we obtain

$$\mathbb{E}[\mathbb{W}_1(X, \hat{X}^n)] = \mathcal{O}(\|\mathbb{E}[\mathbb{W}_1(F_{X_\cdot}, F_\cdot^n)]\|_{L^1(T)} + n^{-\frac{\kappa}{2}}) \tag{I.45}$$

Remark I.4.19. Observe that $\kappa \to \frac{1}{2}$ as $\beta \to 1$ and $\epsilon \to \infty$. Thus, the best convergence rate that can be deduced from the previous theorem is $n^{-\frac{1}{4}}$.

Example I.4.20. We could take empirical marginal distribution functions F^n, that is,

$$F_t^n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i(t)\le x}, \quad x \in \mathbb{R}.$$

By [18, Theorem 1] there exists a constant K_2 (independent of t) such that

$$\mathbb{E}[\|\mathbb{W}_1(F_{X_\cdot}, F_{\hat{X}_\cdot^n})\|_{L^1(T)}] = \int_T \mathbb{E}[\mathbb{W}_1(F_{X_t}, F_{\hat{X}_t^n})]dt \le K_2 \int_T \mathbb{E}[|X_t|^{2+\epsilon}]dtn^{-\frac{1}{2}}.$$

Since X is càdlàg we have $\int_T \mathbb{E}[|X_t|^{2+\epsilon}]dt < \infty$. Thus, Theorem I.4.18 yields

$$\mathbb{E}[\mathbb{W}_1(X, \hat{X}^n)] = \mathcal{O}(n^{-\frac{\kappa}{2}}). \tag{I.46}$$

I.5 Conclusion and Outlook

We introduced a general functional analytic theory for copulas in infinite dimensions and we characterised the main mathematical challenge (the construction problem) compared to the finite-dimensional case. Additionally we pointed out the potential of the framework to serve as a semiparametric alternative in functional data analysis.

The proof of the consistency of the estimator for Gaussian copula models (Theorem I.4.18) is independent of the precise structure of the estimator for Kendalls τ correlation function and better convergence rates can potentially be realised by different estimators. It is also not clear if the rate $n^{-\kappa/2}$ induced by

Theorem I.4.18 is optimal in some cases, for instance in the framework described in Example I.4.20.

Another step is to investigate whether the method described in this subsection transfers to processes with general underlying elliptical copulas in the sense of Example I.4.13. Some more effort is needed to find that out, in particular, in order to find feasible ways to make inference on S (which is not equal to 1, as in the Gaussian case).

These are appealing strands that are left for future research.

Acknowledgements. The authors would like to thank Anton Yurchenko-Tytarenko for providing a clever trick. We are grateful to two anonymous referees for their helpful remarks.

This research was funded within the project STORM: Stochastics for Time-Space Risk Models, from the Research Council of Norway (RCN). Project number: 274410.

References

[1] Alfonsi, A. and Jourdain, B. "A remark on the optimal transport between two probability measures sharing the same copula". In: *Statist. Probab. Lett.* Vol. 84 (2014), pp. 131–134.

[2] Aliprantis, C. D. and Border, K. C. *Infinite Dimensional Analysis - A Hitchhiker's Guide*. 3. Berlin: Springer, 2006.

[3] Berlinet, A. and Thomas-Agnan, C. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer Science & Business Media, 2004.

[4] Bibbona, E., Sacerdote, L., and Torre, E. "A copula-based method to build diffusion models with prescribed marginal and serial dependence". In: *Methodol. Comput. Appl. Probab.* Vol. 18, no. 3 (2016), pp. 765–783.

[5] Bielecki, T. R., Jakubowski, J., and Niewęgłowski, M. "Dynamic modeling of dependence in finance via copulae between stochastic processes". In: *Copula theory and its applications*. Springer, 2010, pp. 33–76.

[6] Bingham, N. H., Goldie, C. M., and Teugels, J. L. *Regular Variation*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1987.

[7] Boente, G., Salibián-Barrera, M., and Tyler, D. E. "A characterization of elliptical distributions and some optimality properties of principal components for functional data". In: *J. Multivar. Anal.* Vol. 131 (2014), pp. 254–264.

[8] Bogachev, V. I. *Weak Convergence of Measures*. American Mathematical Society, 2018.

[9] Bosq, D. *Linear Processes in Function Space: Theory and Applications*. Vol. 149. Springer, 2012.

[10] Buldygin, V. V. and Kharazishvili, A. B. *Geometric aspects of probability theory and mathematical statistics*. Vol. 514. Mathematics and its Applications. Dordrecht: Kluwer Academic Publishers, 2000.

[11] Carlsson, M. "von Neumann's trace inequality for Hilbert–Schmidt operators". In: *Expo. Math.* (2020).

[12] Cherubini, U. et al. *Dynamic Copula Methods in Finance*. John Wiley & Sons, Ltd, 2012.

[13] Constantinescu, C., Hashorva, E., and Ji, L. "Archimedean copulas in finite and infinite dimensions with application to ruin problems". In: *Insur. Math. Econ.* Vol. 49, no. 3 (2011), pp. 487–495.

[14] Cuesta-Albertos, J., Rüschendorf, L., and Tuero-Diaz, A. "Optimal Coupling of Multivariate Distributions and Stochastic Processes". In: *J. Multivar. Anal.* Vol. 46, no. 2 (1993), pp. 335–361.

[15] Darsow, W. F., Nguyen, B., and Olsen, E. T. "Copulas and Markov processes". In: *Illinois J. Math.* Vol. 36, no. 4 (Dec. 1992), pp. 600–642.

[16] Enflo, P. "A counterexample to the approximation problem in Banach spaces". In: *Acta Math.* Vol. 130, no. 1 (1973), pp. 309–317.

[17] Fang, H., Fang, K., and Kotz, S. "The Meta-elliptical Distributions with Given Marginals". In: *J. Multivariate Anal.* Vol. 82, no. 1 (2002), pp. 1–16.

[18] Fournier, N. and Guillin, A. "On the rate of convergence in Wasserstein distance of the empirical measure". In: *Probab. Theory Related Fields* vol. 162, no. 3 (2015), pp. 707–738.

[19] Gatheral, J., Jaisson, T., and Rosenbaum, M. "Volatility is rough". In: *Quantit. Finance* vol. 18, no. 6 (2018).

[20] Hausenblas, E. and Riedle, M. "Copulas in Hilbert spaces". In: *Stochastics* vol. 89, no. 1 (2017), pp. 222–239.

[21] Heil, C. *A Basis Theory Primer. Applied and Numerical Harmonic Analysis*. New York: Birkhäuser/Springer, 2011.

[22] Horváth, L. and Kokoszka, P. *Inference for Functional Data with Applications*. Springer Series in Statistics. Springer, 2012.

[23] Ibragimov, R. "Copula-based characterizations for higher order Markov processes". In: *Econom. Theory* vol. 25, no. 3 (2009), pp. 819–846.

[24] Ibragimov, R. and Lentzas, G. "Copulas and long memory". In: *Probab. Surv.* Vol. 14 (2017), pp. 289–327.

[25] Kallsen, J. and Tankov, P. "Characterization of dependence of multidimensional Lévy processes using Lévy copulas". In: *J. Multivar. Anal.* Vol. 97, no. 7 (2006), pp. 1551–1572.

[26] Klenke, A. *Probability Theory: A Comprehensive Course*. Universitext. London: Springer, 2013.

[27] Kruse, R. L. and Deely, J. J. "Joint continuity of monotonic functions". In: *Am. Math. Mon.* Vol. 76, no. 1 (1969), pp. 74–76.

[28] Lagerås, A. N. "Copulas for Markovian dependence". In: *Bernoulli* vol. 16, no. 2 (2010), pp. 331–342.

[29] Liu, H., Lafferty, J., and Wasserman, L. "The Nonparanormal: Semiparametric Estimation of High Dimensional Undirected Graphs". In: *J. Mach. Learn. Res.* Vol. 10, no. 80 (2009), pp. 2295–2328.

[30] Llave, R. and Obaya, R. "Regularity of the composition operator in spaces of Hölder functions". In: *Discrete Contin. Dyn. Syst.* Vol. 5 (1998), pp. 157–184.

[31] Masarotto, V., Panaretos, V. M., and Zemel, Y. "Procrustes Metrics on Covariance Operators and Optimal Transportation of Gaussian Processes". In: *Sankhya A* vol. 81, no. 1 (2019), pp. 172–213.

[32] McNeil, A. J. and Nešlehová, J. "Multivariate Archimedean copulas, d-monotone functions and l_1-norm symmetric distributions". In: *Ann. Statist.* Vol. 37, no. 5B (2009), pp. 3059–3097.

[33] Moore, D. S. and Spruill, M. C. "Unified Large-Sample Theory of General Chi-Squared Statistics for Tests of Fit". In: *Ann. Statist.* Vol. 3 (1975), pp. 599–616.

[34] Nelsen, R. B. *An Introduction to Copulas.* 2. Springer Series in Statistics. Springer, New York, 2006.

[35] Panaretos, V. M. and Zemel, Y. *An Invitation to Statistics in Wasserstein Space.* SpringerBriefs in Probability and Mathematical Statistics. Springer International Publishing, 2020.

[36] Pettis, B. J. "On integration in vector spaces". In: *Trans. Amer. Math. Soc.* Vol. 44, no. 2 (1938), pp. 277–304.

[37] Rachev, S. T. and Rüschendorf, L. *Mass Transportation Problems. Volume 1: Theory.* Vol. 1. Probability and Its Applications. New York: Springer, 1998.

[38] Ranganath, R. and Blei, D. "Correlated Random Measures". In: *J. Amer. Statist. Assoc.* Vol. 113, no. 521 (2018), pp. 417–430.

[39] Rüschendorf, L. "On the distributional transform, Sklar's theorem, and the empirical copula process". In: *J. Statist. Plann. Inference* vol. 139, no. 11 (2009), pp. 3921–3927.

[40] Schwartz, L. *Radon Measures on Arbitrary Topological Spaces and Cylindrical Measures.* London: Published for the Tata Institute of Fundamental Research, Bombay by Oxford University Press, 1973.

[41] Sempi, C. "Copulæ of Processes Related to the Brownian Motion: A Brief Survey". In: *On Logical, Algebraic, and Probabilistic Aspects of Fuzzy Set Theory.* Ed. by Saminger-Platz, S. and Mesiar, R. Cham: Springer International Publishing, 2016, pp. 173–180.

[42] Singer, I. *Bases in Banach spaces. I.* Die Grundlehren der mathematischen Wissenschaften, Band 154. Springer-Verlag, New York-Berlin, 1970.

[43] Vaart, A. W. van der. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

[44] Villani, C. *Optimal Transport: Old and New*. Vol. 338. Heidelberg: Springer Science & Business Media, 2008.

[45] Wilson, A. G. and Ghahramani, Z. "Copula processes". In: *Adv. Neural Inf. Process. Syst.* 2010, pp. 2460–2468.

[46] Zhang, H., Xu, Y., and Zhang, J. "Reproducing kernel Banach spaces for machine learning". In: *J. Mach. Learn. Res.* Vol. 10, no. Dec (2009), pp. 2741–2775.

Appendix I.A Copulas in Finite Dimensions

We begin with the definition of finite-dimensional cumulative distribution functions.

Definition I.A.1. Let J be a finite set. A function $F : (-\infty, \infty)^J \to [0, 1]$ is a cumulative distribution function on $\mathbb{R}^J$, if

(a) $\lim_{x_j \to \infty, j \in J} F((x_j)_{j \in J}) = 1$;

(b) For each $j_0 \in J$ $\lim_{x_{j_0} \to -\infty} F((x_j)_{j \in J}) = 0$;

(c) For each $i \in J$ the function $t \to F((x_j)_{j \in J \setminus \{i\}}, (t)_{j=i})$ is right-continuous for each $(x_j)_{j \in J \setminus \{i\}} \in \mathbb{R}^{J \setminus \{i\}}$;

(d) The F-volume of a multivariate interval $[a, b] := \times_{j \in J} [a_j, b_j]$

$$V_F([a, b]) := \sum_{v \in \prod_{j \in J} \{a_j, b_j\}} sign(v) F(v) \tag{I.47}$$

is nonnegative, that is, $V_F([a, b]) \geq 0$ for all $[a, b] := \prod_{j \in J} [a_j, b_j] \subset \mathbb{R}^J$. (Recall that the function $sign : \prod_{j \in J} \{a_j, b_j\} \to \{-1, 1\}$ is given by $sign(v) = (-1)^{N(v)}$, where $N(v) = \#\{j \in J : v_j = a_j\}$.)

Definition I.A.2. Let J be an arbitrary finite set. A copula on $\mathbb{R}^J$ is a cumulative distribution function $C : [0, 1]^J \to [0, 1]$ with uniform marginal distributions, i.e. for all $u \in [0, 1]$ we have

$$C((1)_{j \in J \setminus \{i\}}, (u)_{j=i}) = u.$$

Equivalently, each copula C can be uniquely identified with a probability measure μ_C, with cumulative distribution function C.

For the theory of copulas, the most important result is Sklar's Theorem:

Theorem I.A.3 (Sklar's Theorem in finite dimensions)**.** *Let J be an arbitrary finite set. Let F be a cumulative distribution function on $\mathbb{R}^J$ with marginal one-dimensional cumulative distribution functions F_j for each $j \in J$. Then there exists a copula C on $\mathbb{R}^J$, such that for all $(x_j)_{j\in J} \in (-\infty, \infty)^J$ we have*

$$F((x_j)_{j\in J}) = C((F_j((x_j))_{j\in J}). \tag{I.48}$$

If the marginals F_j are continuous for each $j \in J$, C is unique. If conversely C is a copula on $\mathbb{R}^J$ and F_j are one-dimensional cumulative distribution functions for each $j \in J$, then F defined by (I.48) *is a cumulative distribution function on $\mathbb{R}^J$ with marginals F_j for each $j \in J$.*

Proof. See for example Nelsen [34]. ∎

We use the construction of copulas by distributional transforms. The proof of the next Theorem can be found in for instance in [33] and in [39].

Theorem I.A.4. *Let J be a finite set, $F = F_J$ be a cumulative distribution function on $\mathbb{R}^J$ with marginals $F_j, j \in J$. Let $X = (X_j)_{j\in J}$ be a random vector with law F. Let U be uniformly distributed on $[0,1]$ and independent of X. Then a copula of F is given by the cumulative distribution function corresponding to the random vector $(U_1, ..., U_d)$, defined by*

$$U_i = F_i(X_i-) + U(F_i(X_i) - F_i(X_i-)). \tag{I.49}$$

Appendix I.B Proofs of Section I.4

In [14] the authors used the notion that two laws μ and ν on $(\mathbb{R}^I, \otimes_{i\in I}\mathcal{B}(\mathbb{R}))$ have the same dependence structure, if there exist two stochastic processes $X = (X_i)_{i\in I}$ and $Y = (Y_i)_{i\in I}$, such that $X \sim \mu$ and $Y \sim \nu$ on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and X_i and Y_i are similarly ordered ($X_i \overset{\text{s.o.}}{\sim} Y_i$) for all $i \in I$, that is,

$$(X_i(\omega) - X_i(\omega'))\,(Y_i(\omega) - Y_i(\omega')) \geq 0 \quad \mathbb{P}\otimes\mathbb{P} \text{ a.s.}.$$

In finite dimensions this notion is equivalent to the existence of a common underlying copula by virtue of Sklar's Theorem I.A.3. This is also valid in infinite dimensions, as we show next.

Lemma I.B.1. *Two probability measures μ and ν on $\otimes_{i\in I}\mathcal{B}(\mathbb{R})$ have a common underlying copula measure in the sense of (I.2) if and only if $X_i \overset{s.o.}{\sim} Y_i$ for all $i \in I$.*

Proof. Let μ and ν have the same underlying copula measure C and let $U \sim C$ be a corresponding copula process. Define with the notion introduced in (I.3) the random variables $X := (F^{[-1]}_{\mu_i}(U_i))_{i\in I}$ and $Y = (F^{[-1]}_{\nu_i}(U_i))_{i\in I}$. By construction and analogously to the proof of Sklar's Theorem I.2.3, we obtain $X \sim ((F^{[-1]}_{\mu_i})_{i\in I})_*C = \mu$ and $Y \sim ((F^{[-1]}_{\nu_i})_{i\in I})_*C = \nu$. Since the quantile transforms $F_{\mu_i}^{[-1]}$ and $F_{\nu_i}^{[-1]}$ are nondecreasing functions, we obtain that $X_i \overset{\text{s.o.}}{\sim} Y_i$ for all $i \in I$.

Vice versa, let $X \sim \mu$ and $Y \sim \nu$ be two random variables such that $X_i \overset{\text{s.o.}}{\sim} Y_i$ for all $i \in I$. Then by Proposition 2.1 in [14] for each $i \in I$ there exist a uniformly distributed random variable U_i such that $X_i = F_{\mu_i}^{[-1]}(U_i)$ and $Y_i = F_{\nu_i}^{[-1]}(U_i)$ (observe that the proof of this assertion does not need second moments, as stated in Remark 1 in [14]). If C is the law of $U = (U_i)_{i\in I}$, we obtain $\mu = ((F_{\mu_i}^{[-1]})_{i\in I})_* C$ and $\nu = ((F_{\nu_i}^{[-1]})_{i\in I})_* C$. This shows, that X and Y have the same underlying copula measure C. ■

Proof of Theorem I.4.1. Since the proof for the $\mathcal{L}^p(T)$ case is analogous, we will just show the assertion for l^p valued random variables X and Y.

Assume (i) holds. By Corollary I.3.25 we have that $F_X^{[-1]}(U)$ and $F_Y^{[-1]}(U)$ are measurable random variables taking values in l^p for a copula process $U \sim C$. Moreover, they are a coupling, as consequence of Sklar's Theorem I.2.3. To show optimality, observe first that for $X \sim \nu^1$ and $Y \sim \nu^2$ we have

$$\begin{aligned}
\mathbb{W}_p^p(X,Y) &= \inf_{\rho < {}^{\nu^1}_{\nu^2}} \int_{l^p \times l^p} \|x-y\|_p^p \rho(dx,dy) \\
&= \inf_{\rho < {}^{\nu^1}_{\nu^2}} \int_{l^p \times l^p} \sum_{i=1}^{\infty} |x_i - y_i|^p \rho(dx,dy) \\
&\geq \sum_{i=1}^{\infty} \inf_{\rho < {}^{\nu^1}_{\nu^2}} \int_{l^p \times l^p} |x_i - y_i|^p \rho(dx,dy) \qquad \text{(I.50)} \\
&= \sum_{i=1}^{\infty} \inf_{\rho_i < {}^{\nu^1_i}_{\nu^2_i}} \int_{\mathbb{R}\times\mathbb{R}} |x_i - y_i|^p \rho_i(dx_i,dy_i) \\
&= \sum_{i=1}^{\infty} \mathbb{W}_p^p(\nu_i^1,\nu_i^2) = \sum_{i=1}^{\infty} \mathbb{W}_p^p(X_i,Y_i).
\end{aligned}$$

This general lower bound on the Wasserstein distance is actually achieved in our case since, by (I.29), we obtain

$$\begin{aligned}
\sum_{i=1}^{\infty} \mathbb{W}_p^p(X_i,Y_i) =& \sum_{i=1}^{\infty} \int_{[0,1]} |F_{X_i}^{[-1]}(u_i) - F_{Y_i}^{[-1]}(u_i)|^p du_i \\
=& \int_{[0,1]^{\mathbb{N}}} \|(F_X^{[-1]}(u) - F_Y^{[-1]}(u))\|_p^p C(du) \\
=& \|F_X^{[-1]}(U) - F_Y^{[-1]}(U)\|_{L^p(\Omega;l^p)}^p \\
\geq& \mathbb{W}_p^p(X,Y).
\end{aligned}$$

This shows $(i) \Leftrightarrow (ii)$ and $(i) \Rightarrow (iii)$. Since $(ii) \Rightarrow (i)$ is trivial, it is therefore sufficient to show $(iii) \Rightarrow (i)$. Since equality in (I.50) can just hold, if there is an optimal coupling (X,Y), such that $\mathcal{W}_p^p(X_i,Y_i) = \mathbb{E}[|X_i - Y_i|^p]$, we have that

(X_i, Y_i) must also be an optimal coupling for all $i \in \mathbb{N}$. By Proposition 2.1 in [39] we obtain that for all $i \in \mathbb{N}$ we have that $X_i \overset{\text{s.o.}}{\sim} Y_i$. This implies (i) due to Lemma I.B.1. ■

Proof of Theorem I.4.10. By the triangle inequality we have

$$\begin{aligned} &\mathbb{E}\left[\|X - Y\|^p_{L^p(T)}\right]^{\frac{1}{p}} \\ \leq &\mathbb{E}\left[\|X - F_Y^{[-1]}(U_X)\|^p_{L^p(T)}\right]^{\frac{1}{p}} + E\left[\|F_Y^{[-1]}(U_X) - Y\|^p_{L^p(T)}\right]^{\frac{1}{p}} \end{aligned} \tag{I.51}$$

and since the L^p-distance majorizes the Wasserstein-distance

$$\mathbb{W}_p(X, Y) \leq \mathbb{E}\left[\|X - F_Y^{[-1]}(U_X)\|^p_{L^p(T)}\right]^{\frac{1}{p}} + \mathbb{W}_p(F_Y^{[-1]}(U_X), Y). \tag{I.52}$$

From Theorem I.4.1 we know that $(X, F_Y^{[-1]}(U_X))$ is an optimal coupling and

$$\mathbb{E}\left[\|X - F_Y^{[-1]}(U_X)\|^p_{L^p(T)}\right]^{\frac{1}{p}} = \|\mathbb{W}_p(F_{X_\cdot}, F_{Y_\cdot})\|_{L^p(T)}. \tag{I.53}$$

Let us now estimate the second summands. We will first show the assertion for the L^p-distance. Set $\delta := 1 + \frac{p(q+\beta)-q\beta}{(q+\beta)\epsilon} > 1$, such that $\frac{\delta - \frac{q\beta}{(q+\beta)p}}{\delta - 1} = 1 + \frac{\epsilon}{p}$. Then we can estimate for $\gamma = \frac{\delta}{\delta-1}$ using Hölder's inequality

$$\begin{aligned} &\mathbb{E}\left[\|F_Y^{[-1]}(U_X) - Y\|^p_{L^p(T)}\right] \\ =&\int_T \mathbb{E}\left[|F_{Y_t}^{[-1]}(U_t^X) - Y_t|^p\right] dt \\ =&\int_T \mathbb{E}\left[|F_{Y_t}^{[-1]}1(U_t^X) - Y_t|^{\frac{q\beta}{(q+\beta)\delta}} |F_{Y_t}^{[-1]}(U_t^X) - Y_t|^{p - \frac{q\beta}{(q+\beta)\delta}}\right] dt \\ \leq&(\int_T \mathbb{E}\left[|F_{Y_t}^{[-1]}(U_t^X) - Y_t|^{\frac{q\beta}{(q+\beta)}}\right] dt)^{\frac{1}{\delta}} (\int_T \mathbb{E}\left[|F_{Y_t}^{[-1]}(U_t^X) - Y_t|^{\gamma(p - \frac{q\beta}{(q+\beta)\delta})}\right] dt)^{\frac{1}{\gamma}}. \end{aligned} \tag{I.54}$$

Now observe that since Y_t and $F_{Y_t}^{[-1]}(U_t^X)$ share the same distribution and by the elementary inequality $|x+y|^r \leq 2^{r-1}(|x|^r + |y|^r)$ for $r \geq 1$ we have

$$\begin{aligned} \int_T \mathbb{E}\left[|F_{Y_t}(U_t^X) - Y_t|^{\gamma(p - \frac{q\beta}{(q+\beta)\delta})}\right] dt &= \int_T \mathbb{E}\left[|F_{Y_t}(U_t^X) - Y_t|^{p\frac{\delta - \frac{q\beta}{(q+\beta)p}}{\delta - 1}}\right] dt \\ &= \int_T \mathbb{E}\left[|F_{Y_t}(U_t^X) - Y_t|^{p+\epsilon}\right] dt \\ &\leq 2^{p+\epsilon-1} \int_T \mathbb{E}\left[|F_{Y_t}(U_t^X)|^{p+\epsilon} + |Y_t|^{p+\epsilon}\right] dt \\ &= 2^{p+\epsilon} \int_T \mathbb{E}\left[|Y_t|^{p+\epsilon}\right] dt \end{aligned}$$

$$=(2\|Y\|_{L^{p+\epsilon}(\Omega\times T)})^{p+\epsilon} \tag{I.55}$$

This shows that $\int_T \mathbb{E}[|F_{Y_t}(U_t^X) - Y_t|^{\gamma(p-\frac{q\beta}{(q+\beta)\delta})}]dt < \infty$, since Y is assumed to have finite moments up to $p+\epsilon$.

Now observe that

$$\begin{aligned}
&\mathbb{P}\left[(U_t^X, U_t^Y) \in (0,1)\right] \\
=&1 - \mathbb{P}\left[U_t^X \in \{0,1\} \text{ or } U_t^Y \in \{0,1\}\right] \\
\geq&1 - (\mathbb{P}\left[U_t^X \in \{0,1\}\right] + \mathbb{P}\left[U_t^Y \in \{0,1\}\right]) = 1 - 0 = 1
\end{aligned} \tag{I.56}$$

Moreover, by Assumption 1 we further have $0 < f_{Y_t}(F_{Y_t}^{[-1]}(\zeta))$ for $\zeta \in (0,1)$. Hence, since by (I.56)

$$[\min(U_t^X, U_t^Y), \max(U_t^X, U_t^Y)] \subset (0,1)$$

almost surely, we obtain by the inverse function theorem for $\zeta \in [\min(U_t^X, U_t^Y), \max(U_t^X, U_t^Y)]$

$$\frac{d}{dx} F_{Y_t}^{[-1]}(\zeta)) = \left(f_{Y_t}^{-1}(F_{Y_t}(\zeta))\right)^{-1}.$$

Appealing to the mean value theorem and once more Hölder's inequality ($\|fg\|_{L^1(T)} \leq \|f\|_{L^{\frac{r}{r-1}}(T)}\|g\|_{L^r(T)}$ with $r = \frac{(q+\beta)}{\beta}$) we obtain

$$\begin{aligned}
&\int_T \mathbb{E}\left[|F_{Y_t}^{[-1]}(U_t^X) - Y_t|^{\frac{q\beta}{(q+\beta)}}\right] dt \\
\leq &\int_T \mathbb{E}\left[\left(\sup_{\zeta\in\left[\min(U_t^X,U_t^Y),\max(U_t^X,U_t^Y)\right]} \left(f_{Y_t}\left(F_{Y_t}^{[-1]}(\zeta)\right)\right)^{-1} |U_t^X - U_t^Y|\right)^{\frac{q\beta}{(q+\beta)}}\right] dt \\
\leq &\left(\int_T \mathbb{E}\left[\sup_{\zeta\in[\min(U_t^X,U_t^Y),\max(U_t^X,U_t^Y)]} \left(f_{Y_t}\left(F_{Y_t}^{[-1]}(\zeta)\right)\right)^{-\beta}\right] dt\right)^{\frac{q}{\beta+q}} \\
&\quad \times \left(\int_T \mathbb{E}\left[|U_t^X - U_t^Y|^q\right] dt\right)^{\frac{\beta}{\beta+q}}.
\end{aligned} \tag{I.57}$$

We now show that the first factor is finite. Denote the random variables

$$\begin{aligned}
Z &:= \max_{\zeta\in[\min(U_t^X,U_t^Y),\max(U_t^X,U_t^Y)]} \left(f_{Y_t}\left(F_{Y_t}^{[-1]}(\zeta)\right)\right)^{-\beta} \\
\zeta^* &= \operatorname*{arg\,max}_{\zeta\in[\min(U_t^X,U_t^Y),\max(U_t^X,U_t^Y)]} \left(f_{Y_t}\left(F_{Y_t}^{[-1]}(\zeta)\right)\right)^{-\beta}
\end{aligned}$$

and choose x_0^t according to Assumption 1 such that g_t is ultimately monotone on $[-x_0^t, x_0^t]^c$, where without loss of generality $m_t = 0$. We can argue by continuity and monotonicity of cumulative distribution and quantile functions as well as Assumption 1 that

$$\begin{aligned}\mathbb{E}\left[\mathbb{I}_{\zeta^* \in (F_{Y_t}(-x_0^t), F_{Y_t}(x_0^t))} Z\right] \leq & \sup_{\zeta \in (F_{Y_t}(-x_0^t), F_{Y_t}(x_0^t))} \left(f_{Y_t}\left(F_{Y_t}^{[-1]}(\zeta)\right)\right)^{-\beta} \\ \leq & \sup_{\zeta \in (F_{Y_t}(-x_0^t), F_{Y_t}(x_0^t))} \left(g_t\left(F_{Y_t}^{[-1]}(\zeta)\right)\right)^{-\beta} \\ \leq & \mathbb{I}_{(0,\infty)}(x_0^t)\lambda^{-\beta} \end{aligned} \tag{I.58}$$

Without loss of generality we can assume g to be symmetric in the tails, that is, $g(x) = g(-x)$ for $x \geq x_0$. For $\zeta^* \notin [F_{Y_t}(-x_0^t), F_{Y_t}(x_0^t)]$ we have by definition

$$\left[\min(U_t^X, U_t^Y), \max(U_t^X, U_t^Y)\right] \not\subset \left[F_{Y_t}(-x_0^t), F_{Y_t}(x_0^t)\right]$$

and thus, we must have either $U_t^X \in [F_{Y_t}(-x_0^t), F_{Y_t}(x_0^t)]^c$ or $U_t^Y \in [F_{Y_t}(-x_0^t), F_{Y_t}(x_0^t)]^c$. Hence, by Assumption 1 as well as the monotonicity and the symmetry of g, we have

$$\begin{aligned}&\mathbb{I}_{\zeta^* \notin [F_{Y_t}(-x_0^t), F_{Y_t}(x_0^t)]} Z \\ \leq & \mathbb{I}_{\zeta^* \notin [F_{Y_t}(-x_0^t), F_{Y_t}(x_0^t)]} \max_{\zeta \in [\min(U_t^X, U_t^Y), \max(U_t^X, U_t^Y)]} \left(g_t\left(F_{Y_t}^{[-1]}(\zeta)\right)\right)^{-\beta} \\ \leq & \max\left(\left(g\left(F_{Y_t}^{[-1]}\left(U_t^Y\right)\right)\right)^{-\beta}, \left(g\left(F_{Y_t}^{[-1]}\left(U_t^X\right)\right)\right)^{-\beta}\right) \end{aligned} \tag{I.59}$$

Therefore, for any uniformly distributed U on $[0,1]$ we obtain

$$\begin{aligned}&\mathbb{E}\left[\mathbb{I}_{\zeta^* \notin \left[F_{Y_t}(-x_0^t), F_{Y_t}(x_0^t)\right]} Z\right] \\ \leq & \mathbb{E}\left[\max\left(\left(g_t\left(F_{Y_t}^{[-1]}\left(U_t^X\right)\right)\right)^{-\beta}, \left(g_t\left(F_{Y_t}^{[-1]}\left(U_t^Y\right)\right)\right)^{-\beta}\right)\right] \\ \leq & 2\mathbb{E}\left[\left(g_t\left(F_{Y_t}^{[-1]}(U)\right)\right)^{-\beta}\right] \\ = & 2\mathbb{E}\left[\left(g_t(Y_t)\right)^{-\beta}\right]. \end{aligned} \tag{I.60}$$

Thus, (I.58) and (I.60) imply

$$\begin{aligned}&\left(\int_T \mathbb{E}\left[\sup_{\zeta \in [\min(U_t^X, U_t^Y), \max(U_t^X, U_t^Y)]} \left(f_{Y_t}\left(F_{Y_t}^{[-1]}(\zeta)\right)\right)^{-\beta}\right] dt\right)^{\frac{q}{\beta+q}} \\ \leq & \left(\lambda^{-\beta} \int_T \mathbb{I}_{(0,\infty)}(x_0^t) dt + 2\int_T \mathbb{E}\left[\left(g_t(Y_t)\right)^{-\beta}\right] dt\right)^{\frac{q}{\beta+q}} \end{aligned} \tag{I.61}$$

Combining (I.54), (I.55), (I.57) and (I.61) we obtain

$$\mathbb{E}\left[\|F_Y^{[-1]}(U_X) - Y\|_{L^p(T)}^p\right]$$

$$\le \left(\lambda^{-\beta}\int_T \mathbb{I}_{(0,\infty)}(x_0^t)dt + 2\|\left(g_t(Y_t)\right)^{-\beta}\|_{L^1(\Omega\times T)}\right)^{\frac{q}{\delta\beta+\delta q}}$$
$$\times\left(2\|Y\|_{L^{p+\epsilon}(\Omega\times T)}\right)^{\frac{p+\epsilon}{\gamma}}\|U^X-U^Y\|_{L^q(\Omega\times T)}^{\frac{q\beta}{\delta\beta+\delta q}}$$
$$= K^p\|U^X-U^Y\|_{L^q(\Omega\times T)}^{\frac{\epsilon q\beta}{(p+\epsilon)(q+\beta)-q\beta}}.$$

This shows (I.32). To obtain (I.33) we just have to recognize that

$$\mathbb{W}_p(F_Y^{[-1]}(U_X)-Y)\le \inf_{U\sim U_X, V\sim U_Y}\|F_Y^{[-1]}(U)-F_Y^{[-1]}(V)\|_{L^p(\Omega\times T)}.$$

The expression under the infimum can be estimated in complete analogy to the distance $\mathbb{E}\left[\|F_Y^{[-1]}(U_X)-Y\|_{L^p(T)}^p\right]^{\frac{1}{p}}$ and hence, the proof is complete. ■

Proof of Theorem I.4.12. Using the triangle inequality, we obtain

$$\|U^Y-U^X\|_{L^q(T\times\Omega)}\le\|U^Y-F_{\tilde Y}(\tilde X)\|_{L^q(T\times\Omega)}+\|F_{\tilde Y}(\tilde X)-U^X\|_{L^q(T\times\Omega)}$$
$$=: (1)+(2)$$

Then for the first summand we have by the mean value inequality

$$(1)^q=\|F_{\tilde Y}(\tilde Y)-F_{\tilde Y}(\tilde X)\|_{L^q(T\times\Omega)}^q=\int_T\mathbb{E}[|F_{\tilde Y_t}(\tilde Y_t)-F_{\tilde Y_t}(\tilde X_t)|^q]dt$$
$$\le\|f_{\tilde Y}\|_\infty^q\int_T\mathbb{E}[|\tilde Y_t-\tilde X_t|^q]dt$$
$$=\|f_{\tilde Y}\|_\infty^q\|\tilde Y-\tilde X\|_{L^q(T\times\Omega)}^q \tag{I.62}$$

For the second summand we have again by the mean value theorem

$$(2)^q=\|F_{\tilde Y}(F_{\tilde X}^{[-1]}(U_t^X))-U^X\|_{L^q(T\times\Omega)}^q$$
$$=\int_T\int_0^1|F_{\tilde Y_t}(F_{\tilde X_t}^{[-1]}(u))-u|^q du dt$$
$$=\int_T\int_0^1|F_{\tilde Y_t}(F_{\tilde X_t}^{[-1]}(u))-F_{\tilde Y_t}(F_{\tilde Y_t}^{[-1]}(u))|^q du dt$$
$$\le\|f_{\tilde Y}\|_\infty^q\int_T\int_0^1|F_{\tilde X_t}^{[-1]}(u)-F_{\tilde Y_t}^{[-1]}(u)|^q du dt$$
$$=\|f_{\tilde Y}\|_\infty^q\|\mathbb{W}_q^q(F_{\tilde X_\cdot},F_{\tilde Y_\cdot})\|_{L^1(T)}.$$

Moreover, since $\|\mathbb{W}_q^q(F_{\tilde X_t},F_{\tilde Y_t})\|_{L^1(T)}=\mathbb{W}_q^q(\tilde X,F_{\tilde Y}(U^X))$, which by Remark 4.5 can be estimated as,

$$\mathbb{W}_q^q(\tilde X,F_{\tilde Y}(U^X))$$
$$\le\mathbb{W}_q^q(\tilde X,F_{\tilde Y}(U^Y))=\mathbb{W}_q^q(\tilde X,\tilde Y)\le\mathbb{E}[\|\tilde X-\tilde Y\|_{L^q(T)}^q]=\|\tilde X-\tilde Y\|_{L^q(T\times\Omega)}^q,$$

also the second assertion follows. The inequality in terms of the Wasserstein-distance follows also immediately, since

$$\mathbb{W}_q(U^Y, U^X) \leq \mathbb{W}_q(U^Y, F_{\tilde{Y}}(\tilde{X})) + \|F_{\tilde{Y}}(\tilde{X}) - U^X\|_{L^q(T\times\Omega)}.$$

The first summand is

$$\mathbb{W}_q(U^Y, F_{\tilde{Y}}(\tilde{X})) = \inf_{x\sim X, y\sim Y} \|F_{\tilde{Y}}(y) - F_{\tilde{Y}}(x)\|_{L^q(\Omega\times T}$$

and the expression under the infimum can be analogously estimated as (1). ■

Proof of Theorem I.4.15. We can use von Neumann's Trace inequality (see e.g. [11] for a proof in the Hilbert-space case) and get for any positive semidefinite trace-class operator $B = \sum_{j=1}^{\infty} \mu_j f_j^{\otimes 2}$ that (since $\mu_j \geq 0$)

$$\begin{aligned}\|T_{\hat{c}_n} - \hat{C}_n\|_{\text{HS}}^2 = \sum_{j=1, \tilde{\lambda}_{jn}<0}^{\infty} \tilde{\lambda}_{jn}^2 \leq \sum_{j=1, \tilde{\lambda}_{jn}<0}^{\infty} & \tilde{\lambda}_{jn}^2 + \mu_j^2 - 2\tilde{\lambda}_{jn}\mu_j \\ \leq & \|T_{\hat{c}_n}\|_{\text{HS}}^2 + \|B\|_{\text{HS}}^2 - 2\langle T_{\hat{c}_n}, B\rangle_{\text{HS}} \\ = & \|T_{\hat{c}_n} - B\|_{\text{HS}}^2.\end{aligned}$$

Therefore we have

$$\begin{aligned}\|\tilde{C} - \hat{C}_n\|_{\text{HS}} \leq \|\tilde{C} - T_{\hat{c}_n}\|_{\text{HS}} + \|T_{\hat{c}_n} - \hat{C}_n\|_{\text{HS}} \leq & 2\|\tilde{C} - T_{\hat{c}_n}\|_{\text{HS}} \\ = & 2\|\tilde{c} - \hat{c}_n\|_{L^2(T^2)}.\end{aligned}$$

By the mean value theorem we get

$$\begin{aligned}\|\tilde{c} - \hat{c}_n\|_{L^2(T^2)}^2 &= \int_{T^2} (\tilde{c}(t,s) - \hat{c}_n(t,s))^2 dsdt \\ &= \int_{T^2} (\sin(\frac{\pi}{2}\tau(t,s)) - \sin(\frac{\pi}{2}\hat{\tau}_n(t,s)))^2 dsdt \\ &\leq \int_{T^2} \frac{\pi^2}{4}(\tau(t,s) - \hat{\tau}_n(t,s))^2 dsdt\end{aligned}$$

Therefore

$$\|\tilde{C} - \hat{C}_n\|_{\text{HS}}^2 \leq \pi^2 \|\tau - \hat{\tau}_n\|_{L^2(T^2)}^2 = \pi^2 \|T_\tau - T_{\hat{\tau}_n}\|_{\text{HS}}^2.$$

Since $T_{\hat{\tau}_n}$ is bounded in the Hilbert-Schmidt norm, has T_τ as its mean and $\mathbb{E}[\|T_{\hat{\tau}_n} - T_\tau\|^2] \leq 2$, we can conclude from Corollary 2.1 in [9] that almost surely

$$\limsup_{n\to\infty} \sqrt{\frac{\frac{n}{2}}{\log \frac{n}{2}}} \|T_c - T_{\hat{c}_n}\|_{HS} \leq \pi \limsup_{n\to\infty} \sqrt{\frac{\frac{n}{2}}{\log \frac{n}{2}}} \|T_\tau - T_{\hat{\tau}_n}\|_{HS} \leq 2\pi.$$

Moreover, as $sign(X - \tilde{X})$ is bounded by 1 and has therefore all moments, [22, Theorem 2.5] yields

$$\mathbb{E}[\|\tilde{C} - \hat{C}_n\|_{\text{HS}}^2] \leq \pi^2 \mathbb{E}[\|T_\tau - T_{\hat{\tau}_n}\|_{\text{nuc}}^2] \leq \frac{\pi^2}{n} \mathbb{E}\left[\|sign(X - \tilde{X})\|_{L^2(T)}^4\right]$$

$$\leq \frac{\pi^2 \mu(T)^2}{n}.$$

■

Proof of Theorem I.4.18. Observe that the copula process underlying $\hat{X}^n$ is U^n, which is given by $U^n := F_{\mathcal{N}(0,\hat{Q}_n^2 \hat{C}_n)}(\hat{Y}^n)$. Since Assumption 1 is valid for the marginals of X we find K and ρ by Theorem I.4.10 such that

$$\mathbb{W}_1(X, \hat{X}_n^m) \leq \|\mathbb{W}_1(F_{X_\cdot}, F_{(\hat{X}^n)_\cdot})\|_{L^1(T)} + K\mathbb{W}_2(U^X, U^n)^\rho. \tag{I.63}$$

Recall that we denoted $Y = F^{[-1]}_{\mathcal{N}(0,Q^2C)}(F_X(X))$, where C arises from the kernel $c(s,t)$. Q^2C has a corresponding covariance kernel q, such that $s \mapsto q(s,s)$ is continuous and strictly positive and for all $j' \in \mathbb{N}$ we have

$$q(s,s) = \sum_{j=1}^{\infty} \frac{\lambda_j}{b_j^2 j^4} e_j(s)^2 \geq \frac{\lambda_{j'}}{b_{j'}^2 (j')^4} e_{j'}(s)^2.$$

Since by assumption $\tilde{c}(s,s) = \sum_{j=1}^{\infty} \lambda_j e_j(s)^2 = \frac{1}{\sqrt{2\pi}}$ for all $s \in T$, there always exists some $j' \in \mathbb{N}$ such that $\lambda_{j'} e_{j'}(s)^2 > 0$. Hence, for the density function f_Y of Y we obtain $\|f_Y\|_\infty = \sup_{s\in T}(2\pi q(s,s))^{-\frac{1}{2}} < \infty$. By Theorem I.4.12 we have

$$\mathbb{W}_2(U^X, U^n) \leq 2 \sup_{s\in T} \frac{1}{\sqrt{2\pi q(s,s)}} \mathbb{W}_2(Y, \hat{Y}^n). \tag{I.64}$$

The 2-Wasserstein-distance between two centered Gaussians corresponds to the Procrustes distance of their covariances (see [31]), such that

$$W_2(Y, \hat{Y}^n) = \inf_{U \text{unitary}} \|(Q^2C)^{\frac{1}{2}} - U(\hat{Q}^2\hat{C}_n)^{\frac{1}{2}}\|_{\text{HS}} \leq \|(Q^2C)^{\frac{1}{2}} - (\hat{Q}^2\hat{C}_n)^{\frac{1}{2}}\|_{\text{HS}}.$$

Now observe that the square-roots of the covariance operators of Y and $\hat{Y}^n$ are, due to commutativity, given by

$$(Q^2C)^{\frac{1}{2}} = QC^{\frac{1}{2}}, \quad (\hat{Q}^2\hat{C}_n)^{\frac{1}{2}} = \hat{Q}\hat{C}_n^{\frac{1}{2}}.$$

Therefore we can estimate

$$W_2(Y, \hat{Y}^n) \leq \|(Q - \hat{Q})C^{\frac{1}{2}}\|_{\text{HS}} + \|\hat{Q}(C^{\frac{1}{2}} - \hat{C}_n)^{\frac{1}{2}})\|_{\text{HS}}.$$

Define $e'_{jn} := sign(\langle e_{jn}, e_n\rangle)e_j$. Then we have $Q = \sum_{j=1}^{\infty} \frac{1}{j^2 b_j}(e'_{jn})^{\otimes 2}$ and the first summand can be estimated as

$$\begin{aligned}\|(Q - \hat{Q})C^{\frac{1}{2}}\|_{\text{HS}} \leq& \|(Q - \hat{Q})\|_{\text{HS}} \|C^{\frac{1}{2}}\|_{\text{op}} \\ \leq& (\sum_{j=1}^{\infty} \frac{1}{j^2 b_j} \|e_{jn}^{\otimes 2} - (e'_{jn})^{\otimes 2}\|_{\text{HS}}) \|C^{\frac{1}{2}}\|_{\text{op}}.\end{aligned}$$

Since

$$\begin{aligned}\|e_{jn}^{\otimes 2}-(e_{jn}')^{\otimes 2}\|_{\mathrm{HS}} \leq & \|e_{jn}\otimes(e_{jn}-e_{jn}')\|_{\mathrm{HS}}+\|(e_{jn}-e_{jn}')\otimes e_{jn}'\|_{\mathrm{HS}}\\ \leq & \|e_{jn}-e_{jn}'\|_{L^2(T)}\end{aligned}$$

we obtain by Lemma 4.3 in [9] that

$$\begin{aligned}\|(Q-\hat{Q})C^{\frac{1}{2}}\|_{\mathrm{H}S} \leq & (\sum_{j=1}^{\infty}\frac{1}{j^2 b_j}\|e_{jn}-e_{jn}'\|_{L^2(T)})\|C^{\frac{1}{2}}\|_{\mathrm{op}}\\ \leq & \|C-\hat{C}_n\|_{\mathrm{op}}(\sum_{j=1}^{\infty}\frac{2\sqrt{2}}{j^2})\|C^{\frac{1}{2}}\|_{\mathrm{op}}. \end{aligned} \tag{I.65}$$

For the second summand we obtain by [8, Lemma 2.5.1]

$$\|\hat{Q}(C^{\frac{1}{2}}-\hat{C}_n)^{\frac{1}{2}}\|_{\mathrm{HS}} \leq \|\hat{Q}\|_{\mathrm{HS}}\|C^{\frac{1}{2}}-\hat{C}_n^{\frac{1}{2}}\|_{\mathrm{op}} \leq \|Q\|_{\mathrm{HS}}\|C-\hat{C}_n\|_{\mathrm{op}}^{\frac{1}{2}}. \tag{I.66}$$

Combining (I.64), (I.65) and (I.66) and since $\|Q\|_{\mathrm{HS}} \leq \sum_{j=1}^{\infty}\frac{1}{j^2}=\frac{\pi}{6}$ we obtain

$$\begin{aligned}\mathbb{W}_2(U^X,U^n) \leq & 2\sup_{s\in T}\frac{1}{\sqrt{2\pi q(s,s)}}((\sum_{j=1}^{\infty}\frac{2\sqrt{2}}{j^2})\|C^{\frac{1}{2}}\|_{\mathrm{op}}+\|Q\|_{\mathrm{HS}})\|C-\hat{C}_n\|_{\mathrm{op}}^{\frac{1}{2}}\\ \leq & \sup_{s\in T}\frac{1}{\sqrt{q(s,s)}}\frac{\pi\sqrt{\pi}}{3}(2\|C\|_{op}^{\frac{1}{2}}+\frac{1}{\sqrt{2}})\|C-\hat{C}_n\|_{\mathrm{op}}^{\frac{1}{2}}.\end{aligned}$$

This yields the assertion. ■

Paper II

A Topological Proof of Sklar's Theorem in Arbitrary Dimensions

Fred Espen Benth, Giulia Di Nunno, Dennis Schroers

Published in *Dependence Modeling*, April 2022, volume 10, issue 1, pp. 22–28.
DOI: 10.1515/demo-2022-0103

II

Abstract

Copulas are appealing tools in multivariate probability theory and statistics. Nevertheless, the transfer of this concept to infinite dimensions entails some nontrivial topological and functional analytic issues, making a deeper theoretical understanding indispensable toward applications. In this short work, we transfer the well known property of compactness of the set of copulas in finite dimensions to the infinite-dimensional framework. As an application, we prove Sklar's theorem in infinite dimensions via a topological argument and the notion of inverse systems.

Contents

II.1 Introduction

Copulas are widely used and well-known concepts in the realm of statistics and probability theory. This is not least due to the advantages that go along with their often intuitive and flexible handling. To some extent, such practicality is lost in infinite dimensions, as consistency problems may occur and constructions via copulas in topological vector spaces culminate in general in cylindrical, rather than actual probability measures (cf. [10] and [2]). Building up a functional analytic and topological theory for copulas becomes pivotal in order to make the copula approach fully applicable to general stochastic processes.

The keystone of the theory is Sklar's theorem and there is a vast literature solely focussing on different proofs of this fundamental result. Among others, there are proofs based on the distributional transform in [12] and [4] and earlier already in [11], based on mollifiers in [7] or the constructive approach by the extension of subcopulas, as it was proved for the bivariate case in [14] and for the general multivariate case in [17] or [3].

The naive transfer of the subcopula-approach to an infinite-dimensional setting appears to be challenging, since, after the extension of the subcopulas corresponding to the finite-dimensional laws of an infinite-dimensional distribution, one would also have to check that this construction meets the necessary consistency conditions. In contrast, and besides the approach via distributional transforms (as extended to an infinite-dimensional setting in [2]), a nonconstructive proof based on topological arguments in [6] is naturally in tune with an infinite-dimensional setting.

In this paper, we will therefore adopt this ansatz. In contrast to the proof in [2], we prove Sklar's theorem on the level of probability measures and not on the level of random variables. The argument carries the steps in [6] over to an infinite-dimensional setting:

(i) Show that the set of copula measures is compact with respect to the topology of convergence of the finite-dimensional distributions (Definition II.4.1 below).

(ii) Prove the second part of Sklar's theorem (that every copula measure can be merged with any family of marginals to a probability measure). This step is straightforward and as in [2].

(iii) Prove that the operation of merging a copula measure with marginals is a continuous mapping and use the compactness of the set of copulas to conclude that this map has closed image. The second part of Sklar's theorem follows by arguing that this image is also dense in the space of probability measures.

In finite dimensions, the compactness of copulas is described as "folklore" in [7] and it implies some useful applications. Some of these results carry over to the infinite-dimensional setup (cf. Section II.5).

II.2 Short Primer on Topological Inverse Systems

We will frequently use the notation $\bar{\mathbb{R}}$ for the extended real line $[-\infty, \infty]$. For any measure μ on a measurable space $(B, \mathcal{B})$ and a measurable function $f : (B, \mathcal{B}) \to (A, \mathcal{A})$ into another measurable space $(A, \mathcal{A})$ we denote by $f_*\mu$ the pushforward measure with respect to f given by $f_*\mu(S) := \mu(f^{-1}(S))$ for all $S \in \mathcal{A}$. For I an arbitrary index set, $B = \bar{\mathbb{R}}^I$ and $\mathcal{B} = \otimes_{i \in I} \mathcal{B}(\bar{\mathbb{R}})$, we use the shorter notations $\pi_{J*}\mu =: \mu_J$ for a subset $J \subseteq I$ and $\pi_{\{i\}*}\mu =: \mu_i$ for an element $i \in I$, where π_J denotes the canonical projection on $\mathbb{R}^J$. If $J \subset I$ is finite, we denote the corresponding finite-dimensional cumulative distribution functions by F_{μ_J} or F_{μ_i} respectively, where in the latter we used $J = \{i\}$. We use the notation $\mathcal{I}$ for the set consisting of all finite subsets of I. Moreover, for a one-dimensional Borel measure μ_i on $\mathbb{R}$, we use the notation $F_{\mu_i}^{[-1]}$ for the quantile functions

$$F_{\mu_i}^{[-1]}(u) := \inf \{x \in (-\infty, \infty) : F_{\mu_i}(x) \geq u\}.$$

We will refer to the one-dimensional distributions $\mu_i, i \in I$ and equivalently $F_{\mu_i}, i \in I$ as marginals of the measure μ. We denote the set of all probability measures on $(\bar{\mathbb{R}}^I, \otimes_{i \in I} \mathcal{B}(\bar{\mathbb{R}}))$ by $\mathcal{P}(\bar{\mathbb{R}}^I)$. Moreover, for two topological spaces X, Y we write $X \cong Y$ if they are homeomorphic.

The remainder of the section is mainly based on [13]. Let X_J be a set for each $J \in \mathcal{I}$ and

$$(P_{J_1,J_2} : X_{J_2} \to X_{J_1}) \qquad \text{for } J_1 \subseteq J_2, \text{ with } J_1, J_2 \in \mathcal{I}$$

a family of mappings, also called projections, such that

(i) $P_{J,J} = id_J$ is the identity mapping for all $J \in \mathcal{I}$, and

(ii) $P_{J_1,J_3} = P_{J_1,J_2} \circ P_{J_2,J_3}$ for all $J_1 \subseteq J_2 \subseteq J_3$ in $\mathcal{I}$.

The system

$$(X_J, P_{J_1,J_2}, \mathcal{I}) := \Big((X_J)_{J \in \mathcal{I}}, ((P_{J_1,J_2} : X_{J_2} \to X_{J_1})_{\substack{J_1 \subseteq J_2 \\ J_1, J_2 \in \mathcal{I}}}) \Big)$$

is called an *inverse system* (over the partially ordered set $\mathcal{I}$). If (X_J, τ_J) are topological spaces for each $J \in \mathcal{I}$ and (P_{J_1,J_2}) are continuous for all $J_1 \subseteq J_2$ with $J_1, J_2 \in \mathcal{I}$, we call

$$(X_J, \tau_J, P_{J_1,J_2}, J \in \mathcal{I}) := \Big((X_J, \tau_J)_{J \in \mathcal{I}}, ((P_{J_1,J_2} : X_{J_2} \to X_{J_1})_{\substack{J_1 \subseteq J_2 \\ J_1, J_2 \in \mathcal{I}}}) \Big)$$

a *topological inverse system*. A *topological inverse limit* of this inverse system is a space X together with continuous mappings $P_J : X \mapsto X_J, J \in \mathcal{I}$, such that $P_{J_1,J_2} P_{J_2} = P_{J_1}$ for all $J_1 \subseteq J_2$ in $\mathcal{I}$ (that is, the mappings are *compatible*) and the following *universal property* holds: Whenever there is a topological space Y, such that there are continuous mappings $(\psi_J : Y \to X_J)_{J \in \mathcal{I}}$ which are compatible, i.e., $P_{J_1,J_2} \psi_{J_2} = \psi_{J_1}$ for all $J_1 \subseteq J_2$ in $\mathcal{I}$, then there exists a unique continuous mapping

$$\Psi : Y \to X, \tag{II.1}$$

with the property $P_J \Psi = \psi_J$ for all $J \in \mathcal{I}$. We have that

$$\left\{ x = (x_J)_{J \in \mathcal{I}} \in \prod_{J \in \mathcal{I}} X_J : P_{J_1,J_2}(\pi_{J_2}(x)) = \pi_{J_1}(x) \text{ for } J_1 \subseteq J_2 \right\} \subseteq \prod_{J \in \mathcal{I}} X_J \tag{II.2}$$

equipped with the subspace topology with respect to the product topology is an inverse limit of the topological inverse system, induced by the canonical projections $\pi_{J'}((x_J)_{J\in\mathcal{I}}) = x_{J'}$. Each topological inverse limit is homeomorphic to this space and therefore to every topological inverse limit (See the proof of Theorem 1.1.1 in [13]). We write $\lim_{\leftarrow} X_J \subseteq \prod_{J\in\mathcal{I}} X_J$ for the inverse limit as a subset of the product space and we equip it throughout with the induced subspace topology.

Lemma II.2.1. *Let $(X_J, \tau_J, \pi_{J_1,J_2})$ be a topological inverse system (over the poset $\mathcal{I}$) of Hausdorff spaces. Then $\lim_{\leftarrow} X_J$ is a closed subset of $\prod_{J\in\mathcal{I}} X_J$ with respect to the product topology.*

Proof. See [13, Lemma 1.1.2]. ∎

Lemma II.2.2. *Let X be a compact Hausdorff space and $(X_J, \tau_J, \pi_{J_1,J_2})$ be a topological inverse system of compact Hausdorff spaces. Let $\psi_J : X \to X_J$, $J \in \mathcal{I}$ be a family of compatible surjections and Ψ the induced mapping. Then either $\lim_{\leftarrow} X_J = \emptyset$ or $\Psi(X)$ is dense in $\lim_{\leftarrow} X_J$.*

Proof. See [13, Corollary 1.1.7]. ∎

II.3 Copulas and Sklar's Theorem

As they are cumulative distribution functions, copulas in finite dimensions have a one-to-one correspondence to probability measures. In infinite dimensions, we will therefore work with the notion of copula measures as introduced in [2].

Definition II.3.1. A copula measure (or simply copula) on $\bar{\mathbb{R}}^I$ is a probability measure $C \in \mathcal{P}(\bar{\mathbb{R}}^I)$, such that its marginals C_i are uniformly distributed on $[0, 1]$. We will denote the space of copula measures on $\bar{\mathbb{R}}^I$ by $\mathcal{C}(\bar{\mathbb{R}}^I)$.

Sklar's theorem as stated below was proved in [2] by following the arguments for the finite-dimensional assertion in [12]. Here we give an alternative proof for the infinite-dimensional setting using a topological argument as in [6].

Theorem II.3.2 (Sklar's Theorem)**.** *Let $\mu \in \mathcal{P}(\bar{\mathbb{R}}^I)$ be a probability measure with marginal one-dimensional distributions $\mu_i, i \in I$. There exists a copula measure C, such that for each $J \in \mathcal{I}$, we have*

$$F_{C_J}\left(\left(F_{\mu_j}(x_j)\right)_{j\in J}\right) = F_{\mu_J}\left((x_j)_{j\in J}\right) \tag{II.3}$$

for all $(x_j)_{j\in J} \in \bar{\mathbb{R}}^J$. Moreover, C is unique if F_{μ_i} is continuous for each $i \in I$. Vice versa, let C be a copula measure on $\bar{\mathbb{R}}^I$ and let $(\mu_i)_{i\in I}$ be a collection of (one-dimensional) Borel probability measures over $\bar{\mathbb{R}}$. Then there exists a unique probability measure $\mu \in \mathcal{P}(\bar{\mathbb{R}}^I)$, such that (II.3) *holds.*

II.4 Topological Properties of Copulas and a Proof of Sklar's Theorem

The collection $(\mathcal{P}(\bar{\mathbb{R}}^J), J \in \mathcal{I})$, where each $\mathcal{P}(\bar{\mathbb{R}}^J)$ is considered as a topological space with the topology of weak convergence, is a topological inverse system with the projections $\pi_{J_1,J_2}(\mu_{J_2}) = (\mu_{J_2})_{J_1}$ for $\mu_{J_2} \in \mathcal{P}(\bar{\mathbb{R}}^{J_2})$ and $J_1, J_2 \in \mathcal{I}$, $J_1 \subseteq J_2$. Moreover, observe that each $\mathcal{P}(\bar{\mathbb{R}}^J)$ is a Hausdorff space, since it is metrizable by the Prohorov metric (cf. [15, Theorem 4.2.5]). The space $\lim_{\leftarrow} \mathcal{P}(\bar{\mathbb{R}}^J) \subset \prod_{J\in\mathcal{I}} \mathcal{P}(\bar{\mathbb{R}}^J)$ of consistent families of probability measures is a topological inverse limit, equipped with the corresponding inverse limit topology. The space of probability measures on $\otimes_{i\in I}\mathcal{B}(\mathbb{R})$ has via its finite-dimensional distributions a one-to-one correspondence with this family of consistent finite-dimensional distributions, and hence there is a natural bijection between $\lim_{\leftarrow} \mathcal{P}(\bar{\mathbb{R}}^J)$ and $\mathcal{P}(\bar{\mathbb{R}}^I)$.

We equip the space $\mathcal{P}(\bar{\mathbb{R}}^I)$ with the topology of *weak convergence of the finite-dimensional distributions*, which we define as follows:

Definition II.4.1. The topology of convergence of the finite-dimensional distributions on $\mathcal{P}(\bar{\mathbb{R}}^I)$ is defined as the topology such that $\mathcal{P}(\bar{\mathbb{R}}^I) \cong \lim_{\leftarrow} \mathcal{P}(\bar{\mathbb{R}}^J)$.

$\mathcal{P}(\bar{\mathbb{R}}^I)$ with this topology is by definition a topological inverse limit. Define also $\lim_{\leftarrow} \mathcal{C}(\bar{\mathbb{R}}^J) := \lim_{\leftarrow} \mathcal{P}(\bar{\mathbb{R}}^J) \cap \prod_{J\in\mathcal{I}} \mathcal{C}(\bar{\mathbb{R}}^J)$. Certainly, we have

$$\mathcal{C}\left(\bar{\mathbb{R}}^I\right) \cong \lim_{\leftarrow} \mathcal{C}\left(\bar{\mathbb{R}}^J\right) \tag{II.4}$$

with the corresponding topologies.

The following result contains among other things the topological proof of Sklar's theorem II.3.2.

Theorem II.4.2. *The following statements hold.*

II.4.2.1. $\mathcal{P}(\bar{\mathbb{R}}^I)$ *with the topology of weak convergence of the finite-dimensional distributions is a Hausdorff space*

II.4.2.2. *The space of copula measures* $\mathcal{C}(\bar{\mathbb{R}}^I)$ *is compact with respect to the topology of convergence of finite-dimensional distributions.*

II.4.2.3. *For a copula measure* C *on* $\bar{\mathbb{R}}^I$ *and (one-dimensional) Borel probability measures* $(\mu_i)_{i\in I}$ *over* $\bar{\mathbb{R}}$ *the push-forward measure*

$$\mu := ((F_{\mu_i}^{[-1]})_{i\in I})_* C \tag{II.5}$$

satisfies (II.3).

II.4.2.4. *If we equip* $\mathcal{C}(\bar{\mathbb{R}}^I) \times \prod_{i\in I} \mathcal{P}(\bar{\mathbb{R}})$ *with the product topology of weak convergence on each* $\mathcal{P}(\bar{\mathbb{R}})$ *and the topology of convergence of the finite-dimensional distributions on* $\mathcal{C}(\bar{\mathbb{R}}^I)$ *and* $\mathcal{P}(\bar{\mathbb{R}}^I)$*, then the mapping* $\Phi : \mathcal{C}(\bar{\mathbb{R}}^{\mathcal{I}}) \times \prod_{i\in I} \mathcal{P}(\bar{\mathbb{R}}) \to \mathcal{P}(\bar{\mathbb{R}}^I)$ *given by*

$$\Phi(C, (\mu_i)_{i\in I}) := ((F_{\mu_i}^{[-1]})_{i\in I})_* C$$

is continuous and surjective. In particular, Sklar's theorem holds.

Proof. (II.4.2.1) Since products of Hausdorff spaces are Hausdorff and $\mathcal{P}(\bar{\mathbb{R}}^I)$ is homeomorphic to a subset of a product of Hausdorff spaces, it is Hausdorff.

(II.4.2.2) We know by [7, Thm. 3.3] that every $\mathcal{C}(\bar{\mathbb{R}}^J)$ is compact with respect to the topology of weak convergence on $\mathcal{P}(\bar{\mathbb{R}}^J)$. Tychonoff's Theorem guarantees also that $\prod_{J\in\mathcal{I}}\mathcal{C}(\bar{\mathbb{R}}^J)$ is compact with respect to the product topology on $\prod_{J\in\mathcal{I}}\mathcal{P}(\bar{\mathbb{R}}^J)$. Therefore, as $\lim_{\leftarrow}\mathcal{P}(\bar{\mathbb{R}}^J)$ is closed by Lemma II.2.1, we obtain that $\mathcal{C}(\bar{\mathbb{R}}^I)$ is compact, since it is homeomorphic to an intersection of a closed and a compact set in the product topology.

(II.4.2.3) This corresponds to the second part of Sklar's theorem and the proof can be conducted analogously to the one in [2]. Therefore, it is enough to see that

$$\left([0, F_{\mu_j}(x_j)]\right)_{j\in J} \setminus \left(\left(F_{\mu_j}^{[-1]}\right)^{-1}(-\infty, x_1]\right)_{j\in J}$$

is a C_J-nullset for all $(x_j)_{j\in J}\in\bar{\mathbb{R}}^J$, $J\in\mathcal{I}$, since then we immediately obtain

$$\begin{aligned} C_J\left(\left(\left(F_{\mu_j}^{[-1]}\right)^{-1}(-\infty, x_1]\right)_{j\in J}\right) =& C_J\left(\left([0, F_{\mu_j}(x_j)]\right)_{j\in J}\right) \\ =& F_{C_J}\left(F_{\mu_j}\left((x_j)\right)_{j\in J}\right). \end{aligned}$$

(II.4.2.4) Define $\phi_J : \mathcal{C}(\bar{\mathbb{R}}^I)\times\prod_{i\in I}\mathcal{P}(\bar{\mathbb{R}})\to\mathcal{P}(\bar{\mathbb{R}}^J)$ by

$$\phi_J(C, (\mu_i)_{i\in I}) := \Phi(C, (\mu_i)_{i\in I})_J,$$

which is well defined by (II.4.2.3). Since the finite-dimensional distributions of a law are consistent, $(\phi_J, J\in\mathcal{I})$ forms a compatible family. Define analogously for $J\in\mathcal{I}$ also $\tilde{\phi}_J : \mathcal{C}(\bar{\mathbb{R}}^J)\times\prod_{j\in J}\mathcal{P}(\bar{\mathbb{R}})\to\mathcal{P}(\bar{\mathbb{R}}^J)$ by

$$\tilde{\phi}_J(C_J, (\mu_j)_{j\in J}) = (F_{\mu_j}^{[-1]})_{j\in J})_* C_J.$$

This is by Sklar's theorem in finite-dimensions surjective and by [16, Thm. 2] also continuous. Hence $\phi_J = \tilde{\phi}_J\pi_J$ is continuous and surjective, since both, $\tilde{\phi}_J$ and π_J are. Φ must be the uniquely induced continuous mapping by the family $(\phi_J, J\in\mathcal{I})$ by the universality property of the inverse limit. Moreover, since by [15, Corollary 4.2.6] $\mathcal{P}(\bar{\mathbb{R}})$ is compact and by (II.4.2.2) also $\mathcal{C}(\bar{\mathbb{R}}^I)$ is compact, we have that $\mathcal{C}(\bar{\mathbb{R}}^I)\times\prod_{i\in I}\mathcal{P}(\bar{\mathbb{R}})$ is compact by Tychonoff's theorem. The continuity of Φ implies therefore that $\Phi(\mathcal{C}(\bar{\mathbb{R}}^I)\times\prod_{i\in I}\mathcal{P}(\bar{\mathbb{R}}))$ is compact, hence closed. Since moreover Lemma II.2.2 implies that $\Phi(\mathcal{C}(\bar{\mathbb{R}}^I)\times\prod_{i\in I}\mathcal{P}(\bar{\mathbb{R}}))$ is dense, we obtain that Φ is surjective and therefore also the first part of Sklar's theorem holds. The uniqueness of the copulas in the case of continuous marginals follows immediately by Sklar's theorem in finite dimensions via the uniqueness of the finite-dimensional distribution of the corresponding copula measure. ∎

II.5 Applications of the Compactness of Copulas in Infinite Dimensions

Apart from the alternative proof of Sklar's theorem, the compactness of the family of copulas has some useful implications.

Observe that $\mathcal{P}(\bar{\mathbb{R}}^J)$ is a subset of a locally convex Hausdorff space. Indeed, $\mathcal{P}(\bar{\mathbb{R}}^J)$ once equipped with the topology of convergence of the finite-dimensional distributions is topologically embedded in $C_b(\bar{\mathbb{R}}^J)^*$ equipped with the *weak**-topology, where $C_b(\bar{\mathbb{R}}^J)^*$ is the topological dual of the space of bounded continuous functions equipped with the topology induced by the uniform norm. Thus, with respect to the topology of weak convergence for each $J \in \mathcal{I}$, we obtain that also the inverse limit $\mathcal{P}(\bar{\mathbb{R}}^I) \cong \lim_{\leftarrow} \mathcal{P}(\bar{\mathbb{R}}^J)$ is topologically embedded in a locally convex Hausdorff space, as it is isomorphic to a subset of the product $\prod_{J\in\mathcal{I}} \mathcal{P}(\bar{\mathbb{R}}^J) \hookrightarrow \prod_{J\in\mathcal{I}} C_b(\bar{\mathbb{R}}^J)^*$. Since $\mathcal{C}(\bar{\mathbb{R}}^I) \hookrightarrow \mathcal{P}(\bar{\mathbb{R}}^I)$ is convex, we obtain the following result by the Krein-Milman theorem [5, Thm.V.8.4], as mentioned for instance in [8, p.30] for the finite-dimensional case:

Lemma II.5.1. *$\mathcal{C}(\bar{\mathbb{R}}^I)$ is the closure of the convex hull of its extremal points with respect to the topology of weak convergence of finite-dimensional distributions.*

As mentioned in [1] this implies for instance that

$$\sup_{C\in\mathcal{C}(\bar{\mathbb{R}}^I)} g(C) = \sup_{C\in ext(\mathcal{C}(\bar{\mathbb{R}}^I))} g(C)$$

where $ext(C(\bar{\mathbb{R}}^I))$ denotes the set of extremal points of $C(\bar{\mathbb{R}}^I)$ and $g : \mathcal{C}(\bar{\mathbb{R}}^I) \mapsto \mathbb{R}$ is a convex function.

The compactness of copulas might also be of interest for proving limit theorems. In fact, by the compactness of copulas in finite dimensions, we obtain that every sequence $(C^n)_{n\in\mathbb{N}}$ of multivariate copulas (of fixed dimension) has a convergent subsequence. This was for instance used in [9]. If $(C^n)_{n\in\mathbb{N}} \subset \mathcal{C}(\bar{\mathbb{R}}^I)$ is a sequence of copula measures and I is an infinite index set, this also implies that all finite distributions $F_{C^n_J}$ for $J \subset I$ finite have a convergent subsequent that converges weakly. However, it is not clear if this subsequence can be chosen uniformly for all finite $J \in I$, i.e. for all finite-dimensional distributions. Thus, this result is intricate to transfer to the infinite-dimensional setting especially since the notions of compactness and sequential compactness may not coincide and one has to appeal to the notion of nets instead of sequences. For countable index sets, we have at least

Lemma II.5.2. *If I is a countable index set, then $\mathcal{C}(\bar{\mathbb{R}}^I)$ is sequentially compact.*

Proof. $\bar{\mathbb{R}}^I$ is a product of polish spaces and hence polish with respect to the product topology. Thus, the Lévy-Prokhorov metric makes $\mathcal{P}(\bar{\mathbb{R}}^I)$ a metric space, whose topology coincides with the topology of convergence in distribution with respect to the product topology on $\mathbb{R}^I$, which itself coincides with the topology of weak convergence of the finite-dimensional distributions. As a compact set in a metrisable space, $\mathcal{C}(\bar{\mathbb{R}}^I)$ is sequentially compact. ■

With this Lemma it might be easy to prove convergence criteria in some topological vector spaces. Recall that the p-Wasserstein space $\mathcal{W}_p(E)$ over a separable Banach space E is given by

$$\mathcal{W}_p(E) = \left\{ \nu : \nu \text{ is a Borel law on } E, \int_E \|x\|_E^p \nu(dx) < \infty \right\}.$$

Let $E = l^p$ be the sequence space

$$l^p := \left\{ (x_n)_{n\in\mathbb{N}} \subset \mathbb{R}^{\mathbb{N}} : \|(x_n)_{n\in\mathbb{N}}\|_p := \left(\sum_{n=1}^{\infty} |x_n|^p \right)^{\frac{1}{p}} < \infty \right\}$$

for some $p \in [1, \infty)$. Recall that for the case $p = 2$ this class of spaces contains (by isomorphy) all separable Hilbert spaces.

Define by π_i for $i \in \mathbb{N}$ the projection onto the ith component, i.e. $\pi_i(x_1, x_2, ...) = x_i \quad \forall x \in l^p$. Any $\mu \in \mathcal{W}_p(l^p)$ is uniquely specified by the family $\mu_{i_1,...,i_d}$ of finite dimensional distributions given by

$$\mu_{i_1,...,i_d} := \mu \circ (\pi_{i_1}, ..., \pi_{i_d})^{-1}$$

for any finite subset $\{i_1, ..., i_d\} \subset \mathbb{N}$ of the natural numbers. We write $m^p_{\mu_i}$ for the pth absolute moment of the univariate measures μ_i, $i \in \mathbb{N}$, i.e.

$$m^p_{\mu_i} := \int_{\mathbb{R}} |x|^p \mu_i(dx).$$

Corollary II.5.3. *Assume $(\mu^n)_{n\in\mathbb{N}} \subset \mathcal{W}_p(l^p)$. If all for all $i \in \mathbb{N}$ there is a $\mu_i^\infty \in \mathcal{W}_p(l^p)$ such that*

$$\sum_{i=1}^{\infty} m^p_{\mu_i^\infty} < \infty,$$

then there is a subsequence $(\mu^{n_k})_{k\in\mathbb{N}}$ of Borel laws that converges with respect to the topology of weak convergence of finite dimensional distributions to some Borel law μ in $\mathcal{W}_p(l^p)$, such that

$$\mu_i^\infty = \mu_i \quad \forall i \in \mathbb{N}. \tag{II.6}$$

Proof. Recall from section 3.1 in [2] that a basis copula corresponding to a measure $\nu \in \mathcal{W}_1(l^p)$ is the copula measure C_ν in $\mathbb{R}^n$, such that its finite-dimensional distributions are given by the copula measures associated with the finite-dimensional distributions $\nu_{i_1,...,i_d}$. Let $(C_n)_{n\in\mathbb{N}}$ denote the sequence of basis copulas associated with the sequence $(\mu_n)_{n\in\mathbb{N}}$. Then by Lemma II.5.2, there is a subsequence $(C_{n_k})_{k\in\mathbb{N}}$ that converges to a copula measure $C \in \mathcal{C}(\bar{\mathbb{R}}^I)$ with respect to the topology of convergence of the finite-dimensional distributions. Theorem 2 in [16] tells us, that for multivariate random variables weak convergence follows by the convergence of the marginal distributions and the weak convergence of the associated copulas. By assumption, this yields the existence of a Borel law μ in $\mathbb{R}^{\mathbb{N}}$ such that

$$\mu^n \longrightarrow \mu,$$

(II.6) holds and μ has C as its underlying copula measure. That indeed $\mu \in \mathcal{W}_p(l^p)$ holds follows from Corollary 4 in [2]. ■

Remark II.5.4. Certainly, we can identify elements of Banach spaces with a Schauder basis uniquely with elements in $\mathbb{R}^{\mathbb{N}}$. In that way, transferring the assertion of Corollary II.5.3 to this more general situation is possible for instance by appealing to Corollary 3 in [2].

Acknowledgements

This research was funded within the project STORM: Stochastics for Time-Space Risk Models, from the Research Council of Norway (RCN). Project number: 274410.

References

[1] Benes, V. and Stepan, J. "Extremal Solutions in the Marginal Problem". In: *Advances in Probability Distributions with Given Marginals. Mathematics and Its Applications*. Ed. by Dall'Aglio, G., Kotz, S., and Salinetti, G. Vol. 67. Dordrecht: Springer, 1991.

[2] Benth, F., Di Nunno, G., and Schroers, D. "Copula Measures and Sklar's Theorem in Arbitrary Dimensions". In: *Scand. J. Statist. (forthcoming)* (2021).

[3] Carley, H. and Taylor, M. D. "A New Proof of Sklar's Theorem". In: *Distributions With Given Marginals and Statistical Modelling*. Ed. by Cuadras, C. M., Fortiana, J., and Rodriguez-Lallena, J. A. Dordrecht: Springer Netherlands, 2002, pp. 29–34.

[4] Deheuvels, P. "A multivariate Bahadur–Kiefer representation for the empirical copula process". In: *J. Math. Sci. (N. Y.)* Vol. 163 (2009), pp. 382–398.

[5] Dunford, N. and Schwartz, J. T. *Linear operators. Part I.* Wiley Classics Library. General theory, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1988, pp. xiv+858.

[6] Durante, F., Fernández-Sánchez, J., and Sempi, C. "A topological proof of Sklar's theorem". In: *Appl. Math. Lett.* Vol. 26, no. 9 (2013), pp. 945–948.

[7] Durante, F., Fernández-Sánchez, J., and Sempi, C. "Sklar's theorem obtained via regularization techniques". In: *Nonlinear Anal.* Vol. 75, no. 2 (2012), pp. 769–774.

[8] Durante, F. and Sempi, C. *Principles of Copula Theory*. London: CRC/Chapman & Hall, 2015.

[9] Fernández-Sánchez, J., Nelsen, R., and Úbeda-Flores, M. "Multivariate copulas, quasi-copulas and lattices". In: *Statist. Probab. Lett.* Vol. 81, no. 9 (2011), pp. 1365–1369.

[10] Hausenblas, E. and Riedle, M. "Copulas in Hilbert spaces". In: *Stochastics* vol. 89, no. 1 (2017), pp. 222–239.

[11] Moore, D. S. and Spruill, M. C. "Unified Large-Sample Theory of General Chi-Squared Statistics for Tests of Fit". In: *Ann. Statist.* Vol. 3 (1975), pp. 599–616.

[12] Rüschendorf, L. "On the distributional transform, Sklar's theorem, and the empirical copula process". In: *J. Statist. Plann. Inference* vol. 139, no. 11 (2009), pp. 3921–3927.

[13] Ribes, L. and Zalesskii, P. *Profinite groups*. Second. Vol. 40. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]. Springer-Verlag, Berlin, 2010, pp. xvi+464.

[14] Schweizer, B. and Sklar, A. "Operations on distribution functions not derivable from operations on random variables". In: *Studia Math.* Vol. 52 (1974), pp. 43–52.

[15] Schweizer, B. and Sklar, A. *Probabilistic Metric Spaces*. Dover Publications, New York, 2005.

[16] Sempi, C. "Convergence of copulas: critical remarks". In: *Rad. Mat.* Vol. 12, no. 2 (2004), pp. 241–249.

[17] Sklar, A. "Random variables, distribution functions, and copulas—a personal look backward and forward". In: *Distributions with fixed marginals and related topics*. Ed. by Rüschendorf, L., Schweizer, B., and Taylor, M. D. Vol. Volume 28. Lecture Notes–Monograph Series. Hayward, CA: Institute of Mathematical Statistics, 1996, pp. 1–14.

Paper III

A Weak Law of Large Numbers for Realised Covariation in a Hilbert Space Setting

Fred Espen Benth, Dennis Schroers, Almut E. D. Veraart

Published in *Stochastic Processes and their Applications*, March 2022, volume 145, pp. 241–268. DOI: 10.1016/j.spa.2021.12.011

Abstract

This article generalises the concept of realised covariation to Hilbert-space-valued stochastic processes. More precisely, based on high-frequency functional data, we construct an estimator of the trace-class operator-valued integrated volatility process arising in general mild solutions of Hilbert space-valued stochastic evolution equations in the sense of [25]. We prove a weak law of large numbers for this estimator, where the convergence is uniform on compacts in probability with respect to the Hilbert-Schmidt norm. In addition, we determine convergence rates for common stochastic volatility models in Hilbert spaces.

III

Contents

III.1 Introduction

Stochastic volatility and covariance estimation are of key importance in many fields. Motivated in particular by financial applications, a lot of research has been devoted to constructing suitable (co-) volatility estimators and to deriving their asymptotic limit theory in the setting when discrete, high-frequent observations are available. Initially, the main interest was in (continuous-time) stochastic models based on (Itô) semimartingales, where the so-called realised variance and covariance estimators (and their extensions) proved to be powerful tools. Relevant articles include the works by [3, 4, 5, 7] and [32], amongst many others, and the textbooks by [33] and [1].

Subsequently, the theory was extended to cover non-semimartingale models, see, for instance, [22], [9], [8], [23], [21] and the survey by [41], where the proofs of the asymptotic theory rely on Malliavin calculus and the famous fourth-moment theorem, see [37]. The multivariate theory has been studied in [30, 39].

Common to these earlier lines of investigation is the fact that the stochastic processes considered have finite dimensions. In this article, we extend the concept of realised covariation to an infinite-dimensional framework.

The estimation of covariance operators is elementary in the field of functional data analysis and was elaborated mainly for discrete-time series of functional data (see e.g. [42], [27], [46], [16], [31], [38]). However, spatio-temporal data that can be considered as functional might also be sampled densely in time, like forward curves for interest rates or commodities and data from geophysical and environmental applications.

In this paper, we consider a separable Hilbert space H and study H-valued stochastic processes Y of the form

$$Y_t = \mathcal{S}(t)Y_0 + \int_0^t \mathcal{S}(t-s)\alpha_s ds + \int_0^t \mathcal{S}(t-s)\sigma_s dW_s, \quad t \in [0,T], \tag{III.1}$$

for some $T > 0$. Here $(\mathcal{S}(t))_{t\geq 0}$ is a strongly continuous semigroup, $\alpha := (\alpha_t)_{t\in[0,T]}$ a predictable and almost surely integrable H-valued stochastic process, $\sigma := (\sigma_t)_{t\in[0,T]}$ is a predictable operator-valued process, Y_0 with values in H is some initial element and W a so called Q-Wiener process on H (see Section III.2 below for details).

Our aim is to construct an estimator for the integrated covariance process

$$\left(\int_0^t \sigma_s Q \sigma_s^* ds\right)_{t\in[0,T]}.$$

More precisely, we denote by

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} (Y_{t_i} - \mathcal{S}(\Delta_n)Y_{t_{i-1}})^{\otimes 2}, \tag{III.2}$$

the *semigroup-adjusted realised covariation (SARCV)* for an equally spaced grid $t_i := i\Delta_n$ for $\Delta_n = 1/n$, $i = 1, \ldots, \lfloor t/\Delta_n \rfloor$. We prove uniform convergence in probability (ucp) with respect to the Hilbert-Schmidt norm of the (SARCV) to the integrated covariance process. It is in line with the finite-dimensional theory for continuous semimartingales that, apart from the necessary assumptions for stochastic integrability, no assumptions have to be imposed on the stochastic volatility process σ to guarantee the validity of this weak law of large numbers.

In that sense, the (SARCV) can be regarded as a natural generalisation of the well known realised quadratic (co-)variation in finite dimensions (which is a special case) to processes of the form (III.1), which are sometimes coined *mild Itô processes*, cf. [24, Section 2].

Nevertheless, our framework certainly differs from common high-frequency settings mainly due to peculiarities that arise from infinite dimensions. Observe that the main motivation to consider processes in this form, is that a vast amount of parabolic stochastic partial differential equations possess only mild (in opposition to analytically strong) solutions, which are of the form (III.1). That is, Y is (under weak conditions) the mild solution of a stochastic partial differential equation

$$\text{(SPDE)} \quad dX_t = (AX_t + \alpha_t)dt + \sigma_t dW_t, \quad X_0 = Y_0, \quad t \in [0,T],$$

where A is the infinitesimal generator of the semigroup $(\mathcal{S}(t))_{t\geq 0}$ (cf. [25], [40] or [35]).

In contrast to finite-dimensional stochastic diffusions, this is a priori not an H-valued semimartingale, but rather an H-valued Volterra process and under certain conditions on the volatility, the rate of convergence can be affected. For instance, in the case of a constant deterministic volatility, the rate is $\mathcal{O}(\Delta_n^{1/2})$ in the semimartingale case, but might be arbitrarily slow in our infinite-dimensional mild framework, as it is essentially determined by the continuity of the semigroup on the range of the volatility (see Theorem III.3.2 below and the subsequent remark). A discussion around different rates of convergence in various cases is included in Section III.4 of the paper.

Various recent developments related to statistical inference for (parabolic) SPDEs based on discrete observations in time and space have emerged, see e.g. [20], [15], [17], [18].

To the best of our knowledge, our paper is the first one considering high-frequency estimation of (co-) volatility of infinite-dimensional stochastic evolution equations in an operator setting. This is of interest for various reasons. For instance, a simple and important application might be the parameter estimation for H-valued Ornstein-Uhlenbeck process (that is, $\sigma_s = \sigma$ is a constant operator). Elementary techniques such as functional principal component analysis might then be considered on the level of volatility. In a multivariate setting, dynamical dimension reduction was conducted for instance in [2]. Furthermore, it can be used as a tool for inference of infinite-dimensional stochastic volatility models as in [13] or [14]. In the special case of a semigroup that is continuous with respect to the operator norm, the framework also covers the estimation of volatility for H-valued semimartingales.

We organize the paper as follows: First, we recall the main technical preliminaries of our framework in Section III.2. In Section III.3, we establish the weak law of large numbers. In Section III.4, we study the behaviour of the estimator in special cases of semigroups and volatility. We derive convergence rates for particular examples of semigroups in Section III.4.1 and stochastic volatility models in Section III.4.2. Section 5 is devoted to the proofs of our main results, while in Section III.6 we discuss our results and methods in relation to some existing literature and provide some outlook into further developments.

III.2 Notation and Some Preliminary Results

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ denote a filtered probability space satisfying the usual conditions. Consider two separable Hilbert spaces U, H with scalar products denoted by $\langle\cdot,\cdot\rangle_U$, $\langle\cdot,\cdot\rangle_H$ and norms $\|\cdot\|_U$, $\|\cdot\|_H$, respectively. We denote $L(U,H)$ the space of all linear bounded operators $K: U \to H$, and use the shorthand notation $L(U)$ for $L(U,U)$. Equipped with the operator norm, $L(U,H)$ becomes a Banach space. The adjoint operator of a $K \in L(U,H)$ is denoted by K^*, and is an element on $L(H,U)$.

Following [40, Appendix A] we use the following notations: An operator $K \in L(U,H)$ is called *nuclear* or *trace class* if the following representation holds

$$Ku = \sum_k b_k \langle u, a_k\rangle_U, \text{ for } u \in U,$$

where $\{a_k\} \subset U$ and $\{b_k\} \subset H$ such that $\sum_k \|a_k\|_U \|b_k\|_H < \infty$. The space of all nuclear operators is denoted by $L_1(U,H)$; it is a separable Banach space and its norm is denoted by

$$\|K\|_1 := \inf\left\{\sum_k \|a_k\|_U\|b_k\|_H : Ku = \sum_k b_k\langle u, a_k\rangle_U\right\}.$$

We denote by $L_1^+(U,H)$ the class of all symmetric, non-negative-definite nuclear operators from U to H. We write $L_1(U)$ and $L_1^+(U)$ for $L_1(U,U)$ and $L_1^+(U,U)$, respectively.

For $x \in U$ and $y \in H$, we define the tensor product $x \otimes y$ as the linear operator in $L(U,H)$ defined as $x \otimes y(z) := \langle x, z\rangle_U y$ for $z \in U$. We note that $x \otimes y \in L_1(U,H)$ and $\|x \otimes y\|_1 = \|x\|_U \|y\|_H$, see [40, p. 107].

The operator $K \in L(U,H)$ is said to be a *Hilbert-Schmidt operator* if

$$\sum_k \|Ke_k\|_H^2 < \infty,$$

for any orthonormal basis (ONB) $(e_k)_{k\in\mathbb{N}}$ of U. The space of all Hilbert-Schmidt operators is denoted by $L_{\text{HS}}(U,H)$. We can introduce an inner product by

$$\langle K, L\rangle_{\text{HS}} := \sum_k \langle Ke_k, Le_k\rangle_H, \text{ for } K, L \in L_{\text{HS}}(U,H).$$

The induced norm is denoted $\|\cdot\|_{\text{HS}}$. As usual, we write $L_{\text{HS}}(U)$ in the case $L_{\text{HS}}(U,U)$.

We have the following convenient result for the space of Hilbert-Schmidt operators. Although it is well-known, we include the proof of this result for the convenience of the reader:

Lemma III.2.1. *Let U, V, H be separable Hilbert spaces. Then $L_{HS}(U,H)$ is a separable Hilbert space. Moreover, if $K \in L_{HS}(U,V), L \in L_{HS}(V,H)$, then $LK \in L_{HS}(U,H)$ and*

$$\|LK\|_{HS} \le \|L\|_{op}\|K\|_{HS} \le \|L\|_{HS}\|K\|_{HS}, \tag{III.3}$$

where the HS-norms are for the spaces in question.

Proof. It is well-known that $L_{\mathrm{HS}}(U,H)$ is a separable Hilbert space (see e.g. [40, Appendix A.2, p. 356]). Indeed, an orthonormal basis is $(e_i \otimes f_j)_{i,j\in\mathbb{N}}$ where $(e_i)_{i\in\mathbb{N}}$ is an orthonormal basis for U and $(f_j)_{j\in\mathbb{N}}$ for H. Notice that for any $x \in U$, we have for $L \in L_{\mathrm{HS}}(U,H)$

$$\|Lx\|_H^2 = \sum_{i=1}^{\infty}\langle Lx, e_i\rangle_H^2 = \sum_{i=1}^{\infty}\langle x, L^*e_i\rangle_H^2 \le \|x\|_H^2 \sum_{i=1}^{\infty}\|L^*e_i\|_H^2 = \|x\|_U^2\|L^*\|_{\mathrm{HS}}^2,$$

where $(e_i)_{i=1}^{\infty}$ is an orthonormal basis in U and we applied the Cauchy-Schwarz inequality. Hence, $\|L\|_{\mathrm{op}} \le \|L^*\|_{\mathrm{HS}} = \|L\|_{\mathrm{HS}}$. It can be seen directly from the definition of the Hilbert-Schmidt norm that for $L \in L_{\mathrm{HS}}(V,H), K \in L_{\mathrm{HS}}(U,V)$, it holds

$$\|LK\|_{\mathrm{HS}} \le \|L\|_{\mathrm{op}}\|K\|_{\mathrm{HS}} \le \|L\|_{\mathrm{HS}}\|K\|_{\mathrm{HS}},$$

and the claimed algebraic structure of Hilbert-Schmidt operators follows. ■

III.2.1 Hilbert Space-Valued Stochastic Integrals

Fix $T > 0$ and assume that $0 \le t \le T$. Let H and U be separable Hilbert spaces throughout. Recall that a U-valued random variable X is normal with mean $a \in U$ and covariance operator $Q \in L_1^+(U)$ if $\langle X, f\rangle_U$ is a real-valued normally distributed random variable for each $f \in U$, with mean $\langle a, f\rangle$ and

$$E[\langle X, f\rangle_U \langle X, g\rangle_U] = \langle Qf, g\rangle_U,$$

for all $f, g \in U$.

Definition III.2.2. A stochastic process $(W_t)_{t\ge 0}$ with values in U is called a Wiener process with covariance operator $Q \in L_1^+(U)$, if $W_0 = 0$ almost surely, W has independent and stationary increments, and for $0 \le s \le t$, we have $W_t - W_s \sim N(0, (t-s)Q)$.

Throughout let W denote a Wiener process taking values in U with covariance operator $Q \in L_1^+(U)$. To this operator we can assign the reproducing kernel Hilbert space $U_0 := Q^{\frac{1}{2}}U$ equipped with the scalar product $\langle h, g\rangle_0 := \langle Q^{-\frac{1}{2}}h, Q^{-\frac{1}{2}}g\rangle_H$, where $Q^{-\frac{1}{2}}$ is the pseudo-inverse of $Q^{\frac{1}{2}}$. The space $(U_0, \langle\cdot,\cdot\rangle_0)$ forms again a separable Hilbert space (cf. Proposition C.03 in [34]). We define for $T < \infty$ the space $\mathcal{N}_W(0,T;H)$ as the space of all predictable $L_{\mathrm{HS}}(U_0;H)$-valued processes $(\sigma_s)_{s\in[0,T]}$ such that

$$\mathbb{P}\left[\int_0^T \|\sigma_s Q^{1/2}\|_{\mathrm{HS}}^2 ds < \infty\right] = 1. \tag{III.4}$$

Let $\sigma = (\sigma_t)_{t\geq 0}$ denote a stochastic volatility process where $\sigma \in \mathcal{N}_W(0,T;H)$ for some fixed $T < \infty$. The stochastic integral

$$Y_t := \int_0^t \sigma_s dW_s$$

can then be defined as in [34, Chapter 2] and takes values in the Hilbert space H.

We denote the tensor product of the stochastic integral Y by $(Y_t)^{\otimes 2} = Y_t \otimes Y_t$, and define the corresponding stochastic variance term as the *operator angle bracket* (not to be confused with the inner products introduced above!) given by

$$\langle\langle Y \rangle\rangle_t = \int_0^t \sigma_s Q \sigma_s^* ds = \int_0^t (\sigma_s Q^{1/2})(\sigma_s Q^{1/2})^* ds,$$

see [40, Theorem 8.7, p. 114].

Remark III.2.3. As in [25, p. 104], we note that $(\sigma_s Q^{1/2}) \in L_{HS}(U,H)$ and $(\sigma_s Q^{1/2})^* \in L_{HS}(H,U)$. Hence the process $(\sigma_s Q^{1/2})(\sigma_s Q^{1/2})^* = \sigma_s Q \sigma_s^*$ for $s \in [0,T]$ takes values in $L_1(H,H)$.

Remark III.2.4. The integral $\int_0^t \sigma_s Q \sigma_s^* ds$ is interpreted as a Bochner integral in the space of Hilbert-Schmidt operators $L_{\text{HS}}(H)$. Indeed, $\sigma_s Q \sigma_s^*$ is a linear operator on H, and we have almost surely

$$\begin{aligned}\int_0^t \|\sigma_s Q \sigma_s^*\|_{\text{HS}} ds &= \int_0^t \|\sigma_s Q^{1/2} (\sigma_s Q^{1/2})^*\|_{\text{HS}} ds \\ &\leq \int_0^t \|\sigma_s Q^{1/2}\|_{\text{HS}}^2 ds < \infty,\end{aligned}$$

by appealing to Lemma III.2.1 and the assumption that $\sigma \in \mathcal{N}_W(0,T;H)$.

Remark III.2.5. From the existence of a localising sequence of stopping times

$$\tau_N := \{t \in [0,T] : \int_0^t \|\sigma_s Q^{\frac{1}{2}}\|_{\text{HS}}^2 ds > N\},$$

as described in [34, p.36] such that for the stopped process given by $Y_{\min(t,\tau_N)} = \int_0^t \mathbb{I}_{[0,\tau_n]}(s)\sigma_s dW_s$ we have

$$\mathbb{E}\left[\int_0^T \|\mathbb{I}_{[0,\tau_n]}(s)\sigma_s Q^{\frac{1}{2}}\|_{\text{HS}}^2 ds\right] < \infty$$

and appealing to [40, Theorem 8.2, p. 109] we deduce that the process $(M_t)_{t\geq 0}$ with

$$M_t = (Y_t)^{\otimes 2} - \langle\langle Y \rangle\rangle_t$$

is an $L_1(H)$-valued local martingale w.r.t. $(\mathcal{F}_t)_{t\geq 0}$. Thus, the operator angle bracket process can be called the *quadratic covariation process* of Y_t, which we shall do from now on.

We will need the following result, which is a direct corollary of the Hilbert space version of the Burkholder-Davis-Gundy inequality (cf. [36]).

Lemma III.2.6. *Let $\sigma \in \mathcal{N}_W(0,T;H)$. Then there is a positive constant C_4, independent of σ or t, such that*

$$\mathbb{E}\left[\sup_{s\leq t}\left\|\int_0^t \sigma_s dW_s\right\|_H^4\right] \leq C_4 \mathbb{E}\left[\left(\int_0^t \|\sigma_s Q^{\frac{1}{2}}\|_{HS}^2 ds\right)^2\right].$$

This finishes our section with preliminary results.

III.3 The Weak Law of Large Numbers

In this section, we show our main result on the law of large numbers for Volterra-type stochastic integrals in Hilbert space with operator-valued volatility processes. Consider, for some $\mathcal{F}_0$-measurable H-valued Y_0,

$$Y_t := \mathcal{S}(t)Y_0 + \int_0^t \mathcal{S}(t-s)\alpha_s ds + \int_0^t \mathcal{S}(t-s)\sigma_s dW_s, \tag{III.5}$$

where W is a Q-Wiener process on the separable Hilbert space U, σ is an element of $\mathcal{N}_W(0,T;H)$, $(\mathcal{S}(t))_{t\geq 0}$ is a C_0-semigroup on H and α is an almost surely square integrable (in the Bochner sense) predictable process with values in H. We assume that we observe Y at times $t_i := i\Delta_n$ for $\Delta_n = 1/n$, $i = 1, \ldots, \lfloor t/\Delta_n \rfloor$ and define the semigroup-adjusted increment

$$\widetilde{\Delta}_n^i Y := Y_{t_i} - \mathcal{S}(\Delta_n)Y_{t_{i-1}} = \int_{t_{i-1}}^{t_i} \mathcal{S}(t_i - s)\alpha_s ds + \int_{t_{i-1}}^{t_i} \mathcal{S}(t_i - s)\sigma_s dW_s. \tag{III.6}$$

We define the process of the semigroup-adjusted realised covariation (SARCV) as

$$t \mapsto \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} (\tilde{\Delta}_n^i Y)^{\otimes 2}.$$

The aim is to prove the following weak law of large numbers for the SARCV

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} (\tilde{\Delta}_n^i Y)^{\otimes 2} \overset{ucp}{\to} \int_0^t \sigma_s Q \sigma_s^* ds, \qquad \text{as } n \to \infty,$$

in the ucp-topology, that is, for all $\epsilon > 0$ and $T > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{0\leq t\leq T}\left\|\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} (\tilde{\Delta}_n^i Y)^{\otimes 2} - \int_0^t \sigma_s Q \sigma_s^* ds\right\|_{\text{HS}} > \epsilon\right) = 0. \tag{III.7}$$

III.3.1 The Main Result

As we use the notation quite frequently, we will write $\|\cdot\| := \|\cdot\|_H$ and $\langle\cdot,\cdot\rangle := \langle\cdot,\cdot\rangle_H$ in what follows. We will first impose a moment condition to hold for the drift and volatility processes, which will later be weakened by localization:

Assumption 3. Assume that for $T > 0$ the following moment conditions hold:

$$\mathbb{E}\left[\int_0^T \|\alpha_s\|^2\, ds\right] + \mathbb{E}\left[\left(\int_0^T \|\sigma_s Q^{\frac{1}{2}}\|_{\text{HS}}^2 ds\right)^2\right] \leq C(T), \tag{III.8}$$

for some constant $C(T) > 0$.

Remark III.3.1. Using the Cauchy-Schwarz inequality, we can deduce under Assumption 3

$$\mathbb{E}\left[\int_0^t \|\sigma_s Q^{1/2}\|_{\text{HS}}^2 ds\right] \leq \mathbb{E}\left[\left(\int_0^t \|\sigma_s Q^{1/2}\|_{\text{HS}}^2 ds\right)^2\right]^{\frac{1}{2}} \leq \sqrt{C(T)} < \infty.$$

Thus, the integrability condition on $(\sigma_t)_{t\in[0,T]}$ holds for predictable processes satisfying Assumption 3.

Denote for $t \geq 0$

$$M(t) := \sup_{x\in[0,t]} \|\mathcal{S}(x)\|_{op}, \tag{III.9}$$

which is finite by the Hille-Yosida bound on the semigroup. In order to prove the ucp-convergence (III.7) we will first show the following stronger result, which can be used to derive convergence rates under Assumption 3:

Theorem III.3.2. *Assume that Assumption 3 holds for some $T > 0$. Then there exist constants $L_1(T), L_2(T), L_3(T) > 0$ such that*

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\left\|\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}(\tilde{\Delta}_n^i Y)^{\otimes 2} - \int_0^t \sigma_s Q \sigma_s^* ds\right\|_{HS}\right] \leq L_1(T)\Delta_n^{\frac{1}{2}} + L_2(T)a_n(T) + L_3(T)b_n(T), \tag{III.10}$$

where

$$a_n(T) := \mathbb{E}\left[\left(\sup_{i=1,\dots,\lfloor T/\Delta_n\rfloor+1}\int_{t_{i-1}}^{\min(t_i,T)} \left\|\sigma_s Q^{\frac{1}{2}}\right\|_{HS}^2 ds\right)^2\right]^{\frac{1}{4}}, \tag{III.11}$$

$$b_n(T) := \left(\int_0^T \sup_{x\in[0,\Delta_n]} \mathbb{E}[\|(I-\mathcal{S}(x))\sigma_s Q^{\frac{1}{2}}\|_{op}^2]ds\right)^{\frac{1}{2}}, \tag{III.12}$$

and

$$L_1(T) := M(\Delta_n)^2 \left(\Delta_n^{\frac{1}{2}} C(T) + 2C(T)^{\frac{3}{4}} \right), \tag{III.13}$$

$$L_2(T) := M(\Delta_n)^2 C(T)^{\frac{1}{4}} \left(8(1+C_4)\right)^{\frac{1}{2}} + a_n, \tag{III.14}$$

$$L_3(T) := \left(1 + M(\Delta_n)\right) C(T)^{\frac{1}{4}}, \tag{III.15}$$

where C_4 is the universal constant from Lemma III.2.6 and $C(T)$ is the constant from Assumption 3. Moreover, a_n and b_n converge to 0 *and we have*

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{0\le t\le T} \left\| \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} (\tilde{\Delta}_n^i Y)^{\otimes 2} - \int_0^t \sigma_s Q \sigma_s^* ds \right\|_{HS} \right] = 0.$$

Remark III.3.3. The precise forms of $L_1(T)$, $L_2(T)$ and $L_3(T)$ follow by combining equations (III.38), (III.40) and (III.47) below. One should observe that their magnitude can shrink with larger values of n.

That $(a_n(T)))_{n\in\mathbb{N}}$ converges to 0, follows from the integrability condition in Assumption 3 and the implied uniform continuity of the mapping

$$t \mapsto \int_0^t \left\| \sigma_s Q^{\frac{1}{2}} \right\|_{\text{HS}}^2 ds.$$

Observe, that in many cases we may assume the volatility to have integrable fourth moments, i.e.

$$\int_0^T \mathbb{E}\left[\left\| \sigma_s Q^{\frac{1}{2}} \right\|_{\text{HS}}^4 \right] ds < \infty. \tag{III.16}$$

In this case we have $a_n = \mathcal{O}(\Delta_n^{1/4})$, as it is easy to see that

$$a_n(T) \le \left(\int_0^T \mathbb{E}\left[\left\| \sigma_s Q^{\frac{1}{2}} \right\|_{\text{HS}}^4 \right] ds \right)^{\frac{1}{4}} \Delta_n^{\frac{1}{4}}.$$

If we further assume that

$$\mathbb{E}\left[\sup_{s\in[0,T]} \left\| \sigma_s Q^{\frac{1}{2}} \right\|_{\text{HS}}^4 \right] < \infty, \tag{III.17}$$

then we even have $a_n = \mathcal{O}(\Delta_n^{1/2})$ as

$$a_n(T) \le \left(\mathbb{E}\left[\sup_{s\in[0,T]} \left\| \sigma_s Q^{\frac{1}{2}} \right\|_{\text{HS}}^4 \right] \right)^{\frac{1}{4}} \Delta_n^{\frac{1}{2}}.$$

That $(b_n(T))_{n\in\mathbb{N}}$ converges to 0 is an implication of Proposition III.5.1 below. The magnitude of this sequence essentially determines the rate of convergence of the realised covariation by virtue of inequality (III.10). We will come back to the magnitude of the b_n's in specific cases in Section III.4.1.

A localisation argument yields the general law of large numbers

Theorem III.3.4. *Assume $\sigma \in \mathcal{N}_W(0,T;H)$, i.e. it is stochastically integrable, and the drift α is almost surely square integrable, i.e.*

$$\mathbb{P}\left[\int_0^T \|\alpha_s\|^2 ds < \infty\right] = 1.$$

Then

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{0\le s\le t}\left\|\sum_{i=1}^{\lfloor s/\Delta_n\rfloor}(\tilde{\Delta}_i^n Y)^{\otimes 2} - \int_0^s \sigma_u Q\sigma_u^* du\right\|_{HS} > \epsilon\right) = 0, \tag{III.18}$$

We also emphasize, that the following holds:

Corollary III.3.5. *Let $(\bar{Y}_t)_{t\in[0,T]}$ be another process on another separable Hilbert space $\bar{H}$ of the form*

$$\bar{Y}_t := \bar{\mathcal{S}}(t)\bar{Y}_0 + \int_0^t \bar{\mathcal{S}}(t-s)\bar{\alpha}_s ds + \int_0^t \bar{\mathcal{S}}(t-s)\bar{\sigma}_s dW_s, \tag{III.19}$$

where $\bar{Y}_0$ is $\mathcal{F}_0$-measurable with values in H, $\bar{\sigma}$ is an element of $\mathcal{N}_W(0,T;\bar{H})$, $(\bar{\mathcal{S}}(t))_{t\ge 0}$ is a C_0-semigroup on $\bar{H}$ and $\bar{\alpha}$ is an almost surely square integrable (in the Bochner sense) predictable process. We have with respect to the Hilbert-Schmidt norm-topology on $L_{HS}(H,\bar{H})$

$$\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}(Y_{t_i} - \mathcal{S}(\Delta_n)Y_{t_{i-1}})\otimes(\bar{Y}_{t_i} - \bar{\mathcal{S}}(\Delta_n)\bar{Y}_{t_{i-1}}) \overset{ucp}{\to} \int_0^t \bar{\sigma}Q_s\sigma_s^* ds.$$

Proof. We define the process $\hat{Y} := (Y,\bar{Y})^\top$ on the Hilbert space $H\times\bar{H}$ equipped with the scalar product

$$\langle (h,\bar{h})^\top, (g,\bar{g})^\top\rangle_{H\times\bar{H}} := \langle h,g\rangle_H + \langle \bar{h},\bar{g}\rangle_{\bar{H}}.$$

Moreover, define the strongly continuous semigroup

$$\hat{\mathcal{S}}(t) := \begin{pmatrix}\mathcal{S}(t-s) & 0\\ 0 & \bar{\mathcal{S}}(t-s)\end{pmatrix}, \quad t\ge 0$$

on $H\times\bar{H}$. As

$$\hat{Y}_t = \begin{pmatrix}Y_0\\ \bar{Y}_0\end{pmatrix} + \int_0^t \hat{\mathcal{S}}(t-s)\begin{pmatrix}\alpha_s\\ \bar{\alpha}_s\end{pmatrix} ds + \int_0^t \hat{\mathcal{S}}(t-s)\begin{pmatrix}\sigma_s & 0\\ \bar{\sigma}_s & 0\end{pmatrix} d\begin{pmatrix}W_s\\ 0\end{pmatrix}. \tag{III.20}$$

Denote by P_1 the projection from $H\times\bar{H}$ onto the first component given by $P_1(h,\bar{h})^\top = h$ and by P_2 onto $\bar{H}$ given by $P_2(h,\bar{h}) := \bar{h}$. Both P_1 and P_2 are continuous linear projections, and as (III.20) again is a mild Itô process of the form (III.1), the law of large numbers in Thm. III.3.4 is valid. This is why we obtain

$$\begin{aligned}&\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} (Y_{t_i} - \mathcal{S}(\Delta_n)Y_{t_{i-1}}) \otimes (\bar{Y}_{t_i} - \bar{\mathcal{S}}(\Delta_n)\bar{Y}_{t_{i-1}}) - \int_0^t \bar{\sigma}_s Q \sigma_s^* ds \\ =&P_2 \left(\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} (\hat{Y}_{t_i} - \hat{\mathcal{S}}(\Delta_n)\hat{Y}_{t_{i-1}})^{\otimes 2} - \int_0^t \begin{pmatrix} \sigma_s & 0 \\ \bar{\sigma}_s & 0 \end{pmatrix} \begin{pmatrix} Q & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \sigma_s & 0 \\ \bar{\sigma}_s & 0 \end{pmatrix}^* ds \right) P_1^* \\ &\overset{ucp}{\to} 0.\end{aligned}$$

The Corollary follows. ■

III.4 Applications

In this section, we give an overview of potential settings and scenarios for which we can use the techniques described above to infer volatility.

Stochastic integrals of the form (III.5) arise naturally in correspondence to mild or strong solutions to stochastic partial differential equations. Take as a simple example a process given by

$$\text{(SPDE)} \begin{cases} dY_t = AY_t dt + \sigma_t dW_t, & t \geq 0 \\ Y_0 = h_0 \in H, \end{cases} \tag{III.21}$$

where A is the generator of a C_0-semigroup $(\mathcal{S}(t))_{t\geq 0}$ on the separable Hilbert space H, W is a Q-Wiener process on a separable Hilbert space U for some positive semidefinite and symmetric trace class operator $Q : U \to U$ and $\sigma \in \mathcal{N}_W(0,T;H)$.

There are three components in this model, which need to be estimated in practice: the covariance operator Q of the Wiener process, the generator A (or the semigroup $(\mathcal{S}(t))_{t\geq 0}$ respectively) and the stochastic volatility process σ.

III.4.1 Semigroups

The essence of the convergence result in Theorem III.3.2 is that we can infer on Q and σ based on observing the path of Y, given that we *know* the semigroup $(\mathcal{S}(t))_{t\geq 0}$. Even more, in this case, Theorem 3.2 allows us to derive rates of convergence, which are specified by the behaviour of the semigroup on the volatility. We outline some examples below.

III.4.1.1 Martingale Case

For $A = 0$ and $S(t) = I$ and for all $t \geq 0$, we have the solution

$$Y_t = \int_0^t \sigma_s dW_s,$$

for the stochastic partial differential equation (III.21). Clearly in this case we have

$$b_n(T) = 0.$$

III.4.1.2 Uniformly Continuous Semigroups

Assume that $(\mathcal{S}(t))_{t\geq 0}$ is continuous with respect to the operator norm. This is equivalent to $A \in L(H)$ and $\mathcal{S}(t) = e^{tA}$.

Lemma III.4.1. *Let Assumption 3 hold. If the semigroup $(\mathcal{S}(t))_{t\geq 0}$ is uniformly continuous, we have, for b_n given in* (III.12)*, that*

$$b_n(T) \leq \Delta_n \|A\|_{op} e^{\|A\|_{op}\Delta_n} C(T)^{\frac{1}{4}}.$$

Proof. Recall the following fundamental equality from semigroup theory (cf. [26, Lemma II.1.3]):

$$(\mathcal{S}(x) - I)h = \int_0^x A\mathcal{S}(s)hds, \qquad \forall h \in D(A). \tag{III.22}$$

Using (III.22), we get

$$\begin{aligned}\sup_{x\in[0,\Delta_n]} \|(I - \mathcal{S}(x))\|_{\text{op}} &= \sup_{x\in[0,\Delta_n]} \sup_{\|h\|=1} \left\|\int_0^x A\mathcal{S}(s)hds\right\| \\ &\leq \sup_{x\in[0,\Delta_n]} x\|A\|_{\text{op}} e^{\|A\|_{\text{op}}x} = \Delta_n \|A\|_{\text{op}} e^{\|A\|_{\text{op}}\Delta_n}.\end{aligned}$$

It follows that

$$\begin{aligned}b_n^2(T) &= \int_0^T \mathbb{E}[\sup_{x\in[0,\Delta_n]} \|(I - \mathcal{S}(x))\sigma_s Q^{\frac{1}{2}}\|_{\text{op}}^2]ds \\ &\leq \sup_{x\in[0,\Delta_n]} \|(I - \mathcal{S}(x))\|_{\text{op}}^2 \int_0^T \mathbb{E}[\|\sigma_s Q^{\frac{1}{2}}\|_{\text{HS}}^2]ds \\ &\leq \Delta_n^2 \|A\|_{\text{op}}^2 e^{2\|A\|_{\text{op}}\Delta_n} \mathbb{E}\left[\left(\int_0^T \|\sigma_s Q^{\frac{1}{2}}\|_{\text{HS}}^2 ds\right)^2\right]^{\frac{1}{2}},\end{aligned}$$

and the claim follows. ∎

For uniformly continuous semigroups and if (III.17) holds, we obtain a convergence speed of the order $\Delta_n^{1/2}$ for the convergence of the adjusted realized covariation to the quadratic covariation in Theorem III.3.2.

Remark III.4.2. Note that, if the semigroup is uniformly continuous and under Assumption 3, we can get back to a case similar to Section III.4.1.1: As A is continuous, $(Y_t)_{t\in[0,T]}$ is a strong solution to the SPDE (III.21) and therefore takes the form

$$Y_t = Y_0 + \int_0^t AY_s ds + \int_0^t \sigma_s dW_s.$$

As the drift process given by $\alpha_s = AY_s$ is square-integrable, we can choose the semigroup equal to the identity and therefore the law of large numbers holds without any the adjustment, i.e. we have the convergence of the (nonadjusted) realised covariation

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \left(Y_{t_i} - Y_{t_{i-1}}\right)^{\otimes 2} \overset{ucp}{\to} \int_0^t \sigma_s Q \sigma_s^* ds.$$

By definition $b_n(T) = 0$ in this case and if (III.17) holds, the rate of convergence is $\mathcal{O}(\Delta_n^{1/2})$, similar to the case for the adjusted realised covariation.

Let us turn our attention to a case of practical interest coming from financial mathematics applied to commodity markets.

III.4.1.3 Forward Contracts in Commodity and Interest Rate Markets: the Heath-Jarrow-Morton Approach

A case of relevance for our analysis is inference on the volatility for forward prices in commodity markets as well as for forward rates in fixed-income markets. The Heath-Jarrow-Morton-Musiela equation (HJMM-equation) describes the term structure dynamics in both of these settings (see [28] for a detailed motivation for the use in interest rate modelling and [12] its use in commodity markets) and is given by

$$\text{(HJMM)} \begin{cases} dX_t = (\frac{d}{dx} X_t + \alpha_t)dt + \sigma_t dW_t, & t \geq 0 \\ X_0 = h_0 \in H, \end{cases} \tag{III.23}$$

where H is a Hilbert space of functions $f : \mathbb{R}_+ \to \mathbb{R}$ (the *forward curve space*), $(\alpha_t)_{t\geq 0}$ is a predictable and almost surely locally Bochner-integrable stochastic process and σ and W are as before. Conveniently, the states of this *forward curve dynamics* are realized on the separable Hilbert space

$$H = H_\beta = \{h : \mathbb{R}_+ \to \mathbb{R} : h \text{ is absolutely continuous and } \|h\|_\beta < \infty\}, \tag{III.24}$$

for fixed $\beta > 0$, where the inner product is given by

$$\langle h, g \rangle_\beta = h(0)g(0) + \int_0^\infty h'(x)g'(x)\mathrm{e}^{\beta x} dx,$$

and norm $\|h\|_\beta^2 = \langle h, h \rangle_\beta$. This space was introduced and analysed in [28]. As in [28], one may consider more general scaling functions in the inner product than the exponential $\exp(\beta x)$. However, for our purposes here this choice suffices. The suitability of this space is partially due to the following result:

Lemma III.4.3. *The differential operator $A = \frac{d}{dx}$ is the generator of the strongly continuous semigroup $(\mathcal{S}(t))_{t\geq 0}$ of shifts on H_β, given by $\mathcal{S}(t)h(x) = h(x+t)$, for $h \in H_\beta$, such that*

$$M(\Delta_n) = \sup_{t \leq \Delta_n} \|\mathcal{S}(t)\|_{op} \leq e^{\Delta_n}. \tag{III.25}$$

Proof. See for example [28]. For the quasi-contractive property (III.25) compare [12, Theorem 3.5]. ■

The HJMM-equation (III.23) possesses a mild solution (see e.g. [40])

$$f_t = \mathcal{S}(t)f_0 + \int_0^t \mathcal{S}(t-s)\alpha_s ds + \int_0^t \mathcal{S}(t-s)\sigma_s dW_s. \tag{III.26}$$

Since forward prices and rates are often modelled under a risk neutral probability measure, the drift has in both cases (commodities and interest rates) a special form. In the case of forward prices in commodity markets, it is zero under the risk neutral probability, whereas in interest rate theory it is completely determined by the volatility via the no-arbitrage drift condition

$$\alpha_t = \sum_{j\in\mathbb{N}} \sigma_t^j \Sigma_t^j, \quad \forall t \in [0,T], \tag{III.27}$$

where $\sigma_t^j = \sqrt{\lambda_j}\sigma_t(e_j)$ and $\Sigma_t^j = \int_0^t \sigma_s^j ds$ for some eigenvalues $(\lambda_j)_{j\in\mathbb{N}}$ and a corresponding basis of eigenvectors $(e_j)_{j\in\mathbb{N}}$ of the covariance operator Q of W (cf. Lemma 4.3.3 in [28]).

Lemma III.4.4. *Assume that the volatility process $(\sigma_t)_{t\in[0,T]}$ satisfies* (III.17) *and that for each $t \in [0,1]$ the operator σ_t maps into*

$$H_\beta^0 = \{h \in H_\beta : \lim_{x\to\infty} h(x) = 0\}.$$

Then the drift given by (III.27) *has values in H_β, is predictable, and has finite second moments.*

Proof. That the drift is well defined follows from Lemma 5.2.1 in [28]. Predictability follows immediately from the predictability of the volatility. We have by Theorem 5.1.1 from [28] that there is a constant K depending only on β such that

$$\|\sigma_t^j \Sigma_t^j\|_\beta \leq K\|\sigma_t^j\|_\beta^2.$$

Therefore, we get by the triangle inequality that

$$\|\alpha_t\|_\beta \leq K \sum_{j\in\mathbb{N}} \|\sigma_t^j\|_\beta^2 = K\|\sigma_t Q^{\frac{1}{2}}\|_{\text{HS}}^2.$$

We obtain the finite second-moment property by (III.17) as

$$\sup_{t\in[0,T]} \mathbb{E}[\|\alpha_t\|_\beta^2] \leq K^2 \sup_{t\in[0,T]} \mathbb{E}[\|\sigma_t Q^{\frac{1}{2}}\|_{\text{HS}}^4].$$

Moreover, the Bochner integrability follows, since we have

$$\mathbb{E}\left[\int_0^T \|\alpha_t\|_\beta dt\right] \leq \int_0^T \mathbb{E}[\|\alpha_t\|_\beta^2]^{\frac{1}{2}} dt \leq TK \sup_{t\in[0,T]} \mathbb{E}[\|\sigma_t Q^{\frac{1}{2}}\|_{\text{HS}}^4]^{\frac{1}{2}} < \infty.$$

The result follows. ■

Remark III.4.5. Since we know the exact form of the semigroup $(S(t))_{t\geq 0}$, we can recover the adjusted increments $\tilde{\Delta}_n^i f$ efficiently from forward curve data by a simple shifting in the spatial (e.g., time-to-maturity) variable of these curves. Theorem III.3.2 and more generally Theorem III.3.4 can therefore be applied in practice to make inference on σ.

The shift semigroup is strongly, but not uniformly, continuous, leaving us with the question to determine the convergence speed of the estimator established in Corollary III.10. We close this subsection by deriving a convergence bound under regularity conditions of the volatility in the space variable (that is time to maturity).

Observe that by Theorem 4.11 in [12] we know that for all $r \in [0, T]$ there exist random variables c_r with values in $\mathbb{R}$, f_r, g_r with values in H such that $g_r(0) = 0 = f_r(0)$ and $p_r \in L^2(\mathbb{R}_+^2)$ such that we have

$$\sigma_r Q^{\frac{1}{2}} h(x) = c_r h(0) + \langle g_r, h\rangle_\beta + h(0) f_r(x) + \int_0^\infty q_r(x,z) h'(z) dz,$$

where $q_r(x,z) = \int_0^x p_r(y,z) e^{\frac{\beta}{2}(z-y)} dy$. We denote by $C^{1,\gamma}_{\text{loc}} := C^{1,\gamma}_{\text{loc}}(\mathbb{R}_+)$ the space of continuously differentiable functions with locally γ-Hölder continuous derivative for $\gamma \in (0,1]$. The proof of the following result can be found in Section III.5.2.

Theorem III.4.6. *Assume that $f_r, q_r(\cdot, z) \in C^{1,\gamma}_{loc}$ for all $z \geq 0$, $r \in [0,T]$ and that for the corresponding local Hölder constants $L^1_r(x)$ of $e^{\frac{\beta}{2}\cdot} f'_r(\cdot)$ and $L^2_r(x,z)$ of p_r, we have that for all $x \in [0,1]$*

$$|e^{\beta(x+y)} f'_r(x+y) - e^{\beta x} f'_r(x)| \leq L^1_r(x) y^\gamma$$

and

$$|p(y+x,z) - p(x,z)| \leq L^2_r(x,z) y^\gamma.$$

Moreover, we assume that L^1_r and L^2_r are square integrable in x and in (x,z) respectively such that for some $\zeta \in (0,T)$

$$\hat{L} := \left(\int_0^T \mathbb{E}\left[\left(|f'_r(\zeta)| + \sqrt{8}\left(\frac{e^{\frac{\beta+1}{2}}}{\beta}\right)\|f_r\|_\beta + \sqrt{2}\|L^1_r\|_{L^2(\mathbb{R}_+)} + \|L^2_r\|_{L^2(\mathbb{R}^2_+)} + \left(1+\frac{\beta}{2}\right)\|p_r\|_{L^2(\mathbb{R}^2_+)} \right)^2 \right] dr \right)^{\frac{1}{2}} < \infty.$$

Then for $b_n(T)$ as given in (III.12)*, we can estimate*

$$b_n(T) \leq \hat{L} \Delta_n^{\min(\gamma, \frac{1}{2})}.$$

In the next section, we investigate the asymptotic behaviour for different stochastic volatility models.

III.4.2 Stochastic Volatility Models

In this section different models for stochastic volatility in Hilbert spaces are discussed. So far, infinite-dimensional stochastic volatility models are specified by stochastic partial differential equations on the positive cone of Hilbert-Schmidt operators (see [13], [14]). We will check therefore, which models satisfy Assumption 3. Throughout this section, we take $H = U$ for simplicity.

III.4.2.1 Constant Volatility

We start with the simple but important special case of constant volatility, i.e. $\sigma_s = I$ for all $s \in [0,T]$ and we want to make inference on Q. In this case (III.17) is trivially fulfilled and it is easy to see that $C(T) \leq T^2 \mathrm{Tr}(Q)^2$. The convergence rate is thus $\mathcal{O}(\Delta_n^{1/2} + b_n(T))$. The magnitude of $b_n(T)$ is now completely dependent on the range of the square root of the covariance operator $Q^{\frac{1}{2}}$. We define

$$\begin{aligned}\tilde{Z}_n(i) :=& (\tilde{\Delta}_i^n Y)^{\otimes 2} - \int_{t_{i-1}}^{t_i} \mathcal{S}(t_i - s) Q \mathcal{S}(t_i - s)^* ds \\ =& (\tilde{\Delta}_i^n Y)^{\otimes 2} - \int_0^{\Delta_n} \mathcal{S}(\Delta_n - s) Q \mathcal{S}(\Delta_n - s)^* ds. \qquad \text{(III.28)}\end{aligned}$$

It is interesting to note the following: As the sequence $(\tilde{Z}_n(i))_{i\in\mathbb{N}}$ is a centred i.i.d. sequence of random variables, we also obtain a convergence result, if $T \to \infty$ and Δ_n is constant. Namely, the classical law of large numbers in Hilbert spaces, see e.g. [16, Theorem 2.4], yields

$$\lim_{T\to\infty} \frac{1}{\lfloor T/\Delta_n \rfloor} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \tilde{Z}_n(i) \stackrel{u.c.p.}{\longrightarrow} 0.$$

If the semigroup is the identity, this again yields a consistent way of estimating Q, which is analogous to the finite-dimensional case. However, note that, if the semigroup is not equal to the identity, the long time estimator (III.28) estimates $\int_0^{\Delta_n} \mathcal{S}(\Delta_n - s) Q \mathcal{S}(\Delta_n - s)^* ds$, rather than $\Delta_n Q$.

III.4.2.2 Barndorff-Nielsen & Shephard (BNS) Model

The volatility is oftentimes given as the unique positive square-root of a process Σ_t, e.g.,

$$\sigma_t := \Sigma_t^{\frac{1}{2}}, \qquad \text{(III.29)}$$

where Σ takes values in the set of positive Hilbert-Schmidt operators on H. This is for instance the case in the Hilbert space-valued volatility model suggested in [13], extending the BNS-model introduced in [6] to infinite dimensions. There Σ is given by the Ornstein-Uhlenbeck dynamics

$$(BNS) \begin{cases} d\Sigma_t = \mathbb{B}\Sigma_t dt + d\mathcal{L}_t, \\ \Sigma_0 \in L_{\mathrm{HS}}(H), \end{cases}$$

where $\mathbb{B}$ is a positive bounded linear operator on the space of Hilbert-Schmidt operators $L_{\text{HS}}(H)$, $\mathcal{L}$ is a square integrable Lévy subordinator on the same space and Σ_0 is also positive definite. $\mathbb{B}$ is then the generator of the uniformly continuous semigroup given by $\mathbb{S}(t) = \exp(\mathbb{B}t)$ and the equation has a mild solution given by

$$\Sigma_t = \mathbb{S}(t)\Sigma_0 + \int_0^t \mathbb{S}(t-s)d\mathcal{L}_s,$$

which defines a stochastically integrable process in $\mathcal{N}_W(0,T;H)$ (see [13]). We have

$$\sup_{s\in[0,T]} \mathbb{E}[\|\sigma_s\|_{\text{op}}^4] = \sup_{s\in[0,T]} \mathbb{E}[\|\Sigma_s^{\frac{1}{2}}\|_{\text{op}}^4] = \sup_{s\in[0,T]} \mathbb{E}[\|\Sigma_s\|_{\text{HS}}^2].$$

By the Itô isometry, we obtain

$$\begin{aligned}
\sup_{t\in[0,T]} \mathbb{E}[\|\Sigma_t\|_{\text{HS}}^2]^{\frac{1}{2}} &\leq \sup_{t\in[0,T]} \left(\|\mathbb{S}(t)\Sigma_0\|_{\text{HS}} + \mathbb{E}\left[\left\| \int_0^t \mathbb{S}(t-u)d\mathcal{L}_u \right\|_{\text{HS}}^2 \right]^{\frac{1}{2}} \right) \\
&\leq \sup_{t\in[0,T]} \left(\|\mathbb{S}(t)\Sigma_0\|_{\text{HS}} + \left(\int_0^T \|\mathbb{S}(t-u)Q_{\mathcal{L}}^{\frac{1}{2}}\|_{\text{HS}}^2 du \right)^{\frac{1}{2}} \right) \\
&\leq e^{\|\mathbb{B}\|_{\text{op}}T}\|\Sigma_0\|_{\text{HS}} + e^{\|\mathbb{B}\|_{\text{op}}T}\text{Tr}(Q_{\mathcal{L}})^{\frac{1}{2}}T^{\frac{1}{2}},
\end{aligned}$$

where $Q_{\mathcal{L}}$ denotes the covariance operator of $\mathcal{L}$. This yields that we can find an upper bound for the constant $C(T)$ from Assumption 3 according to

$$\begin{aligned}
C(T) &\leq T^2 \sup_{s\in[0,T]} \mathbb{E}\left[\left\| \sigma_s Q^{\frac{1}{2}} \right\|_{\text{op}}^4 \right] \\
&\leq \text{Tr}(Q)^2 T^2 \left(e^{\|\mathbb{B}\|_{\text{op}}T}\|\Sigma_0\|_{\text{HS}} + e^{\|\mathbb{B}\|_{\text{op}}T}\text{Tr}(Q_{\mathcal{L}})^{\frac{1}{2}}T^{\frac{1}{2}} \right)^2.
\end{aligned}$$

Moreover, it is easy to see that (III.16) holds, which is why the rate of convergence in the law of large numbers Thm. III.3.2 becomes $\mathcal{O}(b_n(T) + \Delta_n^{1/4})$. Now we can combine this result with the ones from the previous section (for instance for the term structure models) and obtain explicit expressions for the constants $L_1(T)$, $L_2(T)$ and $L_3(T)$ from Theorem III.3.2.

It is also possible to derive ucp convergence for rough volatility models, which we present in the following section.

III.4.2.3 Rough Volatility Models

In [11] pathwise constructions of Volterra processes are established and suggested for use in stochastic volatility models. In this setting, a process is mostly known to be Hölder continuous almost surely of some particular order.

If H is a Banach algebra (like the forward curve space defined by (III.24)), we can define the volatility process by

$$\sigma_t h := \exp(\mathcal{Y}_t)h. \tag{III.30}$$

This is a direct extension of the volatility models proposed in [29], if we define the rough process $\mathfrak{Y}_t$ as follows: For $\rho > 0$ and a locally Bochner integrable function $f : \mathbb{R}_+ \to H$ define the fractional integral operator $I^\rho \in L(L^1_{\text{loc}}(\mathbb{R}_+, H))$ as

$$I^\rho(f)(t) := \frac{1}{\Gamma(\rho)} \int_0^t (t-s)^{\rho-1} f(s)ds. \tag{III.31}$$

For the special case of $\rho = 0$ we set $I^0 = id_{L^1_{loc}(\mathbb{R}_+;H)}$. Define a noise term $\mathfrak{X}$ as the Gaussian process

$$\mathfrak{X}(t) = \int_0^t (t-s)^{\rho-1} d\mathfrak{W}(s). \tag{III.32}$$

This integral is well defined pathwise via a Sewing Lemma in Banach spaces (see [11, Prop. 14]), for a process $\mathfrak{W}$ with sample paths in space of γ-Hölder continuous functions $C^\gamma([0,T];H)$ on H, such that $\rho + \gamma - 1 > 0$. For an initial condition $y \in H$, an Ornstein-Uhlenbeck process in this framework is considered to be the solution to the integral equation

$$\mathfrak{Y}_t = y + I^\rho(\mathbb{A}\mathcal{Y})_t + \mathfrak{X}_t \quad t \in [0,T], \tag{III.33}$$

where $\mathbb{A} \in L(H)$. It was shown in [11, Prop. 26] that this pathwise integral equation possesses a unique solution $\mathfrak{Y}$ with sample paths in $C^\alpha([0,T];H)$ for $0 < \alpha < \rho + \gamma - 1$. This solution is moreover Gaussian and hence, by virtue of Fernique's theorem, cf. [40, Theorem 3.31], satisfies (III.16), which is why the rate of convergence is $\mathcal{O}(b_n(T) + \Delta_n^{1/4})$. More precisely, the cross-covariance structure is characterized by

$$\begin{aligned} Q_{\mathfrak{Y}}(t,t') &:= \mathbb{E}\left[\mathfrak{Y}_t \otimes \mathfrak{Y}_{t'}\right] \\ &= \int_0^t \int_0^{t'} (t-r)^{\rho-1} E_{\rho,\rho}(\mathbb{A}(t-r)^\rho) d^2 \mathcal{Q}_{\mathfrak{W}}(r,r')(t'-r')^{\rho-1} E_{\rho,\rho}(\mathbb{A}^*(t'-r')^\rho), \end{aligned}$$

where for $\mathbb{B} \in L(H)$

$$E_{\alpha,\beta}(\mathbb{B}) := \sum_{i=1}^\infty \frac{\mathbb{B}^i}{\Gamma(\alpha i + \beta)}$$

is the Mittag-Leffler operator and Γ is the Gamma-function. From these analytic expressions, one can derive again explicit formulas for the constants $L_1(T)$, $L_2(T)$ and $L_3(T)$.

III.5 Proofs

In this section, we will present the proofs of our previously stated results.

III.5.1 Proofs of Results in Section III.3

III.5.1.1 Uniform Continuity of Semigroups on Compact Sets

In order to verify that $b_n(T)$ defined in (III.12) converges to 0 and to prove Theorem III.3.2, we need to establish some convergence properties of semigroups on compacts.

The next proposition follows from Dini's theorem and will be important for our analysis:

Proposition III.5.1. *Let U, H be two separable Hilbert spaces. The following holds:*

(i) *If σ is an almost surely compact random linear operator with values in $L(U,H)$, we get that*

$$\sup_{x\in[0,\Delta_n]} \|(I-\mathcal{S}(x))\sigma\|_{op} \to 0, \qquad \text{as } n\to\infty, \tag{III.34}$$

where the convergence holds almost surely. If furthermore $\sigma \in L^p(\Omega; L(U,H))$ for some $p\in[1,\infty)$, the convergence holds also in $L^p(\Omega;\mathbb{R})$.

(ii) *Assume that $s\mapsto \sigma_s Q^{\frac{1}{2}}$ is a stochastic process of almost surely compact operators, such that*

$$\mathbb{P}\left[\int_0^T \|\sigma_s\|_{op}^p\, ds < \infty\right] = 1,$$

for $p\in[1,\infty)$. Then almost surely

$$\int_0^T \sup_{x\in[0,\Delta_n]} \|(I-\mathcal{S}(x))\sigma_s\|_{op}^p ds \to 0, \qquad \text{as } n\to\infty. \tag{III.35}$$

If $\int_0^T \mathbb{E}\left[\left\|\sigma_s Q^{\frac{1}{2}}\right\|_{op}^p\right] ds < \infty$, then

$$\int_0^T \mathbb{E}[\sup_{x\in[0,\Delta_n]} \|(I-\mathcal{S}(x))\sigma_s\|_{op}^p] ds \to 0, \qquad \text{as } n\to\infty. \tag{III.36}$$

Proof. Let $B_0(1) := \{h\in H : \|h\| = 1\}$ be the unit sphere in H and fix $\omega\in\Omega$, such that $\sigma(\omega)$ is compact. Since $\sigma(\omega)$ is compact, $\mathcal{C} := \overline{\sigma(\omega)(B_0(1))}$ is compact in H. We define the set $F(\omega)$ of functionals of the form

$$f_n := \sup_{x\in[0,\Delta_n]} \|(I-\mathcal{S}(x))\cdot\| : \mathcal{C}\to\mathbb{R}.$$

The functions in $F(\omega)$ are continuous, as

$$\begin{aligned}
&|\sup_{x\in[0,\Delta_n]} \|(I-\mathcal{S}(x))h\| - \sup_{x\in[0,\Delta_n]} \|(I-\mathcal{S}(x))g\||\\
&\quad\le \sup_{x\in[0,\Delta_n]} \|(I-\mathcal{S}(x))(h-g)\|\\
&\quad\le \sup_{x\in[0,\Delta_1]} \|(I-\mathcal{S}(x))\|_H \|h-g\|,
\end{aligned}$$

for all $g, h \in \mathcal{C}$. Hence Dini's theorem (cf. Theorem 7.13 in [44]) yields (III.34) in the almost sure sense. Since the sequence is uniformly bounded by $(1+M(T))\|\sigma\|_{op}$, which has finite pth moment, we obtain $L^p(\Omega;\mathbb{R})$-convergence by the dominated convergence theorem, and therefore (III.34) holds in the L^p-sense.

The convergences (III.35) and (III.36) follow now immediately by appealing to the dominated convergence theorem, as

$$\sup_{x\in[0,\Delta_n]} \|(I-\mathcal{S}(x))\sigma_s\|_{op}^p \leq M(\Delta_n)^p \|\sigma_s\|_{op}^p$$

and $\sup_{x\in[0,\Delta_n]} \|(I-\mathcal{S}(x))\sigma_s\|_{op}^p$, respectively $\mathbb{E}\left[\sup_{x\in[0,\Delta_n]} \|(I-\mathcal{S}(x))\sigma_s\|_{op}^p\right]$, converges to 0 by (III.34). ■

Recall also the following fact:

Lemma III.5.2. *The family $(\mathcal{S}(t)^*)_{t\geq 0}$ of adjoint operators of the C_0-semigroup $(\mathcal{S}(t))_{t\geq 0}$ forms again a C_0-semigroup on H.*

Proof. See Section 5.14 in [26]. ■

Now we can proceed with the proof of our main theorem in the next subsection.

III.5.1.2 Elimination of the Drift

The drift process will not affect the asymptotic behaviour of the realised covariation. This is proved in the next Lemma:

Lemma III.5.3. *To prove Theorem III.3.2, we can without loss of generality assume $\alpha \equiv 0$ and $Y_0 \equiv 0$.*

Proof. That we can assume $Y_0 \equiv 0$ can be seen immediately as

$$\widetilde{\Delta}_n^i Y := Y_{t_i} - \mathcal{S}(\Delta_n)Y_{t_{i-1}} = \int_{t_{i-1}}^{t_i} \mathcal{S}(t_i - s)\alpha_s ds + \int_{t_{i-1}}^{t_i} \mathcal{S}(t_i-s)\sigma_s dW_s$$

is not dependent on the initial condition. We can then argue for the drift as follows: We have

$$\begin{aligned}
&\mathbb{E}\left[\sup_{0\leq t\leq T}\left\|\sum_{i=1}^{\lfloor t/\Delta_n\rfloor} (\tilde{\Delta}_n^i Y)^{\otimes 2} - \int_0^t \sigma_s Q \sigma_s^* ds\right\|_{\text{HS}}\right] \\
\leq &\mathbb{E}\left[\sup_{0\leq t\leq T}\left\|\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\left(\int_{t_{i-1}}^{t_i} \mathcal{S}(t_i-s)\alpha_s ds\right)^{\otimes 2}\right\|_{\text{HS}}\right] \\
&+\mathbb{E}\left[\sup_{0\leq t\leq T}\left\|\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\left(\int_{t_{i-1}}^{t_i} \mathcal{S}(t_i-s)\alpha_s ds\right)\otimes\left(\int_{t_{i-1}}^{t_i} \mathcal{S}(t_i-s)\sigma_s dW_s\right)\right\|_{\text{HS}}\right]
\end{aligned}$$

$$
\begin{aligned}
&+\mathbb{E}\left[\sup_{0\le t\le T}\left\|\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\left(\int_{t_{i-1}}^{t_i}\mathcal{S}(t_i-s)\sigma_s dW_s\right)\otimes\left(\int_{t_{i-1}}^{t_i}\mathcal{S}(t_i-s)\alpha_s ds\right)\right\|_{\text{HS}}\right]\\
&+\mathbb{E}\left[\sup_{0\le t\le T}\left\|\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\left(\int_{t_{i-1}}^{t_i}\mathcal{S}(t_i-s)\sigma_s dW_s\right)^{\otimes 2}-\int_0^t\sigma_s Q\sigma_s^* ds\right\|_{\text{HS}}\right]\\
\le&\sum_{i=1}^{\lfloor T/\Delta_n\rfloor}\mathbb{E}\left[\left\|\int_{t_{i-1}}^{t_i}\mathcal{S}(t_i-s)\alpha_s ds\right\|^2\right]\\
&\qquad+2\left(\sum_{i=1}^{\lfloor T/\Delta_n\rfloor}\mathbb{E}\left[\left\|\int_{t_{i-1}}^{t_i}\mathcal{S}(t_i-s)\sigma_s dW_s\right\|^2\right]\right)^{\frac{1}{2}}\\
&\qquad\qquad\times\left(\sum_{i=1}^{\lfloor T/\Delta_n\rfloor}\mathbb{E}\left[\left\|\int_{t_{i-1}}^{t_i}\mathcal{S}(t_i-s)\alpha_s ds\right\|^2\right]\right)^{\frac{1}{2}}\\
&+\mathbb{E}\left[\sup_{0\le t\le T}\left\|\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\left(\int_{t_{i-1}}^{t_i}\mathcal{S}(t_i-s)\sigma_s dW_s\right)^{\otimes 2}-\int_0^t\sigma_s Q\sigma_s^* ds\right\|_{\text{HS}}\right]\\
=&(1)+(2)+(3).
\end{aligned}
$$

In order to prove the assertion, we have to show that (1) and (2) converge to 0 as $n\to\infty$. We find by Bochner's inequality

$$
\mathbb{E}\left[\left\|\int_{t_{i-1}}^{t_i}\mathcal{S}(t_i-s)\alpha_s ds\right\|^2\right]\le\Delta_n M^2(\Delta_n)\mathbb{E}\left[\int_{t_{i-1}}^{t_i}\|\alpha_s\|^2 ds\right],
$$

and by the Itô isometry

$$
\mathbb{E}\left[\left\|\int_{t_{i-1}}^{t_i}\mathcal{S}(t_i-s)\sigma_s dW_s\right\|_H^2\right]\le M^2(\Delta_n)\mathbb{E}\left[\int_{t_{i-1}}^{t_i}\|\sigma_s Q^{\frac{1}{2}}\|_{HS}^2 ds\right],
$$

where we appealed to the bound (III.9) on the semigroup. Hence, $(1)+(2)=\mathcal{O}(\Delta_n^{\frac{1}{2}})$, so the first two terms will not impact the estimation of the covariation (in the limit). More precisely we have that

$$
\begin{aligned}
&(1)+(2) \qquad\qquad\qquad\qquad \text{(III.37)}\\
\le&\Delta_n M^2(\Delta_n)\int_0^T\mathbb{E}\left[\|\alpha_s\|^2\right]ds\\
&+2\left(\Delta_n M^2(\Delta_n)\int_0^T\mathbb{E}\left[\|\alpha_s\|^2\right]ds\right)^{\frac{1}{2}}\left(M^2(\Delta_n)\int_0^T\mathbb{E}\left[\left\|\sigma_s Q^{\frac{1}{2}}\right\|_{\text{HS}}^2\right]ds\right)^{\frac{1}{2}}
\end{aligned}
$$

and therefore, given Assumption 3

$$(1) + (2) \leq \Delta_n^{\frac{1}{2}} M(\Delta_n)^2 \left(\Delta_n^{\frac{1}{2}} C(T) + 2C(T)^{\frac{3}{4}} \right). \tag{III.38}$$

The Lemma follows. ■

III.5.1.3 Proof of Theorem III.3.2

In view of Lemma III.5.3 we assume in this subsection that the process Y takes the form $Y_t = \int_0^t \mathcal{S}(t-s)\sigma_s dW_s$. The operator bracket process for the semigroup-adjusted increment takes the form

$$\langle\langle \tilde{\Delta}_n^i Y \rangle\rangle = \int_{t_{i-1}}^{t_i} \mathcal{S}(t_i - s)\sigma_s Q \sigma_s^* \mathcal{S}(t_i - s)^* ds. \tag{III.39}$$

We have

Proposition III.5.4. *Let Assumption 3 hold. Then*

$$\mathbb{E}\left[\sup_{t\in[0,T]} \left\| \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} (\tilde{\Delta}_i^n Y)^{\otimes 2} - \langle\langle \tilde{\Delta}_i^n Y \rangle\rangle \right\| \right] \leq M(\Delta_n)^2 C(T)^{\frac{1}{4}} \left(8(1+C_4)\right)^{\frac{1}{2}} a_n(T). \tag{III.40}$$

Proof. We define

$$\tilde{Z}_n(i) := (\tilde{\Delta}_i^n Y)^{\otimes 2} - \langle\langle \tilde{\Delta}_i^n Y \rangle\rangle = (\tilde{\Delta}_i^n Y)^{\otimes 2} - \int_{t_{i-1}}^{t_i} \mathcal{S}(t_i - s)\sigma_s Q \sigma_s^* \mathcal{S}(t_i - s)^* ds.$$

First we show that $\sup_{t\in[0,T]} \| \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \tilde{Z}_n(i) \|_{\text{HS}}$ has finite second moment. By the triangle inequality and Lemma III.2.1

$$\begin{aligned}
\sup_{t\in[0,T]} \left\| \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \tilde{Z}_n(i) \right\|_{\text{HS}} &\leq \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \|\tilde{Z}_n(i)\|_{\text{HS}} \\
&\leq \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \|(\tilde{\Delta}_i^n Y)^{\otimes 2}\|_{\text{HS}} \\
&\quad + \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \left\| \int_{t_{i-1}}^{t_i} \mathcal{S}(t_i - s)\sigma_s Q \sigma_s^* \mathcal{S}(t_i - s)^* ds \right\|_{\text{HS}} \\
&\leq \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \|\tilde{\Delta}_i^n Y\|_H^2 + M(\Delta_n)^2 \int_0^T \left\| \sigma_s Q^{\frac{1}{2}} \right\|_{\text{HS}}^2 ds.
\end{aligned}$$

Considering $\mathbb{E}\left[\sup_{t\in[0,T]} \| \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \tilde{Z}_n(i)\|_{\text{HS}}^2\right]$, we get a finite sum of linear combinations of the following terms

$$\mathbb{E}\left[\left(\int_0^T \|\sigma_s Q^{\frac{1}{2}}\|_{\text{HS}}^2 ds \right)^2 \right] = C(T), \tag{III.41}$$

$$\mathbb{E}\left[\|\tilde{\Delta}_i^n Y\|^2 \|\tilde{\Delta}_j^n Y\|^2\right], \tag{III.42}$$

$$\int_0^T \mathbb{E}\left[\|\tilde{\Delta}_i^n Y\|^2 \|\sigma_s Q^{\frac{1}{2}}\|_{\mathrm{HS}}^2\right] ds. \tag{III.43}$$

The expression in (III.41) is finite by the imposed Assumption 3. The term in (III.42) is finite, since by the Cauchy-Schwarz inequality

$$\begin{aligned}
&\mathbb{E}\left[\|\tilde{\Delta}_i^n Y\|^2 \|\tilde{\Delta}_j^n Y\|^2\right] \\
\leq &\mathbb{E}\left[\|\tilde{\Delta}_i^n Y\|^4\right]^{\frac{1}{2}} \mathbb{E}\left[\|\tilde{\Delta}_j^n Y\|^4\right]^{\frac{1}{2}} \\
\leq &C_4 \mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} \|\mathcal{S}(t_i - s)\sigma_s Q^{\frac{1}{2}}\|_{\mathrm{HS}}^2 ds\right)^2\right]^{\frac{1}{2}} \\
&\qquad \times \mathbb{E}\left[\left(\int_{t_{j-1}}^{t_j} \|\mathcal{S}(t_j - s)\sigma_s Q^{\frac{1}{2}}\|_{\mathrm{HS}}^2 ds\right)^2\right]^{\frac{1}{2}} \\
\leq &M(\Delta_n)^4 \mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} \|\sigma_s Q^{\frac{1}{2}}\|_{\mathrm{HS}}^2 ds\right)^2\right]^{\frac{1}{2}} \mathbb{E}\left[\left(\int_{t_{j-1}}^{t_j} \|\sigma_s Q^{\frac{1}{2}}\|_{\mathrm{HS}}^2 ds\right)^2\right]^{\frac{1}{2}} \qquad \text{(III.44)} \\
\leq &M(\Delta_n)^4 C(T),
\end{aligned}$$

where the second inequality followed from Lemma III.2.6. For (III.43), we apply the Cauchy-Schwarz inequality and argue as for the first two. In conclusion, we obtain a finite second moment as desired.

Now note that $t \mapsto \psi_t = \int_{t_{i-1}}^t \mathcal{S}(t_i - s)\sigma_s dW_s$ is a martingale for $t \in [t_{i-1}, t_i]$. From [40, Theorem 8.2, p. 109] we deduce that the process $(\zeta_t)_{t\in[t_{i-1},t_i]}$ with

$$\zeta_t = (\psi_t)^{\otimes 2} - \langle\langle \psi \rangle\rangle_t$$

is a centred martingale w.r.t. $(\mathcal{F}_t)_{t\in[t_{i-1},t_i]}$ and hence

$$\mathbb{E}\left[\tilde{Z}_n(i)\middle| \mathcal{F}_{t_{i-1}}\right] = \mathbb{E}\left[\zeta_{t_i}\middle| \mathcal{F}_{t_{i-1}}\right] = 0.$$

Also, this shows that $M_m^n := \sum_{i=1}^m \tilde{Z}_n(i)$ defines a discrete-time martingale in $L_{\mathrm{HS}}(H)$ and therefore $\|M_m^n\|_{\mathrm{HS}}$ a positive real-valued submartingale with respect to $(\mathcal{F}_{t_i})_{i=0,\dots}$. This is why by Doob's martingale inequality [43, Corollary (II.1.6)]

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\in[0,T]} \left\|\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \tilde{Z}_n(i)\right\|_{\mathrm{HS}}^2\right] =&\mathbb{E}\left[\max_{m=1,\dots,\lfloor T/\Delta_n \rfloor} \|M_m^n\|_{\mathrm{HS}}^2\right] \\
\leq &4\mathbb{E}\left[\|M_{\lfloor T/\Delta_n \rfloor}^n\|_{\mathrm{HS}}^2\right] \\
=&4\mathbb{E}\left[\left\|\sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \tilde{Z}_n(i)\right\|_{\mathrm{HS}}^2\right].
\end{aligned} \tag{III.45}$$

Moreover, for $j < i$, as each $\tilde{Z}_n(i)$ is $\mathcal{F}_{t_{i-1}}$ measurable and as the conditional expectation commutes with bounded linear operators, and also using the tower property of conditional expectation

$$\begin{aligned}\mathbb{E}\left[\langle \tilde{Z}_n(i), \tilde{Z}_n(j)\rangle_{\mathrm{HS}}\right] =& \mathbb{E}\left[\mathbb{E}\left[\langle \tilde{Z}_n(i), \tilde{Z}_n(j)\rangle_{\mathrm{HS}} | \mathcal{F}_{t_{i-1}}\right]\right] \\ =& \mathbb{E}\left[\langle \mathbb{E}\left[\tilde{Z}_n(i) | \mathcal{F}_{t_{i-1}}\right], \tilde{Z}_n(j)\rangle_{\mathrm{HS}}\right] = 0. \qquad \text{(III.46)}\end{aligned}$$

Combining (III.45) and (III.46) we obtain

$$\mathbb{E}\left[\sup_{t\in[0,T]} \left\| \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \tilde{Z}_n(i) \right\|_{\mathrm{HS}}^2\right] \leq 4 \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \mathbb{E}\left[\|\tilde{Z}_n(i)\|_{\mathrm{HS}}^2\right].$$

Applying the triangle and Bochner inequalities, the basic inequality $(a+b)^2 \leq 2(a^2+b^2)$ and appealing to (III.44), we find

$$\begin{aligned}&\mathbb{E}\left[\|\tilde{Z}_n(i)\|_{\mathrm{HS}}^2\right] \\ &\leq 2\mathbb{E}\left[\|(\tilde{\Delta}_i^n Y)^{\otimes 2}\|_{\mathrm{HS}}^2 + \left(\int_{t_{i-1}}^{t_i} \|\mathcal{S}(t_i - s)\sigma_s Q \sigma_s^* \mathcal{S}(t_i - s)^*\|_{\mathrm{HS}} ds\right)^2\right] \\ &\leq 2\mathbb{E}\left[\|\tilde{\Delta}_i^n Y\|^4 + M(\Delta_n)^4 \left(\int_{t_{i-1}}^{t_i} \|\sigma_s Q^{\frac{1}{2}}\|_{\mathrm{HS}}^2 ds\right)^2\right] \\ &\leq 2M(\Delta_n)^4 (C_4 + 1)\mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} \|\sigma_s Q^{\frac{1}{2}}\|_{\mathrm{HS}}^2 ds\right)^2\right].\end{aligned}$$

Summing up, we have

$$\begin{aligned}&\mathbb{E}\left[\sup_{t\in[0,T]} \left\| \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \tilde{Z}_n(i) \right\|_{\mathrm{HS}}^2\right] \\ \leq& 8(1+C_4)M(\Delta_n)^4 \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} \|\sigma_s Q^{\frac{1}{2}}\|_{\mathrm{HS}}^2 ds\right)^2\right] \\ \leq& 8(1+C_4)M(\Delta_n)^4 \mathbb{E}\left[\int_0^T \left\|\sigma_r Q^{\frac{1}{2}}\right\|_{\mathrm{HS}}^2 dr \sup_{i=1,\dots,\lfloor T/\Delta_n \rfloor} \int_{t_{i-1}}^{t_i} \left\|\sigma_s Q^{\frac{1}{2}}\right\|_{\mathrm{HS}}^2 ds\right] \\ \leq& 8(1+C_4)M(\Delta_n)^4 \mathbb{E}\left[\left(\int_0^T \left\|\sigma_r Q^{\frac{1}{2}}\right\|_{\mathrm{HS}}^2 dr\right)^2\right]^{\frac{1}{2}} \\ &\times \mathbb{E}\left[\left(\sup_{i=1,\dots,\lfloor T/\Delta_n \rfloor} \int_{t_{i-1}}^{t_i} \left\|\sigma_s Q^{\frac{1}{2}}\right\|_{\mathrm{HS}}^2 ds\right)^2\right]^{\frac{1}{2}}\end{aligned}$$

$$\leq 8(1+C_4)M(\Delta_n)^4 C(T)^{\frac{1}{2}} a_n(T)^2.$$

Hence, the proposition follows by application of the Cauchy-Schwarz inequality. ■

The Law of large numbers, Theorem III.3.2, follows now from the following result:

Proposition III.5.5. *Suppose that Assumption 3 holds. Then*

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\left\|\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\langle\langle\tilde{\Delta}_n^i Y\rangle\rangle - \int_0^t \sigma_s Q\sigma_s^* ds\right\|_{HS}\right]$$
$$\leq (1+M(\Delta_n))\, b_n(T)C(T)^{\frac{1}{4}} + a_n(T)^2. \tag{III.47}$$

Proof. Recall the expression for $\langle\langle\tilde{\Delta}_n^i Y\rangle\rangle$ in (III.39). By the triangle and Bochner inequalities, we find,

$$\sup_{t\in[0,T]}\left\|\int_0^{\lfloor t/\Delta_n\rfloor\Delta_n}\sigma_s Q\sigma_s^* ds - \sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\int_{t_{i-1}}^{t_i}\mathcal{S}(t_i-s)\sigma_s Q\sigma_s^*\mathcal{S}(t_i-s)^* ds\right\|_{\text{HS}}$$
$$\leq \sup_{t\in[0,T]}\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\int_{t_{i-1}}^{t_i}\|\sigma_s Q\sigma_s^* - \mathcal{S}(t_i-s)\sigma_s Q\sigma_s^*\mathcal{S}(t_i-s)^*\|_{\text{HS}} ds$$
$$\leq \sum_{i=1}^{\lfloor T/\Delta_n\rfloor}\int_{t_{i-1}}^{t_i}\|\sigma_s Q\sigma_s^* - \mathcal{S}(t_i-s)\sigma_s Q\sigma_s^*\mathcal{S}(t_i-s)^*\|_{\text{HS}} ds.$$

By Lemma III.2.1 and the Cauchy-Schwarz inequality we obtain

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\left\|\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\langle\langle\tilde{\Delta}_n^i Y\rangle\rangle - \int_0^t \sigma_s Q\sigma_s^* ds\right\|_{\text{HS}}\right]$$
$$\leq \sum_{i=1}^{\lfloor T/\Delta_n\rfloor}\int_{t_{i-1}}^{t_i}\mathbb{E}[\|(I-\mathcal{S}(t_i-s))\sigma_s Q\sigma_s^*\|_{\text{HS}}]$$
$$+\mathbb{E}[\|\mathcal{S}(t_i-s)\sigma_s Q\sigma_s^*(I-\mathcal{S}(t_i-s)^*)\|_{\text{HS}}]ds$$
$$+\sup_{t\in[0,T]}\int_{t_n}^t\mathbb{E}[\|\sigma_s Q\sigma_s^*\|_{\text{HS}}]ds$$
$$\leq \sum_{i=1}^{\lfloor T/\Delta_n\rfloor}\int_{t_{i-1}}^{t_i}\mathbb{E}[\|(I-\mathcal{S}(t_i-s))\sigma_s Q^{\frac{1}{2}}\|_{\text{op}}\|Q^{\frac{1}{2}}\sigma_s^*\|_{\text{HS}}]$$
$$+M(\Delta_n)\mathbb{E}[\|\sigma_s Q^{\frac{1}{2}}\|_{\text{HS}}\|Q^{\frac{1}{2}}\sigma_s^*(I-\mathcal{S}(t_i-s)^*)\|_{\text{op}}]ds$$
$$+\sup_{t\in[0,T]}\int_{t_n}^t\mathbb{E}[\|\sigma_s Q^{\frac{1}{2}}\|_{\text{HS}}^2]ds$$

$$\leq (1 + M(\Delta_n)) \left(\int_0^T \sup_{t \leq \Delta_n} \mathbb{E}\left[\left\| (I - \mathcal{S}(t))\sigma_s Q^{\frac{1}{2}} \right\|_{\text{op}}^2 \right] ds \right)^{\frac{1}{2}}$$

$$\times \left(\int_0^T \mathbb{E}\left[\left\| \sigma_s Q^{\frac{1}{2}} \right\|_{\text{HS}}^2 \right] ds \right)^{\frac{1}{2}}$$

$$+ \sup_{t \in [0,T]} \mathbb{E}\left[\left(\int_{t_n}^t \|\sigma_s Q^{\frac{1}{2}}\|_{\text{HS}}^2 ds \right)^2 \right]^{\frac{1}{2}}$$

$$\leq (1 + M(\Delta_n))\, b_n(T) C(T)^{\frac{1}{4}} + a_n(T)^2.$$

This completes the proof. ■

III.5.1.4 Proof of Theorem III.3.4

Proof of Theorem III.3.4. As σ and α are locally square-integrable, we can for all $m \in \mathbb{N}$ define the stopping time

$$\tau_m := \inf \left\{ t \in [0,T] : \int_0^T \left(\|\alpha_s\|^2 + \left\| \sigma_s Q^{\frac{1}{2}} \right\|_{\text{HS}}^2 \right) ds > m \right\}.$$

Define with the notations $\sigma_t^{(m)} := \sigma_t \mathbf{1}_{(0,\tau_m]}(t)$ and $\alpha_t^{(m)} := \alpha_t \mathbf{1}_{(0,\tau_m]}(t)$, such that

$$\begin{aligned} Y_t^{(m)} :=& \mathcal{S}(t) Y_0 + \int_0^{\min(t,\tau_m)} \mathcal{S}(t-s)\alpha_s ds + \int_0^{\min(t,\tau_m)} \mathcal{S}(t-s)\sigma_s dW_s \\ =& \mathcal{S}(t) Y_0 + \int_0^t \mathcal{S}(t-s)\alpha_s^{(m)} ds + \int_0^t \mathcal{S}(t-s)\sigma_s^{(m)} dW_s, \end{aligned}$$

where the last equality holds almost surely for all $t \in [0,T]$ (cf. Lemma 2.3.9 in [34]). We further define

$$\mathcal{Z}_m^n := \sup_{0 \leq s \leq t} \left\| \sum_{i=1}^{\lfloor s/\Delta_n \rfloor} (\tilde{\Delta}_n^i Y^{(m)})^{\otimes 2} - \int_0^s \sigma_u^{(m)} Q \sigma_u^{(m)*} du \right\|_{\text{HS}},$$

$$\mathcal{Z}^n := \sup_{0 \leq s \leq t} \left\| \sum_{i=1}^{\lfloor s/\Delta_n \rfloor} (\tilde{\Delta}_n^i Y)^{\otimes 2} - \int_0^s \sigma_u Q \sigma_u^* du \right\|_{\text{HS}}.$$

Since $\alpha^{(m)}$ and $\sigma^{(m)}$ satisfy Assumption 3 and thus, the conditions of Theorem III.3.2, we obtain that for all $m \in \mathbb{N}$ and $\epsilon > 0$

$$\lim_{n \to \infty} \mathbb{P}[\mathcal{Z}_m^n > \epsilon] = 0. \tag{III.48}$$

We have $\mathcal{Z}_m^n = \mathcal{Z}^n$ on $\Omega_m := \{\tau_m \geq t\}$ and hence

$$
\begin{aligned}
\mathbb{P}[\mathcal{Z}^n > \epsilon] &= \int_{\Omega_m} \mathbf{1}(\mathcal{Z}^n > \epsilon) d\mathbb{P} + \int_{\Omega_m^c} \mathbf{1}(\mathcal{Z}^n > \epsilon) d\mathbb{P} \\
&= \int_{\Omega_m} \mathbf{1}(\mathcal{Z}_m^n > \epsilon) d\mathbb{P} + \int_{\Omega_m^c} \mathbf{1}(\mathcal{Z}^n > \epsilon) d\mathbb{P} \\
&\leq \mathbb{P}[\mathcal{Z}_m^n > \epsilon] + \mathbb{P}[\Omega_m^c],
\end{aligned}
$$

which holds for all $n, m \in \mathbb{N}$. Now, by (III.48) we obtain for all $m \in \mathbb{N}$ that

$$
\limsup_{n\to\infty} \mathbb{P}[\mathcal{Z}^n > \epsilon] \leq \mathbb{P}[\Omega_m^c].
$$

As $\Omega_m \uparrow \Omega$ (due to the local integrability of drift and volatility) and by the continuity of $\mathbb{P}$ from below, $\mathbb{P}[\Omega_m^c]$ converges to 0 as $m \to \infty$ and therefore

$$
\lim_{n\to\infty} \mathbb{P}[\mathcal{Z}^n > \epsilon] = \limsup_{n\to\infty} \mathbb{P}[\mathcal{Z}^n > \epsilon] = 0.
$$

■

III.5.2 Proof of Theorem III.4.6

Proof of Theorem III.4.6. Since for all $h \in H_\beta$ one has $|h(0)| \leq \|h\|_\beta$, we have for $\|h\|_\beta = 1$ that

$$
\begin{aligned}
\|(I - \mathcal{S}(x))\sigma_r Q^{\frac{1}{2}} h\|_\beta \leq& \|(I - \mathcal{S}(x)) f_r\|_\beta + \left\| (I - \mathcal{S}(x)) \int_0^\infty q_r(\cdot, z) h'(z) dz \right\|_\beta \\
=& (1) + (2).
\end{aligned}
$$

The first summand can be estimated as follows: By the mean value theorem, there is a $\zeta \in (0,1)$, such that for $x < 1$ we have

$$
\begin{aligned}
(1) \leq& \left(|f_r(x)|^2 + 2\int_0^\infty ((1 - e^{\frac{\beta}{2}x}) f_r'(y + x))^2 e^{\beta y} dy \right. \\
& \left. + 2\int_0^\infty (e^{\frac{\beta}{2}(x+y)} f_r'(y + x) - e^{\frac{\beta}{2}y} f_r'(y))^2 dy \right)^{\frac{1}{2}} \\
\leq& (|f_r'(\zeta)|^2 x^2 + 2\|\mathcal{S}(x) f_r\|_\beta^2 (\frac{2e^{\frac{\beta}{2}}}{\beta})^2 x^2 + 2x^{2\gamma} \|L_r^1\|_{L^2(\mathbb{R}_+)}^2)^{\frac{1}{2}} \\
\leq& x^\gamma (|f_r'(\zeta)| + \sqrt{8}(\frac{e^{\frac{\beta+1}{2}}}{\beta}) \|f_r\|_\beta + \sqrt{2} \|L_r^1\|_{L^2(\mathbb{R}_+)}). \qquad \text{(III.49)}
\end{aligned}
$$

Here we used in the second inequality $f_r(0) = 0$ and that L_1 is the Hölder constant of $e^{\frac{\beta\cdot}{2}} f'(\cdot)$. In the third inequality we used the subadditivity of the squareroot and that the semigroup is quasi-contractive and satisfies $\|\mathcal{S}(x)\|_{\text{op}} < e^1 = e$ for

$x \leq 1$ (cf. [12, Lemma 3.5]). We can show, using the Hölder inequality, for all $h \in H_\beta$ such that $\|h\|_\beta = 1$, that for some $\zeta' \in (0,1)$

$$
\begin{aligned}
(2) \leq & \left| \int_0^\infty q_r(x,z) h'(z) dz \right| \\
& + \left(\int_0^\infty \left[\partial_y \int_0^\infty (q_r(y+x,z) - q_r(y,z)) h'(z) dz \right]^2 e^{\beta y} dy \right)^{\frac{1}{2}} \\
= & \left| \int_0^x \int_0^\infty p_r(y,z) e^{\frac{\beta}{2}(z-y)} h'(z) dz dy \right| \\
& + \left(\int_0^\infty \left[\int_0^\infty \left(e^{-\frac{\beta}{2}x} p_r(y+x,z) - p_r(y,z) \right) e^{\frac{\beta}{2}(z-y)} h'(z) dz \right]^2 e^{\beta y} dy \right)^{\frac{1}{2}} \\
= & \left(\int_0^x \int_0^\infty e^{\beta(z-y)} h'(z)^2 dz dy \right)^{\frac{1}{2}} \left(\int_0^x \int_0^\infty p_r(y,z)^2 dz dy \right)^{\frac{1}{2}} \\
& + \left(\int_0^\infty \left[\int_0^\infty (e^{-\frac{\beta}{2}x} p_r(y+x,z) - p_r(y,z)) e^{\frac{\beta}{2}z} h'(z) dz \right]^2 dy \right)^{\frac{1}{2}} \\
\leq & \left(x \int_0^\infty e^{\beta z} h'(z)^2 dz \right)^{\frac{1}{2}} \|p_r(\cdot,\cdot)\|_{L^2(\mathbb{R}_+^2)} \\
& + \left(\int_0^\infty \int_0^\infty (e^{-\frac{\beta}{2}x} p_r(y+x,z) - p_r(y,z))^2 dz \|h\|_\beta^2 dy \right)^{\frac{1}{2}} \\
\leq & x^{\frac{1}{2}} \|p_r\|_{L^2(\mathbb{R}_+^2)} + \left(\int_0^\infty \int_0^\infty (e^{-\frac{\beta}{2}x} p_r(y+x,z) - p_r(y,z))^2 dz dy \right)^{\frac{1}{2}}.
\end{aligned}
$$

Now we can estimate, for $x < 1$, using the triangle inequality

$$
\begin{aligned}
(2) \leq & x^{\frac{1}{2}} \|p_r\|_{L^2(\mathbb{R}_+^2)} + \left(\int_0^\infty \int_0^\infty (e^{-\frac{\beta}{2}x} (p_r(y+x,z) - p_r(y,z)))^2 dz dy \right)^{\frac{1}{2}} \\
& + \left(\int_0^\infty \int_0^\infty (e^{-\frac{\beta}{2}x} - 1)^2 p_r(y,z)^2 dx dz \right)^{\frac{1}{2}} \\
\leq & x^{\frac{1}{2}} \|p_r\|_{L^2(\mathbb{R}_+^2)} + x^\gamma \|L_r^2\|_{L^2(\mathbb{R}_+^2)} + |e^{-\frac{\beta}{2}x} - 1| \|p_r\|_{L^2(\mathbb{R}_+^2)} \\
\leq & \left(x^{\frac{1}{2}} \|p_r\|_{L^2(\mathbb{R}_+^2)} + x^\gamma \|L_r^2\|_{L^2(\mathbb{R}_+^2)} + \frac{\beta}{2} x \|p_r\|_{L^2(\mathbb{R}_+^2)} \right) \\
\leq & x^{\min(\gamma, \frac{1}{2})} (\|L_r^2\|_{L^2(\mathbb{R}_+^2)} + (1 + \frac{\beta}{2}) \|p_r\|_{L^2(\mathbb{R}_+^2)}). \qquad \text{(III.50)}
\end{aligned}
$$

Combining (III.49) and (III.50), we obtain, for $\|h\|_\beta = 1$,

$$
\begin{aligned}
& \|(I - \mathcal{S}(x)) \sigma_r Q^{\frac{1}{2}} h\|_\beta \\
& \quad \leq x^{\min(\gamma, \frac{1}{2})} \Bigg[|f_r'(\zeta)| + \sqrt{8} (\frac{e^{\frac{\beta+1}{2}}}{\beta}) \|f_r\|_\beta + \sqrt{2} \|L_r^1\|_{L^2(\mathbb{R}_+)}
\end{aligned}
$$

$$+\|L_r^2\|_{L^2(\mathbb{R}_+^2)} + (1+\frac{\beta}{2})\|p_r\|_{L^2(\mathbb{R}_+^2)}\Bigg].$$

Now we can conclude that

$$\begin{aligned} b_n(T)^2 &\leq \int_0^T \mathbb{E}[\sup_{x\in[0,\Delta_n]}\sup_{\|h\|_\beta=1}\|(I-\mathcal{S}(x))\sigma_r Q^{\frac{1}{2}}h\|_\beta^2]dr \\ &\leq x^{2\min(\gamma,\frac{1}{2})}\int_0^T \mathbb{E}\left[\left(|f_r'(\zeta)| + \sqrt{8}(\frac{e^{\frac{\beta+1}{2}}}{\beta})\|f_r\|_\beta + \sqrt{2}\|L_r^1\|_{L^2(\mathbb{R}_+)}\right.\right. \\ &\qquad \left.\left. +\|L_r^2\|_{L^2(\mathbb{R}_+^2)} + (1+\frac{\beta}{2})\|p_r\|_{L^2(\mathbb{R}_+^2)}\right)^2\right] dr. \end{aligned}$$

This concludes the proof. ■

III.6 Discussion and Outlook

Our paper develops a new asymptotic theory for high-frequency estimation of the volatility of infinite-dimensional stochastic evolution equations in an operator setting. We have defined the so-called semigroup-adjusted realised covariation (SARCV) and derived a weak law of large numbers based on uniform convergence in probability with respect to the Hilbert-Schmidt norm. Moreover, we have presented various examples where our new method is applicable.

Many articles on (high-frequency) estimation for stochastic partial differential equations rely on the so-called spectral approach and assume therefore the applicability of spectral theorems to the generator A (cf. the survey article [19]). This makes it difficult to apply these results on differential operators that do not fall into the symmetric and positive definite scheme, as for instance $A = \frac{d}{dx}$ in the space of forward curves presented in Section III.4.1.3, a case of relevance in financial applications that is included in our framework. Moreover, a lot of the related work assumes the volatility as a parameter of estimation to be real-valued (cf. the setting in [19]). An exception is the spatio-temporal volatility estimation in the recent paper by [17] (see also [18] for limit laws for the power variation of fractional stochastic parabolic equations). Here, the stochastic integrals are considered in the sense of [45] and the generator is the Laplacian. In our analysis, we operate in the general Hilbert space framework in the sense of [25] for stochastic integration and semigroups.

In our framework, we work with high-frequent observations of Hilbert-space valued random elements, hence we have observations, which are discrete in time but not necessarily in space. Recent research on inference for parabolic stochastic partial differential considered observation schemes which allow for discreteness in time and space, cf. [20], [15], [17], [18]. However, as our approach falls conveniently into the realm of functional data analysis, we might reconstruct data in several cases corresponding to well-known techniques for interpolation or smoothing. Indeed, in practice, a typical situation is that the Hilbert space consists of real-valued functions (curves) on $\mathbb{R}^d$ (or some subspace thereof), but we only have access to discrete observations of the curves. We may have data for $Y_{t_i}(x_j)$ at locations $x_j, j = 1, \ldots, m$, or possibly some aggregation of these (or, in more generality, a finite set of linear functionals of Y_{t_i}). For example, in commodity forward markets, we have only a finite number of forward contracts traded at all times, or, like in power forward markets, we have contracts with a delivery period (see e.g. [10]) and hence observations of the average of Y_{t_i} over intervals on $\mathbb{R}_+$. In other applications, like observations of temperature and wind fields in space and time, we may have accessible measurements at geographical locations where meteorological stations are situated, or, from atmospheric reanalysis where we have observations in grid cells regularly distributed in space. From such discrete observations, one must recover the Hilbert-space elements Y_{t_i}. This is a fundamental issue in functional data analysis, and several smoothing techniques have been suggested and studied. We refer to [42] for an extensive discussion of this. However, smoothing introduces another layer of approximation, as we do not recover Y_{t_i} but some approximate version $Y_{t_i}^m$, where the superscript m indicates that we have smoothed based on the m available observations. The construction of a curve from discrete observations is not a unique operation as this is an inverse problem. In future research, it will be interesting to extend our theory to the case when (spatial) smoothing has been applied to the discrete observations.

In addition, there could be cases, in which we do not know a closed form of the semigroup, but rather the generator A. One then has to recover the semigroup adjusted increment in some way. Appealing to finite difference schemes like the implicit Euler method could be one way, which nevertheless, approximates the semigroup just strongly. Mathematically, this opens up an interesting numerical problem, which is left for future research.

Interestingly, when we compare our work to recent developments on high-frequency estimation for volatility modulated Gaussian processes in finite dimensions, see e.g. [41] for a survey, it appears that a scaling factor is needed in the realised (co)variation so that an asymptotic theory for Volterra processes can be derived. This scaling factor is given by the variogram of the associated so-called Gaussian core process, and depends on the corresponding kernel function. However, in our case, due to the semigroup property, we are in a better situation than for general Volterra equations, since we have (or can reconstruct) the data to compute the semigroup-adjusted increments. We can then develop our analysis based on extending the techniques and ideas that are used in the semimartingale case. In this way, the estimator becomes independent of further assumptions on the remaining parameters of the equation. However, the price to pay for this universality is that the convergence speed cannot generally be determined. The semigroup adjustment of the increments effectively forces the estimator to converge at most at the same rate as the semigroup converges to the identity on the range of the volatility as t goes to 0. At first glance, it seems that the strong continuity of the semigroup suggests that we can obtain convergence just with respect to the strong topology. This would make it significantly harder to apply methods from functional data analysis, even for constant volatility processes. Fortunately, the compactness of the operators $\sigma_t Q^{\frac{1}{2}}$ for $t \in [0, T]$

comes to the rescue and enables us to prove that convergence holds with respect to the Hilbert-Schmidt norm. In this case, we obtain reasonable convergence rates for the estimator.

Acknowledgements. F. E. Benth and A. E. D. Veraart would like to thank the Isaac Newton Institute for Mathematical Sciences for support and hospitality during the programme *The Mathematics of Energy Systems* when parts of the work on this paper were undertaken. This work was supported by: EPSRC grant number EP/R014604/1. D. Schroers and F. E. Benth gratefully acknowledge financial support from the STORM project 274410, funded by the Research Council of Norway, and the thematic research group SPATUS, funded by UiO:Energy at the University of Oslo. We thank the editor, associated editor and two anonymous referees for their careful reading and constructive comments.

References

[1] Aït-Sahalia, Y. and Jacod, J. *High-Frequency Financial Econometrics*. Princeton, New Jersey: Princeton University Press, 2014.

[2] Aït-Sahalia, Y. and Xiu, D. “Principal component analysis of high-frequency data”. In: *J. Amer. Statist. Assoc.* Vol. 114, no. 525 (2019), pp. 287–303.

[3] Andersen, T. et al. “Modeling and Forecasting Realized Volatility”. In: *Econometrica* vol. 71, no. 2 (2003), pp. 579–625.

[4] Barndorff-Nielsen, O. E. and Shephard, N. “Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics”. In: *Econometrica* vol. 72, no. 3 (2004), pp. 885–925.

[5] Barndorff-Nielsen, O. E. and Shephard, N. “Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models”. In: *J. R. Statist. Soc. B* vol. 64, no. 2 (2002), pp. 253–280.

[6] Barndorff-Nielsen, O. E. and Shephard, N. “Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in economics”. In: *J. R. Statist. Soc. B* vol. 63, no. 2 (2001). (with discussion).

[7] Barndorff-Nielsen, O. E. and Shephard, N. “Realized power variation and stochastic volatility models”. In: *Bernoulli* vol. 9, no. 2 (Apr. 2003), pp. 243–265.

[8] Barndorff-Nielsen, O. E., Corcuera, J., and Podolskij, M. “Limit theorems for functionals of higher order differences of Brownian semistationary processes”. In: *Prokhorov and Contemporary Probability*. Ed. by Shiryaev, A. E., Varadhan, S. R. S., and Presman, E. Vol. 33. Springer Proceedings in Mathematics and Statistics. 2013, pp. 69–96.

[9] Barndorff-Nielsen, O. E., Corcuera, J., and Podolskij, M. “Multipower variation for Brownian semistationary processes”. In: *Bernoulli* vol. 17, no. 4 (2011), pp. 1159–1194.

[10] Benth, F., Šaltytė Benth, J., and Koekebakker, S. *Stochastic Modeling of Electricity and Related Markets.* World Scientific, 2008.

[11] Benth, F. E. and Harang, F. A. "Infinite dimensional pathwise Volterra processes driven by Gaussian Noise–Probabilistic Properties and Applications". In: *Electron. J. Probab.* Vol. 26 (2021).

[12] Benth, F. E. and Krühner, P. "Representation of infinite-dimensional forward price models in commodity markets". In: *Commun. Math. Statist.* Vol. 2, no. 1 (2014), pp. 47–106.

[13] Benth, F. E., Rüdiger, B., and Süss, A. "Ornstein–Uhlenbeck processes in Hilbert space with non-Gaussian stochastic volatility". In: *Stoch. Proc. Applic.* Vol. 128, no. 2 (2018), pp. 461–486.

[14] Benth, F. E. and Simonsen, I. C. "The Heston stochastic volatility model in Hilbert space". In: *Stoch. Analysis Applic.* Vol. 36, no. 4 (2018), pp. 733–750.

[15] Bibinger, M. and Trabs, M. "Volatility estimation for stochastic PDEs using high-frequency observations". In: *Stoch. Proc. Applic.* Vol. 130, no. 5 (2020), pp. 3005–3052.

[16] Bosq, D. *Linear Processes in Function Space: Theory and Applications.* Vol. 149. Springer, 2012.

[17] Chong, C. "High-frequency analysis of parabolic stochastic PDEs". In: *Ann. Statist.* Vol. 48, no. 2 (Apr. 2020), pp. 1143–1167.

[18] Chong, C. and Dalang, R. C. "Power variations in fractional Sobolev spaces for a class of parabolic stochastic PDEs". In: *E-print arXiv:2006.15817* (2020).

[19] Cialenco, I. "Statistical Inference for SPDEs: an overview". In: *Statist. Inf. Stoch. Proc.* Vol. 20, no. 2 (Dec. 2018), pp. 309–329.

[20] Cialenco, I. and Huang, Y. "A note on parameter estimation for discretely sampled SPDEs". In: *Stoch. Dynamics* vol. 20, no. 03 (2020).

[21] Corcuera, J., Nualart, D., and Podolskij, M. "Asymptotics of weighted random sums". In: *Commun. Appl. Indust. Math.* Vol. 6, no. 1 (2014), e–486, 11.

[22] Corcuera, J. M., Nualart, D., and Woerner, J. H. "Power variation of some integral fractional processes". In: *Bernoulli* vol. 12, no. 4 (Aug. 2006), pp. 713–735.

[23] Corcuera, J. M. et al. "Asymptotic theory for Brownian semi-stationary processes with application to turbulence". In: *Stoch. Proc. Applic.* Vol. 123, no. 7 (2013), pp. 2552–2574.

[24] Da Prato, G., Jentzen, A., and Röckner, M. "A mild Itô formula for SPDEs". In: *Trans. Amer. Math. Soc.* Vol. 372, no. 6 (2019), pp. 3755–3807.

[25] Da Prato, G. and Zabczyk, J. *Stochastic Equations in Infinite Dimensions*. Second. Vol. 152. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, 2014.

[26] Engel, K.-J. and Nagel, R. *One-Parameter Semigroups for Linear Evolution Equations*. Vol. 194. Springer Science & Business Media, 1999.

[27] Ferraty, F. and Vieu, P. *Nonparametric Functional Data Analysis: Theory and Practice*. New York: Springer, 2006.

[28] Filipović, D. "Consistency Problems for HJM Interest Rate Models". In: vol. 1760. Lecture Notes in Mathematics. Berlin: Springer, 2001.

[29] Gatheral, J., Jaisson, T., and Rosenbaum, M. "Volatility is rough". In: *Quantit. Finance* vol. 18, no. 6 (2018).

[30] Granelli, A. and Veraart, A. E. D. "A central limit theorem for the realised covariation of a bivariate Brownian semistationary process". In: *Bernoulli* vol. 25, no. 3 (Aug. 2019), pp. 2245–2278.

[31] Horváth, L. and Kokoszka, P. *Inference for Functional Data with Applications*. Springer Series in Statistics. Springer, 2012.

[32] Jacod, J. "Asymptotic properties of realized power variations and related functionals of semimartingales". In: *Stoch. Proc. Applic.* Vol. 118, no. 4 (2008), pp. 517–559.

[33] Jacod, J. and Protter, P. *Discretization of Processes*. Vol. 67. Stochastic Modelling and Applied Probability. Springer, Heidelberg, 2012.

[34] Liu, W. and Röckner, M. *Stochastic Partial Differential Equations: An Introduction*. Universitext. Springer International Publishing, 2015.

[35] Mandrekar, V. and Gawarecki, L. *Stochastic Differential Equations in Infinite Dimensions*. Probability and Its Applications. Berlin, Heidelberg: Springer, 2011.

[36] Marinelli, C. and Röckner, M. "On the maximal inequalities of Burkholder, Davis and Gundy". In: *Expo. Math.* Vol. 34 (2016), pp. 1–26.

[37] Nualart, D. and Peccati, G. "Central limit theorems for sequences of multiple stochastic integrals". In: *Ann. Probab.* Vol. 33, no. 1 (Jan. 2005), pp. 177–193.

[38] Panaretos, V. M. and Tavakoli, S. "Fourier analysis of stationary time series in function space". In: *Ann. Statist.* Vol. 41, no. 2 (Apr. 2013), pp. 568–603.

[39] Passeggeri, R. and Veraart, A. E. D. "Limit theorems for multivariate Brownian semistationary processes and feasible results". In: *Adv. Appl. Probab.* Vol. 51 (2019), pp. 667–716.

[40] Peszat, S. and Zabczyk, J. *Stochastic Partial Differential Equations with Lévy Noise*. Vol. 113. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, 2007.

[41] Podolskij, M. “Ambit fields: survey and new challenges”. In: *XI Symposium of Probability and Stochastic Processes: CIMAT, Mexico, November 18-22, 2013*. Ed. by Mena, R. H. et al. Vol. 69. Progress in Probability. Springer, 2015, pp. 241–279.

[42] Ramsay, J. and Silverman, B. W. *Functional Data Analysis*. Second. Springer Series in Statistics. Springer, 2005.

[43] Revuz, D. and Yor, M. *Continuous Martingales and Brownian motion*. Third. Vol. 293. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1999, pp. xiv+602.

[44] Rudin, W. *Principles of Mathematical Analysis*. 3rd ed. New York: McGraw-Hill, 1976.

[45] Walsh, J. “An introduction to stochastic partial differential equations”. In: *Ecole d’Eté de Probabilités de Saint-Flour XIV (1984)*. Ed. by Carmona, R., Kesten, H., and Walsh, J. Vol. 1180. Lecture Notes in Mathematics. Springer, 1986, pp. 265–436.

[46] Yao, F., Müller, H.-G., and Wang, J.-L. “Functional data analysis for sparse longitudinal data”. In: *J. Amer. Statist. Assoc.* Vol. 100, no. 470 (2005), pp. 577–590.