

# Signature Methods in Stochastic Portfolio Theory

Christa Cuchiero\*      Janka Möller†

October 8, 2024

## Abstract

In the context of stochastic portfolio theory we introduce a novel class of portfolios which we call linear path-functional portfolios. These are portfolios which are determined by certain transformations of linear functions of a collections of feature maps that are non-anticipative path functionals of an underlying semimartingale. As main example for such feature maps we consider the signature of the (ranked) market weights. We prove that these portfolios are universal in the sense that every continuous, possibly path-dependent, portfolio function of the market weights can be uniformly approximated by signature portfolios. We also show that signature portfolios can approximate the growth-optimal portfolio in several classes of non-Markovian market models arbitrarily well and illustrate numerically that the trained signature portfolios are remarkably close to the theoretical growth-optimal portfolios. Besides these universality features, the main numerical advantage lies in the fact that several optimization tasks like maximizing (expected) logarithmic wealth or mean-variance optimization within the class of linear path-functional portfolios reduce to a convex quadratic optimization problem, thus making it computationally highly tractable. We apply our method also to real market data based on several indices. Our results point towards out-performance on the considered out-of-sample data, also in the presence of transaction costs.

**Keywords:** signature methods, linear feature maps in machine learning, stochastic portfolio theory, growth-optimal portfolio, portfolio selection, convex quadratic optimization

**MSC (2020) Classification:** 91G10, 60L10, 90C20, 62P05

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Preliminaries</b>	<b>5</b>
2.1	The Signature of a Continuous Semimartingale . . . . .	5
2.2	Universal Approximation Theorem for Linear Functions on the Signature . . . . .	7
2.3	Global Universal Approximation Theorem for Linear Functions on the Signature . . . . .	10
2.4	The Johnson-Lindenstrauss Lemma . . . . .	12
2.5	The Market and Important Portfolio Specifications . . . . .	13
<b>3</b>	<b>(Linear) Path-Functional Portfolios</b>	<b>16</b>
<b>4</b>	<b>Approximation Properties of Signature Portfolios</b>	<b>19</b>
4.1	Universal Approximation of the Growth-Optimal Portfolio . . . . .	21
4.2	A Class of Non-Markovian Markets where the Growth-Optimal Portfolio is a Signature Portfolio . . . . .	22

---

\*Vienna University, Department of Statistics and Operations Research, Data Science Uni Vienna, Kolingasse 14-16 1, A-1090 Wien, Austria, christa.cuchiero@univie.ac.at

†Vienna University, Department of Statistics and Operations Research, Kolingasse 14-16, A-1090 Wien, Austria, janka.moeller@univie.ac.at

The authors gratefully acknowledge financial support through grant Y 1235 of the Austrian Science Fund.

<b>5 Optimization Tasks for Linear Path-Functional Portfolios</b>	<b>24</b>
5.1 Mean-Variance Optimization . . . . .	25
5.2 Optimizing the (Expected) Log-(Relative)-Wealth . . . . .	28
5.3 Constraints and Regularization . . . . .	30
<b>6 Transaction Costs for Portfolios with Short-Selling and Without Bank Account</b>	<b>33</b>
<b>7 Numerical Results</b>	<b>34</b>
7.1 Simulated Data . . . . .	35
7.2 Real Market Data . . . . .	37
7.2.1 Details on the Optimization and Investment Procedure . . . . .	37
7.2.2 Rank-based approach: NASDAQ . . . . .	43
7.2.3 Name-based approach: SMI and S&P500 . . . . .	44
<b>A Proofs of Section 2</b>	<b>57</b>
<b>B Proofs of Section 4</b>	<b>59</b>
<b>C Proofs of Section 5</b>	<b>60</b>
<b>D Proofs of Section 6</b>	<b>61</b>

## 1 Introduction

Portfolio theory and portfolio optimization are a core topic in mathematical finance and have been the subject of vivid research for decades. The most prominent example is of course the mean-variance optimization problem as formulated by Harry Markowitz, the founder of *Modern Portfolio Theory* (see [57]). His framework takes into account risk-preferences of investors while offering a tractable form of optimization, namely being convex and quadratic. Although the choice of constant portfolio weights is arguably oversimplified, the elegance of the Markowitz model can not be denied. Ever since, researchers have been striving for more and more realistic models with fewer limiting assumptions while hoping to preserve tractability.

In the spirit of coming up with more realistic assumptions, for instance without the need of claiming a specific form of the hardly measurable drift, Robert Fernholz developed *Stochastic Portfolio Theory* (SPT) (see [27] and also [29]). A key feature of SPT, again in the realm of more realistic modeling assumptions, is the rather relaxed no-arbitrage condition. Indeed, the price process is only assumed to be a (continuous) semimartingale satisfying *No Unbounded Profit with Bounded Risk* and not necessarily the stronger condition of *No free lunch with vanishing risk* (see [24]). It is also central that the portfolios' performance is measured with respect to a benchmark portfolio, which is usually chosen to be the market portfolio, i.e. large indices like the S&P500. At least since the introduction of very liquid ETFs replicating market indices the out-performance of the market is indeed a major challenge for investors. In this context Fernholz introduced *functionally generated portfolios* and derived the so-called *master formula* describing the relative wealth process of such portfolios and allowing for the detection for relative arbitrages with respect to the market portfolio. These portfolios are constructed via the log-gradient of a function of the market weights. While appreciating the beauty of the framework and also its robustness in view of certain optimal worst case long-run growth rates (see [49, 41]), one may point out that the log-gradient form as well as the neglect of past information is somewhat restrictive. To reduce these limitations various generalizations of the original framework have been developed, e.g., towards replacing the market weights by market-to-book ratios [50], considering trading strategies generated by Lyapunov functions [47], adding extra information in

the form of a finite variation process [64] and incorporating past information in a path-dependent setting [65, 46]. In this context it is worth mentioning that the latter two papers do not work in a stochastic setting but rather in a path-wise one, using functional Itô-calculus as in [14, 25] based on Föllmer integration. An alternative path-wise formulation allowing for more general strategies (beyond gradient type) and building on the theory of rough path has been studied in [3].

In this paper, we consider for simplicity the stochastic setting of continuous semimartingales (even though a path-wise formulation in the spirit of [3] would also be possible), take inspiration from the functionally generated portfolios and generalize them to what we call *path-functional portfolios*. These are portfolios constructed via an auxiliary portfolio  $\tau$  and non-anticipative path-functionals with a general semimartingale as input. In doing so, we get rid of any gradient-type form of the functionals and allow for a general benchmark portfolio  $\tau$  and past information of the market or other exogenous relevant factors in form of a general continuous semimartingale. The choice of continuous semimartingales is here only for the ease of exposition, but could be replaced by other (predictable stochastic) processes. This generalization can be achieved by using rough path theory, where similar universal approximation results hold for more general processes, see e.g. [20] for rough paths with jumps.

One of our main interest concerns *tractable portfolio optimization* in the framework of SPT using such path-functional portfolios. To do so, we focus on *linear path-functional portfolios*, which are constructed via a linear combination of possibly path-dependent feature maps and constant optimization parameters.

This allows us to incorporate modern machine learning techniques for time-series data, in particular *signature methods* (see e.g. [44, 6, 33] in the area of optimal control, [9, 62] on applications to data generation/simulation, [55, 17, 20] for topics in financial modelling, [13, 18, 19, 2] on randomization techniques and their application, [52, 51, 12, 4, 16, 1] for other applications in mathematical finance and machine learning), but also other tools, like random neural networks [38, 39, 34, 11]. In the current paper, our focus lies on signature methods, which play an important role in rough path theory (see e.g. [56, 31]) and whose appeal stems from the (*global*) *universal approximation theorem*. It states that linear functions on the signature can approximate continuous (with respect to certain variation distances) path-functionals arbitrarily well on compact sets of paths or even globally when using the setup of weighted spaces (see [22]). This result then motivates the notion of *signature portfolios*, which are linear path-functional portfolios with feature maps being elements of the signature. Indeed, with this choice we can approximate generic path-functional portfolios including functionally generated portfolios and the growth-optimal portfolio in a large class of non-Markovian markets, arbitrary well.

Despite this versatility, signature portfolios and linear path-functional portfolios in general lead to remarkably tractable optimization tasks. More concretely, we show that maximizing the (expected) logarithmic wealth as well as the mean-variance optimization of portfolio returns lead to *convex quadratic optimization problems*. We would like to highlight that the Markowitz portfolio itself is included in the class of signature portfolios and moreover, that any chosen benchmark portfolio  $\tau$  can always be *attained* by signature portfolios constructed via  $\tau$ .

In addition to our theoretical results, we optimize signature portfolios in numerical experiments using simulated and real market data. By means of simulated data generated from the Black-Scholes model, volatility stabilized models [28] and so-called signature market models, we illustrate that the trained signature portfolios (of small degree) are remarkably close to the theoretical growth-optimal portfolios.

In the real market situation we additionally incorporate transaction costs in our optimization by proposing a regularization under which the optimization problems remain convex and quadratic. To deal with high-dimensional market indices such as the NASDAQ and S&P500, we use dimension reduction techniques leading to what we call randomized signature portfolios

and JL-signature portfolios, where the former rely on randomized signatures (see [1, 13, 18, 19] and also [59] for a connection to neural signature kernels), while the latter are based on an explicit Johnson-Lindenstrauss projection. Note that we do not reduce the dimension of the underlying process, but only of its signature. To make the computation of the JL-signature portfolios feasible, we present a memory efficient algorithm to do so. We train the portfolios in both a log-relative wealth and mean-variance optimization in the NASDAQ, the SMI and S&P500 markets. We use randomized- and JL-signature portfolios in the high-dimensional markets (NASDAQ and S&P500) and signature portfolios for the SMI. In the NASDAQ market the portfolios are computed from the ranked market weights, while in the other two cases the name-based market weights are used. We incorporate proportional transaction costs of 1% and 5% respectively in the name-based markets. In almost all configurations, our trained portfolios outperform the market portfolios during the out-of-sample period, even under 5% of proportional transaction costs. It is worth noting that in the setting with transaction costs we still re-balance our portfolios daily whereas the market portfolio *does not* pay any transaction costs.

To provide more context to our findings, we highlight two papers which are particularly related to our research, namely [10] and [33]. In [10], the authors study portfolio optimization in SPT over a family of rank-based functionally generated portfolios parameterized by an exponentially concave function. Their objective is to maximize the relative logarithmic growth rate, which in their parameterization leads to a convex optimization problem. In their empirical study they use (ranked) market data of the 100 largest US stocks, which they manage to out-perform during the out-of-sample period with their trained portfolios. As they do not include transaction costs in the learning procedure (which however could be done by adopting e.g. similar transaction cost treatments as ours), the out-performance does no longer work with 0.45% of proportional transaction costs. This could also be related to the fact that the considered portfolios are long only functionally generated and thus a rather small class. Indeed, the class of portfolios we consider is more general in several respects: we allow for short-selling, the inclusion of information from the past and from exogenous signals as well as for a general benchmark portfolio  $\tau$  in the construction of the portfolios. At the same time the portfolios of [10] can be approximated arbitrarily well using signature portfolios, which is a result of the universal approximation theorem.

Very recently, the work of [33] appeared, where the authors study mean-variance optimization of the wealth process using signature methods. In contrast to our multiplicative setting, they consider an additive approach where the trading strategies correspond to numbers of shares, which are then directly parameterized as linear functions on the signature. This means in particular that the strategies are only self-financing if a bank-account is included, while in our case the strategies are always self-financing *without* a bank-account, since we use portfolio weights. Even though the formulation of [33], where the *wealth process* itself is the quantity of interest, is the standard one in the literature of the continuous-time mean-variance portfolio selection problem (see e.g. [67]), it differs from the original approach of Markowitz [57], where mean-variance optimization of the *returns* was actually the objective. The latter is advantageous when optimizing (in a model-free setup) along the observed trajectory since the returns are more likely to be stationary than the assets themselves. For these reasons we use returns and a multiplicative approach, which allows in particular to include the *classical Markowitz portfolio* (as a very special case) in our setup (which is not possible in [33]). Another difference is that we consider general *linear* path-functionals and not only signature portfolios, while still obtaining a convex quadratic optimization problem. Additionally we include transaction costs and dimension reduction techniques to be able to deal with 500-dimensional price processes, whereas in the numerical implementations of [33] only two-dimensional price processes are considered. Let us also point out that in [33] the signal is augmented with the two-dimensional lead-lag process to express Itô-integrals as linear functions on the corresponding signature. We usually only aug-

ment with time, which – under certain conditions on the quadratic variation – still allows to get linear expressions for the Itô-integrals (see Remark 5.12). As the computation of the signature becomes expensive for higher dimensions, this might be of relevance.

The remainder of the article is structured as follows. In Section 2 we recall the notion of the signature of continuous semimartingales, provide two versions of the universal approximation theorem, state the Johnson-Lindenstrauss Lemma and specify the financial market setting as well as classes of important portfolios. Section 3 is dedicated to the introduction of (linear) path-functional portfolios with signature portfolios as special case, while Section 4 addresses their approximation properties, in particular in view of approximating the growth optimal portfolio. In Section 5 the convex quadratic optimization tasks for the mean-variance and (expected) log-(relative) wealth are discussed and Section 6 explains the treatment of proportional transaction costs. Section 7 concludes the paper with a presentation of the numerical results. Some proofs are gathered in the Appendix.

## 2 Preliminaries

### 2.1 The Signature of a Continuous Semimartingale

In this section, we first introduce the signature of an  $\mathbb{R}^n$ -valued continuous semimartingale  $X$  and then state the so-called *universal approximation theorem* (see Theorem 2.12), which motivates the use of signature methods. In the following, we denote by  $X$  an  $\mathbb{R}^n$ -valued continuous semimartingale defined on some filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]})$  with  $T > 0$  finite. We start by introducing several preliminary notions needed to define the signature. Similar introductions can for instance be found in [54, 31]. Readers who are neither interested in a mathematical introduction of the signature nor the associated approximation theorems of linear functions of the signature in general or signature portfolios in particular, may choose to skip Sections 2.1, 2.2, 2.3 and 4.

**Definition 2.1** ((Truncated) Extended Tensor Algebra and Group). The signature of  $X$ , denoted by  $\mathbb{X}$ , is an element of the extended tensor algebra  $T((\mathbb{R}^n))$  defined as

$$T((\mathbb{R}^n)) := \left\{ (a_0, a_1, \dots) \mid a_k \in (\mathbb{R}^n)^{\otimes k}, k \geq 0 \right\}$$

with  $(\mathbb{R}^n)^{\otimes 0} := \mathbb{R}$ . For  $\mathbf{a}, \mathbf{b} \in T((\mathbb{R}^n))$  we define the following operations:

$$\begin{aligned} \mathbf{a} + \mathbf{b} &= (a_0 + b_0, a_1 + b_1, \dots), \\ \mathbf{a} \otimes \mathbf{b} &= (c_0, c_1, \dots) \text{ for } c_k = \sum_{j=0}^k a_j \otimes b_{k-j}, \\ \lambda \mathbf{a} &= (\lambda a_0, \lambda a_1, \dots) \text{ for } \lambda \in \mathbb{R}. \end{aligned}$$

Note that  $T((\mathbb{R}^n))$  forms a real non-commutative unital algebra under these operations with unit element  $\mathbf{1} = (1, 0, 0, \dots)$ . Moreover,  $T_1((\mathbb{R}^n)) \subset T((\mathbb{R}^n))$ , defined as

$$T_1((\mathbb{R}^n)) := \{ \mathbf{a} \mid \mathbf{a} \in T((\mathbb{R}^n)) \text{ with } a_0 = 1 \}$$

is a group under the operation  $\otimes$  with unit element  $\mathbf{1} = (1, 0, 0, \dots)$ . For  $\mathbf{a} \in T_1((\mathbb{R}^n))$  its inverse is defined as

$$\mathbf{a}^{-1} = \sum_{k \geq 0} (\mathbf{1} - \mathbf{a})^k.$$

We define the *truncated* tensor algebra at level  $N$  by

$$T^N(\mathbb{R}^n) := \left\{ (a_0, a_1, \dots, a_N) \mid a_k \in (\mathbb{R}^n)^{\otimes k}, 0 \leq k \leq N \right\}$$

and denote by  $T_1^N(\mathbb{R}^n)$  the corresponding *truncated* tensor group. We denote by  $(e_1, \dots, e_n)$  the canonical basis of  $\mathbb{R}^n$ . Let  $I = (i_1, \dots, i_m)$  be a multi-index and denote its length by  $|I| = m$ , then  $e_I := e_{i_1} \otimes \dots \otimes e_{i_m}$  a basis element of  $(\mathbb{R}^n)^{\otimes m}$ . We use the convention that  $|I| = 0$  means that  $I = \emptyset$ , the empty word. In this formulation, we can write any  $\mathbf{a} \in T^N(\mathbb{R}^n)$  as

$$\mathbf{a} = \sum_{0 \leq |I| \leq N} \langle e_I, \mathbf{a} \rangle e_I,$$

where  $\langle \cdot, \cdot \rangle$  is defined as the inner product of  $(\mathbb{R}^n)^{\otimes k}$  for each  $k \geq 0$ . Finally we introduce the two canonical projections

$$\begin{aligned} \Pi_k : T((\mathbb{R}^n)) &\rightarrow (\mathbb{R}^n)^{\otimes k} \\ \mathbf{a} &\mapsto a_k, \end{aligned}$$

$$\begin{aligned} \Pi_{\leq N} : T((\mathbb{R}^n)) &\rightarrow T^N(\mathbb{R}^n) \\ \mathbf{a} &\mapsto (a_0, \dots, a_N). \end{aligned}$$

Of course, these projections can also be applied on the truncated tensor algebra, i.e.  $\Pi_k : T^N(\mathbb{R}^n) \rightarrow (\mathbb{R}^n)^{\otimes k}$  for  $k \leq N$  and  $\Pi_{\leq N} : T^M(\mathbb{R}^n) \rightarrow T^N(\mathbb{R}^n)$  for  $M \geq N$ .

We define accordingly  $T_0^N(\mathbb{R}^N) := \{\mathbf{a} \in T^N(\mathbb{R}^d) | a_0 = 0\}$ . In fact, it holds that  $T_1^N(\mathbb{R}^d)$  is a Lie group under the tensor product  $\otimes$  and  $T_0^N(\mathbb{R}^d)$  is a Lie algebra with Lie bracket  $T_0^N(\mathbb{R}^d) \times T_0^N(\mathbb{R}^d) \ni (\mathbf{a}, \mathbf{b}) \mapsto [\mathbf{a}, \mathbf{b}] := \Pi_{\leq N}(\mathbf{a} \otimes \mathbf{b} - \mathbf{b} \otimes \mathbf{a}) \in T_0^N(\mathbb{R}^d)$ . The two are related by the exponential map

$$\begin{aligned} \exp : T_0^N(\mathbb{R}^d) &\rightarrow T_1^N(\mathbb{R}^d) \\ \mathbf{a} &\mapsto \Pi_{\leq N} \left( 1 + \sum_{k=1}^N \frac{\mathbf{a}^{\otimes k}}{k!} \right). \end{aligned}$$

We call  $\mathfrak{g}^N(\mathbb{R}^d) = \mathbb{R}^d \oplus [\mathbb{R}^d, \mathbb{R}^d] \oplus \dots \oplus [\mathbb{R}^d, [\dots, [\mathbb{R}^d, \mathbb{R}^d]]]$  the *free step-N nilpotent Lie algebra* and  $G^N(\mathbb{R}^d) = \exp(\mathfrak{g}^N(\mathbb{R}^d))$  the *free nilpotent group of step N*.

We are now ready to define the signature of a continuous semimartingale. For an introduction of the signature of a continuous semimartingale in a similar spirit, see [17]. To this end, consider a finite time-horizon  $T > 0$  and some filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$ .

**Definition 2.2** (Signature of a Continuous Semimartingale). For an  $\mathbb{R}^n$ -valued continuous semimartingale  $X$ , we denote its signature by  $\mathbb{X}$  and define  $\mathbb{X}$  as the process in  $T_1((\mathbb{R}^n))$  whose value at time  $t$  is given via

$$\Pi_k(\mathbb{X}_t) = \int_{0 \leq u_1 \leq \dots \leq u_k \leq t} \otimes \circ dX_{u_1} \cdots \otimes \circ dX_{u_k},$$

where  $\circ$  denotes Stratonovic integration. Accordingly, we define the truncated signature at level  $N$ , which we denote by  $\mathbb{X}^N$  as

$$\mathbb{X}_t^N = \Pi_{\leq N}(\mathbb{X}_t).$$

Note that  $\mathbb{X}_t^N$  takes values in  $G^N(\mathbb{R}^n)$  for all  $t \in [0, T]$ . We define the increments of the signature  $\mathbb{X}_{s,t}$  by  $\mathbb{X}_{s,t} := \mathbb{X}_s^{-1} \otimes \mathbb{X}_t$  and note that

$$\Pi_k(\mathbb{X}_{s,t}) = \int_{s \leq u_1 \leq \dots \leq u_k \leq t} \otimes \circ dX_{u_1} \cdots \otimes \circ dX_{u_k},$$

in particular,  $\Pi_1(\mathbb{X}_{s,t}) = X_t - X_s$ .

Let us now state a few important and useful properties about the signature.

**Lemma 2.3** (Shuffle Product Property). *For two multiindices  $I, J$  we define the shuffle product  $\sqcup$  as*

$$e_I \sqcup e_\emptyset = e_\emptyset \sqcup e_I = e_I$$

$$(e_I \otimes e_i) \sqcup (e_J \otimes e_j) = (e_I \sqcup (e_J \otimes e_j)) \otimes e_i + ((e_I \otimes e_i) \sqcup e_J) \otimes e_j.$$

Then, for any two multiindices  $I, J$  and any continuous semimartingale  $X$ , it holds that

$$\langle e_I, \mathbb{X} \rangle \langle e_J, \mathbb{X} \rangle = \langle e_I \sqcup e_J, \mathbb{X} \rangle.$$

*Proof.* This follows directly from the properties of the Stratonovic integral.  $\square$

**Lemma 2.4** (Uniqueness of the Signature). *Let  $X$  be a continuous semimartingale with at least one strictly monotone component. Then the signature uniquely determines  $X$  up to vertical translations of the trajectories.*

*Proof.* This follows from the fact, that the signature determines the underlying semimartingale uniquely up to tree-like equivalences, see [8]. A more direct proof, in the case where that component is the time, can be found e.g. in [20].  $\square$

We now define a class of functions which plays a major role in this paper, namely *linear functions on the signature*:

**Definition 2.5** (Linear Functions on the Signature). Let  $X$  be an  $\mathbb{R}^n$ -valued continuous semimartingale and  $\mathbb{X}$  its signature. We call a function  $L : T((\mathbb{R}^n)) \rightarrow \mathbb{R}$  a *linear function on the signature* if  $L$  is given by

$$L(\mathbb{X}_t) = \sum_{0 \leq |I| \leq N} l_I \langle e_I, \mathbb{X}_t \rangle$$

for some coefficients  $l_I \in \mathbb{R}$  and any  $N \in \mathbb{N}$ . It is important to note that all  $l_I$  are *constant in time*. We denote the set of all linear functions on the signature by  $\mathfrak{L}$ .

## 2.2 Universal Approximation Theorem for Linear Functions on the Signature

Linear functions on the signature are dense in a set of certain continuous path-functionals defined on compact sets of paths. We call this the *universal approximation property* and state it rigorously in Theorem 2.12. To this end we need the following preparations which have been formulated in a similar manner in [6, 44, 53].

Let us first define for some  $p \in (2, 3)$  the  $p$ -variation metric of paths  $\mathbf{x}, \mathbf{y}$  with values in the group  $T_1^N(\mathbb{R}^n)$  by

$$d_{p-var;[0,t]}(\mathbf{x}, \mathbf{y}) = \max_{k \in 1, \dots, N} \sup_{D \in \mathcal{D}} \left( \sum_{t_i, t_{i-1} \in D} |\Pi_k(\mathbf{x}_{t_{i-1}, t_i} - \mathbf{y}_{t_{i-1}, t_i})|^{\frac{p}{k}} \right)^{\frac{k}{p}}, \quad (2.1)$$

where  $\mathcal{D}$  is the set of all partitions of  $[0, t]$  and  $|\cdot|$  is the Euclidean distance. The definition of increments of paths with values in the group  $T_1^N(\mathbb{R}^n)$  is as for the signature in Definition 2.2, namely  $\mathbf{x}_{s,t} := \mathbf{x}_s^{-1} \otimes \mathbf{x}_t$ .

**Definition 2.6** (Lifted Paths). Let  $\varphi : [0, T] \rightarrow \mathbb{R}$  be a strictly monotone increasing function. For  $N \geq 2$  and  $t \in [0, T]$  we define by

$$\mathcal{G}_t^N(\varphi, x) = \bigcup_{X \text{ cont. semimart.}} \{ \hat{\mathbb{X}}_{[0,t]}^N(\omega) \mid \hat{X}_s = (\varphi(s), X_s), \omega \in \Omega_X, X_0 = x \}$$

the space of *lifted paths*  $\hat{\mathbb{X}}_{[0,t]}^N(\omega)$  of all  $\mathbb{R}^n$ -valued continuous semimartingales with initial value  $x$  and for each  $X$ ,  $\Omega_X \subset \Omega$  is taken to be a probability-one set on which the metric given in (2.1) is well-defined for all  $t \in [0, T]$ . Note that we denote by  $\hat{\mathbb{X}}$  the signature of  $\hat{X}$  and we use the subscript  $[0, t]$  to emphasize that we take the *entire path* between  $[0, t]$ . We equip  $\mathcal{G}_t^N(\varphi, x)$  with the metric  $d_{p-var;[0,t]}$  (now for paths with values in  $T_1^N(\mathbb{R}^{n+1})$ ).

**Remark 2.7.** We would like to point out that using a strictly monotone increasing function  $\varphi$  other than the identity can be useful in practice. For example, we use the time-augmentation  $\varphi_T(t) := \frac{t}{T}$  for our experiments using real market data in Section 7.2 to account for the different length of the training and testing period.

**Definition 2.8** (Space of Lifted (Stopped) Paths). We define the space of *lifted stopped paths* via

$$\Lambda_T^N(\varphi, x) = \bigcup_{t \in [0, T]} \mathcal{G}_t^N(\varphi, x)$$

We then equip  $\Lambda_T^N(\varphi, x)$  with the metric  $d_\Lambda$  with its form being based on [14, 25]

$$\begin{aligned} d_\Lambda & \left( \hat{\mathbb{X}}_{[0,t]}^N(\omega_1), \hat{\mathbb{Y}}_{[0,s]}^N(\omega_2) \right) \\ &= |t - s| + d_{p-var;[0,t \vee s]} \left( \hat{\mathbb{X}}_{[0,t \vee s]}^N(\omega_1), \hat{\mathbb{Y}}_{[0,t \vee s]}^N(\omega_2) \right), \end{aligned}$$

where we denote by  $\widehat{X}^t$  the process  $(\widehat{X}^t)_u = (\varphi(u), X_u^t)$ , i.e. the process where we stop the semimartingale  $X$  at time  $t$  but not the time-augmentation  $\varphi$  and write  $\widehat{\mathbb{X}}^t$  for the signature of  $\widehat{X}^t$ . See [6] for a similar definition in the setting of general rough paths.

**Remark 2.9.** Note that if  $\hat{\mathbb{X}}_{[0,t]}^N(\omega) \in \Lambda_T^N(\varphi, x)$ , then  $\widehat{\mathbb{X}}_{[0,T]}^N(\omega) \in \mathcal{G}_T^N(\varphi, x)$  and hence  $\widehat{\mathbb{X}}_{[0,T]}^N(\omega) \in \Lambda_T^N(\varphi, x)$ . This is the reason, we do not stop the time-augmentation  $\varphi$ . Indeed,  $(\hat{\mathbb{X}}^N)_{[0,T]}^t(\omega) \notin \mathcal{G}_T^N(\varphi, x)$ , since  $\varphi$  is required to be strictly monotone.

Based on [14] we formulate the notion of non-anticipative path-functionals in our setting:

**Definition 2.10** (Non-Anticipative Path-Functionals). We call functionals  $f : \Lambda_T^2(\varphi, x) \rightarrow \mathbb{R}$  *non-anticipative path-functionals*.

The following continuity result for the signature is crucial for the universal approximation property.

**Corollary 2.11** (Lyon's Lift). *For all  $t \in [0, T]$  and  $N \geq 2$  there exists a unique bijection*

$$S^N : \mathcal{G}_t^2(\varphi, x) \mapsto \mathcal{G}_t^N(\varphi, x)$$

*with inverse  $\Pi_{\leq 2}$  and  $S^N$  is continuous on bounded sets. We call  $S^N$  Lyons' lift. Moreover, for each  $X$  and all  $\omega \in \Omega_X$  it holds that*

$$S^N \left( \hat{\mathbb{X}}_{[0,t]}^2(\omega) \right) = \hat{\mathbb{X}}_{[0,t]}^N(\omega).$$

*Proof.* The first part of the theorem follows from [31, Theorem 8.10, Theorem 9.5 and Corollary 9.11] and the second statement follows from [31, Proposition 17.1 and Exercise 17.2].  $\square$

Relying on two lemmas proved in Appendix A, we can now show the universal approximation theorem for linear functions on the signature. We here state a version that holds uniformly in time. For this we also provide a rigorous proof in Appendix A which is – up to our knowledge – not available in the literature. For similar assertions we however refer to [6, 44]. In the setting of càdlàg rough paths an analogous result has been proved in [20], however with different topologies.

**Theorem 2.12.** Let  $K \subset \mathcal{G}_T^2(\varphi, x)$  be a compact subset. For any continuous non-anticipative path-functional  $f : \Lambda_T^2(\varphi, x) \rightarrow \mathbb{R}$  and for every  $\epsilon > 0$  there exists a linear function on the signature  $L$  such that

$$\sup_{(t, \hat{\mathbb{X}}_{[0,T]}^2(\omega)) \in [0,T] \times K} |f(\hat{\mathbb{X}}_{[0,t]}^2(\omega)) - L(\hat{\mathbb{X}}_t)(\omega)| < \epsilon.$$

Let us point out that, conceptually, the reason for Theorem 2.12 to hold uniformly in time is due to the time-augmentation. To show the applicability of this result let us give some examples of non-anticipative path-functionals.

**Example 2.13** (Examples of Continuous Non-Anticipative Path-Functionals).

1. **Solutions of SDEs driven by  $X$ :** For  $\gamma > 2$ , an  $\mathbb{R}^n$ -valued continuous semimartingale  $X$  and an  $\mathbb{R}^m$ -valued continuous process of bounded variation  $H$ , consider the SDE

$$dY_t = \sum_{i=1}^n V_i(Y_t) \circ dX_t^i + \sum_{j=1}^m W_j(Y_t) dH_t^j, \quad Y_0 = y_0 \in \mathbb{R}^d,$$

where  $\{V_i\}_{1 \leq i \leq n}$  is a collection of vector fields in  $\text{Lip}^\gamma(\mathbb{R}^d)$  and  $\{W_i\}_{1 \leq i \leq m}$  is a collection of  $\text{Lip}^1(\mathbb{R}^d)$  vector fields.<sup>1</sup> Set  $Z = (X, H)$  and  $z = (X_0, H_0)$  and define  $D_Z$  to be the set

$$D_Z = \bigcup_{t \in [0, T]} \left\{ \hat{\mathbb{Z}}_{[0,t]}^2(\omega) \mid \omega \in \Omega_Z \right\} \subset \Lambda_T^2(\varphi, z).$$

Then the components of the solution to this SDE  $(Y_t^i)_{t \in [0, T]}$ , can be seen as a map

$$\begin{aligned} Y_t^i(\omega) : D_Z &\rightarrow \mathbb{R} \\ \hat{\mathbb{Z}}_{[0,t]}^2(\omega) &\mapsto Y_t^i(\omega) \end{aligned}$$

and it holds almost surely (i.e. on  $\Omega_Z$ ) that  $Y_t^i \in C(D_Z; \mathbb{R})$  for every  $t \in [0, T]$ . The continuity follows from the fact that the solution to the corresponding rough differential equation solves the SDE almost surely by [31, Theorem 17.3] and the RDE solution is continuous [31, Corollary 10.28 & Theorem 12.10] in the lifted input path. Hence, we can apply Theorem 2.12 and obtain that for any compact set  $K \subset \{\hat{\mathbb{Z}}_{[0,T]}^2(\omega) \mid \omega \in \Omega_Z\}$  and all  $\epsilon > 0$  it holds that for all  $i \in \{1, \dots, d\}$  there exists a linear function  $L^i$  on the signature such that

$$\sup_{(t \in [0, T], \hat{\mathbb{Z}}_{[0,T]}^2(\omega)) \in K} |Y_t^i(\omega) - L^i(\hat{\mathbb{Z}}_t(\omega))| < \epsilon.$$

2. **Continuous Functions Without Path-Dependence:** Consider an arbitrary but fixed  $\mathbb{R}^n$ -valued continuous semimartingale  $X$  with initial value  $X_0 = x$ . For any  $t \in [0, T]$  let

$$\begin{aligned} h : \mathbb{R}^{d+1} &\rightarrow \mathbb{R} \\ \hat{X}_t(\omega) &\mapsto h(\hat{X}_t(\omega)) \end{aligned}$$

be a continuous function. Such a function can for almost all  $\omega \in \Omega$  be extended to a non-anticipative path-functional  $\tilde{h}$  on  $\Lambda_T^2(\varphi, x)$ . To be precise, simply define  $\tilde{h}$  as

$$\begin{aligned} \tilde{h} : \Lambda_T^2(\varphi, x) &\rightarrow \mathbb{R} \\ \hat{\mathbb{X}}_{[0,t]}^2(\omega) &\mapsto \tilde{h}(\hat{\mathbb{X}}_{[0,t]}^2(\omega)) := h((\hat{X}_{[0,t]})_\odot(\omega)) = h(\hat{X}_t(\omega)). \end{aligned}$$

---

<sup>1</sup>Following the definition in [31, Definition 10.2] a vector field  $V$  is in  $\text{Lip}^\gamma$  if  $V$  is  $\lfloor \gamma \rfloor$  times continuously differentiable and if the supremum norm of  $V$  and the supremum norm of its first  $\lfloor \gamma \rfloor$  derivatives admits a common bound.

In particular,  $\tilde{h} \in C(\Lambda_T^2(\varphi, x); \mathbb{R})$ . To see this, note that for any  $\hat{\mathbf{x}}_{[0,t]}, \hat{\mathbf{y}}_{[0,s]} \in \Lambda_T^2(\varphi, x)$  it holds that

$$d_\Lambda(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \geq |\hat{x}_{0,t} - \hat{y}_{0,s}| = |\hat{x}_t - \hat{y}_s|$$

with  $\hat{x} = \Pi_1(\hat{\mathbf{x}})$ ,  $\hat{y} = \Pi_1(\hat{\mathbf{y}})$ , where we replaced the  $\max_{k \in \{1, \dots, N\}}$  by  $k = 1$  and  $\sup_{D \in \mathcal{D}}$  by  $D = [0, t \vee s]$  in the definition of  $d_{p-var; [0, t \vee s]}$  to obtain the inequality. The last equality follows from the fact that all paths in  $\Lambda_T^2(\varphi, x)$  have the same initial value. Due to the continuity of  $h$ , for every  $\epsilon > 0$  there exists a  $\delta > 0$  such that for all  $t_1, t_2 \in [0, T]$  and all paths  $X(\omega_1), Y(\omega_2)$  satisfying  $|\hat{X}_{t_1}(\omega_1) - \hat{Y}_{t_2}(\omega_2)| < \delta$  we have

$$\left| h(\hat{X}_{t_1}(\omega_1)) - h(\hat{X}_{t_2}(\omega_2)) \right| < \epsilon.$$

By the above estimate,  $d_\Lambda(\hat{\mathbb{X}}_{[0,t_1]}^2(\omega_1), \hat{\mathbb{X}}_{[0,t_2]}^2(\omega_2)) < \delta$  therefore implies

$$\left| \tilde{h}(\hat{\mathbb{X}}_{[0,t_1]}^2(\omega_1)) - \tilde{h}(\hat{\mathbb{X}}_{[0,t_2]}^2(\omega_2)) \right| < \epsilon$$

and thus proves the claim.

### 2.3 Global Universal Approximation Theorem for Linear Functions on the Signature

In this subsection, we state an alternative universal approximation theorem which is due to [22]. We will refer to Theorem 2.20 also as *global* universal approximation theorem because we are no longer limited to compact sets, however this comes at the price of approximating with respect to weighted norms instead of the uniform norm. The following is based on the statements and the results of [22] adapted to the setting considered in this paper. However, we do formulate everything in the path-functional setting here, leading to a approximation *uniformly in time* which has not been done in [22].

Let us start by introducing the following metrics:

**Definition 2.14** (The Carnot-Carathéodory Norm and Homogeneous  $\alpha$ -Hölder Metric).

1. Let  $\mathbf{a} \in G^N(\mathbb{R}^n)$ , we define the *Carnot-Carathéodory* norm  $\|\cdot\|_{cc}$  by

$$\begin{aligned} \|\mathbf{a}\|_{cc} := \inf \Big\{ \int_0^T \|dX_t\| \Big| X : [0, T] \rightarrow \mathbb{R}^d \text{ continuous and of finite variation,} \\ \text{such that } \mathbb{X}_T^N = \mathbf{a} \Big\}. \end{aligned}$$

Moreover, for  $\mathbf{a}, \mathbf{b} \in G^N(\mathbb{R}^n)$  the corresponding metric  $d_{cc}(\cdot, \cdot)$  is induced by

$$d_{cc}(\mathbf{a}, \mathbf{b}) := \|\mathbf{a}^{-1} \otimes \mathbf{b}\|_{cc}.$$

2. Recall the space of lifted path  $\mathcal{G}_t^N(\varphi, x)$  for  $N \geq 2$  and  $t \in [0, T]$ . For two lifted paths  $\hat{\mathbb{X}}_{[0,t]}^N(\omega_1), \hat{\mathbb{Y}}_{[0,t]}^N(\omega_2) \in \mathcal{G}_t^N(\varphi, x)$  we define the *homogeneous  $\alpha$ -Hölder metric*  $d_{\alpha, [0,t]}(\cdot, \cdot)$  for  $0 \leq \alpha < \frac{1}{2}$  by

$$d_{\alpha, [0,t]} \left( \hat{\mathbb{X}}_{[0,t]}^N(\omega_1), \hat{\mathbb{Y}}_{[0,t]}^N(\omega_2) \right) = \sup_{\substack{r, s \in [0, t] \\ r < s}} \frac{d_{cc} \left( \hat{\mathbb{X}}_{r,s}^N(\omega_1), \hat{\mathbb{Y}}_{r,s}^N(\omega_2) \right)}{|r - s|^\alpha}.$$

*Notation 2.15.* In the following we shall denote by  $\mathcal{G}_t^{N,\alpha}(\varphi, x) := (\mathcal{G}_t^N(\varphi, x), d_{\alpha,[0,t]})$  the space of lifted paths equipped with the homogeneous  $\alpha$ -Hölder metric for some  $\frac{1}{3} < \alpha < \frac{1}{2}$ . Moreover, we write  $\Lambda_T^{N,\alpha}(\varphi, x) := (\Lambda_T^N(\varphi, x), d_{\Lambda^\alpha})$  for the corresponding space of lifted stopped paths equipped with

$$d_{\Lambda^\alpha} \left( \hat{\mathbb{X}}_{[0,t]}^N(\omega_1), \hat{\mathbb{Y}}_{[0,s]}^N(\omega_2) \right) := |t - s| + d_{\alpha,[0,t \vee s]} \left( \hat{\mathbb{X}}_{[0,t \vee s]}^N(\omega_1), \hat{\mathbb{Y}}_{[0,t \vee s]}^N(\omega_2) \right).$$

We shall assume without loss of generality that for each continuous semimartingale  $X$ , the associated probability-one set  $\Omega_X$  in the definition of  $\mathcal{G}_t^N(\varphi, x)$  is such that the homogeneous  $\alpha$ -Hölder norm is finite for all lifted paths and all  $0 \leq \alpha < \frac{1}{2}$ .

*Remark 2.16.* Due to the fact that we are now considering different topologies on the space of lifted (stopped) paths, the continuity statements of Lemma A.2, Corollary 2.11 and Lemma A.3 need to be proved again. In fact, they do hold analogously:

- The proof that the map  $\lambda$  defined in Lemma A.2 is again continuous in the  $\alpha$ -Hölder topology works very similar to the one of Lemma A.2, where one ultimately chooses  $\delta > 0$  such that

$$\delta + d_{\alpha,[0,t+\delta]} \left( \hat{\mathbb{Y}}_{[0,t+\delta]}^N(\omega_1), \hat{\mathbb{Y}}_{[0,t+\delta]}^N(\omega_2) \right) < \epsilon.$$

- The continuity of Lyons' lift holds by [31, Theorem 9.5, Corollary 9.11].
- The proof of Lemma A.3 holds analogously in the  $\alpha$ -Hölder topology.

In order to formulate the global universal approximation theorem we need the notion of a weighted space and certain weighted function spaces being generalizations of continuous functions.

**Definition 2.17** (Weighted Space and Weighted Function Space). Let  $(X, \tau_X)$  be a completely regular Hausdorff space and consider a function  $\psi : X \rightarrow (0, \infty)$ . We call the pair  $(X, \psi)$  a *weighted space* if it holds that  $K_R := \psi^{-1}((0, R]) = \{x \in X | \psi(x) \leq R\}$  is  $\tau_X$ -compact for all  $R > 0$ . Moreover, for a weighted space  $(X, \psi)$ , we define

$$B_\psi(X) = \left\{ f : X \rightarrow \mathbb{R} \mid \sup_{x \in X} \frac{|f(x)|}{\psi(x)} < \infty \right\}$$

and equip it with  $\|f\|_{B_\psi(X)} := \sup_{x \in X} \frac{|f(x)|}{\psi(x)}$ . We denote by  $\mathcal{B}_\psi(X)$  the closure of the space of bounded continuous functions  $C_b(X)$  under  $\|f\|_{B_\psi(X)}$ . We refer to  $\mathcal{B}_\psi(X)$  as weighted function space.

The global universal approximation theorem via linear functions on the signature hinges on an application of the weighted Stone-Weierstrass theorem [22, Theorem 3.6]. In contrast to [22, Theorem 5.4] we want to extend the result to be uniform in time and hence to the space of lifted stopped paths. While  $\Lambda_T^2(\varphi, x)$  is *not* a weighted space, we can show that a larger space which contains  $\Lambda_T^2(\varphi, x)$  is a weighted space. In fact, we need to consider the "stopped paths" of all *weakly geometric  $\alpha$ -Hölder rough path* and not just those stemming from a continuous semimartingale. To this end we introduce the space of *weakly geometric  $\alpha$ -Hölder rough paths* for  $\frac{1}{3} < \alpha < \frac{1}{2}$  starting in  $x \in \mathbb{R}^n$  and time-augmented by a continuous and strictly increasing function  $\varphi$  and denote it by

$$\begin{aligned} \hat{C}_t^\alpha(\varphi, x) := \left\{ \hat{\mathbf{X}}_{[0,t]} \in C^\alpha([0, T]; G^2(\mathbb{R}^{n+1})) \mid \langle e_1, \hat{\mathbf{X}}_s \rangle = \varphi(s) \text{ for all } s \in [0, t] \right. \\ \left. \text{and } \langle e_i, \hat{\mathbf{X}}_0 \rangle = x^i \text{ for all } i \in \{2, \dots, n+1\} \right\}. \end{aligned}$$

The paths which are a realization of a continuous semimartingale, i.e.  $\mathcal{G}_t^{2,\alpha}(\varphi, x)$ , is a subspace of  $\hat{C}_t^\alpha(\varphi, x)$ . In the following, we will slightly abuse notation and extend  $d_{\alpha,[0,t]}$  to elements of

$\hat{C}_t^\alpha(\varphi, x)$ , as well as  $d_{\Lambda^\alpha}$  to elements of  $\bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x)$ . To do so, we need a notion of how to stop a weakly geometric rough path such that this is compatible with how we defined the signature of the stopped path in the semimartingale setting.

**Definition 2.18.** For any  $\frac{1}{3} < \alpha < \frac{1}{2}$ ,  $t \in [0, t]$  and  $\hat{\mathbf{X}}_{[0, t]} \in \hat{C}_t^\alpha(\varphi, x)$  we define the associated *stopped rough path*  $\hat{\mathbf{X}}^t_{[0, T]}$  in the following way; for all  $s \in [0, t]$ ,  $(\hat{\mathbf{X}}^t)_s := \hat{\mathbf{X}}_s$  and for all  $r \in [t, T]$

$$\langle e_I, (\hat{\mathbf{X}}^t)_r \rangle := \begin{cases} \varphi(r) & \text{for } I = (1) \\ \frac{1}{2}(\varphi(r))^2 & \text{for } I = (11) \\ \langle e_I, \hat{\mathbf{X}}_t \rangle & \text{for } I = (i) \text{ or } I = (ji) \text{ where } i \in \{2, \dots, n+1\}, j \in \{1, \dots, n+1\} \\ \varphi(r) \cdot \langle e_i, \hat{\mathbf{X}}_t \rangle - \langle e_{(1i)}, \hat{\mathbf{X}}_t \rangle & \text{for } I = (i1) \text{ where } i \in \{2, \dots, n+1\}. \end{cases}$$

Let us point out that the above definition is constructed such that any stopped weakly geometric rough path remains weakly geometric and such that it coincides with how we constructed the lifts of stopped semimartingales.

We are now ready to formulate the following lemma which we prove in Appendix A.

**Lemma 2.19.** Let  $\frac{1}{3} < \alpha < \frac{1}{2}$  and define the weight function

$$\psi(\hat{\mathbf{X}}_{[0, t]}) := \exp\left(\zeta \|\hat{\mathbf{X}}^t_{[0, T]}\|_{\alpha, [0, T]}^\xi\right) \quad (2.2)$$

for  $\hat{\mathbf{X}}_{[0, t]} \in \bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x)$ ,  $\zeta > 0$ ,  $\xi > 2$ ,  $\alpha < \frac{1}{2}$  and  $\|\cdot\|_{\alpha, [0, T]}$  being the norm induced by  $d_{\alpha, [0, T]}$ . Then  $\bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x)$  equipped with  $d_{\Lambda^{\alpha'}}$  for  $0 \leq \alpha' < \alpha$  is a weighted space.

We have thus shown that  $\left(\bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x), d_{\Lambda^{\alpha'}}\right)$  is a weighted space for the weight function given by (2.2) and  $0 \leq \alpha' < \alpha$ . We would like to emphasise that to achieve this result it is crucial to equip  $\bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x)$  with a weaker topology than the  $\alpha$ -Hölder topology. We are finally ready to formulate the global universal approximation theorem for non-anticipative path-functionals. Its proof is also given in Appendix A.

**Theorem 2.20** (Global Universal Approximation Theorem). *Let  $\frac{1}{3} < \alpha < \frac{1}{2}$ ,  $0 \leq \alpha' < \alpha$  and consider the weight function  $\psi$  given in (2.2). Then, for every non-anticipative path-functional  $f \in \mathcal{B}_\psi\left(\left(\bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x), d_{\Lambda^{\alpha'}}\right)\right)$  (i.e.  $\bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x)$  equipped with  $d_{\Lambda^{\alpha'}}$ ) and for every  $\epsilon > 0$  there exists a linear function on the signature  $L$  such that*

$$\sup_{\hat{\mathbb{X}}_{[0, t]}^2(\omega) \in \Lambda_T^{2, \alpha'}(\varphi, x)} \frac{|f(\hat{\mathbb{X}}_{[0, t]}^2(\omega)) - L(\hat{\mathbb{X}}_t)(\omega)|}{\psi(\hat{\mathbb{X}}_{[0, t]}^2(\omega))} < \epsilon.$$

*Remark 2.21.* Note that the approximation stated in Theorem 2.20 is indeed uniform in time, as the supremum is taken over all paths in  $\Lambda_T^{2, \alpha'}(\varphi, x)$ , i.e. over all lifted stopped paths. One could equivalently take the supremum over  $(t, \hat{\mathbb{X}}_{[0, T]}^2(\omega)) \in [0, T] \times \mathcal{G}_T^{2, \alpha'}(\varphi, x)$ . In fact, as shown in the proof of Theorem 2.20, we can even take the supremum over  $\bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x)$ .

## 2.4 The Johnson-Lindenstrauss Lemma

The Johnson-Lindenstrauss Lemma [43] is an important mathematical result, which is interesting for machine learning and data science, as it can be seen as a dimension reduction technique. There are many variants of the Johnson-Lindenstrauss lemma, we will here use a version stated in [58, Theorem 3.1]:

**Theorem 2.22.** Let  $n \in \mathbb{N}$ ,  $\epsilon \in (0, \frac{1}{2})$ ,  $\delta \in (0, 1)$  and set  $k = C\epsilon^{-2} \log(\frac{\delta}{2})$  for a suitable constant  $C$ . Then consider a random matrix  $A$  of dimensions  $k \times n$ , whose entries  $A_{ij}$  are i.i.d. random variables with  $\mathbb{E}[A_{ij}] = 0$ ,  $\text{Var}[A_{ij}] = \frac{1}{k}$  and sub-Gaussian tails. Then it holds for all  $x \in \mathbb{R}^n$  that

$$\mathbb{P}[(1 - \epsilon)\|x\| \leq \|Ax\| \leq (1 + \epsilon)\|x\|] \geq 1 - \delta.$$

*Remark 2.23.* Note that since the above holds for all  $x \in \mathbb{R}^n$ , it immediately follows that it holds in particular for all  $x - y$ , where  $x, y \in \mathbb{R}^n$ . Hence, Theorem 2.22 can be understood to preserve distances up to  $(1 \pm \epsilon)$ .

*Remark 2.24.* In this paper, we will use Theorem 2.22 to obtain a random projection of the truncated signature. We will choose the components  $A_{ij}$  to be i.i.d. with  $A_{ij} \sim \mathcal{N}(0, \frac{1}{k})$ .

## 2.5 The Market and Important Portfolio Specifications

Consider a finite time-horizon  $T > 0$  and some filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$ . We assume that the market consists of  $d$  companies, with market capitalizations given by a vector  $(S_t)_{t \in [0, T]} = (S_t^1, \dots, S_t^d)_{t \in [0, T]}$  of  $d$  positive continuous semimartingales with fixed initial value  $S_0 = s_0$ . We assume this market fulfills the *No Unbounded Profit with Bounded Risk (NUPBR)* [40] condition. Furthermore, we assume that each company has issued a single infinitely divisible share, hence  $S_t^1, \dots, S_t^d$  correspond to the prices of these shares. We denote the relative capitalization of the companies, also called *market weights* by  $\mu_t = (\mu_t^1, \dots, \mu_t^d)$  with

$$\mu_t^i = \frac{S_t^i}{S_t^1 + \dots + S_t^d}.$$

**Definition 2.25.** We call a vector  $\pi_t = (\pi_t^1, \dots, \pi_t^d)$  of predictable processes fulfilling  $\sum_{i=1}^d \pi_t^i \equiv 1$  and being integrable with respect to  $R$ , with  $R_t^i := \int_0^t \frac{dS_u^i}{S_u^i}$ , an *(R-integrable) portfolio*.

The component  $\pi_t^i$  denotes the proportion of wealth invested in stock  $i$  at time  $t$ , hence every portfolio is self-financing. Consider investing with a portfolio  $\pi$  over a time-horizon  $[t_0, t_1]$  for  $0 \leq t_0 < t_1 \leq T$ . We denote its wealth process portfolio  $\pi$  with initial wealth  $w$  by  $(W_t^{\pi, w})_{t \in [t_0, t_1]}$ . Without loss of generality we will assume from now on that the initial investment is always one unit of currency, i.e. that  $w = 1$  and therefore denote  $W_t^\pi := W_t^{\pi, 1}$ . The wealth process of a portfolio fulfills the stochastic differential equation

$$\frac{dW_t^\pi}{W_t^\pi} = \sum_{i=1}^d \pi_t^i \frac{dS_t^i}{S_t^i} \tag{2.3}$$

which is understood in Itô-sense.

A special portfolio is the *market portfolio* whose portfolio weights are given by the relative capitalizations  $\mu_t$ , whence we also refer to  $\mu_t$  as the market weights. The wealth of the market portfolio is given by

$$W_t^\mu = \frac{S_t^1 + \dots + S_t^d}{S_{t_0}^1 + \dots + S_{t_0}^d}$$

since we require  $W_{t_0}^\mu = 1$ .

We define the relative wealth process of a portfolio  $\pi$  to be

$$V_t^\pi := \frac{W_t^\pi}{W_t^\mu}.$$

By using Itô's-formula, it follows directly from (2.3) that

$$\frac{dV_t^\pi}{V_t^\pi} = \sum_{i=1}^d \pi_t^i \frac{d\mu_t^i}{\mu_t^i}. \tag{2.4}$$

Furthermore, by applying Itô's formula again, we obtain

$$d \ln(V_t^\pi) = \sum_{i=1}^d \frac{\pi_t^i}{\mu_t^i} d\mu_t^i - \frac{1}{2} \sum_{i,j=1}^d \frac{\pi_t^i}{\mu_t^i} \frac{\pi_t^j}{\mu_t^j} d[\mu^i, \mu^j]_t, \quad (2.5)$$

where  $[X, Y]_t$  denotes the co-variation process of continuous semimartingales  $X, Y$ .

*Notation 2.26.* We denote by  $\Delta^n$  the unit simplex of  $\mathbb{R}^n$  and by  $\Delta_+^n$  its interior. We call a portfolio  $\pi$  long-only, if  $\pi_t \in \Delta^d$  for all  $t \in [0, T]$  almost surely.

We now turn to some portfolios which are of special interest, those are the *numeraire*, *log-optimal* and *growth-optimal* portfolio.

**Definition 2.27** (Numeraire Portfolio). We call a portfolio  $\rho$  the *numeraire portfolio* if  $\frac{W_T^\pi}{W_T^\rho}$  is a supermartingale for any other portfolio  $\pi$ .

**Theorem 2.28.** *In a semimartingale market the NUPBR condition (see e.g. [40]) holds if and only if the numeraire portfolio  $\rho$  exists and  $W_T^\rho < \infty$  almost surely.*

*Proof.* See [45, Theorem 4.12], where in our case the predictable closed convex constraints  $\mathfrak{C}$  are

$$\mathfrak{C} = \left\{ \pi \in \mathbb{R}^d \mid \sum_{i=1}^d \pi^i = 1 \right\}.$$

Note that [45, Theorem 4.12] considers the setting with bank-account, however our constraint set  $\mathfrak{C}$  prohibits us from investing into the bank account, i.e.  $\pi^0 \equiv 0$  (in the notation of [45]).  $\square$

To connect the numeraire portfolio with the log-optimal one introduced in Definition 2.30 below, we recall Proposition 4.19 from [45].

**Proposition 2.29.** *Let  $U : (0, \infty) \rightarrow \mathbb{R}$  be concave and strictly increasing. The utility maximization problem*

$$\sup_{\pi} \mathbb{E}[U(W_T^\pi)]$$

*has no solution or has infinitely many solutions, if the NUPBR condition does not hold.*

**Definition 2.30.** Consider the log-utility optimization problem

$$\sup_{\pi} \mathbb{E}[\log(W_T^\pi)].$$

If it is well-posed, we call the corresponding optimal portfolio *log-optimal* portfolio.

The following lemma can also be found in [45].

**Lemma 2.31.** *Under the NUPBR condition, the log-utility optimization problem admits a solution, if it is finite, in which case the numeraire portfolio is the log-optimal portfolio.*

Finally, let us introduce the notion of the growth-optimal portfolio.

**Definition 2.32.** Consider a market model

$$dS_t = \text{diag}(S_t)(a_t dt + \Sigma_t dB_t),$$

where  $a_t$  is a  $d$ -dimensional vector,  $\Sigma_t$  a  $d \times m$ -matrix for  $m \geq d$  and  $B$  a  $m$ -dimensional Brownian motion. Assume  $a_t, B_t$  are predictable processes and satisfy

$$\sum_{i=1}^d \int_0^T |a_t^i| dt + \sum_{i=1}^d \sum_{j=1}^m \int_0^T |\Sigma_t^{ij}|^2 dt < \infty \quad \mathbb{P}\text{-a.s.}$$

and  $\Sigma_t \Sigma_t^\top$  is almost surely invertible for all  $t \in [0, T]$ . The log-wealth-process of a portfolio  $\pi$  in this market model is then given by

$$d \ln(W_t^\pi) = \left( \pi_t^\top a_t - \frac{1}{2} \pi_t^\top \Sigma_t \Sigma_t^\top \pi_t \right) dt + \pi_t^\top \Sigma_t dW_t.$$

We call  $g_t^\pi = \pi_t^\top a_t - \frac{1}{2} \pi_t^\top \Sigma_t \Sigma_t^\top \pi_t$  the portfolios *growth-rate* and the portfolio with maximal growth-rate, if it exists, the *growth-optimal* portfolio.

**Lemma 2.33.** *Consider a market model given in Definition 2.32 and let us define the (instantaneous) market price of risk*

$$\theta_t := \Sigma_t^\top (\Sigma_t \Sigma_t^\top)^{-1} a_t.$$

*The growth-optimal portfolio exists if and only if  $\int_0^T \|\theta_t\|^2 dt < \infty$  almost surely. Moreover, the log-optimal portfolio exists if and only if  $\mathbb{E} \left[ \int_0^T \|\theta_t\|^2 dt \right] < \infty$ . Moreover, if the growth-optimal portfolio  $\pi^{(g)}$  exists, it is given by*

$$\pi_t^{(g)} = (\Sigma_t \Sigma_t^\top)^{-1} (a_t - \kappa \mathbf{1}) \text{ for } \kappa = \frac{\sum_i ((\Sigma_t \Sigma_t^\top)^{-1} a_t)_i - 1}{\sum_{i,j} (\Sigma_t \Sigma_t^\top)^{-1}_{i,j}}$$

where  $\mathbf{1} = (1, \dots, 1)^\top$ .

*Proof.* Regarding the existence of the growth-optimal and log-optimal portfolios, see [40], also for the general case of continuous semimartingale markets. The specific form of the growth-optimal portfolio is due to [30].  $\square$

**Lemma 2.34.** *Consider the setting of a market as in Definition 2.32. Then the growth-optimal portfolio exists if and only if the numeraire portfolio exists in which case they are the same<sup>2</sup>. Moreover, if the log-utility maximization problem is finite, then they also coincide with the log-optimal portfolio.*

*Proof.* See [30] and [40]. The last assertion then follows from Lemma 2.31  $\square$

In SPT, there is another class of portfolios of special interest, those are *functionally generated portfolios*, see [27].

**Definition 2.35** (Functionally Generated Portfolios). Let  $U$  be a neighbourhood of  $\Delta^d$  and consider a  $C^2$ -function  $G : U \rightarrow \mathbb{R}_+$  such that  $x_i D_i \log G(x)$  is bounded on  $\Delta^d$ . Then  $G$  defines the functionally generated portfolio via

$$\pi_t^i = \mu_t^i \left( D_i \log G(\mu_t) + 1 - \sum_{j=1}^d \mu_t^j D_j \log G(\mu_t) \right).$$

The function  $G$  is called the *portfolio generating function* and if it is concave,  $\pi$  is a long-only portfolio.

Before we conclude this section, we introduce the *ranked market-weights* which are of particular interest in SPT, due to the remarkable stability of the capital distribution curves, see for example [27, 47].

---

<sup>2</sup>Assuming that their generated wealth is almost surely finite this holds if and only if NUPBR is satisfied (see Theorem 2.25).

**Definition 2.36** (Ranked Capitalization and Market Weight Processes). For an  $\mathbb{R}_+^n$ -valued continuous semimartingale  $X$  we denote its *ranked process* by  $\mathbf{X} = (X^{(1)}, \dots, X^{(n)})$  which is defined as

$$\max_{i \in \{1, \dots, n\}} X_t^i = X_t^{(1)} \geq X_t^{(2)} \geq \dots \geq X_t^{(n-1)} \geq X_t^{(n)} = \min_{i \in \{1, \dots, n\}} X_t^i,$$

where we break ties by allocating a lower rank to smaller labels. To be precise, if  $X_t^i = X_t^j$  with  $i \geq j$ , we set  $X_t^{(r_i)} = X_t^i$  and  $X_t^{(r_j)} = X_t^j$  where  $r_i \geq r_j$ .

Accordingly, we denote by  $\boldsymbol{\mu}$  the *ranked market weights* and by  $\mathbf{S}$  the *ranked capitalization process*.

**Corollary 2.37.** *Given an  $\mathbb{R}^n$ -valued continuous semimartingale  $X$ , its ranked process is again an  $\mathbb{R}^n$ -valued continuous semimartingale.*

*Proof.* See [5, Theorem 2.2]. □

### 3 (Linear) Path-Functional Portfolios

Before we present our application of signature methods in the context of SPT and portfolio optimization, we generalize the concept of functionally generated portfolios by introducing so-called *path-functional portfolios*.

**Definition 3.1** (Path-Functional Portfolios). We introduce *path-functional portfolios* which are generated by a family of non-anticipating path functionals  $\{f^i\}_{1 \leq i \leq d}$  depending on time and the path of  $X$  and which are constructed via an auxiliary portfolio  $\tau$ , where  $\tau$  is required to be uniformly bounded. Moreover, we assume that the non-anticipating path functionals  $\{f^i\}_{1 \leq i \leq d}$  fulfill the necessary integrability conditions such that the following processes  $\pi$  are portfolios in the sense of Definition 2.25. We consider two types of such portfolios:

$$(I) \quad \pi_t^i(\tau, X) = \tau_t^i \left( f^i(t, X_{[0,t]}) + 1 - \sum_{j=1}^d \tau_t^j f^j(t, X_{[0,t]}) \right)$$

$$(II) \quad \pi_t^i(\tau, X) = f^i(t, X_{[0,t]}) + \tau_t^i \left( 1 - \sum_{j=1}^d f^j(t, X_{[0,t]}) \right)$$

where we use the notation of the subscript  $[0, t]$  to make the dependence of on the entire path explicit. Moreover, we refer to the functionals  $\{f^i\}_{1 \leq i \leq d}$  as the *portfolio controlling functions*.

*Remark 3.2* (Relation to (Classical) Stochastic Portfolio Theory). In the above definition, we consider a general continuous semimartingale and a general auxiliary portfolio  $\tau$  in the construction of the path-functional portfolios. In the spirit of SPT, one would choose  $\tau = \mu$  the market portfolio and  $X = \mu$ , the process of market weights. One can then recover the classical functionally generated portfolios from the path-functional portfolios of type I by setting  $X = \mu$  and choosing  $f^i(t, \mu_{[0,t]}) = f^i(\mu_t) = D_i \log(G(\mu_t))$ . We would like to highlight that beyond this relation to classical SPT, constructing the path-functional portfolios in this form (i.e. via additive normalization) is crucial for our results in Sections 4 and 5, in particular in view of getting quadratic optimization problems in the current multiplicative setting where the strategies are in terms fractions of wealth.

*Remark 3.3* (Role of the process  $X$ ). In relation to classical SPT, we will often choose  $X = \mu$ , i.e. construct our portfolios via information on the relative market weights. However, we would like to point out that other information may be of interest such as the absolute capitalization, market-to-book ratios [50] or earnings data, to just name a few. Adding information beyond the market weights to the portfolio construction, has also appeared e.g. in [63] in the context of SPT. However, we would like to point out that in [63] the additional information has to be a continuous process of finite variation, while we allow it to be a general continuous semimartingale.

*Remark 3.4* (Path-Functional Portfolios of Type *II*). A simple example of a path-functional portfolio of type *II* would be any portfolio generated by a deep (recurrent) neural network, see for example [66]. Denote by  $\mathcal{NN}^i(t, X_{[0,t]})$  the  $i$ -th output of a deep (recurrent) neural network. If

- the last layer of the neural network is for example a softmax function and the portfolio weights are defined by  $\pi_t^{(\mathcal{NN}),i} = \mathcal{NN}^i(t, X_{[0,t]})$ , or,
- the portfolio weights are constructed as  $\pi_t^{(\mathcal{NN}),i} = \frac{\mathcal{NN}^i(t, X_{[0,t]})}{\sum_{j=1}^d \mathcal{NN}^j(t, X_{[0,t]})}$ ,

then, the portfolios  $\pi^{(\mathcal{NN})}$  are in both cases path-functional portfolios of type *II* but not of type *I*. This is by identifying  $f^{(\mathcal{NN}),i}(t, X_{[0,t]}) = \mathcal{NN}^i(t, X_{[0,t]})$  or  $f^{(\mathcal{NN}),i}(t, X_{[0,t]}) = \frac{\mathcal{NN}^i(t, X_{[0,t]})}{\sum_{j=1}^d \mathcal{NN}^j(t, X_{[0,t]})}$  respectively. Note that the choice of  $\tau$  is irrelevant because  $\sum_{j=1}^d f^{(\mathcal{NN}),j}(t, X_{[0,t]}) = 1$  by definition.

*Remark 3.5* (Conversion between Path-Functional Portfolios of Type I&II). Note that for a fixed auxiliary portfolio  $\tau$  with weight-processes which are continuous semimartingales and for a continuous semimartingale  $(X_t)_{t \in [0,T]}$  any path-functional portfolio of type *I* with portfolio controlling functions  $f^{(I),i} : X_{[0,t]} \mapsto f^{(I),i}(X_{[0,t]})$  is a path-functional portfolio of type *II* with portfolio controlling function  $f^{(II),i} : (\tau, X)_{[0,t]} \mapsto \tau_t^i f^{(I),i}(X_{[0,t]})$ . This also holds vice versa, if the components of  $\tau$  are all non-zero. Indeed, any path-functional portfolio of type *II* with portfolio controlling function  $f^{(II),i} : X_{[0,t]} \mapsto f^{(II),i}(X_{[0,t]})$  is a path-functional portfolio of type *I* with portfolio controlling function  $f^{(I),i} : (\tau, X)_{[0,t]} \mapsto \frac{f^{(II),i}(X_{[0,t]})}{\tau_t^i}$ . However, note that this conversion always involves adding the process of the auxiliary portfolio  $\tau$  to the input. Not only does this enlarge the dimensions of the inputs, but also is  $\tau$  then required to be a continuous semimartingale. Hence, this conversion is not always possible.

We now introduce a special class of path-functional portfolios which are particularly useful for optimizing path functional portfolios, as we will make more explicit in Section 5. This class is the one of *linear path-functional portfolios*.

**Definition 3.6** (Linear Path-Functional Portfolios). For a path-functional portfolio of type *I* or *II*, we call it a *linear* path-functional portfolio of the corresponding type, if the portfolio controlling functions are of the form

$$f^i(t, X_{[0,t]}) = \sum_{\nu \in \mathcal{V}} l_\nu^i \phi^\nu(t, X_{[0,t]})$$

where  $\{\phi^\nu\}_{\nu \in \mathcal{V}}$  is a collection of feature maps which are themselves non-anticipating path functionals,  $\mathcal{V}$  is a finite set of features and  $l_\nu^i \in \mathbb{R}$  for each  $i, \nu$  are constant (optimization) parameters.

*Remark 3.7* (Significance of the Auxiliary Portfolio  $\tau$ ). Considering linear path-functional portfolios gives us a first idea on how to do portfolio optimization in that context. Namely, optimizing the parameters  $\{l_\nu^i\}_{1 \leq i \leq d, \nu \in \mathcal{V}}$  for a given collection of feature maps. Note, that the auxiliary portfolio  $\tau$  can always be attained in this way by

- setting  $l_\nu^i = l_\nu^j$  for all  $i, j \in \{1, \dots, d\}$ ,  $\nu \in \mathcal{V}$  for portfolios of type *I*
- setting  $l_\nu^i = 0$  for all  $i \in \{1, \dots, d\}$ ,  $\nu \in \mathcal{V}$  for portfolios of type *II*.

Hence, if one wants to learn a linear path-functional portfolio which outperforms a given benchmark portfolio  $\bar{\pi}$ , one should use the benchmark portfolio as the auxiliary portfolio, i.e. set  $\tau = \bar{\pi}$ , since then the benchmark portfolio is included in the family of portfolios one optimizes over. In the context of stochastic portfolio theory, we therefore often choose  $\tau = \mu$ , because we aim to outperform the market portfolio.

Among all linear path-functional portfolios, we are particularly interested in those, whose feature maps are (randomized) elements of the signature, resulting in (*JL- or randomized-signature portfolios*). We make this more precise in the following definition:

**Definition 3.8** (Signature Portfolios, JL- and Randomized-Signature Portfolios). For a given auxiliary portfolio  $\tau$  and any  $\mathbb{R}^n$ -valued continuous semimartingale  $X$ , we call linear path-functional portfolios (*JL- or randomized-signature portfolios*) if the feature maps are elements of the (JL- or randomized-)signature of the time-augmented semimartingale  $\hat{X}$ . More precisely, let  $\hat{X}_t = (\varphi(t), X_t)$  be the time-augmented process of  $X$ , where  $\varphi$  is a strictly increasing function. Recall that  $\hat{\mathbb{X}}^N$  denotes the signature of  $\hat{X}$  truncated at level  $N$ . We then consider the following portfolios:

- *Signature Portfolios:* A *signature portfolio of degree  $N$*  is a linear path-functional portfolio with  $\mathcal{V} = \{I \mid I = (i_1, \dots, i_m) \in \{1, \dots, n\}^m \text{ for } 0 \leq m \leq N\}$  (the set of multiindices up to length  $N$ ) and with feature maps

$$\phi^I(t, X_{[0,t]}) = \langle e_I, \hat{\mathbb{X}}_t^N \rangle = \langle e_I, \hat{\mathbb{X}}_t \rangle.$$

In other words, signature portfolios are path-functional portfolios, where the portfolio controlling functions are linear functions on the signature.

- *JL-Signature Portfolios:* A *JL-signature portfolio of dimension  $(P, N)$*  is a linear path-functional portfolio with  $\mathcal{V} = \{1, \dots, P\}$ , where  $P$  is the dimension of the projection, and for each  $p \in \{1, \dots, P\}$  with feature maps

$$\phi^p(t, X_{[0,t]}) = \langle A^p, \hat{\mathbb{X}}_t^N \rangle,$$

where  $A^p$  is the  $p$ -th column of the Johnson-Lindenstrauss projection introduced in Theorem 2.22. Hence, JL-signature portfolios are path-functional portfolios, where the portfolio controlling functions are linear functions on the Johnson-Lindenstrauss projected signature.

- *Randomized-Signature Portfolios:* Let  $\mathcal{S}$  denote the solution to

$$d\mathcal{S}_t = \sum_{i=1}^{n+1} \sigma(b^{(i)} + A^{(i)} \mathcal{S}_t) \circ d\hat{X}_t^i \quad \mathcal{S}_0 = (1, 0, 0, \dots) \in \mathbb{R}^p,$$

where  $\sigma$  is an activation function, randomly chosen  $b^{(i)} \in \mathbb{R}^p$ ,  $A^{(i)} \in \mathbb{R}^{p \times p}$  for all  $i \in \{1, \dots, n+1\}$  and  $\hat{X}$  is the time-augmented process of an  $\mathbb{R}^n$ -valued continuous semimartingale  $X$ . We call  $\mathcal{S}(\hat{X})$  the randomized signature of dimension  $p$  of  $\hat{X}$ , see [1, 13, 18, 19]. For a connection with neural signature kernels and controlled ResNets we refer to [59]. A *randomized-signature portfolio of dimension  $P$*  is a linear path-functional portfolio with  $\mathcal{V} = \{1, \dots, P\}$ , where  $P$  is the dimension of the randomized signature, and with feature maps

$$\phi^p(t, X_{[0,t]}) = \langle e_p, \mathcal{S}_t \rangle.$$

Hence, randomized-signature portfolios are path-functional portfolios, where the portfolio controlling functions are linear combinations of the elements of the randomized signature.

*Remark 3.9.* We have highlighted signature, JL- and randomized-signature portfolios as linear path-functional portfolios of special interest in this paper. However, we would like to mention some other examples of linear path-functional portfolios whose portfolio controlling functions are

- linear functions on increments of the signature. That is, for a fixed time span  $\Theta$ , we observe a rolling window of length  $\Theta < T$  and forget the information before that, i.e.

$$\phi^I(t, X_{[0,t]}) = \langle e_I, \hat{\mathbb{X}}_{0\vee(t-\Theta),t} \rangle.$$

We will refer to such portfolios as signature portfolios *with rolling windows*. We shall not put particular emphasis on these portfolios in this paper, since our approximation results in Section 4 are tailored to signature portfolios *without* rolling windows and since portfolios with rolling windows are numerically less tractable, as we outline in Section 7.

- random neural networks, which are neural networks where the parameters of the hidden layers are not trained but randomly sampled and only the linear read-out layer is trained, see [38]. Training the read-out layer, exactly amounts to training the parameters  $\{l_\nu^i\}_{1 \leq i \leq d, \nu \in \mathcal{V}}$  of path-functional portfolios. Such neural networks may have an infinite-dimensional input space, corresponding in our context to path spaces, as e.g. considered in [22] or [7].
- given by reservoir computers with linear read-out layers, see for instance [37]. A special class of which are Echo State Networks [42, 36, 35] or Quantum Reservoir Computers [60, 32]. Note, that there is an interesting connection between the Johnson-Lindenstrauss projection of signatures, randomized signature and reservoir computing, which is worked out in [18].
- given by a constant, which leads to the case of constant portfolio weights when using a constant auxiliary portfolio  $\tau$ . In this case we recover the Markowitz portfolio optimization as a very special case in Subsection 5.1. Moreover, since the first element of the signature is constant, this class of portfolios is also included in the class of signature portfolios.

## 4 Approximation Properties of Signature Portfolios

This section is dedicated to prove universal approximation properties of signature portfolios. Let us recall that for a compact subset  $K \subset \mathcal{G}_T^2(\varphi, x)$  the set

$$\mathcal{K}_\Lambda = \lambda([0, T] \times K) = \left\{ \hat{\mathbb{X}}_{[0,t]}^2(\omega) \mid t \in [0, T] \text{ and } \hat{\mathbb{X}}_{[0,T]}^2(\omega) \in K \right\} \quad (4.1)$$

is a compact subset of  $\Lambda_T^2(\varphi, x)$  by the continuity of  $\lambda$ , see Lemma A.2. Furthermore, recall that in Definition 3.1 we have introduced portfolio controlling functions  $\{f^i\}_{1 \leq i \leq d}$  as non-anticipative path-functionals taking as input the paths of some semimartingale  $X$ , i.e. the  $\{f^i\}_{1 \leq i \leq d}$  were of form  $f^i(t, X_{[0,t]})$ . In the following, we will require a more specific form, namely portfolio controlling functionals depending on the lifted path, that is  $f^i(\hat{\mathbb{X}}_{[0,t]}^2)$ , where the dependence on the lifted path is important for the continuity statements. However, this is not a contradiction to the general definition, in particular since  $\hat{\mathbb{X}}^2$  is a continuous semimartingale itself. Note that the functions  $\{f^i\}_{1 \leq i \leq d}$  can of course only depend on the first level  $X_{[0,t]}$  as well.

**Theorem 4.1.** *Consider an arbitrary but fixed auxiliary portfolio  $\tau$  and a path-functional portfolio  $\pi(\tau, \cdot)$  of type I (type II) with portfolio controlling functions  $f^i \in C(\mathcal{K}_\Lambda; \mathbb{R})$  for all  $i \in \{1, \dots, d\}$ . Then for every  $\epsilon > 0$  there exists a signature portfolio  $\pi^*(\tau, \cdot)$  of type I (type II) such that it holds for all  $i \in \{1, \dots, d\}$*

$$\sup_{(t, \hat{\mathbb{X}}_{[0,T]}^2(\omega)) \in [0, T] \times K} |\pi^i(\tau, \hat{\mathbb{X}}_{[0,t]})(\omega) - \pi^{*,i}(\tau, \hat{\mathbb{X}}_{[0,t]})(\omega)| < \epsilon.$$

Moreover, take  $\frac{1}{3} < \alpha < \frac{1}{2}$ ,  $0 \leq \alpha' < \alpha$ , an arbitrary but fixed auxiliary portfolio  $\tau$  and a path-functional portfolio  $\pi(\tau, \cdot)$  of type I (type II) with portfolio controlling functions such that for all

$i \in \{1, \dots, d\}$   $f^i \in \mathcal{B}_\psi \left( (\bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x), d_{\Lambda^{\alpha'}}) \right)$ . Then for every continuous semimartingale  $X$  starting in  $x$  and fulfilling  $\mathbb{E} \left[ \exp(\zeta \|\hat{\mathbb{X}}_{[0, T]}^2\|_{\alpha, [0, T]}^\xi) \right] < \infty$  with  $\zeta > 0$  and  $\xi > 2$ , it holds that for every  $\epsilon, \delta > 0$  there exists a signature portfolio  $\pi^*(\tau, \cdot)$  of type I (type II) such that it holds for all  $i \in \{1, \dots, d\}$

$$\mathbb{P} \left[ \sup_{t \in [0, T]} \left| \pi^i(\tau, \hat{X}_{[0, t]}) - \pi^{*, i}(\tau, \hat{X}_{[0, t]}) \right| > \epsilon \right] < \delta.$$

*Proof.* We present the proof for the case of portfolios of type I, however the proof is analogous for the case of type II.

Let us start by showing the first statement of the theorem. By the universal approximation result (Theorem 2.12), we know that for each  $\{f^i\}_{1 \leq i \leq d}$  there exists a linear function on the signature of  $\hat{X}$  that approximates  $f^i$  arbitrarily well, on compact sets. Those linear functions can be chosen as the portfolio controlling functions  $f^i$ , of a signature portfolio  $\pi^*$ . Hence, there exists for each  $\epsilon' > 0$  a signature portfolio  $\pi^*$  of type I such that for all  $i \in \{1, \dots, d\}$  it holds that

$$\sup_{(t, \hat{\mathbb{X}}_{[0, T]}^2(\omega)) \in [0, T] \times K} |f^i(\hat{\mathbb{X}}_{[0, t]}^2(\omega)) - f^{*, i}(\hat{\mathbb{X}}_{[0, t]}^2(\omega))| < \epsilon' \quad (4.2)$$

for almost all  $\omega \in \Omega$ .

Now, we show that this translates to the portfolio weights. The statement is trivial for all  $\omega \in \Omega$  where  $\hat{\mathbb{X}}_{[0, T]}^2 \notin K$ .

On  $\Omega^{(K)} := \{\omega \mid \hat{\mathbb{X}}_{[0, T]}^2(\omega) \in K\}$ , it is implied by (4.2) that

$$\begin{aligned} \sup_{t \in [0, T]} \left| \pi_t^i - \pi_t^{*, i} \right| &= \sup_{t \in [0, T]} \left| \tau_t^i \left( f^i(\hat{\mathbb{X}}_{[0, t]}^2) + 1 - \sum_{j=1}^d \tau_t^j f^j(\hat{\mathbb{X}}_{[0, t]}^2) \right) \right. \\ &\quad \left. - \tau_t^i \left( f^{*, i}(\hat{\mathbb{X}}_{[0, t]}^2) + 1 - \sum_{j=1}^d \tau_t^j f^{*, j}(\hat{\mathbb{X}}_{[0, t]}^2) \right) \right| \\ &= \sup_{t \in [0, T]} \left| \tau_t^i \left( f^i(\hat{\mathbb{X}}_{[0, t]}^2) - f^{*, i}(\hat{\mathbb{X}}_{[0, t]}^2) \right) - \sum_{j=1}^d \tau_t^j \left( f^j(\hat{\mathbb{X}}_{[0, t]}^2) - f^{*, j}(\hat{\mathbb{X}}_{[0, t]}^2) \right) \right| \\ &\leq \sup_{t \in [0, T]} \left| \tau_t^i \left( f^i(\hat{\mathbb{X}}_{[0, t]}^2) - f^{*, i}(\hat{\mathbb{X}}_{[0, t]}^2) \right) \right| + \sum_{j=1}^d \sup_{t \in [0, T]} \left| \tau_t^j \left( f^j(\hat{\mathbb{X}}_{[0, t]}^2) - f^{*, j}(\hat{\mathbb{X}}_{[0, t]}^2) \right) \right| \\ &\leq M \sup_{t \in [0, T]} \left| f^i(\hat{\mathbb{X}}_{[0, t]}^2) - f^{*, i}(\hat{\mathbb{X}}_{[0, t]}^2) \right| + M \sum_{j=1}^d \sup_{t \in [0, T]} \left| f^j(\hat{\mathbb{X}}_{[0, t]}^2) - f^{*, j}(\hat{\mathbb{X}}_{[0, t]}^2) \right| \\ &< M(d+1)\epsilon' \end{aligned}$$

holds almost surely, where  $M$  is the bound of  $\tau$ . The result follows by choosing  $\epsilon' = \frac{\epsilon}{M(d+1)}$ .

The proof of the second approximation statement follows by similar arguments. Recall Theorem 2.20 and use the same reasoning as above to translate the approximation result from the portfolio controlling functions to the portfolio weights itself, i.e. we obtain that there exists for each  $\epsilon' > 0$  a signature portfolio  $\pi^*$  of type I such that for all  $i \in \{1, \dots, d\}$  it holds that

$$\sup_{t \in [0, T]} \frac{|\pi^i(\tau, \hat{X}_{[0, t]}) - \pi^{*, i}(\tau, \hat{X}_{[0, t]})|}{\exp(\zeta \|\hat{\mathbb{X}}_{[0, T]}^2\|_{\alpha, [0, T]}^\xi)} < \epsilon' \quad (4.3)$$

almost surely. Therefore,

$$\begin{aligned} \mathbb{P} \left[ \sup_{t \in [0, T]} \left| \pi^i(\tau, \hat{X}_{[0, t]}) - \pi^{*, i}(\tau, \hat{X}_{[0, t]}) \right| > \epsilon \right] &\leq \epsilon^{-1} \cdot \mathbb{E} \left[ \sup_{t \in [0, T]} \left| \pi^i(\tau, \hat{X}_{[0, t]}) - \pi^{*, i}(\tau, \hat{X}_{[0, t]}) \right| \right] \\ &< \frac{\epsilon'}{\epsilon} \mathbb{E} \left[ \sup_{t \in [0, T]} \exp(\zeta \|\hat{\mathbb{X}}_{[0, T]}^2\|_{\alpha, [0, T]}^\xi) \right], \end{aligned}$$

where the first inequality is due to Markov's inequality and the second one follows from (4.3). Note that by the properties of  $\|\cdot\|_{cc}$  and  $\|\cdot\|_{\alpha, [0, T]}$  it holds that  $\sup_{t \in [0, T]} \|\hat{\mathbb{X}}_{[0, T]}^2\|_{\alpha, [0, T]} = \|\hat{\mathbb{X}}_{[0, T]}^2\|_{\alpha, [0, T]}$ . Hence the statement is obtained by setting  $\delta = \frac{\epsilon'}{\epsilon} \mathbb{E} \left[ \exp(\zeta \|\hat{\mathbb{X}}_{[0, T]}^2\|_{\alpha, [0, T]}^\xi) \right]$  and choosing  $\epsilon'$  accordingly.  $\square$

*Remark 4.2.* Note, that by Corollary 2.13, Theorem 4.1 also ensures that signature portfolios approximate classical functionally generated portfolios arbitrarily well, on compact sets. In particular, this includes the functionally generated portfolios as introduced by Fernholz [27] and those considered for functional portfolio optimization in [10].

*Remark 4.3.* The requirements of Theorem 4.1 on  $f^i$  is a simple continuity condition and examples for such continuous path-functionals are provided in Example 2.13.

Having obtained a universal approximation theorem of path-functional portfolios, we will now apply it to obtain universal approximation results of the growth-optimal portfolio by signature portfolios in several markets.

## 4.1 Universal Approximation of the Growth-Optimal Portfolio

In this subsection, we present a large class of markets where the growth-optimal portfolio can be regarded as a path-functional portfolio and show that it can therefore be approximated by a signature portfolio.

The following statements are to be understood in the market setting outlined in Definition 2.32 and we therefore assume that the corresponding necessary conditions are satisfied, in particular those ensuring the existence of the growth-optimal portfolio (Lemma 2.33). Moreover, recall also the definition of the compact set  $\mathcal{K}_\Lambda$  given in (4.1).

**Theorem 4.4.** *Consider the following class of markets with  $d$  stocks:*

$$dS_t = \text{diag}(S_t)(a(\hat{\mathbb{X}}_{[0, t]}^2)dt + \Sigma(\hat{\mathbb{X}}_{[0, t]}^2)dB_t), \quad (4.4)$$

*where the components of  $a$ ,  $\Sigma$  are non-anticipative path-functionals in  $C(\mathcal{K}_\Lambda; \mathbb{R})$  and  $X$  some  $\mathbb{R}^n$ -valued semimartingale with  $n \in \mathbb{N}$ . Then for any auxiliary portfolio  $\tau$  there exists a signature portfolio of type II which approximates the weights of the growth-optimal portfolio arbitrarily well on  $\mathcal{K}_\Lambda$ . Moreover, if the components of the auxiliary portfolio  $\tau$  are deterministic, then there also exists a signature portfolio of type I which approximates the weights of the growth-optimal portfolio almost surely arbitrarily well on  $\mathcal{K}_\Lambda$ .*

*Alternatively, let  $\frac{1}{3} < \alpha < \frac{1}{2}$ ,  $0 \leq \alpha' < \alpha$  and let  $a, \Sigma$  be such that the components of  $\pi^{(g)}$  are non-anticipative path-functionals in  $\mathcal{B}_\psi \left( (\bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x), d_{\Lambda^{\alpha'}}) \right)$ .<sup>3</sup> Then for every continuous semimartingale  $X$  starting in  $x$  and fulfilling  $\mathbb{E} \left[ \exp(\zeta \|\hat{\mathbb{X}}_{[0, T]}^2\|_{\alpha, [0, T]}^\xi) \right] < \infty$  with  $\zeta > 0$*

<sup>3</sup>Note that a sufficient condition for this to hold is if the components of  $a$  are non-anticipative path-functionals in  $\mathcal{B}_\psi \left( (\bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x), d_{\Lambda^{\alpha'}}) \right)$  and  $\Sigma$  is constant.

and  $\xi > 2$ , it holds that for every  $\epsilon, \delta > 0$  there exists a signature portfolio  $\pi^*(\tau, \cdot)$  for type II for any auxiliary portfolio  $\tau$  (or of type I for a deterministic portfolio  $\tau$ ) such that it holds for all  $i \in \{1, \dots, d\}$

$$\mathbb{P} \left[ \sup_{t \in [0, T]} \left| \pi^{(g), i}(\tau, \hat{X}_{[0, t]}) - \pi^{*, i}(\tau, \hat{X}_{[0, t]}) \right| > \epsilon \right] < \delta,$$

where  $\pi^{(g)}$  is the growth-optimal portfolio of the respective market.

*Remark 4.5.* Note that the semimartingale  $X$  in (4.4) could be (a functions of)  $S$ , so that we are actually dealing with path-dependent SDEs. In the particular case where  $a$  and  $\Sigma$  are entire functions of the signature as introduced in [23] we are then in the tractable setup of *signature SDEs*.

*Proof of Theorem 4.4.* Recalling Lemma 2.33, it is straightforward that for any auxiliary portfolio  $\tau$ , the growth-optimal portfolio  $\pi^{(g)}$  is a path-functional portfolio of type II with portfolio controlling functions

$$f^{(g, II), i} = \pi^{(g), i}.$$

Likewise,  $\pi^{(g)}$  is a path-functional portfolio of type I for any deterministic auxiliary portfolio and portfolio controlling functions

$$f^{(g, I), i} = \tau^i \cdot \pi^{(g), i}.$$

By Lemma B.1 and the form of  $\pi^{(g), i}$  Theorem 4.1 is applicable for  $\pi^{(g), i}$  (viewed as a path-functional portfolio of type I resp. type II). Hence, the statements follow.  $\square$

*Remark 4.6.* If  $X = \mu$  or  $X = S$ , the growth optimal is also a path-functional portfolio of type I with auxiliary portfolio  $\tau = \mu$ .

*Remark 4.7 (Examples of Markovian Markets).* We here give two prominent examples of markets, to which Theorem 4.4 can be applied.

- Black-Scholes Market:

$$dS_t = \text{diag}(S_t)(adt + \Sigma dB_t),$$

where  $a$  and  $\Sigma$  are constant.

- Volatility Stabilized Markets:

$$\frac{dS_t^i}{S_t^i} = \frac{1+\alpha}{2} \frac{1}{\mu_t^i} dt + \sqrt{\frac{1}{\mu_t^i}} dB_t^i \text{ for all } 1 \leq i \leq d$$

and  $\alpha \geq 0$ . Volatility stabilized markets are of great interest in SPT because they reflect the observation in real markets that smaller stocks tend to have greater volatility than larger stocks, see for example [29]. See also [15, 28] for further properties of volatility stabilized markets.

## 4.2 A Class of Non-Markovian Markets where the Growth-Optimal Portfolio is a Signature Portfolio

Let us now turn to a class of (possibly) non-Markovian markets, where the growth-optimal portfolio can not only be approximated by a signature portfolio but *is* a signature portfolio.

**Theorem 4.8.** Let  $X$  be a continuous semimartingale. For a market of  $d$  stocks, consider the class of Sig-market models

$$dS_t = \text{diag}(S_t)(a(\hat{X}_{[0,t]})dt + \Sigma dB_t)$$

where  $a(\hat{X}_{[0,t]})_i = \sum_{0 \leq |I| \leq N} \alpha_I^{(i)} \langle e_I, \hat{\mathbb{Y}}_t \rangle$  for  $N \geq 0$ ,  $\alpha_I^{(i)} \in \mathbb{R}$ ,  $\Sigma$  is a constant  $d \times m$  matrix and  $B_t$  is an  $m$ -dimensional Brownian motion for  $m \geq d$ . Then, the growth-optimal portfolio is for any auxiliary portfolio a signature portfolio of type II and a signature portfolio of type I for any constant and deterministic auxiliary portfolio  $\tau$ .

*Proof.* The integrability and measurability assumptions required in Definition 2.32 and Lemma 2.33 follow from the integrability and measurability of elements of the signature. Applying Lemma 2.33, it follows that the growth-optimal weights are linear functions on the signature of  $\hat{X}$ . This is for any auxiliary portfolio a signature portfolio of type II, or a signature portfolio of type I for any constant and deterministic auxiliary portfolio  $\tau$ .  $\square$

*Remark 4.9* (Existence of Solutions to Sig-Markets). As addressed in Remark 4.5, the existence of solutions is non-trivial for Sig-markets if  $X = h(S)$ , i.e. if the semimartingale of which we construct the signature is a function of the price process itself. In particular, if  $h$  is a real analytic function, then we deal with signature-SDEs in the spirit of [23]. The study of existence of solutions to such equations goes beyond the subject of this paper, however, we give a simple example below, where the existence is guaranteed. To this end, let us denote by

$$\mathcal{I}_1 := \{I \mid I = (i_1, \dots, i_m) \text{ for } 0 \leq n \leq N \text{ and } i_2 = i_3 = \dots = i_n = 1\}$$

Let  $X = \log S_t$ . If  $\alpha_I^{(i)} = 0$  for all  $I \notin \mathcal{I}_1$ , then

$$dS_t = \text{diag}(S_t) \cdot \left( a((\widehat{\log S})_{[0,t]})dt + \Sigma dB_t \right) \quad (4.5)$$

admits a unique strong solution. To see this, note that (4.5) can be seen as a part of a (larger) system of linear equations. Setting  $Y = \log S$ ,  $\hat{Y} = (t, Y)$  and denoting by  $\hat{\mathbb{Y}}^N|_{\mathcal{I}_1} = (\langle e_I, \hat{\mathbb{Y}} \rangle)_{I \in \mathcal{I}_1}$  the restriction to elements of the signature corresponding to multiindices in  $\mathcal{I}_1$ . Let us denote the vectorization of the truncated signature denoted by  $\text{vec}(\cdot)$  and fix a labelling function  $\mathcal{L} : \mathcal{I}_1 \rightarrow \{1, \dots, |\mathcal{I}_1|\}$  such that  $\text{vec}(\hat{\mathbb{Y}}^N|_{\mathcal{I}_1})_{\mathcal{L}(I)} = \langle e_I, \hat{\mathbb{Y}}^N|_{\mathcal{I}_1} \rangle$ . Moreover, for a multiindex  $I = (i_1, \dots, i_{|I|})$ , we denote by  $I' := (i_1, \dots, i_{|I|-1})$  corresponding multiindex shortened by the last letter. Then said linear system of equations is

$$d \left( \text{vec}(\hat{\mathbb{Y}}^N|_{\mathcal{I}_1})_t \right) = \left( b + A \cdot \text{vec}(\hat{\mathbb{Y}}^N|_{\mathcal{I}_1})_t \right) dt + \tilde{\Sigma} dB_t.$$

where  $\tilde{\Sigma}_{j,r} = \Sigma_{s,r} \delta_{j,\mathcal{L}(s)}$  for  $1 \leq j \leq |\mathcal{I}_1|$ ,  $1 \leq r \leq m$  and for  $k, l \in \{1, \dots, |\mathcal{I}_1|\}$

$$b_k = \begin{cases} -\frac{1}{2} \Sigma^\top \Sigma + a_\emptyset^{(i)} & \text{if } k = \mathcal{L}(i), i \neq 1 \\ 1 & \text{if } k = \mathcal{L}(1) \\ 0 & \text{else} \end{cases}$$

$$A_{k,l} = \begin{cases} a_{\mathcal{L}(J)}^{(i)} & \text{if } k = \mathcal{L}(i), l = \mathcal{L}(J), i \neq 1 \\ 1 & \text{if } |I| \geq 2, k = \mathcal{L}(I), l = \mathcal{L}(I') \\ 0 & \text{else.} \end{cases}$$

*Remark 4.10* (NUPBR in Sig-Markets). The NUPBR condition in the above Sig-markets holds. This follows from the existence of the growth-optimal portfolio by Lemma 2.33, Lemma 2.34 and Theorem 2.28.

## 5 Optimization Tasks for Linear Path-Functional Portfolios

We now want to study some tasks for portfolio optimization and their form for linear path-functional portfolios. To formulate them in the most general way, let us introduce what we call a *universe* of stocks.

**Definition 5.1** (Universe). Consider a market of  $d$  stocks with capitalization process  $S = (S^1, \dots, S^d)$ . We call  $\mathcal{U} \subseteq \{1, \dots, d\}$  a *universe* of stocks with capitalization process  $S^{\mathcal{U}} = (S^u)_{u \in \mathcal{U}}$ . Moreover, we define the universe weights (resp. universe portfolio)  $\mu^{\mathcal{U}}$  to be

$$\mu_t^{\mathcal{U},i} = \begin{cases} \frac{S_t^{\mathcal{U},i}}{\sum_{j \in \mathcal{U}} S_t^{\mathcal{U},j}} & \text{if } i \in \mathcal{U} \\ 0 & \text{otherwise} \end{cases}.$$

Moreover, denote by  $(W_t^{\mathcal{U}})_{t \in [0, T]}$  the wealth process of universe  $\mathcal{U}$ , which is given by

$$W_t^{\mathcal{U}} = \frac{\sum_{i \in \mathcal{U}} S_t^i}{\sum_{i \in \mathcal{U}} S_{t_0}^i}.$$

This notion of a universe of stocks is useful if one does not necessarily want to invest in all the stocks in the market but just in a subset of stocks. Of course, this can always be achieved by fixing certain weights of a portfolio to be zero, however using our notion of a universe of stocks is particularly useful, if one wants to compare the wealth process of a portfolio to the wealth process of the universe. Let us make this precise:

**Corollary 5.2.** *Take a universe  $\mathcal{U} \subseteq \{1, \dots, d\}$  and consider a portfolio  $\pi$  where  $\pi^i \equiv 0$  for  $i \notin \mathcal{U}$ . Then, the relative wealth process of  $\pi$  with respect to  $\mu^{\mathcal{U}}$  is given by*

$$d \left( \frac{W_t^{\pi}}{W_t^{\mathcal{U}}} \right) = \left( \frac{W_t^{\pi}}{W_t^{\mathcal{U}}} \right) \sum_{i \in \mathcal{U}} \pi_t^i \frac{d\mu_t^{\mathcal{U},i}}{\mu_t^{\mathcal{U},i}}. \quad (5.1)$$

Or, equivalently

$$d \log \left( \frac{W_t^{\pi}}{W_t^{\mathcal{U}}} \right) = \sum_{i \in \mathcal{U}} \pi_t^i \frac{d\mu_t^{\mathcal{U},i}}{\mu_t^{\mathcal{U},i}} - \frac{1}{2} \sum_{i, j \in \mathcal{U}} \frac{\pi_t^i}{\mu_t^{\mathcal{U},i}} \frac{\pi_t^j}{\mu_t^{\mathcal{U},j}} d[\mu^{\mathcal{U},i}, \mu^{\mathcal{U},j}]_t. \quad (5.2)$$

*Proof.* Using the definition of  $\pi$  and  $W^{\mathcal{U}}$ , the statement follows directly from (2.3) and applying Itô's formula.  $\square$

**Remark 5.3.** When we consider path-functional portfolios investing in a universe  $\mathcal{U}$ , it is convenient to construct them via the universe portfolio  $\mu^{\mathcal{U}}$ . Note that for path-functional portfolios  $\pi^{(I)}$  of type I it directly follows that  $\pi^{(I),i}(\mu^{\mathcal{U}}, \cdot) = 0$  for  $i \notin \mathcal{U}$ . However, for path-functional portfolios  $\pi^{(II)}$  of type II, we have to require additionally that  $\pi^{(II),i}(\mu^{\mathcal{U}}, \cdot) = 0$  for  $i \notin \mathcal{U}$  by setting  $f^i \equiv 0$  for  $i \notin \mathcal{U}$ .

**Remark 5.4** (Investments in Ranked Markets). Rank-based portfolios are of particular interest in SPT, see [27]. We aim to therefore incorporate those into our considerations. Recall that our very basic assumptions on the financial market were that the stocks' capitalizations are positive continuous semimartingales. Considering a market  $M$  which fulfills these assumptions, we know that there exists another market which also fulfills these assumptions and for which the stocks' capitalizations are the same as the ranked capitalizations of the market  $M$ . This follows simply from the fact that the ranked capitalizations are again positive continuous semimartingales. However, the assumption of NUPBR does not translate directly (see [47, Section 3.2] for a comment on the potential lack of a local martingale deflator in ranked markets and [48] for a connection between the existence of a local martingale deflator and the NUPBR condition). For

the optimization problems to be meaningful in the ranked market, we have to assume NUPBR again additionally. Hence, under the assumption of NUPBR for the ranked market weights, the optimization problems we present in the following hold also for investments in the ranked market.

It is sometimes useful to write optimization tasks in a vectorized form. To this end, let us introduce what we call a *labelling function*.

**Definition 5.5.** For a given linear path-functional portfolio  $\pi$  investing in a universe of stocks  $\mathcal{U}$ , let  $\mathcal{V}$  be the set of features,  $|\mathcal{V}|$  the number of features and  $|\mathcal{U}|$  the number of stocks in a universe. A bijective function  $\mathcal{L}$  of the form

$$\begin{aligned}\mathcal{L} : \mathcal{U} \times \mathcal{V} &\longrightarrow \{1, \dots, |\mathcal{U}| \cdot |\mathcal{V}|\} \\ (i, \nu) &\longmapsto \mathcal{L}(i, \nu)\end{aligned}$$

is called *labelling function*.

Before studying two concrete optimization tasks for linear path-functional portfolios, let us make the following statement about a more general class of optimization tasks, which turn out to be *quadratic optimization problems* for linear path-functional portfolios:

**Proposition 5.6.** Let us denote by  $\mathbb{S}^d$  the space of symmetric matrices of dimension  $d \times d$ . For a universe  $\mathcal{U}$ , consider a class of linear path-functional portfolios  $\Pi$  (of either type) given by a feature space  $\mathcal{V}$ , a collection of feature maps  $\{\phi^\nu\}_{\nu \in \mathcal{V}}$  consisting of non-anticipative path-functionals, a continuous semimartingale  $X$  being the underlying process and an auxiliary portfolio  $\tau$ . Recall that the portfolio controlling functionals  $\{f^i\}_{i \in \mathcal{U}}$  of such portfolios are given by  $f^i(t, X_{[0,t]}) = \sum_{\nu \in \mathcal{V}} l_\nu^i \phi^\nu(t, X_{[0,t]})$  for  $i \in \mathcal{U}$ . Then, any optimization problem of the form

$$\inf_{\pi \in \Pi} \mathbb{E} \left[ \int_{t_0}^t \pi_s^\top C_s \pi_s \lambda_1(ds) - \int_{t_0}^t b_s^\top \pi_s \lambda_2(ds) \right]$$

where  $b, C$  are stochastic processes with values in  $\mathbb{R}^d, \mathbb{S}^d$  respectively and  $\lambda^i$  are signed measures on  $[t_0, t]$ , is a quadratic optimization problem in the linear coefficients  $\{l_\nu^i\}_{\nu \in \mathcal{V}, 1 \leq i \leq d}$ .

*Proof.* Recall the form of linear path-functional portfolios of type I and II respectively. Therefore, it is clear that the parameters  $\{l_\nu^i\}$  appear at most in quadratic terms. Moreover, since they are constant in time and deterministic, they can be pulled outside of the integrals.  $\square$

Let us now, present two prominent optimization tasks which fall into the above class, namely *mean-variance* and *log-wealth optimization*. We assume that the following optimization problems are well-posed and that the expectations are finite.

## 5.1 Mean-Variance Optimization

In this optimization problem, we consider a strategy, where we invest at time  $t$  with portfolio weights  $\pi_t$  of a linear path-functional portfolio into a universe  $\mathcal{U}$  and follow a buy-and-hold strategy over the time-span  $\Delta$ . The (relative) return of this strategy over the time-horizon  $[t, t + \Delta]$  is given by

$$\mathcal{R}_{t,t+\Delta}^{W,\pi} = \frac{W_{t+\Delta}^\pi}{W_t^\pi} - 1 = \sum_{j \in \mathcal{U}} \pi_t^j \frac{S_{t+\Delta}^j}{S_t^j} - 1 \quad (5.3)$$

and respectively the relative return (relative to the universe portfolio  $\mu^{\mathcal{U}}$ )

$$\mathcal{R}_{t,t+\Delta}^{V,\pi} = \frac{V_{t+\Delta}^\pi}{V_t^\pi} - 1 = \sum_{j \in \mathcal{U}} \pi_t^j \frac{\mu_{t+\Delta}^{\mathcal{U},j}}{\mu_t^j} - 1. \quad (5.4)$$

In this setting, we can state the following corollary about the mean-variance optimization problem:

**Corollary 5.7.** *Let  $\lambda \in \mathbb{R}_+$  be a risk-tolerance parameter, specifying the investor's risk-preferences. Then, for a linear path-functional portfolio  $\pi(\tau, X)$  the following two optimization problems are convex quadratic optimization problems in the parameters  $\{l_\nu^i\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}$ . More precisely, for the*

1. mean-variance optimization of returns we have

$$\begin{aligned} & \min_{\pi} \text{Var}(\mathcal{R}_{t,t+\Delta}^{W,\pi}) - \lambda \mathbb{E}(\mathcal{R}_{t,t+\Delta}^{W,\pi}) \\ \Leftrightarrow & \min_{\{l_\nu^i\}} \mathbf{l}^\top \text{Var}(\mathbf{Y}^W) \mathbf{l} - \mathbf{l}^\top (\lambda \mathbb{E}(\mathbf{Y}^W) - \boldsymbol{\sigma}^W) \end{aligned}$$

for  $\mathbf{l}_{\mathcal{L}(i,\nu)} = l_\nu^i$ ,  $\boldsymbol{\sigma}_{\mathcal{L}(i,\nu)}^W = 2\text{Cov}(\mathbf{Y}_{\mathcal{L}(i,\nu)}^W, \mathcal{R}_{t,t+\Delta}^{W,\tau})$ ,

$$\mathbf{Y}_{\mathcal{L}(i,\nu)}^W = \begin{cases} \tau_t^i \phi^\nu(t, X_{[0,t]}) \left( \frac{S_{t+\Delta}^i}{S_t^i} - \mathcal{R}_{t,t+\Delta}^{W,\tau} - 1 \right) & \text{if } \pi \text{ is of type I} \\ \phi^\nu(t, X_{[0,t]}) \left( \frac{S_{t+\Delta}^i}{S_t^i} - \mathcal{R}_{t,t+\Delta}^{W,\tau} - 1 \right) & \text{if } \pi \text{ is of type II} \end{cases}$$

and where  $\text{Var}(\mathbf{Y}^W)$  refers to the covariance matrix of  $\mathbf{Y}^W$ .

2. mean-variance optimization of relative return we have

$$\begin{aligned} & \min_{\pi} \text{Var}(\mathcal{R}_{t,t+\Delta}^{V,\pi}) - \lambda \mathbb{E}(\mathcal{R}_{t,t+\Delta}^{V,\pi}) \\ \Leftrightarrow & \min_{\{l_\nu^i\}} \mathbf{l}^\top \text{Var}(\mathbf{Y}^V) \mathbf{l} - \mathbf{l}^\top (\lambda \mathbb{E}(\mathbf{Y}^V) - \boldsymbol{\sigma}^V) \end{aligned}$$

for  $\mathbf{l}_{\mathcal{L}(i,\nu)} = l_\nu^i$ ,  $\boldsymbol{\sigma}_{\mathcal{L}(i,\nu)}^V = 2\text{Cov}(\mathbf{Y}_{\mathcal{L}(i,\nu)}^V, \mathcal{R}_{t,t+\Delta}^{V,\tau})$  and

$$\mathbf{Y}_{\mathcal{L}(i,\nu)}^V = \begin{cases} \tau_t^i \phi^\nu(t, X_{[0,t]}) \left( \frac{\mu_{t+\Delta}^{\mathcal{U},i}}{\mu_t^{\mathcal{U},i}} - \mathcal{R}_{t,t+\Delta}^{V,\tau} - 1 \right) & \text{if } \pi \text{ is of type I} \\ \phi^\nu(t, X_{[0,t]}) \left( \frac{\mu_{t+\Delta}^{\mathcal{U},i}}{\mu_t^{\mathcal{U},i}} - \mathcal{R}_{t,t+\Delta}^{V,\tau} - 1 \right) & \text{if } \pi \text{ is of type II} \end{cases}$$

and where  $\text{Var}(\mathbf{Y}^V)$  refers to the covariance matrix of  $\mathbf{Y}^V$ .

*Proof.* We proof the above statement for the case of  $\mathcal{R}^{W,\pi}$  and for  $\pi$  of type I. The proof of the other cases is completely analogous.

$$\begin{aligned} \mathcal{R}_{t,t+\Delta}^{W,\pi} &= \sum_{i \in \mathcal{U}} \tau_t^i \left( f^i(t, X_{[0,t]}) + 1 - \sum_{j \in \mathcal{U}} \tau_t^j f^j(t, X_{[0,t]}) \right) \left( \frac{S_{t+\Delta}^i}{S_t^i} \right) - 1 \\ &= \sum_{i \in \mathcal{U}} \tau_t^i f^i(t, X_{[0,t]}) \left( \frac{S_{t+\Delta}^i}{S_t^i} - \mathcal{R}_{t,t+\Delta}^{W,\tau} - 1 \right) + \mathcal{R}_{t,t+\Delta}^{W,\tau} + 1 - 1 \\ &= \mathbf{l}^\top \mathbf{Y}^W + \mathcal{R}_{t,t+\Delta}^{W,\tau} \end{aligned}$$

Hence,

$$\mathbb{E}(\mathcal{R}_{t,t+\Delta}^{W,\pi}) = \mathbf{l}^\top \mathbb{E}(\mathbf{Y}^W) + \mathbb{E}(\mathcal{R}_{t,t+\Delta}^{W,\tau}). \quad (5.5)$$

Note that the rightmost term of (5.5) do not contribute in the optimization problem, as they are independent of  $\mathbf{l}$ . And,

$$\begin{aligned} \text{Var}(\mathcal{R}_{t,t+\Delta}^{W,\pi}) &= \text{Var}(\mathbf{l}^\top \mathbf{Y}^W) + 2\text{Cov}(\mathbf{l}^\top \mathbf{Y}^W, \mathcal{R}_{t,t+\Delta}^{W,\tau}) + \text{Var}(\mathcal{R}_{t,t+\Delta}^{W,\tau}) \\ &= \mathbf{l}^\top \text{Var}(\mathbf{Y}^W) \mathbf{l} + \mathbf{l}^\top \boldsymbol{\sigma}^W + \text{Var}(\mathcal{R}_{t,t+\Delta}^{W,\tau}), \end{aligned} \quad (5.6)$$

where the rightmost term of (5.6) does not contribute in the optimization problem. The fact that the optimization problem is convex follows directly from the fact that the covariance matrix  $\text{Var}(\mathbf{Y}^W)$  is positive semi-definite.  $\square$

*Remark 5.8.* Note that the expression of the optimization problem for the relative return simplifies considerably if we use a linear path-functional portfolio of type  $I$  and choose the universe portfolio as the auxiliary portfolio, i.e.  $\tau = \mu^U$ . To be precise, we obtain

$$\begin{aligned} & \min_{\pi} \text{Var}(\mathcal{R}_{t,t+\Delta}^{V,\pi}) - \lambda \mathbb{E}(\mathcal{R}_{t,t+\Delta}^{V,\pi}) \\ & \Leftrightarrow \min_{\{l_\nu^i\}} \mathbf{l}^\top \text{Var}(\mathbf{Y}^V) \mathbf{l} - \lambda \mathbf{l}^\top \mathbb{E}(\mathbf{Y}^V) \end{aligned}$$

for  $\mathbf{l}_{\mathcal{L}(i,\nu)} = l_\nu^i$  and  $\mathbf{Y}_{\mathcal{L}(i,\nu)}^V = \phi^\nu(t, X_{[0,t]})(\mu_{t+\Delta}^{\mathcal{U},i} - \mu_t^{\mathcal{U},i})$ , since  $\mathcal{R}_{t,t+\Delta}^{V,\mu^U} = 0$ .

*Remark 5.9.* If  $\text{Var}(\mathbf{Y}^V)$  or  $\text{Var}(\mathbf{Y}^W)$  respectively are almost surely invertible, that is the optimization problem is *strictly* convex and hence admits a unique set of optimal parameters, then the optimal vector  $\mathbf{l}^*$  can be found via first-order conditions, i.e.

$$\mathbf{l}^* = \text{Var}(\mathbf{Y}^V)^{-1} (\lambda \mathbb{E}(\mathbf{Y}^V) - \sigma^V)$$

and likewise for the case of optimizing returns. Note that the problem is never strictly convex for *any* linear path-functional portfolio of type  $I$ , because due to the form of such portfolios, the weights are invariant under scalar shifts of the parameters  $\{l_\nu^i\}_{\nu \in \mathcal{V}, 1 \leq i \leq d}$ . More precisely, for linear path-functional portfolios of type  $I$ , let  $\pi^*(\tau, X)$  be optimal portfolio weights corresponding to an optimal set  $\{l_\nu^{*,i}\}_{\nu \in \mathcal{V}, 1 \leq i \leq d}$ , then the set  $\{l_\nu^{*,i} + \alpha\}_{\nu \in \mathcal{V}, 1 \leq i \leq d}$  for  $\alpha \in \mathbb{R}$  results in *the same* portfolio weights  $\pi^*(\tau, X)$ .

*Remark 5.10.* Let us here outline how the mean-variance optimization problem (of the relative returns) of Corollary 5.7 can be related to a (model-free) real market situation where only the observation of one trajectory is available (see also Section 7.2). Tailored to Corollary 5.7 we assume a discrete time situation with time steps of length  $\Delta$  and suppose that the conditional law of  $\mu_{i\Delta}$  is time-homogeneous and depends on the last  $n \wedge i$  values of  $\mu$  for some  $n \in \mathbb{N}$ . More precisely, given  $(\mu_{(i-n)\Delta \wedge 0}, \dots, \mu_{(i-1)\Delta}) = (x_0, \dots, x_{(n-1)\wedge(i-1)}) \in (\Delta^d)^{n \wedge i}$ , we know the conditional law  $\rho(x_0, \dots, x_{(n-1)\wedge(i-1)}, \cdot)$  of  $\mu_{i\Delta}$ . Consider now a portfolio  $\pi$  that is a non-anticipative path functional depending also only on the past  $n$  values of  $\mu$ , i.e. a portfolio with rolling window in the spirit of Remark 3.9. Then for  $i \geq n$

$$\begin{aligned} \mathbb{E} \left[ \frac{V_{i\Delta}^\pi}{V_{(i-1)\Delta}^\pi} \middle| \mathcal{F}_{(i-1)\Delta} \right] &= \mathbb{E}[(R_{(i-1)\Delta, i\Delta}^{V,\pi} + 1) | \mathcal{F}_{(i-1)\Delta}] \\ &= \int_{\Delta^d} \sum_{j=1}^d \pi^j(x_0, \dots, x_{n-1}) \frac{y^j}{x_{n-1}^j} \rho(x_0, \dots, x_{n-1}, dy) \end{aligned}$$

and similarly for  $i < n$ , taking just the history starting from 0 into account. Assume now further that we have an invariant measure for the distribution of the rolling path segments of length  $n$ , which we denote by  $\rho$ . Then under this invariant measure we have for each  $i \geq n$

$$\begin{aligned} \mathbb{E} \left[ \frac{V_{i\Delta}^\pi}{V_{(i-1)\Delta}^\pi} \right] &= \int_{(\Delta^d)^n} \int_{\Delta^d} \sum_{j=1}^d \pi^j(x_0, \dots, x_{n-1}) \frac{y^j}{x_{n-1}^j} \rho(x_0, \dots, x_{n-1}, dy) \rho(dx_0, \dots, dx_{n-1}) \\ &=: M^\pi \end{aligned}$$

and accordingly for  $i < n$  assuming compatible invariant measures for the paths of shorter lengths. Under these assumptions we then obtain the following ergodicity result

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \frac{V_{i\Delta}^\pi}{V_{(i-1)\Delta}^\pi} = 1 + \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N R_{(i-1)\Delta, i\Delta}^{V,\pi} = M^\pi.$$

This can be deduced from Birkhoff's ergodic theorem for discrete time Markov processes (see e.g. [Theorem 2.2, Section 2.1.4][26] and also [21, Section 3] for the standard Markov situation with  $n = 1$ ). Indeed, this becomes applicable by identifying the states of the Markov process as the past path segments of length  $n$ , i.e. a state  $z_{i\Delta}$  is given by  $z_{i\Delta} = (\mu_{(i-n)\Delta}, \dots, \mu_{i\Delta})$ . The transition probabilities from  $z_{i\Delta}$  to  $z_{(i+1)\Delta}$  must then be chosen such that they assign 0 probability to elements where there exists some  $j \in \{1, \dots, n-1\}$  with  $z_{(i+1)\Delta}^j \neq z_{i\Delta}^{j+1}$ .

Clearly the above reasoning also holds true for  $\text{Var}[V_{i\Delta}^\pi / V_{(i-1)\Delta}^\pi] = \text{Var}[R_{(i-1)\Delta, i\Delta}^{V, \pi}]$ . Therefore, under such ergodicity assumptions all the expected values in Corollary 5.7 can be approximated by time averages, which will be important in our numerical implementations as outlined in Section 7.2.

## 5.2 Optimizing the (Expected) Log-(Relative)-Wealth

Similarly as the mean-variance portfolio optimization problem also the log-optimal portfolio over linear path functional portfolios can be found by solving a convex quadratic optimization task. We state it here in form of the relative log-utility optimization problem, i.e. the goal is to solve

$$\sup_{\pi} \mathbb{E} \left[ \log \left( \frac{W_t^\pi}{W_t^{\mathcal{U}}} \right) \right] = \sup_{\pi} \mathbb{E} [\log(V_t^\pi)],$$

where  $V_t^\pi$  denotes again the relative wealth with respect to the universe  $\mathcal{U}$ . Note however that under similar ergodicity conditions as in Remark 5.10 the use of time averages

$$\frac{1}{N} \log(V_{N\Delta}^\pi) = \frac{1}{N} \sum_{i=1}^N \log \left( \frac{V_{i\Delta}^\pi}{V_{(i-1)\Delta}^\pi} \right)$$

is justified to approximate  $\frac{1}{N} \mathbb{E}[\log(V_{N\Delta}^\pi)]$ . This can also be translated to continuous time, where under appropriate assumptions (see e.g. [21, Section 4.2.3 and Theorem 4.9]) we also have that

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log(V_t^\pi) = \lim_{t \rightarrow \infty} \frac{1}{t} \mathbb{E}[\log(V_t^\pi)],$$

whence

$$\frac{1}{t} \log(V_t^\pi) \approx \frac{1}{t} \mathbb{E}[\log(V_t^\pi)].$$

In this case the expected values in the subsequent theorem, i.e.  $\mathbb{E}[\mathbf{Q}(t)]$  and  $\mathbb{E}[\mathbf{c}(t)]$  can therefore be just replaced by  $\mathbf{Q}(t)$  and  $\mathbf{c}(t)$  computed along the observed trajectory. This will play again an important role in our empirical implementations.

**Theorem 5.11** (Relative Log-Utility Optimization). *Consider a universe  $\mathcal{U} \subseteq \{1, \dots, d\}$  and a linear path functional portfolio  $\pi(\mu^{\mathcal{U}}, X)$ . Let  $t_0 \geq 0$  be the time at which we start investing. Denote by  $\mathcal{L}$  an arbitrary but fixed labelling function. Then, for the (relative) log-utility optimization problem we have*

$$\sup_{\{l_\nu^{(i)}\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}} \mathbb{E} \left[ \log \left( \frac{W_t^\pi}{W_t^{\mathcal{U}}} \right) \right] \Leftrightarrow \inf_{\mathbf{l} \in \mathbb{R}^{|\mathcal{U}| \cdot |\mathcal{V}|}} \frac{1}{2} \mathbf{l}^\top \mathbb{E}[\mathbf{Q}(t)] \mathbf{l} - \mathbb{E}[\mathbf{c}(t)^\top] \mathbf{l}. \quad (5.7)$$

where  $\mathbf{l}_{\mathcal{L}(i, \nu)} = l_\nu^i$  for  $i \in \mathcal{U}$ ,  $\nu \in \mathcal{V}$  and

$$(\mathbf{c}(t))_{\mathcal{L}(i, \nu)} = \int_{t_0}^t \gamma_s^i \phi^\nu(s, X_{[0, s]}) d\mu_s^{\mathcal{U}, i}$$

$$(\mathbf{Q}(t))_{\mathcal{L}(i, \nu), \mathcal{L}(j, \rho)} = \int_{t_0}^t \gamma_s^i \gamma_s^j \phi^\nu(s, X_{[0, s]}) \phi^\rho(s, X_{[0, s]}) d[\mu^{\mathcal{U}, i}, \mu^{\mathcal{U}, j}]_s$$

for

$$\gamma_s^i = \begin{cases} 1 & \text{if } \pi \text{ is of type I} \\ (\mu_s^{\mathcal{U},i})^{-1} & \text{if } \pi \text{ is of type II} \end{cases}.$$

Moreover, it is a convex quadratic optimization problem.

*Remark 5.12.* Let us highlight how the expression of  $\mathbf{c}$ ,  $\mathbf{Q}$  simplify in the case of a signature portfolio of type I,  $t_0 = 0$ ,  $\mathcal{U} = \{1, \dots, d\}$ ,  $X = \mu$  and  $\mu$  is such that it holds for some  $N_0 > 0$

$$d[\mu^i, \mu^j]_t = \sum_{0 \leq |I| \leq N_0} \alpha_I^{i,j} \langle e_I, \hat{\mu}_t \rangle,$$

where  $\alpha_I^{i,j} \in \mathbb{R}$  for all  $i, j \in \{1, \dots, d\}$ ,  $0 \leq |I| \leq N_0$ . Then we obtain

$$\begin{aligned} (\mathbf{c}(t))_{\mathcal{L}(i,I)} &= \langle e_I \otimes e_i, \hat{\mu}_t \rangle \\ (\mathbf{Q}(t))_{\mathcal{L}(i,I), \mathcal{L}(j,J)} &= \sum_{0 \leq |K| \leq N_0} \alpha_K^{i,j} \langle (e_I \sqcup e_J) \otimes e_K, \hat{\mu}_t \rangle. \end{aligned}$$

More generally, consider a signature portfolio  $\pi(\mu, \check{\mu})$  of type I where  $\check{\mu}_t := (t, \mu_t, [\mu, \mu]_t)$  for all  $t \in [0, T]$  in a vectorized form, i.e. fix a labelling function  $\kappa(\cdot, \cdot)$  such that  $\check{\mu}^{\kappa(i,j)} = [\mu^i, \mu^j]$  for  $1 \leq i, j \leq d$  and set  $t_0 = 0$ ,  $\mathcal{U} = \{1, \dots, d\}$ . Then the elements of  $\mathbf{c}$  and  $\mathbf{Q}$  are linear functions on the signature of  $\check{\mu}$ , namely

$$\begin{aligned} (\mathbf{c}(t))_{\mathcal{L}(i,I)} &= \langle e_I \otimes e_i, \check{\mu}_t \rangle \\ (\mathbf{Q}(t))_{\mathcal{L}(i,I), \mathcal{L}(j,J)} &= \langle (e_I \sqcup e_J) \otimes e_{\kappa(i,j)}, \check{\mu}_t \rangle. \end{aligned}$$

*Remark 5.13.* Note that although we start trading at time  $t_0 \geq 0$  we feed to the feature maps  $\phi^\nu$  the semimartingale  $X$  starting at time 0. The interpretation of this is that we start “observing” the market at time 0 and but only start trading at time  $t_0 \geq 0$ . This time-difference between observing and starting to invest can be important in non-Markovian settings.

Before we prove Theorem 5.11, let us state the following lemma.

**Lemma 5.14.** *For a path-functional portfolio  $\pi(\mu^{\mathcal{U}}, X)$  it holds that*

$$\begin{aligned} \log \left( \frac{W_t^\pi}{W_t^{\mathcal{U}}} \right) &= \sum_{i \in \mathcal{U}} \int_{t_0}^t \gamma_s^i f^i(s, X_{[0,s]}) d\mu_s^{\mathcal{U},i} \\ &\quad - \frac{1}{2} \sum_{i,j \in \mathcal{U}} \int_{t_0}^t \gamma_s^i \gamma_s^j f^i(s, X_{[0,s]}) f^j(s, X_{[0,s]}) d[\mu^{\mathcal{U},i}, \mu^{\mathcal{U},j}]_s, \end{aligned} \tag{5.8}$$

where

$$\gamma_s^i = \begin{cases} 1 & \text{if } \pi \text{ is of type I} \\ (\mu_s^{\mathcal{U},i})^{-1} & \text{if } \pi \text{ is of type II} \end{cases}.$$

We refer the reader to Appendix C for a proof of Lemma 5.14.

*Proof of Theorem 5.11.* To show the equivalence in (5.7), we use Lemma 5.14 and make use of the fact that we are using linear path-functional portfolios.

$$\log \left( \frac{W_t^\pi}{W_t^{\mathcal{U}}} \right) = \sum_{i \in \mathcal{U}} \sum_{\nu \in \mathcal{V}} l_\nu^i \int_{t_0}^t \gamma_s^i \phi^\nu(s, X_{[0,s]}) d\mu_s^{\mathcal{U},i} \tag{5.9}$$

$$- \frac{1}{2} \sum_{i,j \in \mathcal{U}} \sum_{\nu, \rho \in \mathcal{V}} l_\rho^j l_\nu^i \int_{t_0}^t \gamma_s^i \gamma_s^j \phi^\nu(s, X_{[0,s]}) \phi^\rho(s, X_{[0,s]}) d[\mu^{\mathcal{U},i}, \mu^{\mathcal{U},j}]_s \tag{5.10}$$

Hence, we have shown that

$$\log \left( \frac{W_t^\pi}{W_t^U} \right) = \mathbf{c}(t)^\top \mathbf{l} - \frac{1}{2} \mathbf{l}^\top \mathbf{Q}(t) \mathbf{l} \quad (5.11)$$

from which the equivalence in (5.7) follows directly.

To show that the optimization problem given in (5.7) is convex, we show that the matrix  $\mathbf{Q}(t)$  is positive semidefinite. Consider the stochastic process  $Y_t = \sum_{j=1}^d \int_{t_0}^t \gamma_s^j f^j(s, X_{[0,s]}) d\mu_s^{\mathcal{U},j}$  whose quadratic variation  $[Y, Y]_t$  is given by it holds that

$$[Y, Y]_t = \sum_{i=1}^d \sum_{j=1}^d \int_{t_0}^t \gamma_s^i \gamma_s^j f^i(s, X_{[0,s]}) f^j(s, X_{[0,s]}) d[\mu^{\mathcal{U},i}, \mu^{\mathcal{U},j}]_s.$$

On the other hand

$$\mathbf{l}^\top \mathbf{Q}(t) \mathbf{l} = \sum_{i=1}^d \sum_{j=1}^d \int_{t_0}^t \gamma_s^i \gamma_s^j f^i(s, X_{[0,s]}) f^j(s, X_{[0,s]}) d[\mu^{\mathcal{U},i}, \mu^{\mathcal{U},j}]_s.$$

Hence,  $\mathbf{l}^\top \mathbf{Q}(t) \mathbf{l} = [Y, Y]_t \geq 0$  for all  $t \in [t_0, T]$  and for all  $\mathbf{l} \in \mathbb{R}^{(|\mathcal{U}| \cdot |\mathcal{V}|)}$ , proving that  $\mathbf{Q}(t)$  is positive semidefinite for every  $t \in [t_0, T]$ .  $\square$

*Remark 5.15.* If the log-utility maximization problem is finite, it holds that

$$\sup_{\{l_\nu^{(i)}\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}} \mathbb{E} [\log(W_t^\pi)] \Leftrightarrow \sup_{\{l_\nu^{(i)}\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}} \mathbb{E} \left[ \log \left( \frac{W_t^\pi}{W_t^U} \right) \right]. \quad (5.12)$$

This simply follows from the following equivalences

$$\begin{aligned} \sup_{\{l_\nu^{(i)}\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}} \mathbb{E} [\log(W_t^\pi)] &\Leftrightarrow \sup_{\{l_\nu^{(i)}\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}} \mathbb{E} [\log(W_t^\pi) - \log(W_t^U)] \\ &\Leftrightarrow \sup_{\{l_\nu^{(i)}\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}} \mathbb{E} \left[ \log \left( \frac{W_t^\pi}{W_t^U} \right) \right]. \end{aligned}$$

*Remark 5.16.* Similar to Remark 5.9, the optimal vector  $\mathbf{l}^*$  is given by

$$\mathbf{l}^* = 2\mathbb{E}[\mathbf{Q}(t)]^{-1}\mathbb{E}[\mathbf{c}(t)],$$

if  $\mathbb{E}[\mathbf{Q}(t)]$  is almost surely invertible. As mentioned this can never hold for linear path-functional portfolios of type I.

### 5.3 Constraints and Regularization

Given an optimization task involving a linear path-functional portfolio which is convex and quadratic, we can make the following statements about adding constraints or regularization:

**Proposition 5.17.** *Given an optimization problem*

$$\min_{\mathbf{l}} -\mathbf{l}^\top \mathbf{c}(t) + \frac{1}{2} \mathbf{l}^\top \mathbf{Q}(t) \mathbf{l}$$

*which is convex quadratic, it remains convex quadratic under the following modifications.*

1. **Bounds Constraints:**  $|l_\nu^i|^2 \leq b_{\mathcal{L}(i,\nu)}$  for  $b_{\mathcal{L}(i,\nu)} \geq 0$  for all  $i \in \mathcal{U}, \nu \in \mathcal{V}$ .

2. **L2-Regularization:** we can add an  $L^2$ -regularization term of the form  $\gamma \mathbf{l}^T \mathbf{l}$  such that the optimization problem becomes

$$\min_{\mathbf{l}} -\mathbf{l}^T \mathbf{c}(t) + \frac{1}{2} \mathbf{l}^T \mathbf{Q}(t) \mathbf{l} + \gamma \mathbf{l}^T \mathbf{l}. \quad (5.13)$$

Note that, this corresponds to a new matrix  $\tilde{\mathbf{Q}}(t) = \mathbf{Q}(t) + \gamma \mathbb{I}$ , where  $\mathbb{I}$  is the identity matrix.

Moreover, the above modifications can be combined and the optimization task remains convex.

*Proof.* 1.  $|l_\nu^i| \leq b_{\mathcal{L}(i,\nu)}$  for  $b_{\mathcal{L}(i,\nu)} \geq 0$  corresponds to the constraint

$$\mathbf{l}^T \delta_{\mathcal{L}(i,\nu), \mathcal{L}(i,\nu)} \mathbf{l} - b_{\mathcal{L}(i,\nu)}^2 \leq 0,$$

where  $\delta_{\mathcal{L}(i,\nu), \mathcal{L}(i,\nu)}$  is the matrix

$$\delta_{jk} = \begin{cases} 1 & \text{if } j = k = \mathcal{L}(i, \nu) \\ 0 & \text{otherwise} \end{cases},$$

which is obviously positive semi-definite.

2. Since  $\mathbf{Q}(t)$  and  $\mathbb{I}$  are positive semi-definite and the sum of positive semi-definite matrices is positive semi-definite,  $\tilde{\mathbf{Q}}(t)$  is again positive semi-definite.  $\square$

*Remark 5.18* (Long-Only and Leverage Constraints). We can add several forms of short-selling constraints under which the convex quadratic optimization problems remain convex and quadratic for linear path-functional portfolios. However, we would like to emphasise that these constraints are only effective during training and there is no theoretical guarantee that the portfolio-weights respect such constraints during the out-of-sample period. The reason for this is that different dynamics of the underlying process  $X$  may lead to violation of the constraints during the testing period.

1. **Long-Only:**  $\pi^i(t) \geq 0$  for all  $i \in \mathcal{U}$ ;
2. **Bounds for Short-Selling:**  $-a \leq \pi^i(t) \leq b$  for all  $i \in \mathcal{U}$  and some  $a, b \in \mathbb{R}$ ;
3. **Leverage constraints** (see [61]):  $\pi^i(t) \geq 0$  for all  $i \in \mathcal{U}_L \subseteq \mathcal{U}$  i.e., impose long-only constraints on a subset  $\mathcal{U}_L$  of stocks and impose a leverage constraint  $\sum_{i \in \mathcal{U}_L} \pi_t^i \leq b$  for some  $b > 0$ .

**Proposition 5.19** (Long-Only Path-Functional Portfolios). *When choosing  $\tau$  to be long-only, we can formulate the following sufficient conditions for a path-functional portfolio to be long-only which increase in strength in the sense that (iv)  $\Rightarrow$  (iii)  $\Rightarrow$  (ii)  $\Rightarrow$  (i). The following are required to hold  $\mathbb{P}$ -a.s. for all  $t \in [0, T]$ .*

- (i)  $|f^i(t, X_{[0,t]}) - f^{max}(t, X_{[0,t]})| \leq 1$  where  $f^{max}(t, X_{[0,t]}) := \max\{f^1(t, X_{[0,t]}), \dots, f^d(t, X_{[0,t]})\}$  for all  $i \in \{1, \dots, d\}$
- (ii)  $|f^i(t, X_{[0,t]}) - f^j(t, X_{[0,t]})| \leq 1$  for all  $i, j \in \{1, \dots, d\}$
- (iii)  $0 \leq f^i(t, X_{[0,t]}) \leq 1$  for all  $i \in \{1, \dots, d\}$
- (iv)  $f^i(t, X_{[0,t]}) \in [\alpha, \alpha + 1]$ ,  $\alpha \in \mathbb{R}$  for all  $i \in \{1, \dots, d\}$

*Proof.* By definition, we know that for each  $i \in \{1, \dots, d\}$  for path-functional portfolios of type  $I$  it holds:

$$\pi^i(t) \geq 0 \Leftrightarrow \left( f^i(t, X_{[0,t]}) + 1 - \sum_{j=1}^d f^j(t, X_{[0,t]}) \mu_t^j \right) \geq 0 \Leftrightarrow (f^i(t, X_{[0,t]}) - \sum_{j=1}^d f^j(t, X_{[0,t]}) \mu_t^j) \geq -1.$$

One sufficient condition for this to hold is

$$|f^i(t, X_{[0,t]}) - f^{max}(t, X_{[0,t]})| \leq 1$$

where  $f^{max}(t, X_{[0,t]}) := \max\{f^1(t, X_{[0,t]}), \dots, f^d(t, X_{[0,t]})\}$ . That is because

$$\begin{aligned} \left( f^i(t, X_{[0,t]}) - \sum_{j=1}^d f^j(t, X_{[0,t]}) \mu_t^j \right) &\geq f^i(t, X_{[0,t]}) - f^{max}(t, X_{[0,t]}) \sum_{j=1}^d \mu_t^j \\ &= f^i(t, X_{[0,t]}) - f^{max}(t, X_{[0,t]}). \end{aligned}$$

Similarly, for path-functional portfolio of type  $II$  we find:

$$\pi^i(t) \geq 0 \Leftrightarrow f^i(t, X_{[0,t]}) - \tau_t^i \sum_{j=1}^d f^j(t, X_{[0,T]}) \geq -\tau_t^i \geq -1$$

and also

$$\begin{aligned} f^i(t, X_{[0,t]}) - \tau_t^i \sum_{j=1}^d f^j(t, X_{[0,T]}) &\geq f^i(t, X_{[0,t]}) - d \cdot \tau_t^i f^{max}(t, X_{[0,T]}) \\ &\geq f^i(t, X_{[0,t]}) - f^{max}(t, X_{[0,t]}). \end{aligned}$$

□

The previous sufficient conditions are formulated on the level of portfolio controlling functions. How they translate to the optimization parameters  $\{l_\nu^i\}_{\nu \in \mathcal{V}, 1 \leq i \leq d}$  for linear path-functional portfolios depends on the form of the feature maps and on the properties of  $X$ . The following example holds for a particular case of signature-portfolios with  $X = \mu$ .

**Example 5.20.** Let us choose  $X = \mu$ ,  $\tau = \mu$  and a linear path-functional portfolio of either type  $I$  or  $II$  with feature maps

$$\phi^{i_1 \sqcup \dots \sqcup i_m}(t, \mu_{[0,t]}) = \langle e_{i_1 \sqcup \dots \sqcup i_m}, \hat{\mu}_t \rangle = \hat{\mu}_t^{i_1} \dots \hat{\mu}_t^{i_m}$$

for  $0 \leq m \leq N$  and  $i_1, \dots, i_m \in \{1, \dots, d+1\}$ . Hence, by the form of the above the feature maps are simply polynomials and it holds that

$$0 \leq \phi^{i_1 \sqcup \dots \sqcup i_m}(t, \mu_{[0,t]}) \leq 1 \text{ for all } t \in [0, T] \text{ and all } i_1, \dots, i_m \in \{1, \dots, d+1\}, m \leq N$$

assuming w.l.o.g that the time-augmentation  $\varphi(t) \in [0, 1]$  for all  $t \in [0, T]$  (recalling that  $\hat{\mu}_t^1 = \varphi(t)$ ). In this case it follows that

$$\begin{aligned} |f^j(t, X_{[0,t]}) - f^k(t, X_{[0,t]})| &= \left| \sum_{m=0}^N \sum_{1 \leq i_1 \leq \dots \leq i_m \leq d+1} (l_{i_1 \sqcup \dots \sqcup i_m}^{(j)} - l_{i_1 \sqcup \dots \sqcup i_m}^{(k)}) \langle e_{i_1 \sqcup \dots \sqcup i_m}, \mu \rangle_t \right| \\ &\quad (5.14) \end{aligned}$$

$$\begin{aligned} &\leq \sum_{m=0}^N \sum_{1 \leq i_1 \leq \dots \leq i_m \leq d+1} |l_{i_1 \sqcup \dots \sqcup i_m}^{(j)} - l_{i_1 \sqcup \dots \sqcup i_m}^{(k)}| \cdot |\langle e_{i_1 \sqcup \dots \sqcup i_m}, \mu \rangle_t| \\ &\quad (5.15) \end{aligned}$$

$$\begin{aligned} &\leq \sum_{m=0}^N \sum_{1 \leq i_1 \leq \dots \leq i_m \leq d+1} |l_{i_1 \sqcup \dots \sqcup i_m}^{(j)} - l_{i_1 \sqcup \dots \sqcup i_m}^{(k)}| \\ &\quad (5.16) \end{aligned}$$

We denote by  $\tilde{N}$  the number of words of length  $0 \leq m \leq N$  with letters in  $\{1, \dots, d+1\}$ , where the order does not matter, more precisely,

$$\tilde{N} = 1 + \sum_{m=1}^N \binom{d+m}{m}.$$

Then,  $|l_{i_1 \sqcup \dots \sqcup i_N}^{(j)} - l_{i_1 \sqcup \dots \sqcup i_N}^{(k)}| \leq 1/\tilde{N}$  implies that

$$\sum_{N=0}^n \sum_{1 \leq i_1 \leq \dots \leq i_N \leq d} |l_{i_1 \sqcup \dots \sqcup i_N}^{(j)} - l_{i_1 \sqcup \dots \sqcup i_N}^{(k)}| \leq 1. \quad (5.17)$$

And hence the quadratic constraint, preserving the convex quadratic structure of the optimization problem,

$$(l_{i_1 \sqcup \dots \sqcup i_N}^{(j)} - l_{i_1 \sqcup \dots \sqcup i_N}^{(k)})^2 \leq \frac{1}{\tilde{N}^2}$$

is sufficient to ensure that the signature portfolios are long-only. Of course this is a particular case of a signature portfolio based on multivariate polynomials whose coefficients should not be too far away from each other.

## 6 Transaction Costs for Portfolios with Short-Selling and Without Bank Account

When trading in the presence of (proportional) transaction costs, we need to consider re-balancing at discrete times. Let  $t$  be an arbitrary but fixed time at which we re-balance. At time  $t^-$  (just before re-balancing) our portfolio has weights  $\pi_{t^-}$  and a wealth of  $W_{t^-}^\pi$ . We want to re-balance to some fixed target weights  $\pi_t$ . This situation has been studied in [64] for the case of long-only weights. However, we allow for short-selling in our portfolios which makes the problem significantly more involved, as we illustrate below.

As the transaction costs are not paid on the portfolio weights, but on the dollar amounts invested in a stock, we define  $\psi_{t^-}^i, \psi_t^i$  to be the dollar amounts invested in stock  $i$  directly before and after re-balancing. Note, that of course it holds

$$\psi_t^i = \pi_t^i \cdot W_t^\pi$$

and the transaction costs to be paid are

$$TC_t = c \sum_{i=1}^d |\psi_t^i - \psi_{t^-}^i|$$

where  $c$  is the percentage of transaction costs to be paid. The difficulty lies in solving for  $\psi_t^i$ , since the amount which we can invest in stock  $i$  depends on the amount of transaction costs to be paid *and vice-versa*. This difficulty arises from the fact, that we do not have a bank-account. Since we know the portfolio weights at each time, we need to solve for  $W_t^\pi$ . We can use the self-financing identity which requires

$$W_t^\pi = W_{t^-}^\pi - TC_t = W_{t^-}^\pi - c \sum_{i=1}^d |\psi_t^i - \psi_{t^-}^i|. \quad (6.1)$$

Moreover, it is convenient to define  $\alpha(t) := \frac{W_t^\pi}{W_{t^-}^\pi}$  and rewrite (6.1) as

$$1 - \alpha_t = c \sum_{i=1}^d |\alpha_t \pi_t^i - \pi_{t^-}^i|. \quad (6.2)$$

Hence, we reduced the problem to solving for  $\alpha_t$ . Since we consider an arbitrary but fixed re-balancing time  $t$ , let us set  $\alpha := \alpha_t$  in the following. We can make the following statement about solutions for  $\alpha$ :

**Proposition 6.1.** *Let  $\alpha^*$  be a solution to the equation*

$$1 - \alpha = c \sum_{i=1}^d |\alpha \pi_t^i - \pi_{t-}^i|$$

*and denote by  $L(\alpha) = 1 - \alpha$  and  $R(\alpha) = c \sum_{i=1}^d |\alpha \pi_t^i - \pi_{t-}^i|$ . We can make the following statements about  $\alpha^*$ :*

1. *If  $\sum_{i=1}^d |\pi_{t-}^i| < \frac{1}{c}$  a unique solution  $\alpha^* \in [0, 1]$  exists.*
2. *If  $\sum_{i=1}^d |\pi_{t-}^i| \geq \frac{1}{c}$  there may not exist a solution, or at least none for  $\alpha \in [0, 1]$ . If a solution exists,  $\alpha^* \leq 1$  and it may not be unique.*

*Remark 6.2.* If we have long-only portfolios (and if  $c < 1$  i.e. less than 100% transaction costs) only the first case of Proposition 6.1 is relevant. Hence, in that case one always has a unique solution  $\alpha^* \in [0, 1]$ . However, when we allow for short-selling, the second case of Proposition 6.1 becomes important.

We give a proof of Proposition 6.1 in Appendix D and move directly to the interpretation of situation in the second case of Proposition 6.1.

*Remark 6.3* (Interpretation of No or Non-Unique Solutions for  $\alpha$ ). We want to give some interpretation of the three possible outcomes of the second case of Proposition 6.1.

- (a) *no solution:* The trade is infeasible. In order to pay the transaction costs, we have to reduce the dollar amounts we invest. However, reducing the dollar amounts  $\psi_t$  leads to more transaction costs (because the difference to the previous investment amount grows, i.e.  $|\psi_t - \psi_{t-}|$  increases). If the transaction costs grow faster than the money we gain from reducing the dollar amount  $\psi_t$ , the trade is infeasible.
- (b) *no solution in  $[0, 1]$ :* Again, if we need to pay transaction costs without a bank-account we have to free-up the money by selling stocks. Unlike case a) this might be feasible, but the amount of transaction costs to be paid is more than the total wealth of the portfolio. This leads to a solution with negative  $\alpha$ .
- (c) *no unique solution in  $[0, 1]$ :* Here, we are in the case where the transaction costs do not exceed our total wealth. Since reducing the dollar amount leads to an increase in transaction costs, there might be several strategies with respect to dollar amounts invested leading to the desired portfolio weights after transaction costs. For example, one could do a more expensive trade (smaller  $\psi_t$ ) and pay more transaction costs or a less expensive trade (larger  $\psi_t$ ) where one pays less transaction costs.

In practice, we treat the above cases as follows. We consider cases a) and b) to lead to ruin, because both are caused by strategies which are infeasible given our wealth. If one of those cases occurs, we set our wealth to zero and terminate the investment. In case c) we choose the solution with highest  $\alpha$ . This is what any reasonable investor would do: choose the trade with the least transaction costs.

## 7 Numerical Results

In this section we present our numerical results using simulated and real market data.<sup>4</sup>

---

<sup>4</sup>The code corresponding to this section is available at <https://github.com/janka-moeller/Sig-SPT>

## 7.1 Simulated Data

In this subsection, we aim to learn the signature portfolios which maximize  $\mathbb{E}[\log \frac{W_T^\pi}{W_t^\mu}]$  in three financial market models respectively, by using simulations of those market models. This corresponds to the optimization task described in Theorem 5.11. To apply this in a numerical setting, we use the following Monte-Carlo optimization to approximate the expected values:

**Corollary 7.1.** *Let  $\tilde{\Omega}$  be a probability-one set on which  $\mathbf{Q}(t)$  is positive-semidefinite. Consider a Monte-Carlo type optimization of the expected log-relative wealth. Recall that  $V_t^\pi = \frac{W_t^\pi}{W_t^\mu}$ , hence we optimize for*

$$\max_{\pi} \frac{1}{M} \sum_{m=1}^M \log(V_t^\pi)(\omega_m)$$

for  $\omega_1, \dots, \omega_M \in \tilde{\Omega}$ . Again, this can be formulated as a convex quadratic optimization problem, namely

$$\min_{\mathbf{l}} \frac{1}{2} \mathbf{l}^T \hat{\mathbf{Q}}(t) \mathbf{l} - \mathbf{l}^T \hat{\mathbf{c}}(t), \quad (7.1)$$

where  $\hat{\mathbf{Q}}(t) = \frac{1}{M} \sum_{m=1}^M \mathbf{Q}(t)(\omega_m)$  and  $\hat{\mathbf{c}}(t) = \frac{1}{M} \sum_{m=1}^M \mathbf{c}(t)(\omega_m)$ .

*Proof.* We have shown in the proof of Theorem 5.11 that  $\mathbf{Q}(t)$  is almost surely positive semi-definite for all  $t \in [t_0, T]$ , hence such a set  $\tilde{\Omega}$  exists. Since the sum of positive semi-definite matrices is positive semi-definite,  $\hat{\mathbf{Q}}(t)$  is positive semi-definite and the statement follows.  $\square$

Moreover, we will compare the performance of the learned portfolio with the growth-optimal portfolio of the respective markets. This is justified by Lemma 2.34, which states that the growth-optimal portfolio (which is also the numeraire portfolio) solves the relative log-utility optimization problem.

For each market, we simulate  $M_{train}$  trajectories over the time-horizon  $[0, 1]$  of a given market model and consider investments over the whole time-horizon, i.e.  $t_0 = 0, t = 1$ . We train a signature portfolio using the Monte-Carlo type optimization and the formulas for  $\mathbf{Q}(t)(\omega)$  and the vector  $\mathbf{c}(t)(\omega)$  provided in Theorem 5.11, where we use signature portfolios  $\pi(\mu, \hat{\mu})$ , i.e. choose  $X = \mu$  to construct the portfolios. Note, that calculating the matrix  $\mathbf{Q}(t)(\omega)$  and vector  $\mathbf{c}(t)(\omega)$  for each training-sample can be parallelized and computed offline. Having obtained the respective  $\hat{\mathbf{Q}}(t)$  and  $\hat{\mathbf{c}}(t)$ , we use the `gurobipy.model.optimize()`<sup>5</sup> method to solve the convex quadratic optimization problem. We want to emphasize again, that we only need to solve the convex quadratic optimization task once and we never have to update  $\hat{\mathbf{Q}}(t)$  and  $\hat{\mathbf{c}}(t)$ , which is very beneficial for the computational tractability of the optimization task.

After the optimization, we compare the weights and performance of the trained signature portfolio and the theoretical growth-optimal portfolio on  $M_{test}$  out-of-sample trajectories.

More concretely, we consider the following three market models and signature portfolios:

### 1. Correlated Black-Scholes Model:

$$dS_t = \text{diag}(S_t)(adt + \Sigma dB_t),$$

where  $a \in \mathbb{R}^d$  and  $\Sigma \in \mathbb{R}^{d \times d}$ ,  $B_t$  is a  $d$ -dimensional Brownian motion. We here take  $d = 3$  and train a signature portfolio  $\pi(\mu, \hat{\mu})$  of type I of degree three. We simulate 1'000 time-steps.

---

<sup>5</sup>for more information, see [https://www.gurobi.com/documentation/9.1/refman/py\\_python\\_api\\_overview.html#sec:Python](https://www.gurobi.com/documentation/9.1/refman/py_python_api_overview.html#sec:Python).

## 2. Volatility Stabilized Market:

$$\frac{dS_t^i}{S_t^i} = \frac{1+\alpha}{2} \frac{1}{\mu_t^i} dt + \sqrt{\frac{1}{\mu_t^i}} dB_t^i, \quad 1 \leq i \leq d.$$

We choose again  $d = 3$ , so that  $B$  is a three-dimensional Brownian motion. We take  $\alpha = 10$  and train a signature portfolio  $\pi(\mu, \hat{\mu})$  of type  $I$  of degree three. We simulate 10'000 time-steps.

## 3. Signature Market:

$$dS_t = \text{diag}(S_t)(\mathbf{a}_t dt + \Sigma dB_t).$$

We choose again  $d = 3$ , so that  $B$  is a three-dimensional Brownian motion. Moreover,  $(\mathbf{a}_t)_i = \sum_{0 \leq |I| \leq 2} \alpha_I^{(i)} \langle e_I, \hat{\mu} \rangle_t$ ,  $\alpha_I^{(i)} \in \mathbb{R}$  with the restriction that  $\alpha_I^{(i)} = 0$  if  $I = (i_1, i_2)$  and  $i_2 \neq 1$  and  $\Sigma$  is a constant  $3 \times 3$  matrix. Here we train a signature portfolio  $\pi(\mu, \hat{\mu})$  of type  $II$  of degree two and simulate 1'000 time-steps.

We use  $M_{train} = 100'000$  in-sample trajectories for training and  $M_{test} = 100'000$  out-of-sample trajectories for evaluation, i.e. computing the average logarithmic relative wealth of the respective portfolios on the test-samples. Furthermore, we used the bound constraints  $|l_I^i| \leq 10'000$  for all  $I$  and for all  $1 \leq i \leq d$  for each signature portfolio.

*Remark 7.2.* In a volatility stabilized market, signature portfolios  $\pi(\mu, \hat{\mu})$  of type  $I$ , with  $\hat{\mu} = (t, \mu)$  fall into the first case of Remark 5.12, where

$$d [\mu^i, \mu^j]_t = \begin{cases} \mu_t^i (1 - \mu_t^i) dt & \text{for } i = j \\ -\mu_t^i \mu_t^j dt & \text{for } i \neq j. \end{cases}$$

*Remark 7.3.* In Lemma 2.34 we have already established that the growth-optimal portfolio solves the relative log-utility optimization problem. We now want to outline that the growth-optimal portfolio exists *and* that the relative log-utility optimization for signature portfolios is well-posed, i.e. that the expectation of this optimization task is finite for all signature portfolios. Let us denote the growth-optimal portfolio by  $\pi^{(g)}$ , recall its form given by Lemma 2.33 and note that a sufficient condition for both of the above to be satisfied is if  $|\mathbb{E}[\log W_T^{\pi^{(g)}}]| < \infty$ . We will now present the arguments for the existence of  $\pi^{(g)}$  and the well-posedness of the optimization problem for each of the markets.

1. Correlated Black-Scholes Model: It can easily be checked that for the growth-optimal portfolio  $\pi^{(g)}$  it holds that  $|\mathbb{E}[\log W_T^{\pi^{(g)}}]| < \infty$ .
2. Volatility Stabilized Market: Our optimization problem is well-posed for signature portfolios, because volatility stabilized market models are polynomial processes, see [15] and all coefficients of  $\mathbf{c}$  and  $\mathbf{Q}$  are linear functions on the signature of  $\hat{\mu}$ , as established in Remarks 5.12 and 7.2. Moreover, we know that the truncated signature of a polynomial process is again a polynomial process (see [16]) and hence their expectation is finite. Therefore, the expectation of any coefficient of  $\mathbf{c}$  and  $\mathbf{Q}$  is finite and  $|\mathbb{E}[\frac{W^{\pi}}{W^{\mu}}]| < \infty$  holds for any signature portfolio  $\pi$  of type  $I$ .
3. Signature Market: Assuming that a solution exists, the coefficients  $\mathbf{a}_i$  are all uniformly bounded on  $[0, 1]$ . This follows from the fact that  $\mu^i$  are positive and uniformly bounded for all  $i = 1, \dots, d$ , hence so is  $\int_0^T \mu_t^i dt$ . It thereby follows  $|\mathbb{E}[\log W_T^{\pi^{(g)}}]| < \infty$  for the growth-optimal portfolio  $\rho$ .

*Remark 7.4.* Note that the growth-optimal portfolios of the Black-Scholes and volatility stabilized markets are also functionally generated portfolios in the classical sense. Indeed, in the Black-Scholes case the portfolio generating function is

$$G^{(BS)}(\mu_t) = \prod_{i=1}^d (\mu_t^i)^{c_i},$$

with portfolio weights  $\pi_t^{(BS),i} = c_i$ , where

$$c_i = ((\Sigma^\top \Sigma)^{-1} a)_i - \sum_{j=1}^d ((\Sigma^\top \Sigma)^{-1})_{ij} \left( \frac{\sum_{k=1}^d (\Sigma^\top \Sigma)^{-1} a)_k - 1}{\sum_{j,k=1}^d ((\Sigma^\top \Sigma)^{-1})_{jk}} \right).$$

In the volatility stabilized market case the generating function of the growth optimal portfolio is given by

$$G^{(Vol)}(\mu_t) = \prod_{i=1}^d (\mu_t^i)^{\frac{1+\alpha}{2}} \exp \left( \mu_t^i \frac{d}{2} (\alpha - 1) \right),$$

with portfolio weights

$$\pi_t^{(Vol),i} = \frac{1+\alpha}{2} + \mu_t^i \frac{d}{2} (\alpha - 1).$$

Therefore, these two examples are not only examples of signature portfolios approximating the respective growth-optimal portfolios but also examples for signature portfolios approximating functionally generated portfolios.

Figure 1 shows the trained signature portfolios (right) and the growth-optimal portfolios (left) evaluated at one out-of-sample trajectory in each market respectively (rows). As expected from the theoretical results, the signature weights are very similar to the theoretical growth-optimal weights in all markets. We want to emphasize, that the growth-optimal weights were never shown to the signature portfolio during training, but the signature portfolio was trained to maximize the expected logarithmic relative wealth.

Furthermore, we want to highlight, that although the growth-optimal portfolio in the Black-Scholes market has constant weights, this approximation task is far from trivial for a signature portfolio of type I. Because signature portfolios of type I approximate the controlling functions of the growth-optimal portfolio, the approximation task was actually

$$f^{(BS),i}(t, \mu_{[0,t]}) \approx \frac{c_i}{\mu_t^i}.$$

And likewise, in the volatility stabilized market, the approximation task was

$$f^{(Vol),i}(t, \mu_{[0,t]}) \approx \frac{\alpha+1}{2\mu_t^i} + \frac{d}{2}(\alpha - 1).$$

We quantify the performance by comparing the average logarithmic relative wealth over the  $M_{test}$  test samples of the growth-optimal and signature portfolios, where in both cases we compute the average logarithmic relative wealth numerically on the test samples. This explains why the signature portfolio sometimes even leads to higher values. We present these results in Table 1.

## 7.2 Real Market Data

### 7.2.1 Details on the Optimization and Investment Procedure

Before we present the performance of signature portfolios, as well as, JL- and randomized-signature portfolios in real markets, we want to give some details on the optimization and investment procedure.

Mean log-relative wealth of	growth-optimal portfolio	signature portfolio
Black-Scholes market:	9.0115	9.0122
Volatility stabilized market:	8.7619	8.7417
Signature market:	0.4399	0.4398

Table 1: Mean logarithmic relative wealth of the theoretical growth-optimal and signature portfolios, evaluated on 100'000 test-samples, in each market respectively.

**Approximating Expectations by Time-Averages** When working with real market data, we only have one realization available. Therefore, as already discussed in Remark 5.10 and at the beginning of Section 5.2 we replace the expectation by a time-average. More precisely, consider  $[0, \Delta, 2\Delta, \dots, N\Delta = t]$  to be a equally spaced grid of  $[0, t]$  with distance  $\Delta$ . Then, for  $s \in [0, t - \Delta]$  and analogously to Remark 5.10 we consider the following approximations:

$$\begin{aligned} \mathbb{E} \left[ \log \left( \frac{V_{s+\Delta}^\pi}{V_s^\pi} \right) \right] &\approx \frac{1}{N} \sum_{i=1}^N \log \left( \frac{V_{i\Delta}^\pi}{V_{(i-1)\Delta}^\pi} \right) = \frac{1}{N} \log(V_t^\pi) \\ \mathbb{E}[\mathcal{R}_{s,s+\Delta}^{V,\pi}] &\approx \frac{1}{N} \sum_{i=1}^N \mathcal{R}_{(i-1)\Delta,i\Delta}^{V,\pi} =: \bar{\mathcal{R}}^{V,\pi}, \\ \text{Var}[\mathcal{R}_{s,s+\Delta}^{V,\pi}] &\approx \frac{1}{N-1} \sum_{i=1}^N (\mathcal{R}_{(i-1)\Delta,i\Delta}^{V,\pi} - \bar{\mathcal{R}}^{V,\pi})^2. \end{aligned}$$

Recall that sufficient ergodicity/stationarity conditions guaranteeing that these approximations are justified have been discussed in Remark 5.10.

Note however that we cannot expect that stationarity of the relative returns and log-relative returns is satisfied for *every* portfolio in our optimization class. Consider for example signature portfolios of type *I* with  $\tau = (\frac{1}{d}, \dots, \frac{1}{d})^\top$ ,  $\hat{X}_t = (t, X_t)$  and  $\mathcal{U} = \{1, \dots, d\}$ . It is easy to see that the stationarity assumption can not hold for all portfolios in this class, just by considering the following portfolio  $\pi^1$  where the only parameters that are non-zero are chosen to be  $l_{(1)}^d$ . Then

$$\mathbb{E} \left[ \mathcal{R}_{t,t+\Delta}^{V,\pi^1} \right] = \frac{1}{d} \mathbb{E} \left[ t \cdot \frac{\mu_{t+\Delta}^d}{\mu_t^d} \right] + \left( \frac{1}{d} - \frac{t}{d^2} \right) \mathbb{E} \left[ \sum_{j=1}^d \frac{\mu_{t+\Delta}^j}{\mu_t^j} \right] - \frac{t+1}{d}.$$

However, there are of course classes of linear path-functional portfolios where the stationarity assumptions holds for all portfolios in the class. Indeed, the simplest example is the class of Markowitz-type portfolios, i.e. with  $\tau$  being constant and constant portfolio maps. A far-reaching generalization thereof are signature portfolios with *with rolling windows*, as defined in Remark 3.9 and for which stationarity holds under the conditions of Remark 5.10. Computing these portfolios is slightly more expensive, since one needs to compute the increments of the signature. Therefore we did not consider them in our implementations.

Finally, let us emphasize that the optimization using time-averages does make sense in practice, even if the stationarity of returns may not hold for every portfolio. For example consider  $\Delta = 1$  day, then the mean-variance optimization using time-averages can be seen as looking for a portfolio with high average daily returns (in time) but without the daily returns varying too much over time and likewise for the log-relative wealth optimization. Moreover, we argue that the relative returns and log-relative returns of our optimized portfolios do exhibit some "stability" in time, as almost all the learnt portfolios perform very well in the out-of-sample period, as we will present in the following. Note, that all of the above optimization problems remain convex quadratic optimization problems if we replace expectations by time-averages.

**Optimization under Transaction Costs** In practice, we re-balance our portfolio-weights once a day since our market data is also of daily frequency. In this setting, we would like to incorporate transaction costs during optimization.

As explained in detail in Section 6, including transaction costs is not trivial in our setting. In particular, since we have shown that (6.1) does not necessarily have a (unique) solution, we cannot aim for any closed-form. Therefore, incorporating the exact amount of transaction-costs to be paid during the optimization would render the optimization task infeasible. However, we propose the following penalization which preserves the form of a convex quadratic optimization problem. Moreover, our empirical results in Subsection 7.2.3 verify that this penalization is effective in all cases and indeed useful for learning portfolios which perform well even under 5% of transaction costs.

**Corollary 7.5.** *Given a convex quadratic optimization problem of the form*

$$\min_{\mathbf{l}} \frac{1}{2} \mathbf{l}^T \mathbf{Q}(T) \mathbf{l} + \mathbf{c}(T)^T \mathbf{l}$$

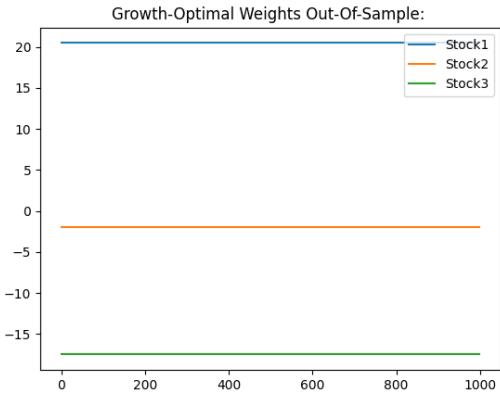
*adding the penalization for linear path-functional portfolios  $\pi$*

$$+ \frac{\beta}{T} \sum_{t=0}^{T-1} \sum_{i \in \mathcal{U}} \left( \frac{\pi_{t+1}^i}{\mu_{t+1}^i} - \frac{\pi_t^i}{\mu_t^i} \right)^2$$

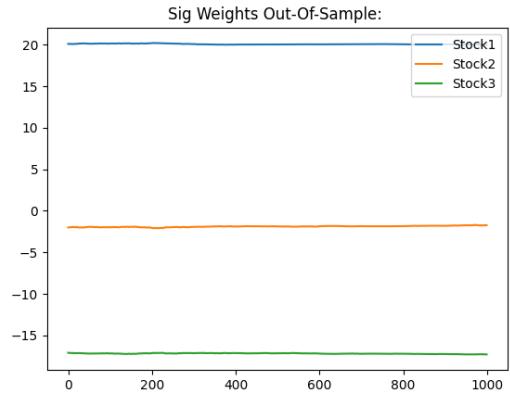
*preserves the form of a convex quadratic optimization problem. Here,  $\beta$  is a hyperparameter that has to be chosen appropriately.*

*Proof.* Clearly the penalization is quadratic in  $\pi$  and hence quadratic in the parameters  $\{l_\nu^i\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}$ . Moreover, the penalization is positive for all  $\{l_\nu^i\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}$  and hence the convexity is preserved as well.  $\square$

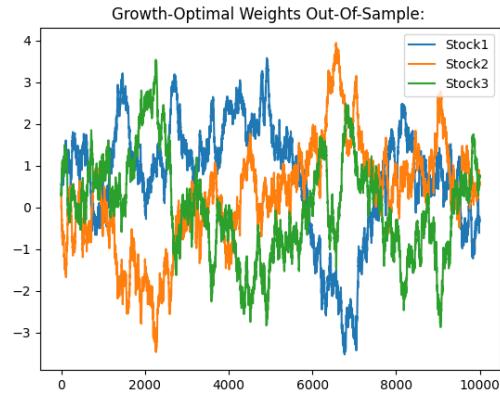
The motivation for this penalization is, first of all, that the universe portfolio  $\mu^{\mathcal{U}}$  is not punished, which should be the case because this portfolio has no transaction costs at all. Moreover, the penalization punishes changes in the weights which exceed changes in the market weights. Note that those are exactly the changes that lead to transaction costs.



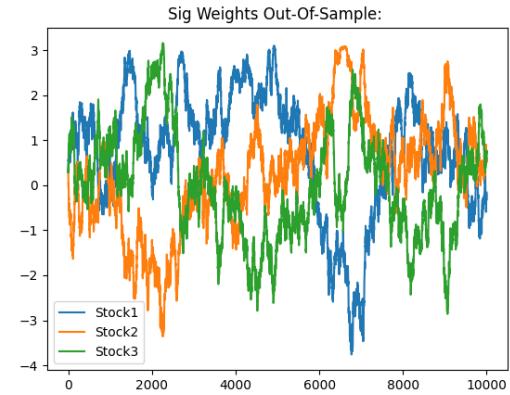
(a) Black-Scholes market: growth-optimal weights



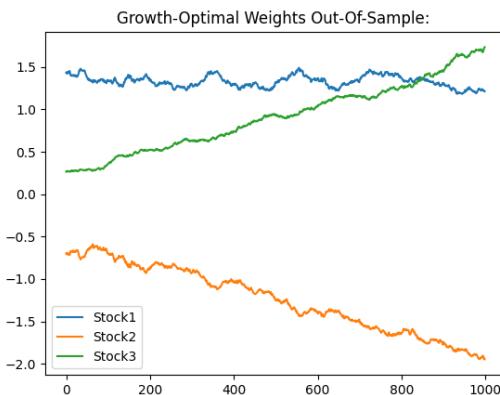
(b) Black-Scholes market: signature weights



(c) Vol. stabilized market: growth-optimal weights



(d) Vol. stabilized Market: signature weights



(e) Signature market: growth-optimal weights



(f) Signature market: signature weights

Figure 1: The theoretical growth-optimal weights (left) and the signature portfolio's weights (right) for the Black-Scholes, volatility stabilized and signature market respectively, evaluated at one test sample.

---

**Algorithm 1** Optimization Under Transaction Costs

---

```

1: function OPTIMIZE_UNDER_TC(beta):
2:    $l \leftarrow \text{SOLVE\_CONVEX\_QUADRATIC}(\beta)$ 
3:   weights  $\leftarrow \text{GET\_INSAMPLE\_WEIGHTS}(l)$ 
4:   wealth  $\leftarrow \text{WEALTH\_UNDER\_TC}(\text{weights})$ 
5:   return  $-1 \cdot \text{wealth}$ 
6: end function
7: while  $(-1 \cdot \text{OPTIMIZE\_UNDER\_TC}(\beta_0)) < 10^{-4}$  do
8:    $\beta_0 += 0.5$ 
9: end while
10: optimal_beta  $\leftarrow \text{MINIMIZE}(\text{OPTIMIZE\_UNDER\_TC}, \beta_0)$ 
11: optimal_l  $\leftarrow \text{SOLVE\_CONVEX\_QUADRATIC}(\text{optimal\_beta})$ 

```

---

The downside of this penalization is, that for a given level of transaction costs, we do not know how to choose  $\beta$ . To find an appropriate  $\beta$ , we propose the procedure described in Algorithm 1. That is, we solve the quadratic optimization problem for a given  $\beta$  and calculate the in-sample performance *under the true transaction costs* for the found optimal portfolio. We then minimize over  $\beta$  in order to find the one which best penalizes for transaction costs over the in-sample period. Note that choosing the initial value  $\beta_0$  can be delicate. Namely, if in a neighbourhood around  $\beta_0$  all optimal strategies lead to ruin under transaction costs, one may not move away from the initial value. Therefore, we test if  $\beta_0$  leads to ruin under transaction cost and, if so, increase it, in Algorithm 1. Note that the condition in Line 8 must be false for some  $\beta_0$  because the universe portfolio itself is included in the set of portfolios we optimize over.

**The Optimization Problems** In the following, we consider optimization problems of various signature-type linear path-functional portfolios with  $L^2$ -regularization, with and without a regularization for transaction costs. The two types of optimization problems are maximizing the expected log-relative wealth

$$\max_{\{l_\nu^{(i)}\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}} \mathbb{E} \left[ \log \left( \frac{W_t^\pi}{W_t^{\mathcal{U}}} \right) \right] - \gamma_{\text{LO}} \sum_{\substack{i \in \mathcal{U} \\ \nu \in \mathcal{V}}} (l_\nu^{(i)})^2 - \frac{\beta}{T} \sum_{t=0}^{T-1} \sum_{i \in \mathcal{U}} \left( \frac{\pi_{t+1}^i}{\mu_{t+1}^{\mathcal{U}, i}} - \frac{\pi_t^i}{\mu_t^{\mathcal{U}, i}} \right)^2 \quad (\text{Log-Opt.})$$

and maximizing mean-variance

$$\max_{\{l_\nu^{(i)}\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}} \lambda \mathbb{E} \left( \mathcal{R}_{t,t+\Delta}^{V,\pi} \right) - \text{Var} \left( \mathcal{R}_{t,t+\Delta}^{V,\pi} \right) - \gamma_{\text{MV}} \sum_{\substack{i \in \mathcal{U} \\ \nu \in \mathcal{V}}} (l_\nu^{(i)})^2 - \frac{\beta}{T} \sum_{t=0}^{T-1} \sum_{i \in \mathcal{U}} \left( \frac{\pi_{t+1}^i}{\mu_{t+1}^{\mathcal{U}, i}} - \frac{\pi_t^i}{\mu_t^{\mathcal{U}, i}} \right)^2 \quad (\text{MV-}\lambda)$$

where in both cases we added bound constraints  $|l_\nu^{(i)}| \leq 10'000$  for all  $i \in \mathcal{U}, \nu \in \mathcal{V}$ .

**Hyperparameters and Cross-Validation** The main hyperparameters of our optimization procedure are:

- $t_0$ : this is the time at which we start to invest, with respect to the time when we start calculating the signature. The interpretation of this is that we can observe the market for some time before we start to invest, which is particularly relevant in the non-Markovian setting. In the following we choose  $t_0 = 100$ , hence, we always start calculating the signature 100 days before the respective investment period starts.
- $\gamma$ : this is the parameter of the  $L^2$ -regularization described in Proposition 5.17. For the case of maximizing the (expected) log-relative wealth, we choose our  $\gamma$  during cross-validation. We do not train this regularization parameter in the mean-variance optimization but fix it apriori to a small value, as the minimization of the variance itself can already be regarded as regularizing.

For performing the cross-validation, we first split our data into a in-sample-period of  $T_{ins}$  days, a cross-validation of  $T_{cv}$  (consecutive to the in-sample period) and a testing period of  $T_{test}$  days, starting at the end of the cross-validation period. Once we found the optimal hyper-parameters during cross-validation, we learn the parameters of our portfolios again on the  $T_{ins}$  days prior to the testing period (using the obtained hyper-parameters). We then evaluate the performance of the learned portfolios during the testing-period.

**Calculating the JL-Signature** In order to calculate the JL-signature, the true truncated signature needs to be calculated before applying the JL-projection. However, calculating the true signature can be computationally too heavy in large markets. We therefore propose the following memory-efficient procedure (Algorithm 2), which is based on the observation that for each component of the signature at level  $l$ , at most  $l$  components of the underlying path appear in the integrals. Hence, one can instead compute the signature of combinations of  $l$  components of the path and apply the projection "batch-wise". It is important to note that by doing so, some words are computed multiple times, for example

$$\int_0^t \int_0^s \circ dX_u^1 \circ dX_s^1$$

can arise from the combination  $(X^1, X^2, X^3)$  and from  $(X^1, X^4, X^5)$  and from many more. Therefore, one needs to be careful which words to keep and our proposed algorithm takes care of that. Moreover, it is important to store the random projection matrix, once a realization is computed.

---

#### Algorithm 2 Calculate JL-Signature

---

```

1: function JL_SIGNATURE(path, level):
2:   Initialize A_list=list()
3:   Initialize JL_sig=0
4:   for  $l \in \{1, \dots, \text{level}\}$  do
5:     combos  $\leftarrow$  COMBNATIONS(dim(path),  $l$ )
6:     for  $c \in \text{combos}$  do
7:       A_slice  $\sim N(0, 1/k)$ 
8:       sig, words  $\leftarrow$  SIGNATURE(path[c], level)
9:       words_to_keep= list()
10:      for word  $\in$  words do
11:        if length(set(word)) ==  $l$  then
12:          words_to_keep.append(word)
13:        end if
14:      end for
15:      sig= sig[words_to_keep]
16:      A_slice= A_slice[ :, words_to_keep]
17:      JL_sig+=matrix_product(A_slice, sig)
18:    end for
19:  end for
20:  return JL_sig
21: end function

```

---

**Parametrization of Time-Components** We parameterize the time-component of the time augmentation in the following way: Let  $\hat{X}_t = (\varphi_T(t), X_t)$ . For a given trading horizon  $T_{hor}$ , we consider the parametrization  $\varphi_{T_{hor}}(t) = \frac{t}{T_{hor}}$ . Concretely, for the in-sample period of 2000 days, we set  $T_{hor} = 2000$  and for the out-of-sample period we set  $T_{hor} = 750$ . This time-augmentation therefore contains information about the amount of time that is left (or has passed) in the current trading period. This is to compensate for the different training and testing periods. Another way to deal with this would be to use signature portfolios with rolling windows as defined in Remark 3.9.

**Performance Metrics** Recall that we re-balance our portfolios once a day and follow a buy-and-hold strategy in between. We evaluate the out-of-sample performance of portfolios by their log-relative wealth

$$\log \left( V_T^{b\&h} \right) = \log \left( \prod_{t=t_0}^{T-1} \sum_{i \in \mathcal{U}} \pi_t^i \frac{\mu_{t+1}^{\mathcal{U},i}}{\mu_t^{\mathcal{U},i}} \right)$$

and their log-wealth

$$\log \left( W_T^{b\&h} \right) = \log \left( \prod_{t=t_0}^{T-1} \sum_{i \in \mathcal{U}} \pi_t^i \frac{S_{t+1}^{\mathcal{U},i}}{S_t^{\mathcal{U},i}} \right).$$

**Survivorship Bias** In our name-based approach we only consider a universe of stocks to invest in which were present in the market for the entire in- and out-of-sample period. This opens up the issue of survivorship bias. Note that the same holds true for the ranked NASDAQ market, since there was always a 100-largest stock present, even though the company holding that rank may have changed over time. As a first measure to reduce this bias, we compare the performance of our trained portfolios only to the universe portfolio of the surviving stocks and not to the entire market portfolio. Nevertheless, our trained portfolios may benefit from the fact that stocks do not go bankrupt in other ways which the universe portfolio can not exploit. Let us address them in more detail:

- **Leverage:** If bankruptcy of stocks is not possible, leverage (i.e. short-selling of some stocks in order to invest more heavily in others) is less risky. We do not impose any leverage constraints in our optimization. Nevertheless we do not observe extreme short-selling during the out-of-sample period. We report the minimum observed out-of-sample weights observed in Tables 2a and 2b, were we highlight those in bold which are long-only. At least for those which do not short-sell we can exclude leveraging as an unfair advantage. In the cases where only small negative weights were observed, we dare to say that an unfair advantage through leveraging is unlikely to have been the driving factor for potential out-performance of the universe portfolio. Aside from leveraging, we would like to point out that regularization for transaction costs reduces short-selling.
- **Over-weighting small stocks:** If stocks' capitalization cannot go to zero, a potential strategy may be to put a lot of weight on stocks with capitalizations close to zero, because the upside is much higher than the downside. To investigate this, we point to Figures 3c and 3d, which show the average out-of-sample weights of the ranked NASDAQ universe. We observe that the trained portfolios *underweight* small stocks on average compared to the universe portfolio.

**Choice of a Benchmark** In the following numerical experiments we will use the universe weights as a auxiliary portfolio  $\tau$ . Recall from Remark 3.7 that the auxiliary portfolio  $\tau$  is the natural benchmark for linear path-functional portfolios because it can always be attained by them. As an additional benchmark we included the equally-weighted portfolio. We want to emphasise that of course other auxiliary portfolios and hence benchmark portfolios could be used.

### 7.2.2 Rank-based approach: NASDAQ

In this part of our empirical analysis we study a rank-based approach to portfolio optimization. We consider the *ranked* NASDAQ market and choose as a universe the stocks with ranks  $1, \dots, 100$ . We obtained the data from the CRSP database.<sup>6</sup> We train JL-signature portfolios

---

<sup>6</sup>The raw/processed data required to reproduce our findings cannot be shared at this time due to legal reasons.

Portfolio Reg. TC	Signature Portfolio			JL-Signature Portfolio			Rand.-Signature Portfolio		
	0%	1%	5%	0%	1%	5%	0%	1%	5%
Log-Opt.	-0.5	-0.45	-0.13	-0.32	-0.32	-0.12	-0.01	-0.01	<b><math>3 \cdot 10^{-3}</math></b>
MV $\lambda = 1$	-3.31	-1.45	-0.16	-3.70	-1.67	-0.18	-3.37	-2.26	-0.20
MV $\lambda = 0.75$	-2.43	-1.72	-0.16	-2.72	-2.10	-0.18	-2.48	-2.22	-0.21
MV $\lambda = 0.5$	-1.56	-1.54	-0.17	-1.75	-1.75	-0.19	-1.59	-1.54	-0.22
MV $\lambda = 0.05$	-0.05	-0.05	-0.04	-0.05	-0.05	-0.04	-0.04	-0.04	-0.03

(a) Minimum portfolio weight observed during the out-of-sample period for the SMI universe for the respective portfolios with and without regularization for transaction costs.

Portfolio Reg. TC	JL-Signature Portfolio			Rand.-Signature Portfolio		
	0%	1%	5%	0%	1%	5%
Log-Opt.	-1.67	-0.66	-0.11	-0.80	-0.63	-0.15
MV $\lambda = 1$	-0.14	-0.14	-0.11	-0.04	-0.04	-0.04
MV $\lambda = 0.75$	-0.10	-0.10	-0.10	-0.03	-0.03	-0.03
MV $\lambda = 0.5$	-0.06	-0.06	-0.06	-0.01	-0.01	-0.01
MV $\lambda = 0.05$	<b><math>8 \cdot 10^{-5}</math></b>					

(b) Minimum portfolio weight observed during the out-of-sample period for the S&P500 universe for the respective portfolios with and without regularization for transaction costs.

Table 2: The tables show the minimum portfolio weights observed during the out-of-sample period of the respective universe. The ones which are positive (i.e. no short-selling) are highlighted in bold.

of dimension  $(50, 3)$  of type  $I$  and randomized signature portfolios of dimension 50 of type  $I$  investing in the universe of the 100 largest stocks and use as input the universe weights i.e.  $\hat{X} = \hat{\mu}^U$  and the universe portfolio as the auxiliary portfolio (i.e.  $\tau = \mu^U$ ). The portfolio is daily re-balanced. We train such portfolios in two ways, once in the log-wealth optimization and once in the mean-variance optimization, where in both optimization problems, we measure the performance of the portfolios by the log-relative wealth and the log-wealth achieved in the out-of-sample period.

We choose  $t_0 = 100$  and take as an in-sample period 2000 trading days and as an out-of-sample period the following 750 trading days. For the mean-variance optimization task we set the  $L^2$ -regularization parameter  $\gamma_{MV} = 10^{-6}$  and for the log-wealth optimization task we found  $\gamma_{LO} = 4.849 \cdot 10^{-3}$  for JL-signature portfolios and  $\gamma_{LO} = 1 \cdot 10^{-2}$  for randomize-signature portfolios during the cross-validation. The grid-search for  $\gamma_{LO}$  was performed over 100 equally-distant points in  $[10^{-6}, 10^{-2}]$ .

We report the out-of-sample performance of the trained portfolios in Table 3 in terms of log-(relative)-wealth with respect to the universe portfolio. All but one of the trained portfolios outperform the universe portfolio.

In Figure 2a and Figure 2b we display the wealth processes of the trained portfolio, the one of the universe portfolio and as a benchmark also of the equally-weighted portfolio. The wealth processes of the mean-variance portfolio are the more volatile, the higher the risk-tolerance is, as one would expect. Moreover, the wealth-processes of the randomized signature portfolios are more tamed than their JL-counterparts.

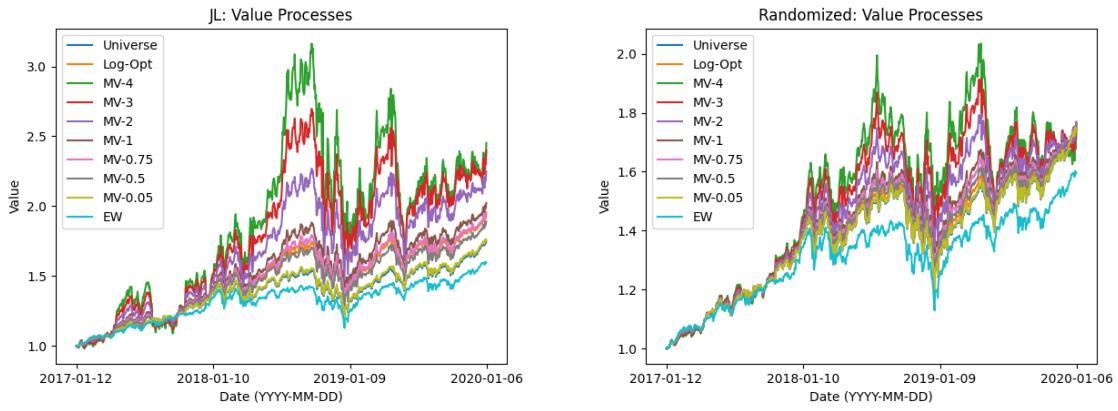
We present the average values (in time) of the portfolio weights for each rank in Figure 3. The trained portfolios mainly take over- and under-weighted positions in the largest stocks. While the positions are extremer, the higher the risk-tolerance. Although we do not enforce any long-only constraints or regularization, we do not observe any extreme short-selling positions.

### 7.2.3 Name-based approach: SMI and S&P500

In the name-based setting we tackle the mean-variance and log-relative wealth optimization problems under transaction costs. We do this in two markets; the Swiss Market Index (SMI)

Optimization Task	JL-Signature $\log V_T^{b\&h}$	$\log W_T^{b\&h}$	Randomized Signature $\log V_T^{b\&h}$	$\log W_T^{b\&h}$
Log-Opt.	0.1045	0.6622	0.0002	0.5579
MV $\lambda = 4$	0.3395	0.8972	-0.0045	0.5531
MV $\lambda = 3$	0.3173	0.8750	0.0081	0.5658
MV $\lambda = 2$	0.2531	0.8108	0.0131	0.5707
MV $\lambda = 1$	0.1472	0.7049	0.0104	0.5680
MV $\lambda = 0.75$	0.1143	0.6720	0.0085	0.5662
MV $\lambda = 0.5$	0.0788	0.6364	0.0061	0.5638
MV $\lambda = 0.05$	0.0083	0.5660	0.0007	0.5584

Table 3: Performance of the trained portfolios over the out-of-sample period in terms of log-(relative)-wealth. We ran the mean-variance optimization task for seven different risk-tolerances  $\lambda$ . All but one of the optimized portfolios out-performed the universe portfolio over the out-of-sample period.



(a) Wealth processes of the learned JL-signature portfolios. (b) Wealth processes of the randomized signature portfolios.

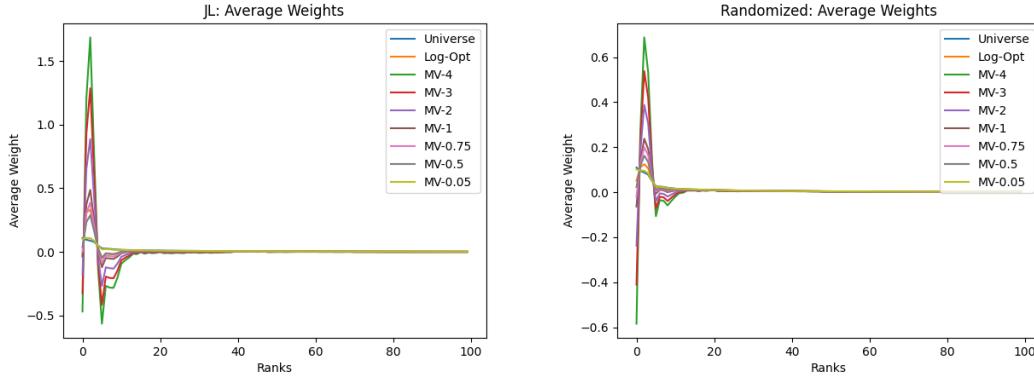
Figure 2: Wealth processes of NASDAQ the learned signature portfolios, the universe portfolio and the equally-weighted portfolio, over the out-of-sample period.

and S&P500 Index. For the SMI, we consider the universe of stocks which survived between 2000-2022 and for the S&P500 those that survived between 2001-2022. This amounts to 17 stock in the SMI universe and to 378 stocks in the S&P500 universe. For both markets, we obtained the data from Reuters Datastream.<sup>7</sup> As inputs and as auxiliary portfolio we use the respective universe weights, i.e.  $X = \tau = \mu^{\mathcal{U}}$ . We compare the performance of the trained portfolios with the universe portfolio. Again, we choose  $t_0 = 100$ , an in-sample period of 2000 days and an out-of-sample period of 750 days. When we include the regularization cost, we choose  $\beta_0 = 0.5$  and enforce a lower bound  $\beta \geq 10^{-8}$ . We want to emphasize, that even with transaction costs our portfolios are re-balanced daily.

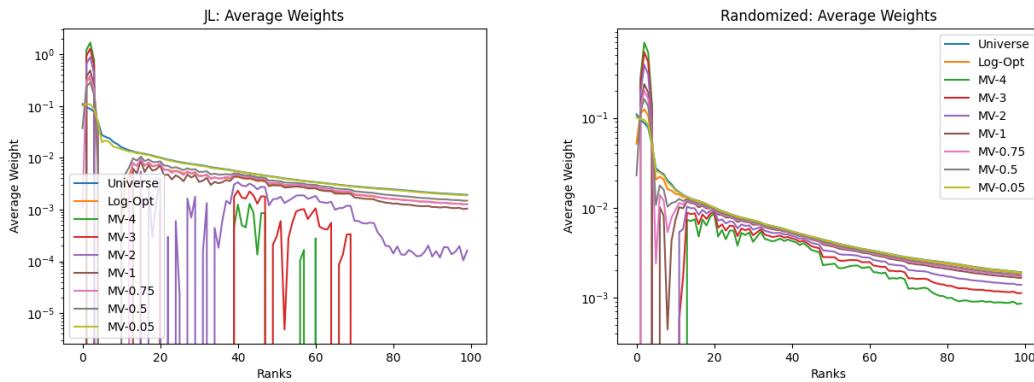
**SMI Market** We train signature portfolios  $\pi(\mu^{\mathcal{U}}, \hat{\mu}^{\mathcal{U}})$  of type *I* with three configurations of feature maps

- the true signature up to degree two;
- the JL-signature of dimension (30,2);
- the randomized signature of dimension 30.

<sup>7</sup>The raw/processed data required to reproduce our findings cannot be shared at this time due to legal reasons.



(a) Average weights of the JL-signature portfolios. (b) Average weights of the randomized signature portfolios.



(c) Plot of the logarithm of the average weights of the JL-signature portfolios. (d) Plot of the logarithm of the average weights of the randomized-signature portfolios.

Figure 3: Average weights in the NASDAQ of the signature portfolios and the universe portfolio, averaged over the out-of-sample period.

We set  $\gamma_{MV} = 10^{-8}$  and found  $\gamma_{LO} = 2.02 \cdot 10^{-3}$  for the true signature,  $\gamma_{LO} = 2.42 \cdot 10^{-3}$  for the JL-signature and  $\gamma_{LO} = 1 \cdot 10^{-2}$  for the randomized signature respectively during cross-validation, where the grid-search was performed over 100 equally-distant points in  $[10^{-8}, 10^{-2}]$ . We present the out-of-sample performance in terms of log-relative wealth in Tables 4a, 4b, 4c as well as the annualized Sharpe-ratios in Table 5a. The first three columns of Tables 4a, 4b, 4c show the performance of portfolios with and without transaction costs, where we added no regularization for transaction costs. The next four columns show the performance with and without transactions costs, but with a regularization for transaction costs at the respective level. The corresponding regularization parameters  $\beta$  which we found during the in-sample training are shown in Table 6. We note that the lower the risk-tolerance  $\lambda$  the less regularization for transaction costs is needed. However, adding such a regularization, the signature portfolios trained under the mean-variance optimization often outperform the universe portfolio under transaction costs during the out-of-sample period. This is remarkable, since the universe portfolio does not pay any transaction costs. The portfolio trained under log-relative wealth optimization does not out-perform the universe portfolio under transaction costs for the true and JL-signature portfolios and neither do some of the portfolios trained using the JL- and randomized signature under mean-variance optimization with higher risk-tolerances. Nevertheless, the regularization for transaction cost proved effective, since all portfolios performed better with the regularization than without, under the respective level of transaction costs.

We show the wealth processes of the signature portfolios for the SMI universe in Figure 4. It

is obvious that the higher  $\lambda$  is, the more volatile the portfolios are, which lead to some signature portfolios under-performing the universe portfolio under transaction costs during some parts of the out-of-sample period. The signature portfolio with  $\lambda = 0.05$ , however, did rather well over the entire period even under transaction costs, which we highlight in Figures 4d & 4e.

Turning to the Sharpe-ratios in Table 5a we see that while the Sharpe-ratios of the optimized portfolios are often higher than the one of the universe portfolio without transaction costs, the universe portfolio is less often out-performed under transaction costs in terms of Sharpe-ratios. Only the mean-variance optimized portfolio with  $\lambda = 0.05$  achieves a consistently higher Sharpe ratio than both universe and equally-weighted portfolio. It is also worth highlighting, that often the Sharpe-ratio increases for higher transaction costs (and thereby higher regularization for transaction costs).

**S&P500 Market** We trained JL-signature portfolios  $\pi(\mu^{\mathcal{U}}, \hat{\mu}^{\mathcal{U}})$  of dimension (30,2) and type  $I$  as well as randomized-signature portfolios  $\pi(\mu^{\mathcal{U}}, \hat{\mu}^{\mathcal{U}})$  of dimension 30 and type  $I$  in the S&P500 market. For the mean-variance optimization, we set  $\gamma_{MV} = 10^{-6}$  and for obtained  $\gamma_{LO} = 1.02 \cdot 10^{-4}$  for both the JL-signature portfolio and the randomized signature portfolio during cross-validation over an equally-spaced grid of 100 points in  $[10^{-6}, 10^{-2}]$ . We report the out-of-sample results of the optimized portfolios in Table 4d & 4e for the JL- and randomized-signature portfolios respectively. It is remarkable that many portfolios also out-performed the universe portfolio without any regularization for transaction costs. Indeed, in many cases the optimization yielded  $\beta = 10^{-8}$ , as we report in Table 6. In the cases where such a regularization was needed, the performance with transaction costs was always improved with the regularization. Apart from the log-wealth optimized portfolio under 5% of transaction costs, the learned portfolio all out-performed the universe portfolio under transaction costs. We show the corresponding wealth processes in Figure 6.

The out-of-sample annualized Sharpe-ratios are reported in Table 5b. The mean-variance optimized portfolios manage to out-perform the universe and equally-weighted portfolios also in terms of Sharpe-ratios, even under transaction costs. Only the log-wealth optimized portfolio fails to achieve a higher Sharpe-ratio than the universe portfolio under transaction costs.

**Conclusion** We would like to offer some conclusion on the performance of our optimized portfolios in particular comparing the signature, JL-signature and randomized signature portfolios. We highlight that computing the true signature portfolio of the S&P500 universe would not have been feasible, since the number of stocks would have been too high. However, the JL-signature portfolios did consistently achieve a better out-of-sample performance both in terms of log-relative wealth and Sharpe-ratios than the randomized signature portfolios. For the SMI universe we are impressed with the performance of the JL- and randomized-signature portfolios compared to those of the true signature portfolio, especially due to the vast difference in number of optimization parameters (510 for JL- & rand.-signature each vs. 10'290 for the true signature portfolios!). In terms of Sharpe-ratio the JL- and randomized-signature portfolios even slightly out-performed the true signature portfolios, but no significant difference between the two randomization approaches was observable. In terms of log-relative wealth, the true signature portfolio performed better than the JL-signature portfolios, which in turn achieved better results than the randomized-signature ones, however we would like to point out that the differences were small. Overall, in terms of feasibility and performance, we consider the JL-signature portfolios to be the most favorable.

Opt. Task Prop. TC	$\log V_T$ without Reg.			$\log V_T$ with Reg. $\beta_{1\%}$		$\log V_T$ with Reg. $\beta_{5\%}$	
	0%	1%	5%	0%	1%	0%	5%
Log-Opt.	0.0458	-0.0494	-0.4333	0.0461	-0.0293	0.0540	-0.0158
MV $\lambda = 1$	1.7896	-1.9354	$-\infty$	1.0514	0.1760	0.0257	0.2804
MV $\lambda = 0.75$	1.4767	-0.6214	-14.6792	1.1547	0.0601	0.2724	0.0247
MV $\lambda = 0.5$	1.0748	0.1155	-4.3536	1.0539	0.1409	0.2620	0.0220
MV $\lambda = 0.05$	0.1240	0.0973	-0.0096	0.1229	0.0979	0.1056	0.0381

(a) Log-relative wealth of the signature portfolios with and without transaction costs trading in the SMI universe.

Opt. Task Prop. TC	$\log V_T$ without Reg.			$\log V_T$ with Reg. $\beta_{1\%}$		$\log V_T$ with Reg. $\beta_{5\%}$	
	0%	1%	5%	0%	1%	0%	5%
Log-Opt.	0.0620	0.0093	-0.2009	0.0619	0.0094	0.0420	-0.0201
MV $\lambda = 1$	1.6656	-1.6336	$-\infty$	0.9654	0.1402	0.2061	0.0187
MV $\lambda = 0.75$	1.3676	-0.4919	-12.2094	1.1439	-0.0583	0.2051	0.0178
MV $\lambda = 0.5$	0.9913	0.1398	-3.7521	0.9913	0.1398	0.2074	0.0144
MV $\lambda = 0.05$	0.1137	0.0894	-0.0072	0.1137	0.0894	0.0947	0.0254

(b) Log-relative wealth of the JL-signature portfolios with and without transaction costs trading in the SMI universe.

Opt. Task Prop. TC	$\log V_T$ without Reg.			$\log V_T$ with Reg. $\beta_{1\%}$		$\log V_T$ with Reg. $\beta_{5\%}$	
	0%	1%	5%	0%	1%	0%	5%
Log-Opt.	0.0397	0.0331	0.0070	0.0396	0.0331	0.0217	0.0050
MV $\lambda = 1$	1.4602	-1.3840	$-\infty$	0.9378	-0.0191	0.1334	-0.0005
MV $\lambda = 0.75$	1.1976	-0.4161	-11.9226	1.0101	-0.1188	0.1381	0.0004
MV $\lambda = 0.5$	0.8668	0.1176	-3.3104	0.8157	0.1504	0.1498	0.0006
MV $\lambda = 0.05$	0.0991	0.0752	-0.0195	0.0970	0.0749	0.0659	0.0177

(c) Log-relative wealth of the randomized-signature portfolios with and without transaction costs trading in the SMI universe.

Opt. Task Prop. TC	$\log V_T$ without Reg.			$\log V_T$ with Reg. $\beta_{1\%}$		$\log V_T$ with Reg. $\beta_{5\%}$	
	0%	1%	5%	0%	1%	0%	5%
Log-Opt.	1.7442	-2.5700	$-\infty$	0.9282	0.0402	0.2287	-0.1564
MV $\lambda = 1$	0.3089	0.2657	0.0944	0.3089	0.2657	0.2292	0.0987
MV $\lambda = 0.75$	0.2342	0.2092	0.1101	0.2342	0.2092	0.2179	0.1078
MV $\lambda = 0.5$	0.1578	0.1460	0.0990	0.1578	0.1460	0.1578	0.0990
MV $\lambda = 0.05$	0.0161	0.0156	0.0137	0.0161	0.0156	0.0161	0.0137

(d) Log-relative wealth of the JL-signature portfolios with and without transaction costs trading in the S&P500 universe.

Opt. Task Prop. TC	$\log V_T$ without Reg.			$\log V_T$ with Reg. $\beta_{1\%}$		$\log V_T$ with Reg. $\beta_{5\%}$	
	0%	1%	5%	0%	1%	0%	5%
Log-Opt.	1.0296	0.1415	-3.9870	0.7823	0.2411	0.1789	-0.0820
MV $\lambda = 1$	0.0988	0.0903	0.0566	0.0988	0.0903	0.0988	0.0566
MV $\lambda = 0.75$	0.0743	0.0686	0.0455	0.0743	0.0686	0.0743	0.0455
MV $\lambda = 0.5$	0.0497	0.0462	0.0319	0.0497	0.0462	0.0497	0.0319
MV $\lambda = 0.05$	0.0050	0.0047	0.0033	0.0050	0.0047	0.0050	0.0033

(e) Log-relative wealth of the randomized-signature portfolios with and without transaction costs trading in the S&P500 universe.

Table 4: The first three columns show the performance with and without transaction costs of signature portfolios with no regularization for transaction costs. The next four column show the performance with and without transaction costs of portfolios trained with a regularization for transaction costs at the respective level.

Portfolio Prop. TC	Signature Portfolio			JL-Signature Portfolio			Rand.-Signature Portfolio		
	0%	1%	5%	0%	1%	5%	0%	1%	5%
Universe	0.6061	0.6061	0.6061	0.6061	0.6061	0.6061	0.6061	0.6061	0.6061
Equally-weighted	<b>0.7487</b>	<b>0.6287</b>	0.1482	<b>0.7487</b>	<b>0.6287</b>	0.1482	<b>0.7487</b>	<b>0.6287</b>	0.1482
Log-Opt.	0.5786	0.4353	0.55	<b>0.6735</b>	0.5459	0.5389	<b>0.7246</b>	<b>0.7052</b>	<b>0.6239</b>
MV $\lambda = 1$	<b>1.282</b>	0.5191	0.5998	<b>1.263</b>	0.4941	<b>0.6144</b>	<b>1.2029</b>	0.3374	0.5805
MV $\lambda = 0.75$	<b>1.3051</b>	0.4052	0.6004	<b>1.2873</b>	0.3081	<b>0.6125</b>	<b>1.2237</b>	0.2585	0.5812
MV $\lambda = 0.5$	<b>1.3396</b>	0.4849	0.5979	<b>1.3226</b>	0.4906	0.6016	<b>1.2513</b>	0.5202	0.5773
MV $\lambda = 0.05$	<b>0.946</b>	<b>0.8718</b>	<b>0.7073</b>	<b>0.9211</b>	<b>0.8515</b>	<b>0.6726</b>	<b>0.8707</b>	<b>0.8028</b>	<b>0.6499</b>

(a) Annualized Sharpe-ratios of the signature portfolios, universe portfolio and equally weighted portfolio with and without transaction costs trading in the SMI universe. Note that for the results with transaction costs, the corresponding regularization for transaction costs was included in the training.

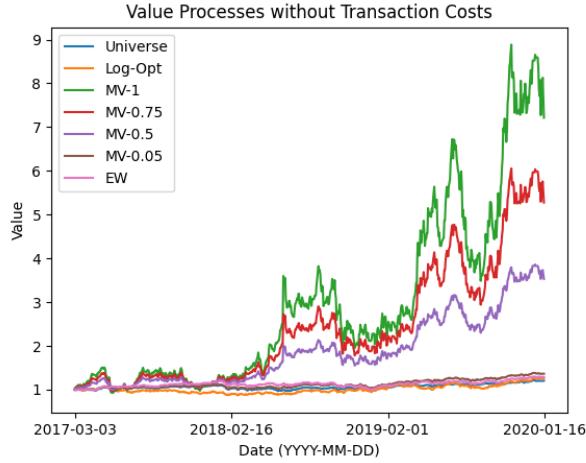
Portfolio Prop. TC	JL-Signature Portfolio			Rand.-Signature Portfolio		
	0%	1%	5%	0%	1%	5%
Universe	0.8530	0.8530	0.8530	0.8530	0.8530	0.8530
Equally-weighted	<b>0.9292</b>	0.7466	0.0164	<b>0.9292</b>	0.7466	0.0164
Log-Opt.	<b>1.223</b>	0.4901	0.3653	<b>1.1904</b>	0.6777	0.5330
MV $\lambda = 1$	<b>1.1817</b>	<b>1.1040</b>	<b>0.8753</b>	<b>1.0077</b>	<b>0.9879</b>	<b>0.9087</b>
MV $\lambda = 0.75$	<b>1.1402</b>	<b>1.0905</b>	<b>0.9048</b>	<b>0.9754</b>	<b>0.9614</b>	<b>0.9054</b>
MV $\lambda = 0.5$	<b>1.0781</b>	<b>1.0521</b>	<b>0.9488</b>	<b>0.9391</b>	<b>0.9302</b>	<b>0.8945</b>
MV $\lambda = 0.05$	<b>0.8834</b>	<b>0.8821</b>	<b>0.8772</b>	<b>0.8625</b>	<b>0.8616</b>	<b>0.8581</b>

(b) Annualized Sharpe-ratios of the signature portfolios, universe portfolio and equally weighted portfolio with and without transaction costs trading in the S&P500 universe. Note that for the results with transaction costs, the corresponding regularization for transaction costs was included in the training.

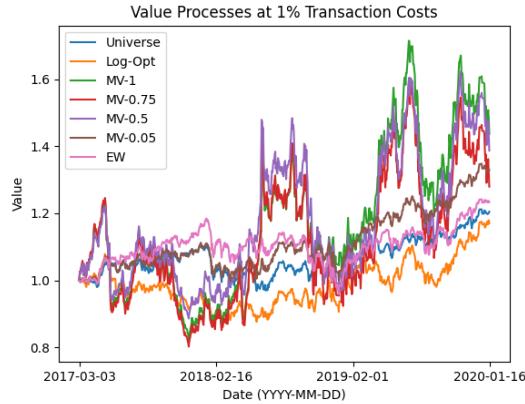
Table 5: The tables show the annualized out-of-sample Sharpe-ratios of the trained signature portfolios as well as universe and equally-weighted portfolio in the SMI and SPX universe respectively. The numbers marked in bold are those that are higher than the Sharpe-ratio of the respective universe portfolio.

Universe Beta	SMI		SMI (JL)		SMI (rand.)		S&P500 (JL)		S&P500 (rand.)	
	$\beta_{1\%}$	$\beta_{5\%}$	$\beta_{1\%}$	$\beta_{5\%}$	$\beta_{1\%}$	$\beta_{5\%}$	$\beta_{1\%}$	$\beta_{5\%}$	$\beta_{1\%}$	$\beta_{5\%}$
Log-Opt.	27.958	847.61	0.5	619.60	0.4999	960.77	9.6437	97.160	2.9908	65.285
MV $\lambda = 1$	0.0879	1.1661	0.0609	0.9611	0.0193	0.5510	$10^{-8}$	0.0059	$10^{-8}$	$10^{-8}$
MV $\lambda = 0.75$	0.0240	0.8063	0.0101	0.6598	0.0040	0.3692	$10^{-8}$	0.0012	$10^{-8}$	$10^{-8}$
MV $\lambda = 0.5$	0.0008	0.4537	$10^{-8}$	0.3645	0.0010	0.1954	$10^{-8}$	$10^{-8}$	$10^{-8}$	$10^{-8}$
MV $\lambda = 0.05$	0.0003	0.0095	$10^{-8}$	0.0076	0.0003	0.0111	$10^{-8}$	$10^{-8}$	$10^{-8}$	$10^{-8}$

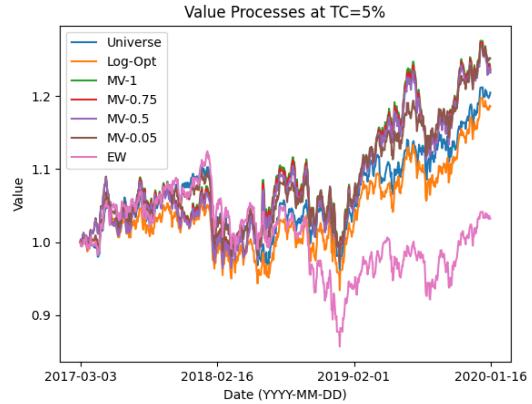
Table 6: Optimal regularization parameters  $\beta$  found during in-sample training.



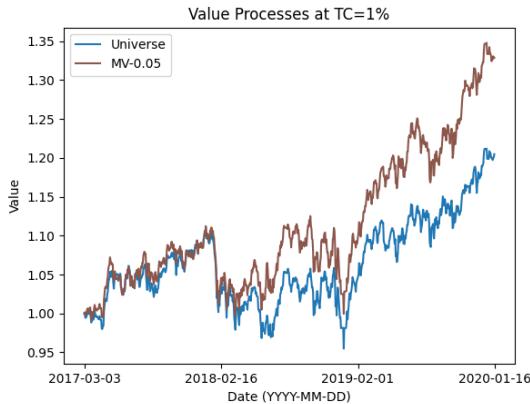
(a) Wealth process without transaction costs



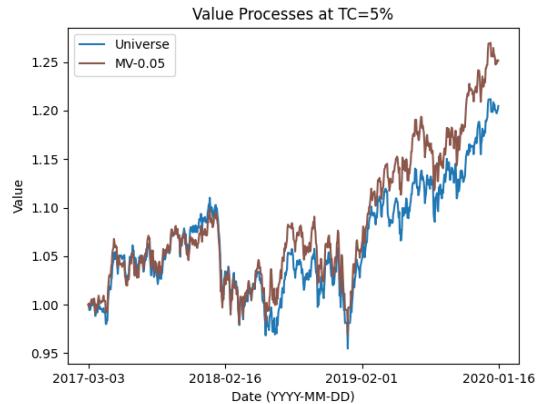
(b) Wealth process with 1% proportional transac- (c) Wealth process with 5% proportional transac-



tion costs

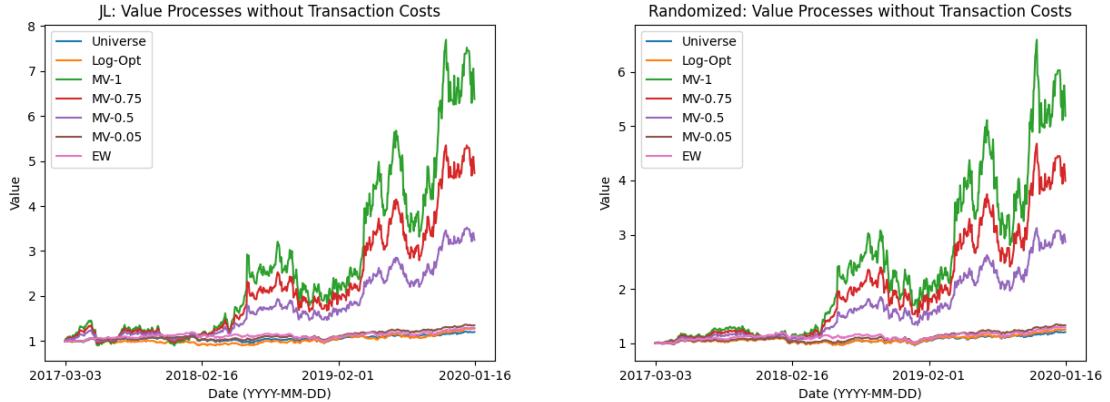


(d) Wealth process with 1% proportional transac- (e) Wealth Process with 5% Proportional Transac-

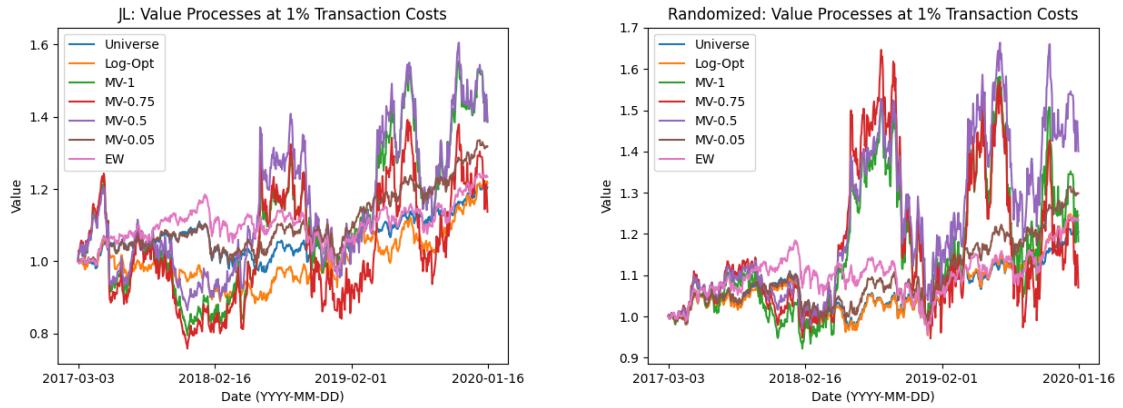


tion Costs

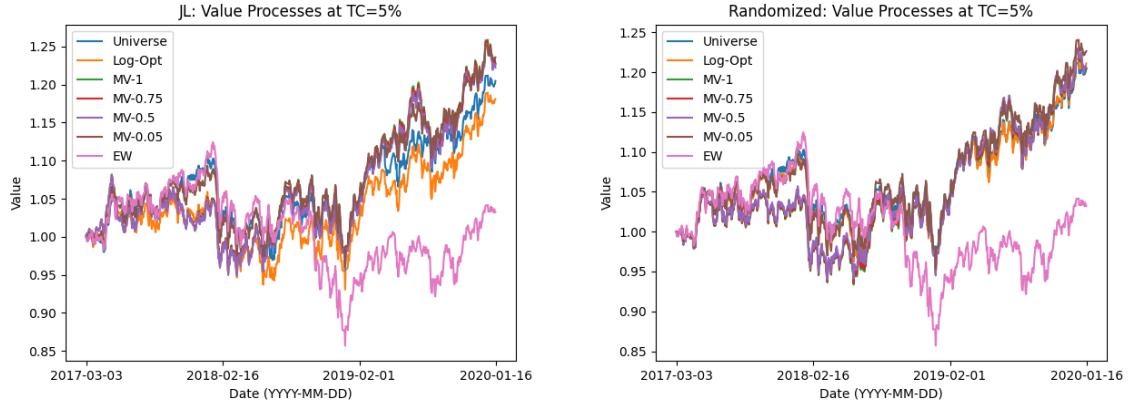
Figure 4: SMI wealth processes of the signature portfolios, the universe portfolio and the equally-weighted portfolio with and without transaction costs. Note that signature portfolios in the settings with transaction cost are those, where we included the regularization for transaction costs.



(a) Wealth process without transaction costs for JL-signature portfolios (left) and randomized-signature portfolios (right)

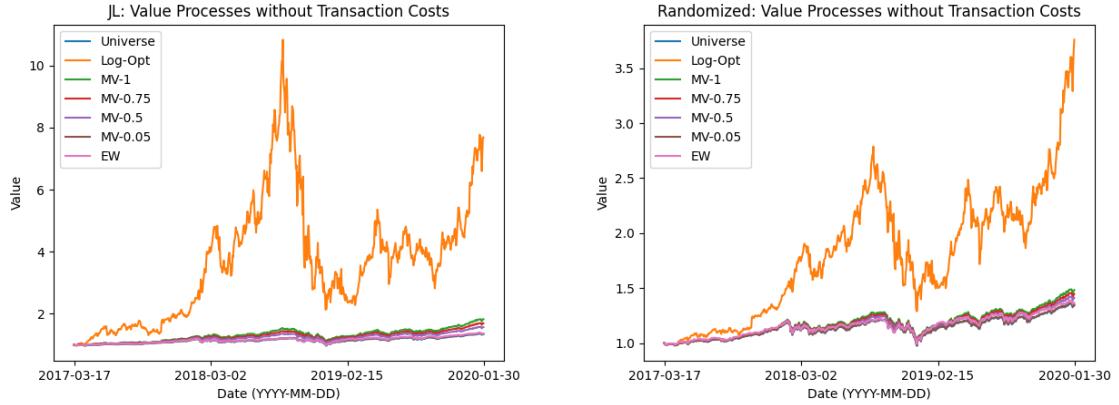


(b) Wealth process with 1% proportional transaction costs for JL-signature portfolios (left) and randomized-signature portfolios (right)

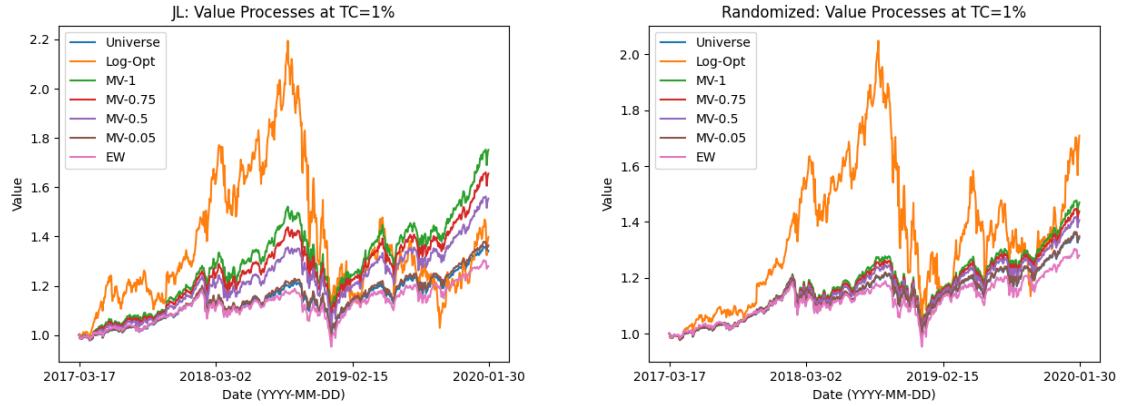


(c) Wealth process with 5% proportional transaction costs for JL-signature portfolios (left) and randomized-signature portfolios (right)

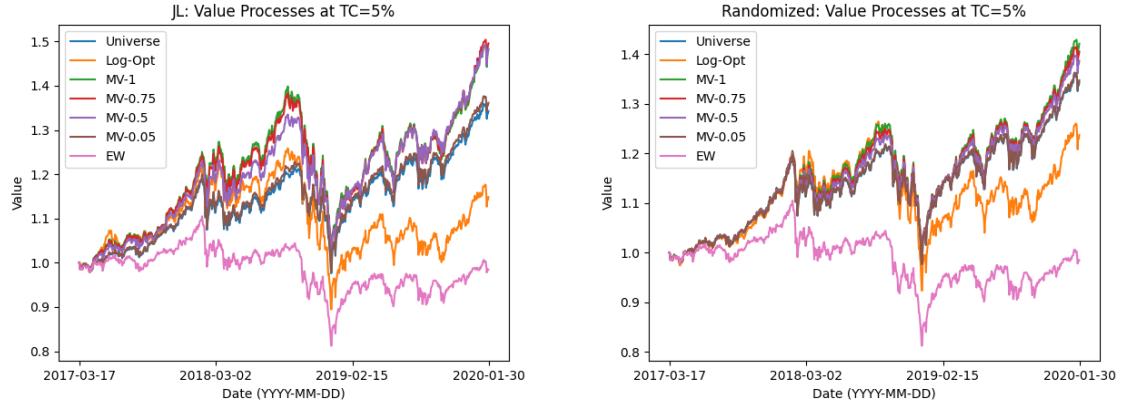
Figure 5: SMI wealth processes of the JL-signature portfolios (left), randomized-signature portfolios (right), universe portfolio and equally-weighted portfolio with and without transaction costs. Note that signature portfolios in the settings with transaction cost are those, where we included the regularization for transaction costs.



(a) Wealth process without transaction costs for JL-signature portfolios (left) and randomized-signature portfolios (right)



(b) Wealth process with 1% proportional transaction costs for JL-signature portfolios (left) and randomized-signature portfolios (right)



(c) Wealth process with 5% proportional transaction costs for JL-signature portfolios (left) and randomized-signature portfolios (right)

Figure 6: S&P500 wealth processes of the JL-signature portfolios (left), randomized-signature portfolios (right), universe portfolio and equally-weighted portfolio with and without transaction costs. Note that signature portfolios in the settings with transaction cost are those, where we included the regularization for transaction costs.

## References

- [1] E. Akyildirim, M. Gambara, J. Teichmann, and S. Zhou. Applications of signature methods to market anomaly detection. *Preprint arXiv:2201.02441*, 2022.
- [2] E. Akyildirim, M. Gambara, J. Teichmann, and S. Zhou. Randomized signature methods in optimal portfolio selection. *Preprint arxiv:2312.16448*, 2023.
- [3] A. Allan, C. Cuchiero, C. Liu, and D. Prömel. Model-free portfolio theory: A rough path approach. *Math. Finance*, 33:709–765, 2023.
- [4] H. Andrès, A. Boumezoued, and B. Jourdain. Signature-based validation of real-world economic scenarios. *ASTIN Bulletin*, 24:410–440, 2024.
- [5] A. Banner and R. Ghomrasni. Local times of ranked continuous semimartingales. *Stochastic Process. Appl.*, 118:1244–1253, 2008.
- [6] C. Bayer, P. Hager, S. Riedel, and J. Schoenmakers. Optimal stopping with signatures. *Ann. Appl. Probab.*, 33(1), 2021.
- [7] F. E. Benth, N. Detering, and L. Galimberti. Neural networks in fréchet spaces. *Ann. Math. Artif. Intell.*, 91:75–103, 2023.
- [8] H. Boedihardjo, X. Geng, T. Lyons, and D. Yang. The signature of a rough path: Uniqueness. *Adv. Math.*, 293:720–737, 2016.
- [9] H. Bühler, B. Horvath, T. Lyons, I. P. Arribas, and B. Wood. A data-driven market simulator for small data environments, 2020.
- [10] S. Campbell and T.-K. L. Wong. Functional portfolio optimization in stochastic portfolio theory. *SIAM J. Financial Math.*, 13(2):576–618, 2022.
- [11] W. Cao, X. Wang, Z. Ming, and J. Gao. A review on neural networks with random weights. *Neurocomputing*, 275:278–287, 2018.
- [12] I. Chevyrev and H. Oberhauser. Signature moments to characterize laws of stochastic processes. *J. Mach. Learn. Res.*, 23(176):1–42, 2022.
- [13] E. M. Compagnoni, L. Biggio, A. Orvieto, T. Hofmann, and J. Teichmann. Randomized signature layers for signal extraction in time series data. *Preprint arxiv:2201.00384*, 2022.
- [14] R. Cont and D. Fournie. A functional extension of the ito formula. *Comptes Rendus Mathematique*, 348(1):57–61, 2010.
- [15] C. Cuchiero. Polynomial processes in stochastic portfolio theory. *Stochastic Process. Appl.*, 129(5):1829–1872, 2019.
- [16] C. Cuchiero, G. Gazzani, J. Möller, and S. Svaluto-Ferro. Joint calibration to spx and vix options with signature-based models. *Math. Finance*, page 1–53, 2024.
- [17] C. Cuchiero, G. Gazzani, and S. Svaluto-Ferro. Signature-based models: Theory and calibration. *SIAM J. Financial Math.*, 14(3):910–957, 2023.
- [18] C. Cuchiero, L. Gonon, L. Grigoryeva, J.-P. Ortega, and J. Teichmann. Discrete-time signatures and randomness in reservoir computing. *IEEE Trans. Neural Netw. Learn. Syst.*, 33(11):6321–6330, 2021.

- [19] C. Cuchiero, L. Gonon, L. Grigoryeva, J.-P. Ortega, and J. Teichmann. Expressive power of randomized signature. In *The Symbiosis of Deep Learning and Differential Equations*, 2021.
- [20] C. Cuchiero, F. Primavera, and S. Svaluto-Ferro. Universal approximation theorems for continuous functions of càdlàg paths and lévy-type signature models. *Preprint arxiv:2208.02293*, 2022.
- [21] C. Cuchiero, W. Schachermayer, and T.-K. L. Wong. Cover's universal portfolio, stochastic portfolio theory and the numeraire portfolio. *Math. Finance*, 29:773–803, 2019.
- [22] C. Cuchiero, P. Schmocke, and J. Teichmann. Global universal approximation of functional input maps on weighted spaces. *Preprint arxiv:2306.03303*, 2023.
- [23] C. Cuchiero, S. Svaluto-Ferro, and J. Teichmann. Signature sdes from an affine and polynomial perspective. *Preprint arxiv:2302.01362*, 2023.
- [24] F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Math. Ann.*, 300(1):463–520, 1994.
- [25] B. Dupire. Functional itô calculus. *Quant. Finance*, 19(5):721–729, 2019.
- [26] A. Eberle. Markov processes. *Lecture Notes at University of Bonn*, 2016.
- [27] R. Fernholz. *Stochastic Portfolio Theory*. Springer, 2002.
- [28] R. Fernholz and I. Karatzas. Relative arbitrage in volatility-stabilized markets. *Ann. Finance*, 1(2):149–177, 2005.
- [29] R. Fernholz and I. Karatzas. Stochastic portfolio theory: an overview, 2010.
- [30] D. Filipovic and E. Platen. Consistent market extensions under the benchmark approach. *Math. Finance*, 19:41–52, 2009.
- [31] P. K. Friz and N. Victoir. *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*. Cambridge University Press, 2010.
- [32] K. Fujii and K. Nakajima. Harnessing disordered-ensemble quantum dynamics for machine learning. *Physical Review Applied*, 8(2), 2017.
- [33] O. Futter, B. Horvath, and M. Wiese. Signature trading: A path-dependent extension of the mean-variance framework with exogenous signals. *Preprint arxiv: 2308.15135*, 2023.
- [34] C. Galicchio and S. Scardapane. Deep randomized neural networks. *Preprint arxiv:2002.12287*, 2021.
- [35] L. Gonon and J.-P. Ortega. Fading memory echo state networks are universal. *Neural networks : the official journal of the International Neural Network Society*, 138:10–13, 2021.
- [36] L. Grigoryeva and J.-P. Ortega. Echo state networks are universal. *Neural networks : the official journal of the International Neural Network Society*, 108:495–508, 2018.
- [37] L. Grigoryeva and J.-P. Ortega. Universal discrete-time reservoir computers with stochastic inputs and linear readouts using non-homogeneous state-affine systems. *J. Mach. Learn. Res.*, 19(24):1–40, 2018.
- [38] C. Herrera, F. Krach, P. Ruyssen, and J. Teichmann. Optimal stopping via randomized neural networks. *Frontiers of Mathematical Finance*, 3(1):31–77, 2024.

- [39] G.-B. Huang, L. Chen, and C.-K. Siew. Universal approximation using incremental constructive feedforward networks with random hidden nodes. *Trans. Neur. Netw.*, 17(4):879–892, 2006.
- [40] H. Hulley and M. Schweizer. *M6—On Minimal Market Models and Minimal Martingale Measures*, pages 35–51. Springer Berlin Heidelberg, 2010.
- [41] D. Itkin, B. Koch, M. Larsson, and J. Teichmann. Ergodic robust maximization of asymptotic growth under stochastic volatility. *Preprint arXiv:2211.15628*, 2022.
- [42] H. Jaeger. The “echo state” approach to analysing and training recurrent neural networks—with an erratum note’. *German National Research Center for Information Technology GMD Technical Report*, 148, 2001.
- [43] W. B. Johnson. Extensions of lipschitz mappings into hilbert space. *Contemporary mathematics*, 26:189–206, 1984.
- [44] J. Kalsi, T. Lyons, and I. P. Arribas. Optimal execution with rough path signatures. *SIAM J. Financial Math.*, 11(2):470–493, 2020.
- [45] I. Karatzas and C. Kardaras. The numéraire portfolio in semimartingale financial models. *Finance Stoch.*, 11(4):447–493, 2007.
- [46] I. Karatzas and D. Kim. Trading strategies generated pathwise by functions of market weights. *Finance Stoch.*, 24(2):423–463, 2019.
- [47] I. Karatzas and J. Ruf. Trading strategies generated by lyapunov functions. *Finance Stoch.*, 21(3):753–787, 2017.
- [48] C. Kardaras. Market viability via absence of arbitrage of the first kind. *Finance and Stoch.*, 16(4):651–667, 2012.
- [49] C. Kardaras and S. Robertson. Ergodic robust maximization of asymptotic growth. *Ann. Appl. Probab.*, 31(4):1787–1819, 2021.
- [50] D. Kim. Market-to-book ratio in stochastic portfolio theory. *Finance Stoch.*, 27(2):401–434, 2023.
- [51] F. J. Kiraly and H. Oberhauser. Kernels for sequentially ordered data. *J. Mach. Learn. Res.*, 20(31):1–45, 2019.
- [52] D. Levin, T. Lyons, and H. Ni. Learning from the past, predicting the statistics for the future, learning an evolving system. *Preprint arXiv:1309.0260*, 2013.
- [53] C. Litterer and H. Oberhauser. On a chen–fliess approximation for diffusion functionals. *Monatshefte für Mathematik*, 175:577–593, 2011.
- [54] T. Lyons, M. Caruana, and T. Lévy. *Differential Equations Driven by Rough Paths*. Springer, 2007.
- [55] T. Lyons, S. Nejad, and I. P. Arribas. Non-parametric pricing and hedging of exotic derivatives. *Appl. Math. Finance*, 27:457 – 494, 2019.
- [56] T. J. Lyons. Differential equations driven by rough signals. *Revista Matemática Iberoamericana*, 14(2):215–310, 1998.
- [57] H. M. Markovitz. *Portfolio selection: Efficient diversification of investments*. John Wiley, 1959.

- [58] J. Matoušek. On variants of the johnson–lindenstrauss lemma. *Random Structures Algorithms*, 33(2):142–156, 2008.
- [59] N. Muca Cirone, M. Lemercier, and C. Salvi. Neural signature kernels as infinite-width-depth-limits of controlled ResNets. *Proceedings of Machine Learning Research*, 202:25358–25425, 2023.
- [60] M. Negoro, K. Mitarai, K. Fujii, K. Nakajima, and M. Kitagawa. Machine learning with controllable quantum dynamics of a nuclear spin ensemble in a solid. *Preprint arxiv:1806.10910*, 2018.
- [61] C. Ni, Y. Li, and P. A. Forsyth. Neural network approach to portfolio optimization with leverage constraints: a case study on high inflation investment. *Quantitative Finance*, 24(6):753–777, 2023.
- [62] H. Ni, L. Szpruch, M. Sabate-Vidales, B. Xiao, M. Wiese, and S. Liao. Sig-Wasserstein GANs for Time Series Generation. *Proceedings of the Second ACM International Conference on AI in Finance*, 2022.
- [63] J. Ruf and K. Xie. Generalised lyapunov functions and functionally generated trading strategies. *Applied Mathematical Finance*, 26(4):293–327, 2019.
- [64] J. Ruf and K. Xie. The impact of proportional transaction costs on systematically generated portfolios. *SIAM J. Financial Math.*, 11(3):881–896, 2020.
- [65] A. Schied, L. Speiser, and I. Voloshchenko. Model-free portfolio theory and its functional master formula. *SIAM J. Finan. Math.*, 9(3):1074–1101, 2018.
- [66] Z. Zhang, S. Zohren, and S. Roberts. Deep learning for portfolio optimization. *The Journal of Financial Data Science*, 2020.
- [67] X. Y. Zhou and D. Li. Continuous-time mean-variance portfolio selection: A stochastic lq framework. *Appl. Math. Optim.*, 42:19–33, 2000.

## A Proofs of Section 2

Since the space of lifted stopped paths as introduced in Definition 2.8 contains paths defined on different time-intervals, we need the following definition.

**Definition A.1** (Projection on the final value). We denote by  $(\cdot)_\odot$  the projection to the value at the final time, i.e. for each  $t \in [0, T]$

$$(x_{[0,t]})_\odot := x_t.$$

For this projection on the final value we have the following continuity result.

**Lemma A.2.** *The function*

$$\begin{aligned} \lambda : [0, T] \times \mathcal{G}_T^N(\varphi, x) &\rightarrow \Lambda_T^N(\varphi, x) \\ (t, \hat{\mathbb{X}}_{[0,T]}^N(\omega)) &\mapsto \hat{\mathbb{X}}_{[0,t]}^N(\omega) \end{aligned}$$

is continuous.

*Proof.* Take an arbitrary but fixed  $(s, \hat{\mathbb{Y}}_{[0,T]}^N(\omega_1))$ . We show that  $\lambda$  is continuous at  $(s, \hat{\mathbb{Y}}_{[0,T]}^N(\omega_2))$  but since this point is arbitrary,  $\lambda$  is continuous everywhere. For simplicity, we write  $\hat{\mathbf{x}}_{[0,t]} = \hat{\mathbb{X}}_{[0,t]}^N(\omega_2)$ ,  $\hat{\mathbf{y}}_{[0,t]} = \hat{\mathbb{Y}}_{[0,t]}^N(\omega_1)$  and  $\hat{\mathbf{y}}^s_{[0,t]} = \widehat{\mathbb{Y}}_{[0,t]}^s(\omega_1)$ . For any  $(t, \hat{\mathbf{x}}_{[0,T]}) \in [0, T] \times \mathcal{G}_T^N$  it holds that

$$\begin{aligned} d_\Lambda(\lambda(t, \hat{\mathbf{x}}_{[0,T]}), \lambda(s, \hat{\mathbf{y}}_{[0,T]})) \\ &= d_\Lambda(\hat{\mathbf{x}}_{[0,t]}, \hat{\mathbf{y}}_{[0,s]}) \\ &\leq d_\Lambda(\hat{\mathbf{x}}_{[0,t]}, \hat{\mathbf{y}}_{[0,t]}) + d_\Lambda(\hat{\mathbf{y}}_{[0,t]}, \hat{\mathbf{y}}_{[0,s]}) \\ &= d_{p-var;[0,t]}(\hat{\mathbf{x}}_{[0,t]}, \hat{\mathbf{y}}_{[0,t]}) + |t - s| + d_{p-var;[0,t \vee s]}(\widehat{\mathbf{y}}_{[0,t \vee s]}, \widehat{\mathbf{y}}_{[0,t \vee s]}) \\ &\leq d_{p-var;[0,t]}(\hat{\mathbf{x}}_{[0,t]}, \hat{\mathbf{y}}_{[0,t]}) + |t - s| + \|\hat{\mathbf{y}}\|_{p-var;[t \wedge s, t \vee s]}, \end{aligned}$$

where  $\|\cdot\|_{p-var;[0,T]}$  is the norm induced by  $d_{p-var;[0,T]}$ . Let now  $\epsilon > 0$  and choose  $\delta > 0$  such that

$$\delta + \|\hat{\mathbf{y}}\|_{p-var;[s-\delta, s+\delta]} < \epsilon.$$

Note that this is possible because  $\|\hat{\mathbf{y}}\|_{p-var;[s-\delta, s+\delta]} \rightarrow 0$  as  $\delta \rightarrow 0$ .

Therefore, for all  $(t, \hat{\mathbf{x}}_{[0,T]}) \in [0, T] \times \mathcal{G}_T^N$  (equipped with the product topology) satisfying

$$|t - s| + d_{p-var;[0,T]}(\hat{\mathbf{x}}_{[0,T]}, \hat{\mathbf{y}}_{[0,T]}) < \delta,$$

we have

$$\begin{aligned} |t - s| + d_{p-var;[0,t]}(\hat{\mathbf{x}}_{[0,t]}, \hat{\mathbf{y}}_{[0,t]}) + \|\hat{\mathbf{y}}\|_{p-var;[t \wedge s, t \vee s]} &< \delta + \|\hat{\mathbf{y}}\|_{p-var;[t \wedge s, t \vee s]} \\ &\leq \delta + \|\hat{\mathbf{y}}\|_{p-var;[s-\delta, s+\delta]} \\ &< \epsilon, \end{aligned}$$

which proves continuity of  $\lambda$ . □

The following lemma can be proved by means of Corollary 2.11.

**Lemma A.3.** *For each multiindex  $I$  the map*

$$\Lambda_T^2(\varphi, x) \ni \hat{\mathbb{X}}_{[0,t]}^2(\omega) \mapsto \langle e_I, S^{|I|}(\hat{\mathbb{X}}_{[0,t]}^2(\omega))_\odot \rangle = \langle e_I, \hat{\mathbb{X}}_t^N(\omega) \rangle \in \mathbb{R}$$

is continuous on bounded sets.

*Proof.* Let us show the statement for an arbitrary but fixed multiindex  $I$ . For  $N = 2$  we apply the same notation as in the proof of Lemma A.2 above, i.e.  $\hat{\mathbf{x}}_{[0,t]} = \hat{\mathbb{X}}_{[0,t]}^2(\omega_x)$ , etc. We show continuity at some fixed  $\hat{\mathbf{y}}_{[0,s]}$  in some bounded set of  $\Lambda_T^2(\varphi, x)$ . Recall that by Corollary 2.11

$$\hat{\mathbf{x}}_{[0,t]} \mapsto \langle e_I, S^{|I|}(\hat{\mathbf{x}}_{[0,t]}) \rangle$$

is continuous on bounded sets of  $\mathcal{G}_t^2(\varphi, x)$  for each  $t \in [0, T]$ . Thus, for every  $\epsilon > 0$  there exists some  $\delta_1 > 0$  such that for all  $\hat{\mathbf{x}}_{[0,t]}$  satisfying

$$d_{p-var,[0,t]}(\hat{\mathbf{x}}_{[0,t]}, \hat{\mathbf{y}}_{[0,t]}) < \delta_1$$

we have

$$\|\langle e_I, S^{|I|}(\hat{\mathbf{x}}_{[0,t]}) \rangle - \langle e_I, S^{|I|}(\hat{\mathbf{y}}_{[0,t]}) \rangle\|_{p-var,[0,t]} < \frac{\epsilon}{2}.$$

Here,  $\|\cdot\|_{p-var,[0,t]}$  denotes the  $p$ -variation distance on  $\mathbb{R}$ . Moreover, the map

$$[0, T] \ni t \mapsto \langle e_I, S^{|I|}(\hat{\mathbf{y}}_{[0,t]}) \rangle \in \mathbb{R}$$

is continuous for each fixed  $\hat{\mathbf{y}}$ . Hence, there exists some  $\delta_2 > 0$  such that for all  $t$  with  $|t - s| < \delta_2$

$$|\langle e_I, S^{|I|}(\hat{\mathbf{y}}_{[0,t]}) \rangle - \langle e_I, S^{|I|}(\hat{\mathbf{y}}_{[0,s]}) \rangle| < \frac{\epsilon}{2}.$$

Choose now  $\delta = \min(\delta_1, \delta_2)$  and let  $\hat{\mathbf{x}}_{[0,t]}$  (in the bounded set) be such that

$$\begin{aligned} \delta > d_\Lambda(\hat{\mathbf{x}}_{[0,t]}, \hat{\mathbf{y}}_{[0,s]}) &= |t - s| + d_{p-var,[0,t \vee s]}(\hat{\mathbf{x}}_{[0,t \vee s]}, \hat{\mathbf{y}}_{[0,t \vee s]}) \\ &\geq |t - s| + d_{p-var,[0,t]}(\hat{\mathbf{x}}_{[0,t]}, \hat{\mathbf{y}}_{[0,t]}^{t \wedge s}). \end{aligned}$$

Then, by the above

$$\begin{aligned} |\langle e_I, S^{|I|}(\hat{\mathbf{x}}_{[0,t]}) \rangle - \langle e_I, S^{|I|}(\hat{\mathbf{y}}_{[0,s]}) \rangle| &\leq |\langle e_I, S^{|I|}(\hat{\mathbf{x}}_{[0,t]}) \rangle - \langle e_I, S^{|I|}(\hat{\mathbf{y}}_{[0,t]}^{t \wedge s}) \rangle| \\ &\quad + |\langle e_I, S^{|I|}(\hat{\mathbf{y}}_{[0,t]}^{t \wedge s}) \rangle - \langle e_I, S^{|I|}(\hat{\mathbf{y}}_{[0,s]}) \rangle| \\ &\leq \|\langle e_I, S^{|I|}(\hat{\mathbf{x}}_{[0,t]}) \rangle - \langle e_I, S^{|I|}(\hat{\mathbf{y}}_{[0,t]}^{t \wedge s}) \rangle\|_{p-var,[0,t]} \\ &\quad + |\langle e_I, S^{|I|}(\hat{\mathbf{y}}_{[0,t]}) \rangle - \langle e_I, S^{|I|}(\hat{\mathbf{y}}_{[0,s]}) \rangle| \\ &< \epsilon, \end{aligned}$$

which proves the assertion. □

We are now prepared to prove Theorem 2.12.

*Proof of Theorem 2.12.* Consider the function  $\lambda$  as introduced in Lemma A.2 and note that for any compact subset  $K \subset \mathcal{G}_T^2(\varphi, x)$  the set

$$\mathcal{K}_\Lambda = \lambda([0, T] \times K) = \left\{ \hat{\mathbb{X}}_{[0,t]}^2(\omega) \mid t \in [0, T] \text{ and } \hat{\mathbb{X}}_{[0,T]}^2(\omega) \in K \right\}$$

is a compact subset of  $\Lambda_T^2(\varphi, x)$  due to the continuity of  $\lambda$  as proved in Lemma A.2. For each  $\hat{\mathbb{X}}_{[0,t]}^2(\omega) \in \Lambda_T^2(\varphi, x)$  and  $N \geq 2$  it holds that  $S^N(\hat{\mathbb{X}}_{[0,t]}^2(\omega))_\odot = \hat{\mathbb{X}}_t^N(\omega)$  a.s. Hence, we can associate with every linear function on the signature  $L$  a non-anticipative path-functional  $f_L : \hat{\mathbb{X}}_{[0,t]}^2(\omega) \mapsto L(\hat{\mathbb{X}}_t(\omega))$ . Moreover, for every linear function on the signature  $L$  it holds that  $f_L \in C(\Lambda_T^2(\varphi, x); \mathbb{R})$ , which follows from Lemma A.3. Let us denote by

$$\mathfrak{F}_{\mathfrak{L}} := \left\{ f_L \mid f_L : \hat{\mathbb{X}}_{[0,t]}^2(\omega) \mapsto L(\hat{\mathbb{X}}_t(\omega)) \text{ for } \hat{\mathbb{X}}_{[0,t]}^2(\omega) \in \Lambda_T^2(\varphi, x), L \in \mathfrak{L} \right\}$$

the set of path functionals induced by linear functions on the signature and by

$$\mathfrak{F}_{\mathcal{L}}|_{\mathcal{K}_{\Lambda}} := \left\{ f_L \mid f_L : \hat{\mathbb{X}}_{[0,t]}^2(\omega) \mapsto L(\hat{\mathbb{X}}_t(\omega)) \text{ for } \hat{\mathbb{X}}_{[0,t]}^2(\omega) \in \mathcal{K}_{\Lambda}, L \in \mathcal{L} \right\}$$

its restriction to lifted paths in  $\mathcal{K}_{\Lambda}$ . We will now apply the Stone-Weierstrass Theorem to show that  $\mathfrak{F}_{\mathcal{L}}|_{\mathcal{K}_{\Lambda}}$  is dense in  $C(\mathcal{K}_{\Lambda}; \mathbb{R})$ . To this end we need to show that  $\mathfrak{F}_{\mathcal{L}}|_{\mathcal{K}_{\Lambda}}$  is a sub-algebra of  $C(\mathcal{K}_{\Lambda}; \mathbb{R})$  which separates points and vanishes nowhere. The algebra property follows from the fact that for each  $\hat{\mathbb{X}}_{[0,t]}^2(\omega) \in \Lambda_T^2(\varphi, x)$  and for any two  $f_{L_1}, f_{L_2} \in \mathfrak{F}_{\mathcal{L}}$  it holds that

$$\begin{aligned} (f_{L_1} \cdot f_{L_2})(\hat{\mathbb{X}}_{[0,t]}^2(\omega)) &= f_{L_1}(\hat{\mathbb{X}}_{[0,t]}^2(\omega)) f_{L_2}(\hat{\mathbb{X}}_{[0,t]}^2(\omega)) \\ &= L_1(\hat{\mathbb{X}}_t(\omega)) L_2(\hat{\mathbb{X}}_t(\omega)) = L(\hat{\mathbb{X}}_t(\omega)) = f_L(\hat{\mathbb{X}}_{[0,t]}^2(\omega)), \end{aligned}$$

where the second last equality follows from the shuffle product property. We now show that  $\mathfrak{F}_{\mathcal{L}}$  separates points. To this end take  $\hat{\mathbb{X}}_{[0,t]}^2(\omega_x), \hat{\mathbb{Y}}_{[0,s]}^2(\omega_y) \in \Lambda_T^2(\varphi, x)$  with  $\hat{\mathbb{X}}_{[0,t]}^2(\omega_x) \neq \hat{\mathbb{Y}}_{[0,s]}^2(\omega_y)$ . If  $t \neq s$ , the two points are separated in the time-component, i.e.  $\langle e_1, \hat{\mathbb{X}}_t(\omega_x) \rangle \neq \langle e_1, \hat{\mathbb{Y}}_s(\omega_y) \rangle$ . If  $t = s$ , the point-separation follows from Lemma 2.4. The algebra vanishes nowhere because  $f_{\emptyset}(\hat{\mathbb{X}}_{[0,t]}^2(\omega)) = \langle e_{\emptyset}, \hat{\mathbb{X}}_t(\omega) \rangle \equiv 1$  for all  $\hat{\mathbb{X}}_{[0,t]}^2(\omega) \in \Lambda_T^2$ , where  $\emptyset$  denotes the empty word. The desired density statement follows by the Stone-Weierstrass Theorem, i.e. for any continuous non-anticipative path-functional  $f : \Lambda_T^2(\varphi, x) \rightarrow \mathbb{R}$  and for every  $\epsilon > 0$  there exists a path-functional  $f_L \in \mathfrak{F}_{\mathcal{L}}$  such that for almost all  $\omega \in \Omega$

$$\sup_{(t, \hat{\mathbb{X}}_{[0,T]}^2(\omega)) \in [0,T] \times K} |f(\hat{\mathbb{X}}_{[0,t]}^2(\omega)) - f_L(\hat{\mathbb{X}}_t(\omega))| < \epsilon,$$

which is equivalent to the claim of Theorem 2.12.  $\square$

*Proof of Lemma 2.19.* We have to prove that  $K_R = \psi^{-1}((0, R])$  is compact for any  $R > 0$ . Let us first note that the map  $\lambda$  can be easily extended to weakly geometric  $\alpha$ -Hölder rough paths and is continuous again by the exact same arguments as before. Moreover, note that the set

$$K'_R = \left\{ \widehat{\mathbf{X}}_{[0,T]}^t \middle| \psi(\widehat{\mathbf{X}}_{[0,t]}) \in [0, R] \right\}$$

equipped with the  $d_{\alpha', [0,T]}$ -topology is for any  $R > 0$  a  $\|\cdot\|_{\alpha, [0,T]}$ -bounded subset of  $\hat{C}_T^{\alpha}(\varphi, x)$  and hence compact for  $0 \leq \alpha' < \alpha$  due to the compact embedding of  $\alpha$ -Hölder spaces into  $\alpha'$ -Hölder spaces, see [22, Theorem A.3]. Moreover,  $K_R = \lambda([0, T] \times K'_R)$  and is therefore compact for any  $R > 0$  due to Tychonoff's theorem and the continuity of  $\lambda$ .  $\square$

*Proof of Theorem 2.20.* Using Lemma 2.19 the proof is analogous to the proof of [22, Theorem 5.4]. In fact it even holds that

$$\sup_{\widehat{\mathbf{X}}_{[0,t]} \in \bigcup_{t \in [0,T]} \hat{C}_t^{\alpha}(\varphi, x)} \frac{|f(\widehat{\mathbf{X}}_{[0,t]}) - L(\hat{\mathbb{X}}_t)|}{\psi(\widehat{\mathbf{X}}_{[0,t]})} < \epsilon$$

and since  $\Lambda_T^2(\varphi, x) \subset \bigcup_{t \in [0,T]} \hat{C}_t^{\alpha}(\varphi, x)$  for every  $\frac{1}{3} < \alpha \frac{1}{2}$  the claim follows.  $\square$

## B Proofs of Section 4

The following lemma is needed in the proof of Theorem 4.4.

**Lemma B.1.** *Let  $D$  be a (subset of a) metric space. Consider a matrix-valued function  $M : D \ni x \mapsto M(x)$  whose coefficients are continuous functions from  $D$  to  $\mathbb{R}$  and which is non-singular for all  $x \in D$ , then the coefficients of its matrix-inverse  $M^{-1}$  are again continuous functions.*

*Proof.* Let us prove the statement by induction:

- Let  $M : D \rightarrow \mathbb{R}^{2 \times 2}$ . Then the inverse of

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is given by

$$M^{-1} = \frac{1}{(ad - bc)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

By assumption the determinant  $(ad - bc)$  is non-zero, which implies continuity of the coefficients of  $M^{-1}$ .

- Assume the statement holds for an arbitrary but fixed  $n \geq 2$ .
- Let  $M : D \rightarrow \mathbb{R}^{(n+1) \times (n+1)}$ . Then we can write  $M$  in block matrix form

$$M = \begin{pmatrix} \tilde{M} & m^1 \\ m^2 & m^3 \end{pmatrix},$$

where  $\tilde{M}$  takes values in  $\mathbb{R}^{n \times n}$ ,  $m^1, (m^2)^\top$  in  $\mathbb{R}^n$  and  $m^3$  in  $\mathbb{R}$ . The inverse is then computed via

$$M^{-1} = \begin{pmatrix} \tilde{M}^{-1} + \frac{1}{\Delta} \tilde{M}^{-1} m^1 & m^2 \tilde{M}^{-1} & -\frac{1}{\Delta} \tilde{M}^{-1} m^1 \\ \frac{1}{\Delta} m^2 \tilde{M}^{-1} & \frac{1}{\Delta} & \frac{1}{\Delta} \end{pmatrix}$$

where  $\Delta = m^3 - m^2 \tilde{M}^{-1} m^1$  with values in  $\mathbb{R} \setminus \{0\}$  by assumption on invertability. As (the components of)  $\tilde{M}^{-1}, m^1, m^2$  and  $m^3$  are all continuous functions, continuity of the components of  $M^{-1}$  follows as well.

□

## C Proofs of Section 5

*Proof of Lemma 5.14.* By Equation (5.2), we know that

$$\begin{aligned} \log \left( \frac{W_t^\pi}{W_t^{\mathcal{U}}} \right) &= \sum_{i \in \mathcal{U}} \int_{t_0}^t \gamma_s^i f^i(s, X_{[0,s]}) d\mu_s^{\mathcal{U},i} \\ &\quad - \frac{1}{2} \sum_{i,j \in \mathcal{U}} \int_{t_0}^t \gamma_s^i \gamma_s^j f^i(s, X_{[0,s]}) f^j(s, X_{[0,s]}) d[\mu_s^{\mathcal{U},i}, \mu_s^{\mathcal{U},j}]_s \\ &\quad + \sum_{i \in \mathcal{U}} \int_{t_0}^t \left( 1 - \sum_{k=1}^d \lambda_s^k f^k(s, X_{[0,s]}) \right) d\mu_s^{\mathcal{U},i} \end{aligned} \tag{C.1}$$

$$- \frac{1}{2} \sum_{i,j \in \mathcal{U}} \int_{t_0}^t \left( 1 - \sum_{k=1}^d \lambda_s^k f^k(s, X_{[0,s]}) \right) \left( 1 - \sum_{m=1}^d \lambda_s^m f^m(s, X_{[0,s]}) \right) d[\mu_s^{\mathcal{U},i}, \mu_s^{\mathcal{U},j}]_s, \tag{C.2}$$

for

$$\lambda_s^k = \begin{cases} \mu_s^{\mathcal{U},k} & \text{if } \pi \text{ is of type I} \\ 1 & \text{if } \pi \text{ is of type II} \end{cases}.$$

Note that the terms in (C.1) and (C.2) are both zero due to  $\sum_{i \in \mathcal{U}} \mu_t^{\mathcal{U},i} \equiv 1$ .

□

## D Proofs of Section 6

*Proof of Proposition 6.1.* 1. This case always holds for long-only portfolios and is discussed in [64]. The argument in the long-only case is that  $L(1) < R(1)$  and  $L(0) > R(0)$ , while both sides are obviously continuous and monotone. While the former statement still holds in the case with short-selling, the monotonicity is no longer true. However, one can show by a tedious case-by-case study that where  $R'(\alpha)$  exists it is piece-wise constant and increasing. Therefore a unique solution for  $\alpha^* \in [0, 1]$  is still assured in this case.

2. Here the above argument does not hold anymore. This is because now  $L(0) \leq R(0)$ , while it still holds that  $L(1) < R(1)$ . Let us give three examples, one for each of the cases a) no solution b) no solution in  $[0, 1]$ , c) no unique solution in  $[0, 1]$ .

- (a) *no solution*: consider a market of two stocks and  $c = 0.05$ , i.e. 5% transaction costs.  
For the portfolio weights

$$\pi_{t^-} = (13.1, -12.1)^T, \quad \pi_t = (13, -12)^T$$

there does not exist a solution for  $\alpha$ .

- (b) *no solution in  $[0, 1]$* : again for  $c = 0.05$  and a market of two stocks. For the portfolio weights

$$\pi_{t^-} = (11, -10)^T, \quad \pi_t = (10, -9)^T$$

the solution is  $\alpha = -1$ .

- (c) *no unique solution in  $[0, 1]$* : consider  $c = 0.05$  and a market of three stocks with portfolio weights

$$\pi_{t^-} = (5, 6, -10)^T, \quad \pi_t = (5.5, 6.5, -11)^T$$

then the solutions are  $\alpha = 0.3333$  and  $\alpha = 0.9535$  (numbers are rounded).

□