

---

# Signature Trading: A Path-Dependent Extension of the Mean-Variance Framework with Exogenous Signals

---

Owen Futter<sup>1</sup>, Blanka Horvath<sup>2,3</sup>, and Magnus Wiese<sup>4</sup>

<sup>1</sup>Imperial College London, Department of Mathematics

<sup>2</sup>University of Oxford, Mathematical Institute and Oxford Man Insititute

<sup>3</sup>The Alan Turing Institute

<sup>4</sup>University of Kaiserslautern, Department of Mathematics

## Abstract

In this article we introduce a portfolio optimisation framework, in which the use of *rough path signatures* [Lyo98] provides a novel method of incorporating path-dependencies in the joint signal-asset dynamics, naturally extending traditional factor models, while keeping the resulting formulas lightweight, tractable and easily interpretable. Specifically, we achieve this by representing a trading strategy as a linear functional applied to the *signature of a path* (which we refer to as “*Signature Trading*” or “*Sig-Trading*”). This allows the modeller to efficiently encode the evolution of past time-series observations into the optimisation problem. In particular, we derive a concise formulation of the dynamic mean-variance criterion alongside an explicit solution in our setting, which naturally incorporates a drawdown control in the optimal strategy over a finite time horizon. Secondly, we draw parallels between classical portfolio stategies and *Sig-Trading* strategies and explain how the latter leads to a pathwise extension of the classical setting via the “*Signature Efficient Frontier*”. Finally, we give explicit examples when trading under an exogenous signal as well as examples for momentum and pair-trading strategies, demonstrated both on synthetic and market data. Our framework combines the best of both worlds between classical theory (whose appeal lies in clear and concise formulae) and between modern, flexible data-driven methods (usually represented by ML approaches) that can handle more realistic datasets. The advantage of the added flexibility of the latter is that one can bypass common issues such as the accumulation of heteroskedastic and asymmetric residuals during the optimisation phase. Overall, *Sig-Trading* combines the flexibility of data-driven methods without compromising on the clarity of the classical theory and our presented results provide a compelling toolbox that yields superior results for a large class of trading strategies.

**Keywords** Mean-Variance Optimisation · Signature Methods · Data-Driven Methods · Dynamic Trading Strategies · Path-Dependent Signals · Statistical Arbitrage · Momentum Strategies · Stochastic Filtering

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Background and Motivation . . . . .	4
1.2	Summary of Contributions . . . . .	7

The authors would like to thank William F. Turner for helpful comments as well as Johannes Muhle-Karbe, Christopher Salvi and Joseph Mulligan for fruitful discussions throughout the writing of this paper. OF and BH thankfully acknowledge the funding of this research by Atlantic House Investments & EPSRC Oxford-ICL Centre of Doctoral Studies in Mathematics of Random Systems.

<b>2</b>	<b>The Modelling Setup of the Sig-Factor Model</b>	<b>7</b>
2.1	The Signature . . . . .	7
2.2	Sig-Factor Model . . . . .	9
<b>3</b>	<b>Main Result</b>	<b>14</b>
3.1	Sig-Factor Model vs Classical Factor Model . . . . .	18
3.2	Optimal Static Portfolio . . . . .	19
3.3	Sig-Factor Model Efficient Frontier . . . . .	20
<b>4</b>	<b>Implementation</b>	<b>21</b>
<b>5</b>	<b>Numerical Results</b>	<b>23</b>
5.1	Synthetic Data . . . . .	23
5.1.1	Pairs Trading . . . . .	23
5.1.2	Incorporating Exogenous Signal . . . . .	25
5.2	Learning Momentum as a Sig-Trading Strategy . . . . .	27
<b>6</b>	<b>Conclusion</b>	<b>29</b>
<b>Appendix A Rough Path &amp; Tensor Algebra Preliminaries</b>		<b>30</b>
A.1	The Tensor Algebra . . . . .	30
A.2	Rough Paths & The Signature . . . . .	31
<b>Appendix B Proofs</b>		<b>34</b>
B.1	Proof of Theorem 2.11 . . . . .	34

## 1 Introduction

The design and construction of trading strategies is a fundamental aspect of finance and subsequently, has been extensively researched in the past decades. Creating a trading strategy can mainly be divided into two areas: extracting *alpha* and allocating the associated *risk*. Methods of extracting alpha depend on the trade horizon or trade frequency and are often achieved through statistical techniques, which can then be incorporated into a parametric model. Meanwhile, allocating the associated risk generally depends on your choice of model and objective criterion; conventionally, optimisation methods are deployed to find a transformation of the signal-asset dynamics that maximises the chosen utility function of the strategy PnL, with respect to the underlying model. The work ([Mar52]) of Markowitz introduced modern portfolio theory based on portfolio allocation determined by investors' preferences to risk and returns, resulting in a well-diversified portfolio. Classical methods consist of fixing a class of probabilistic parametric models and calibrating the model's parameters with respect to the empirical price process. Depending on the choice of parametric model and its corresponding parameters, the optimal portfolio will be different. In well-studied families of models it is possible to find closed-form expressions of the optimal portfolio given certain risk and return preferences, however in more complex parametric models or with more general utility functions, explicit solutions may not be attainable and so numerical techniques are used.



We can observe from the above blueprint, that the possible choices of model are vast - for example all possible ways of feature engineering, return prediction, or the choice of probabilistic model. The choice of optimisation technique will then be tailored to the choice of utility function and the model. While well-studied models provide a useful benchmark for practitioners, many of

the assumptions made are restrictive. Several modelling assumptions do not reflect stylised facts in practice such as non-stationarity, heavy tails and path-dependent volatility ([Fuk21; Das22; GL22; MMB23]) - all of which are exhibited by financial time-series data ([Con01]). The latter considerations are even more relevant when working with multiple assets since dependence structures are highly non-linear, often with cross-sectional trend- and mean-reversion patterns occurring in practice. Recent research activity has brought modelling approaches to the forefront that are inherently data-driven and provide a highly flexible framework that is able to capture a broader set of asset dynamics than currently used classical stochastic models did. Data-driven techniques for trading have increased in recent years due to the rise of machine learning applications in finance, such as in [ZZR20b; Wan19; ZZR20a; Jai+21; Van+21; PZ22; GPZ21; CJC22], as well as in hedging applications in [Büh+18; HTZ21; LH23]. The requirement for complex and accurate synthetic data to train and test trading frameworks has also led to extensive research in developing market generators ([Bue+21; Wie+19; Ni+20; Ni+21; Iss+23]). In this work, we go a step further than just modelling asset dynamics in a model-free way and take the approach of utilising rough path theory ([Lyo14; FV10]) also to develop a trading strategy that is determined by the expected signature of the joint signal-asset process. Since the aforementioned expected signature uniquely determines the law of the stochastic process ([CL13]), it is immediate to see how the techniques simplify to the classical case when applied to traditional stochastic models. By taking a pathwise approach to trading, this naturally alleviates probabilistic restrictions, resulting in a *model-free* or model-agnostic setup ([CC23]). A pathwise setting has been used in [PP16; Rig16a; Rig16b; ACX23], and has more recently been utilised to handle more general portfolios in [All+21], as well as in applications to derivative pricing and calibration in [CGS22] and optimal stopping problems in [Bay+21].

Inspired by the work of Perez et al. in [Arr18; KLA20; LNA19b], we adopt the idea of representing a trading strategy as a linear functional applied to the signature of a path and extend this to incorporate exogenous market signals and multiple assets. We can think of a trading strategy as a decision made using the knowledge of the current state (e.g. the previous price path, plus some exogenous trading signal); in other words as a function from path space to some decision process. In practice, the true driving processes are most likely not (directly) observable and hidden by layers of noise. Hence the mapping from the previous price process to a trading strategy is often done first by removing noise in the system through stochastic filtering [BC09; CLO21] (i.e. an exponentially weighted moving average or Kalman filter) and then optimised based on the resulting prediction. Recently, the authors in [Coh+23] prove how the Kalman filter can in fact be equivalently written as a linear regression on the signature. Work has also been conducted in [DCS21; Dye+22] with relation to approximate Bayesian computation using signatures. This naturally prompts the question - can the class of linear functionals on the signature be seen as a rich enough class of maps that represent such trading strategies? We will show that this is possible due to signatures' capability to approximate continuous functions on paths. Also perhaps crucially, the Sig-Trading framework does not impose the restriction that the underlying asset or signal be Markovian, allowing the Sig-Trader to capture auto-correlation and mean-reverting behaviours within the process that perhaps some more classical methods are not able to exploit. This enables us to incorporate path-dependent considerations into our trading decisions, while still obtaining a closed-form solution, that is easy to compute and simple to analyse.

The paper is organised as follows: In Section 2 we introduce key foundations of Sig-Trading and the concept of the extension to classical factor models. In Section 3 we present our main result, an analytic solution to the dynamic mean-variance criterion for Sig-Trading and compare this to original factor models, whilst introducing a Sig-Trading version of the efficient frontier. Finally, in Section 4 we discuss its implementation and in Section 5 we highlight the advantages of Sig-Trading, provide intuitive examples and demonstrate its possible use cases in practice, such as pairs trading, momentum, and trading under an exogenous signal.

## 1.1 Background and Motivation

### The Objective

In this article, we are concerned with finding an optimal *systematic* and *dynamic* trading strategy such that the trader continuously updates their position as new information filters in. This is done with no discretion and is defined as a function of the market state, which continuously updates through time, leading to a new position for each time  $t$ .

We denote  $T \in \mathbb{R}$  the terminal time. Let  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$  be a filtered probability space. Furthermore, denote by  $X = (X_t)_{t \in [0, T]}$  a non-negative  $\mathbb{R}^d$ -valued stochastic process satisfying  $X_0^m = 1$  for  $m \in \{1, \dots, d\}$ . We are interested in finding an optimal, predictable dynamic trading strategy  $(\xi_t)_{t \in [0, T]}$  that maximises the expected utility of the PnL

$$\max_{(\xi_t)_{t \in [0, T]}} \mathbb{E}(U(V_T)) \quad (1.1)$$

s.t constraints

where  $V_T$  is the terminal value of the trading strategy (i.e the PnL)

$$V_T = \sum_{m=1}^d \int_0^T \xi_t^m dX_t^m. \quad (1.2)$$

Here, the optimal strategy  $\xi$  is a function of the market state (i.e filtration  $\mathcal{F}$  at time  $t$ ) and its optimality is with respect to the constraints and utility function that the trader chooses. In order to solve such an optimisation problem, commonly methods are constructed from the following structure:

1. Fix a framework to model the underlying dynamics of  $X$ ,
2. Choose an objective criterion (utility and constraints),
3. Optimisation with respect to (1) and (2).

Most likely, a trader may be trading under the presence of exogenous information  $f$  to enrich the filtration  $\mathcal{F}$  (and hence the model of the dynamics of  $X$ ). Recently, extensive research in this direction has been conducted in optimal execution literature, e.g. in a more classical setting in [LN19; For+22; SC22; CDM22; BCK23; CMN23] and with machine learning applications in [BDG21; CDO23]. In this work, we do not consider market impact, but refer the reader to [KLA20; CAS22] to work in this area involving signatures.

### Modelling Dynamics

The choice of model in (1) often requires explicitly predicting asset returns through supervised learning techniques. Therefore, a large number of models tend to fall into the *predict-then-optimise* framework, where heavy assumptions are made on the asset returns/market factors that are input into any prediction, leaving the final solution exposed to asymmetric and compounded errors. Solutions to these problems are very well studied with specific assumptions and restrictions on the underlying asset process, however without these assumptions this can be much more difficult. In such stochastic control problems, the asset dynamics are explained by a diffusion process and dynamic programming can then be used to solve the Hamilton-Jacobi-Bellman (HJB) equation. In practice, the future expected returns (the drift of the process  $X$ ),  $\mu_{t+1}$ , are often predicted via supervised learning methods, using trading signals or *factors* as statistical predictors which are embedded into the framework itself ([CR83; NER92; FF93; FF15; SW10; GP13]). A generic (linear) factor model models the asset returns at time  $t$ , as

$$\mu_{t+1} = \mathbb{E}[r_{t+1} | \mathcal{F}_t] = Bf_t + \varepsilon_{t+1} \quad (1.3)$$

where  $r$  is the  $d$ -dimensional asset returns,  $B$  is a  $d \times N$  matrix of factor coefficients,  $f_t$  is a vector of  $N$  factor returns and  $\varepsilon$  is a vector of the  $d$  assets' (unexplained) residuals returns. This is set up as a supervised linear regression on future returns, as a function of the trading signals/factors.

However, when working with financial data, there is generally a very low signal to noise ratio and the residual terms  $\varepsilon_t$  are badly behaved, violating many statistical assumptions. Due to autocorrelation, non-stationarity and path-dependent volatility in the underlying asset  $X$  (and also the signal  $f$ ), this framework can very quickly become problematic, and these asymmetric errors are then compounded in the optimisation phase. In order to capture some of the autocorrelation in the residuals, a trader could incorporate the path into the signal  $f$  via stochastic filtering ([BC09; CLO21]). Traders may also scale returns for volatility in order to remove heteroskedasticity ([Eng01]), log transform to remove asymmetry or winsorise to remove fat tails; in [Das22], the roughness of signals is also discussed. However, there still remains a large amount of discretion in such feature engineering and we will show that the signature can be used efficiently and robustly to tackle these issues in a data-driven manner. To avoid the accumulation of mis-specified error terms from the prediction phase, end-to-end (E2E) approaches using machine learning frameworks have been used in [CI22] and [Zha+21] to ensure robustness by bypassing the prediction stage.

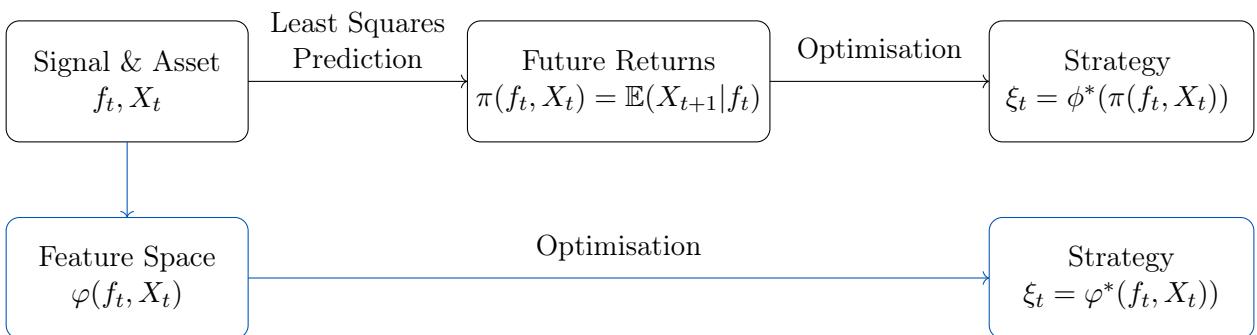


Figure 1: Original framework (above) and end-to-end (E2E) optimisation framework (below).

### Choice of Utility

Once we have a model that characterises the dynamics of the driving signal and the underlying asset, we can proceed to transforming this into a trading strategy position. How one does this depends on a variety of conditions such as the type of strategy, if we are trading multiple assets, risk preferences, trade frequency and investment horizon. Common objective criteria focus on a single trade-by-trade optimisation basis, overlooking the potential path that the trading strategy will take. However as alluded to previously, in practice signals and underlying assets can have strong temporal dependencies and so subsequent trading strategy positions will inherit autocorrelation structure.

As a motivating example in Figure 2, we consider the comparison between a strategy that is mean-reverting (has autocorrelation) vs one that doesn't. Both strategies yield identical daily PnL distributions (and hence Sharpe ratio), but have different distributions at a future time due to the temporal structure within the strategy over time. By optimising with respect to a future time horizon, we are inherently optimising for a dynamic criterion that is dependent on the path. This approach naturally integrates a drawdown control in the optimisation, since drawdowns are a path-dependent characteristic. Achieving a strategy with a small maximum drawdown either requires a strong trend-to-noise ratio (high Sharpe ratio) or relies on the strategy being mean-reverting during volatile periods ([RSB17] provides a neat analysis).

Mean-variance optimisation, perhaps the most widely known choice of utility function, was first introduced in Markowitz's thesis ([Mar52]). He demonstrated that it was natural to construct

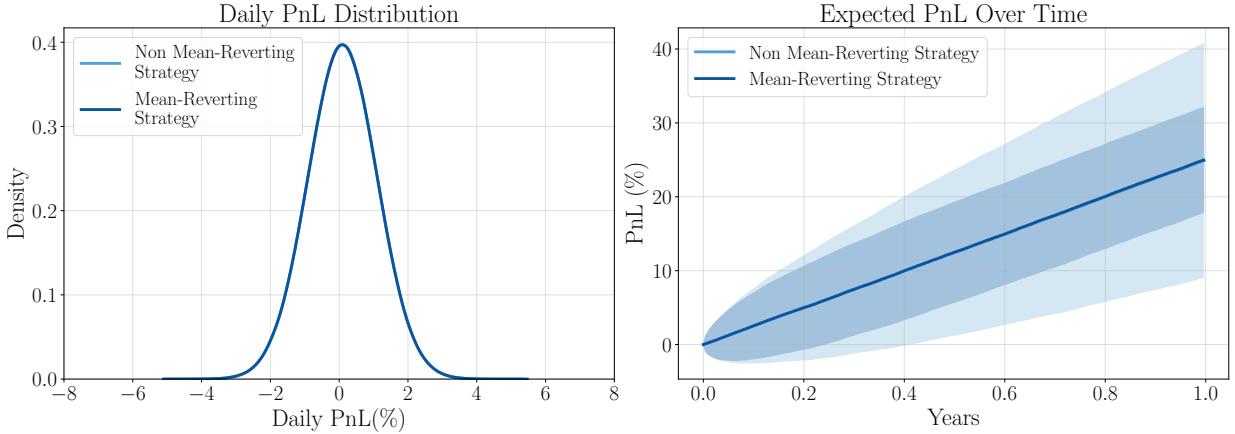


Figure 2: Comparison of the PnL profile through time of a mean-reverting strategy vs a non mean-reverting strategy.

an objective function that rewarded positive returns while penalizing associated risk (via variance). Subsequently, this framework has been researched extensively in the literature, including the CAPM asset pricing framework ([Sha64]). In this work, we extend this approach to integrate path-dependencies and optimise under a *dynamic* mean-variance criterion. We do so by simultaneously capturing path-dependent structure in the dynamics, while managing path-dependent variance through the lifetime of the trade. The general dynamic mean-variance optimisation can be framed as

$$\max_{(\xi_t)_{t \in [0, T]}} \mathbb{E} \left( \sum_{m=1}^d \int_0^T \xi_t^m dX_t^m \right) - \frac{\lambda}{2} \text{Var} \left( \sum_{m=1}^d \int_0^T \xi_t^m dX_t^m \right), \quad (1.4)$$

where the quantity

$$V_T := \sum_{m=1}^d \int_0^T \xi_t^m dX_t^m$$

is the strategy PnL at some future time  $T$ .

## Optimisation

Depending on the choice of underlying model, the optimisation method can be different - for example, reinforcement learning can in fact be used to simultaneously learn the model and the optimal strategy ([WZ19; BT23; Soo+23]). The authors in [KEA19] provide an extensive overview of alternate formulations and methods for mean-variance optimisation. In the case of the predict-then-optimise framework, such as (1.3), the mean-variance solution is given by

$$\xi_t^* = \frac{1}{\lambda} \Sigma_t^{-1} \mu_t \quad (1.5)$$

where  $\Sigma_t$  is the  $d \times d$  covariance matrix of asset returns at time  $t$  and  $\mu_t$  is the  $d \times 1$  vector of expected returns over the next time period  $t \in [0, 1]$ . This solution is highly intuitive and tractable. The performance of such a trading strategy is then mostly reliant on the predictive power of the signal that goes into the model, as well as the well-posedness of the covariance structure, leaving flexibility and responsibility in the hands of the trader. However, as has been highlighted in decades of literature, this framework is highly restrictive, such as the assumption of normally distributed returns and stationarity in the factor signal.

## 1.2 Summary of Contributions

Several stylised facts exhibited in asset prices are not represented in most classical dynamic optimisation frameworks. Many of these stylised facts, such as slow decaying autocorrelation and volatility clustering, are path-dependent properties and so it advantageous to incorporate path-dependence into any factor model. Due to such characteristics in financial time series data, the predict-then-optimise composition in classical factor models can be prone to asymmetric errors that are compounded through the optimisation process. Work from [CI22] and [Zha+21] have provided robust machine learning approaches to help account for this issue by bypassing the direct prediction phase. Alternatively, we are able to exploit such dynamics in the signal and asset and incorporate these not only to increase expected return, but to dynamically reduce variance of PnL. Due to the powerful properties arising in rough path theory we are able to simultaneously incorporate such dynamics into both the modelling framework and the optimisation phase at the same time, bypassing any explicit prediction.

In summary, our contribution is to extend the existing signature trading strategy framework first presented in the thesis of Perez ([Per20]), by deriving a closed form mean-variance optimal trading strategy for multiple assets that accounts for path-dependent dynamics between exogenous trading signals and the underlying assets, thereby providing a pathwise extension to classical factor models. Our framework is simple to implement and does not require heavy machinery while comparative machine learning methods may require extensive network building and hyperparameter tuning every time the trader wants to perform a new optimisation. This is possible due to the mathematical properties of the signature allowing us to linearise the objective function, meaning it is relatively straightforward to solve for an explicit closed-form solution. Sig-Trading is a one-model-fits-all type framework that is flexible enough to adapt to any type of underlying asset and market factor process. Once a linear functional  $\ell$  is obtained from past data samples it is straightforward to unravel this into an implementable trading strategy characterised by the number of units to buy/sell at each time point  $t \in [0, T]$ . Since the strategy is dynamic, as new data arrives the Sig-Trader will continuously compute the signature and update their position accordingly.

## 2 The Modelling Setup of the Sig-Factor Model

Rough path theory has provided many valuable tools, such as the Signature of a path, to help shift the focus from probabilistic to pathwise approaches when working with streams of data. The concept of the signature was first introduced in [Che57; Che77] and has played a crucial role in rough path theory in [Lyo98; LCL07; FV10; FH20]. Recent mathematical finance literature has benefited immensely from its universality property, which states that linear functionals on the signature are dense in the space of continuous functions on compact sets of paths (Theorem A.13). This result allows to approximate a trading strategy as a linear functional on the terms of the signature. Incorporating higher order path-dependent characteristics via the signature allows us to capture stylised facts of financial time series data, without requiring or imposing an explicit probability distribution on the future returns. In this section, we discuss the notion of a *Signature Trading Strategy*, first introduced in [LNA19b], and how this is incorporated into a *model-free* setting, as well as under the presence of exogenous market signals.

### 2.1 The Signature

In this section, we first recall definitions of path augmentations and how these are used in preceding results such as Theorem 2.10, which is crucial for our main result Theorem 3.1. Whilst we introduce fundamental definitions in Section 2.1, we have collected some basic concepts and definitions from rough path theory for convenience in Appendix A as they may aide in understanding of notations and technicalities throughout the paper. For a more thorough introduction to signatures, see for example [Gyu+13; CK16; LLN13; Fer21; LM22] for excellent articles focusing on intuition and

understanding in a practical setting. For specific applications of signatures in finance we refer the reader to [Bon+19; KLA20; ASS20; Bay+21; CGS22; Ald+22; DT23; IH23; WKM23].

Unless stated otherwise, the process  $(X_t)_{t \in [0, T]}$  is a continuous, stochastic process defined on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$ . We often refer to the path trajectories of the process  $(X_t)_{t \in [0, T]}$  as  $X : [0, T] \rightarrow \mathbb{R}^d$ , which we assume can be lifted to geometric rough paths (Definition A.6).

**Definition 2.1.** (Time reparameterisation). Let  $X : [0, T] \rightarrow \mathbb{R}^d$ ,  $\varphi : [0, T] \rightarrow [T_1, T_2]$  a non-decreasing surjection, then the re-parameterised path is denoted as  $X \circ \varphi =: X^\varphi : [T_1, T_2] \rightarrow \mathbb{R}^d$ .

**Definition 2.2.** (Add-time process). Often, we may wish to preserve the temporal structure of a path and so we keep the time parameterisation of the path  $X$  by defining a new process,  $\hat{X}$ . We denote the time-augmented process by  $\hat{X}_t = (t, X_t), t \in [0, T]$  such that  $(t, X_t) =: \hat{X} : [0, T] \rightarrow \mathbb{R}^{d+1}$ .

**Definition 2.3.** (Hoff Lead-Lag Process, [Hof06], [FHL16]). Let  $\hat{X} : [0, T] \rightarrow \mathbb{R}^{d+1}$  be the continuous time-augmented process of  $X$ , discretely sampled at  $t = t_0, \dots, t_{2N}$ . The *Hoff lead-lag transformed path* is defined as the piecewise linear interpolation  $\hat{X}^{LL} : [0, T] \rightarrow \mathbb{R}^{2(d+1)}$  such that

$$(\hat{X}_{t_i}^{LL})_{i=1}^{2N} = (\hat{X}_{t_i}^{\text{lead}}, \hat{X}_{t_i}^{\text{lag}})_{i=1}^{2N}$$

where

$$\begin{aligned} \hat{X}_t^{\text{lead}} &= \begin{cases} \hat{X}_{t_{k+1}}, & \text{if } t \in [2k, 2k+1] \\ \hat{X}_{t_{k+1}} + 2(t - (2k+1))(X_{t_{k+2}} - X_{t_{k+1}}), & \text{if } t \in [2k+1, 2k+\frac{3}{2}] \\ \hat{X}_{t_{k+2}}, & \text{if } t \in [2k+\frac{3}{2}, 2k+2], \end{cases} \\ \hat{X}_{t_j}^{\text{lag}} &= \begin{cases} \hat{X}_{t_i}, & \text{if } t \in [2k, 2k+\frac{3}{2}] \\ \hat{X}_{t_{k+1}} + 2(t - (2k+\frac{3}{2}))(X_{t_{k+1}} - X_{t_k}), & \text{if } t \in [2k+\frac{3}{2}, 2k+2]. \end{cases} \end{aligned}$$

*Remark 2.1.* There also exists a more intuitive and straightforward definition of a lead-lag process, however we decide not to opt for this version and instead consider the so-called *Hoff process* ([Hof06]), due to its fundamental properties in the case when we want to calculate the PnL of our trading strategy (Theorem 2.10). This is due to the powerful, non-trivial result proven in [FHL16] that states that the Itô integral of a function of a process  $X$  against itself, can be recovered via the components of the Hoff lead-lag transformation. This will be made more precise in Section 2.2.

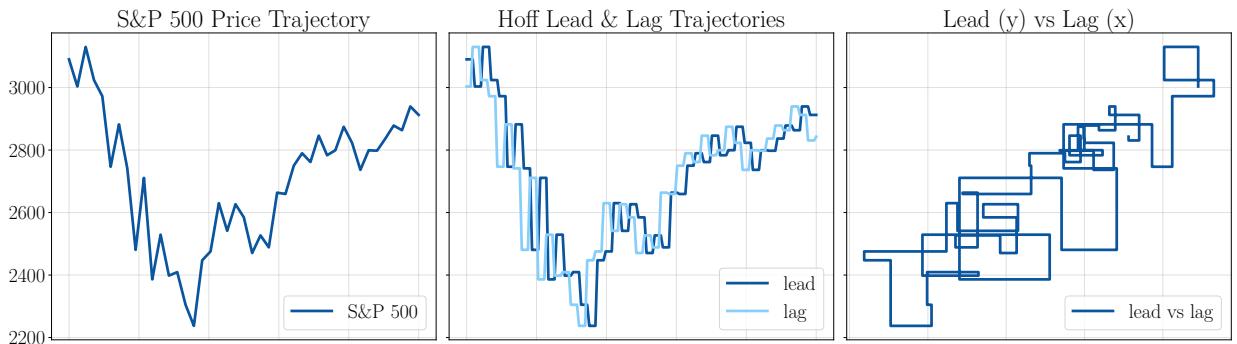


Figure 3: SPY ETF sample price trajectory between 02/03/2020-30/04/2020 (LHS) and its respective Hoff lead-lag transformation (Centre) and the lead vs the lag component (RHS).

**Notation:** We distinguish that any integral of the form  $\int f \circ dx$  refers to Stratonovich integration, meanwhile  $\int f dx$  will refer to Itô integration.

**Definition 2.4.** (Signature Transform). Let  $X : [0, T] \rightarrow \mathbb{R}^d \in C^{1-var}([0, T]; \mathbb{R}^d)$  be a (piecewise) smooth path. Let us define the simplex  $\Delta_T = \{(s, t) : 0 \leq s \leq t \leq T\}$ . The signature of  $X$  between fixed time  $s$  and  $t$  is a map

$$\begin{aligned}\mathbb{X} &: \Delta_T \rightarrow T((\mathbb{R}^d)) \\ (s, t) &\mapsto \mathbb{X}_{s,t} := (1, \mathbb{X}_{s,t}^1, \dots, \mathbb{X}_{s,t}^n, \dots)\end{aligned}$$

where the  $n$ -th order of the signature is defined as

$$\mathbb{X}_{s,t}^n := \int_{s < u_1 < \dots < u_n < t} dX_{u_1} \otimes \dots \otimes dX_{u_n} \in (\mathbb{R}^d)^{\otimes n}.$$

The signature is a  $T((\mathbb{R}^d))$ -valued process, which can be viewed as

$$\mathbb{X}_{0,T}^{<\infty} = \left( \underbrace{\mathbb{X}_{0,T}^{\emptyset}}_{=1}, \underbrace{\begin{pmatrix} \mathbb{X}_{0,T}^{\mathbf{1}} \\ \vdots \\ \mathbb{X}_{0,T}^{\mathbf{d}} \end{pmatrix}}_{\mathbb{X}_{0,T}^1}, \underbrace{\begin{pmatrix} \mathbb{X}_{0,T}^{\mathbf{11}} & \dots & \mathbb{X}_{0,T}^{\mathbf{1d}} \\ \vdots & \ddots & \vdots \\ \mathbb{X}_{0,T}^{\mathbf{d1}} & \dots & \mathbb{X}_{0,T}^{\mathbf{dd}} \end{pmatrix}}_{\mathbb{X}_{0,T}^2}, \dots \right).$$

The signature of a path can be truncated at any finite order  $N \in \mathbb{N}$ . We denote the truncated signature up to order  $N$  as

$$\begin{aligned}\mathbb{X}^{\leq N} &: \Delta_T \rightarrow T^{(N)}(\mathbb{R}^d) \\ (s, t) &\mapsto \mathbb{X}_{s,t} := (1, \mathbb{X}_{s,t}^1, \dots, \mathbb{X}_{s,t}^N).\end{aligned}$$

**Notation:** Throughout, we refer to the (un-truncated) signature between time 0 and time  $T$  as  $\mathbb{X}_{0,T}^{<\infty}$ . We may refer to  $\mathbb{X}_{0,T}^n$  as being the  $n$ -th *level* of the signature and  $\mathbb{X}_{0,T}^{\leq N}$  as the  $N$ -th *order* truncated signature.

**Notation:** We will use bold blue for any word  $\mathbf{w} \in \mathcal{W}(A_d)$ . Words are the multi-indices that can be thought of as the different multi-indices inside the tensor algebra. By using a specific word  $\mathbf{w}$ , this will often equate to referring to the specific multi-index associated to that word.

**Definition 2.5.** (Linear functionals on the tensor algebra). Note that there is a natural pairing between the extended tensor algebra  $T((\mathbb{R}^d))$  and its dual space  $T((\mathbb{R}^d)^*)$ , by which we denote

$$\langle \cdot, \cdot \rangle : T((\mathbb{R}^d)^*) \times T((\mathbb{R}^d)) \rightarrow \mathbb{R}$$

and is given by

$$\langle \ell, \mathbb{X} \rangle = \sum_{\mathbf{w} \in \mathcal{W}(A_d)} \ell_{\mathbf{w}} \mathbb{X}^{\mathbf{w}}$$

for  $\ell \in T((\mathbb{R}^d)^*)$ ,  $\mathbb{X} \in T((\mathbb{R}^d))$ . We make clear that there exists a canonical isomorphism  $(\mathbb{R}^d)^* \cong \mathbb{R}^d$  through the mapping  $(\mathbb{R}^d)^* \ni \langle \ell, \cdot \rangle \mapsto \ell \in \mathbb{R}^d$ .

## 2.2 Sig-Factor Model

In this section, we develop and collect the tools required in order to find optimal trading strategies with respect to a given objective criterion. The key ingredient is that we can approximate a trading strategy as a linear functional on the signature of the path, which is formulated in Theorem 2.8. Then, by embedding the *market factor process* (Definition 2.6) as a geometric rough path via

the signature and using Theorem 2.10, we alleviate most probabilistic restrictions that classical methods may have. Theorem 2.11 subsequently allows us to in fact bypass the Itô integral entirely and express the expected PnL as a linear functional on the expected Hoff lead-lag signature. Hence, all that remains to be done in order to optimise under a given criterion (i.e mean-variance), is to solve an optimisation problem - equating to solving a system of linear equations.

We consider the scenario, as is often the case in practice, that we observe a much larger market state than just the asset trajectories themselves, i.e we observe the natural filtration of a new process that consists of the original price trajectories, enriched with some new market factors  $f_t = (f_t^1, \dots, f_t^N)$ . A *market factor* can be any un-tradable characteristic, classified as a stochastic process  $(f_t)_{t \in [0, T]}$  that may be used in tandem with our framework, in order to enrich the market state and provide predictability. In practice, factors are chosen and understood as a driving signal of the underlying process. Unless stated otherwise, we assume the underlying asset process  $(X_t)_{t \in [0, T]}$  is a continuous, stochastic process whose path trajectories can be lifted to geometric rough paths (Definition A.6). We note that, while our price process may be a semi-martingale, we do not require any restriction on the nature of the exogenous market signals, where as traditional factor models may require stationarity.

**Definition 2.6.** (Market factor process). Let  $X = (X_t)_{t \in [0, T]}$  be a  $d$ -dimensional tradable underlying asset process and  $\{f^i\}_{i=1}^N$  are  $N$  un-tradable exogenous trading signals. Then we define the (time-augmented) *market factor process* as the  $(1 + d + N)$ -dimensional process

$$\hat{Z}_t := (t, X_t, f_t)$$

that induces the natural filtered probability space  $(\Omega, \mathcal{F}^Z, (\mathcal{F}_t^Z)_{t \in [0, T]}, \mathbb{P})$ , where the filtration of  $X$  satisfies  $\mathcal{F}^X \subseteq \mathcal{F}^Z$  such that  $X$  is driven by  $\{f^i\}_{i=1}^N$ . Throughout the remainder of this paper, we maintain such assumptions on  $\hat{Z}$ . We define as the space of all *market factor trajectories* for given assets  $X$  and factors  $f$ .

$$\mathcal{Z}_{0,T}^f := \{ \hat{Z}_t = (t, X_t, f_t) \mid X : [0, T] \rightarrow \mathbb{R}^d, \text{ and } f : [0, T] \rightarrow \mathbb{R}^N \text{ and } Z_0 = (0, 1, 1) \}$$

**Definition 2.7.** (Exogenous Signature Trading Strategy). Let  $\hat{Z}$  be a market factor process as defined in Definition 2.6. Let  $\xi = (\xi_t)_{t \in [0, T]}$  be an adapted,  $\mathcal{F}_t^Z$ -predictable and integrable strategy such that  $\int_0^T \xi_s^2 ds < \infty$ . If the strategy  $\xi$  is then a function of the market state, that is  $\xi_t = \phi(\hat{Z}_{0,t})$ , a continuous function of the previous *market factor trajectory* up to time  $t$ . We can extend  $\xi$  to be a *signature trading strategy*, such that

$$\xi_t^m = \langle \ell_m, \hat{Z}_{0,t} \rangle \approx \phi(t, Z_{0,t}), \quad \forall m = 1, \dots, d.$$

We define the space of all *exogenous signature trading strategies* with respect to market factors  $f$  as

$$\mathcal{A}^{f,\text{sig}} := \left\{ (\xi)_{t \in [0, T]} = (\langle \ell, \hat{Z}_{0,t} \rangle)_{t \in [0, T]} \mid \int_0^T \xi_s^2 ds < \infty, \quad \forall \hat{Z}_{0,T} \in \hat{\mathcal{Z}}_{0,T}^f \text{ and } \ell \in T((\mathbb{R}^{N+d+1})^*) \right\}.$$

*Remark 2.2.* If instead we wanted to trade *endogenously* without any trading signals, we would simply take the market factors to be the null-process  $\emptyset$ , such that  $\hat{X} = \hat{Z}$ , the results would still hold.

Now that we have defined the framework required in order to trade a signature trading strategy embedded with market factors, we can state a key result in order to combine the importance of the Hoff process in Theorem 2.10, that allows us to explicitly state the PnL of the exogenous sig-trader without the need for an integral at all.

**Lemma 2.8.** Let  $\mathcal{Z}_{0,T}^f \subset K \subset C^{1-var}([0,T], \mathbb{R}^d)$  be a compact set of market factor trajectories. Then for any exogenous signature trading strategy  $\xi = \phi(\hat{Z}_{0,T})$  that acts on paths in  $K$  and for every  $\epsilon > 0$ , there exists a linear functional  $\ell \in T((\mathbb{R}^{N+d+1})^*)$ , such that

$$\sup_{Z \in K} \|\phi(\hat{Z}_{0,T}) - \langle \ell, \hat{Z}_{0,T} \rangle\| < \epsilon.$$

*Proof.* We can see that this result follows from the universal approximation of continuous functions on paths by linear functionals acting on the signature (Theorem A.13).  $\square$

**Definition 2.9.** (Trading Strategy PnL). Let  $\hat{Z} = (t, X_t, f_t)$  be a market factor process with  $X$  a tradable underlying process and  $f$  an untradable signal process. If  $\xi \in \mathcal{A}^{f,\text{sig}}$  is a linear signature trading strategy such that  $\xi_t^m = \langle \ell_m, \hat{Z}_{0,t} \rangle$ , for each asset  $m = 1, \dots, d$ , then we define the PnL of the signature trading strategy between time 0 and time  $T$  as

$$V_T = \sum_{m=1}^d \int_0^T \langle \ell_m, \hat{Z}_{0,t} \rangle dX_t^m \quad (2.1)$$

where the integral is understood in the Itô sense.

It is crucial to distinguish the difference between the Itô integral used in this definition versus the Stratonovich integral in (2) of Example A.3. If that integral was in fact an Itô integral then we could describe the above definition of PnL directly in terms of the add-time signature  $\hat{Z}_{0,t}$ , however this is unfortunately not the case! So in order to develop a more friendly version of  $V_T$ , we require the following theorem involving the *Hoff lead-lag process* as seen in Definition 2.3.

**Theorem 2.10.** (Recovery of Itô Integral using the Hoff process, (Theorem 5.1, [FHL16])). Let  $X = (X_t)_{t \in [0,T]}$  be a stochastic process on the filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0,T]}, \mathbb{P})$ . Suppose we observe piecewise smooth streams of  $X$  that are discretely sampled over a sequence of times  $\{t_i\}_{i=0}^N$ . Let  $\hat{X}^{LL} = (\hat{X}_t^{\text{lead}}, \hat{X}_t^{\text{lag}})_{t \in [0,T]}$  be the associated observed Hoff lead-lag transform of  $\hat{X}$  as defined in Definition 2.3. Let  $\phi = (\phi^1, \dots, \phi^d)$  be continuous functions acting on paths of  $X$ , then we have that

$$\begin{aligned} \sum_{m=1}^d \int_0^T \phi^m(\hat{X}_t^{\text{lag}}) d\hat{X}_t^{m,\text{lead}} &\rightarrow \int_0^T \phi(X_t) dX_t := \sum_{m=1}^d \int_0^T \phi^m(X_t) dX_t^m \\ &\text{as } \max_{t_i, t_{i+1}} |t_{i+1} - t_i| \rightarrow 0, \end{aligned}$$

in either probability or  $L^p$ -norm.

*Proof.* The basic idea of the proof is that the areas between the *lead* and *lag* components of the Hoff lead-lag process, (captured by  $\hat{X}^{LL}$ ), introduce a correction factor in the stochastic integral limit, consequently allowing us to recover the Itô, not Stratonovich, integral.  $\square$

This result allows us to show that applying a function to the observed lagged path, against the observed leading path, we recover the true Itô integral against the stochastic process  $X$  as the mesh size goes to zero. In this sense, we see that we are able to approximate asymptotically the true Itô integral as instead an integral of the observed, Hoff lead-lag stream.

*Remark 2.3.* We can clearly see the analogy with Theorem 2.10 and our definition of PnL in Definition 2.9. By transforming our asset price data via the Hoff lead-lag process, now regarded as a geometric rough path, the quadratic variation of the underlying process  $X$  naturally arises. The path-dependent concept of volatility has been researched extensively ([BCD98; GJR14; JL20; GL22]) and this is explicitly embedded in the second order of the signature of the Hoff process.

So not only does this result allow us to capture the path-dependency of volatility, which is mostly endogenous, we can enrich this with other path-dependent market factors.

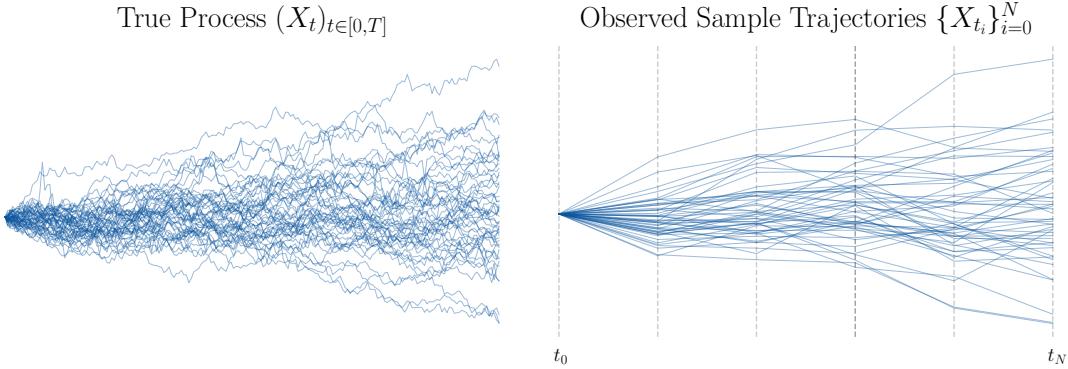


Figure 4: Distinguishing between the true stochastic process  $X$  and the discretely sampled observed process for which we define the Hoff lead-lag process.

We can see from Figure 4 that in order for Theorem 2.10 to hold, we require frequent sampling such that the distance between observations is small. A natural next step is to consider the case where our strategy  $\xi$  is in fact a *signature trading strategy*, i.e.  $\xi_t^m = \langle \ell_m, \hat{Z}_{0,t} \rangle$  for each asset  $m = 1, \dots, d$ . Moreover, how can we find an expression for the PnL  $V_T$  in (2.1), in the Itô integral sense, using Theorem 2.10? We will in fact extend this idea to a much more powerful result, under the presence of exogenous market signals.

**Theorem 2.11.** (*PnL of a d-asset signature trading strategy under exogenous signal*). Let  $\hat{Z} := (t, X_t, f_t)_{t \in [0,T]}$  be the market factor process, where  $X$  is a  $d$ -dimensional tradable stochastic process and  $f$  is a  $N$ -dimensional un-tradable factor process. Let  $\ell_1, \dots, \ell_d \in T((\mathbb{R}^{N+d+1})^*)$  and define our trading strategy as  $\xi_t^m = \langle \ell_m, \hat{Z}_{0,t}^{<\infty} \rangle$  for  $m = 1, \dots, d$ . Then, we have that the PnL of the strategy between time 0 and time  $T$  can be represented as

$$V_T = \sum_{m=1}^d \int_0^T \langle \ell_m, \hat{Z}_{0,t}^{<\infty} \rangle dX_t^m \approx \sum_{m=1}^d \int_0^T \langle \ell_m, \hat{Z}_{0,t}^{\text{lag}, <\infty} \rangle dX_t^{m,\text{lead}} = \sum_{m=1}^d \langle \ell_m \mathbf{f}(m), \hat{Z}_{0,T}^{LL, <\infty} \rangle \quad (2.2)$$

where  $\mathbf{f}(m) : \{1, \dots, d\} \rightarrow \mathcal{W}(A_{2(N+d+1)})$  is a shift operator which is defined as  $\mathbf{f}(m) = \pi(e_{m+N+d+2}^*)$  where  $\pi : T((\mathbb{R}^{2(N+d+1)})^*) \rightarrow \mathcal{W}(A_{2(N+d+1)})$  the canonical isomorphism between the dual space  $T((\mathbb{R}^{2(N+d+1)})^*)$  and the space of all words  $\mathcal{W}(A_{2(N+d+1)})$ .

Now, due to the linearity of expectation, we are able to represent the *expected PnL* at time  $T$  as the following:

$$\mathbb{E}(V_T) = \sum_{m=1}^d \langle \ell_m \mathbf{f}(m), \mathbb{E}(\hat{Z}_{0,T}^{LL, <\infty}) \rangle. \quad (2.3)$$

where  $\mathbb{E}(\hat{Z}_{0,T}^{LL, <\infty})$  is the expected signature of the Hoff lead-lag transformation of  $\hat{Z}$ .

*Proof.* Given in Appendix B.1. □

**Remark 2.4.** This result allows us to express any integral of a linear functional on the signature of a time-augmented market factor path as a newly defined linear functional on the time-augmented lead-lag market factor path instead. We must note that the authors in ([LNA19a], Lemma 3.11), construct a version in the specific case when  $d = 1$ ,  $f = \emptyset$  and so  $\mathbf{f}(m) = \mathbf{4}$ . Here, we extend this case to allow for any exogenous market information, as well as the multi dimensional case for  $d > 1$ .

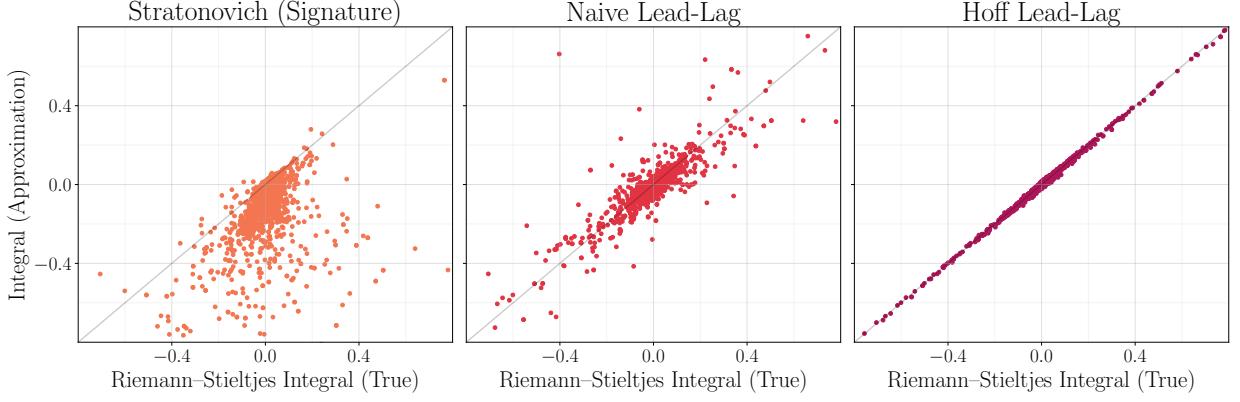


Figure 5: A comparison between different approximations of the true Itô PnL.

*Remark 2.5.* Note that  $\mathbf{f}(m) : \{1, \dots, d\} \rightarrow \mathcal{W}(A_{2(N+d+1)})$  is simply just the shift operator that allows any multi-index  $(j_1, \dots, j_n) \in \{1, \dots, d\}^n$  for the original signature terms to be transformed into a multi-index for the signature terms of the  $2(N+d+1)$ -dimensional time-augmented lead-lag process.

**Example 2.1.** Define the 4-dimensional market factor process  $\hat{Z} := (t, X^1, X^2, f^1)$  for two assets and one corresponding factor. Consider the case where our signature trading strategy is truncated at level  $M = 2$ , then the truncated expected signature,  $\mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, \leq M})$  will have

$$|\mathcal{W}_{d+N+1}^M| = |\mathcal{W}_4^2| = \sum_{k=0}^2 4^k = 21$$

terms. Notably, the 21 *words* associated with these signature terms  $\mathcal{W}_4^2$  are defined as:

$$\mathcal{W}_4^2 = \{\emptyset, \mathbf{0}, \mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{00}, \mathbf{01}, \mathbf{02}, \mathbf{03}, \mathbf{10}, \mathbf{11}, \mathbf{12}, \mathbf{13}, \mathbf{20}, \mathbf{21}, \mathbf{22}, \mathbf{23}, \mathbf{30}, \mathbf{31}, \mathbf{32}, \mathbf{33}\}$$

and so any linear functional associated with the truncated expected signature,  $\mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, \leq M})$ , will have at most 21 non-zero terms. We can see that if we concatenate a linear functional  $\ell$  with the letter  $\mathbf{f}(m)$ , then, it must be applied to the truncated signature  $\hat{\mathbb{Z}}_{0,T}^{LL, \leq M+1}$  and the words associated with  $\ell\mathbf{f}(m)$  will be  $\mathbf{wf}(m)$  where  $\mathbf{w} \in \mathcal{W}_4^2$ . Throughout, we will refer to terms of the truncated signature via *words* such as  $\mathbf{w}, \mathbf{v} \in W_{d+N+1}^M$ .

**Lemma 2.12.** (*Variance of the PnL of a signature trading strategy under exogenous signal*). Let  $X$  be a  $d$ -dimensional tradable stochastic process and let  $f$  be a  $N$ -dimensional un-tradable factor process. Define  $\hat{Z}_t := (t, X_t, f_t)$  as the market factor process and  $\ell_1, \dots, \ell_d \in T((\mathbb{R}^{N+d+1})^*)$  and define our trading strategy as  $\langle \ell_m, \hat{\mathbb{Z}}_{0,s}^{<\infty} \rangle$  for  $m = 1, \dots, d$ . Let the expected PnL of the strategy between time 0 and time  $T$  be defined as in (2.3). Then the Variance of the PnL at time  $T$  is defined as

$$\text{Var}(V_T) = \sum_{m=1}^d \sum_{n=1}^d \langle \ell_m \mathbf{f}(m) \sqcup \ell_n \mathbf{f}(n), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL}) \rangle - \langle \ell_m \mathbf{f}(m), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL}) \rangle \langle \ell_n \mathbf{f}(n), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL}) \rangle.$$

where  $\sqcup$  is the shuffle product defined in Definition A.3.

*Proof.* This results follows simply from Theorem 2.11 and the fact that variance of a random

variable is defined as  $\text{Var}(V_T) = \mathbb{E}(V_T^2) - (\mathbb{E}(V_T))^2$ , where  $\mathbb{E}(V_T)$  is defined in (2.3) and

$$\mathbb{E}(V_T^2) = \sum_{m=1}^d \sum_{n=1}^d \langle \ell_m \mathbf{f}(m) \sqcup \ell_n \mathbf{f}(n), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL}) \rangle \quad (2.4)$$

$$\mathbb{E}(V_T)^2 = \sum_{m=1}^d \sum_{n=1}^d \langle \ell_m \mathbf{f}(m), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL}) \rangle \langle \ell_n \mathbf{f}(n), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL}) \rangle. \quad (2.5)$$

□

### 3 Main Result

In this section we derive our main result, which is an explicit and concise expression for the optimal signature trading strategy in the presence of exogenous market signals considering a pathwise version of the classical mean-variance criterion. Using the trading strategy expected PnL and variance of PnL as defined in Chapter 2, we show that the path-dependent mean-variance optimisation problem is convex in the weights of the linear functionals  $\ell_1, \dots, \ell_d$ .

**Theorem 3.1.** (*Optimal Signature Trading Strategy*). Denote  $T \in \mathbb{N}$  the terminal time. Let  $X$  be a  $d$ -dimensional tradable stochastic process and let  $f$  be a  $N$ -dimensional un-tradable factor process. Define  $\hat{Z}_t := (t, X_t, f_t)$  as the market factor process. Define a signature trading strategy  $\xi_t^m$  through a linear functional on the signature of the market factor process, i.e.  $\xi_t^m = \langle \ell_m, \hat{\mathbb{Z}}_{0,t} \rangle$ . Then, for a given truncation level  $M$ , the mean-variance optimal signature trading strategy  $\ell^* = (\ell_1^*, \dots, \ell_d^*)$ ,  $\ell_m^* \in T^{(M)}((\mathbb{R}^{N+d+1})^*)$ , satisfies

$$\ell_m^* := \underset{\substack{\ell_m \in T^{(M)}((\mathbb{R}^{N+d+1})^*) \\ \text{Var}(V_T) \leq \Delta}}{\operatorname{argmax}} \sum_{m=1}^d \left\langle \ell_m \mathbf{f}(m), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, < \infty}) \right\rangle, \quad \forall m \in \{1, \dots, d\} \quad (3.1)$$

and is given by

$$\langle \ell_m^*, e_{\mathbf{w}} \rangle = \frac{((\Sigma^{\text{sig}})^{-1} \mu^{\text{sig}})_{\mathbf{w}\mathbf{f}(m)}}{2\lambda}, \quad m \in \{1, \dots, d\}, \mathbf{w} \in \mathcal{W}_{N+d+1}^M$$

where the variance-scaling parameter  $\lambda$  is given by

$$\lambda = \frac{1}{2\sqrt{\Delta}} \left( \sum_{m=1}^d \sum_{n=1}^d \sum_{\mathbf{w} \in \mathcal{W}_{N+d+1}^M} \sum_{\mathbf{v} \in \mathcal{W}_{N+d+1}^M} ((\Sigma^{\text{sig}})^{-1} \mu^{\text{sig}})_{\mathbf{w}\mathbf{f}(m)} ((\Sigma^{\text{sig}})^{-1} \mu^{\text{sig}})_{\mathbf{v}\mathbf{f}(n)} \Sigma_{\mathbf{w}\mathbf{f}(m), \mathbf{v}\mathbf{f}(n)}^{\text{sig}} \right)^{\frac{1}{2}}.$$

We define the “Signature PnL attribution” as the  $d \cdot |\mathcal{W}_{N+d+1}^M|$ -length vector  $\mu^{\text{sig}} = (\mu_1^{\text{sig}}, \dots, \mu_d^{\text{sig}})^\top$  as

$$\mu_{\mathbf{w}\mathbf{f}(m)}^{\text{sig}} = \left\langle \mathbf{w}\mathbf{f}(m), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, < \infty}) \right\rangle, \quad \forall \mathbf{w} \in \mathcal{W}_{N+d+1}^M, m \in \{1, \dots, d\} \quad (3.2)$$

and the “Signature PnL covariances” as the  $d \cdot |\mathcal{W}_{N+d+1}^M| \times d \cdot |\mathcal{W}_{N+d+1}^M|$  matrix  $\Sigma^{\text{sig}}$  as

$$\Sigma_{\mathbf{w}\mathbf{f}(m), \mathbf{v}\mathbf{f}(n)}^{\text{sig}} = \left\langle \mathbf{w}\mathbf{f}(m) \sqcup \mathbf{v}\mathbf{f}(n), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, < \infty}) \right\rangle - \left\langle \mathbf{w}\mathbf{f}(m), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, < \infty}) \right\rangle \left\langle \mathbf{v}\mathbf{f}(n), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, < \infty}) \right\rangle \quad (3.3)$$

for all  $\mathbf{w}, \mathbf{v} \in \mathcal{W}_{N+d+1}^M$  and  $m, n \in \{1, \dots, d\}$ .

*Proof.* In order to solve the constrained optimisation problem (3.1), we can introduce the Lagrangian

$$\mathcal{L}(\ell_1, \dots, \ell_d, \lambda) = \mathbb{E}(V_T) - \lambda(\text{Var}(V_T) - \Delta)$$

and we are interested in finding saddle points  $\ell_1^*, \dots, \ell_d^* \in T^{(M)}((\mathbb{R}^{N+d+1})^*)$ ,  $\lambda^0 \in \mathbb{R}$  that satisfy  $\nabla \mathcal{L}(\ell_1^*, \dots, \ell_d^*, \lambda) = 0$ . For this purpose recall Theorem 2.11, where the expected terminal PnL is given by (2.3). Furthermore, we can decompose the variance of the terminal PnL, given in equations (2.4) and (2.4). Next, we can compute for each asset  $m \in \{1, \dots, d\}$  and each word  $\mathbf{w} \in \mathcal{W}_{N+d+1}^M$ , the gradients of (2.3), (2.4) and (2.5) with respect to  $\langle \ell_m, e_{\mathbf{w}} \rangle$

$$\begin{aligned} \frac{\partial \mathbb{E}(V_T)}{\partial \langle \ell_m, e_{\mathbf{w}} \rangle} &= \langle \mathbf{wf}(m), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL,<\infty}) \rangle \\ \frac{\partial \mathbb{E}(V_T^2)}{\partial \langle \ell_m, e_{\mathbf{w}} \rangle} &= 2 \sum_{n=1}^d \sum_{\mathbf{v} \in \mathcal{W}_{N+d+1}^M} \langle \ell_n, e_{\mathbf{v}} \rangle \langle \mathbf{wf}(m) \sqcup \mathbf{vf}(n), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL,<\infty}) \rangle \\ \frac{\partial \mathbb{E}(V_T)^2}{\partial \langle \ell_m, e_{\mathbf{w}} \rangle} &= 2 \sum_{n=1}^d \sum_{\mathbf{v} \in \mathcal{W}_{N+d+1}^M} \langle \ell_n, e_{\mathbf{v}} \rangle \langle \mathbf{wf}(m), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL,<\infty}) \rangle \langle \mathbf{vf}(n), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL,<\infty}) \rangle \end{aligned}$$

Using these computed gradients, we can compute the gradient of the Lagrangian with respect to  $\langle \ell_m, e_{\mathbf{w}} \rangle$

$$\frac{\partial \mathcal{L}(\ell_1, \dots, \ell_d, \lambda)}{\partial \langle \ell_m, e_{\mathbf{w}} \rangle} = \langle \mathbf{wf}(m), b \rangle - 2\lambda \sum_{n=1}^d \sum_{\mathbf{v} \in \mathcal{W}_{N+d+1}^M} \langle \ell_n, e_{\mathbf{v}} \rangle \Sigma_{\mathbf{wf}(m), \mathbf{vf}(n)}^{\text{sig}}$$

where we define  $\mu^{\text{sig}} = (\mu_1^{\text{sig}}, \dots, \mu_d^{\text{sig}})^\top$  as

$$\mu_{\mathbf{wf}(m)}^{\text{sig}} = \langle \mathbf{wf}(m), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL,<\infty}) \rangle, \quad \forall \mathbf{w} \in \mathcal{W}_{N+d+1}^M, m \in \{1, \dots, d\}$$

and the matrix  $\Sigma^{\text{sig}}$  as

$$\Sigma_{\mathbf{wf}(m), \mathbf{vf}(n)}^{\text{sig}} = \langle \mathbf{wf}(m) \sqcup \mathbf{vf}(n), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL,<\infty}) \rangle - \langle \mathbf{wf}(m), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL,<\infty}) \rangle \langle \mathbf{vf}(n), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL,<\infty}) \rangle$$

for all  $\mathbf{w}, \mathbf{v} \in \mathcal{W}_{N+d+1}^M$  and  $m, n \in \{1, \dots, d\}$ .

Using the first order conditions, observe that we obtain the system of linear equations

$$\mu_{\mathbf{wf}(m)}^{\text{sig}} = 2\lambda \sum_{n=1}^d \sum_{\mathbf{v} \in \mathcal{W}_{N+d+1}^M} \langle \ell_n, e_{\mathbf{v}} \rangle \Sigma_{\mathbf{wf}(m), \mathbf{vf}(n)}^{\text{sig}}, m \in \{1, \dots, d\}, \mathbf{w} \in \mathcal{W}_{N+d+1}^M. \quad (3.4)$$

Under truncation, we find that (3.4) is a system of

$$\begin{aligned} d_M &= |I| \cdot |\mathcal{W}_{N+d+1}^M| \\ &= d \sum_{k=0}^M (N+d+1)^k \\ &= (N+d+1)^{M+1} - 1 \end{aligned}$$

equations and  $d_M$  unknowns. Hence, assuming that  $\Sigma^{\text{sig}}$  is invertible, we can solve (3.4) and obtain for  $m \in \{1, \dots, d\}$ ,  $\mathbf{w} \in \mathcal{W}_{N+d+1}^M$

$$\langle \ell_m^*, e_{\mathbf{w}} \rangle = \frac{((\Sigma^{\text{sig}})^{-1} \mu^{\text{sig}})_{\mathbf{wf}(m)}}{2\lambda} \quad (3.5)$$

where we assumed by complementary slackness (KKT) that  $\lambda \neq 0$ . We can then substitute (3.5) into the variance constraint to obtain two solutions for  $\lambda$

$$\begin{aligned}\lambda_{\pm} &= \pm \frac{1}{2\sqrt{\Delta}} \left( \sum_{m=1}^d \sum_{n=1}^d \sum_{\mathbf{w} \in \mathcal{W}_{N+d+1}^M} \sum_{\mathbf{v} \in \mathcal{W}_{N+d+1}^M} \langle \ell_m, e_{\mathbf{w}} \rangle \langle \ell_n, e_{\mathbf{v}} \rangle \Sigma_{\mathbf{w}\mathbf{f}(m), \mathbf{v}\mathbf{f}(n)}^{\text{sig}} \right)^{\frac{1}{2}} \\ &= \pm \frac{1}{2\sqrt{\Delta}} \left( \sum_{m=1}^d \sum_{n=1}^d \sum_{\mathbf{w} \in \mathcal{W}_{N+d+1}^M} \sum_{\mathbf{v} \in \mathcal{W}_{N+d+1}^M} ((\Sigma^{\text{sig}})^{-1} \mu^{\text{sig}})_{\mathbf{w}\mathbf{f}(m)} ((\Sigma^{\text{sig}})^{-1} \mu^{\text{sig}})_{\mathbf{v}\mathbf{f}(n)} \Sigma_{\mathbf{w}\mathbf{f}(m), \mathbf{v}\mathbf{f}(n)}^{\text{sig}} \right)^{\frac{1}{2}}\end{aligned}$$

Hence, we obtain the solution

$$\langle \ell_m^*, e_{\mathbf{w}} \rangle = \frac{(\Delta)^{\frac{1}{2}} ((\Sigma^{\text{sig}})^{-1} \mu^{\text{sig}})_{\mathbf{w}\mathbf{f}(m)}}{\left( \sum_{m,n \in \{1, \dots, d\}} \sum_{\mathbf{w}, \mathbf{v} \in \mathcal{W}_{N+d+1}^N} ((\Sigma^{\text{sig}})^{-1} \mu^{\text{sig}})_{\mathbf{w}\mathbf{f}(m)} ((\Sigma^{\text{sig}})^{-1} \mu^{\text{sig}})_{\mathbf{v}\mathbf{f}(n)} \Sigma_{\mathbf{w}\mathbf{f}(m), \mathbf{v}\mathbf{f}(n)}^{\text{sig}} \right)^{\frac{1}{2}}}$$

for  $m \in \{1, \dots, d\}$ ,  $\mathbf{w} \in \mathcal{W}_{N+d+1}^M$ . □

In the remainder of this section we first give an example how the main theorem can be used and then proceed to draw a parallel to the classical case and explain how our theorem extends classical formulas to the path-dependent case.

*Remark 3.1.* The matrix  $\Sigma^{\text{sig}}$  and vector  $\mu^{\text{sig}}$  are just placeholders for different terms and combinations of  $\mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, <\infty})$ . Hence, all we need in order to know what our optimal functional is, is the expected Hoff lead-lag signature, and this is straightforward to compute for reasonable orders of truncation and number of assets and factors. The following example aims to provide some practical intuition behind how the optimal strategy is calculated and the relation between the signature and words on the tensor algebra.

**Example 3.1.** Let us consider the case of when we have two assets  $X = (X^1, X^2)$ , one factor  $f$ , such that  $d = 2$ ,  $N = 1$ . Then we define the 4-dimensional market factor process  $\hat{Z} := (t, X, f)$ .

Let us fix the truncation level to be  $M = 2$ . We remark that Example 2.1 is of the same form, and so the 2nd level truncated signature  $\hat{\mathbb{Z}}_{0,t}^{\leq 2}$  has  $|\mathcal{W}_4^2| = 21$  terms, namely the associated words are:

$$\mathcal{W}_4^2 = \{\emptyset, \mathbf{0}, \mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{00}, \mathbf{01}, \mathbf{02}, \mathbf{03}, \mathbf{10}, \mathbf{11}, \mathbf{12}, \mathbf{13}, \mathbf{20}, \mathbf{21}, \mathbf{22}, \mathbf{23}, \mathbf{30}, \mathbf{31}, \mathbf{32}, \mathbf{33}\} \quad (3.6)$$

Hence, the optimal linear signature trading strategy will correspond to two linear functionals (one for each asset)  $\ell_1, \ell_2$ , each of length 21, defined as

$$\begin{aligned}\xi_t^1 &= \langle \ell_1, \hat{\mathbb{Z}}_{0,t}^{\leq 2} \rangle \\ \xi_t^2 &= \langle \ell_2, \hat{\mathbb{Z}}_{0,t}^{\leq 2} \rangle\end{aligned}$$

In the above theorem, we can see the vector  $\mu^{\text{sig}} = (\mu_1^{\text{sig}}, \mu_2^{\text{sig}})$  will be defined as follows:

$$\mu^{\text{sig}} := \left\{ \mathbb{E} \left( \hat{\mathbb{Z}}_{\mathbf{w}\mathbf{f}(m)}^{LL, \leq 3} \right) \right\}_{\mathbf{w} \in \mathcal{W}_4^2, m=1,2}.$$

Hence,  $\mu^{\text{sig}}$  will be a  $d \cdot |\mathcal{W}_{N+d+1}^M| = 2 \times 21 = 42$  length vector, containing elements of the expected lead-lag signature of order 3.

Recall that the shift operator  $\mathbf{f}(m)$  is defined for each asset  $m = 1, 2$  as:

$$\begin{aligned}\mathbf{f}(1) &= \mathbf{5} \\ \mathbf{f}(2) &= \mathbf{6}.\end{aligned}$$

which correspond to the 5th and 6th dimensions of the lead-lag process. Hence, we see that  $\mu_1^{\text{sig}}$  contains the expected lead-lag signature terms corresponding to the index of words

$$I_1 := \{\mathbf{5}, \mathbf{05}, \mathbf{15}, \mathbf{25}, \mathbf{35}, \mathbf{005}, \mathbf{015}, \mathbf{025}, \mathbf{035}, \mathbf{105}, \mathbf{115}, \mathbf{125}, \mathbf{135}, \mathbf{205}, \mathbf{215}, \mathbf{225}, \mathbf{235}, \mathbf{305}, \mathbf{315}, \mathbf{325}, \mathbf{335}\}$$

and  $\mu_2^{\text{sig}}$  contains the expected lead-lag signature terms corresponding to the index of words

$$I_2 := \{\mathbf{6}, \mathbf{06}, \mathbf{16}, \mathbf{26}, \mathbf{36}, \mathbf{006}, \mathbf{016}, \mathbf{026}, \mathbf{036}, \mathbf{106}, \mathbf{116}, \mathbf{126}, \mathbf{136}, \mathbf{206}, \mathbf{216}, \mathbf{226}, \mathbf{236}, \mathbf{306}, \mathbf{316}, \mathbf{326}, \mathbf{336}\},$$

such that we have:

$$\mu^{\text{sig}} = \left[ \begin{array}{c} \left\langle \mathbf{5}, \mathbb{E}(\hat{\mathbb{Z}}^{LL, \leq 3}) \right\rangle \\ \vdots \\ \left\langle \mathbf{335}, \mathbb{E}(\hat{\mathbb{Z}}^{LL, \leq 3}) \right\rangle \\ \left\langle \mathbf{6}, \mathbb{E}(\hat{\mathbb{Z}}^{LL, \leq 3}) \right\rangle \\ \vdots \\ \left\langle \mathbf{336}, \mathbb{E}(\hat{\mathbb{Z}}^{LL, \leq 3}) \right\rangle \end{array} \right] \quad \text{42 elements}$$

Intuitively, we can think of each element of  $\mu^{\text{sig}}$  as the expected *attribution* that each signature term has to a given assets future returns. For example, consider the term of the signature corresponding to the word  $\mathbf{3}$ , i.e  $\langle \mathbf{3}, \hat{\mathbb{Z}}_{0,T}^{\leq 2} \rangle$ , which corresponds to the increments of the factor signal, i.e

$$\langle \mathbf{3}, \hat{\mathbb{Z}}_{0,T}^{\leq 2} \rangle = \int_0^T \circ dZ_t^3$$

Then the expected *attribution* that this has on the increment of asset 1, can be defined as

$$\mathbb{E} \left( \int_0^T \langle \mathbf{3}, \hat{\mathbb{Z}}_{0,t}^{\leq 2} \rangle dX_t^1 \right) = \left\langle \mathbf{3}\mathbf{f}(1), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, \leq 3}) \right\rangle = \left\langle \mathbf{35}, \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, \leq 3}) \right\rangle = \mu_{\mathbf{35}}^{\text{sig}},$$

therefore the vector  $\mu^{\text{sig}}$  consists of expected PnL attribution for each of the 21 terms of the signature of the factor process that we observe.

Now, we consider the  $42 \times 42$  matrix  $\Sigma^{\text{sig}}$ , defined element-wise as

$$\Sigma_{\mathbf{wf}(m), \mathbf{vf}(n)}^{\text{sig}} = \left\langle \mathbf{wf}(m) \sqcup \mathbf{vf}(n), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, < \infty}) \right\rangle - \left\langle \mathbf{wf}(m), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, < \infty}) \right\rangle \left\langle \mathbf{vf}(n), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, < \infty}) \right\rangle$$

for all  $\mathbf{w}, \mathbf{v} \in \mathcal{W}_4^2$  and  $m, n = 1, 2$ .

Each element of this matrix represents a covariance term between the PnL *attributions* that we discussed previously. For example, let us observe an arbitrary element of  $\Sigma^{\text{sig}}$ . Let  $w = \mathbf{01}, v = \mathbf{23}, m = 1, n = 2$ . Then  $\mathbf{wf}(m) = \mathbf{015}, \mathbf{vf}(n) = \mathbf{236}$  and the corresponding element in the matrix is given as

$$\Sigma_{\mathbf{015}, \mathbf{236}}^{\text{sig}} = \left\langle \mathbf{015} \sqcup \mathbf{236}, \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, \leq 6}) \right\rangle - \left\langle \mathbf{015}, \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, \leq 3}) \right\rangle \left\langle \mathbf{236}, \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, \leq 3}) \right\rangle$$

where **015**  $\sqcup$  **236** is a sum of 20 different words in  $\mathcal{W}_4^6$ . We refer the reader also to Example A.2 for another example of the shuffle product. It is evident that, while our linear functional is only applied to the second order truncated signature, we require the sixth order signature in order to compute the covariance matrix, which can cause a computational bottleneck in practice.

Piecing this altogether, to obtain our optimal solution  $\ell^* = (\ell_1^*, \ell_2^*)$ , we have

$$\ell^* = (\Sigma^{\text{sig}})^{-1} \mu^{\text{sig}}$$

of which we obtain a 42-dimensional vector  $\ell^*$  consisting of two length 21 vectors  $\ell_1^*, \ell_2^*$ . We can then compute the trading strategy, for each time  $t$ , as

$$\begin{aligned}\xi_t^1 &= \langle \ell_1^*, \hat{\mathbb{Z}}_{0,t}^{\leq 2} \rangle \\ \xi_t^2 &= \langle \ell_2^*, \hat{\mathbb{Z}}_{0,t}^{\leq 2} \rangle.\end{aligned}$$

Note also how we can explicitly calculate the expected PnL and variance of the portfolio explicitly in this framework. Let  $\ell$  be the 42-length vector  $\ell = (\ell_1^*, \ell_2^*)$ , then we have

$$\begin{aligned}\mathbb{E}(V_T) &= \ell^\top \mu^{\text{sig}} \\ \text{Var}(V_T) &= \ell^\top \Sigma^{\text{sig}} \ell.\end{aligned}$$

### 3.1 Sig-Factor Model vs Classical Factor Model

Since  $\hat{Z}$  is embedded with market factors  $f$ , the expected lead-lag signature  $\mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, \leq \infty})$  will contain a wealth of path-dependent characteristics about our assets and how they are driven by the past price trajectory and the past trajectory of the market factors. At this point, we observe this solution is analogous to the classical framing of an optimal factor model under the mean-variance framework, seen in (1.3) and (1.5). In the classical factor model framework, for  $N$  factors  $f = f^1, \dots, f^N$ , to obtain the the  $m$ -th asset position at time  $t$ ,  $\pi_t^m$ , we have

$$\begin{aligned}\pi_t^m &= \left\langle \frac{1}{\lambda} (\Sigma^{-1} B)_m, f_t \right\rangle \\ &= \langle \beta_m, f_t \rangle\end{aligned}\tag{3.7}$$

where  $(\Sigma^{-1} B)_m$  is the  $m$ -th row of the  $d \times N$  matrix  $\Sigma^{-1} B$ . Therefore  $\beta_m := \frac{1}{\lambda} (\Sigma^{-1} B)_m$  represents a risk-weighted transformation of the coefficients used in the prediction step. This sequence of  $N$  coefficients are then applied to the factors via an inner product to obtain the position for the  $m$ -th asset,  $\pi_t^m \in \mathbb{R}$ . We can see just how similar this example is to the sig-factor model.

$$\begin{aligned}\xi_t^m &= \left\langle \frac{1}{2\lambda} ((\Sigma^{\text{sig}})^{-1} \mu^{\text{sig}})_m, \hat{\mathbb{Z}}_{0,t} \right\rangle \\ &= \langle \ell_m, \hat{\mathbb{Z}}_{0,t} \rangle\end{aligned}$$

Here, we have that the vector  $\mu^{\text{sig}}$  consists of the expected lead-lag signature PnL terms, for all  $|\mathcal{W}_{N+d+1}^M|$  terms in the  $M$ -th order truncated signature that end in the letter  $\mathbf{f}(m)$ , i.e..

$$\mu^{\text{sig}} = \left\{ \mathbb{E} \left( \hat{\mathbb{Z}}_{\mathbf{w}\mathbf{f}(m)}^{LL, \leq \infty} \right) \right\}_{\mathbf{w} \in \mathcal{W}_{N+d+1}^M, m \in \{1, \dots, d\}}.$$

and so the linear functional  $\ell_m$  is a risk-adjusted weighting of coefficients that are applied to the signature terms. A sensible question now to ask is - would we produce the same portfolio if the factors  $f = f^1, \dots, f^N$  in (3.7) were the terms of the signature? In this case, the answer is no. The sole reason for this is due to the *prediction* phase in the classical factor model which induces

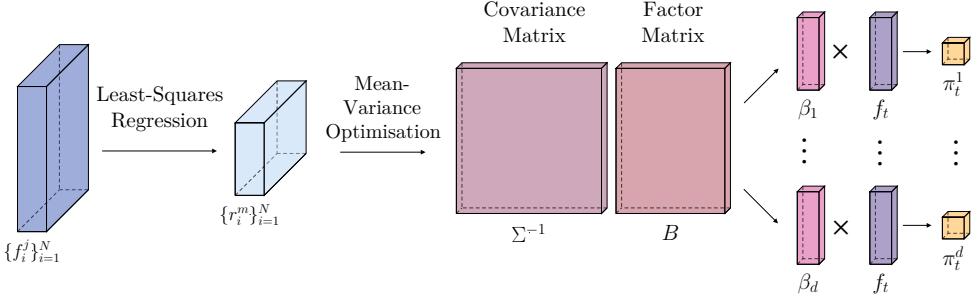


Figure 6: Classical Factor Model

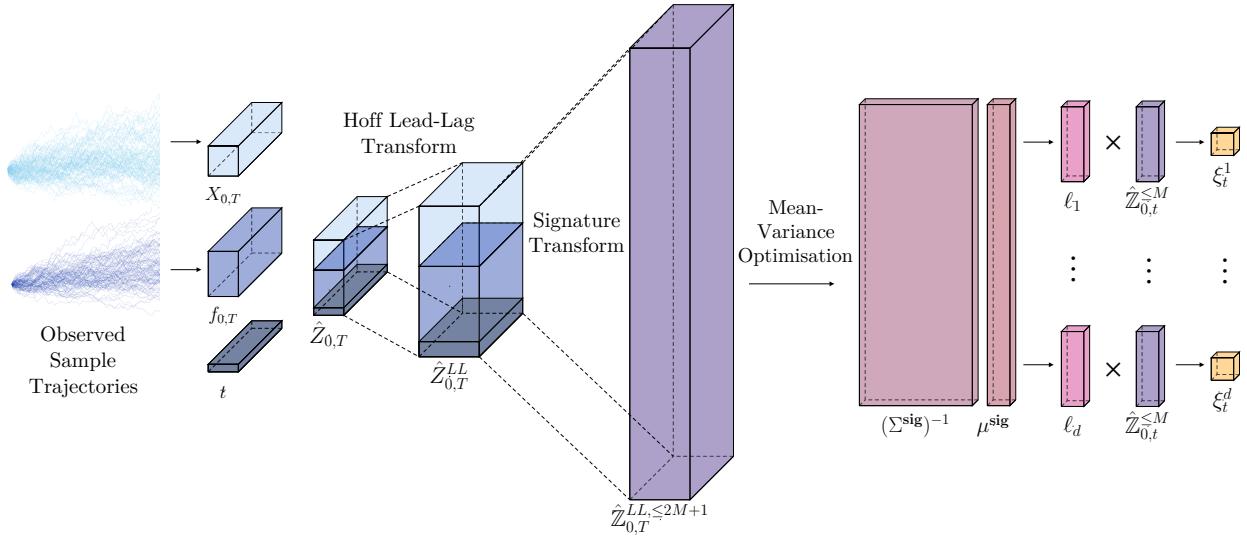


Figure 7: Sig-Factor Model

asymmetric errors that are compounded when applied to the covariance matrix, and subsequently observed factors at time  $t$ , meaning the linear functionals  $\beta_m$  and  $\ell_m$  would not be the same.

Due to powerful results from rough path theory, we are able to bypass the prediction phase by lifting and projecting our market factor process into a much higher dimensional space. Using Theorem 2.3, we can express the expected future PnL as a linear functional on the expected Hoff lead-lag signature and then perform the optimisation (3.1) in this much higher-dimensional space, without inducing any errors originating from a least-squares regression.

### 3.2 Optimal Static Portfolio

By design, a *Signature Trading Strategy*,  $\xi_t = \langle \ell, \hat{\mathbb{Z}}_{0,t} \rangle$ , is a dynamic strategy that continuously updates its position at time  $t$ , depending on the value of  $\hat{\mathbb{Z}}_{0,t}$ . However, we recall that the signature is defined at each level as

$$\hat{\mathbb{Z}}_{0,t} = (1, \hat{\mathbb{Z}}_{0,t}^1, \dots, \hat{\mathbb{Z}}_{0,t}^N, \dots)$$

where the  $k$ -th level has  $d^k$  elements. We observe that in fact the zero-th level of the signature is equal to 1, and so if we choose to only trade depending on this level of the signature, any linear functional  $\ell$  applied to 1, will just return a static position for all time  $t \in [0, T]$ . Therefore, for a  $d$ -asset portfolio, we obtain

$$\begin{aligned} \xi_t^1 &= \langle \ell_1, 1 \rangle = \ell_1 \in \mathbb{R}, \quad \forall t \in [0, T] \\ &\vdots \\ \xi_t^d &= \langle \ell_d, 1 \rangle = \ell_d \in \mathbb{R}, \quad \forall t \in [0, T]. \end{aligned}$$

Using Theorem 3.1, we have that  $\ell = (\ell_1, \dots, \ell_d) \in \mathbb{R}^d$ , where

$$\ell = \frac{1}{2\lambda}(\Sigma^{\text{sig}})^{-1}\mu^{\text{sig}}.$$

Since we are trading with respect to the zero-th order of the signature, then the only word  $\mathbf{w}$  that we are interested in is  $\mathbf{w} = \emptyset$ , therefore the number of words in our linear functional is  $|\mathcal{W}_{N+d+1}^0| = 1$ . Hence, we can define the  $d$ -dimensional vector

$$\begin{aligned}\mu_{\mathbf{wf}(m)}^{\text{sig}} &= \left\langle \mathbf{wf}(m), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, < \infty}) \right\rangle, \quad \forall \mathbf{w} \in \mathcal{W}_{N+d+1}^0, m \in \{1, \dots, d\} \\ &= \left\langle \mathbf{f}(m), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, \leq 1}) \right\rangle, \quad m \in \{1, \dots, d\}.\end{aligned}$$

We can observe that in fact, the elements of the  $d$ -dimensional vector  $\mu^{\text{sig}}$  are simply the expected returns of each asset, i.e. for element corresponding to the  $m$ -th asset,

$$\mu_m^{\text{sig}} = \left\langle \mathbf{f}(m), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, \leq 1}) \right\rangle = \mathbb{E} \left( \int_0^T dX_t^m \right) = \mathbb{E}(X_T^m) - \mathbb{E}(X_0^m).$$

Therefore, we have

$$\mu^{\text{sig}} = \begin{bmatrix} \mathbb{E}(X_T^1) - \mathbb{E}(X_0^1) \\ \vdots \\ \mathbb{E}(X_T^d) - \mathbb{E}(X_0^d) \end{bmatrix}$$

which corresponds to the expected returns vector. Likewise, we obtain the  $d \times d$  matrix  $\Sigma^{\text{sig}}$  as

$$\Sigma_{m,n}^{\text{sig}} = \left\langle \mathbf{f}(m) \sqcup \mathbf{f}(n), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, \leq 2}) \right\rangle - \left\langle \mathbf{f}(m), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, \leq 1}) \right\rangle \left\langle \mathbf{f}(n), \mathbb{E}(\hat{\mathbb{Z}}_{0,T}^{LL, \leq 1}) \right\rangle,$$

which naturally corresponds to the covariance matrix of the returns for each of the  $d$  assets.

We can see how this solution will in fact yield us the same results as the classical Markowitz portfolio, provided we use the same historical period to calculate the expected returns and covariances, of which we provide further evidence by constructing the Sig-Factor Model extension of the efficient frontier.

### 3.3 Sig-Factor Model Efficient Frontier

In Modern Portfolio Theory (MPT), first introduced in [Mar52], the mean-variance optimal portfolio can be represented via the efficient frontier, which contains all portfolios that have the maximal Sharpe ratio. In the Sig-factor model, we can also obtain an efficient frontier that represents the relationship between expected returns and variance. For a given signature trading strategy  $\xi_t^m = (\xi_t^1, \dots, \xi_t^d)$ , where  $\xi_t^m = \langle \ell_m, \hat{\mathbb{Z}}_{0,t} \rangle$ , we can explicitly define the expected PnL and variance of our portfolio in terms of the matrix  $\Sigma^{\text{sig}}$  and vector  $\mu^{\text{sig}}$ , as defined in Theorem 3.1, e.g we have

$$\begin{aligned}\mathbb{E}(V_T) &= \ell^\top \mu^{\text{sig}} \\ \text{Var}(V_T) &= \ell^\top \Sigma^{\text{sig}} \ell.\end{aligned}$$

Therefore, for any linear functional  $\ell$ , we can observe the expected PnL and variance corresponding to it.

In Figure 8, the LHS represents the level-3 Sig-trader efficient frontier curve for a portfolio consisting of the ETFs DBC/BND/VTI. This demonstrates the relationship between portfolio returns and variance, just as Markowitz had posed. The orange circle corresponds to the mean-variance

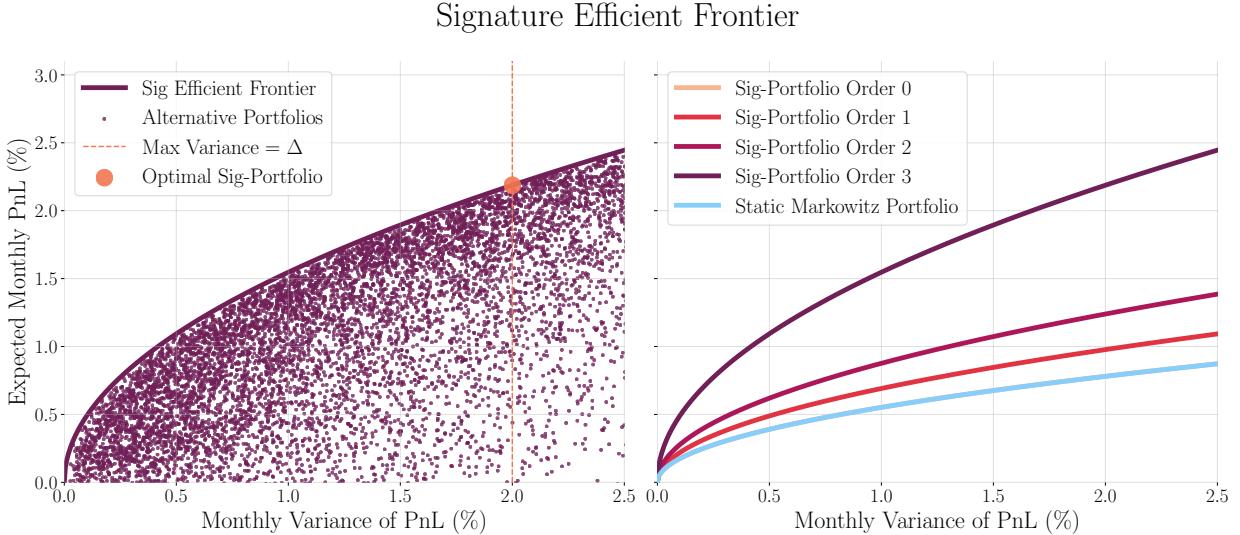


Figure 8: Sig-Factor Model Efficient Frontier for 3-asset portfolio of ETFs DBC/BND/VTI.

optimal sig-trading strategy  $\ell^*$  for a given maximum variance  $\Delta$ , according to the solution provided in Theorem 3.1. Each of the smaller dots then represent different choices of linear functional  $l$ , that are perturbations of the original strategy. We clearly observe that the Sig-trader portfolio maximises the risk-adjusted return (Sharpe ratio), for a given level of risk,  $\Delta$ .

On the RHS, we observe the efficient frontier curve for different orders of Sig-trader, between levels 0 to 3. Firstly, as eluded to previously, we can observe that the zero-th order Sig-trader is in fact identical to the static Markowitz portfolio, meaning that their frontier curves are not distinguishable from one another. We can also observe the improvement that is made to our expected returns, as we introduce more levels of the signature. This shouldn't be a surprise, as we know that more information about the dynamics is explained as we increase the level of truncation, which suggests that non-linear dependency structures are drivers of the future asset returns.

## 4 Implementation

Throughout this work, we use the package `signatory` ([KL21]), alongside PyTorch for calculating and performing functionality related to tensors and the signature transform. Other packages offering signature computation include `esig` and `iisignature`. The mean-variance Sig-Trading framework is very simple to implement and does not require heavy machinery while comparative machine learning methods may require extensive network building and hyperparameter tuning for every time the user wants to perform a new optimisation. Signature-trading is a one-model-fits-all type framework that is flexible enough to adapt to different types of signal and underlying asset process. The algorithm to find an optimal trading strategy can be found in Algorithm 1.

The method is also self-contained in the sense that it requires no user inputted assumptions such as expected returns or covariances and that they are inferred along with characteristics of the whole process within the algorithm itself. Once a linear functional  $\ell$  is obtained from past data samples it is straightforward to unravel this into an implementable trading strategy characterised by the number of units to buy/sell at each time point  $t \in [0, T]$ . Since the strategy is dynamic, as new data arrives the Sig-Trader will continuously compute the signature and update their position accordingly.

Whilst Sig-Trading is data driven and does not require direct probabilistic assumptions on the underlying model, just like most frameworks, there is versatility when deploying Sig-Trading in

practice as there does still remain quantities that need reliably estimating in the fitting procedure. The expected lead-lag signature could be noisy when calibrated through time and so we leave any questions of robustness (with respect to the fitting procedure) to future ongoing extensions of this project, as there remains question marks over how well defined such a solution is, especially in the context of rapidly changing regimes such as in financial markets. In step (4) of Algorithm 1, we suggest taking a Monte Carlo approach by calculating the empirical expected lead-lag signature, such that  $\mathbb{E}(\hat{\mathbb{Z}}^{LL, \leq N}) = \frac{1}{M} \sum_{i=1}^M \hat{\mathbb{Z}}_{t \in [0, T]}^{LL, \leq N}$ . However, there remains a multitude of alternative ways to approach this for some given sample/fitting data, such as cross validation techniques, in order to ensure robustness out of sample. This paper does not directly discuss these approaches but we do point out that any traders favourite statistical estimation and training procedures can work in this setting. It may be that more recent training data is more important for fitting and this may be incorporated via a rolling fitted model - but this may lead to overfitting. Hence, in this paper we leave such statistical procedures directly to the discretion of the user and instead provide a framework in which such techniques can be ensembled.

---

**Algorithm 1** Fitting The Optimal Signature Trading Strategy

---

**Input:** A finite set of  $M$   $d$ -dimensional sample market paths  $\left\{X_{t \in [0, T]}^i\right\}_{i=1}^M$   
A finite set of  $M$   $N$ -dimensional sample factor (signal) paths  $\left\{f_{t \in [0, T]}^i\right\}_{i=1}^M$

**Output:**  $\ell_m \in T((\mathbb{R}^{N+d+1})^*) \quad \forall m \in \{1, \dots, d\}$ : The optimal trading strategy as a linear functional on the signature of the time-augmented path

**Parameters:**  $N \geq 0$ : Truncation level of the signature

$\Delta \geq 0$ : Maximum variance of PnL

- 1: Create the  $(N+d+1)$ -dimensional market factor process  $\hat{Z} = (t, X_t, f_t)$  to obtain  $\left\{\hat{Z}_{t \in [0, T]}\right\}_{i=1}^M$ .
  - 2: Apply time-augmentation and lead-lag transformations to all market factor paths, to obtain  $\left\{\hat{Z}_{t \in [0, T]}^{LL}\right\}_{i=1}^M$ .
  - 3: Compute the truncated signature (at order  $N$ ) of each market factor path  $\left\{\hat{\mathbb{Z}}_{t \in [0, T]}^{LL, \leq N}\right\}_{i=1}^M$ .
  - 4: Calculate the empirical expected signature  $\mathbb{E}(\hat{\mathbb{Z}}^{LL, \leq N}) = \frac{1}{M} \sum_{i=1}^M \hat{\mathbb{Z}}_{t \in [0, T]}^{LL, \leq N}$
  - 5: Populate vector  $\mu^{\text{sig}}$  as a subset of  $\mathbb{E}(\hat{\mathbb{Z}}^{LL, \leq N})$  terms, using (3.2).
  - 6: Populate matrix  $\Sigma_{\mathbf{w}, \mathbf{v}}^{\text{sig}}$  for each word  $\mathbf{w}, \mathbf{v} \in \mathcal{W}_{N+d+1}$ , using (3.3).
  - 7: Solve the system of linear equations as to obtain  $\ell_m \in T((\mathbb{R}^{N+d+1})^*) \quad \forall m \in \{1, \dots, d\}$  where
$$\langle \ell_m^*, e_{\mathbf{w}} \rangle = \frac{((\Sigma^{\text{sig}})^{-1} \mu^{\text{sig}})_{\mathbf{wf}(m)}}{2\lambda}, \quad m \in \{1, \dots, d\}, \mathbf{w} \in \mathcal{W}_{N+d+1}^M$$
  - 8: **return** Linear functional  $\ell_m \quad \forall m \in \{1, \dots, d\}$ .
- 

Likewise, while signature trading strategies do a good job of drawdown control, we do not discuss extra practicalities such as vol-scaling positions in this paper. Due to the nested nature of filtrations in this work (i.e. we are in possession of strictly increasing amounts of information

through time), this may impact performance of a trading strategy differently through the life of the trade, hence it would make practical sense to smooth out Sig-Trading positions through time - this also can reduce any bias that may arise from the exact start date of the trade. We also suggest that, especially at higher frequencies, that we could replace time-augmentation with *volume-augmentation* since this is still a monotonically increasing channel in the process (ensuring the signature remains unique), but represents a new notion of time.

---

**Algorithm 2** Trading The Optimal Signature Trading Strategy, at time  $t$ 


---

**Input:** The previous stopped underlying asset and signal paths  $X_{s \in [0,t]}, f_{s \in [0,t]}$ ,  
Linear functional  $\ell_m \in T((\mathbb{R}^{N+d+1})^*) \quad \forall m \in \{1, \dots, d\}$ .

**Output:**  $\xi_t^m \quad \forall m \in \{1, \dots, d\}$ : The optimal trading strategy

- 1: Create the  $(N+d+1)$ -dimensional market factor paths and apply time-augmentation to obtain  $\hat{Z}_{s \in [0,t]}$ .
- 2: Compute the truncated signature of the stopped market factor path,  $\hat{Z}_{0,t}$ .
- 3: For each asset  $m \in \{1, \dots, d\}$ , obtain the optimal strategy  $\xi_t^m$  by applying an inner product of the terms of  $\ell_m$  to the signature of the time-augmented market path.

$$\xi_t^m = \langle \ell_m, \hat{Z}_{0,t} \rangle$$

- 4: **return**  $\xi_t^m \quad \forall m \in \{1, \dots, d\}$
- 

## 5 Numerical Results

Throughout this section, we aim to demonstrate and highlight some of the capabilities of the Sig-Trading framework, incorporating path-dependencies and exogenous signals.

### 5.1 Synthetic Data

#### 5.1.1 Pairs Trading

First, we explore the scenario in which we have no exogenous trading signal to enrich our trading strategy but we only have access to the underlying time series of the asset process. The true advantage that data-driven methods have over classical parametric frameworks is that they can exploit the inefficiencies present in financial time series data, without specifying the explicit dynamics that they are trying to capture. In this toy example, we aim to isolate one specific aspect that is exhibited by time series data and highlight how the Sig-Trader exploits it. *Pairs Trading* is one of the most famous and original active trading strategies deployed by investors that focuses on trading the joint behaviour between two assets. The sentiment is that while both assets have their own dynamics, the difference (spread) between the prices of the two assets should hold some predictability on future (co-)movements. Classical literature focuses on modelling this relationship using a mean-reverting process such as an Ornstein-Uhlenbeck process. The strategy should then contain a *buy signal* when the spread falls below some threshold and a *sell signal* when the spread is above the threshold, in the anticipation that this spread should converge back to the threshold. More on mean-reversion strategies can be found in [AD02; Vid04; MPW08; CJ15; LL15].

In this experiment we take two assets such that

$$dX_t = \sigma^X dW_t^X$$

$$dY_t = \kappa(X_t - Y_t)dt + \sigma^Y dW_t^Y$$

Where  $X$  is a standard arithmetic Brownian motion with zero drift and volatility  $\sigma^X$ .  $Y$  however is modelled as a mean-reverting process where its drift is proportional to the spread between  $X$  and  $Y$ . We can clearly see in this toy example that the only exploitable *alpha* within the framework is the temporal dependence through mean-reversion in  $Y$  since there is no long term drift in  $X$  or  $Y$ .

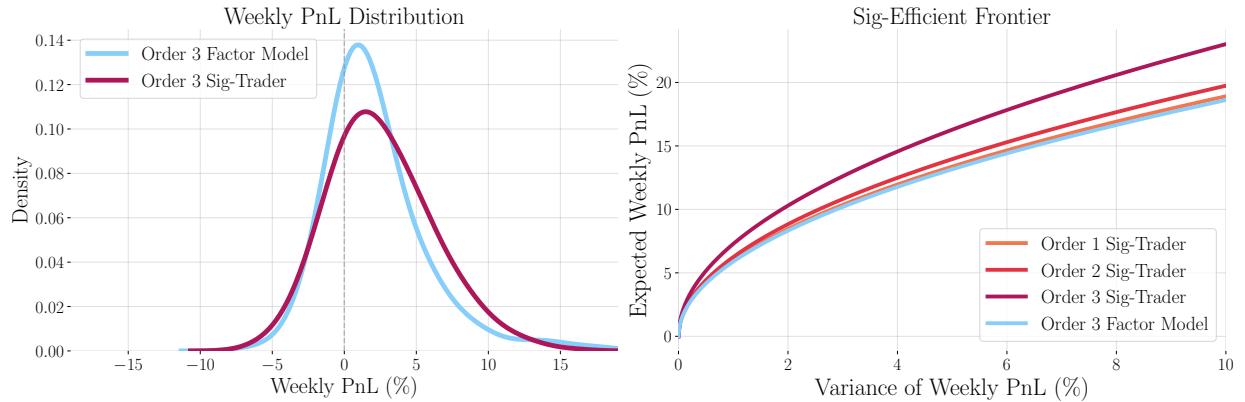


Figure 9: Sharpe ratio distribution (LHS) and Sig-Trader strategy as a function of the spread between the two assets (RHS).

If a trader was to deploy a static buy and hold strategy here, the PnL would be zero on average, however, for higher order Sig-Traders, it is possible to exploit the mean-reversion dynamics. In fact, this example demonstrates how the above-mentioned *alpha* is self-contained within the 1st level of the signature (which is the increment of the path, or the ‘drift’) and so orders of  $N \geq 2$  do not contain any more predictive power on excess returns than that of  $N = 1$ . However, on the right hand side (RHS) of Figure 9, the signature efficient frontier illustrates that when trading with a weekly look-forward horizon, the ratio of return to variance of weekly PnL is greater as we increase the order of the Sig-Trader. This is due to the higher levels of the signature capturing non-linear path-dependencies that can help reduce variance in the weekly PnL distribution. We relate back to Figure 2 to demonstrate that in fact higher order Sig-Traders are able to construct mean-reverting strategies that naturally limit drawdowns, which are an inherently path-dependent characteristic.

In this synthetic example, we compare the Sig-Trader strategy to that of the original factor model. The generic factor model is set up as a supervised linear regression on future returns, as a function of the current signature,

$$\mu_{t+1} = \mathbb{E}[r_{t+1} | \mathcal{F}_t] = B\mathcal{S}_t + \varepsilon_{t+1}$$

where  $r$  is the 2-dimensional asset returns,  $B$  is a  $2 \times N$  matrix of factor coefficients,  $\mathcal{S}_t$  is a vector of the signature values of the and  $\varepsilon$  is a vector of the 2 assets’ (unexplained) residuals returns. We can see that this model uses the same input as the Sig-Trader (the signature), with the same tools (applying a linear function).

The key distinction is that the factor model is optimal for maximising the daily mean-variance PnL profile, which ignores any long term, pathwise dynamics of the strategy itself. Figure 10 demonstrates the difference in nature between the Sig-Trader and the factor model. In the order 1 case, as described above, the exploitable alpha in terms of returns are captured by the 1st level of the signature and so we see a similar position profile for the Sig-Trader and the factor model. However, when the factor model tends to go more short (or long), the higher order Sig-Traders, e.g order 3, tends to reduce its position since this will be more beneficial to minimising the weekly variance of PnL. We can think of this behaviour as being similar to applying a sigmoid function to your position in search of robustness, or applying a stop-loss to avoid becoming too leveraged - which are common practices. By incorporating path-dependencies in the strategy, we have access to a more robust and intuitive extension to classic factor models.



Figure 10: Comparison between the Markowitz Order 3 factor model positions and the corresponding Sig Trader positions, for orders 1,2,3.

### 5.1.2 Incorporating Exogenous Signal

We now consider the case when we are in possession of an exogenous signal that can be used to inform our trading decisions, rather than solely relying on raw asset time series data. Suppose the asset price process  $X$  is driven by some function of the signal  $\phi(t, f_{0,t})$ , e.g

$$dX_t = \phi(t, f_{0,t})dt + \sigma^X dW_t^X.$$

Since  $\phi(t, f_{0,t})$  is a function of the past time series of  $f$ , it may be difficult for a trader to directly model the dynamics of this system directly if they do not know the explicit form of  $\phi$ . In practice, a trader may use a Kalman (or alternative) filter to capture the impact of a noisy signal; in fact, the authors in [Coh+23] prove how the Kalman filter can be equivalently written as a linear regression on the signature. In this example, we propose the following system for the signal process  $f$  and its consequent impact on the underlying asset  $X$ ,

$$\begin{aligned} df_t &= -\kappa f_t + \sigma^f dW_t^f \\ Z_t &= \int_0^t K(t-s)df_s \\ dX_t &= Z_t dt + \sigma^X dW_t^X. \end{aligned}$$

We let the (observable) signal  $f$  be a generic mean-reverting OU process with zero mean, while it has a path-dependent causal impact on some latent process  $Z$  via a time-dependent kernel  $K$  that we do not observe. This kernel could simply be a stochastic filter, for example if the kernel is exponential, we recover an exponentially weighted average of previous increments in the signal. In this example, we take the kernel to be  $K(t, s) = \exp\{-\alpha(t-s)\}$ , which can be understood to be a decaying impact of the signal on the process  $Z$ , i.e more recent observations of  $f$  have a greater impact on the instantaneous drift. The asset price process  $X$  then behaves like an arithmetic Brownian motion with long term zero drift, but short term temporal structure.

The underlying asset process  $X$  does not have a long term drift, so we would expect a static strategy to have no long-term expected return, as seen in the left hand plot of Figure 12. In fact, we notice that any excess return is marginal when the Sig-Trader trades endogenously without access to the signal, which can be seen in the light blue distributions in Figure 12. The red distributions correspond to the simple factor model that takes the value of the signal at time  $t$  and predicts the future (one-step) return, i.e

$$\mu_{t+1} = \mathbb{E}[r_{t+1} | \mathcal{F}_t] = \beta f_t + \varepsilon_{t+1} \quad (5.1)$$

where the position is then scaled according to the expected return  $\mu_{t+1}$ . The dark blue Sharpe ratio distributions correspond to the Sig-Trader for different orders of truncation. Clearly, we

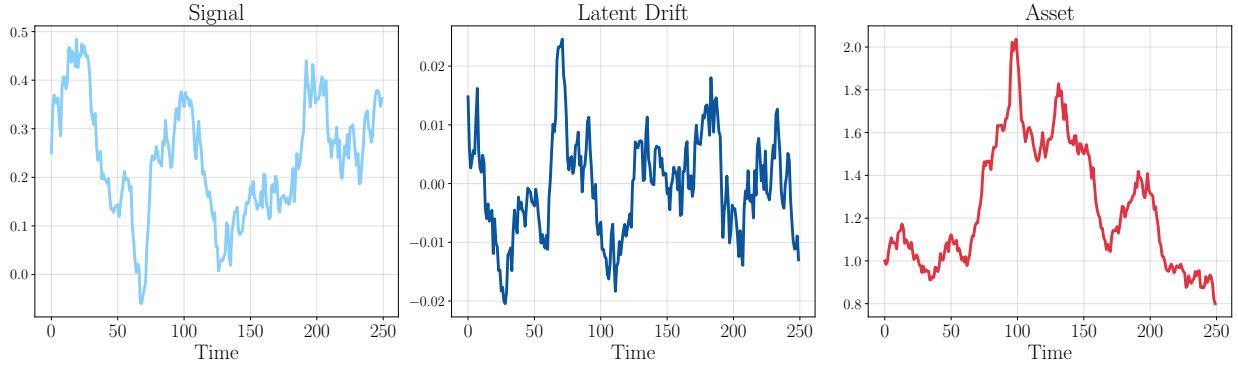


Figure 11: An Example of Signal/Drift/Asset Trajectories.

see that for higher orders of truncation, the Sharpe ratio improves as the Sig-Trader can better approximate the non-linear relationship between the signal and the underlying asset process. In

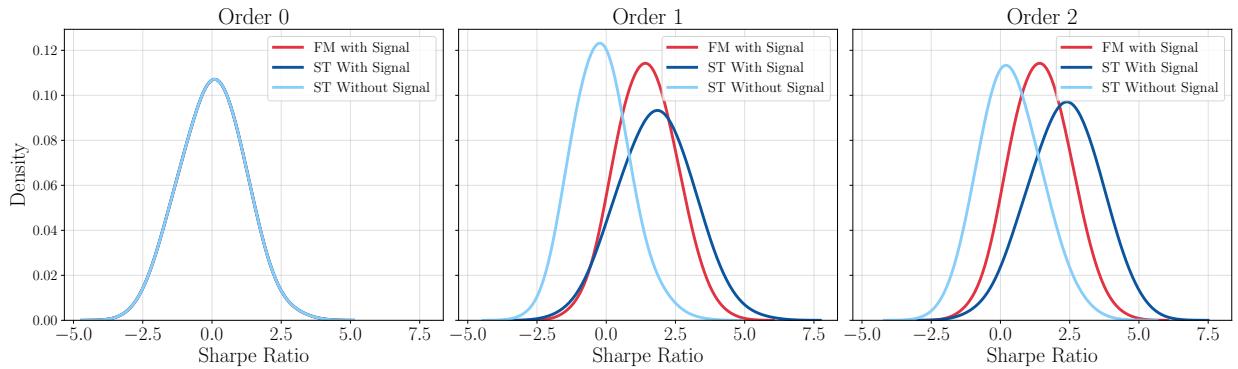


Figure 12: Sharpe ratio distributions for the Sig-Trader of truncation orders 0,1,2, compared to the factor model described in (5.1).

this system there are several layers of noise and complexities to sift through, including the path-dependent impact of the signal, as well as noise from the signal itself  $\sigma^f$  and exogenous noise of the underlying asset,  $\sigma^X$ . In this simple system, this volatility of such randomness is constant through time, but in practice this is likely not the case and this is the type of scenario when the Sig-Trader can outperform the classic predict-then-optimize frameworks. The left hand side of Figure 13 illustrates how the Sig-Trader can transform a signal of varying strengths, into a position that is able to produce strong risk-adjusted returns.

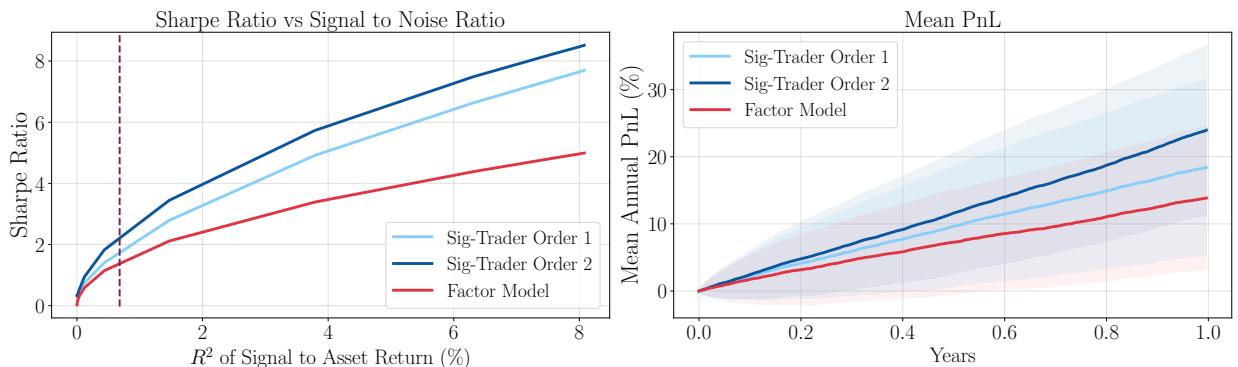
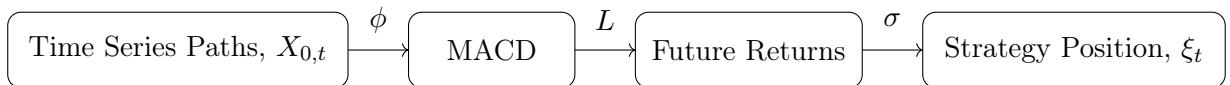


Figure 13: Sharpe ratio against the signal to noise ratio (LHS). The mean PnL and variance of PnL through time (RHS).

## 5.2 Learning Momentum as a Sig-Trading Strategy

As highlighted previously, the class of signature trading strategies, contains most common systematic trading strategies that are functions of the past market time-series. One of the most common and established forms of systematic trading strategy is trend following (and its multiple variants), which generally applies some function (i.e a filter) to the past market time series, to determine the strength and direction of the underlying asset trend. The trader then trades in this direction, adjusting their position according to the strength (alternatively, they may want to trade the opposite direction, which would constitute a mean-reversion strategy). For a high-level overview of the design and mechanics of momentum strategies, we refer the reader to [RD12; RA12], which give a large discussion on both linear and non-linear momentum. For practical applications, see [MOP12; AMP13; BK13; BS15; Lem+14].

Within the broader class of momentum strategies live several variants, and this is dependent on your choice of function that characterises the trend. Variants (and combinations thereof) include RSI indicators, Bollinger bands, MACD (moving average convergence divergence) or any other type of moving average crossover. The key observation is that all of the previous mentioned variants constitute alternative functions (filters) of the past time series, therefore a natural question to ask is what are the best (or in fact optimal) types of function or filter to apply to the past market time series, in order to capture the dynamics of the underlying asset? Since momentum strategies are contained within the space of signature trading strategies, we are able to find the linear functional corresponding to a given momentum strategy. To make this more precise, we focus on capturing the characteristics of a MACD momentum strategy. We recall that  $\text{MACD}(t_1, t_2)$  is the difference between the (exponentially weighted)  $t_1$ -moving average (the fast signal) and the  $t_2$ -moving average (the slow signal), designed to indicate strength of trend. The general setup might look as follows:



This is a very general framework that takes a given filter (i.e the MACD), uses it to predict future returns, and then applies some normalisation (i.e a sigmoid function) in order to retrieve a final strategy position. We note that a momentum strategy should naturally result in a positive regression coefficient of  $L$  since a positive MACD signal indicates that the asset is ‘trending’ (otherwise, if the coefficient was negative, this would imply mean-reversion). We can think of this whole framework as being one continuous function of the past path, i.e

$$\xi_t = \varphi(t, X_{0,t}) = \sigma(L(\phi(t, X_{0,t}))).$$

The goal is therefore to demonstrate that the function  $\varphi$  can be captured via a signature trading strategy such that  $\varphi(t, X_{0,t}) = \langle \ell, \hat{\mathbb{X}}_{0,t} \rangle$ .

Figure 14 displays the corresponding MACD filter, which we denote  $\phi$ , while Figure 15 shows the improvement of learnt linear functionals  $\ell$ , as the order of truncation of the signature increases. We notice that the order 3 signature is able to approximate the function  $\varphi$  with almost 90%  $R^2$  accuracy, even though both the filter  $\phi$  and the normalisation function  $\sigma$  are highly non-linear.

Given that we are able to approximate the momentum trading strategy via a Sig-Trading strategy, we can compare this strategy to the mean-variance optimal order 3 strategy. Using the learnt linear functional  $\ell$ , we are able to construct the efficient frontier via the expected terminal PnL and variance of terminal PnL,

$$\begin{aligned}\mathbb{E}(V_T) &= \ell^\top \mu^{\text{sig}} \\ \text{Var}(V_T) &= \ell^\top \Sigma^{\text{sig}} \ell.\end{aligned}$$

Figure 16 (LHS) demonstrates that the optimal order 3 Sig-Trader has a much better risk-return profile over the chosen trade horizon of 20 days (one month), than that of the learnt momentum trader. It is worth noting however that the Sig-Trader in this example has many similar characteristics to that of the momentum trader, however the optimal Sig-Trader was able to capture further

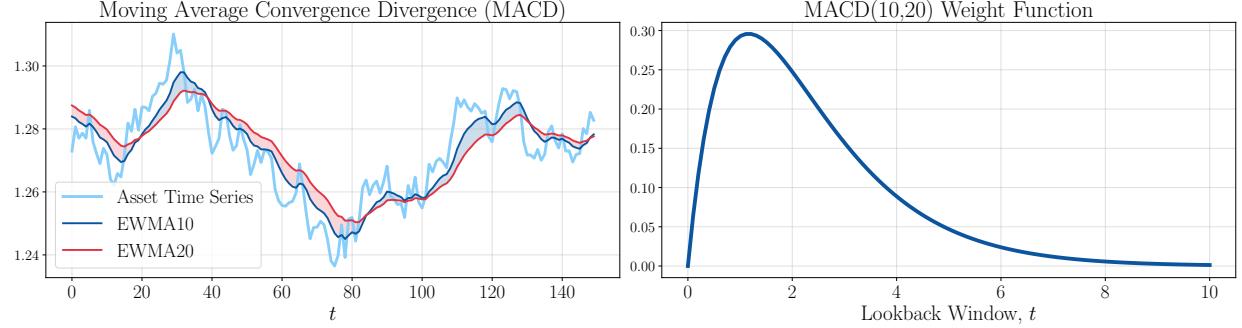


Figure 14: An example of the corresponding slow and fast moving averages of a MACD(10,20) signal on the TLT ETF during 2006 (LHS). The associated weight function  $\phi(t - s)$  with the MACD(10,20) filter (RHS).

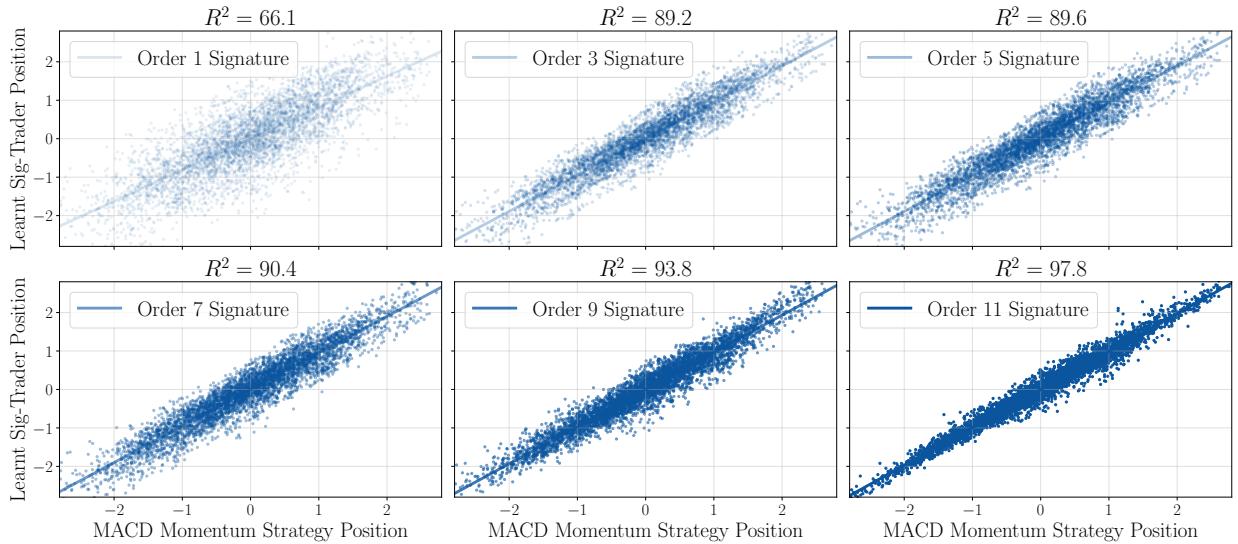


Figure 15: The learnt MACD momentum strategy as a linear functional on the signature for different orders of truncation.

characteristics about the underlying dynamics and factor in the pathwise optimisation to achieve a lower monthly variance. The convex return profile of the order 3 Sig-Trader (RHS of Figure 16) is reminiscent of a trend following strategy that earns positive returns with high conviction in trending (upwards or downwards) markets. This example shows us that in fact, besides from focusing on an optimal Sig-Trading solution, by re-casting other strategies in the same format, it can inform us better on how our strategies work and the various exposures that may exist. It might be such that a given strategy is largely exposed to specific terms in the signature, which can explain more about its characteristics and associated risks. We can also analyse how ‘far away’ our strategies are to an optimal Sig-Trading solution and hence better optimise for dynamic risks.

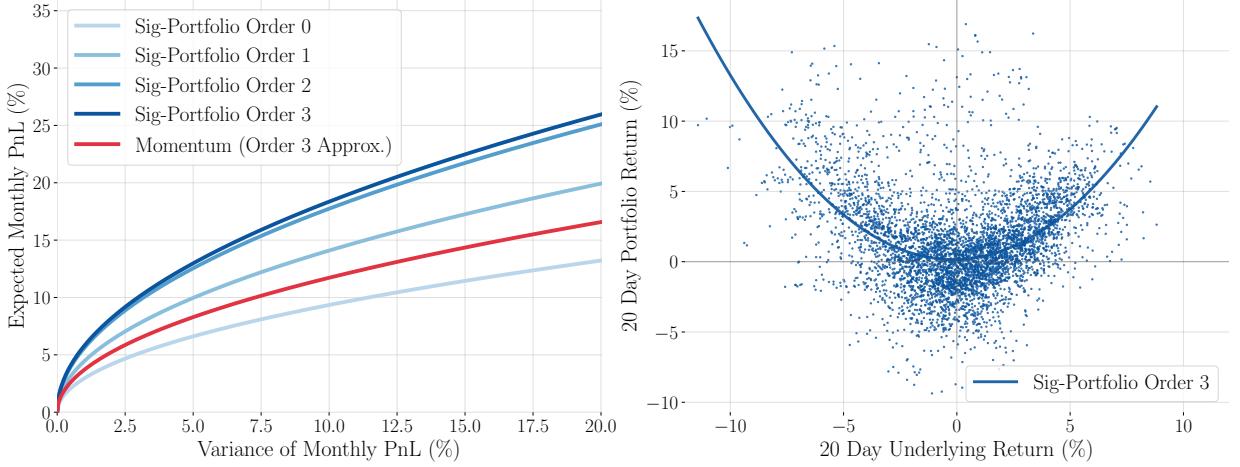


Figure 16: Signature efficient frontier comparison for optimal Sig-Traders and the corresponding learnt 3rd order approximation of the MACD momentum strategy (LHS). The order 3 Sig-Trading strategy return profile compared to the underlying returns (RHS).

## 6 Conclusion

Path dependencies such as non-Markovian data structures, or time series exhibiting temporal correlation are a frequent phenomenon in financial data. Momentum and mean-reversion (of assets or signals) are two of the purest features of time series' data that a trader can exploit, and these features are inherently reliant on the whole path. However, many traditional techniques used for portfolio optimisation are too inflexible to handle such structural complexities of the data and signals. Signature trading strategies, first developed and covered in detail in the thesis of Perez ([Per20]), are a versatile representation of any investment strategy, which have been demonstrated to be a powerful tool in the context of pricing, hedging, and optimal execution.

We observe that in fact their advantages can be carried over to more general portfolio optimisation problems as they encompass many common trading styles that are present in practice (including the before-mentioned momentum and mean-reversion). More specifically, in this paper we extend classical factor models into the Sig-Trading framework, obtaining a closed form solution to the optimal mean-variance Sig-Trading strategy and derive a clear intuition for portfolio managers to navigate and use this formula in a path-dependent context. Furthermore, by lifting the mean-variance optimisation into the lead-lag signature space (See Definition 2.3), we bypass the necessity for any explicit prediction of returns, which is commonly required in traditional settings. This alleviates the accumulation of asymmetric residuals from the prediction phase, which can often be difficult to control. Moreover, the Sig-Trading framework simultaneously captures the joint signal-asset dynamics, whilst performing a dynamic optimisation which naturally incorporates a drawdown control within the objective function.

In summary, the Sig-Trading framework provides an alternative to machine learning methods, in its ability to handle path-dependent, non-linear dynamics and signals in a portfolio optimisation context. Unlike machine learning methods, our framework requires no training or gradient descent optimisation and provides a closed-form solution that is lightweight to work with in practice. Our closed-form solution is tractable, easily implementable and ensures interpretability of the derived optimal strategies. Overall, our results provide more intuition than ML-based methods and establish more discretion in the fitting procedure than classical methods, providing a malleable framework that still solves many challenges faced when working with financial data.

## Appendix A Rough Path & Tensor Algebra Preliminaries

We aim to keep this paper self-contained by recalling the concepts and definitions we explicitly use in this paper. In this appendix we recall necessary fundamental building blocks used in our derivations.

### A.1 The Tensor Algebra

In this section, we define the space on which the signature is defined and introduce notations that are used throughout the paper.

**Definition A.1.** (Tensor Algebra). Let  $d \geq 1$ . We define the extended tensor algebra over  $\mathbb{R}^d$  by

$$T((\mathbb{R}^d)) := \left\{ \mathbf{a} = (a_0, a_1, \dots, a_n, \dots) \mid a_n \in (\mathbb{R}^d)^{\otimes n} \right\}$$

Similarly, we define the truncated tensor algebra of order  $N \in \mathbb{N}$  and the tensor algebra by  $T^{(N)}(\mathbb{R}^d)$  and  $T(\mathbb{R}^d)$  respectively, by

$$\begin{aligned} T^{(N)}(\mathbb{R}^d) &:= \left\{ \mathbf{a} = (a_n)_{n=0}^{\infty} \mid a_n \in (\mathbb{R}^d)^{\otimes n} \text{ and } a_n = 0 \forall n \geq N \right\} \subset T((\mathbb{R}^d)) \\ &= \bigoplus_{n=0}^N (\mathbb{R}^d)^{\otimes n}, \\ T(\mathbb{R}^d) &:= \bigoplus_{n=0}^{\infty} (\mathbb{R}^d)^{\otimes n} \subset T((\mathbb{R}^d)). \end{aligned}$$

Note that the truncated tensor algebra of order  $N$  has dimension  $\sum_{k=0}^N d^k = \frac{d^{N+1}-1}{d-1}$ .

*Remark A.1.* Intuitively, we have that the zero-th level of the tensor algebra,  $(\mathbb{R}^d)^{\otimes 0}$ , is simply the set of all scalars  $a_0 \in \mathbb{R}$ , with dimension  $d^0 = 1$ . At the first level,  $(\mathbb{R}^d)^{\otimes 1}$  is the set of all  $\mathbb{R}$ -valued vectors of length  $d$ , with dimension  $d^1$ . Likewise, at the second level,  $(\mathbb{R}^d)^{\otimes 2}$  is the set of all  $\mathbb{R}$ -valued  $d \times d$  matrices. Then the truncated tensor algebra at order 2,  $T^{(2)}(\mathbb{R}^d)$ , has dimension  $1 + d + d^2$  and contains all  $\mathbb{R}^d$ -valued tensors of order 0, 1, 2.

**Definition A.2.** (Dual Space of the Tensor Algebra). Let  $\{e_1, \dots, e_d\} \subset \mathbb{R}^d$  be a basis for  $\mathbb{R}^d$ , then it has a dual basis  $\{e_1^*, \dots, e_d^*\} \subset (\mathbb{R}^d)^*$  for  $(\mathbb{R}^d)^*$ , the dual space of  $\mathbb{R}^d$ . Recall that this dual space is the space of all linear functions  $\mathbb{R}^d \rightarrow \mathbb{R}$ . We may similarly define a basis for  $T((\mathbb{R}^d))$  and its dual space  $T((\mathbb{R}^d))^*$ .

We identify this dual space of the tensor algebra,  $T((\mathbb{R}^d)^*)$ , with the space of all words. Consider the following alphabet  $A_d := \{\mathbf{1}, \dots, \mathbf{d}\}$ , which consists of  $d$  letters. In order to ease notations, we make the following identification:

$$e_{i_1}^* \otimes \cdots \otimes e_{i_n}^* \in T((\mathbb{R}^d)^*) \leftrightarrow \mathbf{i}_1 \dots \mathbf{i}_n \in \mathcal{W}(A_d), \quad (\text{A.1})$$

where  $\mathcal{W}(A_d)$  is the real vector space of all words with the alphabet  $A_d$ . The empty word will be denoted by  $\emptyset$ . We then have the identification  $T((\mathbb{R}^d)^*) = \mathcal{W}(A_d)$ . That is, that any linear functional  $\ell : T((\mathbb{R}^d)) \rightarrow \mathbb{R}$  can be identified via elements in  $\mathcal{W}(A_d)$ . Hence we can think of *words* as linear functions on the tensor algebra.

**Example A.1.** Let  $\mathbb{X} \in T((\mathbb{R}^d))$  be an element of the tensor algebra. We can view  $\mathbb{X}$  in terms of

its elements within the tensor algebra and each multi-index corresponding to a word, e.g.

$$\mathbb{X} = \left( \underbrace{\mathbb{X}^{\emptyset}}_{\in (\mathbb{R}^d)^{\otimes 0}}, \underbrace{\begin{pmatrix} \mathbb{X}^{\mathbf{1}} \\ \vdots \\ \mathbb{X}^{\mathbf{d}} \end{pmatrix}}_{\in (\mathbb{R}^d)^{\otimes 1}}, \underbrace{\begin{pmatrix} \mathbb{X}^{\mathbf{11}} & \dots & \mathbb{X}^{\mathbf{1d}} \\ \vdots & \ddots & \vdots \\ \mathbb{X}^{\mathbf{d1}} & \dots & \mathbb{X}^{\mathbf{dd}} \end{pmatrix}}_{\in (\mathbb{R}^d)^{\otimes 2}}, \dots \right).$$

The space of all words is defined as

$$\mathcal{W}(A_d) = \{\emptyset, \mathbf{1}, \dots, \mathbf{d}, \mathbf{11}, \dots, \mathbf{dd}, \dots\}.$$

Two algebraic operations on  $\mathcal{W}(A_d)$  are the sum and concatenation. The sum of two words  $\mathbf{w}$  and  $\mathbf{v}$  is just the formal sum  $\mathbf{w} + \mathbf{v} \in \mathcal{W}(A_d)$ . The concatenation of  $\mathbf{w} = \mathbf{i}_1 \dots \mathbf{i}_n, \mathbf{v} = \mathbf{j}_1 \dots \mathbf{j}_m \in \mathcal{W}(A_d)$  is defined by

$$\mathbf{wv} := \mathbf{i}_1 \dots \mathbf{i}_n \mathbf{j}_1 \dots \mathbf{j}_m \in \mathcal{W}(A_d).$$

With some abuse of notation, we will then use the concatenation on  $\mathcal{W}(A_d)$  and  $T((\mathbb{R}^d)^*)$  interchangeably, in the sense that we will sometimes write  $\ell\mathbf{w}$  for  $\ell \in T((\mathbb{R}^d)^*), \mathbf{w} \in \mathcal{W}(A_d)$ .

**Definition A.3.** (Shuffle Product). The shuffle product  $\sqcup : \mathcal{W}(A_d) \times \mathcal{W}(A_d) \rightarrow \mathcal{W}(A_d)$  is defined inductively by

$$\begin{aligned} \mathbf{ua} \sqcup \mathbf{vb} &= (\mathbf{u} \sqcup \mathbf{vb})\mathbf{a} + (\mathbf{ua} \sqcup \mathbf{v})\mathbf{b} \\ \mathbf{w} \sqcup \emptyset &= \emptyset \sqcup \mathbf{w} = \mathbf{w} \end{aligned}$$

for all words  $\mathbf{u}, \mathbf{v}$  and letters  $\mathbf{a}, \mathbf{b} \in \mathcal{W}(A_d)$ .

**Example A.2.** Let  $\mathbf{w} = \mathbf{12}, \mathbf{v} = \mathbf{34}$ , then  $\mathbf{w} \sqcup \mathbf{v}$  is given by

$$\mathbf{12} \sqcup \mathbf{34} = \mathbf{1234} + \mathbf{1324} + \mathbf{1342} + \mathbf{3124} + \mathbf{3142} + \mathbf{3412}$$

*Remark A.2.* We make extensive use of the fact that polynomials of linear functionals can be expressed as shuffle products of the linear functionals themselves.

## A.2 Rough Paths & The Signature

**Definition A.4.** ( $p$ -variation). Let  $p \geq 1$  and  $X : [t', T] \rightarrow \mathbb{R}^d$  be a  $d$ -dimensional continuous path. We say the  $p$ -variation of  $X$  is denoted as the seminorm

$$\|X\|_p = \left( \sup_{\mathcal{P}} \sum_{[s,t] \in \mathcal{P}} \|X_t - X_s\|^p \right)^{\frac{1}{p}}$$

where  $\|\cdot\|$  is any norm on  $\mathbb{R}^d$  and the supremum is taken over all partitions  $\mathcal{P}$  of the interval  $[t', T]$ .

**Definition A.5.** (Space of  $p$ -variation paths). We denote the space of all  $\mathbb{R}$ -valued  $d$ -dimensional paths of finite  $p$ -variation, to be  $C^{p-var}([0, T]; \mathbb{R}^d)$ .

*Remark A.3.* We refer to the paths of *bounded variation* as elements of the space  $C^{1-var}([0, T]; \mathbb{R}^d)$ . Note that all continuous piecewise smooth paths  $X_{0,T} \in C^{1-var}([0, T]; \mathbb{R}^d)$ .

**Definition A.6.** (Geometric  $p$ -rough paths). We define the space of *geometric  $p$ -rough paths*,  $G^{\lfloor p \rfloor}(\mathbb{R}^d)$ , as the closure of the space of signatures of smooth paths, at order  $\lfloor p \rfloor$ , namely

$$G^{\lfloor p \rfloor}(\mathbb{R}^d) := \overline{\left\{ \mathbb{X}^{\leq N} : \Delta_T \rightarrow T^{(N)}(\mathbb{R}^d) \quad | \quad N = \lfloor p \rfloor \right\}}^{d_{p-\text{var}}}$$

where the closure is with respect to the  $p$ -variation metric (defined in [LCL07], Definition 1.5).

**Theorem A.7.** (Extension Theorem, [LCL07], Theorem 3.7). Let  $p > 1$  be a real number and  $\mathbb{X}_{s,t}^{\leq N} \in G^{\lfloor p \rfloor}(\mathbb{R}^d)$  be the truncated signature at level  $N \in \mathbb{N}$  of the path  $X_{s,t} \in \mathcal{X}_{s,t}^d$ . Then for every  $n \geq \lfloor p \rfloor + 1$ , there exists a unique  $\mathbb{X}_{s,t}^n$  such that

$$(s, t) \mapsto \mathbb{X}_{s,t} = (1, \mathbb{X}_{s,t}^1, \dots, \mathbb{X}_{s,t}^{\lfloor p \rfloor}, \dots, \mathbb{X}_{s,t}^n, \dots) \in T((\mathbb{R}^d))$$

has finite  $p$ -variation. We denote  $\mathbb{X}_{s,t}$  as the extension of  $\mathbb{X}_{s,t}^{\leq N}$ .

**Remark A.4.** The signature of a (piecewise) smooth path  $X_{0,T} \in \mathcal{X}_{0,T}^d$  is itself a geometric rough path, namely  $\mathbb{X}_{0,T}^{<\infty} \in G^{[1]}(\mathbb{R}^d)$ .

**Lemma A.8.** (Shuffle product property). Let  $\mathbb{X}_{0,T}^{<\infty} \in G^{[p]}(\mathbb{R}^d)$  be a geometric rough path and let  $\ell_1, \ell_2 \in T((\mathbb{R}^d)^*)$  be elements of the dual space of the tensor algebra, then

$$\langle \ell_1, \mathbb{X}_{0,T}^{<\infty} \rangle \langle \ell_2, \mathbb{X}_{0,T}^{<\infty} \rangle = \langle \ell_1 \sqcup \ell_2, \mathbb{X}_{0,T}^{<\infty} \rangle$$

**Remark A.5.** The *shuffle product property* is heavily used in deriving an explicit representation of the variance of PnL in Section 3.

**Proposition A.9.** (Signature is point-separating, ([HL10], [Boe+16])). Let  $X : [0, T] \rightarrow \mathbb{R}^d$ , then its signature  $\mathbb{X}_{0,T}^{<\infty} \in G^{[p]}(\mathbb{R}^d)$  is unique up to tree like equivalence and translation.

**Corollary A.9.1.** Let  $\hat{X} : [0, T] \rightarrow \mathbb{R}^{d+1}$  be the associated add-time process of  $X$ . Then its signature  $\hat{\mathbb{X}}_{0,T}^{<\infty}$  uniquely determines  $X$  up to translations.

**Definition A.10.** (Probability measure on the space of paths). Let  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$  be a filtered probability space such that  $\mathcal{F} = (\mathcal{F}_t)_{t \in [0, T]}$  is the filtration generated by the non-negative  $\mathbb{R}^d$ -valued stochastic process  $X = (X_t)_{t \in [0, T]}$ . We say its path trajectories  $X_{0,T}$  are sampled under the probability measure  $\mathbb{P}$  and we denote  $\mathcal{P}(\mathcal{X}_{0,T}^d)$  as the set of all such probability measures.

**Definition A.11.** (Expected Signature). Let  $X$  be defined as previous. Then for any such  $\mathbb{R}^d$ -valued random path  $X_{0,T}$ , we can see that taking its signature  $\mathbb{X}_{0,T}$  is also a random variable under  $\mathbb{P} \in \mathcal{P}(\mathcal{X}_{0,T}^d)$  and so hence we can define the notion of an *expected signature* under  $\mathbb{P}$  at each order as

$$\mathbb{E}^{\mathbb{P}} \left[ \mathbb{X}_{0,T}^n \right] := \mathbb{E}^{\mathbb{P}} \left[ \int_{s < u_1 < \dots < u_k < t} \dots \int dX_{u_1} \otimes \dots \otimes dX_{u_k} \right] \in (\mathbb{R}^d)^{\otimes n}$$

where we have

$$\mathbb{E}^{\mathbb{P}} \left[ \mathbb{X}_{0,T}^{<\infty} \right] := (1, \mathbb{E}^{\mathbb{P}}[\mathbb{X}_{0,T}^1], \dots, \mathbb{E}^{\mathbb{P}}[\mathbb{X}_{0,T}^n], \dots) \in T((\mathbb{R}^d)).$$

The map  $\mathbb{P} \mapsto \mathbb{E}^{\mathbb{P}} \left[ \mathbb{X}_{0,T}^{<\infty} \right] \in T((\mathbb{R}^d))$ , which maps the probability measure  $\mathbb{P}$  to its expected truncated signature, is injective ([CO22]).

*Remark A.6.* The *expected signature* allows us to systematically characterize the empirical probability measure on the streams in a model-free sense. The expected signature of paths  $X \sim \mathbb{P}$ , i.e.  $\mathbb{E}^{\mathbb{P}}[\mathbb{X}_{0,T}^{<\infty}]$  can be thought of as the moment generating function of a path-valued random variables  $X$ . We must also note, however, that the assumption that the expected signature even exists at all is a strong assumption for all levels of the signature  $N \in \mathbb{N}$ . For example, the authors in [Bay+21] highlight that this rules out many stochastic volatility models, such as the Heston model.

**Lemma A.12.** (*Factorial Decay*).

The reason we are able to work with the truncated signature with sufficient confidence is due to the factorial decay of the terms of the signature. Let  $X \in \mathcal{X}_{0,T}^d \subset C^{1-var}([0, T], \mathbb{R}^m)$  and  $[s, t] \subset [0, T]$ . Then  $\forall n \geq 1$

$$\|\mathbb{X}_{s,t}^n\| = \left\| \int_{t_0 < u_1 < \dots < u_n < t} \cdots \int dX_{u_1} \otimes \cdots \otimes dX_{u_n} \right\| \leq \frac{(\|X\|_{1,[s,t]})^n}{n!}$$

**Theorem A.13.** (*Universal Approximation*, [LLN13], Theorem 3.1). Let  $K \subset C^{1-var}([0, T], \mathbb{R}^d)$ <sup>1</sup> be a compact subset of paths. For any  $\phi \in C(K, \mathbb{R})$ , and for every  $\epsilon > 0$ , there exists a linear functional  $\ell \in T((\mathbb{R}^d)^*)$  such that

$$\sup_{X \in K} \|\phi(X) - \langle \ell, \mathbb{X}_{0,T}^{<\infty} \rangle\| < \epsilon$$

where the choice of suitable candidate topology is discussed in [CT22].

*Remark A.7.* If we imagine that the optimal adapted dynamic trading strategy  $\xi_t$  is simply a function of the past path, i.e.  $\xi_t : X_{0,t} \mapsto \phi(X_{0,t})$  for some non-linear  $\phi$ , then Theorem A.13 allows us to approximate the non-linear  $\phi$  by  $\phi(X_{0,t}) \approx \langle \ell, \mathbb{X}_{0,t}^{<\infty} \rangle$  where  $\ell$  is linear. This will be made more precise in Section 2.

**Notation:** Throughout, we denote the time-augmented process by  $\hat{X}_t = (t, X_t), t \in [0, T]$ , the signature of  $\hat{X}_{0,T}$  by  $\hat{\mathbb{X}}_{0,T}^{<\infty}$  and the signature of the lead-lag process  $\hat{\mathbb{X}}_{0,T}^{LL, <\infty}$ .

**Example A.3.** (Linear functional on the signature). Let  $X = (X^1, \dots, X^d) : [0, T] \rightarrow \mathbb{R}^d$  be a  $d$ -dimensional path. We previously made the identification that any linear functional on the tensor algebra can be identified via elements in the space of all words. Take for example:

1. Consider any signature term at the second level of the  $i$ -th and  $j$ -th component of  $X$ .

$$\mathbb{X}_{0,T}^{ij} = \iint_{0 < u_i < u_j < T} \circ dX_{u_i}^i \circ dX_{u_j}^j = \int_0^T (X_t^i - X_0^i) \circ dX_t^j = \langle ij, \hat{\mathbb{X}}_{0,T}^{<\infty} \rangle$$

where we can think of  $ij$  as a word in the space  $\mathcal{W}(A_d)$ , which can be identified as a linear functional on the tensor algebra, as shown in (A.1).

2. Consider an arbitrary linear functional  $\ell \in T((\mathbb{R}^d)^*)$  on the signature, then its (Stratonovich) integral against the  $i$ -th component of the  $d$ -dimensional process  $X$  is simply the linear functional concatenated with the letter  $i$ , applied to the signature  $\mathbb{X}_{0,T}^{<\infty}$ , such as:

$$\int_0^T \langle \ell, \mathbb{X}_{0,t}^{<\infty} \rangle \circ dX_t^i = \langle \ell i, \mathbb{X}_{0,T}^{<\infty} \rangle$$

---

<sup>1</sup>  $K$  should be a subset of tree-reduced paths. However, as all of the paths we are concerned with are tree-reduced, this is not an issue.

3. Let

$$\ell = \alpha_0 \mathbf{1} + \alpha_1 \mathbf{2} + \alpha_2 \mathbf{12}.$$

When applied to the signature, we obtain

$$\langle \ell, \mathbb{X}_{0,t}^{<\infty} \rangle = \alpha_0 \mathbb{X}_{0,t}^{\mathbf{1}} + \alpha_1 \mathbb{X}_{0,t}^{\mathbf{2}} + \alpha_2 \mathbb{X}_{0,t}^{\mathbf{12}}$$

and if we consider the concatenation  $\ell\mathbf{34}$ , then we have

$$\langle \ell\mathbf{34}, \mathbb{X}_{0,t}^{<\infty} \rangle = \alpha_0 \mathbb{X}_{0,t}^{\mathbf{134}} + \alpha_1 \mathbb{X}_{0,t}^{\mathbf{234}} + \alpha_2 \mathbb{X}_{0,t}^{\mathbf{1234}}.$$

## Appendix B Proofs

### B.1 Proof of Theorem 2.11

**Theorem 2.11.** (*PnL of a d-asset signature trading strategy under exogenous signal*). *Let  $\hat{Z} := (t, X_t, f_t)_{t \in [0, T]}$  be the market factor process, where  $X$  is a d-dimensional tradable stochastic process and  $f$  is a N-dimensional un-tradable factor process. Let  $\ell_1, \dots, \ell_d \in T((\mathbb{R}^{N+d+1})^*)$  and define our trading strategy as  $\xi_t^m = \langle \ell_m, \hat{Z}_{0,t}^{<\infty} \rangle$  for  $m = 1, \dots, d$ . Then, we have that the PnL of the strategy between time 0 and time  $T$  can be represented as*

$$V_T = \sum_{m=1}^d \int_0^T \langle \ell_m, \hat{Z}_{0,t}^{<\infty} \rangle dX_t^m \approx \sum_{m=1}^d \int_0^T \langle \ell_m, \hat{Z}_{0,t}^{\text{lag}, <\infty} \rangle dX_t^{m, \text{lead}} = \sum_{m=1}^d \langle \ell_m \mathbf{f}(m), \hat{Z}_{0,T}^{LL, <\infty} \rangle \quad (2.2)$$

where  $\mathbf{f}(m) : \{1, \dots, d\} \rightarrow \mathcal{W}(A_{2(N+d+1)})$  is a shift operator which is defined as  $\mathbf{f}(m) = \pi(e_{m+N+d+2}^*)$  where  $\pi : T((\mathbb{R}^{2(N+d+1)})^*) \rightarrow \mathcal{W}(A_{2(N+d+1)})$  the canonical isomorphism between the dual space  $T((\mathbb{R}^{2(N+d+1)})^*)$  and the space of all words  $\mathcal{W}(A_{2(N+d+1)})$ .

*Proof.* Let  $\hat{Z}^{<M}$  be the  $M$ -th order signature of the (time-augmented) market factor process  $\hat{Z}$  and define  $\hat{Y}_t = (\hat{Z}_t, \hat{Z}_t)$  and its  $M$ -th order truncated signature as  $\hat{Y}_t^{<M}$ . In order to prove this theorem, we first state some important results that we shall use as tools throughout. We denote the quadratic co-variation of 2 processes at time  $t$  as  $[\cdot, \cdot]_t$ .

1.  $\int_0^T \langle \ell_m, \hat{Z}_{0,t}^{<M} \rangle \circ dX_t^m = \langle \ell_m \mathbf{f}(m), \hat{Y}_{0,T}^{<M+1} \rangle$
2.  $\int_0^T Y dX = \int_0^T Y \circ dX - \frac{1}{2} [X, Y]$  for 2 stochastic processes  $X, Y$ .
3.  $\mathbb{Z}_{0,T}^{LL, \leq 2} = \hat{Y}_{0,T}^{<2} + \psi_{0,T}$  where

$$\psi_{0,T} = \begin{pmatrix} 0 & -\frac{1}{2} [\hat{Y}]_{0,T} \\ \frac{1}{2} [\hat{Y}]_{0,T} & 0 \end{pmatrix}$$

4.  $\mathbb{Z}_{0,T}^{LL, \leq M} = \int_0^T \hat{Z}_{0,t}^{<M-1} \otimes d\hat{Z}_t$
5.  $\langle \mathbf{m}, \hat{Z}_{0,t}^{<M} \rangle = \langle \mathbf{f}(m), \hat{Y}_{0,t}^{<M} \rangle$
6.  $[\int_0^T \xi dX, Y] = \int_0^T \xi d[X, Y]$

First, we observe that the trading strategy PnL can be decomposed asset-wise such that  $V_T = \sum_{m=1}^d V_T^m$  where  $V_T^m$  is the PnL of the trading strategy of the  $m$ -th asset. Hence, all that needs to be shown is that for arbitrary asset  $m$ ,

$$V_T^m = \int_0^T \langle \ell_m, \hat{Z}_{0,t}^{<\infty} \rangle dX_t^m = \langle \ell_m \mathbf{f}(m), \hat{Z}_{0,T}^{LL, <\infty} \rangle \quad (\text{B.1})$$

where the integral above is in the Itô sense. We also wish to show that this result holds for any truncation level  $M \geq 1$  and for any number of market factors,  $N$ .

Let us fix truncation level  $M \geq 1$ . By (2) we can decompose the Itô integral in (B.1) into a Stratonovich integral and a quadratic variation correction term, i.e

$$\begin{aligned}
\int_0^T \langle l_m, \hat{\mathbb{Z}}_{0,t}^{\leq M} \rangle dX_t^m &= \int_0^T \langle l_m, \hat{\mathbb{Z}}_{0,t}^{\leq M} \rangle \circ dX_t^m - \frac{1}{2} \left[ \langle l_m, \hat{\mathbb{Z}}_{0,\cdot}^{\leq M} \rangle, X^m \right]_T \\
&= \overbrace{\langle l_m \mathbf{f}(m), \hat{\mathbb{Y}}_{0,T}^{\leq M+1} \rangle}^{\text{by (1)}} - \overbrace{\frac{1}{2} \left[ \left\langle l_m, \int_0^{\cdot} \hat{\mathbb{Z}}_{0,t}^{\leq M-1} \otimes d\hat{Z}_t \right\rangle, X^m \right]_T}^{\text{by (4)}} \\
&= \langle l_m \mathbf{f}(m), \hat{\mathbb{Y}}_{0,T}^{\leq M+1} \rangle - \overbrace{\frac{1}{2} \left[ \left\langle l_m \mathbf{f}(m), \int_0^{\cdot} \hat{\mathbb{Y}}_{0,t}^{\leq M-1} \otimes d\hat{Y}_t \right\rangle, X^m \right]_T}^{\text{by (5)}} \\
&= \langle l_m \mathbf{f}(m), \hat{\mathbb{Y}}_{0,T}^{\leq M+1} \rangle - \frac{1}{2} \left\langle l_m \mathbf{f}(m), \left[ \int_0^{\cdot} \hat{\mathbb{Y}}_{0,t}^{\leq M-1} \otimes d\hat{Y}_t, X^m \right]_T \right\rangle \\
&= \langle l_m \mathbf{f}(m), \hat{\mathbb{Y}}_{0,T}^{\leq M+1} \rangle - \frac{1}{2} \left\langle l_m \mathbf{f}(m), \overbrace{\int_0^{\cdot} \hat{\mathbb{Y}}_{0,t}^{\leq M-1} d[\hat{Y}]_t}^{\text{by (6)}} \right\rangle \\
&= \langle l_m \mathbf{f}(m), \hat{\mathbb{Y}}_{0,T}^{\leq M+1} \rangle - \frac{1}{2} \left\langle l_m \mathbf{f}(m), \int_0^T \hat{\mathbb{Y}}_{0,t}^{\leq M-1} \otimes d[\hat{Y}]_t \right\rangle \\
&= \left\langle l_m \mathbf{f}(m), \hat{\mathbb{Y}}_{0,T}^{\leq M+1} - \frac{1}{2} \int_0^T \hat{\mathbb{Y}}_{0,t}^{\leq M-1} \otimes d[\hat{Y}]_t \right\rangle. \tag{B.2}
\end{aligned}$$

Hence, what remains to be shown is that the RHS of (B.1) is equal to (B.2), i.e that

$$\left\langle l_m \mathbf{f}(m), \hat{\mathbb{Z}}_{0,T}^{LL, \leq M+1} \right\rangle = \left\langle l_m \mathbf{f}(m), \hat{\mathbb{Y}}_{0,T}^{\leq M+1} - \frac{1}{2} \int_0^T \hat{\mathbb{Y}}_{0,t}^{\leq M-1} \otimes d[\hat{Y}]_t \right\rangle \tag{B.3}$$

for any truncation level  $M \geq 1$ . For the case when  $M = 1$ , we have that

$$\hat{\mathbb{Z}}_{0,T}^{LL, \leq 2} = \hat{\mathbb{Y}}_{0,T}^{\leq 2} + \psi_{0,T}$$

as defined in (3), which is proven in [FHL16], Theorem 4.1. Therefore, we can clearly see that

$$\begin{aligned}
\left\langle l_m \mathbf{f}(m), \hat{\mathbb{Z}}_{0,T}^{LL, \leq 2} \right\rangle &= \left\langle l_m \mathbf{f}(m), \hat{\mathbb{Y}}_{0,T}^{\leq 2} + \psi_{0,T} \right\rangle \\
&= \left\langle l_m \mathbf{f}(m), \hat{\mathbb{Y}}_{0,T}^{\leq 2} - \frac{1}{2} [\hat{Y}]_T \right\rangle \\
&= \left\langle l_m \mathbf{f}(m), \hat{\mathbb{Y}}_{0,T}^{\leq 2} - \frac{1}{2} \int_0^T 1 d[\hat{Y}]_t \right\rangle \\
&= \left\langle l_m \mathbf{f}(m), \hat{\mathbb{Y}}_{0,T}^{\leq 2} - \frac{1}{2} \int_0^T \hat{\mathbb{Y}}_{0,t}^{\leq 0} \otimes d[\hat{Y}]_t \right\rangle
\end{aligned}$$

Hence, the statement (B.3) holds for  $M = 1$ . By the same proof seen in Lemma 3.2.11 in [Per20], the result follows for all  $M \geq 1$  via induction, taking  $I = \mathbf{i}_1 \mathbf{i}_2 \dots \mathbf{i}_k \in \{1, \dots, d\}^k$  for the multi-dimensional result.

We note here that by setting  $Z = (t, X, f)$ , we lose no strength in this argument, since we are still only integrating against  $X^m$ , for which the result then follows through  $\mathbf{f}(m)$  for  $m = 1, \dots, d$  and

so any extra remaining exogenous factors embedded in  $\mathbb{Z}$  do not change the result. Likewise, since our portfolio PnL is defined asset-wise, the result holds for all assets  $m$  and so the overall trading strategy PnL is defined as

$$V_T = \sum_{m=1}^d \int_0^T \langle l_m, \hat{\mathbb{Z}}_{0,s}^{<\infty} \rangle dX_s^m = \sum_{m=1}^d \langle l_m \mathbf{f}(m), \hat{\mathbb{Z}}_{0,T}^{LL,<\infty} \rangle$$

□

## References

- [Ald+22] Andrew Alden et al. “Model-Agnostic Pricing of Exotic Derivatives Using Signatures”. In: *Proceedings of the 3rd ACM International Conference on AI in Finance, ICAIF 2022*. Association for Computing Machinery, Inc, Nov. 2022, pp. 96–104. ISBN: 9781450393768. DOI: [10.1145/3533271.3561740](https://doi.org/10.1145/3533271.3561740).
- [AD02] Carol Alexander and Anca Dimitriu. “The Cointegration Alpha: Enhanced Index Tracking and Long-Short Equity Market Neutral Strategies”. In: *ISMA Finance Discussion Paper No. 2002-08* (June 2002). DOI: [10.2139/ssrn.315619](https://doi.org/10.2139/ssrn.315619). URL: <https://papers.ssrn.com/abstract=315619>.
- [All+21] Andrew L. Allan et al. “Model-free Portfolio Theory: A Rough Path Approach”. In: *Mathematical Finance* 33.3 (Sept. 2021), pp. 709–765. URL: [http://arxiv.org/abs/2109.01843](https://arxiv.org/abs/2109.01843).
- [ACX23] Anna Ananova, Rama Cont, and Renyuan Xu. “Model-free Analysis of Dynamic Trading Strategies”. In: *SSRN Electronic Journal* (June 2023). ISSN: 1556-5068. DOI: [10.2139/ssrn.4489503](https://doi.org/10.2139/ssrn.4489503). URL: <https://papers.ssrn.com/abstract=4489503>.
- [Arr18] Imanol Perez Arribas. *Derivatives pricing using signature payoffs*. Sept. 2018. URL: [http://arxiv.org/abs/1809.09466](https://arxiv.org/abs/1809.09466).
- [ASS20] Imanol Perez Arribas, Cristopher Salvi, and Lukasz Szpruch. “Sig-SDEs model for quantitative finance”. In: *ICAIF 2020 - 1st ACM International Conference on AI in Finance* (May 2020). DOI: [10.48550/arxiv.2006.00218](https://doi.org/10.48550/arxiv.2006.00218). URL: <https://arxiv.org/abs/2006.00218v2>.
- [AMP13] Clifford S. Asness, Tobias J. Moskowitz, and Lasse Heje Pedersen. “Value and Momentum Everywhere”. In: *Journal of Finance* 68.3 (June 2013), pp. 929–985. ISSN: 15406261. DOI: [10.1111/jofi.12021](https://doi.org/10.1111/jofi.12021).
- [BC09] Alan Bain and Dan Crisan. “Fundamentals of Stochastic Filtering”. In: *Stochastic Modelling and Applied Probability* 60 (2009). DOI: [10.1007/978-0-387-76896-0](https://doi.org/10.1007/978-0-387-76896-0). URL: [http://link.springer.com/10.1007/978-0-387-76896-0](https://link.springer.com/10.1007/978-0-387-76896-0).
- [BK13] Akindynos-Nikolaos Baltas and Robert Kosowski. “Momentum Strategies in Futures Markets and Trend-following Funds”. In: *SSRN Electronic Journal* (Jan. 2013). DOI: [10.2139/ssrn.1968996](https://doi.org/10.2139/ssrn.1968996). URL: <https://papers.ssrn.com/abstract=1968996>.
- [BCK23] Peter Bank, Álvaro Cartea, and Laura Körber. “Optimal execution and speculation with trade signals”. June 2023. URL: <https://arxiv.org/abs/2306.00621v1>.
- [BS15] Pedro Barroso and Pedro Santa-Clara. “Momentum has its moments”. In: *Journal of Financial Economics (JFE)* 116.1 (2015), pp. 111–120. DOI: [10.1016/j.jfineco.2014.11.010](https://doi.org/10.1016/j.jfineco.2014.11.010). URL: [http://dx.doi.org/10.1016/j.jfineco.2014.11.010](https://dx.doi.org/10.1016/j.jfineco.2014.11.010).
- [Bay+21] Christian Bayer et al. “Optimal stopping with signatures”. May 2021. URL: [http://arxiv.org/abs/2105.00778](https://arxiv.org/abs/2105.00778).
- [BDG21] Philippe Bergault, Fayçal Drissi, and Olivier Guéant. “Multi-asset optimal execution and statistical arbitrage strategies under Ornstein-Uhlenbeck dynamics”. In: *SIAM Journal on Financial Mathematics* 13.1 (2021).
- [Boe+16] Horatio Boedihardjo et al. “The signature of a rough path: Uniqueness”. In: *Advances in Mathematics* 293 (Apr. 2016), pp. 720–737. ISSN: 10902082. DOI: [10.1016/j.aim.2016.02.011](https://doi.org/10.1016/j.aim.2016.02.011).

- [Bon+19] Patric Bonnier et al. “Deep Signature Transforms”. In: *Advances in Neural Information Processing Systems* 32 (May 2019). ISSN: 10495258. URL: <https://arxiv.org/abs/1905.08494v2>.
- [BCD98] F. Jay Breidt, Nuno Crato, and Pedro De Lima. “The detection and estimation of long memory in stochastic volatility”. In: *Journal of Econometrics* 83.1-2 (Mar. 1998), pp. 325–348. ISSN: 0304-4076. DOI: [10.1016/S0304-4076\(97\)00072-9](https://doi.org/10.1016/S0304-4076(97)00072-9).
- [BT23] Alessio Brini and Daniele Tantari. “Deep Reinforcement Trading with Predictable Returns”. 2023.
- [Bue+21] Hans Buehler et al. “Generating Financial Markets With Signatures”. In: *Risk* (2021). URL: <https://www.risk.net/cutting-edge/banking/7841726/generating-financial-markets-with-signatures>.
- [Büh+18] Hans Bühler et al. *Deep Hedging*. Feb. 2018. URL: <http://arxiv.org/abs/1802.03042>.
- [CDM22] Alvaro Cartea, Fayçal Drissi, and Marcello Monga. “Execution and Statistical Arbitrage with Signals in Multiple Automated Market Makers”. 2022. URL: <https://ssrn.com/abstract=4388104>.
- [CAS22] Álvaro Cartea, Imanol Pérez Arribas, and Leandro Sánchez-Betancourt. “Double-Execution Strategies Using Path Signatures”. In: <https://doi.org/10.1137/21M1456467> 13.4 (Nov. 2022), pp. 1379–1417. ISSN: 1945497X. DOI: [10.1137/21M1456467](https://doi.org/10.1137/21M1456467). URL: <https://pubs.siam.org/doi/10.1137/21M1456467>.
- [CDO23] Álvaro Cartea, Fayçal Drissi, and Pierre Osselin. “Bandits for Algorithmic Trading with Signals”. In: *SSRN Electronic Journal* (June 2023). ISSN: 1556-5068. DOI: [10.2139/ssrn.4484004](https://doi.org/10.2139/ssrn.4484004). URL: <https://papers.ssrn.com/abstract=4484004>.
- [CJ15] Álvaro Cartea and Sebastian Jaimungal. “Algorithmic Trading of Co-Integrated Assets”. In: *SSRN Electronic Journal* (Aug. 2015). DOI: [10.2139/ssrn.2637883](https://doi.org/10.2139/ssrn.2637883). URL: <https://papers.ssrn.com/abstract=2637883>.
- [CT22] Thomas Cass and William F. Turner. “Topologies on unparameterised path space”. June 2022. URL: <http://arxiv.org/abs/2206.11153>.
- [CR83] Gary Chamberlain and Michael Rothschild. “Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets”. In: *Econometrica* 51.5 (Sept. 1983), p. 1281. ISSN: 00129682. DOI: [10.2307/1912275](https://doi.org/10.2307/1912275).
- [Che77] Kuo Tsai Chen. “Iterated path integrals”. In: *Bulletin of the American Mathematical Society* 83.5 (1977), pp. 831–879. ISSN: 0002-9904. DOI: [10.1090/S0002-9904-1977-14320-6](https://doi.org/10.1090/S0002-9904-1977-14320-6). URL: <https://www.ams.org/bull/1977-83-05/S0002-9904-1977-14320-6/>.
- [Che57] Kuo-Tsai Chen. “Integration of Paths, Geometric Invariants and a Generalized Baker-Hausdorff Formula”. In: *The Annals of Mathematics* 65.1 (Jan. 1957), p. 163. ISSN: 0003486X. DOI: [10.2307/1969671](https://doi.org/10.2307/1969671).
- [CK16] Ilya Chevyrev and Andrey Kormilitzin. *A Primer on the Signature Method in Machine Learning*. Mar. 2016. URL: <http://arxiv.org/abs/1603.03788>.
- [CL13] Ilya Chevyrev and Terry Lyons. “Characteristic functions of measures on geometric rough paths”. In: (July 2013). DOI: [10.1214/15-AOP1068](https://doi.org/10.1214/15-AOP1068). URL: <http://arxiv.org/abs/1307.3580%20http://dx.doi.org/10.1214/15-AOP1068>.
- [CO22] Ilya Chevyrev and Harald Oberhauser. “Signature Moments to Characterize Laws of Stochastic Processes”. In: *Journal of Machine Learning Research* 23 (2022), pp. 1–42. URL: <http://jmlr.org/papers/v23/20-1466.html>.
- [CC23] Henry Chiu and Rama Cont. “A model-free approach to continuous-time finance”. In: *Mathematical Finance* (2023). ISSN: 14679965. DOI: [10.1111/MAFI.12370](https://doi.org/10.1111/MAFI.12370).
- [CJC22] Anthony Coache, Sebastian Jaimungal, and Alvaro Cartea. “Conditionally Elicitable Dynamic Risk Measures for Deep Reinforcement Learning”. 2022.
- [Coh+23] Samuel N Cohen et al. “Nowcasting with signature methods”. 2023.
- [Con01] Rama Cont. *Empirical properties of asset returns: stylized facts and statistical issues*. Tech. rep. 2001. URL: <http://www.cmap.polytechnique.fr/>.

- [CMN23] Rama Cont, Alessandro Micheli, and Eyal Neuman. “Fast and Slow Optimal Trading with Exogenous Information”. 2023.
- [CI22] Giorgio Costa and Garud N. Iyengar. “Distributionally Robust End-to-End Portfolio Construction”. June 2022. URL: <http://arxiv.org/abs/2206.05134>.
- [CLO21] Dan Crisan, Alexander Lobbe, and Salvador Ortiz-Latorre. “Pathwise approximations for the solution of the non-linear filtering problem”. 2021.
- [CGS22] Christa Cuchiero, Guido Gazzani, and Sara Svaluto-Ferro. “Signature-based models: theory and calibration”. July 2022. URL: <http://arxiv.org/abs/2207.13136>.
- [Das22] Purba Das. “Roughness properties of paths and signals”. PhD thesis. University of Oxford, 2022.
- [DT23] Bruno Dupire and Valentin Tissot-Daguette. “Functional Expansions”. 2023.
- [DCS21] Joel Dyer, Patrick Cannon, and Sebastian M Schmon. “Deep Signature Statistics for Likelihood-free Time-series Models”. In: *ICML Workshop on Invertible Neural Networks, Normalizing Flows, and Explicit Likelihood Models* (2021).
- [Dye+22] Joel Dyer et al. “Approximate Bayesian Computation for Panel Data with Signature Maximum Mean Discrepancies”. In: *ICML Time-series Workshop* (2022).
- [Eng01] Robert Engle. “GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics”. In: *Journal of Economic Perspectives* 15.4 (2001), pp. 157–168. ISSN: 0895-3309. DOI: [10.1257/JEP.15.4.157](https://doi.org/10.1257/JEP.15.4.157).
- [FF93] Eugene F Fama and Kenneth R French. “Common risk factors in the returns on stocks and bonds”. In: *Journal of Financial Economics* 33 (1993), pp. 3–56.
- [FF15] Eugene F. Fama and Kenneth R. French. “A five-factor asset pricing model”. In: *Journal of Financial Economics* 116.1 (Apr. 2015), pp. 1–22. ISSN: 0304-405X. DOI: [10.1016/J.JFINECO.2014.10.010](https://doi.org/10.1016/J.JFINECO.2014.10.010).
- [Fer21] Adeline Fermanian. “Embedding and learning with signatures”. In: *Computational Statistics & Data Analysis* 157 (May 2021), p. 107148. ISSN: 0167-9473. DOI: [10.1016/J.CSDA.2020.107148](https://doi.org/10.1016/J.CSDA.2020.107148).
- [FHL16] Guy Flint, Ben Hambly, and Terry Lyons. “Discretely sampled signals and the rough Hoff process”. In: *Stochastic Processes and their Applications* 126.9 (2016), pp. 2593–2614.
- [For+22] Martin Forde et al. “Optimal trade execution for Gaussian signals with power-law resilience”. In: *Quantitative Finance* 22.3 (2022), pp. 585–596. ISSN: 14697696. DOI: [10.1080/14697688.2021.1950919](https://doi.org/10.1080/14697688.2021.1950919). URL: <https://www.tandfonline.com/action/journalInformation?journalCode=rquf20>.
- [FH20] Peter K Friz and Martin Hairer. *A Course on Rough Paths With an introduction to regularity structures*. 2020.
- [FV10] Peter K. Friz and Nicolas B. Victoir. *Multidimensional Stochastic Processes as Rough Paths*. Cambridge University Press, Feb. 2010. DOI: [10.1017/cbo9780511845079](https://doi.org/10.1017/cbo9780511845079).
- [Fuk21] Masaaki Fukasawa. “Volatility has to be rough”. In: *Quantitative Finance* 21.1 (2021), pp. 1–8. ISSN: 14697696. DOI: [10.1080/14697688.2020.1825781](https://doi.org/10.1080/14697688.2020.1825781). URL: <https://ideas.repec.org/a/taf/quantf/v21y2021i1p1-8.html%20https://ideas.repec.org/a/taf/quantf/v21y2021i1p1-8.html>.
- [GP13] Nicolae Gârleanu and Lasse Heje Pedersen. “Dynamic Trading with Predictable Returns and Transaction Costs”. In: *Journal of Finance* 68.6 (Dec. 2013), pp. 2309–2340. ISSN: 15406261. DOI: [10.1111/jofi.12080](https://doi.org/10.1111/jofi.12080).
- [GJR14] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. “Volatility is rough”. In: *Quantitative Finance* 18.6 (Oct. 2014), pp. 933–949. ISSN: 14697696. DOI: [10.48550/arxiv.1410.3394](https://arxiv.org/abs/1410.3394). URL: <https://arxiv.org/abs/1410.3394v1>.
- [GPZ21] Jorge Guijarro-Ordonez, Markus Pelger, and Greg Zanotti. “Deep Learning Statistical Arbitrage”. June 2021. URL: <http://arxiv.org/abs/2106.04028>.
- [GL22] Julien Guyon and Jordan Lekeufack. “Volatility Is (Mostly) Path-Dependent”. 2022. URL: <https://ssrn.com/abstract=4174589>.

- [Gyu+13] Lajos Gergely Gyurkó et al. *Extracting information from the signature of a financial data stream*. July 2013. URL: <http://arxiv.org/abs/1307.7244>.
- [HL10] Ben Hambly and Terry Lyons. “Uniqueness for the signature of a path of bounded variation and the reduced path group”. In: *Annals of Mathematics* 171.1 (2010), pp. 109–167. ISSN: 0003486X. DOI: [10.4007/ANNALS.2010.171.1.109](https://doi.org/10.4007/ANNALS.2010.171.1.109).
- [Hof06] Ben Hoff. “The Brownian Frame Process as a Rough Path”. PhD thesis. University of Oxford, 2006.
- [HTZ21] Blanka Horvath, Josef Teichmann, and Zan Zuric. “Deep Hedging under Rough Volatility”. In: *Swiss Finance Institute Research Paper Series* (Feb. 2021). URL: <http://arxiv.org/abs/2102.01962>.
- [IH23] Zacharia Issa and Blanka Horvath. “Non-parametric online market regime detection and regime clustering for multidimensional and path-dependent data structures”. June 2023. URL: <https://arxiv.org/abs/2306.15835v1>.
- [Iss+23] Zacharia Issa et al. “Non-adversarial training of Neural SDEs with signature kernel scores”. May 2023. URL: <https://arxiv.org/abs/2305.16274v1>.
- [JL20] Antoine Jacquier and Chloé Lacombe. “Path-dependent Volatility Models”. 2020.
- [Jai+21] Sebastian Jaimungal et al. “Robust Risk-Aware Reinforcement Learning”. In: *SIAM Journal on Financial Mathematics* 13 (Aug. 2021), pp. 213–226. URL: <http://arxiv.org/abs/2108.10403>.
- [KEA19] Can B Kalayci, Okkes Ertenlice, and Anil Akbay. “A comprehensive review of deterministic models and applications for mean-variance portfolio optimization”. In: *Expert Systems With Applications* 125 (2019), pp. 345–368. DOI: [10.1016/j.eswa.2019.02.011](https://doi.org/10.1016/j.eswa.2019.02.011). URL: <https://doi.org/10.1016/j.eswa.2019.02.011>.
- [KLA20] Jasdeep Kalsi, Terry Lyons, and Imanol Perez Arribas. “Optimal execution with rough path signatures”. In: *SIAM Journal on Financial Mathematics* 11.2 (May 2020). URL: <http://arxiv.org/abs/1905.00728>.
- [KL21] Patrick Kidger and Terry Lyons. “Signatory: differentiable computations of the signature and logsignature transforms, on both CPU and GPU”. In: *ICLR 2021 - 9th International Conference on Learning Representations* (Jan. 2021). URL: <https://github.com/patrick-kidger/signatory>.
- [LN19] Charles-Albert Lehalle and Eyal Neuman. “Incorporating Signals into Optimal Trading”. In: *Finance and Stochastics* 23 (Apr. 2019), pp. 275–311. URL: <http://arxiv.org/abs/1704.00847>.
- [Lem+14] Yves Lemperiere et al. *Risk Premia: Asymmetric Tail Risks and Excess Returns*. Sept. 2014. DOI: [10.2139/ssrn.2502743](https://doi.org/10.2139/ssrn.2502743). URL: <https://papers.ssrn.com/abstract=2502743>.
- [LL15] Tim Leung and Xin Li. “Optimal Mean Reversion Trading with Transaction Costs and Stop-Loss Exit”. In: *International Journal of Theoretical and Applied Finance* 18.3 (2015).
- [LLN13] Daniel Levin, Terry Lyons, and Hao Ni. *Learning from the past, predicting the statistics for the future, learning an evolving system*. Sept. 2013. URL: <http://arxiv.org/abs/1309.0260>.
- [LH23] Yannick Limmer and Blanka Horvath. “Robust Hedging GANs”. July 2023. URL: <https://arxiv.org/abs/2307.02310v1>.
- [Lyo14] Terry Lyons. “Rough paths, Signatures and the modelling of functions on streams”. In: *Proceedings of the International Congress of Mathematicians, Korea*. May 2014. URL: <http://arxiv.org/abs/1405.4537>.
- [LM22] Terry Lyons and Andrew D. McLeod. *Signature Methods in Machine Learning*. June 2022. URL: <https://arxiv.org/abs/2206.14674v4>.
- [LNA19a] Terry Lyons, Sina Nejad, and Imanol Perez Arribas. “Non-parametric Pricing and Hedging of Exotic Derivatives”. In: *Applied Mathematical Finance* 27 (2019), pp. 457–494.

- [LNA19b] Terry Lyons, Sina Nejad, and Imanol Perez Arribas. *Numerical method for model-free pricing of exotic derivatives using rough path signatures*. May 2019. URL: <http://arxiv.org/abs/1905.01720>.
- [Lyo98] Terry J. Lyons. “Differential equations driven by rough signals.” In: *Revista Matemática Iberoamericana* 14.2 (1998), pp. 215–310. ISSN: 0213-2230.
- [LCL07] Terry J. Lyons, Michael Caruana, and Thierry Lévy. “Differential Equations Driven by Rough Paths”. In: Lecture Notes in Mathematics 1908 (2007). DOI: [10.1007/978-3-540-71285-5](https://doi.org/10.1007/978-3-540-71285-5). URL: <http://link.springer.com/10.1007/978-3-540-71285-5>.
- [Mar52] Harry Markowitz. *Portfolio Selection*. Tech. rep. 1. 1952, pp. 77–91.
- [MMB23] Rudy Morel, Stéphane Mallat, and Jean-Philippe Bouchaud. “Path Shadowing Monte-Carlo”. Aug. 2023. URL: <https://arxiv.org/abs/2308.01486v1>.
- [MOP12] Tobias J. Moskowitz, Yao Hua Ooi, and Lasse Heje Pedersen. “Time series momentum”. In: *Journal of Financial Economics* 104.2 (May 2012), pp. 228–250. ISSN: 0304405X. DOI: [10.1016/j.jfineco.2011.11.003](https://doi.org/10.1016/j.jfineco.2011.11.003).
- [MPW08] Supakorn Mudchanatongsuk, James A. Primbs, and Wilfred Wong. “Optimal pairs trading: A stochastic control approach”. In: *Proceedings of the American Control Conference* (2008), pp. 1035–1039. ISSN: 07431619. DOI: [10.1109/ACC.2008.4586628](https://doi.org/10.1109/ACC.2008.4586628).
- [NER92] Victor Ng, Robert F Engle, and Michael Rothschild. *A multi-dynamic-factor model for stock returns*. Tech. rep. 1992, p. 2455266.
- [Ni+20] Hao Ni et al. “Conditional Sig-Wasserstein GANs for Time Series Generation”. June 2020. URL: <https://arxiv.org/abs/2006.05421v1>.
- [Ni+21] Hao Ni et al. “Sig-Wasserstein GANs for Time Series Generation”. In: *ICAIF 2021 - 2nd ACM International Conference on AI in Finance* (Nov. 2021). DOI: [10.1145/3490354.3494393](https://doi.org/10.1145/3490354.3494393). URL: <https://arxiv.org/abs/2111.01207v1>.
- [Per20] Imanol Perez Arribas. “Signatures in machine learning and finance”. PhD thesis. University of Oxford, 2020.
- [PP16] Nicolas Perkowski and David J. Prömel. “Pathwise stochastic integrals for model free finance”. In: *Bernoulli* 22.4 (Nov. 2016), pp. 2486–2520. ISSN: 13507265. DOI: [10.3150/15-BEJ735](https://doi.org/10.3150/15-BEJ735).
- [PZ22] Ruan Pretorius and Terence van Zyl. “Deep Reinforcement Learning and Convex Mean-Variance Optimisation for Portfolio Management”. Feb. 2022. URL: <http://arxiv.org/abs/2203.11318>.
- [RSB17] Adam Rej, Philip Seager, and Jean-Philippe Bouchaud. “You are in a drawdown. When should you start worrying?” In: *Wilmott* 2018.93 (July 2017), pp. 56–59. DOI: [10.1002/wilm.10646](https://doi.org/10.1002/wilm.10646). URL: <https://arxiv.org/abs/1707.01457v2>.
- [RA12] Richard J Martin and Ali Bana. “Non-linear momentum strategies”. In: *Risk* 25.11 (Oct. 2012), pp. 60–65. URL: <https://www.risk.net/derivatives/structured-products/2221212/non-linear-momentum-strategies>.
- [RD12] Richard J Martin and David Zou. “Momentum trading: ’skews me’”. In: *Risk* 25.8 (July 2012), pp. 84–89. URL: <https://www.risk.net/derivatives/structured-products/2194247/momentum-trading-skews-me>.
- [Rig16a] Candia Riga. *A pathwise approach to continuous-time trading*. Feb. 2016. URL: <http://arxiv.org/abs/1602.04946>.
- [Rig16b] Candia Riga. “Pathwise functional calculus and applications to continuous-time finance”. PhD thesis. Feb. 2016. URL: <http://arxiv.org/abs/1602.04523>.
- [SC22] Leandro Sánchez-Betancourt and Alvaro Cartea. “Brokers and Informed Traders: dealing with toxic flow and extracting trading signals”. 2022. URL: <https://ssrn.com/abstract=4265814>.
- [Sha64] William F. Sharpe. “Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk”. In: *The Journal of Finance* 19.3 (1964), pp. 425–442. ISSN: 15406261. DOI: [10.1111/J.1540-6261.1964.TB02865.X](https://doi.org/10.1111/J.1540-6261.1964.TB02865.X).

- [Soo+23] Srijan Sood et al. “Deep Reinforcement Learning for Optimal Portfolio Allocation: A Comparative Study with Mean-Variance Optimization”. 2023. URL: [www.aaai.org](http://www.aaai.org).
- [SW10] James H Stock and Mark W Watson. “Dynamic Factor Models”. In: *Handbook of Macroeconomics* 2A (2010). ISSN: 1574-0048. URL: <http://dx.doi.org/10.1016/bs.hesmac.2016.04.002>.
- [Van+21] Pieter M Van Staden et al. “A data-driven neural network approach to dynamic factor investing with transaction costs”. 2021. URL: [www.invesco.com..](http://www.invesco.com..)
- [Vid04] Ganapathy. Vidyamurthy. “Pairs trading: Quantitative Methods and Analysis”. In: (2004), p. 210. URL: <https://www.wiley.com/en-gb/Pairs+Trading%3A+Quantitative+Methods+and+Analysis-p-9780471460671>.
- [WZ19] Haoran Wang and Xun Yu Zhou. *Continuous-Time Mean-Variance Portfolio Selection: A Reinforcement Learning Framework*. Apr. 2019. URL: <http://arxiv.org/abs/1904.11392>.
- [Wan19] Kewei Wang. “Portfolio Optimisation under Rough Stochastic Volatility via Machine Learning”. 2019.
- [WKM23] Magnus Wiese, Ralf Korn, and Phillip Murray. *Sig-Splines: universal approximation and convex calibration of time series generative models*. July 2023. URL: <https://papers.ssrn.com/abstract=4514421>.
- [Wie+19] Magnus Wiese et al. “Quant GANs: Deep Generation of Financial Time Series”. In: *Quantitative Finance* 20.9 (July 2019), pp. 1419–1440. DOI: [10.1080/14697688.2020.1730426](https://doi.org/10.1080/14697688.2020.1730426). URL: <http://arxiv.org/abs/1907.06673%20http://dx.doi.org/10.1080/14697688.2020.1730426>.
- [Zha+21] Chao Zhang et al. “A Universal End-to-End Approach to Portfolio Optimization via Deep Learning”. Nov. 2021. URL: <http://arxiv.org/abs/2111.09170>.
- [ZZR20a] Zihao Zhang, Stefan Zohren, and Stephen Roberts. “Deep Learning for Portfolio Optimization”. In: *The Journal of Financial Data Science* 2.4 (2020), pp. 8–20.
- [ZZR20b] Zihao Zhang, Stefan Zohren, and Stephen Roberts. “Deep Reinforcement Learning for Trading”. In: *The Journal of Financial Data Science* 2.2 (2020), pp. 25–40.