UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

DOCTORAL THESIS

Forecasting high-dimensional time series using Manifold and Machine Learning with Applications in Financial Markets and Epidemiology

Author:
Panagiotis PAPAIOANNOU

Supervisor:
Prof. Konstantinos Constantinos SIETTOS

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in

Applied Mathematics and Applications
in the

Dipartimento di Matematica e Applicazioni “Renato Caccioppoli"

March 3, 2022

Declaration of Authorship

I, Panagiotis PAPAIOANNOU, declare that this thesis titled, "Forecasting high-dimensional time series using Manifold and Machine Learning with Applications in Financial Markets and Epidemiology" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

"Dedicated to my wife Agapi and daughter Marion!"

Panagiotis Papaioannou

UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

Abstract

Dipartimento di Matematica e Applicazioni “Renato Caccioppoli"

Doctor of Philosophy

Forecasting high-dimensional time series using Manifold and Machine Learning with Applications in Financial Markets and Epidemiology

by Panagiotis PAPAIOANNOU

The aim of this Ph.D. Thesis is to exploit the arsenal of recent theoretical advances in manifold and machine learning thus developing a new computational framework, integrating also well established mathematical tools from the non-linear time-series analysis theory for the forecasting, i.e. out-of-sample extrapolation of high-dimensional time series. This is a three-fold task: (a) find the dimension of a low-dimensional manifold where the dynamics evolve and embed the high dimensional dataset on this manifold, (b) construct low-dimensional models on the embedded manifold using machine learning with a focus on Gaussian Processes, and (c) forecast, i.e. perform out-of-sample extrapolation in the ambient original space. This is a conceptually different task with respect to the interpolation problem of the model reduction of well-defined dynamical models in the form of Ordinary and or Partial Differential Equations. Applications include Financial and Electricity Markets modelling and forecasting, market inefficiencies identification, trading and portfolio management strategies deployment tested on a wide spectrum of financial assets, the task of challenging the Efficient Market Hypothesis on all of its efficiency forms, as well as the recent COVID-19 pandemic’s dynamics short-term forecasting challenge.

Acknowledgements

This Thesis would not be have been possible without the support, guidance, mentoring, mathematical rigorousness and academic excellence acting always as a great inspiration to me, hands-on contribution, as well as moral elevation open-heartedly offered to me by my supervisor Prof. Konstantinos Siettos. Furthermore, honored and sincerely grateful to Prof. Yannis Kevrekidis, for the opportunity to have been guided, mentored and further intuitively framed around the base concepts of this Thesis. For being my kickstart stepping stone and mentor both in life and academia, many thanks to my father Dr. G. Papaioannou. I like considering this work a continuation and mental “child" of his Thesis, while wishing to have inherited his neverending pursuit of the truth via the wonderful world of mathematics.

The highest level of gratitude and love goes to my wife Agapi, for always being there for me in all battles, both in life and academics, with her support and smile always morally boosting every effort towards progress and always prettifying every aspect of our family life. I would also like to thank her for bringing to life the most wonderful complexity of our life, our daughter and raising her in accordance to the strong principles obtained from our families, her father Theodoros, mother Maria and brother George, as well as my father George, mother Mersini and sister Anastasia.

Finally, i would like to thank my professional colleagues and friends, Mr. Nikolao Parliari, Konstantino Dimaresi and Thoma Dionysopoulo, for their extensive and open-hearted sharing of experience and expertise, all over the years, which acted as insight, intuition and problems’ setting cornerstones, around which this Thesis has also been framed. . . .

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Forecasting is something like the “holy grail" in a wide spectrum of areas of science and engineering, ranging from finance (e.g. forecasting of asset prices) and macro-economics (e.g. forecasting of inflation, GDP growth, unemployment) to geomechanics and earthquake forecasting and from ecology and environmental engineering (e.g. forecasting of climate change and natural catastrophes) to neuroscience (forecasting of epileptic seizures based on EEG signals) and epidemiology, to name just a few. Its origin can be traced back in the early civilizations, around 650 B.C. when the Babylonians were trying to forecast short-term weather changes by monitoring macro-observables such as clouds and halows *NASA* n.d. Around 340 BC, Aristotle wrote a book entitled Meteorologica that was considered to be the reference on weather forecast for almost 2000 years. During the Renaissance, scientists realised that in order to forecast weather at a local scale, out of the many observables, three macroscopic quantities were fundamental: humidity (1450, Nicholas Cusa), temperature (1592, Galileo Galilei) and pressure (1643, Evangelista Torricelli), i.e. that weather forecasting could be successful on a three-dimensional manifold. Currently, a significant part of weather forecasting, at least at a local scale and in the short term, is based on the numerical solution of physical models based on the Navier-Stokes equations occupied with the Coriolis force. However, due to the adjective nasty nonlinear terms, they are particularly difficult to solve and their theoretical understanding (about the existence and smoothness of the solutions) remains one of the Clay Mathematics Institute million dollar problems. Nonetheless, for many complex problems of contemporary interest, such macroscopic dynamical models in the form of Partial or Ordinary Differential Equations, simply do not exist. In such cases, forecasting is mainly data-driven.

1.1 Motivation

An inherent problem of using such modelling/forecasting approaches, when dealing with temporal measurements in a high-dimensional feature space, is the “curse of dimensionality”, Bellmann, 1957. More specifically, it is known that the number of samples necessary towards an arbitrary function’s approximation, within a given level of precision, increases in an exponential pace with respect to the input variables’ respective number, i.e. the inherent dimension of the target function. In such cases, training, i.e. the estimation of the values of the model parameters, requires a large number of snapshots, which increases exponentially with the dimension of the feature space. Thus, a fundamental task in modelling and forecasting is that of dimensionality reduction, or, the “mapping" (transformation) of the original space’s data, to low-dimensional spaces (e.g. low dimensional manifolds), possibly (ideally) of significantly lower dimension than the first, where informative properties of the initial data are retained, and potential data quality and processing related undesirabilities like sparsity and computational intractability.

1.2 State-of-the-art

1.2.1 Time-Series forecasting using the tools of Nonlinear dynamical systems and the concept of embedding

The fundamentals of chaotic nonlinear dynamical systems and chaotic time series analysis theory were the first to be exploited towards the development of tools and methodologies associated with this particular challenge. The associated literature review stands strong on a number of seminal papers in the nonlinear time series analysis and chaos theory. We refer, indicatively, to the works of Kantz, 1994,Schuster and Just, 2005,Strogatz, 2018 Grassberger and Procaccia, 1983,Bradley and Kantz, 2015 Casdagli, 1992,KK, Weigend, and Gershenfeld, 1994,Abarbanel and Gollub, 1996. We refer to Kantz, 1994, Bradley and Kantz, 2015, for an excellent introduction and overview of the main nonlinear time series analysis tools and concepts, as well. In the history of time series analysis, the deterministc chaos and self-similarity, two concepts of the nonlinear behavior found during ‘60s, constitute two seminal events that changed drastically the view the researchers had about complexity. The most pronounced “paradigm shift" was that determinism, compared to stochasticity is not a prerequisite to secure predictability, since in a plethora of nonlinear dynamical systems it is found that very close-by trajectories exhibit exponential divergence or sensitivity on initial conditions. This unpredictability is also manifested by the presence of irregular, non-periodic dynamic evolution of the measured signals, even without the presence of any stochastic component.
Linear time series analysis and modeling consider that the sequence of scalar measures in the time series is a combination of periodic and stochastic components, regarding also any irregularities as noise. This linear perspective is very restrictive, and in general inadequate, in capturing the nonlinear structure in a real-world high dimensional dynamical systems that exhibit complex and unpredictable behavior. Linear time series analysis is based on 1-D scalar measurements of variables of high-dimensional systems, that actually represent a projection of a higher dimensional time evolution on the one-dimensional set of real numbers. This is equivalent to a “compression" of the a high-dimensional information set to a set of much less dimensionality, resulting as a consequence in loss of valuable information of the underlying system under analysis. Also in a linear setting, the way of inferring the high-dimensional dynamics and structural details of the “true" state- or phase-space, by simply relying on 1-D scalar measurements, from an unknown system is not known.
The answer to this formidable problem was given by the publication of two seminal papers, the first by Packard et al., 1980 and the second by Takens, 1981, Sauer, Yorke, and Casdagli, 1991a, with the concept of embedding, being one of the most important advances in the field, upon which many other conceptually similar approaches in theoretical and applied mathematics (in areas like machine and manifold learning) have been based. Both papers constitute the corner-stones for the emergence of the Nonlinear time series field, and are viewed as the ‘boosting’ mechanisms that generated a “flood" of papers focusing on the detection of chaos in time series taken from experiments in laboratories, as well as real world systems. During the very “fruitful" decade of 80s, in the production of theoretical and practical methods regarding the extraction of geometry and strange

(dynamical) structures (attractors), the concepts self similarity and fractal dimensions,Hutchinson, 1994, Mandelbrot, 1985, Graf, 1987, strange attractors Shaw, 1981,Ruelle, 1981,Grassberger and Procaccia, 1983, chaos synchronization Pecora et al., 1997,Rosenblum, Pikovsky, and Kurths, 1996,Rulkov et al., 1995, and finally bifurcations Rand and Holmes, 1980,Gardini, Lupini, and Messia, 1989, Collet, Bragard, and Dauby, 2017 were also developed and strongly enrich and enhance the arsenal of the nonlinear time series tools. Also, in the 80s and 90s, another “boom" in the development of new powerful approaches occurred as Lyapunov exponents Kantz, 1994, information and Correlation dimension Grassberger and Procaccia, 1983, and reccurence plots, Eckmann, Kamphorst, and Ruelle, 1987. The nonlinear time series analysis and chaos measurement tools, include a variety of complexity measures like Maximal Lyapunov, λ_{max} and Hurst H exponents (estimated via e.g. Rescaled Range, Detrenended Fluctuation Analysis and Generalised Hurst Exponent methods), Entropy E, a measure of uncertainty and complexity in a dynamical system.

1.2.2 Manifold Learning/Embedding

Refocusing on the fact that all of the previously outlined nonlinear tools and theoretical approaches - with particular focus on the embedding concept related ones - are closely connected to the main challenge of "escaping the curse of dimensionality", we observe that out of the many features that can be measured in time, one has first to identify (or “learn") the intrinsic dimension of the possible low-dimensional subspace (manifold) and the corresponding vectors that span it (“the coordinate system"), which actually govern the emergent/ coarse observed dynamics of the system under study. Assuming that such coarse dynamics exist and evolve on a smooth enough low-dimensional manifold, an arsenal of linear and nonlinear manifold learning algorithms can be exploited towards this direction. The most popular, well documented and investigated both in theoretical level but also regarding its super-intuitive essence when applied on numerous real-life applications, is the Principal Component Analysis (PCA), appropriate for linear latent factor modelling. Initially introduced by Pearson, 1901 as the "principal axis theorem's" direct analogue in mechanics, (Strang, 1988), its well known name and main theoretical setup was given by Hotelling, 1933. Since then, it further progressed through theoretical but mostly applied research on a wide spectrum of fields, like signal processing, meteorology, structural dynamics and many others, while also depending on each corresponding applied area, small modifications, scope adjustments and also its naming was also changed to e.g. Singular value Decomposition (SVD), Stewart, 1993, Factor Analysis, Eigenvalue Decomposition (EVD), Empirical Eigenfunction Decomposition (EED) and more, (see Jollife and Cadima, 2016 for a detailed view on the method).

Despite its wide application and popularity, PCA, being a linear embedding method, suffers from limitations regarding nonlinear data structures. Extensions targeting improving results in certain real-life applications where nonlinearities are expected to be hidden in the data structures, have been developed over the years and are usually first tested in artificial nonlinear "toy-models" like the "Swiss Roll", (Paul and Chalup, 2017). Such an extension is the PCA direct "upgrade", namely the kernel-PCA, Schölkopf, Smola, and Müller, 1998, which was developed based on the mathematical framework of “kernel methods" (usually referred to as the “kernel trick"), and reproducing kernel Hilbert Spaces (Aronszajn, 1950). The latter inheriting the property of “closeness in norm" in their member functions, (stating that if any two functions f, g in a reproducing kernel Hilbert Space are close to each other in the sense that if $\|f - g\|$ is small, then the distance $\|f(x) - g(x)\|, \forall x$ in the origin set, is also close), offer the mathematically solid and intuitively understandable framework for the several nonlinear manifold learning techniques to be safely grounded. This is in the sense of a connection between two distinct spaces of different dimensions, while maintaining specific properties during the transition from one to the other. The identification and definition of the appropriate set of these properties is the primary goal of the several manifold learning techniques. The reproducing kernel theoretical framework introduced by Stanislaw Zaremba in 1907 (Szarski, 2017) was also solidly grounded by the work of James Mercer in the theory of integral equations and their connection with the reproducing property, Mercer, 1909, and later further developed by Nachman Aronszajn and Stefan Bergman in a more systematic manner, in the 1950 (Aizerman, Braverman, and Rozonoer, 1964). The idea of this “kernel trick" is that hidden structure might not be able to be expressed in data terms (via inferring dependencies between the data themselves), but on “feature mapped" expressions of them. It acts as pre-processing operation (or "filter" - usually applied in “denoising" targeted applications, Chen and Fomel, 2015) targeting on “clustering" the input points into arbitrary hyperplanes. In the case of kernel-PCA, this operation is being performed in terms of the inner product of the input points (Schölkopf, Smola, and Müller, 1998). Several other approaches have been conceptualised and formalised towards the task of the proper kernel formation, most usually specifically designed to meet certain applications' associated tasks. Among others, methods associated with the "kernel trick" are k-nn clustering, Support Vector Machines (SVMs), kernel perceptrons and Gaussian processes, Rasmussen and Williams, 2018.

Topology and Geometry

The concepts of “coordinate system", “clustering" and “Distance" are mathematically associated with the terminologies “manifold learning" and “metric learning" of geometric and topological data analysis. More specifically, the first is targeting the identification of the proper global coordinate system, which optimally preserves local structures, while the second is inducing a proper metric accounting for the similarity between the several individual input points. Both concepts are based on topology and specifically on the definition of differentiable manifolds, i.e. topological spaces with the property that the surrounding neighborhoods of each point, locally resemble Euclidean space. Based on that, the use of calculus on higher dimensional structures, that might in fact be globally non-Euclidean, is enabled. The study of “invariants" under distortions present in the manifold (loops, holes etc), is the target of topological methods, while the bridge between the geometrical perspective and topology's theoretical overlays, (as is the case of the several manifold learning and statistical ranking related techniques), is based on Hodge Theory (Cavalcanti, 2020). Due to its connection to all these mathematical fields, manifold learning extensions were also explored outside vector spaces (real analysis), investigating dominant modes of variation of functional data (functional analysis), as well. For example the functional PCA (FPCA), Viviani, Grön, and Spitzer, 2005, Yang, Yang, and Shang, 2021, Gao and Shang, 2017 is based not on vectors, but on curves.

Nonlinear Manifold Learning Algorithms

In linear manifold learning approaches like the previously discussed PCA, both local and global structures are assumed to be linear, and are modelled under the prism of Euclidean distancing (equivalence between the underlying quadratic form of the PCA optimization problem and the euclidean distance of a point x, as $\boldsymbol{x}^T\boldsymbol{x} = \|\boldsymbol{x}\|^2$). Other Euclidean distancing based approaches include Multidimensional Scaling (MDS), where instead of using as inputs the original space points (as in the PCA), the pairwise distance matrix is given Mead, 1992, instead. As such the main assumption behind its logic is that global structure is preserved, if pairwise distances are preserved, as well. Such linear assumptions regarding the underlying “true" structure of the manifold though, while might be valid “locally", might not be the correct ones, on a global scale. Towards revealing these global nonlinear structures, methods included in the broader category of the so called “kernel eigenmap methods", such as Local Tangent Space Allignment LTSA (Zhang, Ma, and Tan, 2018), Local Linear Embedding (LLE) and its variants, such as Modified LLE (MLLE), or the Hessian LLE (HLLE), (Yao et al., 2017, Prabhakar and Rajaguru, 2017, Roweis and Saul, 2000, Donoho and Grimes, 2003), T-SNE Maaten and Hinton, 2008, Diffusion Maps (DMs), Coifman et al., 2005; Coifman and Lafon, 2006a; Coifman and Lafon, 2006b; Nadler et al., 2006; Coifman et al., 2008, Dsilva et al., 2015, Dsilva et al., 2018, others related to Dynamical Systems and time series modelling, such as works of Lin et al., 2006, Talmon et al., 2015 and Park, Sriram, and Yin, 2010, or the Koopman operator, Mezić, 2013; Williams, Kevrekidis, and Rowley, 2015; Dietrich, Thiem, and Kevrekidis, 2020, and Dynamic Mode Decomposition (DMD) by Proctor, Brunton, and Kutz, 2016; Schmid et al., 2011; Kutz, Fu, and Brunton, 2016; Mann and Kutz, 2016, were developed and proposed over the years.

Indicatively, we outline Local Linear Embedding (LLE), which tries to “bridge" local linear fitting (or “stitching"), with nonlinear global structure recovering. Using the k-nn algorithm, as the first step of its internal mechanism, it represents the several input points as linear combinations of their "nearest neighbors", while also forming a corresponding "distance matrix" based on an induced metric, the Euclidean one. Other methods working towards the same direction, like ISOMAP, Tenenbaum, De Silva, and Langford, 2000; Balasubramanian et al., 2002, extended this framework, by inducing geodesic distances instead of euclidean, while also incorporating the concepts of “Isometry", “shortest-path" discovery algorithms (like e.g. Dijkstra’s (Wahyuningsih and Syahreza, 2018,K. Rohila, P. Gouthami, and M, 2014, Dcsa and Chhillar, 2014) and “graph geodesics" from graph theory. Another graph theory based framework, is the Laplacian Eigenmaps/Spectral Embedding, Belkin and Niyogi, 2003, which uses the Laplacian Matrix (or graph/discrete/Kirchhoff Laplacian), acting as the direct analogue of the Laplacian operator, to represent the associated graph, in a matrix format. This reminds us of finite difference methods, used to numerically solve differential equations, by utilizing the discrete analogues (approximations) of operators defined on the continuous domain, (Smola and Kondor, 2003). The connection between the theory of Partial/Ordinary Differential Equations and differentiable manifolds (Differential topology), is well documented and extensively formalised. Several tools have been proposed for solving a wide spectrum of PDEs and ODEs related problems over the years, highlighting their dependence on linear operators defined on differentiable manifolds, inherenting all related conditions and limitations, (n.d.(b); Raymond O.Wells, 2008). Such an operator is the generalisation of the previously used Laplace operator, namely the Laplace-Beltrami operator. It is a generalization to functions that are defined on Riemannian (and pseudo-Riemannian) manifolds (sub-manifolds of the Euclidean space), Colding, 2020; Boothby, 2003; Willmore, 1987; Craioveanu, Puta, and Rassias, 2001. Diffusion Maps methodology is associated with the approximation of this operator on a manifold. More specifically, it considers Markov matrices' eigenvectors, as appropriate coordinate systems for representing the target datasets. Therefore, data originally modelled as a graph, can be represented as "cloud of points" in the Euclidean space. The associated (diffusion distance) kernel matrix (Coifman and Lafon, 2006a), is designed to reproduce the diffusion introduced by a Fokker-Planck equation of statistical mechanics (Medved, Davis, and Vasquez, 2020), describing the time evolution of the probability density function of transition between one point and another in the available dataset (Markov matrix), further connecting the method with concepts like Brownian Motion, random walks, stochastic integration etc from stochastic calculus and stochastic differential equations. Due to their ability to handle "noisy" data, its applications' spectrum includes among others, face, image and voice recognition, like Barkan and Aronowitz, 2013, Barkan et al., 2013, Sidi et al., 2011, Farbman, Fattal, and Lischinski, 2010, but also multivatiate time series analysis related ones, such as Lian et al., 2015. Staying on the time series analysis space, we highlight Dynamic Mode Decomposition, a dimensionality reduction algorithm combining (among others) the theory of dynamical systems and time series modelling, with manifold learning. As the rest of the previously mentioned methods, it is acting as an approximation of the Koopman (or composition) operator (Koopman, 1931, Nathan Kutz, Proctor, and Brunton, 2018). The operators approximation numerical aspect of all the aforementioned methodologies is usually associated with some form of optimization. As such, progress on the field is also associated with corresponding progress on the field of numerical analysis methods, for convex optimization and eigenproblems solving, (De Bie, Cristianini, and Rosipal, 2005).

1.2.3 Time Series Modelling, Physical Meaning & the Pre-image Problem

In a time series modelling setting, one of the main targets of all the previously outlined embedding approaches is to be able to use any of the time series modelling and forecasting models developed throughout the years, towards explaining future dynamics. To name a few, "mainstream" linear tools as the exponential smoothing and moving average models, conditional mean and volatility measuring models like the ARIMA/(E)GARCH ones, Brockwell et al., 2016, stochastic processes from stochastic calculus, as the Ornstein - Uhlenbeck (OU), Shreve, Chalasani, and Jha, 1997, (Fractional) Brownian Motion (Xu and Luo, 2018) mixes of Stochastic Calculus and nonlinear dynamics (Yannacopoulos et al., 2000; Yannacopoulos, 2008; Yannacopoulos, Frangos, and Karatzas, 2011; Yannacopoulos, 2011), and of course, Bayesian non parametric models including Gaussian Processes Rasmussen, 2003 and other machine learning schemes, such as Forward and Recurrent Neural Networks, Deep Learning, LSTMs, Reinforcement Learning, Reservoir Computing, Fuzzy Systems etc, Granger, 1969; De Gooijer and Hyndman, 2006; Lukoševičius and Jaeger, 2009; Brockwell et al., 2016; Greff et al., 2016; Pathak et al., 2018; Vlachas et al., 2020; Wyffels and Schrauwen, 2010 are utilised towards the direction of creating predictions in the "smoother" lower-dimensional embedded space, rather than in the "noisier" original one. Such an approach though, might incur difficulties on the interpretation of these predictions, or "practicality" issues, if the embedding is performed using nonlinear methodologies. Physical meaning or intuition regarding the embedded space's structure when this is expressed under nonlinear and sometimes "black-box" approaches, is not feasible, and as such even efficient forecasts cannot be easily inferring informative value towards real-life applications. Furthermore, such lack of intuition may be met under both a temporal and/or spacial aspect. More specifically, addressing the first one, consider the case where measurements corresponding to different time instances, are being "clustered" together under a nonlinear "similarity" criterion (induced distance or metric): e.g. temperatures being within the range $25 - 32^oC$ are all clustered together as being "normal", or returns of a stock being lower than 1% are clustered together as "insignificant". The spacial aspect intuitively corresponds to the inferred connection of the term "distance" and "correlation" from statistics, with examples of clustering being e.g. grouping together all stocks which exhibit correlation with each other of greater than 80% or lower than 20% etc. The potential non-practicality of a nonlinear method in this case, would be the replacement of the intuitive covariance-terms-related correlation, of the PCA analogue, with a nonlinear "similarity" measuring counterpart (e.g. nonlinear kernel).

The so called "pre-image problem" is the one associated with this limitation, framing the goal of reconstructing (or "lifting") the observation embedded in low-dimensional manifold to the ambient high-dimensional domain, recovering in a sense, "practicality" and physical meaning in the analysis. Much literature around the topic has been produced throughout the years, framing the problem foundations Aizerman, Braverman, and Rozonoer, 1964, Schölkopf, Smola, and Müller, 1998 and later connecting it to the "kernel trick", as well (Mika et al., 1999a, Mika et al., 1999b, Richard, Bermudez, and Honeine, 2009, Honeine and Richard, 2011). New ideas about the "lifting" operation were introduced from the multidimensional interpolation space, as well, with popular proposed approaches being the Radial Basis Functions, Buhmann and Levesley, 2004, Majdisova and Skala, 2017, Leon-Delgado et al., 2018, Que and Belkin, 2020, and also from Harmonics Analysis, with the Geometric Harmonics approach, Lafon, 2004, Coifman and Lafon, 2006c, Bermanis, Averbuch, and Coifman, 2013, Heimowitz and Eldar, 2018, based on the theoretical framework of Nystrom Extension, Nyström, 1929, having attracted much of attention.

Other Related Works/ Applications

The overall modelling flow has been followed by several research works over the years. For the two-fold task of dimensionality reduction and modelling in the low-dimensional subspace, Chiavazzo et al. Chiavazzo et al., 2014 constructed reduced kinetics models, by extracting the slow dynamics on a manifold globally parametrized by a truncated DM. A comparison of the reconstructed and the original high-dimensional dynamics was also reported. The reconstruction was achieved using Radial Basis Functions (RBFs) interpolation and Geometric Harmonics (GHs). Liu et al.Liu et al., 2014 used DMs to identify coarse variables that govern the emergent dynamics of a particle-based model of animal swarming and, based on these, they constructed a reduced stochastic differential equation. Dsilva et al. Dsilva et al., 2015 used DMs with a Mahalanobis distance as a metric to parametrize the slow manifold of multiscale dynamical systems. Williams et al. Williams, Kevrekidis, and Rowley, 2015 addressed an extension of DMD to compute approximations of the Koopman eigenvalues, thus showing that for large data sets the procedure converges to the numerical approximation one would obtain from a Galerkin method. The performance of the method was tested via the unforced Duffing equation and a stochastic differential equation with a double-well potential that converges to a Fokker-Planck equation. Brunton et al. Brunton, Proctor, and Kutz, 2016 used SVD to embed the dynamics of nonlinear systems in a linear manifold spanned by a few SVD modes, and constructed state-space linear models on the manifold to approximate the full dynamics. The so-called Sparse Identification of the Nonlinear Dynamics (SINDy) method was demonstrated using the Lorenz equations and the 2D Navier–Stokes equations. Bhattacharjee and Matouš, Bhattacharjee and Matouš, 2016 used ISOMAP to construct reduced-order models of heterogeneous hyperelastic materials in the 3D space, whose solutions are obtained by finite elements, while for the construction of the inverse map used RBF interpolation. Wand and Sapsis Wan and Sapsis, 2017 proposed a methodology for forecasting and quantifying uncertainty in a reduced-dimensionality space, based on Gaussian process regression. The efficiency of the proposed approach was validated using data series from the Lorenz 96 model, the Kuramoto-Sivashinsky, as well as a barotropic climate model in the form of a PDE. Kutz et al. Nathan Kutz, Proctor, and Brunton, 2018 proposed a scheme based on the Koopman-operator theory to embed spatio-temporal dynamics of PDEs with the aid of DMD, for approximating the evolution of the high-dimensional dynamics on the reduced manifold; the reconstruction of the states of the original system was achieved by a linear transformation. Chen and Ferguson Chen and Ferguson, 2018 have proposed a molecular enhanced sampling method based on the concept of autoencoders Kramer, 1991 to discover the low-dimensional projection of Molecular Dynamics and then reconstruct back the atomic coordinates. Vlachas et al. Vlachas et al., 2018 proposed an LSTM-based method to predict the high-dimensional dynamics from the information acquired in a low-dimensional space. The embedding space is constructed based on the Takens embedding theorem Takens, 1981. The method is demonstrated through time-series obtained from the Lorenz 96 equations, the Kuramoto–Sivashinsky equation and a barotropic climate model as in Wan and Sapsis, 2017. Mcdonald, Shalizi, and Schervish, 2017, introduce the concept of mixing time series in order to establish nonparametric bounds for time series forecasts. However in the analysis and modeling of nonlinear time series coming from nonlinear dynamical systems (deterministic or stochastic), the identification process has not yet reached the status of the linear case and also is a very complicated work. The scenario of non-mixing time series, is much related to the forecasting of financial time series, via the Martingale condition that must be satisfied in such time series. Max Simchowitz, 2018, introduced a learning without mixing approach the martingale small ball condition and provided evidence for the existence of upper bounds on the least squares estimator, a valid result also for a linear nonstationary (unit root) model, the autoregressive model AR. Lee et al., 2020 used DMs and Automatic Relevance Determination to construct embedded PDEs by the aid of Artificial Neural Networks (ANNs) and Gaussian Processes. The methodology was illustrated with Lattice-Boltzman simulations of the 1D Fitzhugh-Nagumo model. Koronaki, Nikas, and Boudouvis, 2020 used DMs to embed the dynamics of a one-dimensional tubular reactor as modelled by a system of two PDEs on a 2D manifold, and then constructed a feedforward ANN to learn the dynamics on the 2D manifold. The efficiency of the scheme was compared with the original high-dimensional PDE dynamics by lifting the reduced dynamics back to the original space using GHs and RBFs. Recently, Lin and Lu, 2021 used the Koopman and the Mori-Zwanzig projection operator formalism to construct reduced-order models for chaotic and randomly forced dynamical systems, and then based on the Wiener projection they derived Nonlinear Auto-Regressive Moving Average with exogenous input models. The approach was demonstrated through the Kuramoto-Sivashinsky PDE and the viscous Burgers equation with stochastic forcing.

1.3 Aims, Objectives and Contribution of the Thesis

The underlying dynamics governing the evolution of data, change during the course of time, due to a number of internal and external factors of the system generating the time series (e.g. macroeconomic conditions, regulatory and government intervention etc., in case of financial market data), which also have affected the structure of the dynamical system (market). We consider that the research outcome of the Thesis is useful for both theoretical and practical reasons. First, because the “extraction” of the degree of complexity of an “unknown" system, from the complexity of a model capturing the dynamics and volatility of data generated by the system, introduces an innovative way of thinking in quantifying system’s organization or architectural complexity. For this purpose we applied a number of nonlinear tools from stochastic and chaos theory of time series to investigate empirically the impact of the internal mechanisms of dynamical systems on the time evolution of the volatility of related time series. Combining typical or “mainstream" tools like Hurst exponent (estimated by different methods), maximum Lyapunov exponent, Correlation dimension etc with Manifold learning approaches, can be extremely useful and effective towards understanding better the dynamics of the structural complexity of dynamical systems, via analyzing time series of variables of the system. Thus we can capture more reliably the endogenous complexities of the systems due to the interdependency of a plethora of co-evolving factors of both deterministic and stochastic type, affecting the dynamic evolution of the system. Therefore, analyzing time series taken from such complex system, requires the application of very sophisticated tools, such as the ‘hybrid’ ones described in this thesis, built via combining well known tools from both linear and nonlinear time series with innovative approaches from the field of manifold learning and dimensionality reduction. For example, using the Lyapunov maximum Exponent λ_{max} to measure the stability of a dynamic behavior of a variable reflected by a financial time series, we can show whether there is a stable or an unstable (chaotic) dynamics in the used by most of the related applications like trading or portfolio management, log-returns, depending on its sign, for the period of time we are interested. We can combine this information gained by estimating the Lyapunov maximum Exponent, with results produced via using Hurst exponent estimation, hoping to receive consistent results, thus enhancing greatly the reliability of our findings. However, in case the system under examination is high-dimensional (as the financial systems analyzed in this thesis), before ‘embarking’ in any time series analysis using tools as in the examples mentioned before, it is necessary to reduce the dimensionality of such systems, via using manifold learning techniques.

It is therefore of paramount importance, to consider all the typical and main stream linear and nonlinear approaches-modes, under the prism of manifold learning, focusing strongly on how these two approaches complement each other and how both enhance our arsenal of tools in analyzing very complex systems. This consideration or view, will make more pronounced the contribution of this thesis in the field of time series forecasting, by providing strong evidence that the above mentioned complementarity indeed enhances significantly our efficiency and reliability in forecasting nonlinear time series extracted by measuring representative variables of their underlying complex dynamical system. The main objective in such empirical research works to shed light on the interaction or possible "causal" relations among a variety of quantitative measures usually applied to the target modelled time series data, like the stability of the underlying "system" through e.g. maximum Lyapunov exponent (λ_{max}), volatility (σ, conditional and unconditional), the complexity measures of Hurst exponent, modern concept of entropy, as well as how all these could be related, enhanced and combined with the recent developments in machine and manifold learning, with the latter being the main topic of interest of this thesis. Analyzing, modeling and forecasting time series of variables taken from the underlying systems of several applications, will be outlined and thoroughly investigated in the context of the thesis. Moving to applications, for example Financial and Energy markets being the two main fields of focus, generate complex nonlinear-deterministic (chaotic?) behavior, as well as pronounced stochastic behavior in their price volatility, as it is evidenced due to the difficulty in out-of-sample accurate forecasting. This is equivalent to say that the dynamic evolution of nonlinear high-dimensional (or multi-parametric) models are very sensitive to initial conditions, a well-known characteristic of chaotic systems,Papaioannou et al., 2016; Papaioannou and Karytinos, 1995,Kantz, 1994.

It is now well demonstrated, by using the tools of nonlinear dynamics that seemingly random or irregular behavior in natural systems may arise from purely deterministic dynamics with unstable paths or trajectories. Even in appeared randomly distributed data an underlying order of pattern may exist. Dynamical systems that exhibit also sensitivity to initial conditions (IC) (even tiny change in IC can have tremendous effect on the future evolution of dynamics), are known as chaotic systems. Therefore, if the electricity market as a system also processes a chaotic trait, manifested e.g. by a chaotic behavior of its wholesale price, generic stochastic models based on the theory of probability and statistics are not the appropriate and accurate ones to capture this chaotic feature of the market. As it is demonstrated by the references given above, chaos theory can reveal the intrinsic regularities of the electricity price or load, obscured by "noise" or randomness.

The weakness and inefficacy of traditional linear and nonlinear time series models to make a consistent forecast may strongly affect society in a plethora of situations (e.g. growth corporations, country development, stock market crash, natural disaster etc). The 'curse of dimensionality' problem (time series models become fast overparameterized as the dimension grows) is a formidable reason for the difficulty in forecasting of high dimensional time series. The over-parameterization problem, the stationarity assumption and other constraints (see below), have rendered traditional parametric models like Autoregressive Integrated Moving Average (ARIMA) or Vector Autoregressive (VAR) models, very weak in capturing the, mainly, rich nonlinear dynamics contained in time series.

1.3.1 Contribution of the Thesis

The performance of most of the above mentioned approaches in other works has been mainly assessed in terms of interpolation and ergodicity, namely, the data are produced using dynamical models (ODEs, PDEs, SDEs or agent-based models) that were simulated in some specific interval of the high-dimensional domain and their properties do not change over time, then reduced order models were constructed and trained based on the generated data, and finally a reconstruction of the high-dimensional dynamics was implemented within the original interval of the high-dimensional domain.

Here, we address a new three-tier computational scheme to perform forecasting in the high-dimensional space, by Panagiotis Papaioannou, 2021:

(i) reducing the high-dimensional time series into a low-dimensional manifold using manifold learning;

(ii) constructing and training reduced-order models on the manifold, and, based on these, making predictions;

(iii) lifting the predictions in the high-dimensional space.

The performance of this "embed-forecast-lift" scheme is assessed through four sets of time series: three of them are synthetic time series resembling EEG recordings generated by linear and nonlinear discrete stochastic models with different model orders, while the fourth set is a real-world data set containing the time series of 10 key pairs of foreign exchange prices retrieved from the `www.investing.com` open API, spanning the period 03/09/2001-29/10/2020.
This is the first work that addresses the problem of forecasting based on manifold learning, providing a comparison of various embedding, modelling and reconstruction approaches, and assessing their performance on both synthetic time series and a real-world FOREX data set.
The task of forecasting that is attempted here is different from that attempted for the reconstruction of high-dimensional models of dynamical systems in the form of ODEs or PDEs proposed in the previous mentioned works, in four main aspects. First, for real-world data sets, the existence of a relatively smooth low-dimensional manifold is not guaranteed as in the case of well-defined dynamical models. Second, non-stationary dynamics, which in general are not an issue when dealing with the approximation of dynamical systems, pose a major challenge for a reliable/consistent forecasting. It should be noted that the stationarity assumption may be required to hold true even for interpolation problems. For example, the implementation of the MVAR models and Gaussian Process with a Gaussian kernel requires that the stationary covariance function assumption is satisfied (see e.g. the discussion in Rasmussen, 2003; Cheng et al., 2015). Third, when dealing with real-world time series, such as financial time series, the number of available snapshots (even at the intraday frequency of trading) are limited, in contrast to the size of temporal data that one can produce by model simulations. In such cases, the quest for beating the "curse of dimensionality" using the correct (parsimonious) embedding and modelling is stronger. Finally, in the case of smooth dynamical systems, as the dynamics emerges from the same physical laws as expressed by the underlying ODEs, PDEs or SDEs, what is usually sought is a single global manifold (a single geometry). However, in many complex problems such as those in Finance, the parametrization of the manifold (if it exists) may change over time. Thus, one would now seek for a family of (sequential in time) sub-manifolds.
As already mentioned, the related to the topic literature developed throughout the years, is really extensive, on both a theoretical and on an applied level. However, there is always room for the several theoretical fields to be revisited in terms of intuition, as well as to be challenged in terms of empirical evidence validity. As such, several aspects of this Thesis's contribution could be identified under the empirical research perspective. On the mathematical basis framed in the previous sections, the Thesis contributes on the extension of previous research efforts regarding

the efficient design and construction of “Embed-Train-Predict-Lift" approaches for out-of-sample forecasting purposes, further validating their efficiency and opening paths for further research on the fields. It computationally validates the inferred and previously outlined connection between the fields of nonlinear dynamical systems, machine/manifold learning and stochastic calculus, while offering further insight on how tools from one field can be beneficial to the other, in terms of “educated" model setting and parameterisation.

1.3.2 Applications

Furthermore, the Thesis contributes to the field of forecasting of financial and epidemiological time series, fields with a wide and important economic, social and health interest.

Financial Time Series

On a high-level financial theory related basis, as outlined in the following subsection 1.4, the several applications of this Thesis are focusing primarily on the modelling and forecasting of Financial Markets. The Thesis tackles one of the building blocks of financial analysis, namely the Efficient Market Hypothesis (EMH), (Fama, 1970, Malkiel and Fama, 1970), challenging all of its forms to a certain degree, while discussing debatable aspects of the obtained results. The weak form of efficiency is the first to be directly questioned, with the several methodologies displaying exploitable hidden structures in the historical data. Extending the challenge of the EMH to its semi-strong and even strong forms of efficiency, the quantification and empirical evidence report against the statements claiming that any "Public" or "Private" information regarding the markets are not sufficient towards generating excess profits by "informed" investors.

The specific applications are outlined below.

Electricity Markets modelling

In the case of the electricity markets, the important challenges of load (demand) and price forecasting, as thoroughly investigated in Papaioannou et al., 2016, Papaioannou et al., 2018 and application’s related section 3.2, is tackled using our proposed hybrid approach, which to the best of our knowledge is one of the first to combine the particular traditional forecasting approaches with the manifold learning based one, while even more specifically applied in the target Electricity Markets of Greece and Italy. Among others, the reported results related to the detection of complexity in the target electricity markets, the identification and analysis of external factors affecting the underlying load and price dynamics, as well as the more challenging task of price forecasting, are the main contributions of this work. The promising results of the last task also tackles all three forms of efficiencies of EMH, while discussing why the markets are considered to be inefficient both under a purely quantification but also market microstructure analysis perspective.

Foreign Exchange Rate (FOREX) Markets modelling

The most important contribution here, is how the proposed “Embed-train-lift" hybrid approach manages to overcome these physical meaning related pitfalls, extending the set of available manifold learning methodologies suitable for such applications. Focusing particularly on the so called "pre-image problem" and the proper use of multidimensional interpolation methods, several promising results are obtained. To the best of our knowledge such approach has not again been proposed by similar works in the area of trading and constructing financial portfolios with manifold learning. The results show outperformance of the hybrid methodology compared to other mainstream portfolio management techniques, while also displaying how a combination between all approaches can be efficiently structured to generate even greater results and also open the path for further research on the topic.

Another contribution of the Thesis regarding the applications in finance, is the further validation of the findings of Cuturi and D'Aspremont, 2016, regarding the "tradability" of specific PCA embedded portfolios, in a wide set of FOREX markets, as well. Additionally to their work, this application sets the concepts of identifying "Mean-Reverting" and "Trending" trading portfolios, not only under the prism of linear AR models' application, but also under the equivalence of these models with several classes of processes from stochastic calculus, namely the Fractional Brownian Motions (fbm) and Matern Processes. Based on the connection between these classes of processes, the work contributes, on how someone could use measures like the Hurst exponent of fbm and the related α parameter of the Matern processes, to properly form the necessary "belief" behind the selection and form of the kernel used in the Gaussian Process Regression framework (GPR). Finally, although the results are really promising even after transaction costs are incorporated (market frictions), destructing once more the weak form of EMH (and under an even more realistic perspective), the work also displays the limitations of the several linear approaches used, while discussing the need for incorporating nonlinear manifold learning techniques and the pitfalls associated with their outputs' practical (physical) meaning, especially in applications like trading and portfolio management.

Epidemiological Time Series

Recently, the large amount of data on COVID-19 that have been produced and made publicly available has enabled the development of numerous machine learning algorithms, thus resulting in numerous studies Bachtiger, Peters, and Walsh, 2020 mainly for clinical prediction purposes Wang et al., 2020; An et al., 2020; Yan L., 2020; Yadaw et al., 2020; Gao et al., 2020, but also the estimation of the parameters of compartmental SEIR-like models Deng, 2020 and the forecasting of the disease's dynamics as stand-alone models Rustam et al., 2020; Chimmula and Zhang, 2020; Zeroual et al., 2020). However, all of these studies were restricted to very short-time forecasts. In particular, Rustam et al., 2020 used four techniques of machine learning, namely linear regression, LASSO, Support Vector Machine (SVM) and Exponential Smoothing to provide a 10-day forecast of the confirmed cases. Chimmula and Zhang, 2020 used a Long short-term memory (LSTM) network to provide an up to 14-days forecast for Canada with Italy and USA based on the available data until March 31, 2020. Zeroual et al., 2020 used five methods, namely, Recurrent Neural Network (RNN), Bidirectional LSTM (BiLSTM), Gated recurrent units (GRUs) and Variational AutoEncoder (VAE) algorithms to provide a 17-days forecast for the evolution of the pandemic in Italy, Spain, France, China, and the USA based on the available data until June 1, 2020. Katris, 2021 used five methods, namely Exponential Smoothing models, Auto Regressive Integrated Moving Average (ARIMA), Multivariate Adaptive Regression

Splines (MARS) and Feed-Forward Artificial Neural Networks (ANN) to provide a 7-day forecast for Greece based on the reported cases until August 14, 2020.

Here, we use three machine learning schemes, namely, shallow FNNs, Gaussian Processes and LSTMs to forecast over short-time intervals the evolution of the pandemic in all regions of Italy on the basis of publicly available data of the confirmed cases and regional mobility reports. The Gaussian Process is fed with delayed time series of the infected cases from all regions of Italy to forecast the number of infected cases in each region. The training, validation and test of the models is performed on the basis of available data of the confirmed infected cases. The prediction models and the corresponding yielded multi-step ahead (daily) predictions, act as the main contribution of this work, while also the overall modelling flow and output, for the several Italian Regions under focus is also the main component of the custom COVID-19 Monitor web-application "#combatcovid19uninafisr", 2021, where useful real-time and daily updated statistics and predictions for the pandemic's evolution are displayed.

1.4 Thesis Overview

The rest of the Thesis is structured in two main parts, the “Methodology", where the mathematical background and rigorous formulation of the several forecasting, embedding and lifting modelling concepts are developed and the "Applications", in which the several applications' motivations, specifications, used tools and the results obtained from the applied hybrid schemes are outlined.

More specifically, in Section 2.1 we setup the main mathematical theoretical framework behind forecasting with regression models. We present the traditional Linear Gaussian Models, members of which are the classical ARIMA models, while also detailing the mechanics of their upgrades and multidimensional extensions, i.e. the SARIMAX, Exponential Smoothing (ES) and Multivariate Autoregressive Models, respectively. We then move into the Machine Learning space, outlining the mathematical concepts and mechanics of Artificial Neural Networks (ANN), Support Vector Machines (SVM), Recurrent Neural Networks (RNN) and also Gaussian Process Regression (GPR) models from the Bayesian space. Much focus is concentrated on the last one, with deeper mathematical rigorousness and theory outlined in the relevant subsection.

In section 2.2, we discuss the parts comprising the Nonlinear Time Series Analysis space. The concepts of state space reconstruction, (optimum) embedding dimension, method of delays (MOD), (Generalised) Hurst Exponent and Detrended Fluctuation, as well as Lyapunov Exponents and Tsallis Entropy are thoroughly investigated. The connection of all these complexity measures, acting as the "parents" of manifold learning (as discussed in the motivation subsection earlier), lead us on the following section, 2.3.2, where the fundamentals, as well as the several selected linear and nonlinear manifold learning techniques are detailed. In section 2.4 the Pre-Image problem is formulated, while the two used approaches, namely the Radial Basis Functions Chiavazzo et al., 2014 and Geometric Harmonics Coifman and Lafon, 2006b, are mathematically formalised.

Moving to the applications' part of the Thesis, in Papaioannou et al., 2016, we propose a new hybrid model, a combination of the manifold learning Principal Components (PC) technique and the traditional multiple regression (PC-regression), for short and medium term forecasting of daily, aggregated, day-ahead, electricity system-wide load in the Greek Electricity Market for the period 2004–2014. PC-regression is shown to effectively capture the intraday, intraweek and annual patterns of load. We compare our model with a number of classical statistical approaches (Holt-Winters exponential smoothing of its generalizations Error-Trend-Seasonal, ETS models, the Seasonal Autoregressive Moving Average with exogenous variables, Seasonal Autoregressive Integrated Moving Average with eXogenous (SARIMAX) model as well as with the more sophisticated artificial intelligence models, Artificial Neural Networks (ANN) and Support Vector Machines (SVM). Using a number of criteria for measuring the quality of the generated in-and out-of-sample forecasts, we have concluded that the forecasts of our hybrid model outperforms the ones generated by the other models, with the SARMAX model being the next best performing approach, giving comparable results. Our approach contributes to studies aimed at providing more accurate and reliable load forecasting, prerequisites for an efficient management of modern power systems. The application in Papaioannou et al., 2018, aims to detect the impact of fundamentals and regulatory reforms on the Greek Wholesale Electricity Market, applying SARMAX and GARCH models and using data from January 2004 to December of 2014. The System Marginal Price (SMP) is considered as the footprint of an underlying stochastic and nonlinear process, reflecting not only the effects of endogenous/fundamental market factors but also the effects of exogenous variables as well as regulatory reforms, which also affect the market dynamics. To capture the dynamics of the conditional mean and volatility of SMP, a number of SARMAX/GARCH models have been estimated using as regressors an extensive set of fundamental factors in the Greek Electricity Market (GEM), as well as dummy variables that mimic the history of Regulator's interventions. The best found model captures adequately the dependency of the spot price to the regulatory reforms that have imposed on the GEM. The findings show the sign and the magnitude of the effect of fundamentals and of the reforms in the dynamic evolution of SMP. The dynamics of the conditional volatility generated by the model can be useful to the decision makers, towards restructuring the GEM in compliance with the requirements of the European Target Model. The findings can be associated with the "Private information" term stated in the strong-form of EMH efficiency, by casting the "regulatory reforming" as a pattern generating mechanism for "insiders" (agents who might have access on the exact timing of the several applied policies) to be able to exploit towards profit making purposes. Another contribution of this work is the empirical results reporting regarding the several nonlinear tools application, not only on the "immature" Greek Electricity Market, but also on the different in structure and mature Italian Electricity market (PUN), in the application's related section 3.2, as well. Really informative patterns and statistics related to complexity and predictability of the price dynamics of the two markets are reported.

The bridging section 3.3 addresses the main financial theories and concepts, upon which the several financial markets modelling applications are based on, while also discusses how the several research ideas were born and what limitations had to be handled for each one of them. Following the rationale of one of our previous works, Papaioannou et al., 2013, we further challenge the validity of one of the building blocks of finance, namely the Efficient Market hypothesis. Focusing particularly on its weak form, we extend the target financial assets' space into a broaded one in the application of Section 3.4. By monitoring the time evolution of the most liquid Futures contracts traded globally as acquired using the Bloomberg API from 03 January 2000 until 15 December 2014 we were able to forecast the S&P 500 index, beating the "Buy and Hold" trading strategy. The proposed approach is a trend following trading strategy based on convolution computations of 42 of the most liquid Futures contracts of four basic financial asset classes, namely, equities, bonds, commodities and foreign exchange. Simulations provide empirical evidence of directional predictability of the S&P500 Index, thus indeed proving the invalidity of the EMH's weak form of efficiency, while also enhancing the financial assets' forecasting toolkit of quantitative trading academics and professionals.

The next application detailed in Section 3.5 is focused on examining predictable patterns in the FOREX markets, exploiting them in order to construct profitable trading strategies under the same EMH challenging spirit of the previous two applications. Based on the work of Cuturi and D'Aspremont, 2016, "Mean-Reverting Portfolios: Tradeoffs Between Sparsity and Volatility", we develop a hybrid scheme based on the PCA embedding and several time series forecasting models, as the traditional Exponential Moving Average (EMA), Autoregressive (AR) models, Gaussian Process Regressors (GPR) and Recurrent Neural Networks (RNN). We validate their findings regarding the "Mean-Reverting" nature of specific embedded portfolios, discussing several connections of their dynamics with a specific set of stochastic processes, namely the Ornstein Uhlenbeck, Fractional Brownian Motion and Matern process, respectively. We connect all these, with the concepts of Hurst Exponent and Matern kernel setting in the Gaussian Process Regression framework, while also discussing their practical meaning on the specific FOREX markets' trading application setup. Additionally to the PCA, we introduce Local Linear Embedding (LLE) manifold learning technique into the hybrid scheme, using it in order to create trading indicators, potentially capable of capturing the several PCA costructed portfolios, as well. Finally, we discuss the limitations of the approach while preparing the ground for the next application's more advanced context on the same theme.

In the application of Section 3.6, we address a three-tier numerical framework based on manifold learning for the forecasting of high-dimensional time series. At the first step, we embed the time series into a reduced low-dimensional space using a nonlinear manifold learning algorithm such as Locally Linear Embedding and Diffusion Maps, but on a modified context compared to the previous application's approach. At the second step, we construct reduced-order regression models on the manifold, in particular Multivariate Autoregressive (MVAR) and/or Gaussian Process Regression (GPR) models, to forecast the embedded dynamics. At the final step, we lift the embedded time series back to the original high-dimensional space using Radial Basis Functions interpolation and Geometric Harmonics. For our illustrations, we first display the proposed numerical scheme's results on forecasting three artificial (synthetic) stochastic datasets : resembling EEG signals produced from linear and nonlinear stochastic models with different model orders, while finally applying it on FOREX trading, as well. The forecasting performance of the proposed numerical scheme is assessed using the combinations of manifold learning, modelling and lifting approaches. We also provide a comparison with Principal Component Analysis using the leading components as variables as well as with the naive random walk model and the MVAR and GPR models trained and implemented directly in the high-dimensional space.

In the final Section 3.7, we move away from the financial modelling space. As many research teams from all over the world, we dedicated a substantial workload towards the combat COVID-19 initiatives that emerged during the pandemic's period. From our standing, we decided to contribute in this battle, by applying the several time series models on the predicting the pandemic's evolution. We created a prediction scheme based on Gaussian Process Regression, in order to forecast the evolution of the pandemics in several regions of Italy. The model is based on the historical dynamics of the infected cases in the several regions, as reported on a daily basis by the responsible state authorities. We report significant multiple-days-ahead predictive efficiency of the model in terms of normal-plots of the corresponding residuals. The outlined in this Thesis GPR model, is part of a wider armory of forecasting tools deployed and acting as the basis of the "#combatcovid19uninafisr", 2021 web application, designed as a "monitor", reporting daily updates of several pandemic's associated statistics (like the effective reproduction number R_t), as well as the corresponding models' calculated predictions regarding its evolution, in the several Italian regions.

Chapter 2

Methodology

The methodological framework proposed and its elements, are members of both the “traditional" (linear and nonlinear) time series analysis methods, but also from the more recent machine learning advances. The list includes among others, the ARIMA, ARIMAX, SARIMAX, Exponential Smoothing and GARCH, Multivariate Autoregressive (MVAR) methodologies from the "traditional" approaches, the ANNs, SVMs, RNNs, LSTMs and GPRs from the supervised machine learning space of models, while particular focus is directed on the manifold learning (unsupervised machine learning) techniques like the linear Principal Component Analysis (PCA), but also more importantly, the nonlinear ones, like the Local Linear Embedding (LLE) and Diffusion Maps (DM) and their specific use on the "hybrid" schemes and prediction applications. The applications vary from Energy Markets System Marginal Price and Load forecasting, anomalies' and/or potential market distortions detection due to regulatory and legal framework's changes to traditional financial markets' assets trading and portfolio management applications (focusing on the asset class of Foreign Exchange (FOREX)).

The flow diagram of Figure 2.1 clusters the several methodological time series analysis frameworks and models developed throughout the years, while positioning our proposed hybrid approaches proposed in this Thesis.

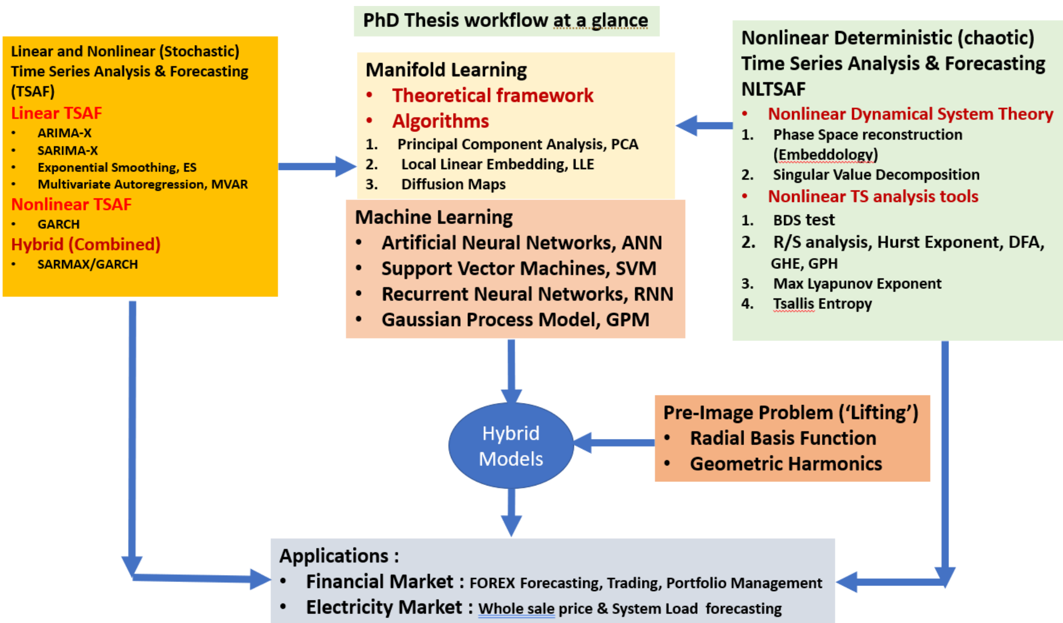

FIGURE 2.1: Schematic of the State-of-the Art Methodologies for Forecasting Time Series

The section starts by setting the mathematical framework of forecasting using regression models, while discussing the underlying concepts, assumptions and limitations of linear modelling and forecasting. It then discusses the differences between linear and nonlinear time series methodologies, related to the concepts of "the curse of dimensionality" and complexity, while also presenting the several nonlinear tools' mathematical foundations, in greater detail. Besides the classical nonlinear tools associated with complexity, chaos detection, fractals, time delays and embedding dimension calculation, the also nonlinear in nature Machine Learning models applied in the several applications are investigated, while allocating much space on the Gaussian Process Regression framework. Furthermore, the section connects nonlinear time series analysis and manifold learning, discussing the the need of "hybrid" approaches developed upon the concepts of dimensionality reduction and embedding. Finally, towards the "Embed-Train-Lift" hybrid approach, the pre-image problem is defined and investigated, with the Radial Basis Functions and Geometric Harmonics framework

2.1 Forecasting with Regression Models

Let us assume that the time series are being generated by a (nonlinear) law of the form

$$\boldsymbol{y}_t = \boldsymbol{\phi}(\boldsymbol{z}, \boldsymbol{\mu}) + \boldsymbol{e}_t, \tag{2.1}$$

where $\boldsymbol{y}_t \in \mathbb{R}^d$ denotes the vector containing the measured values of the response variables at time t, $\boldsymbol{\phi} : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^d$ is a function encapsulating the law that governs the system dynamics, which contains the parameters and exogenous variables represented by the vector $\mu \in \mathbb{R}^q$, $\boldsymbol{z} \in \mathbb{R}^p$ is the vector containing the explanatory input variables (predictors), which may include past values of the response variables at certain times, say $t-1, t-2, \ldots, t-m$ and $\boldsymbol{e}_t$ is the vector of the unobserved noise at time t.

In general, the forecasting problem (here in the low-dimensional space) using regression models can be written as a minimization problem of the following form:

$$\underset{g \in G,\ \boldsymbol{\theta} \in \mathbb{R}^l}{\operatorname{argmin}} \ \mathcal{L}(\boldsymbol{y}_{t+k} - \boldsymbol{g}(\boldsymbol{z}, \boldsymbol{\theta})), \tag{2.2}$$

where k is the prediction horizon, $\boldsymbol{g} : \mathbb{R}^p \times \mathbb{R}^l \to \mathbb{R}^d$ is the regression model

$$\hat{\boldsymbol{y}}_{t+k} = \boldsymbol{g}(\boldsymbol{z}, \boldsymbol{\theta}), \tag{2.3}$$

with parameters $\boldsymbol{\theta} \in \mathbb{R}^l$, G is the space of available models $\boldsymbol{g}$, $\hat{\boldsymbol{y}}_t$ is the output of the model at time t, and $\mathcal{L}$ is the loss function that determines the optimal forecast. We note that the forecasting problem can be posed in two different modes Marcellino, Stock, and Watson, 2006: the iterative and the direct one. In the iterative mode, one trains one-step ahead models based on the available time series and simulates/iterates the models to predict future values. In the direct mode, forecasting is performed in a sliding-window framework, where the model is retrained with the data contained within the sliding window to provide a multiperiod-ahead value of the dependent variables.

2.1.1 Linear Gaussian Models (LGM)

We start with two definitions :

Definition : A time series X_t, is a family of real – valued random variables indexed by $t \in Z$, where Z denotes the set of integers.
Definition : X_t is said to be stationary if, for any $t_1, t_2, \dots t_n \in Z$, any $k \in Z$ and $n = 1, 2, \dots,$

$$F_X(X_{t_1}, X_{t_2}, \dots, X_{t_n}) = F_X(X_{t_1+k}, X_{t_2+k}, \dots, X_{t_n+k}) \tag{2.4}$$

where F is the distribution function of the set of random variables. The term strictly stationary describes more often 2.4. Weakly stationary, or second – order stationary or covariance stationary are used for the expressions $E(X_{t_1}) = E(X_{t_1} + k)$, $Cov(X_{t_1}, X_{t_2}) = Cov(X_{t_1+k}, X_{t_2+k})$, $\forall t_1, t_2, k \in Z$. We assume the existence of covariances.

For almost 70 years LGMs have dominated the "building" of time series models. Actually, the area of linear time series began with the AR (Autoregressive) models of (Yule, 2012; Brockwell et al., 2016). The AR is written as :

$$X_t = \alpha_0 + \sum_{j=1}^{k} \alpha_j X_{t-j} + \epsilon_t \tag{2.5}$$

where α_j are real constants ($\alpha_k \neq 0$), k is a finite positive integer or the order of the AR model, and ϵ_t are zero – mean uncorrelated random variables, so – called white noise, with common variance $\sigma_\epsilon^2 (< \infty)$. We usually express 2.5 by $X_t \sim AR(k)$. By replacing ϵ_t with a weighted mean of the $\epsilon_t, \epsilon_{t-1}, \dots, \epsilon_{t-l}$:

$$X_t = \alpha_0 + \sum_{j=1}^{k} \alpha_j X_{t-j} + \sum_{j=0}^{l} \beta_j \epsilon_{t-j} \tag{2.6}$$

by real constants $\beta_j \neq 0, \quad b_0 = 1$ (usually).

Relation 2.6 is the so called autoregressive / moving average (ARMA) model, symbolized also as $X_t \sim ARMA(k, l)$, with l being the order of the ARMA model. If $k = 0$, we get the moving – average (MA) model of l order, the MA(l). The conditions on the ARMA models (Brockwell et al., 2016) are :

Condition 1 : The roots of polynomials :

$$A(z) = z^k - \sum_{j=1}^{k} \alpha_j z^{k-j} \tag{2.7}$$

$$B(z) = \sum_{j=0}^{l} \beta_j z^{l-j}, \quad (b_0 = 1) \tag{2.8}$$

with modulus less than one. The function of $A(z)$ and $B(z)$ are called the AR generating function and MA generating functions, respectively.

Condition 2 : ϵ_t is a sequence of i.i.d. (independent and identically distributed) random variables, following $\sim \mathcal{N}(0, \sigma^2)$, distributions. It is called the Gaussian white noise of the process.

Let the mean and variance of X_t being given by :

$$\mu_x = \frac{a_0}{1 - \sum_{j=1}^{k} \alpha_j} \tag{2.9}$$

$$\sigma_x^2 = \frac{\sigma_\epsilon^2}{2\pi}\int_{-\pi}^{\pi} |\frac{B(e^{-i\omega})}{A(e^{-i\omega})}| d\omega \tag{2.10}$$

Now, lets $X_0 \sim \mathcal{N}(\mu_x, \sigma_x^2)$. Then by using the above conditions, we have :

1. $X_t : \quad t = 1, 2, \ldots$ is stationary
2. $\forall t_1, t_2, \ldots, t_k \in Z_+, \quad Z_+$ the set of non-negative integers, and $\forall k \in Z_+$, then $(X_{t_1}, X_{t_2}, \ldots, X_{t_k})$ is joinlty Gaussian - $X_t : t = 1, 2, \ldots$ is a Gaussian sequence.
3. X_t admits the linear (one-sided) model expression

$$X_t = \mu_x + \sum_{j=0}^{\infty} \beta_j \epsilon_j \tag{2.11}$$

with $\sum_{j=0}^{\infty} \beta_j \quad < \infty$, $\beta_0 = 1$ and $\mu_x = E(X_t)$

ϵ_t also admits a linear model in terms of X_s, $\quad s \leq t$, which is called the invertibility of X_t. All ARMA models are assumed to satisfy the above conditions. If a model is of the form 2.11, in which ϵ_t is an i.i.d. process, then is called a linear Gaussian model.

Now let C being the class of continuous integrable fanctions Priestley and Rao, 1969. An ARMA model has the very important property of having a finite number of parameters. In the process of model building, the class of ARMA models is considered a useful choise of C, if and only if the autocoraviances are considered an important feature. ARMA models have the following advantages :

1. ARMA are linear difference equations, for which a complete mathematical theory exists. Also, the Gaussian sequence is well and easily understood, from the probabilistic point of view, with well developed statistical inference. In terms of the predictor space and the Markovian representation, suggested by Akaike, 1974, the stationary Gaussian ARMA models have an elegant geometric characterization, viewed in control system theory perspective. An important result is that a stationary Gaussian time series, has a stationary Gaussian ARMA representation, if and only if, its predictor space is finite dimensional.
2. Fitting an ARMA model on a given time series data requires very little time (in the most cases), and over the years a large amount of experience has been accumulated in applying the model Box, Jenkins, and Reinsel, 2013
3. An ARMA model is a successful practical tool for analysis modeling and forecasting of time series, representing reality with a good first approximation.

However, ARMA models have the following drawbacks :

1. If we set the ϵ_t to be constant, or $var(\epsilon_t) = 0$, Equation 2.11 is transformed to a deterministic difference equation in X. A stable limit point state is attained under Condition 1 as $t \to \infty$, since $X_t \to const$, independent of the initial value. If $|A(z) > 1|$, then $|X_t| \to \infty$ as $t \to \infty$, i.e. an unstable state is generated. If some of the roots of $A(z)$ are equal to 1 and the other are less than 1, then X_t oscillates depending on the initial value. This is a neutrally stable situation. Thus, based on all the above, linear difference equations do not permit stable periodic solutions independent of the initial value.

2. When the time series data exhibit sudden bursts of very large amplitudes at irregular time instances, ARMA models are not the ideal ones to capture this dynamics behavior. Instead, ARMA models are more suitable for time series exhibiting very small probability of very high crossing. If is Known from statistics that the probability of large excursions is associated with the existence of moments, which means that if the set of possible values of X_t is bounded from both sides, then its distribution has the moment property, i.e. moments of all orders exist. A Gaussian random variable has the moment property. We rote here that models with out the moment property may be built by just perturbing ARMA models. For example the bilinear models $X_t = (\alpha + \beta\epsilon_t)X_t + \epsilon_t$, with ϵ_t Gaussian noise (Zero mean, unit variance), have the moment property, The coefficient of X_t is no longer a constant. We prefer these model to the ARMA, when we have time series with sadden bursts.

3. Stationary Gaussian ARMA models are not ideally fitted well on data exhibiting strong asymmetry, since these models have symmetric joint distributions.

4. ARMA models are usually weak in capturing strong cyclicity present in time series data. This is because due to its joint normality, ARMA models see all aspects of the joint distributions of (X_t, X_{t-j}) as being linear via the autocovariances γ_j. However, other vital aspects as the information contained in the regression function $E(X_t|X_{t-j})$ at the j-th lag, is not captured by γ_j, It is also noted that since the ACF (autocorrelation function) is strongly cyclical if the data are strongly cyclical at lags where ACF is very large in modulus, the regression function is well approximated by linear functions. At lags however, where ACF is very small, the above approximation is under question i.e. the linear approximation may fail. Thus, a nonlinear approximation for the regression function may be more appropriate at the lags where ACF values are very small, in order to capture the nonlinear effects.

5. For data exhibiting time irreversibility, ARMA models are not ideally suited. By introducing higher – order spectra, we can gain further insights about how time irreversibility affects the probabilish structure of X_t.

The concept of Autocorrelation function (ACF), the power spectra and the coefficients of a linear model, contain the same information about its response (dynamic behavior) due to its excitation (or diving) by an uncorrelated white noise. Therefore, in order to characterize the features of a time series via e.g. an ARMA model, one would need to focus on the corresponding power spectrum, as an efficient and useful fool for such characterization purposes. However the power spectrum can completely fail due to existence of even simple non-linearities in the time series, which make power spectra very complex. A deterministic (noise free) nonlinear system with a small number of degrees of freedom, and a linear system driven stochastically by external noise, can produce very similar broadband spectra, and the natural question, is how we can distinguish the two different systems by a linear (as the ACF or power spectrum) tool. An example of such systems is the well known model of May, 1976 population dynamics, where the next value of a series is generated by the current one, according to the rule $x_{t+1} = \lambda x_t(1 - x_t)$, i.e. a parabola called the logistic map. The parameter λ can generate a fixed point (λ small) behavior, as well as deterministic chaos (divergence of nearly trajectories). Such a map generates a broadband power spectrum. Also the map $x_t = 2x_{t-1}(mod 1)$ can be implemented in a simple physical system consisting of a classical billiard ball and reflecting surfaces. The x_t provides the successive positions where the ball crosses a given line, (Moore and Spiegel, 1966). Even though both the above systems are totally deterministic, since the evolutionary dynamics is completely determined by the initial condition X_0), they can generate signals with broadband power spectra. If we analyze the two systems with an ARMA model then we must attribute the broadband spectra to an external stochastic factor that drives this linear model. However, in the case of these two models, the internally generated nonlinearities (due to parabola and straight line) resulted in broadband spectra therefore the ARMA model is not a good choice for describing these two maps. This is a manifestation of failure of a linear model. In 1980, Tong and Lim, 2009 suggested that instead of making a one globally linear model (function) for predicting time series, they used two functions, the so-called threshold autoregressive model (TAR), which is a globally nonlinear model consisting of adopting one or two local linear AR models, depending on the value of the state of the underlying system. In order for a nonlinear model to be efficient in capturing both linear and nonlinear structures "hidden" in the time series, a process that can detect such structures and help the construction of the model, must be available. The application of nonlinear time series is limited if such a process (which sheds light in the hidden nonlinearites) is absent. The construction of models based on state-space reconstruction connecting frameworks (neural networks) and manifold learning (ML) approaches (one of the main topics of this thesis), offer complementary solutions to the problems of linear and nonlinear (stochastic and deterministic) models, as mentioned before.

As a conclusion, the above drawbacks – limitations of the linear Gaussian models are serious and strongly limit their capability in capturing the rich nonlinear effects hidden in specific time series data. One remedy would be the incorporation in the model, features different than the autocovariances, i.e. to allow the white noise to be non-Gaussian and retain the ARMA framework. Another remedy would be to abandon completely the assumption of linearity. In the first case, limitations 2, 3, 4 and 5 of Gaussian ARMA could be removed by a more wise choice of the distributions of ϵ_t, which would result, for example, a non-Gaussian MA(l). However, we point out here that even the non-Gaussian ARMA models fail to remove limitation 1, so there is a need for a paradigm shift, to adopt nonlinear models capable of capturing much richer dynamical properties. In this thesis we show how the combination of tools taken from nonlinear determinstic or stochastic time series analysis and forecasting fields, as well as from the manifold learning unsupervised machine learning subfield, can lead to "hybrid" models and schemes, that have this capability.

2.1.2 Seasonal ARIMA(X) and traditional ARIMA(X) Models

The seasonal (or multiplicative) ARIMA model (SARIMA) is an extension of the Autoregressive Integrated Moving Average (ARIMA) model, Brockwell et al., 2016, when the series contains both seasonal and non-seasonal behavior, as our time series load does. This behavior of load makes the typical ARIMA model inefficient to be used. This is because it may not be able to capture the behavior along the seasonal part of the load series and therefore mislead to a wrong selection for non-seasonal component.

Let d and D are nonnegative integers, the series x_t is a seasonal ARIMA(p,d,q)(P,D,Q) process (Brockwell et al., 2016) with seasonality or period s if the differenced series:

$$y_t = (1 - B)^d (1 - B^s)^D x_t \tag{2.12}$$

is a casual ARMA process (Brockwell and Davis, 1991), defined by:

$$\phi_p(B)\Phi_P(B^s)y_t = \theta_q(B)\Theta_Q(B^s)\epsilon_t, \epsilon_t \sim WN(0, \sigma^2) \tag{2.13}$$

where B is the difference or lag operator, and $\phi(x), \Phi(x), \theta(x)$ and $\Theta(x)$ polynomials defined as follows:

$$\phi_p(x) = 1 - \phi_1 x - \cdots - \phi_p x^p, \Phi_P(x) = 1 - \Phi_1 x - \cdots - \Phi_P x^P \tag{2.14}$$

$$\theta_q(x) = 1 + \theta_1 x + \cdots + \theta_q x^q, \Theta_Q(x) = 1 + \Theta_1 x + \cdots + \Theta_Q x^Q \tag{2.15}$$

In the ARIMA form (p, d, q) (P, D, Q)S the term (p, d, q) is the non-seasonal and (P, D, Q)S is the seasonal part. We note here that the process y_t is casual if and only if $\phi(x) \neq 0$ and $\phi(x) = 0$ for $|x| \leq 1$, while in applications d is rarely more than 1, $D = 1$ and P and Q are typically less than 3, Brockwell et al., 2016. Also, if the fitted model is appropriate, the rescaled residuals should have properties similar to those of a white noise ϵ_t driving the ARMA process. All the diagnostic tests that follow, on the residuals, after fitting the model, are based on the expected properties of residuals under the assumption that the fitted model is correct and that $\epsilon_t \sim WN(0, \sigma^2)$.

In order to incorporate exogenous variables in the SARIMA model, to obtain a SARIMAX model (Brockwell et al., 2016) the previous model is modified as follows:

$$\phi_p(B)\Phi_P(B^s)y_t = c + \beta X_t + \theta_q(B)\Theta_Q(B^s)\epsilon_t \tag{2.16}$$

where y_t as previously is the univariate response series, in our case load X_t is the row t of X, which is the matrix of predictors or explanatory variables, β is a vector with the regression coefficients corresponding to predictors, c is the regression model intercept, and ϵ_t is a white noise innovation process. The ϕ_p, ϕ_P, θ_q and Θ_Q are as previously defined i.e., the lag operator polynomials, seasonal and nonseasonal:

1. $\phi_p(B)$: nonseasonal AR
2. $\Phi_P(B^s)$: seasonal AR(SAR)
3. $(1 - B)^d$: nonseasonal integration
4. $(1 - B^s)^D$: seasonal integration
5. c : intercept
6. βX_t : exogenous variables
7. $\theta_q(B)$: nonseasonal MA
8. $\Theta_Q(B^s)$: seasonal MA

2.1.3 Exponential Smoothing (ES)

The exponential smoothing approaches date back to the 1950s. However, procedures for model selection were not developed until relatively during our days. All exponential smoothing methods (including non-linear methods) have been shown to be optimal forecasts from innovations state space models Rob J Hyndman and George, 2018. Pegels, 1969 was the first to make a taxonomy of the exponential smoothing methods. Rob J Hyndman and George, 2018 modified the work of Gardner, 2006

TABLE 2.1: Classification of fifteen exponential smoothing methods, by Taylor, 2003

	Trend Component	Seasonal Component N (None)	A (Additive)	M (Multiplicative)
N	None	N,N (SES)	N,A	N,M
A	Additive	A,N Holt's linear method	A,A Additive Holt's Winter	A,M Multiplicative Holt's Winter)
A_d	Additive damped	A_d,N Damped trend method	A_d,A	A_d,M
M	Multiplicative	M,N	M,A	M,M
M_d	Multiplicative damped	M_d,N	M_d,A	M_d,M

who had previously extended the work of Pegel's classification table. Finally, Taylor, 2003 provided the following Table 2.1 with all fifteen methods of exponential smoothing.

However, a subset of these methods is better known with other names. For instance, the well-known simple exponential smoothing (or SES) method corresponds to cell (N, N). Similarly, cell (A, N) is Holt's linear method, and the damped trend method is given by cell (Ad, N). Also, the additive Holt-Winters' method is given by cell (A, A) while and the multiplicative Holt-Winters' method is given by cell (A, M). As it is described in Weron, 2006 and Taylor, 2012, exponential smoothing is the traditional model used in forecasting load data series. Due to its satisfactory results it has been adopted by lots of utilities. However, it is pointed out in the above paper that seasonal multiplicative smoothing, SMS, or a variant of it Holt-Winter's SMS are very useful in capturing the seasonal (periodicity) behavior. Centra, 2011, has applied a Holt-Winter's SMS model to forecast hourly electricity load. The given time series Y may be decomposed into a trend (T), seasonal (S) and error (E) components. The trend reflects the long-term movement of Y, the seasonal corresponds to a pattern with known periodicity and the error is the uncertain, unpredictable constituent of the series. In a purely additive model we have: $Y = T + S + E$ or $Y = S + T$, while in a pure multiplicative model: $Y = T \cdot S \cdot E$ or $Y = T \cdot E$. Hybrid models are: $Y = (T \cdot S) + E$ or $Y = (T + S)(1 + E)$. Let now the forecasted trend T h periods out, by these models, be decomposed i nto a level term (l) and a growth term (b) in a number of ways. Let also $0 < \phi < 1$ be the damping parameter. The five different trend types, according to different assumptions about the growth term.

Let also s represent the included seasonal terms, we can define the general p-dimensional state vector, $x_t \equiv (l_t, b_t, s_t, s_{t-1}, \ldots, s_{t-m})'$. A nonlinear dynamic model representation of the ES equations using state space model with a common error part (Ord et al., 1997):

$$y_t = h(x_{t-1}, \theta) + k(x_{t-1}, \theta)\epsilon_t \tag{2.17}$$

Additive	$T_h = l + b \cdot h$
Additive damped	$T_h = l + b \cdot \sum_{s=1}^{h} \phi^s$
Multiplicative	$T_h = l \cdot b^h$
Multiplicative damped	$T_h = l \cdot b^{\phi}$
None	$T_h = l$

$$x_t = f(x_{t-1}, \theta) + g(x_{t-1}, \theta)\epsilon_t \tag{2.18}$$

where h, k continuous scalar functions, f and g continuous functions with continuous derivatives, $R^p \to R^P$ and $\epsilon_t \sim WN(0, \sigma^2)$, independent past realizations of y and x. θ is a parameter set. Intuitively, y_t reflects the way the various state variables (l_{t-1}, b_{t-1}, s_{t-m}) are combined to express $\hat{y}_t = h(x_{t-1}, \theta)$ and ϵ_t. With additive errors we have $k = 1$, so:

$$y_t = h(x_{t-1}, \theta) + \epsilon_t \tag{2.19}$$

with multiplicative errors, $k = h$, so:

$$y_t = h(x_{t-1}, \theta)(1 + \epsilon_t) \tag{2.20}$$

For the ETS models we applied in this work, we consider the updating smoothing equations as being weighted average of a term depending on the current prediction error (and prior states), and one depending on the prior states. The general form of the resulting state equations is:

$$l_t = \alpha P(x_{t-1}, \epsilon_t) + (1 - \alpha)Q(x_{t-1}) \tag{2.21}$$

$$b_t = \beta R(x_{t-1}, \epsilon_t) + (1 - \beta)\phi_1 b_{t-1}^{\phi_2} \tag{2.22}$$

$$s_t = \gamma T(x_{t-1}, \epsilon_t) + (1 - \gamma)s_{t-m} \tag{2.23}$$

where $P_t = P(x_{t-1}, \epsilon_t)$, R_t and T_t are functions of the forecasting error and lagged states and $Q_t = Q(x_{t-1})$ is a function of the lagged states. ϕ_1, ϕ_2 are the damping parameters for linear trend and multiplicative trend models respectively (they are 1, in the absence of damping). The exact form of the above depends on the specific ETS specifications. In Hand, 2009, all the 30 possible specifications are listed.

Simple Exponential Smoothing (A,N,N)

By inserting $x_t = l_t$ and $h(x_{t-1}, \theta) = l_{t-1}$, $P_t = l_{t-1} + \epsilon_t$ and $Q_t = l_{t-1}$ in previous Equations the ETS representation of this (A,N,N) model, in the form of (10–12) is : $y_t = l_{t-1} + \epsilon_t$ and $l_t = l_{t-1} + \alpha\epsilon_t$. Similarly, the ETS expressions for the models Holt's Method with multiplicative Errors (M,A,N) and Holt-Winters method with multiplicative errors and seasonal (M,A,M), and fully multiplicative method with damping (M,Md,M), are given in the Appendix.

2.1.4 Multivariate Autoregressive (MVAR) model

An MVAR model (Brockwell et al., 2016) can then be written as

$$\boldsymbol{y}_t = \boldsymbol{\theta}_0 + \sum_{j=1}^{m} \boldsymbol{y}_{t-j}\boldsymbol{\Theta}_j + \boldsymbol{e}_t, \tag{2.24}$$

where $\boldsymbol{y}_t = [y_{t1}, y_{t2}, \ldots, y_{td}]^T \in \mathbb{R}^d$ is the vector of the response time series at time t, m is the model order, i.e. the maximum time lag, $\boldsymbol{\theta_0} \in \mathbb{R}^d$ is the regression intercept, $\boldsymbol{\Theta}_j \in \mathbb{R}^{d\times d}$ is the matrix containing the regression coefficients of the MVAR model, and $\boldsymbol{e}_t = [e_t^{(1)}, e_t^{(2)}, \ldots, e_t^{(d)}]^T$ is the vector of the unobserved errors at time t, which are assumed to be uncorrelated random variables with zero mean and constant variance σ^2. In a more general form, in view of (2.1), the MVAR model can be written as

$$y_{ik} = \theta_{0k} + \sum_{j=1}^{m} \theta_{jk} z_{ij} + e_{ik}, \quad i = 1, 2, \ldots d, \quad k = 1, 2, \ldots N, \tag{2.25}$$

where y_{ik} is the model output for the i-th variable at the k-th time instant, $\theta_{0k} \in \mathbb{R}$ is the corresponding regression intercept and θ_{jk} the corresponding j-th regression coefficient, z_{ij} is the j-th predictor of the i-th response variable (e.g. the time-delayed time series). According to the Gauss-Markov theorem, the best unbiased linear estimator of the regression coefficients is the one that results from the solution of the least-squares (LS) problem

$$\arg\min_{\theta_{0k},\theta_{jk}} \sum_{i=1}^{d} \sum_{k=1}^{N} \left(y_{ik} - \theta_{0k} - \sum_{j=1}^{m} \theta_{jk} z_{ij} - e_{ik} \right)^2,$$

given by

$$\hat{\boldsymbol{\Theta}} = (\boldsymbol{Z}^T \boldsymbol{Z})^{-1} \boldsymbol{Z}^T \boldsymbol{Y}, \tag{2.26}$$

where $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots \boldsymbol{y}_d] \in R^{N\times d}$ and $\boldsymbol{Z} = [\mathbf{1}_N, \boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_m] \in \mathbb{R}^{N\times(m+1)}$. Assuming that the unobserved errors are i.i.d. normal random variables, then by the maximum likelihood estimate of the error covariance, one can also estimate the forecasting intervals of a new observation.

2.1.5 ANNs and SVMs

Artificial Neural Networks (ANN) being the building block of supervised machine learning regression models, aim to mimic brain activity and approach solution to a problem by applying a laws that the network has ''learned'' as a result of a training process Vapnik, 1999; Goodfellow, Bengio, and Courville, 2016. Much like a biological brain, the ANN is a set of individual operators, called neurons that co-operate to accomplish a task. Neurons are structured in layers and typically an ANN has an input layer, a number of hidden layers and an output layer. The number of neurons in each layer can vary and usually is selected by an optimization algorithm. Each neuron has one input and its output is connected with all neurons of the next layer, creating the connections of the network. There are various neuron types, most common being the nonlinear sigmoid operator (Goodfellow, Bengio, and Courville, 2016) :

$$f(x) = \frac{1}{exp(-ax + b) + c} \tag{2.27}$$

An ANN is functional only after the training process is complete, thus the parameters of each neuron are estimated. There exist numerous training algorithms, some benchmark methods being Hebbian training and error back-propagation.

Support Vector Machine (SVM) is a machine learning algorithm which was first used in recognition and classification problems. Through years the same concept has been modified in order to be used also in regression problems. SVM projects the given time series (input data) into a kernel space and then builds a linear model in this space. First a Training set must be defined : $D = [x_i, y_i] \in R^n \times R, i = 1, \dots, l$, where x_i and y_i are the input and output vectors, respectively. Then a mapping function $\Phi(x)$ is defined which maps inputs x_i into a higher dimensional space, $\Phi(x) : R^n \to R^f, x \in R^n \to \Phi(x) = [\Phi_1(x)\Phi_2(x)\dots\Phi_n(x)]^T \in R^f$. The mapping converts the nonlinear regression problem into a linear regression one, of the form : $f(x, w) = w^T\boldsymbol{\Phi}(x) + b$. In the former equation w represents the weight factor and b is a bias term. Both parameters can be determined by minimizing the error function R (Kecman, 2002; Goodfellow, Bengio, and Courville, 2016) :

$$R_{w,\xi,\xi_*} = \frac{1}{2}ww^T + C\sum_{i=1}^{l}(\xi_i + \xi_{i*}) \tag{2.28}$$

where ξ_i is the upper limit of the training error and ξ_i^* the lower one. C parameter is a constant which determines the trade-off between the model complexity and the approximation error. The above equation is subjected to the following constraints:

$$y_i - w^T x_i - b \leq \epsilon + \xi_i \tag{2.29}$$

$$w^T x_i + b - y_i \leq \epsilon + \xi_i^* \tag{2.30}$$

with $i = 1, \dots, l$, $\xi \geq 0$ and $\xi^* \geq 0$. In the above equation ϵ defines the Vapnik's insensitive loss function which determines the width of the ϵ-tube.

The former optimization problem can be solved with the help of Lagrangian multipliers:

$$w = \sum_{i=1}^{l}(a_i - a_i^*)\boldsymbol{\Phi}(x_i) \tag{2.31}$$

Thus:

$$f(\boldsymbol{x}, \boldsymbol{w}) = \boldsymbol{w}^T\boldsymbol{\Phi}(\boldsymbol{x}) + b = \sum_{i=1}^{l}(a_i - a_i^*)\boldsymbol{\Phi}(\boldsymbol{x_i})\boldsymbol{\Phi}(\boldsymbol{x}) + b = \sum_{i=1}^{l}(a_i - a_i^*)K(\boldsymbol{x_i}, \boldsymbol{x_j}) + b \tag{2.32}$$

where a_i and a_i^* are the Lagrange multipliers and $K(\boldsymbol{x_i}, \boldsymbol{x_j})$ represents the Kernel function. Kernel functions are used in order to reduce the computational burden since $\Phi(x)$ has not to be explicitly computed. By maximizing the Lagrange multipliers the Karush-Kuhn-Tucker (KKT) conditions for regression are applied (Kecman, 2002) : Maximization of:

$$L_d(\boldsymbol{a}, \boldsymbol{a^*}) = -\epsilon + \sum_{i=1}^{l}(a_i - a_i^*)y_i - \frac{1}{2}\sum_{Lj=1}^{l}(a_i - a_i^*)(a_j - a_j^*)K(\boldsymbol{x_i}, \boldsymbol{x_j}) \tag{2.33}$$

Subject to the following constraints:

$$0 \leq a_i^*, \quad a_i \leq C, \quad \sum_{i=1}^{l}(a_i - a_i^*) = 0 \tag{2.34}$$

where $i = 1, \dots, l$. The performance of the SVM regression model is strongly dependent on the selection of the parameter C, ϵ as well as the chosen kernel function.

2.1.6 Recurrent Neural Networks (RNNs)

The Recurrent Neural Network (RNN) model (Goodfellow, Bengio, and Courville, 2016) is defined following the general mathematical formulation of (Hochreiter and Schmidhuber, 1997) and (Gers, Schmidhuber, and Cummins, 2000):

$$f_t = \sigma_g(\tilde{W}_f \pi_t + \tilde{U}_f h_{t-1} + b_f) \tag{2.35}$$

$$i_t^* = \sigma_g(\tilde{W}_{i^*} \pi_t + \tilde{U}_{i^*} h_{t-1} + b_{i^*}) \tag{2.36}$$

$$o_t = \sigma_g(\tilde{W}_o \pi_t + \tilde{U}_o h_{t-1} + b_o) \tag{2.37}$$

$$c_t^* = f_t \circ c_{t-1} + i_t^* \circ \sigma_c(\tilde{W}_c \pi_t + \tilde{U}_c h_{t-1} + b_c) \tag{2.38}$$

$$h_t = o_t \circ \sigma_h(c_t^*), \tag{2.39}$$

where $\pi_t \in R_d$ acts as the input vector in the RNN's unit, $f_t \in R_h$ is the activation vector of the forget gate, $i_t \in R_h$ is the input or update gate's activation vector, $o_t \in R_h$ is the activation vector of the output gate, $h_t \in R_h$ is the hidden state vector, and $c_t^* \in R_h$ is the cell state vector. The training consists of learning the weights of the matrices $\tilde{W} \in R_{hxd}$, $\tilde{U} \in R_{hxh}$ and the bias vector's parameters set $b \in R_h$, with d and h being the input features' and hidden units' numbers, respectively. Finally, σ_g is the sigmoid function, σ_h is the hyperbolic tangent function or $\sigma_h(x) = x$ ((Gers and Schmidhuber, 2001), (Gers, Schraudolph, and Schmidhuber, 2003)) and σ_c is the hyperbolic tangent function.

2.1.7 Gaussian Process Model - Regression (GPR)

This section goes through the Gaussian Process Model's mechanics, specifications but also intuition behind the several underlying mathematical concepts. The applications outlined in the related sections are heavily based on the usage of the Gaussian process model for solving several regression tasks. As such particular focus and deeper investigation on the internal mathematical blocks is given.

Bayesian Approach

The regression task is the one trying to find the target unknown function $\boldsymbol{\phi}$ which optimally explains the structure and connection of sets of training data points, namely between the input ("regressors" or "predictors" variables) and the target ("response" or "predicted") variables. Since the Gaussian Process modelling is based on the Bayesian (or "frequentist") approach, the terms "prior probability distribution", "likelihood" and "posterior probability distribution" should be introduced and intuitively examined first, since they constitute the main building blocks of the whole modelling rationale. More specifically, under this approach the model assumes that there is sufficient information to estimate the several model's parameters present in the training set, and if careful combination of the so called "prior belief" of the model's structure (i.e. parameters) and the data's "likelihood" is achieved, then using the Bayes theorem, a "posterior belief" is retrieved about those parameters, which should be acting as the "optimal" one, in the sense of explaining "as best as possible" the training data themselves, but also any other data of interest (test set points).

We will later examine the statement "as best as possible" in the model evaluation section below. As a crucial note at this higher-level step though, we should highlight that, opposed to the case of the OLS framework of the standard linear regression problem, the model is able to identify those "optimal" parameters, by using or "carrying information" of the training data along the optimization procedure, while in the OLS standard fitting process, once the optimal parameters' set is identified the model will no longer need to remember anything from the previous training data structure. This difference between the two approaches is one the main thinking divergences of the parametric versus the non-parametric modelling frameworks.

Setting the model's framework in a mathematical context, we start by defining the input of the model, which consists of the training data Z and the test data Z_*, both of which could be one-or-multiple dimensional, according to each target application. There are several ways to address the regression task as per each subcategory of input/output dimensions as "one-to-one" problem, or "many-to-one" and "many-to-many" ones etc. The different problem setups and regressions come along with specific mathematical structures, limitations, complexities, but also applications' frameworks, as well, for example multidimensional interpolation with "Krigging" (Williams, 1998). The specific training and test sets are the ones that will be used for the model's parameters identification since they are the ones that a modeller has in-hand, meaning that the main assumption is that the identified function $\boldsymbol{\phi}$ should be extendable on any arbitrary data point, as well. We denote this arbitrary set of points, which are not visible by the modeler at the stage of the model setup, as Z_a, (or "Z_{all}" - *"a" acts as a notation letter and not as an evaluated subscript variable*).

The Bayes Theorem relates all the building blocks for a particular parameter θ and arbitrary dataset Z_a, in the following formula

$$p(\theta|Z_a) = \frac{p(Z_a|\theta)p(\theta)}{p(Z_a)} \tag{2.40}$$

where $p(\theta|Z_a)$ is the posterior, $p(Z_a|\theta)$ is the likelihood, $p(\theta)$ is the prior and $p(Z_a)$ is the marginal likelihood, acting as the "data's proof" or "empirical information gain" on the rest of the model's terms.

Under the Bayesian modelling framework, the function $\boldsymbol{\phi}$ is considered to be a random variable assigned on each input location $z_a \in Z_a$, and as such in the training and test sets, Z and Z_*, respectively. Under this context, the random variable $\boldsymbol{\phi}$, evaluated at the location $\boldsymbol{z}$, will yield a "potential" output of the function at this point, and as such it actually defines a distribution of outputs around this particular location. The Gaussian Process model assumes that all those distributions are Gaussian, and as such it completely parameterizes all of them simply by their mean and variance terms. Following this rationale, the actual output at each arbitrary location, is then a "sample" obtained by this particular underlying distribution.

Gaussian Process Regression

Quick Overview

For the implementation of GPR (Goodfellow, Bengio, and Courville, 2016) it is assumed that the unknown function $\boldsymbol{\phi}$ in (2.1) can be modelled by d single-output Gaussian distributions given by

$$p(\boldsymbol{\phi}_i|\boldsymbol{z}) = \mathcal{N}(\boldsymbol{\phi}_i|\boldsymbol{\mu}, \boldsymbol{K}(\boldsymbol{z}, \boldsymbol{z}|\boldsymbol{\theta})), \tag{2.41}$$

where $\boldsymbol{\phi}_i = [\phi_i(z_1), \phi_i(z_2), \ldots, \phi_i(z_N)]$, and ϕ_i is the i-th component of $\boldsymbol{\phi}$, $\boldsymbol{\mu}(\boldsymbol{z})$ is the vector with the expected values of the function, and $\boldsymbol{K}(\boldsymbol{z}, \boldsymbol{\theta})$ is a $N \times N$ covariance matrix formed by a kernel. The prior mean function is often set to $\boldsymbol{\mu}(\boldsymbol{z}) = \mathbf{0}$ with appropriate normalization of the data. The reason is that usually, i.e. at the majority of the applications, no prior "safe" educated guess or "knowledge" of a correct mean function is easily inferred. Although several applications mostly related to physical or biological applications could give better insights on this term, like e.g. weather related quantities as temperature and humidity or blood pressures and body dimensions etc, where physical limitations exist and are extremely rarely violated.

Predictions at a new point in the test set, say z_*, are made by drawing $\boldsymbol{\phi}_{i*}$ from the posterior distribution $p(\boldsymbol{\phi}_i|(\boldsymbol{Z}, \boldsymbol{y}_i))$, where $\boldsymbol{Z} = [z_1, z_2, \ldots, z_N]^T \in \mathbb{R}^{N\times p}$, and by assuming a Gaussian distributed noise, and appropriate normalization of the data points is given by the joint multivariate normal distribution:

$$\begin{bmatrix} \boldsymbol{y}_i \\ \phi_i(\boldsymbol{z}_*) \end{bmatrix} \sim \mathcal{N}\begin{pmatrix} \boldsymbol{0} \\ 0 \end{pmatrix}, \quad \begin{bmatrix} \boldsymbol{K}(\boldsymbol{Z}, \boldsymbol{Z}|\boldsymbol{\theta}) + \sigma^2 \boldsymbol{I} & \boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{z}_* \boldsymbol{\theta}) \\ k(\boldsymbol{z}_*, \boldsymbol{Z}|\boldsymbol{\theta}) & k(\boldsymbol{z}_*, \boldsymbol{z}_*|\boldsymbol{\theta}) \end{bmatrix}, \quad i = 1, 2, \ldots, d. \tag{2.42}$$

It can be shown that the posterior conditional distribution

$$p(\phi_i(\boldsymbol{z}_*)|\boldsymbol{y}_i, \boldsymbol{Z}, \boldsymbol{z}_*) \tag{2.43}$$

can be analytically derived and the expected value and covariance of the estimation are given by:

$$\bar{\boldsymbol{\phi}}_i(\boldsymbol{z}_*) = k(\boldsymbol{z}_*, \boldsymbol{Z}|\boldsymbol{\theta}) \left(\boldsymbol{K}(\boldsymbol{Z}, \boldsymbol{Z}|\boldsymbol{\theta}) + \sigma^2 \boldsymbol{I}\right)^{-1} \boldsymbol{y}_i, \quad i = 1, 2, \ldots, d \tag{2.44}$$

$$\sigma_*^2 = k(\boldsymbol{z}_*, \boldsymbol{z}_*|\boldsymbol{\theta}) - k(\boldsymbol{z}_*, \boldsymbol{Z}|\boldsymbol{\theta})[\boldsymbol{K}(\boldsymbol{Z}, \boldsymbol{Z}|\boldsymbol{\theta}) + \sigma^2 \boldsymbol{I}]^{-1} k(\boldsymbol{Z}, \boldsymbol{z}_*|\boldsymbol{\theta}) \tag{2.45}$$

The hyperparameters in the above equations are estimated by minimizing the marginal likelihood that is given by

$$\begin{aligned} \ell &= -\log p(\boldsymbol{y}_i|\boldsymbol{Z}, \boldsymbol{\theta}) \\ &= \frac{1}{2}\boldsymbol{y}_i^T[\boldsymbol{K}(\boldsymbol{Z}, \boldsymbol{Z}|\boldsymbol{\theta}) + \sigma^2 \boldsymbol{I}]^{-1}\boldsymbol{y}_i + \frac{1}{2}\log[\boldsymbol{K}(\boldsymbol{Z}, \boldsymbol{Z}|\boldsymbol{\theta}) + \sigma^2 \boldsymbol{I}] + \frac{N}{2}\log 2\pi. \end{aligned}$$

The above formulation frames an overview of the Gaussian process modelling and summarizes the overall rationale on how it models and predicts. We will now get to the specifics of each building block of its internal mechanics, in greater detail below, following the formalism expressed by Dablander, 2019 and Yi, 2019.

Modelling Distributions Dependence

One of the key points in the Bayesian approach is the assumption that all the underlying distributions used across the modelling flow, should be **dependent on each other**. The reason is that as outlined earlier, the model is identifying "optimal" structure by the use of the Bayes Theorem ("rule"), which in turn is dependent on previously "seen" or "available" datasets, acting as the "ground truth" in the overall process - the training set - and carries any informational gain along any new "test" point, as well. As such, in the absence of such dependence, no information gain would be retrieved from "prior knowledge" and as such, the Bayesian approach would be of no particular value towards the regression task handling. Gaussian process modelling frames this dependence by using the multivariate Gaussian distribution's marginalization rule, which states that if we marginalize out the underlying variables of a multivariate Gaussian distribution, then the result would also be a Gaussian distribution - namely the Gaussian distribution is considered to be "closed under marginalisation".

Towards the formulation of the marginalization rule in our context, we define the following sub-sets of focus :

1. (available) training set locations : $\boldsymbol{Z}$, with dimensions $N \times p$, consisting of N p-dimensional vectors
2. (potential) test set locations : $\boldsymbol{Z_*}$, with dimensions $d \times p$, consisting of d p-dimensional vectors. Those are the so called "query" points or locations, at which we wish to predict the modelling function's values
3. all "potential" locations set : $\boldsymbol{Z_a}$: a set of infinite length since there are infinite number of potential locations different than the ones of the training and test set in the real line

Getting all these into the multivariate Gaussian distribution framework, we have

$$\begin{bmatrix} \boldsymbol{\phi(Z)} \\ \boldsymbol{\phi(Z_*)} \\ \boldsymbol{\phi(Z_a)} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_Z \\ \mu_{Z_*} \\ \mu_{Z_a} \end{bmatrix}, \begin{bmatrix} \boldsymbol{k(Z,Z)} & \boldsymbol{k(Z,Z_*)} & \boldsymbol{k(Z,Z_a)} \\ \boldsymbol{k(Z,Z_*)} & \boldsymbol{k(Z_*,Z_*)} & \boldsymbol{k(Z_*,Z_a)} \\ \boldsymbol{k(Z,Z_a)} & \boldsymbol{k(Z_*,Z_a)} & \boldsymbol{k(Z_a,Z_a)} \end{bmatrix} \right) \tag{2.46}$$

where for instance :

$$\boldsymbol{k(Z,Z)} = \begin{bmatrix} \boldsymbol{k(z_1,z_1)} & \boldsymbol{k(z_1,z_2)} & \dots & \boldsymbol{k(z_1,z_N)} \\ \boldsymbol{k(z_2,z_N)} & \boldsymbol{k(z_2,z_2)} & \dots & \boldsymbol{k(z_2,z_N)} \\ \dots & & & \\ \boldsymbol{k(z_N,z_1)} & \boldsymbol{k(z_N,z_2)} & \dots & \boldsymbol{k(z_N,z_N)} \end{bmatrix} \tag{2.47}$$

is the $N \times N$ matrix containing all covariances for all potential pairs of locations in the set Z. The same formulation can be applied for the other covariance matrices for locations Z_* and Z_a, respectively, while though the dimensions of the latter would be infinite - which is the main reason why the marginalization rule is required in order to proceed, as well. More specifically, we will show how we can obtain the distribution of $\begin{bmatrix} \boldsymbol{\phi(Z)} \\ \boldsymbol{\phi(Z_*)} \end{bmatrix}$, i.e. removing any information from the infinite dimensional Z_a set.

We cite Blitzstein and Hwang, 2014, and the definition expressed in pages pp. 309-310 for $p = 3$:

<u>Definition</u> : A random $\boldsymbol{X} = (X_1, X_2, X_3)^T$ has a multivariate Gaussian distribution if <u>every linear</u> combination of its components has a (multivariate) Gaussian distribution $X \sim N(\mu, \Sigma)$, if and only if $A\boldsymbol{X} \sim N(A\mu, A\Sigma A^T)$.

As such, for our case of the three random variables of interest, $\boldsymbol{\phi(Z_a)}, \boldsymbol{\phi_Z}$ and $\boldsymbol{\phi(Z_*)}$, and with $\rho_{Z,Z_*} = \rho(\boldsymbol{\phi(Z)}, \boldsymbol{\phi(Z_*)})$ and $\rho_{Z_*,Z_a} = \rho(\boldsymbol{\phi(Z_*)}, \boldsymbol{\phi(Z_a)})$, we can write :

$$\begin{pmatrix} \boldsymbol{\phi(Z)} \\ \boldsymbol{\phi(Z_*)} \\ \boldsymbol{\phi(Z_a)} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_Z \\ \mu_{Z_*} \\ \mu_{Z_a} \end{pmatrix}, \begin{pmatrix} \sigma_Z^2 & & \\ \rho_{Z,Z_*}\sigma_Z\sigma_{Z_*} & \sigma_{Z_*}^2 & \\ \rho_{Z,Z_a}\sigma_Z\sigma_{Z_a} & \rho_{Z_*Z_a}\sigma_{Z_*}\sigma_{Z_a} & \sigma_{Z_a}^2 \end{pmatrix} \right) \tag{2.48}$$

Plugging in matrix $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ in the expression of the definition, we get :

$$\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix}\begin{pmatrix}\boldsymbol{\phi}(\boldsymbol{Z})\\\boldsymbol{\phi}(\boldsymbol{Z}_*)\\\boldsymbol{\phi}(\boldsymbol{Z}_a)\end{pmatrix}$$

$$\sim \mathcal{N}\left(\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix}\begin{pmatrix}\boldsymbol{\mu}_{\boldsymbol{Z}}\\\boldsymbol{\mu}_{\boldsymbol{Z}_*}\\\boldsymbol{\mu}_{\boldsymbol{Z}_a}\end{pmatrix},\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix}\begin{pmatrix}\sigma^2_Z&&\\\rho_{Z,Z_*}\sigma_Z\sigma_{Z_*}&\sigma^2_{Z_*}&\\\rho_{Z,Z_a}\sigma_Z\sigma_{Z_a}&\rho_{Z_*,Z_a}\sigma_{Z_*}\sigma_{Z_a}&\sigma^2_{Z_a}\end{pmatrix}\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix}\right) \tag{2.49}$$

which yields :

$$\begin{pmatrix}\boldsymbol{\phi}(\boldsymbol{Z})\\\boldsymbol{\phi}(\boldsymbol{Z}_*)\end{pmatrix}\sim \mathcal{N}\left(\begin{pmatrix}\boldsymbol{\mu}_{\boldsymbol{Z}}\\\boldsymbol{\mu}_{\boldsymbol{Z}_*}\end{pmatrix},\begin{pmatrix}\sigma^2_Z&\\\rho_{Z,Z_*}\sigma_Z\sigma_{Z_*}&\sigma^2_{Z_*}\end{pmatrix}\right) \tag{2.50}$$

Following the above formulation and also utilizing the connection formula between covariance and correlation of e.g. one of the pairs of the defined random variables of the analysis, $\rho_{z,z_*} = \frac{k(Z,Z_*)}{\sigma_z\sigma_{z_*}}$, we can also represent the previously expressed probability distribution in the format of the corresponding probability density function. This expression is actually the Gaussian Process "prior", representing the joint distributions of the random variables under focus.

$$\begin{aligned} p\left(\begin{bmatrix}\boldsymbol{\phi}(\boldsymbol{Z})\\\boldsymbol{\phi}(\boldsymbol{Z}_*)\end{bmatrix}\right) &= \frac{1}{(2\pi)^{\frac{N+p}{2}}det(\boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}_*))^{\frac{1}{2}}} \\ exp(-\frac{1}{2}&\left(\begin{bmatrix}\boldsymbol{\phi}(\boldsymbol{Z})\\\boldsymbol{\phi}(\boldsymbol{Z}_*)\end{bmatrix}-\begin{bmatrix}\mu_Z\\\mu_{Z_*}\end{bmatrix}\right)^T \boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}_*)^{-1}\left(\begin{bmatrix}\boldsymbol{\phi}(\boldsymbol{Z})\\\boldsymbol{\phi}(\boldsymbol{Z}_*)\end{bmatrix}-\begin{bmatrix}\mu_Z\\\mu_{Z_*}\end{bmatrix}\right)) \end{aligned} \tag{2.51}$$

The same marginal distribution formulation for all other pairs of random variables following multivariate Gaussian distributions could also be obtained.

Following the same rationale, we proceed with the formulation of the second important term of the Gaussian model, namely the "likelihood" and subsequently the "marginal likelihood", as well. The likelihood term is actually expressing the probability of observing a specific set of "targets" values, $\boldsymbol{Y}$, at the training set locations $\boldsymbol{Z}$. The reason for explicitly defining the likelihood in the training set's locations, is that by doing so in any other locations' set, e.g. in the test set, a contradiction would be generated regarding how a model could actually been instructed to "learn" valuable information or obtain "knowledge" from "unseen" or "not available" locations. Under the main bayesian rationale already followed throughout the previously described framework, we again formalise this term by introducing the random variable vector $\boldsymbol{y}(\boldsymbol{Z})$, incorporating this "available" information in the dataset of focus, and we are focusing in expressing its density, in terms of the $\boldsymbol{\phi}(\boldsymbol{Z})$:

$$p(\boldsymbol{y}(\boldsymbol{Z})|\boldsymbol{\phi}(\boldsymbol{Z})) = \mathcal{N}(\boldsymbol{y}(\boldsymbol{Z});\boldsymbol{\phi}(\boldsymbol{Z}),\sigma^2\boldsymbol{I_N}) \tag{2.52}$$

or using the Gaussian linear transformation form by :

$$\boldsymbol{y}(\boldsymbol{Z}) = \boldsymbol{I_N}\boldsymbol{\phi}(\boldsymbol{Z}) + \epsilon, \epsilon \sim \mathcal{N}(0,\sigma^2\boldsymbol{I_N}) \tag{2.53}$$

The likelihood probability density function is formalised as :

$$p(\boldsymbol{y}(\boldsymbol{Z})|\boldsymbol{\phi}(\boldsymbol{Z})) = \frac{1}{(2\pi)^{n/2}det(\boldsymbol{k}(\boldsymbol{Y},\boldsymbol{Z}) + \sigma^2\boldsymbol{I_N})^{(1/2)}}$$
$$exp(-\frac{1}{2}(\boldsymbol{y}(\boldsymbol{Z}) - \boldsymbol{\phi}(\boldsymbol{Z}))^T(\boldsymbol{k}(\boldsymbol{Y},\boldsymbol{Z}) + \sigma^2\boldsymbol{I_N})^{-1}(\boldsymbol{y}(\boldsymbol{Z}) - \boldsymbol{\phi}(\boldsymbol{Z}))) \tag{2.54}$$

Using the multivariate Gaussian marginalization rule towards isolating the distribution for $\boldsymbol{\phi}(\boldsymbol{Z})$, we get that $\boldsymbol{\phi}(\boldsymbol{Z}) \sim \mathcal{N}(\boldsymbol{\mu_Z}, \boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}))$, which combined with the linear transformation rule also yield the distribution of $\boldsymbol{\phi}(\boldsymbol{z})$, i.e. the target expression of the **marginal likelihood** :

$$\begin{aligned} \boldsymbol{y}(\boldsymbol{Z}) &\sim \mathcal{N}(\boldsymbol{I_N}\boldsymbol{\mu_Z}\boldsymbol{I_N}^T, \boldsymbol{k}(\boldsymbol{Y},\boldsymbol{Z}) + \sigma^2\boldsymbol{I_N}) \\ &\sim \mathcal{N}(\boldsymbol{\mu_Z}, \boldsymbol{k}(\boldsymbol{Y},\boldsymbol{Z}) + \sigma^2\boldsymbol{I_N}) \end{aligned} \tag{2.55}$$

$$p(\boldsymbol{y}(\boldsymbol{Z})) = \frac{1}{(2\pi)^{n/2}det(\boldsymbol{k}(\boldsymbol{Y},\boldsymbol{Z}) + \sigma^2\boldsymbol{I_N})^{(1/2)}}$$
$$exp(-\frac{1}{2}(\boldsymbol{y}(\boldsymbol{Z}) - \boldsymbol{\mu_Z})^T(\boldsymbol{k}(\boldsymbol{Y},\boldsymbol{Z}) + \sigma^2\boldsymbol{I_N})^{-1}(\boldsymbol{y}(\boldsymbol{Z}) - \boldsymbol{\mu_Z})) \tag{2.56}$$

Finally, we reached the point of computing the third and final main term of the Gaussian model, the **posterior distribution**, using the :

1. prior distribution: with constituents being the two (latent) random variables, $\boldsymbol{\phi}(\boldsymbol{Z})$ and $\boldsymbol{\phi}(\boldsymbol{Z_*})$ and
2. the likelihood : $p(\boldsymbol{y}(\boldsymbol{Z})|\boldsymbol{\phi}(\boldsymbol{Z}))$

As in the case of the rest of the distributions outlined and expressed in the previous steps, the posterior could be defined for all possible locations of the arbitrary inputs' domain $Z \cup Z_*$, but of particular interest is the ones in the test set. The reason is that the overall target is to make predictions on these "query" locations, i.e. the ones we can actually "observe" and also evaluate the predictions' efficiency upon. Any other "potential" or "non-visible" location has no intuitive value in any way. As such, the target is to compute the $p(\boldsymbol{\phi}(\boldsymbol{Z_*})|\boldsymbol{y}(\boldsymbol{Z}))$, using the already formulated distributions

1. $\boldsymbol{y}(\boldsymbol{Z}) \sim \mathcal{N}(\boldsymbol{\mu_Z}, \boldsymbol{k}(\boldsymbol{Y},\boldsymbol{Z}) + \sigma^2\boldsymbol{I_N})$ and
2. $\boldsymbol{\phi}(\boldsymbol{Z_*}) \sim \mathcal{N}(\boldsymbol{\mu_{z_*}}, \boldsymbol{k}(\boldsymbol{Z_*},\boldsymbol{Z_*}))$, obtained again from the marginalization rule.

As formulated in the quick overview introduction of the current section, we can now make predictions at any new points from the test set, by drawing samples from the posterior distribution $p(\boldsymbol{\phi}|(\boldsymbol{Z},\boldsymbol{Y}))$, with the parameters' set $\boldsymbol{\theta}$ denoting the model's hyperparameters, which will be identified after proper optimization. Assuming a Gaussian distributed noise, and appropriate normalization of the data points, we consider that in order to be able to express the joint distribution of $\boldsymbol{\phi}(\boldsymbol{Z_*})$ and $\boldsymbol{y}(\boldsymbol{Z})$ following the previous logic, we only miss the computation of their covariance, $Cov(\boldsymbol{\phi}(\boldsymbol{Z_*}), y_Z)$. We write :

$$\begin{bmatrix} \boldsymbol{\phi}(\boldsymbol{Z_*}) \\ \boldsymbol{y}(\boldsymbol{Z}) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu_{Z_*}} \\ \boldsymbol{\mu_Z} \end{bmatrix}, \begin{bmatrix} \boldsymbol{k}(\boldsymbol{Z_*},\boldsymbol{Z_*}) & Cov(\boldsymbol{\phi}(\boldsymbol{Z_*}), \boldsymbol{y}(\boldsymbol{Z})) \\ Cov(\boldsymbol{\phi}(\boldsymbol{Z_*}), \boldsymbol{y}(\boldsymbol{Z}))^T & \boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}) + \sigma^2\boldsymbol{I_N} \end{bmatrix}\right) \tag{2.57}$$

with assumed zeroed means of the prior and the covariance term of focus, $Cov(\boldsymbol{\phi}(\boldsymbol{z}_*), \boldsymbol{y}(\boldsymbol{Z}))$ be given by :

$$\begin{aligned}
&Cov(\boldsymbol{\phi}(\boldsymbol{Z}_*), \boldsymbol{y}(\boldsymbol{Z})) = \\
&= \mathbb{E}[(\boldsymbol{\phi}(\boldsymbol{Z}_*) - \mu_{\boldsymbol{Z}_*})(\boldsymbol{y}(\boldsymbol{Z}) - \mu_{\boldsymbol{Z}})] \\
&= \mathbb{E}[(\boldsymbol{\phi}(\boldsymbol{Z}_*) - \mu_{\boldsymbol{Z}_*})(\boldsymbol{\phi}(\boldsymbol{Z}) + \epsilon - \mu_{\boldsymbol{Z}})] \\
&= \mathbb{E}[(\boldsymbol{\phi}(\boldsymbol{Z}_*) - \mu_{\boldsymbol{Z}_*})(\boldsymbol{\phi}(\boldsymbol{Z}) - \mu_{\boldsymbol{Z}}) + (\boldsymbol{\phi}(\boldsymbol{Z}_*) - \mu_{\boldsymbol{Z}_*})\epsilon] \\
&= \mathbb{E}[(\boldsymbol{\phi}(\boldsymbol{Z}_*) - \mu_{\boldsymbol{Z}_*})(\boldsymbol{\phi}(\boldsymbol{Z}) - \mu_{\boldsymbol{Z}})] + \mathbb{E}[(\boldsymbol{\phi}(\boldsymbol{Z}_*) - \mu_{\boldsymbol{Z}_*})\epsilon] \\
&= Cov(\boldsymbol{\phi}(\boldsymbol{Z}_*), \boldsymbol{\phi}(\boldsymbol{Z})) + \mathbb{E}[(\boldsymbol{\phi}(\boldsymbol{Z}_*) - \mu_{\boldsymbol{Z}_*})\epsilon] \\
&= K(\boldsymbol{\phi}(\boldsymbol{Z}_*), \boldsymbol{\phi}(\boldsymbol{Z})) + \mathbb{E}[(\boldsymbol{\phi}(\boldsymbol{Z}_*) - \mu_{\boldsymbol{Z}_*})\epsilon] \\
&= \boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z}) + 0
\end{aligned} \tag{2.58}$$

As such, equation 2.57 becomes :

$$\begin{bmatrix} \boldsymbol{\phi}(\boldsymbol{Z}_*) \\ \boldsymbol{y}(\boldsymbol{Z}) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_{z_*} \\ \mu_{\boldsymbol{Z}} \end{bmatrix}, \begin{bmatrix} \boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z}_*) & \boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z}) \\ \boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z})^T & \boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I}_N \end{bmatrix} \right) \tag{2.59}$$

In order to calculate the target quantity $p(\boldsymbol{\phi}(\boldsymbol{Z}_*)|\boldsymbol{y}(\boldsymbol{Z}))$, we need to use the conditional rule of the multivariate Gaussian distribution. We simplify the covariance matrix in 2.59 as :

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z}_*) & \boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z}) \\ \boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z})^T & \boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I}_N \end{bmatrix} \tag{2.60}$$

$$\begin{aligned}
&p(\boldsymbol{\phi}(\boldsymbol{Z}_*) \mid \boldsymbol{y}(\boldsymbol{Z})) = \frac{p(\boldsymbol{\phi}(\boldsymbol{Z}_*), \boldsymbol{y}(\boldsymbol{Z}))}{p(\boldsymbol{y}(\boldsymbol{Z}))} \\
&= \frac{(2\pi)^{-(N+d)/2} |\boldsymbol{C}|^{-1/2} \exp\left(-\frac{1}{2} \left[\begin{pmatrix} \boldsymbol{\phi}(\boldsymbol{Z}_*) \\ \boldsymbol{y}(\boldsymbol{Z}) \end{pmatrix}^T \boldsymbol{C}^{-1} \begin{pmatrix} \boldsymbol{\phi}(\boldsymbol{Z}_*) \\ \boldsymbol{y}(\boldsymbol{Z}) \end{pmatrix} \right] \right)}{(2\pi)^{-d/2} |(\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I}_N)|^{-1/2} \exp\left(-\frac{1}{2} \left[\boldsymbol{y}(\boldsymbol{Z})^T (\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I}_N)^{-1} \boldsymbol{y}(\boldsymbol{Z}) \right] \right)} \\
&= (2\pi)^{-N/2} |\boldsymbol{C}|^{-1/2} |(\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I}_N)|^{1/2} \\
&\exp(-\frac{1}{2} \left[\begin{pmatrix} \boldsymbol{\phi}(\boldsymbol{Z}_*) \\ \boldsymbol{y}(\boldsymbol{Z}) \end{pmatrix}^T \begin{pmatrix} (\boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z}_*)) & (\boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z})) \\ (\boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z})^T) & (\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I}_N) \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\phi}(\boldsymbol{Z}_*) \\ \boldsymbol{y}(\boldsymbol{Z}) \end{pmatrix} \right] \\
&+ \boldsymbol{y}(\boldsymbol{Z})^T (\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I}_N)^{-1} \boldsymbol{y}(\boldsymbol{Z})) \\
&= (2\pi)^{-n/2} |\boldsymbol{C}|^{-1/2} |(\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I}_N)|^{1/2} \\
&\exp(-\frac{1}{2} [\begin{pmatrix} \boldsymbol{\phi}(\boldsymbol{Z}_*) \\ \boldsymbol{y}(\boldsymbol{Z}) \end{pmatrix}^T \begin{pmatrix} (\boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z}_*)) & (\boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z})) \\ (\boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z})^T) & (\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I}_N) \end{pmatrix}^{-1} \\
&\begin{pmatrix} \boldsymbol{\phi}(\boldsymbol{Z}_*) \\ \boldsymbol{y}(\boldsymbol{Z}) \end{pmatrix} - \boldsymbol{y}(\boldsymbol{Z})^T (\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I}_N)^{-1} \boldsymbol{y}(\boldsymbol{Z})]) \ .
\end{aligned} \tag{2.61}$$

Re-notating $\boldsymbol{C}^{-1}$, as :

$$\boldsymbol{C}^{-1} = \boldsymbol{\eta} = \begin{pmatrix} \boldsymbol{\eta_{11}} & \boldsymbol{\eta_{12}} \\ \boldsymbol{\eta_{21}} & \boldsymbol{\eta_{22}} \end{pmatrix} \tag{2.62}$$

with :

$$\boldsymbol{\eta_{11}} = \Big(\boldsymbol{k}(\boldsymbol{Z_*}, \boldsymbol{Z_*}) - \boldsymbol{k}(\boldsymbol{Z_*}, \boldsymbol{Z})(\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I_N})^{-1} \boldsymbol{k}(\boldsymbol{Z_*}, \boldsymbol{Z})^T \Big)^{-1} \tag{2.63}$$

$$\begin{aligned} \boldsymbol{\eta_{12}} = &- \Big(\boldsymbol{k}(\boldsymbol{Z_*}, \boldsymbol{Z_*}) - \boldsymbol{k}(\boldsymbol{Z_*}, \boldsymbol{Z})(\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I_N})^{-1} \boldsymbol{k}(\boldsymbol{Z_*}, \boldsymbol{Z}) \Big)^{-1} \\ &\boldsymbol{k}(\boldsymbol{Z_*}, \boldsymbol{Z})(\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I_N})^{-1} \end{aligned} \tag{2.64}$$

$$\begin{aligned} \boldsymbol{\eta_{21}} &= -(\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I_N})^{-1} \boldsymbol{k}(\boldsymbol{Z_*}, \boldsymbol{Z})^T \\ &\Big(\boldsymbol{k}(\boldsymbol{Z_*}, \boldsymbol{Z_*}) - \boldsymbol{k}(\boldsymbol{Z_*}, \boldsymbol{Z})(\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I_N})^{-1} \boldsymbol{k}(\boldsymbol{Z_*}, \boldsymbol{Z}) \Big)^{-1} \end{aligned} \tag{2.65}$$

$$\begin{aligned} \boldsymbol{\eta_{22}} &= (\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I_N})^{-1} + (\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I_N})^{-1} \boldsymbol{k}(\boldsymbol{Z_*}, \boldsymbol{Z})^T \\ &\Big(\boldsymbol{k}(\boldsymbol{Z_*}, \boldsymbol{Z_*}) - \boldsymbol{k}(\boldsymbol{Z_*}, \boldsymbol{Z})(\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I_N})^{-1} \boldsymbol{k}(\boldsymbol{Z_*}, \boldsymbol{Z}) \Big)^{-1} \\ &\boldsymbol{k}(\boldsymbol{Z_*}, \boldsymbol{Z})(\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I_N})^{-1} \end{aligned} \tag{2.66}$$

we proceed, as follows :

$$
\begin{aligned}
&p(\boldsymbol{\phi}(\boldsymbol{Z}_*) \mid \boldsymbol{y}(\boldsymbol{Z})) = \\
&(2\pi)^{-N/2}|\boldsymbol{C}|^{-1/2}|(\boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}) + \sigma^2\boldsymbol{I}_N)|^{1/2} \\
&\exp\left(-\frac{1}{2}\left[\begin{pmatrix}\boldsymbol{\phi}(\boldsymbol{Z}_*)\\ \boldsymbol{y}(\boldsymbol{Z})\end{pmatrix}^T \begin{pmatrix}\boldsymbol{\eta}_{11} & \boldsymbol{\eta}_{12}\\ \boldsymbol{\eta}_{21} & \boldsymbol{\eta}_{22}\end{pmatrix} \begin{pmatrix}\boldsymbol{\phi}(\boldsymbol{Z}_*)\\ \boldsymbol{y}(\boldsymbol{Z})\end{pmatrix} - \boldsymbol{y}(\boldsymbol{Z})^T(\boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}) + \sigma^2\boldsymbol{I}_N)^{-1}\boldsymbol{y}(\boldsymbol{Z})\right]\right) \\
&= (2\pi)^{-N/2}|\boldsymbol{C}|^{-1/2}|(\boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}) + \sigma^2\boldsymbol{I}_N)|^{1/2} \\
&\exp\left(-\frac{1}{2}\left[\begin{pmatrix}\boldsymbol{\phi}(\boldsymbol{Z}_*)^T\boldsymbol{\eta}_{11} + \boldsymbol{y}(\boldsymbol{Z})^T\boldsymbol{\eta}_{21}\\ \boldsymbol{\phi}(\boldsymbol{Z}_*)^T\boldsymbol{\eta}_{12} + \boldsymbol{y}(\boldsymbol{Z})^T\boldsymbol{\eta}_{22}\end{pmatrix}^T \begin{pmatrix}\boldsymbol{\phi}(\boldsymbol{Z}_*)\\ \boldsymbol{y}(\boldsymbol{Z})\end{pmatrix} - \boldsymbol{y}(\boldsymbol{Z})^T(\boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}) + \sigma^2\boldsymbol{I}_N)^{-1}\boldsymbol{y}(\boldsymbol{Z})\right]\right) \\
&= (2\pi)^{-N/2}|\boldsymbol{C}|^{-1/2}|(\boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}) + \sigma^2\boldsymbol{I}_N)|^{1/2} \\
&\exp(-\frac{1}{2}[\boldsymbol{\phi}(\boldsymbol{Z}_*)^T\boldsymbol{\eta}_{11}\boldsymbol{\phi}(\boldsymbol{Z}_*) + \boldsymbol{y}(\boldsymbol{Z})^T\boldsymbol{\eta}_{21}\boldsymbol{\phi}(\boldsymbol{Z}_*) + \\
&\boldsymbol{\phi}(\boldsymbol{Z}_*)^T\boldsymbol{\eta}_{12}\boldsymbol{y}(\boldsymbol{Z}) + \boldsymbol{y}(\boldsymbol{Z})^T\boldsymbol{\eta}_{22}\boldsymbol{y}(\boldsymbol{Z}) - \boldsymbol{y}(\boldsymbol{Z})^T(\boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}) + \sigma^2\boldsymbol{I}_N)^{-1}\boldsymbol{y}(\boldsymbol{Z})]) \\
&= (2\pi)^{-N/2}|\boldsymbol{C}|^{-1/2}|(\boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}) + \sigma^2\boldsymbol{I}_N)|^{1/2} \\
&\exp(-\frac{1}{2}[\boldsymbol{\phi}(\boldsymbol{Z}_*)^T\boldsymbol{\eta}_{11}\boldsymbol{\phi}(\boldsymbol{Z}_*) + 2\boldsymbol{\phi}(\boldsymbol{Z}_*)^T\boldsymbol{\eta}_{12}\boldsymbol{y}(\boldsymbol{Z}) + \boldsymbol{y}(\boldsymbol{Z})^T \\
&\left(\boldsymbol{\eta}_{22} - (\boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}) + \sigma^2\boldsymbol{I}_N)^{-1}\right)\boldsymbol{y}(\boldsymbol{Z})])
\end{aligned}
\tag{2.67}
$$

which since :

$$
\boldsymbol{\eta}_{22} - (\boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}) + \sigma^2\boldsymbol{I}_N)^{-1} = \boldsymbol{\eta}_{21}{\boldsymbol{\eta}_{11}}^{-1}\boldsymbol{\eta}_{12} \tag{2.68}
$$

and

$$
|\boldsymbol{C}| = |\boldsymbol{k}(\boldsymbol{Z}_*,\boldsymbol{Z}_*) - \boldsymbol{k}(\boldsymbol{Z}_*,\boldsymbol{Z})(\boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}) + \sigma^2\boldsymbol{I}_N)^{-1}\boldsymbol{k}(\boldsymbol{Z}_*,\boldsymbol{Z}_*)^T| \times |(\boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}) + \sigma^2\boldsymbol{I}_N)| \tag{2.69}
$$

we get the final expression as :

$$
\begin{aligned}
&p(\boldsymbol{\phi}(\boldsymbol{Z}_*) \mid \boldsymbol{y}(\boldsymbol{Z})) = (2\pi)^{-N/2}|\boldsymbol{C}|^{-1/2}|(\boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}) + \sigma^2\boldsymbol{I}_N)|^{1/2} \\
&\exp\left(-\frac{1}{2}\left(\boldsymbol{\phi}(\boldsymbol{Z}_*) + {\boldsymbol{\eta}_{11}}^{-1}\boldsymbol{\eta}_{12}\boldsymbol{y}(\boldsymbol{Z})\right)^T\boldsymbol{\eta}_{11}\left(\boldsymbol{\phi}(\boldsymbol{Z}_*) + {\boldsymbol{\eta}_{11}}^{-1}\boldsymbol{\eta}_{12}\boldsymbol{y}(\boldsymbol{Z})\right)\right) \\
&= (2\pi)^{-N/2}|\boldsymbol{k}(\boldsymbol{Z}_*,\boldsymbol{Z}_*) - \boldsymbol{k}(\boldsymbol{Z}_*,\boldsymbol{Z})(\boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}) + \sigma^2\boldsymbol{I}_N)^{-1}\boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}_*)|^{-1/2} \\
&\exp(-\frac{1}{2}\left(\boldsymbol{\phi}(\boldsymbol{Z}_*) - \boldsymbol{k}(\boldsymbol{Z}_*,\boldsymbol{Z})(\boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}) + \sigma^2\boldsymbol{I}_N)^{-1}\boldsymbol{y}(\boldsymbol{Z})\right)^T \\
&(\boldsymbol{k}(\boldsymbol{Z}_*,\boldsymbol{Z}_*) - \boldsymbol{k}(\boldsymbol{Z}_*,\boldsymbol{Z})(\boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}) + \sigma^2\boldsymbol{I}_N)^{-1}\boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}_*))^{-1} \\
&(\boldsymbol{\phi}(\boldsymbol{Z}_*) - \boldsymbol{k}(\boldsymbol{Z}_*,\boldsymbol{Z})(\boldsymbol{k}(\boldsymbol{Z},\boldsymbol{Z}) + \sigma^2\boldsymbol{I}_N)^{-1}\boldsymbol{y}(\boldsymbol{Z})))
\end{aligned}
\tag{2.70}
$$

which is actually a conditional distribution with conditional mean (under the assumption of zero $\boldsymbol{\mu}(\boldsymbol{Z}_*)$)

$$\boldsymbol{\mu}_* = \boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z})(\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I_N})^{-1} \boldsymbol{y}(\boldsymbol{Z}) \tag{2.71}$$

and conditional variance

$$\boldsymbol{\sigma_*}^2 = \boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z}_*) - \boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z})(\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma^2 \boldsymbol{I_N})^{-1} \boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}_*)^T \tag{2.72}$$

Key note here is that the conditional variance is indeed always non-negative. By assuming a kernel function of the form :

$$k(\boldsymbol{z_i}, \boldsymbol{z_j}) = \sigma_k^2 exp(-\frac{(\boldsymbol{z_i} - \boldsymbol{z_j})^2}{2l^2}) \tag{2.73}$$

for two arbitrary arguments $z_i, z_i \in \boldsymbol{Z_a}$, where σ_k and l are optimizable hyperparameters, i.e. part of the hyperparameters' set $\boldsymbol{\theta}$: the $\sigma_k = \theta_1$ is usually called the "signal variance" and $l = \theta_2$ the length-scale, while also the third hyperparameter "noise variance" $\sigma = \theta_3$ is set to zero for simplicity of the following calculations :

$$\begin{aligned}
\boldsymbol{\sigma_*}^2 &= \boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z}_*) - \boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z})(\boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}) + \sigma_k^2 \boldsymbol{I_N})^{-1} \boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z}_*)^T \\
&= \boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z}_*) - \boldsymbol{k}(\boldsymbol{Z}, \boldsymbol{Z})^{-1} \boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z})^2 = \\
&= \boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z}_*) - \frac{1}{\sigma_k^2} \boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z})^2 \\
&= \sigma_k^2 - \frac{1}{\sigma_k^2} \boldsymbol{k}(\boldsymbol{Z}_*, \boldsymbol{Z})^2 \\
&= \sigma_k^2 - \frac{1}{\sigma_k^2} \sigma_k^4 exp(-\frac{2(\boldsymbol{Z}_* - \boldsymbol{Z})^2}{2l^2}) \\
&= \sigma_k^2 (1 - exp(-\frac{2(\boldsymbol{Z}_* - \boldsymbol{Z})^2}{2l^2})) \geq 0
\end{aligned} \tag{2.74}$$

The final step is to calculate the posterior distribution for the "observed" random variable $\boldsymbol{y}(\boldsymbol{Z}_*) \mid \boldsymbol{y}(\boldsymbol{Z}) = \boldsymbol{I_d}(\boldsymbol{\phi}(\boldsymbol{Z}_*) \mid \boldsymbol{y}(\boldsymbol{Z})) + \boldsymbol{\epsilon}_*$, with $\epsilon_* \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I_d})$. By applying again the multivariate Gaussian transformation rule and also by knowing the distribution of $\boldsymbol{\phi}(\boldsymbol{Z}_*) \mid \boldsymbol{y}(\boldsymbol{Z})$, we get :

$$p(\boldsymbol{y}(\boldsymbol{Z}_*) \mid \boldsymbol{y}(\boldsymbol{Z})) = \mathcal{N}(\boldsymbol{I_d} \boldsymbol{\mu}_*, \boldsymbol{I_d} \sigma^2 \boldsymbol{I_d}^T + \sigma_k^2 \boldsymbol{I_d}) = \mathcal{N}(\boldsymbol{\mu}_*, \boldsymbol{\sigma_*}^2 + \sigma^2 \boldsymbol{I_d}) \tag{2.75}$$

Optimization and parameter learning

At this point, all necessary ingredients are in hand towards constructing the regression mechanism from specific training data points, but also for making predictions at any potential point from the test set. However, we still have to identify the optimal set of hyperparameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) = (\sigma_k, l, \sigma)$, which best "explains" the "observed" data $\boldsymbol{Y}$, using information from the also "observed" training set $\boldsymbol{Z}$. Furthermore, the target objective function used towards the optimal parameters' set identification, should be dependent on these three quantities only : $\boldsymbol{Y}$, $\boldsymbol{Z}$ and $\boldsymbol{\theta}$. As such, we can see that the most suitable term for these requirements, across all terms formulated at this section's context, is the marginal likelihood one, by plugging in the observed values Y, i.e. by setting Y in equation 2.56 and also by taking the logarithm of the corresponding expression to cancel out the exponential (Goodfellow, Bengio, and Courville, 2016) :

$$\begin{aligned} &log(p(\boldsymbol{y}(\boldsymbol{Z}))) = log(\frac{1}{(2\pi)^{n/2}det(\boldsymbol{k}(\boldsymbol{Y},\boldsymbol{Z}) + \sigma^2\boldsymbol{I_N})^{(1/2)}} \\ &exp(-\frac{1}{2}(\boldsymbol{y}(\boldsymbol{Z}) - \boldsymbol{\mu_Z})^T(\boldsymbol{k}(\boldsymbol{Y},\boldsymbol{Z}) + \sigma^2\boldsymbol{I_N})^{-1}(\boldsymbol{y}(\boldsymbol{Z}) - \boldsymbol{\mu_Z}))) \\ &= \frac{1}{2}log(det(\boldsymbol{k}(\boldsymbol{Y},\boldsymbol{Z}) + \sigma^2\boldsymbol{I_N})) \\ &- \frac{1}{2}(\boldsymbol{y}(\boldsymbol{Z}) - \boldsymbol{\mu_Z})^T(\boldsymbol{k}(\boldsymbol{Y},\boldsymbol{Z}) + \sigma^2\boldsymbol{I_N})^{-1}(\boldsymbol{y}(\boldsymbol{Z}) - \boldsymbol{\mu_Z}) \\ &- \frac{N}{2}log(2\pi) \end{aligned} \tag{2.76}$$

2.2 Nonlinear time series Analysis

Nonlinear time-series analysis in the reconstructed state space is a powerful and useful idea, but it does have some practical limitations. These limitations are by no means fatal, but one has to be aware of them in order to report correct results. Theoretically, the method of delays (MOD) for setting the coordinates in the embedded phase-space is a valid and reliable approach only for an assumed and 'ideal' situation: the series of observations must be infinitely long and noise-free taken from a single dynamical system. This assumption in practice, poses a number of problems:

1. the issue of nonstationarity : embedding a time series recorded from a system which is under often bifurcations, for example, will generate a topological stew of those different attractors. Therefore if we want to compute the "Invariants" from such a topological structure, we must take into account that the reconstructed phase space may not be the appropriate one, since it will not accurately capture any of the 'true' dynamical regimes of the original phase space. However, the introduction of manifold learning (ML) approaches for learning ('embedding') the high-dimensional time series to a low dimensional phase-space, can alleviate the above problem of topological stew of various attractors formed due to non-stationarity and other factors. Traditionally, when using MOD for embedding, the above problem is faced by e.g., repeating the analysis on different sections of the time series (subsequences of the data) and examining if the results change. A powerful 'visual' tool, the recurrence plots Eckmann, Kamphorst, and Ruelle, 1987 can be also used in these situations, that allows us to detect quickly if different sections of the time series exhibit different dynamical patterns. Examining different parts of the time series has a number of other uses besides detecting nonstationarity.

2. The length of data: an important parameter in the reconstruction of phase space, can be examined too. Embedding theorems require an infinite amount of data, but this strict requirement that makes the MOD useless in practice has been alleviated and depending on the time series at hand, looser bounds have

since been established Tsonis, Elsner, and Georgakakos, 1993,Smith, 1988,Eckmann and Ruelle, 1992. Therefore, in choosing MOD, one should to know and be careful about data limits. For example the attempt to compute, say, the largest Lyapunov exponent of a 5-dimensional system from a single time series of say 500 points, will result in a wrong and unreliable value of this invariant measure, which quantifies the system's stability. Therefore if someone wants to use the Lyapunov exponent as a diagnostic tool, he must be careful when repeating analyses on subsets of the his time series, since the changes in the values of the exponent can simply be occur due to short data lengths.

3. Dimension: it is a crucial practical issue, not just because it is not known a-priori but also because can be a challenge to estimate. The dimension of the underlying dynamical system scales -often quite badly- with the length of the time series, a fact that adversely affects the success of nonlinear time series methods. This is an even more serious problem in analyzing time series of spatially extended systems, which have very high-dimensional or even infinite state space and their dynamics is spatio-temporal. In such cases the MOD fails to reconstruct the full attractor. To 'solve' the problem,Bär, Hegger, and Kantz, 1999,Parlitz and Merkwirth, 2000, have suggested exploiting homogeneity of the system. Now if the dynamics is translationally invariant, local dynamics can be reconstructed and used for predictions. But detecting homogeneity in the phase space and translationally invariant dynamics are only some of the main capabilities of the manifold learning approaches, an observation that makes the need for embedding via ML more pronounced.

4. Noise effects: these also scale with dimension, and this is because in an m-dimensional embedding, any noisy time-series point will affect m of the points. Noise detection and reduction via filtering strategies, can alleviate noise problems. Also analysis of different sections can be employed to examine whether the length of the data is adequate to perform the analysis.

5. Method Of Delays embedding requires the time series data to be evenly sampled in time. In case they are not, constructing the delay vector is impossible without interpolation, but this will introduce spurious dynamics. This problem however is not so intense in analyzing discrete time series data like inter-spike data. Sauer, 1995 has analyzed the interspike interval embedding in chaotic signals. If the spikes can be considered to be the result of an integrate-and-fire process, then their spacing is an effective proxy for the integral of the corresponding internal variable, and that is a wholly justifiable quantity to embed. Even in the absent of integrate-and-fire dynamics (main pattern in neurons firing), we can interpret interspike intervals as a specific Poinar´e map, therefore justifying their embedding.

6. Using the same delay time τ : even though in practical applications keeping the same τ is quite practical, doing so in between successive elements of a delay vector may not be optimal. If we want to capture dynamics of evolving in different times scales, by implementing delay vectors of the form $y(t), y(t-\tau_1), y(t-\tau_1-\tau_2), \ldots, y(t-\tau_1-\tau_2-\cdots-\tau_{m-1})$, where τ_i is non-negative, we can introduce more time scales into the reconstruction, a very useful approach for many cases Pecora et al., 2007,Holstein and Kantz, 2009. The above strategy can also be used in tackling signals from multi-scale dynamics. In such cases, a fixed τ may be too large to resolve the short scales,

while is too small to resolve the long scales. For example in the analysis of EEG signals, we may have, say 30, channels (electrodes) located in different brain (scalp) regions to measure the brain activation during a task. The different regions may generate dynamics of different scale so using the same delay time to embed those signals is not an appropriate choice. Another example is the analysis of ECG (electrocardiogram): in embedding an ECG signal by using standard delay vectors, we can either unfold the QRS complex (corresponding to some time scale) or represent the t-wave as a loop (another time scale dynamics), but not both dynamics. However, when we use ML embedding we do not risk to lose dynamics of different time scales, because the ML tools can 'learn' by construction such patterns, a fact that again points out the advantage of using ML embedding techniques.

We make extensive use of non-linear methods in analyzing time series, based on chaos theory, i.e. the theory of deterministic systems with apparently random evolution due to sensitive dependence on small changes in the initial conditions Kugiumtzis, 1996. In terms of chaos theory, geometrical objects formed by the system trajectories (so-called attractors) are characterized by fractal dimension. For attractors from regular deterministic systems, e.g. limit cycles or tori, the fractal dimension is equal to their topological dimension, but for attractors from chaotic systems (strange attractors) it is typically a non-integer that indicates their fractal structure. Among various measures of the fractal dimension the most common is the correlation dimension due to its computational simplicity Kugiumtzis, 1996, Theiler, 1990.

2.2.1 Reconstruction of the state space

Nonlinear dynamical systems exhibit diverse complex dynamic behavior. However, when the systems come form real world, the application of nonlinear TS methods give rise to critical problems, regarding the reliability of making inferences, from one-dimensional scalar time series, about high-dimensional systems. Solution to this problem are provided by the papers of Packard et al., 1980 and Takens, 1981.

Given a time series, an attractor can be embedded in a multidimensional state space. It is known that under certain conditions the reconstructed attractor preserves the topology of the original attractor of the system that generated the data,Takens, 1981, Sauer, Yorke, and Casdagli, 1991a. Working with particular time series data from high dimensional systems (e.g. Markets), the expression "the original attractor" refers to the interactivity of the global variables of the system. Carefully chosen reconstruction schemes are needed in order to maintain the equivalence of the two attractors.

We know from the theory of dynamical systems that a flow on a topological space X is a continuous function $F : R \times X \to X$, satisfying the following conditions :

$$\begin{cases} F(0,x) = x \\ F(t_1+t_2,x) = F(t_1,F(t_2)), \quad \forall, \quad t_1,t_2 \in R, \quad x \in X \end{cases} \tag{2.77}$$

We say that a flow on a topological space defines a continuous-time dynamical system. Given a flow, let F_x is the trajectory $F_x : R \to X$, given by $F_x(t) = F(t,x)$, and by F^t we mean the homeomorphism $F^t : X \to X$ (or diffeomorphism in the smooth case) given by $F^t(x) = F(t,x)$. Of course, we have $F^t(x) = F_x(t)$. Given a smooth vector field H on a smooth manifold on X, we set the differential equation $\frac{dx}{dt} = H(x)$.

Let's take a dynamical system X, F consisting of the d-dimensional state space $X \subset R^d$ and the flow $F : R \times X \to X$, which maps a point in the state space at a given time to the next point on the trajectory at the next time instant. A scalar time series $s_t \subset R^1$, can be obtained via a measurement function $h : X \to R$, which maps discrete locations $x_t \in X$ of the d-dimensional flow to discrete, time-ordered scalar values (i.e. time series) such that $s_t = h(x_t)$, where $x_t = F^t(x_0)$ and x_0 is the initial state of the system at time $t = 0$. In this setting, we can define a delay coordinates map Parlitz and Merkwirth, 2000, $G : X \to R^m, \quad m < d$, such that $G(x_t) = y_t$, where $y_t = (s_t, s_{t+\tau}, s_{t+2\tau}, \ldots, s_{t+(m-1)\tau}), \quad \tau \in R$.

In the above framework, G maps a point x_t on the flow to a point y of the reconstructed phase space in R^m, via using two very important parameters in 'embeddology' parlance, the 'time delay' τ, and the 'embedding dimension' m. By a landmark theorem of Takens, 1981, it is proved that G has the generic property that it is an embedding of X, F in R^d when the condition $m \geq 2d + 1$ is satisfied. The fact that it is an embedding has the consequence that $G : X \to G(X) \subset R^m$ is a diffeomorphism, while the 'generic' term indicates, that the subset of pairs of parameters (h,τ) that guarantee an embedding constitute an open and dense subset in the set of all pairs of (h,τ). Taken's theorem is, in principle, a path-breaking result, this does not work in practice in the analysis of actual high dimensional real-world systems. This is because the embedding dimension has to be greater than twice the dimension of the phase space, which means that the length of the scalar time series have to be extremely large in order to facilitate an embedding in such a high dimension. A solution in this problem is provided by the work of Sauer, Yorke, and Casdagli, 1991b, widely known as the paper on 'embedology', according to which the condition $m \geq 2d + 1$ is further constrained so that now, the parameter m is needed to be greater only than twice the box-counting dimension of the attractor $A \subset X$, ensuring in this way that it is an embedding of X, F . Embeddings of such forms are not 'generic' but rather "prevalent" meaning that 'almost all' pairs of (h,τ) guarantee the formation of an embedding. The construction of the work of Sauer et al., has a tremendous practical consequence, since a plethora of nonlinear dynamical systems are, in fact, dissipative systems exhibiting a dimensionality of their drastically smaller than the phase space X. This facilitates a lot the work of embedding since it is possible to analyze and model the equilibrium dynamics of a high dimensional system X,F within a smaller phase space (since all initial conditions would ultimately settle to the dynamics on A of much smaller volume) reconstructed by using a much smaller dimensional embedding G as given above.

2.2.2 Embedding dimension

Kennel, Brown, and Abarbanel, 1992 proposed the method of false nearest neighbors (FNN) as a tool to determine the 'optimum' embedding dimension, given a time delay τ, chosen as described earlier. The base of this approach is the geometric consideration that given an embedding y_i^m in dimension m, the differentiation between 'true' and 'false' neighbors of points on the reconstructed trajectory, is possible. A crucial step in implementing this method, is the definition of the concept 'neighborhood', in a mathematically accepted way, so we can detect the neighbors of all points on the trajectory R^m. We proceed with detecting the false neighbors, i.e. those neighbors that are not neighbors in $m + 1$ dimensions, i.e., when we consider the trajectory $y_i^{m+1} \in R^{m+1}$. The false neighbors are neighbors in the lower dimensional embedding only due to the fact that the attractor does not 'live' in a space large enough to accommodate all its trajectories, or it is not properly unfolded. In

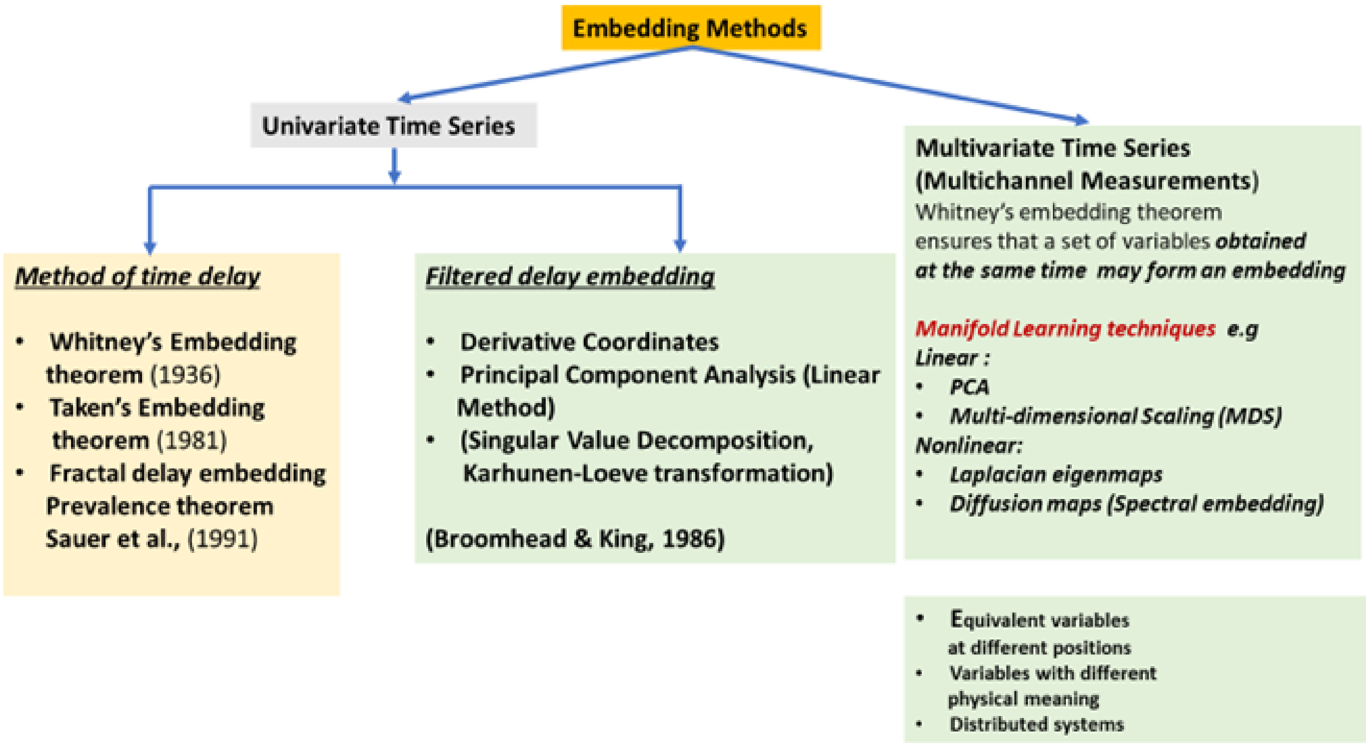

FIGURE 2.2: Embedding Methods

this restricted space, we are actually looking at a projection of the attractor rather than the attractor itself. Consider the example of a 2-D limit cycle trajectory on a circle: opposite points located almost on the same vertical line will be mistakenly seen as neighbors, in case the same dynamic is projected on to the horizontal axis, i.e., the 1-D real line R^1. When the attractor is properly expanded, however, the number of false neighbors would go to zero. The notion of FNN, in practice, is implemented as follows :

$$FNN(r) = \frac{\sum_{i=1}^{N-m\tau} \Theta\left(\frac{\|y_i^{m+1} - y_{k(i,m)}^{m+1}\|}{\|y_i^m - y_{k,(i,m)}^m\|} \parallel -r \right) \Theta(\frac{\sigma}{r} - \parallel y_i^m - y_{k(i,m)}^m) \parallel}{\sum_{i=1}^{N-m\tau} \Theta(\frac{\sigma}{r} - \parallel y_i^m - y_{k(i,m)}^m \parallel)} \tag{2.78}$$

where $\Theta(\cdot)$ indicates the Heaviside function (1 for all positive arguments and zero otherwise), $\parallel \cdot \parallel$ symbolizes the maximum norm and $k(i,m)$ is the index of the point closest to y_i^m placed in the m-dimensional embedding, based on maximum norm. The first term in the numerator counts all those cases when the distance between a point and its closest neighbor increases by more than a factor of r in going from m dimensions to $m+1$ dimensions. In order to avoid counting those cases where the points are already far apart in m dimensions, a weight must be applied, used also as a normalization factor, and this is accounted for by the second term in the numerator. The second term is also used in order to ensure that we count only the points for which the closest neighbor in m dimensions is closer than σ/r (σ the standard deviation of the data). Therefore, the 'optimum' embedding dimension me corresponds to the smallest value of m for which the fraction $FNN(m)$ is zero.

2.2.3 Method of Delays

The sampled time series or signal is denoted by $\{s_i\}_{i=1}^{N} = \{s(i\tau_s)\}_{i=1}^{N}$, where τ_s is the sampling time and N is the length of the time series. The reconstructed state vector with MOD is formed directly from the scalar measurements : $s_i = [s_i, s_{i+\tau}, \ldots, s_{i+(m-1)\tau}]^{\tau}$, where τ is the delay time, given as a multiple integer of the sampling time τ and m is the embedding dimension of the reconstructed space. The parameters τ and m give the time window τw of length $(m-1)\tau$. The $N-(m-1)\times m$ matrix $s_i = [s_1, s_2, \ldots, s^T_{N-(m-1)\tau]}$ is then the constructed trajectory matrix. The delay time is usually estimated as the value giving the least correlation among the components of s_i. Linear de-correlation is obtained by choosing the first zero of the autocorrelation function $R(\tau)$, $R(\tau_0) = 0$ or the correlation time τc corresponding to $R(\tau) = 1/e$, while general de-correlation is determined as the first local minimum of the mutual information function $I(\tau)(minI(\tau) = I(\tau m))$, Fraser and Swinney, 1986. However, these two functions often provide very large estimates of τ for a given time series under analysis. This may result in a poor description of the details of the geometric shape of the attractor.

For the embedding dimension m, a lower limit for the reconstruction to be valid is given by Takens' theorem $m \geq 2d+1$, Takens, 1981. Takens originally proposed d to be the topological dimension, but recent theoretical results relax this criterion assigning d to the fractal dimension of the underlying attractor, Sauer, Yorke, and Casdagli, 1991a. In practice, however, lower values for m are often sufficient to reconstruct the attractor successfully, and one typically searches for the minimum embedding dimension m^*. A method often used to estimate m^* is the False Nearest Neighbors (FNN), Kennel, Brown, and Abarbanel, 1992. This method applies a geometrical criterion – based on the behavior of spatially near neighbors – to the attractor embedded in successively higher dimensions until a limit m^* is reached where the criterion is fulfilled. The implementation of this method to time series data does not give unique m^* but shows a dependence of m^* on τ. In the correlation dimension estimation with MOD reconstruction, only τ is the critical parameter since m is increased in order to investigate if there is a saturation of the scaling of $logC(r)$ vs. $log(r)$.

2.2.4 Singular Spectrum Approach

Using SSA for the reconstruction, the state vector $s_i = [s_i, s_{i+\tau}, \ldots, s_{i+(p-1)}]^T$ is formed for p large and $\tau = 1$, where the mean value of the time series is first subtracted from s_i. The trajectory in R^p is then the $(N-p+1)\times p$ matrix $S = [s_1, s_2, \ldots, s_{(N-p+1)}]^T$. A new basis for R^p is formed by the p singular vectors of S computed with the Singular Value Decomposition (SVD), Broomhead and King, 1986. The trajectory is then projected onto the m first singular vectors ranked according to the variance they explain, which correspond to the m largest singular values σ_i of S. The final $(N-p+1)\times m$ trajectory matrix is $Y = [s_1^T C s_2^T C, \ldots, s^T_{(N-p+1)} C]^T$, where the $p \times m$ matrix C has the m first singular vectors as columns. The parameters of this method are the initial embedding dimension p and the final embedding dimension m. The initial embedding dimension p determines the time window length $\tau w = (p-1)$. For the estimation of m, the cut-off of the spectrum of singular values $(\sigma_1 > \cdots > \sigma m^* >> \sigma m^* + 1 > \ldots \sigma_p)$ has been proposed, but for signals from nonlinear systems such a cut-off is not guaranteed.

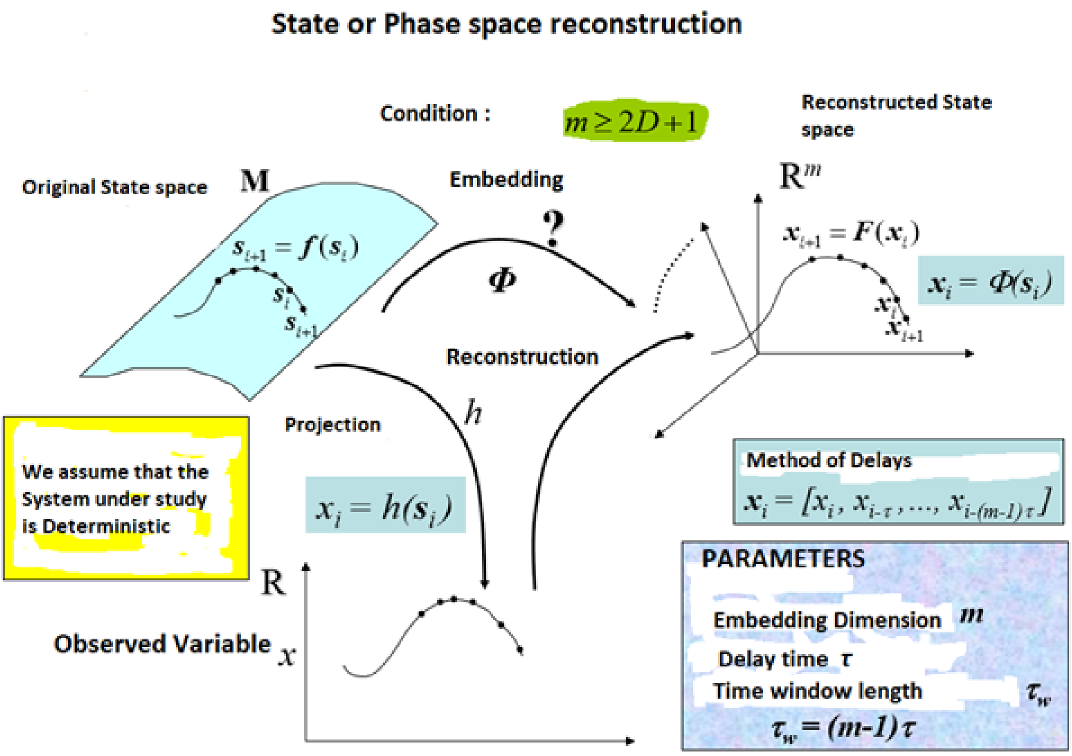

FIGURE 2.3: Phase Space Reconstruction

2.2.5 Finding the best time delay and embedding dimension parameters

Determining the time delay and the embedding dimension is considered as one of the most important steps in nonlinear time series modeling and prediction. A number of methods have been developed in determining the time delay and the minimum embedding dimension since the early beginning of nonlinear time series study. Here we will describe and apply several of them to the foreign exchange time series data sets.

The first step in phase space reconstruction is to choose an optimum delay parameter τ. Different prescriptions have appeared in the literature to choose τ but they are all empirical in nature and do not necessarily provide appropriate estimates.

1. First pass through zero of the autocorrelation function: in earlier works it was suggested to use the value of τ for which the autocorrelation function

$$C(\tau) = \sum_n (s_i - \hat{s})(s_{i+\tau} - \hat{s}) \tag{2.79}$$

first passes through zero which is equivalent to requiring linear independence. The application of the zero crossing of the autocorrelation function gives quite high values for both series.

2. First minimum of the Average mutual information: Fraser and Swinney, 1986; Kantz, 1994 suggested to use the average mutual information (AMI) function $I(\tau)$, as a kind of nonlinear correlation function to determine when the values of s_i and $s_{i+\tau}$ are independent enough (as per the threshold-ed definition and discussion on the distance between s_i and $s_{i+\tau}$ in (Hegger, Kantz, and Schreiber, 1999)) of each other to be useful as coordinates in a time delay vector

but not so independent as to have no connection with each other at all. This is our preferred method following the works of Kliková and Raidl, 2011; Vlachos and Kugiumtzis, 2009; Hegger, Kantz, and Schreiber, 1999 on its documented efficiency. For a discrete time series, $I(\tau)$ can be calculated as :

$$I(\tau) = \sum_{n,n+T} P(s_i, s_{i+\tau}) log_2 [\frac{P(s_i, s_{i+\tau})}{P(s_i)P(s_{i+\tau})}] \tag{2.80}$$

where $P(s_i)$ refers to individual probability and $P(s_i, s_{i+\tau})$ is the joint probability density. Following the method developed by Abarbanel and Gollub, 1996, to determine $P(s_i)$ we simply project the values taken from s_i versus i back onto the s_i axis and form a histogram of the values. once normalized, this gives us $P(s_i)$. For the join distribution of s_i and $s_{i+\tau}$ we form the two-dimensional histogram in the same way.

In general, the time lag provided by $I(\tau)$ is normally lower than the one calculated with the $C(\tau)$, $\tau_{AMI} \geq \tau_{correl}$ and provides the appropriate characteristic time scales for the motion. Even through $C(\tau)$ is the optimum linear choice from the point of view of predictability in a least square sense of $s_{i+\tau}$ from knowledge of s_i, it is not clear why it should work for nonlinear systems and it has been shown that in some cases it does not work at all. We plot $I(\tau)$ versus τ (the AMI) for the first difference or returns of the time series. The first minimum of the curve provides the best value of τ used as input to calculate the minimum embedding dimension in the algorithm of FNN.

2.2.6 BDS Test

The BDS test by Broock et al., 1996 is a statistical test for nonlinearity dependence, although not designed as a direct test for this purpose, it can distinguish between deterministic nonlinear systems, random process, and stochastic Opong et al., 1999a. It is a nonparametric technique that allows us to distinguish whiteness from an unspecified alternative Yousefpoor, Esfahani, and Nojumi, 2008, Barnett and Serletis, 2000. However we must stress here that the BDS represents a necessary but not sufficient condition for chaos Barnett et al., 1997, Barnett and Hinich, 1992,Barnett and Serletis, 2000. It is a widely used test in most studies analyzing chaotic behavior in financial time series Papaioannou and Karytinos, 1995. The null hypothesis H_0 is that the series is an independent and identically distributed (i.i.d) process and the BDS statistic is given by

$$W(N, m, \epsilon) = \sqrt{N} \frac{C(N, m, \epsilon) - C(N, m, \epsilon)^m}{\hat{\sigma}(N, m, \epsilon)} \tag{2.81}$$

where $C(N, m, \epsilon)$ is the correlation integral of the embedding dimension m, expressed by the following relation:

$$C(N, m, \epsilon) = \frac{1}{M(M-1)} \sum_{i=1}^{M} \sum_{j \neq i+1}^{M} H(\epsilon - \parallel X_i - X_j \parallel) \tag{2.82}$$

where $X = [X_1 X_2, \ldots, X_M]^T$ is the constructed trajectory and X_i the state of the system at discrete time i. Let our series have N data $\{x_1, x_2, \ldots, x_N\}$,$X_i$ is given by $[X_i X_{i+J}, \ldots, X_{i+(m-1)J}]$, where L is the lag or the reconstruction delay. Thus, we symbolize by X the $M \times m$ matrix, where M is given by $M = N - (m-1)L^3$. Let also denoted by $\hat{\sigma}(N,m,\epsilon)$ the estimate of the asymptotic standard deviation of the expression $C(N,m,\epsilon)^m - C(N,1,\epsilon)$. Under H_0 of i.i.d or whiteness, the quantity $\hat{\sigma}^2(N,m,\epsilon)$ is calculated as follows:

$$\hat{\sigma}^2(N,m,\epsilon) = 4[K^m(N,\epsilon) + 2\sum_{j=2}^{m-1} K^{m-j}(N,\epsilon)C^{2j}(N,1,\epsilon) + (m-1)^2C^{2m}(N,1,\epsilon) - m^2K(N,\epsilon)C^{2m-2}(N,1,\epsilon)] \tag{2.83}$$

where

$$C_m(N,\epsilon) = \frac{2}{N(N-1)(N-2)}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\sum_{K=j+1}^{N} H(\epsilon - \parallel X_i - X_j \parallel)H(r - \parallel X_j - X_K \parallel)) \tag{2.84}$$

and the Heaviside function is

$$H(\epsilon - \parallel X_i - X_j \parallel) = \begin{cases} 1, & \text{if} \quad \epsilon - \parallel X_i - X_j \parallel > 0 \\ 0, \text{otherwise} \end{cases} \tag{2.85}$$

Brock, 1986, have demonstrated that the $W(N,m,\epsilon)$ statistics follows, asymptotically, the standard normal distribution. A rejection of the null i.i.d hypothesis in this test is an indication of nonlinear dependence in the data. Furthermore , a positive and significant BDS statistics hints that certain patterns within the reconstructed trajectory have the probability to emerge more frequently , a behavior not shown in a random process. On the opposite, a negative and significant BDS indicate the infrequent appearance of certain patterns compared with a random process. Brock (1986) proposed a residual test in order to validate the results of the BDS test as well as to make the statistic a valuable tool in distinguishing between deterministic chaos and nonlinear stochastic processes. This test is applied to the standardized residual series from an ARCH-type model Broock et al., 1996,Adrangi et al., 2001. To test the null hypothesis of i.i.d. and whiteness, BDS is a powerful tool. However, the rejection of the null hypothesis does not lead directly to the conclusion that the series may be chaotic, since this rejection can be due to many alternatives like nonstationarity (e.g. due to changing macroeconomic, regulatory etc. factors), low complexity chaotic dynamics etc. Yousefpoor, Esfahani, and Nojumi, 2008,HSIEH, 1991.

Therefore, we point out here that rejecting the H_0 of i.i.d. does not lead us necessarily to accept the chaotic dynamic behaviour of our data. We restate the null and alternative assumptions as follows :

1. Null Hypothesis H_0: The observations of the sample are i.i.d.

2. Alternative H_1: The observations are not i.i.d.

In the BDS test statistic m is the embedding dimension and e is the distance between two observations, defined in terms of σ (St.dev). Since the BDS statistics are positive, if the p values are very significant ($p = 0.000$) the null hypothesis is rejected. Also in case that the BDS coefficients are less than 2.0 and 3.0 then according to Opong et al., 1999a, the test rejects the H_0 of i.i.d. process. The BDS test is also applied on the standardized residuals from the ARIMA/EGARCH models of conditional mean and volatilities, in order to check for any remaining structure in the residuals so assessing the quality of the ARIMA/EGARCH models. Via this we can detect the presence or not presence of nonlinear dependence in the residuals, so assessing the quality of the model. The combination of results of BDS and R/S-Hurst analysis analysed in the following sections, provide a firm evidence that the returns of an ARIMA/EGARCH model deviate strongly from random walk. We stress here, however, that the BDS is a necessary but not sufficient condition of the presence of a chaotic behaviour and also that it is consistent with other direct test for chaos, e.g. maximum Lyapunov exponent, correlation dimension etc. We note here that there are opposing views about the reliability of R/S-Hurst and BDS tests in detecting chaos :Dakhlaoui and Aloui, 2013 have deduced that the R/S-Hurst and BDS test work complementarily in the verification of the necessary conditions for the presence of chaos while Opong et al., 1999a argue that these two test fail to provide such an evidence.

2.2.7 Hurst Exponent in nonlinear time series analysis

Hurst developed a method called the rescaled range (R/S) for distinguishing completely random time series from correlated time series, Hurst, 1951. The procedure in Hurst analysis consists of the steps, starting with dividing a time series (of returns) of length L into d subseries of length n. Next for each subseries $k = 1, \ldots, d$; a) find the mean (E_k) and standard deviation (S_k); b) normalize the data ($Z_{i,k}$) by subtracting the sample mean $X_{i,k} = Z_{i,k} - E_k$ for $i = 1, \ldots, n$; c) create a cumulative time series $Y_{i,k} = \sum_{j=1}^{i} X_{j,k}$, for $i = 1, \ldots, n$; d) find the range $R_k = maxY_{1,k}, \ldots, Y_{n,k} - minY_{1,k}, \ldots, Y_{n,k}$; e) rescale the range R_k/S_k . Finally, the mean value of the rescaled range for subseries of length n is $(R/S)_n = (1/d)\sum_{k=1}^{d} R_k/S_k$. Equivalently, we can plot the $(R/S)_n$ statistics against n, on a double-logarithmic paper. If the returns process is white noise the plot is roughly a straight line with slope 0.5. If the process is persistent then the slope is > 0.5; if it is anti-persistent then the slope is < 0.5. the significance level is usually chosen to be $\sqrt{(1/N)}$ – the standard deviation of a Gaussian white noise. However, it should be noted that for small n there is a significant deviation from the 0.5 slope. For this reason the theoretical (i.e. for white noise) values of the (R/S) statistics are approximated by Weron and Przybylowicz, 2000a.

$$E(R/S)_n = \begin{cases} \frac{(n-\frac{1}{2})}{n} \frac{\Gamma(\frac{n-1}{2})}{\sqrt{\pi}\Gamma(\frac{n}{2})} \sum_{i=1}^{n-1} \sqrt{\frac{n-i}{i}}, n \leq 340 \\ \frac{(n-\frac{1}{2})}{n} \frac{1}{\sqrt{n\frac{\pi}{2}}} \sum_{i=1}^{n-1} \sqrt{\frac{n-i}{i}}, n > 340 \end{cases} \tag{2.86}$$

R/S is shown to follow asymptomatically the relation $R/S_n \sim cn^H$ thus by taking logs of both sides are have $logR/S_n = logc + Hlogn$. Therefore the value of H can be estimated by simple regression. The Hurst component $0 \leq H \leq 1$ is equal to 0.5 for random walk time series < 0.5 for anticorrelated series, and > 0.5 for positively correlated series. The Hurst exponent is directly related to the "fractal dimension", which gives a measure of the roughness of a surface. The relationship between the fractal dimension, D and the Hurst exponent, H, is $D = 2 - H$. Hurst exponents quantify the correlation of a **fractional Brownian motion**. A fractional Brownian motion (fBm) is a random walk with a Hurst exponent different from 0.5 and thus with a **memory**. The decaying of spectral density of an fBm has a relationship with the Hurst exponent as follows : P(f)=spectral density $\propto 1/f^{\beta}$, where $\beta = 2H + 1$, is the power spectrum exponent. β can be deduced from the slope of a curve generated by a linear fit of the data of f vs spectral density. Financial time series have been found to exhibit some universal characteristics that resemble the scaling laws typical of natural systems in which large number of units interacts. For instance, the Hurst exponent has been extensively applied by (Peters, 2015, Hsieh and Peters, 1993), to various capital markets and in most of the cases he has found persistent memory.

In an anti-persistent time series (also known as a mean-reverting series) an increase in values will most likely be followed by a decrease or vice-versa (i.e., values will tend to revert to a mean). This means that future values have a tendency to return to a long-term mean. In a persistent time series an increase in values will most likely be followed by an increase in the short term and a decrease in values will most likely be followed by an increase in the short term. If the Hurst exponent $H > 0.5$, then it reveals that the price process is correlated or that a price increase in the past is more likely to be followed by an additional increase than a price decrease. Hense, daily price movements are persistent and subject to trends. We can also interpret H as a measure of the bias in the fractional Brownian motion (fBm). Thus, the deviation from random walk provides interesting information on the volatility and risk inherent in financial markets. A relatively high H underpins the relatively smooth trend and less or controlled volatility. The persistent effect when $H > 0.5$ is stronger than when $H < 0.5$, i.e. the anti-persistent or anti-correlation or mean-reversion process of price movements. All the above information is related to the market efficiency. Higher H values correspond to emerging or developing markets in which the EMH is not satisfied. We know that EMH roughly stated, "argues" that in competitive markets all information is instantaneously reflected in prices and it is possible to systematically beat the market. A consequence of this postulate is that prices should follow a random walk ($H = 0.5$) and market cannot be forecasted. The longer H deviates from 0.5, the less noise in system or equivalently the more mean-reverting (or anti-correlated) is the price time series, therefore the ability of a model to forecast it is larger. Expressed differently, the closer the H value is to 0.00, the stronger is the tendency for the time series to revert to its long-term mean value. We can use Hurst exponents to classify markets: a) those where prices exhibit a strong mean-reverting mechanism ($H < 0.5$) and b) those where prices behave almost like Brownian motion ($H \approx 0.5$). Due to mean reversion behavior, a market is characterized as incomplete since there may exist infinitely many risk-neutral probability measures Oum, Oren, and Deng, 2006. Market incompleteness arises due to real-world market frictions, and this constitute a significant factor in making the system (market) high dimensional, nonlinear and complex. In such cases, the need to apply dimensionality reduction methods becomes crucial, making the manifold learning techniques such as those described and used in this thesis, a real prerequisite in order to analyze, model and forecast time series taken from such a complex system.

2.2.8 Generalized Hurst Exponent GHE and weighted GHE

The generalized Hurst exponent (GHE) is related to the long-term statistical dependence of a certain time series $X(t)$, with $t = (1, 2, \ldots, k, \ldots, \Delta t)$, defined over a time-window Δt with time-steps of one-unit. GHE measures the correlation persistence, thus it has to be related to fundamental statistical quantities, actually to the qth-order moments of the distribution of the increments of the time series, defined as

$$K_q(\tau) = \frac{\langle |X(t+\tau) - X(t)|^q \rangle}{\langle |X(t)|^q \rangle} \tag{2.87}$$

Barabsi and Vicsek, 1991, where τ can vary between 1 and τ_{max} and $\langle \cdot \rangle$ symbolizes the sample average over the time-window. We point out here that for $q = 2$, $K_q(\tau)$ is analogous to the autocorrelation function: $C(t, \tau) = \langle X(t + \tau)X(t)\rangle$. From the scaling behavior of $K_q(\tau)$, the generalized Hurst exponent is then defined given that the following relation holds: $K_q(\tau) \propto \tau^{(qH(q))}$. Based on the aforementioned behavior, processes exhibiting this scaling can be categorized into two groups: (a) Processes with $H(q) = H$, i.e. independent of q, i.e they are uniscaling (or unifractal) and their behavior is uniquely determined by the constant H (Hurst exponent or self-affine index), (b) Processes with varying $H(q)$, called multiscaling (or multifractal) and each moment scales with a different exponent. Matteo, Aste, and Dacorogna, 2005 have pointed out how financial data exhibit multi-fractal scaling behaviors. Due to the fact that all the information about the scaling properties of a time series is contained in the scaling exponent $H(q)$, the analysis based on GHE is quite simple. If we consider that the recent past is more important than the remote past we can use the "exponential smoothing" technique to take into account that the informational relevance of observations decays exponentially the further we move to the past. This is obtained by defining smoothing as :

$$w_s = w_0 exp(\frac{-s}{\theta}), \forall s \in 0, 1, 2, \ldots, \Delta t - 1 \tag{2.88}$$

where θ is the weights characteristic time and $a = \frac{1}{\theta}$ is the exponential decay factor. According to Pozzi, Matteo, and Aste, 2012, the parameter w_0 is $w_0(a) = \frac{1-e^{-a}}{1-e^{-a\Delta t}}$. In general, for any quantity $f(u_t)$, the weighted average over the time-window $[t - \Delta t + 1, t]$ is given by $\langle f\rangle_w(t) = \sum_{s=0}^{\Delta t-1} w_s f(u_{t-s})$, and the weighted GHE (wGHE) is therefore obtained by substituting the normal with weighted averages:

$$K_q^w(t, \tau) = \frac{\langle |X(t + \tau) - X(t)|^q\rangle_w(t)}{\langle |X(t)|^q\rangle_w(t)} \tag{2.89}$$

Using the scaling law we reach to the linear equation :

$$ln(K_q^w(t, \tau)) = qH^w(q)ln(\tau) + const \tag{2.90}$$

which provides the wGHE. The nonlinear behavior can be detected by ploting the variation of $qH(q)$ versus q.

Autocorrelation function ACF and Hurst

A long memory process is a process with a random component, where a past event has a decaying effect on future events. The process has some memory of past events, which is "forgotten" as time moves forward. The mathematical definition of long memory processes is given in terms of autocorrelation. When a data set exhibits autocorrelation, a value x_i at time t_i is correlated with a value x_{i-d} at time t_{i-d} where d is some time increment in the future. In a long memory process, autocorrelation decays over time and the decay follows a power law, i.e. $p(k) = Ck^(- \delta)$, where C is a constant and $p(k)$ is the autocorrelation function with lag k. The Hurst exponent is related to the exponent δ by $H = 1 - \frac{\delta}{2}$. DFA algorithm analysed in the following section, is appropriate for detecting weak autocorrelations (AC) in nonstationary data. It has the ability to distinguish intrinsic AC associated with memory effects in the underlying dynamics from effects that are attributed to exogenous factors.

2.2.9 Detrended fluctuation Analysis (DFA)

DFA is an alternative tool to estimate long-range dependence, proposed by Peng et al., 1994, and its advantage over R/S analysis is that it avoids the artifacts of non-stationarity expressed as apparent long-range correlation. To estimate DFA we divide again a time series (of returns) of Length L into d subseries of length n. For each subseries then, $k = 1, \ldots, d$; a) we create a commutative time series $Y_{i,k} = \sum_{j=1}^{i} X_{j,k}$ for $i = 1, \ldots, k$; b) we fit a least squares line $(Y_k(x) = \hat{a}_k x + b_k$ to $Y_{1,k}, \ldots, Y_{n,k}$ and finally c) compute the RMS fluctuation (i.e. standard deviation) of the integrated and detrended time series

$$F(k) = \sqrt{(\frac{1}{n}\sum_{i=1}^{n}(Y_{i,k} - a_k i - b_x)^2} \tag{2.91}$$

The mean value of RMS fluctuation for all subseries of length n is

$$\hat{F}(k) = \frac{1}{d}\sum_{k=1}^{d} F(k) \tag{2.92}$$

As is the case of Hurst analysis, we plot on a double-logarithmic "paper" $\hat{F}(k)$ against n and we get the power-law scaling $F(k) \sim k^{\alpha}$ where α is the fractal DFA exponent. If the returns process is white noise then the slope is roughly 0.5. If the slope is > 0.5 the process is persistent while for an antipersistent process the slope is <0.5. For various processes, index a takes the following values: $a < 1/2$ (anti-correlated), $a \approx 1/2$, (white noise, uncorrelated), $a > 1/2$ (correlated), $a \approx 1$ ($1/f$ -noise, pink noise), $a > 1$ (non-stationary, unbounded) and $a \approx 3/2$ (Brownian noise). The classical Hurst exponent corresponds to the second moment for stationary cases $H = \alpha$ and to the second moment -1 for nonstationary cases $H = a - 1$ Stanley et al., 1999. Both measures H and DFA have a major drawback, the fact that no asymptotic distribution theory has been derived. However, Weron, 2002 has obtained empirical confidence intervals for both measures via a Monte Carlo study.

2.2.10 Periodogram Regression (GPH)

This technique denoted as GPH was originally proposed by Geweke and Zhou, 1996, in order to estimate the Hurst exponent. This method is based on observations of the slope of the spectral density around the angular frequency ($w = 0$) of a fractionally integrated model. The spectral density of a general fractionally integrated model with differencing parameter d is identical to that of a fractional Gaussian noise with Hurst exponent $H = d + 0.5$. This estimation procedure of the Hurst exponent begins with calculating the periodogram. Then a linear regression model is developed such as:

$$log(I_k(\omega_k)) = \alpha - dx_t + \epsilon_k = \alpha - dlog(4sin(\omega_k/2)^2) + \epsilon_k \tag{2.93}$$

at low frequencies ω_k, $k = 1, 2, \ldots, K \leq L/2$. The least squares estimate of the slope yields the differencing parameter d, hence the Hurst exponent is given by the following formula: $H = 0.5 + \widehat{d}$. The choice of k differs in the sense that some authors recommend $K = L^0.5$, however other values have been also suggested; $K = L^0.45$ or $L^0.2 \leq K \leq L^0.5$. Parameter $\widehat{d}$ is calculated according to the formula:

$$\widehat{d} \sim \mathcal{N}\left(d, \frac{\pi^2}{6\sum_{k=1}^{K}(x_t - \hat{x})^2}\right) \tag{2.94}$$

If a time series is generated by anti-persistent, mean-reverting processes far away from the Geometric Brownian Motion for which H=0.5, then the process is non-Markovian and anti-correlated Shreve, Chalasani, and Jha, 1997. This result is of great importance in pricing options or futures in financial markets. If this behavior of an index in a market is not taken into account, then it increases the risk exposure for someone which "writes" the Option contract. It is also well known, at least in the financial markets, that the fundamental assumption in the Black-Scholes-Merton (BSM) model for option pricing Merton, 1980,Eydeland and Wolyniec, 2003 is that the underline asset follows a Geometric Brownian Motion, GBM (H=0.5). However, for some markets, as the electricity market, this assumption is far from being satisfied. It is very important, therefore, to understand why this anti-persistent, mean-reverting dynamics is shown to be so robust with time. In a stock market we also observe such a process, but it is shortly lived because the agents take advantage of this, they "learn" about it quickly, so ultimately this arbitrage opportunity is vanished. In order to shed light on this peculiar feature, we must note the specific characteristics of the market that is analyzed. For example, a commodity market (electricity is a peculiar commodity), exhibits a prolonged mean-reverting behavior, mixed with periods of spikes, making the behavior very complex.

We now show the relations between indices β (power spectral index), DFA index α index in decaying ACF δ, and the Hurst exponent H. Using Wiener-Khinchin theorem (Cohen, 1998) we have :

$$\delta = 2 - 2\alpha \tag{2.95}$$

$$\beta = 2\alpha - 1 \tag{2.96}$$

$$\delta = 1 - \beta \tag{2.97}$$

Therefore, α is connected to the slope of the power spectrum β, useful in describing the color noise by the following relation

$$\alpha = ((\beta + 1))/2 \tag{2.98}$$

For fractional Gaussian noise (FGN), $\beta \in [-1, 1]$ therefore $\alpha = [0, 1]$ and $\beta = 2H - 1$. For fractional Brownian motion (FBM), $\beta \in [1, 3]$ so $\alpha = [1, 2]$ and $\beta = 2H + 1$. The index α is equal to $H + 1$ for FBM. Therefore, FBM is the cumulative sum or the integral of FGN, which equivalently means that they have power spectra $\beta \neq 2$.

2.2.11 Quantifying Complexity via Entropy measure. The Tsallis Entropy as an alternate measure of volatility

In the previous sections we examined a number of analyzing techniques of Hurst exponents (classical R/S, GHE, DFA and AWC). All these methods rely on the scaling of some kind of fluctuation measure, as the standard deviation or variance, within a rolling window and measure the correlation scaling exponent, based on the assumption that the series analyzed follows a Gaussian distribution, as for example in the case of fractional Brownian motion, fbm. However, in the majority of cases of analyzing real time series the Gaussanity assumption is not valid. Therefore the correct correlation scaling exponent (Hurst Exponent) can be reliably estimated by a method that consider not only Gaussian but also non-Gaussian (e.g Levy walks) as processes that both contributing or driving the dynamics of electricity prices. Therefore a method that determines and separates the contribution to the scaling from both correlations and non-Gaussian process is needed. Entropy analysis is such a method and is based on the thermodynamics view of a time series.

Tsallis Entropy is a useful tool for the analysis of a broad range of signals including complex, nonlinear as well as nonstationary signals. Although these concepts are most often named after Tsallis due to his work in the area, they had been studied by others long before him. In general, entropy in physics is defined as the measure of "disorder" or uncertainty, but in finance entropy is associated with a time dependent probability distribution function. Entropy is defined as a quantitative representation of "disorder" (or measure of uncertainty). It is actually a relationship between macroscopic and microscopic quantities that quantify the dispersion of energy. The equivalence of the dispersion in financial and energy markets is the way the number of states (or regimes) translates themselves into a probability distribution of the aggregate sentiment in these markets.

The most well-known entropy is the Shannon information measure $S_n(P)$ of a probability measure p on a finite set X, given by (in discrete setting) :

$$S_n(P) = -\sum_{i=1}^{p_i} lnp_i \tag{2.99}$$

where $\sum_{i=1}^{n} p_i = 1, \quad p_i \geq 0$, $0ln0 = 0$ and $i = 1, \ldots, n$ the number of states, p_i the propability of outcome i. The entropy is the sum over the product of probability of outcome (p_i) times the logarithm of the inverse p_i, which is also called i's surprisal and the entropy of x is the expected value of its outcome's surprisal. We point out here that if two states A and B are independent from each other, then $p(A \cup B) = p(A)p(B)$, then $S_n(A) + S_n(B) = S_n(A \cup B)$. Let x is a stochastic variable, for example the pair of an FX market, within the general context of anomalously diffusing systems, in which the above mentioned index "live" having a time-dependent probability distribution $f(x,t)$ of x. In the non-extensive TE the parameter a, as we have noted, characterizes the non-extensivity of the entropy, with which the following constraints are associated :

$$\int f(x,t)dx = 1 \tag{2.100}$$

$$\langle x - \hat{x}(t) \rangle)_a \equiv \int [x - \hat{x}(t)] f(x,t)^a dx = 0 \tag{2.101}$$

$$\langle (x - \hat{x}(t))^2 \equiv \int [x - \hat{x}(t)]^2 f(x,t)^a dx = \sigma_a(t)^2 \tag{2.102}$$

Relation 2.100 is just the normalization of probability. In 2.101 and 2.101 $f(x,t)$ is raised to the power a. Unless $a = 1$ these are not the usual constraints that load to the mean and variance, so σ_a^2 (the "a -variance") is not the usual variance. Notice that the Tsallis parameter a does not depend on time. Using the constraints above, the maximization of TE for fixed a yields (Tsallis, 1988):

$$f(x,t) = \frac{1}{z(t)} 1 + \beta(t)(a-1)[x - \hat{x}(t)]^{2\frac{-1}{(a-1)}} \tag{2.103}$$

where $z(t)$ a normalization constant and β the Langrage multipliers, related to the first and third constraints. The ordinary variance of $f(x,t)$ is given by :

$$\sigma^2(t) = \langle (x - \hat{x}(t))^2 \rangle_{a=1} = \begin{cases} \frac{1}{(5-3a)\beta(t)}, & a < \frac{5}{3} \\ \infty, & a \geq \frac{5}{3} \end{cases} \tag{2.104}$$

Therefore for application to stochastic data, with finite variance, the Tsallis parameter a must be within the range $1 < a < 5/3$. In continuous setting :

$$H = -\int_{-\infty}^{\infty} f(x) ln f(x) dx, \quad \int_{-\infty}^{\infty} f(x) dx = 1, \quad f(x) \geq 0 \tag{2.105}$$

where $f(x)$ is the density function of a continuous probability distribution, evaluated for all x values. Such a case is the Shannon entropy applied to a probability density function $f(x)$ or in a discrete set where is defined as the product of the probability of the outcome p_i times the logarithm of the inverse of p_i. But, Shannon entropy cannot be extended to account for long-range interactions. Nonextensive entropy also known as Tsallis Entropy (TE) is used instead, because it generalizes the classical and quantum statistics. More specifically if we consider any positive real number a, the Tsallis entropy of order a of a probability measure p on a finite set X is defined as (Tsallis, 1988):

$$H_a(p) = \begin{cases} \frac{1}{a-1}(1 - \sum_{i \in X} p_i^a), & \text{if} \quad a \neq 1 \\ -\sum_{i \in X} p_i ln(p_i), & \text{if} \quad a = 1 \end{cases} \tag{2.106}$$

As we have mentioned before, the characterization of the Tsallis entropy is the same as that of the Shannon entropy except that for the Tsallis entropy, the degree of nonadditivity is a instead of 1. The exponent a is a measure of nonadditivity such that $H_a(A \cup B) = H_a(A) + H_a(B) - (1-a)H_a(A)H_a(B)$. Larger values of α emphasize longe-range interactions between states and can be interpreted as long memory-parameter $H_a(P)$ recovers the shannon entropy if $a \to 1$.

From the above, TE is a bounded measure, where the lower corresponds to a total lack of dispersion while the upper to the maximum dispersion at a given point in time. The idea behind using TE in financial markets is that the sentiments in them can be captured through the aggregation of the subjective expectations of market agents. Extreme price movements (large volatility) are less likely to occur in case the expectations of market agents are highly dispersed and independent (i.e the entropy is low). The opposite happens when market agents have highly dependent and less dispersed expectations, in which case the aggregate market sentiment can drive the prices to very high levels. By using this particular measure we place emphasis on long-range, time-dependent, interactive instability in the market. Considering now the following constraints:

$$\int f(x,t) dx = 1 \tag{2.107}$$

$$\frac{\int x^2 f(x)^a dx}{\int f(y)^2 dy} = \sigma^2 \tag{2.108}$$

We can yield the a-Gaussian probability density function from the maximum entropy principle for TE:

$$f(x) = \frac{exp_a(-\beta_a x^2)}{\int exp_a(-\beta_a x^2) dx} \propto \frac{1}{z}[1 + (1-a)(-\beta_a x^2)]^{\frac{1}{(1-a)}} \tag{2.109}$$

where β_a and Z are a function of a and the a-exponential function is defined as:

$$exp_a(x) = \begin{cases} [1 + (1-a)x]^{\frac{1}{1-a}}, & 1 + (1-a)x > 0 \\ 0, & \text{otherwise} \end{cases} \tag{2.110}$$

For $a \to 1$, a-Gaussian distribution is simply the usual Gaussian distribution. The parameter a itself measures the non-extensivity of the system and it is not a measurement of system's complexity. The relationship between TE and ARCH-GARCH processes has been reported in literarure by linking the dynamic parameters γ_1 and a_1 as well as the noise mature with the entropic index a (Burlaga and Viñas, 2005, Tsallis, Gell-Mann, and Sato, 2005, Tsekouras and Tsallis, 2005). There are cases (Balasis et al., 2008) where high entropy values indicate a low degree of organization which is reflected in the low volatility of the system, whilst lower entropy values indicate a high degree of organization or equivalently high volatility of the system. In this sense, TE is regarded as an alternative risk indicator capable of measuring and forecasting market volatility. Thus without loss of generality (Balasis et al., 2008) there is a stong link between GARCH processes and Tsallis Entropy.

Rolling Tsallis Entropy (TE) can also be used as an alternative approach to rolling Lyapunov algorithm, due to inherent weakness of the later (non-robustness and heavy computational burden). Moreover TE successfully captures the higher moments in the entire spectrum of our analysis and thus can be also compared against the unconditional volatility of both time series. The mathematical definition of entropy reveals also the strong and clear link between the entropy and the predictability of a system. Positive entropy means that the future state of the system is unpredictable, thus independently of the time period of the previous observations we cannot obtain knowledge of the future state(s) of the system.

On the other hand, Lyapunov's exponents can be used to determine the rate of predictability in the sense that positive Lyapunov's exponents indicate **chaos**; the bigger the largest Lyapunov exponent is, the less predictable is the system under consideration. Predictability thus, can be expressed as the reciprocal of Lyapunov exponent according to the formula, Predictability $= \frac{1}{\lambda}$, where λ is the Lyapunov exponent. There is a strong relation between entropy and Lyapunov's exponents according to the following formula:

$$H(T) \leq \sum_{(i,\lambda_i>0)} \lambda_i \tag{2.111}$$

We see that the entropy is upper bounded by the sum of the positive Lyapunov exponents and is therefore finite and positive. Since positive Lyapunov's exponents reveal instability the same intepretation is also valid for the positive entropy. On the other hand, zero entropy means that if we observe long enough a system's state(s) we are certain about its future one(s). Entropy as a measure of uncertainty has been extensively used in literature in order to quantify the behavioral aspects of systemic risks, as well as to gain insight into the evolution of the aggregate market expectations (Gradojevic and Gençay, 2011, Gradojevic and Caric, 2017, Gençay, Selçuk, and Whitcher, 2003).

In the following section, we investigate the Lyapunov exponents and more specifically the importance of the maximal Lyapunov Exponent on the rise of chaotic behavior in a time series dynamics.

2.2.12 The maximal Lyapunov Exponent

Chaos arises from the exponential growth of infinitesimal perturbations associated simultaneously with a global folding process in order to secure boundedness of the trajectory evolution. Lyapunov exponents characterize this exponential instability Eckmann and Ruelle, 1985. Suppose that we assume a local decomposition of the phase space into directions with different stretching or contraction rates. It is natural then to consider that the spectrum of all exponents is the proper average of these local rates over the whole invariant set (there are as many exponents as there are space directions). Since we do not really know the exact number of degrees of freedom of the dynamics we investigate (more often the physical phase space is unknown), it creates a difficult problem in time series analysis, so instead the spectrum of Lyapunov exponents is computed in some embedding space, meaning that the number of exponents depends on the dimensions of the reconstructed space, and might be larger than in the actual or physical phase space. Therefore, the additional exponents (corresponding to extra dimensions) are spurious. Stoop and Parisi, 1991, provide practical suggestions to either avoid or to identify them. We point out here that Lyapunov exponents are invariant under smooth transformations and are thus independent of the measurement function or the embedding procedure. They carry a dimension of an inverse time and have to be normalized to the sampling interval Kantz, 1994. We can determine the maximal Lyapunov exponent without the explicit construction of a model for the time series. Checking explicitly the independence of embedding parameters and the exponential law for the growth of distances secures a reliable characterization. The presence of noise in a data series is the main problem in detecting Chaos. Therefore in any test used in searching for chaos the noise effect must be taken into consideration.

We consider the time series data as a trajectory in the reconstructed embedding space, of dimension m and delay time τ.

Let us assume also that we observe in a neighbor u_i, of diameter ϵ, a very close return s_i', to a previously visited point s_i. We now can consider the distance $\Delta_0 = s_i - s_i'$ as a small perturbation, which should grow exponentially in time. Its future dynamic evolution some time $t = I$ ahead can be monitored by plotting the time series: $\Delta_I = s_{i+I} - s_{i+I}'$. If we find that $|\Delta_I| \approx \Delta_0 e^{\lambda I}$ then λ is (with probability one) the maximal Lyapunov exponent. Because of many effects, in practice there will be fluctuations. Therefore, due to fluctuations we can derive a robust, consistent and unbiased estimator for the maximal Lyapunov exponent given as follows

$$S(\epsilon, m, t) = \langle \left(\frac{1}{|u_i|} \sum_{s_i' \in u_i} |s_{i+\tau} - s_{i+\tau}'| \right) \rangle \tag{2.112}$$

If $S(\epsilon, m, t)$ exhibits a linear increase with identical slope for all m larger than some m_0 and for a reasonable range of ϵ, then the slope can be taken as an estimate of the maximal exponent λ_{max}. We point out here that the r.h.s of equation 2.112 is the average divergence or separations of Nearest Trajectories NT, and is the basis of the method developed by Rosenstein, Collins, and Luca, 1993, in estimating λ_{max}. The Lyapunov exponent measures the average exponential divergence (in case of (+) exponent) or convergence (in case of (-) exponent) rate between two adjacent trajectories or Nearest Trajectories (NT) within a time distance that differ in initial conditions just by an extremely tiny amount. If $\lambda < o$ then the orbits attracts to a stable fixed point or stable periodic orbit. Dissipative or non-conservative dynamical systems have negative Lyapunov exponents, exhibiting asymptotic stability. The more negative the exponent, the greater the stability. At the extreme side, $\lambda = -\infty$ corresponds to super-stable fixed or periodic points. If $\lambda = o$ the orbit is actually a neutral fixed point so the system is in a steady state mode and it is called a conservative system, exhibiting Lyapunov stability. Such a system with $\lambda \neq o$ is near the "transition to chaos". If $\lambda > o$, the trajectory is unstable and chaotic. Neighboring points, no matter how “close” to each other are, will diverge, in time, with an arbitrary distance. Therefore, the entire phase space will be “filled” with visiting points that are unstable. The maximum Lyapunov exponent test is considered to be the natural way for chaos detection and constitutes the most direct approach for its estimation Yousefpoor, Esfahani, and Nojumi, 2008. According to Barnett and Serletis, 2000, Iseri, Caglar, and Caglar, 2008 and Yousefpoor, Esfahani, and Nojumi, 2008, maximum Lyapunov exponent, λ_{max}, is the most appropriate statistical invariant in detecting the sensitive dependence on initial conditions characteristic of a chaotic signal. As we seen abobe, max Lyapunov exponent, λ_{max}, measures the average rate of exponential separation (positive value) or contraction (negative value) between two nearby trajectories in the initial conditions (Barnett and Serletis, 2000, Kyrtsou and Terraza, 2002, Yousefpoor, Esfahani, and Nojumi, 2008, Orzeszko, 2008). In general, a (+) λ_{max}, indicates a chaotic system while a (-) λ_{max} indicates a stochastic system.

For the calculation of λ_{max} three methods are available : a) The direct method developed by Wolf et al., 1985 is the older one, also used by Rosenstein, Collins, and Luca, 1993, b) the Jacobian or regression method, developed by Nychka et al., 1992, McCaffrey et al., 1992, according to which λ_{max} is calculated by first estimating the Jacobians and then using neural network models, Barnett and Serletis, 2000, Yousefpoor, Esfahani, and Nojumi, 2008 and finally c) the third method based on a new algorithm using neural network methodology suggested by Gençay, Selçuk, and Whitcher, 2003 and Kyrtsou and Terraza, 2002.

The direct method requires long time series containing a large number of observations Brock, 1986; Barnett and Serletis, 2000 and it is very sensitive in the presence of noise, so the detection of chaotic dynamics requires a delicate and tedious work and the result cannot be easily accepted as a reliable one, especially for noisy financial series (Kyrtsou and Terraza, 2002), since in most cases gives overestimated exponents (Kyrtsou and Terraza, 2002, Barnett and Serletis, 2000). The problem of noise sensitivity of the first method is solved by the third method. Overall, according to existing literature in empirical studies of max Lyapunov exponent, the Jacobian approach provides the most reliable estimation despite the presence of noise in the time series.

In our applications we adopt the Jacobian method which is briefly described below following the work of Barnett and Serletis, 2000; based on the method as suggested by Nychka et al., 1992.

According to Jacobian method, in order to estimating λ_{max} we must first estimate the individual Jacobian matrixes:

$$J_t = \frac{\partial F(S_t)}{\partial S^T} \tag{2.113}$$

where S_t is the state-space representation of s_t, given as

$$S_t = F(S_{t-\tau}) + E_t, \quad F : R^m \to R^m \tag{2.114}$$

S_t can be expressed as a function of s_t as $S_t = (s_t, s_{t-\tau}, \ldots, s)$, where s_t is the sequence of variables $\{s_t\}$ produced by a nonlinear AR model, given below. For $1 \leq t \leq N$, let parameter τ is the time-delay and m is the length of the autoregression, so we have

$$s_t = f(s_{t-\tau}, s_{t-2\tau}, \ldots, s_{t-m\tau}) + e_t \tag{2.115}$$

Therefore,

$$F(S_{t-\tau}) = (f(s_{t-\tau}, \ldots, s), s, \ldots, s_{t-m\tau+\tau})^T \tag{2.116}$$

and

$$E_t = (e_t, 0, \ldots, 0)^T \tag{2.117}$$

where f is a smooth unknown function and $\{e_t\}$ is a random process, independent random variables with zero mean and unknown constant variance. The optimal combination of the parameters in the calculation of max Lyapunov exponent, of the triplet (τ, m, q), is found by minimizing the Bayesian information criterion Nychka et al., 1992. So, the estimate of the maximum Lyapunov exponent $\hat{\lambda}_{max}$ is given by :

$$\hat{\lambda}_{max} = \frac{1}{2N} log(\theta_1(N)) \tag{2.118}$$

where $\theta_1(N)$ is the largest eigenvalue of the matrix $\hat{T}_N^T \hat{T}_N$ and
$\hat{T}_N = \hat{J}_N \hat{J}_{N-1} \ldots \hat{J}_1$.
$\hat{J}_t$, the nonparametric estimator of J_t, can be written as

$$\hat{J}_t = \frac{\partial \hat{F}(X_t)}{\partial X^T} \tag{2.119}$$

A relatively fast algorithm in estimating λ_{max} based on the above method and Neural Networks is provided by BenSaïda, 2015. In his algorithm, the test hypothesis H are : null hypothesis $H0 : \lambda \geq 0$, which indicate presence of Chaos ; and alternative hypothesis $H1 : \lambda < 0$, which indicates absence of chaos. The crucial parameters in applying this method on noisy real data are the embedding dimension, m, the delay time, τ, and the optimum (according to some criteria described in BenSaida's paper) number of hidden layers, q, in the Neural Network. The best triplet (τ, m, q) is found by the algorithm, based on preliminary max orders of these parameters, estimated by trial and error until a saturation of the effects of changes of their combination on λ is attained. The algorithm then finds the optimum orders of (τ, m, q) that maximizes λ, computed from all $\tau \times m \times q$ estimates Bensaïda and Litimi, 2013; BenSaïda, 2015.

2.3 Manifold Learning

Manifold learning techniques can be viewed as unsupervised learning algorithms in the sense that they“learns" from the available data a low-dimensional representation of the input high-dimensional space, thus providing an “optimal" (under certain assumptions) embedded information. The term “optimal" means preserving or optimizing a specific information criterion of the initial data-set's distribution or geometry (Ian Goodfellow, Yoshua Bengio, 2015). Here, we briefly present elements of the theory of the manifold learning algorithms that we use to embed the high-dimensional data as well as the lifting operators used to reconstruct the original high dimensional space.

2.3.1 Fundamentals of Manifold Learning

Combining typical or 'mainstream' tools like Hurst exponent (estimated by different methods), BDS, maximum Lyapunov exponent, Correlation dimension, Tsallis entropy etc. with Manifold learning approaches, can be extremely useful and effective towards understanding better the dynamics of the structural complexity of dynamical systems, via analyzing time series of variables of the system. Thus we can capture more reliably the endogenous complexities of the systems due to the inter dependency of a plethora of co-evolving factors of both deterministic and stochastic type, affecting the dynamic evolution of the system. Therefore, analyzing time series taken from such complex system, requires the application of very sophisticated tools, such as the 'hybrid' ones described in this thesis, built via combining well known tools from both linear and nonlinear time series with innovative approaches from the field of manifold learning and dimensionality reduction. For example, using the Lyapunov maximum Exponent λ_{max} to measure the stability of a dynamic behavior of a variable reflected by a time series, we can show whether there is a stable or no unstable (chaotic) dynamics in the return of our time series, depending on the sign of λ_{max}, for the period of time we are interested. We can combine this information gained by estimating the Lyapunov maximum Exponent, with results produced via using Hurst exponent estimation, hoping to receive consistent results, thus enhancing greatly the reliability of our findings. However, in case the system under examination is high-dimensional (as the systems analyzed in this thesis), before 'embarking' in any time series analysis using tools as as in the examples mentioned before, it sis necessary to reduce the dimensionality of such systems, via using manifold learning techniques.

It is therefore of paramount importance, to consider all the typical and main stream linear and nonlinear approaches-models, under the prism of manifold learning, focusing strongly on how these two approaches complement each other and how both enhance our arsenal of tools in analyzing very complex systems. This consideration or view, will make more pronounced the contribution of this thesis in the field of time series forecasting, by providing strong evidence that the above mentioned complementarity indeed enhances significantly our efficiency and reliability in forecasting nonlinear time series extracted by measuring representative variables of their underlying complex dynamical system.

Mathematical Framework

Let us denote by $\boldsymbol{x}_i \in \mathbb{R}^D$ the vector containing the features at the i-th snapshot of the time series, and by $\boldsymbol{X} \in \mathbb{R}^{D \times N}$ the matrix having as columns the vectors spanning a time window of N observations. The assumption is that the data lie on a "smooth enough" low-dimensional (say of dimension d) manifold that is embedded in the high-dimensional $\mathbb{R}^D$ space. Our aim is to exploit manifold learning algorithms to forecast the time series in the high-dimensional feature space. Thus, our purpose is to bypass the "curse of dimensionality" associated with both the dimension of the original data and, importantly, with the limited size of their snapshots in time that usually characterize real-world data (such as financial data). Towards this aim, we propose a numerical scheme that consists of three steps.

At the first step, we employ embedding algorithms (LLE and DMs) to construct nonlinear maps from the high-dimensional to a low-dimensional subspace that preserve as much as possible the intrinsic geometry of the manifold (i.e. maps assuring that neighborhoods of points in the high-dimensional space are mapped to neighborhoods of the same points in the low-dimensional manifold, and the notion of distance between the points is maintained as much as possible). In general it is expected that the manifold learning algorithms will provide an approximation to the intrinsic embedding dimension, which will differ from the "true" one. This depends on the threshold that one puts for the selection of the number of embedded vectors that span the low-dimensional manifold.

A central concept here is the Riemannian metric, which defines the properties of this map. At the second step, based on the resulting embedding features that span the low-dimensional subspace, we train a class of regression models (MVAR and GPR) to predict the evolution of the embedded time-series on the manifold based on their past values. The final step implements the lifting map. The aim here is to form the inverse map from the out-of-sample forecasted embedded points to the reconstruction of the features in the high-dimensional space. On one hand, the existence of such an inverse map is theoretically guaranteed by the assumption that a low-dimensional "sufficiently smooth" manifold exists. At this point, we should note that the embedding generated using noisy data may produce a distorted, and therefore unreliable geometry (see e.g. the discussion in Gajamannage, Paffenroth, and Bollt, 2019). On the other hand, the data-driven derivation of the corresponding inverse map is neither unique Lee, 2013 nor trivial as for example when using PCA. In general, one has to solve a nonlinear least squares problem requiring the minimization of an objective function that can be formed using different criteria for the properties of the neighborhood of points; this results in different inverse maps whose performance has to be validated.

Next, for the completeness and clarity of the presentation, we present basic concepts of the manifold theory, useful to understand the steps of the proposed numerical methodology.

Manifold learning techniques can be viewed as unsupervised learning algorithms in the sense that they "learn" from the available data a low-dimensional representation of the original high-dimensional space, thus providing an "optimal" (under certain assumptions) embedded subspace where the information available in the high-dimensional space is preserved as much as possible. Here, we briefly present some basic elements of the theory of manifolds and manifold learning (see e.g. Berger and Gostiaux, 2012; Kühnel, 2015; Lee, 2013; Wang, 2012). Let us start with the definition of a d-dimensional manifold.

<u>Manifold</u>

A set $M \subset \mathbb{R}^n$ is called a d-dimensional manifold (of class C^∞) if for each point $\boldsymbol{p} \in M$, there is an open set $W \subset M$ such that there exists a C^∞ bijection $\boldsymbol{f} : U \to W$, $U \subset \mathbb{R}^d$ open, with C^∞ inverse $\boldsymbol{f}^{-1} : W \to U$ (i.e. W is C^∞-diffeomorphic to U). The bijection $\boldsymbol{f}$ is called a local parametrization of M, while $\boldsymbol{f}^{-1}$ is called a local coordinate mapping, and the pair $(W, \boldsymbol{f}^{-1})$ is called a chart (or neighborhood) of $\boldsymbol{p}$ on M.

Thus, in a coordinate system $(W, \boldsymbol{f}^{-1})$, a point $\boldsymbol{p} \in W$ can be expressed by the coordinates $(f_1^{-1}, f_2^{-1}, \ldots, f_n^{-1})$, where f_i^{-1} is the i-th element of $\boldsymbol{f}^{-1}$. A chart $(W, \boldsymbol{f}^{-1})$ is centered at $\boldsymbol{p}$ if and only if $\boldsymbol{f}^{-1}(\boldsymbol{p}) = \mathbf{0}$. Thus, we always assume that the manifold satisfies the Hausdorff separation axiom, stating that every pair of distinct points on M have disjoint open neighborhoods. Summarizing, a manifold is a Hausdorff (Separated/T_2) space, in which every point has a neighborhood that is diffeomorphic to an open subset in $\mathbb{R}^d$.

<u>Tangent space and tangent bundle to a manifold</u>

Let $\boldsymbol{p} \in M \subset \mathbb{R}^n$. The tangent space to M at $\boldsymbol{p}$ is the vector subspace $T_{\boldsymbol{p}}M$ formed by the tangent vectors at $\boldsymbol{p}$ defined by

$$T_{\boldsymbol{p}}M = D\boldsymbol{f}_{\boldsymbol{u}}(T_{\boldsymbol{u}}\mathbb{R}^d), \quad \boldsymbol{f}(\boldsymbol{u}) = \boldsymbol{p},$$

where $T_{\boldsymbol{u}}\mathbb{R}^d$ is the tangent space at $\boldsymbol{u}$ and

$$D\boldsymbol{f}_{\boldsymbol{u}}(T_{\boldsymbol{u}}\mathbb{R}^d) = \left[\frac{\partial f_i(\boldsymbol{u})}{\partial u_j}\right] = \begin{bmatrix} \frac{\partial f_1}{\partial u_1} & \frac{\partial f_1}{\partial u_2} & \cdots & \frac{\partial f_1}{\partial u_d} \\ \frac{\partial f_2}{\partial u_1} & \frac{\partial f_2}{\partial u_2} & \cdots & \frac{\partial f_2}{\partial u_d} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_n}{\partial u_1} & \frac{\partial f_n}{\partial u_2} & \cdots & \frac{\partial f_n}{\partial u_d} \end{bmatrix},$$

with $rank(D\boldsymbol{f}) = d$. Thus, $S = \left\{\frac{\partial f}{\partial u_1}, \frac{\partial f}{\partial u_2}, \dots, \frac{\partial f}{\partial u_d}\right\}$ is a basis for $T_{\boldsymbol{p}}M$. The union of tangent spaces

$$TM = \cup_{\boldsymbol{p}\in M} T_{\boldsymbol{p}}M$$

is called the tangent bundle of M.

Riemannian manifold and metric

A Riemannian manifold is a manifold endowed with a positive definite inner product $g_{\boldsymbol{p}}$ defined on the tangent space $T_{\boldsymbol{p}}M$ at each point $\boldsymbol{p}$. The family of inner products $g_{\boldsymbol{p}}$ is called a Riemannian metric and the Riemannian manifold is denoted (M, g).

The above definitions refer to the case of continuum limit – infinite number of points. However, for real data with a limited number of observations, one can only try to approximate the manifold; the analytic charts of the Riemannian metric that actually define the geometry of the manifold are simply not available. Thus, a consistent manifold learning algorithm constructs in a numerical way a map that approximates in a probabilistic way the continuous one when the sample size $N \to \infty$. Thus, a fundamental preprocessing step of a class of manifold learning algorithms such as Laplacian Maps, LLE, ISOMAP and DMs is the construction of a weighted graph, say $\mathcal{G}(\boldsymbol{X}, E)$, where E denotes the set of weights between points. This construction is based on a predefined metric (such as the Gaussian kernel and the k-NN algorithm), which is used to appropriately connect each point $\boldsymbol{x} \in \mathbb{R}^D$ with all the others. This defines a weighted graph of neighborhoods of points. Theoretically, with an appropriate choice of the metric and its parameters, the graph is guaranteed to be connected, i.e. there is a path between every pair of points in $\boldsymbol{X}$ Lee, 2013.

For the completeness of the presentation, in the following sections we briefly discuss the manifold learning algorithms, the regression models and the lifting techniques used in this work.

2.3.2 Manifold learning algorithms

Principal Component Analysis (PCA)

After centering the input dataset by subtracting the mean of $\boldsymbol{X}$ we compute the covariance matrix of the centered matrix $\tilde{\boldsymbol{X}}$, $\boldsymbol{C} = \tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^T \in R^{D\times D}$ and calculate the eigenvectors $\boldsymbol{v}_j \in R^D$, $j = 1, 2, \dots, L$ and the corresponding eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_D$ of

C by solving the eigenvalue problem:

$$\boldsymbol{C}\boldsymbol{V} = \boldsymbol{\Lambda}\boldsymbol{V} \tag{2.120}$$

We then select the L eigenvectors corresponding to the largest eigenvalues. In this way, we form the matrix $\boldsymbol{V}_{D\times L}$, with columns the L leading eigenvectors $\boldsymbol{v}_j \in R^D$, $j = 1, 2, \ldots, L$. Thus, the projection of each point $\boldsymbol{x}_t \in R^D$, $t = 1, 2, \ldots N$ onto the low L-dimensional space at a point $\boldsymbol{y}_t \in R^L$, $L < D$ reads:

$$\boldsymbol{y}_{(PCA)t} = \boldsymbol{V}^T_{DxL}\boldsymbol{x}_t \tag{2.121}$$

Thus, the projection of all the data set $\boldsymbol{X} = (\boldsymbol{x}_1 \ldots \boldsymbol{x}_N)$ on the principal components subspace spanned by $\boldsymbol{V}_{DxL} = (\boldsymbol{v}_1 \ldots \boldsymbol{v}_L)$ is given by the matrix $\boldsymbol{Y}_{(PCA)} = (\boldsymbol{y}_{1(PCA)} \cdots \boldsymbol{y}_{N(PCA)}) \in R^{L\times N}$.

Resolving the Sign Ambiguity Problem

The intrinsic ambiguity in the sign of the loadings (scores) in the calculations of the principal components it is a well-known problem which (if not properly handled) may lead to significant misinterpretations of the outcomes. We follow the approach addressed by (Bro, Acar, and Kolda, 2008). The underlying idea behind the proposed implementation is that the principal components should point in the same direction of the data. This approach is implemented in the “sklearn" python module, (Pedregosa et al., 2011b).

Locally Linear Embedding (LLE)

The LLE algorithm (L.K.Saul and S.T.Roweis, 2003) is a nonlinear manifold learning algorithm which constructs $\mathcal{G}(\boldsymbol{X}, E)$ based on the k-NN algorithm. It is assumed that the neighborhood forms a basis for the reconstruction of any point in the neighborhood itself. Thus, every point is written as a linear combination of its neighbors. The weights of all pairs are then estimated with least squares, minimizing the L_2 norm of the difference between the points and their neighborhood-based reconstruction. The same rationale is assumed for the low-dimensional subspace, keeping the same estimates of the weights in the high-dimensional space. In the low-dimensional subspace, one now seeks for the coordinates of the points that minimize the L_2 norm of the difference between the coordinate of the points on a d-dimensional manifold and the weighted neighborhood reconstruction. The minimization problem is represented by an eigenvector problem, whose first d bottom non-zero eigenvectors provide an ordered set of orthogonal coordinates that span the manifold. It should be noted that in order to obtain a unique solution to the minimization problem, the number of k nearest neighbors should not exceed the dimension of the input space (L.K.Saul and S.T.Roweis, 2003).

Thus, the LLE algorithm can be summarized in the following three basic steps (L.K.Saul and S.T.Roweis, 2003).

1. Identify the k nearest neighbors for all $\boldsymbol{x}_i \in R^D$, $i = 1, 2, \ldots, N$ (with $k \geq d+1$). For each $\boldsymbol{x}_i$ this forms a set $K\{\boldsymbol{x}_i\} \subset \mathbb{R}^{k\times D}$ containing the k nearest neighbors of $\boldsymbol{x}_i$.

2. Write each $\boldsymbol{x}_i$ in terms of $K\{\boldsymbol{x}_i\}$ as

$$\boldsymbol{x}_i = \sum_{j \in K\{\boldsymbol{x}_i\}} w_{ij} \boldsymbol{x}_j, \tag{2.122}$$

and find the matrix $\boldsymbol{W} = [w_{ij}] \in \mathbb{R}^{N \times N}$ of the unknown weights by minimizing the objective function

$$\mathcal{L}(\boldsymbol{W}) = \sum_{i=1}^{N} \left\| \boldsymbol{x}_i - \sum_{j=1, j \neq i}^{N} w_{ij} \boldsymbol{x}_j \right\|_{L_2}^2, \tag{2.123}$$

with the constraint

$$\sum_{j=1}^{N} w_{ij} = 1. \tag{2.124}$$

3. Embed the points $\boldsymbol{x}_i \in \mathbb{R}^D$, $i = 1, 2, \ldots, N$, into a low-dimensional space with coordinates $\boldsymbol{y}_i \in \mathbb{R}^d$, $i = 1, 2, \ldots, N$, $d \ll D$. This step in LLE is accomplished by computing the vectors $\boldsymbol{y}_i \in \mathbb{R}^d$ that minimize the objective function:

$$\phi(\boldsymbol{Y}) = \sum_{i=1}^{N} \left\| \boldsymbol{y}_i - \sum_{j=1, j \neq i}^{N} w_{ij} \boldsymbol{y}_j \right\|_{L_2}^2, \tag{2.125}$$

where the weights w_{ij} are fixed at the values found by solving the minimization problem (2.123). The embedding vectors are required to be centered at the origin with an identity covariance matrix (L.K.Saul and S.T.Roweis, 2003) The embedded vector $\boldsymbol{y}_i$ are constrained so that they have a zero mean and a unit covariance matrix.

The cost function (2.125) is quadratic and can be stated as

$$\phi(\boldsymbol{Y}) = \sum_{i=1}^{N} \sum_{j=1}^{N} q_{ij} \langle \boldsymbol{y}_i, \boldsymbol{y}_j \rangle, \tag{2.126}$$

involving the inner products of the embedding vectors and a symmetric square matrix $\boldsymbol{Q}$ with elements

$$q_{ij} = \delta_{ij} - w_{ij} - w_{ji} + \sum_{k} w_{ki} w_{kj}, \tag{2.127}$$

$\delta_{ij} = 1$ if $i = j$, and 0 otherwise. In matrix form, $\boldsymbol{Q} \in \mathbb{R}^{N \times N}$ can be written as

$$\boldsymbol{Q} = (\boldsymbol{I} - \boldsymbol{W})^T (\boldsymbol{I} - \boldsymbol{W}). \tag{2.128}$$

In practice, this representation of $\boldsymbol{Q}$ gives rise to a significant reduction of the computational cost, especially when N is large, since left multiplication by $\boldsymbol{Q}$ can be performed as

$$\boldsymbol{Q}\boldsymbol{v} = (\boldsymbol{v} - \boldsymbol{W}\boldsymbol{v}) - \boldsymbol{W}^T (\boldsymbol{v} - \boldsymbol{W}\boldsymbol{v}), \quad \boldsymbol{v} \in \mathbb{R}^N, \tag{2.129}$$

requiring just two multiplications by the sparse matrices $\boldsymbol{W}$ and $\boldsymbol{W}^T$.

Minimizing the cost function (2.125) can be performed via the eigenvalue decomposition of $\boldsymbol{Q}$. Specifically, the eigenvector associated with the smallest (zero) eigenvalue is the vector with all 1's and it trivially minimizes (2.125) (L.K.Saul and S.T.Roweis, 2003); it is disregarded because it leads to a constant (degenerate) embedding. The optimal embedding is therefore obtained by the d eigenvectors of $\boldsymbol{M}$, denoted as $\boldsymbol{q}_k \in R^N$, $k = 1, 2, \ldots, d$, corresponding to the next d smallest eigenvalues. Thus, the coordinates of $\boldsymbol{x}_i$, $i = 1, 2, \ldots, N$, in the embedded space are given by the vector

$$\boldsymbol{\mathcal{R}}(\boldsymbol{x}_i) = \boldsymbol{y}_i = [q_{i1}, q_{i2}, \ldots, q_{id}]^T, \tag{2.130}$$

where q_{ij} denotes the j-th element of eigenvector $\boldsymbol{q}_i$.

Diffusion Maps (DMs)

Here, the construction of the affinity matrix is based on the computation of a random walk on the graph $\mathcal{G}(\boldsymbol{X}, E)$. The first step is to construct a graph using a kernel function, say $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$. The kernel function can be chosen as a Riemannian metric, so that it is symmetric and positive definite. Standard kernels, such as the Gaussian kernel, typically define a neighborhood of each point $\boldsymbol{x}_i$, i.e. a set of points $\boldsymbol{x}_j$ having strong connections with $\boldsymbol{x}_i$, namely, large values of $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

At the next step, one constructs a Markovian transition matrix, say $\boldsymbol{P}$, whose elements correspond to the probability of jumping from one point to another in the high-dimensional space. This transition matrix defines a Markovian (i.e. memoryless) random walk X_t by setting

$$p_{ij} = p(\boldsymbol{x}_i, \boldsymbol{x}_j) = \text{Prob}(X_{t+1} = \boldsymbol{x}_j | X_t = \boldsymbol{x}_i).$$

For a graph constructed from a sample of finite size N, the weighted degree of a point (node) is defined by

$$deg(\boldsymbol{x}_i) = \sum_{j=1}^{N} k(\boldsymbol{x}_i, \boldsymbol{x}_j), \tag{2.131}$$

and the volume of the graph is given by

$$vol(\mathcal{G}) = \sum_{i=1}^{N} d(\boldsymbol{x}_i).$$

Then, the random walk on such a weighted graph can be defined by the transition probabilities

$$p_{ij} = p(\boldsymbol{x}_i, \boldsymbol{x}_j) = \frac{k(\boldsymbol{x}_i, \boldsymbol{x}_j)}{deg(\boldsymbol{x}_i)}. \tag{2.132}$$

Clearly, from the above definition, we have that $\sum_{j=1}^{N} p(\boldsymbol{x}_i, \boldsymbol{x}_j) = 1$.

We note that in the continuum the above can be described as a continuous Markov process on a probability space $(\Omega, \mathcal{H}, \mathcal{P})$, where Ω is the sample space, $\mathcal{H}$ is a σ-algebra of events in Ω, and $\mathcal{P}$ is a probability measure. Let μ be the density function of the points in the sample space Ω induced from the probability measure, $\mu : \mathcal{H} \to \mathbb{R}$ with $\mu(\Omega) = 1$. Then, using the kernel function k, the transition probability from a point $\boldsymbol{x} \in \Omega$ to another point $\boldsymbol{y} \in \Omega$ is given by

$$p(\boldsymbol{x}, \boldsymbol{y}) = \frac{k(\boldsymbol{x}, \boldsymbol{y})}{\int_{\boldsymbol{\Omega}} k(\boldsymbol{x}, \boldsymbol{y}) d\mu(\boldsymbol{\Omega})}, \tag{2.133}$$

which is the continuous-space counterpart of (2.132).

The above procedure defines a row-stochastic transition matrix, $\boldsymbol{P} = [p_{ij}]$, which encapsulates the information about the neighborhoods of the points. Note that by raising $\boldsymbol{P}$ to the power of $t = 1, 2, \ldots$ we get the jumping probabilities in t steps. This way, the underlying geometry is revealed through high or low transition probabilities between the points, i.e. paths that follow the underlying geometry have a high probability of occurrence, while paths away from the "true" embedded structure have a low probability. Note that the t-step transition probabilities, say $p_t(\boldsymbol{x}_i, \boldsymbol{x}_j)$, satisfy the Chapman-Kolmogorov equation:

$$p_{t_1+t_2}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sum_{\boldsymbol{x}_k \in \boldsymbol{X}} p_{t_1}(\boldsymbol{x}_i, \boldsymbol{x}_k) p_{t_2}(\boldsymbol{x}_k, \boldsymbol{x}_j). \tag{2.134}$$

The Markov process defined by the probability matrix $\boldsymbol{P}$ has a stationary distribution given by

$$\pi(\boldsymbol{x}_i) = \frac{deg(\boldsymbol{x}_i)}{\sum_{\boldsymbol{x}_j \in \boldsymbol{X}} deg(\boldsymbol{x}_j)}, \tag{2.135}$$

and it is reversible, i.e.

$$\pi(\boldsymbol{x}_i) p(\boldsymbol{x}_i, \boldsymbol{x}_j) = \pi(\boldsymbol{x}_j) p(\boldsymbol{x}_j, \boldsymbol{x}_i), \quad \forall \boldsymbol{x}_i, \boldsymbol{x}_i \in \boldsymbol{X}. \tag{2.136}$$

Furthermore, if the kernel is appropriately chosen so that the graph is connected, then the Markov chain is ergodic and irreducible. The Brouwer Fixed Point Theorem Kellogg, Li, and Yorke, 1976 implies that the transition matrix of an ergodic process has a stationary vector $\boldsymbol{\pi}$ such that:

$$\boldsymbol{P}^T \boldsymbol{\pi} = \boldsymbol{\pi}, \tag{2.137}$$

and hence the matrix of a Markov ergodic chain has always an eigenvalue 1 corresponding to the stationary state. From the Perron-Frobenius Theorem, we know that its geometric multiplicity is 1. It can be shown that all other eigenvalues have a magnitude smaller than 1 Coifman and Lafon, 2006a. From (2.132) we get

$$\boldsymbol{P} = \boldsymbol{D}^{-1} \boldsymbol{K}, \quad \boldsymbol{D} = \operatorname{diag}\left(\sum_{j=1}^{N} k_{ij}\right), \tag{2.138}$$

where $k_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$. We note that the transition matrix $\boldsymbol{P}$ is similar to the symmetric positive definite matrix $\boldsymbol{S} = \boldsymbol{D}^{-1/2} \boldsymbol{K} \boldsymbol{D}^{-1/2}$. Thus, the transition matrix $\boldsymbol{P}$ has a decomposition given by

$$\boldsymbol{P} = \sum_{i=1}^{N} \lambda_i \boldsymbol{u}_i \boldsymbol{v}_i^T, \tag{2.139}$$

where $\lambda_i \in \mathbb{R}$ are the (positive) eigenvalues of $\boldsymbol{P}$, $\boldsymbol{u}_i \in \mathbb{R}^N$ are the left eigenvectors, and $\boldsymbol{v}_i \in \mathbb{R}^N$ are the right eigenvectors, such that $\langle \boldsymbol{u}_i, \boldsymbol{v}_j \rangle = \delta_{ij}$.

The set of right eigenvectors $\boldsymbol{v}_i$ establishes an orthonormal basis for the subspace $R(\boldsymbol{P}^T)$ of $\mathbb{R}^D$ spanned by the rows of $\boldsymbol{P}$. Row i represents the transition probabilities from point $\boldsymbol{x}_i$ to all the other points of the graph. According to the Eckart-Young-Mirsky Theorem Eckart and Young, 1936; Mirsky, 1960, the best d-dimensional low-rank approximation of the row space of $\boldsymbol{P}$ in the Euclidean space $\mathbb{R}^d$ is provided by its d right eigenvectors corresponding to the d largest eigenvalues.

It has been shown that asymptotically, when the number of data points uniformly sampled from a low dimensional manifold goes to infinity, the matrix $\frac{1}{\sigma}(\boldsymbol{I} - \boldsymbol{P})$ (where σ is the kernel function scale) approaches the Laplace-Beltrami operator of the underlying Riemannian manifold Coifman et al., 2005; Nadler et al., 2006. This allows to consider the eigenvectors corresponding to the largest few eigenvalues of $\boldsymbol{P}$ as discrete approximations of the principal eigenfunctions of the Laplace-Beltrami operator. Since the principal eigenfunctions of the Laplace-Beltrami operator establish an accurate embedding of the manifold Jones, Maggioni, and Schul, 2008, this result promotes the usage of the eigenvectors of $\boldsymbol{P}$ for practical (discrete) data embedding.

At this point, the so-called diffusion distance, which is an affinity distance related to the reachability between two points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, is given by

$$D_t^2(\boldsymbol{x}_i, \boldsymbol{x}_j) = ||p_t(\boldsymbol{x}_i, \cdot), p_t(\boldsymbol{x}_j, \cdot)||^2_{L_2, 1/deg} = \sum_{i=1}^{N} \frac{(p_t(\boldsymbol{x}_i, \boldsymbol{x}_k) - p_t(\boldsymbol{x}_j, \boldsymbol{x}_k))^2}{deg(\boldsymbol{x}_k)}, \tag{2.140}$$

where $p_t(\boldsymbol{x}_i, \cdot)$ is the i-th row of $\boldsymbol{P}^t$. The embedding of the t-step transition probabilities is achieved by forming a family of maps (DMs) of the N points $\boldsymbol{x}_i \in \boldsymbol{X}$ in the Euclidean subspace of $\mathbb{R}^d$ defined by

$$\boldsymbol{\mathcal{R}}_t(\boldsymbol{x}_i) = \boldsymbol{y}_i = \left[\lambda_1^t \boldsymbol{v}_1(i), \lambda_2^t \boldsymbol{v}_2(i), \ldots, \lambda_d^t \boldsymbol{v}_d(i)\right]^T, i = 1, 2, \ldots N. \tag{2.141}$$

A useful property of DMs is that the Euclidean distance in the embedded space $||\boldsymbol{\mathcal{R}}_t(\boldsymbol{x}_i) - \boldsymbol{\mathcal{R}}_t(\boldsymbol{x}_j)||_{L_2}$ is the best d-dimensional approximation of the diffusion distance, given by (2.140), where the equality holds for $d = N$.

2.4 The Pre-Image Problem ("Lifting")

The "Lifting" operation is performed, in order to reconstruct the high-dimensional space from measurements on the manifold, i.e., to “lift” the predictions made by the reduced-order models on the manifold back to the original high-dimensional space. The main motivation is based on the fact that if nonlinear embedding methodologies have been performed towards the discovery of the proper underlying manifold, then the structure of the mapped data points might be of no use (i.e. might no have physical (practical) meaning) in applications’ specific problems’ solving.

In the linear case of PCA, this task is trivial, as the solution to the reconstruction problem, given by

$$\underset{\boldsymbol{U}_d \in \mathbb{R}^{D\times d}}{\operatorname{argmin}} \sum_{i=1}^{N} ||\boldsymbol{x}_i - \boldsymbol{U}_d \boldsymbol{U}_d^T \boldsymbol{x}_i||^2_{L_2}, \quad \boldsymbol{U}_d^T \boldsymbol{U}_d = I, \tag{2.142}$$

is just a linear transformation which maximizes the variance of the data on the linear manifold, and is given by the first d principal eigenvectors of the covariance matrix, while in the case of nonlinear manifold learning algorithms, (such as LLE and DMs), we want to learn the inverse map (lifting operator):

$$\boldsymbol{\mathcal{L}} \equiv \boldsymbol{\mathcal{R}}^{-1} : \boldsymbol{\mathcal{R}}(\boldsymbol{X}) \rightarrow \boldsymbol{X}, \tag{2.143}$$

for new samples on the manifold $\boldsymbol{y}_* \notin \boldsymbol{\mathcal{R}}(\boldsymbol{X})$. This inverse problem is referred as “out-of-sample extension pre-image” problem. Following an excellent overview of the problem in the work of Honeine and Richard, 2011 "Preimage problem in kernel-based machine learning", Honeine and Richard, 2011, we dive a bit deeper to its formulation.

One of the first interests and proposed approaches towards tackling this problem, was initiated by Aizerman, Braverman, and Rozonoer, 1964, who introduced the so called "kernel trick", based on the idea that the data of analysis might be modelled with greater efficiency if a similarity measure is applied on them on a pre-processing stage (actually an effort towards "denoising" the initial data). Such a similarity measure can be the inner product, or any other reproducing kernel with specific properties and conditions regarding its form. The first condition is for the kernel to be a positive semi-definite symmetric function, which is a condition that actually shifts the whole problem from the original to the feature space, usually without extra computational complexity costs compared to the linear approaches. Secondly, the kernel must be defining an inner product in some space, which is sufficienclty ensured by Mercer's theorem Mercer, 1909, as explicitly stated in Honeine and Richard, 2011 :

Theorem : Any positive semi-definite kernel can be expressed as an inner product in some space, where the positive semi-definiteness of a kernel $k : X \times X \to R$ is determined by the property

$$\sum_{i,j} \alpha_i \alpha_j k(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq 0 \tag{2.144}$$

$\forall \alpha_i \alpha_j \in R$ and $\boldsymbol{x}_i, \boldsymbol{x}_j \in X$ (X denoting the original space). A unique reproducing kernel Hilbert Space, with the inner product $\langle \cdot, \cdot \rangle$ being the conditioned by 2.144 kernel, as well, is also proved to be associated with the kernel as per Aronszajn, 1950. Again in Honeine and Richard, 2011 is stated that in this space of functions, point evaluations are bounded, and as such the reproducing kernel is guaranteed to exist and be unique, by Riesz representation Theorem, as well. These conditions have provided the necessary mathematical rigorousness and formalism, towards the reproducing kernels concepts to be able to be used in several applications related to time series modelling and interpolation, E Parzen, 1970, Kailath, 1971, S. and Wahba, 1991.

Mathematically formulating this, let X be the original (compact) data space and H be a Hilbert space of functions defined on it and let $\psi(\boldsymbol{x}) \in H$ be an evaluation function bounded $\forall \boldsymbol{x} \in X$. The theorem states that there exists a unique function $\phi(\boldsymbol{x}) \in H$ s.t.

$$\psi(\boldsymbol{x}) = \langle \psi, \phi(\boldsymbol{x}) \rangle_H \tag{2.145}$$

and also regarding the reproducing kernel, the function also satisfies the :

$$k(\boldsymbol{x_i}, \boldsymbol{x_j}) = \langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j) \rangle_H \tag{2.146}$$

for any $\boldsymbol{x}_i, \boldsymbol{x}_j \in X$ and under the "kernel trick" approach, we also have :

$$\begin{aligned} &\|\phi(\boldsymbol{x}_i) - \phi(\boldsymbol{x}_j)\|_H^2 = \\ &\langle \phi(\boldsymbol{x}_i) - \phi(\boldsymbol{x}_j), \phi(\boldsymbol{x}_i) - \phi(\boldsymbol{x}_j) \rangle_H = \\ &k(\boldsymbol{x_i}, \boldsymbol{x_i}) - 2k(\boldsymbol{x_i}, \boldsymbol{x_j}) + k(\boldsymbol{x_j}, \boldsymbol{x_j}) \end{aligned} \tag{2.147}$$

with $\| \cdot \|_H^2$ being the induced norm in the reproducing kernel Hilbert space.

While kernels are more usually restricted to vectorial inputs in vectorial spaces, endowed with the Euclidean input product $\langle \boldsymbol{x_i}, \boldsymbol{x_j} \rangle = \boldsymbol{x_i}^T \boldsymbol{x_j}$ and associated norm $\|\boldsymbol{x_i}\|$, many other categories have been introduced over the last years of machine learning revolution. Projective kernels, i.e. functions of inner product such as :

1. the monomial kernel :
$$(\langle \boldsymbol{x_i}, \boldsymbol{x_j} \rangle)^p \tag{2.148}$$

2. the polynomial kernel :

$$(c + \langle \boldsymbol{x_i}, \boldsymbol{x_j} \rangle)^p \tag{2.149}$$

3. the exponential kernel :

$$exp(\frac{\langle \boldsymbol{x_i}, \boldsymbol{x_j} \rangle}{2\sigma^2}) \tag{2.150}$$

4. the sigmoid kernel (perceptron) :

$$tanh(\frac{\langle \boldsymbol{x_i}, \boldsymbol{x_j} \rangle}{\sigma} + c) \tag{2.151}$$

and Radial kernels, with the list including :

1. the Gaussian kernel :

$$exp(-\frac{\|\boldsymbol{x_i} - \boldsymbol{x_j}\|^2}{2\sigma^2}) \tag{2.152}$$

2. the Laplacian kernel :

$$exp(-\frac{\|\boldsymbol{x_i} - \boldsymbol{x_j}\|}{2\sigma^2}) \tag{2.153}$$

3. the Multiquadratic kernel :

$$\sqrt{\|\boldsymbol{x_i} - \boldsymbol{x_j}\|^2 + c} \tag{2.154}$$

4. the Inverse Multiquadratic kernel :

$$\frac{1}{\sqrt{\|\boldsymbol{x_i} - \boldsymbol{x_j}\|^2 + c}} \tag{2.155}$$

One of the main pioneering theorems around kernel-based machine learning is also the so called "Representer Theorem", again thoroughly investigated in Honeine and Richard, 2011 :

Theorem : Given a set of input observations $\boldsymbol{x}_i \in X, \quad i = 1, \ldots, n$ and targets (labels) $y_i, \quad i = 1, \ldots, n$, and denoting as $f(\cdot)$ a cost function and $g(\cdot)$ a monotonically increasing function on R_+, then for any function $\psi \in H$, minimizing the (regularized) cost function :

$$\sum_{i=1}^{n} f(y_i, \psi(\boldsymbol{x}_i)) + \eta g(\|\psi\|_H^2) \tag{2.156}$$

the function ϕ is expressed in terms of the given data points, as follows :

$$\psi = \sum_{i=1}^{n} \alpha_i \phi(\boldsymbol{x}_i) \tag{2.157}$$

displaying that the data processing can now be performed in the subspace spanned by the mapped versions of the training given data.

In general, the solution of the pre-image problem can be posed as:

$$\arg\min_{\boldsymbol{c}} ||\boldsymbol{y}_* - \boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{L}}(\boldsymbol{y}_*)|\boldsymbol{c}))||, \tag{2.158}$$

subject to a constraint, where $\boldsymbol{\mathcal{L}}(\cdot)|\boldsymbol{c}$ is the lifting operator depending on some parameters $\boldsymbol{c}$.

In the case, where the kernel are represented as invertible functions, like e.g. the case of projective kernels (e.g. polynomial kernels with odd degree and also the sigmoid one), the solution to the pre-image optimization problem defined in 2.158 exists and is usually easily computed in exact form.

More specifically, if $h : R \to R$ defines the inverse function and :

$$h(k(\boldsymbol{x}_i, \boldsymbol{x}_j)) = \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle \tag{2.159}$$

then every element $\boldsymbol{x} \in X$ can be expressed as :

$$\boldsymbol{x} = \sum_{j=1}^{N} \langle \boldsymbol{e}_j, \boldsymbol{x} \rangle \boldsymbol{e}_j = \sum_{j=1}^{N} h(k(\boldsymbol{e}_j, \boldsymbol{x}))\boldsymbol{e}_j \tag{2.160}$$

for any given orthonormal basis $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_N\}$ in the input space.

In that case the exact solution x_* to the pre-image problem associated with 2.157, or expressed as $\phi(\boldsymbol{x}_*) = \psi$, is given by :

$$\boldsymbol{x}_* = \sum_{j=1}^{N} h(\sum_{i=1}^{n} \alpha_i k(\boldsymbol{e}_j, \boldsymbol{x}_i))\boldsymbol{e}_j \tag{2.161}$$

Particularly, focusing on kernels being invertible functions of the distance, like the radial kernels and using the polarization identity

$$4\langle \boldsymbol{x}_*, \boldsymbol{e}_j \rangle = \|\boldsymbol{x}_* + \boldsymbol{e}_j\|^2 - \|\boldsymbol{x}_* - \boldsymbol{e}_j\|^2 \tag{2.162}$$

given by Schölkopf, Smola, and Müller, 1998, a similar solution to 2.161 is obtained.

In the case exact solutions cannot be explicitly found, an armory of numerical methods has been developed towards exploring approaximations of them, by solving the optimization problem in 2.158, including the Gradient descent technique, Fixed-point iteration method, MDS-based technique, the Conformal map approach or even purely data driven learning machines with training elements being the feature space's (manifold) mapped points (from the embeddings) and the estimated values the input space's (original space) ones. All of these approaches are thoroughly discussed and investigated in Honeine and Richard, 2011 and Kwok and Tsang, 2004.

Below, we describe the reconstruction methods that we used in this work, namely, Radial Basis Functions (RBFs) interpolation and Geometric Harmonics (GH), also giving some insight about their implementation and pros and cons. While RBFs are based on the previously outlined theoretical framework of reproducing kernels, GH is also based on the so called "out-of-sample extension" problem, usually referring to the direct problem instead, i.e. that of learning the direct embedding map (i.e. the restrictions to the manifold) $\boldsymbol{\mathcal{R}}(\boldsymbol{X}) : \boldsymbol{X} \to \boldsymbol{\mathcal{R}}$, for new samples in the input space $\boldsymbol{x}_* \notin \boldsymbol{X}$. Towards this aim, a well established methodology is the Nyström extension Nyström, 1929. For a review and a comparison of such methods in the framework of chemical kinetics see Chiavazzo et al., 2014.

2.4.1 Radial Basis Function (RBF) Interpolation

The lifting operator is constructed with interpolation through RBFs among the corresponding set of neighbors of the new state $\boldsymbol{y}_*$ in the ambient space. The lifting operator is defined by Chiavazzo et al., 2014:

$$\mathcal{L}(\boldsymbol{y}_*) = x_{i*} = \sum_{j=1}^{k} c_{ji}\psi(||\boldsymbol{y}_* - \boldsymbol{y}_j||), \quad i = 1, 2, \ldots, D, \tag{2.163}$$

where x_{i*} is the i-th coordinate of $\boldsymbol{x}_*$, the $\boldsymbol{y}_j$'s are the neighbors of the unseen sample $\boldsymbol{y}_*$ on the manifold, and ψ is the radial basis function.

Similarly, the restriction operator can be written as:

$$\mathcal{R}(\boldsymbol{x}_*) = y_{i*} = \sum_{j=1}^{k} c_{ji}\psi(||\mathcal{L}(\boldsymbol{y}_*) - \mathcal{L}(\boldsymbol{y}_j)||), \quad i = 1, 2, \ldots, D, \tag{2.164}$$

where $\mathcal{L}(\boldsymbol{y}_j) = \boldsymbol{x}_j$ is the known image of $\boldsymbol{y}_j$ (the neighbors of $\boldsymbol{y}_*$ in the ambient space). The unknown coefficients of the lifting operator can be computed by setting at the left hand side of (2.163) the coordinates of the one-to-one corresponding neighbors of $\boldsymbol{y}_*$ in the ambient space. Here, we consider a special class of RBFs known as radial powers, given by

$$\psi(||\mathcal{L}(\boldsymbol{y}_*) - \mathcal{L}(\boldsymbol{y}_j)||) = ||\mathcal{L}(\boldsymbol{y}_*) - \mathcal{L}(\boldsymbol{y}_j)||_{L_2}^{p},$$

where p is an odd integer. By doing so, the unknown coefficients c_{ji} of the lifting operator are given by the solution of the following linear system:

$$\boldsymbol{A}\begin{bmatrix} c_{1i} \\ c_{2i} \\ \vdots \\ c_{ki} \end{bmatrix} = \begin{bmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ki} \end{bmatrix}, \quad i = 1, 2, \ldots, D, \tag{2.165}$$

where

$$\boldsymbol{A} = \begin{bmatrix} 0 & ||\boldsymbol{y}_1 - \boldsymbol{y}_2||_{L_2}^{p} & \ldots & ||\boldsymbol{y}_1 - \boldsymbol{y}_k||_{L_2}^{p} \\ ||\boldsymbol{y}_2 - \boldsymbol{y}_1||_{L_2}^{p} & 0 & \ldots & ||\boldsymbol{y}_2 - \boldsymbol{y}_k||_{L_2}^{p} \\ \vdots & \vdots & \ldots & \vdots \\ ||\boldsymbol{y}_k - \boldsymbol{y}_1||_{L_2}^{p} & ||\boldsymbol{y}_k - \boldsymbol{y}_2||_{L_2}^{p} & \ldots & 0 \end{bmatrix}, \tag{2.166}$$

and x_{ji} is the i-th coordinate of the j-th point in the ambient space, whose restriction on the manifold is the j-th nearest neighbor of $\boldsymbol{y}_*$. For our computations, we used $p = 1$. Then, (2.163) can be used to find the coordinates of $\boldsymbol{x}_*$ in the ambient space.

It has been shown by Poggio and Girosi, 1989, Poggio and Girosi, 1990, Girosi and Poggio, 1990 and Girosi, Jones, and Poggio, 1993 that regularization principles as in Equation 2.156 lead to approximation schemes which are equivalent to one-hidden layered networks, also called "Regularization Networks (RN)". More specifically, the Radial Basis Functions, (Micchelli, 1986, Buhmann and Levesley, 2004, Biancolini, 2018), can be considered to be single-layered artificial neural networks, usually called Radial Basis Function Networks (RBFN), with the activation function being given by the radial basis function expression. In that case the coefficients in 2.165 are "learned" using the armory of iterative methods for neural networks, respectively.

Several types of RBFs have been proposed over the years, being used in several applications in engineering, computer graphics, signal processing (Buhmann and Levesley, 2004), as well as time series prediction with RBFN, (Hutchinson, 1994, Alexandridis, Famelis, and Tsitouras, 2015, Leung, Lo, and Wang, 2001, Hutchinson, 1994) with an indicative list including :

1. Gaussian :
$$\phi(\|x_i - x_j\|) = e^{-(\epsilon\|x_i - x_j\|)^2} \tag{2.167}$$

2. Multiquadric :
$$\phi(\|x_i - x_j\|) = \sqrt{1 + (\epsilon\phi(\|x_i - x_j\|))^2} \tag{2.168}$$

3. Inverse Quadratic :
$$\phi(\|x_i - x_j\|) = \frac{1}{1 + (\epsilon\|x_i - x_j\|)^2} \tag{2.169}$$

4. Inverse Multiquadric :
$$\phi(\|x_i - x_j\|) = \frac{1}{\sqrt{1 + (\epsilon\phi(\|x_i - x_j\|))^2}} \tag{2.170}$$

5. Polyharmonic Spline :
$$\begin{aligned} \phi(\|x_i - x_j\|) &= (\|x_i - x_j\|)^k, \quad k = 1,3,5,\dots \\ \phi(\|x_i - x_j\|) &= (\|x_i - x_j\|)^k ln(\|x_i - x_j\|), \quad k = 2,4,6,\dots \end{aligned} \tag{2.171}$$

6. Thin plate spline :
$$\phi(\|x_i - x_j\|) = (\|x_i - x_j\|)^2 ln(\|x_i - x_j\|) \tag{2.172}$$

7. Bump :
$$\phi(\|x_i - x_j\|) = \begin{cases} e^{(}-\frac{1}{1-(\epsilon\|x_i - x_j\|)^2}), & \text{for} \quad r < \frac{1}{\epsilon} \\ 0, & \text{otherwise} \end{cases} \tag{2.173}$$

with ϵ being the so called "shape parameter" of the radial kernels (Micchelli, 1986, Buhmann and Levesley, 2004, Biancolini, 2018).

We note radial powers are better for the task of interpolation compared to Gaussian kernels, as they do not suffer from ill conditioning. The matrix A is invertible with the assumption that the centers y_j are distinct (see, e.g., Monnig, Fornberg, and Meyer, 2014; Amorim et al., 2015). However, A can be rank deficient from a numerical point of view, e.g. because the points may be very close to each other. Thus, the solution of (2.166) using a method such as the LU factorization of A may result in numerical instabilities.

2.4.2 Geometric Harmonics (GHs)

GHs are a set of functions that allow the extension of the embedding of new unseen points on the manifold $x_* \notin X$, which are not given in the set of points used for building the embedding Coifman and Lafon, 2006b. Their derivation is based on the

Nyström (or quadratic) extension method Nyström, 1929, which has been used for the numerical solutions of integral equations Delves and Mohamed, 1988; Press and Teukolsky, 1990, and in particular the Fredholm equation of second kind, reading:

$$f(t) = g(t) + \mu \int_a^b k(t,s) f(s) ds, \tag{2.174}$$

where $f(t)$ is the unknown function, while $k(t,s)$ and $g(t)$ are known. The Nyström method starts with the approximation of the integral, i.e.

$$\int_a^b y(s) ds \approx \sum_{j=1}^{N} w_j y(s_j), \tag{2.175}$$

where s_j are N appropriately chosen collocation points and w_j are the corresponding weights, which are determined, e.g., by the Gauss-Jacobi quadrature rule. Then, by using (2.175) in (2.174) and evaluating f and g at the N collocation points, we get the following approximation:

$$(\boldsymbol{I} - \mu \tilde{\boldsymbol{K}}) \hat{\boldsymbol{f}} = \boldsymbol{g}, \tag{2.176}$$

where the matrix $\tilde{\boldsymbol{K}}$ has elements $\tilde{k}_{ij} = k(s_i, s_j) w_j$. Based on the above, the solution of the homogenous Fredholm problem ($\boldsymbol{g} = 0$) is given by the solution of the eigenvalue problem

$$\tilde{\boldsymbol{K}} \hat{\boldsymbol{f}} = \frac{1}{\mu} \hat{\boldsymbol{f}}, \tag{2.177}$$

i.e.

$$\sum_{j=1}^{N} w_j k(s_i, s_j) \hat{f}_j = \frac{1}{\mu} \hat{f}_i, \quad i = 1, 2, \ldots, N, \tag{2.178}$$

where $\hat{f}_i = \hat{f}(s_i)$ is the i-th component of $\hat{\boldsymbol{f}}$. The Nyström extension of $f(t)$, using a set of N sample (collocation) points, at an arbitrary point x in the full domain is given by:

$$\boldsymbol{\mathcal{E}}(f(x)) = \hat{f}(x) = \mu \sum_{j=1}^{N} w_j k(x, s_j) \hat{f}_j. \tag{2.179}$$

Within the framework of DMs, we seek for the out-of-the-sample (filtered) extension of a real-valued function f defined at N sample points $\boldsymbol{x}_i \in \boldsymbol{X}$ to one or more unseen points $\boldsymbol{x}_* \notin \boldsymbol{X}$. The function f can be for example a DM coordinate $\lambda_j^t \boldsymbol{v}_j(\boldsymbol{x}_i)$, $j = 1, 2, \ldots d$, or another function representing the output of a regression model (see also the discussion in Thiem et al., 2020; Rabin et al., 2016.

Recalling that the eigenvectors $\boldsymbol{v}_j$ form a basis, the extension is implemented (see Coifman and Lafon, 2006b) by first expanding $f(\boldsymbol{x}_i)$ in the first d parsimonious eigenvectors $\boldsymbol{v}_l$ of the Markovian matrix $\boldsymbol{P}^t$:

$$\hat{f}(\boldsymbol{x}_i) = \sum_{l=1}^{d} a_l \boldsymbol{v}_l(\boldsymbol{x}_i), \quad i = 1, 2, \ldots N,$$

where $a_l = \langle \boldsymbol{v}_l, \boldsymbol{f} \rangle$ are the projection coefficients of the function on the first d parsimonious eigenvectors, and $\boldsymbol{f} \in \mathbb{R}^N$ is the vector containing the values of the function f at the N points $\boldsymbol{x}_i$. Then, one computes the Nyström extension of $\hat{f}$ at $\boldsymbol{x}_*$ using the same projection coefficients as

$$\mathcal{E}(\hat{f}(\boldsymbol{x}_*)) = \sum_{l=1}^{d} a_l \hat{\boldsymbol{v}}_l(\boldsymbol{x}_*), \tag{2.180}$$

where

$$\hat{\boldsymbol{v}}_l(\boldsymbol{x}_*) = \frac{1}{\lambda_l^t} \sum_{j=1}^{N} k(\boldsymbol{x}_j, \boldsymbol{x}_*) \boldsymbol{v}_l(\boldsymbol{x}_j), \quad l = 1, 2, \ldots, d. \tag{2.181}$$

are the corresponding GHs. Scaling up the (filtered) extension of the function f to a set of say L new points can be computed using the following matrix product (Coifman and Lafon, 2006b; Thiem et al., 2020):

$$\mathcal{E}(\hat{\boldsymbol{f}}) = \boldsymbol{K}_{L\times N} \boldsymbol{V}_{N\times d} \boldsymbol{\Lambda}_{d\times d}^{-1} \boldsymbol{V}_{d\times N}^{T} \boldsymbol{f}_{N\times 1}, \tag{2.182}$$

where $\boldsymbol{K}_{L\times N}$ is the corresponding kernel matrix, $\boldsymbol{V}_{N\times d}$ is the matrix with columns the d parsimonious eigenvectors $\boldsymbol{v}_l$, $\boldsymbol{\Lambda}_{d\times d}$ is the diagonal matrix with elements λ_l^t.

The above direct approach provides a map from the ambient space to the reduced-order space (restriction) and vice versa (lifting).

Chapter 3

Applications

3.1 Electricity Markets - Load Demand and Price Forecasting

Electricity Price and Load forecasting, since the early 1990s, has increasingly and steadily become the two most crucial not only for researchers but also for the people working in Electricity companies, since these two fundamental operational parameters play a vital role in decision making. Since that period mentioned before, the landscape of traditionally monopolistic and government-controlled power sectors has drastically changed to a competitive market. This change has a significant economic impact on producers and consumers. Worldwide, electricity is now traded under market rules that resemble those of a financial market, using spot (wholesale) price and derivative contracts. However, there are major differences between electricity and financial markets, and this is so because electricity is a very special commodity with very particular stylized facts: it is economically non-storable, and the power system that generates, transmits and distributes electricity must ensure a constant balance of production and consumption, a characteristic that is not present in the financial markets.

'Spot market' market is typically a day-ahead (DA) market that does not allow for continuous trading, a constraint imposed by the System Operators which require an advance notice so they can verify that the generation-consumption schedule is feasible and within the transmission constraints. Thus, in the DA market, agents submit their bids and offers for the physical delivery of power during each hour load period of the next day. Therefore, considering forecasting of intraday electricity prices, we must keep in our mind that in the majority of electricity markets, prices for all contracts of the next day are determined at the same time using the same available information (for example on the load forecasts for next day).

Electricity load (demand), similarly, depends heavily on a number of uncontrolled, exogenous and extremely stochastic factors, as weather temperature, wind speed, precipitation etc., as well as on factors related to business cycles and intensity (e.g. peak vs off-peak hours, work days vs. weekends, holidays vs regular days etc.). Those stochastic and seasonal coexisting factors make the dynamic behavior of both electricity price and load unique and specific that no other market shares. The interaction of stochastic and seasonal (at the daily, weekly, annual time-scales) factors generate very often unpredictable price spikes of tremendous impact on the operation of various agents competing in an electricity market, making them vulnerable since they cannot pass their cost on the retail consumers. These costs for buying (via over- or under-contracting) and then selling power in the real-time market can become so high (due to previously mentioned stochastic and unpredictable variables) that they can force electricity companies to exhibit huge economic losses or even collapse. Another very peculiar stylized fact of the electricity price is its volatility that can be up to two order of magnitude higher than that of a financial commodity, a fact that forces the players of the electricity market to use derivative product to hedge not only against quantity-volume risk but also against the price volatility. This situation makes obvious the need for reliable price and load forecasting to help power portfolio managers in their decision making process. For example, a generating (electricity) company or a large industrial (consumer) company will have huge benefits from a reliable and accurate forecasting of the price volatility since they will be possible to adjust their bidding strategies and generation or consumption program more effectively, aiming at minimizing the risks and maximizing profits in day-ahead trading. Another fact that makes electricity price and load forecasting extremely important is the recently observed massive Renewable Energy Sources, RES (mainly wind and solar) penetration in the generation mixture of the power systems. For example, as wind power forecasting depends on the aforementioned stochastic weather factors of wind speed, precipitation, temperature, etc., this definitely increases the stochasticity, complexity and difficulty in making predictions thus bearing some inauspicious effects on the market operations. So once again, the improvements of the forecasting methods in this market are necessary and attract a large number of researchers, market participants as well as Governmental and Regulating bodies Policy and decision makers. The complexity of forecasting electricity markets' price and load depends on the increasing number of both endogenous and exogenous variables employed as inputs in constructing 'sophisticated' models for more accurate and reliable predictions, with various modeling approaches of varying degree of success, as follows:

1. Multi-agent (game theoretic, multi-agent simulations) models which simulate the interaction of a variety of heterogeneous 'players' in the operation of the power system, that forms the price dynamics by equilibrating supply and demand

2. fundamental (structural) methods , which describe the price dynamic evolution by incorporating a variety of fundamental physical, regulatory and economic factors that affects, via their interaction, the price

3. reduced-form (quantitative, stochastic models) , describing the time-varying properties of electricity prices

4. Statistical (econometric, technical analysis), methods, which in fact are direct applications of statistical methods of load forecasting , or the use of well-known econometric models in the electricity markets, and finally 5) Computational intelligence (artificial intelligence-based, non-parametric, non-linear statistical) approaches , appropriate in describing complex dynamic systems with an 'intelligent' way combining elements of learning, fuzziness and evolution

In these works we apply a number of price and load forecasting models that are pure members of the above main classification groups or 'hybrid' models made by combining models belonging to different groups. More specifically, our models of price and load forecasting are based on the approaches enumerated above.

In the paper Papaioannou et al., 2016, regarding load forecasting, we aimed at analyzing, modelling and forecasting the daily electricity system wide load (demand) in the Greek electricity market, using data from the period 2004-2014. To this end, we utilized univariate (as well multivariate, Principal Components regression) methods in order to capture the intra-day, intra-week and annual seasonal patterns. Due to this seasonality capturing target, it was natural to adopt some typical representative models, available for this purpose, as the Holt-Winters exponential smoothing or its generalization Error-Trend-Seasonal, ETS, and the double seasonal ARMA models that are shown to be competitive. In summary, the forecasting models considered in this thesis include, Artificial Neural Networks (ANN), Support Vector Machines (SVM), ETS models, the Seasonal Autoregressive Integrated Moving Average with exogenous variables SARIMAX and the hybrid nonparametric Principal Component Regression model, which is a combination of the linear tool of manifold learning Principal Component with a multiple linear regression (MLR). This is a manifestation of the previously stated 'hybrid' approach in modeling, by using models of the different classification groups, described above. We have used a number of criteria for measuring the quality of the produced forecasts, actually by comparing the forecasting effectiveness of models by assessing their predictions made on different estimation and Out-of-Sample time periods. The main finding of the paper was that the 'hybrid' models of SARIMAX and Principal Components (PC) regression, produce forecasts that clearly outperform in quality and accuracy the forecasts of typical, 'main-stream' models, working in isolation, as the exponential smoothing, SARIMAX, ETS and SVM models (which behave poorly in forecasting when fitted on daily data). The contribution of the paper is that 'blending' traditional linear stochastic models with manifold learning models produces more efficient models in forecasting load (demand) in electricity markets.

The paper, Papaioannou et al., 2018, is about detecting the impact on the dynamic evolution of the wholesale or System Marginal Price (SMP) in the Greek electricity market GEM, of fundamental (endogenous, of the system) factors as well as of factors related to various regulatory reforms of the Regulator and/or Government, as well as factors related to other, co-evolving markets, as the Carbon dioxide European market, that exerts a strong influence in the electricity price formation, via affecting the costs of the thermal generating units. We considered the SMP as the footprint of an underlying stochastic and nonlinear process, reflecting not only the effects of endogenous/fundamental market factors but also the effects of exogenous variables as well as regulatory reforms, which also affect the market dynamics. We applied a number of SARMAX/GARCH (of different orders) using data from January 2004 to December of 2014. In order to capture the dynamic behavior of conditional volatility, an extremely crucial parameter in 'writing' derivative contracts , we have combine the SARMAX model to capture conditional mean, with the GARCH model to capture volatility. The 'X' letter in SARMAX stand for the exogenous variables as quantitative (numerical) regressors-predictors, as well as dummy variables that mimic the history of Regulator's interventions. The main findings of this work is that the best found model (based on a number of criteria, as AIC , BIC information criteria) captures adequately the dependency of the spot price to the regulatory reforms that have imposed on the GEM, revealing the impact that this reforms have on the dynamics of SMP. Another finding is that the model reliably 'predicts' the sign and the magnitude of the effect of quantitative fundamentals on the dynamic evolution of SMP. The model, as it provides good evidence of the dynamics of the conditional volatility, is a promising and useful tool (after adjusting to new information from the market) to the decision makers, towards restructuring the GEM in compliance with the requirements of the European Target Model, a Pan-European energy market that homogenizes the various member states energy markets.

3.2 Using the Hurst Exponent, GHE, DFA and rolling Tsallis Entropy (TE) to assess the efficient market hypothesis (EMH) for the Greek and Italian electricity markets.

In this application, we extend the previous framework of Papaioannou et al., 2016 and Papaioannou et al., 2018, by assessing the complexity of the Greek wholesale electricity market, but also of the Italian, by using Hurst Exponent and Tsallis Entropy measures. Thus, in an attempt to interpret more accurately the dynamics of their spot prices (SMP, and PUN respectively), we must first discuss briefly the structure of their respective electricity markets and their different operational principles. The two aforementioned countries operate under the European Directives, which have deregulated the electricity sector, albeit at different levels of maturity. In brief, the so-called Target Model for the electricity markets in Europe comprises of the following four different markets:

1. Forward and future market
2. Day-ahead (or spot) market
3. Intra-day market
4. Balancing market (For a detailed description for the current electricity market design in Europe we refer the interested reader to kuleuven, n.d.)

It is natural, therefore, to anticipate a significant level of complexity of the future Pan-European (target model), since it will be structured by incorporating the four aforementioned markets of each European member state, of different maturity and 'depth', as well as the separate national idiosyncrasies. Greece's liberalized electricity market was established according to the European Directive 96/92/EC. Currently only Day-Ahead market is operational in Greece, which sets Greece as the less mature of the electricity markets analyzed in this study. The Greek electricity market (GEM) is currently in a transitional period, during which the market structure evolves towards its final design, namely the Target Model. The wholesale electricity market is a day ahead mandatory pool which is subject to inter-zonal transmission constraints, unit technical constraints, reserve requirements, the interconnection Net Transfer Capacities (NTCs) and in general all system constraints. More specifically, based on forecasted demand, generators' offers, suppliers' bids, power stations' availabilities, unpriced or must-run production (e.g., hydro power mandatory generation, cogeneration and RES outputs), schedules for interconnection as well as a number of transmission systems' and power stations' technical constraints, an optimization process is followed in order to dispatch the power plant with the lower cost, both for energy and ancillary (i.e. supporting) services. The Italian wholesale electricity market, commonly called the Italian Power Exchange (IPEX), is run by the Gestore del Mercato Elettrico (GME), which consists of the following:

The Spot Electricity Market (MPE) which includes:

1. The Day-Ahead Market (MGP), where all market participants (producers, wholesalers, consumers) trade energy blocks for the next day.
2. The Intra-Day Market (MI) where the initial schedules can be modified by the market participants.

TABLE 3.1: Descriptive Statistics of prices and returns of wholesale electricity prices (2005-2013): SMP (Greece), PUN (Italy)

Descriptive Stats	SMP	SMPret	PUN	PUNret
min	10.23792	-0.96787	21.13875	-0.61298
max	124.99330	1.07460	142.55580	0.79215
Mean	54.74035	0.00034	69.99168	0.00032
Standard Dev	18.24662	0.19893	14.24931	0.16078
Skewness	0.50643	0.14083	0.39766	0.68259
Kurtosis	2.86262	5.32397	3.84237	5.32344
Sample Size	3287	3286	3287	3286

3. The Ancillary Services Market (MSD), which consists of the ex-ante MSD which is a scheduling substage of MSD and of the Balancing Market (MB).

The Forward Electricity Market (MTE), which is the venue where market participants can sell/purchase future electricity supplies via forward electricity contracts. The platform for physical delivery of financial contracts concluded on Italian Derivatives Energy Exchange (IDEX) (CDE).

Data description, descriptive Statistics and tests

We have collected data on the Italian and the Greek electricity markets. The data consists of daily spot prices, spanning over the course of nine years, from January 2005 to December 2013. The daily prices were calculated as the average of the 24 hourly prices for each day. For the Italian market, we have the hourly day-ahead system marginal price, so-called Prezzo Unico Nazionale, or PUN, quoted and publicly available on the Italian Power Exchange IPEX, website (www.mercatoelettrico.org). For the Greek market the hourly spot, day-ahead System wide Marginal Price, SMP, is available at the official site of the Greek Independent Power Transmission Operator, IPTO, ADMIE S.A. (www.admie.gr). In order to conduct a preliminary analysis of the data, we calculated various descriptive statistics for the two markets, both for prices and returns, which are shown in table 3.2 below.

By examining the above table 3.1, is evident that the logarithmic returns deviate significantly from normal distribution. The Italian return series exhibit fat tail (kurtosis is over 3), a phenomenon not exhibited by the Greek returns. In addition, both series are right skewed. The positive skewness of the electricity prices is a well-known phenomenon; this is due to the effect of positive spikes evident in electricity prices, which happens in periods of shortage in the supply. In Table 3.2 below, we perform a series of formal statistical tests on the return series in order to validate our assumptions.

The Jaque-Bera and the Kolmogorov-Smirnov statistical tests, test the null hypothesis of normality. As expected the null hypothesis for both tests is strongly rejected for all the return series at 1% significance level. According to the ADF test, the results confirm to Random Walk (RW) because the null H0 of unit-root cannot be rejected for the series of prices while it is rejected for returns with a level of significance greater that 99%. The prices appear to be integrated of order one (difference stationary), a prerequisite for RW. RW, as it is broadly known, is the ideal model for an efficient market, therefore the first intuitive step taken when examining the efficiency of a market is to search for serial autocorrelation, AC, in the returns. The

TABLE 3.2: Statistical tests and respective p-values for logarithmic returns

Statistical tests	SMP	PUN
Jaque- Berra (normality)	750.32***	994.31***
p-val	0.00	0.00
Kolmogorov-Smirnov (normality)	0.32***	0.36***
p-val	0.00	0.00
Augmented Dickey–Fuller (unit root) ADF Test	-79.68***	-70.24***
p-val	0.00	0.00
Ljung-Box Q-test (residual autocorrelation)	903.61***	2569.30***
p-val	0.00	0.00
Engle's ARCH test (homoscedasticity)	319.41***	85.87***
p-val	0	0

*** Indicates significance at the 1% confidence interval.

returns should be uncorrelated at all leads and lags, according to RW hypothesis, so the presence of AC is per se already a sign of predictability of future return, therefore inefficiency. The autocorrelation function $\rho(k)$ at lag k ($\rho(k)$ is the AC coefficient) should be equal to zero for every $k > 0$, in a process with perfectly uncorrelated returns. In order to test whether all the autocorrelation are zero at the same time, we use the Ljung and Box (1978) statistic. The null hypothesis is H_0: the returns have no serial correlation, with alternative H1: the returns are correlated, i.e. reject the RW hypothesis. So, rejecting H_0 is equivalent to rejecting the RW assumption. The results are shown in 3.2, indicating that in both markets the returns are correlated, since the null is rejected at 1% confidence level, so they are not RW process, a further evidence of inefficient markets.

As we observe in the graphs of returns (Figures 3.1, 3.2) the presence of volatility clustering pattern appears evident for all returns. This pattern, very common in financial markets, is not per se a sign of market inefficiency, according to random walk required form of Efficient Market Hypothesis (EMH). However, the volatility clustering may affect the results of EMH test, leading therefore to erroneous inferences about the efficiency of a wholesale electricity market. Therefore, the actual presence of conditional heteroskedasticity is tested by using the Engle's ARCH test (or test for ARCH effect). The null hypothesis here is of no ARCH effect. The results are shown on Table 3.2 where the test for all returns reject the null hypothesis, confirming the presence of heteroskedasticity, therefore all the EMH tests that follow have been carried out by taking into account this results.

We start presenting our results by first reporting the Hurst exponent estimations on the returns of the prices of each market. As mentioned earlier, any deviation of the Hurst exponent value from the theoretical one of $H = 0.5$, practically means that the analyzed time-series exhibit either mean-reverting or persistent behavior, in the cases of a negative or a positive deviations respectively. However, it seems worthy to present first the Hurst exponent estimations on some representative artificial data for comparison purpose as well as for showing the potential of the tools adopted here to estimate this important exponent. To this end, a simulation has been conducted in order to show the Hurst exponent values based on the DFA approach for indicative processes, namely the white noise, the $1/f$ and the fractional Brownian noise, fBm. In particular, Figure 3.3 displays the Hurst exponent derived for each of these processes, where in the white noise case, the Hurst exponent coincides with the

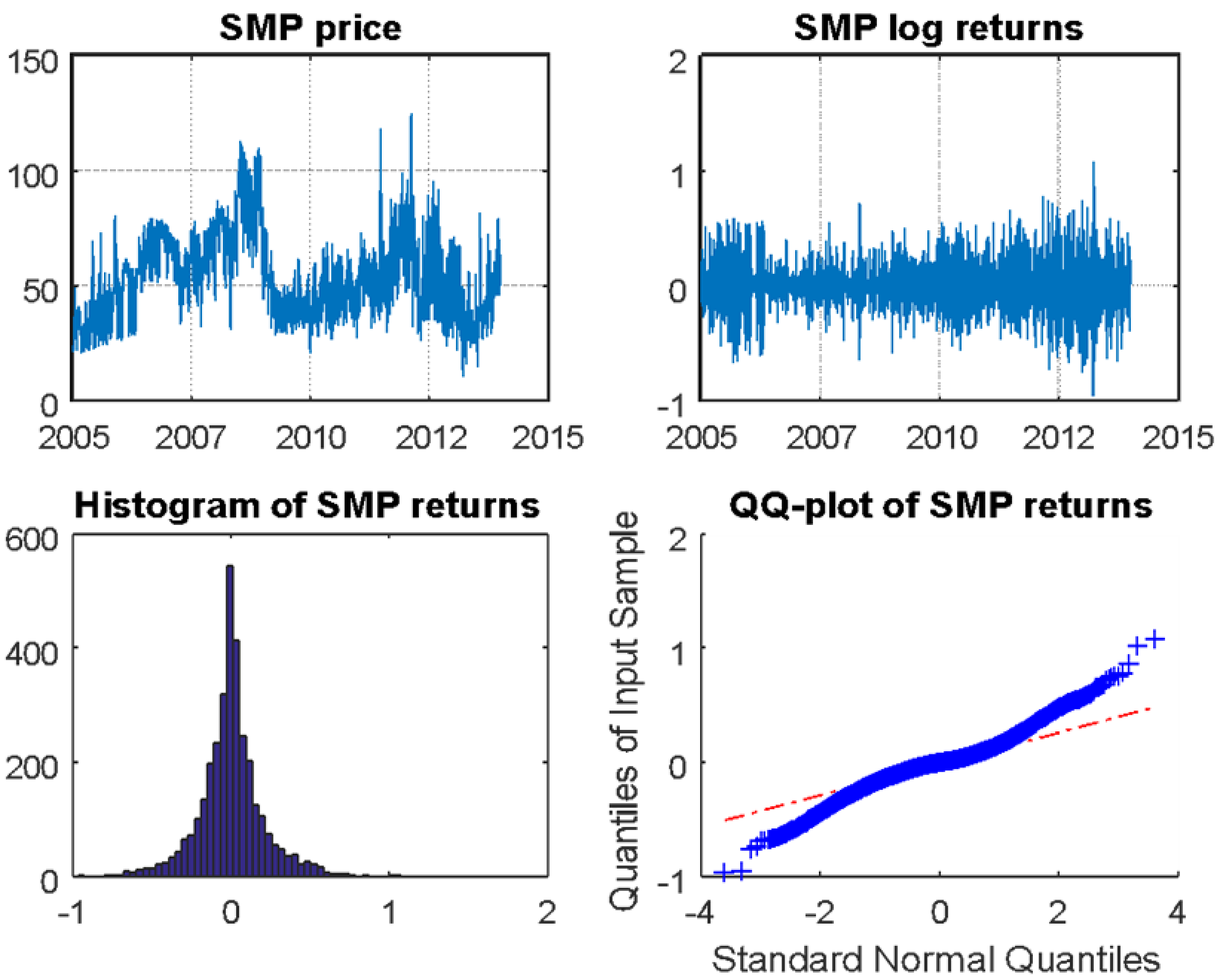

FIGURE 3.1: The time series of SMP prices and log returns with its histogram and QQ-plot

ideal value of $H = 0.5$ displaying a random walk process. However in the other two considered cases a deviation from the line with 0.5 gradient is observed, suggesting that these processes display persistent characteristics.

Calculations for the Hurst exponent based on the DFA approach are graphically represented in Figure 3.4 and Figure 3.5, respectively. As it can be easily observed, the slope of the Hurst exponent line presented in both cases significantly deviates from the theoretical line with a slope of 0.5. In fact, this means that the price returns of the two markets cannot be characterized by Random Walk process, hence according to the weak form of RWH they both are characterized as inefficient.

In addition, Figures 3.6 and 3.7 depict the Power spectrum of SMP and PUN returns. We can easily determine the DFA index α and the power spectral index β by the use of the Hurst exponent (H). We must keep in our mind that any deviation from the value of $H = 0.5$ suggests that the processes which describe the markets diverge from the RW process. In particular, the two markets have $\beta = 1.20$ (Greek) and $\beta = 1.18$ (Italian), while by using the formulas above, the DFA index for the Greek market is found to be 1.10 and the respective index for the Italian market is obtained as $\alpha = 1.09$. The derived indices suggest that the examined markets are inefficient.

Moreover, Table 3.3 summarizes the results on the Hurst exponent, estimated by

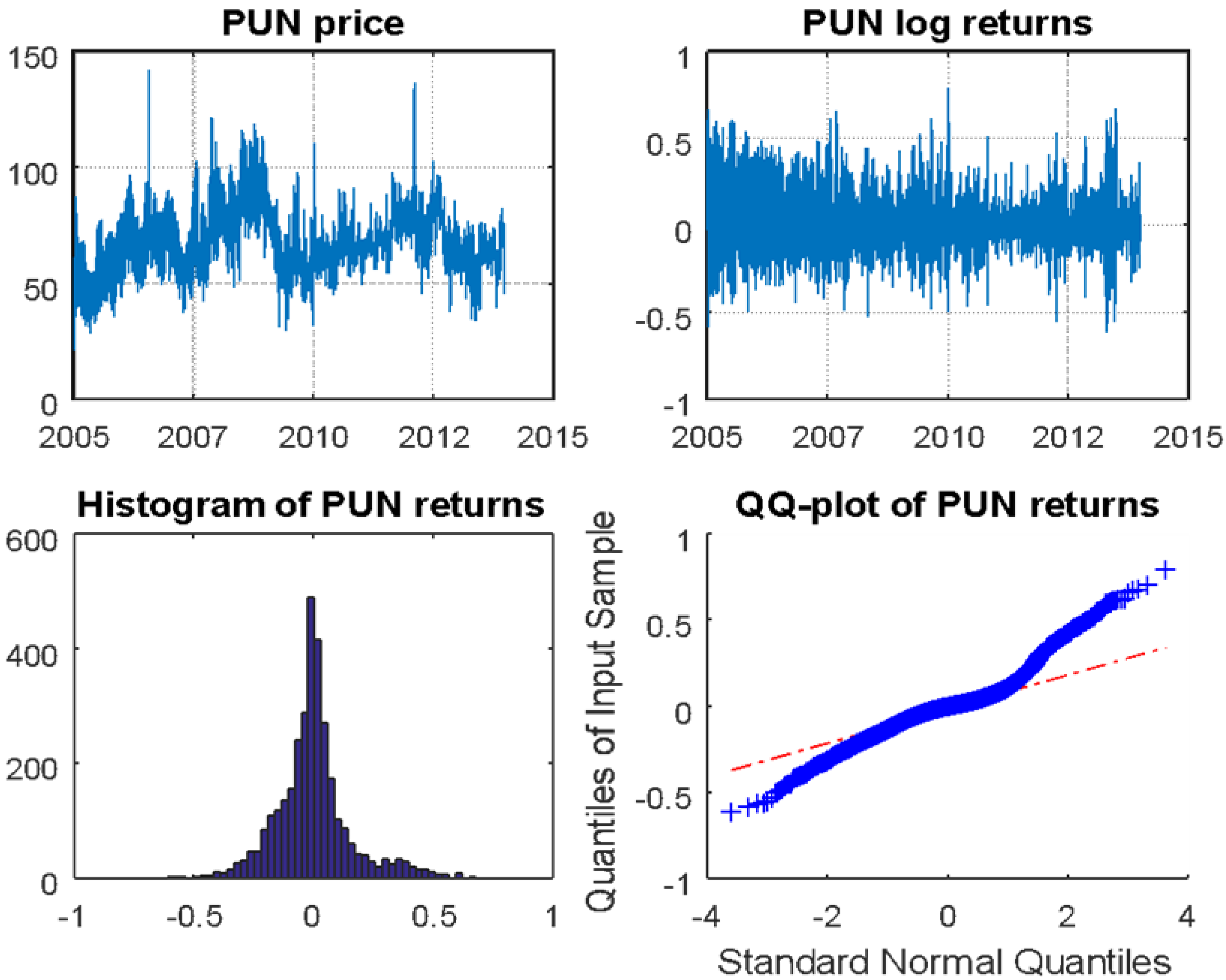

FIGURE 3.2: The time series of PUN prices and log returns with its histogram and QQ-plot

all methods described in the related Models section, for the returns of SMP, PUN. For comparison purposes, we provide also results of Hurst exponent estimations for the General Index of the financial markets of Greece (ASE) and Italy (FTMIB). It is apparent from the table below that all wholesale prices are generated by anti-persistent, mean-reverting processes far away from the Random Walk or Geometric Brownian Motion for which $H = 0.5$.

From Table 3.3, we conclude that the electricity price process is non-Markovian Shreve, Chalasani, and Jha, 1997; Eydeland and Wolyniec, 2003; Benth, Benth, and Koekebakker, 2008 which means that if the instantaneous price moves away from the (time-varying) fundamental price level, it will tend to return back to this level again (mean reverting process). Finally, we use the Tsallis entropy (TE) to assess the complexity of the two markets. Figure and show the rolling TE, with a window of 50 days, for the same period 2005-2014. We have used an optimization algorithm in order to find the optimum values 1.412 and 1.483 of index a or Tsallis exponent in (38), for the Greek and Italian markets, respectively. The top part for the figures show the time series, the middle part the rolling TE and lower part the unconditional rolling standard deviation (volatility). It is well obvious that rolling TE exhibits a rich amount of volatility information compared to the 'coarse-grained' one of the typical unconditional volatility used in financial analysis. An interesting result is

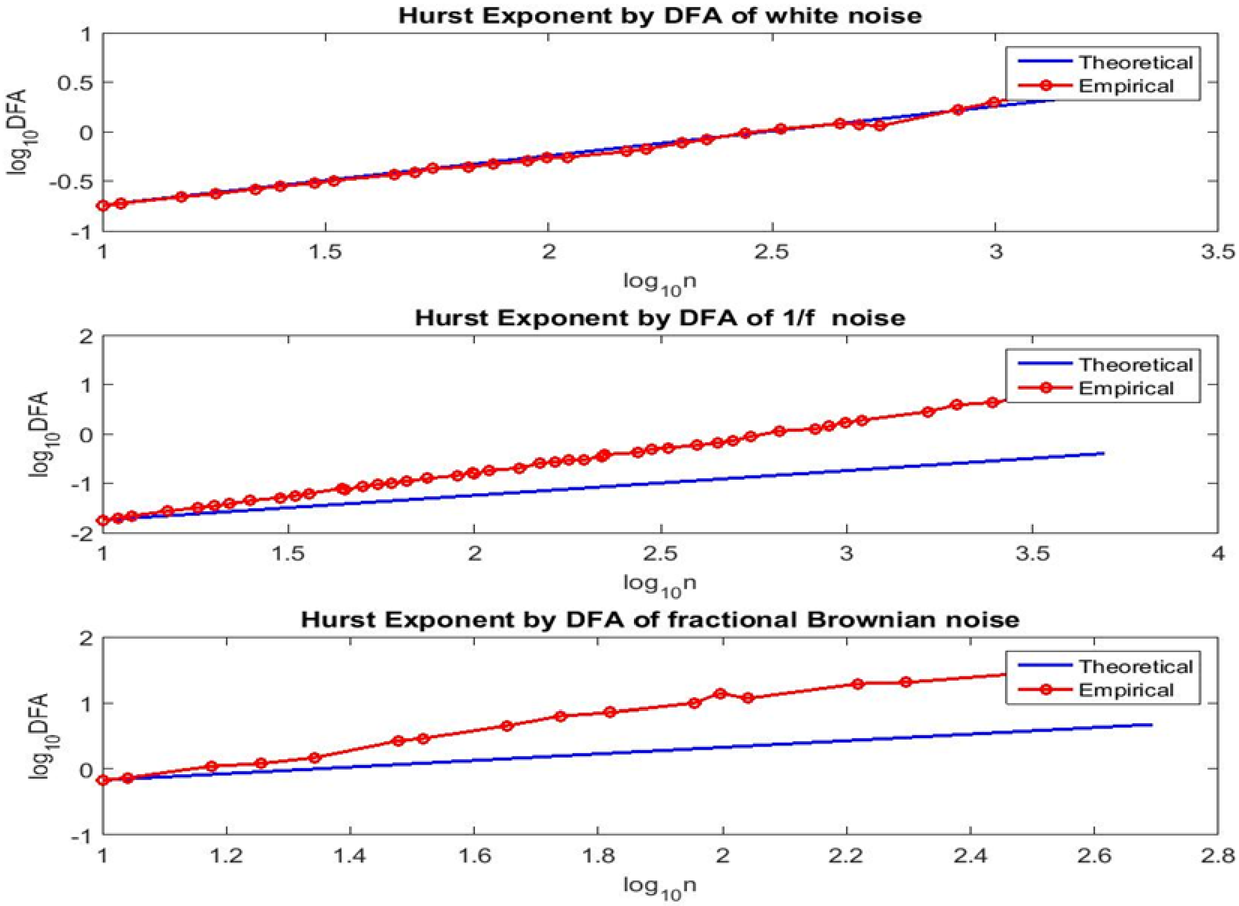

FIGURE 3.3: Hurst exponent representation based on DFA approach for different simulated processes.

that the level of TE of SMP (Greek market) is higher than that of the PUN (Italian market), indicating that the volatility in the former market, for the entire period considered, is higher than in the second market. This is because the Greek market is less developed, less matured, and shallower (fewer agents), which is equivalent to less dispersed and highly dependent agents' expectations, therefore a market in which its aggregate sentiment can drive the prices to higher levels. The rolling TE is therefore a very powerful measure which places emphasis on long-range, time dependent, interactive instability in the market.

Application of Nonlinear Deterministic (Chaotic) Time Series Analysis methods on the wholesale price in the Greek and Italian electricity markets

The BDS test statistic results for the stationary log returns of SMP and PUN are given in Tables 3.4 and 3.5 respectively. Here, m is the embedding dimension and e is the distance between two observations, defined in terms of σ (St.dev). We took for e/σ the value 0.5 and $m = 1$ to 6. Because the BDS statistics are positive and the p values are very significant ($p = 0.000$) the null hypothesis is rejected (also the BDS coefficients are less than 2.0 and 3.0 and, according to Opong et al., 1999a, the test rejects the H_0 of i.i.d. process). So both the SMP and PUN returns, with

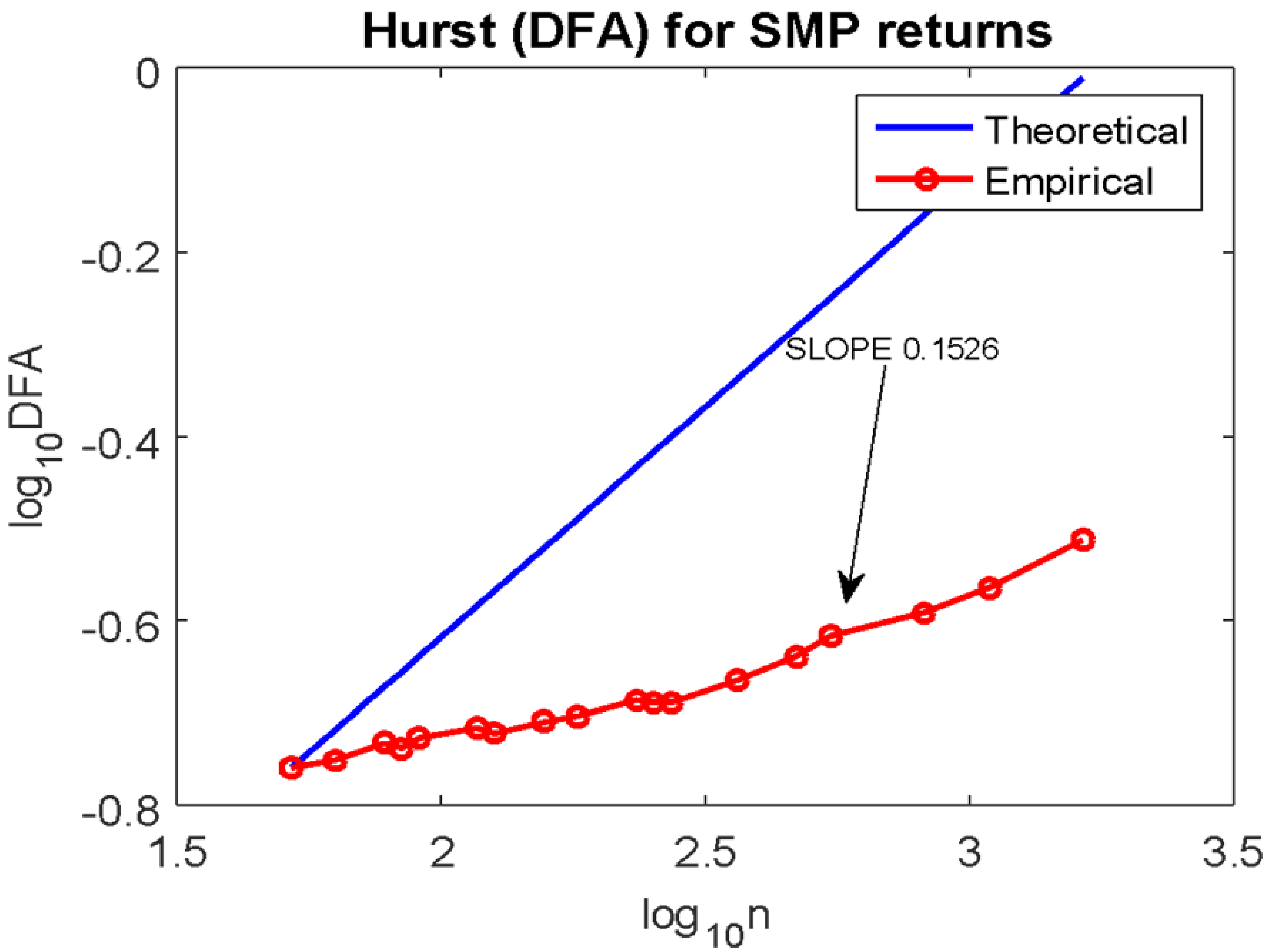

FIGURE 3.4: Hurst exponent derived through DFA approach for the Greek market. The slope of the fitted line (red color) deviates significantly from the theoretical one (blue, Random Walk Process)

95% and 99% confidence, cannot be considered i.i.d, and clearly exhibit a nonlinear dependence, which is one of the four “footprints” of chaos. This finding indicate that SMP and PUN returns do not reflect the historical price information, i.e. the efficiency of the two markets is against the weak-form EMH. The BDS test is also applied on the standardized residuals from the ARIMA/EGARCH (EGARCH stands for the Exponential GARCH) models of conditional mean and volatilities, in order to check for any remaining structure in the residuals so assessing the quality of the ARIMA/EGARCH models. The results (not reported here) provide little support for the presence of nonlinear dependence in the residuals enhancing the good quality of the model. The combination of results of BDS and R/S-Hurst analysis, prove that both returns deviated strongly from random walk. We stress here, however, that the BDS is a necessary but not sufficient condition of the chaotic behaviour and also that it is consistent with other direct test for chaos, e.g. maximum Lyapunov exponent. we have deduced the same result as Dakhlaoui and Aloui, 2013 that, contrary to existing literature Opong et al., 1999a; Opong et al., 1999b the R/S-Hurst and BDS test work complementarily in the verification of the necessary conditions of chaos.

TABLE 3.3: Estimation of Hurst Exponent for Electricity (SMP, PUN) and financial markets (Greece, Italy)

Time Series (Returns)	R/S (AL)	DFA	GHE-H(1)	GHE-H(2)	Mean of 1st and 2nd columns
Electricity Market					
SMP	0.214	0.152	0.169	0.149	0.183
PUN	0.243	0.106	0.105	0.085	0.174
Financial Market					
ASE	0.605	0.650	0.556	0.542	0.627
FTMIB	0.530	0.5410	0.494	0.467	0.535

TABLE 3.4: The BDS statistics applied to the daily returns of SMP time series

BDS Test for SMPRET					
Included observations: 3287					
Dimension	BDS Statistic	Std. Error	z-Statistic	Prob.	
2	0.023611	0.001126	20.96881	0.0000	
3	0.025776	0.000891	28.91836	0.0000	
4	0.018807	0.000530	35.46008	0.0000	
5	0.011913	0.000277	43.04621	0.0000	
6	0.007327	0.000134	54.75693	0.0000	
Raw epsilon	0.099448				
Pairs within epsilon	3722168.	V-Statistic	0.344716		
Triples within epsilon	5.36E+09	V-Statistic	0.151097		
Dimension	C(m,n)	c(m,n)	C(1,n-(m-1))	c(1,n-(m-1))	$c(1, n-(m-1))^k$
2	767177.0	0.142229	1857733.	0.344409	0.118618
3	359438.0	0.066678	1857343.	0.344547	0.040902
4	177181.0	0.032888	1855821.	0.344474	0.014081
5	90321.00	0.016775	1855631.	0.344649	0.004863
6	48468.00	0.009008	1855418.	0.344819	0.001681

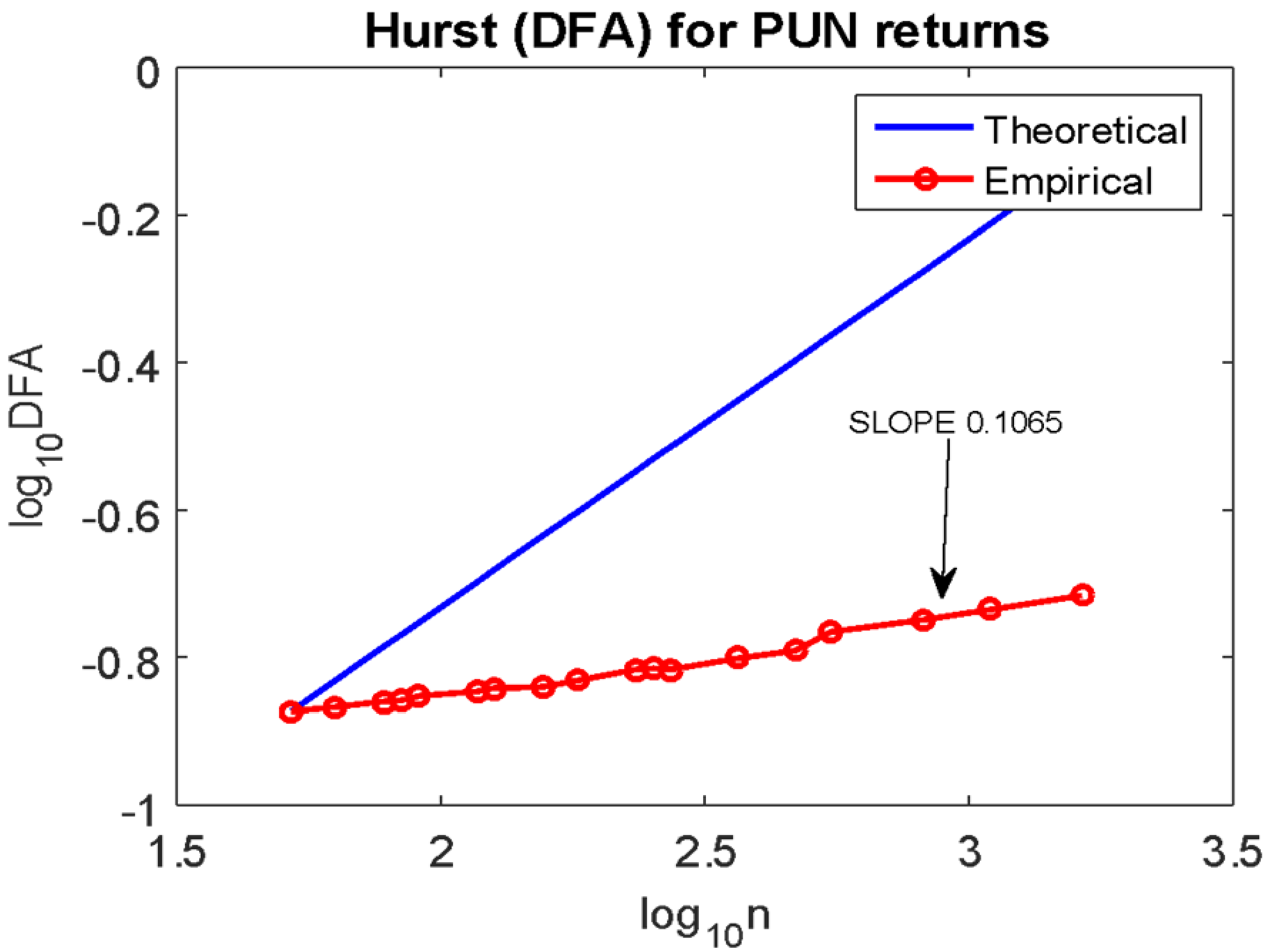

FIGURE 3.5: Hurst exponent representation for the Italian market. The slope of the fitted line (red color) deviates significantly from the theoretical one (blue, Random Walk Process)

Estimating 'best' parameters, delay time and embedding dimension

For both time series, the FNN method suggest an embedding dimension of 6 setting the "active" dimensionality of the manifold on which the system evolves, see fig.3.10, 3.11. The increase of the number of FNN after certain d_E is an indication of the presence of noise in the signal.

Tables and provide the estimation of Hurst exponent by using different methods.

It is apparent from the tables above that both SMP and PUN are generated by anti-persistent, mean-reverting processes far away from the Geometric Brownian Motion for which $H = 0.5$.

Estimating maximal Lyapunov Exponent

The preliminary found max orders of the triplet (τ,m,q) are (2,7,3), respectively, for both SMP and PUN, while the optimum ones are as shown in Table 3.8 below :

TABLE 3.5: The BDS statistics applied to the daily return of PUN returns

BDS Test for PUNRET				
Included observations: 3287				
BDS Statistic	Std. Error	z-Statistic	Prob.	
0.024502	0.001258	19.47594	0.0000	
0.021843	0.001045	20.90815	0.0000	
0.011911	0.000652	18.26255	0.0000	
0.005480	0.000357	15.34633	0.0000	
0.003659	0.000181	20.20040	0.0000	
Raw epsilon	0.080376			
Pairs within epsilon	3903186.	V-Statistic	0.361480	
Triples within epsilon	5.92E+09	V-Statistic	0.166721	
C(m,n)	c(m,n)	C(1,n-(m-1))	c(1,n-(m-1))	$c(1, n-(m-1))^k$
835937.0	0.154976	1948368.	0.361212	0.130474
371701.0	0.068952	1946911.	0.361162	0.047109
155864.0	0.028931	1945899.	0.361194	0.017020
62551.00	0.011618	1944117.	0.361083	0.006138
31653.00	0.005883	1943969.	0.361276	0.002223

TABLE 3.6: Annual evolution of Hurst exponent computed by different methods (2005-2013). * Note: Anis-Lloyd correction for small sample bias Weron and Przybylowicz, 2000b; Annis and Lloyd, 1976; Peters, 2015

Year	R/S Hurst (traditional)	R/S Hurst with Anis-Lloyd modification for sample size*	DFA Hurst	GPH Hurst	GHE Hurst (q=1)
	SMP,PUN	SMP,PUN	SMP,PUN	SMP,PUN	SMP,PUN
2005	0.222,0.205	0.134,0.253	0.116,0.043	0.121,0.335	0.150,0.092
2006	0.207,0.252	-0.069,0.223	0.193,0.048	0.054,0.386	0.223,0.097
2007	0.279,0.258	0.282,0.248	0.153,0.100	0.0550,0.579	0.198,0.148
2008	0.315,0.258	0.277,0.179	0.186,0.081	0.0060,0.090	0.192,0.125
2009	0.271,0.323	0.225,0.151	0.109,0.128	0.0880,0.267	0.133,0.134
2010	0.227,0.247	0.125,0.213	0.123,0.095	-0.0150,0.380	0.140,0.119
2011	0.290,0.248	0.257,0.152	0.147,0.100	0.1990,0.086	0.1480,0.107
2012	0.204,0.347	0.142,0.326	0.126,0.106	-0.2440,0.000	0.1130,0.071
2013	0.302,0.230	0.232,0.220	0.122,0.095	0.0460,0.337	0.1540,0.094
2005-2013	0.267,0.254	0.211,0.231	0.143,0.097	0.329,0.254	0.1660,0.107

TABLE 3.7: Hurst exponents computed by various methods for SMP, PUN raw and deseasonalized returns (2005-2013).*Note: for the Hurst values computed by GHE method we use prices and deseasonalized prices, not returns

SMP		
Method	Original returns	Deseasonalized returns*
Hurst GPH	0.330	0.263
Hurst R/S	0.211	0.212
Hurst DFA (α)	0.143	0.158
Hurst GHE (prices) (q=1)	0.166	0.189
PUN		
Method	Original returns	Deseasonalized returns*
Hurst GPH	0.254	0.190
Hurst R/S	0.231	0.246
Hurst DFA (α)	0.097	0.124
Hurst GHE (prices) (q=1)	0.107	0.138

TABLE 3.8: Max Lyapunov Exponent, λ_{max}, for SMP and PUN daily raw returns, 2005-13, using BenSaid algorithm, which implements Nychka et al., 1992 method.

Series	Optimum Orders (τ,m,q)	λ_{max}	p-Value	Confidence Interval, CI	Hypothesis
SMP	(2,7,3)	-0.1663	0.000	[-0.1774, inf]	H_1
PUN	(1,7,1)	-0.0694	0.000	[-0.0793,inf]	H_1

TABLE 3.9: Maximum Lyapunov Exponent, computed for each separate year, for SMP and PUN returns hourly data (based on 8760 data and estimated by BenSaid A. algorithm BenSaïda, 2015, using the optimum orders of parameters (τ,m,q).

Year	Time delay, τ (by Mutual Information)	Embedding, m (by False Nearest Neighbors, FNN)	Maximum Lyapunov Exponent, λ_{max}	Dynamical System stability or chaos
	SMP,PUN	SMP,PUN	SMP,PUN	SMP,PUN
2005	3,2	10,6	0.1102,-0,257	chaos,stability
2006	3,2	10,6	0.6550,-0.2090	chaos,stability
2007	3,2	10,6	-0.0180,-0.1600	stability,stability
2008	3,3	10,6	-0.0076,0.1158	stability,stability
2009	3,3	10,6	0.3957,-0.1280	stability,stability
2010	3,3	10,6	-0.0577,-0.1560	stability,stability
2011	3,3	10,6	0.1140,-0.1266	chaos,stability
2012	3,3	10,6	-0.1034,0.0740	stability,stability
2013	3,3	10,6	0.0369,-0.0835	chaos,stability

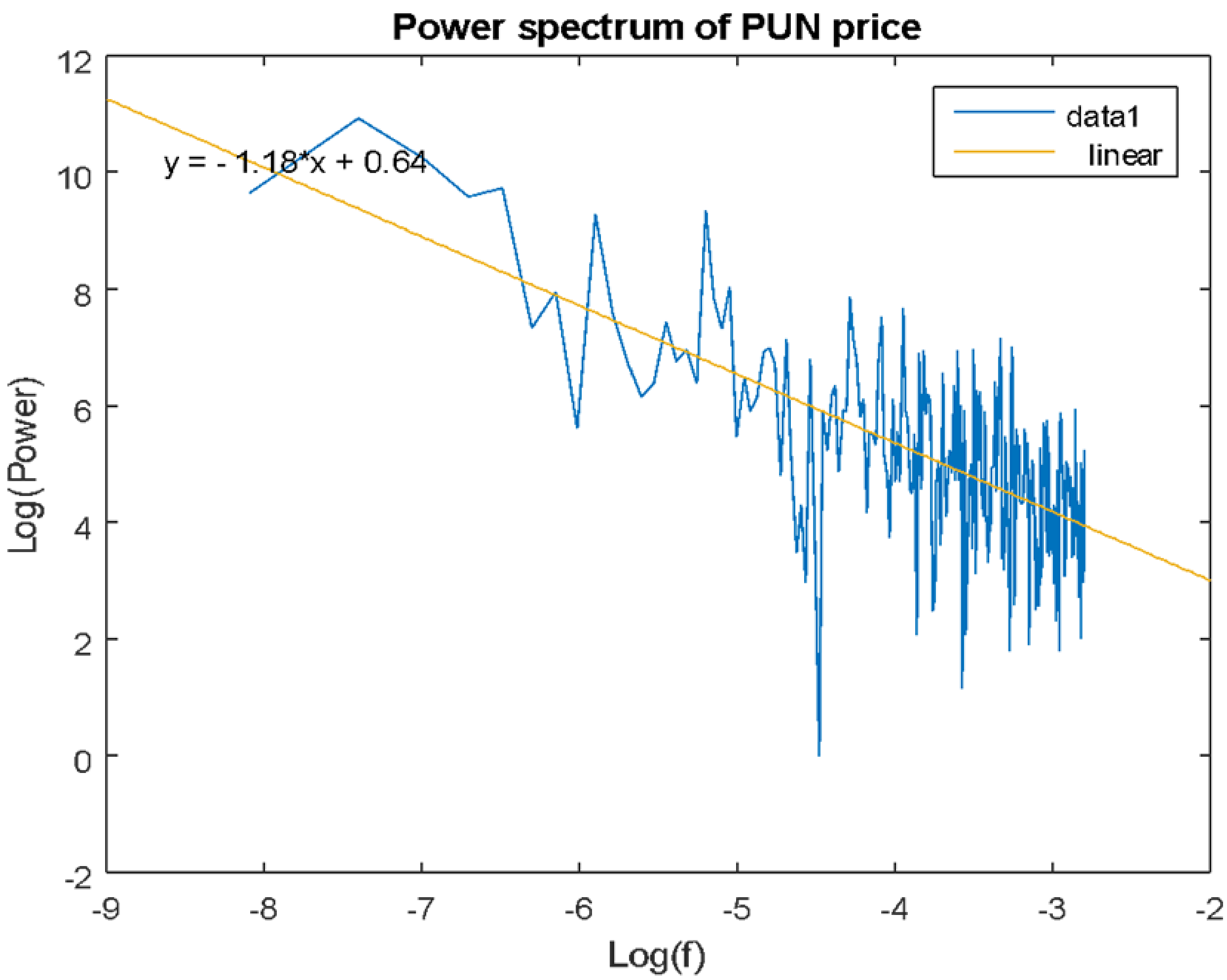

FIGURE 3.6: Power Spectrum of PUN returns with power spectrum index $\beta = 1.18$

For the calculations in Table , we used hourly data (8760 values per year) providing an adequate date set size, so securing good statistics in λ_{max} estimation. We observe that λ_{max} is negative for the entire period 2005-13, indicatind a stable dunamical system, while for the yearly computations PUN is a stable system while SMP returns for years 2005, 2006, 2011 and 2013 show positive λ_{max}, indicating that the market was, in that period, in a "chaotic" state, however not permanent. Overall, we can state that both markets are not chaotic (λ_{max}<0), since there is no evidence of persistent chaotic dynamics. This result is consistent with the mean-reverting dynamic behavior indicated by Hurst exponent and the SARIMA/EGARCH model for the conditional mean and variance of both wholesale prices. In figure we plot the maximum Lyapunov exponent as a function of embedding dimension, computed by the Jacobian method, using BenSaida's algorithm. The figure reveals that λ_{max} for both returns converge to the interval [-0.05, -0.15], implying the market systems generating the returns converge to an asymptotic stability, more pronounced in the case of Italy.

From the above tables and figure 3.14 it is clear that there is no unstable (chaotic) dynamics in both SMP and PUN returns since the largest Lyapunor exponent is negative. However it is also clear that the SMP returns have a more negative λ_{max} than the PUN returns, suggesting a more stable dynamic process (a more stable 'market')

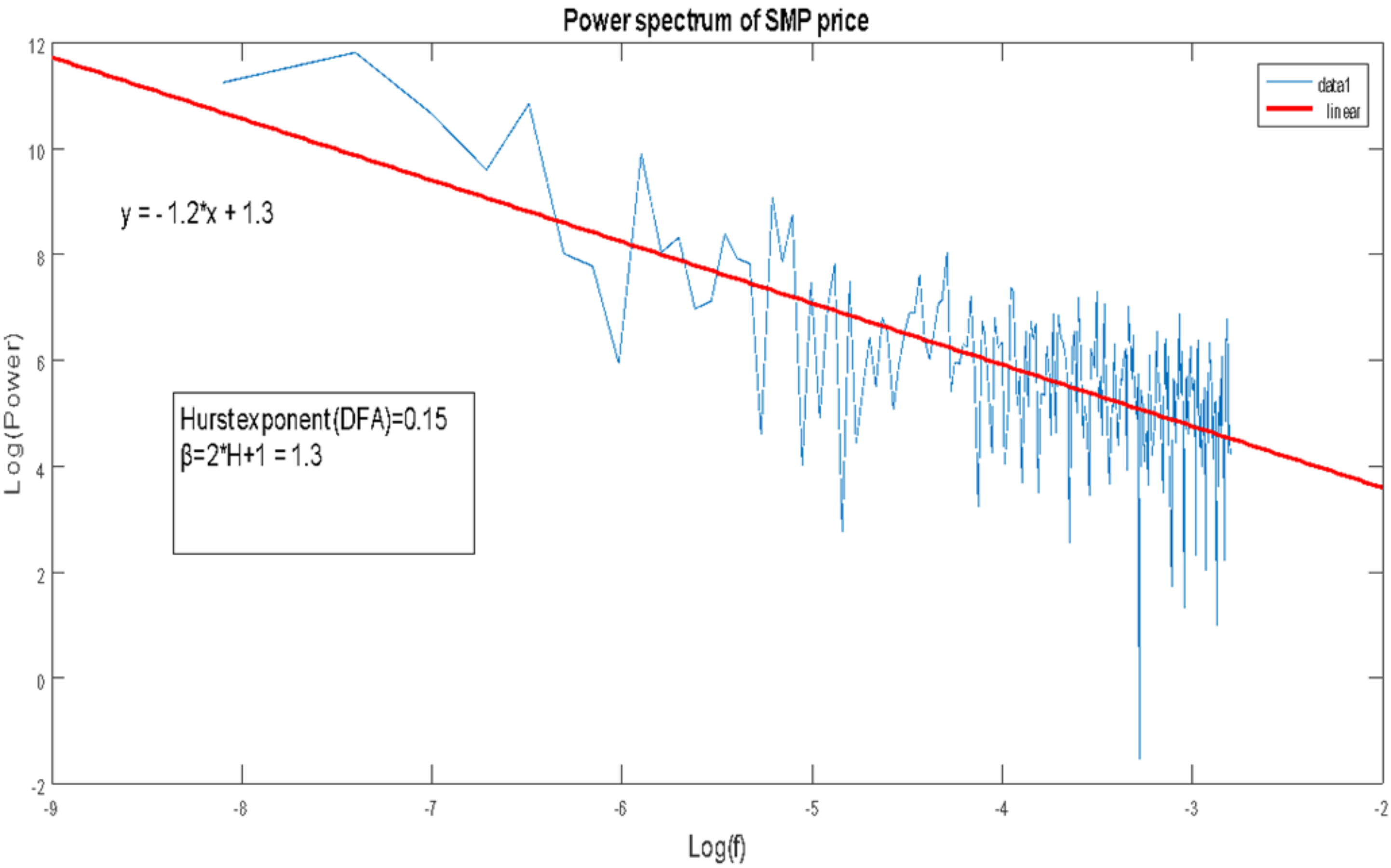

FIGURE 3.7: Power Spectrum of SMP returns with power spectrum index $\beta = 1.20$

that generates the returns. We now compute the changes in max Lyapunov exponent from year to year.

From Table 3.10 we can track the direction of changes of λ_{max}, and conditional volatility according to the following rational:

1. If λ_{max} becomes more negative (MN) then the stability is increased (I) so the volatility σ must decrease (D).

2. The further Hurst exponent, H moves from 0.5, either towards to 0 or to 1, the less efficient the market becomes (the more mean-reverting or more persistent, respectively, is the time series), which is equivalent to say, more predictable but also more volatile, due to inefficiency. So, stability and volatility must move in opposite directions (i.e they are negatively correlated). From the Table above the change directions of λ_{max} and σ are "reasonable" for only a number of periods indicated by bold .

It is proved that as the dominant and stable eigenvalue decreases (hence its distance from zero increases), the system becomes more stable. Closely related to stability is the notion of volatility. We can draw conclusions on the stability of the market by observing the behavior of its respective volatility. It is clear that in almost all cases of the Greek Electricity Market, a larger negative eigenvalue leads to reduced volatility (and vice versa), meaning that as the system becomes more stable,

TABLE 3.10: Directions of changes of λ_{max},H and σ for SMP and PUN returns

Year	λ_{max}	$\Delta\lambda_{max}$	Stability	Conditional Variance σ_t	$\Delta\sigma_t$	Direction of change of Volatility
SMP						
2005	0.1102	-	-	0.0890(0.298)	-	-
2006	0.655	LN	decrease	0.00746(0.086)	-0.212	decrease
2007	-0.0180	MN	increase	0.01153(0.107)	0.021	increase
2008	-0.0076	LN	decrease	0.00904(0.095)	-0.012	decrease
2009	0.3957	LN	decrease	0.02177(0.147)	0.052	increase
2010	-0.0577	MN	increase	0.0108(0.103)	-0.044	decrease
2011	0.1140	LN	decrease	0.004618(0.067)	-0.036	decrease
2012	-0.1034	MN	increase	0.009197(0.0959)	0.029	increase
2013	0.0369	LN	decrease	0.01047(0.1023)	0.006	increase
PUN						
2005	-0.257	-	-	0.0317(0.1780)	-	-
2006	-0.209	LN	decrease	0.0163(0.127)	-0.051	decrease
2007	-0.160	LN	decrease	0.0075(0.086)	-0.041	decrease
2008	-0.116	LN	decrease	0.0077(0.087)	0.001	increase
2009	-0.128	MN	increase	0.0140(0.118)	0.031	increase
2010	-0.156	MN	increase	0.0225(0.15)	0.032	increase
2011	-0.126	LN	decrease	0.01851(0.136)	-0.014	decrease
2012	-0.074	LN	decrease	0.0439(0.209)	0.073	increase
2013	-0.0835	MN	increase	0.03109(0.176)	-0.033	decrease

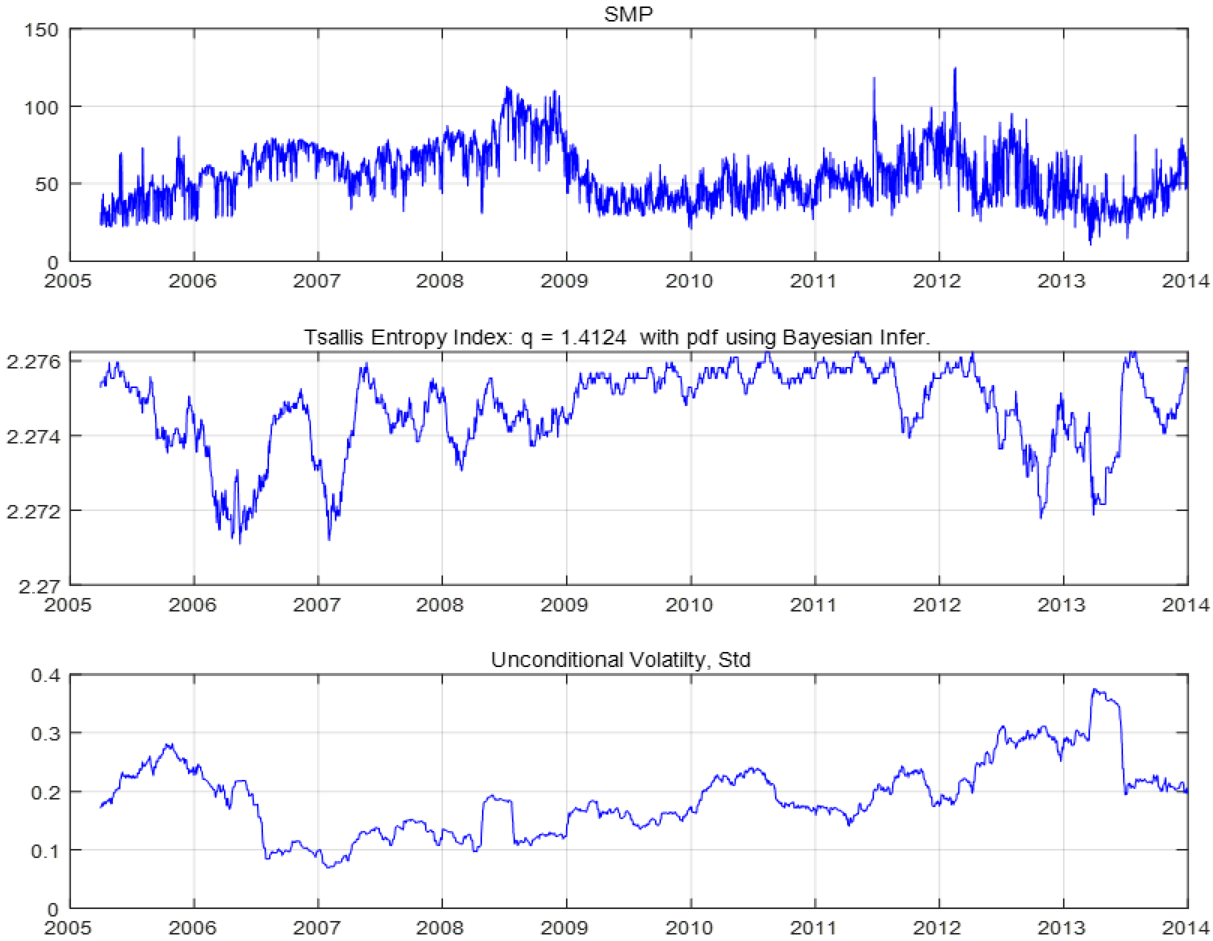

FIGURE 3.8: The rolling Tsallis Entropy (TE) of the whole sale price (System Marginal Price) of the Greek electricity market.

it also becomes less volatile. In the case of PUN however this induction cannot be applied. We observe an almost opposite behavior, showing increased volatility as the dominant eigenvalue sinks further into the negative plane. This could be explained as increased betting from the side of the electricity traders during periods of established certainty and conversely, as more conservative activity during periods of uncertainty. This can be combined with the observation drawn earlier, that the Italian market experiences long memory effect, so traders know the previous state of the system and try to profit from its current status. Additionally, we know that the Italian market is well interconnected with external markets and during the period of financial crisis, it is possible that the phenomenon of volatility spillover occurred and resulted to unwanted increase of the volatility of electricity prices. There are also the first two years, when the PUN showed a positive eigenvalue, which represents an unstable system. The regulatory authorities quickly took countermeasures to dispel the uncertainty linked with instability and drag the market to stability and after two years the market showed negative eigenvalues.

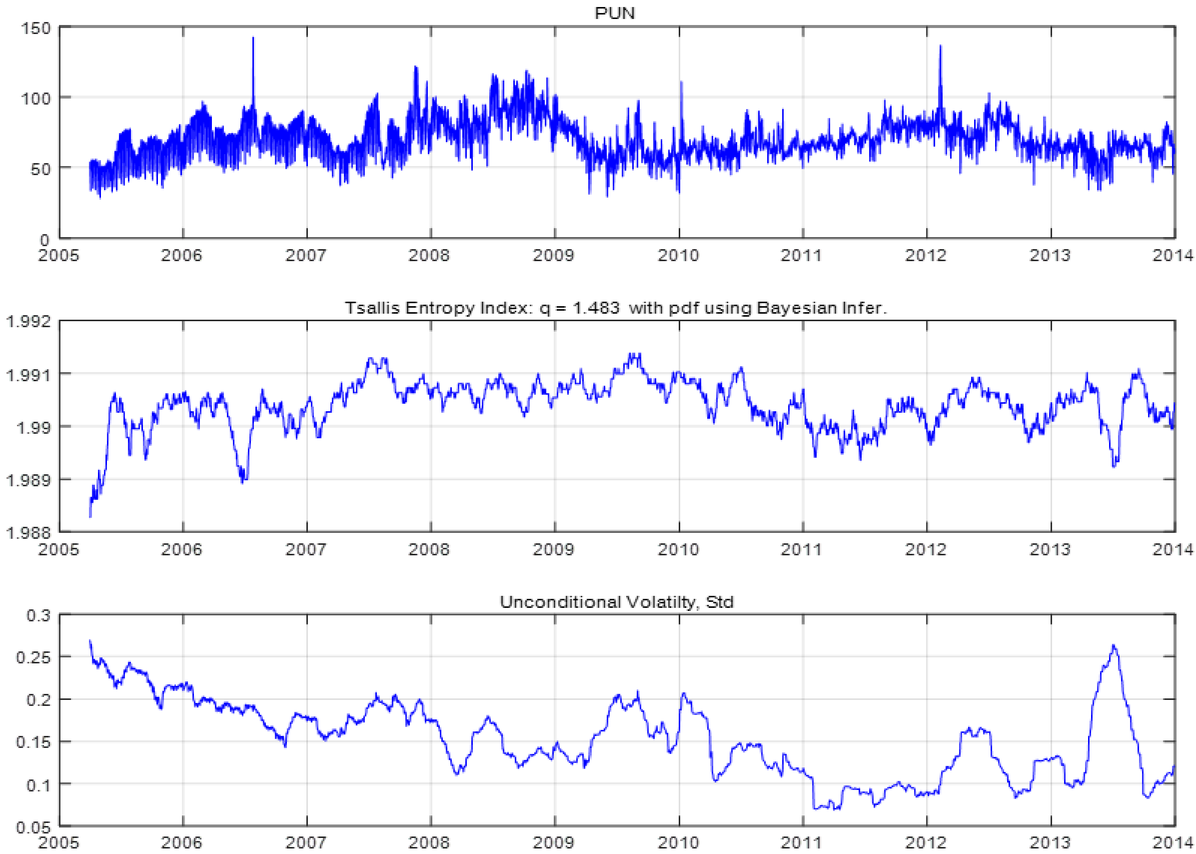

FIGURE 3.9: The rolling Tsallis Entropy (TE) of the whole sale price (System Marginal Price) of the Italian electricity market

Application of tools on Conditional Volatility

Here we display the results of the application of conditional volatility models, applied in the two Electricity markets.

Tables and summarize conditional mean and conditional volatility of the optimal model for both SMP and PUN returns, respectively.

Substituting the values of coefficients K, γ_1, α_1, ξ_1, the final models for the Conditional Volatility of PUN and SMP are written respectively:

$$log(\sigma_t^2) = -0.1432 + 0.9085 \quad log\sigma_{t-1}^2 + 0.347[\frac{|\epsilon_{t-1}|}{\sigma_{t-1}} - E[\frac{|\epsilon_{t-1}|}{\sigma_{t-1}}]] - 0.066(\frac{\epsilon_{t-1}}{\sigma_{t-1}}) \tag{3.1}$$

$$log(\sigma_t) = -0.111 + 0.971 \quad log\sigma_{t-1}^2 + 0.238[\frac{|\epsilon_{t-1}|}{\sigma_{t-1}} - E[\frac{|\epsilon_{t-1}|}{\sigma_{t-1}}] - 0.0853(\frac{\epsilon_{t-1}}{\sigma_{t-1}}) \tag{3.2}$$

where the required condition for the persistence $\gamma < 1$ holds for both models. Since $\xi_i < 0$ for both series the leverage effect and the asymmetric impact conditions are confirmed. This result for PUN is in the opposite direction with that found in the

TABLE 3.11: Conditional Mean and Conditional Volatility of the optimal model for the SMP returns

		Conditional Mean	
Variable	Symbol	Coefficient	Probability
AR(1)	ϕ_1	0.2230	0.0000
SAR(7)	Φ_1	-1.0420	0.0000
SAR(14)	Φ_2	-0.7990	0.0000
SAR(21)	Φ_3	0.0343	0.0000
MA(1)	θ_1	-0.7200	0.0000
SMA(7)	Θ_1	0.1610	0.0000
SMA(14)	Θ_2	-0.1610	0.0000
SMA(21)	Θ_3	-0.8041	0.0000
		Conditional Volatility	
Variable	Symbol	Coefficient	Probability
K	k	-0.4320	0.0000
GARCH(1)	γ_1	0.9085	0.0000
ARCH(1)	α_1	0.3477	0.0000
Leverage(1)	ξ_1	-0.0662	0.0000

TABLE 3.12: Conditional Mean & Conditional Volatility of the optimal model for the PUN returns

Conditional Mean			
Variable	Symbol	Coefficient	Probability
AR(1)	ϕ_1	1.2090	0.0000
AR(2)	ϕ_2	-0.2810	0.0000
SAR(7)	Φ_1	0.6780	0.0000
SAR(14)	Φ_2	0.2030	0.0000
MA(1)	θ_1	-1.6240	0.0000
MA(2)	θ_2	0.6260	0.0000
MA(3)	θ_3	0.0070	0.0000
SMA(7)	Θ_1	-0.6350	0.0000
SMA14)	Θ_2	-0.1840	0.0000
Conditional Volatility			
Variable	Symbol	Coefficient	Probability
K	k	-0.1110	0.0000
GARCH(1)	γ_1	0.9710	0.0000
ARCH(1)	α_1	0.2380	0.0000
Leverage(1)	ξ_1	-0.0853	0.0000

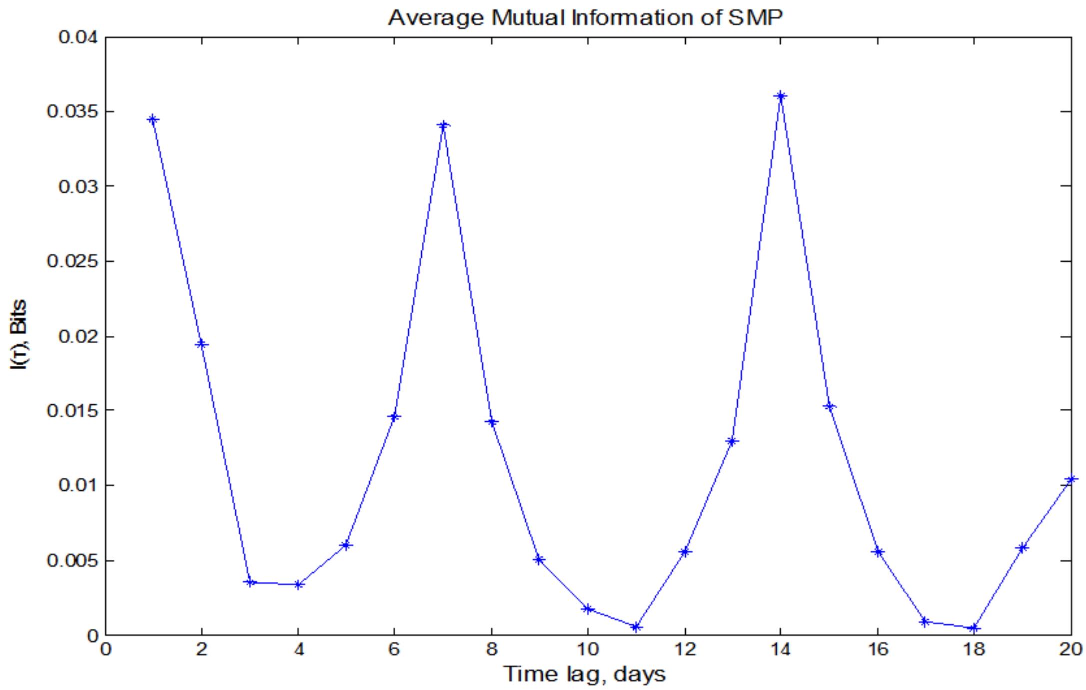

FIGURE 3.10: Average Mutual Information (AMI) of SMP returns. The first minimum is 4 days, defining the time delay τ in the reconstruction vector by MOD

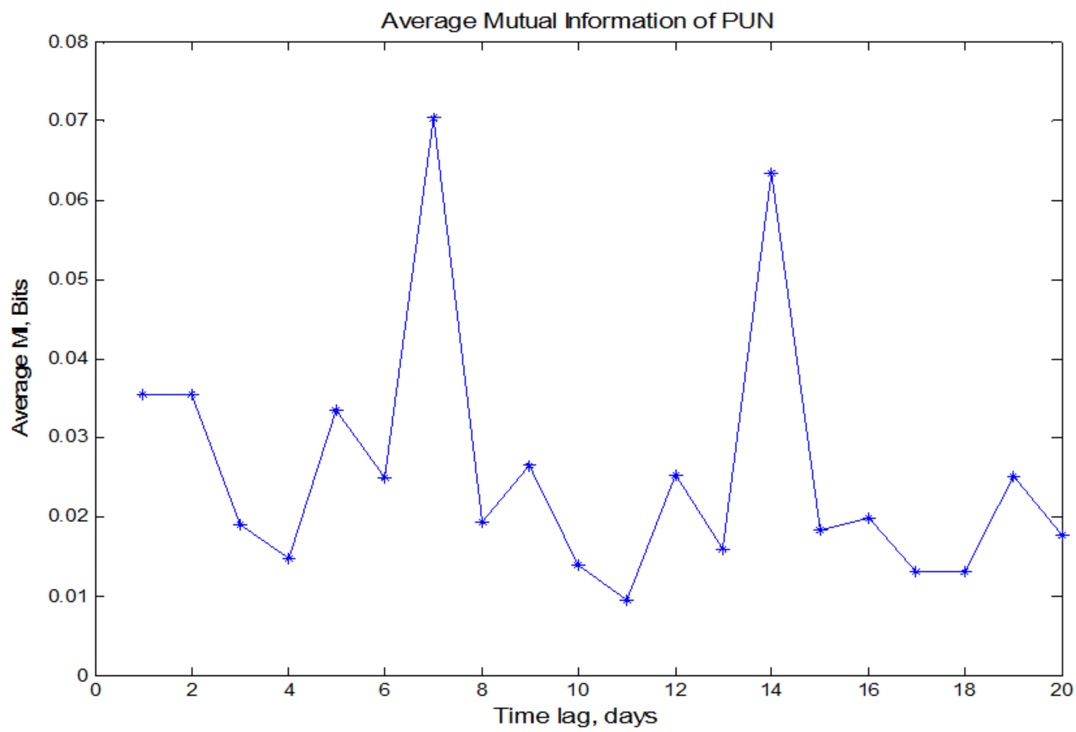

FIGURE 3.11: Average MI (AMI) of PUN returns. The first minimum is 4 days, defining the time delay τ in the reconstruction vector by MOD

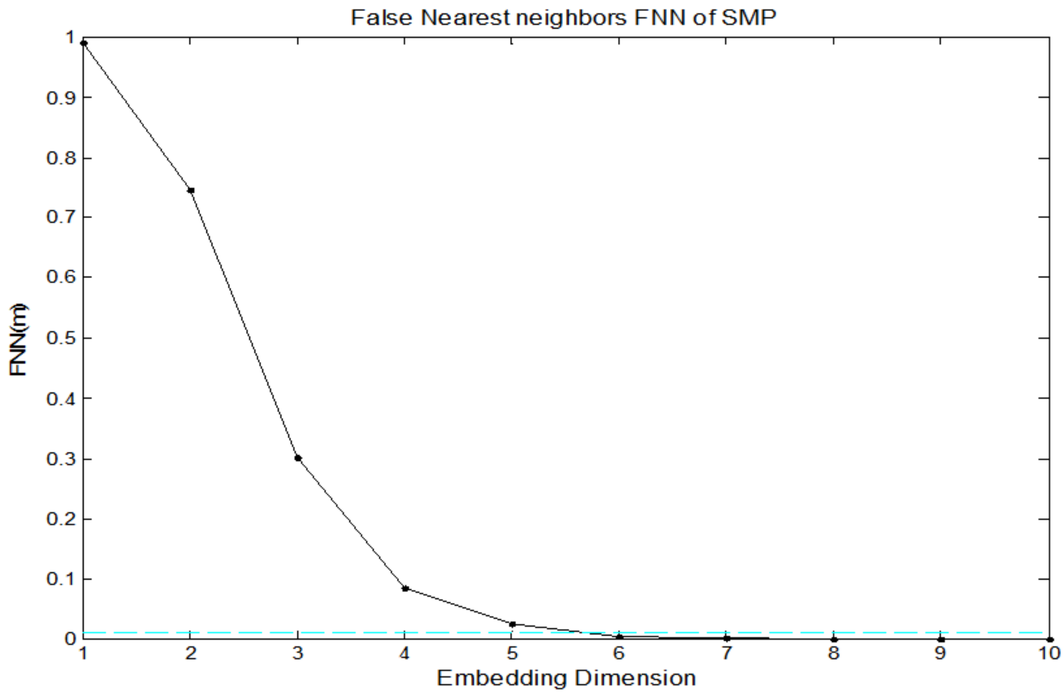

FIGURE 3.12: False Nearest Neighbors for estimating the minimum embedding dimension of the reconstructed phase space for SMP returns

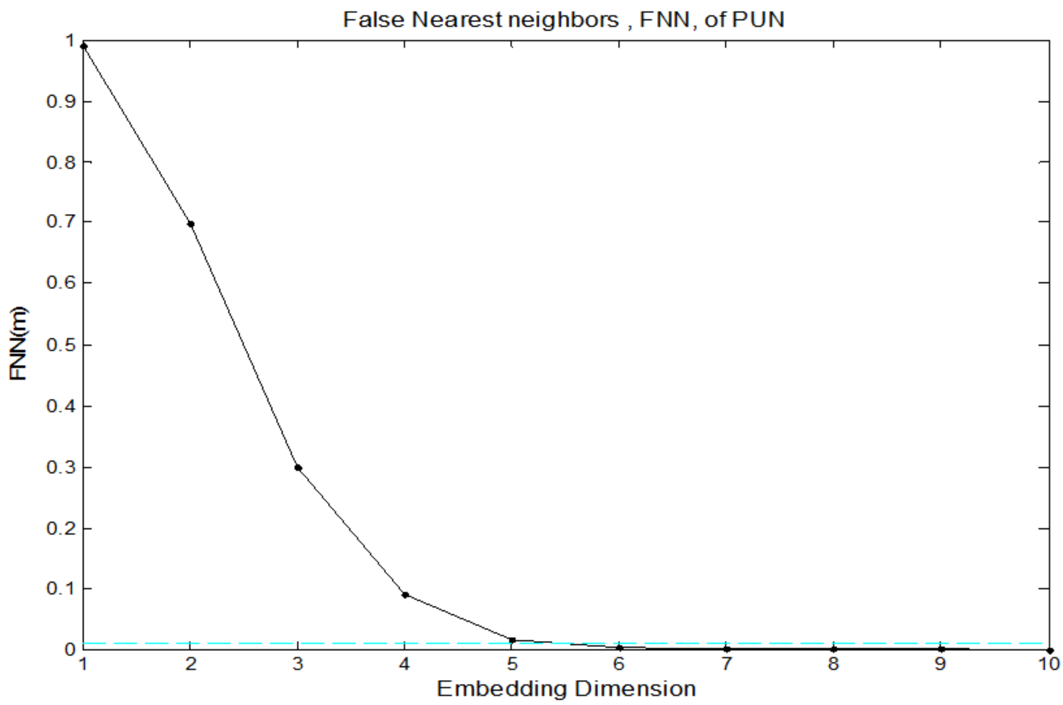

FIGURE 3.13: False Nearest Neighbors for estimating the minimum embedding dimension of the reconstructed phase space for PUN returns

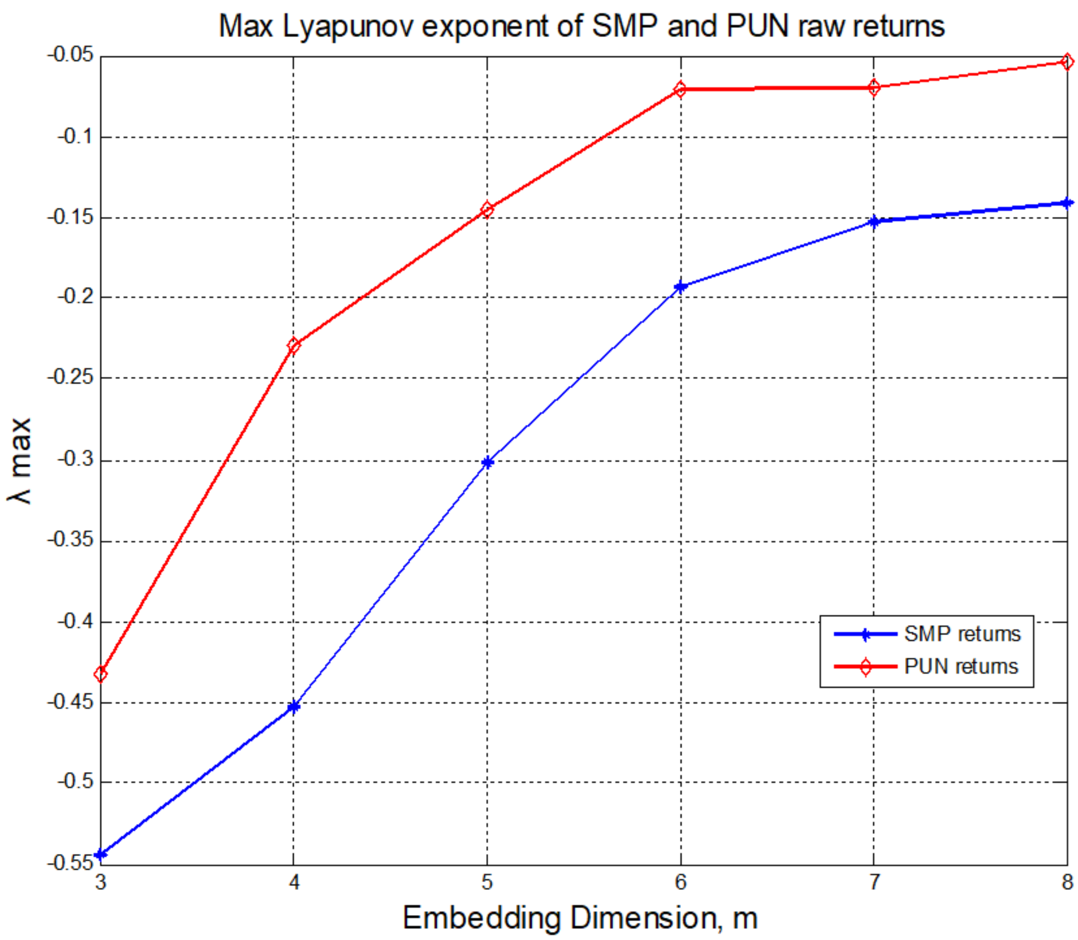

FIGURE 3.14: Max Lyapunov exponent versus the embedding dimension, m, for SMP and PUN raw returns, computed by BenSaida's algorithm.

work of Petrella and Sapio, 2012, possibly due to different sampling periods or due to the fact that the PUN series, in our case, has undergone a strong filtering before fitted to the model. Figures 3.15 and 3.16 below, show the conditional volatilities and residuals from the models provided in tables 3.15 and 3.16 respectively.

Table , shows the inter-annual GARCH coefficients as well as the theoretical Unconditional Variance $\sigma\epsilon$.

Theoretical Unconditional Variance $\sigma\epsilon$ can be easily calculated from GARCH coefficients with good accuracy. In general, this assumption is valid as long as the preferred model has inherent symmetry; this is not applicable in our case where the best found model is the EGARCH(1,1). At this point, we must introduce the reader to a new and very important variable that 'measures' the degree of completion of a market, the Herfindahl-Hirschman Index (HHI), for which we aim to examine how it is related to the volatility and complexity of a market.

Table 3.2 provides information on the 'degree" of competition in the generation "market" in Greek and Italian Markets measured by the HHI index. HHI stands for Hefindahl-Hishman index, a measure of market concentration (United States Department of Justice, link: HHI, 2018). If HHI=10000 (upper limit) the market is considered a monopoly, while for $HHI > 5000$ the market is over-concentrated, for

TABLE 3.13: Inter-Annual GARCH Coefficients γ_1 and theoretical Unconditional Variance $\sigma\epsilon$

	γ_1(SMP)	$\Delta\gamma_1$(SMP)	$\sigma\epsilon$(SMP)	$\Delta\sigma\epsilon$(SMP)	γ_1(PUN)	$\Delta\gamma_1$(PUN)	$\sigma\epsilon$(PUN)	$\Delta\sigma\epsilon$(PUN)
2005	0.928	-	0.055	-	0	-	0.038	-
2006	0.000	-0.928	0.007	-0.0543	0.8077	0.807	0.009	-0.026
2007	0.1414	0.1414	0.013	0.006	0.2136	-0.594	0.071	0.062
2008	0.852	0.710	0.010	-0.003	0.9628	0.749	0.006	-0.065
2009	0.778	-0.074	0.027	0.017	0.6436	-0.319	0.014	0.008
2010	0.0204	-0.757	0.011	-0.016	0.753	0.109	0.023	0.009
2011	0.651	0.630	0.004	-0.007	0.950	0.197	0.017	-0.006
2012	0.865	0.214	0.008	0.004	0.183	-0.767	0.042	0.025
2013	0	-0.865	0.012	0.004	0.897	0.714	0.032	-0.01

TABLE 3.14: Herfindahl-Hirschman Index (HHI) in Greek and Italian Power Generation 2005-2014 as appeared in the National reports of Regulators sent to EC

Year	HHI (Greek Market)	HHI (Italian Market)
2005	9409	1900
2006	8905	1643
2007	8390	1440
2008	8390	1380
2009	8427	1280
2010	6844	1097
2011	5746	953
2012	6983	884
2013	6553	830
2014	8091	908

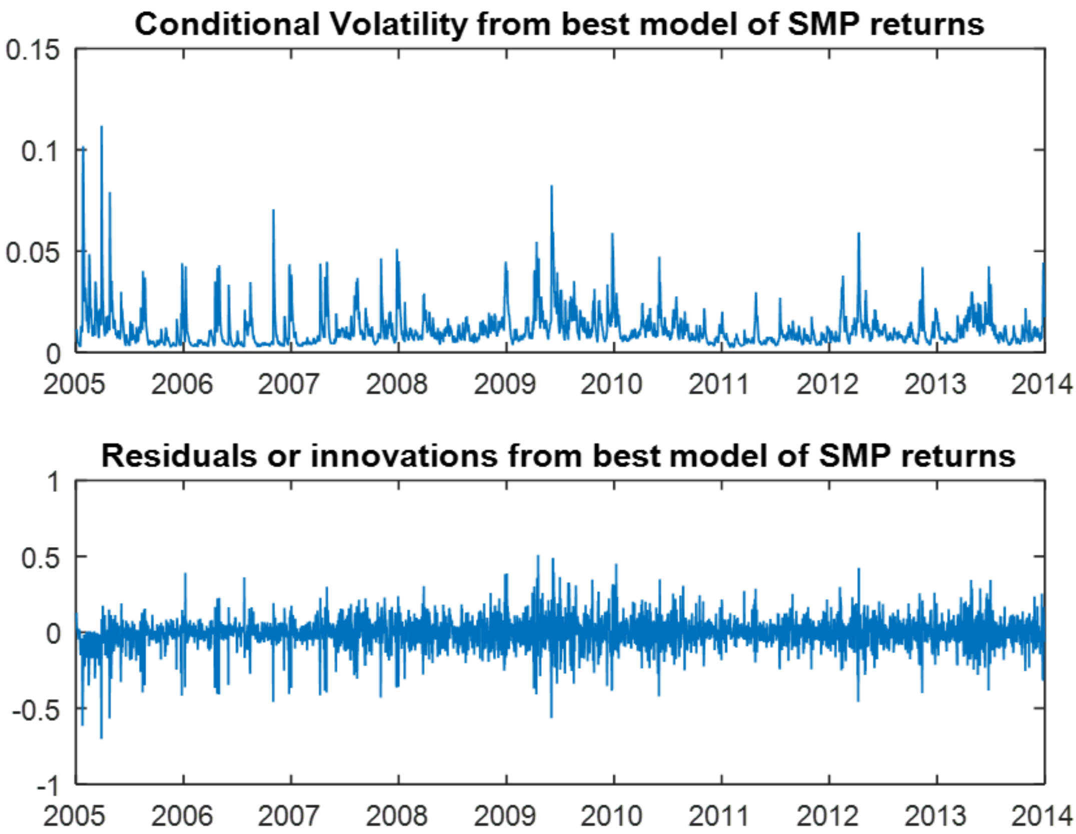

FIGURE 3.15: The graphs of conditional volatility and residuals of SMP returns from model given in table 3.11.

H>1800 concentrated, for 1000<HHI<1800 efficiently or moderately competitive and finaly for $HHI \leq 1000$ the market is low concentrated or competitive. The Index is calculated as follows:

$$HHI_n = 10000 \sum_{i=1}^{n} s_i^2 \tag{3.3}$$

where s_i is the market share of generator i.

According to Table , we can see that concentration levels for Italy and Greece have decreased, but the Italian market at the start of the period of analysis 2005, was a medium concentration market, while the Greek market was close to the monopoly top limit. However, the Greek market, even with a very slow rate, gradually becomes more and more less concentrated with the lowest HHI index value in 2011. However, it still remains a highly concentrated generation market. In the opposite direction, the Italian market decreased the level of concentration very fast and since 2007 went beyond the 1500 threshold. The reason for this fast moving can be attributed to various causes: 1) During 2001-2002, ENEL sold out 1500 MW, while at the same time ENEL's competitors increase their investment for installed power capacity, very intensively in the last years, causing a rise in green energy consumption. 2) Recent EU Energy policies, enhanced the energy purchasing from green energy generators. 3) The steady increase in the number of marginal generators (producers) which invested in wind and solar technologies, transforming the Italian generation

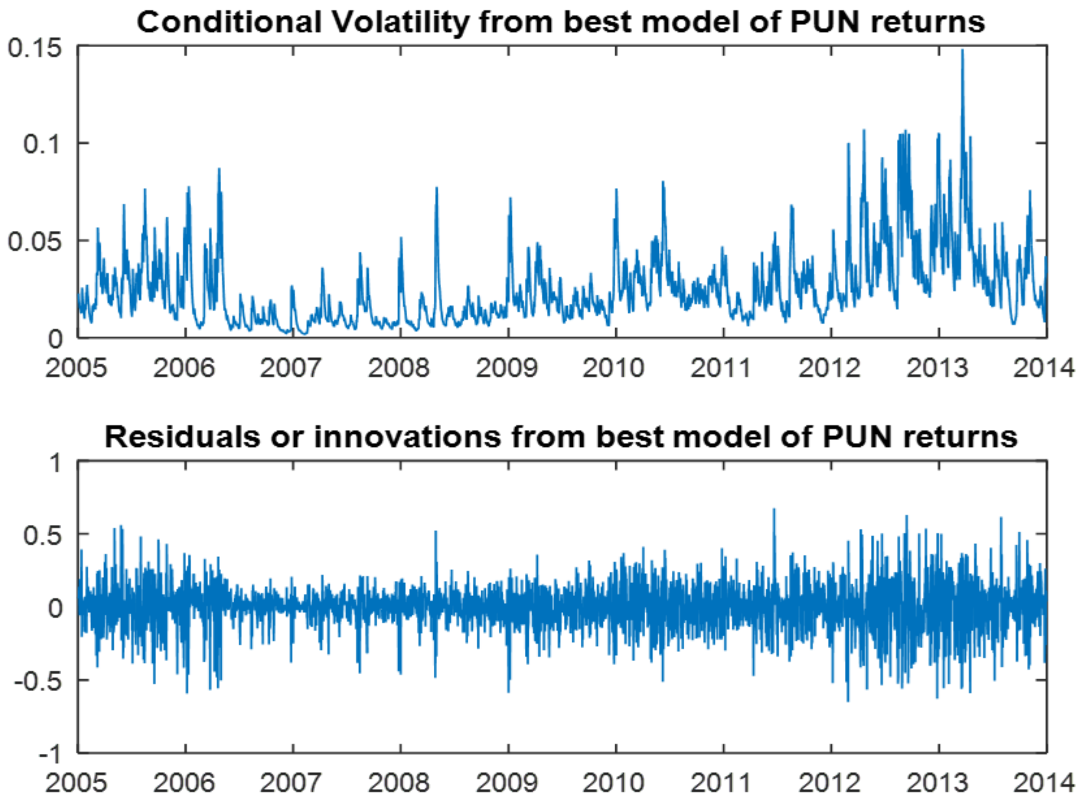

FIGURE 3.16: The graphs of conditional volatility and residuals of PUN returns from model given in table 3.12.

"market" to the most competitive one. Table 3.15 summarizes the unconditional volatility on raw returns and log returns for both SMP and PUN series. Also HHI, Hurst and Lyapunon exponents have been calculated for both markets.

From table we conclude that less price volatility in a market means that the dynamical system generating the price series is more stable i.e the max Lyapunov exponent λ_{max} is more negative, therefore the correlation coefficient between σ and λ_{max} is expected to be negative $corr(\sigma_t, \lambda_{max}) < 0$. If the market competition increases which is equivalent to say that market concentration reduces or HHI decreases, then the distance Hurst exponent from the value of 0.5 (EMH) is expected also to decrease (since the market becomes more and more efficient, and Hurst moves toward 0.5). Therefore the correlation between Hurst distance from 0.5 and HHI is expected to be positive i.e Corr (Hurst distance, HHI) > 0. The more efficient becomes the market (i.e Hurst distance from 0.5 is getting less and less) the less volatile is the market, so Hurst_distance and σ are positively correlated : Corr(Hurst distance, σ) > 0. The former conclusions are summarized on table E7; from this table we extract useful conclusions regarding the correlation of conditional and unconditional volatility, HHI, Hurst and, Lyapunov exponents as well as the correlation between Hurst distance from 0.5, for both markers; these are designated as: uv_smp for the unconditional volatility of the SMP, uv_des_smp for the deseasonalized unconditional volatility of the SMP, C_VOLAT_SMP for the conditional volatility of the SMP, HHI_G is the HHI for the Greek Market, Hurst_RS_smp is the Hurst exponent of the SMP, H_smp_dist

TABLE 3.15: Annual mean values of volatility HHI, Hurst and maximum Lyapunov exponents for the Greek and the Italian markets

Year	Unconditional Volatility (St.dev.) raw returns	Unconditional Volatility (St.dev.) fully deseasonalized log-returns	HHI (Greek Market)	HHI (Italian Market)	Hurst-R/S	Lyapunov, λ_{max}
	SMP,PUN	SMP,PUN			SMP,PUN	SMP,PUN
2005	0.226,0.230	0.172,0.251	9409	1900	0.2220,0.2050	0.1102,-0.2570
2006	0.145,0.179	0.114,0.164	8905	1643	0.2070,0.2520	0.6550,-0.2090
2007	0.123,0.179	0.092,0.173	8390	1440	0.2790,0.2580	-0.0180,-0.1600
2008	0.134,0.139	0.105,0.130	8390	1380	0.3250,0.2580	-0.0076,-0.1158
2009	0.161,0.164	0.137,0.213	8427	1280	0.2710,0.3230	0.3960,-0.1280
2010	0.201,0.127	0.178,0.095	6844	1097	0.2270,0.2470	-0.0577,-0.1560
2011	0.194,0.087	0.160,0.132	5746	953	0.2900,0.2480	0.1140,-0.1266
2012	0.275,0.130	0.254,0.145	6983	864	0.2040,0.3470	-0.1034,-0.0740
2013	0.277,0.145	0.237,0.087	6553	830	0.3020,0.2300	0.2130,-0.0835
2005-2013	0.199,0.160	0.167,0.164				

TABLE 3.16: Correlation matrix between volatility σ and λ_{max}

Volability σ	$\lambda_{max}(SMP_{lyap})$	$\lambda_{max}(PUN_{lyap})$
Unconditional Volatility (raw)	-0.260	-0.734
Unconditional Volatility (deseasonalized)	-0.287	-0.643
Conditional Volatility (deseasonalized)	-0.0049	0.122 (not expected)

is the Hurst distance from the 0.5 of the SMP, and smp_lyap is the Lyapunov exponent of the SMP. The respective designations for the Italian Market use PUN instead of SMP and IT instead of G. For instance uv_pun is the unconditional volatility of the PUN whereas, HHI_IT is the HHI for the Italian Electricity Market.

3.3 Financial Trading and Portfolio Management

Having applied the several proposed forecasting schemes in the applications related to the electricity markets of the former sections, we extend their spectrum to the classical but extremely complex and complicated to model and predict financial markets. Additionally, to the efficient modelling of the underlying assets of each specific dataset under focus, the goals of each respective application is to extend and test the capacity of each proposed scheme, under a trading and portfolio management perspective, as well. While we have tackled all traditional asset classes, particular interest is given to the Foreign Exchange Markets (FOREX). The developed trading strategies and constructed portfolios performance efficiency is accessed mostly by the Sharpe Ratio metric Sharpe, 1994.

The motivations and ideas' generation of each application, was based on the challenging of the validity of one of the main building blocks of financial theory : the Efficient Market Hypothesis (EMH) Malkiel and Fama, 1970; Fama, 1970 with its separate forms : weak, semi-strong and strong. At this point, we should also mention the more recently developed theory of the Adaptive Market Hypothesis (AMH), by Andrew Lo (Lo and MacKinlay, 1988; Lo, 2002; Lo and Wu, 2004), which also asserts the aspect that behavioral factors should also be migrated in the EMH framework, as for example competition, adaptation and natural selection, which though is not evaluated in the current applications' context, mostly due to their lack of behavioral perspective in their modelling rationale.

In a previous analysis of ours, Papaioannou et al., 2013, we tried to identify exploitable trading patterns in the EUR/USD's price dynamics under the so called "social trading" framework, incorporating information from historical prices, as well. The main assumption was that past historical price dynamics can be useful on efficiently forecasting the future movements of the pair, is tackling the "weak" form of EMH. Additionally the "semi-strong" from was also challenged, by the assumption that the publicly available information, in the form of "Tweets" or messages displayed in the Twitter social media platform and sent by individual traders, members of the platform, can adequately explain future behavior of the pair, as well. Although, modelling financial assets' dynamics by monitoring as higher-frequency related price recordings as possible, seems to be the intuitively best approach to go, such vast datasets are not always available and easily accessible (in terms of both money and even computational capacity/storage/processing) for a wide range of instruments. Furthermore, extending the database by retrieving e.g. even more Twitter messages from millions of users corresponding to even more assets, as well, or additionally try to find new datasets in other platforms, as well (e.g. Google, Yahoo, Facebook and other platforms), would require an enormous computational and storage usage capacity. Since our primary research goals was mainly the testing of time series modelling and forecasting capacity of the several proposed "hybrid" approaches and not the deeper connection of them to this specific financial and economical aspect of the EMH, we focused our subsequent applications on the tackling of just the EMH's weak form of efficiency alone. Towards that direction, the first application outlined in Section 3.4, targets on the deconstruction of the EMH's weak form of efficiency, by focusing into one of the most popular and liquid Stock Market Indexes of the global financial exchanges : the S&P 500. The shift to the Stock markets or Equities asset class, was due to the fact that the Stock Market is considered to be the "corner stone" of financial investing and trading, with tones of publications connecting its dynamics with the movements of all other financial assets' price fluctuations, as well. Its popularity, with even Wall Street movies having been devoted to their effect on global economies and expectations, along with its easily accessible structure, has led millions traders, both professional and retail ones, to become members of its vast ecosystem. As such, by using S&P500, as the most representative one of its class, the assumption that it is incorporating all available information, by all participating traders' beliefs, trading not only in the Equities Markets, but across all financial markets globally, is not sounding so unrealistic. Upon this rationale, the target of this application is to test if S&P500, and more specifically its Generic Futures Contract, is "predictable". The "predictability" test is accessed not just in terms of time series modelling capacity, but also under the possibility of an efficient and profitable trading strategy construction perspective. As such, this enforces the overall scheme to a "directional predictability" perspective, i.e. on the identification of "up" or "down" movements in the underlying modelled trading asset dynamics and the formation of "Buy" and "Sell" trading signal upon the corresponding predictions. The main tested assumption is that under the proper identification and use of information related to its historical price history and its connection (correlation) to the historical prices of a wide list of other popular and liquid trading assets, a profitable and efficient (in terms of Sharpe Ratio scores) trading strategy can be implemented, disproving the weak form of EMH. Furthermore, we compare the performance metrics with the classical "Buy and Hold" approach of financial investing, while discussing the pitfalls and limitations of the approach, connecting them with the models and methodologies developed in the rest of the applications, respectively. As a key note here, we mention that no transaction costs analysis is being performed in this application's framework, since we are only interested on tackling the EMH's weak form of efficiency under a pure data and pattern recognition related perspective. However, we extend our work by also incorporating market frictions in the subsequent applications' analysis, as carefully outlined and discussed throughout the following paragraphs.

Having proven existence of "tradability" (at least under the purely data and pattern recognition related perspective discussed above), we accept the statement that the implementation of (consistently) winning trading strategies with an appealing risk/return trade-off (as depicted by the Sharpe Ratio metric) is probable, and also that specific correlation based approaches can help us towards such strategies design. As we comment in the relevant application though, except from this correlation existence and "tradability" validation, the specific S&P500 approach lacks realistic justification, in the sense that the selected as "best forecasting" correlations between the several assets and the S&P500 are sorted and ranked in an "ex-post" manner. This "ex-post" approach's limitations and drawbacks, which are thoroughly discussed in the relevant section of the application, will be the ones that we will try to overcome under the proposed "hybrid" methodologies of the subsequent applications. Furthermore, in these applications we extend our approaches on a multivariate time series modelling and forecasting regime, moving in the field of portfolio construction, selection and management of financial investing. More specifically, instead of focusing on the trading of specific single trading assets (like the EUR/USD of the first application and the S&P500 Index in the second), we test our modelling schemes on their capacity to construct, select and trade financial portfolios comprised of several underlying traded assets, ideally able to outperform the standard (and even the modern) practices and theories of financial portfolio management literature.

While this analysis could be performed using all assets comprising the dataset used for the S&P500 trading application, we limit our investable space in only a specifically selected list of FOREX pairs. Being accessible from a historical **"spot"** prices' dataset perspective, we evaluate the FOREX markets' data as exhibiting richer and more trustworthy "data quality", in terms of realistic execution assumptions and final obtained performance results. More specifically, regarding making realistic assumptions about the potential markets' frictions and costs affecting the results of the analysis, FOREX is much easier to handle than the Derivatives Markets' Futures contracts counterparts. Additionally to assumptions related to market Spread (Bid/Offer) and Slippage, which should be also made for the FOREX markets as well, potential Rollover and Volatility effects need to be properly incorporated in the analysis for the case of the rest of the Futures Contracts, as well. For that reason, since Transaction Cost Analysis (TCA) seemed to be "safer" for the case of the FOREX Markets we focused the whole approach only on them. The whole dataset could be used, in the case where we would be granted access in databases with such Transaction Costs related data, e.g. average execution costs reports' data of a vast majority of large institutional players like for instance, Banks, Brokerage Houses, and Market Makers from the so called "sell" (liquidity offering) side, displaying the effect of such market frictions on the quoted prices. Being "OTC" (Over the counter) in nature and not centralised-exchange based as the Derivatives markets, FOREX markets' average daily turnover is the world's largest OTC one, with its level being around $5 trillion. The constantly increasing in number electronic liquidity providers, offering services to one-another, to institutional "Buy"-side professional investors, but also to retail traders utilising the potential of "leverage" trading, have made the access to those markets via their electronic platforms extremely easy and richer in data quality. Average costs in terms of commisions and/or markups, as well as slippage effects incurred per transaction, are displayed for free even in real time and on a tick-by-tick basis, via numerous online publicly available forex-data related web platforms, in order for traders to be able to distinguish, filter and select the best performing (or less charging) vendor for their trading needs. We exploited this information to perform the Transaction Costs Analysis (TCA) in our applications, as well, with the details being outlined in the relevant section below.

However, the same "hybrid" modelling approaches presented in the FOREX trading respective applications, could of course be easily extended and applied on the wider cross-asset-classes' dataset, as well, while though careful consideration on how the peculiarities related to the several investable derivatives instruments' pricing (quotes) and trading execution difficulties and frictions (costs) would affect the obtained performance results. The previously discussed effects and execution limitations should be carefully incorporated in these results, in order for them to have realistic meaning and be of any true value to practitioners.

3.4 S&P500 Forecasting and trading using convolution analysis of major asset classes

Ref. Paper : Papaioannou et al., 2017

3.4.1 Motivation

Corporate finance theorists, macroeconomists, behavioral psychologists, quantitative finance mathematicians have teamed up to find the "holy grail" of the financial markets: that of consistently beating the market. Towards this aim methods can be categorized into (a) fundamentals, and (b) financial engineering. On the one hand, fundamentals include models most of which try to approximate the dynamics of macroeconomic observables such as the aggregate demand and supply, investment volume and consumption, risk premia etc. On the other hand, financial engineering and quantitative finance models try to forecast the price action of several financial instruments such as the S&P 500 Index on a data driven basis.

Both approaches focus basically on several sub-targets such as the forecasting of a single asset as well as asset allocation methods among the four basic asset classes, namely, Equities, Commodities, Bonds, and Foreign Exchange Markets (Bekkers, Doeswijk, and Lam, 2012; and, Husain, and Bowman, 2004; Diebold and Li, 2006; Papaioannou et al., 2013). Several models have been proposed to forecast their future dynamics based on historical data and information acquired from almost all possible sources: price action, fundamentals as well as behavioral sentiment analysis and e-social platforms (Polk, Thompson, and Vuolteenaho, 2006). Even though success stories have been reported, these models are still prone to failure in situations undergoing structural changes and market crashes Zellner and Chetty, 1965; Klein and Bawa, 1977; Kandel and Stambaugh, 1996; Barberis, 2000; Caballero, Farhi, and Gourinchas, 2008.

Here we propose an approach to forecast the E-mini S&P500 Futures contract exploiting convolution analysis of 42 of the most liquid Futures contracts of four basic financial asset classes:

1. equities
2. bonds
3. commodities
4. foreign exchange (FOREX)

Our choice of convolution is motivated by the fact that we want to use a "non-memory" approach, one that does not need training datasets (e.g. ANNs etc.) and parameters (i.e. the length of the signals for correlation calculation, or the "lag" of the correlation operation) and which would provide a point wise, time dependent and out-of-sample trading strategy, even from the first few observations of the historical dataset. Based on the proposed approach, we managed to successfully forecast the S&P500 Futures Contract beating the "Buy and Hold" benchmark trading strategy.

3.4.2 Methodology

The question of how financial uncertainty gets incorporated in the risk premia offered by several financial assets and how it formulates the investors' preferences and their corresponding portfolios allocation is fundamental in contemporary financial research. Previous studies have shown that allocation decisions made by fund managers on behalf of the investors are shaped by several bounds regarding the portfolios weights of each asset class in order to ensure portfolio robustness, lower transaction costs (turnovers), longer investment horizon, smaller concentrations etc.Roncalli, 2012; Roncalli and Weisang, 2016; Jagannathan, Ma, and Zhang,

2019. Building up on these studies, our main hypothesis is that the shifts between the combinations of asset classes must be smooth and relatively slow, in order to achieve solid, safe and robust allocation of capital across economies and markets. An argument that supports this hypothesis is that the investors rebalance their portfolios via trading orders at global financial exchanges. Another argument towards the belief that the global trading microstructure affects the market dynamics is that, responsible execution venues that govern the majority of financial transactions worldwide, have to fulfill restrictions imposed by state financial authorities. Under normal market conditions, the shifts in the investments between asset classes that are influenced predominantly by supervised institutions must be relatively slow and smooth, as well as interconnected under some invariant internal properties of the entire market itself. The strong structural limitations imposed by the state authorities to the most influential market participants suggests that there must be a pattern that depicts the effects of several invariant properties of the biggest asset classes' dynamics. In addition, market participants follow the rationale of the risk diversification benefits among the several asset classes. This basic investing rationale supports the hypothesis that the intercorrelation dynamics of asset classes are themselves smooth. Within this framework, we used convolution analysis to find patterns among the representatives of the four biggest financial asset classes that would provide "good" forecasters for indices such as the S&P500. For our analysis we used Bloomberg tickers downloaded via the Bloomberg API from 03 January 2000 until 15 December 2014. We considered the most liquid futures contracts of 42 financial assets traded globally. The assets were selected on the basis of the global GDP ranking across countries worldwide according to the lists published by the International Monetary Fund (IMF). These lists traditionally include the following countries: EU, USA, China, Japan, Germany, UK, France, Brazil, Italy, India, Russia, Canada, Australia, South Korea, Spain and Mexico. Additional asset selection criteria were gleaned from the monthly reports of the World Federation of Exchanges depicting the volume ranking of the global stock exchanges by market capitalization, as well as the volume ranking of the global total value of bonds trading. Following common practices for a representative aggregation of data worldwide, we selected the basic assets for the USA aggregating the America region, the same basic assets for the EU aggregating the European region and the basic assets for China, Japan, Australia and New Zealand for the Asia region. Due to lower trading activity and volume, the emerging markets' dynamics is targeted solely in the FX market (see e.g. investopedia.com, world's most traded currencies by value 2012. 2013; Bank for International Settlements, Triennial Central Bank Survey, 2013). Our heuristic mining approach to try and efficiently forecast the underlying dynamics of ES1 Index resulted to the following list :

Equities Markets: E-Mini S&P 500 (ES1 Index), E-Mini Nasdaq 100 (NQ1 Index), Eurex DJ Euro Stoxx 50 (VG1 Index), Cac40 (CF1 Index), Eurex Dax (GX1 Index), Ftse 100 (Z1 Index), Ibovespa (BZ1 Index), Swiss Market Index (SM1 Index), Mexican Market Index (IS1 Index), Australia Market Index (XP1 Index), Nikkei (NK1 Index), Topix (TP1 Index), Hang Seng (HI1 Index)

Bonds Markets: 10 Year U.S. T Note (TY1 Comdty), 2 Year U.S. T-Note (TU1 Comdty), Canadian Government 10 Year Note (CN1 Comdty), France Government Bond Future (CF1 Comdty), Eurex Euro Bund (RX1 Comdty), Eurex Euro Schatz (DU1 Comdty), Gilt UK (G1 Comdty), 5 Year T-Note (FV1 Comdty), Japan 10 Year Bond Futures (BJ1 Comdty), Australian 10 Year Bond (XM1 Comdty), Australian 3

Year Bond (YM1 Comdty), Euro Bobl (OE1 Comdty)

Commodities Markets: Natural Gas (NG1 Comdty), Gold (GC1 Comdty), Silver (SI1 Comdty), Crude Oil (CL1 Comdty), Corn (ZC1 Comdty)

FX Markets: EURUSD Curncy, USDJPY Curncy, EURJPY Curncy, EURCHF Curncy, GBPUSD Curncy, EURGBP Curncy, AUDUSD Curncy, AUDJPY Curncy, NZDUSD Curncy, EURAUD Curncy, USDRUB Curncy, USDCNH Curncy, USD-MXN Curncy, USDINR Curncy

The dataset records are daily futures closing prices, while the analysis was based on log-returns :

$$r_{i,t} = log(\frac{P_{i,t}}{P_{i,t-1}}) = dlog(P_{i,t}) \tag{3.4}$$

with $P_{i,t}$ denoting the price of the i-th asset in the dataset, at time t, with $i = 1, 2, \ldots, 42$. Writing the cumulative return at time $t+1$, of a trading strategy applied in an investable asset (in our case the ES1 Index), evaluated up until time τ, as :

$$\Pi_{t+1} = \sum_{t=0}^{\tau} W_t r_{ES,t+1} \tag{3.5}$$

with the W_t being the "trading weight" or "signal", we observe that it actually reminds us of the following three frameworks of inferring "relationship" between W_t and $r_{ES,t}$:

1. Signal Processing : a convolution relationship between the variables W_t and $r_{ES,t}$.
2. Statistics : correlation between the two time series
3. Geometry : dot product between the vectors $\vec{\boldsymbol{W}}$ and $\boldsymbol{r}_{\vec{ES}}$, (where "time" and sequential order has been removed)

More specifically, geometrically interpreted, the two signals W_t and $r_{ES,t}$ can also be considered the vectors, $\vec{\boldsymbol{W}}$ and $\boldsymbol{r}_{\vec{ES}}$ and equation 3.5 is actually their dot product :

$$\begin{aligned} \vec{\Pi} &= \vec{\boldsymbol{W}} \cdot \boldsymbol{r}_{\vec{ES}} \\ &= \|\vec{\boldsymbol{W}}\| \|\boldsymbol{r}_{\vec{ES}}\| cos\theta \end{aligned} \tag{3.6}$$

with $cos\theta$ denoting the cosine of their angle.

Convolution and correlation are two similar concepts and operations, with the only difference that in the correlation process between two arbitrary signals, the rotation of one the two might be necessary in order for "rotational symmetry" to be guaranteed. As such, correlation can be considered a convolution operation between two signals, with the one of the two being the "functional inverse". Following this rationale, we could also express the correlation between the two signals as :

$$\begin{aligned} \rho_t &= W_t * r_{ES,t} \\ &= \int_{-\infty}^{\infty} W_\tau r_{ES,(\tau-t)} d\tau \end{aligned} \tag{3.7}$$

where this expression is addressing the operation a continuous time regime, and $*$ denotes the correlation operation, while in the discrete time regime, it reads as follows :

$$\rho_n = \sum_{k=-\infty}^{k=\infty} W[k] r_{ES}[k-n] \tag{3.8}$$

Correspondingly, convolution is expresses as :

$$\begin{aligned} h(t) &= W_t \otimes r_{ES,t} \\ &= \int_{-\infty}^{\infty} W(\tau) r_{ES,(\tau-t)} d\tau \end{aligned} \tag{3.9}$$

where $\otimes$ represents the convolution operation., which again in expresses the operation in the continuous regime and in the discrete one it reads :

$$h(n) = \sum_{k=-\infty}^{k=\infty} W[k] r_{ES}[n-k] \tag{3.10}$$

Based on the above framework, exploring the "best choice" for W_t could be associated with identifying the best "kernel" in the case of convolution, or "highest correlation" in terms of statistics or even a vector being "the closest to parallel" to the vector of the ES1 Index returns. Under this logic, towards expressing the trading weight in terms of a cross-asset information set, we should think as "find the highest correlated assets" with S&P500 under statistics, or the assets "with the lowest or ideally zero angles". Such an approach would require a threshold setting for both the correlation (if e.g. we used the Pearson Correlation Coefficient) level above which we should characterize an asset as "highly correlated" or an angle level below which the asset is "close to parallel" with the ES1. Since the convolution operation is the only one without such a parameterization requirement and since our primary target in this application is to prove existence of useful patter regarding the "tradability" of ES1 using information from all other assets in the dataset, we use this one for the construction of the following W_t proposed candidate :

1. We calculate all individual convolutions between the several assets and the ES1 Index :
$$g_{i,t} = r_{i,t} \otimes r_{ES,t}, \quad i = 1, 2, \ldots, 42 \tag{3.11}$$
where $g_{i,t}$ denotes the convolution of the i-th asset's returns with the respective returns of ES1 Index.

2. We form a "first candidate" strategy using this corresponding individual expressed convolution under the assumption that one of the two assets is a "driver" or "leader" regarding the dynamics of the other. To replicate this relationship, we use a trading rationale used by many technical traders, as well as quantitative analysts, trying to identify underlying trending or mean reverting periods in the price dynamics, meaning the use of Exponential Moving Averages. We use three different moving averages approaches targeting a specific period pattern "cyclicality". In particular we used Exponential Moving Averages (EMA) with three separate time lags (trading days) as inputs: 3, 25 and 500 days. Using these three lags we aimed at capturing different convolution patterns of the ES_t relative to the whole other assets' universe targeting daily,

monthly and 2-year trading activity, respectively. The respective trading signal is then :

$$s_{i,t} = sign(EMA_L(r_{i,t})), \quad i = 1, 2, \ldots, 42 \tag{3.12}$$

where EMA_L denotes the exponential moving average with L being the lag period : $L = 3, \quad 25, \quad 500$ trading days. The sign function is used to "filter" noise effects of skewness and kurtosis emerged by the assets' probability distribution, thus focusing only on the positive or negative jumps of the price movement in a binary sense (1 for positive, -1 for negative, 0 for neutral).

3. We calculate the Sharpe Ratios of sub-strategies, which trade the ES1 Index, as :

$$\pi_{i,t} = s_{i,t} r_{ES,t}, \quad i = 1, 2, \ldots, 42 \tag{3.13}$$

as (Sharpe, 1994) :

$$sh_i = \frac{\mu_\pi}{\sigma_\pi} \tag{3.14}$$

with μ_π being the average and σ_π being the standard deviation of the strategy's returns

4. We sort and select the top N Forecasters, i.e. the N assets corresponding to the highest N ("absolute") Sharpe Ratio scoring sub-strategies.

   ****This step is inferring an "ex-post optimization" or "ex-post selection" rationale, which in turn yields an undesirable "cherry picking" perspective in the overall trading framework. However, as already stated in the beginning of the modelling process, it is sufficient on acting as a data-driven empirical evidence vehicle, against the EMH's weak form efficiency and more specifically on the cross-asset analysis undertaken in the above process.*

5. Calculate the coarse grained variable, acting as our proposed "best kernel" or "best trading weight", based on the N top forecasters as :

$$W_t = \sum_{j=1}^{N} s_{j,t}, \quad j = 1, 2, \ldots, N \tag{3.15}$$

which is the final trading signal applied on ES1 Index, and which comprises the cross-asset information of the N "Best" forecasters in the dataset.

6. Risk Calibration : Calculate of the total strategy's returns (Π_t), as per Equation 3.5 and the corresponding rolling annualized volatility at each time step t, σ_Π. Divide the trading signal with this rolling volatility, to replicate the approach as per Perchet, Carvalho, and Moulin, 2014, i.e. target a 1% volatility in the overall strategy's final returns, (risk parity approach).

7. Report the final risk parity-filtered performance in terms of Sharpe Ratio.

We show that this strategy can be used to trade the S&P500 future contract with better results than the "Buy and Hold" benchmark strategy, offering therefore a relatively simpler approach compared to other quantitative trading methods.

TABLE 3.17: The Sharpe Ratios as computed using the proposed trading approach with respect to different EMA lags (3, 5, 500), as well as different number of forecasters (10, 20, 30, 42).

# Forecasters	EMA(3)	EMA(25)	EMA(500)
10	0.75	0.55	0.82
20	0.43	0.45	0.80
30	0.48	0.58	0.60
42	0.67	0.9	0.70

3.4.3 Results and Discussion

The performance of the trading strategy for the ES1 Index - future contract was tested using four different sets of forecasters ($N = 10, 20, 30, 42$) and three time lags ($L = 3, 25, 500$). Table 3.17 summarizes the Sharpe ratios as obtained by the proposed methodology under the different scenarios with respect to different EMA lags as well as to different number of forecasters. Figure 3.17 shows the cumulative returns when we used the entire proposed dataset in order to forecast the ES1 Index. Figures 3.19, 3.21, 3.23, 3.25, show the corresponding weights for this "complete" portfolio. The results suggest that the cumulative performances for all three portfolios (corresponding to the three different lags) are stable. robust, beating the ES1 Index's performance. As it is shown all three portfolios stay close to each other in terms of the cumulative returns. The Buy and Hold strategy's Sharpe ratio of the ES1 Index defined on the raw returns of the index itself is **0.10** for the period under study. Hence, the proposed methodology outperforms consistently the "Buy and Hold" benchmark.

Thus, these results depict that combinations of long term dynamics/basket of forecasters produce a robust and consistent way of generating significant returns when trading the ES1 Index. The values of the Sharpe ratios can also help indicate the best combination using fundamental analysis on the optimal forecasters.

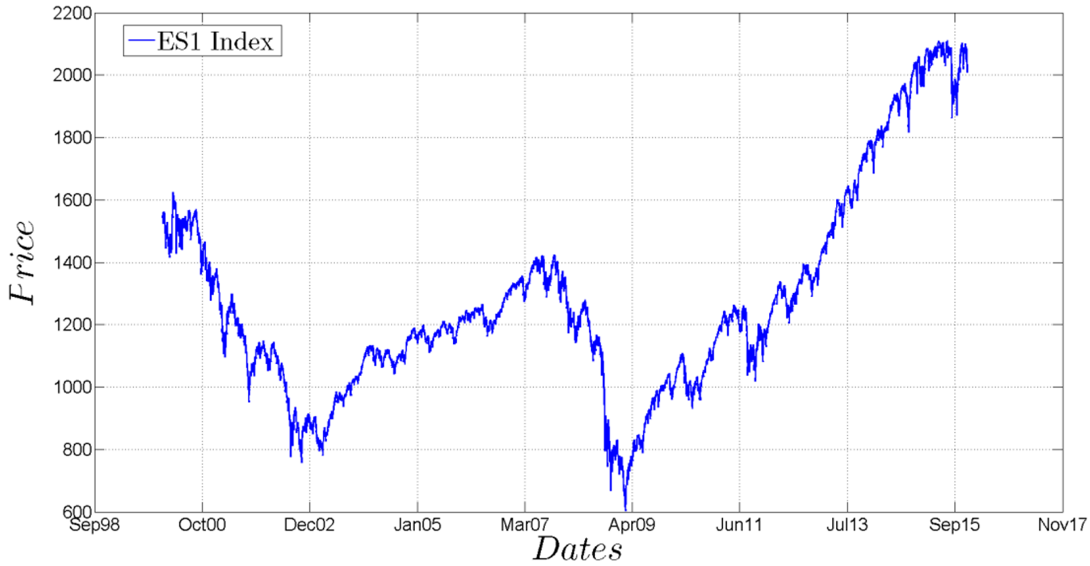

FIGURE 3.17: The ES1 Index spot price in the time period 03 January 2000 -15 December 2014.

FIGURE 3.18: The Cumulative Returns performance, Π_t, of the strategy trading the ES1 Index using 10 optimal assets as forecasters for the three different time lags (3, 25, 500 days).

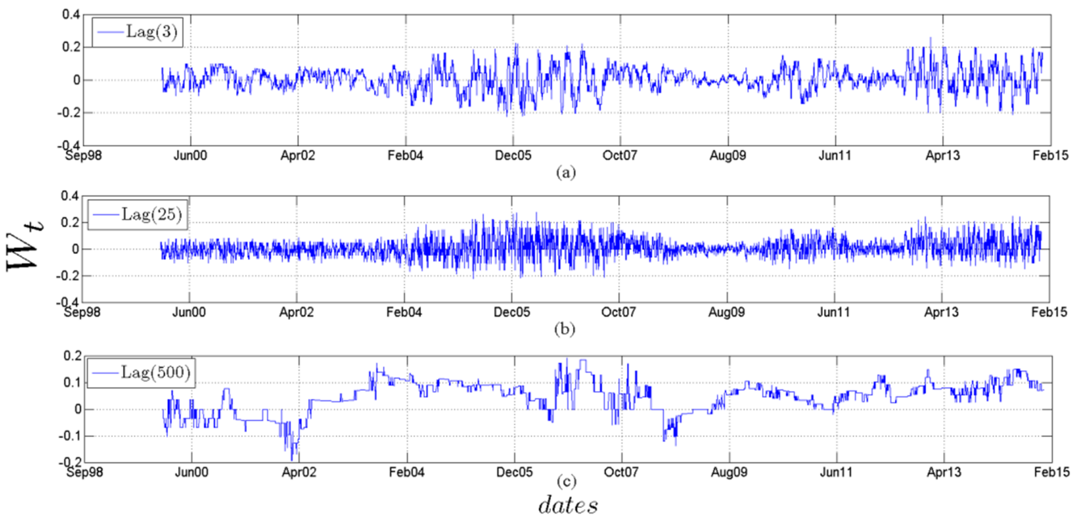

FIGURE 3.19: The weights, W_t of the strategy trading the ES1 Index using 10 optimal forecasters for a time lag of (a) 3 days, (b) 25 days, (c) 500 days.

Even though, the challenge to identify a cross-asset information based, excess returns' generating mechanism was achieved, we should highlight that an "ex-post" approach has been followed throughout the strategy creation, with the "Best Sharpe" approach inferring a "cherry picking" unrealistic pitfall, as discussed earlier. Another path could be to try to identify on a solely data-driven based approach the governing features. We extend all above, with the use of Data mining techniques and more specifically with the hybrid approaches outlined in the following applications, towards identifying the parsimonious features under time varying realistic

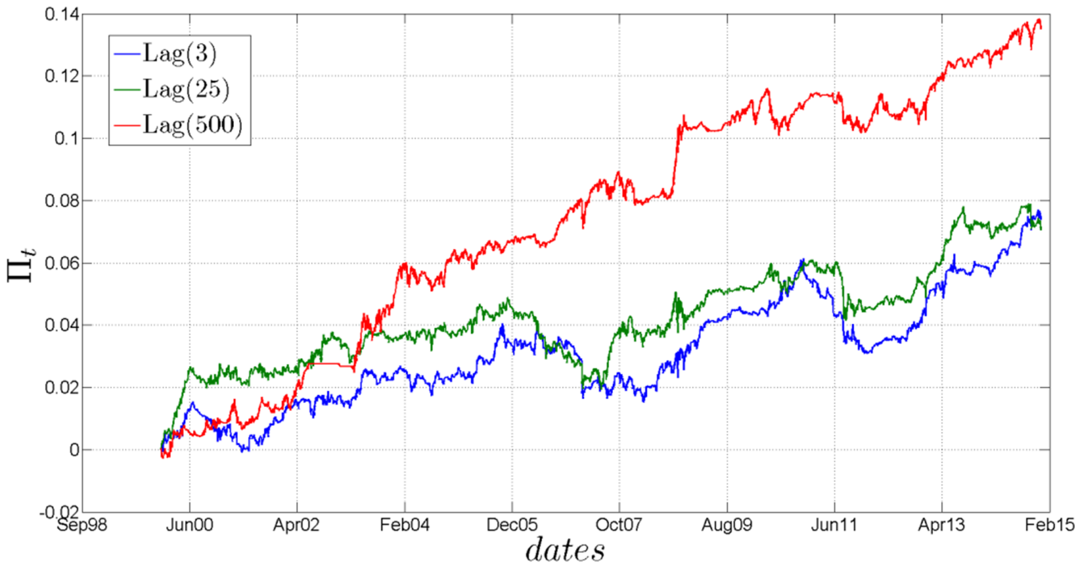

FIGURE 3.20: The Cumulative Returns performance, Π_t, of the strategy trading the ES1 Index using 20 optimal assets as forecasters for the three different time lags (3, 25, 500 days).

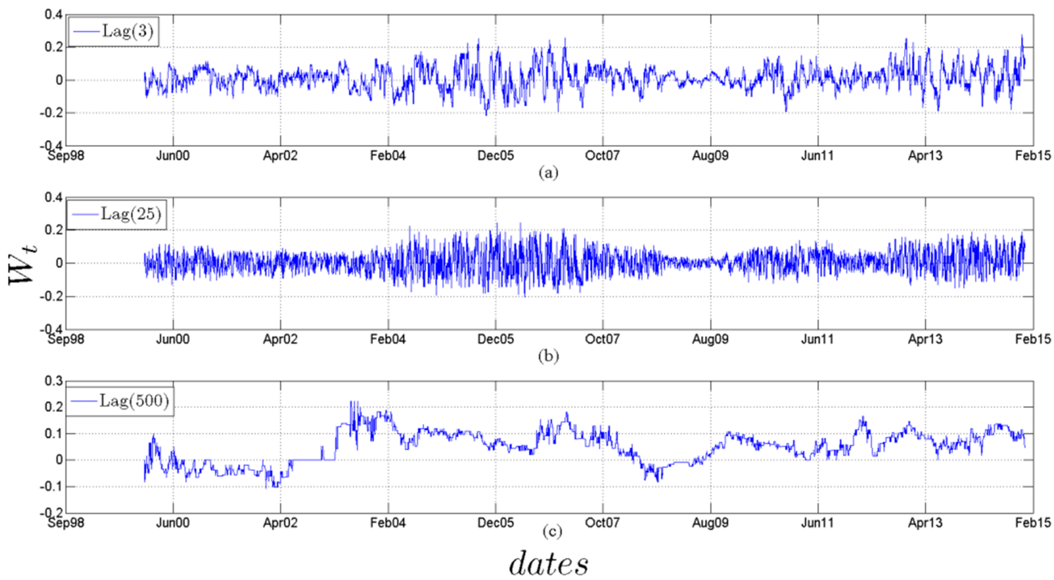

FIGURE 3.21: The weights, W_t, of the strategy trading the ES1 Index using 20 optimal forecasters for lag of (a) 3 days, (b) 25 days, (c) 500 days

out-of-sample frameworks.

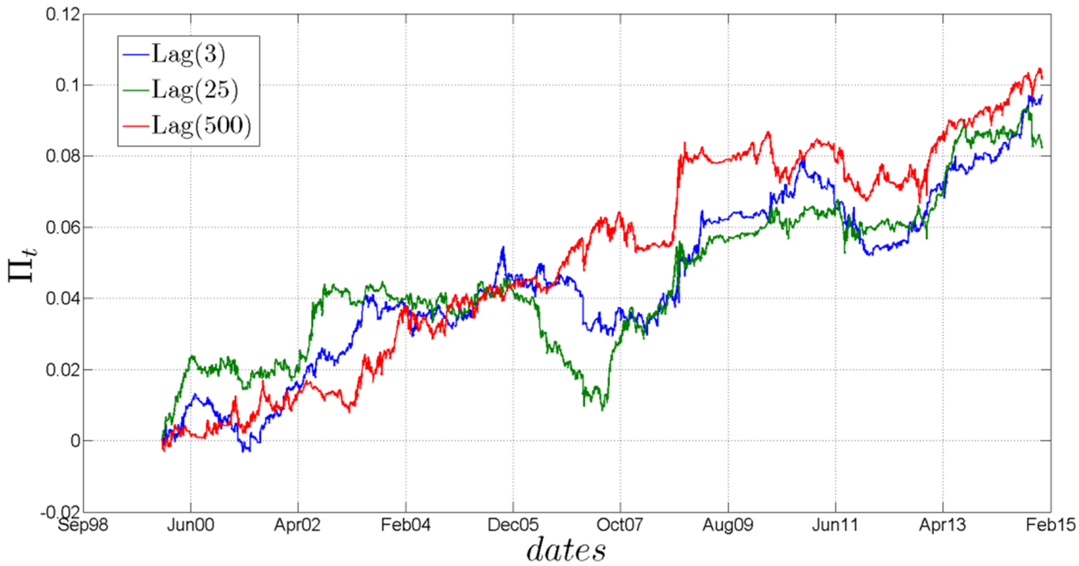

FIGURE 3.22: The Cumulative Returns performance, Π_t, of the strategy trading the ES1 Index using 30 optimal assets as forecasters for the three time lags (3, 25, 500 days).

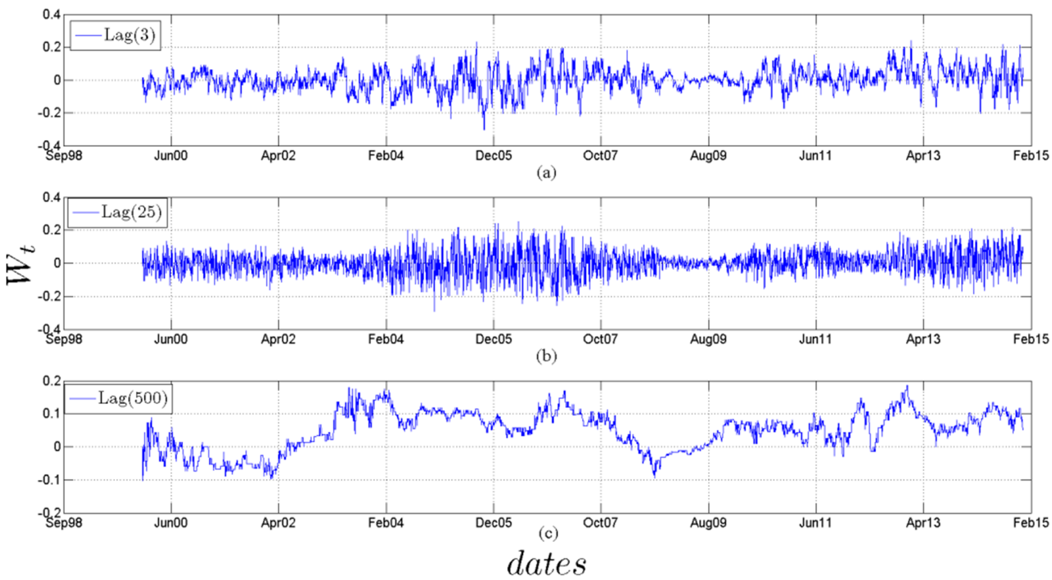

FIGURE 3.23: The weights, W_t, of the strategy trading the ES1 Index using 30 optimal forecasters for a time lag of (a) 3 days, (b) 25 days, (c) 500 days.

3.5 Portfolio Management and Trading in the FOREX markets using PCA and LLE - Sparsity vs Volatility

3.5.1 Motivation

Much work has already been performed on the applications of Supervised Machine Learning using Neural Networks on the stock markets, but also on the FX as well.

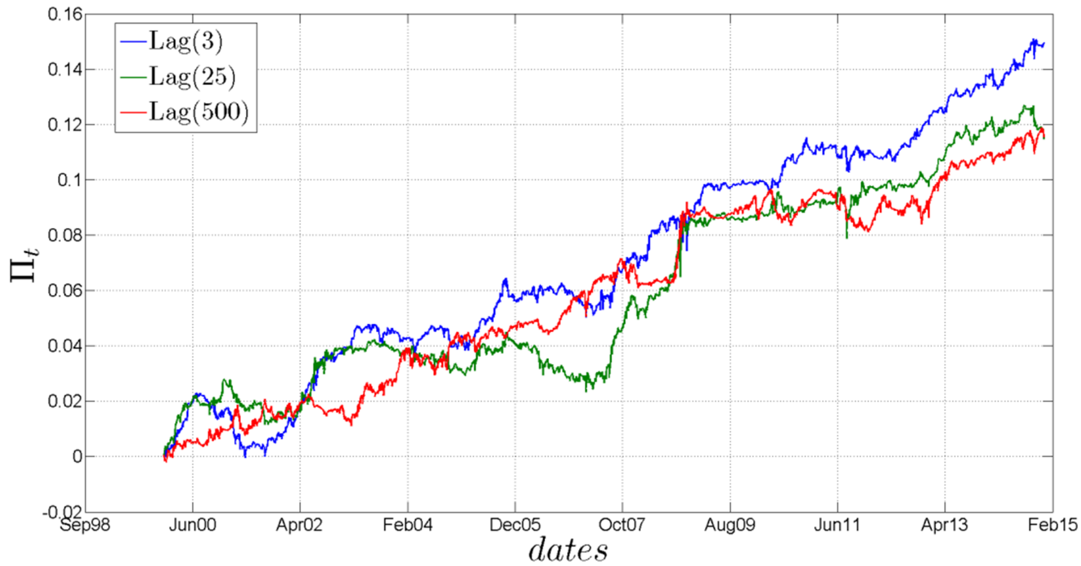

FIGURE 3.24: The Cumulative Returns performance, Π_t, of the strategy trading the ES1 Index using the entire assets universe as forecasters for the three different time lags (3, 25, 500 days).

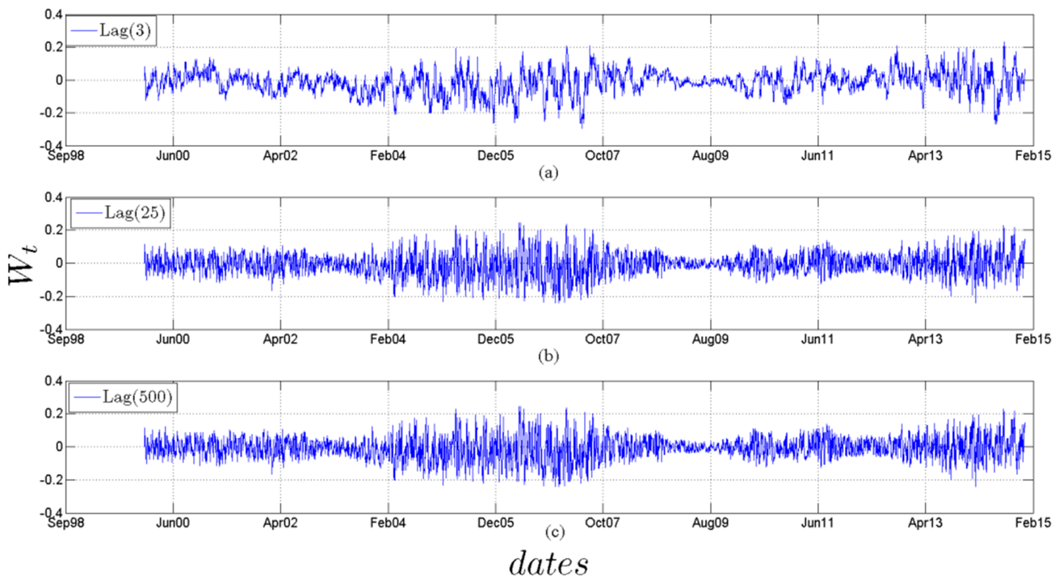

FIGURE 3.25: The weights, W_t, of the strategy trading the ES1 Index using the entire assets universe as optimal forecasters for a time lag of (a) 3 days, (b) 25 days, (c) 500 days.

We cite Byun et al., 2015 who applied PCA and an Artificial Neural Network classifier on a multicurrency dataset, reporting satisfactory trading results, but also Dautel et al., 2020 who applied LSTMs and GRUs in several simple recurrent networks' and feedforward networks' architectures, reporting that the more complex the architecture gets, the worse the performance in trading efficiency terms becomes. They also comment on the existence of inconsistency and divergence between the several machine learning models' fitting performance and efficiency versus the financial and economical one, in terms of trading profits and Sharpe Ratios' scores, which has already been commented and outlined by several other recent research works, as well, (Leitch and Tanner, 1991 and Fischer and Krauss, 2018). Much academic and practical debate has been around the proper way of applying each time series modelling algorithm, under a time varying regime.

Focusing even more in trading related papers, we cite one of the core and most inspiring papers, for our work here as well, namely the "Mean-Reverting Portfolios: Tradeoffs Between Sparsity and Volatility" paper by Cuturi and D'Aspremont, 2016. Its main target is to construct, model and trade mean-reverting pairs-trading portfolios, (Gatev, Goetzmann, and Rouwenhorst, 2006, Krauss, 2017, Liu and Timmermann, 2013), and report their efficiency, as compared to other methodologies, as well. They deploy PCA, reporting that its last eigenvectors are strongly connected to the existence of stationary, co-integrated and predictable (under the predictability definition of Box and Tiao, 1977), mean-reverting pairs-trading portfolios in the stock markets. They also report that these embedded portfolios are easily modeled and predicted as random walk with drift processes, and as such models like the Auto-regressive (AR(1)) ones, could easily forecast the portfolios returns' time series dynamics over time. They also refer and apply the mean-reverting trading strategies developed and proposed by Jurek and Yang in their "Dynamic Portfolio Selection in Arbitrage" paper (2011),Jurek and Yang, 2011, suggesting that they manage to score high Sharpe Ratios, and significantly reduce the overall portfolio risk for the investor. In the current context, we extend their analysis, by applying LLE additionally to the PCA methodology, testing for potential existence other hidden patterns in the pairs' dynamics and the respective benchmark portfolios, as well, easily exploitable by the several trading strategies developed in the paper's framework. We believe that the relative value between two stocks traded under the "Pair Trading" rationale of Cuturi et.al. is also easily extended intuitively in the FX Pairs, as well, since they already formulate a relative-value context themselves, as each pair acts as a financial vehicle incorporating two currencies' values in a single format (a ratio). We furthermore test the efficiency of the application of the AR(1) models on the generated embedded portfolios' trading framework outlined in the following sections, as well as the Gaussian Process Regressors Rasmussen and Williams, 2018. We validate Cuturi et. al's findings on the proper dynamics' modelling of the sparsity dominant components of the PCA embedding as a strongly persistent mean-reverting process, while digging deeper on this behavior by calculating other metrics of the underlying forex pairs' time series related to it, like the Hurst Exponent and the classical ADF and KPSS stationarity criteria. We also identify the expected theoretical equivalence of the AR(1) model and the Ornstein-Uhlenbeck (OU) processes from stochastic calculus (Borodin et al., 1996), and more specifically with their long-memory setting related natural extensions, namely the fractional Ornstein-Uhlenbeck (fOU) processes (Høg et al., 2006), and their respective eminent mean-reverting (or trending) dynamics. The AR(1) or ARFIMA (1,$H-\frac{1}{2}$,0) (with H denoting the Hurst exponent and $H-1/2$ the degree of "persistence") in the fractional setting, being the discrete analogues of the (continuous-time expressed) OU and fOU, respectively, are also extended to the more modern Gaussian Process Regression (GPR) framework of the Machine Learning and Bayesian inference context. More specifically, we display empirical evidence that the theoretical equivalence of the GPRs structured using the RBF kernel or its generalisation - the Matern Kernel (Lilly et al., 2017), with the fOU processes, also holds in this application's context, with the GPRs also being able to identify the "mean-reverting" dynamics of the sparsity targeting components of the PCA embedding. Finally, we extend the forecasting base on the rest of the PCA components, as well, exploring predictable patterns in the entire PCA eigenspace, connecting them with "Trending" (fractional) Ornstein Uhlenbeck processes (Mejia and Zapata Quimbayo, 2018) as inferred by the several calculated Hurst exponents. We discuss the difficulties on modelling the potentially highly stochastic, non-linear and higher-order dependent first - maximum volatility targeting - coordinates of the PCA embedding, while empowering our forecasting toolkit using information from the nonlinear Local Linear Embedding (LLE) technique in order to forecast these hard-to-model components of the PCA. We test the PCA and LLE methodologies, under several rolling window frameworks of different time length each, in order to construct PCA embedded portfolios from a daily data-set of 20 FX pairs obtained by the online investing.com Database (www.investing.com) spanning the time period between 03-01-2000 and 29-10-2020, but also trade them using the LLE embeddings' information set, as well. Towards resolving the problem of sign ambiguity of the loadings we used the numerical algorithm and concept addressed by Bro, Acar, and Kolda, 2008. Furthermore, we use forecasting models like the Exponential Moving Averages, Auto-regressive Models, Gaussian Processes Regression (GPR) and Recurrent Neural Networks (RNNs) to forecast the dynamics of the FX Pairs' as well as the several embedded portfolios' returns across time. We compare the efficiency of each of the PCA and LLE trading frameworks, against the classical Equally Weighted (or "Long Only") as well as the Risk Parity portfolio, defined and formulated in the methodology section, under the Sharpe Ratio performance metric. The efficiency of the proposed approach was accessed based on trading simulations or “backtesting". One of the main concepts of Forex markets' investing is the “carry trade effects", which has been thoroughly studied in several works, on both portfolio management, (Clarida, Davis, and Pedersen, 2009, Menkhoff et al., 2012), as well as risk management perspectives (Burnside, 2012, Brunnermeier, Nagel, and Pedersen, 2008, Menkhoff et al., 2012). In our analysis, we firstly focus on incorporating these carry trade effects in our financial modelling and forecasting context of the FX Dataset, in order to yield realistic results. More specifically, in order to construct the framework for the approximations of the “carry-trade" effects, we used the Short Term Interest Rates data as acquired from the online open source database of OECD iLibrary, OECD, 2014; these data depict the rates at which short-term government papers are issued and traded on the market, which in turn are heavily affecting the several financial institutions' borrowings. The data are daily averages of the several countries in the list of the OECD Short Term Interest Rates database and are reported in percentage terms of the three-month money market rates for each country. Using the available options of the countries listed in the dataset, we selected the ones that are additionally members of the G20 countries' list, while we additionally included the Colombian pesso even if it does not belong to the G20 list, investigating potential performance effects as per the paper of Burnside et al., 2011 on carry trade and peso problems connection. The selected countries are United States of America, European Union, Great Britain, Australia, New Zealand, Japan, Canada, Switzerland, Sweeden, Norway, Denmark, South Africa, Russia, Poland, Mexico, China, South Korea, India, Indonesia, Hungary, Colombia. Upon this context, our target dataset consists of the 20 FX pairs corresponding to those countries' currencies' pairs.

3.5.2 Quotation Format

Since a Forex pair corresponding to two countries currencies, like for example the United States Dollar (USD) and the Euro (EUR) can be quoted in two potential ways, i.e. as "USD/EUR" or as "EUR/USD", we need to clarify and select one of them for our final dataset construction. We select all the pairs to have each corresponding country's currency as its base currency, while the Quote currency being the Dollar (USD) for all of them. This choice was based on the following intuition. Investments in Equity Indices or simple stocks like e.g. the Apple stock on NASDAQ stock exchange or the Deutsche Bank Stock in the Xetra Stock Exchange, are also inherently connected to the foreign exchange markets, since each investment is denominated in the corresponding exchange's country's currency. More specifically, we need US Dollars to buy Apple stocks and Euros to buy Deutsche Bank stocks, so we could easily use the quoting format of the Forex pairs to also quote those assets in terms of their respective currency. Denoting as **AAPL** the Apple stocks and **DB** the Deutsche Bank stocks in our example, we could quote their closing value as of 29/10/2020 (downloaded again from investing.com database) as for example:

1. Close price of **AAPL** = 115.32 USD, i.e. we need 115.32 US Dollars to buy 1 unit of Apple's equity, i.e. measured in number of shares (1 stock). As such, following the forex quotation format, we could display the value as **AAPL/USD**

2. **DB** = 7.884 Euros i.e. we need 7.884 Euros to buy 1 unit of Deutsche Bank's equity measured in number of shares (1 stock). Again, we could display the quotation as **DB/EUR**.

Since we are particularly interested in modelling the dynamics of the several currencies' values across time, we need a common reference currency to measure their "strength" or "weakness". Thus, we quoted all of our Forex pairs' in USD terms, and as such, all forex pairs could now be considered as investments in "artifacts-US stocks". As a straight forward example, the "EUR/USD" quote of e.g. 1,1674 as of 29/10/2020 means that we need to make an investment of 1,1674 USD to buy 1 unit of Euro currency at that time, and the same logic can be applied to all other currencies pairs as well. While, intuitively, this value measuring concept of the Forex pairs, seems to be almost identical with the one of the common stocks investing examples above, it is still lacking some theoretical validity, on what is the main value incentive behind each asset class investment decision. Equities investments provide dividends, which in the real world are positively valued (otherwise the stock's price should be negative) and as such they comprise the main driver of the investors' incentives, towards the more classical "Buy and Hold" portfolios holding (Fama, 1970, Fama and French, 1993, Fama, 1998). The Forex pair investing analogue on this mechanism is the so called "Long carry-trade", which is studied in many research papers (Brunnermeier, Nagel, and Pedersen, 2008, Menkhoff et al., 2012) and which is based on the interest rate differential between two countries' currencies (Choie, 1993). We will thoroughly investigate and define the "carry-trade" framework in the following section. Following this rationale, we formalised the target Forex pairs' quotation as follows. We used the "investpy 0.9.14", Canto, n.d. module of python to download the USD based FX pairs daily spot closing prices, from the "www.investing.com" open API, spanning the period from 19/09/2001 till 29/10/2020, totalling to a 5002 trading days set. These 20 FX pairs are listed below:
USD Quoted FX pairs =
1. EUR/USD, 2. GBP/USD, 3. AUD/USD, 4. NZD/USD, 5. JPY/USD, 6. CAD/USD, 7. CHF/USD, 8. SEK/USD, 9. NOK/USD, 10. DKK/USD, 11. ZAR/USD, 12. RUB/USD, 13. PLN/USD, 14. MXN/USD, 15. CNY/USD, 16. KRW/USD, 17. INR/USD, 18. IDR/USD, 19. HUF/USD, 20. COP/USD

The time period was particularly selected to contain crucial market crashes like the 2008 Lehman related stock market crash, 2010 sovereign debt crisis in the Eurozone and even the most recent Covid-19 related market crash of 2020. For our analysis we computed the compounded daily returns of the FX Pairs as:

$$r_{i,t} = log(\frac{P_{i,t}}{P_{i,t-1}}) = dlog(P_{i,t}), \tag{3.16}$$

where $P_{i,t}$, is the spot price of the i-th FX pair ($i = 1, 2, \ldots 20$) in the data-set above, and $r_{i,t}$ is each corresponding compounded return at date t. These returns though, still lack the necessary adjustments for the so called "carry" effects. More specifically, as an example, we consider the case of an investor who buys a specific amount of Euros against USD, at the end of day t and for a short investment horizon of 1 trading day. By making this trade, the investor should be aware of two main factors that determine the net profit or loss of the trade at the end of day $t + 1$. The first one is the market price dynamics, i.e. the anticipation that the EURUSD's market price level will rise during the upcoming trading day $t + 1$, or not. Secondly, the investor also has to consider the foregone interest that his USD dollars would generate on his USD bank account, relative to the one that will be gained or lost on his euros, in a corresponding Euro account, for a period of 1 trading day. The net Profit or Loss on the transaction, will have to account for both effects in order to be precise and realistic. The daily price returns defined in equation 3.16 only account for the first effect, i.e. the market price dynamics of each of the FX Pairs in our Dataset. In order to approximate the second factor, i.e. the one related to the interest differentials of the investments in each of the currencies which comprise each FX Pair, we use the short term interest rates data retrieved from the OECD database. Regarding our EURUSD example, we consider the approximation of the interest that would be gained on the Euro account of the investor, as the European Union's short term interest rate data in the OECD dataset and the "foregone" interest in the corresponding USD account, as the United States of America's short term interest rate data in the OECD dataset. Since the short term interest rates data are reported on a monthly basis, we need to construct daily approximations of them, so as to make them comparable to the market price daily returns defined in Equation 3.16. Towards that aim, we matched those market price daily returns' time series, with the OECD dataset's monthly short term interest rates' records, on dates. This step created null values for the trading days where no short term interest rates records were available. To fill these null values, we linearly interpolated through the available records, using the linear interpolation methodology of the "Pandas" module of Python, McKinney, 2010. This process created a daily dataset of equal length to that of the time series of the FX Pairs' market price returns of equation 3.16. Furthermore, since the OECD's short term interest rates' records are reported on an annual percentage format, we divided their values by 365 and 100 respectively, in order to reformat them on a daily basis. We denote the time series of the USA short term interest rate's daily approximations, formulated in the previous steps, as $IR_{USA,t}$. We denote the corresponding time series for the remaining 20 countries of the OECD countries' set, as $IR_{i,t}$, with $i = 1, 2, \ldots, 20$ denoting each individual country. Under this notation and since all of our FX Pairs are quoted against the US Dollar, we calculate the Interest Rates Differentials's time series, between each of the OECD countries' short term interest rates daily approximations $IR_{i,t}$ and the $IR_{USA,t}$. More specifically, we write

$$IRD_{i,t} = IR_{i,t} - IR_{USA,t}, \tag{3.17}$$

with $IRD_{i,t}$ denoting the interest rate differential between country i and the USA at date t, with $i = 1, 2, \ldots, 20$.

Under this notation, we compute the "carry" adjusted FX pairs' returns, as:

$$x_{i,t} = r_{i,t} + IRD_{i,t}, \tag{3.18}$$

where $x_{i,t}$ is the i-th FX Pair's "carry" adjusted return, with $i = 1, 2, \ldots, 20$, corresponding to its raw "unadjusted" market price return, $r_{i,t}$ defined in Eq. (3.16), and $IRD_{i,t}$ is the daily interest rates differential between the short term interest rate daily time series of the i-th pair's base currency's country (e.g. the Eurozone for the case of EURUSD FX pair) and the USA's one, as formulated in Eq.(3.17). We show the $x_{i,t}$ time series in Figure 3.26.

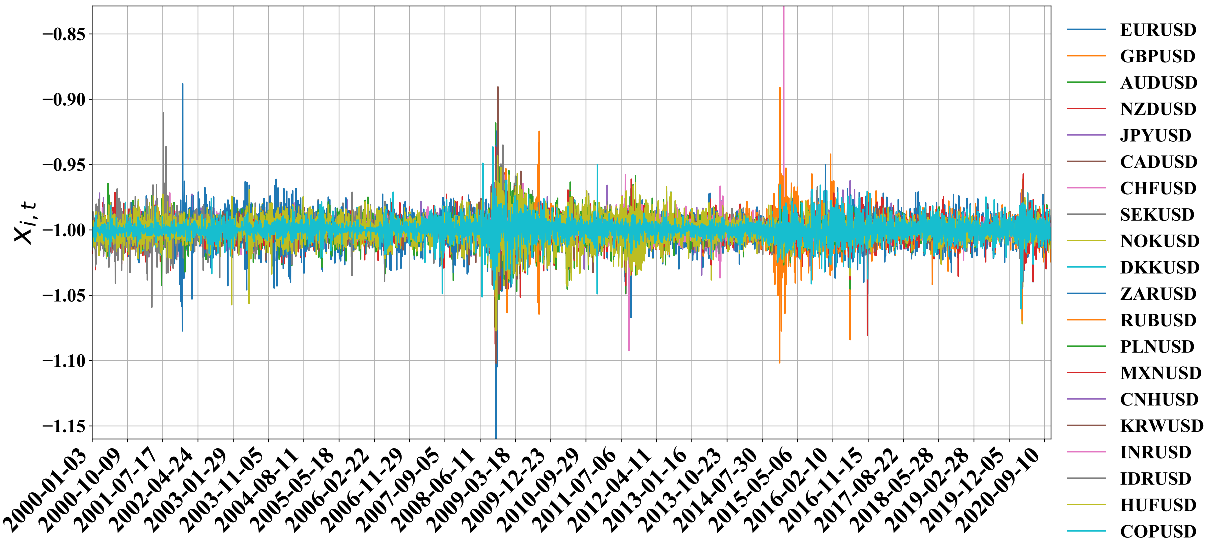

FIGURE 3.26: Carry Adjusted FX Returns time series, $x_{i,t}$ for the $i = 1, 2, \ldots, 20$ FX Pairs in our Dataset.

3.5.3 Cumulative and Portfolios Log-Returns

Since we are using log-returns, as defined in equation 3.16, we need to define under which context we formulate the several constructed portfolios' returns, as well as their cumulative performance, in respect to their constituents (the FX Pairs returns) over time. As thoroughly investigated in Miskolczi, 2017, differences between simple and log-returns in financial data, seem to be present and significant under some specific conditions. More specifically, if we consider the case of a single asset's growth over time, i.e. the i-th FX Pair's log-returns in our dataset, the multi-period cumulative return or "growth" of it for a period of e.g. 3 days, would be $r_{i,t=3} + r_{i,t=2} + r_{i,t=1} = (log(P_3) - log(P_2)) + (log(P_2) - log(P_1)) + (log(P_1) - log(P_0)) = log(P_3) - log(P_0) = log(\frac{P_3}{P_0})$. As such, in order to get the true asset's price appreciation or depreciation across time, we need to remove the natural logarithm from the respective expression. As such we would get back the intuitively expected growth level of $\frac{P_3}{P_0}$. All of the single-asset or single-portfolio related cumulative returns calculations performed in This work's framework, will also apply this logarithm filter. The cumulative returns of $x_{i,t}$ log-returns time series are shown in Figure 3.26. For the calculation of the several constructed portfolios' returns, we consider the logic behind the equivalence of the log-returns and the simple ones, i.e. we assume that the conditions in the framework thoroughly described and analysed in Miskolczi, 2017, are adequately satisfied. More specifically, following this rationale, we consider the “carry" adjusted returns $x_{i,t}$. They are in fact formatted using the log-returns $r_{i,t}$ and the small carry-adjustments related to the interest rate differentials between the several $i = 1, 2, \ldots, 20$ FX pairs in our dataset (as defined in equation 3.18). We construct an arbitrary portfolio consisting of those pairs, and we denote its logarithmic returns at time t, as R_t. We also denote the corresponding portfolio weights for each of the pairs as $\tilde{u}_{i,t}$. Thus:

$$R_t = ln(\sum_{i=1}^{20} \tilde{u}_{i,t-1} e^{x_{i,t}}) \tag{3.19}$$

As formulated in equation 3.19, the portfolio returns are not written in the intuitively expected format of the sum of the weighted returns $x_{i,t}$, for the $i = 1, 2, \ldots, 20$ FX pairs in the dataset. However, as suggested in Miskolczi, 2017, if we additionally make the assumption that the log-returns under consideration are close to zero, then by using the approximation $ln(1 + x) \approx x$ and the definition of the exponential function, we could write the approximated aggregated portfolio returns, as:

$$R_t \approx \sum_{i=1}^{20} \tilde{u}_{i,t-1} x_{i,t}. \tag{3.20}$$

We assume that the approximation error introduced in Eq.(3.20) is small enough so that it can be neglected from our analysis, and as such we choose to formulate all of the constructed portfolios under the approximation rule in consideration. Towards the justification of this choice, we will also display some demonstrative empirical results between the two different approaches of portfolio log-returns’ calculations, for the case of the Long Only portfolio, in the relevant section 3.5.5.

3.5.4 Rolling Window Configuration

For our illustrations, we compared each of the constructed portfolio, using the rolling window of 25, 100, 150 and 250 trading days configurations, corresponding to 1 month, 4 months (a quarter), 6 months and 1 year of active trading days, with a step of 1 trading day for each one of them. As such, we denote the length of time instances at each time window configuration as N, where $N : N_t = 25,100,150$ or 250, respectively.

3.5.5 Long Only Portfolio

At this section we define the classical Equally Weighted or “Long Only" (LO) (Clarke, De Silva, and Thorley, 2006, Clarke, De Silva, and Thorley, 2011, DeMiguel, Garlappi, and Uppal, 2009) and Equally Risk or “Risk Parity" Portfolios (RP) (Braga, 2015, Chaves et al., 2012, Qian, 2011). More specifically, the Long Only portfolio is constructed by investing the same $1/D$ trading weight on each of the $D = 20$ assets in our FX Dataset, and as such we write its returns’ time series at time $t + 1$ as:

$$y_{LO,t+1} = \frac{1}{D} \sum_{i=1}^{D} x_{i,t+1}. \tag{3.21}$$

TABLE 3.18: Statistics over the approximation rule application

Description	Value
Time Period spanned in Years = 20.81	≈ 21
Average proper portfolio Log-Returns	≈ 0.0032 %
Average of the Approximated Returns	≈ 0.0016 %
Average Annual Log-Returns	≈ 5.7289 %
Average Annual Approximated Returns	≈ 5.2555 %
Average Log-Returns vs Approximated Returns Difference	≈ 0.0015 %
Total Log vs Approximated Returns Difference	≈ 8.6241 %
Average Annual Log vs Approximated Returns Difference : (8.6241 % / 21 Years)	≈ 0.4141 %

where, $x_{i,t+1}$ is the i-th FX Pair's return at $t+1$, for the $i = 1, 2, \ldots, D$ and $D = 20$ FX Pairs in our Dataset. As mentioned and analysed above, the Eq.(3.21), which defines the overall portfolio's returns makes the basic assumption that the simple returns under consideration are close to zero, and as such the approximation rule described in the same framework could be applied. In fact, the proper portfolio equation, would be defined as:

$$\tilde{y}_{LO,t+1} = ln(\sum_{i=1}^{D} \frac{1}{D} \exp x_{i,t+1}). \tag{3.22}$$

Towards the quantification of the potential divergence between the two approaches, and in order to get an even better intuition of the relative impact of this approximation rule onto the portfolio returns under consideration, some key statistics are reported in Table 3.18.

The approximated returns seem to be “under"-estimating the performance of the assets and portfolios under consideration, something that can be identified in Sharpe Ratio terms, as well. As such, the results obtained by the methodologies outlined in the following sections could be considered to operate on a “conservative" basis. As such, in order to simplify the overall calculations in the frameworks analysed below, and since the target of this workis to compare the several portfolio construction mechanisms' with each other on the same dataset and under several different time window configurations, we applied this approximation rule to construct the several proposed portfolios. However, we greatly encourage further research to potentially explore the corresponding impact of this approximation mechanism on any or all of the constructed portfolios outlined in the paper's following sections, as well. The Long Only portfolio's Sharpe Ratio is 0.15, with its returns having a mean value of 1.07 % (CI 95%: -1.98%, 4.14%) and standard deviation of 6.94.

3.5.6 Risk Parity Portfolio

For the Risk Parity portfolio, we follow the naive risk parity investing approach, thoroughly investigated, analyzed and compared with even more sophisticated risk parity investing rationales in Braga, 2015. Under this type of investment methodology, each asset is allocated with a portfolio weight which is proportional to its inverse risk. As such less volatile assets will be assigned higher trading weights, while smaller portfolio weights will be assigned to more volatile (risky) ones. While there are several ways for quantifying risk, such as volatility, variance, VaR, maximum DrawDown etc, the analysis here is performed using the classical variance for quantifying the risk for both the assets comprising the portfolio and the total portfolio's risk itself.

Here, since the target of the overall analysis is to compare the PCA and LLE embedded portfolio construction mechanisms with the classical approaches, as well as with each other, we follow the naive Risk Parity approach for constructing the corresponding risk parity portfolio. This means that we do not incorporate the correlations of the assets across time into the adjusting process. We follow the logic outlined in the steps outlined below to construct the Risk Parity portfolio :

1. We compute the rolling and expanding window volatilities, for each of the $i = 1, 2, \ldots, 20$ assets in our FX Dataset. We denote these time varying volatility under a rolling window of 250 trading days, as $\sigma_{i,t} = std(x_i, 250)$ (rounding a trading year of 252 business trading days to 250 and also to match the corresponding 250 trading days rolling window configuration of the PCA and LLE embedding methods, respectively).

2. In order to measure the volatility effects on an annual basis, as followed by many academics but also practitioners in the real markets, we annualized the corresponding time series, by multiplying them with the $\sqrt{252}$ (for the 252 (business) trading days in a year) and also by 100 to format them as percentages, thus ending up with the quantity $\hat{\sigma}_{i,t} = 100\sqrt{252}\sigma_{i,t}$, denoting the corresponding annualised volatility of the i-th asset's returns in our FX Dataset (see also Weber, 2017).

The portfolio allocation mechanism described above, practically means investing $1/\sigma_{i,t}$ at each of the $i = 1, 2, \ldots, 20$ FX Pairs in our dataset, with corresponding carry-adjusted returns $x_{i,t}$, at the end of the t trading day. For instance, considering the example of calculating the annualised rolling risk of the i-th FX Pair carry adjusted returns over the last 250 trading days, and getting a 20% value, then this would mean investing 20% of our portfolio balance in this FX Pair, respectively. As such the "profit" or "loss" at the next day $(t + 1)$ (i.e. next trading day's portfolio return) will be:

$$y_{RP,t+1} = \sum_{i=1}^{D} \frac{x_{i,t+1}}{\hat{\sigma}_{i,t}}, \tag{3.23}$$

where $\hat{\sigma}_{i,t}$ is the annualised rolling volatility of the i-th FX pair's returns, under a rolling window of 250 trading days configuration, with $i = 1, 2, \ldots, D$. $x_{i,t+1}$ is the i-th FX Pair's return at time $t + 1$, and $y_{(RP),t+1}$ is the risk parity portfolio's return at time $t + 1$. The aggregated Risk Parity portfolio's Sharpe Ratio is 0.37, with its returns having a mean value of 5.04% (CI, 95% : -1.02%, 12.96%) and standard deviation of 15.87%.

Our target is to investigate whether the PCA and LLE methodologies can construct more predictable trading portfolios' returns' dynamics, for our time series analysis tools to capture, while also comparing to the ones generated by the two classical approaches.

3.5.7 PCA and LLE portfolio trading framework

Under the portfolio construction context described above, several portfolios may be defined. More specifically, as any linear combination of the $D = 20$ assets in our FX Dataset can be considered as a portfolio consisting of these particular assets, then additionally to the classical "Long Only", "Risk Parity", the "global" PCA-Embedded portfolio, $\boldsymbol{Y}_{(PCA)}$, we can also consider each one of the PCA coordinates, as separate individual “sub"-portfolios themselves. As “global" we refer to the construction of portfolios using the first or the last L eigenvectors from the PCA eigen-spectrum, respectively. Furthermore, regarding the LLE case, each embeddings' information, acting as directional indicator of the FX Pairs' and the traditional portfolios' general dynamics, as well, can be considered as a trading strategy itself, targeting trending or mean-reverting trading rationales of the several target traded assets. In this analysis, we test 2 values for the corresponding L parameter, i.e. $L = 3, 5$ corresponding to both the first L and the last L eigenvectors set of each embedding method. This way we test for existence of mean-reverting or trending sub-portfolios in each eigenvectors' subset, respectively. All of those portfolios can now be analysed under the Markowitz framework, Markowitz, 1952, and we can perform a comparison analysis based on the Sharpe Ratio terms (Sharpe, 1994). Furthermore, they can be used for forecasting purposes using e.g. Exponential Moving Averages (EMA), AR models, Gaussian Processes Regression (GPR) and Recurrent Neural Networks (RNN), towards potential high level Sharpe Ratio generation (Smith, Elton, and Gruber, 1982, Guercio and Reuter, 2014, Bennett et al., 2016, Drew, Veeraraghavan, and Wilson, 2005).

We construct the PCA-embedded portfolios in a rolling time window approach, each of which spanning N time points. The rolling window approach targets on capturing more short-term or “local" market dynamics, analysing at each time window a subset of neighboring time points. We wish to test the assumption that this short-term captured statistics, will be sufficient enough to provide a consistently predictable behavior of the FX Pairs' returns, in the long-run, as well.

In sections 2.3.2 and 2.3.2 which describe the PCA and LLE frameworks, each method's projection algorithm was formulated. Each of them projected the carry adjusted returns X, from the input high dimensional space with dimension $D \times N$ (with N corresponding to each of the varying time window configurations defined in section 3.5.4), to a lower dimensional manifold consisting of L coordinates, with dimension $L \times N$. The calculated eigenvectors v_j for the PCA case act as the trading weights applied in the FX assets comprising the respective embedded portfolios and the q_j of the respective LLE one, act as directional dynamics indicators, with $j = 1, 2, \ldots L$, respectively.

In our analysis, we aim at simulating the necessary portfolio construction calculations, as well as trading actions, by a trader or portfolio manager, who utilizes the developed trading strategies to invest in real markets, under a constantly time evolving market environment. Towards this, we follow an "out-of-sample" approach, which steps are listed below :

1. We split the entire data set of 5437 trading days, into separate time windows, each of which contain N time points, where N is defined as per the relevant section 3.5.4 of time windows configuration setting.

2. We perform the PCA and LLE projection methodologies, at each of these time windows, and store the resulted PCA eigenvectors $\boldsymbol{v}_j \in R^D$ and LLE eigenvectors $\boldsymbol{q}_j \in R^D$, corresponding to the j-th projection's coordinate of each method, with $j = 1, 2, \ldots, L$ at each time step t.

3. Having constructed the time series of the PCA eigenvectors for the entire set of time windows and for each of the rolling and expanding window configurations, we write the PCA embedded portfolios returns at time $t + 1$ for each case, as:

$$y_{PCA,j,t+1} = \boldsymbol{v}_{j,t}^T \boldsymbol{x}_{t+1} \tag{3.24}$$

where $\boldsymbol{x}_{t+1} \in R^D$ is the vector containing the FX Pairs' returns at time $t+1$ and under each time window configuration defined in the relevant section 3.5.4, $y_{PCA,j,t+1} \in R^L$, $L < D$, is the j-th embedded portfolios returns at time $t+1$, using the j-th PCA eigenvector's values, $\boldsymbol{v}_{j,t}$, as the trading weights applied on each asset at the end of day t.

4. We also write the Cumulative returns of the PCA embedded portfolios, after adjusting for the exponential filtering, denoting them as $\Sigma y_{PCA,j,t}$, with $j = 1, 2, \ldots, L$, as :

$$\Sigma y_{PCA,j,t} = exp(\sum_{t=1}^{t} y_{PCA,j,t}) - 1 \tag{3.25}$$

5. We denote the "global" PCA-embedded portfolios' returns time series, corresponding to the embedded portfolios defined in the previous step, as:

$$Y_{PCA,t+1} = \sum_{j=1}^{L} \boldsymbol{v}_{j,t}^T \boldsymbol{x}_{t+1} \tag{3.26}$$

with the corresponding Cumulative Returns being:

$$\Sigma Y_{PCA,t} = exp(\sum_{t=1}^{t} Y_{PCA,t}) - 1 \tag{3.27}$$

Note : All embedded portfolios returns and corresponding cumulative returns in the following sections' Figures, are displayed in percentage (%) format, meaning they are all multiplied by 100 for better visualisation purposes.

6. As a final step we enhance the target trading strategies' returns, by applying time series forecasting into the PCA coordinates' based embedded portfolios $y_{PCA,j,t}$, as well as the corresponding global one, $Y_{PCA,t}$, while also applying the same models into the respective generated LLE eigenvectors' dynamics over time. The prediction models are the EMAs, ARs, GPRs and RNNs, targeting both linear and nonlinear time series prediction rationales. We use the LLE embeddings' predictions, as trading signals which are then applied on the benchmark (general trend) related portfolios' returns time series, and the PCA embedded portfolios, as well.

3.5.8 Markowitz Framework

Having formulated the PCA-embedded portfolios and the LLE trading signals related strategies' returns time series, we are now in the position to analyse them under the Markowitz framework Markowitz, 1952, and compare their efficiency respective to each other, but also relative to the (raw) "Long Only" and "Risk Parity" portfolios. The comparison is performed in Sharpe Ratio terms Sharpe, 1994. Towards that, we use a general notation to denote the set of forecasted time series, i.e. the PCA-embedded portfolios, as well as the LLE eigenvectors acting as trading indicators in our context, and as such all of them are denoted simply as π_t. The respective trading weights' vectors, allocated on each of the $D = 20$ FX Pairs at time t, under each of the trading frameworks defined in each relevant section, are simply denoted as $\boldsymbol{u}_t \in R^D$. As such, with $i = 1, 2, \ldots, 20$ and $j = 1, 2, \ldots, 5$, we write :

$$\pi_{t+1} = \begin{cases} y_{LO,t+1} = \frac{1}{D}\sum_{i=1}^{D} x_{i,t+1} & \text{for the Long Only portfolio} \\ y_{RP,t+1} = \sum_{i=1}^{D} \frac{x_{i,t+1}}{\hat{\sigma}_{i,t}} & \text{for the Risk Parity portfolio} \\ y_{PCA,j,t+1} = \boldsymbol{v}_{j,t}^{T}\boldsymbol{x}_{t+1} & \text{for the PCA coordinate based embedded portfolios} \\ Y_{PCA,t+1} = \sum_{j=1}^{L} \boldsymbol{v}_{j,t}^{T}\boldsymbol{x}_{t+1} & \text{for the global PCA embedded portfolios} \\ \boldsymbol{q}_{j,t+1} & \text{for the LLE eigenvectors time series} \end{cases}$$

3.5.9 Forecasting Models and Trading Strategies

The main rationale behind the application of the several time series forecasting methodologies outlined in the following sections is to manage to create a systematic approach on forecasting the several portfolios returns' dynamics over time. As per Cuturi and D'Aspremont, 2016, we define "predictability" for the portfolio π_t defined in the previous section (Box and Tiao, 1977), as follows :

$$\pi_{t+1} = \hat{\pi}_t + \epsilon_{t+1} \tag{3.28}$$

where $\hat{\pi}_t$ is the predicted value of the portfolio's return at time $t+1$ using the previous steps' portfolio returns up until time t, while ϵ_t denotes a zero-mean Gaussian noise with covariance Σ. We focus on the univariate case of the predictability definition, since we are interested in forecasting each portfolio individually. As such, denoting the variance of π_t as σ_π and of $\hat{\pi}$, as $\hat{\sigma}_\pi$, we follow Box and Tiao's definition of portfolio π_t's dynamics "predictability", by the ratio

$$\lambda_\pi = \frac{\sigma_\pi}{\hat{\sigma}_\pi} \tag{3.29}$$

and intuitively state that if the ratio is small, then the variance of the noise dominates the one of the predictor, so in general noise will dominate the dynamics of π_t; while in the case where the ratio is large, the predictor's value $\hat{\pi}_t$ dominates the noise term and the predictions for π_t will be accurate on average. In the trading application of the models in this work, this accurate on average prediction of portfolio π_t's will be considered as additional input on the decision making process of the trader/portfolio manager, trading under the corresponding constructed strategies, on whether she should trust or not the several portfolio weights generated by each embedding method, at each time instance in the period we are analysing.

3.5.10 Exponential Moving Averages (EMA)

We define the Exponential Moving Average (EMA) portfolio π_t's predictor value $\hat{\pi}_t$, following the corresponding definition of Equation 3.28, as:

$$\hat{\pi}_t = \frac{EMA_t(1-\tilde{h}) - EMA_t + 1}{\tilde{h}} \tag{3.30}$$

where t=time index, $\tilde{h} = 2/(t_d+1)$ a smoothing - scaling parameter, connected to the t_d total number of trading days (points) used for the EMA calculation at each window of calculations.

3.5.11 Auto-regressive Models (AR(p))

The second forecasting tool is the Auto-regressive model with order p, AR(p). We define the corresponding portfolio π_t's predictor value as :

$$\hat{\pi}_t = c + \phi_1 \pi_t + ... + \phi_p \pi_{t-p}, \tag{3.31}$$

where $\phi_1, \phi_2, \dots, \phi_p$ are the target fitting parameters in the AR model.

3.5.12 Gaussian Process Regression (GPRs)

As a third forecasting method of the dynamics of the several embedded portfolios' returns over time, we apply the already discussed and thoroughly investigated Gaussian Processes Regressors (GPR) model. As per the theoretical framework of the Gaussian process modelling in section 2.1.7, a "prior" belief about the "shape" or "behavior" of the function expressing the underlying dynamics of the portfolios' returns, should be determined at the very start of the modelling process. In our time series forecasting framework, this means that we need to select a particular covariance function $k(\pi_t, \pi_{t'})$, connecting the modelled portfolio's returns time series at time t, π_t, with any previous steps' returns, $\pi_{t'}$, with $t' \neq t$, while also trying to depend the selected covariance structure, with a set of "fine tuning" hyperparameters $\boldsymbol{\theta}$, which should be able to handle any potential "a-priori" knowledge, "best guess" or "belief" about the underlying process's dynamics. In our specific FOREX trading application, this "belief" is actually related to our pattern recognition goal of extracting the potential "Mean-Reverting" or "Trending" behavior of the constructed embedded portfolios' returns, extending the work of Cuturi and D'Aspremont, 2016 as per all the relevant analysis in the previous sections. More intuitively set, what we "know" is that the underlying portfolios' dynamics will either be "Mean-Reverting" or "something else", meaning either "Trending" or experiencing a "hybrid" behavior between the two.

Kernel Selection

Under this logic, we selected our covariance function to be represented by the Matern kernel, with its expression tailored to our portfolio trading application, as follows :

$$k(\pi_t, \pi_\tau) = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\frac{\sqrt{2\nu}}{l} \|\pi_t - \pi_\tau\|_{L_2} \right)^{\nu} K_\nu \left(\frac{\sqrt{2\nu}}{l} \|\pi_t - \pi_\tau\|_{L_2} \right) \tag{3.32}$$

with l being the kernel's specific "length-scale", K_ν is a modified Bessel function Abramowitz, Stegun, and Miller, 1965 and ν the main parameter of interest regarding the underlying process' dynamics categorization, specifically designed to control its "smoothness" or "roughness", as formulated and investigated in the rest of the section. The term $\|\pi_t - \pi_\tau\|_{L_2} = d(\pi_t, \pi_\tau) = d$ declares the euclidean distance.

The Matern kernel, thoroughly outlined and documented in the "Covariance Functions" section of Rasmussen, 2003 (p. 84) offers an intuitive parameterisation for our particular identification and categorization framework of target portfolios returns' dynamic's, since its structure encapsulates such "behaviors" via the "fine tuning" and specific level setting of the parameter ν. Depending on the ν's value, we are allowed to connect the modelled time series to an interesting set of functions and stochastic processes, for which a wide theoretical background is available and more mathematically rigorous insight on the modelled dynamics' properties can be retrieved. More specifically, as per the relevant section 4.2 (Rasmussen, 2003), by following Abramowitz, Stegun, and Miller, 1965 and setting $\nu = p + \frac{1}{2}$, $p \in N^+$ the kernel can be written as :

$$k_{p+\frac{1}{2}}(d) = exp\left(\frac{\sqrt{2p+1}d}{l}\right)\frac{p!}{2p!}\sum_{i=0}^{p}\frac{(p+i)!}{i!(p-i)!}\left(\frac{2\sqrt{2p+1}d}{l}\right)^{p-i} \tag{3.33}$$

Using the above formula, several expressions of interest are obtained for specific values of the ν parameter:

1. $\underline{\nu = \frac{1}{2} \quad (p=0)}: k_{\frac{1}{2}}(\pi_t, \pi_\tau) = exp\left(\frac{\|\pi_t - \pi_\tau\|_{L_2}}{l}\right)$, which is the expression of the "absolute exponential" kernel
2. $\underline{\nu = \frac{3}{2} \quad (p=1)}: k_{\frac{3}{2}}(\pi_t, \pi_\tau) = \left(1 + \frac{\sqrt{3}\|\pi_t - \pi_\tau\|_{L_2}}{l}\right)exp\left(-\frac{\sqrt{3}\|\pi_t - \pi_\tau\|_{L_2}}{l}\right)$
3. $\underline{\nu = \frac{5}{2} \quad (p=2)}:$

   $$k_{\frac{5}{2}}(\pi_t, \pi_\tau) = \left(1 + \frac{\sqrt{5}\|\pi_t - \pi_\tau\|_{L_2}}{l} + \frac{5\|\pi_t - \pi_\tau\|_{L_2}^2}{3l^2}\right)exp\left(-\frac{\sqrt{5}\|\pi_t - \pi_\tau\|_{L_2}}{l}\right)$$
4. $\underline{\nu \to \infty}: k_{\frac{1}{2}}(\pi_t, \pi_\tau) = exp\left(\frac{\|\pi_t - \pi_\tau\|_{L_2}^2}{2l^2}\right)$, i.e. the covariance converges on the "squared exponential" kernel

In the current FOREX trading approach, we set the $\nu = \frac{1}{2}$ in the Matern kernel, since as per Lilly et al., 2017, Appendix E :

$$\nu = \alpha - \frac{1}{2} \tag{3.34}$$

The selection is based on the fact that the yielded "absolute exponential" kernel is documented Rasmussen, 2003 to be closely connected to the Ornstein-Uhlenbeck stochastic process (OU), heavily investigated and used as the main tool of several financial pricing and trading applications in the financial engineering academics and practitioners community, over the years. More specifically, the work of Lilly et al., 2017, demonstrates in a really detailed and mathematically rigorous framework, the connection of the Matern process to the Fractional Brownian Motion, being the broader class of stochastic processes with "memory", member of which is the (fractional) Ornstein-Uhlenbeck (Høg et al., 2006), as well. The next part of this section is dedicated on the mathematics behind this particular connection, inferred by the two processes' corresponding autocovariance functions' forms and respective behavior under specific conditions of the internal parameters controlling "smoothness" or "roughness" of the underlying modelled process.

Relationship between Matern Processes & Fractional Brownian Motion

We start with the general (second order) covariance function form of a random process x_t :

$$k(t,\tau) \equiv E(x_t x_{t+\tau}) \tag{3.35}$$

which can also be expressed in term of the inverse Fourier relationship :

$$k(\tau) = \frac{1}{2x}\int_{-\infty}^{\infty} e^{i\omega\tau} S_x(\omega) d\omega \tag{3.36}$$

where $S_x(\omega)$ is the so called spectrum, containing all available statistical information of the process. Based on this expression, the stochastic process x_t can be classified as "diffusive", "sub-diffusive" or "super-diffusive" and its connection with the "memory" property is formulated in terms of long-term decay effects, as follows :

$$k(\tau) \sim |\tau|^{-\mu}, \quad 0 < \mu \leq 1, \quad |\tau| \to 1 \tag{3.37}$$

If this decay rule is valid, then the process is said to possess "long-range dependence" of being classified as a "long-memory process", Beran and Terrin, 1996; Gneiting, Kleiber, and Schlather, 2010, while in case the autocovariance fades faster than the $|1/\tau|$ term, the process is classified as a "short memory" one.

Based on a specific spectrum's form, a first intuitive relationship between the fbm and Matern processes is apparent. More specifically, the form

$$S_x(\omega) = A^2 \tag{3.38}$$

corresponds to a White Noise, the

$$S_x(\omega) = \frac{A^2}{|\omega|^{2\alpha}} \tag{3.39}$$

corresponds to the Fractional Brownian Motion, with α denoting the "slope parameter" Mandelbrot and Wallis, 1968, being conventionally related to the Hurst exponent (as already revisited previously) as :

$$H = \alpha - \frac{1}{2}, \quad 0 < H < 1 \tag{3.40}$$

while the

$$S_x(\omega) = \frac{A^2}{(\omega^2 + \lambda^2)^{\alpha}} \tag{3.41}$$

to that of Matern processes Guttorp and Gneiting, 2006, which as shown below is the so called "damped" version of the Fractional Brownian Motion one, with the parameter λ controlling the corresponding damping timescale of the underlying oscillations.

The autocovariance function of the Fractional Brownian Motion is given by Mandelbrot and Wallis, 1968 and Barton and Poor, 1992:

$$k_{fbm}(t,\tau) = \frac{V_\alpha}{2} A^2 [|t+\tau|^{2\alpha-1} + |t|^{2\alpha-1} - |\tau|^{2\alpha-1}] \tag{3.42}$$

where :

$$V_\alpha = \frac{\Gamma(2-2\alpha) sin(\pi\alpha)}{\pi(\alpha - 1/2)} \tag{3.43}$$

which using the Barton and Poor, 1992 more symmetric formality is expressed as :

$$V_\alpha = \frac{1}{\pi} \frac{\Gamma(\alpha - \frac{1}{2})\Gamma(\frac{3}{2} - \alpha)}{\Gamma(2\alpha)} \tag{3.44}$$

is defined as the variance at time $t = 1$ under a unity set amplitude parameter A. $\Gamma(x)$ is the gamma function - more on the role of the Γ function's properties on the evolution dynamics of the process is extensively discussed in section 3.2 of Lilly et al., 2017.

The variance of the fbm is then given :

$$\sigma^2_{fbm} = E(|z(t)|^2) = k_{fbm}(t, 0) = V_\alpha A^2 |t|^{2\alpha-1} \tag{3.45}$$

which shows an unbounded increase with time, i.e. inferring the non-stationary nature of the process.

We can see that for really small time shifts $|\tau|$ compared to global time t, i.e. $|\tau|/t << 1$ 3.42 becomes :

$$k_{fbm}(t, \tau) \approx \sigma^2_{fbm}(t) - \frac{1}{2} V_\alpha A^2 |\tau|^{2\alpha-1} \tag{3.46}$$

with $\sigma^2_{fbm}(t) \equiv k_{fbm}(t, 0) = V_\alpha A^2 |\tau|^{2\alpha-1}$.

As displayed n the several spectra forms above, the Matern process additionally introduces another parameter $\lambda > 0$, which controls the extra mentioned "damping" effect. With the formulation of the Matern process under the 3.41 equation, and with $\alpha > \frac{1}{2}$, we display below its two asymptotic behaviors (Wolpert and Taqqu, 2005; Eab and Lim, 2006 towards Matern and OU process connections) :

1. $|\omega| >> \lambda \rightarrow S_x(\omega) \approx \frac{A^2}{|\omega|^{2\alpha}}$, meaning that higher frequencies, yield a power-law decay (fbm) dynamics

2. $|\omega| << \lambda \rightarrow S_x(\omega) \approx \frac{A^2}{|\lambda|^{2\alpha}}$, meaning that for lower frequencies, a constant is approached, which if $\frac{|\omega|}{\lambda} << 1$ is said to generate "locally white" dynamics

According to these limits, we can see the Matern process as the (controled by the parameter λ) "continuum" between the two dynamics. More specifically, again as per Lilly et al., 2017 the Matern spectrum can be rewritten in terms of the process's variance :

$$S_M(\omega) = \frac{\lambda^{2\alpha-1}}{c_\alpha} \frac{\sigma^2}{(\omega^2 + \lambda^2)^\alpha} \tag{3.47}$$

with σ_M being given by :

$$\sigma^2_M = c_\alpha \frac{A^2}{\lambda^{2\alpha-1}} \tag{3.48}$$

with the introduced normalised constant being given by :

$$c_\alpha \equiv \frac{1}{2\pi} B(\frac{1}{2}, \alpha - \frac{1}{2}) = \frac{1}{2\pi} \frac{\Gamma(\frac{1}{2})\Gamma(\alpha - \frac{1}{2})}{\Gamma(\alpha)} \tag{3.49}$$

and where $B(x, y) \equiv \Gamma(x)\Gamma(y)/\Gamma(x + y)$ is the beta function. The autocovariance of this spectrum formulation, is given by :

$$k_M(\tau) = \frac{2\sigma_M^2}{\Gamma(\alpha - \frac{1}{2})2^{(\alpha-\frac{1}{2})}} |\lambda\tau|^{a-\frac{1}{2}} K_{\alpha-\frac{1}{2}}(\lambda|\tau|) \tag{3.50}$$

with $K_\nu(x)$ being the modified Bessel function of second order ν.
Observing the behavior of 3.50, as $|\tau| >> \frac{1}{\lambda}$, we get :

$$k_M(\tau) \approx \frac{\sqrt{2\pi}\sigma_M^2}{\Gamma(\alpha - \frac{1}{2})2^{\alpha-\frac{1}{2}}} |\lambda\tau|^{\alpha-1} e^{-\lambda|\tau|} \tag{3.51}$$

i.e the process the covariance function exhibits exponential decay (short-memory process), while for $|\tau| << \frac{1}{\lambda}$ (short-time dynamics) and $1/2 < \alpha < 3/2$ we have :

$$k_M(\tau) \approx \sigma_M^2[1 - (\frac{\lambda\tau}{2})^{2\alpha-1}\frac{\Gamma(3/2-\alpha)}{\Gamma(\alpha+1/2)}] \tag{3.52}$$

which in turn noting that :

$$\frac{c_\alpha}{2^{2\alpha-1}}\frac{\Gamma(3/2-\alpha)}{\Gamma(\alpha+1/2)} = \frac{1}{2}V_\alpha \tag{3.53}$$

is finally expressed as :

$$k_M(\tau) \approx \sigma_M^2 - \frac{1}{2}V_\alpha A^2|\tau|^{2\alpha-1} \tag{3.54}$$

We can see that

$$k_{fbm}(t,\tau) \approx \sigma_M^2 - \frac{1}{2}V_\alpha A^2|\tau|^{2\alpha-1} \tag{3.55}$$

which matches the expression of the fbm process in 3.46, for $|\tau|/t << 1$.

We observe that both expressions are equivalent after focusing on "short time" offsets compared to the "global time" of the whole dynamics. Intuitively, since under the fbm formulation the $x(0) = 0$ is assumed, non-stationarity is inferred, something that is "smoothed out" or "fogrotten" (section 4, Lilly et al., 2017), for nonzero values of λ, i.e. stationary dynamics will emerge, while if instead $\lambda = 0$ it stays on for the entire period of time. Python's module sklearn's Gaussian Process Regressor submodule (Rasmussen and Williams, 2018, Pedregosa et al., 2011b) was used for the several computations.

3.5.13 Random Walk model

As any time series forecasting related paper, we also compare the several forecasting methodologies with the Random Walk model. Since the overall analysis is being performed on the carry-adjusted returns of Equation 3.18, we define a naive Random Walk trading strategy, for both the case of the individual FX Pairs dynamics, but also the several constructed portfolios described in the previous sections. We wish to test the assumption that the FX Pairs themselves or the constructed portfolios might follow a random walk process or not, with or without a drift. If no drift is present, the forecast-ability of the target time-series is of no value at all, since the process is generated by a completely random (white-noise) distribution, while in the case of a drift term's presence, persistent or anti-persistent medium to long term trends might be present in the modelled time series dynamics. This trend or mean-reverting behavior could be the main factor that could be exploited by e.g. AR(1) models with constant terms, towards producing consistent profits over time. Furthermore, if such drift terms are present in the individual FX Pairs' dynamics, they could be explaining fundamental factors emerged by the corresponding countries' currencies or economies. More specifically, they might be connected with monetary policy actions by several Central Banks, like the recent Quantitative Easing policies by the Fed or the ECB since 2008 financial crisis and on-wards, rising prices or inflation shocks, demand for currency by individual or institutional investors, economic growth, as well as export prices. All of these macroeconomic factors' effects, might not be easy to be identified if searched on the individual FX Pairs' dynamics, though. Instead they might be more easily spotted in the several constructed portfolios' dynamics, acting as coarse-grained variables, which are able to incorporate hidden information in the data and emerge it in a more coherent way for the several time series modelling tools to capture in a more effective and robust manner. Following this rationale, we define the Random Walk model trading strategy as :

1. "Buy" ("Long") the portfolio, if the portfolio's return is positive at time t
2. "Sell" ("Short") the portfolio, if the portfolio's return is negative at time t

and as such, the π_t's predictor value $\hat{\pi}_t$ of Equation 3.28, is expressed as :

$$\hat{\pi}_t = sign(\pi_t) \tag{3.56}$$

3.5.14 Numerical Results

We wish to compare the PCA and LLE-embedding trading frameworks relative to the classical portfolio construction approaches, i.e. the Long Only and Risk Parity ones, on Sharpe Ratio terms, not only on their "standalone" context, i.e. with no forecasting model applied on their generated portfolios returns, but also after applying the time series analysis tools on these portfolios' dynamics over time. Towards this aim, we follow the steps described below :

1. We apply the EMA, AR, GPR and RNN models on all of the target forecasted time series π_t defined in the previous steps, using as initial training set a pre-determined number of observations, pre-specified at each model's description in the following sections.
2. We evaluate the several trading strategies' performance results in annualised Sharpe Ratio terms (taking the risk-free rate as zero, Sharpe, 1994) under each model's specifications and tested parameters. More specifically, we calculate them, using the standardised formula, connecting them with the Annualised Mean and Standard deviation quantities of the several portfolios' returns time series, as :

$$sh = \frac{E(\pi_T) * 252}{\sigma_{\pi_T} * \sqrt{252}} = \sqrt{252}\frac{E(\pi_T)}{\sigma_{\pi_T}} \tag{3.57}$$

where $E(\pi_T$ denotes the mean of the portfolio's returns, and the σ_{π_T} their standard deviation. As formulated, the average returns are annualised by multiplying them with the term 252, as there are 252 trading days in a year, and also the standard deviation with the term $\sqrt{252}$ as already explained in the case of Risk Parity, as well. All of the performance calculations in this context will be based on annualised formats, for all respective statistics (mean, std and Sharpe Ratios) of the several portfolios' and assets returns time series. At this point, we should cite Andrew Lo's commentary paper, Lo, 2002, on the potential overestimation of the corresponding Sharpe Ratios in case the several portfolios' returns exhibit serial auto-correlations, but since the main scope of this work is the comparison of the several performance statistics between the several analysed portfolio allocation mechanisms, we move forward with the standardised "traditional" calculation of the Sharpe Ratios, leaving room for further research to test their efficiency with other performance metrics, as well, and comment any potential discrepancies.

3. We apply a one-sample Student's t-test on the trading strategies' returns, towards testing significance of the respective mean value against the zeroed null hypothesis. More specifically, the t-test is defined as :

$$TT = \frac{\tilde{\mu}}{\frac{\tilde{\sigma}}{\sqrt{\tilde{n}}}} \tag{3.58}$$

where $\tilde{\mu}$ is the target population's (the target strategy's returns' mean value), with a $\tilde{n}$ total number of elements and $\tilde{\sigma}$ its standard deviation. We apply this test in both the returns generated by the several applied trading strategies in the target portfolios, but also in the respective returns generated by the naive Random Walk trading strategy applied on the same portfolios, for consistency and confidence comparison purposes.

4. We apply the Welch's t-test (WT) on each of the constructed trading strategies, applied on each of the target portfolios against the respective returns generated by the naive Random Walk trading strategy, towards testing for the respective mean values potential equality. The WT t-test is defined as,

$$WT = \frac{\hat{\mu_1} - \hat{\mu_2}}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \tag{3.59}$$

where μ_1 and μ_2 are the sample means for the two populations, σ_1 and σ_2 are the respective standard deviations and N_1, N_2 are the corresponding sample sizes.

5. We filter all the resulted trading strategies and corresponding portfolios, keeping and displaying the respective performance statistics only for the ones that score a WT test p-value less than 0.05. We additionally display the corresponding naive Random Walk trading strategy's performance statistics towards further reference and comparison.

3.5.15 Exponential Moving Averages (EMA) Trading

In this section, we will use simple Exponential Moving Averages, to trade the several constructed embedded portfolios. Inspired by Jurek and Yang's trading strategy approach in Jurek and Yang, 2011, we implement a "Mean-Reverting" trading strategy using the information from the several Exponential Moving Averages, which are used to estimate the underlying model's mean value, on which each of the target portfolios' returns time series is expected to converge. We target a directional predictability regime, rather than straight returns' levels prediction, meaning we trade the respective portfolios using binary generated signals, as follows :

1. We use four different moving averages, targeting a specific period pattern cyclicality each time. In particular we used Exponential Moving Averages (EMA) with several different values of the time lag parameter t_d, i.e. t_d =2, 3, 5, 10, 15, 25, 50, 100, 150, 200 and 250 trading days, respectively. Using these values, we aim at capturing different market dynamics' and cyclicality patterns of the FX Dataset, in a trading mid-week, week, month, bi-month, quarter, semi-annual, annual etc. level.

2. We apply the EMA model on portfolio π_t as defined in Equation 3.30 in order to get an estimate of the EMA value (EMA(π_t)) which is considered to be the long-term mean, upon which the portfolio's level will be either be "Reverting" or "Following/Trending". More specifically, we define two respective sub-strategies' rationales, the one being consistent with the "Mean-Reverting" portfolio trading of the aforementioned work of Cuturi and D'Aspremont, 2016, but also extending its main concept and also testing for so called "Trending" portfolios in the PCA eigenspace, as well. As such, we define the trading signal $s_{EMA,MR,t}$ of the respective "Mean-Reverting" trading strategy, as :

   EMA "Mean-Reverting" (MR)

   (a) $s_{EMA,t}$ = 1 : "Buy" ("Long"), if the EMA of π_t is trending down (moving averages slope negatively)

   (b) $s_{EMA,t}$ = -1 : "Sell" ("short"), if the EMA of π_t is trending up (moving averages slope positively)

   (c) $s_{EMA,t}$ = 0 : neutral positioning

   and for the "Trending" one, using the intuitively opposite logic, as well :

   EMA "Trending" (TR)

   (a) $s_{EMA,t}$ = 1 : "Buy" ("Long"), if the EMA of π_t is trending up (moving averages slope positively)

   (b) $s_{EMA,t}$ = -1 : "Sell" ("short"), if the EMA of π_t is trending down (moving averages slope negatively)

   (c) $s_{EMA,t}$ = 0 : neutral positioning

In the case of PCA-embedded portfolios, forecasting their returns' dynamics over time, intuitively means that we are answering the question of a practitioner, on whether she/he should "trust" or not, the generated by each method portfolio weights, across time. If the trader trusts the model as it is, i.e. if she believes that the constructed portfolio returns will continue to perform well in the near future, she should trade the trading weights as they are produced by the model, without trying to 'time' the resulting dynamics.

In the case of LLE trading framework, applying the EMA's resulted trading signal into the benchmark and PCA embedded portfolios' returns, intuitively means that the way LLE embeddings' emerging dynamics acts as a leading indicator of the target assets' dynamics over time, and as such yields exploitable by a market participant information. As such, we write the time series of the LLE EMA trading strategy's returns, when applied in the target portfolios' returns, as :

$$y_{LLE,j,t+1} = s_{EMA,t} * \begin{cases} y_{LO,t+1} & \text{for the Long Only portfolio} \\ y_{RP,t+1} & \text{for the Risk Parity portfolio} \\ y_{PCA,j,t+1} & \text{for the PCA coordinate based embedded portfolios} \\ Y_{PCA,t+1} & \text{for the global PCA embedded portfolios} \end{cases}$$

where in this case, $i = 1, 2, \ldots, D$ FX pairs in the dataset, $s_{EMA,t}$ the trading signal generated by the EMA predictor applied on the j-th LLE eigenvector, $q_{j,t}$, with $j = 1, 2, \ldots, 5$, respectively. We display the efficiency of the EMA Trading strategy, when applied in each of the PCA coordinate based and global embedded portfolios, in Table 3.5.19. Since there are numerous traded portfolios vs tested EMA strategies' combinations, only a top performing in Sharpe Ratios' terms subset of them is displayed, ranked and sorted accordingly. The same results are displayed in Table 3.24 for the case of LLE EMA trading signals applied on the individual assets' returns and the two benchmark portfolios, as well. Following Cuturi and D'Aspremont, 2016, our expectation is that that last PCA eigenvectors' based portfolios, will be performing better under the "Mean-Reverting" trading rationale, no such a-priori knowledge or educated guess exists about the expected trading behavior related to using the EMA estimated LLE trading indicators (eigenvectors), respectively. No correlation (or connection) of the several LLE eigenvectors' dynamics and the several target traded portfolios is theoretically documented or practivally tested to the best of our knowledge in a previous work. As such, we also infer this correlation between the several LLE indicators and the target portfolios' dynamics, as either positively leading or negatively leading in the framework of a successful "Mean-Reverting" or "Trending" trading rationale of the respective EMA trading strategy. The success is again measured in terms of Sharpe Ratios of the corresponding strategies. Additionally, we also compare the several trading returns against the ones generated by the naive Random Walk trading strategy, in terms of annualised mean and Standard Deviation levels. Finally, we test the respective values against the zeroed null hypothesis, as well as against each other, using the Welch's t-test, and display the corresponding tests' p-values.

Further to the PCA and LLE trading frameworks, we also apply the EMA Trading Strategies on the Long Only and Risk Parity portfolios themselves, in order to obtain the corresponding performance statistics, towards further comparison and reference. For the case of "Mean-Reverting" EMA trading sub-strategy, we report negative (annualised) Sharpe Ratios for both benchmark portfolios, indicating that their returns' dynamics is not "reverting" to a medium-or-long term mean, sufficiently captured and estimated by the EMA respective predictor. For the case of the "Trending" EMA sub-strategy though, we report an (annualised) Sharpe Ratio of 0.74 for the Long Only portfolio traded with an EMA($t_d = 2$) trading strategy and a 1.12 for the Risk Parity portfolio, when traded using the same model, respectively. The annualised mean returns are 5.11% (CI, 95% : [8.16%, 2.05%) and 17.71% (CI, 95% : 24.69%, 10.74%), with corresponding standard deviations of 6.94% and 15.85%, respectively. This "Trending" underlying trading behavior has again been addressed by Clare et al., 2016, who applied several moving average based Trend following trading models on Risk Parity portfolios comprised by trading assets coming from several traditional asset classes, (Equities, Bonds etc.) and reported significant trading performance results, as well. In the context of the FX trading framework of This work, we comment on this reported trend following rationale, as probably being connected with the USD devaluation over the years, since the USD has been the main funding currency behind all major Foreign Direct Investments (FDI) in e.g. emerging countries of the rest of the world. Applying the risk parity portfolio construction rationale, seems to be providing a "smooth" way of risk identification and as such proper portfolio allocation in each of the respective FX Pairs, yielding an at-least in short-term horizons, predictable portfolio dynamics. However, as we will show in the Transaction Costs Analysis (TCA) section below, 3.5.20, the corresponding strategies' Sharpe Ratios are heavily deteriorated when market frictions (market spread, brokers' commisions, slippage etc) are incorporated in the "net" returns' calculations. As such the strategies are reported to be non-significant in real market conditions as in the "gross" returns' calculation scenario, i.e. when no costs are assumed at all. In the same section, we report that this is not the case for some of the PCA embedded portfolios and some of the LLE indicators' related trading strategies, respectively. For the analysis up to section 3.5.20, we report the "Gross" annualised Sharpe Ratios, i.e. the ones without incorporating transaction costs at all, focusing only on ranking and sorting out the best performing ones, which will then be tested against the several transaction costs' scenarios, respectively.

The results indicate that the expected "Mean-Reverting" behavior of the last PCA component, as per the corresponding findings of Cuturi and D'Aspremont, 2016, are also cross-verified in our analysis, as well. We heuristically define a well-performing strategy, as a trading strategy that produces a Sharpe Ratio of at least 0.5. As another note, we comment on the existence of only one (insignificant) trading performance that manages to score a Sharpe Ratio equal to our heuristically set threshold, which further indicates the inability of the EMA Trend based trading strategies to capture any short, medium or long-term trending portfolios among the PCA embedded ones. However, the overall analysis and characterization of each respective EMA strategy, as "Mean-Reverting" or "Trending" can only be performed under an "ex-post" evaluation rationale, i.e. only after each respective end-strategy's performance results (e.g. Sharpe Ratio) have been evaluated during a specific time period and are classified as significant enough. More intuitively set, a technical analyst which wishes to trade a specific asset or portfolio using the same strategy's rationale, could not know in advance (i.e. when he starts trading) if she should trade the respective asset as "Mean-Reverting" or "Trending", correspondingly. This main "pitfall" of the classical technical analysis EMA model used in this framework, emerges from the fact that no (ex-ante) inherent hyper-parameter estimation could provide information on whether the several portfolios are to be labeled and traded as "Mean-Reverting" or "Trending", respectively. To overcome this, we extend the predictability analysis by applying the Auto-regressive models in the following section, which we believe that at each time step t of the fitting process, will be able to inherently classify each analysed portfolio as "Mean-Reverting" or "Trending", depending on the sign of the respective fitted coefficients applied on the several portfolios' returns at previous time instances. More intuitively examined, we expect that the AR models will "learn" the underlying portfolio's class, being either "Mean-Reverting" or "Trending" depending on the fitted respective hyper-parameters (corresponding AR coefficients).

TABLE 3.19: Top performing PCA coordinates' based and global embedded portfolios' statistics, traded using the **EMA** "Mean-Reverting" or "Trending" Trading strategies. The several PCA embeddings are being performed in rolling window time configurations. They are also compared against the corresponding Random Walk strategy's returns, in terms of annualised mean and standard deviation levels. For each portfolio, the strategy's and Random Walk's respective one's returns mean values, are tested against the zeroed null hypothesis (**TT** and **TT(R.W.)** are the corresponding t-tests p-values), as well as against each other (**WT** : Welch's t-test p-value). **TW**: Time Window Configuration, **Co.Set** : Set of Coordinates used to construct the corresponding PCA embedded portfolio: "Head" means using the first $L = 2, 3, 4, 5$ eigenvectors for the global embedded portfolio construction, while "Tail" means using the last $L = 2, 3, 4, 5$ eigenvectors for the same embedded portfolios construction. When $L = 1$ the respective single coordinate's number is shown. **Sh** : EMA strategy's Sharpe Ratio, **Mean** : EMA strategy's returns' Mean, **Std** : EMA strategy's returns' Standard Deviation, **Sh(R.W.)** : Random Walk strategy's returns' Sharpe Ratio, **Mean(R.W.)** : Random Walk strategy's returns' Mean, **Std(R.W.)** : Random Walk strategy's returns' Standard Deviation. All respective statistics are in annualised format.

TW	Co.Set	t_d	Sh	Mean, CI(95%)	Std	TT	Sh(R.W.)	Mean(R.W.), CI(95%)	Std(R.W.)	TT(R.W.)	WT
Co.(MR)											
250	20	5	1.88	2.51 (1.92, 3.1)	1.34	0	-1.94	-2.6 (-3.19, -2.01)	1.34	0	0
250	20	2	1.82	2.44 (1.85, 3.03)	1.34	0	-1.94	-2.6 (-3.19, -2.01)	1.34	0	0
250	20	3	1.79	2.39 (1.8, 2.98)	1.34	0	-1.94	-2.6 (-3.19, -2.01)	1.34	0	0
250	20	15	1.74	2.33 (1.74, 2.92)	1.34	0	-1.94	-2.6 (-3.19, -2.01)	1.34	0	0
250	20	10	1.72	2.31 (1.72, 2.89)	1.34	0	-1.94	-2.6 (-3.19, -2.01)	1.34	0	0
150	20	2	1.6	2.38 (1.73, 3.04)	1.49	0	-1.45	-2.17 (-2.82, -1.51)	1.49	0	0
150	20	3	1.58	2.35 (1.69, 3.0)	1.49	0	-1.45	-2.17 (-2.82, -1.51)	1.49	0	0
150	20	5	1.53	2.28 (1.63, 2.94)	1.49	0	-1.45	-2.17 (-2.82, -1.51)	1.49	0	0
150	20	10	1.44	2.14 (1.49, 2.8)	1.49	0	-1.45	-2.17 (-2.82, -1.51)	1.49	0	0
250	20	25	1.43	1.92 (1.33, 2.51)	1.34	0	-1.94	-2.6 (-3.19, -2.01)	1.34	0	0
Global(MR)											
100	4-Tail	2	0.6	5.79 (1.51, 10.07)	9.72	0.01	-0.27	-2.57 (-6.85, 1.71)	9.72	0.24	0.01
250	4-Tail	5	0.55	4.98 (0.97, 8.98)	9.09	0.01	-0.35	-3.15 (-7.15, 0.86)	9.1	0.12	0
Co.(TR)											
100	1	2	0.5	16.75 (1.91, 31.59)	33.72	0.03	-0.04	-1.26 (-16.1, 13.59)	33.74	0.87	0.15

3.5.16 Auto-regressive Models (AR) Trading

We apply 3 AR(p) models, i.e. AR(1), AR(2) and AR(3) incorporating the more advanced modelling power of the coefficients' estimations of the p past points in time, which the AR models provide compared to the static and non-adjusted respective EMA mechanism. As already mentioned, we wish them to dynamically "learn" the underlying analysed portfolio's class, being either "Mean Reverting" or "Trending", and inherently generate the out-of-sample prediction, accordingly. In case the coefficients are negative, will indicate an underlying "Mean-Reverting" behavior of the portfolio's returns, while in the opposite case, the portfolio will be traded as "Trending". All of the coefficients, are tested for significance at each learning (re-fitting) step. More specifically, we train the very first instance of the AR models on an initial training set of 500 time observations (i.e. $\sim$ 10% of the entire data history). Then, at each new time point, we feed the latest 250 data points, back into the model for refitting and one-step ahead prediction generation, which then was translated to a trading signal for the time point $t+1$. No further optimization on the length of this latest history data subset was performed, since it was not one of the primary scopes of the overall analysis in this context. In order to be in line with the proper application procedure of the AR models, in each of the time windows outlined in our framework, i.e. both in rolling and expanding window cases, we performed the necessary stationarity tests, using the ADF test methodology. We used the *tsa* submodule of the statsmodels python module, Seabold and Perktold, 2010a, version v0.10.2, which automatically performs the test in each of the corresponding processed time windows. Where and if stationarity was not satisfied for a specific data window, the forecast of the model was set to zero, i.e. no trading activity was performed. We apply the AR model as per Equation 3.31 on portfolio π_t returns' time series, as in the case of the EMA trading framework described in the previous section, and we define the corresponding AR(p) directional prediction targeting trading strategy as the sign of the AR model's one-step prediction of the next portfolio's return at time $t+1$. If the AR(p) predicts that $\pi_{t+1} > \pi_t$, then a "Buy" ("Long") signal is generated, $s_{AR(p),t} = 1$, while a "Sell" ("Short") trading signal is generated if the predicted outcome would be $\pi_{t+1} < \pi_t$, $s_{AR(p),t} = -1$. A neutral position is suggested in any other case, $s_{AR(p),t} = 0$. We also define the LLE ARIMA trading strategy in the same way as in the case of the EMA trading framework in the previous section, using the previously defined ARIMA trading signal, instead.

We also apply the AR trading models in the (raw) Benchmark portfolios, as well. The best performing reported result, which manages to score a Sharpe Ratio of higher than the 0.5 threshold, is in the case of the Risk Parity portfolio, traded using an AR($p=2$) model, yielding a ("Gross") Sharpe Ratio of 0.75, with the returns mean being 12.5% (CI, 95% : 4.82%, 20.18%) and a standard deviation of 16.5%, correspondingly. As in the case of the EMA trading strategy, we show in the Transaction Cost section below (3.5.20), that the performance of this prevailing strategy, heavily worsens under the several transaction costs' scenarios, as well.

The corresponding performance statistics for the PCA embedding method, are displayed in Tables 3.20 for the coordinate based and global embedded portfolios, where the trading results are displayed in sorted (and higher than 0.5) Sharpe Ratios format. No Sharpe Ratio higher than the desired threshold was scored in any of the global embedded portfolios traded using the AR trading strategy, and as such the corresponding trading results were omitted. The results for the LLE AR trading strategy are displayed in Table 3.24 again in sorted top performing Sharpe Ratios' format for the several target traded assets as is in the case of EMA trading framework

TABLE 3.20: Top performing PCA coordinates' based and global embedded portfolios' statistics, traded using the **AR** Trading strategy. The several PCA embeddings are being performed in rolling window time configurations. They are also compared against the corresponding Random Walk strategy's returns, in terms of annualised mean and standard deviation levels. For each portfolio, the strategy's and Random Walk's respective one's returns mean values, are tested against the zeroed null hypothesis (**TT** and **TT(R.W.)** are the corresponding t-tests p-values), as well as against each other (**WT** : Welch's t-test p-value). **TW**: Time Window Configuration, **Co.Set** : Set of Coordinates used to construct the corresponding PCA embedded portfolio: "Head" means using the first $L = 2, 3, 4, 5$ eigenvectors for the global embedded portfolio construction, while "Tail" means using the last $L = 2, 3, 4, 5$ eigenvectors for the same embedded portfolios construction. When $L = 1$ the respective single coordinate's number is shown. **Sh** : AR strategy's Sharpe Ratio, **Mean** : AR strategy's returns' Mean, **Std** : AR strategy's returns' Standard Deviation, **Sh(R.W.)** : Random Walk strategy's returns' Sharpe Ratio, **Mean(R.W.)** : Random Walk strategy's returns' Mean, **Std(R.W.)** : Random Walk strategy's returns' Standard Deviation. All respective statistics are in annualised format.

TW	Co.	p	Sh	Mean, CI(95%)	Std	TT	Sh(R.W.)	Mean(R.W.), CI(95%)	Std(R.W.)	TT(R.W.)	WT
Coordinates											
250	20	1	**1.63**	2.21 (1.58, 2.84)	1.35	0	-1.99	-2.69 (-3.31, -2.06)	1.35	0	0
250	20	2	**1.44**	1.96 (1.33, 2.59)	1.35	0	-1.99	-2.69 (-3.31, -2.06)	1.35	0	0
250	20	3	**1.40**	1.9 (1.27, 2.53)	1.35	0	-1.99	-2.69 (-3.31, -2.06)	1.35	0	0
150	20	2	**1.17**	1.73 (1.05, 2.42)	1.47	0	-1.48	-2.18 (-2.86, -1.49)	1.47	0	0
150	20	3	**1.11**	1.64 (0.95, 2.33)	1.47	0	-1.48	-2.18 (-2.86, -1.49)	1.47	0	0
150	20	1	**1.10**	1.63 (0.95, 2.32)	1.47	0	-1.48	-2.18 (-2.86, -1.49)	1.47	0	0
100	20	3	0.76	1.22 (0.47, 1.97)	1.61	-1.21	0	-1.96 (-2.71, -1.21)	1.61	0 0	
100	20	2	0.70	1.13 (0.38, 1.88)	1.61	-1.21	0	-1.96 (-2.71, -1.21)	1.61	0	0
100	20	1	0.60	0.97 (0.22, 1.72)	1.61	-1.21	0.01	-1.96 (-2.71, -1.21)	1.61	0	0

in the previous section, as well.

When applied on the PCA coordinates' based embedded portfolios (Table 3.20), the highest Sharpe Ratio of 1.63, is scored for the embedded portfolio which is based on the last PCA coordinate (eigenvector) constructed under a rolling window embedding of 250 trading days. The corresponding AR(1) model's coefficients are reported to be negative, with the relevant statistics being displayed in the final section of the paper and significantly different than 1. This fact leads us on further arguing that they manage to efficiently capture the "Mean-Reverting" nature of the underlying traded portfolio, and as such acting as further empirical evidence of the findings of Cuturi and D'Aspremont, 2016. Furthermore, examining the global PCA embedded portfolios, we report no existence of Sharpe Ratios higher than 1 for any embedding and any coordinate set.

The AR trading results indicate that the AR models successfully inferred the "Mean-Reverting" behavior of the last PCA coordinate based embedded portfolio, constructed under the 250 and 150 trading days' rolling window configurations, scoring satisfactorily predictable (Box and Tiao, 1977) trading results, with generated Sharpe Ratios of even higher than 1.5. While these results match the ones obtained by the previously analysed EMA "Mean-Reverting" trading strategy, their added-value compared to them is that they successfully "learned" the underlying "Mean-Reverting" behavior of the respective portfolios in an "ex-ante" manner and traded them accordingly. Finally, the AR models failed on generating significant Sharpe Ratios when applied on the global PCA embedded portfolios, leading us on trying to explain this fact, as their failure to efficiently capture any potential hidden (short-medium-or long term) trends in the target FX Markets. We connect this assumption, with the thoroughly documented and analysed link between the first principal components of a PCA embedding and the respective modelled dynamical system's time derivative across time, as per Stanimirovic et al., 2006. Further exploring this note, we cite Pasini, 2017 on Principal Component Analysis for the stock market, further stating that the embedded portfolios constructed using a subset of the first eigenvectors of the PCA, i.e. using the coordinates of the corresponding embedding which explaining most of the data's variance, are actually highly connected to the 'Market' (Long Only) portfolio itself. As such, we state that the aforementioned weak generated AR models' trading results actually mean their failure to capture the 'Market' portfolio's dynamics across time.

For that reason, we further explore predictable behavior existence in the embedded portfolios, by applying additional, more sophisticated forecasting tools, like the Gaussian Process Regressors (GPR) and Recurrent Neural Networks (RNN). We target on the potential identification of predictable behavior in terms of both "Mean-Reverting" and "Trending" portfolios' dynamics, in the sense that the models will successfully manage to infer ("learn") this behavior themselves, and trade the portfolios accordingly, as in the case of the successful AR models of this section.

3.5.17 Gaussian Processes Regression (GPR)

As in the case of the RNN trading framework, we also apply Gaussian Process Regressors (GPR) towards "learning" the target portfolios' underlying behavior, again in an "ex-ante" manner. During the fitting process, the hyperparameters optimization are expected to successfully derive the trading style appropriate for each target portfolio, i.e. properly trade it as "Mean-Reverting" or "Trending", respectively, generating the corresponding predictions at each step of the analysis.

Targeting comparison consistency with the RNN models and corresponding trading strategies outlined in the relevant section, we used the same "training" and "trading" sub-periods configuration and with exactly the same input sequence parameter $\tilde{l}$'s values.

In Table 3.21, we show the top ranked respective Sharpe Ratios, after applying the described GPR trading strategy on the PCA embedded portfolios, while the LLE GPR trading framework's performance results are displayed in Table 3.24, as well. The corresponding Sharpe Ratios for the GPR trading strategy applied on the Long Only and Risk Parity portfolios are -0.17 and 0.22, respectively, again indicating a failure of the strategy to extract any informative patterns from the benchmark portfolios, as in the case of the RNN strategy, respectively.

3.5.18 Expanding Window Analysis

The results obtained in the previous sections have been focused on rolling window time varying configurations, for both the PCA and LLE embedding frameworks, as well as for the several forecasting methodologies on the embedded and Benchmark portfolios, respectively. In this section, towards further comparison and reference

TABLE 3.21: GPR Trading on the PCA embedded portfolios, sorted by Sharpe Ratios generated by the corresponding **GPR** trading strategy and compared against the corresponding Random Walk strategy's respective statistics. For each portfolio, the GPR strategy's and Random Walk's respective one's returns mean values, are tested against the zeroed null hypothesis (**TT** and **TT(R.W.)** are the corresponding t-tests p-values), as well as against each other (**WT** : Welch's t-test p-value). **TW**: Time Window Configuration, **Co** : Coordinate of the embedding, **Sh** : GPR strategy Sharpe Ratio, **Mean** : GPR strategy Mean, **Std** : GPR strategy Std, **Sh(R.W.)** : Random Walk strategy Sharpe Ratio, **Mean(R.W.)** : Random Walk strategy Mean, **Std(R.W.)** : Random Walk strategy Std. $\tilde{l}$ is the 'lookback' input parameter. All respective statistics are in annualised format.

TW	Co.	$\tilde{l}$	Sh	Mean, CI(95%)	Std	TT	Sh(R.W.)	Mean(R.W.), CI(95%)	Std(R.W.)	TT(R.W.)	WT
Coordinates											
250	20	5	1.00	1.38(0.76, 2.0)	1.37	0	-2	-2.74(-3.36, -2.12)	1.37	0	0
250	20	10	0.99	1.37(0.74, 1.99)	1.38	0	-2	-2.75(-3.37, -2.13)	1.37	0	0
250	20	25	0.91	1.25(0.63, 1.88)	1.38	0	-1.99	-2.74(-3.36, -2.11)	1.37	0	0
250	20	3	0.76	1.05(0.43, 1.67)	1.38	0	-1.99	-2.73(-3.35, -2.11)	1.37	0	0
100	20	5	0.69	1.12(0.39, 1.86)	1.62	0	-1.31	-2.13(-2.87, -1.4)	1.62	0	0
150	6	5	0.66	6.12 (1.93, 10.3)	9.24	0	0.03	0.26 (-3.93, 4.46)	9.24	0.9	0.05
150	17	10	0.61	3.67(0.94, 6.4)	6.02	0.01	-0.04	-0.26(-2.99, 2.47)	6.02	0.85	0.05
150	9	25	0.59	5.04(1.19, 8.9)	8.48	0.01	0.09	0.75(-3.11, 4.6)	8.49	0.7	0.12
100	20	10	0.58	0.94(0.21, 1.68)	1.62	0.01	-1.32	-2.15(-2.88, -1.41)	1.62	0	0
250	16	10	0.57	3.61(0.75, 6.47)	6.3	0.01	-0.35	-2.19(-5.05, 0.67)	6.31	0.13	0
150	10	5	0.56	4.82(0.93, 8.71)	8.57	0.02	0.07	0.57(-3.32, 4.45)	8.58	0.78	0.13
Global											
150	5-Tail	1	0.7	7.85(2.82, 12.87)	11.09	0	-0.04	-0.46(-5.49, 4.57)	11.1	0.86	0.02

purposes, we also perform the same methodologies in an Expanding Window time configuration, testing the assumption that useful information about the local dynamics of the several FX Pairs in our dataset, might be present in older time periods in the past, older even than the predetermined length of rolling windows of the previous sections, respectively. We speculate that the Expanding window configuration might be able to capture these interconnections between the several time periods, and efficiently project the dynamics in a predictable by the several trading models manner, as well. Further to the expected "Mean-Reverting" behavior of the last PCA coordinate based embedded portfolio, we again test the existence of potential predictable behavior in the rest of the PCA eigenspace based embedded portfolios, under the Expanding Window configuration, as well. Interestingly, the Expanding window embedding seems to be generating embedded portfolios, highly predictable and yielding increased trading performance in terms of Sharpe Ratios as compared to the rolling window embedding case. More specifically, in Table 3.22, we show the respective Sharpe Ratios of the top performing and worth mentioning strategies, developed under the Expanding window analysis, focusing on displaying the several trading models' (EMA, AR, RNN and GPR) forecasting efficiency, but also the existence of highly predictable embedded portfolios based on other than the expected last PCA eigenvector, as well. The latter is reported to be scoring the top "Gross" annualised Sharpe Ratio of 3.07, while traded under the EMA "Mean-Reverting" trading strategy in the same time window configuration, with the second best being the PCA embedded portfolio, which is based on the third eigenvector of the same embedding, traded under the same "Mean-Reverting" rationale and generating a "Gross" annualised Sharpe Ratio of 1.65, correspondingly. We highlight this finding as worth-mentioning, since it seems that the embedding under the Expanding Window configuration, projected the several FX Pairs' dynamics into the respective coordinate (3rd), by adding useful information in the resulted embedded portfolio, which in turn was captured by the several forecasting models, in a more efficient way. We further speculate that this third coordinate might be connected to a hidden underlying factor affecting the several pairs' dynamis, possibly related to a macroeconomic factor like e.g. interest rates differentials, inflation and/or growth expectations' differentials etc, of the several economies corresponding to the FX currencies in our dataset. Exploring this potential connection or relationship could be the topic of a future research work but falls out of the scope of the current one. The same portfolio's dynamics is also captured by the "Dynamically Learning" models AR($p = 3$), RNN($\tilde{l}$=1), and GPR($\tilde{l}$=10), as well, with the respective "Gross" annualised Sharpe Ratios being equal to 1.55, 0.95 and 0.56, respectively. Even though all three trading models seem to be capturing the same underlying dynamics, we comment on the underperformance of the GPR and RNN relative to the AR one, as the outcome of a potential "wrong" choice of the associated with each model hyperparameters' sets, e.g. GPR kernel, RNN epochs or number and type of Layers etc. We encourage further research to test and report the results of other hyperparameters' sets, as well. Furthermore, we also note that the "Dynamic Learning" desired property of the three models (AR, GPR and RNN) as opposed to the brute-forced (either "Mean-Reverting" or "Trending") rationale of the EMA trading framework seems to be information-ally "costly", as there is a loss in the respective generated Sharpe Ratios when compared to the EMA ones. Finally, for the case of the global PCA embedded portfolios trading, no significant performances are reported for the GPR and RNN models, while the top performing one, with a "Gross" annualised Sharpe Ratio of 1.01, is the global embedded portfolio which is based on the first four eigenvectors of the Expanding Window embedding, traded using an

TABLE 3.22: Trading on the PCA embedded portfolios, under an Expanding Window time varying configuration, sorted by Sharpe Ratios generated by the corresponding EMA, AR, RNN and GPR trading strategies, compared against the corresponding Random Walk strategy's respective statistics, as well. **DL** : "Dynamic Learning" of the hyperparameters inferring 'Mean-Reversion' or 'Trending' trading behavior

Predictor	Co.	Sh	Mean, CI(95%)	Std	TT	Sh (R.W.)	Mean (R.W.)	,CI(95%)	Std (R.W.)	TT (R.W.)	WT	DL
Coordinates												
EMA(MR,$t_d = 2$)	20	3.07	3.98 (3.41, 4.55)	1.3	0	-2.92	-3.79	(-4.37, -3.22)	1.3	0	0	No
EMA(MR,$t_d = 2$)	3	1.65	17.82 (13.07, 22.57)	10.79	0	-1.25	-13.55	(-18.31, -8.79)	10.82	0	0	No
EMA(TR,$t_d = 25$)	7	0.58	4.72 (1.14, 8.3)	8.13	0.01	-0.02	-0.18	(-3.76, 3.4)	8.14	0.92	0.08	No
AR($p = 1$)	20	2.82	3.55 (2.97, 4.14)	1.25	0	-2.97	-3.73	(-4.31, -3.15)	1.25	0	0	Yes
AR($p = 3$)	3	1.55	17.06 (11.96, 22.17)	10.97	0	-1.42	-15.65	(-20.76, -10.54)	10.98	0	0	Yes
GPR($\tilde{l}$=5)	20	2.23	2.85 (2.28, 3.43)	1.28	0	-2.99	-3.8	(-4.37, -3.22)	1.27	0	0	Yes
GPR($\tilde{l}$=10)	3	0.95	10.48 (5.54, 15.41)	10.88	0	-1.38	-15.09	(-20.02, -10.17)	10.86	0	0	Yes
RNN($\tilde{l}$=1)	20	1.89	2.4 (1.82, 2.98)	1.28	0	-2.98	-3.8	(-4.38, -3.23)	1.27	0	0	Yes
RNN($\tilde{l}$=1)	3	0.56	6.18 (1.24, 11.11)	10.88	0.01	-1.37	-15.02	(-19.94, -10.1)	10.85	0	0	Yes
Global												
EMA(MR,$t_d = 5$)	3-Tail	0.74	5.39 (2.17, 8.61)	7.31	0	-0.56	-4.07	(-7.29, -0.85)	7.32	0.01	0	No
EMA(TR,$t_d = 2$)	4-Head	0.81	32.88 (14.91, 50.84)	40.83	0	0.77	31.37	(13.4, 49.33)	40.84	0	0	No
AR($p = 2$)	4-Head	1.01	42.04 (22.75, 61.32)	41.41	0	0.83	34.5	(15.2, 53.79)	41.43	0	0.59	Yes
GPR($\tilde{l}$=1)	3-Tail	0.53	3.94 (0.61, 7.26)	7.34	0.02	-0.46	-3.44	(-6.77, -0.12)	7.34	0.04	0	Yes
RNN($\tilde{l}$=1)	2-Tail	0.55	2.61 (0.47, 4.76)	4.73	0.02	-0.62	-2.97	(-5.11, -0.82)	4.73	0.01	0	Yes

AR($p = 2$) trading model, and more specifically, traded under a "Trending" trading rationale. Similar Sharpe Ratio level is obtained when trading the same portfolio using an EMA($t_d = 2$) "Trending" trading strategy, where we note that the "information cost" of the "Dynamic Learning" property discussed in the previous case, is not applied, and more specifically the exact opposite effect is observed. It seems that in the global embedding framework, the brute-forced trading behavior seems to be hitting the strategy's performance, as opposed to the AR model, whose "learning" property seems to be "protecting" it, accordingly.

For the case of the LLE trading framework, all respective trading performance results are displayed in Table 3.24. Following the rationale of the previous analysis, regarding the PCA embedded portfolios' trading, we show the respective LLE trading results, focusing on displaying the several LLE trading indicators' efficiency in respect to the two main points below :

1. Their capacity to optimally forecast the dynamics of other than the expected-to-perform-well PCA embedded portfolio, which is based on the last PCA eigenvector and traded under a "Mean-Reverting" trading rationale, ideally "Trending" target portfolios, whose dynamics failed to be infered and traded accordingly, by the several trading models under the PCA only based framework earlier.

2. Their capacity to generate significant Sharpe Ratios (higher than the heuristically set threshold of 0.5), under the rolling window and/or the Expanding Window time configurations.

Mathematically, the trading strategies that are implemented under the LLE framework are formalised as follows :

$$y_{LLE,j,t+1} = sign(\tilde{M}(q_{j,t})) * \begin{cases} y_{LO,t+1} & \text{for the Long Only portfolio} \\ y_{RP,t+1} & \text{for the Risk Parity portfolio} \\ y_{PCA,j,t+1} & \text{for the PCA coordinate based embedded portfolios} \\ Y_{PCA,t+1} & \text{for the global PCA embedded portfolios} \end{cases} \tag{3.60}$$

where $\tilde{M}$ denotes the applied trading model $\tilde{M}$: EMA, AR, GPR, RNN on the j-th LLE coordinate (eigenvector time series) which acts as the main tested trading indicator, $j = 1, 2, \ldots, 5$, at time t, and $y_{LLE,j,t+1}$ the corresponding strategy's return at time $t + 1$. As such, and following the logic in section 3.5.15 for the LLE EMA trading strategy's formalisation, we can write :

$$sign(\tilde{M}(q_{j,t})) = \begin{cases} s_{EMA,t} & \text{for the } \tilde{M} : \text{EMA LLE trading framework} \\ s_{AR,t} & \text{for the } \tilde{M} : \text{AR LLE trading framework} \\ s_{GPR,t} & \text{for the } \tilde{M} : \text{GPR LLE trading framework} \\ s_{RNN,t} & \text{for the } \tilde{M} : \text{RNN LLE trading framework} \end{cases} \tag{3.61}$$

where $s_{EMA,t}, s_{AR,t}, s_{GPR,t}, s_{RNN,t}$ are the corresponding "Buy" and "Sell" (binary) trading signals of each respective section. Under this logic, a successful LLE trading strategy, i.e. one generating highly significant Sharpe Ratio scores, intuitively means that an implied leading correlation between the LLE trading indicator's and target traded portfolio's dynamics is being inferred by the analysis. However, while to the best of our knowledge, no theoretical background exists on the expected sign of this correlation, i.e. regarding the expected correlation between a specific LLE eigenvector's predictor's dynamics, captured by the term $sign(\tilde{M}(q_{j,t}))$ and the respective target traded portfolio's one ($y_{LO,t}$,$y_{RP,t}$,$y_{PCA,t}$ and $Y_{PCA,t}$), we show both the ("ex-post") inferred positive and negative correlations of all tested combinations of strategies. That way, we wish to highlight the fact that, while in the pure PCA based analysis of the previous sections, the several forecasting models were tested on their capacity to capture the dynamics of the target portfolios' dynamics, themselves, i.e. translating their predictions in investment returns directly, in the case of the LLE trading framework, the models are used simply as dynamics predictors of the several LLE eigenvectors over time, which are not directly connected to investment returns as in the previous case. Instead, they are used for investment decisions (trading) at a second stage, inferring potential correlation between the target portfolios' dynamics and their predictors' one, yielding respective investments returns, as well. In a future approach we will try to estimate this correlation in a rolling and/or expanding window manner as in the context of This work's embedding analysis, as well, evaluating the respective correlation capturing model's efficiency and reporting the corresponding results. At the current one, we simply display the inferred leading correlations between the several strategies, targeting empirical evidence reporting of the LLE trading indicators and the several target portfolios, especially the PCA embedded ones which are not easily modeled and traded by the trading strategies outlined in the previous sections.

The main tested assumption behind this approach, is a potential lagging/leading relationship existence between the several LLE trading indicators' dynamics and the respective one of the target traded portfolios. In that case, they could be used as "pre-signals" regarding the respective portfolios' dynamics - at least in short-term horizon, ideally yielding significant performances in terms of Sharpe Ratios for portfolios which fall outside the direct PCA approach, as well. Interestingly, the reported LLE trading results in Table 3.24, indicate that such inter-connection between the two dynamics is indeed apparent, and act as empirical evidence of some worth that indeed the set of "predictable" portfolios can be extended. Moreover, both "Mean-Reverting" and "Trending" portfolios are identified. First in the list, is the one based on the first PCA eigenvector under the 250 trading days rolling window configuration, which failed to generate a significant performance under the previous PCA only based trading framework, and which seems to be satisfactorily predicted (with a positive inferred correlation) with the "Mean-Reverting" EMA($t_d = 100$) predictor of the first LLE eigenvector's dynamics, embedded under the same time window configuration, generating a "Gross" annualised Sharpe Ratio of 0.82. Similar finding is the one related to the first PCA coordinate embedded portfolio, constructed under the Expanding Window configuration and again satisfactorily traded using the first LLE eigenvector based EMA($t_d = 50$) predictor, but under a "Trending" trading style, instead. We highlight both results, since the respective LLE trading indicators seem to be offering valuable additional informative value towards the more efficient forecasting of the respective FX "Trend" related PCA coordinate based embedded portfolio's dynamics over time. Furthermore, additional predictive capacity seems to be offered by several LLE trading indicators on PCA global embedded portfolios, as well, which are also associated with a "Trending" underlying behavior of the FX markets related to the FX Pairs in our dataset. More specifically, the combination ($Y_{PCA,100}$(3-Head), $q_5(25)$) generates a "Gross" annualised Sharpe Ratio of 0.84, if traded under the "Trending" EMA($t_d = 150$) predictor, while several other similar connections can be implied by the AR, GPR and RNN models, respectively. Additionally to these results, several other embedded portfolios, different than the first or last PCA coordinates' based embedded portfolios, are satisfactorily forecasted and traded under the LLE trading framework, with interesting correlations emerging between the several PCA coordinates' and the LLE predictors' dynamics, respectively.

Even more interestingly, when exploring the Benchmark portfolios' LLE trading related results, we see that that same EMA($t_d = 50$) predictor of the first LLE eigenvector's dynamics, which was able on adequately capturing the $y_{PCA,EW,1}$ portfolio's returns over time under a "Mean-Reverting" trading logic, is also capable of satisfactorily forecast the Long Only portfolio's dynamics over time, as well, however using a "Trending" trading rationale, instead. The inferred correlations between the three portfolios dynamics, actually imply a significant negative correlation between the Long Only portfolio and the PCA first coordinate based embedded portfolio under the Expanding Window configuration, since they are both satisfactorily forecasted by the same leading LLE indicator. Furthermore, no such apparent strong correlation seems to be the case for the correlation between the Long Only and the PCA first coordinate based embedded portfolio under the 250 trading days rolling window configuration, instead. Towards further reference and testing of this assumption, we also calculate the Pearson correlation coefficient ($\tilde{\rho}$ between the Long Only and the two PCA first coordinate based embedded portfolios, (a) the one under the Expanding Window configuration and (b) the second being constructed under the 250 trading days' rolling window, respectively. Indeed, a strong negative correlation of -0.87 between the $y_{PCA,EW,1}$ and the Long Only portfolio is reported, while no such strong in absolute correlation terms relationship seems to be the case for the $y_{PCA,250,1}$ embedded portfolio, with a small correlation of 0.10 with the Long Only portfolio, respectively. We highlight the findings as really interesting, in the sense that while a close connection between the first PCA coordinate based embedded portfolio and the market's trend related Long Only portfolio dynamics, is expected, such a connection is validated at least in absolute terms, only under the Expanding Window configuration, while the rolling window embedding fails to correlate itself with either of the two. As such, we display the importance of the embedding time window configuration in the trading framework outlined in this work, while also commenting that the negative sign of the correlation between the $y_{PCA,EW,1}$ and the Long Only portfolio dynamics could also be explained by the sign ambiguity correction algorithm used to impose deterministic outcome of the respective embeddedings' loadings, as per Bro, Acar, and Kolda, 2008, already discussed and outlined in the beginning of the manuscript. Towards even greater insight on the time evolution of the respective PCA embedded portfolios' FX Pairs' allocations (loadings), we plot them in Figure 3.27, where we observe two main points :

1. The negatively signed allocations suggested by the Expanding Window PCA, further validating its negative correlation with the Long Only portfolio, where all FX Pairs are allocated an equal $1/D$ portfolio weight accordingly.

2. The "stability" and "smoothness" of the corresponding portfolio allocations of the Expanding Window approach, as opposed to the 250 trading days rolling window related ones.

Upon these two findings, we base our statement that the time window configuration under which the embeddings are performed, matters in a great extend. The stability, in the sense that no frequent and high-leveled "jumps" between positive and negative allocations are generated by the Expanding Window analysis, could be one of the main information related added values of this approach relative to the rolling window ones. We highlight the only "switch" of the respective allocations, which happened during the 2008' market crash period, where the embedding identified a potential "structural break" on the underlying dynamics and suggested a "protection" against the new regime. No other such switch exist in the rest of the time series. Careful consideration and treatment of the respective allocations is needed, especially after applying the sign ambiguity algorithm by Bro, Acar, and Kolda, 2008, in the sense that it might yield negatively signed allocations, even though "Trending" portfolios as the one under consideration is expected to be constructed by positive signed ones. In financial trading applications, such treatment is tremendously important, since it is directly related to the "Buy" and "Sell" corresponding investment decisions taken by the market practitioner. Upon this note, we comment that in the trading strategies' framework outlined in the previous sections, this "intuitively inconsistent" sign of the respective allocations played no role in the final obtained performance results, since the several "Dynamically Learning" models (AR, GPR, RNN) successfully inferred this sign reversal and generated corresponding trading signals capturing the (intuitively) reversed emerged dynamics. Furthermore, the corresponding EMA trading rationales also displayed this inconsistency implicitly via the generated "ex-post" Sharpe Ratios, as well.

On another note, we also highlight the significance and informative added value of the LLE trading indicators in the underlying dynamics' forecasting of the portfolios under consideration, i.e. the PCA first coordinate based embedded portfolio under the Expanding Window configuration versus the rolling window ones, focusing for simplicity on the 250 trading days' rolling window configuration, respectively. More specifically, the LLE trading indicators sufficiently forecast the target portfolios "Trending" dynamics, with the respective scored Sharpe Ratios being highly significant and beating the respective threshold, while the several trading models (EMA, AR, GPR, RNN) applied directly on the same "Trending" portfolios, failed to generate worth-mentioning performances. More specifically, all models' "Gross" annualised Sharpe Ratios average, is close to 0.05, while the ones under the LLE trading indicators' framework approach even territories close to 1. Furthermore, worth-noting fact as well, is the one that the LLE trading framework also performed well, when applied in on the Long Only portfolio itself, (Table 3.24), satisfactorily capturing the general market's trend related portfolio, too.

Under the above results' analysis, we state that the combination of a general market's trend related portfolio, as the traditional Long Only portfolio or the "Trending" $y_{PCA,EW,1}$ one, with a time series model capable of capturing the drift (mean) of the respective returns' time series, can yield a good approximation of the long-term trend of the underlying markets, at least in a "Gross" returns level. We select and test the most significant portfolios of the previous sections, on their resilience under several transaction costs' scenarios application, in the following section, further validating (or not) their efficiency on an environment where the real market's frictions are also applied.

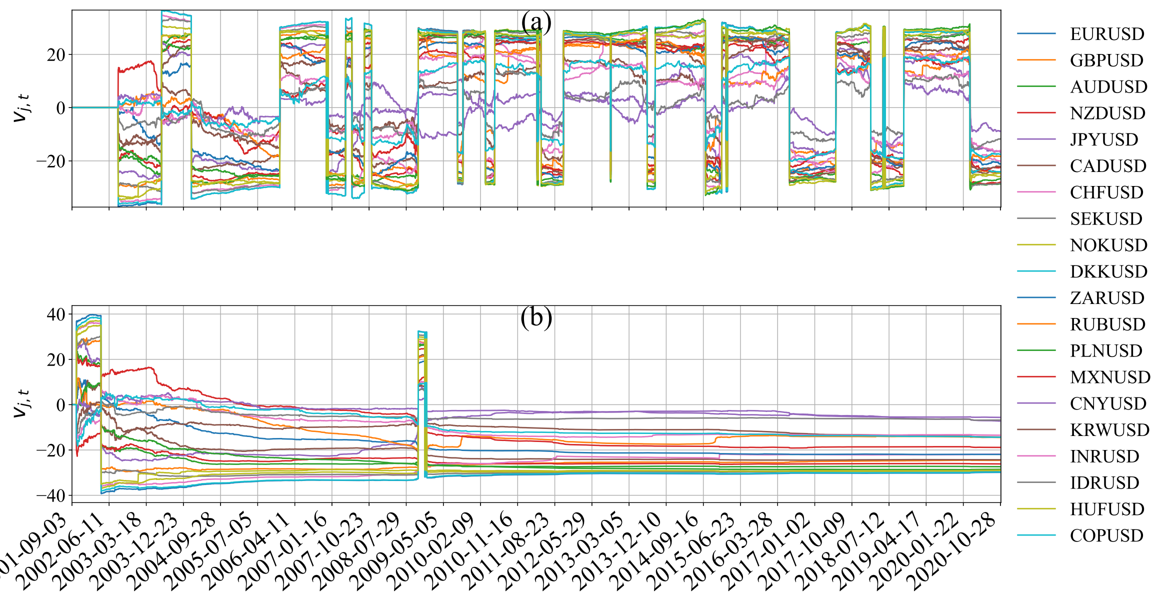

FIGURE 3.27: PCA first coordinate based embedded portfolios' allocations (a) for the 250 trading days' rolling window and (b) for the Expanding Window analysis

TABLE 3.23: Correlation Matrix of the Long Only, Risk Parity and PCA first coordinate based embedded portfolios under the Expanding Window but also under the 250 trading days' rolling window, respectively.

$\tilde{\rho}$	$y_{PCA,250,1}$	$y_{PCA,EW,1}$	Long Only
$y_{PCA,250,1}$	1.00	0	0.10
$y_{PCA,EW,1}$	0	1.00	-0.87
Long Only	0.10	-0.87	1.00

For completeness, additionally to the above, we also need to define the "direct" LLE indicators' related trading strategies, as :

$$y_{LLE_d,j,t+1} = sign(q_{j,t}) * \begin{cases} y_{LO,t+1} & \text{for the Long Only portfolio} \\ y_{RP,t+1} & \text{for the Risk Parity portfolio} \\ y_{PCA,j,t+1} & \text{for the PCA coordinate based embedded portfolios} \\ Y_{PCA,t+1} & \text{for the global PCA embedded portfolios} \end{cases} \tag{3.62}$$

with the associated top-5 best results, being displayed in Table 3.25. Commenting on these performances the resulted Sharpe Ratio levels are all below 1, for both the individual coordinate based ones as also for the global ones, while though as in Table 3.24 above, several other portfolios different from the last-PCA coordinate based one are filtered out as being "tradable" either under a "Mean-Reverting" or "Trending" regime.

LLE Trading Framework Criticism

Even though the reported LLE trading related results, act as empirical evidence that more hidden structures seem to exist in the target FOREX data, potentially exploitable towards further expanding the already identified predictable portfolios under the direct PCA framework list, no theoretical background (or even intuitive sense) about the defined by Equations 3.60 and 3.62 linear projecting relationships between the (linear) PCA and (the nonlinear) LLE embeddings exists. Furthermore, such nonlinear approaches are not easily associated with any financial or economical implicit interpretation, as well, and as such we simply display the corresponding results mostly for demonstrative purposes, in order to act as empirical evidence towards the need to extend our manifold learning based hybrid scheme to nonlinear manifold learning techniques, as well. The necessary theoretical rigorousness and mathematical formalism, for such nonlinear techniques (as the LLE), to be properly used in trading and portfolio management related applications, as the current, is associated with the "Pre-Image Problem" and reconstruction (or "lifting") approaches, analysed in Section 2.4. The subsequent FOREX trading application analysed in Section 3.5.22, is dedicated on developing such a mathematically solid framework, via the so called "Embed-Train-Lift" approach.

TABLE 3.24: Top performing LLE based trading strategies' (Equation 3.60) performance metrics, per trading model (EMA, AR, GPR and RNN) applied on the traded target portfolios, i.e. Long Only, Risk Parity and PCA embedded portfolios, respectively. Also compared against the corresponding Random Walk trading strategy's returns, in terms of annualised mean and standard deviation levels. **TP**: Trading Portfolio, **TI**: Trading Indicator

TP	TI	$\tilde{M}$	IC	Sh	Mean,CI(95%)	Std	TT	Sh (R.W.)	Mean (R.W.)	CI(95%)	Std (R.W.)	TT (R.W.)	WT
EMA													
$y_{PCA,250,1}$	$q_1(250)$	MR,$t_d = 100$	+	0.82	27.81 (13.06, 42.55)	33.38	0	0	0.15	(-14.61, 14.91)	33.42	0.98	0.01
$y_{PCA,100,9}$	$q_5(250)$	MR,$t_d = 2$	+	0.85	7.08 (3.46, 10.69)	8.19	0	0.02	0.13	(-3.55, 3.81)	8.33	0.95	0.01
$y_{PCA,EW,1}$	$q_1(EW)$	TR,$t_d = 50$	-	0.76	25.66 (10.93, 40.37)	33.33	0	-0.16	-5.36	(-20.11, 9.4)	33.39	0.48	0.06
$Y_{PCA,100}$ (3-Head)	$q_5(25)$	TR,$t_d = 150$	-	0.84	33.73 (16.18, 51.26)	39.7	0	-0.08	-3.18	(-20.77, 14.42)	39.82	0.72	0.02
LO	$q_1(EW)$	MR,$t_d = 50$	+	0.89	6.26 (3.21, 9.3)	6.9	0	0.42	2.92	(-0.14, 5.97)	6.92	0.06	0.13
RP	$q_3(250)$	MR,$t_d = 150$	+	0.62	10.03 (3.04, 17.01)	15.81	0	0.88	14.03	(7.04, 21.01)	15.8	0	0.43
AR													
$y_{PCA,250,6}$	$q_1(25)$	$p = 2$	+	0.87	7.65 (3.83, 11.47)	8.65	0	-0.44	-3.93	(-7.86, -0.0)	8.89	0.05	0
$y_{PCA,150,16}$	$q_3(EW)$	$p = 3$	-	0.76	4.68 (1.98,7.38)	6.11	0	-0.32	-2.04	(-4.84, 0.76)	6.33	0.15	0.18
$y_{PCA,EW,14}$	$q_1(250)$	$p = 3$	-	0.75	5.36 (2.26,8.46)	7.01	0	-0.16	-1.22	(-4.46, 2.02)	7.33	0.46	0.07
$Y_{PCA,100}$ (5-Head)	$q_4(25)$	$p = 2$	+	0.82	34.61 (16.31, 52.88)	41.39	0	-0.03	-1.48	(-20.2, 17.24)	42.37	0.88	0.01
LO	$q_3(250)$	$p = 3$	+	0.59	4.06 (1.07, 7.05)	6.76	0.01	0.42	2.92	(-0.14, 5.97)	6.92	0.06	0.6
RP	$q_3(250)$	$p = 3$	+	0.64	10.24 (3.32, 17.15)	15.66	0	0.88	14.03	(7.04, 21.01)	15.8	0	0.45
GPR													
$y_{PCA,EW,8}$	$q_5(150)$	$\tilde{l}$=5	+	0.83	8.99 (3.92, 13.15)	10.72	0	0.18	1.91	(-2.8, 6.62)	10.67	0.43	0.05
$y_{PCA,250,7}$	$q_4(250)$	$\tilde{l}$=2	-	0.83	7.85 (3.44,11.45)	9.31	0	-0.15	-1.35	(-5.36, 2.67)	9.08	0.51	0.04
$y_{PCA,EW,17}$	$q_1(100)$	$\tilde{l}$=25	-	0.82	4.48 (1.92,6.56)	5.41	0	-0.64	-3.49	(-5.87, -1.11)	5.38	0	0.66
$Y_{PCA,100}$ (4-Tail)	$q_1(100)$	$\tilde{l}$=3	+	0.88	8.73 (4.06, 12.51)	9.82	0	-0.26	-2.57	(-6.85, 1.71)	9.69	0.24	0
LO	$q_4(250)$	$\tilde{l}$=3	-	0.74	5.24 (1.95, 8.0)	7.02	0	0.42	2.92	(-0.14, 5.97)	6.92	0.06	0
RP	$q_4(250)$	$\tilde{l}$=3	-	0.73	12.02 (4.43,18.39)	16.22	0	0.88	14.03	(7.04, 21.01)	15.8	0	0
RNN													
$y_{PCA,100,11}$	$q_2(25)$	$\tilde{l}$=1	+	0.86	7.37 (3.38, 10.63)	8.42	0	-0.17	-1.48	(-5.18, 2.23)	8.39	0.43	0
$y_{PCA,100,8}$	$q_1(250)$	$\tilde{l}$=3	+	0.80	7.41 (3.12, 10.95)	9.1	0	-0.17	-1.51	(-5.48, 2.45)	8.98	0.45	0
$y_{PCA,EW,18}$	$q_3(EW)$	$\tilde{l}$=1	-	0.84	4.69 (2.07,6.83)	5.53	0	-0.33	-1.8	(-4.23, 0.62)	5.49	0.15	0.13
$Y_{PCA,EW}$ (3-Tail)	$q_3(EW)$	$\tilde{l}$=1	-	1.01	7.47 (3.93,10.25)	7.33	0	-0.55	-4.07	(-7.29, -0.85)	7.29	0.01	0.19
LO	$q_2(100)$	$\tilde{l}$=5	-	0.69	4.94 (1.67, 7.71)	7.02	0	0.42	2.92	(-0.14, 5.97)	6.92	0.06	0
RP	$q_1(150)$	$\tilde{l}$=3	-	0.55	9.05 (1.6,15.57)	16.23	0.02	0.88	14.03	(7.04, 21.01)	15.8	0	0

TABLE 3.25: Top performing "direct" LLE based trading strategies' (Equation 3.62) performance metrics, applied on the traded target portfolios, i.e. Long Only, Risk Parity and PCA embedded portfolios, respectively. Also compared against the corresponding Random Walk trading strategy's returns, in terms of annualised mean and standard deviation levels. **TP**: Trading Portfolio, **TI**: Trading Indicator

TP	TI	IC	Sh	Mean, CI (95%)	Std	TT	Sh (R.W.)	Mean (R.W.) ' CI (95%)	Std (R.W.)	TT (R.W.)	WT
$y_{PCA,EW,9}$	q_4(EW)	+	0.7916	7.84(3.5,12.19)	9.83	0	-0.15	-1.45 (-5.8, 2.9)	9.85	0.51	0
$Y_{PCA_1 50}$(5-Tail)	q_2(150)	+	0.7390	8.13(3.31,12.94)	10.91	0	-0.1	-1.1(-5.93,3.72)	10.92	0.65	0.01
$y_{PCA,150,2}$	q_2(250)	+	0.6667	10.51 (3.6, 17.41)	15.64	0	-0.12	-1.88(-8.86, 5.09)	15.79	0.6	0.01
$Y_{PCA,100}$(4-Tail)	q_2(EW)	+	0.66206	6.46(2.18, 10.73)	9.68	0	-0.26	-2.57(-6.85, 1.71)	9.69	0.24	0
$y_{PCA,EW,9}$	q_2(250)	+	0.6616	6.33(2.13, 10.52)	9.49	0	-0.15	-1.45(-5.8, 2.9)	9.85	0.51	0.01

3.5.19 Stationarity, FBM-Matern Processes and "Tradability"

Towards even greater exploration of the possibility of Random Walk intrinsic generators in the several FX Pairs' and constructed portfolios' returns dynamics, we perform and report the results of the Augmented Dickey-Fuller (ADF) test, Dickey and Fuller, 1979, testing for the potential existence of Unit Root in the relevant time series data (essential and also necessary for the proper application of the several time series forecasting tools outlined in the previous sections) and also the Hurst Exponent (H) and correspondingly the α parameter connecting the fbm and the Matern processes via the Equation 3.40 investigated earlier on the GPR modelling framework. Based on these, we discuss the meaning of the calculated values regarding the "tradability" of the several pairs and target portfolios, under the "Mean-Reversion" vs "Trending" trading regime discussed in the current context, while displaying the reasons of the current "hybrid" framework's failure to capture the "Trending" ones. Upon that fact, we state the need of constructing a different manifold learning based trading approach, towards identifying and efficiently trading these non-captured "Trending" portfolios, and claim that this new approach is one developed and outlined in the subsequent and final financial trading related application of the Thesis.

For the stationarity testing purposes, we calculate the ADF statistic values, critical threshold values and corresponding p-values, for each of the constructed portfolios, (both the benchmark ones, PCA and LLE based embedded ones), all denoted as π_t for simplicity. The ADF test is written as (see Supplementary Material section as well A):

$$\pi_t = c + \delta t + \phi \pi_{t-1} + \beta_1 \Delta \pi_{t-1} + ... + \beta_\tau \Delta \pi_{t-\tau} + \epsilon_t, \tag{3.63}$$

where Δ is the difference operator, τ is the number of the lagged difference terms (modeler's choice), ϵ_t is an innovation process with zero mean. The null hypothesis is: $H_0 : \phi = 1$ with the alternative being $H1 : \phi < 1$. If the test, defined as $DF_\tau = \frac{\hat{\phi}}{SE(\hat{\phi})}$, fails to reject the null hypothesis, or more specifically when the computed test statistic value is less than the relevant critical value for the Dickey–Fuller test, it fails to reject the possibility of existence of unit root. For demonstrative purposes, we show the 250 days rolling window ADF test's p-values, as calculated on the global PCA embedded portfolios, which were constructed under the time window configurations mentioned in section 3.5.4, in Figure 3.28. We can observe a break of stationarity (p-value greater than 0.0005) in the global PCA embedded portfolios constructed in the different time window configurations, during 2008 financial crisis, 2015-2016 period and in the recent 2020 Covid-19 market crash, as well. We comment that the break of stationarity during the period 2015-2016 might be caused by the first time ever movement from zero (and even negative) valued interest rates to positive territories, across the major economies. We also observe that some of the stationarity breaks for the global LLE embedded portfolio case happen in 2010-2011 period, potentially capturing the effect of the European sovereign debt crisis, as well. (As a note, we skip the first 500 values in the graphs, avoiding the few first calculations on the very first few data). All of them are below the 0.01 threshold level, for all embedded portfolios, with a small spike during the 2008-2009 mortgage financial market crash in the US.

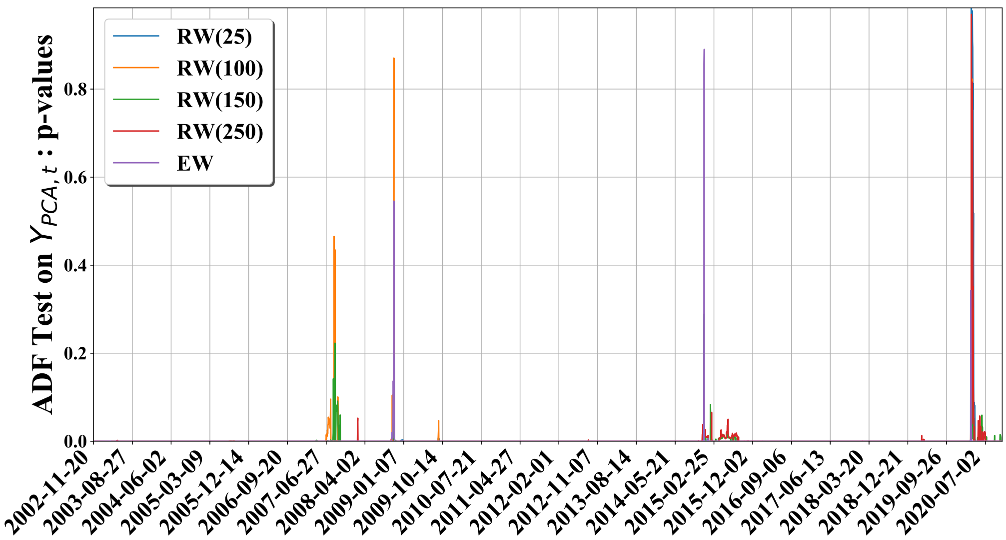

FIGURE 3.28: (Blue) ADF test p-values on the 25 days rolling window PCA global embedded portfolio. Break of stationarity during 2020 Covid-19 period. (Orange) ADF test p-values on the 100 days rolling window PCA global embedded portfolio. Break of stationarity during the 2008 market crash and 2020 Covid-19 periods. (Green) ADF test p-values on the 150 days rolling window PCA global embedded portfolio. Break of stationarity during the 2008, 2015-16 period and 2020 Covid-19 market crash period. (Red) ADF test p-values on the 250 days rolling window PCA global embedded portfolio. Break of stationarity during 2020 Covid-19 period. (Purple) ADF test p-values on the expanding window PCA global embedded portfolio. Break of stationarity during the 2008, 2015-16 period and 2020 Covid-19 market crash period.

Furthermore, we recall the form of the Hurst exponent $\tilde{H}$, already thoroughly investigated in previous sections, as being the slope of the fitted line of the power law :

$$E(\frac{\tilde{R}(N)}{\tilde{S}(N)}) = \tilde{C}N^{\tilde{H}} \tag{3.64}$$

as $N \to \infty$, where N the data points in the time series under analysis, $\tilde{R}(N)$ denotes the range of the first N cumulative deviations from the mean, $\tilde{S}(N)$ is the their standard deviations' sum, E the expected value operator and $\tilde{C}$ is a constant. It is

TABLE 3.26: Hurst Exponents ($\tilde{H}$) of the FX Pairs' carry-adjusted returns under the Standalone (Long Only) and Risk Parity portfolio frameworks

Asset	Hurst(LO)	α(LO)	Hurst(RP)	α(RP)
EURUSD	0.5441	1.0441	0.5558	1.0558
GBPUSD	0.5222	1.0222	0.5321	1.0321
AUDUSD	0.5719	1.0719	0.5591	1.0591
NZDUSD	0.5770	1.0770	0.5508	1.0508
JPYUSD	0.4818	0.9818	0.5130	1.0130
CADUSD	0.5386	1.0386	0.5776	1.0776
CHFUSD	0.5383	1.0383	0.5229	1.0229
SEKUSD	0.5085	1.0085	0.5399	1.0399
NOKUSD	0.5203	1.0203	0.5403	1.0403
DKKUSD	0.5398	1.0398	0.5525	1.0525
ZARUSD	0.5225	1.0225	0.5017	1.0017
RUBUSD	0.5529	1.0529	0.7046	1.2046
PLNUSD	0.5341	1.0341	0.5734	1.0734
MXNUSD	0.4691	0.9691	0.5276	1.0276
CNYUSD	0.5318	1.0318	0.2990	0.7990
KRWUSD	0.5079	1.0079	0.5177	1.0177
INRUSD	0.5421	1.0421	0.5748	1.0748
IDRUSD	0.5893	1.0893	0.5646	1.0646
HUFUSD	0.5411	1.0411	0.5435	1.0435
COPUSD	0.5530	1.0530	0.5579	1.0579
Aggregated	0.5782	1.0782	0.6212	1.1212

based on the work of Mandelbrot and Wallis in Mandelbrot and Wallis, 1968 and Mandelbrot and Wallis, 1969. The threshold level that we are interested in, is the $\tilde{H} = 0.5$, which indicates Random Walk (Brownian Motion) existence. Above that, $\tilde{H} > 0.5$ a process with long-term positive auto-correlation (long-range persistence or dependence) is indicated, while the case $\tilde{H} < 0.5$ indicates a long lasted switch between high and low values of the recent history of the time series ("short memory" fbm). As such, the process can be characterised as "trending" under the $\tilde{H} > 0.5$ scenario, and as a mean-reverting one in case $\tilde{H} < 0.5$. We show the Hurst (and associated α) values of the several FX Pairs in our dataset, under the Long Only and the Risk Parity frameworks, in Table 3.26. The same values for the PCA coordinate based embedded portfolios are displayed in Table 3.27 and for the global PCA embedded portfolios in Table 3.28, respectively. (Note : the Hurst values under the Long Only portfolio, are actually the ones of the pairs' "raw" dynamics, respectively).

We observe that in the case of the ("raw") individual pairs' (i.e. under the Long Only framework), the Hurst levels are distributed around the random walk noise critical level of 0.5, with the max values being under 0.6 and the lowest above 0.45. We consider these levels as an indication of no "strongly Trending" or "strongly Mean-Reverting" behaviors existence in the several underlying processes dynamics. However, in the case of the corresponding levels under the Risk Parity approach, interesting levels are obtained for e.g. the RUB/USD pair with a Hurst value of 0.7046 and that of CNYUSD with a Hurst value of 0.2990. The first one indicates a significant "long-range" dependence or "memory" present in the respective dynamics, which could characterise it as "Trending" related, while on the other hand for

TABLE 3.27: Hurst Exponents ($H(\tilde{T}W)$) of the PCA coordinate based embedded portfolios constructed under the time varying windows TW, which correspond to the rolling windows of 25,100,150 and 250 trading days and also the expanding window one denoted as (EW)

Co.	Hurst(25)	α(25)	Hurst(100)	α(100)	Hurst(150)	α(150)	Hurst(250)	α(250)	Hurst(EW)	α(EW)
1	0.4732	0.9732	0.5911	1.0911	0.5547	1.0547	0.5600	1.0600	0.5361	1.0361
2	0.4733	0.9733	0.6082	1.1082	0.4548	0.9548	0.5388	1.0388	0.5863	1.0863
3	0.5361	1.0361	0.6075	1.1075	0.4718	0.9718	0.4514	0.9514	0.4884	0.9884
4	0.4779	0.9779	0.5508	1.0508	0.5309	1.0309	0.4663	0.9663	0.5755	1.0755
5	0.5009	1.0009	0.4624	0.9624	0.4889	0.9889	0.5198	1.0198	0.5627	1.0627
6	0.5157	1.0157	0.5006	1.0006	0.5484	1.0484	0.5701	1.0701	0.5335	1.0335
7	0.4585	0.9585	0.5537	1.0537	0.4949	0.9949	0.4601	0.9601	0.4981	0.9981
8	0.5707	1.0707	0.4822	0.9822	0.4963	0.9963	0.4519	0.9519	0.4936	0.9936
9	0.5099	1.0099	0.4667	0.9667	0.5585	1.0585	0.5407	1.0407	0.5713	1.0713
10	0.5217	1.0217	0.4960	0.9960	0.5097	1.0097	0.4888	0.9888	0.5329	1.0329
11	0.5347	1.0347	0.6227	1.1227	0.5139	1.0139	0.5249	1.0249	0.5436	1.0436
12	0.5158	1.0158	0.6304	1.1304	0.4636	0.9636	0.4654	0.9654	0.5214	1.0214
13	0.5138	1.0138	0.5087	1.0087	0.4997	0.9997	0.4794	0.9794	0.5093	1.0093
14	0.5076	1.0076	0.4830	0.9830	0.4880	0.9880	0.5172	1.0172	0.4628	0.9628
15	0.5878	1.0878	0.5167	1.0167	0.5114	1.0114	0.5771	1.0771	0.5196	1.0196
16	0.5358	1.0358	0.5192	1.0192	0.4908	0.9908	0.5573	1.0573	0.3819	0.8819
17	0.4909	0.9909	0.4882	0.9882	0.5369	1.0369	0.4740	0.9740	0.4260	0.9260
18	0.5165	1.0165	0.4816	0.9816	0.5389	1.0389	0.4613	0.9613	0.4710	0.9710
19	0.5485	1.0485	0.4802	0.9802	0.5173	1.0173	0.5213	1.0213	0.6412	1.1412
20	0.5041	1.0041	0.4765	0.9765	0.4870	0.9870	0.4292	0.9292	0.3785	0.8785

TABLE 3.28: Hurst Exponents ($H(\tilde{T}W)$) of the PCA "global" embedded portfolios constructed under under the time varying windows TW, which correspond to the rolling windows of 25,100,150 and 250 trading days and also the expanding window one denoted as (EW)

Co.Set	Hurst(25)	α(25)	Hurst(100)	α(100)	Hurst(150)	α(150)	Hurst(250)	α(250)	Hurst(EW)	α(EW)
2(Head)	0.4677	0.9677	0.6203	1.1203	0.5307	1.0307	0.5498	1.0498	0.5770	1.0770
2(Tail)	0.5217	1.0217	0.4869	0.9869	0.5255	1.0255	0.5183	1.0183	0.6340	1.1340
3(Head)	0.4887	0.9887	0.6424	1.1424	0.5306	1.0306	0.5573	1.0573	0.5784	1.0784
3(Tail)	0.5387	1.0387	0.5133	1.0133	0.4703	0.9703	0.4683	0.9683	0.5684	1.0684
4(Head)	0.5060	1.0060	0.6381	1.1381	0.5284	1.0284	0.5650	1.0650	0.5830	1.0830
4(Tail)	0.4999	0.9999	0.5213	1.0213	0.5222	1.0222	0.4855	0.9855	0.5464	1.0464
5(Head)	0.5055	1.0055	0.6341	1.1341	0.5288	1.0288	0.5695	1.0695	0.6029	1.1029
5(Tail)	0.5386	1.0386	0.4706	0.9706	0.4854	0.9854	0.4654	0.9654	0.5222	1.0222

the second one a strong "anti-persistent", or "Mean-Reverting" related dynamics is inferred. (Under a financial perspective, we comment on these levels as highly connected to the carry-effects related to the underlying currencies comprising the pairs (Auer and Hoffmann, 2016). Interestingly the same Hurst level increase happens in the aggregated portfolios' level, as well, with the Hurst of the Long Only portfolio being 0.5782 getting increased to 0.6212 after the Risk Parity approach is performed. Intuitively this can be commented as a move towards "smoothness" after removing the several variance (risk) effects, at least under the current adaptive inverse risk rolling window framework's approach.

Due to the fact that the majority of the pairs did not provide such strong indication towards "Trending" or "Mean-Reverting" familiar dynamics, we move on the corresponding results obtained under the PCA embedding portfolio construction framework. More specifically, we performed the same Hurst levels calculations on the PCA embedded portfolios constructed under all time windows under focus, i.e. the rolling window ones of 25, 100, 150 and 250 trading days, as well as for the expanding window based one. Interesting levels corresponding to Hurst values greater than 0.6 (inferring "Trending" behavior), are obtained for the 2nd, 3rd, 11-th and 12-th coordinate based portfolios under the 100 rolling window approach, while no other portfolio constructed under the other rolling window frameworks generated Hurst values larget than this threshold, or even below 0.4, indicating stronger "Mean-Reverting" tendencies compared with even the individual pairs. This fact changes though under the expanding window approach, where two interesting Hurst levels are obtained. The Hurst exponent of the 20-th (last PCA eigenvector) coordinate based embedded portfolio is 0.3785, and that of the 16-th coordinate, 0.3819, indicating strong relationship with "Mean-Reverting" behavior, while the 19-th one, is producing a Hurst value of 0.6412, inferring a "Trending" related behavior, instead. Those levels are associated with the promising trading results obtained after applying the proposed hybrid approach based on the EMA, AR, GPR and RNN forecasting methodologies on the respective portfolios validating the existence of the previously mentioned "tradable" "Mean-Reverting" or "Trending" trading patterns, under the associated specifications towards the underlying manifold identification. For instance, the success of the AR(1) model on the 20-th coordinate based portfolio (as also suggested by the inspiring work of Cuturi et. al) is associated and connected with the "Mean-Reverting" inferring level of the corresponding Hurst, while the construction of the respective GPR forecasting model using the Matern kernel as discussed in the relevant section, is also associated with the "memory" induction of the underlying fbm process. . The same rationale applies on the rest of the "successful" portfolios, as well, with the inference of fbm existence meaning that the underlying markets are "beatable" at least "locally". More specifically, it seems that the embedding process (i.e. the exploitation of cross-feature correlation patterns in the original space data), induces "local memory" (fbm property) in the form of stationarity with nonlinear variance in the embedded space "mapped" data, and as such extra informational power for the several forecasting models to consume is offered, as opposed to the traditional approach on applying them directly on the original space. (Note : In the several tables of Hurst exponents calculations, we also display the associated α parameter's levels, due to its more direct relationship with the Matern processes as per Lilly et al., 2017.

Intuitively, the models applied will not be able to produce significant trading performances for the portfolios that have Hurst values close to 0.5 (with the "significant" range of 0.4-0.6 having been set heuristically, since the proper identification of is is not under the scope of our analysis). Connecting this with the Matern processes and the associated kernel used in the GPR forecasting framework, with the parameter ν being set to 0.5 - absolute exponential kernel associated with the "mean-reverting" (or "trending") (f)OU process - we expect their performance to also be better when applied on the portfolios experiencing the previously high (>0.6) or low(<0.4) Hurst related dynamics, however, due to their capacity to act as the "continuum" between White noise and power-laws, could potentially identify any further nonlinear hidden patterns (extra "memory") in the not so easily characterisable as "Mean-Reverting" or "Trending" (under the "ex-post" Hurst levels analysis) portfolios, as well. We can easily infer their "success" or "failure" towards this task, by the several models' trading performance evaluation in terms of ranked (best-to-worse) Sharpe Ratios, meaning that if no significant Sharpe Ratio levels are obtained, then the models fail to identify such "tradability" in the constructed portfolios.

Towards this, we focus on Tables , 3.20 and 3.21, corresponding to the obtained Sharpe Ratios of the (linear) EMA, AR and (nonlinear) GPR based strategies, as applied on the PCA embedded portfolios constructed under the rolling window approach. We observe that the first two frameworks' best results, are indeed all generated by the trading of the last PCA coordinate (20-th) based embedded portfolio, experiencing the "max" "Mean-Reverting" inferred dynamics, as also indicated by the corresponding low Hurst value discussed earlier. However, this is not the case for the Matern kernel constructed GPR model, whose mix of "best performing" portfolios (additionally to the 20-th PCA coordinate based one), include embedded portfolios constructed based on other PCA coordinates, as well, i.e. the coordinates 6,9,10,16 and 17, (constructed under the 150 and 250 trading days rolling windows frameworks). The corresponding Hurst levels for all those portfolios are "insignificant" for the "tradability" under the framework discussed in the current sections, since their levels are between the heuristically set range band of 0.4-0.6. As such, based on the obtained trading results, it seems that the GPR models managed to capture some "local memory" rich patterns under the rolling window approach on trading the target portfolios, which are probably highly nonlinear in nature, otherwise the linear models should be able to capture them, as well. The fact that these patterns might be present only "locally" can also be inferred by the fact that no such "successful" performance was obtained under the Expanding Window framework, for the same trading models and configurations. We observe that all modelling frameworks ranked as best (under the Sharpe metric) the strategies trading the same PCA expanding window embedded portfolios (namely the 20-th and the 3rd coordinates). Intuitively, the expanding window approach, seems to act as a "convergence towards noise", as also indicated by the "ex-post" "non-significant" Hurst values associated with these specific portfolios in Table 3.27, and as such no information or "memory" rich patterns can be identified by the several forecasting models, as in the case of the rolling window approaches earlier.

3.5.20 Best Models and Transaction Cost Analysis (TCA)

In order to make the trading results even more realistic, we select the best performing models of the previous sections and perform a transaction cost analysis (TCA), testing their efficiency and "Net" returns, under 6 scenarios of transaction costs' schemes. The selection process is focused on portfolios and trading strategies, which managed to score significant Sharpe Ratios, higher than the heuristically set threshold of 0.5, but also on ones which offer direct financial and economical meaning, as well. As such, the traditional portfolios, Long Only and Risk Parity, whose intuitive and practical trading sense is already defined and discussed in the relevant sections, 3.5.5 and 3.5.6, are already included to the list of the target portfolios, which along with the several LLE trading strategies described in the previous sections provide a good approximation of the general market's trend. The respective TCA results are also compared on the standalone portfolios' themselves, as well, towards greater insight on the informative added value of the proposed trading strategies. Additionally, we also focus on the top performing PCA embedded portfolios, traded under either the several time series based models (EMA, AR, GPR and RNN) or the LLE trading framework, as well, discussing the potential "loss" of performance in terms of Sharpe Ratio scores' deterioration, accordingly.

As a next step, we wish to construct an aggregated portfolio of the trading strategies which are more resilient in the market costs' impact, exploring the potential of yielding even higher Sharpe Ratios than the individual strategies' contributions, due to probable diversification effects.

In Table 3.29, we show the transaction costs scenarios for each pair, which will be the base of our TCA, spanning pricing profiles from "heavily aggressive" to "traditional - conservative" ones. More specifically, the format of the transaction costs is treated and displayed in percentage (%) format, representing that way, any "foregone" return due to the portfolios' "re-balancing", or otherwise stated, positioning adjustments over time. The price difference between 'top-sellers' (Ask or Offer) and 'top-buyers' (Bid), defines the so called market (or "Bid-Offer") spread. In our analysis, this Bid-Offer market cost impact is treated and displayed in terms of percentage (%) of each subsequent trade's value (or "trade notional") at each of the traded FX Pairs traded in this context. For example in the case of EURUSD, getting a snapshot in the market with a Bid Quote of 1.19317 and an Ask Quote of 1.19318, then the market spread's impact - cost in a trade that would buy or sell the pair in the corresponding time would be, (1.19318-1.19317/1.19317) = 0.00084%, and the same rationale would apply in any other FX pair in our dataset. We apply the same market spread costs charging rationale, in all of the FX pairs, after taking a snapshot of a normal trading day's quotes from a set of several retail and institutional online brokers and market makers, whose "raw-spreads" are publicly available in the website, www.forex.com. Other similar sites could also be used, but extensive TCA in terms of market depth, liquidity effect and number of Liquidity Providers etc, is not in the scope of the current work, though we heavily encourage future work to dig deeper in the effects of other transaction costs' schemes in the several portfolios' returns, as well.

This snapshot, will consist the main assumed market cost scenario, notated as "Main Scenario (MS)" in Table 3.29. We also construct 5 more scenarios based on this one, targeting further exploration of respective "break-even" levels of the analysed strategies' net returns over time. More specifically, we construct an even more "aggressive" (meaning "cheaper") cost scheme than the main one, by heuristically reducing the corresponding recorded costs by 25%, and also 4 more conservative ones, by multiplying the same individual costs by 1.25, 1.5, 2 and 3, implying a costlier overall impact of 25%, 50%, 100% and 200%, respectively. We comment that the main cost scheme scenario, could be the one, upon which a well established market player (market maker, trading house, hedge fund, bank etc) would operate and perform each trading transactions. The more aggressive one (MS * 0.75) would correspond to a market player with even "smarter" transactions' execution systems and mechanisms (e.g. a big investment bank acting as a major electronic trading venue), while the more conservative one (MS * 1.25) would be replicating a slightly worse cost based market player than the other two, towards greater insight and comparison between the three cost schemes in the overall performance of the aggregated strategies' portfolios which will be constructed in the next step. The rest of the transaction cost schemes would replicate a "retail" investor, trading under several retail brokers' trading platforms and structures. The several transaction costs are applied only on each subsequent asset's position change in the portfolio, meaning only when a corresponding trade (portfolio position change) has been performed. Mathematically formalised, the total transaction cost, TTC, incurred after a trade in asset i, on which the portfolio has a position u_t at time t, will be

$$TTC_{i,t} = (u_{i,t} - u_{i,t-1}) * ITC_i \tag{3.65}$$

where ITC_i denotes the "individual transaction cost" in percentage format (%) for each asset, as per each transaction cost scenario displayed in Table 3.29. This cost is the foregone return which in turn is subtracted from the respective portfolio's return at time t for each target portfolio. We denote as "Resilience Level" the number of transaction costs' scenarios that a respective strategy manages to overcome ("survive"), i.e. a strategy with $RL = 6$ will be the one that manages to provide a significant "Net" Sharpe Ratio higher than 0.5, even under the scenario which is three times worse than the Main one (MS * 3). We select the sub-set of the most resilient trading systems out of the top performing and TCA filtered ones displayed in Table 3.29, in order to perform a so called "Resilience Contribution" analysis in the next section. More specifically, we will combine the corresponding "Mean-Reverting" and "Trending" trading systems, towards the exploration of "Break-Even" points in the corresponding aggregated portfolios' "Resilience Levels". That way, we are trying to tackle the challenge of combining the respective trading strategies, towards "enhancing" the more sensitive-to-transaction-costs ones, which as observed in the corresponding Table (3.29) are the ones that are connected to "Mean-Reverting" trading styles, with the more resilient to costs ones, which again from the same Table are reported to be the "Trending" trading style connected ones. We show that indeed such "resilience strengthening" against deteriorated transaction costs' schemes (like the ones assumed in the current context) is possible, up to a certain degree, but of course this resilience gain comes with an additional information burden, in terms of respective Sharpe Ratios' values.

The top performing portfolios of the several trading frameworks of the previous sections, are displayed in Table 3.30, where also the several "Gross" (without transaction costs) and "Net" (transaction costs are incorporated) Sharpe Ratios under the several costs' scenarios, are calculated. (<u>Notations</u> : **TP**: Target Portfolio : Benchmark Portfolios (LO, RP), PCA embedded portfolios (PCA). **Strategy** : Either trading model ($\tilde{M}$: EMA, AR, GPR, RNN), or LLE trading indicator (i.e. $\tilde{M} : q_j(TW)$ on the j-th LLE eigenvector constructed under the TW respective time window configuration), applied on the target traded portfolio. **Sys.No** : The number of each corresponding trading system. **Sh** : "Gross" Sharpe Ratio of the respective trading strategy. **MS** : Main Scenario of transaction costs. **RL** : "Resilience Level", i.e. the number of transaction costs' scenarios under which the strategy's (net) Sharpe Ratio manages to stay significant, i.e. above the heuristically set threshold of 0.5. We highlight the corresponding "Net" Sharpe Ratio levels in **bold**).

Towards the selection of the end strategies, which will be comprising the final suggested portfolios for next section's "Resilience Contribution" analysis, we note and comment upon some interesting results of Table 3.30. More specifically, we observe that the majority of the trading systems lost almost all of their trading returns' performance, even under the Main scenario's transaction costs, while only one

TABLE 3.29: Transaction costs in percentage (%) terms for each of the FX Pairs in our Dataset.

Pair	MS * 0.75	MS	MS * 1.25	MS * 1.5	MS * 2	MS * 3
EURUSD	0.0006	0.0008	0.0011	0.0013	0.0017	0.0025
GBPUSD	0.0022	0.0029	0.0036	0.0043	0.0057	0.0086
AUDUSD	0.0020	0.0026	0.0033	0.0039	0.0053	0.0079
NZDUSD	0.0021	0.0028	0.0035	0.0042	0.0056	0.0085
JPYUSD	0.0014	0.0018	0.0023	0.0027	0.0036	0.0054
CADUSD	0.0024	0.0033	0.0041	0.0049	0.0065	0.0098
CHFUSD	0.0025	0.0033	0.0041	0.0049	0.0065	0.0098
SEKUSD	0.0062	0.0083	0.0103	0.0124	0.0165	0.0248
NOKUSD	0.0071	0.0094	0.0118	0.0141	0.0188	0.0283
DKKUSD	0.0024	0.0032	0.0040	0.0048	0.0064	0.0096
ZARUSD	0.0322	0.0430	0.0537	0.0644	0.0859	0.1289
RUBUSD	0.0147	0.0195	0.0244	0.0293	0.0391	0.0586
PLNUSD	0.0179	0.0239	0.0298	0.0358	0.0477	0.0716
MXNUSD	0.0205	0.0273	0.0341	0.0410	0.0546	0.0819
CNYUSD	0.0035	0.0046	0.0058	0.0070	0.0093	0.0139
KRWUSD	0.0226	0.0301	0.0376	0.0451	0.0602	0.0903
INRUSD	0.0091	0.0121	0.0151	0.0181	0.0242	0.0363
IDRUSD	0.0258	0.0345	0.0431	0.0517	0.0689	0.1034
HUFUSD	0.0592	0.0790	0.0987	0.1185	0.1579	0.2369
COPUSD	0.0050	0.0067	0.0083	0.0100	0.0133	0.0200

(Sys.No = 30) managed to overcome all 6 ones. Interestingly, the corresponding trading strategy, was based on the LLE trading framework, while also it traded one of the more "challenging" portfolios, namely the 250 trading days rolling window first PCA coordinate based embedded portfolio, which as already discussed, is related to a "Trending" trading rationale and no trading model among the ones tested in this context ($\tilde{M}$: EMA, AR, GPR and RNN) applied on the portfolio's returns itself, managed to forecast their respective dynamics, generating significant trading performance results in terms of Sharpe Ratio levels. However, the trading system using information from the LLE embedding and more specifically from the dynamics of the first LLE eigenvector, constructed under the same time window configuration, and approximated using the EMA($t_d = 100$) predictor, was really encouraging, with the scored "Net" Sharpe Ratios being additionally highly resilient under all transaction costs' schemes. The respective strategy is the first one that we select and add in the final list of strategies towards the next section's analysis, while we also speculate that its trading dynamics is highly connected to the underlying FX markets' general trend. Furthermore, we also suggest that its resilience explanation, could be the topic of a future research work, with a potentially good starting point being the non-frequent signal change (or portfolio re-balancing) of the respective LLE trading indicator, as opposed to the trading signals created using information by the portfolio's dynamics itself. There might be the case, that the LLE embedding and the respective coordinate is not as easily "confused" by the underlying portfolio's volatility (or "noise" in the respective statistical moments), as opposed to the predictors that were solely based on its own respective dynamics. As such, more efficient minimization of the corresponding "noise" effects in the predictor's output (as per Box and Tiao, 1977) was accomplished. Another worth-mentioning observation, is

TABLE 3.30: Top Performing Strategies' Sharpe Ratios, under each forecasting methodology and each Transaction Cost scenario, as well

Sys.No	TP	Strategy	Sh	MS * 0.75	MS	MS * 1.25	MS * 1.5	MS * 2	MS * 3	RL
1	LO	EMA(TR,$t_d = 2$)	0.73	-4.67	-5.96	-6.98	-7.79	-8.9	-10.06	0
2	RP	EMA(TR,$t_d = 2$)	1.17	**0.68**	**0.53**	0.38	0.23	-0.06	-0.64	2
3	LO	EMA(MR, $t_d = 50$):q_1(EW)	0.89	-0.82	-1.33	-1.78	-2.19	-2.84	-3.7	0
4	RP	EMA(MR, $t_d = 150$):q_3(250)	0.62	**0.57**	**0.55**	**0.53**	**0.51**	0.46	0.38	4
5	LO	AR($p = 3$):$q_3(250)$	0.59	-4.89	-6.18	-7.19	-7.98	-9.07	-10.18	0
6	RP	AR($p = 3$):$q_3(250)$	0.64	-0.36	-0.67	-0.98	-1.28	-1.86	-2.92	0
7	LO	GPR($\tilde{l}$=3):$q_4(250)$	0.74	-6.57	-8.26	-9.57	-10.57	-11.93	-13.28	0
8	RP	GPR($\tilde{l}$=3):$q_4(250)$	0.73	0.11	-0.09	-0.3	-0.5	-0.89	-1.67	0
9	LO	RNN($\tilde{l}$=5):$q_2(100)$	0.69	-6.61	-8.29	-9.59	-10.59	-11.95	-13.3	0
10	RP	RNN($\tilde{l}$=3):$q_1(150)$	0.55	-0.11	-0.33	-0.55	-0.76	-1.19	-2.01	0
11	$y_{PCA,250,20}$	EMA(MR,$t_d = 5$)	1.88	**1.21**	**1.00**	**0.77**	**0.55**	0.11	-0.76	4
12	$y_{PCA,150,20}$	EMA(MR,$t_d = 2$)	1.60	**0.69**	0.39	0.08	-0.22	-0.81	-1.97	1
13	$y_{PCA,100,4-Tail}$	EMA(MR,$t_d = 2$)	0.60	-1.47	-2.14	-2.8	-3.44	-4.66	-6.82	0
14	$y_{PCA,100,1}$	EMA(TR,$t_d = 2$)	0.5	0.22	0.13	0.04	-0.05	-0.24	-0.6	0
15	$y_{PCA,EW,20}$	EMA(MR,$t_d = 2$)	3.07	**2.27**	**2.01**	**1.74**	**1.47**	**0.94**	-0.11	5
16	$y_{PCA,EW,3}$	EMA(MR,$t_d = 2$)	1.65	**0.65**	0.32	-0.01	-0.34	-1.0	-2.26	1
17	$y_{PCA,EW,4-Head}$	EMA(MR,$t_d = 2$)	0.81	0.4	0.26	0.12	-0.01	-0.28	-0.82	0
18	$y_{PCA,250,20}$	AR($p = 1$)	1.63	**0.65**	0.32	-0.01	-0.34	-0.98	-2.23	1
19	$y_{PCA,150,20}$	AR($p = 2$)	1.17	0.25	-0.05	-0.36	-0.66	-1.26	-2.41	0
20	$y_{PCA,100,20}$	AR($p = 3$)	0.76	-0.21	-0.52	-0.84	-1.16	-1.77	-2.95	0
21	$y_{PCA,EW,20}$	AR($p = 1$)	2.82	**1.95**	**1.66**	**1.36**	**1.07**	0.49	-0.66	4
22	$y_{PCA,EW,3}$	AR($p = 3$)	1.55	**0.53**	0.19	-0.15	-0.49	-1.16	-2.45	1
23	$y_{PCA,EW,4-Head}$	AR($p = 2$)	1.01	**0.65**	**0.53**	0.41	0.29	0.05	-0.43	2
24	$y_{PCA,250,20}$	GPR($\tilde{l} = 5$	1.00	0.05	-0.26	-0.57	-0.88	-1.49	-2.67	0
25	$y_{PCA,150,6}$	GPR($\tilde{l} = 5$)	0.66	-0.4	-0.75	-1.1	-1.45	-2.13	-3.47	0
26	$y_{PCA,EW,20}$	GPR($\tilde{l} = 5$)	2.23	**1.35**	**1.08**	**0.81**	**0.54**	0.01	-1.06	4
27	$y_{PCA,EW,20}$	GPR($\tilde{l} = 10$)	1.00	-0.12	-0.49	-0.86	-1.23	-1.96	-3.35	0
28	$y_{PCA,250,20}$	RNN($\tilde{l} = 5$)	0.82	-0.14	-0.46	-0.78	-1.1	-1.73	-2.95	0
29	$y_{PCA,EW,20}$	RNN($\tilde{l} = 5$)	1.89	**1.07**	**0.79**	**0.52**	0.24	-0.3	-1.37	3
30	$y_{PCA,250,1}$	EMA(MR, $t_d = 100$):q_1(250)	0.82	**0.78**	**0.76**	**0.74**	**0.73**	**0.69**	**0.62**	6
31	$y_{PCA,100,9}$	EMA(MR, $t_d = 2$):q_5(250)	0.85	-0.33	-0.72	-1.11	-1.5	-2.26	-3.73	0
32	$y_{PCA,EW,1}$	EMA(TR, $t_d = 50$):q_1(EW)	0.76	**0.69**	**0.67**	**0.64**	**0.62**	**0.57**	0.47	5
33	$Y_{PCA,100,3-Head}$	EMA(TR, $t_d = 150$):q_5(25)	0.84	**0.72**	**0.68**	**0.63**	**0.59**	**0.51**	0.34	5
34	$y_{PCA,250,6}$	AR($p = 2$):q_1(25)	0.87	0.2	-0.03	-0.26	-0.48	-0.93	-1.81	0
35	$y_{PCA,150,16}$	AR($p = 3$):q_3(EW)	0.76	-0.41	-0.8	-1.19	-1.57	-2.31	-3.71	0
36	$y_{PCA,EW,14}$	AR($p = 3$):$q_1(250)$	0.75	-0.23	-0.55	-0.88	-1.2	-1.83	-3.02	0
37	$Y_{PCA,100}$(5-Head)	AR($p = 2$):$q_4(25)$	0.82	0.48	0.36	0.24	0.12	-0.11	-0.58	0
38	$y_{PCA,EW,8}$	GPR($\tilde{l}$=5):$q_5(150)$	0.83	-0.28	-0.64	-1.0	-1.36	-2.06	-3.41	0
39	$y_{PCA,250,7}$	GPR($\tilde{l}$=2):$q_4(250)$	0.83	-0.29	-0.65	-1.02	-1.38	-2.1	-3.49	0
40	$y_{PCA,EW,17}$	GPR($\tilde{l}$=25):$q_1(100)$	0.82	-0.54	-0.99	-1.43	-1.86	-2.7	-4.26	0
41	$Y_{PCA,100}$(4-Tail)	GPR($\tilde{l}$=3):$q_1(100)$	0.88	-1.48	-2.24	-2.98	-3.71	-5.08	-7.49	0
42	$y_{PCA,100,11}$	RNN($\tilde{l}$=1):$q_2(25)$	0.86	-0.43	-0.85	-1.28	-1.69	-2.51	-4.07	0
43	$y_{PCA,100,8}$	RNN($\tilde{l}$=3):$q_1(250)$	0.80	-0.44	-0.84	-1.25	-1.65	-2.45	-3.98	0
44	$y_{PCA,EW,18}$	RNN($\tilde{l}$=1):q_3(EW)	0.84	-0.42	-0.83	-1.23	-1.62	-2.37	-3.71	0
45	$Y_{PCA,EW}$(3-Tail)	RNN($\tilde{l}$=1):q_3(EW)	1.01	-1.34	-2.09	-2.81	-3.5	-4.77	-6.85	0

the fact that the same target traded portfolio and LLE trading indicator combination, constructed under the Expanding Window configuration and forecasted using an EMA($t_d = 50$) predictor, also scored similarly significant "Net" Sharpe Ratios under the 5 first trsnaction costs' schemes ($RL = 5$). This fact indicates again the same underlying connection between the respective embeddings' coordinates. For this reason, we additionally add this strategy in the final list of the ones which will be forwarded to the analysis of the next section, as well.

Another interesting connection is reported, as well, between the fifth LLE coordinate based dynamics (approximated using an EMA($t_d = 150$) predictor) and the (again "Trending" trading style related) global embedded PCA portfolio, constructed under the 100 trading days' rolling window configuration and using the first 5 respective eigenvectors, respectively. However, since we lack any intuitively or theoretically based financial interpretation of the corresponding connection, we omit it from the final list of selected strategies.

As a final note on the LLE trading indicators' related strategies, we observe the failure of any "Dynamically Learning" model like the AR, GPR or RNN one, to efficiently forecast the dynamics of the respective top ranked and selected portfolios. We speculate and state that, by observing the (previously analysed) successful EMA predictors' hyperparameters' values, i.e. $t_d = 100$ and $t_d = 50$ respectively, there seems to be a useful and "cyclical" information pattern, describing the respective traded portfolio's dynamics over time, with a duration of 1 to 2 trading months, which failed to be captured by the more short-term "viewing" AR, GPR and RNN models, as per their corresponding hyperparameters indicate, i.e. the p order for the AR models and the $\tilde{l}$ "look-back" parameter for the GPRs and RNNs, respectively.

For our next choice and addition in the final portfolio of strategies of the next section, we also focus on the trading systems with numbers 11,15, 21 and 26, with "Resilience Levels" of at least 4, respectively, meaning systems which managed to survive even the transaction cost scenario that was 50% worse than the main one (MS * 1.5). The corresponding strategies are all trading the last PCA coordinate based embedded portfolio, under the "Mean-Reverting" (expected by Cuturi and D'Aspremont, 2016) trading style, as well, either under EMA, AR or GPR predictors, (Table 3.30). We add all 4 strategies in the list of systems which will be analysed in the next section.

As a final comment, we also report an interesting connection related to the trading system with number 4, i.e. the one trading one of the benchmark portfolios, the Risk Parity one, under the LLE trading framework, yielding a resilience level of 4 and a corresponding "Net" Sharpe Ratio of 0.51 (i.e. slightly higher than the 0.5 threshold of our analysis). However, since again we cannot associate any financially interpret-able connection between the respective LLE trading indicator and the respective benchmark portfolio's dynamics, we omit it from our final list forwarded to the next section.

3.5.21 Resilience Contribution

As already discussed in the previous section, at this final one, we wish to tackle the challenge of efficiently combining the respective top-performing and selected trading systems of the previous section, in order to increase the Resilience levels of the several sensitive-to-transaction-costs ones. We are going to accomplish this by combining them with more the more resilient to costs ones, i.e. the "Trending" trading style strategies of the same list.

The "Trending" systems in our final list, are the ones with numbers 30 and 32 of the reference Table 3.30, which as discussed in the previous section, as well, are considered the more resilient to the several transaction costs schemes applied in the current context, while the "weaker" "Mean-Reverting" ones are the ones with numbers 11, 15, 21 and 26 referenced in the same Table. Since each strategy has its respective volatility (or risk) of returns, different from any of the others in the list, we follow an "Equal variance" approach towards the target strategies' combination. More specifically, we divide all strategies' returns time series, with the unconditional sample standard deviation of the respective returns' time series, across the entire time period of analysis. That way, all strategies are guaranteed to have identical risks, i.e. returns' variances equal to 1, and can now be combined with the focus being on the average returns dynamics over time (drift), after incorporating the several transaction costs in the overall aggregated portfolio performance, as well. The combination is simply the sum of the respective trading systems' returns time series, e.g. the "Mean-Reverting" trading system with number 11, combined with the "Trending" trading system with number 30 etc. The "Resilience Gain" is defined to be the case where the "Mean-Reverting" strategy's "Net" Sharpe Ratio, under a specific transaction costs scenario, is lower than the respective "Net" Sharpe Ratio of the combination of this particular trading system with one of the "Trending ones, i.e. the systems with numbers 30 and 32, correspondingly.

In Figures 3.29, we show the respective combinations' "Net" Sharpe Ratios under the 6 transaction costs' scenarios, of the "Mean-Reverting" trading systems with numbers 11,15,21 and 26, respectively, with the "Trending" system with number 30. The same corresponding results are displayed for the combinations of the same systems and the "Trending" system with number 32.

We report a successful "Resilience Gain" for both cases. More specifically, the "Trending" trading system, 30, greatly improves the overall trading performance when combined with the "Mean-Reverting" system 11, from the very beginning, contributing positive excess returns in the final aggregated portfolio, while "protecting" the overall performance even in the case of the worst possible transaction costs' scenario (Scenario 6), where while the purely "Mean-Reverting" system generated a highly negative "Net" Sharpe Ratio of -0.76, the combined strategies' "Net" Sharpe was almost "flat", namely -0.09, i.e. protecting as much as possible the capital risked in the respective investment. Interestingly, the same gain seems to be the outcome of the other combinations, even more easily observed under the most "costly" scenarios, as well. Even positively signed "Net" Sharpe Ratios were observed, as in the case of the "Mean-Reverting" system with number 15, in Figure , (b). This "Resilience Gain" though seems to come along with a respective information "cost" or "burden", related to the "foregone" Sharpe Ratio "lost" by the combined portfolio of strategies relative to the standalone "Mean-Reverting" ones, especially in the cases where the transaction costs' scenarios are "cheaper". This practically means that a trader or portfolio manager would actually gain the aforementioned "Resilience Gain" or "protection" against severe transaction costs scenarios, only in the cases where the transaction costs were "costlier" than a threshold level. In that case, the trade-off between this "foregone" gain would justify the "payment" of the respective information "burden", in terms of profitability and Sharpe Ratio score, against the trading performance of the respective standalone "Mean-Reverting" system itself. The same results and rationale applies on the combinations of the several "Mean-Reverting" trading systems with the "Trending" one with number 32, as well. As such, we could state that "Resilience" could be gained in the expense of higher risk undertaking by the trader.

Under this trade-off logic, we again cite the work of (Cuturi and D'Aspremont, 2016) on the same trade-off rationale between sparsity and volatility of the traded assets in their respective work, while we suggest exploring "protection" solutions or "work-arounds" of this trade-off on "Trending" trading systems like the ones designed and demonstrated in the current context. However, while this "protection" or "Resilience Gain" could be worth "co-traded" along with the respective top-performing "Mean-Reverting" strategies, the trader should be really careful on whether the associated "information cost" (loss of "foregone" Sharpe Ratio of the combined strategies' system, as opposed to single "Mean-Reverting" strategy trading) is justified under the transaction costs' scheme applied on his/her real market trading conditions (i.e. brokerage commisions, liquidity shortage, slippage etc.). We comment on this trade-off as the one which favorites "stronger" market participants in the "eyes of the market", e.g. an institutional investor from a retail individual trader, with the first being able to exploit "economies of scale" in order to reduce the marginal transaction costs for the target trading activity, while the latter, being "insignificant" in size and turnover levels, being incapable of exploiting the same transaction costs' schemes. This leaves the "weaker" player in a disadvantageous position relative to the first one. Future work could also shed light on the capacity and "scalability" of the respective trading strategies, while also further exploring and testing the validation of the argument that the "economies of scale" and the corresponding marginal transaction costs' reduction for the "more significant" players in the market, actually enables capital accumulation and distinguishes them against the more "insignificant" ones, in terms of profitability and competitive advantage in the information terrain.

3.5.22 Extending to the "Embed-Train-Lift" approach

The results of the previous framework are promising, even under a realistically set Transaction Costs environment, simulating real market conditions. This application acts as an upgrade of the previous approach, by targeting the extension of the previously identified as "best" (under the "tradability" rationale explained above) embedded portfolios' set. We believe that information-ally richer patterns in the FX Pairs' dynamics is present, but the linear nature of the PCA framework applied in the previous context, was not able to capture any potential nonlinearities hidden in the pairs' dynamics and/or any nonlinear dependencies of the behavior of each pair with all others. Otherwise stated, it failed to efficiently capture their "cross-connections", which might not be express-able via the underlying covariance matrix used in the PCA approach, assumed to be the correct affinity matrix related to the low-dimensional manifold. Towards that aim, a hybrid modelling framework is developed, in which additionally to the PCA and LLE, the Diffusion Maps (DM) manifold learning technique is on-boarded. Furthermore, the necessary for the overall trading framework embedding step is modified, in order to be compatible with the nonlinear nature of the selected embedding approaches. This powerful armory of potential nonlinear dependencies' capturing tools, will be combined again with time series forecasting tools, namely the MVAR and GPRs, with the latter being also modified compared to the previous approach. More specifically, it is constructed using a different kernel than the previous framework, namely a mix of kernels which are more associated with "Trending" dynamics, i.e. "smoother" behavior of the underlying modelled processes. As such, part of the mix is the previously mentioned as "max smoothness" convergent "squared exponential" kernel, as well. The reason for this particular choice is that, in case (hopefully) the several

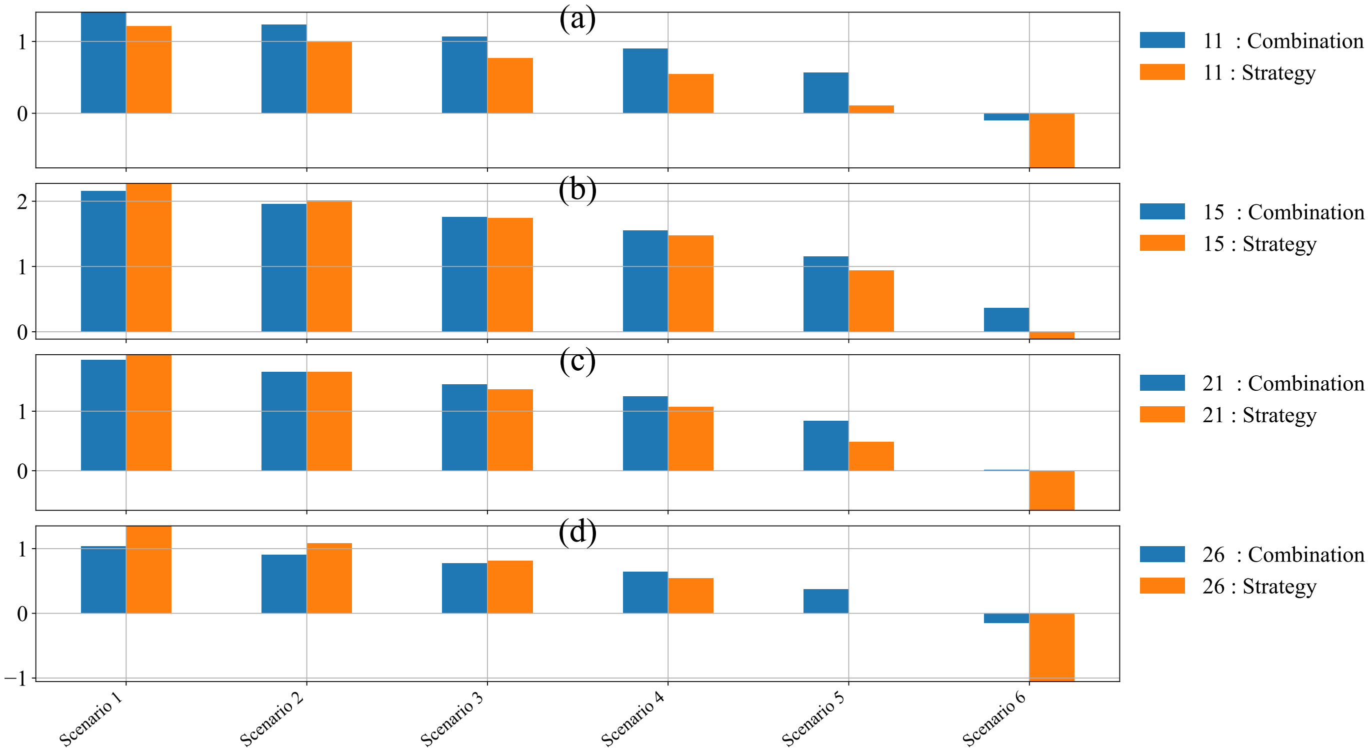

FIGURE 3.29: "Net" Sharpe Ratios of the "Mean-Reverting" trading systems, with numbers (a) 11, (b) 15, (c) 21 and (d) 26, as per Table 3.30 (orange colored bar-plot), against the respective results for the systems which combine them with the "Trending" trading system with number 30 referenced in the same Table (blue colored bar-plot)

potential nonlinearities present in the original space FX Pairs' dynamics and/or their corresponding cross-correlations, are adequately "learned" by our selected nonlinear manifold learning techniques, then, under the modified version of the trading framework, the embedded dynamics should be explainable by "smoother" (under e.g. mean square differentiability terms,Limits and Mean Square, 1973) stochastic processes, than the "noisier" ones obtained under the previous approach. As such, regarding the case of the GPR modelling, we observe that by moving towards values greater than 0.5 for the parameter ν of the Matern kernel, (with the critical level of $\frac{7}{2}$ yielding really noisy dynamics), higher-order differentiable underlying processes are expected to be obtained, i.e. "smoother" shapes of the underlying dynamics paths are anticipated. A key range for the below defined α parameter, which is connected to all of the processes under analysis, (fbm-matern discussed connection in section 3.5.12), but also with the quantities, hurst and fractal dimension of the previous sections, is the $[1/2, 3/2]$. At infinity, the "squared exponential" kernel is obtained, where maximum "smoothess" is yielded. Upon this logic, we again frame our "belief" on the appropriate kernel selection for this particular application (as we did for the previous one in the GPR kernel selection section 3.5.12, as well) and we form a mix of "smoother", "Trending" behavior targeting kernels. We expect the several medium-to-long term "cyclicalities" ("trends"), are associated with "long range dependence" on previous values of the process, to be identified by this approach and generate another set of significant under Sharpe Ratio terms trading strategies. Finally, we also incorporate a White Noise kernel in

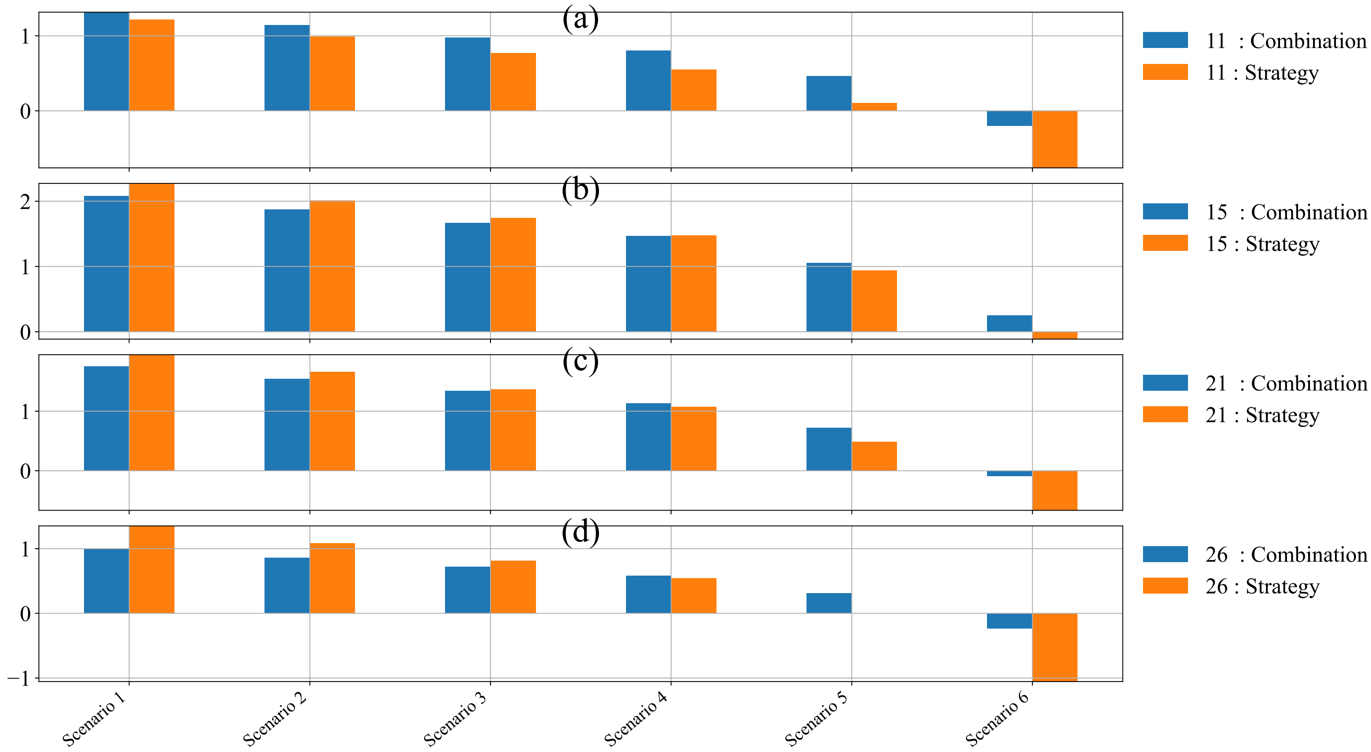

FIGURE 3.30: "Net" Sharpe Ratios of the "Mean-Reverting" trading systems, with numbers (a) 11, (b) 15, (c) 21 and (d) 26, as per Table 3.30 (orange colored bar-plot), against the respective results for the systems which combine them with the "Trending" trading system with number 32 referenced in the same Table (blue colored bar-plot)

the mix, in order to capture any potential affecting stochastic terms in the modelled processes (which under the rolling window approach outlined in the scheme, might appear and try to distort the already "learned" by the models' underlying data structure).

While the PCA embedding framework of the previous analysis starts with the $D \times D$ symmetric covariance matrix of the several assets in our dataset and for each specific rolling window under focus, the nonlinear manifold learning techniques counterparts used in this work, namely the LLE and Diffusion Maps, are built upon the $N \times N$ corresponding symmetric matrices associated with each method, which is specifically designed to optimize a particular geometric property assumed to be key one that optimally identifies the structure of the underlying low-dimensional manifold. This different approach on the target affinity matrices in our FOREX investing context can be interpreted from a trader's point of view, as one targeting on identifying interconnections between the several assets' correlations across time, or how the assets' returns at different time instances are correlated (or somehow "connected") with other time instances in the past. This way the latter approach handles and "weights" the several asset returns in different time instances differently than the linear approach of the traditional PCA. This approach was already performed and outlined in the Greek Electricity market application in the previous related section, where this cross-temporal interconnection between the several time instances' importance on explaining the underlying modelled System Marginal Price and Energy

Load was targeted. We refer to the embeddings using affinity matrices of dimensions $D \times D$ like the PCA framework of the previous section, as **"Spacial"**, while the one targeting the inter-temporal connection of the several returns' vectors in the different time instances, as **"Temporal"**. While the PCA framework is mostly documented on being performed under the "Spacial" embedding environment in the majority of the related to financial analysis, risk management, trading and portfolio management applications, due to its ability to offer a more intuitive essence on the produced outcomes, we also perform PCA under the previously defined "Temporal" manner in order to be able to compare its capacity and efficiency. Our base assumption here is, that if "Temporal" PCA performs adequately well or even outperforms the rest of the nonlinear techniques in this setup, then the underlying dynamics should not exhibit distressed nonlinearities or being perturbed by highly stochastic unpredictable components, and then simple linear prediction models like the ARIMA ones should be able to be sufficient on satisfactorily forecast it. We again compare the several embedding and forecasting tools' trading related results, against the traditional Long Only and Risk Parity portfolio construction approaches, as well as against each other.

3.6 Time Series Forecasting Using Manifold Learning : "Embed-Train-Lift" Scheme

Ref. Paper : Panagiotis Papaioannou, 2021

3.6.1 Application in synthetic datasets

For demonstrating the performance of the proposed numerical framework and comparing the various embedding, modelling and reconstruction approaches, we used three synthetic time series resembling EEG recordings produced by linear and nonlinear stochastic discrete models.

The synthetic time series

Our synthetic stochastic signals resemble EEG time series (see e.g. Baccalá and Sameshima, 2001; Nicolaou and Constandinou, 2016) produced by linear and nonlinear five-dimensional discrete stochastic models with white noise.

The linear five-dimensional stochastic discrete model is given by the following equations:

$$\begin{aligned}
y_t^{(1)} &= 0.2y_{t-1}^{(1)} - 0.4y_{t-1}^{(2)} + w_t^{(1)}, \\
y_t^{(2)} &= -0.5y_{t-1}^{(1)} + 0.15y_{t-1}^{(2)} + w_t^{(2)}, \\
y_t^{(3)} &= -0.14y_{t-1}^{(2)} + w_t^{(3)}, \\
y_t^{(4)} &= 0.5y_{t-1}^{(1)} - 0.25y_{t-1}^{(2)} + w_t^{(4)}, \\
y_t^{(5)} &= 0.15y_{t-1}^{(1)} + w_t^{(5)},
\end{aligned} \tag{3.66}$$

where for the training process the model order is assumed to be known (here equal to 1).

A second (nonlinear) model is given by the following equations (see e.g. Nicolaou and Constandinou, 2016):

$$
\begin{aligned}
y_t^{(1)} &= 3.4y_{t-1}^{(1)}(1-{y_{t-1}^{(1)}}^2)\exp(-{y_{t-1}^{(1)}}^2) + w_t^{(1)},\\
y_t^{(2)} &= 3.4y_{t-1}^{(2)}(1-{y_{t-1}^{(2)}}^2)\exp(-{y_{t-1}^{(2)}}^2) + 0.5y_{t-1}^{(1)}y_{t-1}^{(2)} + w_t^{(2)},\\
y_t^{(3)} &= 3.4y_{t-1}^{(3)}(1-{y_{t-1}^{(3)}}^2)\exp(-{y_{t-1}^{(3)}}^2) + 0.3y_{t-1}^{(2)} + 0.5{y_{t-1}^{(1)}}^2 + w_t^{(3)},\\
y_t^{(4)} &= 0.5y_{t-1}^{(1)} - 0.25y_{t-1}^{(2)} + w_t^{(4)},\\
y_t^{(5)} &= 0.15y_{t-1}^{(1)} + w_t^{(5)}.
\end{aligned}
\tag{3.67}
$$

Finally, the proposed scheme was also validated through the time series produced by a linear stochastic model with a model order greater than 1, given by the following equations:

$$
\begin{aligned}
y_t^{(1)} &= 0.1y_{t-1}^{(1)} - 0.6y_{t-3}^{(2)} + w_t^{(1)},\\
y_t^{(2)} &= -0.15y_{t-3}^{(1)} + 0.8y_{t-3}^{(2)} + w_t^{(2)},\\
y_t^{(3)} &= -0.45y_{t-3}^{(2)} + w_t^{(3)},\\
y_t^{(4)} &= 0.45y_{t-3}^{(1)} - 0.85y_{t-3}^{(2)} + w_t^{(4)},\\
y_t^{(5)} &= 0.95y_{t-2}^{(1)} + w_t^{(5)}.
\end{aligned}
\tag{3.68}
$$

In all the above models, the time series $w_t^{(i)}$, $i = 1,2,3,4,5$, are uncorrelated normally-distributed noise with zero mean and unit standard deviation.

3.6.2 Numerical Results

For the implementation of the numerical algorithms, we used the `datafold`, `sklearn` and `statsmodels` packages of Python Lehmberg et al., 2020; Pedregosa et al., 2011a; Seabold and Perktold, 2010b. The selection of the eigensolver was based on the sparsity and size of the input matrix: ARPACK was used if the size of the input matrix was greater than 200 and $n+1 < 10$ (where n is the number of requested eigenvalues), otherwise, the "dense" option was used. The ARPACK package Lehoucq, Sorensen, and Yang, 1998 implements implicitly restarted Arnoldi methods, using a random default starting vector for each iteration, with a tolerance of $tol = 10^{-6}$ and a maximum number of 100 iterations. This approach is implemented by the function `scipy.sparse.linalg.eigsh` of the `scipy` module (upon which the `sklearn` one depends) Virtanen, Gommers, and al., 2020. The "dense" eigensolver is implemented by the function `eigh` of the `scipy` module and returns the eigenvalues and eigenvectors computed using the LAPACK routine `syevd` using the divide and conquer algorithm Cuppen, 1980; Anderson et al., 1999. We used the default value of the tolerance of the Newton-Raphson iterations, which is of the order of the floating-point precision, and the maximum number of iterations was set to $30N$ iterations, where N is the size of the matrix.

The DM embedding on the training set of all data sets was performed using d parsimonious eigenvectors Dsilva et al., 2018. Here, for the construction of the graph at the first step, we used the standard Gaussian kernel, defined by

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2_{L_2}}{\sigma}\right), \tag{3.69}$$

where σ is a scaling parameter that controls the size of the neighborhood (or the connectivity of the graph). For its derivation, we follow the systematic approach provided by Singer et al., 2009. The full kernel is used for the calculations.

The VAR models were trained using the VAR class of the `statsmodels.tsa.vector_ar.var_model` routine, using the OLS default method for the parameters estimation. The corresponding LAPACK function used to solve the least-squares problem is the default `gelsd`, which exploits the QR or LQ factorization of the input matrix.

For our GPR regression model's computations, we have used a mixed kernel composed by adding four fundamental components (see e.g. Corani, Benavoli, and Zaffalon, 2021) :

1. the squared exponential kernel:

$$k_{SE}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \theta_1^2 \exp\left(-\frac{||\boldsymbol{x}_i - \boldsymbol{x}_j||^2_{L_2}}{2\theta_2^2}\right), \tag{3.70}$$

with $\theta_1 = \sigma_k$ the signal variance parameter, $\theta_2 = l$ the length-scale parameter and $\theta_3 = \sigma$ the noise variance parameter, as already examined in the theoretical framework earlier

2. a linear kernel:

$$k_L(\boldsymbol{x}_i, \boldsymbol{x}_j) = \theta_3^2 + \theta_4^2 \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle \tag{3.71}$$

, where θ_3 (usually expressed as σ_b) is a parameter determining how far from zero the function's height will be at the zero location and θ_4 (usually expressed as σ_v) another signal variance term specifically scaling the related dot product term.

3. a periodic kernel:

$$k_P(\boldsymbol{x}_i, \boldsymbol{x}_j) = \theta_5^2 \exp\left(-\frac{2\sin^2\left(\pi||\boldsymbol{x}_i - \boldsymbol{x}_j||_{L_2}/\theta_6\right)}{\theta_7^2}\right) \tag{3.72}$$

, where θ_5 the kernel's specific signal variance term, θ_6 (usually expressed as p) is the periodicity parameter, useful on determining periodical patterns in the data, and θ_7 the kernel's specific length-scale (l_p).

4. and a white noise kernel:

$$k_W(\boldsymbol{x}_i, \boldsymbol{x}_j) = \theta_8^2 \delta_{ij}. \tag{3.73}$$

dependent only the kernel's specific signal variance term θ_8.

and as such the final mixed kernel will be dependent on the 8 hyperparameters' set $\boldsymbol{\theta}$ outlined in the subsequent kernels : $k(\boldsymbol{\theta}) = k_{SE}(\theta_1, \theta_2) + k_L(\theta_3, \theta_4) + k_P(\theta_5, \theta_6, \theta_7) + k_W(\theta_8)$. The reason for selecting this particular kernel mix is to go beyond the previous "Mean-Reverting" targeting trading perspective in the FOREX markets, actively pursued by the selected Matern kernel and its inherent connection with the fOU process. More specifically, we expect this mix to be able to capture any potential medium-to-longer term "cyclicalities" and periodicities in the several pairs' returns dynamics and as such sufficiently extracting any additional "Trend" related factors.

The hyperparameters of the GPR model were optimized using the (default) L-BFGS-B algorithm of the `scipy.optimize.minimize` method Byrd et al., 1995; Zhu et al., 1997. The gradient vector was estimated using forward finite differences with the numerical approximations of the Jacobian being performed with a default step size $eps = 10^{-8}$. The iterations were stopped either when

$$\frac{(f^k - f^{k+1})}{\max\{|f^k|, |f^{k+1}|, 1\}} < 10^{-9},$$

f^k being the value of the loss function at step k, or when

$$\max_i |\text{proj}(g)_i| < 10^{-5},$$

$\text{proj}(g)_i$ being the i-th component of the projected gradient at the current iterate. We used the default values for the maximum number of function evaluations and the maximum number of iterations, i.e. 15,000, as well as the default value for the maximum number of line searches per iteration, i.e. 20.

For the lifting task, 50 nearest neighbors were considered for interpolation by all the methods. The underlying k-NN algorithm is based on the algorithm proposed in Maneewongvatana and Mount, 2002. Using different values of the number of nearest neighbors within the range 20-100 did not change the outcomes of the analysis. In the case of RBF interpolation, the underlying linear system of equations was solved by the LAPACK `dgesv` routine from `scipy.linalg.lapack.dgesv`, which implements the default method of the LU decomposition with partial pivoting. For the GH approach, we used the Gaussian kernel. In effect, we are performing “double" DM here - computing diffusion maps on the leading retained diffusion map components for the reduced embedding. This procedure, suggested in Chiavazzo et al., 2014 can actually be performed “once and for all" globally (Evangelou et al., 2021).

The computations were performed using a system with an Intel Core i7-8750H CPU @2.20 GHz and 16GB of RAM.

For both the linear and nonlinear models, we produced 2000 points. We used 1500 points for learning the manifold and for training the various models, and 500 points to test the forecasting performance of the various schemes. The forecasting performance was tested using the iterative mode, i.e. by training the models for one-step ahead predictions and then simulating the trained model iteratively to predict future values. The performance was measured using the Root Mean Square Error (RMSE) of the residuals, reading:

$$\text{RMSE} = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(\hat{\boldsymbol{x}}_i - \boldsymbol{x}_i)^2}, \tag{3.74}$$

where $\hat{\boldsymbol{x}}_i$ are the predictions and $\boldsymbol{x}_i$ the actual data. To quantify the forecasting intervals due to the stochasticity of the models, we performed 100 runs for each combination of manifold learning algorithms (DMs and LLE), models (MVAR and GPR) and lifting methods (RBFs and GHs), reporting the median and the 5-th and 95-th percentiles of the resulting RMSE. The RMSE values obtained with the naive random walk model, as well as with the MVAR and GPR models trained in the original space, are also given for comparison purposes.

In Table 3.31, we report the forecasting statistics of the time series produced with the linear model given by (3.66) as obtained over 100 runs. As it is shown, both the MVAR and GPR models trained in the original five-dimensional space outperform the naive random walk. The RMSEs of the MVAR and GPR models suggest a good match with the stochastic model, with the residuals being approximately within one standard deviation of the distribution of the noise level. In the same table, we provide the corresponding forecasting statistics as obtained with the proposed "embed-forecast-lift" scheme for the various combinations of manifold learning algorithms, regression models and lifting approaches.

TABLE 3.31: Linear stochastic model (3.66). RMSE statistics (median, 5-th and 95-th percentiles over 100 runs) for each of the five variables as obtained by: (a) training MVAR and GPR models with model order one in the original 5D space (MVAR(OS), GPR(OS)), and (b) the proposed "embed-forecast-lift" scheme for all combinations of the manifold learning algorithms (DMs and LLE) with two coordinates, models (MVAR and GPR) and lifting approaches (RBFs and GHs). For comparison purposes the RMSEs obtained with the naive random walk is also reported.

Model/Variable	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(3)}$	$y_t^{(4)}$	$y_t^{(5)}$
Random Walk	1.374 (1.287,1.466)	1.446 (1.333,1.528)	1.427 (1.339,1.497)	1.551 (1.455,1.654)	1.425 (1.335,1.512)
MVAR(OS)	1.151 (1.077,1.218)	1.180 (1.113,1.269)	1.010 (0.966,1.062)	1.224 (1.156,1.281)	1.011 (0.960,1.063)
GPR (OS)	1.151 (1.077,1.218)	1.179 (1.113,1.272)	1.010 (0.966,1.061)	1.224 (1.154,1.284)	1.011 (0.959,1.064)
DM-GPR-GH	1.153 (1.082,1.220)	1.184 (1.117,1.273)	1.012 (0.966,1.061)	1.228 (1.160,1.287)	1.013 (0.959,1.065)
LLE-GPR-GH	1.155 (1.077,1.218)	1.182 (1.115,1.272)	1.011 (0.966,1.065)	1.230 (1.157,1.292)	1.014 (0.962,1.064)
DM-GPR-RBF	1.260 (1.111,2.201)	1.390 (1.149,1.980)	1.163 (0.997,2.331)	1.365 (1.173,1.939)	1.166 (0.997,1.853)
LLE-GPR-RBF	1.197 (1.103,1.452)	1.234 (1.131,1.461)	1.076 (0.993,1.582)	1.270 (1.175,1.468)	1.103 (0.990,1.553)
DM-MVAR-GH	1.151 (1.078,1.220)	1.180 (1.113,1.269)	1.011 (0.966,1.063)	1.225 (1.155,1.288)	1.011 (0.959,1.063)
LLE-MVAR-GH	1.155 (1.077,1.218)	1.182 (1.115,1.271)	1.011 (0.966,1.065)	1.229 (1.158,1.293)	1.014 (0.962,1.064)
DM-MVAR-RBF	1.289 (1.124,2.132)	1.312 (1.146,2.018)	1.135 (0.996,1.635)	1.341 (1.186,1.933)	1.106 (0.99,1.601)
LLE-MVAR-RBF	1.187 (1.106,1.477)	1.230 (1.14,1.477)	1.080 (0.995,1.602)	1.264 (1.177,1.479)	1.096 (0.989,1.558)

For our illustrations, we have chosen the first two parsimonious DM coordinates, and the corresponding two LLE eigenvectors. As shown, the best performance is obtained with the GH lifting operator. Using GHs for lifting and any combination of the selected manifold learning algorithms and models outperforms all other combinations, thus resulting in practically the same RMSE values when compared with the predictions made in the original space. This suggests that the proposed "embed-forecast-lift" scheme applied in the 5D feature space provides a very good reconstruction of the predictions made in the original space. On the other hand, lifting with RBF interpolation with LU decomposition, generally resulted in poor reconstructions of the high-dimensional space, thus, in many cases, giving wide forecasting intervals that contained the median value of the naive random walk RMSE.

Next, in Table 3.32, we report the forecasting statistics of the time series for the nonlinear stochastic model (3.67) as obtained over 100 runs. As in the case of the linear stochastic model, both MVAR and GPR trained in the original 5D space outperform the naive random walk. The resulting RMSE values suggest a good match with the nonlinear stochastic model. Yet, the match is poorer than the one obtained for the linear model.

TABLE 3.32: Nonlinear stochastic model (3.67). RMSE statistics (median, 5-th and 95-th percentiles over 100 runs) for each of the five variables as obtained by: (a) training MVAR and GPR in the original 5D space (MVAR(OS), GPR(OS)) with model order one, (b) using the proposed "embed-forecast-lift" scheme for all the combinations of the manifold learning algorithms (DMs and LLE), models (MVAR and GPR), and lifting approaches (RBFs and GHs). For the embedding, three coordinates were used.

Model/Variable	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(3)}$	$y_t^{(4)}$	$y_t^{(5)}$
Random Walk	1.711 (1.597,1.815)	2.085 (1.939,2.258)	2.292 (2.147,2.437)	1.731 (1.62,1.819)	1.435 (1.354,1.527)
MVAR(OS)	1.181 (1.123,1.234)	1.438 (1.359,1.555)	1.565 (1.463,1.648)	1.212 (1.151,1.272)	1.021 (0.967,1.076)
GPR(OS)	1.182 (1.123,1.233)	1.437 (1.360,1.555)	1.684 (1.540,1.799)	1.213 (1.151,1.272)	1.020 (0.966,1.076)
DM-GPR-GH	1.181 (1.123,1.234)	1.440 (1.363,1.56)	1.573 (1.477,1.654)	1.214 (1.151,1.276)	1.022 (0.968,1.077)
LLE-GPR-GH	1.182 (1.126,1.236)	1.451 (1.363,1.563)	1.579 (1.473,1.675)	1.214 (1.151,1.273)	1.023 (0.97,1.076)
DM-GPR-RBF	1.488 (1.172,8.88)	1.618 (1.392,9.95)	2.060 (1.642,6.102)	1.710 (1.203,16.868)	1.199 (1.012,9.108)
LLE-GPR-RBF	1.214 (1.133,1.557)	1.449 (1.365,1.561)	1.587 (1.502,1.692)	1.254 (1.159,1.446)	1.088 (0.989,1.5)
DM-MVAR-GH	1.181 (1.123,1.234)	1.438 (1.365,1.555)	1.571 (1.472,1.649)	1.213 (1.151,1.274)	1.022 (0.968,1.076)
LLE-MVAR-GH	1.182 (1.125,1.235)	1.452 (1.364,1.563)	1.586 (1.468,1.669)	1.214 (1.152,1.275)	1.023 (0.97,1.076)
DM-MVAR-RBF	1.387 (1.169,6.743)	1.534 (1.368,4.634)	1.902 (1.609,4.5)	1.560 (1.199,9.294)	1.131 (0.99,4.655)
LLE-MVAR-RBF	1.213 (1.135,1.507)	1.445 (1.366,1.561)	1.570 (1.484,1.657)	1.253 (1.159,1.621)	1.091 (0.998,1.674)

We also provide the corresponding forecasting statistics as obtained with the proposed "embed-forecast-lift" scheme. For embedding in the reduced space, we have taken three (parsimonious) coordinates. Again, the best performance is obtained with the GHs lifting operator for any combination of manifold learning algorithms and models. Importantly, the reconstruction errors between the forecasts with the "full" MVAR and GPR models trained directly in the original 5D space and the ones obtained with the proposed "embed-forecast-lift" scheme are negligible up to a three-digit accuracy for all the five variables. As with the previous case, lifting with RBFs resulted in a poor reconstruction of the high-dimensional space, with forecasting intervals containing the median RMSE value of the naive random walk.

Finally, in Table 3.33, we report the RMSE statistics for the time series produced with the linear model with a model order three (see (3.68)) as obtained over 100 runs from (a) the naive random walk model applied to the original 5D space, (b) the MVAR models trained in the original 5D data set with model orders one (MVAR(1)) and three (MVAR(3)), and (c) the proposed "embed-forecast-lift" method with the embedding applied to the original 5D data set using DM and LLE for embedding with three coordinates. In the reduced-order space we have trained MVAR models with model orders 1 and 3 and used GHs for lifting. The best results were obtained

TABLE 3.33: Linear stochastic model with model order three (see (3.68)). RMSE statistics (median, 5-th and 95-th percentiles over 100 runs) for 100 simulations, for each of the 5 variables as obtained by: (a) training MVAR(1) and MVAR(3) models in the original 5D feature space, and (b) the proposed scheme with DMs and LLE, MVAR(1) and MVAR(3) models and GHs for lifting. The embedding with DMs and LLE was implemented using three coordinates.

Model/Variable	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(3)}$	$y_t^{(4)}$	$y_t^{(5)}$
Random Walk	2.138 (1.898,2.461)	2.910 (2.412,3.551)	1.930 (1.755,2.152)	3.499 (2.978,4.195)	2.477 (2.253,2.739)
MVAR(1)	1.629 (1.466,1.851)	2.121 (1.817,2.515)	1.382 (1.275,1.517)	2.567 (2.244,2.997)	1.840 (1.699,2.031)
MVAR(3)	1.611 (1.449,1.833)	2.092 (1.787,2.483)	1.371 (1.265,1.501)	2.531 (2.206,2.963)	1.824 (1.689,2.019)
DM-MVAR(1)-GH	1.630 (1.467,1.854)	2.121 (1.825,2.515)	1.383 (1.273,1.516)	2.569 (2.244,2.998)	1.843 (1.697,2.037)
LLE-MVAR(1)-GH	1.651 (1.472,1.905)	2.159 (1.832,2.567)	1.397 (1.281,1.545)	2.619 (2.269,3.080)	1.866 (1.705,2.084)
DM-MVAR(3)-GH	1.621 (1.455,1.839)	2.103 (1.791,2.484)	1.376 (1.267,1.509)	2.546 (2.217,2.972)	1.835 (1.694,2.028)
LLE-MVAR(3)-GH	1.649 (1.473,1.902)	2.149 (1.820,2.572)	1.393 (1.277,1.550)	2.608 (2.252,3.065)	1.862 (1.703,2.094)

when using the proposed scheme with DMs for embedding and a model order 3 for the training of the MVAR model in the corresponding manifold. Importantly, the implementation of DM-MVAR(3)-GH succeeds in reproducing quite well the results obtained by training MVAR with a maximum delay of three in the original space.

3.6.3 Application in FOREX Trading

In this FOREX trading application, we reduce the investable universe to the so called "Majors" FX Pairs : EUR/USD, GBP/USD, AUD/USD, NZD/USD, JPY/USD, CAD/USD, CHF/USD, SEK/USD, NOK/USD, DKK/USD.

The first reason for this restriction, relative to the previous application's dataset, is based on purely financial trading criteria. More specifically, at least on the daily trading frequency level that the two applications are focusing on, and by stating that the results obtained in the current context can be considered realistic, under "safely" associating a "near-to-zero" (or neglect-able) incurred transaction costs' environment (at least on an institutional trading level, as well), then for the selected "Majors" Pairs, Transaction Costs Analysis (TCA), as performed in Section 3.5.20, is not necessary, and as such is ommited from this approach. For the case where Emerging Markets currencies are included, as the previous application, market frictions as market Spread (bid/offer), charged commissions and also the several "carry" effects, might heavily distort the reported trading performance.

The second reason is data driven. More specifically, trying to tackle a "follow-up" question regarding the first reason of restriction outlined in the previous paragraph, on whether this restriction actually changes the underlying structure of the "new mix" of assets, and maybe makes it more "predictable" than the previous one, we refer to some of the excluded pairs' (like RUBUSD or CNYUSD) calculated Hurst exponents, as displayed in Table 3.26, indicating strongly-"persistent" or strongly-"anti-persistent" behavior relative to historical dynamics, easily captured by even linear tools, which as such infers "contributed" predictability to the overall mix, rather than the opposite. In that sense, we expect their exclusion from the available universe to make the forecasting task even harder, and as such highlighting even more the importance and value of the current approach towards achieving the FOREX trading and portfolio management related goals. Towards further justifying this view, we again refer on the Hurst values of the selected "Majors", as well, in Table 3.26, commenting that they are all close to the critical "noise" level of 0.5, even after the Risk Parity (RP) "filter" is applied (i.e. reducing volatility-stochasticity effects to some extend), with the highest reported value being 0.5776 for the CADUSD (under the RP), and the lowest being the 0.4818 for the JPYUSD. This indicates that no "persistence" or "dependence" on past dynamics is present for any of the target pairs. We expect such "dependence" to be present on a cross-asset basis, discover-able only by manifold learning based approaches as the ones outlined here, while in case it is linear, the PCA approach should be able to capture it, deflating the necessity and added value of the proposed nonlinear rivals. Regarding the "Spacial" PCA framework, as performed in the previous application, we also calculate and indicatively display the Hurst exponents of the constructed, by the PCA scheme performed under a rolling window of 250 trading days (a year), embedded portfolios, both in the individual coordinate based level, but also on the global one, in Table 3.34. No "significant" level is again reported, yielding again the same skepticism regarding its capacity to explain the potential existent cross-asset nonlinearities. The case of "Temporal" PCA is compared against the current nonlinear hybrid scheme, in this application's context below.

We show that as expected, the target cross-asset dependence is indeed nonlinear for this particular dataset, as well, with the hybrid scheme outperforming the PCA rival, in terms of trading efficiency. More specifically, we again assess the performance of the proposed forecasting and FOREX trading framework, under the annualized Sharpe (SH) ratio Sharpe, 1994 of a constructed risk parity portfolio. The returns are computed by (3.23).

As already outlined in the previous FX forecasting and trading framework, we consider the underlying dynamics of the FOREX market as in general non-autonomous, at least over a long time period. In principle, there are time-varying exogenous factors including macroscopic economic indices, social interaction and

TABLE 3.34: Hurst exponents and α values for the "Spacial" PCA embedded portfolios comprised of "Majors" FX Pairs only

Portfolio	Hurst	α
$y^{Majors}_{PCA,250,1}$	0.5502	1.0502
$y^{Majors}_{PCA,250,2}$	0.5701	1.0701
$y^{Majors}_{PCA,250,3}$	0.5483	1.0483
$y^{Majors}_{PCA,250,4}$	0.5066	1.0066
$y^{Majors}_{PCA,250,5}$	0.4955	0.9955
$y^{Majors}_{PCA,250,6}$	0.4971	0.9971
$y^{Majors}_{PCA,250,7}$	0.5353	1.0353
$y^{Majors}_{PCA,250,8}$	0.5261	1.0261
$y^{Majors}_{PCA,250,9}$	0.4969	0.9969
$y^{Majors}_{PCA,250,10}$	0.4778	0.9778
$Y^{Majors}_{PCA,250}$	0.5174	1.0174

mimesis (see e.g. Papaioannou et al., 2013, where machine learning has been used to forecast FOREX taking into account Twitter messages), but also seasonal factors and rare events (such as the recent COVID19 pandemic), which influence FOREX over time. This comes in contrast to the synthetic time series that we have examined here, and also other autonomous models that have been considered in other studies that served as benchmarks to assess the performance of the various schemes. Hence, to cope with the such changes of the FOREX market and in general of financial assets, one would set up a sliding/rolling window, then train models within the rolling window, and perform (usually) one-day ahead forecasts for trading purposes.

For our illustrations, we assessed the performance of the proposed scheme based on the previously defined risk parity trading strategy, using a 1-year (250 trading days) embedding rolling window and the last 50 or 100 days of the 250-day rolling window for training. The forecasting performance of the proposed scheme using DMs and LLE for embedding with 3 coordinates, MVAR and GPR for prediction, and GHs for lifting was compared against the naive random walk and the full MVAR and GPR models trained and implemented directly in the original space. A comparison against the linear embedding provided by PCA with the same number of principal components was also performed.

Figure 3.31 depicts the SH ratios obtained with the various methods. As it is shown, the proposed schemes using the DM and LLE algorithms for embedding outperform all the other schemes, when considering the same size of the training window. In particular, the highest SH ratios ($\sim$0.83) are obtained with the combination of LLE and MVAR within the 100-day training window, followed by the combination of DMs and GPR in the 100-day training window, resulting in an SH ratio ~ 0.76. The third (~ 0.73) and fourth (~ 0.72) larger SH ratios result again from the implementation of the DMs and MVAR within the 50-day and 100-day training windows, respectively. On the other hand, the naive random walk (black bar) resulted in a negative SH ratio ($\sim\ -0.42$). Negative SH ratios (of the order of -0.35) resulted also from the implementation of PCA with GPR (yellow bars) for both sizes of the training window, while for the 50-day training window, the combination of PCA with MVAR resulted in an almost zero (but still negative) SH ratio. With PCA, a (small) positive SH value (~ 0.23) was obtained with MVAR within the 100-day training window. The full MVAR and GPR models, trained and implemented directly in the original space, produced positive SH ratios, but still smaller than those of the DMs and LLE schemes. In particular, within the 50-day training window the full MVAR (GPR) model resulted in an SH ratio of ~ 0.39 (~ 0.04), while within the 100-day training window the full MVAR (GPR) model resulted in an SH ratio of ~ 0.68 (~ 0.3). Finally, we also tested the forecasting performance of the full MVAR and GPR models using the 250-day training window. Within this configuration, the MVAR model yielded an SH ratio of ~ 0.49, while the GPR model an SH ratio of ~ 0.2.

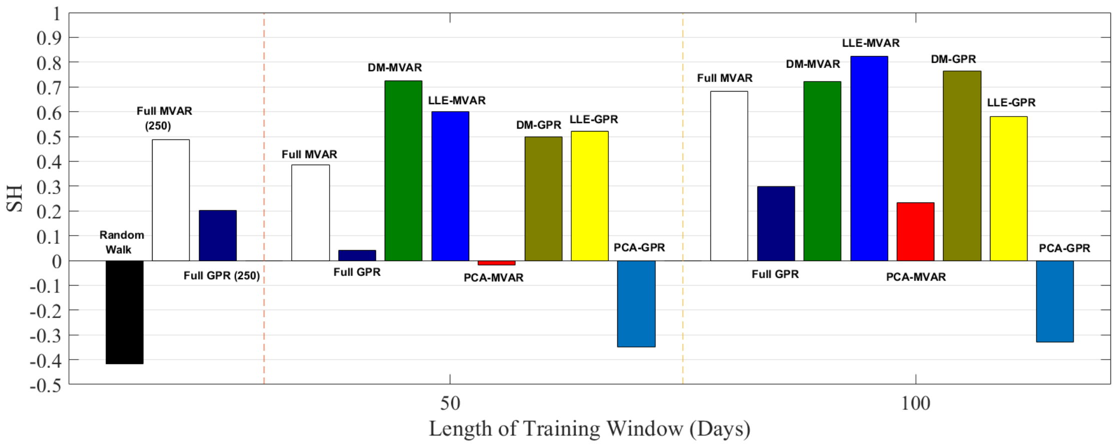

FIGURE 3.31: FOREX trading. One-day-ahead predictions. Sharpe Ratios obtained with the proposed framework (using DMs and LLE for embedding, MVAR and GPR for prediction at the embedded space, and GHs for lifting), as well as with PCA, random walk, and MVAR and GPR models implemented directly in the original space. A rolling window of 250 trading days and 3 vectors were used for the embedding, while the MVAR and GPR models in the embedded space were trained using the last 50 or 100 points of the rolling window.

3.7 Covid-19 pandemic forecasting

Background. Currently, Italy is overwhelmed by a second much stronger and wider wave, which sweeps through all of its regions. Since early November, Italy's health system is under a big pressure and a second national lockdown was reinstated on November 6. As a long and difficult winter is ahead, the accurate region by region now-casting of the pandemic arises as one of the most important and challenging tasks for the design of efficient regional public policies to mitigate the negative effects in both health and economy. Toward to this aim, we address a new machine learning model that on the basis of publicly available data now-casts accurately the evolution of the pandemic in regions of Italy.

Methods. We have designed a Gaussian Process Regression (GPR) model to nowcast and forecast the number of confirmed cases in ten regions of Italy. The training of the network was based on available data of the confirmed infected cases and google mobility reports for the period February 24-October 21 2020 of five metropolitan regions, namely, Lombardy, Lazio, Campania, Emiglia-Romagnia and

Piemonte. The proposed model takes as exogenous variables the percent change from the baseline prior to February 14 of the retail and recreation, park, transit station and workplace mobility and as delayed feedback output, the reported infected cases.

Findings. Based on the trained GPR, we were able to accurately now-cast the evolution of the confirmed infected cases from February 25, i.e. the beginning of the pandemic in Italy, until now, November 28, 2020. Furthermore, we were able to accurately nowcast the dynamics of the pandemic not only in the five regions of Italy, but also in other five regions (Veneto, Friuli-Venezia, Abruzzo, Tuscany and Sicily). Importantly, our findings suggest that one can accurately nowcast and therefore assess the development of the disease by monitoring the change in the mobility of the four above activities.

3.7.1 Motivation

Italy has been ravaged by the pitiless first wave of the novel coronavirus disease (COVID-19) pandemic being after China the world's epicenter for more than two months (March-April 2020). After 44 days of a draconian national-level lockdown, on May 4 when the government announced the relaxation of the measures and the gradual re-opening of activities, the total number of confirmed cases were 211,938 while the death toll was 29,079. During this period, Italy's north was hit hard: in Lombardy, where the first confirmed case of the disease was reported on February 21, more than 14,000 died. On the other hand, the southern Italy experienced a much less severe situation as the spread of the disease from then north was controlled due to the strict intra-regional mobility restrictions. On June 3, the date which Italy became the first European country to reopen its borders to european countries, the pandemic seemed to be under control. After the summer period during which some regions went for days without a case, currently, Italy is overwhelmed by a second much stronger and wider wave, now sweeping through all of its regions. As of November 28, the total cases raised up tp Since early November, the health system was already under a big pressure. Thus, a second national lockdown was reinstated on November 6, now based on a three-tier color code (red (high risk), orange (medium risk), and yellow (low-risk)) for dividing regions according to the intensity of the pandemic. These classifications are updated on a weekly basis according to each region's situation. As of November 28, the regions are divided as follows: red zones: Tuscany, Abruzzo, Campania, Valle d'Aosta, Bolzano; orange zones: Puglia, Calabria, Lombardy, Piedmont, Emilia Romagna, Marche, Umbria, Basilicata, Friuili-Venezia-Giulia; yellow zones: Liguria, Sicily, Sardinia, Lazio, Molise, Veneto, Trento.

While the world waits for the first doses of the vaccine, as a long and difficult winter is ahead, the accurate now-casting of the pandemic, arises as one of the most important and challenging tasks for the design of efficient regional public policies to mitigate the negative effects in both health and economy. Toward to this aim, mathematical models and systems theory are valuable tools for policy makers, Bertozzi et al., 2020. Thus, since the first reported case of the COVID-19 disease in China on January 11, there was an intense research effort for the development of new epidemiological mathematical models and tools to deal with the uncertainties that the newly emerged pandemic posed Bertozzi et al., 2020. Models can be categorized in different levels of increasing realism including very simple simple exponential growth models for the estimation of the basic reproduction number R_0 (see e.g. Zhao et al., 2020; Remuzzi and Remuzzi, 2020), mechanistic SEIR-like compartmental models (Imai et al., 2019; Anastassopoulou et al., 2020; Giordano et al., 2020; Ming, Huang, and Zhang, 2020; Leung et al., 2020) for the estimation of the effective reproduction number, forecasting and estimating the level of under-reporting of infected cases Russo et al., 2020, metapopulation models Wu, Leung, and Leung, 2020; Kucharski et al., 2020; Yuan et al., 2020; De Salazar et al., 2020; Chinazzi et al., 2020; Gilbert et al., 2020; Rossa et al., 2020 incorporating details of travel network and mobility data, and stochastic agent-based models Hoertel et al., 2020; Silva et al., 2020 that take into account details such as demographics of age and income and health, transportation and health infrastructure of the simulated area, to access the health and economic effects of social distancing interventions.

Recently, the large amount of data on COVID-19 that have been produced and made publicly available has enabled the development of numerous machine learning algorithms, thus resulting to numerous studies Bachtiger, Peters, and Walsh, 2020 mainly for clinical prediction purposes Wang et al., 2020; An et al., 2020; Yan L., 2020; Yadaw et al., 2020; Gao et al., 2020, but also the estimation of the parameters of compartmental SEIR-like models Deng, 2020 and the forecasting of the disease's dynamics as stand-alone models Rustam et al., 2020; Chimmula and Zhang, 2020; Zeroual et al., 2020).

However, all of these studies were restricted to very short-time forecasts. In particular, Rustam et al., 2020 used four techniques of machine learning, namely linear regression, LASSO, Support Vector Machine (SVM) and Exponential Smoothing to provide a 10-day forecast of the confirmed cases. Chimmula and Zhang, 2020 used a Long short-term memory (LSTM) network to provide an up to 14-days forecast for Canada with Italy and USA based on the available data until March 31, 2020. Zeroual et al., 2020 used five methods, namely, Recurrent Neural Network (RNN), Bidirectional LSTM (BiLSTM), Gated recurrent units (GRUs) and Variational AutoEncoder (VAE) algorithms to provide a 17-days forecast for the evolution of the pandemic in Italy, Spain, France, China, and the USA based on the available data until June 1, 2020. Katris, 2021 used five methods, namely Exponential Smoothing models, Auto Regressive Integrated Moving Average (ARIMA), Multivariate Adaptive Regression Splines (MARS) and Feed-Forward Artificial Neural Networks (ANN) to provide a 7-day forecast for Greece based on the reported cases until August 14, 2020.

Here, we perform Gaussian Process Regression (GPR), specifically targeting on the prediction of the evolution of the 7-day moving averaged dynamics of the several italian regions infected cases levels. The predicted values are also displayed in the web application #CombatCovid19 Monitor (3.32), as part of the COMBAT-COVID19 research project funded by the Fondo Integrativo Speciale per la Ricerca (FISR). Other metrics like the R_t effective reproduction numbers, Systrom, 2017, as well as any other statistic published by the official authorities, updated on a daily basis, as well.

3.7.2 GPR Forecasting

Towards formulating the GPR forecasting problem for the evolution of the 7-day moving averaged dynamics of the pandemics in the several regions of Italy, we denote as $I^{(7)}_{j,t}$ the 7-day moving averaged infected cases level at time t for the j-th region out of the $j = 1, 2, \dots, 21$ regions analysed. Moreover, we denote as $\boldsymbol{I}^{(7)}$

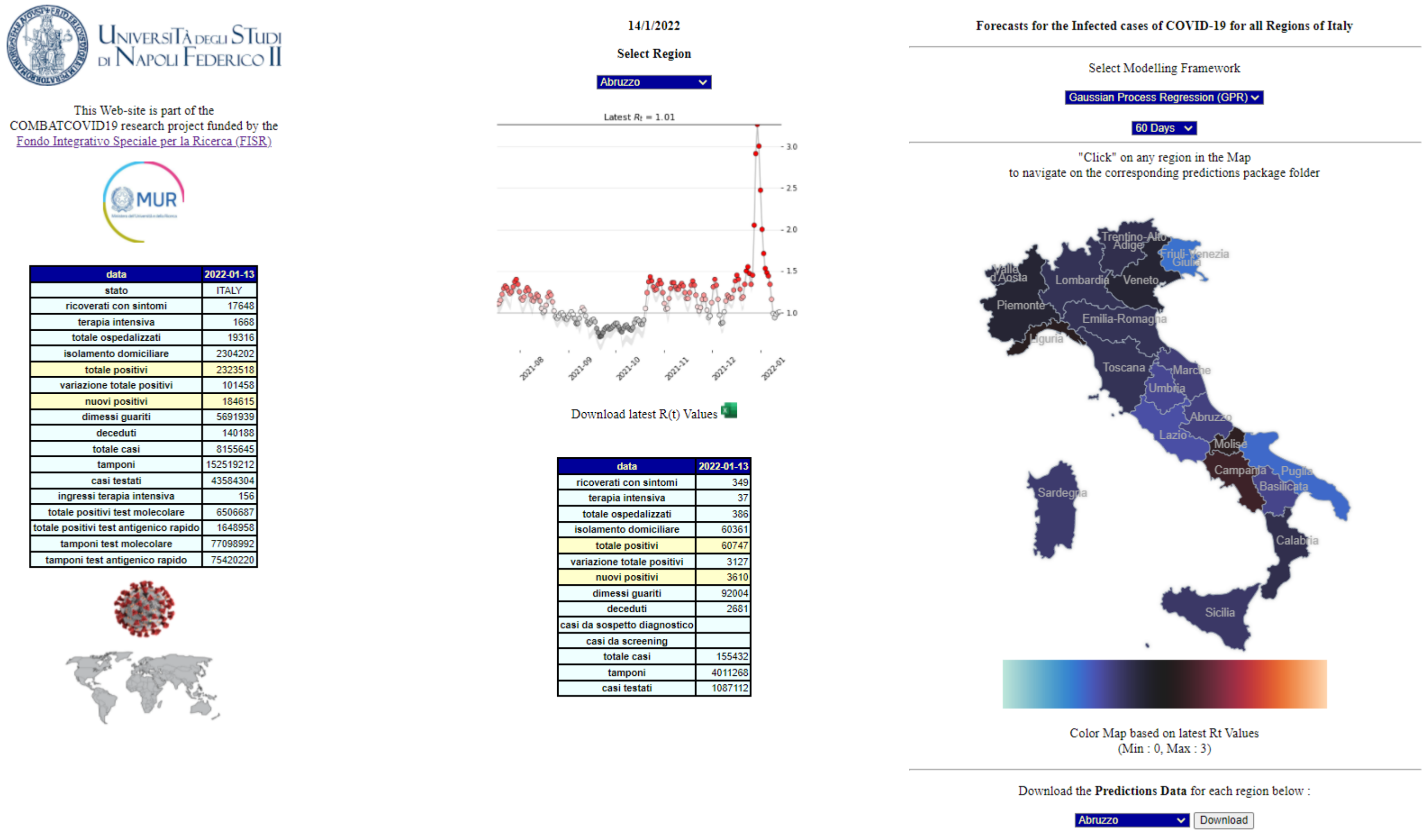

FIGURE 3.32: Snapshot of the Unina/FISR "Combat Covid 19" Monitor web platform

the matrix containing all regions' corresponding time series as columns. The predictions are then produced by regressing the several regions' $I_{j,t}^{(7)}$ time series (the targets) with "lagged" versions of matrix $\boldsymbol{I}^{(7)}$, and more specifically for the selected lags $L = 1, 2, 3, 5, 7$.

As such, we formulate the matrix $\boldsymbol{M_t} = [\boldsymbol{I}_{t-1}^{(7)}, \boldsymbol{I}_{t-3}^{(7)}, \boldsymbol{I}_{t-5}^{(7)}, \boldsymbol{I}_{t-7}^{(7)}]$ with dimensions $T \times 105$ (5 lags multiplied by the 21 regions in the dataset), containing the inputs (features) of the regression.

According to this logic, the forecasting problem can be expressed as :

$$I_{j,t+1}^{(7)} = \phi_j(\boldsymbol{M_t}), \quad j = 1, 2, \ldots, 21 \tag{3.75}$$

with the ϕ_j function being approximated in terms of the posterior distribution distribution of the GPR framework, as per equation 2.42. As such, the expected value and covariance of the corresponding predictions, will be given as :

$$\hat{\phi}_j(I_{j,t+1}^{(7)}) = k(I_{j,t+1}^{(7)}, \boldsymbol{M_t}|\boldsymbol{\theta})(\boldsymbol{K}(\boldsymbol{M_t}, \boldsymbol{M_t}|\theta) + \sigma^2 \boldsymbol{I})^{-1} I_{j,t}^{(7)} \tag{3.76}$$

$$\sigma_*^2 = k(I_{j,t+1}^{(7)}, I_{j,t+1}^{(7)}|\boldsymbol{\theta}) - k(I_{j,t+1}^{(7)}, \boldsymbol{M_t}|\boldsymbol{\theta})[\boldsymbol{K}(\boldsymbol{M_t}, \boldsymbol{M_t}|\boldsymbol{\theta}) + \sigma^2 \boldsymbol{I}]^{-1} k(\boldsymbol{M_t}, I_{j,t+1}^{(7)}|\boldsymbol{\theta}) \tag{3.77}$$

Again the hyperparameters set $\boldsymbol{\theta}$ is given by optimizing the marginal log-likelihood function (section 2.1.7):

$$l = -logp(I_{j,t}^{(7)} | \boldsymbol{M_t}, \boldsymbol{\theta}) \tag{3.78}$$

For the kernel selection, we follow the logic outlined in section 3.5.12 by observing that the modelled time series is a "smoothed" (moving-averaged) version of the raw infected cases, i.e. a good choice should be one of the "smoothness" targeting Matern kernels (ν parameter being larger than the 0.5 critical level discussed again in section 3.5.12). We set $\nu = \frac{3}{2}$ and we mathematically express the kernel for two points in the features containing matrix $\boldsymbol{M_t}$, $\boldsymbol{m}_i, \boldsymbol{m}_j$ as :

$$k_{\frac{3}{2}}(\boldsymbol{m}_i, \boldsymbol{m}_j) = \left(1 + \frac{\sqrt{3}\|\boldsymbol{m}_i - \boldsymbol{m}_j\|_{L_2}}{l}\right) exp\left(-\frac{\sqrt{3}\|\boldsymbol{m}_i - \boldsymbol{m}_j\|_{L_2}}{l}\right) \tag{3.79}$$

The final step is to iteratively produce multi-step predictions ahead in time, by appending the predicted values (prediction vectors) in the features matrix and re-calculating (re-fitting and re-predicting) any subsequent forecasts $I_{j,\tau}^{(7)}$, $\tau > t$ for a specific time horizon h. In our case, the predicted horizon is set to $h = 5$ days. The above process is being performed under a rolling window approach, meaning the length of points forming the features matrix each time, is limited to a past history of 60, 90 and 120 days, i.e. 2,3 and 4 months of historical daily data.

Indicatively, in Figures 3.33, 3.35 and 3.38, we display the "good" or "satisfactory" norm-plots of the GPR predictions corresponding to the regions Veneto, Umbria and Lombardia, while in Figures 3.39 and 3.41 the same norm-plots are shown for the "bad" or "very bad" predictions corresponding to regions Toscana and Piemonte, respectively. The reason for the difference in the efficiency of the several GPR models in the several regions, is that in some of them, sudden "gaps" in the infected cases' level was experienced, due to e.g. lockdown policies applied in them, and as such the "smoothness" targeting covariance kernels of the models were not able to predict those "shocks" in the dynamics effectively. Also for demonstrative purposes, the latest actual predictions associated with these regions are displayed in Figures 3.34, 3.36, 3.38, 3.42, as well. All Figures displayed results are associated with the 60 days rolling window GPR prediction framework. Daily updates of the several predictions and associated norm-plots can be found in the "#combatcovid19uninafisr", 2021 Monitor app (3.32), as well.

3.7.3 Effective Reproduction Number R_t

Further to the predictions regarding the cases dynamics in the several regions of Italy, the #CombatCovid19 Monitor app also calculates the widely accepted as one of the main tools towards controlling pandemic crises Bettencourt and Ribeiro, 2008, effective reproduction number R_t. R_t estimates the number of people who become infected at time t per each already infected person in a given population. It is a dynamically calculated metric, adapting usually in real time (daily) to any new information about the infected cases' level. Critical level of the R_t is being around 1, with lower values suggesting "normality" and values greater than 1 triggering alerts that a large part of the population might be infected soon.

In our case, we follow the work of Bettencourt and Ribeiro, 2008 as outlined by Systrom, 2017, towards calculating the levels of the R_t on a daily basis. The suggested model is based on the Bayesing approach and more specifically on Baye's rule :

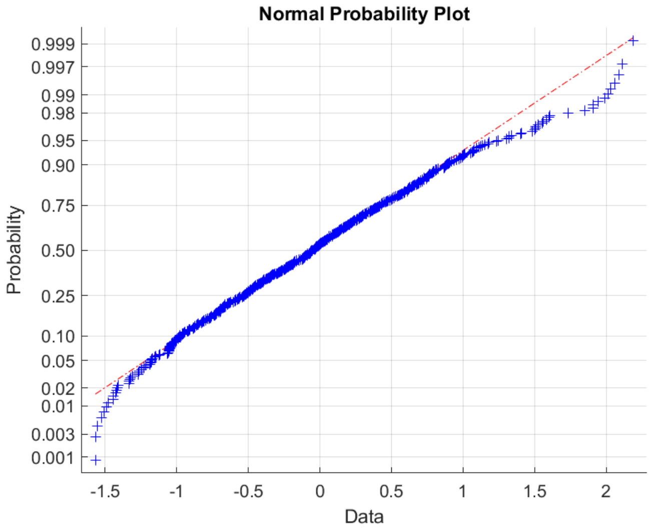

FIGURE 3.33: Norm-plot corresponding to the 5-days GPR predictions for Veneto.

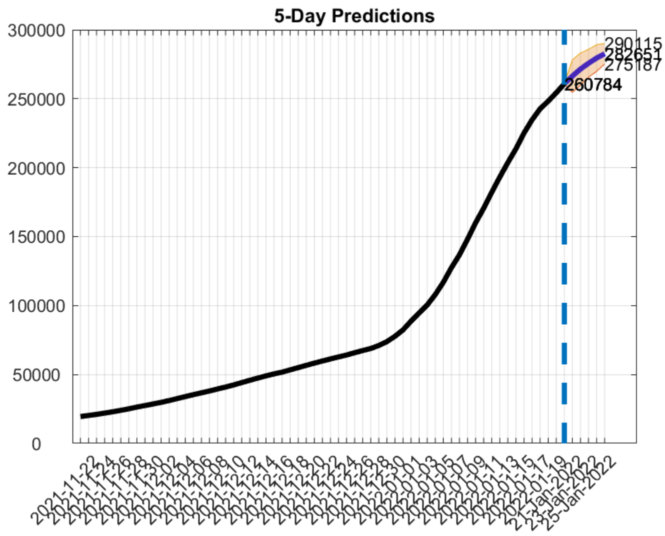

FIGURE 3.34: The latest 5-days GPR predictions for Veneto.

$$P(R_t, k) = \frac{P(k|R_t)P(R_t)}{P(k)} \tag{3.80}$$

with $P(R_t, k)$ being the likelihood function, $P(R_t)$ the prior, k the observed cases ("ground truth").

Upon this formalisation, an iterative approach is followed, based on the assumption that $P(R_t|R_{t-1}) = \mathcal{N}(R_{t-1}, \sigma)$, formulated as :

$$\begin{cases} P(R_1|k_1) \propto P(R_1)l(R_1|k_1), & \text{for day 1} \\ P(R_2|k_1, k_2) \propto P(R_2)l(R_2|k_2) = \sum_{R_1} P(R_1|k_1)P(R_2|k_2)l(R_2|k_2), & \text{for day 2} \\ \ldots \text{etc} \end{cases} \tag{3.81}$$

where the chosen likelihood function is set to be the Poisson Distribution with

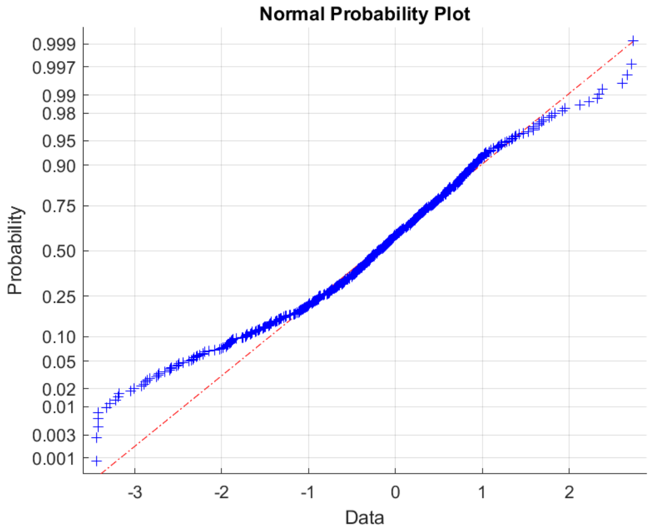

FIGURE 3.35: Norm-plot corresponding to the 5-days GPR predictions for Umbria.

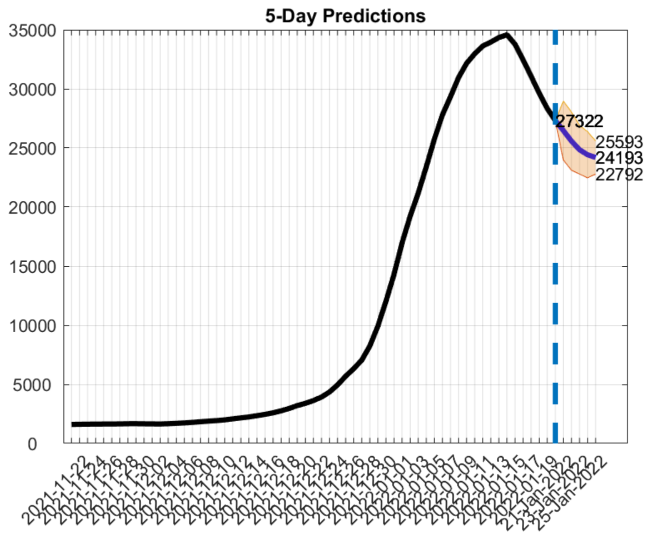

FIGURE 3.36: The latest 5-days GPR predictions for Umbria.

an average arrival rate of $\lambda = k_{t-1}e^{\gamma(R_t-1)}$, expressing the potential new cases per day and the parameter γ being defined in Sanche et al., 2020. It is expressed as :

$$P(k|R_t) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{3.82}$$

Indicatively, we show the calculated R_t values for the previous regions in Figures 3.43, 3.44, 3.45, 3.46 and 3.47, as well.

The calculated R_t values for all regions on 20-01-2022, are displayed in Table 3.35. As observed by the red-colored highlighted values for several regions, as well as from the recent dynamics of the R_t levels in the previous Figures, the situation in Italy is still "risky", while though declining towards "normality" trends are shaped at a slow pace.

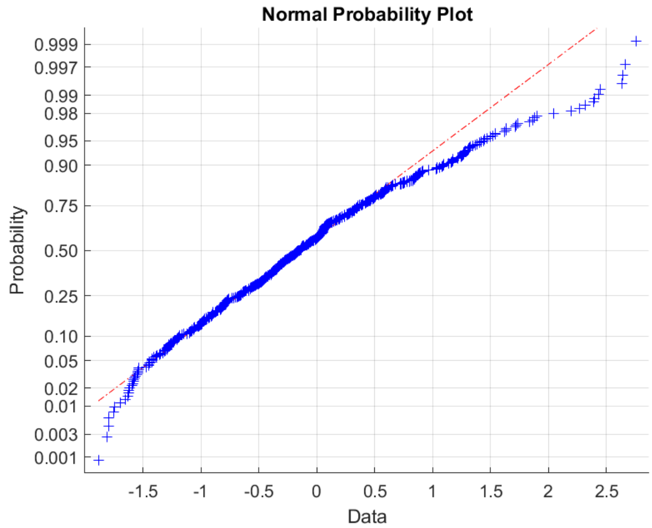

FIGURE 3.37: The latest 5-days GPR predictions for Lombardia.

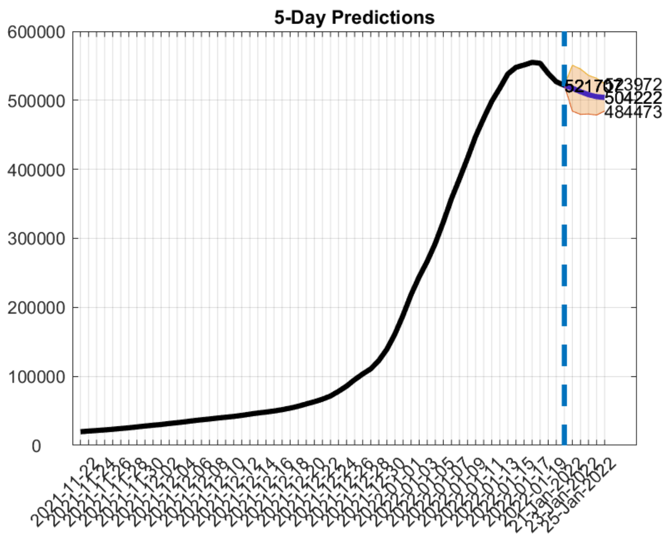

FIGURE 3.38: The latest 5-days GPR predictions for Lombardia.

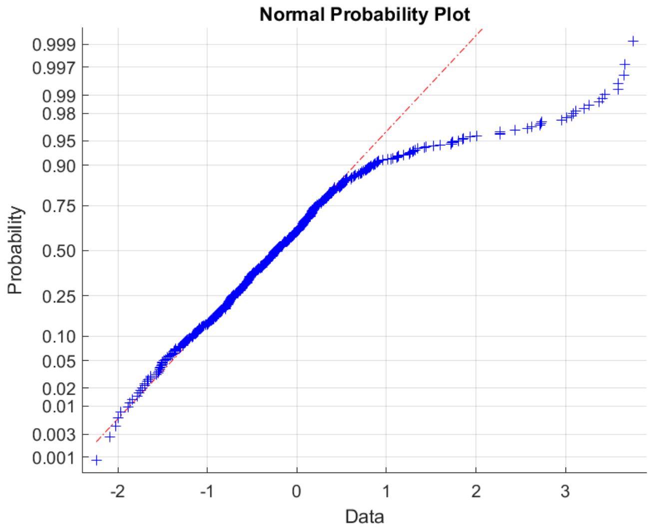

FIGURE 3.39: The latest 5-days GPR predictions for Toscana.

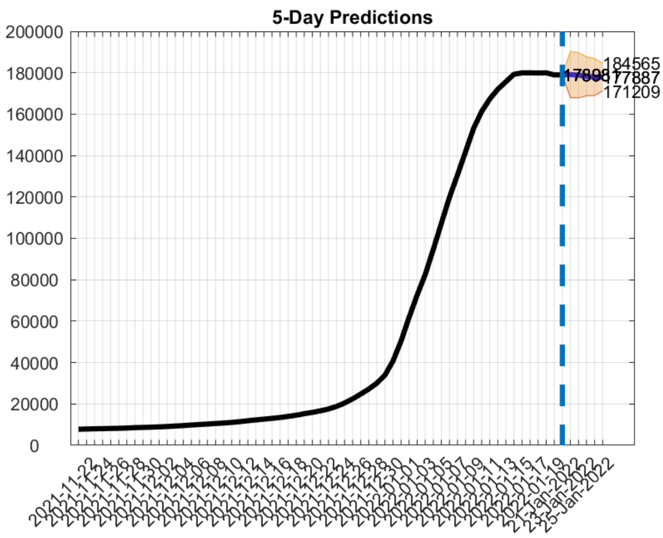

FIGURE 3.40: The latest 5-days GPR predictions for Toscana.

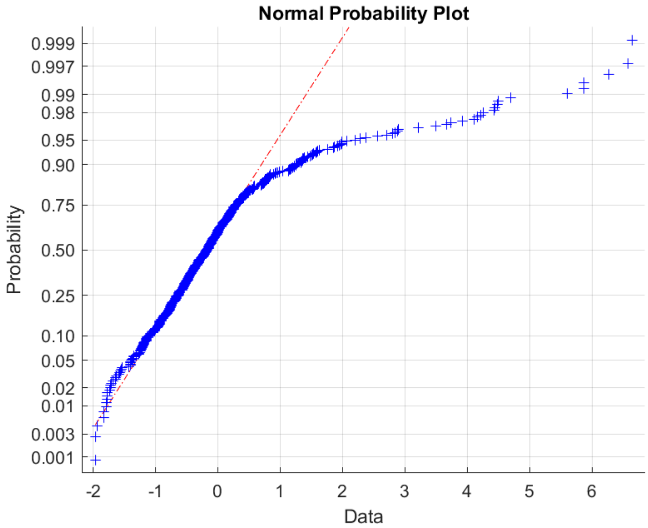

FIGURE 3.41: The latest 5-days GPR predictions for Piemonte.

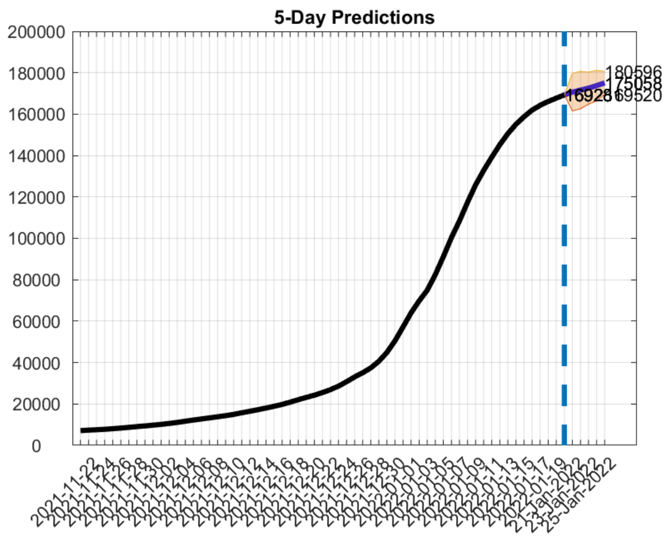

FIGURE 3.42: The latest 5-days GPR predictions for Piemonte.

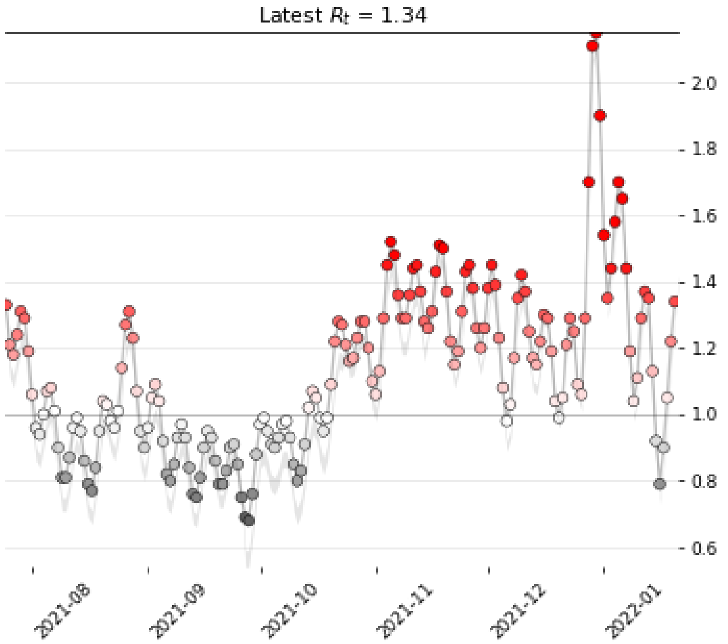

FIGURE 3.43: R_t values for Veneto.

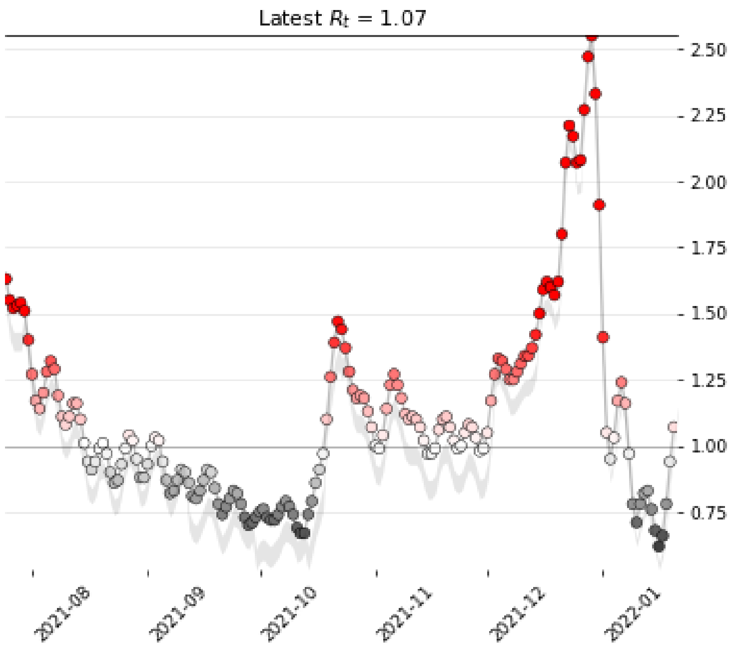

FIGURE 3.44: R_t values for Umbria.

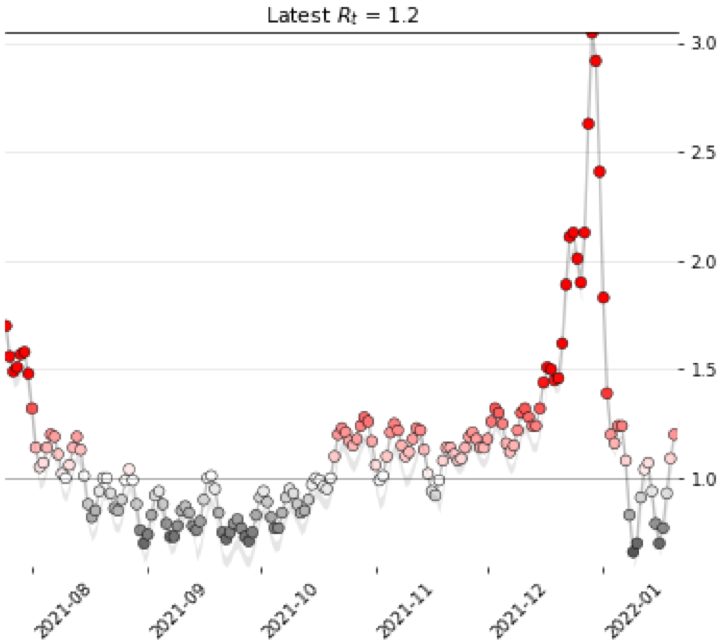

FIGURE 3.45: R_t values for Toscana.

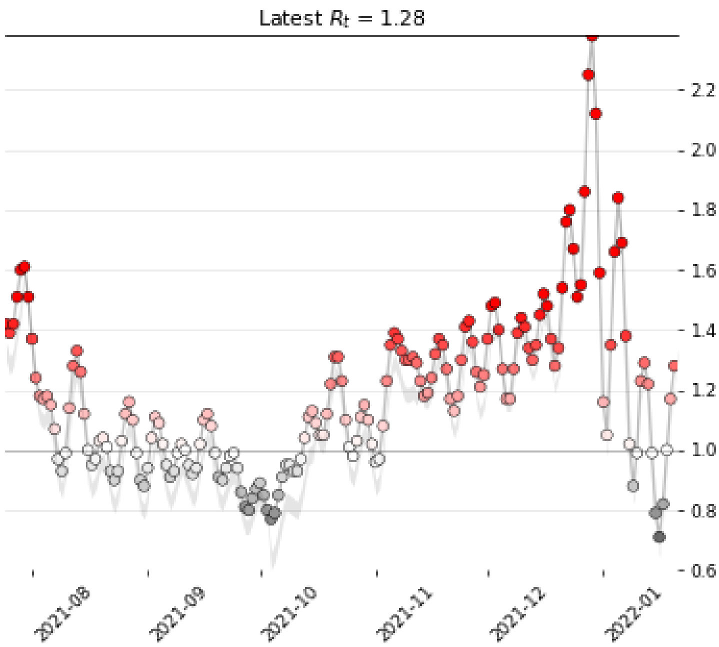

FIGURE 3.46: R_t values for Piemonte.

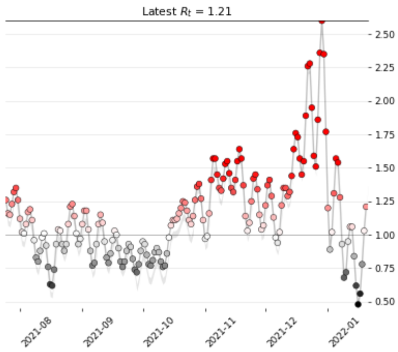

FIGURE 3.47: R_t values for Lombardia.

TABLE 3.35: All regions' R_t values as per 20-01-2022. Highlighted with red color are the R_t values greater than the critical level of 1.

Region	R_t
Calabria	0.85
Toscana	1.2
Piemonte	1.28
Abruzzo	1.14
Molise	0.85
Marche	1.39
Lazio	1.07
Puglia	1.13
Basilicata	1.46
Umbria	1.07
Friuli Venezia Giulia	0.7
Liguria	1.3
Valle d'Aosta	0.89
P.A. Bolzano	1.36
Sicilia	0.85
P.A. Trento	1.25
Campania	0.98
Lombardia	1.21
Veneto	1.34
Emilia-Romagna	1.1
Sardegna	0.94

Chapter 4

Conclusions and Future Research

We proposed a numerical approach based on manifold/machine learning for forecasting of high-dimensional time series. The proposed scheme is composed of three basic steps: (a) embedding of the original high-dimensional data in a low-dimensional manifold, (b) construction of reduced-order machine-learning/based models and in particular Gaussian Process and forecasting on the manifold, (c) reconstruction of the predictions in the high-dimensional space.

As mentioned in the introduction, the task of forecasting is different from that attempted for the reconstruction of high-dimensional models of dynamical systems in the form of ODEs or PDEs, in four main aspects. First, for real-world data sets, the existence of a relatively smooth low-dimensional manifold is not guaranteed as in the case of well-defined dynamical models. Second, non-stationary dynamics, which in general are not an issue when dealing with the approximation of dynamical systems, pose a major challenge for a reliable/consistent forecasting. It should be noted that the stationarity assumption may be required to hold true even for interpolation problems. For example, the implementation of the MVAR models and Gaussian Process with a Gaussian kernel requires that the stationary covariance function assumption is satisfied (see e.g. the discussion in Rasmussen, 2003; Cheng et al., 2015). Third, when dealing with real-world time series, such as financial time series, the number of available snapshots (even at the intraday frequency of trading) are limited, in contrast to the size of temporal data that one can produce by model simulations. In such cases, the quest for beating the "curse of dimensionality" using the correct (parsimonious) embedding and modelling is stronger. Finally, in the case of smooth dynamical systems, as the dynamics emerges from the same physical laws as expressed by the underlying ODEs, PDEs or SDEs, what is usually sought is a single global manifold (a single geometry). However, in many complex problems such as those in Finance, the parametrization of the manifold (if it exists) may change over time. Thus, one would now seek for a family of (sequential in time) sub-manifolds.

Here, we assessed the performance of the proposed numerical scheme by implementing and comparing different combinations of manifold learning algorithms (LLE and DM), regression models (MVAR and GPR) and reconstruction operators (SI, RBF and GH) on various problems uncluding synthetic stochastic data series and a real-world data set of FOREX time series. By doing so, we first showed that for synthetic time series the proposed "embed-predict-lift" scheme, and in particular the one implementing DM for embedding and GH for lifting, reconstructs almost perfectly the predictions obtained in the high-dimensional space with the full regression models. We showed and explained why RBF operators should be in general avoided for lifting, as they may result to an ill-posed problem.

To this end, we conclude with two remarks. The first one is that in order to cope with the generalization property and the topological instability issues (see e.g. Balasubramanian et al., 2002) that arise in implementing kernel-based manifold learning algorithms when the data set is not dense on the manifold, and/or in the presence of "strong" stochasticity, one can resort to techniques such as the constant-shift one to appropriately adjust the graph metric and techniques for the removal of outliers from the data set (see e.g. Choi and Choi, 2007; Wang, 2012). We intend to explore the efficiency of such approaches in a future work. Finally, we note that one can attempt to link the proposed framework with Takens embedding and Whitney immersion theorems Takens, 1981; Whitney, 1944, e.g. by creating an extended data set which contains shifted copies of the original time series. Towards this aim, very recently the Koopman operator and DM with a time delay embedding were used in Lehmberg, Dietrich, and K′oster, n.d. to model the dynamics of pedestrian traffic in Melbourne. Below, we comment challenges and possible future work on the applications that this Thesis confronted.

4.0.1 On the Financial Markets Applications

A long-lasted debate regarding works as the current Thesis, related to the inferred "predictability" of the several time series analysed in the framed applications, and more specifically on whether it is "robust" or "preserved" as time progresses. Can we infer that the patterns extracted under the proposed schemes, will also be exploitable in the future, as well? What could change? How the models react or should react to such changes? Scepticism related to such logic, mathematically is related to notions like "invariance" and "conservation laws", potentially extensible in practical applications, as well, while this is not always feasible. In financial theory, this relates to the challenged by the Thesis, Efficient Market Hypothesis (EMH), respectively. Are the models stating "superiority" and increased ability to time the markets effectively able to "adapt" in order to preserve their status as successful in potential market structure's changes? Adding on this direction, a potential layer of arguments would be connecting this to the Adaptive Market Hypothesis of Andrew Lo, Kaserer, 2019, which infers "adaptability" and market "evolution" as progressing into the future, with agents beliefs' changing and evolving. Under that perspective, will any hidden patterns possibly extracted from "adapting" models, that now seem to be invalidating the EMH, be present in future market structures, as well ? On the "Quant" counter-debating side, "adaptable" machine learning based approaches are being conceptualised as vehicles incorporating this evolution in their parameterizations and associated "learning" mechanisms. While linear modelling, as a tool towards explaining the highly complex and increased stochasticity related market structures, with millions of participants competing each other (or even more interestingly aligning and acting in competing groups), injecting tones of hardly quantifiable information in the financial system, is the first to suffer from the "curse of dimensionality" concept analysed in the Thesis, nonlinear tools and machine learning revolution, seem to be working towards "escaping" this pitfall, with the concepts of dimensionality reduction and embedding.

My view on the topic, is associated with the notions of "control", "stability" and "security" in financial systems, acting as some of the main principles upon which they operate. More specifically, regulatory authorities and states (i.e. the financial stability's "guards" and "ensurers") in their effort to control money supply (central banking policies - interest rates setting), impose regulatory limitations and even propose policies "incentivising" targeted investing allocations and positioning in specific financial assets (sovereign debt related approaches), might actually increase the market's "predictability". Skepticism regarding e.g. the Quantitative Easing approaches of the several central banks globally during the last decades, Krishnamurthy and Vissing-Jorgensen, 2011; Wang, 2021; Beck, Duca, and Stracca, 2021; Gagnon et al., 2012; Kuttner, 2018; Dell'Ariccia, Rabanal, and Sandri, 2018, especially referring to it as "unconventional monetary policies" has given rise to many debates regarding their effect on global wealth distribution and well-being. By working on the direction of "smoothing" any potential divergence in the form of "structural change" in financial markets, usually referring to these as "distortions away from normal market conditions", they might actually also "smooth", mitigate, reduce or even eliminate the "uncertainty", "risk" or "fear" (Akerlof and Shiller, 2010) related to investment decision making. Risk considered one of "fuels" of financial management is responsible for "controlling" bad or dangerous investment decisions, giving incentives for more prudent and cautious players, taking informed decision in favor of the companies and for their clients in case we are focusing on financial services. Furthermore, without risk no value assessment can be performed, in the sense that the essence of financial pricing is containing and incorporating potential risk in the formalisation, and also develop tools towards "Controlling" it within reasonable and "healthy" levels. Moreover, in case "unconventional monetary policies" induce "bias" towards specific asset classes, then potential "moral hazard" Hébert, 2018 might arise in investments from the private sector, in the sense that investing towards the inferred "safety" in one asset class (or even speculating on "informed" in the form of "biased" massive flows), might potentially deprive allocations (capital) from another, possibly more productive and crucial to the economy and well-being than the other. "Bias" also is the exact opposite of what efficient markets are trying to achieve for all above reasons, while inducing "predictability" for maybe well-informed market players to exploit "risk-free". On the other hand though, "Invisible Hand" (Adam Smith, 1759) related approaches and policies have proved to be inefficient and even dangerous practices towards global well-being (Greenspan, 2001) and prosperity. As such, the correct "mix of control and normality" should be located somewhere in between, and mathematics can be of course a really powerful armory towards identifying this balance. Critics on whether financial markets and financial innovation are improving well-being with "smarter" and more efficient capital allocation and resources distribution mechanisms, or not, have been expressed and are fueling several socio-economico-philosophical debates on economics and financial theory for many years. Under the rationale that indeed financial innovation conceptually might indeed achieve such endeavors (Shiller, 2013a; Shiller, 2013b) careful and prudent use of its really powerful tools and mechanisms should be undertaken by both market players and regulatory bodies, as well.

On the Electricity Markets Analysis

There is an increasing emphasis in accurate load forecasting, as electricity markets are transformed into complex structures, facilitating energy commodities and financial-like instruments. In this work, we examined the Greek Electricity Market and attempted to build statistical models that limit the prediction error and provide valuable information to all market participants, from power generators to energy traders. Before embarking in building forecasting models for system load series, we first applied the Principal Component Analysis (PCA) approach, a linear tool of the manifold learning field, on the load data. We considered the given series as a projection-measurement (or realization) taken from a high-dimensional dynamical object, the "load manifold", since the underlying dynamics has a large number of degrees of freedom. This is equivalent to say that load if influenced by a very large number of factors in the electricity market. After applying PCA on load series we obtained a new low dimensional object, a 3D manifold on which the underlying dynamics of load is now influenced by only three factors or Principal Components, PCs (linear combinations of all the original variables, in our case the 24 hourly time series), that in total explain or capture 96% of the total variance of load. These PCs are independent, uncorrelated (since they are orthogonal to each other as they are extracted from PCA "filter") and "ideal" predictors in a multiple regression model, where the response variable is the load series. The PC regression model we have built is in fact a hybrid model, as the one used in the work of Hong and Wu, 2012 in predicting prices. In their work a hybrid PCA outperformed an ANN model. In particular, we have examined a Seasonal ARIMA model with eXogenous parameters, thus constructing a SARIMAX model, the Holt-Winters' exponential smoothing method and a Regressive model using Principal Components Analysis. To the best of our knowledge, there has not been a study in the Greek Electricity Market involving PCA or the augmented ARMA model, SARIMAX. We have not chosen the ANN models that have been used extensively in the GEM Kiartzis, Bakirtzis, and Petridis, 1995; Pappas et al., 2010; Lee and Ko, 2011, with known prons and cons, and instead we have chosen SARIMAX models that have not yet used at all (although ARMA and ARIMA ones have been applied) Cortes and Vapnik, 1995. Our results indicate that these methods are superior to the conventionally applied Holt-Winters' model and produce more reliable results (when applied to daily load data); however they need significantly longer estimation sets than the former method. All considered models are compared by means of the most out-of-sample accuracy measurements. All the measures (MAE, RMSE, MAME and Direction Movement statistics) confirm that two of the models analyzed in This work, i.e. SARIMAX and PC regression, perform quite well along the whole forecasting period. However, focusing on MAPE which is the "benchmark measure" in the forecasting "business" we conclude that PC Regression outperforms ETS, with SARIMAX the next best performing. As it is widely accepted, MAPE on the forecasting period is a crucial measure of prediction reliability and is often an extremely important measure for estimating the financial consequences that result from the supply and demand (load) interaction. In summary, in day-ahead forecasting of system load series of past daily data, we have found that the hybrid PC regression model has drastically outperformed a variety of ETS models used and slightly outperformed the next best SARIMAX model. We however point out here that the ETS models work extremely well on hourly data, especially of the type of ETS that Taylor describe in his works Taylor and McSharry, 2007, the Double Seasonal Holt Winters Exponential smoothing models that outperform both PCA and SARMA. It is our intention, in the near future, to examine the application of other novel techniques of the manifold learning group of models, as the LLE (Locally Linear Embedding) and Isomap (Isometric Feature Mapping) on load and price daily as well as hourly data, methods that have not used extensively as other more traditional ones have.

We have also examined, by using a SARMAX/GARCH approach, the effects of the regulatory reforms implemented over the period 2004-2014 on the structure of GEM. The resulted gradual restructuring of the market has affected the dynamic evolution of both level and volatility of SMP. Our SARMAX/GARCH modeling of SMP significantly contributes to the existing literature regarding the GEM since this approach has not been used before, at least in respect to the plethora of exogenous variables used. In our work we have included almost all relevant fundamentals and reforms, except variables describing the strategic behavior of market agents, which were left for a future next paper. Our findings are in alliance to the results of Petrella and Sapio, 2012 for the Italian market, enhancing the general view that changes in the market structure may have significant effects on the dynamics of SMP of them possibly unintentional as they may distort the market.

A drawback of using dummy variables for the reforms is that the response of the market's agents to these stimuli is not smoothed but abrupt like a step function. Furthermore, if there is a time lapse between the announcement of a reform and its implementation, this allows agents in advance, to take advantage of the reforms by adjusting themselves according to their position in the market.

From the above, we conclude that some of the reforms that have launched permanent or transitional mechanism, in order to enhance competitiveness, have created unintentionally opposite results. However, we have to point out that the phase out of any transitional mechanism without first a detailed and in depth examination of the side effects this might have on the current stability of the market, due to the strong interlinks and coupling nature these mechanisms have with fundamental variables (creating all these years as a consequence the current market microstructure), may have an adverse impact on the available long term generation capacity. Care must be taken therefore in securing future adequate availability by avoiding a forced canceling of mechanism ready in place that may put in danger the current equilibrium in the market.

As a general comment on policy design for market reforms, we can make -based on the results of our work- is that any required regulatory reforms that have to be applied, must be designed in such a way that facilitate the gradual development of competitive environment, allowing at the same time the minimization of the State's intervention (in areas that this is necessary) to the market as well as of the total cost of market transactions. The reforms also must create a secure environment for all participants. For this reason, any reforms that have a transitional characteristic (as some of the RMR considered in this study) must be evaluated according to the holistic or systemic effects on the market due to their interdependence and the fact that they are inherently coupled with the fundamental variables of the market, as the modeling results described in this work has revealed. The ultimate target, therefore, of any regulatory reform is to create motives and signals to the agents of the market, so they have a collective movement to more competitive behaviors and attitudes and suitable investment decisions. Towards this end, this work is believed to contribute to the current efforts of the Regulator and other relevant market bodies, towards restructuring the market as well as in helping developing tools/processes/rules etc. that will allow the fair pricing of product and services available in the Greek Electricity market as well as in attracting investors in the market.

The modeling results presented here can enhance the efforts of the decision in detecting more accurately the factors of uncertainty that influence heavily the Greek electricity market and must be taken into account when designing regulatory measures to face the continuously changing internal and external environment. The results of this work can be further improved by including the player's strategic behavior in the set of exogenous variables. Also, it is in our interest to use a completely different methodology on exactly the same GEM data, namely a Structural Vector Autoregression (SVAR) approach, and through Impulse Response functions and Granger Causality to detect the impact of both fundamentals and regulatory reforms that have shape the evolutionary dynamics of GEM.

On the Social Microblogging Applications

Over the last years it has been demonstrated that social media such as web based search engines and recently Twitter can be used to forecast certain future complex events. In a similar fashion, an unresolved problem in contemporary monetary and economical management research, the foreign currency exchange rate forecasting (Forex) in the short run, might be also benefit from the use of social microblogging. We believe that our study is of significant importance, as the contemporary research in the area trying to contradict the EMH is focused on the uncovering of market anomalies as these may arise by the information flow provided by the market players' "beliefs" in the social networks. Towards this direction, the development of several psychological indexes that are related to market's consensus, is already in great use by players as well as market's regulators. A previous analysis Papaioannou et al., 2013 showed that the rate exchange forecasting at level, based on people's beliefs as these can be data-mined through microblogging may carry significant information that can used to outperform the random walk hypothesis and AR models that do not include such information but solely past values of the exchange rates, in the very short run. This was also demonstrated through moving average trading simulations. However, we should note that this behaviour should be attributed to the underlying trend of the data. This is apparent when encountering larger forecasting horizons. In addition, evaluating forecasting efficiency and accuracy remain an important issue for further research. For example, out-of-sample statistical measures can be improved using rolling-origin evaluations and re-calibration of optimal coefficients based on new data sets (see Tashman, 2000 for a review and critical discussion). Concluding, we believe that social networks can provide the basis for further advances in the field and thus enable the formalization of the experimental side of the market's psychology as this is shaped by the human behavior. Towards this aim detailed Agent-based models analysed by state-of-the-art multiscale techniques (see e.g. Tsoumanis et al., 2010; Siettos, Gear, and Kevrekidis, 2012) have the potential to facilitate computational modeling and exploration - and thus our understanding and forecasting market's complex dynamics.

On the FOREX Trading Applications

PCA/LLE, Section : 3.5

We constructed and traded several portfolios, using both traditional portfolio construction mechanisms and trading rationales, as well as more sophisticated and modern ones, based on recent developments in the fields of unsupervised (manifold learning) and supervised machine learning. The target traded assets was a set of 20 FX Pairs, spanning a period of almost 21 years. The traditional benchmark portfolios, were the standard Long Only portfolio and the naive Risk Parity one. Additionally to these two, we constructed PCA embedded portfolios using either each individual or a set of PCA coordinates' (eigenvectors') elements, as the trading weights allocated to each FX Pair in the portfolio, but also using trading indicators constructed using information from another manifold learning embedding methodology, namely the Local Linear Embedding (LLE). More specifically, while the PCA coordinates' elements were themselves the portfolio weights corresponding to each FX pair in the portfolio, the LLE eigenvectors were used as leading trading indicators of the respective PCA embedded and benchmark portfolios' dynamics over time. Both embeddings were performed in both a rolling and an expanding window configuration.

One of the main targets of our analysis, was to explore predictability of the several constructed portfolios, as per the rationale by Box and Tiao, 1977. We explored this predictability, by using four different time series analysis based trading models' rationales, in order to forecast the dynamics of the several portfolios' returns over time. The time series models used, were the simple Exponential Moving Average (EMA) trading indicator, inspired from the technical analysis trading approach, the Autoregressive AR(p) model from the classical Time Series Analysis, Gausian Process Regressors (GPR) and the Recurrent Neural Networks (RNNs) framework from the Supervised Machine Learning field. Based on these forecasting models, we defined specific trading strategies which were then applied on the target portfolios in order to try to generate excess returns. We evaluated and compared the several trading strategies' performance, in terms of (annualised) Sharpe Ratio levels. We heuristically set a satisfactory Sharpe Ratio level for a well-performing trading strategy to be at least 0.5. Based on Cuturi and D'Aspremont, 2016, we characterized the several modeled and traded portfolios, as "Mean-Reverting" or "Trending", while also highlighting the importance and contribution on the predictability target of the embeddings' time window configuration, namely the rolling and expanding window approaches. We also defined the "Dynamic Learning" property that a specific time series trading model might possess or not, depending on the whether the model is specified by a hyperparameter, which is continuously computed and optimized, capable of identifying this underlying "Mean-Reverting" or "Treding" behavior of the traded portfolio, respectively. We also highlight the importance and significance of this property on real time trading and portfolio management applications.

While PCA has been extensively used for the construction of portfolios and forecasting, nonlinear manifold algorithms have been. The projection of a large-scale data-set on the dominant principal components results to a dimensionality reduction, or more intuitively set, to fewer linear combinations of the initial input assets returns. These linear combinations have been considered as “investment portfolios", and as such, modelling and forecasting of these PCA-constructed (principal) portfolios became a main tool for exploring market dynamics. Within this context, using PCA, (Trzcinka, 1986) found that a one-factor (one principal component) model may be used to explain adequately the pricing dynamics of securities. (Connor and Korajczyk, 1993) suggested that the first six principal components seem to be the most representative for the modelling of a data-set of the New York stock exchange and the American stock exchange returns. (Geweke and Zhou, 1996) found that only the first principal component is related to the clustered portfolio returns of a dataset of industry and market capitalization on a monthly frequency sampling period. Inspired by the work of (Connor and Korajczyk, 1993), (Jones, 2001) found that improved results on the quality of the extracted principal components can be obtained if heteroskedasticity is present in the data. (Avellaneda and Lee, 2010) applied several trading strategies using PCA on a broad set of US Securities and stock returns of the Exchange Traded Funds (ETFs). They modeled the idiosyncratic returns as mean reverting processes and proposed trading strategies which managed to generate satisfactory levels of Sharpe ratios (Sharpe, 1994) on a dataset spanning from 1997 to 2007 incorporating transaction costs into the reported results. The reported average Sharpe ratio of the PCA-based trading strategies was 0.9 for the period 2003 to 2007, while the average annual Sharpe ratio was 1.44 for the entire period, with stronger trading performance metrics during the periods prior to 2003. For the ETFs PCA strategies, they reported a Sharpe ratio of 1.1 from 1997-2007, while they showed improved trading performances after incorporating information from the corresponding trading volumes of the ETFs under consideration, with the resulting Sharpe ratio being 1.51 for the period 2003 to 2007. Finally, they studied the behavior of the generated strategies during the financial crisis in 2007 connecting the mean reverting dynamics with the stock market's cycles. Using PCA and Granger causality, (Granger, 1969; Billio et al., 2010) found strong interconnections between banks, brokers, insurance companies and hedge funds during the financial crisis of 2007 to 2009. They also found that those four types of financial structures were less liquid over the period under consideration. In the above PCA-based studies, positive loadings (positive elements of the principal eigenvectors) reflect long positions while negative loadings reflect short positions of the constructed principal portfolios. (Boyle et al., 2018) and (Avellaneda and Lee, 2010), (Kim and Omberg, 1996) tried to analyze and interpret also negative loadings for trading purposes. At this point, we should underline that one has to be cautious about the interpretation of the sign of the elements of the principal eigenvectors. It is trivial to show that the signs of the loadings are arbitrary and may switch sign during computations and-or using different software (Jollife and Cadima, 2016). Hence, before the extraction of any conclusions or interpretation the sign ambiguity has first to be resolved. (Bro, Acar, and Kolda, 2008). On this context, to the best of our knowledge, a nonlinear “analogue" of the PCA, namely the Local Linear Embedding (LLE) algorithm has not been used so far, but only for the identification of crises and market crashes (Huang, Kou, and Peng, 2017). Thus, it was of great interest to study how LLE compares or enhances the results obtained by PCA in a trading simulation setting.

We reported predictable behavior on the last PCA eigenvector based embedded portfolio, both under the rolling window as well as the expanding window configuration, traded using mean-reverting trading models, validating the work and findings of Cuturi and D'Aspremont, 2016 in the FX Markets, as well. However, interesting results were also obtained in terms of "Trending" portfolios' significant performance, predicted under the LLE framework, as well. More specifically, specific LLE trading indicators were capable to identify and accordingly trade the respective "Trending" related PCA embedded portfolios, yielding higher than the heuristically set threshold of 0.5, annualised Sharpe Ratio scores.

All of the above results, were also tested against real market's frictions, i.e. transaction costs. Transaction Costs Analysis (TCA) was performed in all of the top-ranked, in terms of "Gross" Sharpe Ratio levels, trading strategies. The total market cost impact (market spread and/or brokers' commissions and markups) was approximated under several transaction cost scenarios, each one representing a potential cost scheme, under which a market participant might be able to operate in real market conditions. We defined the term transaction cost "Resilience Level", which is the maximum number of those transaction costs scenarios that a specific trading strategy manages to overcome, and stay above the threshold level of 0.5 in terms of Sharpe Ratios. The corresponding results in Table 3.30 indicate that the majority of the well-performing trading systems in terms of "Gross" Sharpe Ratio levels failed to overcome the incurred costs, generating negative returns even under the most "aggressive" (or favorable to the trading performance) transaction costs' scenarios. Further really interesting results include the high resilience that was reported in the "Trending" related PCA embedded portfolios, forecasted using information from the LLE embedding framework. More specifically, the information extracted by the corresponding leading LLE trading indicators was powerful enough to "protect" the performance of the corresponding "Trending" PCA portfolios, even under all six transaction costs' scenarios. The same "Resilience gain" was also reported, when several "Mean-Reverting" based trading systems, like the most well-performing one, trading the last PCA coordinate embedded portfolio, using either an EMA, AR or GPR predictor, were combined with the more resilient in the several transaction costs, "Trending" ones, like the one presented in the previous example. While the top "Mean-Reverting" strategies scored in average higher Sharpe Ratio levels than the corresponding "Trending" ones, their performance was highly sensitive to the transaction costs' increase as depicted in the rapid Sharpe Ratio deterioration in Table 3.30, as well. This deterioration was partially or completely contained (avoided) after the combination of the respective strategies with either of the two most resilient "Trending" ones, thoroughly analysed and discussed in the relevant section 3.5.21. This "Resilience Gain" though comes in the expense of valuable information in terms of Sharpe Ratio level, if compared to the "Net" Sharpe Ratio that a trader could obtain if he simply traded the "Mean-Reverting" strategy alone, and not in combination with any of the "Trending" strategies, under a specific threshold of transaction costs incurred in the overall trading activity. Upon this note, we commented on the fact that the transaction cost schemes are greatly connected to the capacity of the market participant towards optimal and smarter price discovery technology setups, as well as "economies of scale" gains, based on its size and significance in the market. As an example, we discussed on the case of a retail trader/investor versus a large investment bank, hedge fund or market maker, and their average incurred transaction cost in the overall trading activity, discussing the higher competitive advantage of the latter, in terms of greater profitability and information gain if both players choose to trade the same potential "market anomally" displayed in the current context, but under different transaction costs schemes, respectively.

"Embed-Train-Lift", Section : 3.6

The main proposed methodology of the Thesis, namely the "Embed-Train-Lift" scheme, acted as the extension of the previous FOREX trading framework. The proposed scheme with Diffusion Maps (DM) and Local Linear Embedding (LLE), Multivariariate Auto-regressive (MVAR) and Gaussian Process Regression (GPR) models, along with the Geometric Harmonics (GH) methodology for lifting (reconstruction to the original space), outperformed all other "conventional schemes", i.e. the full MVAR and GPR models implemented directly at the original space, as well as the scheme that used PCA for the task of embedding. A combination of the "Embed-Train-Lift" scheme and the risk parity portfolio construction mechanism, performed within a rolling window framework, reported higher Sharpe Ratio scores (Sharpe, 1994) compared to all other approaches, extending the "predictable" embedded portfolios' set, under the manifold learning based frameworks presented in the Thesis.

4.0.2 On the Epidemic Modelling and Forecasting Applications

From our perspective, the tools used for time series modelling and forecasting for any of the other applications, could also be tested in the context of predicting the dynamics of target variables in the case of the recent pandemic, as well. In this Thesis the Gaussian Process Regression approach is displayed, with the results being really promising and seem to be able to capture a multi-step ahead evolution of the infected cases' time series of several Italian Regions. For some regions where the results are worse than the other, "sudden gaps" in the underlying modelled time series is observed, probably due to the fact that policies' related effects might have "distorted" a "smooth" - either upward or downward trending - dynamics, and as such making it harder for the several GPRs to capture. Daily monitoring of the several predictions in the deployed web-application "#combatcovid19uninafisr", 2021 specifically designed to be part of the COMBATCOVID19 research project by the Fondo Integrativo Speciale per la Ricerca (FISR), offers useful statistics and insights about the pandemic, while also acting as a good tool for academic and educational purposes, regarding the comparison of the performance of the several models applied. Several other models are also designed and constructed towards the same forecasting purpose, as for example Feed-Forward Neural Networks and LSTM, additionally to the "Embed-Train-Lift" scheme proposed in the Thesis, as well. However, still being under the process of proper parameterisation fine-tuning and towards fulfilling the Thesis's submission deadlines, their pandemic related results are omitted from the current context. Soon the whole armory will be launched and publicly available in terms of efficiency and real-time predictions in the web-application "#combatcovid19uninafisr", 2021 for further reference and comparison.

Future Research

Several other different schemes such as the autoencoders (Kramer, 1991; Chen and Ferguson, 2018), reservoir computing (Lukoševičius and Jaeger, 2009; Pathak et al., 2018) and LSTMs Greff et al., 2016; Vlachas et al., 2018 could also be used in the same context and applications' set as the current Thesis. We aim at performing an extensive comparison of the above methodologies in a future work. We note that in order to cope with the generalization property and the topological instability issues (see e.g. Balasubramanian et al., 2002) that arise in implementing kernel-based manifold learning algorithms when the data set is not dense on the manifold, and/or in the presence of "strong" stochasticity, one can resort to techniques such as the constant-shift one to appropriately adjust the graph metric, and techniques for the removal of outliers from the data set and the construction of smooth geodesics (see e.g. Choi and Choi, 2007; Wang, 2012; Gajamannage, Paffenroth, and Bollt, 2019). We intend to explore the efficiency of such approaches in a future work. One can also attempt to link the proposed framework with Takens embedding and Whitney immersion theorems Takens, 1981; Whitney, 1944, e.g. by creating an extended data set which contains shifted copies of the original time series.

Regarding the financial aspect, the previously discussed intervetions and imposed "rules", limitations and "biased" incentives in the markets, could be associated with macroeconomic variables which act as the base of information for the several policies to be framed. This connection is usually based on models constructed in econometrics in the form of statistical regressions trying to infer optimal (target) levels for quantities like inflation, interest rates, unemployment and growth for the economy to be assumed to work "normally". Future research, using the time series and manifold learning related techniques and approaches, as the current, could shed light on how these macro-models' (IS-LM, AD-AS etc) related variables and their connection with policy making (Watanabe, 2021, Hiermeyer, 2019) might be inferring "smoothness" and "predictability" of financial assets' dynamics. Some work has already been initiated, Guirao, García-Rubio, and Vera, 2012, targeting the dynamic modelling of this interconnection between the macro space and tools from nonlinear dynamical systems, with the benefits also extending beyond finance and economics, to e.g. challenges like crisis management as the recent pandemic outburst and the implications of macroeconomic policy to the "health" of the several economies globally, (Buklemishev et al., 2021). A really interesting and useful path to build upon.

Finally, a tremendous need to connect economical and financial modelling with epidemiology was emerged during the recent COVID-19 pandemic, (Kuniya, 2021; Mahi et al., 2021; Bloom, Cadarette, and Sevilla, 2018; Gong et al., 2020; Murray, 2020). The severe effects in global economies and more generally in societies, made it clear that all research fields are and should be contributing to each other. In a more general context, they are all acting under the same prism of achieving the same long-term ethical goals' set, the prosperity, well-being and first of all health of all societies in the world. Towards that goal, the manifold learning approach's perspective as outlined in the current Thesis can be further extended. Multiple other tools, out of the vast and wide variety of embedding, forecasting and lifting methodologies can be further used under the same "Embed-Train-Lift" approach suggested in the Thesis, enhancing and widening the arsenal of potential such "Embed-Train-Lift" schemes, potentially yielding even more efficient combinations and corresponding results.

Appendix A

Appendix

A.1 Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE)

1. Mean Absolute Error (MAE) :

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \tag{A.1}$$

2. Root Mean Squared Error (RMSE) :

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{A.2}$$

3. Mean Absolute Percentage Error (MAPE) :

$$MAPE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{y_t}} \tag{A.3}$$

A.2 Direction Statistics

In energy commodities price forecasting, improved decisions usually depend on correct forecasting of directions, of actual price, y_t, and forecasted price $\hat{y}_t$. The ability to predict movement direction can be measured by a directional statistic (Dstat) Yu and Xu, 2014, which can be expressed as

$$D_{stat} = \frac{1}{N}\sum_{t=1}^{n}a_t 100\% \tag{A.4}$$

with

$$a_t = \begin{cases} 1, & \text{if}(y_{t+1} - y_t)(\hat{y}_{t+1} - \hat{y}_t) \geq 0 \\ 0, \text{otherwise} \end{cases} \tag{A.5}$$

A.3 Durbin-Watson statistics

The Durbin-Watson statistics is defined by:

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \tag{A.6}$$

and it ranges in value from 0 through 4, with an intermediate value 2. If DW is near 2 (2 is the ideal value for white noise residuals), there are no autocorrelations in the residuals. For $DW < 1$ there are positive autocorrelations and for $DW > 1$ there are negative autocorrelations.

A.4 Theil's Inequality Coefficient

Theil's Inequality Coefficient provides a way of measuring the fitting of the in-sample estimation to the actual time series of interest. It is defined by:

$$TIC = \frac{\sqrt{(\sum_{t=T+1}^{T+h} \frac{\hat{y}_t - y_t)^2}{h}}}{\sqrt{\sum_{t=T+1}^{T+h} \frac{\hat{y}_t^2}{h}} + \sqrt{\sum_{t=T+1}^{T+h} \frac{y_t^2}{h}}} \tag{A.7}$$

TIC takes values between 0 and 1, with 0 indicating a perfect fit and 1 no match at all.

A.5 Jarque-Bera Test (JB test) for Normality

The Jarque-Bera is a statistical test that examines if the series in question comes from a normal distribution. More specifically, let H_0 be the null hypothesis : H_0: the time series comes from a normal distribution process. The Jarque-Bera test (JB) for normality is based on two measures, skewness and kurtosis, the first referring to how symmetric the data are around the mean. Perfectly symmetric data will have a skewness of zero. We check how these two values are sufficiently different for zero and 3 respectively. The Jarque-Bera statistics is given by:

$$JB = \frac{N}{6}(s^2 + \frac{(k-3)^2}{4}) \tag{A.8}$$

where N is the sample size, s is skewness and k is kurtosis. Therefore, large values of s and/or $k \neq 3$ will lead to a large value of JB. When the data are normally distributed, the JB has a chi-squared distribution with two degrees of freedom. We reject the null hypothesis of normally distributed data if JB exceeds a critical value: 5.99 (for 5%) and 9.21 (for 1%).

A.6 Augmented Dickey-Fuller test (ADF) for unit root (stability)

The Augmented Dickey-Fuller test for a unit root assesses the null hypothesis of a unit root using the model

$$y_t = c + \delta t + \phi y_{t-1} + \beta_1 \Delta y_{t-1} + \cdots + \beta_p \Delta y_{t-p} + \epsilon_t \tag{A.9}$$

where,

1. Δ is the differencing operator, such that $\Delta y_t = y_t - y_{t-1}$.

2. The number of lagged difference terms, p, is user specified.

3. ϵ_t is a mean zero innovation process.

The null hypothesis of a unit root is : H_0: the series has a unit root ϕ=1. Under the alternative hypothesis, $\phi < 1$. Variants of the model allow for different growth characteristics. The model with $\delta = 0$ has no trend component, and the model with $c = 0$ and $\delta = 0$ has no drift or trend. A test that fails to reject the null hypothesis, fails to reject the possibility of a unit root. In our case unit root is rejected, indicating the stability of the original series.

A.7 Kwiatkowski-Phillips-Schmidt-Shin test (KPSS) for stationarity

The Kwiatkowski-Phillips-Schmidt-Shin test for stationarity assesses the null hypothesis that the series is stationary, using the model:

$$y_t = \beta t + (r_t + \alpha) + \epsilon_t \tag{A.10}$$

$r_t = r_{t-1} + u_t$ is a random walk, the initial value $r_0 = \alpha$ serves as an intercept, t is the time index, u_t are i.i.d. The simplified version of the model without the time trend component is also used to test level stationarity. The null hypothesis is stated as follows:

H_0: the series is stationary.

Rejecting the null hypothesis, means that there is indication of nonstationarity, which deters us from applying the regressive models of interest.

A.8 ARCH test for heteroscedasticity of the residuals

Engle's ARCH test assesses the null hypothesis that a series of residuals (r_t) exhibits no conditional heteroscedasticity (ARCH effects), against the alternative that an ARCH(L) model describes the series. The ARCH(L) model has the following form:

$$r_t^2 = a_0 + a_1 r_{2t-1} + \cdots + a_L r_{2t-L} + e_r \tag{A.11}$$

where there is at least one $a_j \neq 0, \quad j = 0, \ldots, L$. The test statistic is the Lagrange multiplier statistic TR2, where: T is the sample size and R^2 is the coefficient of determination from fitting the ARCH(L) model for a number of lags (L) via regression. Under the null hypothesis, the asymptotic distribution of the test statistic is chi-square with L degrees of freedom.

A.9 Ljung-Box Q-test (LBQ) for autocorrelation of the residuals

The Ljung-Box Q-test is a "portmanteau" test that assesses the null hypothesis that a series of residuals exhibits no autocorrelation for a fixed number of lags L, against the alternative that some autocorrelation coefficient $\rho(k), \quad k = 1, \ldots, L$, is nonzero.

The test statistic is

$$Q = T(T+2)\sum_{k=1}^{L}\left(\frac{(\rho(k)^2)}{(T-k)}\right) \tag{A.12}$$

where T is the sample size, L is the number of autocorrelation lags, and $\rho(k)$ is the sample autocorrelation at lag k. Under the null hypothesis, the asymptotic distribution of Q is chi-square with L degrees of freedom.

Bibliography

(N.d.[b]). In: ().

Abarbanel, Henry D. I. and Jerry P. Gollub (1996). “ Analysis of Observed Chaotic Data ”. In: *Physics Today*. ISSN: 0031-9228. DOI: 10.1063/1.881528.

Abramowitz, Milton, Irene A. Stegun, and David Miller (1965). “Handbook of Mathematical Functions With Formulas, Graphs and Mathematical Tables (National Bureau of Standards Applied Mathematics Series No. 55)”. In: *Journal of Applied Mechanics*. ISSN: 0021-8936. DOI: 10.1115/1.3625776.

Adam Smith, 1759 (1759). “3. THE THEORY OF MORAL SENTIMENTS”. In: *Adam Smith*. DOI: 10.1515/9781400873487-006.

Adrangi, Bahram et al. (2001). “Chaos in oil prices? Evidence from futures markets”. In: *Energy Economics*. ISSN: 01409883. DOI: 10.1016/S0140-9883(00)00079-7.

Aizerman, M A, E A Braverman, and L Rozonoer (1964). “Theoretical foundations of the potential function method in pattern recognition learning.” In: *Automation and Remote Control,*

Akaike, Hirotugu (1974). “A New Look at the Statistical Model Identification”. In: *IEEE Transactions on Automatic Control*. ISSN: 15582523. DOI: 10.1109/TAC.1974.1100705.

Akerlof, George A. and Robert J. Shiller (2010). *Animal spirits: How human psychology drives the economy, and why it matters for global capitalism*. ISBN: 9781400834723. DOI: 10.5455/ey.20020.

Alexandridis, Alex, Ioannis Th Famelis, and Charalambos Tsitouras (2015). “Long-term time-series prediction using radial basis function neural networks”. In: *AIP Conference Proceedings*. ISBN: 9780735412873. DOI: 10.1063/1.4912958.

Amorim, Elisa et al. (2015). “Facing the high-dimensions: Inverse projection with radial basis functions”. In: *Computers & Graphics* 48, pp. 35–47.

An, Chansik et al. (2020). “Machine learning prediction for mortality of patients diagnosed with COVID-19: a nationwide Korean cohort study”. In: *Scientific Reports* 10.1. DOI: 10.1038/s41598-020-75767-2. URL: https://doi.org/10.1038\%2Fs41598-020-75767-2.

Anastassopoulou, Cleo et al. (2020). “Data-Based Analysis, Modelling and Forecasting of the COVID-19 outbreak”. In: DOI: 10.1101/2020.02.11.20022186. URL: https://www.medrxiv.org/content/early/2020/03/12/2020.02.11.20022186.

and, Aasim M. Husain, and Chakriya Bowman (2004). “Forecasting Commodity Prices: Futures Versus Judgment”. In: *IMF Working Papers*. ISSN: 1018-5941. DOI: 10.5089/9781451846133.001.

Anderson, E. et al. (1999). *LAPACK Users’ Guide*. Third. Philadelphia, PA: SIAM. ISBN: 0-89871-447-8 (paperback).

Annis, A. A. and E. H. Lloyd (1976). “The expected value of the adjusted rescaled hurst range of independent normal summands”. In: *Biometrika*. ISSN: 00063444. DOI: 10.1093/biomet/63.1.111.

Aronszajn, N. (1950). “Theory of Reproducing Kernels”. In: *Transactions of the American Mathematical Society*. ISSN: 00029947. DOI: 10.2307/1990404.

Auer, Benjamin R. and Andreas Hoffmann (2016). "Do carry trade returns show signs of long memory?" In: *Quarterly Review of Economics and Finance*. ISSN: 10629769. DOI: 10.1016/j.qref.2016.02.007.

Avellaneda, Marco and Jeong Hyun Lee (2010). "Statistical arbitrage in the US equities market". In: *Quantitative Finance*. ISSN: 14697688. DOI: 10.1080/14697680903124632.

Baccalá, Luiz A. and Koichi Sameshima (2001). "Partial directed coherence: a new concept in neural structure determination". In: *Biological Cybernetics* 84.6, pp. 463–474.

Bachtiger, Patrik, Nicholas S Peters, and Simon LF Walsh (2020). "Machine learning for COVID-19—asking the right questions". In: *The Lancet Digital Health* 2.8, e391–e392. DOI: 10.1016/s2589-7500(20)30162-x. URL: https://doi.org/10.1016\%2Fs2589-7500\%2820\%2930162-x.

Balasis, Georgios et al. (2008). "Dynamical complexity in Dst time series using non-extensive Tsallis entropy". In: *Geophysical Research Letters*. ISSN: 00948276. DOI: 10.1029/2008GL034743.

Balasubramanian, Mukund et al. (2002). "The Isomap algorithm and topological stability". In: *Science* 295.5552, pp. 7–7.

Barabsi, Albert Lszl and Tams Vicsek (1991). "Multifractality of self-affine fractals". In: *Physical Review A*. ISSN: 10502947. DOI: 10.1103/PhysRevA.44.2730.

Barberis, Nicholas (2000). "Investing for the long run when returns are predictable". In: *Journal of Finance*. ISSN: 00221082. DOI: 10.1111/0022-1082.00205.

Barkan, Oren and Hagai Aronowitz (2013). "Diffusion maps for PLDA-based speaker verification". In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. ISBN: 9781479903566. DOI: 10.1109/ICASSP.2013.6639149.

Barkan, Oren et al. (2013). "Fast high dimensional vector multiplication face recognition". In: *Proceedings of the IEEE International Conference on Computer Vision*. ISBN: 9781479928392. DOI: 10.1109/ICCV.2013.246.

Barnett, William A. and Melvin J. Hinich (1992). "Empirical chaotic dynamics in economics". In: *Annals of Operations Research*. ISSN: 02545330. DOI: 10.1007/BF02071045.

Barnett, William A. and Apostolos Serletis (2000). "Martingales, nonlinearity, and chaos". In: *Journal of Economic Dynamics and Control*. ISSN: 01651889. DOI: 10.1016/S0165-1889(99)00023-8.

Barnett, William A. et al. (1997). "A single-blind controlled competition among tests for nonlinearity and chaos". In: *Journal of Econometrics*. ISSN: 03044076. DOI: 10.1016/S0304-4076(97)00081-X.

Barton, Richard J. and H. Vincent Poor (1992). "On generalized signal-to-noise ratios in quadratic detection". In: *Mathematics of Control, Signals, and Systems*. ISSN: 09324194. DOI: 10.1007/BF01211977.

Beck, Roland, Ioana Duca, and Livio Stracca (2021). "Medium Term Treatment and Side Effects of Quantitative Easing: International Evidence". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3325867.

Bekkers, Niels, Ronald Q. Doeswijk, and Trevin W. Lam (2012). "Strategic Asset Allocation: Determining the Optimal Portfolio with Ten Asset Classes". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.1368689.

Belkin, Mikhail and Partha Niyogi (2003). "Laplacian eigenmaps for dimensionality reduction and data representation". In: *Neural computation* 15.6, pp. 1373–1396.

Bellmann, R (1957). "Dynamic programming". In: *Princeton University Press* 8.

Bennett, Scott et al. (2016). "Alpha generation in portfolio management: Long-run Australian equity fund evidence". In: *Australian Journal of Management*. ISSN: 13272020. DOI: 10.1177/0312896214539815.

BenSaïda, Ahmed (2015). "A practical test for noisy chaotic dynamics". In: *SoftwareX*. ISSN: 23527110. DOI: 10.1016/j.softx.2015.08.002.

Bensaïda, Ahmed and Houda Litimi (2013). "High level chaos in the exchange and index markets". In: *Chaos, Solitons and Fractals*. ISSN: 09600779. DOI: 10.1016/j.chaos.2013.06.004.

Benth, Fred Espen, Jurate Šaltyte Benth, and Steen Koekebakker (2008). *Stochastic modeling of electricity and related markets*. ISBN: 9789812812315. DOI: 10.1142/6811.

Beran, Jan and Norma Terrin (1996). "Testing for a change of the long-memory parameter". In: *Biometrika*. ISSN: 00063444. DOI: 10.1093/biomet/83.3.627.

Berger, Marcel and Bernard Gostiaux (2012). *Differential Geometry: Manifolds, Curves, and Surfaces*. Vol. 115. Springer Science & Business Media.

Bermanis, Amit, Amir Averbuch, and Ronald R. Coifman (2013). "Multiscale data sampling and function extension". In: *Applied and Computational Harmonic Analysis*. ISSN: 10635203. DOI: 10.1016/j.acha.2012.03.002.

Bertozzi, Andrea L. et al. (2020). "The challenges of modeling and forecasting the spread of COVID-19". In: *Proceedings of the National Academy of Sciences* 117.29, pp. 16732–16738. ISSN: 0027-8424. DOI: 10.1073/pnas.2006520117. eprint: https://www.pnas.org/content/117/29/16732.full.pdf. URL: https://www.pnas.org/content/117/29/16732.

Bettencourt, Luís M.A. and Ruy M. Ribeiro (2008). "Real time bayesian estimation of the epidemic potential of emerging infectious diseases". In: *PLoS ONE*. ISSN: 19326203. DOI: 10.1371/journal.pone.0002185.

Bhattacharjee, Satyaki and Karel Matouš (2016). "A nonlinear manifold-based reduced order model for multiscale analysis of heterogeneous hyperelastic materials". In: *Journal of Computational Physics* 313, pp. 635–653.

Biancolini, Marco Evangelos (2018). *Fast radial basis functions for engineering applications*. ISBN: 9783319750118. DOI: 10.1007/978-3-319-75011-8.

Billio, Monica et al. (2010). "Measuring Systemic Risk in the Finance and Insurance Sectors". In: *MIT Sloan School Working Paper*. ISSN: 1098-6596. DOI: 10.1017/CBO9781107415324.004. arXiv: arXiv:1011.1669v3.

Blitzstein, Joseph K. and Jessica Hwang (2014). *Introduction to probability*. ISBN: 9781466575592. DOI: 10.1201/b17221.

Bloom, D E, D Cadarette, and J P Sevilla (2018). "Epidemics and Economics: New and resurgent infectious diseases can have far-reaching economic repercussions". In: *Finance & Development*. ISSN: 0015-1947.

Boothby, William M. (2003). *An Introduction to Differentiable Manifolds and Riemannian Geometry*. ISBN: 0121160505.

Borodin, Andrei N. et al. (1996). "Ornstein-Uhlenbeck Process". In: *Handbook of Brownian Motion — Facts and Formulae*. DOI: 10.1007/978-3-0348-7652-0_13.

Box, G. E.P. and G. C. Tiao (1977). "A canonical analysis of multiple time series". In: *Biometrika*. ISSN: 00063444. DOI: 10.1093/biomet/64.2.355.

Box, George E.P., Gwilym M. Jenkins, and Gregory C. Reinsel (2013). *Time series analysis: Forecasting and control: Fourth edition*. ISBN: 9781118619193. DOI: 10.1002/9781118619193.

Boyle, Phelim et al. (2018). "Short Positions in the First Principal Component Portfolio". In: *North American Actuarial Journal*. ISSN: 10920277. DOI: 10.1080/10920277.2017.1387573.

Bradley, Elizabeth and Holger Kantz (2015). "Nonlinear time-series analysis revisited". In: *Chaos*. ISSN: 10897682. DOI: 10.1063/1.4917289.

Braga, Maria Debora (2015). "Risk parity versus other µ-free strategies: a comparison in a triple view". In: *Investment Management and Financial Innovations* 12.1. URL: https://businessperspectives.org/pdfproxy.php?item{_}id:6497.

Bro, R., E. Acar, and Tamara G. Kolda (2008). "Resolving the sign ambiguity in the singular value decomposition". In: *Journal of Chemometrics*. ISSN: 08869383. DOI: 10.1002/cem.1122.

Brock, W. A. (1986). "Distinguishing random and deterministic systems: Abridged version". In: *Journal of Economic Theory*. ISSN: 10957235. DOI: 10.1016/0022-0531(86)90014-1.

Brockwell, Peter J and Richard A Davis (1991). *Time Series: Theory and Methods (Springer Series in Statistics)*. ISBN: 1441903194.

Brockwell, Peter J et al. (2016). *Introduction to time series and forecasting*. Springer.

Broock, W. A. et al. (1996). "A test for independence based on the correlation dimension". In: *Econometric Reviews*. ISSN: 15324168. DOI: 10.1080/07474939608800353.

Broomhead, D. S. and Gregory P. King (1986). "Extracting qualitative dynamics from experimental data". In: *Physica D: Nonlinear Phenomena*. ISSN: 01672789. DOI: 10.1016/0167-2789(86)90031-X.

Brunnermeier, Markus K., Stefan Nagel, and Lasse H. Pedersen (2008). "Carry trades and currency crashes". In: *NBER Macroeconomics Annual*. ISSN: 08893365. DOI: 10.1086/593088.

Brunton, Steven L, Joshua L Proctor, and J Nathan Kutz (2016). "Discovering governing equations from data by sparse identification of nonlinear dynamical systems". In: *Proceedings of the national academy of sciences* 113.15, pp. 3932–3937.

Buhmann, M D and Jeremy Levesley (2004). "Radial Basis Functions: Theory and Implementations". In: *Mathematics of Computation*.

Buklemishev, Oleg V. et al. (2021). "Macroeconomic policy in a pandemic era: What does the is-lm model show?" In: *Voprosy Ekonomiki*. ISSN: 00428736. DOI: 10.32609/0042-8736-2021-2-35-47.

Burlaga, L. F. and A. F. Viñas (2005). "Triangle for the entropic index q of nonextensive statistical mechanics observed by Voyager 1 in the distant heliosphere". In: *Physica A: Statistical Mechanics and its Applications*. ISSN: 03784371. DOI: 10.1016/j.physa.2005.06.065.

Burnside, Craig (2012). "Carry Trades and Risk". In: *Handbook of Exchange Rates*. ISBN: 9780470768839. DOI: 10.1002/9781118445785.ch10.

Burnside, Craig et al. (2011). "Do peso problems explain the returns to the carry trade?" In: *Review of Financial Studies*. ISSN: 08939454. DOI: 10.1093/rfs/hhq138.

Byrd, Richard H. et al. (1995). "A Limited Memory Algorithm for Bound Constrained Optimization". In: *SIAM Journal on Scientific Computing*. ISSN: 1064-8275. DOI: 10.1137/0916069.

Byun, Hyun Woo et al. (2015). "Using a principal component analysis for multi-currencies-trading in the foreign exchange market". In: *Intelligent Data Analysis*. ISSN: 15714128. DOI: 10.3233/IDA-150738.

Bär, Markus, Rainer Hegger, and Holger Kantz (1999). "Fitting partial differential equations to space-time dynamics". In: *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*. ISSN: 1063651X. DOI: 10.1103/PhysRevE.59.337.

Caballero, Ricardo J., Emmanuel Farhi, and Pierre Oliver Gourinchas (2008). "An equilibrium model of global imbalances and low interest rates". In: *American Economic Review*. ISSN: 00028282. DOI: 10.1257/aer.98.1.358.

Canto, Alvaro Bartolome del (n.d.). *investpy - Financial Data Extraction from Investing.com with Python*. URL: https://pypi.org/project/investpy/ (visited on 03/15/2020).

Casdagli, Martin (1992). "Chaos and Deterministic Versus Stochastic Non-Linear Modelling". In: *Journal of the Royal Statistical Society: Series B (Methodological)*. ISSN: 0035-9246. DOI: 10.1111/j.2517-6161.1992.tb01884.x.

Cavalcanti, Gil R. (2020). "Hodge theory of SKT manifolds". In: *Advances in Mathematics*. ISSN: 10902082. DOI: 10.1016/j.aim.2020.107270.

Centra, M. (2011). "Hourly electricity load forecasting: An empirical application to the Italian railways". In: *World Academy of Science, Engineering and Technology*. ISSN: 2010376X. DOI: 10.5281/zenodo.1082583.

Chaves, Denis et al. (2012). "Efficient Algorithms for Computing RiskParity Portfolio Weights". In: *The Journal of Investing*. ISSN: 1068-0896. DOI: 10.3905/joi.2012.21.3.150.

Chen, Wei and Andrew L Ferguson (2018). "Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration". In: *Journal of computational chemistry* 39.25, pp. 2079–2102.

Chen, Yangkang and Sergey Fomel (2015). "Random noise attenuation using local signal-and-noise orthogonalization". In: *Geophysics*. ISSN: 19422156. DOI: 10.1190/GEO2014-0227.1.

Cheng, Changqing et al. (2015). "Time series forecasting for nonlinear and non-stationary processes: a review and comparative study". In: *Iie Transactions* 47.10, pp. 1053–1071.

Chiavazzo, Eliodoro et al. (2014). "Reduced models in chemical kinetics via nonlinear data-mining". In: *Processes* 2.1, pp. 112–140.

Chimmula, Vinay Kumar Reddy and Lei Zhang (2020). "Time series forecasting of COVID-19 transmission in Canada using LSTM networks". In: *Chaos, Solitons & Fractals* 135, p. 109864. DOI: 10.1016/j.chaos.2020.109864. URL: https://doi.org/10.1016\%2Fj.chaos.2020.109864.

Chinazzi, Matteo et al. (2020). "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak". In: *Science*, eaba9757. DOI: 10.1126/science.aba9757. URL: https://doi.org/10.1126\%2Fscience.aba9757.

Choi, Heeyoul and Seungjin Choi (2007). "Robust kernel Isomap". In: *Pattern Recognition* 40.3, pp. 853–862.

Choie, Kenneth S. "Nicholas" (1993). "Currency Exchange Rate Forecast and Interest Rate Differential". In: *The Journal of Portfolio Management*. ISSN: 0095-4918. DOI: 10.3905/jpm.1993.409435.

Clare, Andrew et al. (2016). "The trend is our friend: Risk parity, momentum and trend following in global asset allocation". In: *Journal of Behavioral and Experimental Finance*. ISSN: 22146369. DOI: 10.1016/j.jbef.2016.01.002.

Clarida, Richard, Josh Davis, and Niels Pedersen (2009). "Currency carry trade regimes: Beyond the Fama regression". In: *Journal of International Money and Finance*. ISSN: 02615606. DOI: 10.1016/j.jimonfin.2009.08.010.

Clarke, Roger, Harindra De Silva, and Steven Thorley (2006). "Minimum-variance portfolios in the U.S. equity market". In: *Journal of Portfolio Management*. ISSN: 00954918. DOI: 10.3905/jpm.2006.661366.

— (2011). "Minimum-variance portfolio composition". In: *Journal of Portfolio Management*. ISSN: 00954918. DOI: 10.3905/jpm.2011.37.2.031.

Cohen, Leon (1998). "Generalization of the Wiener-Khinchin Theorem". In: *IEEE Signal Processing Letters*. ISSN: 10709908. DOI: 10.1109/97.728471.

Coifman, Ronald R and Stéphane Lafon (2006a). "Diffusion maps". In: *Applied and computational harmonic analysis* 21.1, pp. 5–30.

Coifman, Ronald R. and Stéphane Lafon (2006c). "Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions". In: *Applied and Computational Harmonic Analysis*. ISSN: 10635203. DOI: 10.1016/j.acha.2005.07.005.

Coifman, Ronald R and Stéphane Lafon (2006b). "Geometric harmonics: a novel tool for multiscale out-of-sample extension of empirical functions". In: *Applied and Computational Harmonic Analysis* 21.1, pp. 31–52.

Coifman, Ronald R et al. (2005). "Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps". In: *Proceedings of the National Academy of Sciences* 102.21, pp. 7426–7431.

Coifman, Ronald R et al. (2008). "Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems". In: *Multiscale Modeling & Simulation* 7.2, pp. 842–864.

Colding, Tobias Holck (2020). "Book Review: Introduction to Global Analysis. Minimal Surfaces in Riemannian Manifolds". In: *Bulletin of the American Mathematical Society*. ISSN: 0273-0979. DOI: 10.1090/bull/1689.

Collet, A., J. Bragard, and P. C. Dauby (2017). "Temperature, geometry, and bifurcations in the numerical modeling of the cardiac mechano-electric feedback". In: *Chaos*. ISSN: 10541500. DOI: 10.1063/1.5000710.

"#combatcovid19uninafisr" (2021). URL: http://combatcovid19uninafisr.com.

Connor, Gregory and Robert A. Korajczyk (1993). "A Test for the Number of Factors in an Approximate Factor Model". In: *The Journal of Finance*. ISSN: 15406261. DOI: 10.1111/j.1540-6261.1993.tb04754.x.

Corani, Giorgio, Alessio Benavoli, and Marco Zaffalon (2021). *Time series forecasting with Gaussian Processes needs priors*. arXiv: 2009.08102 [stat.ML].

Cortes, Corinna and Vladimir Vapnik (1995). "Support-Vector Networks". In: *Machine Learning*. ISSN: 15730565. DOI: 10.1023/A:1022627411411.

Craioveanu, Mircea, Mircea Puta, and Themistocles M. Rassias (2001). "Introduction to Riemannian Manifolds". In: *Old and New Aspects in Spectral Geometry*. DOI: 10.1007/978-94-017-2475-3_1.

Cuppen, J J M (1980). "A divide and conquer method for the symmetric tridiagonal eigenproblem". In: *Numerische Mathematik*. ISSN: 09453245. DOI: 10.1007/BF01396757.

Cuturi, Marco and Alexandre D'Aspremont (2016). "Mean-Reverting Portfolios: Tradeoffs between Sparsity and Volatility". In: *Financial Signal Processing and Machine Learning*. ISBN: 9781118745540. DOI: 10.1002/9781118745540.ch3. arXiv: 1509.05954.

Dablander, Fabian (2019). URL: https://fabiandablander.com/statistics/Two-Properties.html.

Dakhlaoui, Imen and Chaker Aloui (2013). "The US oil spot market: A deterministic chaotic process or a stochastic process?" In: *Journal of Energy Markets*. ISSN: 17563615. DOI: 10.21314/JEM.2013.084.

Dautel, Alexander Jakob et al. (2020). "Forex exchange rate forecasting using deep recurrent neural networks". In: *Digital Finance*. ISSN: 2524-6984. DOI: 10.1007/s42521-020-00019-x.

Dcsa, Pooja Singal and Rohatk R S Chhillar (2014). "Dijkstra Shortest Path Algorithm using Global Positioning System". In: *International Journal of Computer Applications*.

De Bie, Tijl, Nello Cristianini, and Roman Rosipal (2005). "Eigenproblems in Pattern Recognition". In: *Handbook of Geometric Computing*. DOI: 10.1007/3-540-28247-5_5.

De Gooijer, Jan G and Rob J Hyndman (2006). "25 years of time series forecasting". In: *International journal of forecasting* 22.3, pp. 443–473.

De Salazar, Pablo M et al. (2020). "Using predicted imports of 2019-nCoV cases to determine locations that may not be identifying all imported cases". In: DOI: 10.1101/2020.02.04.20020495. URL: https://www.medrxiv.org/content/early/2020/02/11/2020.02.04.20020495.

Dell'Ariccia, Giovanni, Pau Rabanal, and Damiano Sandri (2018). "Unconventional monetary policies in the Euro Area, Japan, and the United Kingdom". In: *Journal of Economic Perspectives*. DOI: 10.1257/jep.32.4.147.

Delves, Leonard Michael and Julie L Mohamed (1988). *Computational methods for integral equations*. CUP Archive.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal (2009). "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?" In: *Review of Financial Studies*. ISSN: 14657368. DOI: 10.1093/rfs/hhm075.

Deng, Qi (2020). "Dynamics and Development of the COVID-19 Epidemic in the United States: A Compartmental Model Enhanced With Deep Learning Techniques". In: *Journal of Medical Internet Research* 22.8, e21173. DOI: 10.2196/21173. URL: https://doi.org/10.2196\%2F21173.

Dickey, David A. and Wayne A. Fuller (1979). "Distribution of the Estimators for Autoregressive Time Series With a Unit Root". In: *Journal of the American Statistical Association*. ISSN: 01621459. DOI: 10.2307/2286348.

Diebold, Francis X. and Canlin Li (2006). "Forecasting the term structure of government bond yields". In: *Journal of Econometrics*. ISSN: 03044076. DOI: 10.1016/j.jeconom.2005.03.005.

Dietrich, Felix, Thomas N Thiem, and Ioannis G Kevrekidis (2020). "On the Koopman operator of algorithms". In: *SIAM Journal on Applied Dynamical Systems* 19.2, pp. 860–885.

Donoho, David L. and Carrie Grimes (2003). "Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data". In: *Proceedings of the National Academy of Sciences of the United States of America*. ISSN: 00278424. DOI: 10.1073/pnas.1031596100.

Drew, Michael E., Madhu Veeraraghavan, and Vanessa Wilson (2005). "Market timing, selectivity and alpha generation: Evidence from Australian equity superannuation funds". In: *Investment Management and Financial Innovations*. ISSN: 18129358.

Dsilva, Carmeline J et al. (2015). "Data-driven reduction for multiscale stochastic dynamical systems". In: *arXiv preprint arXiv:1501.05195*.

Dsilva, Carmeline J et al. (2018). "Parsimonious representation of nonlinear dynamical systems through manifold learning: a chemotaxis case study". In: *Applied and Computational Harmonic Analysis* 44.3, pp. 759–773.

E Parzen (1970). *Statistical inference on time series by rkhs methods*.

Eab, Chai Hok and S. C. Lim (2006). "Path integral representation of fractional harmonic oscillator". In: *Physica A: Statistical Mechanics and its Applications*. ISSN: 03784371. DOI: 10.1016/j.physa.2006.03.029.

Eckart, Carl and Gale Young (1936). "The approximation of one matrix by another of lower rank". In: *Psychometrika* 1.3, pp. 211–218.

Eckmann, J. P., O. Oliffson Kamphorst, and D. Ruelle (1987). "Recurrence plots of dynamical systems". In: *EPL*. ISSN: 12864854. DOI: 10.1209/0295-5075/4/9/004.

Eckmann, J. P. and D. Ruelle (1985). "Ergodic theory of chaos and strange attractors". In: *Reviews of Modern Physics*. ISSN: 00346861. DOI: 10.1103/RevModPhys.57.617.

— (1992). "Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems". In: *Physica D: Nonlinear Phenomena*. ISSN: 01672789. DOI: 10.1016/0167-2789(92)90023-G.

Evangelou, N et al. (2021). *Double Diffusion Maps for Scientific Computation in Latent Space*.

Eydeland, Alexander and Krzysztof Wolyniec (2003). *Energy and Power Risk Management, New Developments in Modeling, Pricing, and Hedging*. ISBN: 3175723993.

Fama, Eugene F. (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work". In: *The Journal of Finance*. ISSN: 00221082. DOI: 10.2307/2325486.

— (1998). "Market efficiency, long-term returns, and behavioral finance". In: *Journal of Financial Economics*. ISSN: 0304405X. DOI: 10.1016/s0304-405x(98)00026-9.

Fama, Eugene F. and Kenneth R. French (1993). "Common risk factors in the returns on stocks and bonds". In: *Journal of Financial Economics*. ISSN: 0304405X. DOI: 10.1016/0304-405X(93)90023-5.

Farbman, Zeev, Raanan Fattal, and Dani Lischinski (2010). "Diffusion maps for edge-aware image editing". In: *ACM Transactions on Graphics*. ISBN: 9781450304399. DOI: 10.1145/1866158.1866171.

Fischer, Thomas and Christopher Krauss (2018). "Deep learning with long short-term memory networks for financial market predictions". In: *European Journal of Operational Research*. ISSN: 03772217. DOI: 10.1016/j.ejor.2017.11.054.

Fraser, Andrew M. and Harry L. Swinney (1986). "Independent coordinates for strange attractors from mutual information". In: *Physical Review A*. ISSN: 10502947. DOI: 10.1103/PhysRevA.33.1134.

Gagnon, Joseph et al. (2012). "Large-Scale Asset Purchases by the Federal Reserve: Did They Work?" In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.1952095.

Gajamannage, Kelum, Randy Paffenroth, and Erik M Bollt (2019). "A nonlinear dimensionality reduction framework using smooth geodesics". In: *Pattern Recognition* 87, pp. 226–236.

Gao, Yuan and Han Lin Shang (2017). "Multivariate functional time series forecasting: Application to age-specific mortality rates". In: *Risks*. ISSN: 22279091. DOI: 10.3390/risks5020021.

Gao, Yue et al. (2020). "Machine learning based early warning system enables accurate mortality risk prediction for COVID-19". In: *Nature Communications* 11.1. DOI: 10.1038/s41467-020-18684-2. URL: https://doi.org/10.1038\%2Fs41467-020-18684-2.

Gardini, L., R. Lupini, and M. G. Messia (1989). "Hopf bifurcation and transition to chaos in Lotka-Volterra equation". In: *Journal of Mathematical Biology*. ISSN: 03036812. DOI: 10.1007/BF00275811.

Gardner, Everette S. (2006). "Exponential smoothing: The state of the art-Part II". In: *International Journal of Forecasting*. ISSN: 01692070. DOI: 10.1016/j.ijforecast.2006.03.005.

Gatev, Evan, William N. Goetzmann, and K. Geert Rouwenhorst (2006). *Pairs trading: Performance of a relative-value arbitrage rule*. DOI: 10.1093/rfs/hhj020.

Gençay, Ramazan, Faruk Selçuk, and Brandon Whitcher (2003). "Systematic risk and timescales". In: *Quantitative Finance*. ISSN: 14697688. DOI: 10.1088/1469-7688/3/2/305.

Gers, F. A. and J. Schmidhuber (2001). "LSTM recurrent networks learn simple context-free and context-sensitive languages". In: *IEEE Transactions on Neural Networks*. ISSN: 10459227. DOI: 10.1109/72.963769.

Gers, Felix A., Jürgen Schmidhuber, and Fred Cummins (2000). "Learning to forget: Continual prediction with LSTM". In: *Neural Computation*. ISSN: 08997667. DOI: 10.1162/089976600300015015.

Gers, Felix A., Nicol N. Schraudolph, and Jürgen Schmidhuber (2003). "Learning precise timing with LSTM recurrent networks". In: *Journal of Machine Learning Research*. ISSN: 15324435. DOI: 10.1162/153244303768966139.

Geweke, John and Guofu Zhou (1996). "Measuring the Pricing Error of the Arbitrage Pricing Theory". In: *Review of Financial Studies*. ISSN: 08939454. DOI: 10.1093/rfs/9.2.557.

Gilbert, Marius et al. (2020). "Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study". In: *The Lancet* 395.10227, pp. 871–877. DOI: 10.1016/s0140-6736(20)30411-6. URL: https://doi.org/10.1016\%2Fs0140-6736\%2820\%2930411-6.

Giordano, Giulia et al. (2020). "Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy". In: *Nature Medicine* 26.6, pp. 855–860. DOI: 10.1038/s41591-020-0883-7. URL: https://doi.org/10.1038\%2Fs41591-020-0883-7.

Girosi, F. and T. Poggio (1990). "Networks and the best approximation property". In: *Biological Cybernetics*. ISSN: 03401200. DOI: 10.1007/BF00195855.

Girosi, Federico, Michael Jones, and Tomaso Poggio (1993). "Priors, Stabilizers and Basis Functions: from regularization to radial, tensor and additive splines". In: *A.I. Memo No.1430*.

Gneiting, Tilmann, William Kleiber, and Martin Schlather (2010). "Matérn cross-covariance functions for multivariate random fields". In: *Journal of the American Statistical Association*. ISSN: 01621459. DOI: 10.1198/jasa.2010.tm09420.

Gong, Mengchun et al. (2020). "Conducting clinical studies during the epidemics of communicable diseases: perspectives of methodology and health economics". In: *Nan fang yi ke da xue xue bao = Journal of Southern Medical University*. ISSN: 16734254. DOI: 10.12122/j.issn.1673-4254.2020.03.11.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. http://www.deeplearningbook.org. MIT Press.

Gradojevic, Nikola and Marko Caric (2017). "Predicting Systemic Risk with Entropic Indicators". In: *Journal of Forecasting*. ISSN: 1099131X. DOI: 10.1002/for.2411.

Gradojevic, Nikola and Ramazan Gençay (2011). "Financial applications of nonextensive entropy". In: *IEEE Signal Processing Magazine*. ISSN: 10535888. DOI: 10.1109/MSP.2011.941843.

Graf, Siegried (1987). "Statistically self-similar fractals". In: *Probability Theory and Related Fields*. ISSN: 01788051. DOI: 10.1007/BF00699096.

Granger, C. W. J. (1969). "Investigating Causal Relations by Econometric Models and Cross-spectral Methods". In: *Econometrica*. ISSN: 00129682. DOI: 10.2307/1912791.

Grassberger, Peter and Itamar Procaccia (1983). "Characterization of strange attractors". In: *Physical Review Letters*. ISSN: 00319007. DOI: 10.1103/PhysRevLett.50.346.

Greenspan, Alan (2001). "Testimony of Federal Reserve Officials". In: *Federal Reserve Bulletin*. ISSN: 0014-9209.

Greff, Klaus et al. (2016). "LSTM: A search space odyssey". In: *IEEE transactions on neural networks and learning systems* 28.10, pp. 2222–2232.

Guercio, Diane Del and Jonathan Reuter (2014). "Mutual fund performance and the incentive to generate alpha". In: *Journal of Finance*. ISSN: 15406261. DOI: 10.1111/jofi.12048.

Guirao, Juan L.G., Raquel García-Rubio, and Juan A. Vera (2012). "On the dynamics of an inflation IS-LM model". In: *Economic Modelling*. ISSN: 02649993. DOI: 10.1016/j.econmod.2012.07.007.

Guttorp, Peter and Tilmann Gneiting (2006). "Studies in the history of probability and statistics XLIX on the Matérn correlation family". In: *Biometrika*. ISSN: 00063444. DOI: 10.1093/biomet/93.4.989.

Hand, David J. (2009). "Forecasting with Exponential Smoothing: The State Space Approach by Rob J. Hyndman, Anne B. Koehler, J. Keith Ord, Ralph D. Snyder". In: *International Statistical Review*. ISSN: 03067734. DOI: 10.1111/j.1751-5823.2009.00085_17.x.

Hébert, Benjamin (2018). "Moral hazard and the optimality of debt". In: *Review of Economic Studies*. ISSN: 1467937X. DOI: 10.1093/restud/rdx080.

Hegger, Rainer, Holger Kantz, and Thomas Schreiber (1999). "Practical implementation of nonlinear time series methods: The TISEAN package". In: *Chaos*. ISSN: 10541500. DOI: 10.1063/1.166424. arXiv: 9810005 [chao-dyn].

Heimowitz, Ayelet and Yonina C. Eldar (2018). "The nystrom extension for signals defined on a graph". In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. ISBN: 9781538646588. DOI: 10.1109/ICASSP.2018.8462408.

HHI (2018). URL: http://www.justice.gov/atr/public/guidelines/hhi.html.

Hiermeyer, Martin (2019). "An Improved IS-LM Model To Explain Quantitative Easing". In: *MPRA*.

Hochreiter, Sepp and Jürgen Schmidhuber (1997). "Long Short-Term Memory". In: *Neural Computation*. ISSN: 08997667. DOI: 10.1162/neco.1997.9.8.1735.

Hoertel, Nicolas et al. (2020). "A stochastic agent-based model of the SARS-CoV-2 epidemic in France". In: *Nature Medicine* 26.9, pp. 1417–1421. DOI: 10.1038/s41591-020-1001-6. URL: https://doi.org/10.1038\%2Fs41591-020-1001-6.

Høg, Esben P et al. (2006). "The fractional Ornstein-Uhlenbeck process: term structure theory and application". In: *Applied Economics*.

Holstein, Detlef and Holger Kantz (2009). "Optimal Markov approximations and generalized embeddings". In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*. ISSN: 15393755. DOI: 10.1103/PhysRevE.79.056202.

Honeine, Paul and Cedric Richard (2011). "Preimage problem in kernel-based machine learning". In: *IEEE Signal Processing Magazine*. ISSN: 10535888. DOI: 10.1109/MSP.2010.939747.

Hong, Ying Yi and Ching Ping Wu (2012). "Day-ahead electricity price forecasting using a hybrid principal component analysis network". In: *Energies*. ISSN: 19961073. DOI: 10.3390/en5114711.

Hotelling, H. (1933). "Analysis of a complex of statistical variables into principal components". In: *Journal of Educational Psychology*. ISSN: 00220663. DOI: 10.1037/h0071325.

HSIEH, DAVID A. (1991). "Chaos and Nonlinear Dynamics: Application to Financial Markets". In: *The Journal of Finance*. ISSN: 15406261. DOI: 10.1111/j.1540-6261.1991.tb04646.x.

Hsieh, David A. and Edgar E. Peters (1993). "Chaos and Order in the Capital Markets: A New View of Cycles, Prices, and Market Volatility." In: *The Journal of Finance*. ISSN: 00221082. DOI: 10.2307/2329084.

Huang, Yan, Gang Kou, and Yi Peng (2017). "Nonlinear manifold learning for early warnings in financial markets". In: *European Journal of Operational Research*. ISSN: 03772217. DOI: 10.1016/j.ejor.2016.08.058.

Hurst, H. E. (1951). "Long-Term Storage Capacity of Reservoirs". In: *Transactions of the American Society of Civil Engineers*. ISSN: 0066-0604. DOI: 10.1061/taceat.0006518.

Hutchinson, J.M. (1994). "A radial basis function approach to financial time series analysis". In: *Doctor*.

Ian Goodfellow, Yoshua Bengio, Aaron Courville (2015). "Deep Learning Book". In: *Deep Learning*. ISSN: 1437-7780. DOI: 10.1016/B978-0-12-391420-0.09987-X. arXiv: arXiv:1011.1669v3.

Imai, N et al. (2019). "Report 3: Transmissibility of 2019-nCoV". In: *Int J Infect Dis*. DOI: 10.1016/j.ijid.2020.01.050. URL: https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-2019-nCoV-transmissibility.pdf.

Iseri, Müge, Hikmet Caglar, and Nazan Caglar (2008). "A model proposal for the chaotic structure of Istanbul stock exchange". In: *Chaos, Solitons and Fractals*. ISSN: 09600779. DOI: 10.1016/j.chaos.2006.09.041.

Jagannathan, Ravi, Tongshu Ma, and Jiaqi Zhang (2019). "A Note on "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps"". In: *Journal of Finance*. ISSN: 15406261. DOI: 10.1111/jofi.12824.

Jollife, Ian T. and Jorge Cadima (2016). *Principal component analysis: A review and recent developments*. DOI: 10.1098/rsta.2015.0202.

Jones, Christopher S. (2001). "Extracting factors from heteroskedastic asset returns". In: *Journal of Financial Economics*. ISSN: 0304405X. DOI: 10.1016/S0304-405X(01)00079-4.

Jones, Peter W, Mauro Maggioni, and Raanan Schul (2008). "Manifold parametrizations by eigenfunctions of the Laplacian and heat kernels". In: *Proceedings of the National Academy of Sciences* 105.6, pp. 1803–1808.

Jurek, Jakub W. and Halla Yang (2011). "Dynamic Portfolio Selection in Arbitrage". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.882536.

K. Rohila, P. Gouthami, and Priya M (2014). "Dijkstra ' s Shortest Path Algorithm". In: *International Journal of Innovative Research in Computer and Communication Engineering*.

Kailath, Thoms (1971). "RKHS Approach to Detection and Estimation Problems-Part I: Deterministic Signals in Gaussian Noise". In: *IEEE Transactions on Information Theory*. ISSN: 15579654. DOI: 10.1109/TIT.1971.1054673.

Kandel, Shmuel and Robert F. Stambaugh (1996). *On the predictability of stock returns: An asset-allocation perspective*. DOI: 10.1111/j.1540-6261.1996.tb02689.x.

Kantz, Holger (1994). "A robust method to estimate the maximal Lyapunov exponent of a time series". In: *Physics Letters A*. ISSN: 03759601. DOI: 10.1016/0375-9601(94)90991-1.

Kaserer, Christoph (2019). "Lo, Andrew W.: Adaptive markets: financial evolution at the speed of thought". In: *Journal of Economics*. ISSN: 0931-8658. DOI: 10.1007/s00712-018-0627-z.

Katris, Christos (2021). "A time series-based statistical approach for outbreak spread forecasting: Application of COVID-19 in Greece". In: *Expert Systems with Applications* 166, p. 114077. DOI: 10.1016/j.eswa.2020.114077. URL: https://doi.org/10.1016\%2Fj.eswa.2020.114077.

Kecman, Vojislav (2002). *Learning and Soft Computing: Support Vector Machines, Neural Networks and Fuzzy Logic Models by V Kecman*. ISBN: 0262112558.

Kellogg, R Bruce, Tien-Yien Li, and James Yorke (1976). "A constructive proof of the Brouwer fixed-point theorem and computational results". In: *SIAM Journal on Numerical Analysis* 13.4, pp. 473–483.

Kennel, Matthew B., Reggie Brown, and Henry D.I. Abarbanel (1992). "Determining embedding dimension for phase-space reconstruction using a geometrical construction". In: *Physical Review A*. ISSN: 10502947. DOI: 10.1103/PhysRevA.45.3403.

Kiartzis, S. J., A. G. Bakirtzis, and V. Petridis (1995). "Short-term load forecasting using neural networks". In: *Electric Power Systems Research*. ISSN: 03787796. DOI: 10.1016/0378-7796(95)00920-D.

Kim, Tong Suk and Edward Omberg (1996). "Dynamic Nonmyopic Portfolio Behavior". In: *Review of Financial Studies*. ISSN: 08939454. DOI: 10.1093/rfs/9.1.141.

KK, Andreas S. Weigend, and Neil A. Gershenfeld (1994). "Time Series Prediction: Forecasting the Future and Understanding the Past. Proceedings of the NATO Advanced Research Workshop on a Comparative Time Series Analysis Held in Santa Fe, New Mexico, 14-17 May 1992." In: *Journal of the American Statistical Association*. ISSN: 01621459. DOI: 10.2307/2290964.

Klein, Roger W. and Vijay S. Bawa (1977). "The effect of limited information and estimation risk on optimal portfolio diversification". In: *Journal of Financial Economics*. ISSN: 0304405X. DOI: 10.1016/0304-405X(77)90031-9.

Kliková, B. and A. Raidl (2011). "Reconstruction of Phase Space of Dynamical Systems Using Method of Time Delay". In: *Proceedings of the 20th Annual Conference of Doctoral Students - WDS 2011*.

Koopman, B. O. (1931). "Hamiltonian Systems and Transformation in Hilbert Space". In: *Proceedings of the National Academy of Sciences*. ISSN: 0027-8424. DOI: 10.1073/pnas.17.5.315.

Koronaki, ED, AM Nikas, and AG Boudouvis (2020). "A data-driven reduced-order model of nonlinear processes based on diffusion maps and artificial neural networks". In: *Chemical Engineering Journal* 397, p. 125475.

Kramer, Mark A (1991). "Nonlinear principal component analysis using autoassociative neural networks". In: *AIChE journal* 37.2, pp. 233–243.

Krauss, Christopher (2017). "STATISTICAL ARBITRAGE PAIRS TRADING STRATEGIES: REVIEW AND OUTLOOK". In: *Journal of Economic Surveys*. ISSN: 14676419. DOI: 10.1111/joes.12153.

Krishnamurthy, Arvind and Annette Vissing-Jorgensen (2011). "The Effects of quantitative easing on interest Rates: Channels and implications for policy". In: *Brookings Papers on Economic Activity*. ISSN: 15334465. DOI: 10.1353/eca.2011.0019.

Kucharski, Adam J et al. (2020). "Early dynamics of transmission and control of COVID-19: a mathematical modelling study". In: *The Lancet Infectious Diseases* 20.5, pp. 553 –558. ISSN: 1473-3099. DOI: https://doi.org/10.1016/S1473-3099(20)30144-4. URL: http://www.sciencedirect.com/science/article/pii/S1473309920301444.

Kugiumtzis, D. (1996). "State space reconstruction parameters in the analysis of chaotic time series - The role of the time window length". In: *Physica D: Nonlinear Phenomena*. ISSN: 01672789. DOI: 10.1016/0167-2789(96)00054-1.

Kühnel, Wolfgang (2015). *Differential geometry*. Vol. 77. American Mathematical Soc.

kuleuven (n.d.). URL: https://set.kuleuven.be/ei/factsheets.

Kuniya, Toshikazu (2021). "Structure of epidemic models: toward further applications in economics". In: *Japanese Economic Review*. ISSN: 14685876. DOI: 10.1007/s42973-021-00094-8.

Kuttner, Kenneth N. (2018). "Outside the box: Unconventional monetary policy in the great recession and beyond". In: *Journal of Economic Perspectives*. DOI: 10.1257/jep.32.4.121.

Kutz, J. Nathan, Xing Fu, and Steven L. Brunton (2016). “Multiresolution dynamic mode decomposition”. In: *SIAM Journal on Applied Dynamical Systems*. ISSN: 15360040. DOI: 10.1137/15M1023543. arXiv: 1506.00564.

Kwok, James Tin Yau and Ivor Wai Hung Tsang (2004). “The pre-image problem in kernel methods”. In: *IEEE Transactions on Neural Networks*. ISSN: 10459227. DOI: 10.1109/TNN.2004.837781.

Kyrtsou, Catherine and Michel Terraza (2002). “Stochastic chaos or ARCH effects in stock series? A comparative study”. In: *International Review of Financial Analysis*. ISSN: 10575219. DOI: 10.1016/S1057-5219(02)00067-4.

Lafon, Stephane (2004). “Diffusion Maps and Geometric Harmonics”. PhD thesis.

Lee, Cheng Ming and Chia Nan Ko (2011). “Short-term load forecasting using lifting scheme and ARIMA models”. In: *Expert Systems with Applications*. ISSN: 09574174. DOI: 10.1016/j.eswa.2010.11.033.

Lee, John M (2013). “Smooth manifolds”. In: *Introduction to Smooth Manifolds*. Springer, pp. 1–31.

Lee, Seungjoon et al. (2020). “Coarse-scale PDEs from fine-scale observations via machine learning”. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 30.1, p. 013141.

Lehmberg, Daniel, Felix Dietrich, and Gerta Ḱoster (n.d.). *Modeling Melburnians - Using the Koopman operator to gain insight into crowd dynamics*.

Lehmberg, Daniel et al. (2020). “Datafold: data-driven models for point clouds and time series on manifolds”. In: *Journal of Open Source Software* 5.51, p. 2283. DOI: 10.21105/joss.02283. URL: https://doi.org/10.21105/joss.02283.

Lehoucq, Richard B, Danny C Sorensen, and Chao Yang (1998). *ARPACK users' guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods*. SIAM.

Leitch, Gordon and J Tanner (1991). “Economic Forecast Evaluation: Profits versus the Conventional Error Measures”. In: *The American Economic Review*. ISSN: 0002-8282. DOI: 10.2307/2006520.

Leon-Delgado, Homero de et al. (2018). “Multivariate statistical inference in a radial basis function neural network”. In: *Expert Systems with Applications*. ISSN: 09574174. DOI: 10.1016/j.eswa.2017.10.024.

Leung, Henry, Titus Lo, and Sichun Wang (2001). “Prediction of noisy chaotic time series using an optimal radial basis function neural network”. In: *IEEE Transactions on Neural Networks*. ISSN: 10459227. DOI: 10.1109/72.950144.

Leung, Kathy et al. (2020). “First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment”. In: *The Lancet* 395.10233, pp. 1382–1393. DOI: 10.1016/s0140-6736(20)30746-7. URL: https://doi.org/10.1016\%2Fs0140-6736\%2820\%2930746-7.

Lian, Wenzhao et al. (2015). “Multivariate time-series analysis and diffusion maps”. In: *Signal Processing*. ISSN: 01651684. DOI: 10.1016/j.sigpro.2015.04.003.

Lilly, Jonathan M. et al. (2017). “Fractional Brownian motion, the Matérn process, and stochastic modeling of turbulent dispersion”. In: *Nonlinear Processes in Geophysics*. ISSN: 16077946. DOI: 10.5194/npg-24-481-2017. arXiv: 1605.01684.

Limits, Stochastic and Calculus in Mean Square (1973). “Stochastic Limits and Calculus in Mean Square”. In: *Mathematics in Science and Engineering*. ISSN: 00765392. DOI: 10.1016/S0076-5392(08)60159-9.

Lin, Kevin K and Fei Lu (2021). “Data-driven model reduction, Wiener projections, and the Koopman-Mori-Zwanzig formalism”. In: *Journal of Computational Physics* 424, p. 109864.

Lin, Ruei Sung et al. (2006). "Learning nonlinear manifolds from time series". In: ISBN: 3540338349. DOI: 10.1007/11744047_19.

Liu, Jun and Allan Timmermann (2013). "Optimal convergence trade strategies". In: *Review of Financial Studies*. ISSN: 08939454. DOI: 10.1093/rfs/hhs130.

Liu, Ping et al. (2014). "Coarse-grained variables for particle-based models: diffusion maps and animal swarming simulations". In: *Computational Particle Mechanics* 1.4, pp. 425–440.

L.K.Saul and S.T.Roweis (2003). "Think globally, fit locally: unsupervised learning of low dimensional manifolds". In: *Journal of Machine Learning Research*.

Lo, Andrew W. (2002). "The Statistics of Sharpe Ratios". In: *Financial Analysts Journal*. ISSN: 0015198X. DOI: 10.2469/faj.v58.n4.2453.

Lo, Andrew W. and A. Craig MacKinlay (1988). "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test". In: *Review of Financial Studies*. ISSN: 0893-9454. DOI: 10.1093/rfs/1.1.41.

Lo, K. L. and Y. K. Wu (2004). "Analysis of relationships between hourly electricity price and load in deregulated real-time power markets". In: DOI: 10.1049/ip-gtd:20040613.

Lukoševičius, Mantas and Herbert Jaeger (2009). "Reservoir computing approaches to recurrent neural network training". In: *Computer Science Review* 3.3, pp. 127–149.

Maaten, Laurens Van Der and Geoffrey Hinton (2008). "User ' s Guide for t-SNE Software". In: *Structure*.

Mahi, Masnun et al. (2021). "A bibliometric analysis of pandemic and epidemic studies in economics: future agenda for COVID-19 research". In: *Social Sciences & Humanities Open*. ISSN: 25902911. DOI: 10.1016/j.ssaho.2021.100165.

Majdisova, Zuzana and Vaclav Skala (2017). "Radial basis function approximations: comparison and applications". In: *Applied Mathematical Modelling*. ISSN: 0307904X. DOI: 10.1016/j.apm.2017.07.033. arXiv: 1806.07705.

Malkiel, Burton G. and Eugene F. Fama (1970). "EFFICIENT CAPITAL MARKETS: A REVIEW OF THEORY AND EMPIRICAL WORK*". In: *The Journal of Finance*. ISSN: 0022-1082. DOI: 10.1111/j.1540-6261.1970.tb00518.x.

Mandelbrot, Benoit B. (1985). "Self-affine fractals and fractal dimension". In: *Physica Scripta*. ISSN: 14024896. DOI: 10.1088/0031-8949/32/4/001.

Mandelbrot, Benoit B. and James R. Wallis (1968). "Noah, Joseph, and Operational Hydrology". In: *Water Resources Research*. ISSN: 19447973. DOI: 10.1029/WR004i005p00909.

— (1969). "Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence". In: *Water Resources Research*. ISSN: 19447973. DOI: 10.1029/WR005i005p00967.

Maneewongvatana, Songrit and David M Mount (2002). "Analysis of approximate nearest neighbor searching with clustered point sets". In: *Data Structures, Near Neighbor Searches, and Methodology* 59, pp. 105–123.

Mann, Jordan and J. Nathan Kutz (2016). "Dynamic mode decomposition for financial trading strategies". In: *Quantitative Finance*. ISSN: 14697696. DOI: 10.1080/14697688.2016.1170194. arXiv: 1508.04487.

Marcellino, Massimiliano, James H Stock, and Mark W Watson (2006). "A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series". In: *Journal of econometrics* 135.1-2, pp. 499–526.

Markowitz, H. (1952). "Portfolio Selection Harry Markowitz". In: *The Journal of Finance*.

Matteo, T. Di, T. Aste, and Michel M. Dacorogna (2005). "Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development". In: *Journal of Banking and Finance*. ISSN: 03784266. DOI: 10.1016/j.jbankfin.2004.08.004.

Max Simchowitz Horia Mania, Stephen Tu Michael I. Jordan Benjamin Recht (2018). "Learning Without Mixing: Towards A Sharp Analysis of Linear System Identification". In: *arxiv*. URL: https://arxiv.org/abs/1802.08334.

May, Robert M. (1976). "Simple mathematical models with very complicated dynamics". In: *Nature*. ISSN: 00280836. DOI: 10.1038/261459a0.

McCaffrey, Daniel F. et al. (1992). "Estimating the Lyapunov exponent of a chaotic system with nonparametric regression". In: *Journal of the American Statistical Association*. ISSN: 1537274X. DOI: 10.1080/01621459.1992.10475270.

Mcdonald, Daniel J., Cosma Rohilla Shalizi, and Mark Schervish (2017). "Nonparametric risk bounds for time-series forecasting". In: *Journal of Machine Learning Research*. ISSN: 15337928.

McKinney, Wes (2010). "Data Structures for Statistical Computing in Python". In: *Proceedings of the 9th Python in Science Conference*. DOI: 10.25080/majora-92bf1922-00a.

Mead, A. (1992). "Review of the Development of Multidimensional Scaling Methods". In: *The Statistician*. ISSN: 00390526. DOI: 10.2307/2348634.

Medved, Andrei, Riley Davis, and Paula A. Vasquez (2020). "Understanding fluid dynamics from Langevin and Fokker–Planck equations". In: *Fluids*. ISSN: 23115521. DOI: 10.3390/fluids5010040.

Mejia, Carlos and Carlos Zapata Quimbayo (2018). "The Trending Ornstein-Uhlenbeck Process: A Technical Note". In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.3263789.

Menkhoff, Lukas et al. (2012). "Carry Trades and Global Foreign Exchange Volatility". In: *Journal of Finance*. ISSN: 00221082. DOI: 10.1111/j.1540-6261.2012.01728.x.

Mercer, J. (1909). "XVI. Functions of positive and negative type, and their connection the theory of integral equations". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*. ISSN: 0264-3952. DOI: 10.1098/rsta.1909.0016.

Merton, Robert C. (1980). "On estimating the expected return on the market: An exploratory investigation". In: *Topics in Catalysis*. ISSN: 0304405x. DOI: 10.1016/0304-405X(80)90007-0.

Mezić, Igor (2013). "Analysis of fluid flows via spectral properties of the Koopman operator". In: *Annual Review of Fluid Mechanics* 45, pp. 357–378.

Micchelli, Charles A. (1986). "Interpolation of scattered data: Distance matrices and conditionally positive definite functions". In: *Constructive Approximation*. ISSN: 01764276. DOI: 10.1007/BF01893414.

Mika, Sebastian et al. (1999a). "Fisher discriminant analysis with kernels". In: *Neural Networks for Signal Processing - Proceedings of the IEEE Workshop*. DOI: 10.1109/nnsp.1999.788121.

Mika, Sebastian et al. (1999b). "Kernel PCA and de-noising in feature spaces". In: *Advances in Neural Information Processing Systems*. ISBN: 0262112450.

Ming, Wai-Kit, Jian Huang, and Casper J. P. Zhang (2020). "Breaking down of the healthcare system: Mathematical modelling for controlling the novel coronavirus (2019-nCoV) outbreak in Wuhan, China". In: *Int J Infect Dis*. DOI: 10.1101/2020.01.27.922443. URL: https://doi.org/10.1101\%2F2020.01.27.922443.

Mirsky, Leon (1960). "Symmetric gauge functions and unitarily invariant norms". In: *The quarterly journal of mathematics* 11.1, pp. 50–59.

Miskolczi, Panna (2017). "Note on simple and logarithmic return". In: *Applied Studies in Agribusiness and Commerce*. ISSN: 1789-221X. DOI: 10.19041/apstract/2017/1-2/16.

Monnig, Nathan D, Bengt Fornberg, and Francois G Meyer (2014). "Inverting nonlinear dimensionality reduction with scale-free radial basis function interpolation". In: *Applied and Computational Harmonic Analysis* 37.1, pp. 162–170.

Moore, Derek W. and Edward A. Spiegel (1966). "A Thermally Excited Non-Linear Oscillator". In: *The Astrophysical Journal*. ISSN: 0004-637X. DOI: 10.1086/148562.

Murray, Eleanor J (2020). "Calls for Epidemic-Related Economics". In: *The Journal of Economic Perspectives*.

Nadler, Boaz et al. (2006). "Diffusion maps, spectral clustering and reaction coordinates of dynamical systems". In: *Applied and Computational Harmonic Analysis* 21.1, pp. 113–127.

NASA (n.d.). URL: https://earthobservatory.nasa.gov/features/WxForecasting/wx2.php.

Nathan Kutz, J, Joshua L Proctor, and Steven L Brunton (2018). "Applied Koopman theory for partial differential equations and data-driven modeling of spatio-temporal systems". In: *Complexity* 2018.

Nicolaou, Nicoletta and Timothy G Constandinou (2016). "A nonlinear causality estimator based on non-parametric multiplicative regression". In: *Frontiers in neuroinformatics* 10, p. 19.

Nychka, Douglas et al. (1992). "Finding Chaos in Noisy Systems". In: *Journal of the Royal Statistical Society: Series B (Methodological)*. ISSN: 0035-9246. DOI: 10.1111/j.2517-6161.1992.tb01889.x.

Nyström, Evert Johannes (1929). *Über die praktische auflösung von linearen integralgleichungen mit anwendungen auf randwertaufgaben der potentialtheorie*. Akademische Buchhandlung.

OECD (2014). *Short-term interest rates*. DOI: https://doi.org/https://doi.org/10.1787/2cc37d77-en. URL: https://www.oecd-ilibrary.org/content/data/2cc37d77-en.

Opong, Kwaku K. et al. (1999a). "The behaviour of some UK equity indices: An application of Hurst and BDS tests". In: *Journal of Empirical Finance*. ISSN: 09275398. DOI: 10.1016/S0927-5398(99)00004-3.

— (1999b). "The behaviour of some UK equity indices: An application of Hurst and BDS tests". In: *Journal of Empirical Finance*. ISSN: 09275398. DOI: 10.1016/S0927-5398(99)00004-3.

Ord, J. K. et al. (1997). "Estimation and prediction for a class of dynamic nonlinear statistical models". In: *Journal of the American Statistical Association*. ISSN: 1537274X. DOI: 10.1080/01621459.1997.10473684.

Orzeszko, Witold (2008). "The new method of measuring the effects of noise reduction in chaotic data". In: *Chaos, Solitons and Fractals*. ISSN: 09600779. DOI: 10.1016/j.chaos.2007.06.059.

Oum, Yumi, Shmuel Oren, and Shijie Deng (2006). "Hedging quantity risks with standard power options in a competitive wholesale electricity market". In: *Naval Research Logistics*. ISSN: 0894069X. DOI: 10.1002/nav.20184.

Packard, N. H. et al. (1980). "Geometry from a time series". In: *Physical Review Letters*. ISSN: 00319007. DOI: 10.1103/PhysRevLett.45.712.

Panagiotis Papaioannou Ronen Talmon, Ioannis Kevrekidis Constantinos Siettos (2021). "Time Series Forecasting Using Manifold Learning". In.

Papaioannou, G. and A. Karytinos (1995). "NONLINEAR TIME SERIES ANALYSIS OF THE STOCK EXCHANGE: THE CASE OF AN EMERGING MARKET". In: *International Journal of Bifurcation and Chaos*. ISSN: 0218-1274. DOI: 10.1142/s0218127495001186.

Papaioannou, George et al. (2016). "Analysis and Modeling for Short- to Medium-Term Load Forecasting Using a Hybrid Manifold Learning Principal Component Model and Comparison with Classical Statistical Models (SARIMAX, Exponential Smoothing) and Artificial Intelligence Models (ANN, SVM): The Case of Greek Electricity Market". In: *Energies*. ISSN: 1996-1073. DOI: 10.3390/en9080635.

Papaioannou, George P. et al. (2018). "Detecting the impact of fundamentals and regulatory reforms on the Greek wholesale electricity market using a SARMAX/GARCH model". In: *Energy*. ISSN: 03605442. DOI: 10.1016/j.energy.2017.10.064.

Papaioannou, Panagiotis et al. (2013). "Can social microblogging be used to forecast intraday exchange rates?" In: *NETNOMICS: Economic Research and Electronic Networking*. ISSN: 13859587. DOI: 10.1007/s11066-013-9079-3.

Papaioannou, Panagiotis et al. (2017). "S&P500 Forecasting and trading using convolution analysis of major asset classes". In: *Procedia Computer Science*. DOI: 10.1016/j.procs.2017.08.307.

Pappas, S. Sp et al. (2010). "Electricity demand load forecasting of the Hellenic power system using an ARMA model". In: *Electric Power Systems Research*. ISSN: 03787796. DOI: 10.1016/j.epsr.2009.09.006.

Park, Jin Hong, T. N. Sriram, and Xiangrong Yin (2010). "Dimension reduction in time series". In: *Statistica Sinica*. ISSN: 10170405.

Parlitz, Ulrich and Christian Merkwirth (2000). "Prediction of Spatiotemporal Time Series Based on Reconstructed Local States". In: *Physical Review Letters*. ISSN: 00319007. DOI: 10.1103/PhysRevLett.84.1890.

Pasini, G. (2017). "PRINCIPAL COMPONENT ANALYSIS FOR STOCK PORTFOLIO MANAGEMENT". In: *International Journal of Pure and Apllied Mathematics*. ISSN: 1311-8080. DOI: 10.12732/ijpam.v115i1.12.

Pathak, Jaideep et al. (2018). "Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach". In: *Physical review letters* 120.2, p. 024102.

Paul, Rahul and Stephan K. Chalup (2017). "A study on validating non-linear dimensionality reduction using persistent homology". In: *Pattern Recognition Letters*. ISSN: 01678655. DOI: 10.1016/j.patrec.2017.09.032.

Pearson, Karl (1901). " LIII. On lines and planes of closest fit to systems of points in space ". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. ISSN: 1941-5982. DOI: 10.1080/14786440109462720.

Pecora, Louis M. et al. (1997). "Fundamentals of synchronization in chaotic systems, concepts, and applications". In: *Chaos*. ISSN: 10541500. DOI: 10.1063/1.166278.

Pecora, Louis M. et al. (2007). "A unified approach to attractor reconstruction". In: *Chaos*. ISSN: 10541500. DOI: 10.1063/1.2430294.

Pedregosa, F. et al. (2011a). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Pedregosa, Fabian et al. (2011b). "Scikit-learn: Machine learning in Python". In: *Journal of Machine Learning Research*. ISSN: 15324435. arXiv: 1201.0490.

Pegels, C. Carl (1969). "Exponential Forecasting : Some New Variations". In: *Management Science*. ISSN: 00251909.

Peng, C. K. et al. (1994). "Mosaic organization of DNA nucleotides". In: *Physical Review E*. ISSN: 1063651X. DOI: 10.1103/PhysRevE.49.1685.

Perchet, Romain, Raul Leote de Carvalho, and Pierre Moulin (2014). "Intertemporal risk parity: a constant volatility framework for factor investing". In: *The Journal of Investment Strategies*. ISSN: 20471238. DOI: 10.21314/jois.2015.036.

Peters, Edgar. E (2015). *Fractal market analysis : applying chaos theory to investment and economics*.

Petrella, Andrea and Alessandro Sapio (2012). "Assessing the impact of forward trading, retail liberalization, and white certificates on the Italian wholesale electricity prices". In: *Energy Policy*. ISSN: 03014215. DOI: 10.1016/j.enpol.2011.10.011.

Poggio, Tomaso and Federico Girosi (1989). *A Theory of Networks for Approximation and Learning*. Tech. rep.

— (1990). "Networks for Approximation arid Learning". In: *Proceedings of the IEEE*. ISSN: 15582256. DOI: 10.1109/5.58326.

Polk, Christopher, Samuel Thompson, and Tuomo Vuolteenaho (2006). "Cross-sectional forecasts of the equity premium". In: *Journal of Financial Economics*. ISSN: 0304405X. DOI: 10.1016/j.jfineco.2005.03.013.

Pozzi, F., T. Di Matteo, and T. Aste (2012). "Exponential smoothing weighted correlations". In: *European Physical Journal B*. ISSN: 14346028. DOI: 10.1140/epjb/e2012-20697-x.

Prabhakar, Sunil Kumar and Harikumar Rajaguru (2017). "Factor analysis, Hessian Local Linear Embedding and Isomap for epilepsy classification from EEG". In: *1st International Conference - EECon 2016: 2016 Electrical Engineering Conference*. ISBN: 9781509053957. DOI: 10.1109/EECon.2016.7830929.

Press, William H and Saul A Teukolsky (1990). "Fredholm and Volterra integral equations of the second kind". In: *Computers in Physics* 4.5, pp. 554–557.

Priestley, M. B. and T. Subba Rao (1969). "A Test for Non-Stationarity of Time-Series". In: *Journal of the Royal Statistical Society: Series B (Methodological)*. ISSN: 2517-6161. DOI: 10.1111/j.2517-6161.1969.tb00775.x.

Proctor, Joshua L., Steven L. Brunton, and J. Nathan Kutz (2016). "Dynamic mode decomposition with control". In: *SIAM Journal on Applied Dynamical Systems*. ISSN: 15360040. DOI: 10.1137/15M1013857. arXiv: 1409.6358.

Qian, Edward E. (2011). "On the Financial Interpretation of Risk Contribution: Risk Budgets Do Add Up". In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.684221.

Que, Qichao and Mikhail Belkin (2020). "Back to the Future: Radial Basis Function Network Revisited". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*. ISSN: 19393539. DOI: 10.1109/TPAMI.2019.2906594.

Rabin, Neta et al. (2016). "Earthquake-explosion discrimination using diffusion maps". In: *Geophysical Journal International* 207.3, pp. 1484–1492.

Rand, R. H. and P. J. Holmes (1980). "Bifurcation of periodic motions in two weakly coupled van der Pol oscillators". In: *International Journal of Non-Linear Mechanics*. ISSN: 00207462. DOI: 10.1016/0020-7462(80)90024-4.

Rasmussen, Carl Edward (2003). "Gaussian processes in machine learning". In: *Summer school on machine learning*. Springer, pp. 63–71.

Rasmussen, Carl Edward and Christopher K. I. Williams (2018). *Gaussian Processes for Machine Learning*. DOI: 10.7551/mitpress/3206.001.0001.

Raymond O.Wells, Jr. (2008). "Differential Analysis on Complex Manifolds". In: *Springer-Verlag*.

Remuzzi, Andrea and Giuseppe Remuzzi (2020). "COVID-19 and Italy: what next?" In: *The Lancet*. DOI: 10.1016/s0140-6736(20)30627-9. URL: https://doi.org/10.1016\%2Fs0140-6736\%2820\%2930627-9.

Richard, Cédric, José Carlos M. Bermudez, and Paul Honeine (2009). "Online prediction of time series data with kernels". In: *IEEE Transactions on Signal Processing*. ISSN: 1053587X. DOI: 10.1109/TSP.2008.2009895.

Rob J Hyndman and Athanasopoulos George (2018). "Forecasting: Principles and Practice". In: *Principles of Optimal Design*.

Roncalli, T. and G. Weisang (2016). "Risk parity portfolios with risk factors". In: *Quantitative Finance*. ISSN: 14697696. DOI: 10.1080/14697688.2015.1046907.

Roncalli, Thierry (2012). "Understanding the Impact of Weights Constraints in Portfolio Theory". In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.1761625.

Rosenblum, Michael G., Arkady S. Pikovsky, and Jürgen Kurths (1996). "Phase synchronization of chaotic oscillators". In: *Physical Review Letters*. ISSN: 10797114. DOI: 10.1103/PhysRevLett.76.1804.

Rosenstein, Michael T., James J. Collins, and Carlo J. De Luca (1993). "A practical method for calculating largest Lyapunov exponents from small data sets". In: *Physica D: Nonlinear Phenomena*. ISSN: 01672789. DOI: 10.1016/0167-2789(93)90009-P.

Rossa, Fabio Della et al. (2020). "A network model of Italy shows that intermittent regional strategies can alleviate the COVID-19 epidemic". In: *Nature Communications* 11.1. DOI: 10.1038/s41467-020-18827-5. URL: https://doi.org/10.1038\%2Fs41467-020-18827-5.

Roweis, Sam T and Lawrence K Saul (2000). "Nonlinear dimensionality reduction by locally linear embedding". In: *science* 290.5500, pp. 2323–2326.

Ruelle, David (1981). "Small random perturbations of dynamical systems and the definition of attractors". In: *Communications in Mathematical Physics*. ISSN: 00103616. DOI: 10.1007/BF01206949.

Rulkov, Nikolai F. et al. (1995). "Generalized synchronization of chaos in directionally coupled chaotic systems". In: *Physical Review E*. ISSN: 1063651X. DOI: 10.1103/PhysRevE.51.980.

Russo, Lucia et al. (2020). "Tracing day-zero and forecasting the COVID-19 outbreak in Lombardy, Italy: A compartmental modelling and numerical optimization approach". In: *PLOS ONE* 15.10. Ed. by Alberto d'Onofrio, e0240649. DOI: 10.1371/journal.pone.0240649. URL: https://doi.org/10.1371\%2Fjournal.pone.0240649.

Rustam, Furqan et al. (2020). "COVID-19 Future Forecasting Using Supervised Machine Learning Models". In: *IEEE Access* 8, pp. 101489–101499. DOI: 10.1109/access.2020.2997311. URL: https://doi.org/10.1109\%2Faccess.2020.2997311.

S., L. L. and Grace Wahba (1991). "Spline Models for Observational Data." In: *Mathematics of Computation*. ISSN: 00255718. DOI: 10.2307/2938687.

Sanche, Steven et al. (2020). "RESEARCH High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2". In: *Emerging Infectious Diseases*. ISSN: 10806059. DOI: 10.3201/eid2607.200282.

Sauer, Tim (1995). "Interspike interval embedding of chaotic signals". In: *Chaos*. ISSN: 10541500. DOI: 10.1063/1.166094.

Sauer, Tim, James A. Yorke, and Martin Casdagli (1991a). "Embedology". In: *Journal of Statistical Physics*. ISSN: 00224715. DOI: 10.1007/BF01053745.

— (1991b). "Embedology". In: *Journal of Statistical Physics*. ISSN: 00224715. DOI: 10.1007/BF01053745.

Schmid, P. J. et al. (2011). "Applications of the dynamic mode decomposition". In: *Theoretical and Computational Fluid Dynamics*. ISSN: 09354964. DOI: 10.1007/s00162-010-0203-9.

Schölkopf, Bernhard, Alexander Smola, and Klaus Robert Müller (1998). "Nonlinear Component Analysis as a Kernel Eigenvalue Problem". In: *Neural Computation*. ISSN: 08997667. DOI: 10.1162/089976698300017467.

Schölkopf, Bernhard, Alexander Smola, and Klaus-Robert Müller (1998). "Nonlinear component analysis as a kernel eigenvalue problem". In: *Neural computation* 10.5, pp. 1299–1319.

Schuster, H. G. and Wolfram Just (2005). *Deterministic Chaos: An Introduction: Fourth Edition*. ISBN: 3527404155. DOI: 10.1002/3527604804.

Seabold, Skipper and Josef Perktold (2010a). "Statsmodels: Econometric and Statistical Modeling with Python". In: *Proceedings of the 9th Python in Science Conference*. DOI: 10.25080/majora-92bf1922-011.

— (2010b). "Statsmodels: econometric and statistical modeling with python". In: *9th Python in Science Conference*.

Sharpe, William F. (1994). "The Sharpe Ratio". In: *The Journal of Portfolio Management*. ISSN: 0095-4918. DOI: 10.3905/jpm.1994.409501.

Shaw, Robert (1981). "Strange Attractors, Chaotic Behavior, and Information Flow". In: *Zeitschrift fur Naturforschung - Section A Journal of Physical Sciences*. ISSN: 18657109. DOI: 10.1515/zna-1981-0115.

Shiller, R. J. (2013a). "Shiller, R. J. (2012). Finance and the Good Society". In: *Revista Icade. Revista de las Facultades de Derecho y Ciencias Económicas y Empresariales*. ISSN: 2341-0841.

Shiller, Robert J. (2013b). *Capitalism and financial innovation*. DOI: 10.2469/faj.v69.n1.4.

Shreve, Steven, Prasad Chalasani, and Somesh Jha (1997). "Steven Shreve: Stochastic Calculus and Finance". In: *FinancialEbooks.NET*. ISSN: 01678019. DOI: 10.1007/bf00052459.

Sidi, Oana et al. (2011). "Unsupervised co-segmentation of a set of shapes via descriptor-space spectral clustering". In: *ACM Transactions on Graphics*. ISSN: 07300301. DOI: 10.1145/2024156.2024160.

Siettos, C. I., C. W. Gear, and I. G. Kevrekidis (2012). "An equation-free approach to agent-based computation: Bifurcation analysis and control of stationary states". In: *EPL*. ISSN: 02955075. DOI: 10.1209/0295-5075/99/48007.

Silva, Petrônio C.L. et al. (2020). "COVID-ABS: An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions". In: *Chaos, Solitons & Fractals* 139, p. 110088. DOI: 10.1016/j.chaos.2020.110088. URL: https://doi.org/10.1016\%2Fj.chaos.2020.110088.

Singer, Amit et al. (2009). "Detecting intrinsic slow variables in stochastic dynamical systems by anisotropic diffusion maps". In: *Proceedings of the National Academy of Sciences* 106.38, pp. 16090–16095.

Smith, Keith V., E. J. Elton, and M. J. Gruber (1982). "Modern Portfolio Theory and Investment Analysis." In: *The Journal of Finance*. ISSN: 00221082. DOI: 10.2307/2327857.

Smith, Leonard A. (1988). "Intrinsic limits on dimension calculations". In: *Physics Letters A*. ISSN: 03759601. DOI: 10.1016/0375-9601(88)90445-8.

Smola, Alexander J. and Risi Kondor (2003). "Kernels and regularization on graphs". In: *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*. DOI: 10.1007/978-3-540-45167-9_12.

Stanimirovic, Olja et al. (2006). "Linking PCA and time derivatives of dynamic systems". In: *Journal of Chemometrics*. ISSN: 08869383. DOI: 10.1002/cem.980.

Stanley, H. E. et al. (1999). "Econophysics: Can physicists contribute to the science of economics?" In: *Physica A: Statistical Mechanics and its Applications*. ISSN: 03784371. DOI: 10.1016/S0378-4371(99)00185-5.

Stewart, G. W. (1993). "On the early history of the singular value decomposition". In: *SIAM Review*. ISSN: 00361445. DOI: 10.1137/1035134.

Stoop, R. and J. Parisi (1991). "Calculation of Lyapunov exponents avoiding spurious elements". In: *Physica D: Nonlinear Phenomena*. ISSN: 01672789. DOI: 10.1016/0167-2789(91)90082-K.

Strang, G (1988).

Strogatz, Steven H. (2018). *Nonlinear Dynamics and Chaos*. DOI: 10.1201/9780429492563.

Systrom, Kevin (2017). URL: https://github.com/k-sys/covid-19/blob/master/Realtime\%20R0.ipynb.

Szarski, J. (2017). "Stanisław Zaremba (1863-1942)". In: *Wiadomości Matematyczne*. ISSN: 2080-5519. DOI: 10.14708/wm.v5i1.2887.

Takens, Floris (1981). "Detecting strange attractors in turbulence". In: *Dynamical systems and turbulence, Warwick 1980*. Springer, pp. 366–381.

Talmon, Ronen et al. (2015). "Manifold Learning for Latent Variable Inference in Dynamical Systems". In: *IEEE Transactions on Signal Processing*. ISSN: 1053587X. DOI: 10.1109/TSP.2015.2432731.

Tashman, Leonard J. (2000). "Out-of-sample tests of forecasting accuracy: An analysis and review". In: *International Journal of Forecasting*. ISSN: 01692070. DOI: 10.1016/S0169-2070(00)00065-0.

Taylor, James W. (2003). "Exponential smoothing with a damped multiplicative trend". In: *International Journal of Forecasting*. ISSN: 01692070. DOI: 10.1016/S0169-2070(03)00003-7.

— (2012). "Short-term load forecasting with exponentially weighted methods". In: *IEEE Transactions on Power Systems*. ISSN: 08858950. DOI: 10.1109/TPWRS.2011.2161780.

Taylor, James W. and Patrick E. McSharry (2007). "Short-term load forecasting methods: An evaluation based on European data". In: *IEEE Transactions on Power Systems*. ISSN: 08858950. DOI: 10.1109/TPWRS.2007.907583.

Tenenbaum, Joshua B, Vin De Silva, and John C Langford (2000). "A global geometric framework for nonlinear dimensionality reduction". In: *science* 290.5500, pp. 2319–2323.

Theiler, James (1990). "Statistical precision of dimension estimators". In: *Physical Review A*. ISSN: 10502947. DOI: 10.1103/PhysRevA.41.3038.

Thiem, Thomas N et al. (2020). "Emergent spaces for coupled oscillators". In: *Frontiers in Computational Neuroscience* 14, p. 36.

Tong, H. and K. S. Lim (2009). "Threshold autoregression, limit cycles and cyclical data". In: *Exploration of a Nonlinear World: An Appreciation of Howell Tong's Contributions to Statistics*. ISBN: 9789812836281. DOI: 10.1142/9789812836281_0002.

Trzcinka, Charles (1986). "On the Number of Factors in the Arbitrage Pricing Model". In: *The Journal of Finance*. ISSN: 15406261. DOI: 10.1111/j.1540-6261.1986.tb05041.x.

Tsallis, Constantino (1988). "Possible generalization of Boltzmann-Gibbs statistics". In: *Journal of Statistical Physics*. ISSN: 00224715. DOI: 10.1007/BF01016429.

Tsallis, Constantino, Murray Gell-Mann, and Yuzuru Sato (2005). "Asymptotically scale-invariant occupancy of phase space makes the entropy Sq extensive". In:

Proceedings of the National Academy of Sciences of the United States of America. ISSN: 00278424. DOI: 10.1073/pnas.0503807102. arXiv: 0502274 [cond-mat].

Tsekouras, G. A. and Constantino Tsallis (2005). "Generalized entropy arising from a distribution of q indices". In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*. ISSN: 15393755. DOI: 10.1103/PhysRevE.71.046144. arXiv: 0412329 [cond-mat].

Tsonis, A. A., J. B. Elsner, and K. P. Georgakakos (1993). "Estimating the dimension of weather and climate attractors: important issues about the procedure and interpretation". In: *Journal of the Atmospheric Sciences*. ISSN: 00224928. DOI: 10.1175/1520-0469(1993)050<2549:ETDOWA>2.0.CO;2.

Tsoumanis, A. C. et al. (2010). "Equation-free multiscale computations in social networks: From agent-based modeling to coarse-grained stability and bifurcation analysis". In: *International Journal of Bifurcation and Chaos*. ISSN: 02181274. DOI: 10.1142/S0218127410027945.

Vapnik, Vladimir N. (1999). *An overview of statistical learning theory*. DOI: 10.1109/72.788640.

Virtanen, Pauli, Ralf Gommers, and Travis E. Oliphant et al. (2020). "SciPy 1.0: fundamental algorithms for scientific computing in Python". In: *Nature Methods*. ISSN: 15487105. DOI: 10.1038/s41592-019-0686-2. arXiv: 1907.10121.

Viviani, Roberto, Georg Grön, and Manfred Spitzer (2005). "Functional principal component analysis of fMRI data". In: *Human Brain Mapping*. ISSN: 10659471. DOI: 10.1002/hbm.20074.

Vlachas, Pantelis R et al. (2018). "Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 474.2213, p. 20170844.

Vlachas, Pantelis R et al. (2020). "Backpropagation algorithms and reservoir computing in recurrent neural networks for the forecasting of complex spatiotemporal dynamics". In: *Neural Networks* 126, pp. 191–217.

Vlachos, I. and D. Kugiumtzis (2009). "State space reconstruction from multiple time series". In: *Topics on Chaotic Systems - Selected Papers from CHAOS 2008 International Conference*. ISBN: 9814271330. DOI: 10.1142/9789814271349_0043.

Wahyuningsih, Delpiah and Erzal Syahreza (2018). "Shortest Path Search Futsal Field Location With Dijkstra Algorithm". In: *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*. ISSN: 1978-1520. DOI: 10.22146/ijccs.34513.

Wan, Zhong Yi and Themistoklis P Sapsis (2017). "Reduced-space Gaussian Process Regression for data-driven probabilistic forecast of chaotic dynamical systems". In: *Physica D: Nonlinear Phenomena* 345, pp. 40–55.

Wang, Dingding et al. (2020). "An efficient mixture of deep and machine learning models for COVID-19 diagnosis in chest X-ray images". In: *PLOS ONE* 15.11. Ed. by Jeonghwan Gwak, e0242535. DOI: 10.1371/journal.pone.0242535. URL: https://doi.org/10.1371\%2Fjournal.pone.0242535.

Wang, Jianzhong (2012). *Geometric structure of high-dimensional data and dimensionality reduction*. Vol. 5. Springer.

Wang, Rui (2021). "Evaluating the Unconventional Monetary Policy of the Bank of Japan: A DSGE Approach". In: *Journal of Risk and Financial Management*. ISSN: 1911-8074. DOI: 10.3390/jrfm14060253.

Watanabe, Toshio (2021). "Reconsideration of the IS–LM model and limitations of monetary policy: a Tobin–Minsky model". In: *Evolutionary and Institutional Economics Review*. ISSN: 1349-4961. DOI: 10.1007/s40844-020-00189-8.

Weber, Arno (2017). "Annual Risk Measures and Related Statistics". In: *SSRN*. URL: http://dx.doi.org/10.2139/ssrn.3054517.

Weron, Rafał (2002). "Estimating long-range dependence: Finite sample properties and confidence intervals". In: *Physica A: Statistical Mechanics and its Applications*. ISSN: 03784371. DOI: 10.1016/S0378-4371(02)00961-5. arXiv: 0103510 [cond-mat].

Weron, Rafal and Beata Przybylowicz (2000a). "Hurst analysis of electricity price dynamics". In: *Physica A: Statistical Mechanics and its Applications*. ISSN: 03784371. DOI: 10.1016/S0378-4371(00)00231-4.

— (2000b). "Hurst analysis of electricity price dynamics". In: *Physica A: Statistical Mechanics and its Applications*. ISSN: 03784371. DOI: 10.1016/S0378-4371(00)00231-4.

Weron, Rafał (2006). *Modeling and forecasting electricity loads and prices: A statistical approach*. ISBN: 9781118673362. DOI: 10.1002/9781118673362.

Whitney, Hassler (1944). "The singularities of a smooth n-manifold in (2n-1)-space". In: *Annals of Mathematics*, pp. 247–293.

Williams, C. K. I. (1998). "Prediction with Gaussian Processes: From Linear Regression to Linear Prediction and Beyond". In: *Learning in Graphical Models*. DOI: 10.1007/978-94-011-5014-9_23.

Williams, Matthew O, Ioannis G Kevrekidis, and Clarence W Rowley (2015). "A data–driven approximation of the koopman operator: Extending dynamic mode decomposition". In: *Journal of Nonlinear Science* 25.6, pp. 1307–1346.

Willmore, T. J. (1987). "AN INTRODUCTION TO DIFFERENTIABLE MANIFOLDS AND RIEMANNIAN GEOMETRY (Second edition)". In: *Bulletin of the London Mathematical Society*. DOI: 10.1112/blms/19.5.489.

Wolf, Alan et al. (1985). "DETERMINING LYAPUNOV EXPONENTS FROM A TIME SERIES Alan WOLF -, Jack B. SWIFT, Harry L. SWINNEY and John A. VASTANO". In: *Physica D: Nonlinear Phenomena*.

Wolpert, Robert L. and Murad S. Taqqu (2005). "Fractional Ornstein-Uhlenbeck Lévy processes and the Telecom process: Upstairs and downstairs". In: *Signal Processing*. ISSN: 01651684. DOI: 10.1016/j.sigpro.2004.09.016.

Wu, Joseph T, Kathy Leung, and Gabriel M Leung (2020). "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study". In: *The Lancet*. DOI: 10.1016/s0140-6736(20)30260-9. URL: https://doi.org/10.1016\%2Fs0140-6736\%2820\%2930260-9.

Wyffels, Francis and Benjamin Schrauwen (2010). "A comparative study of reservoir computing strategies for monthly time series prediction". In: *Neurocomputing* 73.10-12, pp. 1958–1964.

Xu, Liping and Jiaowan Luo (2018). "Stochastic differential equations driven by fractional Brownian motion". In: *Statistics and Probability Letters*. ISSN: 01677152. DOI: 10.1016/j.spl.2018.06.012.

Yadaw, Arjun S et al. (2020). "Clinical features of COVID-19 mortality: development and validation of a clinical prediction model". In: *The Lancet Digital Health* 2.10, e516–e525. DOI: 10.1016/s2589-7500(20)30217-x. URL: https://doi.org/10.1016\%2Fs2589-7500\%2820\%2930217-x.

Yan L. Zhang HT., Goncalves J. et al. (2020). "An interpretable mortality prediction model for COVID-19 patients". In: *Nature Machine Intelligence* 2, pp. 283–288. DOI: 10.1038/s42256-020-0180-7.

Yang, Yang, Yanrong Yang, and Han Lin Shang (2021). “Feature extraction for functional time series: Theory and application to NIR spectroscopy data”. In: *Journal of Multivariate Analysis*. ISSN: 10957243. DOI: 10.1016/j.jmva.2021.104863.

Yannacopoulos, A. N. (2011). “Stochastic saddle paths and economic theory”. In: *Springer Proceedings in Mathematics*. ISSN: 21905622. DOI: 10.1007/978-3-642-14788-3_52.

Yannacopoulos, A. N. et al. (2000). “Conditions for soliton trapping in random potentials using lyapunov exponents of stochastic ODEs”. In: *Physics Letters, Section A: General, Atomic and Solid State Physics*. ISSN: 03759601. DOI: 10.1016/S0375-9601(00)00348-0.

Yannacopoulos, Athanasios N. (2008). “Rational expectations models: An approach using forward-backward stochastic differential equations”. In: *Journal of Mathematical Economics*. ISSN: 03044068. DOI: 10.1016/j.jmateco.2007.05.005.

Yannacopoulos, Athanasios N., Nikolaos E. Frangos, and Ioannis Karatzas (2011). “Wiener chaos solutions for linear backward stochastic evolution equations”. In: *SIAM Journal on Mathematical Analysis*. ISSN: 00361410. DOI: 10.1137/090750652.

Yao, Beibei et al. (2017). “Modified local linear embedding algorithm for rolling element bearing fault diagnosis”. In: *Applied Sciences (Switzerland)*. ISSN: 20763417. DOI: 10.3390/app7111178.

Yi, Wei (2019). URL: https://towardsdatascience.com/understanding-gaussian-process-the-socratic-way-ba02369d804.

Yousefpoor, P., M. S. Esfahani, and H. Nojumi (2008). “Looking for systematic approach to select chaos tests”. In: *Applied Mathematics and Computation*. ISSN: 00963003. DOI: 10.1016/j.amc.2007.08.070.

Yu, Feng and Xiaozhong Xu (2014). “A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network”. In: *Applied Energy*. ISSN: 03062619. DOI: 10.1016/j.apenergy.2014.07.104.

Yuan, Hsiang-Yu et al. (2020). “Estimating the risk on outbreak spreading of 2019-nCoV in China using transportation data”. In: DOI: 10.1101/2020.02.01.20019984. URL: https://www.medrxiv.org/content/early/2020/02/04/2020.02.01.20019984.

Yule, G. U. (2012). “ On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer’s Sunspot Numbers (Philosophical Transactions of the Royal Society of London , A, vol. 226, 1927, pp. 267–73) ”. In: *The Foundations of Econometric Analysis*. DOI: 10.1017/cbo9781139170116.013.

Zellner, Arnold and V. Karuppan Chetty (1965). “Prediction and Decision Problems in Regression Models from the Bayesian Point of View”. In: *Journal of the American Statistical Association*. ISSN: 1537274X. DOI: 10.1080/01621459.1965.10480817.

Zeroual, Abdelhafid et al. (2020). “Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study”. In: *Chaos, Solitons & Fractals* 140, p. 110121. DOI: 10.1016/j.chaos.2020.110121. URL: https://doi.org/10.1016\%2Fj.chaos.2020.110121.

Zhang, Sumin, Zhengming Ma, and Hengliang Tan (2018). “On the Equivalence of HLLE and LTSA”. In: *IEEE Transactions on Cybernetics*. ISSN: 21682267. DOI: 10.1109/TCYB.2017.2655338.

Zhao, Shi et al. (2020). “Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak”. In: *Int J Infect Dis*. DOI: 10.1101/2020.01.23.916395. URL: https://doi.org/10.1101\%2F2020.01.23.916395.

Zhu, Ciyou et al. (1997). “Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization”. In: *ACM Transactions on Mathematical Software*. ISSN: 00983500. DOI: 10.1145/279232.279236.