UNIVERSITAT DE BARCELONA

Facultat de Matemàtiques
i Informàtica

GRAU DE MATEMÀTIQUES

Treball final de grau

THE EXPECTED SIGNATURE APPROACH FOR FINANCIAL TIME SERIES FORECASTING

Autor: Òscar Romero Alarcón

Director: Dr. Josep Vives i Santa Eulàlia
Realitzat a: Departament
de Matemàtiques i Informàtica

Barcelona, 10 de juny de 2024

Abstract

In recent years, rough path theory, which models the interactions between highly oscillatory and non linear systems, has emerged as a prominent topic in mathematical finance. The study builds upon the concept of the signature of a path, a mathematical object introduced by Terry Lyons in his foundational works on rough path theory ([14] and [13]).

In this work, we bring this theory to a particular model called Expected Signature (ES) model, presented in [3], which utilizes the signature of a path to perform regression between stochastic paths, treating them as input and output variables, to identify the functional that relates this variables. It involves defining the expected signature as an element within a probability space to establish the conditional distribution of the dependent response, facilitating regression-based forecasting.

Key theoretical foundations behind the construction of the ES model are established during the thesis, including the critical theorem that linear functionals can approximate continuous functions over compact sets of bounded variation paths. This means the ES model can effectively transform intricate relationships within the data into linear ones. As a result, the model can make more reliable and accurate predictions. A case study validates the ES model's practical application by showing its potential to outperform traditional time series models like ARIMA in certain scenarios.

Resum

En els darrers anys, la teoria dels camins accidentats (*rough paths*), que modela les interaccions entre sistemes altament oscil·latoris i no lineals, ha sorgit com un tema destacat en investigació i recerca a l'àmbit de la matemàtica financera. L'estudi es basa en el concepte de la signatura d'un camí, un objecte matemàtic introduït per Terry Lyons en les seves petjades inicials al voltant d'aquesta teoria ([14] i [13]).

En aquest treball, portem aquesta teoria a un model particular anomenat model de la Signatura Esperada (model ES), presentat a [3], que utilitza la signatura d'un camí per fer regressió entre camins estocàstics, tractant-los com a variables d'entrada i sortida, per identificar el funcional que relaciona aquestes variables. Es tracta de definir la signatura esperada com a un element dins d'un espai de probabilitat per establir la distribució condicional de la resposta dependent, facilitant la previsió de futures dades a partir d'un model de regressió.

Durant la tesi s'estableixen els fonaments teòrics clau subjacents a la construc-

ció del model ES, inclòs el crí teorema que estableix que els funcionals lineals poden aproximar funcions contínues sobre conjunts compactes de trajectòries de variació acotades. Fet que significa que el model ES pot transformar eficaÃ§ment les relacions intrínseques de les dades en relacions lineals i, com a resultat, fer prediccions més fiables i precises. Un estudi de cas final valida l'aplicació pràctica del model ES mostrant-ne el potencial per superar, en determinades situacions, models tradicionals de sèries temporals, com el model SARIMA.

2020 Mathematics Subject Classification. 60L10, 60L20, 62M10, 62M15

Acknowledgments

First and foremost, I want to express my deepest gratitude to my advisor, Dr. Josep Vives, for his invaluable guidance and supervision of this thesis. His trust in my work to explore a topic of mutual interest, his availability to address my doubts whenever necessary, and his direction towards the research field have been instrumental. I have enjoyed every minute spent on this journey.

I extend my sincere thanks to Pere Díaz (University of Oslo) and Hao Ni (University College London) for their advice and theoretical insights during the most challenging moments of my thesis.

I am profoundly grateful to my family and friends for their unwavering interest in my academic journey and for providing the best possible environment for my success. They have never hesitated to support my ideas, no matter how unconventional, and have always made time for me when I needed it most. Even when I enthusiastically explained my research, they listened attentively and stood by me, sharing my excitement. They are the driving force of my daily life.

Specifically, I want to thank Ana Victoria, a fellow faculty member who has been really close in my research. As she was conducting a similar study for her master's thesis, we exchanged and shared ideas that propelled us both towards better results.

Òscar Romero

Contents

Introduction **2**

1 Preliminaries **1**

1.1 Tensor product 1

1.2 Tensor algebra 2

1.3 Path Integrals 4

1.3.1 Paths 4

1.3.2 Iterated Integrals 6

2 Signature of a path **9**

2.1 Definition 9

2.2 Interpretation of the signature coordinates 10

2.3 Properties of the signature 12

3 The expected signature framework **23**

3.1 Law on signatures 23

3.2 The expected signature model 24

3.2.1 Model calibration and prediction 26

4 The ES approach as a time series model **27**

4.1 Time series embedding 27

4.1.1 Time joined embedding 30

4.2 The signature of a time series 31

4.3 Expected Signature model for time series 34

4.4 Classical time series models as ES-model special cases 35

5 Using the ES model for financial time series prediction **39**
5.1 Unveiling Computational Insights: AR as a particular case of the ES approach . . . 39
5.2 ES model for financial time series forecasting . . . 44
5.2.1 Catfish sales and forecasting . . . 44

6 Conclusions **49**

Bibliography **51**

Appendix **53**
6.1 Traditional time series models . . . 53
6.2 Time series autocorrelation . . . 54

Chapter 1

Preliminaries

In order to tackle the formalization of the signature of a path, it is important to introduce some formalization and definitions. We will present the space where the signature lie and the operations that will be present on its computation. These basic concepts can also be found in [4] and [13].

1.1 Tensor product

Definition 1.1. (Tensor product) Given two vector spaces V and W, the tensor product $V \otimes W$ is a vector space to which is associated a bilinear map $V \times W \to V \otimes W$ that maps a pair (v,w), $v \in V$ and $w \in W$, to an element $v \otimes w \in V \otimes W$.

Example 1.2. As a basic example, suppose $\mathbf{u} = [u_1, u_2]$ and $\mathbf{v} = [v_1, v_2, v_3]$. The tensor product $\mathbf{u} \otimes \mathbf{v}$ would result in a 2×3 matrix:

$$\mathbf{u} \otimes \mathbf{v} = \begin{pmatrix} u_1 \cdot v_1 & u_1 \cdot v_2 & u_1 \cdot v_3 \\ u_2 \cdot v_1 & u_2 \cdot v_2 & u_2 \cdot v_3 \end{pmatrix}$$

Consider B_v and B_w the basis of V and W respectively, the set $\{v \otimes w \mid v \in B_v, w \in B_w\}$ is straightforwardly a basis of $V \otimes W$, which is called the tensor product of the bases B_V and B_W.

Intuitively, the tensor product is a way to create a new space consisting of combinations of vectors from V and W. Without delving into the specific formal definition of the operation $v \otimes w$, we can state the following:

$$dim(V \otimes W) = (dimV)(dimW) \tag{1.1}$$

We will firstly generalize our definition using a real Banach space with dimension d, $\mathbb{E}^d$, the space in which the path is defined. Over this space, we can define the tensor product with itself as it follows.

Definition 1.3. (Tensor power) The n-th tensor power of $\mathbb{E}^d$ is defined as:

$$(\mathbb{E}^d)^{\otimes n} := \underbrace{\mathbb{E}^d \otimes \mathbb{E}^d \otimes \cdots \otimes \mathbb{E}^d}_{n \text{ times}} \tag{1.2}$$

For simplicity, we will simply use $\mathbb{E}$ when considering the d-dimensional space $\mathbb{E}^d$. Moreover, as our paths are defined in a finite dimensional space, its basis are $\{e_1, e_2, \ldots, e_d\}$, so the tensor power $\mathbb{E}^{\otimes n}$ has the elements $\{e_{i_1} \otimes e_{i_2} \otimes ... \otimes e_{i_d}\}$ as basis, where $i_1, ..., i_d \in \{1, ..., d\}$.

Taking the previous definitions, we can trivially identify the n-th tensor power with the space of non-commuting polynomials of degree n in d variables. So, $\mathbb{E}^{\otimes n}$ is isomorphic to the space spanned by indexes of length n in the possible values in the set $\{1, ..., d\}$.

Remark 1.4. $\mathbb{E}^{\otimes 0}$ is defined as the underlying scalar field of the vector space $\mathbb{E}$. In our case, we can state that $\mathbb{E}^{\otimes 0} = \mathbb{R}$.

We will now define the conditions in which the tensor powers is endowed with an admissible norm in order to extend properties from the original space into its tensor powers.

Definition 1.5. Considering $\mathbb{E}$, the previously mentioned Banach space, its tensor powers are endowed with an admissible norm $|.|$, if:

1. For each $n \geq 1$, the symmetric group S_n acts by isometry on $\mathbb{E}^{\otimes n}$, i.e.

$$|\sigma v| = |v|, \forall v \in \mathbb{E}^{\otimes n}, \forall \sigma \in S_n. \tag{1.3}$$

2. $\forall n, m \geq 1$,

$$|v \otimes w| \leq |v||w|, \forall v \in \mathbb{E}^{\otimes n}, w \in \mathbb{E}^{\otimes m}. \tag{1.4}$$

1.2 Tensor algebra

Definition 1.6. We define a formal $\mathbb{E}$-tensor series as a sequence of tensors $\left(a_i \in \mathbb{E}^{\otimes i}\right)_{i \in \mathbb{N}}$, denoted by $a = (a_0, a_1, \ldots)$. If we consider the space $\mathbb{E}$ as $\mathbb{R}^d$, the formal tensor series can be called as (non-commuting) formal power series and its elements are of the form

$$\sum_{k=0}^{\infty} \sum_{i_1, \ldots, i_k \in \{1, \ldots d\}} \lambda_{i_1, \ldots, i_k} e_{i_1} \ldots e_{i_k}, \tag{1.5}$$

where the second summation runs over all multi-indexes $(i_1, \ldots, i_k)$, $i_1, \ldots, i_k \in \{1, \ldots, d\}$, and $\lambda_{i_1, \ldots, i_k}$ are real numbers.

In order to construct the algebra, the binary operations we can define on E-tensor series are the addition + and product $\otimes$:

Definition 1.7. Consider the $\mathbb{E}$-tensor series $\mathbf{a} = (a_0, a_1, \dots)$ and $\mathbf{b} = (b_0, b_1, \dots)$, the addition is defined as

$$\mathbf{a} + \mathbf{b} = (a_0 + b_0, a_1 + b_1, \dots) \tag{1.6}$$

and the product as

$$\mathbf{a} \otimes \mathbf{b} = (c_0, c_1, \dots) \tag{1.7}$$

where each term of the resulting $\mathbb{E}$-tensor series is the sum of all the possible combinations of tensor products between the coordinates of both **a** and **b**. For each $i \geq 0$,

$$c_i = \sum_{k=0}^{i} a_k \otimes b_{i-k}. \tag{1.8}$$

We will use the notation **1** for the series $(1, 0, \dots)$, and **0** for the series $(0, 0, \dots)$. Given a scalar $\lambda \in \mathbb{R}$, then we define $(\lambda a_0, \lambda a_1, \dots)$ as $\lambda\mathbf{a}$.

Following the previous notation when considering $\mathbb{E} = \mathbb{R}^d$, the addition can be seen as

$$\begin{aligned}\left(\sum_{k=0}^{\infty} \sum_{i_1,\dots,i_k \in \{1,\dots d\}} \lambda_{i_1,\dots,i_k} e_{i_1} \dots e_{i_k}\right) + \left(\sum_{k=0}^{\infty} \sum_{i_1,\dots,i_k \in \{1,\dots d\}} \mu_{i_1,\dots,i_k} e_{i_1} \dots e_{i_k}\right) \\ = \sum_{k=0}^{\infty} \sum_{i_1,\dots,i_k \in \{1,\dots d\}} \left(\lambda_{i_1,\dots,i_k} + \mu_{i_1,\dots,i_k}\right) e_{i_1} \dots e_{i_k}\end{aligned} \tag{1.9}$$

and the first terms of the product as

$$\begin{aligned}&\left(\sum_{k=0}^{\infty} \sum_{i_1,\dots,i_k \in \{1,\dots d\}} \lambda_{i_1,\dots,i_k} e_{i_1} \dots e_{i_k}\right) \otimes \left(\sum_{k=0}^{\infty} \sum_{i_1,\dots,i_k \in \{1,\dots d\}} \mu_{i_1,\dots,i_k} e_{i_1} \dots e_{i_k}\right) \\ &= \lambda_0\mu_0 + \sum_{i=1}^{d} \left(\lambda_0\mu_i + \lambda_i\mu_0\right) e_i + \sum_{i,j=1}^{d} \left(\lambda_0\mu_{i,j} + \lambda_i\mu_j + \lambda_{i,j}\mu_0\right) e_i e_j + \cdots\end{aligned} \tag{1.10}$$

Definition 1.8. (Tensor algebra) We define $T((E))$ as the vector space of all formal $\mathbb{E}$-tensors series. $T((E))$ with the operations + and $\otimes$ and the action of $\mathbb{R}$ is an associative and unital algebra over $\mathbb{R}$, called the tensor algebra. An element $\mathbf{a} = (a_0, a_1, \dots)$ of $T((E))$ is invertible if and only if $a_0 \neq 0$. In particular, the subset $\{\mathbf{a} \in T((E)) \mid a_0 = 1\}$ forms a group.

The group $\{\mathbf{a} \in T((E)) \mid a_0 = 1\}$ will be useful when defining the signature of a path, as it will be defined as an element of this group.

Concretely, if $\mathbb{E} = \mathbb{R}^d$, it is called the non-commutative tensor algebra of $\mathbb{R}$, where the elements e_1e_2 are distinct than e_2e_1.

As we will see, it will be interesting to look only at finitely many terms of $T((E))$. In order to define elements of this kind, we will use the space $B_n = \{\mathbf{a} = (a_0, a_1, \ldots) \mid a_0 = \ldots = a_n = 0\}$, $n \geq 0$, consisting of the formal series with no monomials of degree less or equal to n. Using this we define the truncated tensor algebra:

Definition 1.9. Let $n \geq 1$ be an integer. The truncated tensor algebra of order n of $\mathbb{E}$ is defined as the quotient algebra

$$T^{(n)}(E) = T((E))/B_n. \tag{1.11}$$

Definition 1.10. We define the canonical homomorphism $T((E)) \longrightarrow T^{(n)}(E)$ as ρ_n. Remarking that $T^{(n)}(E)$ is embedded in $T((E))$ as a linear subspace, but not as a sub-algebra.

We can trivially see that $T^{(n)}(E)$ is isomorphic to $\bigoplus_{k=0}^{n} E^{\otimes k}$ equipped with the product

$$(a_0, \ldots, a_n) \otimes (b_0, \ldots, b_n) = (c_0, \ldots, c_n), \tag{1.12}$$

where, for all $k \in \{0, \ldots, n\}$, $c_k = a_0 \otimes b_k + a_1 \otimes b_{k-1} + \ldots + a_k \otimes b_0$. Which means that the homomorphism ρ_n doesn't take into consideration all the combinations $a_i \otimes b_j$ with $i + j > n$ (terms of degree greater than n).

1.3 Path Integrals

1.3.1 Paths

In order to move on with the theory of this work, it is important to present the definition of the mathematical elements that will be studied, the paths. Basic definitions of paths in $\mathbb{R}^d$ can be found in [5].

Definition 1.11. (Path) A path X in $\mathbb{E}$ is a continuous mapping from some time interval $[a, b]$ to $\mathbb{R}^d$, written as $X : [a, b] \mapsto \mathbb{R}^d$.

For our discussion, we will focus on non-smooth paths that are piecewise differentiable. By "non-smooth paths," we mean paths that do not have derivatives of all orders, but if the path is divided into segments in a certain way, each segment is itself differentiable. This property is what we mean by *piecewise differentiable*.

We now present a concept that is very useful in this theory as it ensures the well-definedness of the signature, the *p-variation* of a path, which quantifies the roughness or variability of a path over a given interval. A more extended definition can be found in [13] and [3].

Definition 1.12. (p-variation of a path) Let $p \geq 1$ be a real number and $X : J \rightarrow \mathbb{E} := \mathbb{R}^d$ be a d-dimensional path where J is a compact interval. X is of finite p-variation for certain p if

$$\|X\|_{p,J} = \left(\sup_{\mathcal{D}_J \subset J} \sum_l \|X_{t_l} - X_{t_{l-1}}\|^p \right)^{1/p} < \infty, \tag{1.13}$$

taking the supremum over all possible finite partitions of the interval J. Let $\mathcal{V}^p(J, \mathbb{R}^d)$ denote the set of any continuous path $X : J \rightarrow \mathbb{R}^d$ of finite p-variation.

Definition 1.13. (p-variation norm) The p-variation norm of a path $X \in \mathcal{V}^p(J, \mathbb{R}^d)$ is defined as

$$\|X\|_{p-var} = \|X\|_{p,j} + \sup_{t \in J} |X_t|$$

In order to further define properties of the signature for non-smooth piece-wise differentiable paths, we need to define some algebraic properties of paths:

Definition 1.14. Given two continuous paths $X : [0,s] \longrightarrow \mathbb{E}$ and $Y : [s,t] \longrightarrow \mathbb{E}$, their concatenation is defined as the path

$$(X * Y)_u = \begin{cases} X_u, & \text{if } u \in [0,s]; \\ X_s + Y_u - Y_s, & \text{if } u \in [s,t]. \end{cases} \tag{1.14}$$

Remark 1.15. This is an associative operation between continuous paths with domains defined over consecutive intervals.

From this point forward, our discussion will be limited to paths that are elements of the set $\mathcal{V}^1(J, E)$, commonly referred to as *1-variation paths* or paths of *bounded variation*.

This kind of paths are exactly the set of functions whose first derivatives exist almost everywhere. It is therefore not a particularly restrictive assumption, as it contains, for example, all Lipschitz functions. In particular, if X is continuously differentiable, and $\dot{X}$ is its first derivative with respect to t, then

$$\|X\|_{1-var} = \int_0^1 \|\dot{X}\| dt.$$

By these means, we observe that it is possible to define Riemann-Stieltjes integrals along paths using the bounded variation property.

1.3.2 Iterated Integrals

We now present the definition of path integral against a fixed function, for simplicity we will set $\mathbb{E} := \mathbb{R}^d$ as done in [5]. For a one-dimensional path X: $[a,b] \mapsto \mathbb{R}$ and $f : \mathbb{R} \mapsto \mathbb{R}$, the path integral of X against f is defined as the usual Riemann integral

$$\int_a^b f\left(X_t\right) dX_t = \int_a^b f\left(X_t\right) (dX_t/dt)dt. \tag{1.15}$$

Generally, we can integrate any path $X^1 : [a,b] \mapsto \mathbb{R}$ against another path $X^2 : [a,b] \mapsto \mathbb{R}$. Following this, we can define

$$\int_a^b X_t^1 dX_t^2 = \int_a^b X_t^1 (dX_t^2/dt)dt \tag{1.16}$$

An intuitive example that will be really useful in our work is the following:

Example 1.16. Consider the path $X_t = \{X_t^1, X_t^2\} = \{1, X_t^2\}$ for all $t \in [a,b]$. It follows that $dX_t^1 = 1$ for all $t \in [a,b]$, and so the path integral of X^1 against X^2 is the increment of X_t^2 at the interval $[a,b]$:

$$\int_a^b X_t^1 dX_t^2 = \int_a^b (dX_t^2/dt)dt = X_b^2 - X_a^2 \tag{1.17}$$

Example 1.17. Consider a linear path $X : [0,1] \to \mathbb{R}^d$, i.e.,

$$X_t = \left(X_t^1, \ldots, X_t^d\right) = \left(a_1 + b_1 t, \ldots, a_d + b_d t\right), \quad 0 \le t \le 1, \quad a_1, \ldots, a_d, b_1, \ldots b_d \in \mathbb{R}.$$

Then, given $t_1, t_2 \in [0,1]$

$$\int_{t_1}^{t_2} dX_u = \int_{t_1}^{t_2} \dot{X}_u du = \begin{pmatrix} \int_{t_1}^{t_2} b_1 du \\ \vdots \\ \int_{t_1}^{t_2} b_d du \end{pmatrix} = \begin{pmatrix} b_1(t_2 - t_1) \\ \vdots \\ b_d(t_2 - t_1) \end{pmatrix}$$

A simple calculation is now presented in order to understand the signature definition, as done in [5]. If we consider a path $X : [a,b] \to \mathbb{R}^d$, where, at time t, its components are written as $(X_t^1, ..., X_t^d)$. Taking $i \in \{1, .., d\}$ we define the following:

$$S(X)_{a,t}^{(i)} = \int_{a<s<t} dX_s^i = X_t^i - X_a^i, \tag{1.18}$$

or, equivalently, the increment of the coordinate X^i of the path on the interval $[a,t]$.

Now for any pair $i, j \in \{1, \dots, d\}$, we can define

$$S(X)^{(i,j)}_{a,t} = \int_{a<s<t} S(X)^i_{a,s} dX^j_s = \int_{a<r<s<t} dX^i_r dX^j_s, \tag{1.19}$$

which correspond to integration over a simplex (i.e. triangle in dimension 2).

We can continue recursively and define it in by any collection of indexes.

Definition 1.18. *Given $k \geq 1$ integer and $i_1, \dots, i_k \in \{1, \dots, d\}$, the k-fold iterated integral of X along the indexes $i_1, \dots, i_k$ is defined as*

$$S(X)^{(i_1,\dots,i_k)}_{a,t} = \int_{a<s<t} S(X)^{(i_1,\dots,i_{k-1})}_{a,s} dX^{i_k}_s = \int_{a<t_k<t} \cdots \int_{a<t_1<t_2} dX^{i_1}_{t_1} \dots dX^{i_k}_{t_k}. \tag{1.20}$$

The previous iterated integrals are all well defined because each $S(X)^{(i_1,\dots,i_{k-1})}_{a,t}$ is itself a real valued path.

Example 1.19. Let's see an example with the path $X_t = \{X^1_t, X^2_t\} = \{2t, 3+t\}$, $dX_t = \{2, 1\}$. Using basic integration rules we can compute some:

$$S(X)^{(1)}_{0,3} = \int_{0<t<3} dX^1_t = X^1_3 - X^1_0 = 6, \tag{1.21}$$

$$S(X)^{(1,1)}_{0,3} = \int_{0<s<t<3} dX^i_s dX^j_t = \int_0^3 [\int_0^t 2ds]2dt = 18, \tag{1.22}$$

$$S(X)^{(2,1)}_{0,3} = \int_{0<s<t<3} dX^2_s dX^1_t = \int_0^3 [\int_0^t ds]2dt = 9, \tag{1.23}$$

Chapter 2

Signature of a path

2.1 Definition

After defining the fundamental concepts necessaries for the study of path signatures, we will now turn our attention to the formal definition of the signature for a path of bounded variation, along with its informal interpretation. This approach aims to provide deeper insights into potential applications of the signature. It is important to maintain the formalities when defining the signature in order to be able to prove its properties later on. Following this, a similar definition as the one given at [4] is presented.

Definition 2.1. (Signature of a path) Let I denote a time interval in $\mathbb{R}^+$. Let $X : I \longrightarrow \mathbb{E}^d$ be a path of bounded variation or a rough path of finite $p-$ variation such that the prior integration makes sense. The signature $S(X)$ (or $\mathbf{X}$) of X is an element of $T((E))$ defined as $S_I(X) = \bigoplus_{n=0}^{\infty} S_I^n(X)$ where

$$S_I^0(X) \equiv 1 \quad \text{and} \quad S_I^n(X) = \int_{a<t_1<\ldots<t_n<t} \cdots \int_{t_1} dX_{t_1} \otimes \ldots \otimes dX_{t_n} \quad \text{for} \quad n \geq 1$$

Where each element $S_I^n(X)$ lies on $(\mathbb{E})^{\otimes n}$ and consist of the iterated integrals considering the possible combinations of indexes of order n, i.e. if $n = 2$ and $d = 3$, the combinations are $(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3)$. If we check all the coordinates of the signature, we have the following infinite vector

$$S_I(X) = (1, S_I(X)^{(1)}, ..., S_I(X)^{(d)}, S_I(X)^{(1,1)}, S_I(X)^{(1,2)}, ...)$$

with its coordinates defined as before, but with the tensor product (more general)

$$S_I(X)^{(i_1,\ldots,i_k)} = \int_{a<t_k<t} \cdots \int_{a<t_1<t_2} dX_{t_1}^{i_1} \otimes \ldots \otimes dX_{t_k}^{i_k}.$$

which superscripts run along the set

$$W = \{(i_1, ..., i_k) \mid k \geq 1,\ i1, ..., ik \in \{1, ..., d\}\}$$

and a, t are the limits of the interval I.

Definition 2.2. (Truncated signature) The truncated signature of X of order n is denoted by $\rho_n(S(X))$ and defined as

$$\rho_n(S(X)) = (1, S^1, S^2, ..., S^n) = (1, S^{(1)}, ..., S^{(d)}, ..., S^{(i_1,...,i_n)}),$$

for every integer $n \geq 1$. Basically, it means taking all coordinates of the signature until the last possible combination of order n, i.e. until last $(i_1, ..., i_n)$ with $i_1, ..., i_n \in \{1, ..., d\}$. The truncated algebra is an element of the truncated tensor algebra $T^n(E)$

2.2 Interpretation of the signature coordinates

The interpretation of the coordinates of the signature is something necessary in order to generate a visual idea of this characteristic object. High order terms of the signature are complex to understand, so we will focus only until order 3 terms. See [9] and [6].

As already mentioned, when considering the first order signature coordinate, i.e. S^i for $i = 1, ..., d$, we have

$$S(X)^{(i)}_{s,t} = X^i_t - X^i_s,$$

the first order increments of each component X^i. This component will be really useful in our work, but it only provides limited information regarding the path.

Now, if we take the second order coordinates of the signature, its interpretation is not straightforward. For the second order coordinates of the kind $S(X)^{i,i}_{s,t}$, it can be seen as

$$S(X)^{(i,i)}_{s,t} = \frac{(X^i_t - X^i_s)^2}{2},$$

which comes from the definition of the iterated integrals of second order, and a special case of the shuffle product 2.18 that we will present later on.

Nevertheless, when the superscripts are different, i.e. $S(X)^{(i,j)}_{s,t}$, this coordinates are related to the *Lévy area* (shown in 2.1) $A^{i,j}_{s,t}$ by

$$A^{i,j}_{s,t} := \frac{1}{2}\left(\int_{s<u_1<u_2<t} \mathrm{d}X^i_{u_1}\, \mathrm{d}X^j_{u_2} - \int_{s<u_1<u_2<t} \mathrm{d}X^j_{u_1}\, \mathrm{d}X^i_{u_2}\right) = \frac{1}{2}\left(S(X)^{(i,j)}_{s,t} - S(X)^{(j,i)}_{s,t}\right)$$

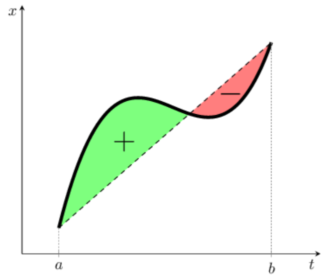

Figure 2.1: The *Lévy area* is the signed area of the path with respect to the chord joining its endpoints

which determine the signed area between the curve $u \mapsto (X_u^i, X_u^j)$ for $i \in [s,t]$ and the cord that connects the initial and final values of the coordinates (X_u^i, X_u^j) at the interval. We consider different possibilities regarding the sign of the area depending on the direction in which we travel along the path and the trajectory of the path in comparison with the cord we previously mentioned. We can see the six general cases displayed in 2.2

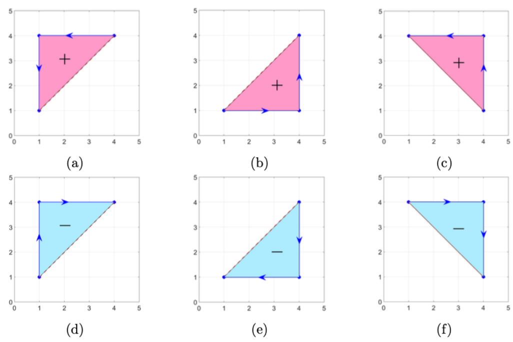

Figure 2.2: The six possibilities of signed area

With the information that the *Lévy area* provides, we can easily realise that paths similar than both in 2.3 have *Lévy area* equal to 0, so there is no possibility of differentiation between each other. In order to be able to tell the difference between these two paths using signature coordinates, one has to look into the third order terms and the second order area appears.

$$A_{s,t}^{i,(i,j)} := \frac{1}{2}\left(\int_{s<u_1<u_2<t} \mathrm{d}X_{u_1}^i \, \mathrm{d}A_{s,u_2}^{i,j} - \int_{s<u_1<u_2<t} \mathrm{d}A_{s,u_2}^{i,j} \, \mathrm{d}X_{u_2}^i\right)$$
$$= \frac{1}{2}\left(S(X)s,t^{(i,i,j)} - S(X)s,t^{(i,j,i)}\right)$$

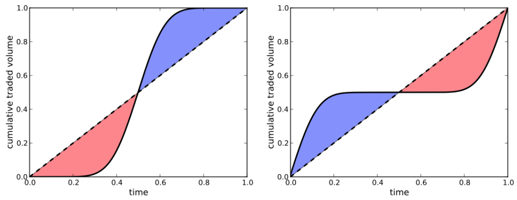

Figure 2.3: Example of two paths with *Lévy area* $= 0$

Basically, it represents the signed area of the path $u \mapsto (X^i_u, A^{i,j}_{s,u})$, the evolution of the value of the *Lévy area* with respect to the coordinate X^i_u of the path, see

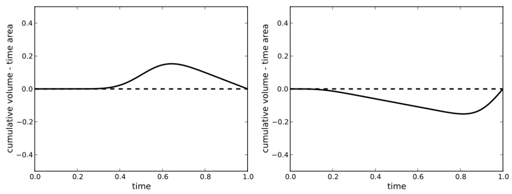

Figure 2.4: Second order area, considering the 2.3 paths

In brief, the signature of a path may be interpreted as the extraction of the order and area of one coordinate with respect to some collection of other coordinate paths. Similar recursive arguments using the shuffle product property (2.18) prove that higher order areas are related to other linear combinations of iterated integrals. For the formal and extended demonstration, the reader is refereed to section 2.2.3 of [13].

2.3 Properties of the signature

Apparently, the signature of a path seems to contain relevant information of the path, nevertheless, in order to state whether this stream of features (signature coordinates) define the path itself, some properties need to be defined. These properties are the basic fundamentals where rough path theory has been constructed, and it has been proven for all kind of p-variation paths, nevertheless, we will only focus on bounded variation paths, which lie on the scope of this thesis.

The following fundamental Chen's theorem [8] asserts that the signature is indeed an homomorphism, taking the concatenation of paths as the multiplicative operation between them.

Theorem 2.3. (Chen) *Let* $X : [0, s] \longrightarrow E$ *and* $Y : [s, t] \longrightarrow E$ *be two continuous paths of* $\mathcal{V}^1(J, \mathbb{R}^d)$ *. Then*

$$S(X * Y) = S(X) \otimes S(Y).$$

Proof. (Theorem 2.9. [13]). □

Let's see a practical application of this theorem, as done in [3], which will be important in the study and computation of the signature when considering a piece-wise linear path, relevant when studying time series. Firstly, we want to see in a practical example what is the signature of a linear path.

Example 2.4. Let $X : I \to \mathbb{E}$ be a linear path, $i = [0, T]$. It implies that for any $t \in [0, T], X_t = X_0 + \frac{X_T - X_0}{T} t$. For any integer $n \geq 1$ and any $(i_1, \ldots, i_n)$

$$S_{[0,T]}^{(i_1,\ldots,i_n)}(X) = \int_{0<t_n<T} \cdots \int_{0<t_1<t_2} d\left(\frac{(X_T - X_0)^{(i_1)}\, t_1}{T}\right) \ldots d\left(\frac{(X_T - X_0)^{(i_n)}\, t_n}{T}\right)$$
$$= \prod_{j=1}^{n} \frac{\left(X_T^{(i_j)} - X_0^{(i_j)}\right)}{T^n} \int_{0<t_n<T} \cdots \int_{0<t_1<t_2} dt_1 \cdots dt_n$$

And if we calculate the integrals

$$\int_{0<t_n<T} \cdots \int_{0<t_1<t_2} dt_1 \cdots dt_n = \frac{T^n}{n!},$$

$S_{[0,T]}^{(i_1,\ldots,i_n)}(X)$ can be simplified to

$$\frac{1}{n!} \prod_{j=1}^{n} \left(X_T^{(i_j)} - X_0^{(i_j)}\right)$$

the product of all i_j-coordinate increments within the time interval, divided by $n!$ Now, if we consider all signature coordinates, the signature ends up with the expression

$$S(X) = \exp\left(X_T - X_0\right),$$

where

$$\exp\left(X_T - X_0\right) = \sum_{n \geq 0} \frac{1}{n!} (X_T - X_0)^{\otimes n}$$

For the curious reader, as a generalization of the example, note that the signature of a path it is indeed the solution of the differential equation

$$dS(x)_{a,t} = S_{a,t}(x) \otimes dxt, \quad S_{a,a} = (1, 0, ...) \in T((E)), \tag{2.1}$$

see [13] for more details.

Now that we have seen how the signature of a linear path can be understood, we can try to apply Chen's theorem for piece-wise linear paths. Let X be a piece-wise linear path, $X : [0,T] \mapsto \mathbb{E} := \mathbb{R}$, that connects the points $x_1, ..., x_k \in \mathbb{R}$, we can use the Chen's identity considering

$$X = X_1 * X_2 * ... * X_{k-1},$$

where each X_i is a linear path connecting x_i with x_{i+1}, and the Example 2.5 to obtain

$$S(X) = \bigotimes_{i=1}^{k-1} \exp(X_i) \tag{2.2}$$

The second important property of the signature is the invariance under time reparametrizations of the path (Lemma 1.6 [13] and Lemma 2.12 [3]).

Lemma 2.5. Let $X \in \mathcal{V}^1(J, \mathbb{R}^d)$ and $\lambda : [0,T] \to [a,b]$ be a non-decreasing surjection (images maintain order and every element in the codomain has at least one preimage in the domain) and define $X_t^\lambda := X_{\lambda_t}$ for the reparametrization of X under λ. Then, for every $s,t \in [0,T]$,

$$S(X)_{\lambda_s,\lambda_t} = S(X^\lambda)_{s,t}$$

Proof. (For simplicity, we prove it only considering smooth reparametrizations, although it is not strictly necessary) Given $\lambda : [0,T] \to [a,b]$, a continuous non-decreasing surjection, a reparametrization, if we define the path coordinates reparametrized $\widetilde{X}_t^i = X_{\lambda_t}^i$ and $\widetilde{X}_t^j = X_{\lambda_t}^j$, observe that

$$d\widetilde{X}_t^i = dX_{\lambda_t}^i d\lambda(t),$$

and it follows that

$$\int_a^b \widetilde{X}_t^j \, d\widetilde{X}_t^i = \int_0^T X_{\lambda_t}^j \, dX_{\lambda_t}^i \, d\lambda(t) \, dt = \int_0^T X_u^j \, dX_u^i$$

where $u = \lambda(t)$. With that path integrals are invariant under time reparametrization, hence the signature is invariant under time reparametrizations, using this result recursively. □

The next proposition gives a notion of how the signature of a time reversed path is related to the signature of the initial path, see [13] for more details.

Proposition 2.6. Let $X : [0,T] \longrightarrow E$ be a path of bounded-variation. Let $\overleftarrow{X}$ be the path X run backwards, i.e. the path defined by $\overleftarrow{X}_t = X_{T-t}, t \in [0,T]$. Then

$$S(\overleftarrow{X}) = S(X)^{-1}$$

In particular, we can state that the range of $S : \mathcal{V}^p([0,T],E) \longrightarrow T((E))$ is a group.

Proof. (Proposition 2.14. [13]). □

Corollary 2.7. As a particularity of the fact that the range of the signature is a group, we have that given $X : [0,T] \longrightarrow E$, $S(X) \otimes S(\overleftarrow{X}) = \mathbf{1}$.

We have already seen several reasons why $S(X) = S(Y)$ does not imply $X = Y$. For example, $S(X)$ does not depend on the parametrization of X. Or, the signature of the constant path is equal to $\mathbf{1}$, but, $S(X) \otimes S(\overleftarrow{X}) = S(X * \overleftarrow{X}) = \mathbf{1}$. Thus, the constant path and $X * \overleftarrow{X}$ have the same signature, but $X * \overleftarrow{X}$ cannot be reparametrised to be constant.

In order to explore to what extent the signature of a path determine the path, we first make some definitions.

Definition 2.8. **(Tree-like paths)** Given $X : [0,T] \to \mathbb{E}$, we say X is a tree-like path in $\mathbf{E}$ if there exists a positive real valued continuous function h defined on $[0,T]$ such that $h(0) = h(T) = 0$, and such that

$$\|X_t - X_s\|_{\mathbf{E}} \leq h(s) + h(t) - 2 \inf_{u \in [s,t]} h(u). \tag{2.3}$$

The function h is called a height function for X. We say X is a Lipschitz tree-like path if h can be chosen to be of bounded variation.

Lemma 2.9. $X : [0,T] \to \mathbb{E}$ is a tree like path if and only if it is a null path as a control, i.e. the trajectories are completely canceled out by themselves.

Proof. See Appendix at [2] for the complete proof. □

As an intuitive example, if X, Y, Z are non-constant paths, then $X * Y * \overleftarrow{Y} * Z * \overleftarrow{Z} * \overleftarrow{X}$ is a tree-like path. As explained in [13], "Tree-like paths are those which can be reduced to a constant path by removing possibly infinitesimal pieces of the form $W * \overleftarrow{W}$". With this example, we note that the tree-like paths X are not strictly of the kind $X = Y * \overleftarrow{Y}$

Definition 2.10. Let $X, Y \in BV(V)$. We say $X \sim Y$ if the concatenation of X and Y 'run backwards' is a *Lipschitz tree-like path*.

We now focus on $\mathbb{E} = \mathbb{R}^d$ and state some results that won't be proved as they are out of the scope of this work. See [2].

Theorem 2.11. Let $X \in BV\left(\mathbb{R}^d\right)$. The path X is tree-like if and only if the signature of X is $\mathbf{0} = (1, 0, 0, \ldots)$.

As we have seen, the map $X \to \mathbb{R}$ is a homomorphism, and running a path backwards gives the inverse for the signature in $T(\mathbb{R})$, so an immediate consequence of the previous theorem is

Corollary 2.12. If $X, Y \in BV\left(\mathbb{R}^d\right)$, then $\mathbf{X} = \mathbf{Y}$ if and only if the concatenation of X and ' Y run backwards' is a Lipschitz tree-like path.

Taking all of these together, we can state the following theorem.

Theorem 2.13. (Uniqueness of the signature) Let $X \in BV\left(\mathbb{R}^d\right)$, then $S(X)$ determines X up to the tree-like equivalence.

As the signature determines the path up to sections on which the path exactly retraces itself, if we have a path in which it has a monotone component, its trajectories will never cancel out, so the following lemma holds.

Lemma 2.14. Let $\mathcal{A}$ be a set of continuous paths with bounded variation such that all paths have the same initial value and have at least one monotone coordinate. Then the signature of a path in A completely determines it in A.

At this point, we have seen which are the conditions to identify a path for each signature. When doing our practical application, these conditions can not be missed, as the results are strictly tied to these last lemma, before every computation of the signature of a path, we will apply the transformations:

1. **Base-point augmentation** (X^*): As a concatenation of the path $X_t, t \in [0, T]$ with the linear path defined between $t = -1$ and $t = 0$, between 0 and x_0, being x_0 the value of $X(0)$:

$$X_t^* = \begin{cases} (t+1)x_0 & t \in [-1, 0), \\ X_t & t \in [0, T]. \end{cases}$$

2. **Time-augmentation** ($\hat{X}^*$): Adding the time as an extra coordinate of the path. $\hat{X}_t^* = (t, X_t^*)$.

Proposition 2.15. *(Lyons et al, [13], Proposition 2.2)* Let $X : [0,T] \mapsto \mathbb{R}^d$ be an element of $\mathcal{V}^1(J, \mathbb{R}^d)$. For any $k \in \mathbb{N}$ we have

$$|\int_{a<t_k<t} \cdots \int_{a<t_1<t_2} dX_{t_1}^{i_1} \otimes \ldots \otimes dX_{t_k}^{i_k}| \leq \frac{\|X\|_{1,[0,T]}^k}{k!}$$

with $\|X\|_{1,[0,T]}^k$ being the p-variation norm previously presented ($p = 1$).

From this, we can deduce that taking the first elements of the signature until a given level k is a good way to approximate the signature of a path, i.e. the truncated signature of order k: $\rho_k(S(X))$.

Corollary 2.16. Let $X \in \mathcal{V}^1([0,T], \mathbb{R}^d)$, then for every $\epsilon > 0$ there exists a $N \in \mathbb{N}$ such that

$$\|S(X) - \rho_N(S(X))\| \leq \epsilon.$$

Furthermore, if we restrict X to a compact set $\mathcal{K} \in \mathcal{V}^1([0,T], \mathbb{R}^d)$, this convergence is uniform.

Proof. Consider $X \in \mathcal{V}^1([0,T], \mathbb{R}^d)$. We know from Proposition 2.16 that each of the elements of the signature of order k verify the following inequation

$$\|S^{i_1,\ldots,i_k}(X)\| \leq \frac{\|X\|_{1,[0,T]}^k}{k!} \tag{2.4}$$

The difference between the signature of X and the truncated signature of order k as $\Delta_N = S(X) - \rho_N(X)$. So,

$$\|\Delta_N\| = \|S(X) - \rho_N(X)\|$$

and 2.4,

$$\|\Delta_N\| \leq \left(\sum_{k=N+1}^{\infty} \frac{\|X\|_{k,1-\text{var}}}{k!} \right)$$

Since $\|X\|_{k,1-\text{var}}$ is bounded for all k, the series converges, and we can make the second term as small as we want by choosing a sufficiently large N. Thus, for every $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that $\|\Delta_N\| \leq \epsilon$.

Consider $\mathcal{K}$ compact subset, and x_n a sequence in $\mathcal{K}$. By sequential compactness, there exists a subsequence x_{n_k} such that x_{n_k} converge to some x as $n_k \to \infty$.

To demonstrate the uniform convergence, we need to show that for any $\epsilon > 0$ there exists an N such that for all $N_k > N$, $\|\Delta_{N_k}\| \leq \epsilon$. We have seen that for all k, $\|\Delta_{N_k}\| \leq \epsilon/2^{N_k}$, so if we choose an N such that $\epsilon/2^N < \epsilon$,

$$\|\Delta_{N_k}\| \leq \epsilon/2^{N_k} < \epsilon/2^N < \epsilon.$$

□

One of the fundamental properties of the signature, originally shown by Ree at [10], is that the product of two terms $S(X)^{(i_1,\dots,i_k)}$ and $S(X)^{(j_1,\dots,j_m)}$ can always be expressed as a sum of another collection of terms of $S(X)$. Note that when doing analysis, it is really important to find a core of real functions on the space we are working on, and this core, ideally, should be an algebra. We will see in this section that this real functions are induced by linear forms on $T((E))$.

We first introduce the definition of the dual basis of the tensor space as done in [12]. Considering $\mathcal{B} = \{e_i\}_{i=1}^d$ the d-dimensional basis of $\mathbb{E}$, for every $n \in \mathbb{N}$, $\mathcal{B}$ determines the basis

$$\mathcal{B}^{\otimes n} := \{e_{\mathbf{K}} = e_{k_1} \otimes \ldots \otimes e_{k_n} : \mathbf{K} = (k_1, \ldots, k_n) \in \{1, \ldots, d\}^n\}$$

for $\mathbb{E}^{\otimes n}$, and also determines the corresponding dual basis

$$(\mathcal{B}^*)^{\otimes n} := \left\{e^*_{\mathbf{K}} = e^*_{k_1} \otimes \ldots \otimes e^*_{k_n} : \mathbf{K} = (k_1, \ldots, k_n) \in \{1, \ldots, d\}^n\right\}$$

for $(\mathbb{E}^*)^{\otimes n}$, forming a basis of $T(E^*)$, if we think of an element as a non commuting power series in the letters $e_1, ..., e_d$, then $e^*_{(i_1,\dots,i_n)}$ picks up the coefficient of the monomial $e_{i_1}...e_{i_n}$. Considering this, we see that we have a linear action of $(\mathbb{E}^*)^{\otimes n}$ on elements of $(\mathbb{E})^{\otimes n}$, which we can extend naturally to a linear map $(\mathbb{E}^*)^{\otimes n} \to T((E))^*$ defined by

$$\phi(A) := \phi(a^n), \text{ when } A = (a^0, a^1, a^2, ...) \in T((E)) \text{ and } a^n \in \mathbb{E}^{\otimes n} \tag{2.5}$$

By letting n vary between 0 and ∞, we get the linear mapping

$$T(E^*) = \bigotimes_{n=0}^{\infty} (E^*)^{\otimes n} \mapsto T((E))^* \tag{2.6}$$

So we can trivially see that $T(E^*) \subset T(E)^*$ by linearity. Basically, we have extended a linear action over each of the elements of the basis of the tensor algebra to a linear action over an element in $T(E)$, as it is A. In particular, if $\phi = \bigoplus_{n=0}^{\infty} \phi_n \in T(E^*)$ and $v = \bigoplus_{n=0}^{\infty} v_n \in T(E)$, then $\phi(v) = \sum_{n=0}^{\infty} \phi_n(v_n)$.

Now we define the inner product over the tensor algebra and relate it with the action of the linear forms of the dual basis of T(E). This inner product comes from the choice of the basis of $\mathcal{B}$ by

$$\langle e_i, e_j \rangle_{\mathbb{E}} := \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{2.7}$$

extending to the whole of space $\mathbb{E}$ bilinearly.

The extension into $\mathbb{E}^{\otimes n}$ is done by defining

$$\langle e_{i_1} \otimes \ldots \otimes e_{i_n}, e_{j_1} \otimes \ldots \otimes e_{j_n} \rangle_{\mathbb{E}^{\otimes n}} := \prod_{k=1}^{n} \langle e_{i_k}, e_{j_k} \rangle_{\mathbb{E}} = \delta_{i_1 j_1} \ldots \delta_{i_n j_n} \tag{2.8}$$

and binearly extending it to the whole space. Finally, if we want to define the inner product on the whole $T((E))$ space, we have to set

$$\langle x, y \rangle_{T((E))} = \sum_{n=0}^{\infty} \langle \rho_n(x), \rho_n(y) \rangle \tag{2.9}$$

for $x, y \in T((E))$, having $(T((E)), \langle \cdot, \cdot \rangle)$ as an inner product space. Hence, we can define the completion of $T((E))$, $\overline{T((E))}$, by adding all limit points, in order to ensure that all Cauchy sequences converge within the space. $\overline{T((E))}$ is a Hilbert space with this inner product (allowing us to apply geometric interpretations and rigorous analysis).

If we want to understand the elements of the dual basis $(\mathcal{B}^*)^{\otimes n}$ in terms of the inner product, we have that for every $n \in \mathbb{N}$, they are given by

$$e_K^*(\cdot) = \langle \cdot, e_K \rangle, \tag{2.10}$$

which means that the vectors of the dual basis that correspond to the tensor space $T((E))$ are equivalent to computing the inner product with its corresponding vector of the base of $T((E))$, i.e. the projection.

With this theoretical framework, we are now totally able to define the coordinate iterated integral (signature coordinate).

Definition 2.17. Following the notation provided at 2.10, given a word $\mathbf{I} = (i_1, \ldots i_n) \in \{1, \ldots, d\}^n$, if $e_{\mathbf{I}} = e_{i_1} \otimes \ldots \otimes e_{i_n} \in \mathbb{E}^{\otimes n}$, we define the *coordinate iterated integral* $\pi^{\mathbf{I}}(S_{a,b}(X))$ by

$$\pi^{\mathbf{I}}(S_{a,b}(X)) := \langle S_{a,b}(X), e_{\mathbf{I}} \rangle_{T(E)} = e_{\mathbf{I}}^* \left(S_{a,b}^n(X) \right).$$

Interpreting it as the projection of the component $S_{a,b}^n(X) \in \mathbb{E}^{\otimes n}$ into the basis $e_{\mathbf{I}}$, which will lead to the coordinate of the signature in the tensor space that corresponds to the index **I**

$$\pi^{\mathbf{I}}(S_{a,b}(X)) = \int_{a \leq t_n \leq t_n \leq b} \cdots \int_{0 \leq t_1 \leq t_2} \langle dx_{t_1}, e_{i_1} \rangle_{\mathbb{E}} \cdots \langle dx_{t_n}, e_{i_n} \rangle_{\mathbb{E}} \,.$$

and all the other coordinates (or iterated integrals) of the signature vanish due to the inner product operator.

Having understood the coordinate iterated integral, or signature coordinate, we want to understand how the product operation between two coordinate iterated integrals is done. This product of linear forms is actually a quadratic form (easy to proof), but we want to see that it also is a linear form on the range of the signature. By this means we introduce the *Shuffle product* between two multi-indexes.

First, a permutation of the set $1, ..., k+m$ is called a (k,m)-shuffle if $\sigma^{-1}(1) < ... < \sigma^{-1}(k)$ and $\sigma^{-1}(k+1) < ... < \sigma^{-1}(k+m)$, and $(\sigma(1), ..., \sigma(k+m))$ is called a shuffle of $(1, ..., k)$ and $(k+1, ..., k+m)$. We will denote Shuffles(k,m) the set of all (k,m)-shuffles, and interpret them as the list of all the ways that two words, of length k and m respectively, can be combined into a single word, of length k + p, while preserving the order in which the letters of each original word appear.

Definition 2.18. (Shuffle product). Consider two multi-indexes $I = (i_1, ..., i_k)$ and $J = (j_1, ..., j_m)$ with $i_1, ..., i_k, j_1, ..., j_m \in 1, ..., d$. And define

$$(r_1, ..., r_k, r_{k+1}, ..., r_{k+m}) = (i_1, ..., i_k, j_1, ..., j_m).$$

The shuffle product of I and J, $I \sqcup\!\sqcup J$, is defined as the following finite set

$$I \sqcup\!\sqcup J := \{(r_{\sigma(1)}, ..., r_{\sigma(k)}, r_{\sigma(k+1)}, ..., r_{\sigma(k+m)}) \mid \sigma \in \text{Shuffles}(k, m)\}$$

Returning to the coordinate iterated integrals, if we consider $\pi^{\mathbf{I}}(S_{a,b}(X))$ and $\pi^{\mathbf{J}}(S_{a,b}(X))$, their product is given by the following theorem (Theorem 2.15. in [13]).

Theorem 2.19. (Shuffle product property). For any $I_1, I_2 \in A^*$, $A^* = \{I = (i_1, ..., i_n)$ where $i_j \in \{1, ..., d\}, \forall j \in \{1, ..., n\}, \forall n \in \mathbb{N}\}$, it holds that for any path X of bounded variation

$$\pi^{I_1}(S(X))\pi^{I_2}(S(X)) = (\pi^{I_1} \sqcup\!\sqcup \pi^{I_1})(S(X)).$$

Corollary 2.20. *(Theorem 2.15 [13])*The linear forms on $T((E))$ induced by $T(E^*)$, when restricted to the range $S(V^1([0,T],E))$ of the signature, form an algebra of real-valued functions.

Proof. It is a direct implication of the Shuffle product property for bounded variation paths. See [13] for the whole proof. □

Example 2.21. To clarify it, we will introduce a simple example. Consider a two-dimensional path $X : [a,b] \to \mathbb{R}^2$. If we apply the shuffle product property we have

$$\pi^{(1)}(S(X))\pi^{(2)}(S(X)) = \pi^{(1,2)}(S(X)) + \pi^{(2,1)}(S(X))$$
$$\pi^{(1)}(S(X))\pi^{(2,1)}(S(X)) = 2\pi^{(2,1,1)}(S(X)) + \pi^{(1,2,1)}(S(X))$$

Note that the shuffle product computes all the possible combinations without altering the order of the indices of the initial words. The combination (1,1,2) is not possible as 2 needs to be before 1, when considering the indices of the initial word (2,1).

As we have seen, the signature of a path embodies a multitude of essential properties crucial for path analysis. Firstly, Chen's Identity serves as a pivotal tool for concatenating paths, allowing for the seamless integration of path segments. Remarkably, the signature is independent of how fast or slow a path is traversed, ensuring that paths with the same shape retain the same signature regardless of timing. This property, known as time−invariance, is crucial for robust analysis across varying speeds. Furthermore, the signature demonstrates unique characteristics under tree-like paths, guaranteeing distinct representations for different path structures. Notably, the uniqueness property extends to paths featuring monotone increasing components, which allows reconstruction of the entire path from its signature if the starting point is known and there exists a monotone increasing coordinate. The signature's efficiency is another strength, achieved through a property where higher order terms have a diminishing effect, making it computationally friendly and paves the way for its application in machine learning tasks. To top it off, the product property simplifies the multiplication of complex path interactions by expressing the product of signature terms as a sum based on their multi−indexes. These properties collectively underscore the versatility and power of the signature as a comprehensive tool for path analysis and processing.

Chapter 3

The expected signature framework

3.1 Law on signatures

In our work, we aim to construct an expected signature model which forecast the signature of stochastic paths. In order to do so, it is important to properly define the expected signature as an element defined in a probability space and measure, as done in [3]. Basic knowledge on measure theory and stochastic processes is assumed for this definition. For simplicity, from this point on, we will denote $\mathbf{X} := S(X)$.

Definition 3.1. Given a probability space $(\Omega, P, \mathcal{F})$ and a $\mathbb{R}^d$-valued stochastic process X, for every $w \in \Omega$, $\mathbf{X}(w)$ is well defined almost surely and its expectation $(\mathbb{E}[\mathbf{X}(w)])$ is finite under the probability measure P. We call $\mathbb{E}[\mathbf{X}(w)]$ the expected signature of X.

Chevyrev and Lyons at [7] define the characteristic function of a random signature $\mathbf{X}$ and assert, as its main result, that $\phi_{\mathbf{X}}$ determines $\mathbf{X}$. By the following theorem, we provide sufficient conditions to state that the expected signature determines the law of the signature, under certain conditions (we will not deep into these conditions).

Theorem 3.2. (Proposition 6.1 [7]). Let X and $\hat{X}$ be two random paths of bounded variation, $\mathbf{X}$ and $\hat{\mathbf{X}}$ their corresponding signatures as random variables. If $\mathbb{E}[\mathbf{X}] = \mathbb{E}[\hat{\mathbf{X}}]$ and they both have infinite convergence radius, then $\mathbf{X} = \hat{\mathbf{X}}$.

3.2 The expected signature model

Basically, our expected signature model will consist in a regression framework based on time series data. This data is discrete, so, in order to properly use this framework, it is necessary to transform the discrete information of the time series into a continuous path. In order to tackle this problem, there are many possible approaches, which comparison is out of our scope, but we can, for example, think about piece-wise linear interpolation between data-points.

Our goal is to formulate the effects of data streams as a dependent variable of a regression problem with an explanatory and dependent variables being paths. Imagine we have observations of the input and output paths $\{X_i, Y_i\}_{i=0}^{n-1}$ of the theoretical regression, both X_i and Y_i of bounded variation taking values in $E = \mathbb{R}^d$. This linear relation can be understood as

$$Y_i = f(X_i) + \epsilon_i \; \forall i = 0, ..., n-1$$

Obviously, the functional form of the relationship is unknown or really complex, so it is necessary to identify specific features of the observed data to linearize its functional relationship. To achieve this, we will present a way to do it using the signature of the path as features on the space where the path is defined.

We first present a very important theorem where this theory is sustained, which states that it suffices to look for linear functions f.

Theorem 3.3. (Signature approximation). Let $S(\mathcal{V}^p(J, E))$ be the set of bounded variation paths, and $S_1 \subset S(\mathcal{V}^p(J, E))$ a compact subset. Then, given $\varepsilon > 0$ and a continuous function $f : S_1 \to \mathbb{R}$, there exists a linear functional $L \in T((E))^*$ such that for every $a \in S_1$

$$|f(a) - L(a)| \leq \varepsilon.$$

In order to proof this result we present the following well-known theorem.

Theorem 3.4. (Stone-Weierstrass theorem). Given X, a compact Hausdorff space, and A a subalgebra of $C(X, R)$, which contains a non-zero constant function, then, A is dense in $C(X, R)$ if and only if it separates points.

Proof. (Signature approximation) Let $\mathcal{L}(S_1)$ denote a family of all linear functions in $T((E))^*$ restricted to S_1, by the shuffle product property of signatures and Corollary 2.20, $\mathcal{L}(S_1)$ is an algebra. Since the 0^{th} term of the signature is always

1, this algebra contains constant functions. Moreover it separates the points (this assumption comes from Corollary 2.16 in [13]). Now, using the Stone-Weierstrass theorem, we can say that $\mathcal{L}(S_1)$ is dense in the space of continuous functions on S_1. □

Corollary 3.5. If S_1 is the set of signatures of any finite number of sample paths, Theorem 3.3 holds.

Proof. Trivial, as this finite set is a compact subset of $S(\mathcal{V}^p(J,E))$. □

As we presented at the beginning of this section, we want to understand the conditional distribution of Y given the information of X, and we now can state that, in our framework, this is equivalent to saying that we are looking for $\mathbb{E}[\mathbf{Y}|\mathbf{X}]$ due to the following results that we have already seen:

Given X a bounded variation path,

1. $\mathbf{X}$ uniquely determines X up to the tree-like equivalence.
2. According to 3.2, under certain conditions (will be assumed that are in our work), the expected signature of stochastic process determines the measure on the random signatures.

Finally, we are be able to present our model with the conclusion that, restricting to the case where $E[\mathbf{Y}|\mathbf{X}]$ is a continuous function of X, by Theorem 3.3, $E[\mathbf{Y}|\mathbf{X}]$ can be well approximated by a linear function on X locally, as each coordinate of the signature can be itself well approximated by a linear functional, and we have seen that these coordinates $(\pi^{\mathbf{I}}(\mathbf{X}))$ form an algebra of real valued functions and serve basis functions to represent any smooth function on signatures locally.

Definition 3.6. (Expected Signature Model [3]). Let X and Y be two stochastic processes taking values in $\mathbb{E}$ and $\mathbb{W}$ respectively. Suppose that the $\mathbf{X}$ and $\mathbf{Y}$ are well defined a.s.. We assume that

$$\mathbf{Y} = L(\mathbf{X}) + \varepsilon, \tag{3.1}$$

considering $\mathbb{E}[\varepsilon|\mathbf{X}] = 0$, and L a linear functional mapping $T((E))$ to $T((W))$.

We will note $\mu_{\mathbf{X}}$ the conditional expectation $\mathbb{E}[\mathbf{Y}|\mathbf{X}]$ and $\sum_{\mathbf{X}}^2$ the conditional covariance defined as

$$\begin{aligned} \Sigma_{\mathbf{X}}^2 : A^* \times A^* &\to \mathbb{R} \\ (I,J) &\mapsto \operatorname{Cov}\left(\pi^I(\mathbf{Y}), \pi^J(\mathbf{Y})|\mathbf{X}\right). \end{aligned} \tag{3.2}$$

The following lemma from [3] won't be explicitly used during this work, but it is an interesting property that we wanted to highlight, as it is a direct implication of the shuffle product property. It asserts that $\Sigma^2_{\mathbf{X}}$ is determined by $\mu_{\mathbf{X}}$.

Lemma 3.7. *Let $\mu_{\mathbf{X}}$ and $\Sigma^2_{\mathbf{X}}$ be defined as before. Then for every $I, J \in A^*$,*

$$\Sigma^2_{\mathbf{X}}(I,J) = \left(\pi^I \sqcup\!\sqcup\, \pi^J\right)(\mu_{\mathbf{X}}) - \pi^I(\mu_{\mathbf{X}})\,\pi^J(\mu_{\mathbf{X}}).$$

Proof. For each $I, J \in A^*$, conditional variance states that

$$\Sigma^2_{\mathbf{X}}(I,J) = \mathbb{E}\left[\pi^I(\mathbf{Y})\pi^J(\mathbf{Y})|\mathbf{X}\right] - \mathbb{E}\left[\pi^I(\mathbf{Y})|\mathbf{X}\right]\mathbb{E}\left[\pi^J(\mathbf{Y})|\mathbf{X}\right].$$

Using the shuffle product property of the signature,

$$\mathbb{E}\left[\pi^I(\mathbf{Y})\pi^J(\mathbf{Y})|\mathbf{X}\right] = \mathbb{E}\left[\left(\pi^I \sqcup\!\sqcup \pi^J\right)(\mathbf{Y})|\mathbf{X}\right] = \left(\pi^I \sqcup\!\sqcup \pi^J\right)(\mathbb{E}[\mathbf{Y}|\mathbf{X}]),$$

and we obtain:

$$\Sigma^2_{\mathbf{X}}(I,J) = \left(\pi^I \sqcup\!\sqcup \pi^J\right)(\mu_{\mathbf{X}}) - \pi^I(\mu_{\mathbf{X}})\,\pi^J(\mu_{\mathbf{X}}).$$

□

3.2.1 Model calibration and prediction

After understanding how the model is theoretically constructed, now, it is necessary to explain how it will be implemented with real world data. First, note that this approach is defined over the space where the signature of a path lies, an infinite dimensional space, which is clearly impossible to handle in terms of computational methods. Hence, we will make use of the Proposition 2.15 and we will limit our problem to estimate the expected signature of Y given information of X by estimating the truncated signature of the order m $\rho_m(\mathbb{E}[X|Y])$, in other terms, estimating $\rho_m \circ f$. Given a large number of samples $\{X_i, Y_i\}_{i=1}^N$, this problem will be reduced to a linear regression problem, where the coordinate iterated integrals of $\rho_m(\mathbf{Y})$ are estimated by the explanatory variables $\rho_n(\mathbf{X})$. We will call this the ES approach and will be used further in this work in a practical setup.

Chapter 4

The ES approach as a time series model

When considering real-life time series, such as financial data, the data collected is discrete (e.g., tick-by-tick or minute-by-minute). It might seem effective to approximate this data by sampling it at very fine intervals. However, interpreting the information using discrete data or linear functionals like Fourier transforms has proven inefficient. This approach generates a lot of redundant data and leads to dimensionality issues when dealing with high sampling rates. In such cases, non-linear feature sets, such as signature transformations, become important. The signature of a path, computed from the embedded discrete time series, is not particularly sensitive to the number of time steps. This makes it a more effective method for capturing the essential features of the data.

While financial data is typically visualized as univariate time series (a sequence of data points over time for a single variable), computing informative features from this data requires embedding it into a continuous path that preserves key characteristics. We won't delve into finding the optimal embedding method here, but we'll explore some theoretical options and discuss some of their pros and cons.

4.1 Time series embedding

Consider $\{(t, r_i)\}_{i=0}^{N}$ an univariate time series. Our goal is to transform it into a continuous function (path). There are many possibilities, and here we present some of them:

1. Piece-wise linear interpolation.

2. Rectilinear interpolation.

3. Lead-lag transformation.

The first two are the most basic and can be seen at Figure 4.1, with red dots being the discrete data and the blue lines its embedding. Piece-wise linear interpolation simply interpolate datapoints linearly and rectilinear interpolation lead to a piece-wise linear path with each linear section parallel to an axis.

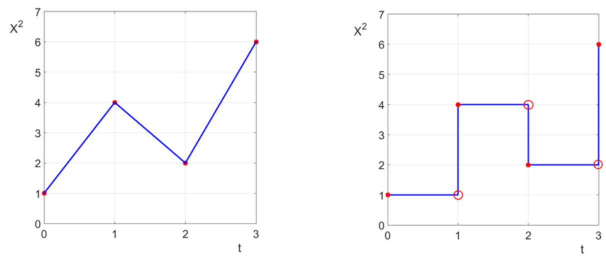

Figure 4.1: Different interpolations [5]. Piece-wise linear interpolation (left) and Rectilinear interpolation (right).

The Lead-lag transformation is a more complex approach, as it augments the dimension of the path that contains the data. In our case, it maps our one-dimensional path into a two dimensional one. Let's see with a practical example how it is computed: Consider a one dimensional path with the following time-indexed values:

$$\{X_i\}_{i=0}^{4} = \{1, 4, 2, 6\},$$

The corresponding LeadLag transformation is:

$$X_i = \{1, 4, 2, 6\} \mapsto \begin{cases} X^{\text{Lead}} & = \{1, 4, 4, 2, 2, 6, 6\} \\ X^{\text{Lag}} & = \{1, 1, 4, 4, 2, 2, 6\} \end{cases}$$

Constructing a 2-dimensional path with a leading component, and a lagged component. Generally, the Lead Lag transformation with lag 1 is defined by

$$(X^{lead}, X^{lag}) = (X_t, \lim_{\epsilon \to 0} X_{t-\epsilon})$$

and, then, a linear interpolation of the 2-dimensional points is made, as seen in 4.2.

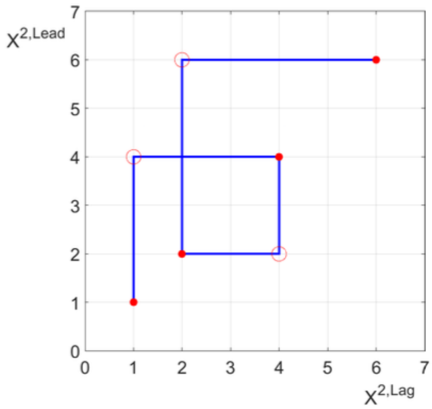

Figure 4.2: [5] Lead-lag transform of 1-dimensional data $\{X_i\}_i$.

Let's briefly state some properties of each different embedding. Regarding piece-wise linear interpolation, when considering a 1-dimensional time-series, it is a really straightforward computation, but the path will only encode information related to increments, leading us to lose some essential information. Contrary, lead-lag transformation provide enough information to read the volatility of the path directly from the second term of the signature (see Equation 4.1), a very important parameter when dealing with financial data. By Chen's theorem, we can straightforwardly see this assumption by computing the signature coordinates of a of a lead-lag embedding:

$$
\begin{aligned}
S^{(1)} = S^{(2)} &= \sum_{i}^{N-1} (X_{i+1} - X_i) \\
S^{(1,1)} = S^{(2,2)} &= \frac{1}{2}\left(\sum_{i}^{N-1} (X_{i+1} - X_i)\right)^2 \\
S^{(1,2)} &= \frac{1}{2}\left[\left(\sum_{i}^{N-1} (X_{i+1} - X_i)\right)^2 + \sum_{i}^{N-1} (X_{i+1} - X_i)\right], \\
S^{(2,1)} &= \frac{1}{2}\left[\left(\sum_{i}^{N-1} (X_{i+1} - X_i)\right)^2 - \sum_{i}^{N-1} (X_{i+1} - X_i)\right]
\end{aligned}
$$

and the following holds

$$
\text{Mean}(X) = \frac{1}{N} S^{(1)} \; ; \; \text{Var}(X) = -\frac{N+1}{N^2} S^{(1,2)} + \frac{N-1}{N^2} S^{(2,1)}. \tag{4.1}
$$

Finally, regarding the rectilinear path, we can highlight that it has an structure with an easy way to extract the path from its signature. As we are aiming to compare the Expected Signature model with some traditional approaches for financial time series, we are going to consider a special case of the rectilinear interpolation called the *Time-joined transformation*.

Remark 4.1. For the curious reader, if one is considering to deal with high frequency data and wants to summarize the information ignoring time re-parameterizations, one should consider piece-wise and lead-lag methods. They prevent over-fitting, as they allow the signature to summarize and incorporate high frequency information using only a few parameters.

4.1.1 Time joined embedding

Definition 4.2. (Time joined transformation [3]). Let $\{(t_i, r_i)\}_{i=m}^{n}$ be a uni-variate time series. Let $R : [2m, 2n+1] \mapsto \mathbb{R} + \times \mathbb{R}$ be a 2-dimensional time-joining path of $\{(t_i, r_i)\}_{i=m}^{n}$, which is defined as :

$$R(s) = \begin{cases} t_m e_1 + r_m (s - 2m) e_2, & \text{if } s \in [2m, 2m+1); \\ [t_i + (t_{i+1} - t_i)(s - 2i - 1)]\, e_1 + r_i e_2, & \text{if } s \in [2i+1, 2i+2); \\ t_{i+1} e_1 + [r_i + (r_{i+1} - r_i)(s - 2i - 2)]\, e_2, & \text{if } s \in [2i+2, 2i+3) \end{cases} \tag{4.2}$$

with $\{e_1, e_2\}$ an orthonormal basis of $\mathbb{R}^2$ and $i = m, m+1, \ldots, n-1$.

R is an specific case of a rectilinear embedding. For a better understanding of this procedure let's see it in an example.

Example 4.3. Consider the following time series:

$$\{(t_i, r_i)\} = \{(2,3), (3,4), (4,7), (5,5), (6,9)\}.$$

The 2-dimensional path $R(s)$ is the piece-wise linear interpolation of the 2-dimensional points

$$\{(2,0), (2,3), (3,3), (3,4), (4,4), (4,7), (5,7), (5,5), (6,5), (6,9)\}.$$

The transformed path is plotted as seen in figure 4.3.

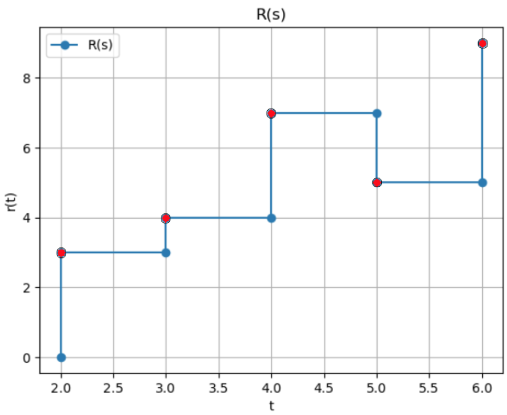

Figure 4.3: Time-joined transformation of example time series. The red points refer to the initial values of the time series

4.2 The signature of a time series

As we have seen, to utilize the Expected Signature model for time series analysis, it is crucial to first embed the data points into a path. This path can be constructed in various ways, depending on the requirements of the experiment and the features that need to be analyzed. Before making further assumptions in our time series model, we must describe the signature of a time series and understand its key properties. Let's consider a time series $\{(t_i, r_i)\}_{i=m}^{n}$ and $R(s)$ its time-joined transformation (recall, that these results can be made for other type of embeddings, but we focus on the previously presented).

Remark 4.4. The time-joined transformation of a time series is a two dimensional path with an non-decreasing coordinate (time) and a fixed starting point $(t_m, 0)$.

Lemma 4.5. With properties described at Remark 4.4, we can state that the signature of the path constructed by the time-joined transformation ($R(s)$) of a time series $\{(t_i, r_i)\}_{i=m}^{n}$ uniquely determines the time series $\{(t_i, r_i)\}_{i=m}^{n}$.

Proof. Given $\{(t_i, r_i)\}_{i=m}^{n}$ the time series and R its time-joined embedding, by construction, the first coordinate of R is non-decreasing. Hence, R is not possibly tree-like. Moreover, all paths R contain the same base-point $(t_m, 0)$, so we are under the conditions of Lemma 2.14 to state that there is a correspondence between the signatures of the paths R generated by $\{(t_i, r_i)\}_{i=m}^{n}$ and time series $\{(t_i, r_i)\}_{i=m}^{n}$. □

Definition 4.6. We use the expression "the signature of a time series" when considering the signature of its time-joined transformed path R, and denote it by $S(\{(t_i, r_i)\}_{i=m}^{n})$.

At this point, the reader must be wondering if it is possible to find the values $\{r_i\}$ of a time series $\{(t_i, r_i)\}_{i=m}^{n}$ given its signature. The answer is that, if time series is embedded using the time-joined transformation, it is possible. Lemma 4.9 is presented at [3] and provides the concrete procedure.

First, we present Vandermonde matrices and characterize its determinant, as they are necessary for the proof of the Lemma proof.

Definition 4.7. *A square matrix of size n, A, is a Vandermonde matrix if there are scalars,* $x_1, x_2, x_3, ..., x_n$ *such that* $[A_{i,j}] = x_i^{j-1}$, $1 \leq i \leq n, 1 \leq j \leq n$.

Theorem 4.8. *Suppose we have A a square Vandermonde matix of size n, constructed by the coefficients* $x_1, x_2, ..., x_n$*, then the determinant of A can be expressed as*

$$det(A) = \prod_{1 \leq i \leq j \leq n} (x_j - x_i)$$

Lemma 4.9. Let $\mathbf{X}$ denote the signature of a time series $\{(t_i, r_i)\}_{i=1}^{n}$ for a known $\{t_i\}_{i=1}^{n}$. Then,

$$\Delta R = T^{-1} S,$$

where

$$S := \begin{pmatrix} 0!\pi^{2}(\mathbf{X}) \\ 1!\pi^{12}(\mathbf{X}) \\ \vdots \\ (n-1)!\pi^{1\ldots12}(\mathbf{X}) \end{pmatrix}, T := \begin{pmatrix} 1 & 1 & \ldots & 1 \\ t_1 & t_2 & \ldots & t_n \\ \vdots & \vdots & \ddots & \vdots \\ t_1^{n-1} & t_2^{n-1} & \ldots & t_n^{n-1} \end{pmatrix},$$

$$\Delta R := \begin{pmatrix} r_1 \\ r_2 - r_1 \\ \vdots \\ r_n - r_{n-1} \end{pmatrix}.$$

Thus, $\{r_i\}$ can be represented as a linear functional on the signature of the time series.

Proof. To understand how the equation of the Lemma is constructed and proof it, is important to analyse the coordinate iterated integrals of of the path R. The information we have at this point is the values of the signature coordinates and the time steps that the time series took values. As we want to get the equation to find the values $\{r_i\}_i$, we need to find which coordinates of the signature are directly related with them. Actually, if we compute the following signature coordinate, considering the coordinates of R given by $R_s^{(1)}$ and $R_s^{(2)}$, time and $r(t)$ respectively,

$$\pi^{(1,2)}(\mathbf{X}) = \int_0^{2n-1} \frac{1}{2}(\pi^{(1)}(\mathbf{X}_s))^2 dR_s^{(2)}.$$

And, as $R(s)$ is constructed over an orthonormal basis of $\mathbb{R}^2$, we can simplify the calculations and note that in the time intervals of the kind $(2i+1, 2(i+1))$, the differential of the coordinate $R^{(2)}$ is null, so we have the expression

$$\pi^{(1,2)}(\mathbf{X}) = \sum_{i=0}^{n-1} \int_{2i}^{2i+1} \frac{1}{2} t_i^2 d\left(\left(r_{t_{i+1}} - r_{t_i}\right) s\right) = \sum_{i=1}^{n-1} \frac{1}{2} t_i^2 \left(r_{t_{i+1}} - r_{t_i}\right). \tag{4.3}$$

With $r_0 = 0$, as seen in the time-joined transformation. Note that it is clearly a linear combination of the $\{r_i\}_i$ values. In order to get the other equations to solve the system, we consider the general expression, obtained similarly than equation 4.3,

$$\begin{aligned}\pi^{(1,\dots,1,2)}(\mathbf{X}) &= \int_0^1 \frac{1}{k!} t_1^k d\left(r_{t_1} s\right) + \sum_{i=1}^{n-1} \int_{2i}^{2i+1} \frac{1}{k!} t_i^k d\left(\left(r_{t_{i+1}} - r_{t_i}\right) s\right) \\ &= \frac{1}{k!} t_1^k r_{t_1} + \sum_{i=1}^{n-1} \frac{1}{k!} t_i^k \left(r_{t_{i+1}} - r_{t_i}\right).\end{aligned}$$

In consequence, if we put all the equations in a matrix form, given $n \geq 2$, we get:

$$S = T\Delta R,$$

with the previous notation

$$S := \begin{pmatrix} 0!\pi^2(\mathbf{X}) \\ 1!\pi^{12}(\mathbf{X}) \\ \vdots \\ (n-1)!\pi^{1\dots12}(\mathbf{X}) \end{pmatrix}, T := \begin{pmatrix} 1 & 1 & \dots & 1 \\ t_1 & t_2 & \dots & t_n \\ \vdots & \vdots & \ddots & \vdots \\ t_1^{n-1} & t_2^{n-1} & \dots & t_n^{n-1} \end{pmatrix},$$

$$\Delta R := \begin{pmatrix} r_1 \\ r_2 - r_1 \\ \vdots \\ r_n - r_{n-1} \end{pmatrix}.$$

Now, continuing with the proof, since T is the transpose of a square Vandermonde matrix with all $x_i = t_i$ different from each other, T is also invertible. So, we can state that

$$\Delta R = T^{-1}S.$$

□

4.3 Expected Signature model for time series

When analyzing time series models, it is crucial to understand the conditions under which the problem operates. One of the most common conditions is the stationarity of the time series, or more generally, the stationarity of a stochastic process. Let's define this concept. Refeer to [11] for a deeper insight of time series theory.

Definition 4.10. We say that a process $\{X_k, k \in \mathbb{Z}\}$ is strictly stationary if for any $k_1, ..., k_n$ and l the vectors

$$(X_{k_1}, ..., X_{k_n}) \text{ and } (X_{k_1+l}, ..., X_{k_n+l})$$

have the same law. In particular, taking $n = 1$, all variables have the same law.

Definition 4.11. We say that a process $\{X_k, k \in \mathbb{Z}\}$ is weakly stationary if

1. $\mathbb{E}[X_k] = \mu \in \mathbb{R}$, for any $k \in \mathbb{Z}$,
2. $\mathcal{C}(X_k, X_k + l) = \gamma(l) \forall k, l \in \mathbb{Z}$ where γ is defined on $\mathbb{N}$, due to symmetry of covariance.

The primary assumption in defining the ES model is the stationarity of the time series $\{r_i\}_i$. This means that the distribution of any $(r_k, ..., r_{k+n})$ remains consistent throughout the entire time series. A clear conclusion on the previous definitions is that strict stationarity implies weak stationarity.

Following this, let's denote the present time as t_k, and $\mathcal{F}_k$ the information we have until t_k. Also, denote $\mathbf{X}_k$ as the signature of the past steps, $S(\{(t_i, r_i)\}_{i=k-p+1}^{k})$, and $\mathbf{Y}_k$ as the signature of the future steps, $S(\{(t_i, r_i)\}_{i=k}^{k+q})$.

When applying the ES model to an univariate time series $\{(t_i, r_i)\}_{i=1}^{N}$, we will try to predict the signature of q-future values given the signature of the last p values, given $p, q \in \mathbb{N}$. By these means, there must exist a linear relation between the signature of the past and the signature of the future.

Definition 4.12. (Expected Signature model ES(p,q,n,m)). Given an stationary time series $\{r_i\}_{i=1}^N$. We say that it satisfies the assumptions of the expected signature model with parameters p, q, n and m, $ES(p,q,n,m)$, if the following holds. There exists a linear function $f : T^n(\mathbb{R}^2) \mapsto T^m(\mathbb{R}^2)$ such that

$$\rho_m(\mathbf{Y}_k) = f(\rho_n(\mathbf{X}_k)) + \epsilon_k,$$

where $\mathbb{E}[\epsilon_k|\mathcal{F}_k] = 0$ and $N >= p + q + 1$.

Now, we denote as μ_k the expectation of $\mathbf{Y}_k$ conditional on the information up to the time t_k, i.e.

$$\mu_k = \mathbb{E}[\mathbf{Y}|\mathcal{F}_k] \tag{4.4}$$

which means that μ_k is a function f on $\mathbb{X}_k$, i.e.

$$\begin{aligned} f : T((\mathbb{R}^2)) &\mapsto T((\mathbb{R}^2)) \\ \mathbf{X}_k &\mapsto \mu_k \end{aligned} \tag{4.5}$$

And the conditional covariance function of $\mathbf{Y}_k$

$$\begin{aligned} \Sigma^2 : A^* \times A^* &\mapsto \mathbb{R} \\ (I,J) &\mapsto \text{Cov}\left(\pi^I(\mathbf{Y}_k), \pi^J(\mathbf{Y}_k)|\mathcal{F}_k\right) \end{aligned} \tag{4.6}$$

Note that the ES model utilizes Corollary 2.15 and assumes that considering the higher-order terms of the signature is sufficient for the linear regression to hold. Therefore, the most important information is contained within these terms, providing facilities to manipulate the model computationally. It also assumes that the distribution of the future values on condition of the current information $\mathcal{F}_k$ depends uniquely on the truncated signature of the last p+1 values of the time series. The ES model does not assume that future values of the time series, given the current information, follow a specific distribution. Instead, it describes the likely distribution of future data points in a flexible, non-parametric manner, enabling the description and prediction of future outcomes based on the present information.

4.4 Classical time series models as ES-model special cases

In classical time series analysis, the foundational models that often come to mind are AR (Autoregressive), ARMA (Autoregressive Moving Average), and ARCH (Autoregressive Conditional Heteroskedasticity). They have been extensively studied and widely applied, particularly in the financial sector, to explain and predict patterns in data such as logarithmic prices. They have been settled as standard tools in the analysis of the dynamics and volatilities inherent in time-dependent data. See complete definitions in the Appendix 6.1.

As prediction-based models, they aim to forecast two key parameters of future values based on the information available up to time t_k: the conditional expectation of the future value r_{k+1} and its variance. In this section, we will explore how these models are specific instances of the ES model. To achieve this, we will derive the same parameters that traditional time series models estimate using a specific ES model, denoted as ES(p, q, n, m). Let's denote this parameters as

$$\begin{aligned} m_k &:= \mathbb{E}[r_{k+1}|\mathcal{F}_k] \text{ and,} \\ \sigma_k^2 &:= \mathrm{Var}[r_{k+1}|\mathcal{F}_k]. \end{aligned} \tag{4.7}$$

In the previous section, we presented the ES model, and how the signature of future values could be obtained from the signature of the p-lag values. This idea is related to the traditional time series approach. Nevertheless, in this case, we would like to uniquely predict the expectation of the future value given present information at time t_k, i.e. μ_k, how can this value be obtained from the signature of the future q steps? In order to answer to this question, we will recall Lemma 4.9 at Lemma 4.13.

Lemma 4.13. Given a time series $\{(t_i, r_i)\}_{i=1}^n$ and its signature $\mathbf{X}$ (as the signature of the join-transformed path of the time series), then

$$r_n = \pi^{(2)}(\mathbf{X}) \tag{4.8}$$

Proof. It is straightforward by using Lemma 4.9. □

Remark 4.14. When considering the ES(p,q,n,m) model to predict the immediate future step on condition of $\mathcal{F}_k$, we will assign $q = 1$. Concretely, the ES(p,1,n,m) model provides the conditional expectation of the future time series $\{(t_{k+i}, r_{k+i})_{i=0}^1\}$, on condition of the signature of the p-lagged values.

Corollary 4.15. The ES(p,1,n,m) model is used to forecast the next step of a time series at a given the information at time t_k by taking the coordinate $\pi^{(2)}$ of $\mathbf{Y}_k$, i.e. $m_k = \mathbb{E}[r_{k+1}|\mathcal{F}_k] = \pi^{(2)}(\mathbf{Y}_k)$, we name μ_k as the truncated signature obtained by the ES(p,1,n,m) model. Hence, by Lemma 3.7 and the shuffle product property, the following holds

$$\begin{aligned} m_k &= \pi^{(2)}(\mu_k), \\ \sigma_k^2 &= 2\pi^{(2,2)}(\mu_k) - (\pi^{(2)}(\mu_k))^2. \end{aligned} \tag{4.9}$$

Having explored how to use the ES model to forecast the immediate future step in a time series, it is now time to demonstrate how traditional time series models are special cases of the ES(p, q, n, m) model. To do this, we first need to present a preliminary result. Let's present a modification of Lemma 4.11. in [3].

Lemma 4.16. Suppose that a time series $\{r_k\}_k$ satisfies ARMA(p,q) conditions. Then,

$$\mu_k = \phi_0 + \sum_{i=1}^{p} \phi_i r_{k-i},$$

being ϕ a constant vector. Suppose $\epsilon = \{\epsilon_0, ..., \epsilon_q\}$ is the white noise error term, with mean 0 and variance σ^2. Then, for $n = \{1,2\}$, $\mathbb{E}[r_{k+1}^n | \mathcal{F}_k]$ is a polynomial of lagged (p+q) values of $\{r_k\}$.

Proof. Given $n \in \{1,2\}$,

$$\begin{aligned} \mathbb{E}[r_k^n | \mathcal{F}_{k-1}] &= \mathbb{E}[(\mu_k + \sum_{j=1}^{q} \theta_j \epsilon_{k-j} + \epsilon_k)^n | \mathcal{F}_{k-1}] \\ &= \mathbb{E}[(\mu_k + \sum_{j=1}^{q} \theta_j \epsilon_{k-j} + \epsilon_k)^n | \mathcal{F}_{k-1}] \\ &= \sum_{l=0}^{n} \mathbb{E}[C_n^l \mu_k^{n-l} (\sum_{j=1}^{q} \theta_j \epsilon_{k-j} + \epsilon_k)^l | \mathcal{F}_{k-1}] \\ &= \sum_{l=0}^{n} C_n^l \mu_k^{n-l} \mathbb{E}[(\sum_{j=1}^{q} \theta_j \epsilon_{k-j} + \epsilon_k)^l | \mathcal{F}_{k-1}] \end{aligned} \tag{4.10}$$

Then, by using the definition of the error term $\epsilon_{k-j} = r_{k-j} - \mu_{k-j}$, we compute $\mathbb{E}[(\sum_{j=1}^{q} \theta_j \epsilon_{k-j} + \epsilon_k)^l | \mathcal{F}_{k-1}]$ taking $l = 1$,

$$\mathbb{E}[\sum_{j=1}^{q} \theta_j \epsilon_{k-j} + \epsilon_k | \mathcal{F}_{k-1}] = \sum_{j=1}^{q} \theta_j \epsilon_{k-j} = \sum_{j=1}^{q} \theta_j (r_{k-j} - \mu_{k-j}), \tag{4.11}$$

and $l = 2$,

$$\mathbb{E}[(\sum_{j=1}^{q} \theta_j \epsilon_{k-j} + \epsilon_k)^2 | \mathcal{F}_{k-1}] = \sum_{j=1}^{q} \theta_j^2 (r_{k-j} - \mu_{k-j})^2 + \sigma_k^2. \tag{4.12}$$

Finally, considering that

$$\mu_{k-j} = \phi_0 + \sum_{i=1}^{p} \phi_i r_{k-j-i}, \forall j \in \{1, ...q\},$$

we can see that $\mathbb{E}[r_k^n|\mathcal{F}_{k-1}]$ is a polynomial of (p+q) lagged values of $\{r_k\}_k$ for both $n = 1$ and $n = 2$. □

Theorem 4.17. Suppose that a time series $\{r_k\}$ satisfies ARMA(p,q) conditions, and its mean equation is

$$\mu_k = \phi_0 + \sum_{i=1}^{p} \phi_i r_{k-i},$$

there exist an integer N such that the time series satisfies the assumptions of ES($p + q$, 1, N, 2), hence, ARMA(p,q) is a particular case of the ES model.

Proof. This result is equivalent to proof that the first and second moment, $\mathbb{E}[r_k|\mathcal{F}_{k-1}]$ and $\mathbb{E}[r_k^2|\mathcal{F}_{k-1}]$, can be expressed as a linear functional on the signature of the (p+q) lagged values of r_k, which have been proven at Lemma 4.16 to be true.

We know that linear forms on the signature of (p+q)-lagged values of r_k are dense in the space of smooth functions on (p+q)-lagged values of r_k, and that the signature of a time series completely determines the time series. Hence, there exists a linear functional f_0 such that

$$\mathbb{E}[r_k|\mathcal{F}_{k-1}] = f_0(S(\{(t_{k-i}, r_{k-i})\}_{i=1}^{p+q})),$$

and a linear functional f_1 such that

$$\mathbb{E}[r_k^2|\mathcal{F}_{k-1}] = f_1(S(\{(t_{k-i}, r_{k-i})\}_{i=1}^{p+q})).$$

And these two equations are enough to obtain a precise approximation of the ARMA(p,q) model parameters, mean and variance. □

This result can be proven for other different traditional time series models, as the ARCH (see [3]), GARCH or AR (specific case of the last theorem, with q=0). Considering these results, one can conclude that the ES approach provides at least as good predictions of time series data as traditional time series models, if the correct parameters p, q, n, m are chosen.

Chapter 5

Using the ES model for financial time series prediction

This chapter will investigate the conditions under which AR and ARMA models are special cases of the ES model. Additionally, it will be demonstrated how the ES model can enhance the results obtained from financial data traditionally analyzed using AR or ARMA models. While the main focus of this thesis is establishing path signatures as a new approach to time series analysis, providing computational examples strengthens the theoretical foundation. The code itself will be included in the appendix for interested readers. This experiments will explore in the conditions that: AR and ARMA are special cases of the ES model, and the ES model provides an improvement in the results that can be obtained from financial data, when considering data that was traditionally studied using the AR or ARMA. The full code is available in the (Github repository).

5.1 Unveiling Computational Insights: AR as a particular case of the ES approach

As recently explained, traditional time series models can be seen as specific instances of the expected signature model. To accomplish this, we will generate a time series by using AR conditions. Subsequently, we will employ the ES model in order to approximate them in optimal precision, and check that the information that generates the AR time series is encapsulated within the ES model. Following this, an analysis of accuracy statistics and computational outcomes will facilitate our conclusions. Recall that the AR (AutoRegressive) model is an ARMA (AutoRegressive Moving Average) with q=0.

We consider the time series $\{r_t\}_t$ generated by an AR model with length $n = 4000$, verifying the equation:

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + \epsilon_t, \tag{5.1}$$

with $\phi = \{0, 0.4, -0.5, -0.2\}$ as constant parameters and ϵ_t a white noise with variance $\sigma^2 = 2$. The time series generated can be seen in figure 5.1.

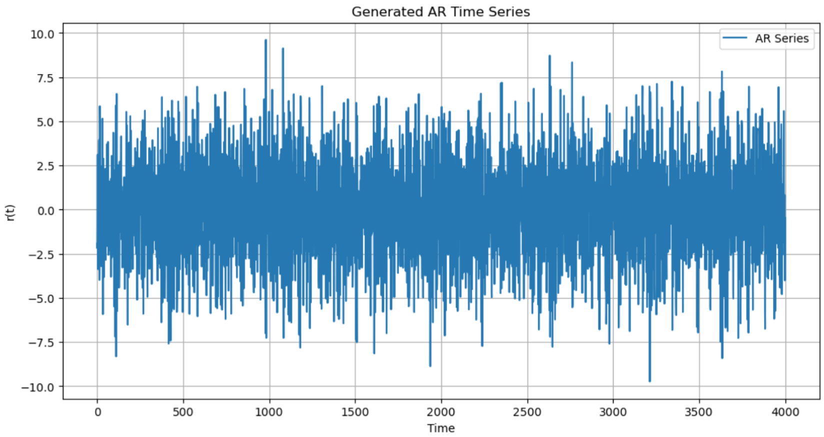

Figure 5.1: AR series

Remark 5.1. An AR(p) time series with $p > 1$ is stationary if the polynomial $1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p = 0$ lie outside the unit circle in the complex plane. Refer to [11] for more details. It can be checked that the generated AR(3) time series with parameters $\phi = \{0, 0.4, -0.5, -0.2\}$ is stationary.

As our main goal is to check that the AR model is a special case of the ES model, we aim to obtain the same expected future value and conditional variance given information up to time $\mathcal{F}_k$. It will be sufficient to obtain the same $\mathbb{E}[r_{k+1}|\mathcal{F}_k]$ and $\mathbb{E}[r_{k+1}^2|\mathcal{F}_k]$ as the AR model. First, we need to determine the suitable p, q, n and m parameters of the ES(p,q,n,m) model. As each step of the AR time series is constructed by using a linear combination of the prior $p = 3$ values, we can use Theorem 4.17 to state that $p = 3$ in our ES model. Then, in order to compute the immediate expected future step of a time series, we noted in Corollary 4.15 that q parameter must be set to 1, this allows us to use the second coordinate of the expected signature obtained by the model, i.e. $\pi^{(2)}(\mathbb{E}[\mathbf{Y}|\mathcal{F}_k])$, to predict the future step k+1.

The other two parameters to determine are n and m, the order of the truncated signatures that will be used in the ES regression. We have seen how the factorial decay property of the signature is key to understand that the most important information that the signature of a path contains is within its low order coordinates. As we will use the coordinates of $\mathbf{X}_k$ as features in our regression model, its important to use the minimum necessary to avoid non-informative noise and computational complexity. As seen in Theorem 4.17, there exists a sufficiently large $n = N$ which approximates the ES(3,1,N,2) model to the AR in the most suitable manner. We will begin with $n = 4$, and modify if results are unsuccessful. Ultimately, as we have demonstrated, the most relevant information contained in the future steps signature is found in the first and second order coordinates. Therefore, we will set a truncation order $q = 2$.

Overall, $f : T^4(\mathbb{R}^2) \mapsto T^2(\mathbb{R}^2)$ is the linear functor that represents the regression in the ES(3,1,4,2) model as

$$\rho_2(\mathbf{Y}_k) = f(\rho_4(\mathbf{X}_k)) + \epsilon_k, \tag{5.2}$$

and $\mathbb{E}[\epsilon_k|\mathcal{F}_k] = 0$.

In order to learn the function f, we are going to use linear regression, concretely, LASSO regression, a regression method suitable for a dataset which contains a high number of features. Another benefit of LASSO regression is its ability to prevent overfitting. By penalizing the use of many coefficients in the regression, LASSO encourages simpler models that generalize better to new data.

After obtaining the AR series of length $n = 4000$, we train the regression model using the signatures of the windows of the past p-values as independent variables and the signatures of the future values as dependent variables. The data processing that has been followed to construct the ES model has been:

1. Compute all the rolling windows of size $p = 3$ of the AR time series, as these will be the windows used to compute $S(\{(t_i, r_i)\}_{i=k-p+1}^{k})$.

2. Compute all the rolling windows corresponding to the future steps of the AR time series. These rolling windows will be of size 2, taking the present value and the immediate future step ($q = 1$).

3. Calculate the Expected AR time series, which consists of calculating, for each step of the AR time series, the next step expected value using the formula:

   $$\mu_k = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i}$$

   The Expected AR time series will be used to compare the expected future value predicted by the ES model in comparison with the expected future value underlying on the AR.

4. For each of the windows, the following modifications are applied, in order to meet the conditions of this theoretical model:

    (i) Time augmentation: Transform the 1-dimensional windows $\{r_t\}_t$ into the 2-dimensional windows $\{(t, r_t)\}_t$.

    (ii) Base-point augmentation: Add the base-point $(-1, 0)$ at the beginning of each window.

    These two modifications lead to a unique correspondence between the windows and their signatures, as seen in Lemma 2.14, providing meaningful data for the regression.

5. Compute the signature of both past and future windows. To this mean, we have used the *esig* python package developed by T.Lyons and his team. This package provides functions to compute the signature of the transformed path from given time series as data points.

6. Split the data into train/test and train the model: In the model, the signatures of the past p-values are used as independent variables (X) and the signatures of the future as dependent variables (Y). This data is split into two sets, train and test (80% train and 20%test), obtaining X_train, X_test, Y_train and Y_test. Next, LASSO regression is applied for training with a regularization parameter (α) set to 0.01.

7. Test the model using X_test as input variables. The output data obtained during testing consists of the expected signatures of future steps for each input in the test data, as generated by the ES model.

8. For each output, take the signature coordinate $\pi^{(2)}$ as the expected future step obtained by the ES(3,1,5,2) model during the test phase, and take 2 times the signature coordinate $\pi^{(2,2)}$, as we have proven in Lemma 3.7 that it is equivalent to $\mathbb{E}[r_{k+1}^2 | \mathcal{F}_k]$.

Once obtained the predictions of the ES model we check the accuracy using two methods: first, overlapping the last 100 steps of the ES predictions with the Expected AR time series, and second, by creating a regression line between the expected values that the AR model computes in each step and our ES predictions (Figure 5.2).

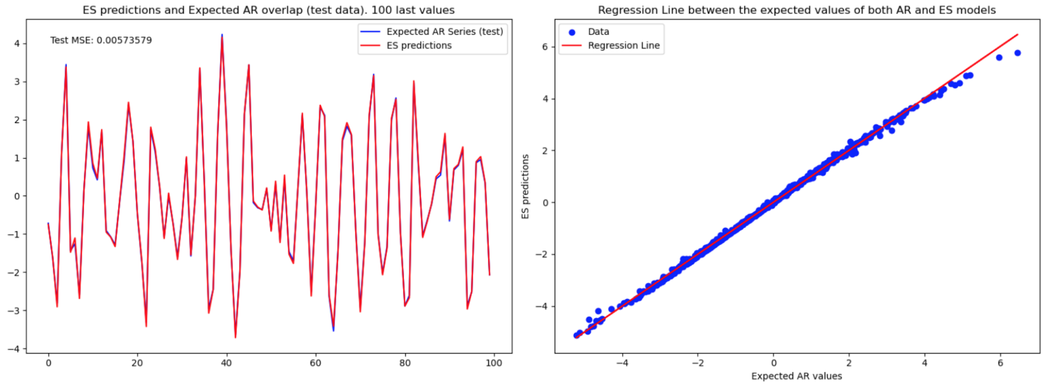

Figure 5.2: (left) Overlap of the ES model predictions and Expected AR Time Series. Last 100 values. (right) Regression line

In addition to the visual results, the model yields a coefficient of determination of $R^2 = 0.99$ and a mean squared error (MSE) of 0.0057, confirming that the $\mathbb{E}[r_{k+1}|\mathcal{F}_k]$ obtained by the ES(3,1,4,2) is nearly identical to the $\mathbb{E}[r_{k+1}|\mathcal{F}_k]$ used to generate the time series. Now, we do a similar procedure to show how the $\mathbb{E}[r_{k+1}^2|\mathcal{F}_k]$ is also well predicted. We plot a regression line between the expected conditional squared r_{k+1} generated by the AR and the ones obtained by our model. The values corresponding to the AR model can be computed using Lemma 4.16, specifically, $\mathbb{E}[r_{k+1}^2|\mathcal{F}_k]$ for the AR model is equal to $\mu_k^2 + \sigma_k^2$. The coefficient of determination obtained is $R^2 = 0.994$. See Figure 5.3.

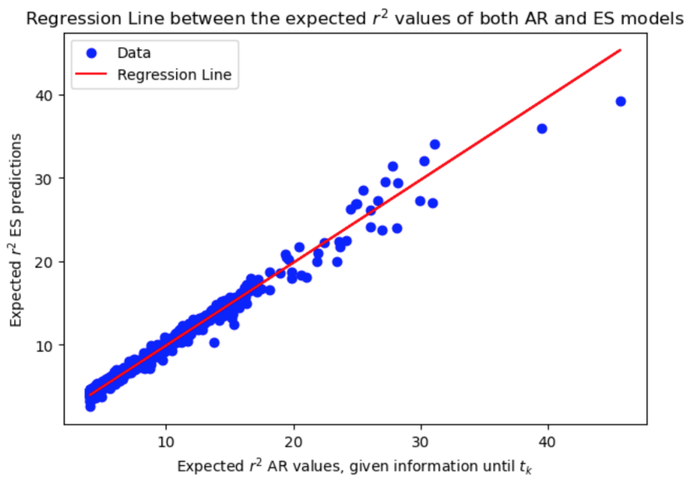

Figure 5.3: Regression line between the expected squared future values for both AR and ES

Remark 5.2. For an AR time series, we have seen in 4.9 that the conditional variance can be directly derived from $\mathbb{E}[r_{k+1}|\mathcal{F}_k]$ and $\mathbb{E}[r_{k+1}^2|\mathcal{F}_k]$.

Given the accuracy with which the ES(3,1,4,2) model approximated the mean and conditional variance of the AR time series, considering the expected error from truncating the signatures and the inherent error in the LASSO regression, we can conclude that our theoretical framework has been validated by these practical results and the AR model is a specific case of the Expected Signature model. This approach can be similarly reproduced for an ARMA time series with $q > 0$.

5.2 ES model for financial time series forecasting

During this thesis we have presented an alternative to the traditional time series models. Actually, we have proven how traditional time series are special cases of the ES model, highlighting the extensive information embedded within this model. Nevertheless, it is clear that financial data is inherently complex and influenced by many factors, making it really challenging to model and predict. Many models, such as ARMA, ARIMA or GARCH time series models, aim to capture underlying patterns and trends, but the battle against randomness is often unsuccessful. By these means, we aim to explore whether the ES model improves modelling and forecasting of real financial data in comparison to traditional time series models.

When modelling financial data, it is important to understand which patterns and features characterize the time series, in order to use the most suitable theoretical approach. Nevertheless, time series data is the most of the time unprepared for its analysis, some techniques must be applied before moving to forecasting.

5.2.1 Catfish sales and forecasting

We are going to analyze an open-source dataset which contains historical data of catfish monthly sales (catfish sales 1986-2012). Concretely, we will study the data within the dates "1992-1-1" and "2012-1-1". This study is an extension of [15].

Before comparing the Expected Signature (ES) model with traditional time series models, it is essential to comprehend key features of the data. This understanding aids in formulating assumptions about which models might be more appropriate. The first interesting approach is to make a STL decomposition of the data, which means, to decompose the time series into three components: season, trend and residual. See figure 5.4.

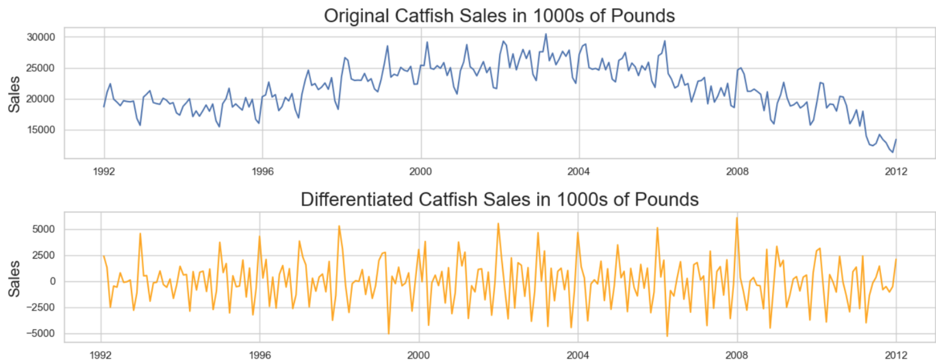

Figure 5.4: Original and differentiated catfish monthly sales in 1000s of Pounds

It is appreciable that the sales time series has an upwards-downwards trend, and a 12-month seasonal pattern. Then, in order to get information about how past data influences future data, i.e. p-lag values, there are some methods to determine them, for instance, the Autorrelation and Partial Autocorrelation function plots (6.5 for the formal definitions). See the results in Figure 5.5. The shaded area represents the confidence intervals for the autocorrelation values.

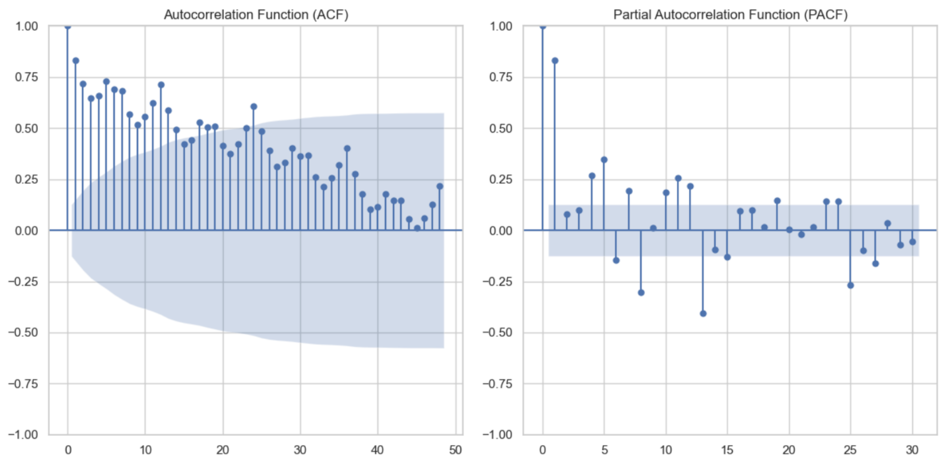

Figure 5.5: ACF and PACF plots

From the plots 5.5, we can note that there is a significant spike around 6 and 12 months lag, and a sinusoidal behaviour in PACF, which tell us that there might be a seasonal effect in our data. Taking all this assumptions into consideration, we can estimate which will be the best traditional time series model for to tackle this problem. As the PACF plot exhibits a seasonal character, it suggests that the data may be well modeled using the SARIMA model (refer to the Annex).

There exists a useful function called *auto_arima* from package *pmdarima*, which employs a stepwise approach to search through multiple combinations of ARIMA (and SARIMA) and selects the model with the lowest AIC (Akaike Information Criterion) score. In our code, the outcome of this function confirms our prior assumptions by suggesting that the model SARIMAX(0,1,0)x(1,0,[1],12).

Now, we will split our data into train/test (80% train and 20% test), fit the model to the training data, and then predict the future data and compare it with the test. The results obtained will be analyzed by considering the time series paths comparison and a regression line (and R^2 coefficient) between actual and predicted differentiated data (for a proper comparison with the ES model). The actual test values vs the predicted values using the SARIMA model can be seen in Figure 5.6. The explanation of the training data appears to be quite accurate and improves at each iteration, but the testing fit is confusing and results in significant errors in many of the steps. The corresponding linear regressions can be seen in Figure 5.7, which yield a coefficient of determination $R^2 = 0.46$ for the training data and $R^2 = 0.26$ for the test data.

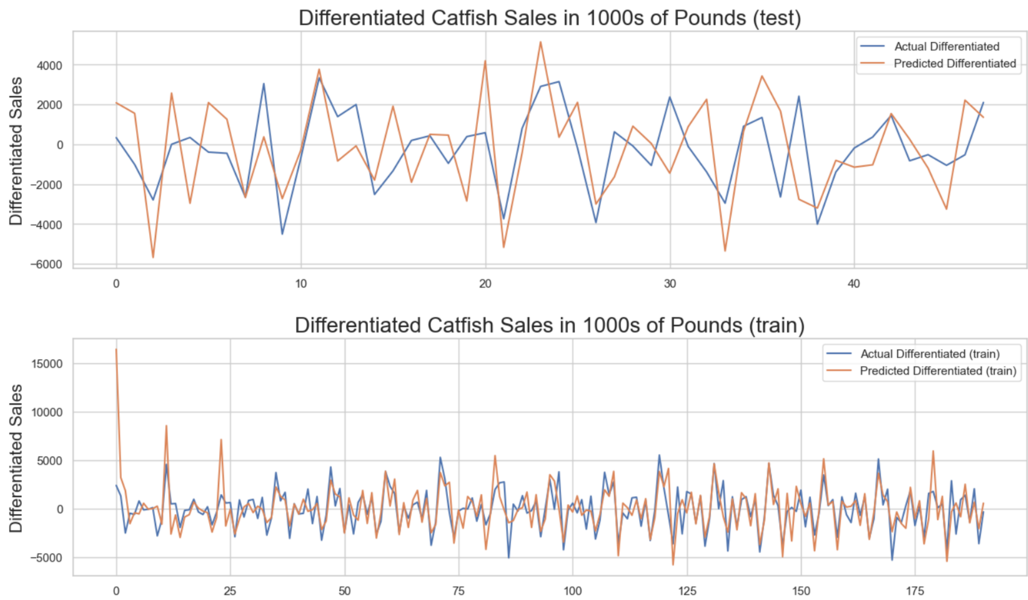

Figure 5.6: (Top) Predicted vs. actual differentiated data using SARIMA model (test data) - (Bottom) Predicted vs actual differentiated data using SARIMA model (train data)

It is time to compare this results with the model we have been developing throughout this thesis, the ES(p,q,n,m) model. To be under the conditions of the

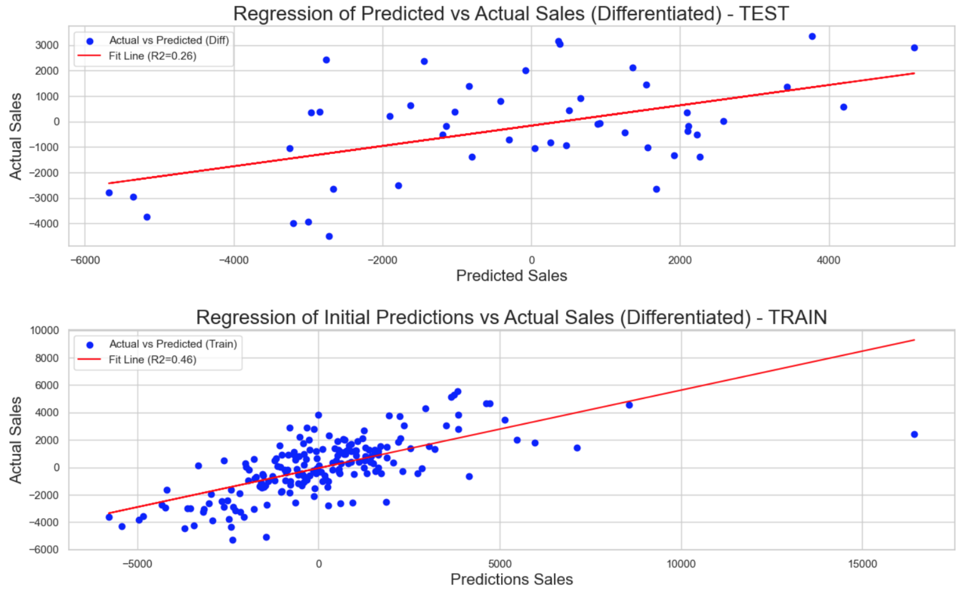

Figure 5.7: (Top) Linear regression on predicted vs. actual differentiated data using SARIMA model (test data) - (Bottom) Linear regression on predicted vs actual differentiated data using SARIMA model (train data)

ES model, the data that will be used will be the differentiated time series, as it improves the stationary of the data. It seems quite obvious to define a p-lag of 12, as we encountered that there exists a seasonality of length 12. The parameter will be set by choosing the best one fitting the training and test data. Finally, as it is a model to predict immediate future step of a time series, we set $q = 1$ and $m = 2$. The results we are considering are the same that the previous model, paths comparison and linear regression (coefficient of determination). In Figure 5.8, the predicted vs actual paths are displayed, and it is appreciable a better approximation of the predictions and a better understanding of the patterns in the training data, these results are sustained by the linear regressions and its corresponding R^2, which can be seen in Figure 5.9.

The results obtained by the ES model are better, with coefficients of determination $R^2 = 0.34$ for the test data and $R^2 = 0.61$ for the training data. This indicates that the model provides a superior explanation of the patterns underlying the training data and offers more accurate predictions for new data.

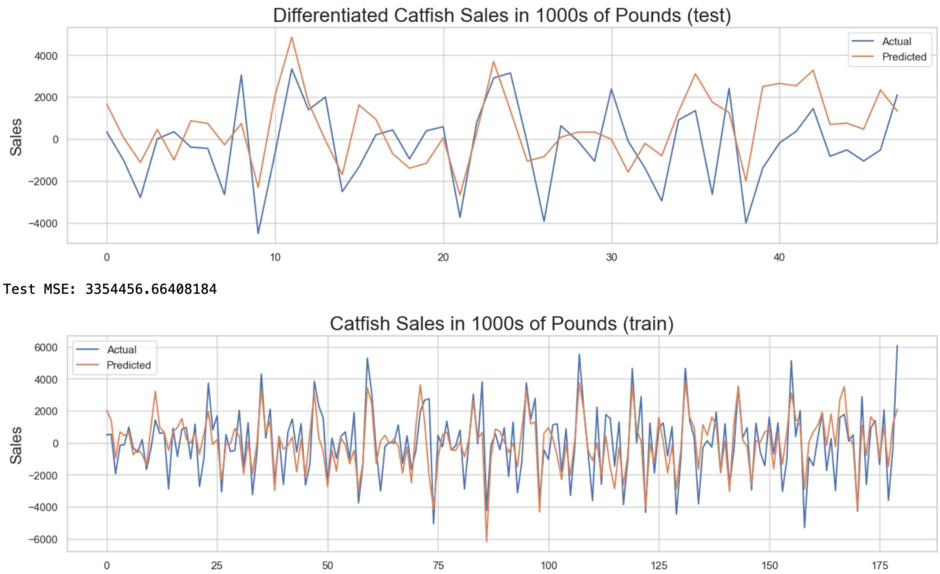

Figure 5.8: (Top) Predicted vs. actual differentiated data using the ES(12,1,4,2) model (test data) - (Bottom) Predicted vs. actual differentiated data using the ES(12,1,4,2) model (train data)

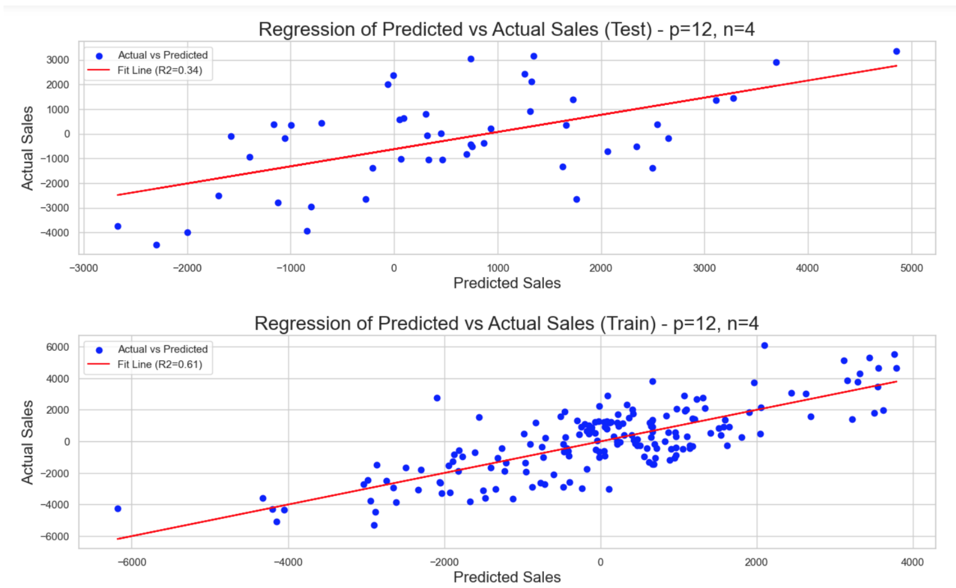

Figure 5.9: (Top) Linear regression on predicted vs. actual differentiated data using the ES(12,1,4,2) model (test data) - (Bottom) Linear regression on predicted vs. actual differentiated data using the ES(12,1,4,2) model (train data)

Chapter 6

Conclusions

After exploring the theoretical framework and practical applications of the ES model conclusions can be drawn. Firstly, it has been demonstrated how the the traditional time series models can be seen as a specific case of the ES model in some defined conditions. This might lead to a misconception that traditional time series are obsolete since they can be generalized within a broader model. However, this is not the case, traditional time series models are still necessary in order to obtain a proper explanation of how the data is modelled, as they offer concrete equations underlying the time series.

The ES model is a non-parametric model that employs machine learning techniques, leading to the phenomenon known as the "black box", where the relation between the income and the outcome is uncertain. We have seen how the ES model performed significantly better in both modelling the training data and predicting unknown data. Contrary to the common criticism that ML models are very prone to overfitting and often fail in real-world prediction scenarios, our model exhibited strong performance in the prediction phase. This improvement in accuracy comes from the ES model's ability to capture data patterns similarly to traditional time series while incorporating additional features in the different signature coordinates, providing extra information beneficial for prediction.

Historically, traditional time series models have been seen as a good way to model and display the equations and patterns within the data, but they have been quite unsuccessful when referring to the prediction phase. Hence, this counterpart can be obtained by the Expected Signature model, which doesn't contain an equation that models each step of the data, but performs better in prediction. When comparing and analyzing time series models, it is crucial to ensure that the data satisfies the conditions of the models used; if noisy, uncleaned, or unprocessed data is used as input, none of the models will yield satisfactory results.

I want to highlight that throughout the completion of this thesis, I thoroughly enjoyed every specific phase I was involved in, from gathering information to drawing conclusions and finalizing my writing. It has been a journey that surpassed my initial expectations. Contrary to what I had anticipated, this experience has opened the door to further research phases in my life, as it has been truly satisfying.

Regarding further research in rough path theory, I am very excited to start my master's program to delve deeper into this complex field and gain a comprehensive understanding of the theory. I am aware of its difficulty and the years of study it might require, but I believe every step and contribution to this research area will be insightful.

Bibliography

[1] A. Fermanian, *Embedding and learning with signatures*, Computational Statistics and Data Analysis, Vol. 157, pp.107148, May 2021.

[2] B.M. Hambly, T. Lyons, *Uniqueness for the signature of a path of bounded variation and the reduced path group*, Annals of Mathematics, Vol. 171 109â167, 2010.

[3] D. Levin, T. Lyons, H. Ni, *Learning from the past, predicting the statistics for the future, learning an evolving system*, arXiv:1309.0260v6 [q-fin.ST], 22 Mar 2016.

[4] H. Ni, L. Margaret, *The expected signature of a stochastic process*, Oxford University Research Archive, 2012.

[5] I. Chevyrev, A. Kormilitzin, *A primer on the signature method in Machine Learning*, arXiv:1603.03788v1 [stat.ML], 11 Mar 2016

[6] I. Pérez, *Rough Path Theory and Signatures Applied To Quantitative Finance - Part 1*, QuantStart articles.

[7] I. Chevyrev, T.Lyons, *Characteristic functions of measures on geometric rough paths*, Ann. Probab. 44 (6) 4049 - 4082, November 2016.

[8] K. T. Chen, *Integration of paths - a faithful representation of paths by noncummutative formal power series*, Transactions of the American Mathematical Society Vol. 89, No. 2, pp. 395-407, Nov., 1958.

[9] L. G. Gyurkó, T. Lyons, M. Kontkowski, and J. Field, *Extracting information from the signature of a financial data stream*, arXiv:1307.7244v2 [q-fin.ST], 15 Jul 2014.

[10] R. Ree, *Lie elements and an algebra associated with shuffles*, Annals of Mathematics, Second Series, Vol. 68, No. 2, pp. 210-220, Sep. 1958.

[11] R. H Shumway, D. S Stoffer, and D. S Stoffer, Time series analysis and its applications, Springer, Vol.3, 2000.

[12] T. Lyons, A.D. McLeod, *Signature Methods in Machine Learning*, arXiv:2206.14674v5 [stat.ML] 26 Jan 2024.

[13] T. Lyons, T. Lévy, M. Caruana, *Differential Equation driven by Rough Paths*, Springer, 2006.

[14] T. Lyons, *Differential equations driven by rough signals*, Revista MatemÃ¡tica Iberoamericana 14.2): 215-310, 1998.

[15] A. Bajaj, *ARIMA & SARIMA: Real-World Time Series Forecasting*, MLOps Blog, neptune.ai, August 2023.

Appendix

6.1 Traditional time series models

Definition 6.1. (ARMA model). Let $\{r_t\}_t$ be a time series. We note ARMA(p,q) as an autoregressive moving average model of order p and q. This ARMA(p,q) model with parameters $\phi = [\phi_0, ..., \phi_p]$, $\theta = [\theta_1, ..., \theta_q]$ and $p, q \in \mathbb{N}$ is defined as

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t, \tag{6.1}$$

where ϕ and θ are the autorregressive and moving average parameters, respectively. Also, p and q are non-negative integers, and ϵ_t is a white noise with mean 0 and variance σ_ϵ^2.

Definition 6.2. (ARCH model). Let $\{r_t\}_t$ be a time series. We note ARCH(q) as an autoregressive conditional heteroskedasticity model of order q. This ARCH(q) model with parameters $\alpha = [\alpha_0, ..., \alpha_q]$ and $q \in \mathbb{N}$ is defined as

$$\epsilon_t = \sigma_t z_t, \tag{6.2}$$

where z_t is a white noise process with mean 0 and variance 1, and

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2, \tag{6.3}$$

where $\alpha_j \geq 0$ for all j and $\alpha_0 > 0$.

Definition 6.3. (ARIMA model). Let $\{r_t\}_t$ be a time series. We note ARIMA(p,d,q) as an autoregressive integrated moving average model of order p, d, and q. This ARIMA(p,d,q) model with parameters $\phi = [\phi_1, ..., \phi_p]$, $\theta = [\theta_1, ..., \theta_q]$ and $p, d, q \in \mathbb{N}$ is defined as

$$\Delta^d r_t = \phi_0 + \sum_{i=1}^{p} \phi_i \Delta^d r_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t, \tag{6.4}$$

where Δ^d is the differencing operator applied d times, and ϕ and θ are the autoregressive and moving average parameters, respectively. Also, p, d, and q are non-negative integers, and ϵ_t is a white noise with mean 0 and variance σ_ϵ^2.

Definition 6.4. (SARIMAX model). Let $\{r_t\}_t$ be a time series. We note

$$\text{SARIMAX}(p,d,q)(P,D,Q)_s$$

as a seasonal autoregressive integrated moving average model with exogenous regressors, of order p, d, q, P, D, and Q with seasonality s. This model with parameters $\phi = [\phi_1, ..., \phi_p]$, $\theta = [\theta_1, ..., \theta_q]$, $\Phi = [\Phi_1, ..., \Phi_P]$, $\Theta = [\Theta_1, ..., \Theta_Q]$ and $p, d, q, P, D, Q, s \in \mathbb{N}$ is defined as

$$\Phi(B^s)\phi(B)\Delta^d\Delta_s^D r_t = \Theta(B^s)\theta(B)\epsilon_t + X_t\beta, \tag{6.5}$$

where B is the backshift operator, Δ^d is the non-seasonal differencing operator applied d times, Δ_s^D is the seasonal differencing operator applied D times, $\phi(B)$ and $\theta(B)$ are the non-seasonal autoregressive and moving average polynomials of orders p and q, respectively, $\Phi(B^s)$ and $\Theta(B^s)$ are the seasonal autoregressive and moving average polynomials of orders P and Q, respectively, X_t represents the exogenous regressors, and β represents their coefficients. Also, p, d, q, P, D, Q, and s are non-negative integers, and ϵ_t is a white noise with mean 0 and variance σ_ϵ^2.

6.2 Time series autocorrelation

Definition 6.5. (Autocorrelation). Let $\{r_t\}_t$ be a time series with mean μ and variance σ^2. The autocorrelation function (ACF) at lag k, denoted by ρ_k, is defined as

$$\rho_k = \frac{\mathbb{E}[(r_t - \mu)(r_{t-k} - \mu)]}{\sigma^2}, \tag{6.6}$$

where $\mathbb{E}[\cdot]$ denotes the expected value operator. The autocorrelation ρ_k measures the linear relationship between r_t and r_{t-k}.

Definition 6.6. (ACF Plot). The Autocorrelation Function (ACF) plot is a graphical representation of the autocorrelations of a time series $\{r_t\}_t$ at different lags k. The ACF plot helps in identifying the extent of correlation between values of the time series separated by different time lags.

Definition 6.7. (PACF Plot). The Partial Autocorrelation Function (PACF) plot is a graphical representation of the partial autocorrelations of a time series $\{r_t\}_t$ at different lags k. The partial autocorrelation at lag k, denoted by ϕ_k, measures the correlation between r_t and r_{t-k} after removing the linear influence of all the intervening lags. The PACF plot helps in identifying the number of significant lags in an autoregressive model.