



**MACQUARIE**  
University

EXTENSION OF PERMUTATION ENTROPY  
AND ORDINAL PATTERN ANALYSIS WITH  
APPLICATION TO FINANCIAL TIME SERIES  
ANALYSIS

Xin Huang

A thesis submitted for the degree of PhD  
in the  
Macquarie Business School  
Department of Actuarial Studies and Business Analytics

November 16, 2021

# Contents

<b>Declaration</b>	<b>iv</b>
<b>Acknowledgements</b>	<b>v</b>
<b>Abstract</b>	<b>vi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Thesis outline . . . . .	3
<b>2 Literature review</b>	<b>4</b>
2.1 Rationale behind permutation entropy (PE) and its previous applications . .	5
2.1.1 Measuring complexity and predictability . . . . .	7
2.1.2 Detecting dynamical changes . . . . .	9
2.1.3 Identification of delayed feedback . . . . .	9
2.1.4 Measuring dependence structures for multivariate time series . . . . .	9
2.2 Empirical properties of financial time series . . . . .	10
2.2.1 Heavy tailed distribution of returns . . . . .	10
2.2.2 Insignificant correlation of returns . . . . .	11
2.2.3 Volatility clustering and long-memory properties . . . . .	11
2.2.4 Non-linearity . . . . .	12
2.2.5 Non-stationary . . . . .	14
2.2.6 Other properties . . . . .	15
2.2.7 Properties of high-frequency financial returns . . . . .	16
2.3 Financial time series models . . . . .	17
2.3.1 Linear models . . . . .	18
2.3.2 Conditional heteroscedastic models . . . . .	20
2.3.3 Support vector regression . . . . .	24
2.3.4 Gaussian process regression . . . . .	26
<b>3 PE and its variants</b>	<b>28</b>
3.1 Definition of PE . . . . .	29
3.2 Advantages of PE . . . . .	31
3.3 Relation of PE and ACF in a Gaussian process . . . . .	31
3.4 PE variant permutation dependence (PD) . . . . .	34
3.5 Comparison between PE, PD and ACF . . . . .	37
3.6 PE, PD and ACF under non-stationarity . . . . .	44

3.7	PE, PD and ordinal pattern distribution in ARMA and GARCH models . . .	45
3.8	PD visualization plot to reveal the deterministic relation captured by the forecasting model . . . . .	54
3.9	Multivariate PE . . . . .	57
<b>4</b>	<b>The PD model sufficiency test</b>	<b>59</b>
4.1	Conventional model comparison and evaluation approaches . . . . .	59
4.2	Limitations of Brock, Dechert and Scheinkman (BDS) test . . . . .	63
4.3	Rationale and specification of the PD model sufficiency test . . . . .	66
4.4	Simulation studies for the new test . . . . .	73
<b>5</b>	<b>Empirical application to high-frequency currency exchange series and sea surface temperature data</b>	<b>81</b>
5.1	Data description . . . . .	82
5.2	Modifications on PE and PD for equal entries . . . . .	84
5.3	Temporal dependence of 10-minute EUR/USD returns . . . . .	85
5.3.1	Preliminary analysis to test against randomness . . . . .	86
5.3.2	Discretization effect . . . . .	89
5.3.3	Temporal dependence in returns caused by structures in return magnitudes	95
5.3.4	Signs of returns . . . . .	96
5.3.5	Bivariate dependence relation between signs of return and return magnitude . . . . .	99
5.3.6	Summary . . . . .	102
5.4	Temporal dependence of EUR/USD 10-minute squared returns and 1-hour realized volatilities . . . . .	104
5.4.1	Preliminary analysis on 10-minute EUR/USD squared returns . . . .	106
5.4.2	Periodicity in squared returns . . . . .	107
5.4.3	Discretization effect in squared return dynamics and on deseasonalization procedures . . . . .	109
5.4.4	Temporal dependence structure in deseasonalized squared returns . .	112
5.4.5	Temporal dependence structure in 1-hour realized volatilities . . . .	115
5.4.6	Summary . . . . .	119
5.5	Modelling and forecasting of 10-minute returns . . . . .	120
5.5.1	The sufficiency of MA(1) model . . . . .	121
5.5.2	Estimated MA(1) dynamics versus real return dynamics . . . . .	122
5.5.3	Summary . . . . .	124

5.6	Modelling and forecasting of 10-minute squared returns and 1-hour aggregate volatilities . . . . .	126
5.6.1	Evaluation of various models in predicting 10-minute squared returns and 1-hour realized volatilities . . . . .	127
5.6.2	Reasons behind the insufficient predicting performance of the employed models . . . . .	136
5.6.3	Main obstacles to predicting return volatilities . . . . .	141
5.7	Ordinal analysis on sea surface temperature data . . . . .	142
<b>6</b>	<b>Conclusion</b>	<b>151</b>
6.1	Methodological contributions . . . . .	151
6.2	Main results in empirical analysis . . . . .	153

## Declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma in any university or equivalent institution, and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Xin Huang

## Acknowledgements

I would like to begin by thanking my supervisors, Professor Piet de Jong, Professor Hanlin Shang, Professor Deb Kane, Professor Paul Smith and Professor David Pitt for their constant support and academic guidance. Their wisdom, immense knowledge and kindness have guided me through my PhD journey.

I am grateful for the financial support of the International Macquarie University Research Excellence Scholarship, a Macquarie PhD Scholarship during my candidature, and the generous provision of financial data from the Securities Industry Research Centre of Asia-Pacific (SIRCA).

Most of all, I would like to express my deepest gratitude to my beloved parents for their constant and unconditional love, encouragement and support.

# Abstract

This thesis explores and extends the use of [Permutation Entropy \(PE\)](#), a complexity measure that has been extensively used in the biomedical and physical fields but less so in the area of financial time series. I make both methodological and empirical contributions to the literature, specifically in relation to financial time series analysis.

First, I establish the connection of PE with the mainstream financial time series models by demonstrating the relation of PE with the [Autocorrelation Function \(ACF\)](#) in the Gaussian process and the parametrizations of the [Autoregressive-moving-average \(ARMA\)](#) and the [Generalized Autoregressive Conditional Heteroskedastic \(GARCH\)](#) models. Additionally, I examine the way that PE responds to a number of commonly observed features of financial data, such as high-kurtosis and non-stationarity, in order to provide appropriate interpretation of this measure when it is used in empirical applications.

Second, I develop a new temporal dependence measure, [Permutation Dependence \(PD\)](#) by adjusting the specification of PE. PD is proposed as a remedy to the major drawbacks of PE, such as its insensitivity to slowly diminishing temporal structures and its non-monotonic corresponding to the strength of temporal dependence at the chosen delay/lag. I show that the PD measure indicates the temporal dependence of the observed time series at the selected delay/lag and it resembles the ACF in linear processes but is not grounded in detecting linear structures. Just like the ACF, by plotting PD against increasing delays, it can be used to reflect the evolution of temporal dependence relations as the lag between entries increases. Additionally, with the help of the new PD measure, I develop a visualization plot for the time series investigated to reveal the deterministic structures captured by different models. The visualization plot provides a universal framework to facilitate comparison between different deterministic relations postulated by both parametric and non-parametric models. Therefore it helps to explain the superior and suboptimal prediction performance of various models when applied to data with different properties.

My third methodological contribution is proposing a model sufficiency test using the ordinal pattern concept in PE to study a given model's point prediction accuracy. Compared to some classical model sufficiency tests, such as the [Broock et al. \(1996\)](#) test, our proposal does not require a sufficient model to eliminate all structures exhibited in the estimated residuals. When the additive innovations in the investigated data's underlying dynamics show a certain structure, such as higher-moment serial dependence, the [Broock et al. \(1996\)](#) test can lead to erroneous conclusions about the sufficiency of point predictors. Due to the structured innovations, inconsistency between the model sufficiency tests and prediction

accuracy criteria can occur. Our proposal overcomes this lack of cohesion between model and prediction evaluation approaches and remains valid when the underlying process has non-white additive innovation.

This thesis also provides critical empirical contributions. By making use of PE, PD and the new model sufficiency test on 10-minute EUR/USD exchange rate return and volatility series, I successfully apply the ordinal based analysis to real-world financial data. In particular, I investigate the existence and nature of the temporal dependence underlying the observed return and volatility dynamics. Additionally, by applying the newly proposed sufficiency test on high-frequency exchange rate return and volatility series, I study the sufficiency of a number of commonly used models in point forecasting the real-world financial series. The model sufficiency evaluation studies also lead me to uncover the main reasons behind each prediction model's sufficient/insufficient performance. My empirical results have profound implications for future studies as they identify the main obstacles to accurate prediction and replication of the dynamics of financial time series, thereby pointing out how to construct more capable and better performing models.

I also present a brief analysis using the ordinal pattern based tools on monthly sea surface temperature data to demonstrate the potential applications of the PE and PD measures in a wider range of disciplines.

# 1 Introduction

Permutation Entropy (PE) is a computer intensive tool for analyzing dependence in long sequences of data, including time series data. It has received little attention in the field of applied finance despite being widely and successfully used in biomedical and physical fields.

The PE statistic is based on relative ranks within increasingly sparsely spaced out subsequences derived from the original sequence. Using theoretical arguments one can establish the behaviour of the expected value of the PE statistic if the sequence is “random”. Non-randomness shows up in defined and potentially informative ways. Converting to ranks implies the analysis is not grounded in linear or quasi-linear analysis of relationships. Instead, it may be termed “non-parametric” since it does not require any prior knowledge or assumptions about the functional form of the system.

PE was originally proposed as a variant entropy measure intended to indicate the complexity and predictability of the dynamics underlying a time series characterized by a continuous probability space. By measuring the complexity and predictability of various signals, one can distinguish between time series generated from a chaotic (low dimensional deterministic) process and that from a pure random process. Over the past 18 years, since PE was first introduced, researchers have been exploring and extending the applications of PE. It has been applied in many fields, including biomedical systems and laser dynamics. The previous literature documents PE’s properties, such as robustness to noise and the ability to deal with highly non-linear complex dynamics. These make PE an exceptional tool in capturing the characteristics of complex systems, such as brain (electrical) activity ([Costa et al. 2002](#)) and heart rate rhythms ([Rubinov & Sporns 2010](#)). Among the few existing applications of PE in economic areas, PE is primarily used as a predictability measure to assess the efficiency of financial markets and indicate the level of market development ([Zunino et al. 2009](#)).

This thesis explores and examines the potential use of PE analysis in financial time series analysis, and makes both methodological and empirical contributions to the existing literature. Methodologically, I establish the connection between PE and the mainstream financial time series models by relating PE with the ACF in the Gaussian process and the parametrization in the Autoregressive-moving-average (ARMA) model and the Generalized- autoregressive-conditional-heteroscedastic (GARCH) model. In addition, I investigate how PE responds to a number of commonly observed properties of financial time series such as high kurtosis and non-stationarity and identify appropriate interpretations of results in PE analysis conducted on financial time series by incorporating mainstream financial time series perspectives. Since PE has a complex relation with the ACF even in a linear process, I develop a new measure,

Permutation Dependence (PD), by adjusting the specification of PE to directly measure the temporal dependence structures present in the dynamics underlying the observed time series at a selected delay/lag. I show that the newly proposed dependence measure PD resembles the ACF in the linear process but is not grounded in detecting linear structures. By plotting PE and PD against increasing delays, they can be used together to detect the temporal dependence structures (both linear and non-linear) underlying the observed time series and reflect how dependence structures diminish as the entries become further apart. Moreover, PD can indicate the existence of long-memory and aid in selecting optimal lag in constructing predicting models.

In addition to these methodological contributions, with the aid of the newly proposed statistic PD, I invent a visualization plot to reveal the deterministic structures captured by different models fitted to an observed time series. The newly proposed visualization requires no prior assumptions and knowledge of the predicting model's functional form, and provides a universal framework to facilitate comparison between different deterministic relations postulated by both parametric and non-parametric models. The visualization technique is of particular importance in understanding the prediction constructed by the non-parametric models that have no closed-form formula. Our invention helps to open the "Black Box" as many non-parametric models, such as machine learning approaches, are often referred to and sheds light in explaining their superior and suboptimal performance when applied to data with different properties. Lastly, enlightened by the ordinal pattern concept, I develop a new approach to evaluate the sufficiency of a model/predictor in terms of its ability to exploit the point prediction potential of the investigated time series.

Empirically, the most critical contribution of this thesis is in providing detailed demonstrations of how PE and the related ordinal pattern analysis reveal important information about the dynamics underlying the real-world financial series. I choose 10-minute interval EUR/USD exchange rate return series as my object of investigation. EUR/USD is the most heavily traded bilateral currency pairs, constituting around 24% of global foreign exchange transactions. Analysis of the exchange return series can be considered a prototype because it has many features and properties in common with other financial data. Further, the results obtained can be easily learned and used as references for a wide range of related studies. Past studies, including [Belaire-Franch & Opong \(2005\)](#) and [Chortareas et al. \(2011\)](#), conducted in the area of market efficiency strongly indicate the lack of temporal dependence in returns over a longer period, such as daily, weekly or monthly, thus intraday returns provide greater predicting and structure exploration potentials ([Chordia et al. 2005](#), [Lien & Xiang 2010](#)). With the aid of PE and our newly proposed measures, I am able to answer a number of fundamental

questions about the data investigated in this study:

- Do intraday EUR/USD return and volatility follow a purely random process or have temporal dependence structures?
- If temporal dependence structure is present in observed intraday returns and volatilities, what is the nature of the detected structure and what contributes to it?
- How predictable (point and density) is intraday return and volatility without using any exogenous variables as inputs?
- Could conventional models such as the ARMA model and the GARCH-class models fully replicate the underlying dynamics in the intraday EUR/USD return and volatilities?
- How well do the mainstream models such as the ARMA model, the GARCH model, [Gaussian Process Regression \(GPR\)](#) and [Support Vector Regression \(SVR\)](#) point predict EUR/USD intraday returns and volatilities? What is the reason behind their sufficient/insufficient point forecasting performance?

The answers to the above questions are crucial to understanding the dynamics of high-frequency financial time series, moreover they develop valuable insight into the high-frequency trading behaviours and the micro-structures in financial markets. Six non-overlapping 1-year period EUR/USD return series are used for analysis in order to search for more general and persistent conclusions about the dynamics of EUR/USD intraday returns across different time periods.

Additionally, PE and PD measures are also applied on monthly [sea surface temperature \(SST\)](#) data. The empirical analysis results showed that the ordinal based measures are not only very informative and insightful in the area of applied finance, but can be easily extended to a wide range of fields on various types of time series. The empirical analysis reflects the level of seasonality and predictability of the SST recorded in different Pacific areas, and helps to identify the maximum time frame in which they can be predicted. Moreover, the bivariate PE can even quantify the level of dependence between the SST indexes in different regions thereby aiding in the selection of criteria for identifying and characterizing El Niño and La Niña events.

## 1.1 Thesis outline

The thesis is structured as follows.

**Chapter 2** summarizes prior studies relevant to the scope of this thesis. The studies

summarized in chapter 2 include the initial rationale for inventing PE and its previous applications, the empirical properties of financial time series and the commonly used models in financial time series analysis.

**Chapter 3** focuses on introducing all theoretical aspects of PE, including providing its definitions and advantages, as well as relating it to the ACF in the Gaussian process. Additionally, the new temporal dependence measure PD is also proposed in this chapter, followed by a comparison between statistics PE, PD and ACF. Moreover, simulation studies of PE and PD in the ARMA and GARCH models are included to examine the way that PE and PD respond to the change of key features of time series dynamics. PE and PD are also tested in the presence of non-stationarity. Lastly, I propose a new visualization technique to reveal the deterministic relation postulated by a given prediction model and provide another example of extension of PE as a multivariate dependence measure.

**Chapter 4** identifies the limitations of the conventional model comparison and evaluation approaches, and introduces a new model sufficiency test, the PD sufficiency test. I explain the rationale and incentive behind the new test and how it overcomes the limitations of the existing methods. Furthermore, I demonstrate its validity by applying it to a number of simulated examples.

**Chapter 5** outlines the empirical application of PE and the relevant ordinal pattern tools introduced in previous chapters to analyze EUR/USD 10-minute returns, EUR/USD 10-minute squared returns, aggregated 1-hour volatility series and the monthly SST time series. For the financial time series, PE and PD are used to detect and investigate the temporal dependent structures present in the investigated time series. The new model sufficiency test is used to evaluate the sufficiency of the one-step-ahead prediction performance of the considered models. In addition, I display the deterministic relation postulated by each prediction model via the new visualization plot to provide explanations of each forecasting model's superior and suboptimal performances. For the SST time series, PE and PD measures are used to reflect the level of seasonality and predictability. Bivariate PE is employed to quantify the causality dependence between the SST index in different regions.

**Chapter 6** concludes the thesis by summarizing the main findings and results.

## 2 Literature review

The literature review of this thesis consists of three parts, each covering a different but important field for our study. The first part provides the rationale behind the initial

innovation of PE and summarizes its previous applications in various areas for a range of different purposes. The first part categorizes previous studies and applications of PE in fields other than financial time series analysis. The second part summarizes the empirical properties commonly observed in financial time series especially in high-frequency financial time series. This part is included since the way that the PE measure responds to and is affected by the characteristic properties of financial data determines the correct interpretation of this relatively novel technique when conducted on financial time series. In the later chapters of this thesis, I examine the behaviour of PE in reacting to a number of typical properties of financial series such as non-stationarity and high-kurtosis. More importantly, it aids in guiding and interpenetrating the empirical analysis conducted in Section 5. The last part of the literature review lists some most commonly used parametric and non-parametric models in financial time series analysis. The overview of various models is needed as PE measures will be tested later in simulation studies to connect their behaviour with the parametrization of the most popular ARMA and GARCH models, to establish a link between PE and the specifications of simulated models. In that way it is possible to better comprehend the type of information PE reveals and neglects, and relate PE with the mainstream finance literature. The included models will be used later in empirical analysis as competing models in fitting and predicting real-world financial data.

## 2.1 Rationale behind permutation entropy (PE) and its previous applications

In the field of information theory, entropy is a quantity measuring the level of unpredictability of the information content of the dynamics of the underlying system. Shannon's entropy is often considered as the foundational and most natural measure of entropy. Given a system governed by a discrete probability distribution  $P = \{p_i : i = 1, \dots, M\}$ , Shannon's entropy is defined as

$$S(P) = - \sum_{i=1}^M p_i \log p_i,$$

where  $M$  represents for the number of possible outcomes, and  $p_i$  measures the probability of each outcome. If the dynamic of a process is perfectly predictable, then we have complete certainty of which outcome will take place, that is,

$$P = \begin{cases} p_i = 1; & i = a, \\ p_i = 0; & i \neq a, \end{cases}$$

the value of Shannon’s entropy reaches its minimum at 0. In contrast,  $S(P)$  is maximized at  $\log M$  if the process follows a uniform distribution so that the process is completely unpredictable. Therefore entropy can be considered as a measure of predictability, uncertainty and system complexity.

Shannon’s entropy suffers several drawbacks when applied in empirical time series analysis. First, this measure neglects temporal dependence relationships between the values in the time series. For instance, the measure yields the same value for time series  $X_1 = \{0, 0, 1, 1\}$  and the time series  $X_2 = \{1, 0, 1, 0\}$ . Second, the definition of Shannon’s entropy is not directly applicable to empirical data. It requires prior knowledge of the probability space of the process underlying the observed data, which is impossible to acquire in advance. To obtain an entropy estimate for discrete time series, the required probability distribution can be obtained by simply counting the frequencies of every possible outcome. However, “extracting” the probability distribution from continuous time series is not a trivial task. Many methods have been proposed to obtain the required probability distribution  $P$  for continuous time series, including binary symbolic-dynamics (Mischaikow et al. 1999), amplitude-statistics (De Micco et al. 2008), Fourier analysis (Powell & Percival 1979), wavelet transform (Blanco et al. 1998), partition entropies (Ebeling et al. 2001) and discrete entropies (Amigó et al. 2007). However these methods capture different aspects of the dynamics, and many of them require the investigated time series to satisfy the assumption of stationarity and low noise contamination. Hence, Bandt & Pompe (2002) propose PE as a more robust and data-driven way to construct the probability distribution required for the computation of entropy, while enhancing the ability to capture the causality dependence relationships.

Bandt & Pompe’s (2002) approach constructs the probability distribution  $P$  by considering the relative magnitudes of the neighbouring entries to symbolize the partitions of the time series. This is best illustrated with an example: if we choose the first three entries in a given time series  $X = \{x_t : t = 1, \dots, N\}$  to form a partition  $s_1 = (x_1, x_2, x_3)$  where  $(x_1, x_2, x_3) = (0, 3, 1)$ , and rank the entries in  $s_1$  in ascending order, the first entry is the smallest, the second entry the biggest and the third entry the second. We have the partition  $s_1$  belonging to the ordinal pattern  $x_2 > x_3 > x_1$ . For a three entry length partition, there are total  $3! = 6$  possible ordinal patterns, excluding equal value circumstances. If we repeat the same procedures to map every partition, that is,  $s_2 = (x_2, x_3, x_4), \dots, s_{N-2} = (x_{N-2}, x_{N-1}, x_N)$ , into one of the six ordinal patterns, we can obtain the required discrete probability distribution by computing the frequencies of each ordinal patterns. In fact, by considering the ordinal pattern, PE extracts the information of the relative magnitudes of the entries in the time series, while not considering the information with regards to their marginal magnitudes. This design provides

PE with many desirable properties. First, PE is extremely robust to both observational and dynamical noise (Bandt & Pompe 2002). Second, since the marginal distribution of time series is not taken into account, PE can deal with time series where variances are infinite or with heavy tailed distributions, as usually arises in economics and high-frequency financial time series. Third, PE is extremely fast in computation and can deal with complex systems with high non-linearity (Zunino et al. 2010). Lastly, PE is invariant to any monotonic transformations. However, the loss of marginal magnitude may cause some inadequacy of the PE measure in terms of its abilities to capture the features of certain dynamical system. Such inadequacy can be remedied by employing weighted PE as supplementary (Fadlallah et al. 2013).

Just like Shannon's entropy, the original use of PE is to distinguish a chaotic process from a random process, and measure the dynamical complexity and the predictability of the underlying time series. The value of PE is bounded in the range of 0 to 1. If a time series follows a completely stochastic process, we expect equal probabilities assigned to each possible ordinal pattern and PE will have a value of 1. But if all the segments in the investigated time series follow the same permutation pattern, the time series is obviously completely deterministic and predictable, and the permutation entropy will have a value of 0.

The PE measure requires no prior information or assumptions about the system underlying the observed time series, but only two pre-chosen parameters: the segment length  $D$  and the delay  $\tau$ . Segment length  $D$  determines the number of entries contained in every constructed segment, thereby the total number of possible ordinal patterns. Clearly, with more ordinal patterns, the measure is more capable of capturing the complex dynamics underlying the observed time series. The choice of delay  $\tau$  provides PE with the flexibility to investigate the structure of time series over short-term and long-term dynamics.

Over the past 18 years, since PE was first introduced, an increasing number of studies has been conducted to explore and extend its applications. The applications of PE are spread across many fields including biomedical systems and economics. The applications of PE can be roughly categorized into four branches as summarized below.

### **2.1.1 Measuring complexity and predictability**

PE was originally proposed as an alternative measure of complexity and predictability of the dynamics underlying an observed time series. By measuring the complexity of various signals, one can distinguish between time series generated from a chaotic (low dimensional deterministic) process and time series from a purely random process despite their seemingly

identical characteristics. In addition, the level of complexity is associated with the level of disorder and irregularity of the dynamical system. Generally, a lower-level complexity indicates the system is more likely to be governed by deterministic structures and has more potential to be captured and predicted, whereas a high-level complexity reflects more disordered or more complex controlling rule for the underlying dynamics, which might be more unpredictable and more difficult to comprehend. Thanks to the aforementioned advantages of PE, such as robustness to noise and the ability to deal with highly non-linear complex systems, the applications of PE as a complexity and predictability measure have been very successful in capturing the characteristics of biological systems, such as brain (electrical) activity and heart rate rhythms.

By investigating brain [Electroencephalographic \(EEG\)](#) data, the studies of PE on brain activity are mainly conducted in the area of epilepsy research (see [Cao et al. 2004](#), [Keller & Wittfeld 2004](#), [Veisi et al. 2007](#), [Li et al. 2007](#), [Bruzzo et al. 2008](#), [Ouyang et al. 2010](#), [Bian et al. 2012](#)) and anaesthesiology (see [Jordan et al. 2008](#), [Li et al. 2008](#), [Silva et al. 2010, 2011](#), [Olofsen et al. 2008](#)). For instance, [Veisi et al. \(2007\)](#) present PE as a diagnostic tool to distinguish normal and epileptic EEG recordings. [Li et al. \(2007\)](#) show that PE can detect pre-seizure state, thereby predicting the occurrence of an upcoming seizure. In anaesthesiology, PE is used to efficiently discriminate between different levels of consciousness during anaesthesia, providing an index of the anaesthesia drug effect ([Olofsen et al. 2008](#)). In the field of cardiology, PE can identify different physiological and pathological conditions when applied to heart rhythm [Beat-to-beat Interval \(BBI\)](#) data, and recognize patients suffering from congestive heart failure ([Parlitz et al. 2012](#)).

Among the few existing applications of PE in economics, PE is primarily used as a predictability measure to assess the efficiency of financial markets and indicate the level of market development ([Zunino et al. 2009](#), [Siokis 2018](#)).

In addition to the standard PE, [Humeau-Heurtier et al. \(2015\)](#), [Aziz & Arif \(2005\)](#), [Shang et al. \(2019\)](#), [Yin & Shang \(2014\)](#) also proposed a multiscale permutation entropy (MPE) and a number of MPE variants to show structures on multiple spatio-temporal scales. They applied the MPE and its variants on physiological time series and financial time series, and demonstrated the added capacity of their measures to standard PE in characterizing the complexity of time series. [Jiang et al. \(2017\)](#) proposed a Gini-Simpson index PE (GPE) to better facilitate the complexity measure of financial time series.

### 2.1.2 Detecting dynamical changes

Instead of computing PE on the complete data series, recording the variations of PE on (overlapping or non-overlapping) sliding blocks of investigated time series can detect dynamical changes in a complex system. [Cao et al. \(2004\)](#) demonstrate that the variation of PE as a function of time can accurately indicate bifurcations in the simulated transient logistic map time series and transient Lorenz system. Moreover, by employing the same technique on EEG data, the change of PE can track dynamical changes in brain activities, thereby indicating the onset of epileptic seizures. In the finance field, a similar approach has been applied to a number of financial time series to identify periods of time where noise prevailed over the deterministic behaviour of the market ([Hanas & Curtis 2008](#)). [Yan et al. \(2012\)](#) applied PE in bearing vibration analysis to detect dynamical changes of the machine working status, thereby monitoring the working status of rolling bearings.

### 2.1.3 Identification of delayed feedback

Many dynamical systems involve delayed feedback mechanisms. A typical example of such systems is the famous Mackey-Glass oscillator, which is presented by the differential-delay equation

$$\frac{dx_t}{dt} = \beta \frac{x_{t-\tau}}{1 + (x_{t-\tau})^n} - \gamma x_t, \quad \gamma, \beta, n > 0.$$

From the equation it is clear that the rate of change in the target variable depends on its lagged historical values. [Zunino et al. \(2010\)](#) apply PE on a time series generated from the Mackey-Glass oscillator. The rationale of using PE in detecting the time delay of the system is that the underlying dynamics should be more predictable and simple when the chosen parameter delay  $\tau$  in the PE coincides with the characteristic time delay of the system, resulting in a minimal in the function of PE over varying delays. Furthermore, it is also shown that this method is capable of revealing multiple time delays if these exist in the system. Many economic models include delay variables to account for the expectations from market participants. Hence the identification of the possible delayed iteration present in the system underlying the observed time series is critical to studies in many scientific areas.

### 2.1.4 Measuring dependence structures for multivariate time series

The extension of multivariate PE has been used to measure the dependence between two or more time series. [Matilla-García et al. \(2014\)](#) extend PE to a bivariate case to study the causal relationship between trading volume and security prices. They demonstrate that their dependence test performs well in detecting linear and non-linear causality. In the

work of Bahraminasab et al. (2008), PE is used with conditional mutual information for the detection of causal relationships between two time series. Under the simulation of van der Pol oscillators, their method demonstrates a good tolerance to external noise. Zhao et al. (2013) proposed the permutation mutual information to achieve the similar purpose as the two aforementioned measures.

## 2.2 Empirical properties of financial time series

In this section, I list a set of properties of financial time series, specifically of financial return time series. These properties have been commonly observed across various financial assets, markets and sampling periods by independent empirical studies, and are sometimes referred to as the “stylized facts” of financial time series. Being aware of and understanding these features is extremely important when conducting financial time series analysis, not only since a great proportion of the financial time series models are designed to capture one or more than one of these documented properties, but also some of these properties impose statistical challenges for certain statistical approaches, hence conclusions and interpretations need to be drawn with caution. By further organizing and summarizing the review papers composed by Sewell (2011), Cont (2001) and other relevant literature I briefly characterize the main empirical features of asset returns, explaining their causes and stating relevant potential problems that might arise when conducting certain statistical analysis.

### 2.2.1 Heavy tailed distribution of returns

The distribution of returns is documented as being approximately symmetric and having high kurtosis. High kurtosis indicates the distribution of asset returns is heavy tailed compared with the normal distribution, which also means extreme values are more likely to occur in the asset returns compared to from a normal distribution. Furthermore, the distribution is increasingly heavy-tailed for higher-frequency (smaller intervals  $\Delta t$ ) returns. Volatility clustering could explain some proportion of heavy tailed distributions in asset returns, however, even after correcting returns for volatility clustering, the residual time series still exhibits a heavy tail, but to a lesser extent.

Even though the heavy tailed distributions of asset returns are widely documented, the precise form of the tails remains an open question. Cont (2001) points out that the (unconditional) distribution of returns seems to display a power law or Pareto-like tail, with a tail index higher than two and less than five for most data set studies. But the precise form of the tails is difficult to determine.

### 2.2.2 Insignificant correlation of returns

Empirical studies suggest that autocorrelations (ACF) of asset returns

$$C(\tau) = \text{CORR}(r_{t,\Delta t}, r_{t+\tau \times \Delta t, \Delta t})$$

are often insignificant, except that high-frequency ( $\Delta t \leq 20$  minutes) returns sometimes exhibit negative autocorrelations at very short delays  $\tau$  (normally only at  $\tau = 1$ ). The insignificance of autocorrelation implies that no linear relations can be found in the dynamics of asset returns, and the absence of linear relations of asset returns is consistent with the “efficient market hypothesis”, where arbitrage (profit) can’t be obtained by simply extrapolating and formulating linear functions of historical returns. Moreover, traditional tools based on second-order properties, such as ARMA modelling and Fourier analysis, are not directly applicable to asset returns. Therefore, nonlinear dependence measures are required to characterize the dependence properties of asset returns. As for the causes of the negative autocorrelation of high-frequency return, [Campbell et al. \(1997\)](#) point out that the negative correlation exhibited in the high-frequency return of transaction prices can be attributed to the bid-ask bounce, where “transaction prices may take place either close to the ask or closer to the bid price and tend to bounce between these two limits”. One may also observe the negative correlation in the bid or ask returns at very short delays. This might be due to the habit of the market makers who tends to set the bid and ask quotes to be mean-reverting (see [Goodhart & O’Hara 1997](#)).

However, the sample ACF of asset returns needs to be interpreted with caution. The aforementioned heavy tailed marginal distributions of asset returns makes the asymptotic standard variance of sample ACFs greater than the classical ones, and the situation is even more difficult for sample ACFs of the squared returns.

### 2.2.3 Volatility clustering and long-memory properties

Despite the fact that the ACFs of asset returns are often insignificant, the ACFs for the absolute and squared returns are always positive and significant. This feature supports the well-known phenomenon in the finance literature, called “volatility clustering”: large price variations are likely to be followed by large price variations. Not only do empirical studies find the ACFs of the absolute and squared returns decay slowly for increasing delays, where ACFs can remain significant for delays of several days, even weeks. [Bollerslev & Mikkelsen \(1996\)](#), [Liu et al. \(1997\)](#), [Andersen & Bollerslev \(1997\)](#) and [Andersen & Bollerslev \(1998\)](#) have remarked that the decays of ACFs of the powers of returns are well described by power-law

decay:

$$\text{corr}(|r_{t,\Delta t}|^\alpha, |r_{t+\tau \times \Delta t, \Delta t}|^\alpha) \sim A\tau^\beta$$

where  $A$  and  $\beta$  are constants, suggesting the existence of long-memory property.

The volatility clustering feature observed in asset returns is the main motivation of ARCH (GARCH) class models, which estimate the future squared returns as a linear function of the past squared returns plus innovations. In order to incorporate the long-range dependence property in volatilities, the [Fractionally Integrated Generalized Autoregressive Conditional Heteroscedastic \(FIGARCH\)](#) model is constructed by introducing a fractional differencing operator into the GARCH model to account for the impacts from long-lagged past squared returns on current volatilities.

However, researchers have shown that the non-stationary short-memory time series (time series with inconstant mean and/or variances) display many of the same properties as a long-memory time series, causing the slow decay of ACF and driving the sum of the estimated parameters  $\alpha$  and  $\beta$  under the GARCH(1,1) model very close to one, inducing “spurious long-memory” effects (see [Mikosch & Stărică 2004](#), [Perron & Qu 2010](#)). [Tzouras et al. \(2015\)](#) propose an alternative test for long-memory that is robust to non-stationarity, concluding that the S&P500 returns exhibit long-memory structures that are not attributed to non-stationarity.

#### 2.2.4 Non-linearity

A stochastic process  $Y = \{y_t | t = 1, \dots, T\}$  is said to be linear if it can be written in the form

$$y_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}, \quad \text{or} \quad y_t = \mu + \sum_{i=-\infty}^{\infty} \psi_i a_{t-i} \quad (1)$$

where  $\mu$  is a constant,  $\psi_i$  are real numbers with  $\psi_0 = 1$ , and  $\{a_t\}$  is a sequence of independent and identically distributed (iid) random variables with a well-defined distribution function. The equation above can be extended to include some exogenous variables in the mean of  $y_t$ , such as economic fundamentals and periodic functions to account for exogenous driving forces and seasonalities. The most commonly used linear model in modelling financial time series is the ARMA process:

$$y_t = \mu + \sum_{i=1}^p \alpha_i y_{t-i} + \sum_{i=1}^q \beta_i \varepsilon_{t-i} + \varepsilon_t,$$

where  $\{\varepsilon_t\}$  is a zero mean iid random variables with finite variance. Any process that violates the condition defined in (1) belongs to a nonlinear process. However, there is no general form

of nonlinear process since there are many possible specifications. In the finance literature, it is generally accepted that there are two main forms of nonlinearity: nonlinearity in the mean (additive nonlinearity)

$$y_t = f(y_{t-1}, \dots, y_{t-k}, \varepsilon_{t-1}, \dots, \varepsilon_{t-k}) + \varepsilon_t,$$

where  $\varepsilon_t$  is a zero mean error term, and  $f(\cdot)$  is a nonlinear function; nonlinearity in the variance (multiplicative nonlinearity)

$$y_t = \sqrt{h(y_{t-1}, \dots, y_{t-k}, \varepsilon_{t-1}, \dots, \varepsilon_{t-k})} \times \varepsilon_t,$$

where  $h(\cdot)$  is nonlinear or time variant. A hybrid process where nonlinearity exists both for the mean and the variance is also possible. An example of an additive nonlinear model is the Nonlinear Moving Average model of [Robinson \(1977\)](#):

$$y_t = \varepsilon_t + \beta \varepsilon_{t-1} \varepsilon_{t-2},$$

while the GARCH (1,1) model of [Bollerslev \(1986\)](#)

$$\begin{aligned} y_t &= \varepsilon_t \times \delta_t, \\ \delta_t^2 &= \omega + \alpha \varepsilon_{t-1}^2 + \beta \delta_{t-1}^2. \end{aligned}$$

is an example of a multiplicative nonlinear model. In the two examples above,  $\varepsilon_t$  is a sequence of iid random variables.

Since both additive and multiplicative nonlinear processes could generate returns that are correlated with their own lags, in order to distinguish between different forms of nonlinearities, [Hsieh \(1989\)](#) suggests a test specifically for multiplicative nonlinearity. The null hypothesis of the test is  $E(\varepsilon_t \varepsilon_{t-i} \varepsilon_{t-j}) = 0$  with the test statistics  $\hat{E}(\varepsilon_t \varepsilon_{t-i} \varepsilon_{t-j}) = \frac{1}{T} \sum u_t u_{t-i} u_{t-j}$ , where  $u_t$  is the filtered residual from the linear regression of its lagged returns, and  $T$  is the length of the time series. The test is based on the property of the multiplicative nonlinearity that

$$E(\varepsilon_t | y_{t-1}, \dots, y_{t-k}, \varepsilon_{t-1}, \dots, \varepsilon_{t-k}) = 0,$$

hence by approximating  $f(\cdot)$  using a second-order Taylor series expansion around zero and obtaining terms such as  $\varepsilon_{t-i} \varepsilon_{t-j}$ ,  $y_{t-i} \varepsilon_{t-i}$  and  $y_{t-i} y_{t-j}$ , we expect  $\varepsilon_t$  is uncorrelated with these terms.

[Ashley et al. \(1986\)](#) summarize various tests for detecting nonlinearity. The most commonly

used non-linear test is the BDS test (Brock et al. 1991), the null hypothesis of which is that the observations are independent and identically distributed, so in order to detect nonlinearity, other possible sources of violation of iid, such as linear dependencies or non-stationarity, must first be eliminated. Hinich (1982) proposed the bispectrum test to test nonlinearity. The bispectrum test is relatively insensitive to many possible forms of nonlinearity, such as ARCH or GARCH. This is because bispectrum tests use bicorrelations, or third-order moments, which are zero-valued for GARCH processes. Therefore, any significant results with this test are a strong indication of a nonlinear temporal dependence structure in a form that is likely to be more complex than simple ARCH or GARCH dependence. Moreover, as Ashley et al. (1986) show, the bispectrum test can be used to test for nonlinearity even in the presence of linear dependencies, with no loss of power.

Empirical studies (see Hsieh 1989, Brock et al. 1991, Ammermann & Patterson 2003, Lim et al. 2008) suggest nonlinearity is widely detected in the mean and (especially) the variance of return dynamics across various financial markets and time intervals. Hsieh (1989) investigates daily returns in five major exchange rate time series, finding no linear correlation but substantial multiplicative nonlinearity. They further conclude that the GARCH model can explain a large part of the nonlinearities detected. Ammermann & Patterson (2003) use the bispectrum test to test the nonlinearity of the daily return of six different stock market indices across the world, and the stocks trading on the Taiwan Stock Exchange, which can be regarded as an isolated market. They show that nonlinear serial dependence is a fundamental aspect and inherent feature of financial time series. However, by conducting the nonlinear test on a sliding window of the data set, they find the nonlinearity is not persistent, that is, nonlinearity dependence showed up at random intervals for a brief period of time but then seemed to disappear again before it could be exploited. Abhyankar (1995) find clear evidence of non-linearity in the 1-minute returns on the Financial Times Stock Exchange (FTSE) index.

### 2.2.5 Non-stationary

Financial markets are prominent examples of highly non-stationary systems. Sample averaged observables such as means and variances strongly depend on the time window in which they are evaluated. This undermines the cornerstone of many conventional approaches that presume past behaviours will stay unchanged into the future.

Despite widespread acceptance of the existence of non-stationarity, there is no consensus on the exact form and explanations of the time-varying behaviours. Even though real-world systems exhibit myriad forms of non-stationarity, the most commonly seen forms of non-

stationarity in financial time series include upwards, downwards and seasonal trends in the mean and variance, unit root non-stationarity, piecewise stationary and arbitrary abrupt change. Unit root non-stationarity is caused by the fact that sometimes the impact from past shocks does not decay over time, but has a permanent effect on the future series. For example, a random walk process

$$x_t = x_{t-1} + \varepsilon_t,$$

where  $\{\varepsilon_t\}$  is a white noise series. Piecewise stationary refers to the behaviour where although the time series as a whole exhibits non-stationarity, weak stationarity rules are satisfied in some local regimes and the transition between those regimes is instantaneous. An example is a Markov Switching model. The following formulae give a simple example of a Markov Switching model:

$$x_t = \begin{cases} a_1 x_{t-1} + \varepsilon_t, & \text{if } s_t = 1, \\ a_2 x_{t-1} + \varepsilon_t, & \text{if } s_t = 2, \end{cases}$$

where  $s_t$  denotes the regime of the system and the transition between two regimes is governed by a hidden two-state Markov chain with transition probabilities

$$P(s_t = 2 | s_{t-1} = 1) = p_1, \quad P(s_t = 1 | s_{t-1} = 2) = p_2.$$

The innovation series  $\{\varepsilon_t\}$  is a sequence of iid random variables with mean zero and finite variance. Arbitrary abrupt change non-stationarity is the most complex type of non-stationarity in a way that the relationship among the intrinsic and exogenous variables can change arbitrarily and abruptly as opposed in a systematic manner as in the aforementioned forms of non-stationarity. Among all types of non-stationarity, systematic trends in mean or variance and unit root non-stationarity in time series are most easily identified and can be treated by de-trending, deseasonalizing and differencing.

### 2.2.6 Other properties

- **Leverage effect:** Another commonly observed feature of asset return is the leverage effect where the volatility of an asset return is observed to be negatively correlated with the return of that asset, when prices decrease, volatility increases; when prices increase, volatility decreases but to a lesser extent. [Threshold Generalized Autoregressive Conditional Heteroscedastic \(TGARCH\)](#) and [Exponential Generalized Autoregressive Conditional Heteroscedastic \(EGARCH\)](#) models are typical examples to account for the leverage effect in the modelling of financial returns. Changes in low-frequency

volatility have more impact on subsequent high-frequency volatility than the opposite. The intensity of this relation depends on the level of volatility at long horizons.

- **Scaling effect:** Scaling laws describe mean absolute and mean squared returns as functions of their time intervals (varying from a few minutes to one or more years). These quantities are proportional to a power of the interval size. The distributions of return and volatility scale for a range of time intervals, which means the distributions of return and volatility for various choices of  $\Delta t$ , ranging from 1 minute up to 1 month, have similar functional forms
- **Inconclusive evidence of low dimensional chaos** is found in financial returns.
- **Intermittency:** Return display intermittency, that is, the time series alternates between periods of seemingly periodic and chaotic fluctuations, evidenced from the presence of irregular bursts. Intermittency is marked by random switching of system trajectories between a relatively regular laminar phase and irregular turbulent phase.
- **Gain/loss asymmetry:** The large drawdowns are more commonly observed in stock prices than equally large upward movements. But exchange rate exhibits a higher symmetry in up/down movements.

### 2.2.7 Properties of high-frequency financial returns

The properties of financial asset returns depend strongly on the selection of time interval  $\Delta t$ . Research into stylized facts for high-frequency data shows that the accepted empirical regularities of daily data and weekly data often do not hold for intraday data ([Gençay et al. 2001](#)). I summarize four special characteristics of high-frequency returns that generally do not exist in returns over longer intervals.

1. **Negative first-order ACF:** High-frequency returns normally display an extremely high negative first-order ACF. The negative first-order ACF can be attributed to the non-synchronous trading phenomenon or bid-ask bounce, and they both belong to the microstructures of market behaviours. The nonsynchronous trading phenomenon refers to the fact that based on the trading intensities the actual time of the last transaction used to compute the return over fixed time interval can vary from time to time. [Campbell et al. \(1997\)](#) propose a model to replicate this phenomenon and show that the actual unequally spaced returns can lead to erroneous conclusions about the asset returns, including inducing first-order negative autocorrelations.
2. **Extremely high kurtosis:** The distribution of the high-frequency returns is generally

more heavy-tailed than lower frequency returns.

3. Intraday seasonality in volatility: For stocks, returns are shown to be more volatile at the beginning and closing of trading hours, and relatively quiet during lunch hour, resulting in an U-shape intraday periodicity in its volatility. For the foreign exchange rate, the seasonal pattern is determined by the differences in trading times in the global foreign exchange markets.
4. Discretization: The discretization comes from the existence of the smallest possible price increment unit in the financial markets. This market convention is often referred to as the tick size in the stock and futures markets, or the pip in the exchange rate market. Since price can only jump in an integer multiple of a certain unit, the possible values of each return are subject to a pre-defined range that varies according to the price at the start of the return interval.

## 2.3 Financial time series models

In order to characterize the dynamics of a time series process, countless parametric and non-parametric models have been proposed to capture the deterministic structures (if exists) underlying the observations and on which future predictions are based. This section summarizes and revises the most commonly used parametric and non-parametric models in financial time series analysis. By presenting a review of the most commonly used models, I aim to reveal the widely accepted benchmark understandings of financial time series dynamics, and the shared and individual challenges each model encounters in its empirical application. The stream of parametric models I choose in this section includes the ARMA model, the GARCH model and its variants: EGARCH, [Glosten-Jagannathan-Runkle Generalized Autoregressive Conditional Heteroscedastic \(GJR-GARCH\)](#) and FIGARCH. The representative non-parametric models I select include the support vector regression (SVR) model and the Gaussian process regression (GPR) model. They are some of the most commonly used approaches in financial time series analysis. The selected models demonstrate mainstream perceptions of modelling financial time series. Linear models focus on capturing the linear serial dependence in the return or in the price dynamics, GARCH and its variants try to capture the volatility clustering features commonly observed in the financial time series dynamics. The FIGARCH model aims to incorporate long-memory properties into the GARCH models. The non-parametric models, SVR and GPR models, have no close form thus are more flexible compared to the parametric models. They have the additional capability of replicating nonlinear structures, but generally involve more parameters to estimate and require a larger data size to produce accurate predictions. In addition, non-parametric models

are more likely to over-fit the training data than parametric models, thus leading to poor generalization when predicting an unseen set of data of interest.

### 2.3.1 Linear models

A stochastic process  $Y = \{y_t | t = 1, \dots, T\}$  is said to be linear if it can be written in the form

$$y_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}, \quad \text{or} \quad y_t = \mu + \sum_{i=-\infty}^{\infty} \psi_i a_{t-i}$$

where  $\mu$  is a constant,  $\psi_i$  are real numbers with  $\psi_0 = 1$ , and  $\{a_t\}$  is a sequence of independent and identically distributed (iid) random variables with a well-defined distribution function. A typical example of a linear model in a time series is the ARMA model. A general ARMA( $p, q$ ) model is in the form

$$y_t = \phi_0 + \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i},$$

where  $\varepsilon_t$  is a white noise series with mean zero and variance  $\sigma_\varepsilon^2$  and  $p$  and  $q$  are non-negative integers. Alternatively, the above equation can be written as:

$$[1 - \phi(L)]y_t = \phi_0 + [1 - \theta(L)]\varepsilon_t. \quad (2)$$

We require that there are no common factors between the polynomial  $1 - \phi(L)$  and  $1 - \theta(L)$ , otherwise the order ( $p, q$ ) of the model can be reduced. For an ARMA( $p, q$ ) process to be weakly stationary, all zeros of  $1 - \phi(L)$  must lie outside the unit circle. ARMA models can be written in two other representations, namely AR representation and MA representation. The alternative representations serve different purposes and can lead to a better understanding of the model. AR representation:

$$y_t = \frac{\phi_0}{1 - \theta_1 - \dots - \theta_q} + \pi_1 y_{t-1} + \pi_2 y_{t-2} + \dots + \varepsilon_t \quad (3)$$

shows the dependence of the current entries on the past entries, whereas MA representation:

$$y_t = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_q} + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots + \varepsilon_t$$

reflects the impact of past shocks  $\varepsilon_{t-i}$  on the current target observations. The parameter  $\pi_k$  and  $\psi_k$  are determined by equating the coefficients in

$$\frac{\phi(L)}{\theta(L)} = 1 - \pi_1 L - \pi_2 L^2 - \dots \equiv \pi(L).$$

and

$$\frac{\theta(L)}{\phi(L)} = 1 + \psi_1 L + \psi_2 L^2 + \dots \equiv \psi(L)$$

When  $p = 1$  and  $q = 1$ , which corresponds to the most used ARMA model ARMA(1,1),

$$\begin{aligned} \pi_k &= (\phi_1 - \theta_1)\theta_1^{k-1}, & k = 1, 2, \dots, \infty, \\ \psi_k &= (\phi_1 - \theta_1)\phi_1^{k-1}, & k = 1, 2, \dots, \infty. \end{aligned}$$

For a stationary ARMA(1,1) process (stationary constraint:  $|\phi_1| < 1$ ), the unconditional mean and unconditional variance of the model is given below:

$$\begin{aligned} E(y_t) &= \frac{\phi_0}{1 - \phi_1}, \\ \gamma_0 &= \frac{\sigma_\varepsilon^2(1 + \theta_1 - 2\phi_1\theta_1)}{1 - \phi_1^2}. \end{aligned}$$

By employing the Yule-Walker equation, it is straightforward to derive the formulae of the ACFs of the ARMA(1,1) process:

$$\begin{aligned} \rho_1 &= \phi_1 - \frac{\theta_1\sigma_\varepsilon^2}{\gamma_0}, \\ \rho_l &= \phi_1\rho_{l-1}, & l = 2, 3, \dots \end{aligned} \tag{4}$$

According to the ACF formulae, the ARMA(1,1) model assumes the dependence relation exhibited in the time series decays exponentially at the rate of  $\phi_1$ . Moreover, since the value of term  $(1 + \theta_1 - 2\phi_1\theta_1)/(1 - \phi_1^2)$  is close to one, thereby  $\rho_1 \approx \phi_1 - \theta_1$ . Hence  $\phi_1 - \theta_1$  roughly determines the dependence between the adjacent entries.

Linear models are the most used and simplest approach to model and forecast time series. However, they are only rough approximations of real-world complex systems and often inadequate in capturing the complex dynamic required to make accurate prediction of a real-world financial time series due to the non-linearity (see section 2.2.4) and non-stationarity (see section 2.2.5) of the real-world system.

### 2.3.2 Conditional heteroscedastic models

The conditional heteroscedastic model refers to the stream of statistical models that aim to model the volatility of an asset return. More specifically, they model the volatility of an asset return as a function of past volatilities and past returns. The main motivation of the conditional heteroscedastic model is to capture the widely observed stylized fact about the financial time series called “volatility clustering”, as outlined in section 2.2.3: large price variations are likely to be followed by large price variations. Empirical evidence of volatility clustering shows that, while the ACFs of asset returns are often insignificant, the ACFs for the absolute and squared returns are always positive and significant. The most popular conditional heteroscedastic models are GARCH and its variants EGRACH, GJR-GARCH and FIGARCH.

Volatility can be an ambiguous term in the statistics and finance literature. Here volatility is defined as the conditional standard deviation of the target variable, that is,

$$\sigma_t = \text{s.d.}(y_t | F_{t-1})$$

where  $F_{t-1}$  denotes the all available information at time  $t - 1$ . Since there is only one realization at each time  $t$ , volatility  $\sigma_t$  is not directly observable and normally reflected in squared returns.

The GARCH model estimates future volatility as a linear function of the past volatilities plus past innovations. Specifically,  $y_t$  follows a GARCH( $p, q$ ) model if

$$\begin{aligned} y_t &= \sigma_t z_t, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^q \alpha_i y_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \\ &= \alpha_0 + \alpha(L) y_t^2 + \beta(L) \sigma_t^2, \end{aligned} \tag{5}$$

where  $z_t$ 's are normally distributed uncorrelated random variables with  $E(z_t) = 0$  and  $\text{Var}(z_t) = 1$ . Rearranging (5), an equivalent ARMA type representation of the GARCH( $p, q$ ) model is given by

$$y_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i y_{t-i}^2 + \sum_{j=1}^p \beta_j y_{t-j}^2 - \sum_{j=1}^p \beta_j v_{t-j} + v_t,$$

or

$$[1 - \alpha(L) - \beta(L)]y_t^2 = \alpha_0 + [1 - \beta(L)]v_t, \tag{6}$$

where  $v_t = y_t^2 - \sigma_t^2$  is the “innovation” in the squared returns.  $\alpha(L)$  is the ARCH polynomial  $Q$  and  $\beta(L)$  is the GARCH polynomial  $P$ . To ensure the non-negativity of the conditional variance process, the model requires  $\alpha_0 > 0, \alpha_i \geq 0, i = 1, \dots, q, \beta_j \geq 0, j = 1, \dots, p$ . In order for the process to be stationary, an additional constraint:  $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1$  must be satisfied. If  $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j = 1$ , the process is an **Integrated Generalized Autoregressive Conditional Heteroscedastic (IGARCH)**( $p, q$ ) specification, where a shock in the squared returns will lead to an infinite persistent effect on the conditional variance.

It is worth mentioning that the reason that the GARCH model is not a linear model despite linear relations being assumed between future and past volatilities is due to the existence of temporal dependence between the innovations  $v_t$ . From the ARMA representation of the GARCH model given in (6), the distribution of the innovation term

$$\begin{aligned} v_t &= y_t^2 - \sigma_t^2 \\ &= \sigma_t^2 z_t^2 - \sigma_t^2 \\ &= \sigma_t^2 (z_t^2 - 1). \end{aligned}$$

is affected by  $\sigma_t^2$ . Since the volatility  $\sigma_t^2$  depends on past squared returns  $y_{t-1}^2$ , consequently is affected by  $v_{t-1}$ . Therefore, it is easy to deduce that the squared return “innovations”  $\{v_t\}$  are not independent of each other. They are correlated in a way that large past innovations  $v_{t-i}$  in volatility tend to lead to greater variance of the distribution of the current innovation  $v_t$ .

Determining the order  $p$  and  $q$  of a GARCH model is not easy. Common practice includes estimating all possible subset models of GARCH( $p, q$ ) and choosing the best according to an information criterion (**Akaike Information Criterion (AIC)**, **Bayesian Information Criterion (BIC)**), or diagnosing the properties of the post-fitting squared residuals for remaining serial correlations (Ljung-Box test). In fact only lower order GARCH models are used in most applications, for instance, GARCH(1,1), GARCH(2,1), and GARCH(1,2) models.

In order to better interpret the implications of GARCH parameters, we transform the GARCH(1,1) model into an AR-like representation. From equation (6), by treating  $y_t^2$  in GARCH(1,1) model as the target variable and  $v_t$  as the innovation term, the equation of a GARCH(1,1) model is just like an ARMA(1,1) process with  $\phi_1 = \alpha_1 + \beta_1$  and  $\theta_1 = \beta_1$  given in equation (2). Therefore, by replacing  $\phi_1$  with  $\alpha_1 + \beta_1$  and  $\theta_1$  with  $\beta_1$ , the AR representation of GARCH(1,1) model is given below:

$$y_t^2 = \frac{\alpha_0}{1 - \beta_1} + \pi_1 y_{t-1}^2 + \pi_2 y_{t-2}^2 + \dots + v_t \quad (7)$$

where  $\pi_k = \alpha_1 \beta_1^{k-1}$ ,  $k = 1, 2, \dots$ . Accordingly, the unconditional mean of  $y^2$

$$E[y_t^2] = \frac{\alpha_0}{1 - \alpha_1 - \beta_1},$$

and the ACF of a GARCH(1,1) model with  $y_t^2$  as the target variable can be obtained from the ACF equation of an ARMA(1,1) model given in equation (4). The final results are

$$\rho_1 = (\alpha_1 + \beta_1) - \beta_1 \frac{[1 - (\alpha_1 + \beta_1)^2]}{1 - 2\alpha_1\beta_1 - \beta_1^2},$$

$$\rho_l = (\alpha_1 + \beta_1)\rho_{l-1}, \quad l = 2, 3, \dots$$

Since the value of  $\frac{[1 - (\alpha_1 + \beta_1)^2]}{1 - 2\alpha_1\beta_1 - \beta_1^2}$  is close to one, we can make the approximation that

$$\rho_1 \approx \alpha_1,$$

$$\rho_l \approx (\alpha_1 + \beta_1)^{l-1} \alpha_1, \quad l = 2, 3, \dots$$

Therefore, for a GARCH(1,1) model, the parameter  $\alpha_1$  determines the extent to which a volatility shock today impacts the next period's volatility whereas  $(\alpha_1 + \beta_1)$  measures the rate that this effect diminishes over time.

The EGARCH model introduced by Nelson (1991) is a GARCH variant that models the logarithm of the volatility to relax the non-negativity constraints previously required by the GARCH model on its coefficients. In addition to modelling the logarithm, the EGARCH model has additional leverage terms to allow for asymmetry effects on volatility between lagged positive and negative asset returns. The model specification is given below:

$$y_t = \sigma_t z_t,$$

$$\log \sigma_t^2 = \omega + \sum_{i=1}^p b_i \log \sigma_{t-i}^2 + \sum_{j=1}^q a_j \left[ \frac{|y_{t-j}|}{\sigma_{t-j}} - E\left\{ \frac{|y_{t-j}|}{\sigma_{t-j}} \right\} \right] + \sum_{j=1}^q \xi_j \left( \frac{y_{t-j}}{\sigma_{t-j}} \right),$$

where  $z_t$ 's are uncorrelated with  $E(z_t) = 0$  and  $\text{Var}(z_t) = 1$ , and  $\xi_j$ 's are leverage coefficients, which are usually negative to reflect the server impact on future volatility from negative returns. To ensure stationarity, all roots of the GARCH lag operator polynomial  $(1 - b_1 L - \dots - b_p L^p)$  must lie outside the unit circle. EGARCH suffers from a major drawback in that it produces biased forecasts, since by Jensen's inequality

$$E(\sigma_t^2) \geq \exp\{E(\log \sigma_t^2)\}.$$

Similar to EGARCH, the GJR-GARCH model is another variant of GARCH that includes an additional term to account for the asymmetry effect. It has the specification:

$$y_t = \sigma_t z_t,$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2 + \sum_{j=1}^q \alpha_j y_{t-j}^2 + \sum_{j=1}^q \zeta_j I[y_{t-j} < 0] y_{t-j}^2,$$

where  $z_t$ 's are uncorrelated with  $E(z_t) = 0$  and  $\text{Var}(z_t) = 1$ . To ensure the positivity of the conditional variance process,  $\alpha_0 > 0$ ,  $\alpha_i \geq 0, i = 1, \dots, q$ ,  $\beta_j \geq 0, j = 1, \dots, p$  and  $\alpha_i + \zeta_i \geq 0, i = 1, \dots, q$ . In order for the process to be stationary, an additional constraint:  $\sum_{i=1}^p \beta_i + \sum_{j=1}^q \alpha_j + \frac{1}{2} \sum_{j=1}^q \zeta_j < 1$  must be met.

The FIGARCH model of [Baillie et al. \(1996\)](#) is an extension of the GARCH model to capture the long-memory properties observed in the volatility dynamics. The FIGARCH model is constructed by incorporating a fractional differencing operator  $(1 - L)^d$  with  $0 < d < 1$  into the GARCH model given in (2):

$$[1 - \alpha(L) - \beta(L)](1 - L)^d y_t^2 = \alpha_0 + [1 - \beta(L)] v_t,$$

or

$$\phi(L)(1 - L)^d y_t^2 = \alpha_0 + [1 - \beta(L)] v_t, \quad (8)$$

where  $v_t = \varepsilon_t^2 - \sigma_t^2$  is the ‘‘innovation’’ in the squared returns,  $d$  is a fraction  $0 < d < 1$ , and all the roots of  $\phi(L)$  and  $[1 - \beta(L)]$  lie outside the unit circle. The degree of persistence depends on the  $d$  parameter. When  $d$  gets closer to one the memory of the process increases. Note that the fractional differencing operator  $(1 - L)^d$  can be written as:

$$(1 - L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k - d)}{\Gamma(k + 1) \Gamma(-d)} L^k.$$

Rearranging (8), an alternative form of the FIGARCH( $p, d, q$ ) model can be obtained as

$$\begin{aligned} \sigma_t^2 &= \alpha_0 [1 - \beta(1)]^{-1} + \{1 - [1 - \beta(L)]^{-1} \phi(L) (1 - L)^d\} y_t^2 \\ &= \alpha_0 [1 - \beta(1)]^{-1} + \lambda(L) y_t^2, \end{aligned}$$

where  $\lambda(L) = \sum_{k=1}^{\infty} \lambda_k L^k$ . To ensure the non-negativity of the conditional variance process all the  $\lambda_k$ 's must be non-negative, that is,  $\lambda_k \geq 0$  for  $k = 1, 2, \dots$ . Take the FIGARCH(1,  $d$ , 1) model as an example: the non-negativity constraint can be derived by expanding the

coefficients in the  $\lambda(L)$ , and set all  $\lambda_k$ s greater or equal to zero. By equating

$$1 - (1 - \beta_1 L)^{-1}(1 - \phi_1 L)(1 - L)^d = \sum_{k=1}^{\infty} \lambda_k L^k$$

it is possible to derive

$$\begin{aligned} \lambda_1 &= \phi_1 - \beta_1 + d \\ \lambda_2 &= (d - \beta_1)(\beta_1 - \phi_1) + \frac{d(1-d)}{2} \\ \lambda_3 &= \beta_1 \left[ d\beta_1 - d\phi_1 - \beta_1^2 + \beta_1\phi_1 + \frac{d(1-d)}{2} \right] + d\frac{1-d}{2} \left( \frac{2-d}{3} - \phi_1 \right) \\ &\vdots \\ \lambda_k &= \beta_1 \lambda_{k-1} + \left( \frac{k-1-d}{k} - \phi_1 \right) \delta_{d,k-1}, \quad k = 4, 5, \dots, \end{aligned} \tag{9}$$

where  $\delta_{d,k} = \delta_{d,k-1}(k-1-d)k^{-1}$  and  $\delta_{d,0} = 1$ . [Baillie et al. \(1996\)](#) and [Bollerslev & Mikkelsen \(1996\)](#) provide the sufficient non-negativity constraint for the FIGARCH(1,  $d$ , 1) model:

$$\beta_1 - d \leq \phi_1 \leq \frac{2-d}{3} \quad \text{and} \quad d \left( \phi_1 - \frac{1-d}{2} \right) \leq \beta_1 (d - \beta_1 + \phi_1),$$

where  $\phi_1 = \alpha_1 + \beta_1$ . For the similar lower order models, such parameter restrictions can be derived similarly, however, for higher order models such restrictions cannot be obtained easily ([Caporin 2003](#)).

### 2.3.3 Support vector regression

Support vector regression (SVR) is a popular machine learning tool for regression, first developed by Vladimir Vapnik and his colleagues in 1992 ([Vapnik 1998](#)). SVR implements the structural risk minimization principle as opposed to the empirical risk minimization principle commonly employed in many traditional artificial neural network models. In essence, SVR achieves non-linear regression by trying to find the best regression hyperplane of minimum upper bound of generalization error with the help of the kernel function. One of the most popular types of SVRs is  $\varepsilon$ -SVR, which locates the hyperplane with an  $\varepsilon$ -insensitive loss function. The SVR function is formulated as follows:

$$f(\mathbf{x}) = \boldsymbol{\omega}^\top \phi(\mathbf{x}) + b$$

where  $\phi(\mathbf{x})$  is a nonlinear function mapping the input vector to the feature space,  $\boldsymbol{\omega}$  is a vector of weight coefficients and  $b$  is a bias constant.

In the case of time series prediction, consider a set of training points  $\{(\mathbf{x}_i, z_i) : i = 1, \dots, N-d\}$  in a given time series  $\{x_t : t = 1, \dots, N\}$ , where  $\mathbf{x}_i = (x_i, \dots, x_{i+d-1}) \in R^d$  is the input vector constructed by  $d$  lagged entries and  $z_i = x_{i+d}$  is the target output. The goal is to find a function  $f(\mathbf{x})$  that deviates from  $z_i$  by a value no greater than  $\varepsilon$  for each training point  $\mathbf{x}_i$  and at the same time is as flat as possible, that is,

$$\begin{aligned} & \min_{\boldsymbol{\omega}, b} \frac{1}{2} \boldsymbol{\omega}^\top \boldsymbol{\omega} \\ & \text{subject to } \boldsymbol{\omega}^\top \phi(\mathbf{x}_i) + b - z_i \leq \varepsilon, \\ & \quad z_i - \boldsymbol{\omega}^\top \phi(\mathbf{x}_i) - b \leq \varepsilon, \\ & \quad i = 1, \dots, N-d. \end{aligned}$$

Frequently no such function  $f(\mathbf{x})$  exists to satisfy these constraints for all points. Therefore slack variables  $\xi_i$  and  $\xi_i^*$  are introduced for each point to deal with potentially infeasible constraints. The slack variables allow regression error to exist up to the value of  $\xi_i$  and  $\xi_i^*$  yet still satisfy the required conditions. Under the given parameter  $C > 0$  and  $\varepsilon > 0$ , the standard form of  $\varepsilon$ -SVR (Vapnik 1998) is

$$\begin{aligned} & \min_{\boldsymbol{\omega}, b, \boldsymbol{\xi}, \boldsymbol{\xi}^*} \frac{1}{2} \boldsymbol{\omega}^\top \boldsymbol{\omega} + C \sum_{i=1}^{N-d} \xi_i + C \sum_{i=1}^{N-d} \xi_i^* \\ & \text{subject to } \boldsymbol{\omega}^\top \phi(\mathbf{x}_i) + b - z_i \leq \varepsilon + \xi_i, \\ & \quad z_i - \boldsymbol{\omega}^\top \phi(\mathbf{x}_i) - b \leq \varepsilon + \xi_i^*, \\ & \quad \xi_i, \xi_i^* \geq 0, i = 1, \dots, N-d, \end{aligned}$$

where parameter  $C$  determines the trade-off between model complexity and training error. To find the optimal solutions of the above equation, a dual formula of the above equation is introduced for simpler computation Vapnik (2013):

$$\begin{aligned} & \min_{\boldsymbol{\alpha}, \boldsymbol{\alpha}^*} \frac{1}{2} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)^\top Q (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) + \varepsilon \sum_{i=1}^{N-d} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{N-d} z_i (\alpha_i - \alpha_i^*) \\ & \text{subject to } \sum_{i=1}^{N-d} (\alpha_i - \alpha_i^*) = 0, \\ & \quad 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \dots, N-d, \end{aligned}$$

where  $Q_{i,j} = K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j)$ . After taking the Lagrangian and conditions for optimality, we can find the model solution in the above dual representation,

$$f(\mathbf{x}) = \sum_{i=1}^{N-d} (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b$$

where  $\alpha_i, \alpha_i^*$  are non-zero Lagrangian multipliers and the solution for the dual problem.  $K(\mathbf{x}_i, \mathbf{x})$  is the kernel function, which represents the inner product  $\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}) \rangle$ . A popular choice of kernel function is the [Radial Basis Function \(RBF\)](#) because of its capacities and simple implementation ([Huang et al. 2010](#)):

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (10)$$

where  $\gamma$  is the width parameter of RBF kernel. Other popular alternative choices of kernel functions include

- linear:  $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle$ ,
- polynomial:  $K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \langle \mathbf{x}_i, \mathbf{x}_j \rangle + r)^d$ ,
- sigmoid:  $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \langle \mathbf{x}_i, \mathbf{x}_j \rangle + r)$ .

### 2.3.4 Gaussian process regression

Gaussian process regression (GPR) is a kernel-based Bayesian approach to regression that not only provides a point prediction of the target object but also the associated predictive distribution to indicate the prediction's uncertainty. Consider a training set  $\{(\mathbf{x}_i, z_i) : i = 1, \dots, N - d\}$  in a given time series  $\{x_t : t = 1, \dots, N\}$ , where  $\mathbf{x}_i = (x_i, \dots, x_{i+d-1}) \in R^d$  is the input vector constructed by  $d$  lagged entries and  $z_i = x_{i+d} \in R$  is the target output. The objective is to capture the dependence structures between  $\mathbf{x}_i$  and  $z_i$  by treating them as  $N - d$  jointly Gaussian distributed random variables. Specifically, we want to obtain the function  $f(x)$  such that

$$z_i = f(\mathbf{x}_i) + \varepsilon_i, \quad i = 1, \dots, N - d$$

where  $\varepsilon_i$  is normally distributed with mean zero and variance  $\delta^2$ . Defining  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{N-d}]^\top$  and  $\mathbf{Z} = [z_1, z_2, \dots, z_{N-d}]^\top$ , we have

$$\mathbf{Z} \sim \mathcal{N}(m(\mathbf{X}), K(\mathbf{X}, \mathbf{X}) + \delta^2 \mathbf{I}),$$

where  $m(\mathbf{X}) = E[f(\mathbf{X})]$  denotes the mean function and  $K(\mathbf{X}, \mathbf{X})$  the covariance matrix with elements  $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ . Usually, for notational simplicity we will take the mean function to be zero since GPRs are flexible enough to model the mean arbitrarily well, although this need not to be done (see [Williams & Rasmussen 2006](#)), thereby a simpler representation of the distribution of  $\mathbf{Z}$  is

$$\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, K(\mathbf{X}, \mathbf{X}) + \delta^2 \mathbf{I}).$$

And similarly

$$f(\mathbf{X}) \sim \mathcal{N}(\mathbf{0}, K(\mathbf{X}, \mathbf{X})).$$

The elements in the covariance matrix are often estimated with the aid of kernel functions. A common choice of kernel function is the radial basis functions given in (10), thereby the covariance matrix is given by

$$k(\mathbf{x}_i, \mathbf{x}_j) = \delta_s^2 \exp\{-\gamma(\mathbf{x}_i, \mathbf{x}_j)^T(\mathbf{x}_i, \mathbf{x}_j)\}$$

where  $\delta_s^2$  denotes the signal variance and  $\gamma$  is the width parameter of RBF kernel function. Intuitively, the covariance matrix is constructed by assigning more value to the covariance entries whose corresponding training input vectors pairs are more alike. Thus, if points  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are considered to be similar by the kernel function values,  $f(\mathbf{x}_i)$  and  $f(\mathbf{x}_j)$  are also expected to be close.

When we want to make a prediction of  $z_*$  based on an unseen new input vector  $\mathbf{x}_*$ , we assume function of the new input  $f(\mathbf{x}_*)$  and the observed target output are also jointly Gaussian and can be represented by

$$\begin{bmatrix} f(\mathbf{x}_*) \\ \mathbf{Z} \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} k(\mathbf{x}_*, \mathbf{x}_*) & k(\mathbf{x}_*, \mathbf{X}) \\ k(\mathbf{X}, \mathbf{x}_*) & K(\mathbf{X}, \mathbf{X}) + \delta^2 \mathbf{I} \end{bmatrix}\right).$$

We are interested in the conditional distribution of  $f(\mathbf{x}_*)$  given the input vector  $\mathbf{x}_*$  and the knowledge learned in the observed training set  $(\mathbf{X}, \mathbf{Z})$ . The conditional distribution of  $f(\mathbf{x}_*)|\mathbf{x}_*, (\mathbf{X}, \mathbf{Z})$  can be obtained through the formulae that gives the conditional distribution of a joint Gaussian distribution using Bayesian Theorem, namely if

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right).$$

then

$$X_1|X_2 \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}).$$

By substituting the above formula, we have

$$f(\mathbf{x}_*)|\mathbf{x}_*, (\mathbf{X}, \mathbf{Z}) \sim N(\mu_*, \Sigma_*)$$

where

$$\begin{aligned}\mu_* &= k(\mathbf{x}_*, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \delta^2 \mathbf{I}]^{-1} \mathbf{Z} \\ \Sigma_* &= k(\mathbf{x}_*, \mathbf{x}_*) - k(\mathbf{x}_*, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \delta^2 \mathbf{I}]^{-1} k(\mathbf{X}, \mathbf{x}_*).\end{aligned}$$

The value of  $\mu_*$  is in general used as the prediction of the target output  $z_*$ , and  $\Sigma_*$  indicates the uncertainty on the prediction.

The optimal value of hyper-parameters  $\delta^2$ ,  $\delta_s^2$  and the parameters for the kernel function  $\gamma$  can be estimated by maximizing the log marginal likelihood using common optimization procedures such as Quasi-Newton methods.

### 3 PE and its variants

This chapter provides the formal definition of the PE measure and summarizes its properties. In addition, several variant measures originated from PE are also included in this chapter. The main objective of this chapter is to explore and extend the use of PE in analyzing temporal dependence structures in time series, and to propose a new measure, Permutation Dependence (PD) which tracks and indicates the deterioration of the temporal dependence between observations at increasing lags.

Concepts based on entropy are closely connected and are well suited for studying temporal dependencies in time series. Any kind of statistical serial correlation, not only linear correlation, is part of the analysis of temporal dependence structures. Traditionally, linear serial correlation is quantified by the ACF, and nonlinear correlations have been studied through higher-order moments. As an example, [Hinich \(1982\)](#) proposed the bispectrum test based on bicorrelations, or third-order moments to test nonlinear temporal dependence. The entropy measure investigates the temporal dependence from the aspect of information content. The strength of the temporal dependence structure is reflected in the measure of entropy so that for a random process with no temporal dependence structure, the entropy measure should reach its maximum. On the other hand, processes with strong temporal dependence are more predictable and thus expected to have a relatively low value of entropy. Compared to traditional methods, the use of PE in the analysis of the temporal dependence structures is advantageous since it is not grounded in linear or quasi-linear analysis of relationships.

Instead, it may be termed non-parametric since it does not require any prior knowledge and assumptions about the functional form of the system. Additionally, using ordinal patterns as a symbolic measure is potentially better than using actual time series values in the sense that it helps to preserve the weak causality dependence structures, especially nonlinear structures in the presence of highly volatile noise (Bandt & Pompe 2002, Parlitz et al. 2012, Bandt 2005, Groth 2005).

In light of the relation between PE and the linear correlation function in a Gaussian process derived by Bandt & Shiha (2007), PE at delay  $\tau$  is not only determined by the temporal dependence at the selected delay but also the temporal dependence at  $2\tau, 3\tau, \dots, (D-1)\tau$ , where  $D$  is the chosen segment length parameter required in the computation of PE. By modifying the specification of PE, I propose a new measure, Permutation Dependence (PD), so that PD has one-to-one correspondence to the temporal dependence at the selected delay. The new measure PD is easier to comprehend and interpret for finance practitioners because its behaviour is similar to the most commonly used serial dependence measure, ACF, but not grounded to detecting linear structures as is ACF. In addition, I find that PE tends to neglect the evidence of determinism in a process in which there is a significant but slowly varying dependence structure for increasing lags. PD remedies this limitation of PE and is proven to be an important complementary tool to PE in later empirical analysis.

### 3.1 Definition of PE

Given a univariate time series  $X = \{x_t; t = 1, \dots, N\}$ , for chosen segment length  $D$  and delay  $\tau$ , we compose a vector consisting of the first and the next  $D$ -th subsequent entries with every chosen entry  $\tau$  interval apart:

$$x_{1,\tau}^D = (x_1, x_{1+\tau}, \dots, x_{1+(D-1)\tau}).$$

Likewise we can construct total  $N - (D-1)\tau$  vectors of the same length by sliding the partition window one step ahead per time along the observed time series:

$$x_{t,\tau}^D = (x_t, x_{t+\tau}, \dots, x_{t+(D-1)\tau}), \quad t = 1, 2, \dots, N - (D-1)\tau. \quad (11)$$

For each constructed vector  $x_{t,\tau}^D$ , by comparing the value of the entries in the vector, each segment  $x_{t,\tau}^D$  is mapped into one of the  $D!$  ordinal patterns  $\{\pi_i; i = 1, \dots, D!\}$  excluding equal entry circumstances (equal entry circumstances are discussed in section 5.2). For instance, if

$D = 3$ , there are 6 distinct ordinal patterns, which are:

Ordinal Pattern	$(a_x b_x c_x)$	Condition
$\pi_1$	(321)	$x_t > x_{t+\tau} > x_{t+2\tau}$ ,
$\pi_2$	(312)	$x_t > x_{t+2\tau} > x_{t+\tau}$ ,
$\pi_3$	(231)	$x_{t+\tau} > x_t > x_{t+2\tau}$ ,
$\pi_4$	(213)	$x_{t+2\tau} > x_t > x_{t+\tau}$ ,
$\pi_5$	(132)	$x_{t+\tau} > x_{t+2\tau} > x_t$ ,
$\pi_6$	(123)	$x_t < x_{t+\tau} < x_{t+2\tau}$ .

The probability of each ordinal pattern is estimated by the number of the constructed segments that are mapped into  $\pi_i$ , divided by the total number of the segments, that is,

$$p_\tau^D(\pi_i) = \frac{\#\{x_{t,\tau}^D | x_{t,\tau}^D \text{ has ordinal pattern } \pi_i\}}{N - (D - 1)\tau}, \quad i = 1, \dots, D!.$$

Inheriting the definition of Shannon's entropy, the permutation entropy for the univariate time series  $X$ , with chosen vector length  $D$  and delay  $\tau$  is defined as

$$\text{PE}_\tau^D(X) = \sum_{i=1}^{D!} \frac{-p_\tau^D(\pi_i) \ln(p_\tau^D(\pi_i))}{\ln D!}.$$

The value of permutation entropy is bounded between 0 and 1, where 0 indicates a completely predictable dynamic and 1 indicates completely stochastic dynamic. We use

$$1 - \text{PE}_\tau^D(X)$$

to measure the level of departure from randomness of the dynamics underlying the observed time series. [Matilla-García & Marín \(2008\)](#) show that under the assumption that the time series  $X$  is iid,

$$2\{[N - (D - 1)\tau] \ln D!\} [1 - \text{PE}_\tau^D(X)]$$

is asymptotically  $\chi_{D!-1}^2$  distributed.

Note that the way the PE constructs the segment in  $X$  originates from Taken's theorem, where a time-delayed version of one generic signal would suffice to embed the  $n$ -dimensional manifold. The segment length  $D$  is identical to Takens's embedding dimension. Taken's embedding theorem is particularly used in reconstructing the attractor in the domain of chaos. According to [Bandt & Pompe \(2002\)](#),  $D$  is recommended to be chosen in the range of  $3 \leq D \leq 7$ . The choice of parameter  $D$  needs to be subject to the restriction:  $N - (D - 1)\tau \gg D!$  ([Rosso](#)

et al. 2007, Kowalski et al. 2007). The rationale behind the recommended range of  $D$  is that, since we partition the observed time series into numbers of overlapping segments and map the constructed segments into  $D!$  possible ordinal patterns, the constructed number length,  $N - \tau - 1$ , should excessively exceed the number of possible pattern categories. If data are relatively short with a sample size of 1000, we recommend  $D = 3$ . If data are relatively long with a sample size of 10000, we recommend  $D = 4$  or 5. For the sake of simplicity and computational efficiency, I choose  $D = 3$  to be the default segment length for all the PE analysis in this paper. I also conduct the analysis for  $D = 4$  with similar results.

### 3.2 Advantages of PE

The most advantages of PE reside in its two main attributes. First, is its model-free nature that allows it to be not grounded to any form of relations and requires no prior assumptions and knowledges about the underlying dynamics of the investigated data. Second, is its ability to unveil the determinism in the dynamics of a system in the presence of highly volatile noise and non-stationarity (Kreuzer et al. 2014). As I illustrate later in the thesis, PE can uncover many structures that other tools are likely to overlook. Using an ordinal pattern as a symbolic measure is advantageous compared to using actual time series values because it helps preserve the weak causality dependence structures, especially nonlinear structures, in the presence of highly volatile noise (Bandt & Pompe 2002, Parlitz et al. 2012, Bandt 2005, Groth 2005). Moreover, Bandt & Shiha (2007) prove that all stationary ergodic processes can be recovered from the one-dimensional distribution and the order structure of the time series. Therefore PE neglects the one-dimensional distribution of the investigated data, but preserves the casual dependence structures underlying the governing dynamics by only considering the order structure. Additionally, PE has the advantage of invariance to monotonic transformation. Accordingly, drifts or scaling artificially introduced by a measurement device will not modify the quantifiers estimation, a nice property if one deals with experimental data or financial series where a number of functional forms of the objects of interest are often considered in conjunction, for example, absolute and squared returns.

### 3.3 Relation of PE and ACF in a Gaussian process

PE was originally proposed as a measure of complexity. Complexity can be an unfamiliar and rarely used term by researchers in the finance area. To better comprehend the type of information that PE reflects, in the section I will illustrate the connection between PE and ACF in a Gaussian process to establish the link between PE and the most familiar terms in financial time series analysis.

Bandt & Shiha (2007) derive the formula that connects the value of PE with the autocorrelation function in the classical model of time series analysis: Gaussian processes. PE is computed from the probability of ordinal patterns. For chosen segment length 3 ( $D = 3$ ), the probability of ordinal pattern  $\pi_{123}$  of delay  $\tau$  in a stationary Gaussian process is given below:

$$p_{\tau}^{D=3}(\pi_{123}) = \frac{1}{\pi} \arcsin \left( \frac{1}{2} \sqrt{\frac{1 - \rho_{2\tau}}{1 - \rho_{\tau}}} \right), \quad (12)$$

where  $\rho_{\tau}$  is its autocorrelation function of Gaussian process at lag  $\tau$ ,

$$p_{\tau}^{D=3}(\pi_{123}) = p_{\tau}^{D=3}(\pi_{321}), \quad (13)$$

and the remaining patterns have equal probabilities. Therefore by knowing  $p_{\tau}^{D=3}(\pi_{123})$ , the probabilities of all ordinal patterns and the value of PE can be determined. According to equation (12), the value of PE depends on two components: the value of  $\rho_{\tau}$  itself, and the ratio between the correlation  $\rho_{\tau}$  and  $\rho_{2\tau}$ . In other words, the ordinal pattern distribution of a Gaussian process reflects two components: the strength of dependence between the current and the lag  $\tau$  period's entries in a given time series; and the rate that the dependence structure changes from lag  $\tau$  to  $2\tau$ . Similarly, Bandt & Shiha (2007) also determine the formula for patterns of length 4 ( $D = 4$ ). However, for patterns of length longer than 5 there are no closed formulas.

Based on the results from Bandt & Shiha (2007), for a chosen pattern length of 3, all stationary Gaussian processes have equally probable monotonic ordinal patterns ( $\pi_{123}, \pi_{321}$ ) and non-monotonic ordinal patterns ( $\pi_{213}, \pi_{132}, \pi_{231}, \pi_{312}$ ) at any arbitrary delay. The reason behind the equally probable monotonic patterns in a Gaussian process is understandable. The equality is governed by two features of the underlying dynamics. First, the symmetrical distributions of Gaussian process that lead to equal probabilities of positive and negative increments:

$$p(X_t - X_{t-\tau} > 0) = p(X_t - X_{t-\tau} < 0).$$

Second, the deterministic relation that ensures the impartial responds to positive and negative increments:

$$p(X_t - X_{t-\tau} > 0 | X_{t-\tau} - X_{t-2\tau} > 0) = p(X_t - X_{t-\tau} < 0 | X_{t-\tau} - X_{t-2\tau} < 0),$$

where  $\tau$  is an arbitrary lag. As a result

$$p(X_t > X_{t-\tau} > X_{t-2\tau}) = p(X_t < X_{t-\tau} < X_{t-2\tau}).$$

If the correlation at lag  $\tau$  is positive, the monotonic ordinal patterns are always more probable than non-monotonic ordinal patterns.

In fact, not only the Gaussian process, but any stationary time series with a linear structure and independent symmetrical distributed noise has equally probable monotonic patterns for any arbitrage delay  $\tau$ . The proof is given below. For any stationary time series  $\{X_t\}$ ,

$$\mathbb{E}(X_t - X_{t-\tau}) = 0.$$

Suppose every stationary time series  $X_t$  can be written in a form of

$$X_t = c + g(X_{t-1}, \dots, X_{t-d}) + \varepsilon_t,$$

where  $c$  is a constant,  $g(\cdot)$  is a time-invariant function and  $\varepsilon_t$  is the noise term with both conditional and unconditional mean of zero, we have

$$X_t - X_{t-\tau} = f(X_{t-1}, \dots, X_{t-d-\tau}) + \varepsilon_t - \varepsilon_{t-\tau},$$

where

$$f(X_{t-1}, \dots, X_{t-d-\tau}) = g(X_{t-1}, \dots, X_{t-d}) - g(X_{t-1-\tau}, \dots, X_{t-d-\tau}).$$

Since

$$\mathbb{E}[X_t - X_{t-\tau}] = \mathbb{E}[f(X_{t-1}, \dots, X_{t-d-\tau}) + \varepsilon_t - \varepsilon_{t-\tau}] = 0,$$

and  $\mathbb{E}(\varepsilon_t) = 0$ ,  $\mathbb{E}(\varepsilon_{t-\tau}) = 0$ , thereby

$$\mathbb{E}[f(X_{t-1}, \dots, X_{t-d-\tau})] = 0.$$

when  $\{X_t\}$  has linear structure and independent symmetric noise,

$$\mathbb{E}[(X_t - X_{t-\tau})^3] = \mathbb{E}[f(X_{t-1}, \dots, X_{t-d-\tau})^3] + \mathbb{E}[(\varepsilon_t - \varepsilon_{t-\tau})^3].$$

Therefore, the kurtosis of  $X_t - X_{t-\tau}$  is zero so that its distribution is symmetric around its mean 0. If we assume  $X_t$  is a continuous time series so that ties are negligible, the symmetrical distribution and zero mean of  $X_t - X_{t-\tau}$  lead to

$$p(X_t > X_{t-\tau}) = p(X_t < X_{t-\tau}).$$

Since the response to increments and decrements in linear relation is symmetric, thereby

$$p(X_t > X_{t-\tau} | X_{t-\tau} > X_{t-2\tau}) = p(X_t < X_{t-\tau} | X_{t-\tau} < X_{t-2\tau}).$$

In addition,

$$\begin{aligned} p_\tau^{D=3}(\pi_{123}) &= p(X_t > X_{t-\tau} | X_{t-\tau} > X_{t-2\tau})p(X_t > X_{t-\tau}), \\ p_\tau^{D=3}(\pi_{321}) &= p(X_t < X_{t-\tau} | X_{t-\tau} < X_{t-2\tau})p(X_t < X_{t-\tau}) \end{aligned}$$

we have  $p_\tau^{D=3}(\pi_{123}) = p_\tau^{D=3}(\pi_{321})$ .

The formulas of ordinal pattern probabilities in time series that are non-stationary, that have non-linear dependence structures or that have asymmetrical noise are more difficult to derive. However, as I demonstrate in the later sections, their ordinal pattern distributions show up in defined and informative ways.

### 3.4 PE variant permutation dependence (PD)

In the previous section, I demonstrated that for a Gaussian process, the value of  $PE_\tau^{D=3}$  is positively correlated with the ACF over lag  $\tau$ , but is offset by the ACF at lag  $2\tau$ . The relation between the ACF and PE in the most classic linear process indicates that  $PE_\tau^D$  represents a composite of multiple temporal dependence relation over  $(D - 1)$  integer multiple of the selected delay. In other words, the value of  $PE_\tau^D$  not only reflects the temporal dependence structure of the selected delay, but is also influenced by, more specifically, negatively related to the temporal dependence structure over integer multiple of selected delay:  $2\tau, 3\tau$  up to  $(D - 1)\tau$ . The absence of one-to-one correspondence between  $PE_\tau^D$  and the temporal dependence structure at  $\tau$  complicates its interpretation and implication. Moreover, if a time series is generated from a process with significant but slowly diminishing dependence structure for increasing lags, PE might not be able to detect it. In order to fix these limitations of PE, I propose a new measure, permutation dependence (PD). The new measure  $PD_\tau^D$  indicates the strength of temporal dependence at the selected delay  $\tau$  and is not affected by the dependence structures of other delays.

The new dependence measure  $PD_\tau^D$  is an extension of PE. Unlike PE, the value of which reflects multiple offsetting components,  $PD_\tau^D$  only reflects the dependence relation at the selected delay  $\tau$ . The computation of the new dependence measure  $PD_\tau^D$  requires three steps. First, we compute the probability of each ordinal pattern by extracting relative ranks of entries within every constructed segments  $x_{t,\tau}^D = (x_t, x_{t+1}, \dots, x_{t+D-2}, x_{t+D-2+\tau})$ ,  $t = 1, 2, \dots, N - D + 2 - \tau$

from the investigated time series  $X = \{x_t; t = 1, \dots, N\}$ . Then we replace the value of  $x_{t+D-2+\tau}$  in every constructed segment with a randomly generated random variable  $r$ , which has the same distribution as the original time series  $X$  using kernel functions (see Peter D (1985), Altman & Leger (1995), Jones (1990) for more details in estimating distribution using kernel functions), and compute the new ordinal pattern probabilities. Lastly, by repeating the second step a sufficient number of times (for this study I repeated 500 times),  $\text{PD}_\tau^D$  is equal to the discrepancy between ordinal pattern probabilities in the original time series and the mean of that in the simulated segments, standardized by the standard deviation of simulated probabilities. The formal specification of new temporal structure measure  $\text{PD}_\tau^D$  is given below. The new dependence measure  $\text{PD}_\tau^D$  is computed by:

$$\text{PD}_\tau^D(X) = \sum_{\pi_i} \left( \frac{p_\tau(\pi_i) - \text{E}[p_\tau^{\text{rand}}(\pi_i)]}{\text{s.d.}[p_\tau^{\text{rand}}(\pi_i)]} \right)^2$$

where  $(\pi_i : i = 1, \dots, D!)$  represents the collection of  $D!$  number of possible distinct ordinal patterns of length  $D$ , and

$$p_\tau(\pi_i) = \frac{\#\{x_{t,\tau}^D | x_{t,\tau}^D \text{ has ordinal pattern } \pi_i\}}{N - (D - 1)\tau}, \quad i = 1, \dots, D!,$$

$$p_\tau^{\text{rand}}(\pi_i) = \frac{\#\{x_{t,\tau}^{D,\text{rand}} | x_{t,\tau}^{D,\text{rand}} \text{ has ordinal pattern } \pi_i\}}{N - (D - 1)\tau}, \quad i = 1, \dots, D!,$$

and

$$\begin{aligned} x_{t,\tau}^D &= (x_t, x_{t+1}, \dots, x_{t+D-2}, x_{t+D-2+\tau}), \quad t = 1, 2, \dots, N - D + 2 - \tau, \\ x_{t,\tau}^{D,\text{rand}} &= (x_t, x_{t+1}, \dots, x_{t+D-2}, r_{t+D-2+\tau}), \quad t = 1, 2, \dots, N - D + 2 - \tau, \end{aligned}$$

where  $\{r_t; t = 1, \dots, N\}$  is generated from a purely random process and has the same distribution as  $\{x_t; t = 1, \dots, N\}$ .

The rationale behind the new dependence measure  $\text{PD}_\tau^D$  is that  $\text{PD}_\tau^D$  measures the indication power of the ordinal patterns in the  $\tau$  lagged consecutive historical observations in predicting the upcoming entry. If the time series under study follows completely stochastic dynamics, we expect no difference can be detected between the ordinal pattern distribution in the partitioned segments  $x_{t,\tau}^D = (x_t, x_{t+1}, \dots, x_{t+D-2}, x_{t+D-2+\tau})$  from the original data and simulated segments  $x_{t,\tau}^{D,\text{rand}} = (x_t, x_{t+1}, \dots, x_{t+D-2}, r_{t+D-2+\tau})$ , since in both cases the ordinal pattern in the past

entries have no impact and hence no indicating power for the upcoming  $x_{t+D-2+\tau}$  or randomly generated entry  $r_{t+D-2+\tau}$ . Thereby  $\text{PD}_\tau^D$  will be close to zero. On the contrary, if the time series has a temporal dependence structure at delay  $\tau$ , significant difference will be detected between the ordinal patterns in the original series and those in the simulations. We are using sum of squares of the standardised differences to measure the discrepancy between  $p_\tau$  and  $p_\tau^{\text{rand}}$ . By choosing the squared difference rather than absolute difference, the measure may inflate the large values of standardized differences. However, since the measure is based on ordinal patterns, such design ensures the measure to be robust to outliers. The squared function is chosen to enhance the sensitivity of the measure to detect deterministic structures.

It is worth mentioning that the constructed vectors  $x_{t,\tau}^D$  in PD are not exactly the same as those in PE. The observations in the constructed vectors in PE are  $\tau$  interval apart, whereas, in PD, the first  $D - 1$  observations in  $x_{t,\tau}^D$  are consecutive, only the last entry is  $\tau$ -step ahead. The purpose of the modifications made into the constructed vectors in PD is to ensure that PD exclusively accounts for the temporal dependence at the selected delay. Additionally, the temporal structure between the consecutive entries is always the strongest in most real-world time series. The design of the constructed segment also enhances the sensitivity of PD in detecting deterministic structures.

By computing and plotting  $\text{PD}_\tau^D$  over increasing delay  $\tau$ , the evolution of the dependence structure between further sparse entries in a given time series can be revealed, just like the plot of a sample ACF.  $\text{PD}_\tau^D$  is advantageous compared to the sample ACF since the sample ACF only extracts linear temporal dependence relations, whereas  $\text{PD}_\tau^D$  is not restricted by any form of dependence structures.

The required parameter of statistic PD is identical to that in the computation of PE, namely the pattern length  $D$  and delay  $\tau$ . For the same reason as the requirements of the parameters in the PE, segments length  $D$  and delay parameter  $\tau$  needs to be subject to the restriction:  $N - (D - 1)\tau \gg D!$  in measure PD. In general, I recommend to choose  $D$  to be 3, 4 or 5 depending on the length of the investigated data.

I generate a time series with a has significant but slowly diminishing dependence structure to show how the newly proposed measure  $\text{PD}_\tau^D$  compensates for the weakness in PE. The generated time series is from an ARMA(1,1) process

$$y_t = 0.2 + 0.98y_{t-1} - 0.95\varepsilon_{t-1} + \varepsilon_t$$

where  $\text{Var}(\varepsilon_t) = 1$ . As shown in the sample and theoretical ACF plot of the simulated process (Figure 1(a)), the generated time series has significant but slowly diminishing linear temporal

dependence structures. Figure 1(b) displays the plot of  $1 - \text{PE}_\tau^{D=3}$  against increasing delays of the generated time series. The value of  $1 - \text{PE}_\tau^{D=3}$  fluctuates at an insignificant level, indicating the PE's inability to detect a slowly diminishing dependence structure. Figure 1(c) illustrates the fact that the new dependence measure  $\text{PD}_\tau^{D=3}$  successfully captures the weak and persistent dependence structure in the simulated ARMA(1,1) time series. Moreover, the shape of the plot of new dependence measure  $\text{PD}_\tau^D$  over increasing delays resembles the plot of theoretical ACF of the generated ARMA(1,1) time series indicating the consistency of the proposed measure with the ACF for time series governed by linear dynamics.

In fact, PE and PD are complementary to each other. When PE is insignificant, it is not possible to tell whether the investigated series is subject to lack of dependence structures or has slowly varying dependence. The new dependence measure  $\text{PD}_\tau^D$  helps to differentiate between them. However, as I am going to show in the next section,  $\text{PD}_\tau^D$  suffers from the problem of low resolution due to the fact that statistic  $\text{PD}_\tau^D$  has relatively large variance. As a result, two time series having different strength of structures can end up with similar value of  $\text{PD}_\tau^D$ . In that case, PE can provide better resolution for differentiating different levels of determinism. Besides, the PD measure is susceptible to non-stationarity whereas PE is relatively robust to it.

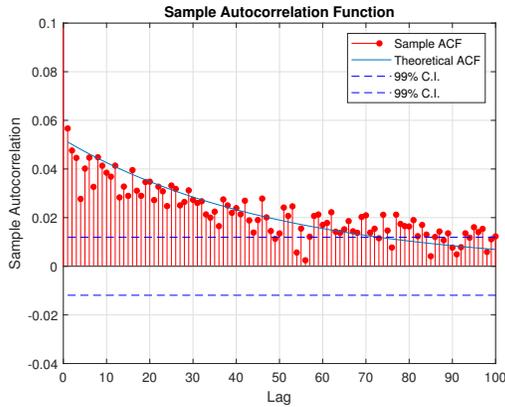
### 3.5 Comparison between PE, PD and ACF

PE, PD and ACF all aim at detecting and capturing temporal dependence structure underlying an investigated time series, and are all used as preliminary tools to explore and provide guidance for later modelling and forecasting procedures. However, their implications and interpretations are quite different when conducted on the same time series. The plots of PE, PD and ACF over increasing delays/lags place emphasis on different aspects of the dependence structures underlying the observed time series. In this section, I demonstrate their similarities and differences with the aid of a number of simulation studies.

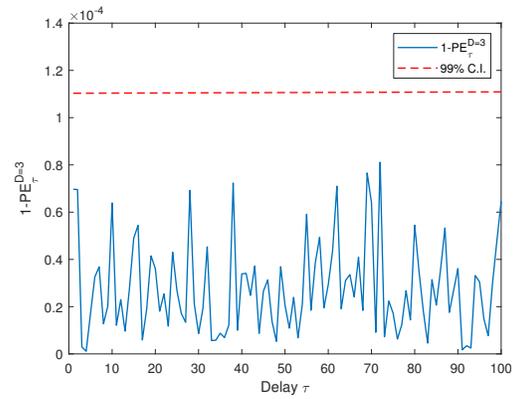
The ACF is the most widely used tool in time series analysis to measure the linear dependence of a process with a delayed copy of itself. The ACF of a given time series  $X = \{x_t : t = 1, \dots, n\}$  at delay  $\tau$  can be estimated by the sample ACF coefficient:

$$r_\tau = \frac{\sum_{t=1}^{n-\tau} (x_t - \bar{x})(x_{t+\tau} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}$$

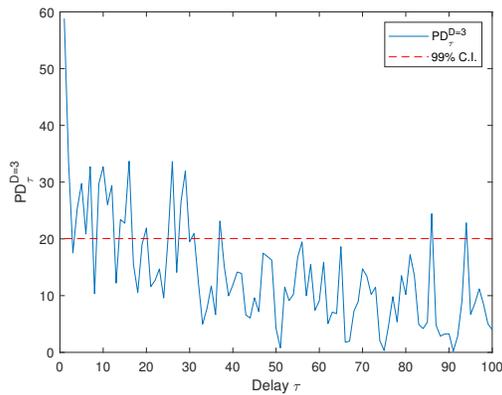
where  $\bar{x}$  is the sample mean of the given time series. In financial time series analysis, the plot of ACFs, together with plot of [partial autocorrelation function \(PACF\)](#) on returns and squared returns, is commonly used in detecting trend and periodicities exhibited in the financial



(a)



(b)



(c)

Figure 1: (a) Plot of ACF against increasing lags of the simulated series generated from an ARMA(1,1) model with significant but slowing diminishing linear dependence structures. (b) Plot of  $1 - PE_{\tau}^{D=3}$  against increasing delays of the simulated series. (c) Plot of  $PD_{\tau}^{D=3}$  against increasing delays of the simulated series.

asset dynamics and in choosing the order parameters  $p, q$  of ARIMA( $p, q$ ) and GARCH( $p, q$ ) models. The simplicity of the tool makes it the most commonly used preliminary exploratory tool in time series analysis. However, it has several significant shortcomings. First, it only reflects linear relations, thus may neglect the non-linear structures if they exist. Second, in order to work properly, ACF requires the investigated time series to be at least weakly stationary and not too far from normality. Mikosch & Stărică (2004) show that it fails to distinguish between long-memory time series from non-stationary short-memory time series.

Unlike the ACF that is grounded in detecting linear dynamics, PE and PD measures are not restricted by any form of temporal dependence structures. They can be considered as universal structure detectors. As explained in the previous section, PE and PD reveal the underlying dependence structures differently. PE reflects the joint effect of two contributing components: the dependence structure at the chosen delay and the dependence varying rate for increasing delays, whereas PD is designed to modify this drawback of PE so that it only quantifies the temporal dependence relation at the selected delay.

To visually illustrate how  $PE_{\tau}^D$ ,  $PD_{\tau}^D$  and ACF reveal a time series temporal dependence structure differently, I simulate 500 time series of length 38160 from two ARMA models, model 1:

$$y_t = 0.2 + 0.5y_{t-1} - 0.3\varepsilon_{t-1} + \varepsilon_t$$

where  $\text{Var}(\varepsilon_t) = 0.01$ , and model 2:

$$y_t = 0.16 + 0.6y_{t-1} - 0.41\varepsilon_{t-1} + \varepsilon_t$$

where  $\text{Var}(\varepsilon_t) = 0.01$ , then compute the average of ACF,  $1 - PE_{\tau}^{D=3}$  and  $PD_{\tau}^{D=3}$  against increasing lags/delays of the generated times series from the above two models, along with their respective 2.5% and 97.5% percentiles. The time series governed by the above two models have the same expected unconditional mean and variance. Moreover, I specifically select the parameters of the above two ARMA processes so that they have very close ACF at lag 1 but decay at different rates, resulting in unequal ACFs at longer lags (see Figure 2(a)).

As shown in Figure 2(a) and Figure 2(c), for a linear process, the ACF plot and  $PD_{\tau}^{D=3}$  plot are in close resemblance, except that the ACF has narrower variance. That means for a linear process, ACF is better at differentiating the small discrepancy between two processes when compared to the universal dependence detector PD. In contrast with ACF and PD, PE cannot indicate the temporal dependence structure at individual delays. This weakness of PE is clearly illustrated in Figure 2(b), which shows that the value of  $PE_{\tau=1}^{D=3}$  of a time series generated from model 1 differs significantly from that from model 2 despite the fact that

they have similar strength of dependence relation at the corresponding delay. Nevertheless, the PE plot is advantageous in revealing different serial dependence diminishing rates of the two simulated models. As illustrated in section 3.3, the value of PE is positively correlated with two components: the strength of dependence between the current and the lag  $\tau$  period's entries in a given time series; and the rate that the dependence structure changes from lag  $\tau$  to  $2\tau$ . A rapidly diminishing rate would lead to larger discrepancies in both components that contributes to the value of PE.

Despite the advantages of higher resolutions, ACF lacks the ability to reveal non-linear dependence structures. This is best illustrated by comparing the plot of the three temporal dependence measures on simulations from an ARMA(1,1) model and a GARCH(1,1) model that share exactly the same linear serial correlations. The ARMA(1,1) model stays the same as the specification given in equation (3.5), and the simulated GARCH model is specified as follows:

$$\begin{aligned} y_t &= \sigma_t z_t, \\ \sigma_t^2 &= 0.2 + 0.2y_{t-1}^2 + 0.3\sigma_{t-1}^2, \end{aligned}$$

where  $\{z_t\}$  is a sequence of iid random variables with mean 0 and variance 1. The simulated GARCH(1,1) model can be arranged in the form of:

$$y_t^2 = 0.2 + (0.2 + 0.3)y_{t-1}^2 - 0.3v_{t-1} + v_t,$$

where  $v_t = y_t^2 - \sigma_t^2 = \sigma_t^2(z_t^2 - 1)$ . Unlike the innovation term in the ARMA process,  $\{v_t\}$  is not a sequence of iid random variables. Even though  $v_t$  has time-invariant zero conditional mean, its variance depends on the past histories  $v_{t-1}$  in a way that larger previous innovation is likely to lead to more volatile subsequent innovations. Therefore a GARCH model can be regarded as an ARMA process with innovations that have dependent structures (see section 2.3.2 for more detail). Figure 4 compares the sample ACF plot, PE plots and PD plots for the simulated ARMA and GARCH series over increasing delays/lag. It is clear that the simulated ARMA and GARCH series share the same ACF plots, but their PE plots and PD plots significantly differ. The value of PE and PD on the simulated ARMA series is significantly greater than that on the simulated GARCH series, especially for small delays, suggesting the simulated ARMA process has an overall stronger temporal dependence structures, compared to the simulated GARCH process. The reason for the weaker temporal dependence structures of the GARCH process is that the existence of the nonlinear dynamical structures in the innovation terms in the GARCH process weakens the determinism of the

linear serial dependence in the conditional mean of the process. The simulation study shows while the ACF plot is unable to differentiate the two simulated process, the PE and PD measures successfully unveil the existence of the additional non-linear temporal dependent structures exhibited in the innovation term of the GARCH process.

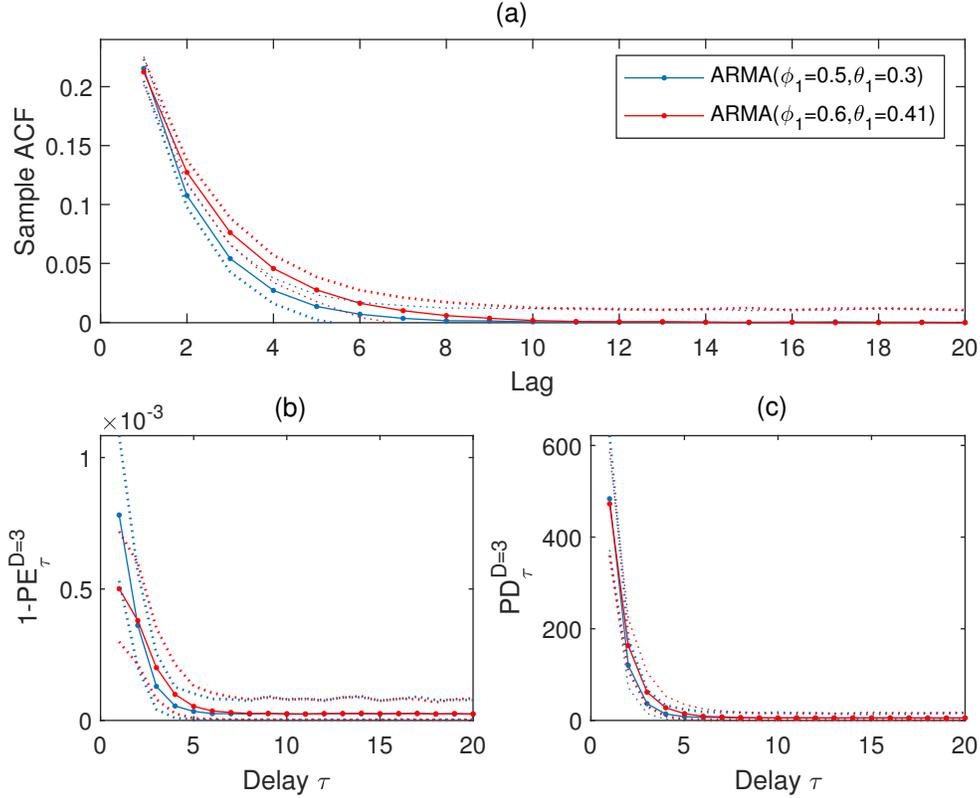


Figure 2: Comparison between average of sample ACF,  $1 - PE_{\tau}^{D=3}$  and  $PD_{\tau}^{D=3}$  on 500 paths of simulated time series following ARMA(1,1) model with  $\phi_0 = 0.2$ ,  $\phi_1 = 0.5$ ,  $\theta_1 = 0.3$ ,  $\sigma_{\epsilon} = 0.1$ , and following ARMA(1,1) model with  $\phi_0 = 0.16$ ,  $\phi_1 = 0.6$ ,  $\theta_1 = 0.41$ ,  $\sigma_{\epsilon} = 0.1$ , along with their respective 2.5% and 97.5% percentiles.

To ensure that the observed strengths and weaknesses of the three dependence measures persist across a broader set of underlying models, I also examine the behavior of PE, PD and ACF on simulations generated from a well-known nonlinear time series model, bilinear model. I simulate 500 paths of time series following the bilinear model specified below

$$y_t = 0.1 + 0.6\sigma_{t-1}\sigma_t + \sigma_t \quad (15)$$

where  $\sigma_t$  is i.i.d. normally distributed innovations with mean 0 and variance 1. Figure 4 provides the average, 2.5 percentile and 97.5 percentile of the sample ACF, 1-PE and PD computed on the simulated time series. Figure 4 shows that the sample ACFs of the simulated

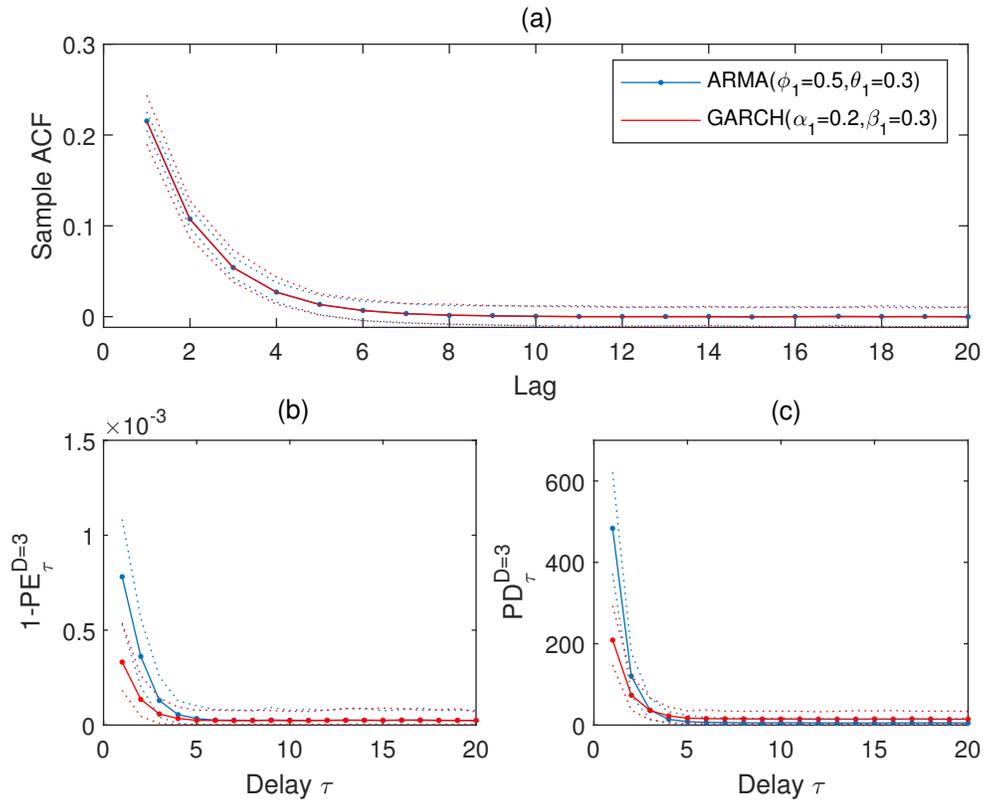


Figure 3: Comparison between average of sample ACF,  $1 - PE_{\tau}^{D=3}$  and  $PD_{\tau}^{D=3}$  on 500 paths simulation generated from ARMA(1,1) model with  $\phi_0 = 0.2$ ,  $\phi_1 = 0.5$ ,  $\theta_1 = 0.3$ , and from GARCH(1,1) model with  $\alpha_0 = 0.2$ ,  $\alpha_1 = 0.2$ ,  $\beta_1 = 0.3$ , along with their respective 2.5% and 97.5% percentiles.

time series are very close to zero, suggesting the absence of linear serial correlation in the simulated process. However, the significant value of 1-PE and PD, especially over the short-term, indicates that despite the insignificant ACF, the simulated time series is not independent. There are nonlinear temporal dependent structures present in the dynamics.

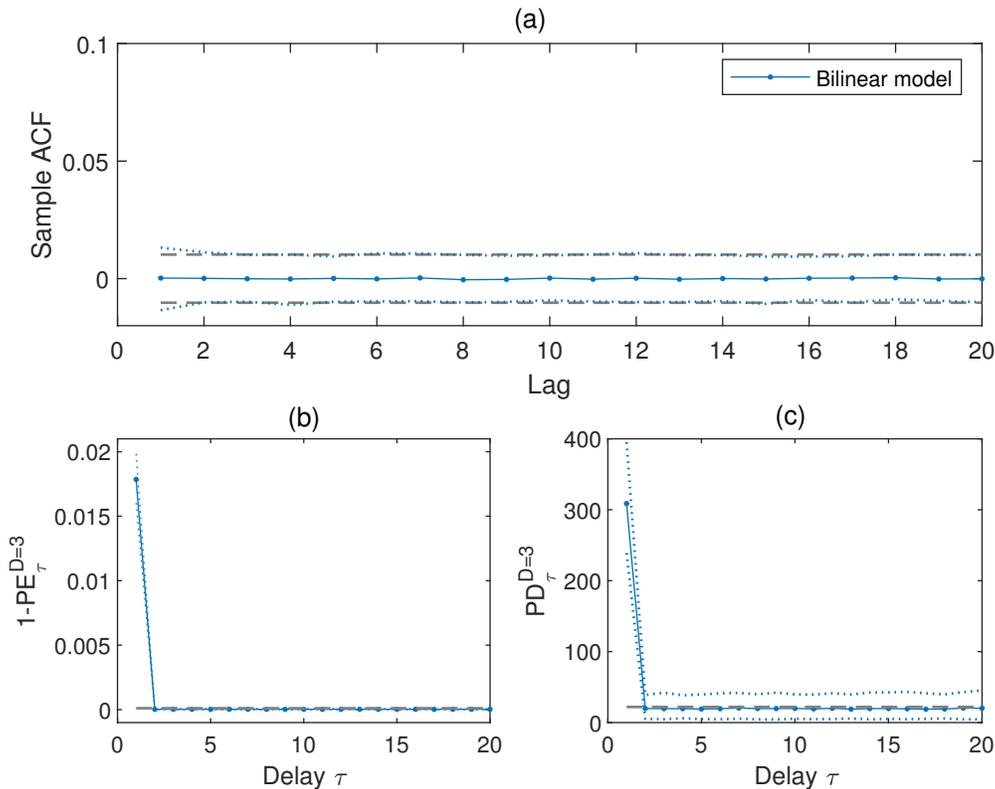


Figure 4: Comparison between average of sample ACF,  $1 - \text{PE}_{\tau}^{D=3}$  and  $\text{PD}_{\tau}^{D=3}$  on 500 paths simulation generated from bilinear model given in 15, along with their respective 2.5% and 97.5% percentiles (blue dashed line). The grey line represents the 99% C.I. of each measure under randomness.

To sum up, for a linear process, the PE and PD measures can both be expressed as a function of a system's ACF, with PE capturing the temporal dependence decay rate better than PD, but PD identifying the individual dependence strength at selected delay better than PE. However, they both have greater variance compared to ACF. If the given series follows a non-linear process, PE and PD are advantageous when compared to ACF as they are capable of characterizing and detecting the existence of non-linear temporal dependence structures. A summary table is given in Table 1 to display the respective characteristics of PE, PD and ACF.

Table 1: Summary of main characteristics of PE, PD and ACF.

	PE	PD	ACF
Detecting nonlinear structures	✓	✓	
Robust to nonstationarity	✓		
Quantifying the dependence at fixed lag		✓	✓
Moderate uncertainties			✓

### 3.6 PE, PD and ACF under non-stationarity

This section investigates the behaviours of PE, PD and ACF under non-stationarity. Non-stationary behaviour refers to the time-varying nature of the underlying dynamics of the system over time. Most time series display some level of non-stationarity, especially financial markets are prominent examples of highly non-stationary systems, where sample averaged moments such as means, variances and correlation coefficients strongly depend on the time window in which they are evaluated. The non-stationary nature of time series is often considered as the most challenging obstacle to many forecasting models and conventional analysis tools. Therefore, it is important to investigate and compare the impacts of non-stationarity on the PE and PD measures in order to avoid drawing erroneous inferences based on them.

In order to investigate how PE and PD are affected by non-stationarity, I examine their respective behaviours on simulated non-stationary time series. Mikosch & Stărică (2004) simulate a simple case of non-stationarity by concatenating time series generated from several distinct GARCH models, and test how the ACF responds to the non-stationarity exhibited in the simulated time series. I employ their approach and compute the values of PE, PD and ACF over increasing delays on the concatenated series as well as the individual simulated GARCH series. The strength of the temporal dependence structure underlying the concatenated series is expected to be equal to, or close to, the average temporal dependence strength of the individual contributing stationary GARCH process. Accordingly, the impact of the non-stationarity on the measures of interest is indicated by comparing their value on the simulated non-stationary time series with their average value on the individual stationary series generated from the constituting models.

The specification of the four GARCH models I use in the simulations are as follows: model 1:  $\alpha_0 = 0.2$ ,  $\alpha_1 = 0.2$ ,  $\beta_1 = 0.3$ , model 2:  $\alpha_0 = 0.24$ ,  $\alpha_1 = 0.4$ ,  $\beta_1 = 0.3$ , model 3:  $\alpha_0 = 0.06$ ,  $\alpha_1 = 0.3$ ,  $\beta_1 = 0.5$ , model 4:  $\alpha_0 = 0.08$ ,  $\alpha_1 = 0.5$ ,  $\beta_1 = 0.1$ . The detailed description of the GARCH model is given in section 2.3.2. Figure 5 displays a single example of the concatenated time series. To reduce sampling error, I generate 500 GARCH series from each

specified model, and 500 concatenated series, and take an average of the values of PE, PD and ACF on the GARCH model simulations and concatenated simulations. All generated series are of length 38160, which is the the same length as the real-world financial data I investigate later in the empirical studies. I plot the mean of the  $1 - PE_{\tau}^{D=3}$ ,  $PD_{\tau}^{D=3}$  and ACF of the individual models and the concatenated series to visually display the impact of non-stationarity on each temporal dependence measure. The comparison between the  $1 - PE_{\tau}^{D=3}$ ,  $PD_{\tau}^{D=3}$ , ACF value on concatenated series and the average  $1 - PE_{\tau}^{D=3}$ ,  $PD_{\tau}^{D=3}$ , ACF on the respective individual stationary GARCH series is provided in Figure 6.

From the plots, it is clear that PE is the least affected by non-stationarity among the three. Non-stationarity only deviates PE at delay 1 slightly from the “correct” level. Other than that, the PE plot of the concatenated series accurately reflects the average temporal dependence structure of the four distinct GARCH models and successfully indicates the lack of dependence structure after delay 10. This result coincides with the earlier finding in [Kreuzer et al. \(2014\)](#) that PE requires minimal stationarity of the underlying process. Sample ACF and PD on the other hand are severely affected and distorted by non-stationarity. In the presence of non-stationarity, ACF tends to display “spurious long-memory”. “Spurious long-memory” refers to the phenomenon in which non-stationarity causes the ACF in the short-memory times series to decay much more slowly than the actual temporal dependence diminish rate, and remains positive and significant for larger lags than it should. A similar impact from non-stationarity also can be found in the PD measure. Non-stationary time series tend to bring up the value of PD and lead to a spurious significant value of PD at very large delays for dynamics that do not exhibit long-term dependence structures. Even after I adjust the PD plot so that the value of PD at very large delays fluctuates around an insignificant level, the value of PD at short-term delay is still significantly greater than the correct level.

In summary, PE is a much more reliable measure when analyzing a non-stationary time series when compared to ACF and PD. In the presence of non-stationarity, the ACF and PD plots can lead to inaccurate inferences. Therefore additional caution needs to be paid when drawing inferences based on them, especially when the value of ACF or PD remains positive and significant for very large lags/delays.

### 3.7 PE, PD and ordinal pattern distribution in ARMA and GARCH models

In this section, I investigate the behaviour of the PE and PD measures on simulations from ARMA and GARCH models with different parametrizations to exam how PE and PD respond

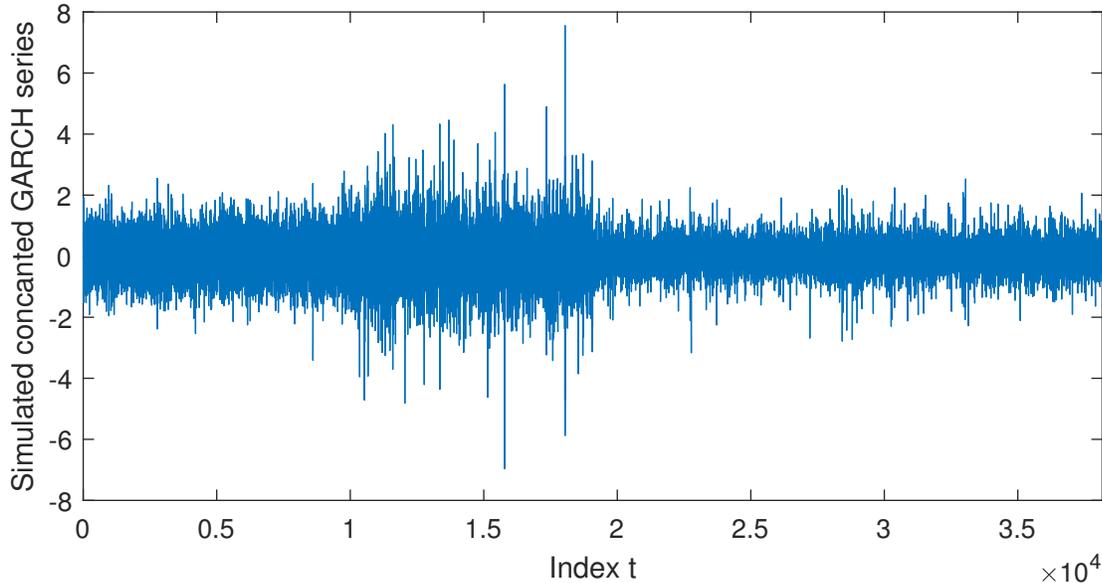


Figure 5: Plot of a single path of concatenated non-stationary time series

to a number of key characteristics in a dynamical process. By changing the parameters in the specified models, I test how PE and PD are affected by the strength of the postulated serial dependence relations, the noise to signal ratio of the process and the dynamical structures and distributions of the innovations. The ARMA and GARCH models are the most used models for practitioners in the field of financial time series analysis. The link between PE and PD measures and the parametrization of the most familiar models is important in building bridges between the relatively novel measures and the mainstream finance literature.

Before we move to the simulation studies, I make an assumption for the general form of time series dynamics. I assume any stationary time series  $\{x_t : t = 1, 2, \dots, N\}$  can be represented or be transformed into the form

$$x_t = c + g(x_{t-1}, x_{t-2}, \dots, x_{t-\tau_T}) + \varepsilon_t \quad (16)$$

where  $\tau_T$  is the furthest lag in which past entries affect the current entries,  $g(\cdot)$  is the deterministic function that connects past observations to the expected value of the current entry, and  $\varepsilon_t$  is the innovation term with  $E(\varepsilon_t) = 0$  and  $E(\varepsilon_t | F_{t-1}) = 0$ . Both the ARMA model and GARCH model, and many other financial time series models, can be written in this form. The differences between various dynamics and models reside in the constant term  $c$ , the form of  $g(\cdot)$  function, the distribution of innovation  $\varepsilon_t$  and the dependence structures within  $\varepsilon_t$  (if exists).

The ARMA model (see equation (3)) assumes linear deterministic relations  $g(\cdot)$  and iid

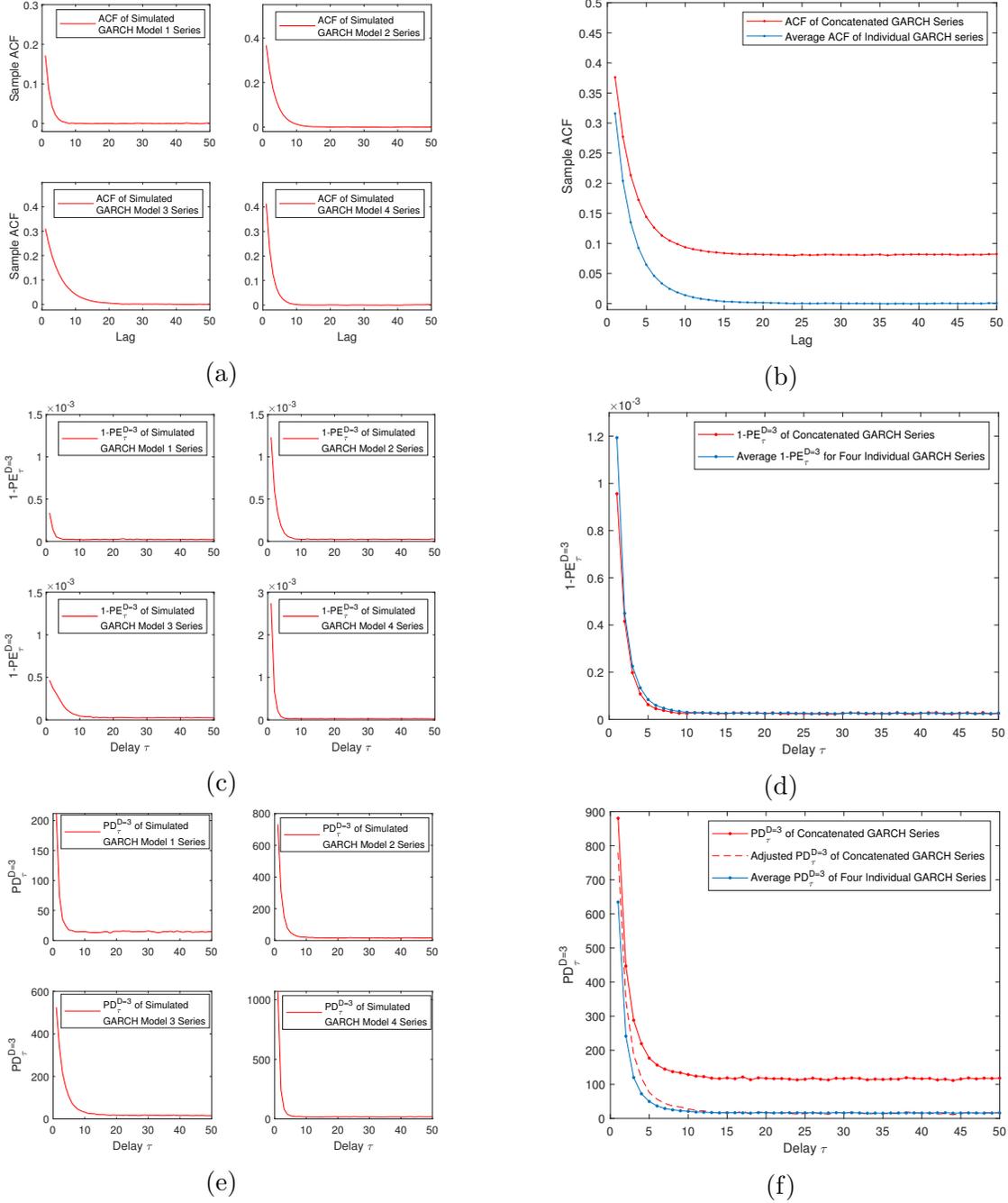


Figure 6: (a) Plot of ACF of four individual GARCH model simulated series. (b) Plot of ACF of the concatenated GARCH Series. (c) Plot of  $1 - PE_{\tau}^{D=3}$  of four individual GARCH model simulated series. (d) Plot of  $1 - PE_{\tau}^{D=3}$  of the concatenated GARCH Series. (e) Plot of  $PD_{\tau}^{D=3}$  of four individual GARCH model simulated series. (f) Plot of  $PD_{\tau}^{D=3}$  of the concatenated GARCH Series.

innovations with Gaussian or Student's  $t$  distribution, whereas in the GARCH model, past observations still affect the current entry  $x_t$  according to a linear governing function  $g(\cdot)$ , but innovations  $\varepsilon_t$  are right skewed distributed and are dependent on each other. More specifically, the GARCH model, as illustrated in section 2.3.2, assumes positive linear serial relations between current and past squared returns (see equation (7)) and the innovation term is dependent in such a way that large lagged innovations tend to lead to large dispersion of the distribution of future innovation and vice versa. In the most complicated cases, a time series dynamic can have a non-linear deterministic relation and dynamical innovations.

The simulation study includes 500 time series generated from 12 different ARMA(1,1) models and 12 different GARCH(1,1) models with various parameter specifications. I specifically choose the parameters of simulated models to cover different level of linear serial dependence, and different dependence decaying rate while satisfying the stationary constraints. All the ARMA and GARCH simulations have the same constant parameter  $\theta_0 = 0.2$ . All ARMA simulations have the innovation variance  $\sigma_\varepsilon^2 = 0.01$ . The length of the simulated series is 38160 which is the length of the real-world empirical data that I analyze in chapter 5 so that the results in the simulation can be used as inference and for comparison afterwards. The plots of average values of  $1 - \text{PE}_\tau^{D=3}$ ,  $\text{PD}_\tau^{D=3}$  and the sample ACF of the 500 simulated series of length  $N = 38160$  and the corresponding ordinal pattern probabilities at delay 1 of the 12 different parametrization of ARMA models and GARCH models are given in Figures 7 and 8. In addition to the standard ARMA(1,1) and GARCH(1,1), I also generate simulations from another 12 ARMA(1,1)/ GARCH(1,1) models with Student's  $t$  distribution of various degrees of freedom to demonstrate how heavy tail distributed innovations affect the value of PE and PD. The simulation studies reveal how the constant  $c$ , the strength of deterministic relation  $g(\cdot)$ , the distribution of innovations and the structures within innovations (if exists) affect the ordinal pattern distributions, thus the value of PE and PD.

Obviously, the value of constant parameter  $\phi_0$  in the ARMA model or  $\alpha_0$  in the GARCH model do not affect the value of PE and PD since they only affect the unconditional mean of the simulated process. PE and PD are based on ordinal patterns and therefore only consider the relative magnitude of the entries and are invariant to monotonic transformation. As for the innovation variance  $\sigma_\varepsilon^2$  of the ARMA processes, even though it does not alter the expected value of PE and PD for the simulated series, they are positively related to the variance of PE and PD. It is worth noting that the fact that the expected value of measure PE and PD are invariant to the variance of innovations in the ARMA and GARCH model simulations further supports one of the most desired property of PE. Since processes with the same specifications except the different noise to signal ratios are expected to generate the

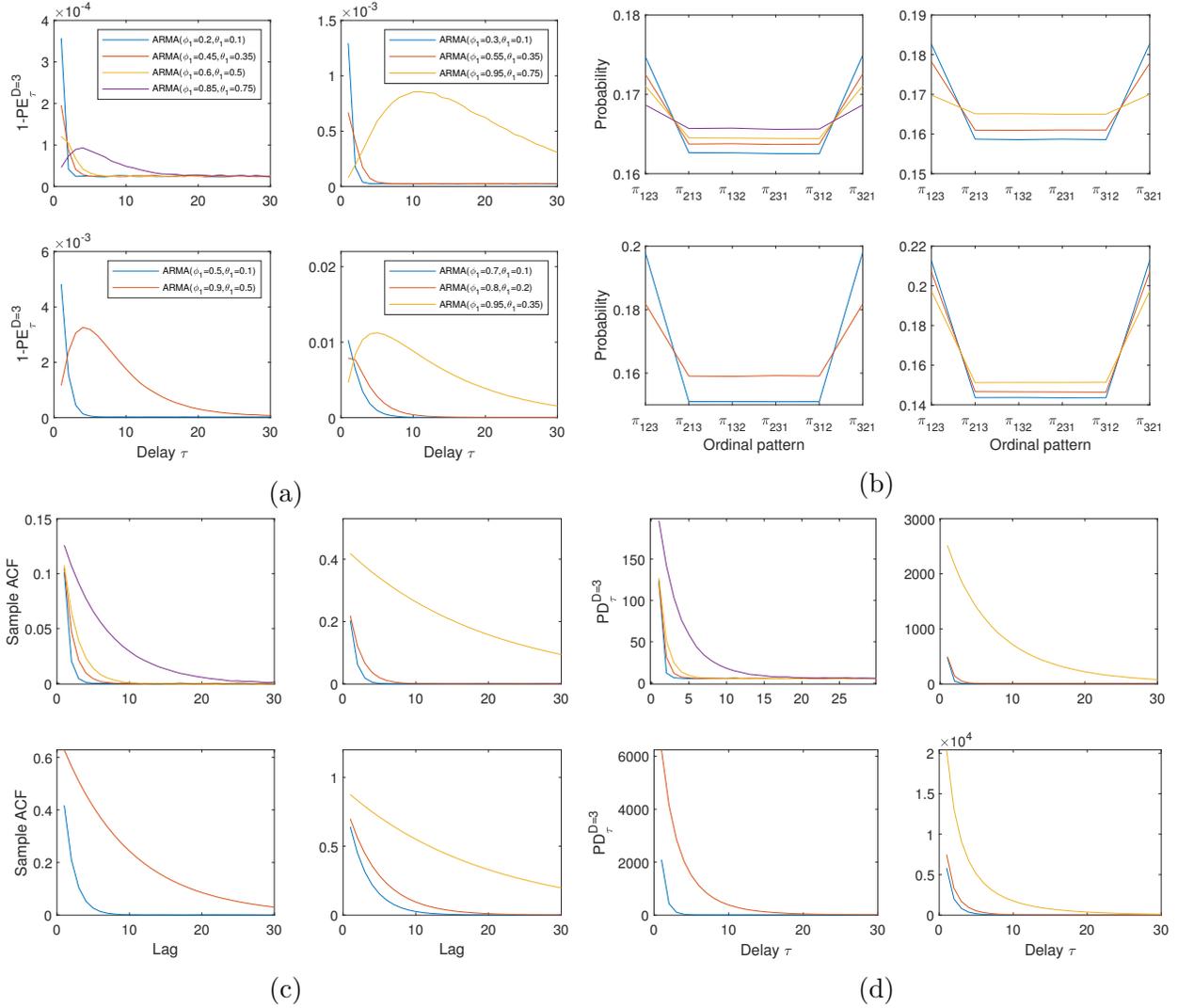


Figure 7: (a) Plot of average of  $1 - \text{PE}_\tau^{D=3}$  as function of delay on 500 simulated series from 12 different parameter combinations of ARMA model. (b) Plot of the average of each ordinal pattern probability on each simulated ARMA model contributing to  $1 - \text{PE}_\tau^{D=3}$  at delay  $\tau = 1$ . (c) Plot of average of ACF on each simulated ARMA model. (d) Plot of average of  $\text{PD}_\tau^{D=3}$  on each simulated ARMA model.

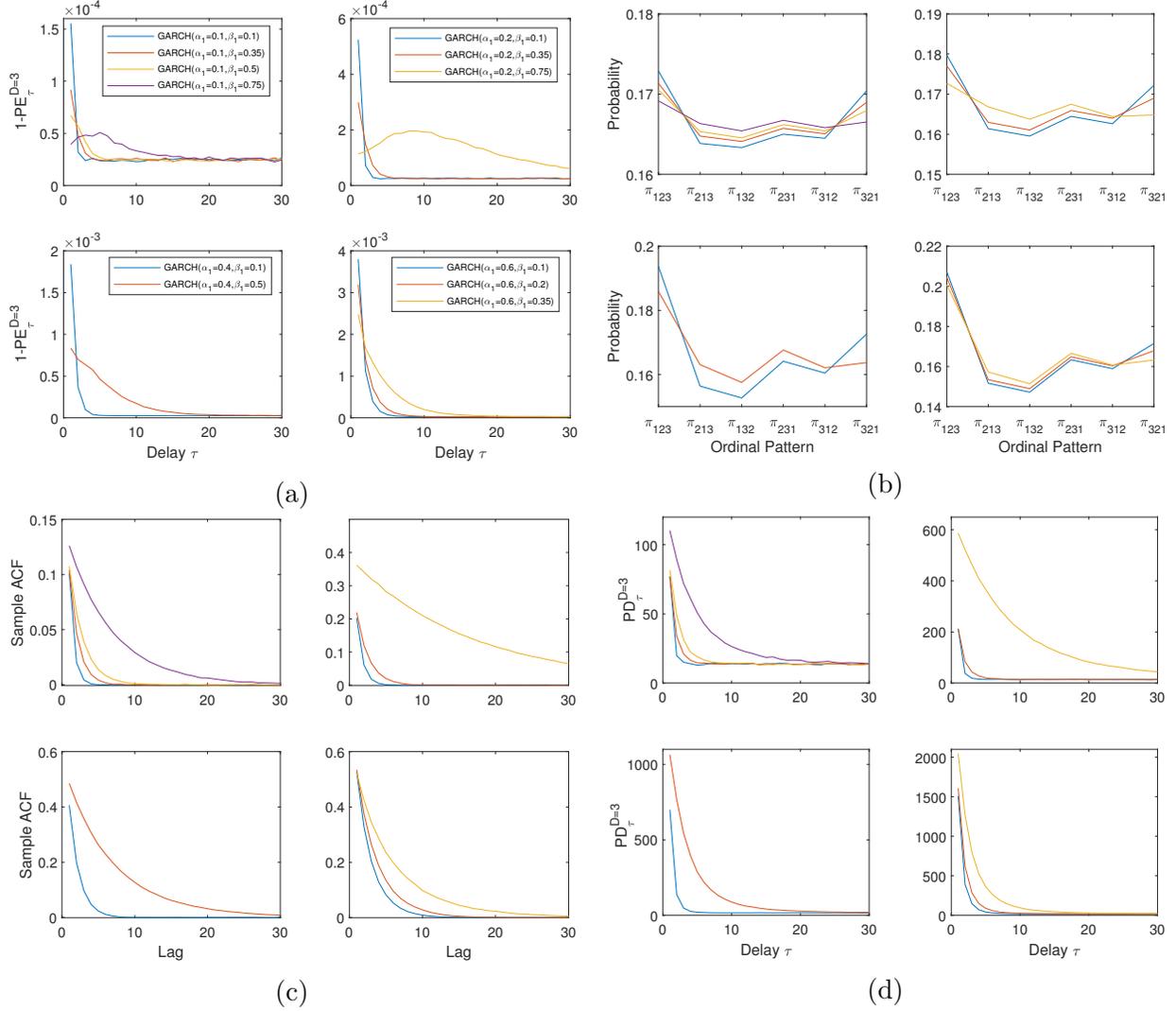


Figure 8: (a) Plot of average of  $1 - PE_{\tau}^{D=3}$  as a function of delay on 500 simulated series from 12 different parameter combinations of GARCH model. (b) Plot of the average of each ordinal pattern probability on each simulated GARCH model contributing to  $1 - PE_{\tau}^{D=3}$  at delay  $\tau = 1$ . (c) Plot of average sample ACF on each simulated GARCH model. (d) Plot of average of  $PD_{\tau}^{D=3}$  on each simulated GARCH model.

same value of PE and PD, the PE and PD measures are robust to volatile noise.

Additionally, the ARMA simulations with Gaussian innovations are, in fact, an empirical replication and verification of the results described in section 3.3. In section 3.3 I illustrated the link between PE with ACF in a Gaussian process. Since the ARMA model with Gaussian innovation is a special case of Gaussian process, the conclusions from section 3.3 are also applicable here. Consistent with the demonstration in section 3.3, all generated ARMA time series display similar ordinal pattern probability ranks of

$$p_\tau(\pi_{123}) = p_\tau(\pi_{321}),$$

and

$$p_\tau(\pi_{213}) = p_\tau(\pi_{132}) = p_\tau(\pi_{231}) = p_\tau(\pi_{312}).$$

This is a common feature throughout all ARMA simulations with Gaussian innovations regardless of the chosen parametrizations. As explained previously, the equal occurrence of all monotonic ordinal patterns and equal probabilities of all non-monotonic ordinal patterns reflect two fundamental characteristics of the ARMA model: the symmetrical probability of positive and negative increments and the linear deterministic relation that ensures the impartial response to positive and negative increments. Due to the existence of these universal features in the ordinal pattern distribution, they can be used as an indicator of the appropriateness of ARMA models to model observed time series even without the need to actually fit them. Additionally, section 3.3 shows that the value of  $1 - \text{PE}_\tau^{D=3}$  is determined by ACF at  $\tau$  and at  $2\tau$ . Moreover, ACFs at  $\tau$  and at  $2\tau$  are offsetting components in determining the value of PE. This explains why, while the ARMA model has strictly diminishing temporal dependence structures for increasing delays, the plot of  $1 - \text{PE}_\tau^{D=3}$  on simulations from ARMA dynamics sometimes exhibit a bell shape as delay increases. This is because the value of  $1 - \text{PE}_\tau^{D=3}$  is maximized at the delays  $\tau$  where the corresponding autocorrelation  $\rho_\tau$  and the structure diminishing rate reflected by the ratio between  $\rho_{2\tau}$  and  $\rho_\tau$  is maximized collectively. On the contrary, measure  $\text{PD}_\tau^{D=3}$  is strictly decreasing as a function of delay just like ACF since statistic  $\text{PD}_\tau^{D=3}$  is specially designed to eliminate the influence from the temporal dependence structures other than the selected delay. Moreover, by comparing Figure 7(c) and 7(d),  $\text{PD}_\tau^{D=3}$  is shown to closely resemble the sample ACF in all ARMA simulations. It successfully distinguishes various levels of serial dependence structure and diminishing rates specified in the simulated processes.

For GARCH simulations, the GARCH model can be regarded as an ARMA model with right-skewed and dynamical innovations if the target objectives are squared returns. Section

3.5 shows that the value of  $1 - \text{PE}_\tau^{D=3}$  and  $\text{PD}_\tau^{D=3}$  in GARCH models are significantly smaller than that in ARMA models with the same ACFs. The discrepancy in the value of PE and PD between ARMA and GARCH processes with the exact same theoretical linear serial dependence relation suggest that beside the deterministic function  $g(\cdot)$  in equation (16), the distribution and structures within innovation can affect the value of PD and PE. The distinct reduction of  $\text{PD}_\tau^{D=3}$  and  $1 - \text{PE}_\tau^{D=3}$  in the GARCH process compared to that in the ARMA process is sensible as it implies that the temporal dependence structure in the innovations reduces the level of predictability and the strength of the deterministic structures in the underlying dynamics compared to the original simple process with independent innovations. Another observation of the GARCH simulation studies is that for all simulated GARCH models, the ordinal pattern distributions are strikingly similar regardless of the chosen parameters. The most prominent feature in the ordinal pattern in GARCH simulations is the extra likelihood of the monotonic increasing patterns over the monotonic decreasing patterns. In the case of pattern length 3, the ordinal pattern  $\pi_{123}$  is more likely to occur than the ordinal pattern  $\pi_{321}$  as opposed to the equal likelihood of  $\pi_{123}$  and  $\pi_{321}$  in the ARMA model. The discrepancy between the monotonic increasing pattern and the monotonic decreasing pattern in  $X_t : t = 1, \dots, N$  where  $X_t$  represents the squared returns in GARCH models, is caused by the unconditional right-skewed distribution of the increments in  $X_t$ , that is,

$$\tilde{\mu}_3(X_t - X_{t-1}) > 0,$$

and asymmetry responds to decrements and increments. Proving all GARCH(1,1) models have right-skewed distribution of  $\nabla X_t = X_t - X_{t-1}$  is straightforward. Due to the additional likelihood of positive increments over negative increments, that is,

$$p(\nabla X_t > 0) > p(\nabla X_t < 0)$$

the probability of ordinal pattern  $\pi_{12}$  is more likely to occur than ordinal pattern  $\pi_{21}$ . Additionally, an increment in  $X_t$  would lead to increased dispersion in the distribution of future squared return  $X_{t+1}$ . In contrast, a decrease in  $X_t$  would cause a reduction of variance in  $X_{t+1}|F_t$  where  $F_t$  denotes all available information at time  $t$ , thus resulting in a more concentrated distribution of  $X_{t+1}|F_t$ . Therefore, an increment in the historical squared returns is more like to lead to an increase in future volatility than vice versa, and hence, altogether resulting in the extra likelihood of ordinal pattern  $\pi_{123}$  over  $\pi_{321}$ .

Other than the deterministic function  $g(\cdot)$  in the generalized form of time series dynamics given in equation (16), the symmetry/asymmetry, the dependence structures and the kurtosis

of the innovations  $\varepsilon_t$  can also alter the probability distribution of the ordinal patterns, and consequently the value of PE and PD. By comparing the value of  $1 - \text{PE}_\tau^{D=3}$ , the ordinal pattern probabilities, the value of ACF and  $\text{PD}_\tau^{D=3}$  on ARMA model and GARCH model simulations with Student's  $t$  innovation for various level of degree of freedom, we can track the way PE and PD respond to increasingly fatter tails in the innovations of a process. The simulation study suggests that the kurtosis of innovations can affect all the above mentioned measures, but the extent and the form of its influence varies depending on the strength and nature of the temporal dependence structures exhibited in the time series dynamics. The simulation results summarized in plot 9 suggest that in ARMA models, increasing kurtosis in the innovation does not alter the value of ACF, but can cause increments in  $1 - \text{PE}_\tau^{D=3}$  and  $\text{PD}_\tau^{D=3}$ . In the GARCH model, as indicated in Figure 10, increasing kurtosis in the innovation reduces the value of  $1 - \text{PE}_\tau^{D=3}$ ,  $\text{PD}_\tau^{D=3}$  and ACF.

From the simulation studies, I find that the distribution of ordinal pattern required in the computation of PE, as well as the plot of  $1 - \text{PE}_\tau^D$  and  $\text{PD}_\tau^D$  against increasing delays, has the ability to reflect many important features of the underlying dynamics of a given time series, such as the functional form of the governing deterministic relation, the asymmetry, kurtosis and the existence of dynamic structures present in the innovations. Moreover, systems with similar characteristics shares similar features in their ordinal pattern probability ranks regardless of the parametrizations. For example, time series following a linear process with Gaussian innovations are expected to have equal probabilities of monotonic and non-monotonic ordinal patterns, whereas GARCH class models with Gaussian innovations are expected to exhibit an extra likelihood of a monotonic increasing ordinal pattern over a monotonic decreasing ordinal pattern. The invariant features in similar time series dynamics aid in distinguishing different processes. Additionally, by comparing the ordinal pattern distributions, PE and PD plot of the observed time series and that from fitted models, one might be able to identify the reasons behind the insufficiency of the selected model. For instance, if we fit an observed time series to an ARMA or GARCH model, and observe significant discrepancies between the 1-PE and PD in the observed time series and that in simulations from the fitted ARMA or GARCH models. A number of possible sources might contribute to it. The fitted model might misspecified the deterministic relation  $g(\cdot)$ , neglect/overestimate of the dependence in the innovation dynamics or overlook/overestimate the heavy tail in the innovation distributions. The ordinal pattern distribution can help identify which of the above reasons is a more likely contributor, and the excess/insufficient value of PE or PD aids in identifying whether the potential misspecified factor is underestimated or overstated. To sum up, the value of PE and PD, and the ordinal pattern probability rank, can aid in identifying the most suitable parametric models to be applied to a time series, and point out any possible inadequacy in

the selected model’s specification.

### 3.8 PD visualization plot to reveal the deterministic relation captured by the forecasting model

With the aid of the newly proposed statistic PD, I develop a new approach to visually show the deterministic relation captured by the estimated model. The new visual representation is critical in model interoperation and providing explanations as to why certain models outperform others, and suggests of promising improvements.

All forecasting models can be written in a generalized form:

$$\hat{x}_t = \hat{c} + \hat{g}(x_{t-1}, x_{t-2}, \dots, x_{t-\tau_T})$$

where  $\hat{c}$  is a constant, and  $\hat{g}(\cdot)$  is the deterministic function captured by the employed model in its effort to connect past observations to future unseen entries.

Different models assume different functional forms for  $\hat{g}(\cdot)$  in their effort to fulfil a prediction task. Without a universally applicable standard, it is hard to compare various models in terms of their respective estimated deterministic relations, especially when comparing parametric models with defined functions and non-parametric models with no closed-form formula. Conventionally, when evaluating and comparing the performance of a number of candidate forecasting models, practitioners establish which one produces better forecasting accuracy, but not understand why.

In order to construct a clear summary that extracts the essence of every estimated deterministic function  $\hat{g}(\cdot)$  postulated by each model, I employ the previously proposed PD measure to reflect the importance or the level of relevance of a lagged observation  $x_{t-\tau}$  in the postulate function  $\hat{g}(\cdot)$ .

Specifically, in order to visually show the deterministic relation captured by any prediction model, I first generate simulated time series  $X = \{x_t\}$  following the relation

$$x_t = c + \hat{g}(x_{t-1}, x_{t-2}, \dots, x_{t-\tau_T}) + \varepsilon_t$$

where  $\hat{g}(\cdot)$  is the estimated function postulated by the considered model, and  $\{\varepsilon_t\}$  is a series of randomly generated innovations that follows normal distribution  $\varepsilon_t \sim \mathcal{N}(0, \delta_\varepsilon)$ . Due to the invariance to monotonic transformation property of the ordinal pattern measure, the

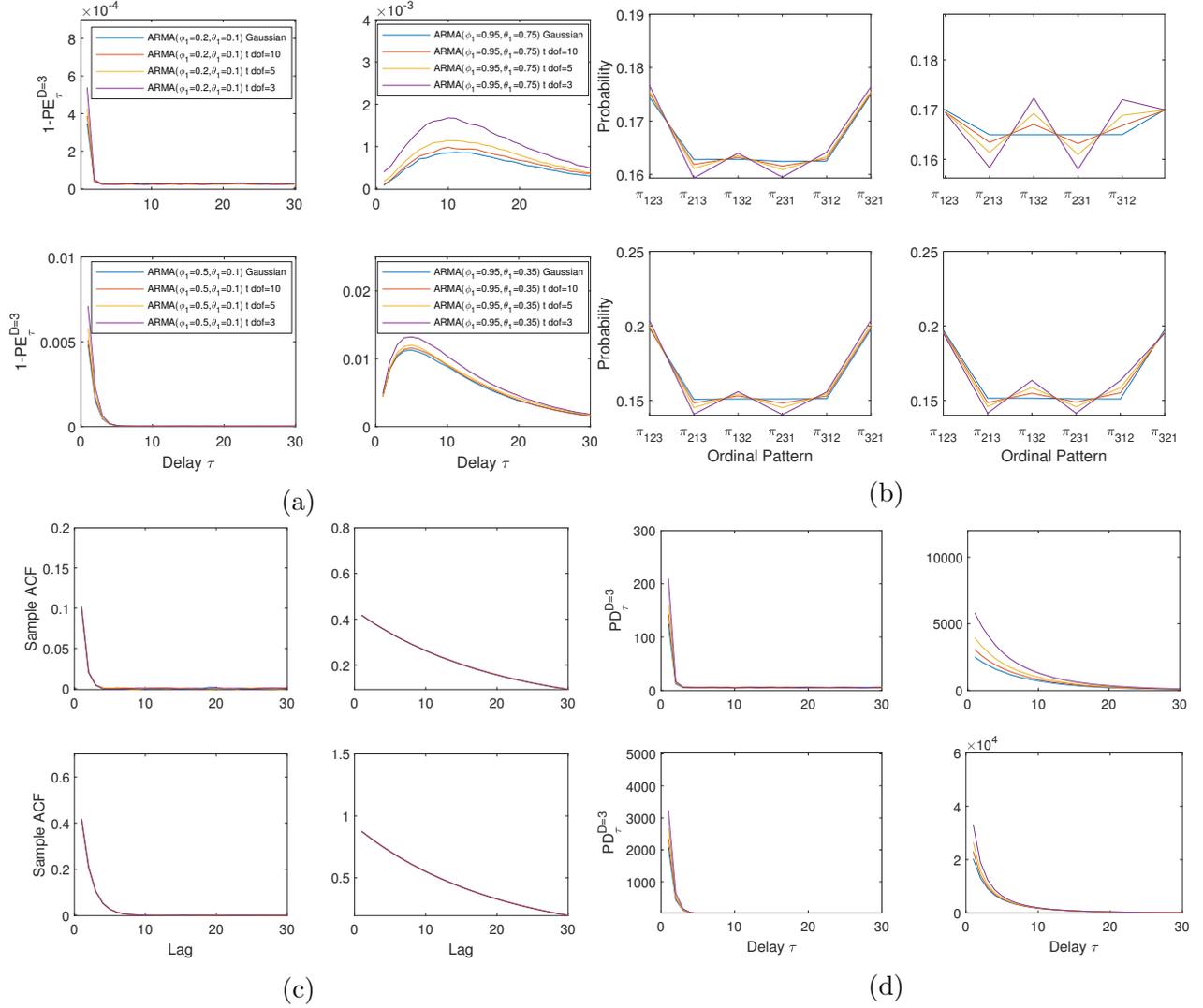


Figure 9: (a) Plot of average  $1 - PE_\tau^{D=3}$  as a function of delay on 500 simulated series of length  $N = 38160$  from 16 ARMA models for various level of kurtosis in the innovation. (b) Plot of the corresponding ordinal pattern probability of simulated series contributing to  $1 - PE_\tau^{D=3}$  at delay  $\tau = 1$ . (c) Plot of average sample ACF as a function of lag on each simulated models. (d) Plot of average  $PD_\tau^{D=3}$  as a function of delay on each simulated models.

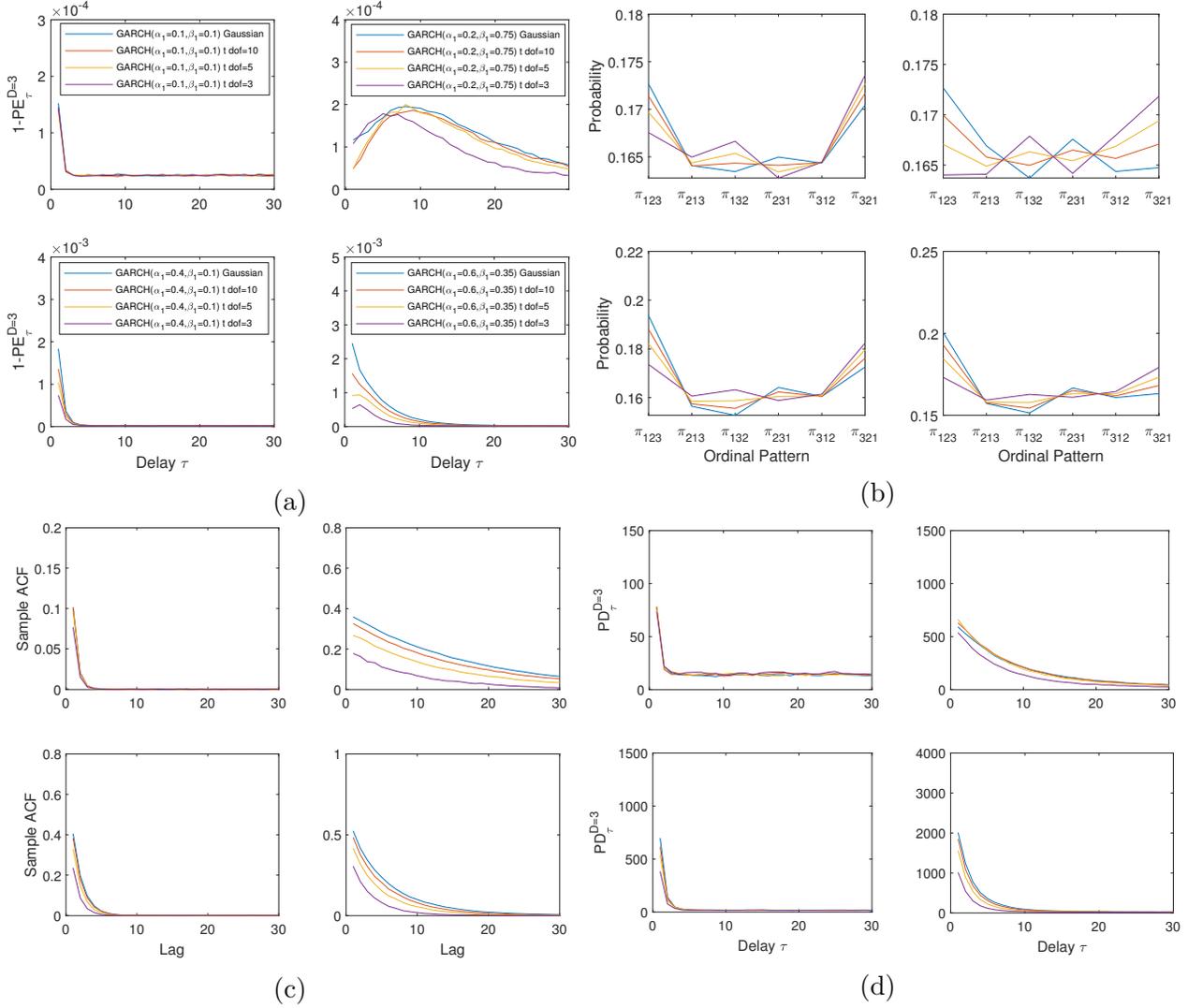


Figure 10: (a) Plot of average  $1 - PE_{\tau}^{D=3}$  as a function of delay on 500 simulated series of length  $N = 38160$  from 16 GARCH models for various level of kurtosis in the innovation. (b) Plot of the corresponding ordinal pattern probability of simulated series contributing to  $1 - PE_{\tau}^{D=3}$  at delay  $\tau = 1$ . (c) Plot of average sample ACF as a function of lag on each simulated models. (d) Plot of average  $PD_{\tau}^{D=3}$  as a function of delay on each simulated models.

variance of the innovations  $\delta_\varepsilon$  and the constant  $c$  do not affect the expected value of the statistic PD, so choosing them relies on personal judgment. However, in many non-linear functions, the “weight” of  $x_{t-\tau}$  in  $\hat{g}(\cdot)$  varies depending on the domain of  $x_{t-\tau}$ . Therefore, we can intentionally select  $c$  and  $\delta_\varepsilon$  to match the domain of the simulated series to that of the investigated series. In order to reduce the simulation error, I generate 100 paths of simulations  $X$  following the same procedures, and plot the average value of  $\text{PD}_\tau^D(X)$  against increasing delays to reflect the “weight” of lagged observations in the postulated function  $\hat{g}(\cdot)$ .

The empirical applications of the PD visualization plot can be found in section 4.4 and section 5.6.2. It is used in the prediction model comparison studies to explain why some models produce better prediction accuracy than others.

### 3.9 Multivariate PE

The concept of PE can also be extended to measure the bivariate dependence between two distinct time series.

Given two equal length time series  $X = \{x_t; t = 1, \dots, N\}$  and  $Y = \{y_t; t = 1, \dots, N\}$ , for a chosen pattern length  $D$  and delay  $\tau$ , we construct  $N - (D - 1)\tau$  segment pairs  $(x_{t,\tau}^D, y_{t,\tau}^D)'$ :

$$\begin{pmatrix} x_{t,\tau}^D \\ y_{t,\tau}^D \end{pmatrix} = \begin{pmatrix} x_t, x_{t+\tau}, \dots, x_{t+(D-1)\tau} \\ y_t, y_{t+\tau}, \dots, y_{t+(D-1)\tau} \end{pmatrix} : \quad t = 1, 2, \dots, N - (D - 1)\tau.$$

Vector  $x_{t,\tau}^D$  and  $y_{t,\tau}^D$  are mapped into one of the  $D!$  ordinal patterns  $\{\pi_i; i = 1, \dots, D!\}$ , then each segment pair is categorized into one of the  $(D!)^2$  joint ordinal patterns  $\{\pi_{i,j}; i = 1, \dots, D!; j = 1, \dots, D!\}$ . For  $D = 3$ , the joint ordinal patterns are defined as follow:

Ordinal Pattern	$(a_x, b_x, c_x; a_y, b_y, c_y)$	Condition
$\pi_{11}$	$(3, 2, 1; 3, 2, 1)$	If $x_t > x_{t+\tau} > x_{t+2\tau}$ and $y_t > y_{t+\tau} > y_{t+2\tau}$ ,
$\pi_{21}$	$(3, 1, 2; 3, 2, 1)$	if $x_t > x_{t+2\tau} > x_{t+\tau}$ and $y_t > y_{t+\tau} > y_{t+2\tau}$ ,
$\pi_{31}$	$(2, 3, 1; 3, 2, 1)$	if $x_{t+\tau} > x_t > x_{t+2\tau}$ and $y_t > y_{t+\tau} > y_{t+2\tau}$ ,
$\pi_{41}$	$(2, 1, 3; 3, 2, 1)$	if $x_{t+2\tau} > x_t > x_{t+\tau}$ and $y_t > y_{t+\tau} > y_{t+2\tau}$ ,
$\pi_{51}$	$(1, 2, 3; 3, 2, 1)$	if $x_{t+2\tau} > x_{t+\tau} > x_t$ and $y_t > y_{t+\tau} > y_{t+2\tau}$ ,
$\pi_{61}$	$(1, 3, 2; 3, 2, 1)$	if $x_{t+\tau} > x_{t+2\tau} > x_t$ and $y_t > y_{t+\tau} > y_{t+2\tau}$ ,
$\vdots$	$\vdots$	$\vdots$

The probability of the joint ordinal patterns  $\pi_{i,j}$  is then estimated by the number of segment

pairs  $(x_{t,\tau}^D, y_{t,\tau}^D)'$  that are mapped to  $\pi_{i,j}$ , divided by the total number of constructed pairs:

$$p_\tau^D(\pi_{i,j}) = \frac{\#\left\{(x_{t,\tau}^D, y_{t,\tau}^D)' \mid (x_{t,\tau}^D, y_{t,\tau}^D)' \text{ has joint ordinal pattern } \pi_{i,j}\right\}}{N - (D-1)\tau}, i = 1, \dots, D!; j = 1, \dots, D!.$$

Hence, the **Bivariate Permutation Entropy (BIPE)** of time series  $X$  and  $Y$ , with chosen segment length  $D$  and delay  $\tau$  is defined by

$$\text{BIPE}_\tau^D(X, Y) = \sum_{i=1}^{D!} \sum_{j=1}^{D!} \frac{-p_\tau^D(\pi_{i,j}) \log p_\tau^D(\pi_{i,j})}{\log(D!)^2}.$$

If time series  $X$  and  $Y$  are independent,

$$\text{BIPE}_\tau^D(X, Y) = \frac{1}{2} [\text{PE}_\tau^D(X) + \text{PE}_\tau^D(Y)].$$

Therefore the dependence between time series  $X$  and  $Y$  is measured by the departure from independence, and is defined to be:

$$h_\tau^D(X, Y) = [\text{PE}_\tau^D(X) + \text{PE}_\tau^D(Y)] - 2\text{BIPE}_\tau^D(X, Y).$$

In light of [Joe's \(1989\)](#) work, I standardized the above dependence measure to be:

$$H_\tau^D(X, Y) = \frac{[\text{PE}_\tau^D(X) + \text{PE}_\tau^D(Y)] - 2\text{BIPE}_\tau^D(X, Y)}{\min\{\text{PE}_\tau^D(X), \text{PE}_\tau^D(Y)\}},$$

so that  $0 \leq H_\tau^D(X, Y) \leq 1$ , with  $H_\tau^D(X, Y) = 0$  iff time series  $X$  and  $Y$  are independent, and  $H_\tau^D(X, Y) = 1$  iff  $y_t$  is a fixed function of  $x_t$  for all  $t$ .

The measure  $H_\tau^D(X, Y)$  quantifies the bivariate dependence between time series  $X$  and  $Y$ . From its specification, we can see that the value of  $H_\tau^D(X, Y)$  is free of the influence of the temporal dependence structure of each of the series involved. No matter by how much the ordinal pattern frequencies in any of  $X$  and  $Y$  deviate from the even distribution, only the co-movement of the ordinal patterns from both  $X$  and  $Y$  contributes to the value of  $H_\tau^D(X, Y)$ . Additionally, due to the use of ordinal pattern, the measure is also invariant to the marginal distributions of the time series involved.

The Permutation Distribution Matrix containing the probability of all joint ordinal patterns

is defined below:

$$P_{\tau}^D(X, Y) = \left( p_{\tau}^D(\pi_{1,1}) \quad p_{\tau}^D(\pi_{1,2}) \quad \cdots \quad p_{\tau}^D(\pi_{D!,D!}) \right)^{\top}.$$

Since high joint probability  $p_{\tau}^D(\pi_{ij})$  does not necessarily indicate strong dependence between pattern  $i$  in  $X$  and pattern  $j$  in  $Y$ , it might just be a result of high marginal probability  $p_{\tau}^D(\pi_i)$  or  $p_{\tau}^D(\pi_j)$ . To eliminate the effect of marginal probability, I define a quantity  $d_{\tau}^D(\pi_{ij})$  as the ratio of the joint probability of any joint ordinal pattern  $\pi_{ij}$  to the product of their respective marginal probabilities, in order to reflect the relation between pattern  $i$  in time series  $X$  and pattern  $j$  in time series  $Y$ , that is,

$$d_{\tau}^D(\pi_{ij}) = \frac{p_{\tau}^D(\pi_{ij})}{p_{\tau}^D(\pi_i) p_{\tau}^D(\pi_j)}, i = 1, \dots, D!, j = 1, \dots, D!.$$

If the pattern  $i$  in  $X$  and pattern  $j$  in  $Y$  are independent,  $d(\pi_{ij}) = 1$ ; if the pattern  $i$  in  $X$  and pattern  $j$  in  $Y$  are negatively correlated,  $d(\pi_{ij}) < 1$ ; if the pattern  $i$  in  $X$  and pattern  $j$  in  $Y$  are positively correlated,  $d(\pi_{ij}) > 1$ .

## 4 The PD model sufficiency test

In the previous chapter, I proposed a new statistic PD to measure the temporal dependence relations in a given univariate time series. In this chapter, I use PD and an extension of PD in a bivariate time series to construct a new prediction sufficiency test, inheriting the ordinal pattern concept. I name it the PD model sufficiency test. Compared to conventional approaches, the PD model sufficiency test evaluates the sufficiency of a model/predictor in terms of its ability to exploit the point prediction potential in the investigated time series. Additionally, the newly proposed test overcomes the limitation of the existing model sufficiency test which generally requires iid innovations of the investigated series.

### 4.1 Conventional model comparison and evaluation approaches

Model selection and evaluation is a necessary step in empirical time series analysis. After fitting and constructing forecasts using a number of competing models/approaches, practitioners need to find out how well the models they employ capture and replicate the dynamics underlying the observed data and how closely the constructed forecast predicts future movements. A lot of criteria have been proposed to compare and evaluate forecasting models. I roughly categorize them into four groups.

The first and most used approaches are based on measures that reflect the departure of the model fitting/estimated values from the realization with or without an additional term for parsimonious control. Typical examples include goodness of fit measures R-square, AIC, BIC, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the Theil Inequality Coefficient. Given an observed time series  $\{y_i : i = 1, 2, \dots, n\}$ , and  $\hat{y}_i$  denotes the forecast/fitted value, these measures are defined as follows:

$$\begin{aligned} \text{MSE} &= \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \\ \text{RMSE} &= \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}, \\ \text{MAE} &= \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \\ \text{MAPE} &= \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \\ R^2 &= 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \end{aligned}$$

and

$$\begin{aligned} \text{AIC} &= \log(\text{MSE}) + 2k/n, \\ \text{BIC} &= \log(\text{MSE}) + \log(n)k/n, \end{aligned}$$

where  $k$  is the number of parameters in the model, and Theil Inequality Coefficient

$$U = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2} + \sqrt{\frac{1}{n} \sum_{i=1}^n \hat{y}_i^2}}$$

The second group of approaches incorporate a loss function and test whether two or more competing models have equal loss differentials, such as the Diebold-Mariano (DM) test. For a pre-specified loss function  $g(\cdot)$ , the DM test is constructed on the null hypothesis that two forecasts have the same prediction accuracy:

$$H_0 : E(d_t) = 0 \quad \forall t$$

versus the alternative hypothesis

$$H_1 : E(d_t) \neq 0$$

where  $d_t = g(e_{1t}) - g(e_{2t})$  and  $e_{1t}, e_{2t}$  are residuals from the two competing models at index  $t$ . Under  $H_0$  and assuming the loss differential series  $\{d_t; t = 1, \dots, T\}$  is covariance stationary and short-memory, the test statistics of the DM test is

$$\frac{\bar{d}}{\sqrt{\frac{2\pi\hat{f}_d(0)}{T}}} \longrightarrow N(0, 1)$$

where

$$\hat{f}_d(0) = \frac{1}{2\pi} \sum_{k=-(T-1)}^{T-1} l\left(\frac{k}{h-1}\right) \hat{\gamma}_d(k),$$

$$\hat{\gamma}_d(k) = \frac{1}{T} \sum_{t=|k|+1}^T (d_t - \bar{d})(d_{t-|k|} - \bar{d})$$

and

$$l\left(\frac{k}{h-1}\right) = \begin{cases} 1, & \text{for } \left|\frac{k}{h-1}\right| \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

The third group of approaches is based on economic criteria that use a simple trading strategy guided by forecasting to test relative pay-offs generated by different forecasting models. For instance, a simple investment rule can be constructed based on the out-of-sample one-ahead-period return forecast. For an initial endowment of fixed monetary unit, an investor must decide whether to maintain this wealth in cash or in assets, depending on whether the return predicted for the next period exceeds a threshold given by the transaction cost. In the end, the final portfolio value will be compared between competing models.

The last group of approaches does not involve comparing the models under consideration and setting them in competition against each other. Instead, tests are conducted in the residual to search for any remaining structures that indicate the sufficiency/insufficiency of the models. The most popular test used for residual analysis is the BDS test. The BDS test, developed by [Broock et al. \(1996\)](#), is the most used test on residuals to check whether the residuals are independent, and hence the estimated predictors are sufficient. The BDS test is, in fact, a portmanteau test for temporal dependence in a series, and can be used for testing against a variety of possible deviations from independence, including linear and non-linear dependence, or chaos. The test statistic in the BDS test is derived from the correlation integral and has its origins in deterministic non-linear dynamics and chaos theory. According

to Packard et al. (1980) and Takens (1981), the method of delays can embed a scalar time series  $\{x_i : i = 1, 2, \dots, N\}$  into a  $m$ -dimensional space as follows

$$\vec{x}_i = (x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau}), \quad \vec{x}_i \in R^m$$

where  $\tau$  is the pre-chosen lag. Accordingly, Grassberger & Procaccia (1983) propose correlation integral as a measure of fractal dimension of deterministic data as it records the frequency with which temporal patterns are repeated in the data. The correlation integral at the embedding dimension  $m$  is given by

$$C(m, N, r) = \frac{2}{M(M-1)} \sum_{1 \leq i < j \leq M} \Theta(r - \|\vec{x}_i - \vec{x}_j\|), \quad r > 0$$

$$\Theta(a) = 0, \quad \text{if } a \leq 0$$

$$\Theta(a) = 1, \quad \text{if } a > 0$$

where  $N$  is the size of the data sets,  $M = N - (m - 1)\tau$  is the number of embedded points in  $m$ -dimensional space,  $r$  is the distance used for testing proximity of the data points and  $\|\cdot\|$  denotes the sup-norm. In essence,  $C(m, N, r)$  measures the fraction of the pairs of points  $\vec{x}_i, i = 1, 2, \dots, M$ , the sup-norm separation of which is not greater than  $r$ . If the data is generated by a strictly stationary stochastic process, then there exists a limit of  $C(m, N, r)$  as  $N \rightarrow \infty$  for each  $r$ , and we write the limit as  $C(m, r) = \lim_{N \rightarrow \infty} C(m, N, r)$ . Following Brock et al. (1996), the BDS statistic for  $m > 1$  is defined as

$$\text{BDS}(m, r) = \frac{\sqrt{M}}{\sigma} [C(m, r) - C^m(1, r)] \quad (17)$$

where  $M = N - (m - 1)\tau$ , the standard deviation  $\sigma$  and

$$C(m, r) = \int \int \Theta(r - \|u - v\|) dF_m(u) dF_m(v),$$

and  $F_m(\cdot)$  is the joint distribution of  $x_1$  to  $x_m$ . Under the iid hypothesis  $\text{BDS}(m, M, r)$  has a limiting standard normal distribution as  $M \rightarrow \infty$ . In order to ensure the BDS statistic is well approximated by the asymptotic distribution, Brock et al. (1991) suggest the data sets of interest need to have 500 or more observations, and the common choice of embedding dimension  $m$  is in the range of  $2 \leq m \leq 5$ . Then, the value of  $r$  is generally selected as a fraction of the standard deviation of the data sets, for instance:  $0.25\text{s.d.}(x)$ ,  $0.5\text{s.d.}(x)$ ,  $0.75\text{s.d.}(x)$ ,  $\text{s.d.}(x)$  and  $1.25\text{s.d.}(x)$ .

Every model evaluation approach has its emphasis in addition to its own pros and cons, and there is no universally accepted criteria for evaluating a model/forecast. The first group of approaches focuses on the accuracy of point prediction where interest is only in the accuracy of the prediction of conditional mean of the target object. The second group of approaches considers the entire probability density of the prediction. Depending on the selected loss function, tests such as the DM test account for the misspecification of the prediction model not only in the conditional mean but also in its higher moments. The third group examines whether and by how much the prediction generates arbitrage profit. All three categories of methods are sensible comparison criteria to differentiate model performance. However, they share a common weakness in not considering the maximum prediction potential of the investigated series. Without knowing the prediction ceiling, a model that is significantly superior to other competing models can still actually be an unsatisfactory selection as its performance can be far away from the optimal level. On the other hand, for a series with minimal predictability, a model that cannot beat or be close to a random process might be the best prediction one can construct.

The last approach, the independence test on residuals, can indicate the sufficiency of individual models. However, it suffers a major limitation in evaluating point forecasts as it presumes that a sufficient model is expected to generate independent innovations. For a series of innovations that exhibit higher-moment dependent structures, the residuals formed by the difference between the actual realizations and their oracle point predictors can violate white noise.

## 4.2 Limitations of Brock, Dechert and Scheinkman (BDS) test

As mentioned above, the BDS test is a non-parametric test which needs minimal assumptions and previous knowledge about the investigated data sets. When the BDS test is applied to model residuals, the asymptotic distribution of its statistic given in (17) is independent of estimation errors under certain sufficient conditions. Specifically, [de Lima \(1996\)](#) shows that for linear additive models or models that can be transformed into that form, the BDS test is nuisance parameter free and does not require any adjustment when applied to fitted model residuals.

Despite its popularity, the BDS test suffers a number of limitations as a model sufficiency test. First, as mentioned in the previous section, it tests the null hypothesis that the residuals are independent and identically distributed against an unspecified alternative. If the governing process of the investigated series originally has non-white innovations, the rejection of the BDS test on residuals is not necessarily affirmative evidence of an insufficient predictor,

especially in evaluating the sufficiency of non-parametric models where only point forecasting accuracy matters. Second, the BDS statistic requires the assumed model to be in a linear additive format or able to be transformed into that format. Otherwise, adjustment is needed to ensure the BDS statistics have the right size under the null hypothesis. For example, when evaluating the sufficiency of the GARCH model, if the BDS test is applied directly to the standardized residuals  $Y_t/\hat{\sigma}_t$ , earlier studies (see Brock et al. 1991) suggest the BDS statistic needs to be adjusted to have the right size. An easier approach is to apply the test on natural logarithms of squared standardized residuals  $\ln(Y_t/\hat{\sigma}_t)$  so that the logarithmic transformation casts the GARCH model into a linear additive model (Fernandes & Preumont 2012). However, as I demonstrate subsequently, the logarithm transformation may deteriorate the uncaptured structures left in standardized residuals and mislead the BDS test to incorrectly accept an insufficient predictor/model.

The following simulation provides an example of how and why the BDS test on a logarithm transformed standardized residuals fails to reject an inadequate predictor/model, especially when the investigated time series has a relatively low signal to noise ratio. I generate a time series from a GARCH(1,1) model with the parametrization  $\alpha_0 = 0.18$ ,  $\alpha_1 = 0.16$  and  $\beta_1 = 0.74$ . Figure 11 plots the generated time series and its optimal predictor computed from the governing formulas of the GARCH model. In addition, in the same graph, I plot an inadequate predictor based on the deterministic relation of a GARCH model  $\alpha_0 = 0.18$ ,  $\alpha_1 = 0.03$  and  $\beta_1 = 0.87$ . Apparently, the inadequate predictor underestimates the linear series correlations in the simulated time series under study and differs significantly from the optimal predictor. Table 2 provides the value of the BDS statistics for the logarithm of a squared standardized residual of the inadequate predictor for varying embedding dimension  $m$  from 2 to 10 and various distance  $r$ :  $0.25\text{s.d.}(x)$ ,  $0.5\text{s.d.}(x)$ ,  $0.75\text{s.d.}(x)$ ,  $\text{s.d.}(x)$  and  $1.25\text{s.d.}(x)$ . Under the iid null hypothesis, the BDS statistic has a limiting standard normal distribution. From the table, it is clear that except for  $M = 10$  and  $r = 0.25\text{s.d.}(x)$ , none of the BDS statistics are significant enough to reject the hypothesis of iid residuals. The reason that the logarithm transformation tends to mislead the BDS test is that the logarithm on residuals would convert a small value that is close to zero to a significant negative valued entry, and in the meanwhile shrink the relatively distinct gaps between the predictions and the actual realizations. Consequently, the uncaptured structure becomes less evident after the logarithm transformation. Figure 12 compares the standardized residuals versus the logarithm of the standardized residuals of a randomly selected sub-period of the target simulation series so as to display the effect of the logarithm transformation on the standardized residuals.

Due to the above limitations, the acceptance and rejection of the BDS test can lead to an

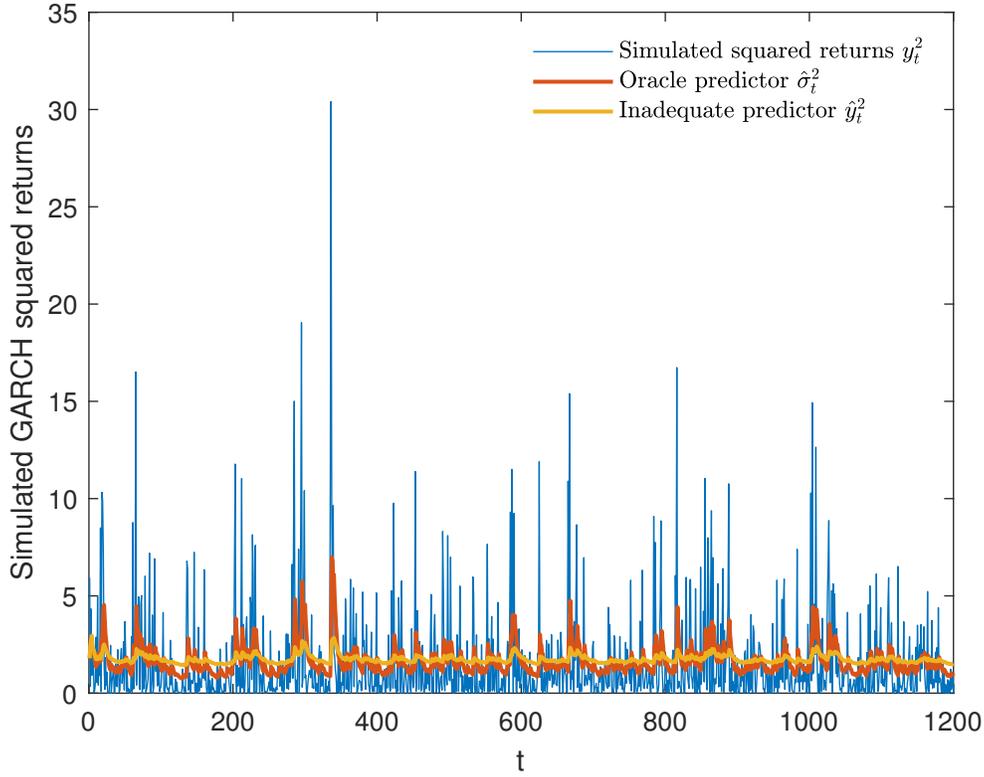


Figure 11: Plot of simulated squared return time series generated from GARCH(1,1) model with parameter  $\alpha_0 = 0.18$ ,  $\alpha_1 = 0.16$ ,  $\beta_1 = 0.74$ , its optimal predictor and an inadequate predictor based on GARCH(1,1) model with parameter  $\alpha_0 = 0.18$ ,  $\alpha_1 = 0.03$ ,  $\beta_1 = 0.87$ .

Table 2: The BDS statistics on  $\ln(y_t^2/\widehat{y}_t^2)$  of embedding dimension  $M$  with distance of  $r$ . Under null hypothesis of iid, the BDS statistic is expected to follow standard normal distribution. \* indicates the iid hypothesis is rejected at 5% significance, and \*\* indicates the iid hypothesis is rejected at 1% significance.

$M$	$r$				
	0.25s.d.( $X$ )	0.5s.d.( $X$ )	0.75s.d.( $X$ )	s.d.( $X$ )	1.25s.d.( $X$ )
2	0.43	0.24	-0.06	-0.26	-0.27
3	-0.08	-0.18	-0.54	-0.69	-0.75
4	-0.34	-0.40	-0.66	-0.88	-1.00
5	-0.24	-0.32	-0.46	-0.70	-0.84
6	0.38	-0.38	-0.31	-0.58	-0.72
7	-1.67	-0.53	-0.22	-0.52	-0.65
8	-0.46	-0.56	-0.15	-0.51	-0.64
9	-2.32*	-0.61	-0.18	-0.51	-0.69
10	-9.95**	-0.29	-0.34	-0.55	-0.73

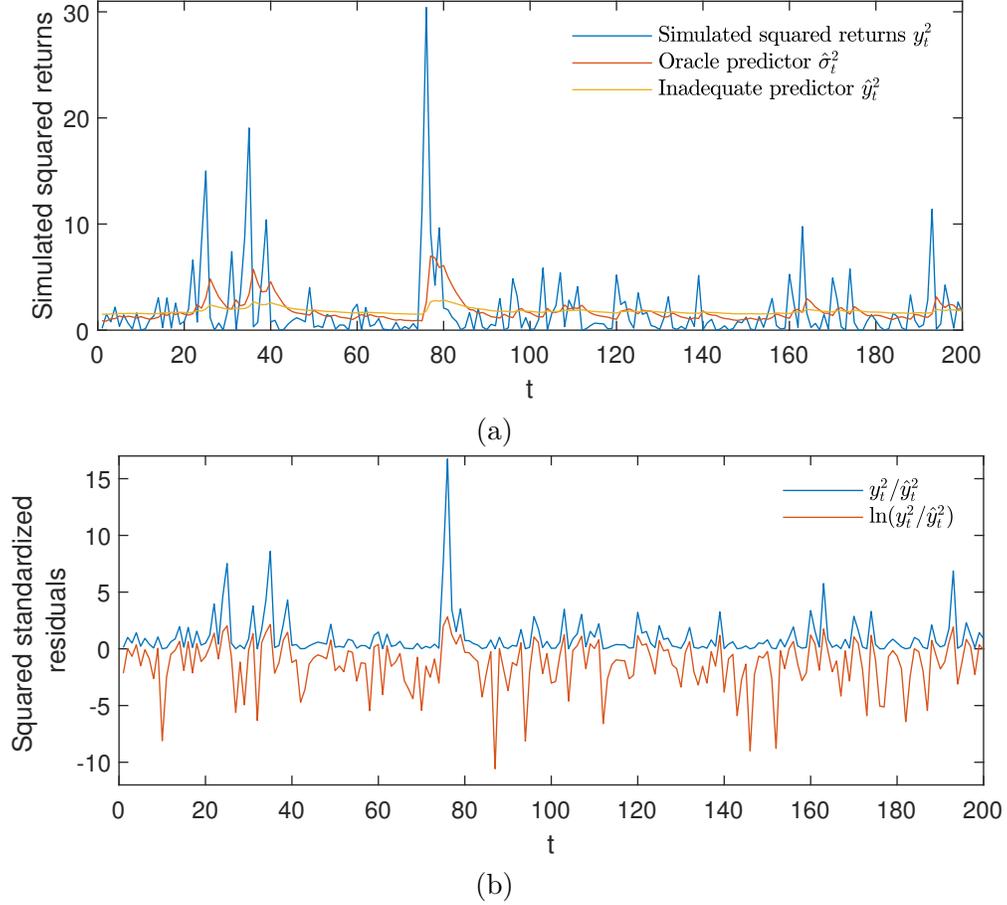


Figure 12: (a) Plot of a randomly selected sub-period of the simulated GARCH squared return series  $y_t^2$ , and the corresponding optimal predictors  $\hat{\sigma}_t^2$  and inadequate predictors  $\hat{y}_t^2$ . (b) Plot of the corresponding squared standardized residuals of inadequate predictors  $y_t^2/\hat{y}_t^2$  and the logarithm of squared standardized residual of inadequate predictors  $\ln(y_t^2/\hat{y}_t^2)$ .

incorrect conclusion in relation to a sufficient or an inadequate model fitted to a given time series. Therefore, I develop a new model sufficiency test that can remedy the limitations of the existing tools.

### 4.3 Rationale and specification of the PD model sufficiency test

Before we move to the construction and introduction of the new model sufficiency indicators, I make an assumption of the general form of time series dynamics. Assume any stationary time series  $\{x_t : t = 1, 2, \dots, N\}$  can be represented or be transformed into the form

$$x_t = c + g(x_{t-1}, x_{t-2}, \dots, x_{t-\tau_T}) + \varepsilon_t \quad (18)$$

where  $\tau_T$  is the furthest lag in which past entries affect the current entry  $x_t$ ,  $g(\cdot)$  is a deterministic function that connects past observations to the expected value of the current entry,  $E(\varepsilon_t) = 0$  and  $E(\varepsilon_t|F_{t-1}) = 0$  so that innovations cannot be further exploited to improve the forecasting accuracy in conditional mean of the current entry and  $F_{t-1}$  denotes the data set available at time  $t - 1$ . The representative additive form of time series in (18) separates the first-moment structures from the higher-moment structures exhibited in the underlying dynamics. More specifically, the deterministic function  $g(\cdot)$  determines the maximum potential of the point prediction capacity of the underlying process. Accordingly, the oracle one-step-ahead point predictions of the future variable  $x_t$  is

$$\hat{x}_t^{\text{oracle}} = E(x_t|F_{t-1}) = c + g(x_{t-1}, x_{t-2}, \dots, x_{t-\tau_T}).$$

The optimal point predictor of  $x_t$  is required to perfectly replicate  $g(\cdot)$ , but it does not need to correctly specify the structures (if existing) within  $\varepsilon_t$  term. The additive innovation term  $\varepsilon_t$  can violate from white noise when there exists higher-order dependence structures that do not have impacts on  $E(x_t|F_{t-1})$  but affect the higher-moment of the distribution of the response.

To be able to assess the sufficiency of a given model in fulfilling the point prediction task, particularly for time series with innovations with dynamical structures, I propose a new method for model sufficiency evaluation by extending the ordinal pattern concept used in the computation of PE.

The rationale behind the proposed test is built on the indicative relation between the strength of dependence within residuals to that between residuals and lagged observations under the oracle and sub-optimal point predictors. Given a univariate time series  $X = \{x_t; t = 1, \dots, N\}$ , I create two statistics  $PD_\tau^D(\hat{\varepsilon}_t)$  and  $PD^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t)$  to measure the temporal dependence structures within residuals and the bivariate dependence between the residuals and the lagged observations, respectively. Segment length  $D$  and delay  $\tau$  are pre-chosen parameters.  $\hat{\varepsilon}_t$  represents the residual at time  $t$  formed by the difference between the observation  $x_t$  and the constructed point predictor  $\hat{x}_t$  from any model, that is  $\hat{\varepsilon}_t = x_t - \hat{x}_t$ .

The permutation dependence measure  $PD_\tau^D$  defined in section 3.4 quantifies the strength of temporal dependence in univariate time series under study. Its specification records the predictability of an upcoming entry based on the ordinal pattern in the  $\tau$  lagged consecutive observations. The predictability of an upcoming entry is reflected by the level of disparity of the ordinal patterns' distribution between the segments  $\hat{\varepsilon}_{t,\tau}^D$  partitioned in the original data with segments  $\hat{\varepsilon}_{t,\tau}^{D,rand}$  which contain an independent random variable as the last entry. If

the time series under study follows a purely random process, no difference can be detected between segments  $\hat{\varepsilon}_{t,\tau}^D$ 's and  $\hat{\varepsilon}_{t,\tau}^{D,rand}$ 's ordinal pattern distribution. In both cases the ordinal pattern in the past entries have no impact, thus no indicating power for the upcoming  $\hat{\varepsilon}_{t+D-2+\tau}$  or randomly generated entry  $r$ . Thereby the value of  $PD_\tau^D$  will be insignificant. However, if the time series has deterministic temporal dependence structure at delay  $\tau$ , there will be a significant discrepancy between the level of regularities of ordinal patterns in the segments partitioned in the original series and that involving independent random variables.

Following the similar rationale,  $PD^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t)$  is specified to measure the bivariate dependence between two separate time series. The bivariate dependence is quantified through the discrepancy of the ordinal regularity in segments formed by  $(x_{t-(D-2)}, \dots, x_{t-\tau-1}, x_{t-\tau}, \hat{\varepsilon}_t)$  with that in the same segments except replacing  $\hat{\varepsilon}_t$  in every segment with an independent random variable  $r$ . Similarly, the random variable  $r$  is specified to have the same empirical distribution as the residual series  $\{\hat{\varepsilon}_t\}$ . The formal definition of  $PD^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t)$  is given below:

$$PD^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t) = \sum_{\pi_i} \left( \left| \frac{p_\tau(\pi_i) - E(p_\tau^R(\pi_i))}{\text{s.d.}(p_\tau^R(\pi_i))} \right|^2 \right)$$

where

$$p_\tau(\pi_i) = \frac{\#\{s_{t,\tau}^D | s_{t,\tau}^D \text{ has ordinal pattern } \pi_i\}}{N - D + 2 - \tau}, \quad i = 1, \dots, D!, \quad (19)$$

$$p_\tau^R(\pi_i) = \frac{\#\{s_{t,\tau}^{D,R} | s_{t,\tau}^{D,R} \text{ has ordinal pattern } \pi_i\}}{N - D + 2 - \tau}, \quad i = 1, \dots, D!,$$

and

$$\begin{aligned} s_{t,\tau}^D &= (x_{t-\tau-(D-2)} - \bar{x}, \dots, x_{t-\tau-1} - \bar{x}, x_{t-\tau} - \bar{x}, \hat{\varepsilon}_t - \bar{\varepsilon}), \quad t = D - 1 + \tau, \dots, N, \\ s_{t,\tau}^{D,R} &= (x_{t-\tau-(D-2)} - \bar{x}, \dots, x_{t-\tau-1} - \bar{x}, x_{t-\tau} - \bar{x}, r - \bar{r}), \quad t = D - 1 + \tau, \dots, N. \end{aligned}$$

The reason I subtract the mean of the time series in the constructed segments is to ensure the comparability of the entries from the two distinct series. Both of the measures  $PD_\tau^D(\hat{\varepsilon}_t)$  and  $PD^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t)$  are strictly positive. The greater their value the stronger dependence they indicate.

The  $PD_\tau^D(\hat{\varepsilon}_t)$  and  $PD^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t)$  statistics approximate the value of

$$\text{Dependence}(\hat{\varepsilon}_{t-\tau}^{D-1}, \hat{\varepsilon}_t) \equiv \int_y \int_x \left[ p_{(\hat{\varepsilon}_{t-\tau}^{D-1}, \hat{\varepsilon}_t)}(x, y) - p_{\hat{\varepsilon}_{t-\tau}^{D-1}}(x) p_{\hat{\varepsilon}_t}(y) \right]^2 dx dy, \quad (20)$$

and

$$\text{Dependence}(x_{t-\tau}^{D-1}, \hat{\varepsilon}_t) \equiv \int_y \int_x \left[ p_{(x_{t-\tau}^{D-1}, \hat{\varepsilon}_t)}(x, y) - p_{x_{t-\tau}^{D-1}}(x) p_{\hat{\varepsilon}_t}(y) \right]^2 dx dy, \quad (21)$$

respectively, where  $\hat{\varepsilon}_{t-\tau}^{D-1} = (\hat{\varepsilon}_{t-\tau-(D-2)}, \dots, \hat{\varepsilon}_{t-\tau-1}, \hat{\varepsilon}_{t-\tau})$  and  $x_{t-\tau}^{D-1} = (x_{t-\tau-(D-2)}, \dots, x_{t-\tau-1}, x_{t-\tau})$ .  $p_{(\hat{\varepsilon}_{t-\tau}^{D-1}, \hat{\varepsilon}_t)}$  and  $p_{(x_{t-\tau}^{D-1}, \hat{\varepsilon}_t)}$  represent the joint probability distribution between  $\hat{\varepsilon}_{t-\tau}^{D-1}$  and  $\hat{\varepsilon}_t$ , and between  $x_{t-\tau}^{D-1}$  and  $\hat{\varepsilon}_t$ .  $p_{\hat{\varepsilon}_{t-\tau}^{D-1}}$  and  $p_{x_{t-\tau}^{D-1}}$  are the marginal probability distributions of  $\hat{\varepsilon}_{t-\tau}^{D-1}$  and  $x_{t-\tau}^{D-1}$ .  $p_{\hat{\varepsilon}_t}$  is the marginal probability distribution of  $\hat{\varepsilon}_t$ . The term dependence refers to the more general definition of dependence that is not limited to linear dependence like the correlation coefficient. More specifically, the overall level of dependence between two variables is measured by the disparity between their joint distribution and the product of their respective marginal distributions.

One might be confused about the difference between the bivariate dependence measure  $H_\tau^D(\cdot, \cdot)$  specified in section 3.9 and the newly proposed bivariate measure  $\text{PD}^{*,D}(\cdot, \cdot)$ , as they both measure the bivariate dependence. Given a pair of time series  $X = \{x_t\}$ ,  $Y = \{y_t\}$ , it is worth clarifying that  $H_\tau^D(\cdot, \cdot)$  measures the level of co-movement between  $(x_t, x_{t+\tau}, x_{t+2\tau})$  and  $(y_t, y_{t+\tau}, y_{t+2\tau})$ , whereas,  $\text{PD}^{*,D}(\cdot, \cdot)$  quantifies how the patterns of  $(x_t, x_{t+1})$  affects  $y_{t+1+\tau}$ . The reason that we propose  $\text{PD}^{*,D}(\cdot, \cdot)$  and not use  $H_\tau^D(\cdot, \cdot)$  for the model sufficiency test is that the test requires the employed bivariate dependence measure to be comparable to the PD measure. In particular, the employed bivariate measure on independent time series pair is required to have identical statistical property to PD computed on random series.  $H_\tau^D(\cdot, \cdot)$  does not meet such requirement. That is why we need to propose a new measure as a part of the test statistics of the hypothesis test.

The values of the expressions given in (20) and (21) are hard to obtain, especially in empirical analysis where the underlying process of  $\{x_t : t = 1, 2, \dots, N\}$  and  $\{\hat{\varepsilon}_t : t = 1, 2, \dots, N\}$  are unknown.  $\text{PD}_\tau^D(\hat{\varepsilon}_t)$  and  $\text{PD}^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t)$  provide a way to approximate them by using the ordinal pattern. By symbolizing the constructed segments according to their ordinal patterns, we convert the complex expression given in (20) and (21) which involve integrals and comparisons between unknown continuous joint distributions and products of continuous marginal distributions into simple comparisons of easily accessible discrete distributions.

By grouping the constructed segments by their ordinal patterns, I only extract the information relating to the relative magnitudes of the involved entries. By considering the relative magnitudes instead of the actual observed values, I lose some information relating to time series under consideration, such as its marginal distribution. But the essential features of the dynamics are kept, such as the temporal dependence structure of the data generating process. The primary advantages of the ordinal-pattern based measures reside in their flexibilities

of accounting for any form of structures, minimal requirement of prior assumptions and knowledges, and the robustness to both dynamical and stochastic noises.

From the generalized form of time series dynamics in (18), the residual generated from a predictor  $\hat{y}_t = \hat{c} + \hat{g}(x_{t-1}, x_{t-2}, \dots, x_{t-\tau_p})$  can be presented in the form of

$$\hat{\varepsilon}_t = c^r + g^r(x_{t-1}, x_{t-2}, \dots, x_{t-\tau^*}) + \varepsilon_t. \quad (22)$$

where  $\tau_p$  is the furthest lag used in the prediction,  $\tau_T$  is the furthest lag in which past entries affects the current entry  $x_t$ ,  $\tau^* = \max(\tau_T, \tau_p)$ ,  $c^r = c - \hat{c}$  and  $g^r(x_{t-1}, \dots, x_{t-\tau^*}) = g(x_{t-1}, \dots, x_{t-\tau^*}) - \hat{g}(x_{t-1}, \dots, x_{t-\tau^*})$ . If the estimated predictor is oracle, the resulting residuals are equal to the “true” additive innovations, that is,

$$\hat{\varepsilon}_t = \varepsilon_t. \quad (23)$$

Otherwise, the estimated residual contains three components, which are the remaining constant term, the residual deterministic function and the “true” additive innovation term as specified in (22).

Consequently, by substituting  $\hat{\varepsilon}_t$  with the generalized form of residual in (22) and (23), and neglecting the constant term, under the condition that the estimated residuals are generated from an optimal point predictor,

$$\begin{aligned} \text{PD}_\tau^D(\hat{\varepsilon}_t) &\equiv \text{Dependence}(\varepsilon_{t-\tau}^{D-1}, \varepsilon_t) \\ \text{PD}^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t) &\equiv \text{Dependence}(x_{t-\tau}^{D-1}, \varepsilon_t). \end{aligned}$$

Otherwise

$$\begin{aligned} \text{PD}_\tau^D(\hat{\varepsilon}_t) &\equiv \text{Dependence}\{[g^r(x_{t-D+1}^{\tau^*}) + \varepsilon_{t-D+2}, \dots, g^r(x_{t-1}^{\tau^*}) + \varepsilon_t], g^r(x_{t+\tau-1}^{\tau^*}) + \varepsilon_{t+\tau}\}, \\ \text{PD}^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t) &\equiv \text{Dependence}[x_t^{D-1}, g^r(x_{t+\tau-1}^{\tau^*}) + \varepsilon_{t+\tau}], \end{aligned}$$

where  $x_{t-1}^{\tau^*} = (x_{t-1}, x_{t-2}, \dots, x_{t-\tau^*})$ ,  $\tau_p$  is the furthest lag used in the prediction,  $\tau_T$  is the furthest lag in which past entries  $x_{t-\tau}$  affects the current entry  $x_t$ ,  $\tau^* = \max(\tau_T, \tau_p)$ . Assuming the temporal dependence within innovation process  $\{\varepsilon_t\}$  deteriorates for increasing lags, the expected value and the relation between  $\text{PD}_\tau^D(\hat{\varepsilon}_t)$  and  $\text{PD}^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t)$  under different scenarios are summarized in Table 3.

Under the oracle point forecast there are two possible scenarios. When the innovation terms are independent of each other, both  $\text{PD}_\tau^D(\hat{\varepsilon}_t)$  and  $\text{PD}^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t)$  will be insignificant. This

Table 3: Expected behaviours of statistic  $\text{PD}_\tau^D(\hat{\varepsilon}_t)$  and  $\text{PD}^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t)$  when the evaluated point forecast is oracle and inadequate, in the presence of independent and dependent innovations.

Point forecast	$\varepsilon_t$ is independent	$\varepsilon_t$ is dependent
Oracle	$\text{PD}_\tau^D(\hat{\varepsilon}_t) = 0$ and $\text{PD}^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t) = 0$	$\text{PD}_\tau^D(\hat{\varepsilon}_t) > 0$ , $\text{PD}^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t) > 0$ and $\text{PD}_\tau^D(\hat{\varepsilon}_t) \geq \text{PD}^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t)$
Inadequate	$\text{PD}_\tau^D(\hat{\varepsilon}_t) > 0$ , $\text{PD}^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t) > 0$ and $\text{PD}^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t) > \text{PD}_\tau^D(\hat{\varepsilon}_t)$	$\text{PD}_\tau^D(\hat{\varepsilon}_t) > 0$ , $\text{PD}^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t) > 0$ and $\text{PD}_\tau^D(\hat{\varepsilon}_t)$ can be less or greater than $\text{PD}^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t)$

is due to the fact that no dependence structure exist between past and current innovations or between any historical observation in  $\{x_t\}$  and the current innovation. However, when the innovation is dependent, both statistics  $\text{PD}_\tau^D(\hat{\varepsilon}_t)$  and  $\text{PD}^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t)$  will be significant. Their significance arises from the temporal dependence within  $\{\varepsilon_t\}$ . Besides, the value of  $\text{PD}^{*,D}(x_{t-\tau}, \hat{\varepsilon}_t)$  will be smaller than  $\text{PD}_\tau^D(\hat{\varepsilon}_t)$  since the only connection between  $x_t^{D-1}$  and  $\varepsilon_{t+\tau}$  is through the intermediate of the temporal dependence of  $\{\varepsilon_t\}$  quantified by  $\text{PD}_\tau^D(\hat{\varepsilon}_t)$ .

I name the newly proposed test the PD model sufficiency test. Suppose the data generating process of an observed time series  $\{x_t; t = 1, 2, \dots, N\}$  can be written in the form

$$x_t = c + g(x_{t-\tau}, x_{t-\tau-1}, \dots) + \varepsilon_t,$$

where  $E(\varepsilon_t) = 0$ ,  $E(\varepsilon_t | F_{t-\tau}) = 0$  and  $F_{t-\tau}$  denotes the information set available at time  $t - \tau$ . Therefore  $c + g(x_{t-\tau}, x_{t-\tau-1}, \dots)$  constitutes the oracle  $\tau$ -step-ahead point forecast of  $x_t$ . A given model provides a  $\tau$ -step-ahead point forecast  $\hat{x}_{\tau,t}$  to predict  $x_t$ , and  $\hat{x}_{\tau,t}$  can be written in the form

$$\hat{x}_{\tau,t} = \hat{c} + \hat{g}(x_{t-\tau}, x_{t-\tau-1}, \dots).$$

The null hypothesis of the PE model sufficiency test is

$$\mathcal{H}_0 : \hat{g}(\cdot) = g(\cdot),$$

versus the alternative hypothesis that

$$\mathcal{H}_1 : \hat{g}(\cdot) \neq g(\cdot).$$

The test shows whether the deterministic relation postulated by the predictor coincides with the “true” deterministic function governing the first-order moment of the investigated process.

The acceptance condition of  $\mathcal{H}_0$  is

$$\text{PD}_\tau^D(\hat{\varepsilon}_t) = 0 \quad \text{and} \quad \text{PD}^{*,D}(\hat{\varepsilon}_t, x_{t-\tau}) = 0, \quad (24)$$

and the rejection criteria of  $\mathcal{H}_0$  is

$$\text{PD}^{*,D}(\hat{\varepsilon}_t, x_{t-\tau}) > 0 \quad \text{and} \quad \text{PD}^{*,D}(\hat{\varepsilon}_t, x_{t-\tau}) - \text{PD}_\tau^D(\hat{\varepsilon}_t) > 0. \quad (25)$$

The acceptance condition specified in (24) corresponds to the case summarized in Table 3 when  $\varepsilon_t$  is independent, and the point forecast is equal to the oracle point forecast. The rejection criteria in (25) eliminates the two possible scenarios of oracle point forecast summarized in Table 3 when  $\varepsilon_t$  is independent or dependent.

To determine whether  $\text{PD}_\tau^D(\hat{\varepsilon}_t)$  and  $\text{PD}^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$  are significantly greater than zero, I need to compare their values with the critical value of  $\text{PD}_\tau^D$  ( $\text{PD}^{*,D}$ ) under independence. Since both of the measures  $\text{PD}_\tau^D$  and  $\text{PD}^{*,D}$  are ordinal-based, they neglect and are invariant to the empirical distribution of the investigated series. The statistical property of  $\text{PD}_\tau^D$  or  $\text{PD}^{*,D}$  is identical for all random processes with any form of continuous univariate distribution. Therefore, an estimate of the 95% C.I. of  $\text{PD}_\tau^D$  (or  $\text{PD}^{*,D}$ ) can be obtained by taking the average of the 95th percentile of the measure computed on a sufficient number of paths of simulated iid series. The generated series needs to be of the same length as the investigated time series, and can be specified to follow any arbitrary continuous distribution.

Apart from assessing the significance of  $\text{PD}_\tau^D(\hat{\varepsilon}_t)$  and  $\text{PD}^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$ , the test needs to determine whether the measure  $\text{PD}^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$  is significantly greater than  $\text{PD}_\tau^D(\hat{\varepsilon}_t)$  when  $\text{PD}^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$  shows significant value. However, when the independence assumptions cannot be met, the distribution of  $\text{PD}_\tau^D(\hat{\varepsilon}_t)$  and  $\text{PD}^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$ , especially their variance, vary depending on the data generating process of  $\{\hat{\varepsilon}_t\}$  and  $\{x_t\}$ . Therefore I choose to use the block bootstrapping method to estimate the critical values of the test statistics  $\text{PD}^{*,D}(\hat{\varepsilon}_t, x_{t-\tau}) - \text{PD}_\tau^D(\hat{\varepsilon}_t)$  when  $\text{PD}^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$  exceed its critical value under independence assumptions.

One might notice, the acceptance and rejection conditions given in (24) and (25) do not exhaustively cover all possible scenarios. There is one particular scenario in which the new test cannot ascertain whether the constructed predictor is sufficient or not. That is when

$$\text{PD}^D(\hat{\varepsilon}_t, x_{t-\tau}) > 0 \quad \text{and} \quad \text{PD}^{*,D}(\hat{\varepsilon}_t, x_{t-\tau}) - \text{PD}_\tau^D(\hat{\varepsilon}_t) < 0.$$

Such a scenario can arise either the prediction is oracle but the innovation exhibits strong structures, or the prediction is inadequate but the dependence structure within the innovations

is too strong and overweights the strength of uncaptured deterministic relations in the original dynamics. In that case, our test cannot make an affirmative conclusion of the sufficiency of the constructed predictor. This is a major weakness of our proposed test and an opportunity of future studies.

As for the selection of the segment length and delay parameter  $D$  and  $\tau$ , following the guidelines of common choice of  $D$  in computing PE, I recommend the required segment length parameter  $D$  to be chosen in the range of  $3 \leq D \leq 7$  so the number of total partitioned segments is excessively greater than the number of possible ordinal patterns  $D!$  (Rosso et al. 2007, Kowalski et al. 2007). The choice of delay  $\tau$  provides practitioners the flexibility to investigate the structure of time series over short-term and long-term dynamics. For most financial dynamics, the temporal dependence structure is generally strongest between entries that are closest to each other, therefore in the later applications of the PD model sufficiency test, I mainly evaluate the test statistics at  $\tau = 1$  to draw inferences about the sufficiency of the employed models.

#### 4.4 Simulation studies for the new test

I conduct simulation studies to demonstrate several applications of our newly proposed test. The PD model sufficiency test will be applied to six different simulated time series that correspond to different time series dynamics all transformed into the general additive form of time series process given in (18). An overview of the six forms of simulating time-series dynamics is given below.

- $X1$ : linear deterministic function  $g(\cdot)$  with iid normally distributed innovations  $\varepsilon_t$  (ARMA)
- $X2$ : linear deterministic function  $g(\cdot)$  with iid asymmetrically distributed innovations  $\varepsilon_t$
- $X3$ : linear deterministic function  $g(\cdot)$  with asymmetrically distributed structural innovations  $\varepsilon_t$  (GARCH)
- $X4$ : nonlinear deterministic function  $g(\cdot)$  with iid normally distributed innovations  $\varepsilon_t$
- $X5$ : nonlinear deterministic function  $g(\cdot)$  with iid asymmetrically distributed innovations  $\varepsilon_t$
- $X6$ : nonlinear deterministic function  $g(\cdot)$  with asymmetrically distributed structural innovations  $\varepsilon_t$ .

In the simulation studies, the PD model sufficiency test is used to indicate the sufficiency of various point forecasts constructed by different models. The prediction models considered include the ARMA model, GARCH model, Gaussian Process regression (GPR) and support vector regression (SVR). The specifications of the employed models are covered in section 2.3. Each simulated series is of length 6360, which corresponds to the number of 1-hour intervals of a one-year financial time series excluding weekends. The selected data length coincides with the length of the empirical realized volatility data that will be investigated later in this thesis. The first 5160 observations (around 80% of total length) are used to estimate the model parameters, and the last 1200 observations are used in assessing out-of-sample prediction accuracy. The non-parametric approaches, namely GPR and SVR, require the pre-chosen hyperparameters. For non-parametric models, the first 4000 observations are used to train the prediction model. The subsequent 1160 observations are used to compare and select the optimal hyperparameters and determine the number of lagged observations  $d$  used as input variables. The last 1200 observations are used in assessing out-of-sample prediction accuracies.

The validity and credibility of our proposed test are verified by contrasting the inferences drawn from the PD model sufficiency test with the prediction error rate of each model under consideration. The prediction error rate reflects the “true” point forecasting performance, and is measured by the distance between the constructed predictor  $\hat{x}_t$  from the oracle predictor  $\hat{x}_t^{\text{oracle}}$  of  $x_t$  relative to  $\hat{x}_t^{\text{oracle}}$ 's variance, that is,

$$\text{prediction error rate} = \frac{\frac{1}{N} \sum_{t=1}^N \left[ (\hat{x}_t - \hat{x}_t^{\text{oracle}}) - \frac{1}{N} \sum_{t=1}^N (\hat{x}_t - \hat{x}_t^{\text{oracle}}) \right]^2}{\text{var}(\hat{x}_t^{\text{oracle}})}.$$

I subtract the systematic bias  $\frac{1}{N} \sum_{t=1}^N (\hat{x}_t - \hat{x}_t^{\text{oracle}})$  in evaluating point forecasting performance because our ordinal-based sufficiency test cannot unveil the derivation in the estimation of the constant term. In other words, the systematic bias introduced by the predictor cannot be captured by the dependence measures used in our test. Due to the nature of the simulation study, the data generating process of the simulated series is known in advance, unlike in the empirical analysis. Therefore we can acquire the oracle point predictor for every simulated series. Based on the general form of time series given in (18)

$$\hat{x}_t^{\text{oracle}} = \text{E}(x_t | F_{t-1}) = x_t - \varepsilon_t,$$

where  $F_{t-1}$  denotes the information set available at time  $t - 1$ ,  $\text{E}(\varepsilon_t) = 0$  and  $\text{E}(\varepsilon_t | F_{t-1}) = 0$ .  $\hat{x}_t^{\text{oracle}}$  is the best one-step-ahead forecast one can construct of  $x_t$ .

Due to the inevitable uncertainties in the model estimation procedures, even the most optimal model would not eliminate prediction error. I divide the prediction error rate into three categories to indicate the cases, namely: (1) when the postulated point predictions are reasonably close; (2) within a moderate distance; (3) and significantly away from the oracle predictor relative to the overall variations of the underlying dynamics. Based on the value of the prediction error rate, we classify the point prediction performance of the considered models into three categories: sufficient (prediction error rate  $\leq 5\%$ ); acceptable ( $5\% < \text{prediction error rate} \leq 15\%$ ); and inadequate (prediction error rate  $> 15\%$ ).

In addition to examining our new test's validity, the objective of the simulation study is to investigate the point forecasting capability of the models in response to various potentially challenging properties commonly observed in financial time series, such as non-normality, non-linearity and dynamical innovations. The specifications of the data generating process of the six simulated time series are given below, all transformed into the additive form given in (18).

- $X1$ : simulated ARMA(1,1) series with parameter  $(\phi_0, \phi_1, \theta_1, \delta^2) = (0.18, 0.9, 0.74, 9)$ . The deterministic function of the data generating process is  $g(x_{t-1}, x_{t-2}, \dots, x_{t-\tau_T}) = 0.69 + \sum_{i=1}^{\infty} 0.16 \times 0.74^{i-1} x_{t-i}$ . Innovation terms  $\varepsilon_t$  are iid with normal distribution of mean 0 and variance 9.
- $X2$ : simulated ARMA(1,1) series with the same  $g(\cdot)$  as in  $X1$ . However, the innovation terms  $\varepsilon_t$  are not following a normal distribution. Instead,  $\varepsilon_t$  follows the empirical asymmetric distribution as the unconditional distribution of the innovation of the simulated GARCH(1,1) series  $X3$ .
- $X3$ : simulated GARCH(1,1) series where the squared returns is the target objective, with parameter  $(\alpha_0, \alpha_1, \beta_1) = (0.18, 0.16, 0.74)$ . The parameters of the GARCH model is specially selected so that the deterministic function  $g(\cdot)$  in  $X3$  is identical to that in  $X1$  and  $X2$ . The innovation terms  $\varepsilon_t$  are asymmetric distributed and dependent to each other.
- $X4$ : a kernel function formed nonlinear  $g(\cdot)$  with the same innovation term as in  $X1$ .
- $X5$ : same nonlinear  $g(\cdot)$  as in  $X4$  with the same innovation term as in  $X2$ .
- $X6$ : same nonlinear  $g(\cdot)$  as in  $X4$  with the innovation term as in  $X3$ .

The specifications of the six simulated time series are motivated by the estimated models fitted to empirical data in section 5.6 so that the simulated time series have similar characteristics as the real-world financial time series. Figure 13 plots the simulated series  $X1$  to  $X6$  and their

respective oracle one-step-ahead point forecast  $\hat{x}_t^{\text{oracle}}$ . Figure 14 provides a scatter plot to display the relation between the oracle point forecast  $\hat{x}_t^{\text{oracle}}$  and the nearest lagged observation  $x_{t-1}$ . Additionally, the distances between the prediction  $\hat{x}_t$  made from the considered models and the oracle point forecast are revealed. I also employ the new visualization plot introduced in section 3.8 to reveal the deterministic structures  $\hat{g}(\cdot)$  captured by each considered model and compare it with the “true” deterministic relations governing the first-order structure of the simulated time series. The true  $g(\cdot)$  and the estimated functions  $\hat{g}(\cdot)$  of each estimated models using the PD visualization plot is given in Figure 15.

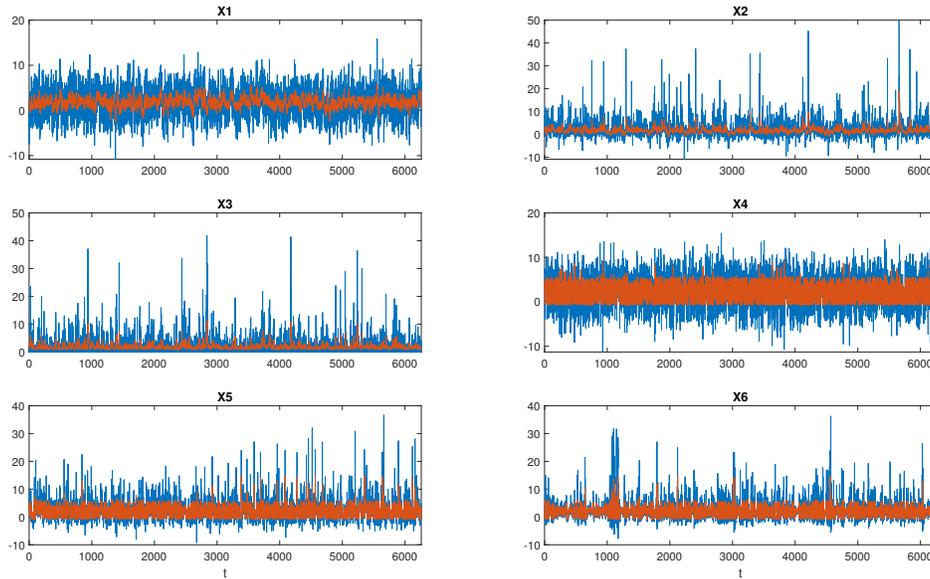


Figure 13: Plot of simulated series X1 to X6 and the conditional mean (oracle prediction) of every observation in each simulated series.

All relevant statistics reflecting the point forecast performance of the models are summarized in Table 4. The first row records the signal-to-noise ratio of the simulated series to reflect their noise contamination level, that is,

$$\text{signal-to-noise ratio} = \frac{\sum_{t=1}^N (\hat{x}_t^{\text{oracle}})^2}{\sum_{t=1}^N (\varepsilon_t^2)}.$$

Subsequently, the value of the prediction error rate is given to reflect the “true” level of the point forecasting performance of the competing models.

In addition, Table 4 provides several forecasting performance metrics that do not require knowledge of the underlying data generating process and thus can be computed in the empirical analysis. I split the widely-used prediction accuracy metric mean squared error (MSE) into two components, namely the adjMSEratio and the bias ratio. The adjMSEratio

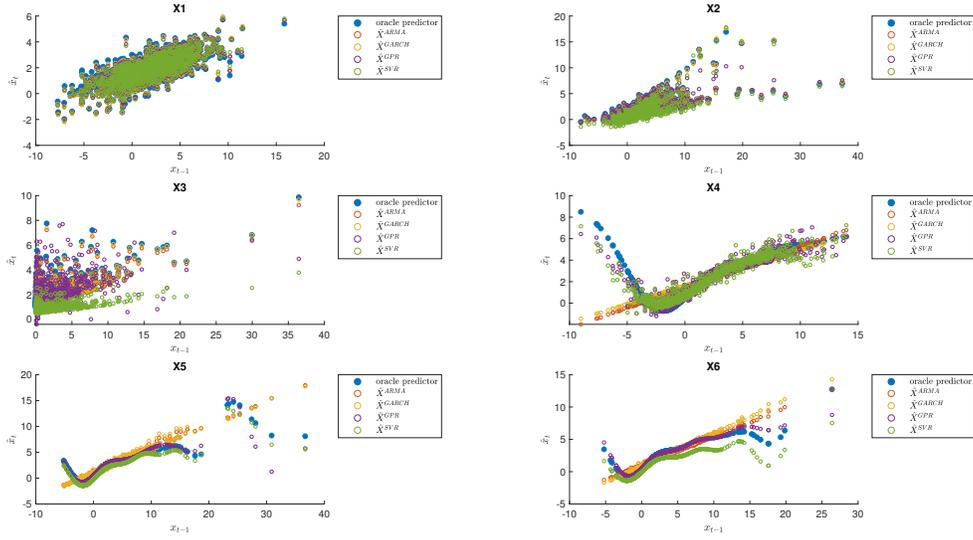


Figure 14: Scatter plot of the oracle one-step-ahead predictions  $\hat{x}_t^{\text{oracle}}$  and prediction constructed by ARMA, GARCH, GPR and SVR models fitted to each simulated series  $X1$  to  $X6$  against the nearest lagged observations  $x_{t-1}$

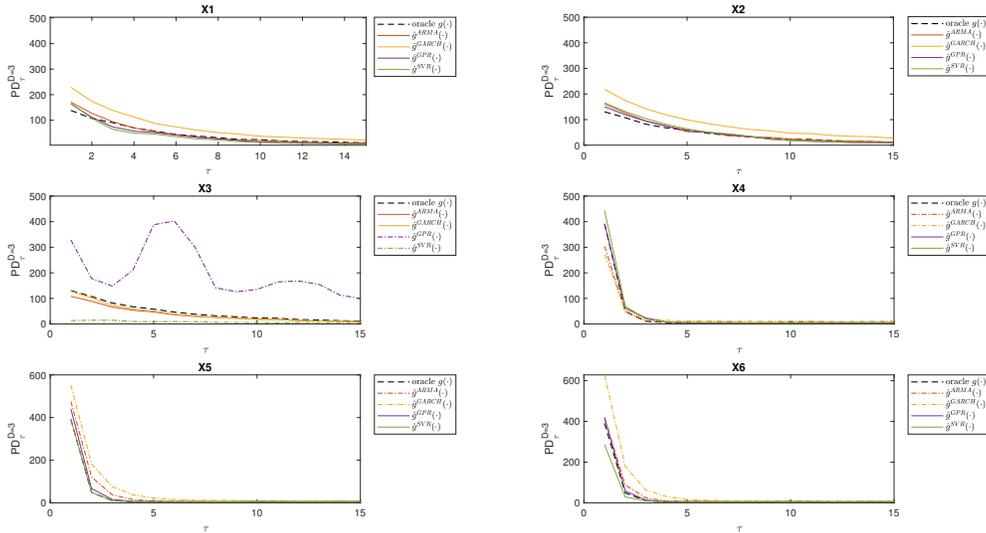


Figure 15: PD visualization plot of the estimated function  $\hat{g}(\cdot)$  postulated by each considered models fitted to the six simulated series  $X1$  to  $X6$ .

and bias ratio are specified as follows

$$\text{adjMSEratio} = \frac{\frac{1}{N} \sum_{t=1}^N \left[ (x_t - \hat{x}_t) - \frac{1}{N} \sum_{t=1}^N (x_t - \hat{x}_t) \right]^2}{\text{var}(x_t)},$$

$$\text{bias ratio} = \frac{\left[ \frac{1}{N} \sum_{t=1}^N (x_t - \hat{x}_t) \right]^2}{\text{var}(x_t)}.$$

The former indicates how well the model replicates the deterministic function governing the investigated data, whereas the latter accounts for the systematic bias in estimating the constant term. The denominator, namely  $\text{var}(x_t)$ , is included to make the statistics more informative. The metric  $\text{adjMSEratio}$  indicates the proportion of unexplained variation in the investigated data left by the postulate predictor. Similarly, the metric  $\text{bias ratio}$  reflects the systematic bias introduced by the considered model relative to the investigated data variance.

The PD model sufficiency test is based on the value of statistics  $\text{PD}_\tau^D(\hat{\varepsilon}_t)$  and  $\text{PD}^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$ . Table 4 records the value of  $\text{PD}_\tau^D(\hat{\varepsilon}_t)$  and  $\text{PD}^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$  of the competing models fitted to each simulated series with segment length  $D = 4$  and delay  $\tau = 1$ . Using Monte Carlo simulations, we construct an estimate of the 95% C.I. of  $\text{PD}_\tau^D(\hat{\varepsilon}_t)$  ( $\text{PD}^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$ ) under independence. The simulation indicates the 95% C.I. of  $\text{PD}_\tau^D(\hat{\varepsilon}_t)$  ( $\text{PD}^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$ ) with  $D = 4$  under independence around 40. Therefore, if

$$\text{PD}_\tau^D(\hat{\varepsilon}_t) < 40 \quad \text{and} \quad \text{PD}^{*,D}(\hat{\varepsilon}_t, x_{t-\tau}) < 40,$$

we conclude the point forecast is sufficient. When  $\text{PD}^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$  is greater than 40, we use the block bootstrapping method to estimate a critical value of  $\text{PD}^{*,D}(\hat{\varepsilon}_t, x_{t-\tau}) - \text{PD}_\tau^D(\hat{\varepsilon}_t)$  to determine whether  $\text{PD}^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$  is significantly greater than  $\text{PD}_\tau^D(\hat{\varepsilon}_t)$ . If that is the case, we reject the null hypothesis and conclude the point forecast is inadequate. Elsewhere, no affirmative conclusions can be drawn. To contrast our proposed test with the BDS test, the inferences of the BDS test are also included in the table as a comparison.

According to the results provided in Table 4, several conclusions can be drawn. Firstly, around 80% of the time, the PD model sufficiency test can successfully indicate the sufficiency/insufficiency of the employed model. Around 20% of the time, we find that no confirmatory inferences can be drawn. The proposed test can be inconclusive when the dependence relation within innovations is too strong, which excessively overweights the residual deterministic relation  $g^r(\cdot)$  left by the predictor. Additionally, by comparing the in-sample and out-of-sample inferences drawn from the PD model sufficiency test, the result implies that the credibility of our proposed test increases with the data size. Consequently,

the sufficiency test conducted on training data with longer data length is more capable of distinguishing sufficient and inadequate predicting models compared to that conducted on the test set. Lastly, by contrasting the inferences made from the conventional BDS test with that from our proposed test, as we expected, the BDS test conclusions are valid when the additive innovations of the underlying dynamics are iid. However, when the additive innovations are non-white, the BDS test can erroneously reject the model that generates the oracle point forecast.

In terms of the point forecasting ability of the ARMA, GARCH, GPR, and SVR models, each model has its strengths and weaknesses. Linear models, ARMA and GARCH, cannot replicate nonlinear deterministic relations as evidenced by their poor performance in predicting  $X4$ ,  $X5$  and  $X6$ , which are governed by nonlinear deterministic function  $g(\cdot)$ . However, the more complex and non-parametric model SVR can be susceptible to the asymmetric distribution of additive innovations. In the SVR model’s attempt to predict simulated series with asymmetrically distributed innovations ( $X2$  and  $X6$ ), the asymmetry leads to substantial derivations in estimating the constant term. More importantly, both GPR and SVR models are severely undermined by the dynamical structures present in the innovations (shown in their poor performance in predicting  $X3$ ), especially when the investigated series has high noise contamination ratios. In contrast, the simpler models, such as ARMA and GARCH, are robust to dynamic innovations.

Table 4: Prediction performance ARMA, GARCH, GPR and SVR models on six simulated time series generated through linear and nonlinear deterministic relation and with different form of innovations. Prediction error rate indicates the “true” predictor accuracy: prediction error rate =  $\frac{1}{N} \sum_{t=1}^N \left[ (\hat{x}_t - \hat{x}_t^{\text{oracle}}) - \frac{1}{N} \sum_{t=1}^N (\hat{x}_t - \hat{x}_t^{\text{oracle}}) \right]^2 / \text{var}(\hat{x}_t^{\text{oracle}})$ . Based on the value of prediction error rate, the performance of various models is classified into three categories: sufficient (prediction error rate  $\leq 5\%$ ) blue cell; acceptable ( $5\% < \text{prediction error rate} \leq 15\%$ ) yellow cell and inadequate (prediction error rate  $> 15\%$ ) red cell. adjMSEratio measures the prediction accuracy without the knowledge of generating process: adjMSEratio =  $\frac{1}{N} \sum_{t=1}^N \left[ (x_t - \hat{x}_t) - \frac{1}{N} \sum_{t=1}^N (x_t - \hat{x}_t) \right]^2 / \text{var}(x_t)$ . bias ratio indicates the bias of each model introduces to the prediction relative to the variance of the investigated data: bias ratio =  $\left[ \frac{1}{N} \sum_{t=1}^N (x_t - \hat{x}_t) \right]^2 / \text{var}(x_t)$ .  $d$  denotes the optimized number of lagged observations used as input variables in constructing GPR and SVR.  $\text{PD}_{\tau}^D(\hat{\varepsilon}_t)$  and  $\text{PD}^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$  with  $D = 4$  and  $\tau = 1$  are the statistics required in the PD model sufficient test. They are computed in both training set and test set. The inferences drawn from the BDS test are also included to compared with that from the PD model sufficiency test.

	X1	X2	X3	X4	X5	X6
signal-to-noise ratio	55.12%	43.91%	45.02%	97.00%	89.16%	79.17%
Oracle						

Continued on next page

	X1	X2	X3	X4	X5	X6
prediction error rate	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
adjMSEratio	92.86%	89.28%	95.58%	74.23%	74.41%	73.24%
bias ratio	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
$PD_{\tau}^D(\hat{\varepsilon}_t)$ in-sample	22.77	31.60	342.16	28.83	26.85	307.90
$PD^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$ in-sample	21.80	28.89	179.55	29.44	25.79	89.98
PD sufficient test	Accept	Accept	Inconclusive	Accept	Accept	Inconclusive
$PD_{\tau}^D(\hat{\varepsilon}_t)$ out-of-sample	30.31	29.97	78.00	29.55	21.34	90.40
$PD^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$ out-of-sample	30.64	18.96	36.11	19.67	22.39	45.85
PD sufficiency test	Accept	Accept	Accept	Accept	Accept	Inconclusive
BDS test	Accept	Accept	Reject	Accept	Accept	Reject
ARMA						
prediction error rate	1.47%	0.52%	0.80%	32.17%	17.90%	16.48%
adjMSEratio	93.00%	89.26%	95.15%	83.12%	78.45%	79.75%
bias ratio	0.00%	0.12%	0.00%	0.00%	0.04%	0.00%
$PD_{\tau}^D(\hat{\varepsilon}_t)$ in-sample	22.71	27.31	305.43	48.22	135.00	404.59
$PD^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$ in-sample	25.61	26.81	174.23	80.81	155.13	288.11
PD sufficiency test	Accept	Accept	Inconclusive	Reject	Inconclusive	Inconclusive
$PD_{\tau}^D(\hat{\varepsilon}_t)$ out-of-sample	34.26	28.13	69.49	22.75	47.05	121.88
$PD^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$ out-of-sample	33.11	22.88	41.44	27.74	70.20	120.70
PD sufficiency test	Accept	Accept	Inconclusive	Accept	Inconclusive	Inconclusive
BDS test	Accept	Accept	Reject	Reject	Reject	Reject
GARCH						
prediction error rate	1.37%	0.48%	0.59%	34.73%	18.34%	16.49%
adjMSEratio	92.98%	89.26%	95.31%	83.62%	78.73%	80.16%
bias ratio	0.00%	0.11%	0.00%	0.00%	0.04%	0.01%
$PD_{\tau}^D(\hat{\varepsilon}_t)$ in-sample	22.12	25.86	308.66	77.40	128.25	127.59
$PD^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$ in-sample	27.47	25.85	183.12	110.17	154.34	148.01
PD sufficiency test	Accept	Accept	Inconclusive	Reject	Inconclusive	Inconclusive
$PD_{\tau}^D(\hat{\varepsilon}_t)$ out-of-sample	33.44	28.36	69.80	33.82	43.24	112.87
$PD^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$ out-of-sample	32.07	22.31	44.07	34.83	64.29	105.93
PD sufficiency test	Accept	Accept	Inconclusive	Accept	Reject	Inconclusive
BDS test	Accept	Accept	Reject	Reject	Reject	Reject
GPR						
$d$	9	8	12	4	1	1

Continued on next page

	X1	X2	X3	X4	X5	X6
prediction error rate	2.62%	11.09%	73.26%	8.24%	2.66%	1.43%
adjMSEratio	93.12%	90.95%	105.29%	75.94%	75.49%	73.69%
bias ratio	0.00%	0.20%	0.00%	0.01%	0.11%	0.00%
$PD_{\tau}^D(\hat{\varepsilon}_t)$ in-sample	19.77	24.23	114.51	37.35	23.98	361.2
$PD^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$ in-sample	18.76	28.84	162.46	25.39	21.66	87.77
PD sufficiency test	Accept	Accept	Reject	Accept	Accept	Inconclusive
$PD_{\tau}^D(\hat{\varepsilon}_t)$ out-of-sample	31.56	31.77	44.08	18.09	17.77	97.37
$PD^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$ out-of-sample	32.82	23.44	48.32	21.14	20.14	41.49
PD sufficiency test	Accept	Accept	Inconclusive	Accept	Accept	Inconclusive
BDS test	Accept	Accept	Reject	Reject	Reject	Reject
SVR						
$d$	8	8	8	4	1	1
prediction error rate	4%	1.26%	43.12%	11.58%	0.25%	5.35%
adjMSEratio	93.04%	89.38%	96.51%	76.87%	74.51%	74.15%
bias ratio	0.00%	2.95%	10.58%	0.00%	4.44%	6.48%
$PD_{\tau}^D(\hat{\varepsilon}_t)$ in-sample	22.92	24.24	156.90	29.36	29.87	203.73
$PD^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$ in-sample	22.97	26.88	135.99	28.84	19.43	32.20
PD sufficiency test	Accept	Accept	Inclusive	Accept	Accept	Accept
$PD_{\tau}^D(\hat{\varepsilon}_t)$ out-of-sample	39.12	27.43	43.51	22.43	20.74	80.20
$PD^{*,D}(\hat{\varepsilon}_t, x_{t-\tau})$ out-of-sample	38.13	22.56	51.87	31.19	23.45	50.23
PD sufficiency test	Accept	Accept	Inconclusive	Accept	Accept	Inconclusive
BDS test	Accept	Accept	Reject	Reject	Reject	Reject

## 5 Empirical application to high-frequency currency exchange series and sea surface temperature data

In this chapter, I conduct an empirical analysis on real-world financial time series using all the ordinal pattern based tools covered in the previous chapters. The time series I choose is the 10-minute interval EUR/USD exchange rate return. There are a number of reasons for my choice. First, unlike other financial markets, the foreign exchange market is a 24 hour **over the counter (OTC)** market, and it is only closed during the weekend: from 10:00 pm GMT Friday (New York) to 10:00 pm GMT Sunday (Sydney). Therefore exchange rate returns suffer less discontinuity issues compared to other financial assets which stop trading

at the end of each day. Second, EUR/USD is the most heavily traded bilateral currency pairs constituting around 24% of the global foreign exchange transactions. The analysis conducted on it can be considered a prototype, where the series under study exhibit many shared features and properties of other financial data. Also, the results obtained can be easily learned and used as references for a wide range of related studies. Lastly, I choose a 10-minute interval as the return frequency since PE analysis is advantageous for data with relatively large size and requires a certain length of data to produce valid results. However, return series of higher frequencies have lower signal-to-noise ratios and a higher proportion of zero returns. I consider a 10-minute interval to be the optimal balance between sufficient data size and the drawbacks caused by a shorter extracting return interval. In addition, past studies conducted in the area of market efficiency strongly indicate the lack of dependence in returns over a longer interval, such as daily, weekly or monthly, thus intraday returns provide more prediction and structure exploration potentials (Reboredo et al. 2013).

In addition to empirical financial time series analysis, I also include a brief section about the empirical application of PE and PD measures on sea surface temperature time series to demonstrate the varied applicability of the investigated measures on various types of data from different disciplines.

## 5.1 Data description

The return series of interest is obtained from computing the simple return of close bid of EUR/USD exchange rate in every 10-minute interval. Let  $X = \{x_t; t = 1, \dots, N\}$  denote the close bid rate of EUR/USD, the return series  $R = \{r_t; t = 1, \dots, N - 1\}$  is computed through

$$r_t = \frac{x_{t+1}}{x_t} - 1,$$

and the squared returns are denoted by  $\text{Rsqr} = \{r_t^2; t = 1, \dots, N - 1\}$ . The close bid price of the EUR/USD exchange rate is provided by Thomson Reuters Tick History (TRTH). I choose to compile data over a six year period, dividing the data set into six non-overlapping 1-year periods. The data set is recorded from 21:00 GMT 16/06/2013 to 20:50 GMT 21/06/2019 with the weekend entries removed (from 21:00 GMT Friday to 20:50 GMT Sunday inclusive). Each period's data starts from the 3rd Sunday of June and ends at the 3rd Friday of June the next year. There are 38160 entries in each price series X, and 38159 entries in each return and squared return series. Every price series X contains missing values, which are public holidays that are not during weekends. On average, missing entries constitute 0.2% of the total series so they can be considered negligible. Figures 16 and 17 show the plots

of time series  $X$  and  $R$  for the six investigated 1-year periods. For the sake of convenience, X2013-14, X2014-15, X2015-16, X2016-17, X2017-18 and X2018-19 represent the 1-year period EUR/USD rate recorded between the specified consecutive years, and R2013-14, R2014-15, R2015-16, R2016-17, R2017-18 and R2018-19 denote the corresponding 10-minute returns. Similarly, Rsq2013-14, Rsq2014-15, Rsq2015-16, Rsq2016-17, Rsq2017-18 and Rsq2018-19 represent the corresponding 10-minute squared return series. I demonstrated earlier in section 3.6 that the PE, PD and ACF measures can be misled by non-stationarity. By observing the six return series plots under study, only R2014-15 displays pronounced non-stationarity of increasing variance along with the indexed time. Therefore, in order to avoid possible distortion, I de-trend the increasing variance in R2014-15 prior to further analysis.

I use simple return instead of log-return because in high-frequency returns equal absolute magnitude entries, especially equal absolute magnitude adjacent entries, are more likely to occur in the log-return than simple returns. For instance, if the three consecutive prices of the target asset are  $x_t = 1.3347$ ,  $x_{t+1} = 1.3349$  and  $x_{t+2} = 1.3348$ , the recorded prices will result in two close but different valued simple returns, or two opposite signed equal magnitude logarithm returns. Note that it is a convention to record most of the exchange rate to four decimal places. When computing the value of PE or other ordinal pattern based measures on squared return series, the occurrence of equal entries will lead to ties in the constructed segments, thus reducing the number of contributing segments and potentially introducing bias into PE and PD. If using logarithm returns in the computation of  $PE_{\tau=1}^{D=3}$  on Rsq2013-14, there will be 9732 segments that have equal entries out of 38158 constructed segments, which accounts for nearly 1/3 of the total segments. In contrast, simple returns only have a total of 3786 tie segments out of 37149, which accounts for just 10% of overall segments. Among all tie segments, 3240 are caused by the occurrence of zero returns. Other than the difference in the number of ties segments, all the results in this thesis for the investigated simple series are also applicable to log-returns, since for short-intervals the value of simple return and log-return is very close.

It is worth noting that the absolute price changes during the weekends are on average around three times the weekday 10-minute price changes. Hence, the cross-weekend returns might distort the continuity in the return dynamic underlying the observed time series. Therefore, I choose to convert them into missing values in the original time series. Nevertheless, due to the limited number of cross-weekend returns compared to the whole data size, the inclusion or removal of cross-weekend returns has minimal effect on the analysis conducted in this thesis.

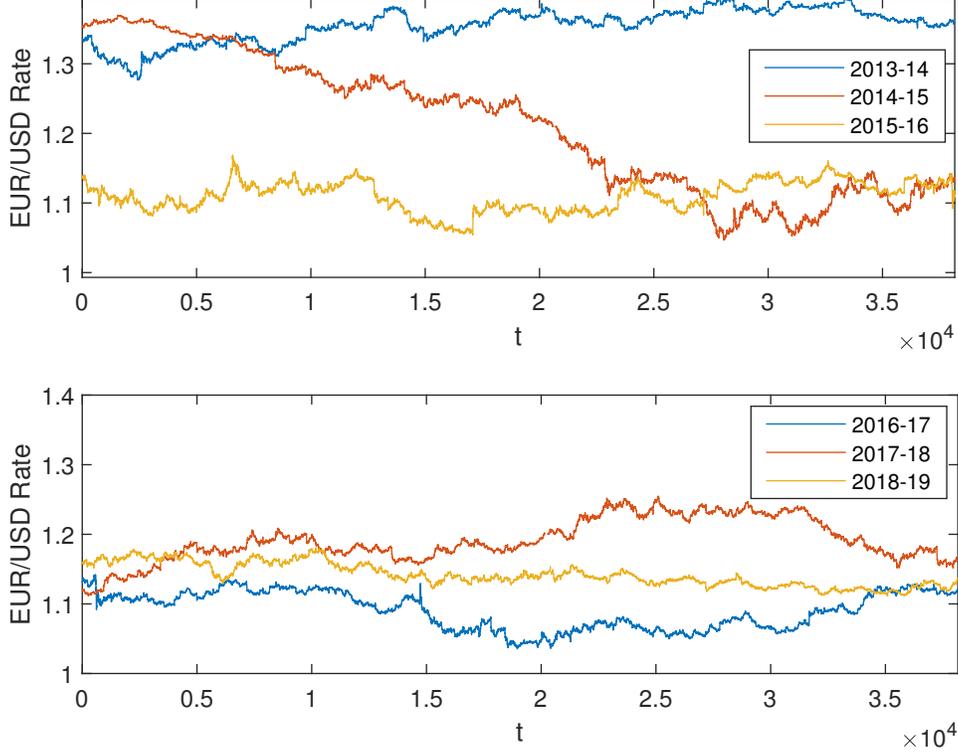


Figure 16: Plots of six non-overlapping 1-year period EUR/USD close bid rate at the end of each 10-minutes interval.

## 5.2 Modifications on PE and PD for equal entries

The definition of PE originally proposed by [Bandt & Pompe \(2002\)](#) neglected the segments with equal entries, probably because the probability of equal entries is negligible for a continuous time series. In the case of segments with equal entries, [Bandt & Pompe \(2002\)](#) proposed to rank equalities in the constructed segments according to their order of occurrence or to eliminate them by adding small random perturbations to the original time series. However, due to the existence of tick size, that is, the minimum price movement unit of a financial instrument, and the low resolution of financial asset recorded price, the probability of equal values in the financial returns, especially in the high-frequency financial returns, can be very high. Besides, the occurrence of equal returns and absolute returns is sometimes structural due to the micro-structure exhibited in the high-frequency financial returns. When the segments with ties are not randomly distributed, the aforementioned approaches of dealing with ties introduces bias into the PE measure. Therefore the computation of PE and PD used in this paper excludes the segments  $s_{t,\tau}^D$  that have equal entries to prevent possible bias

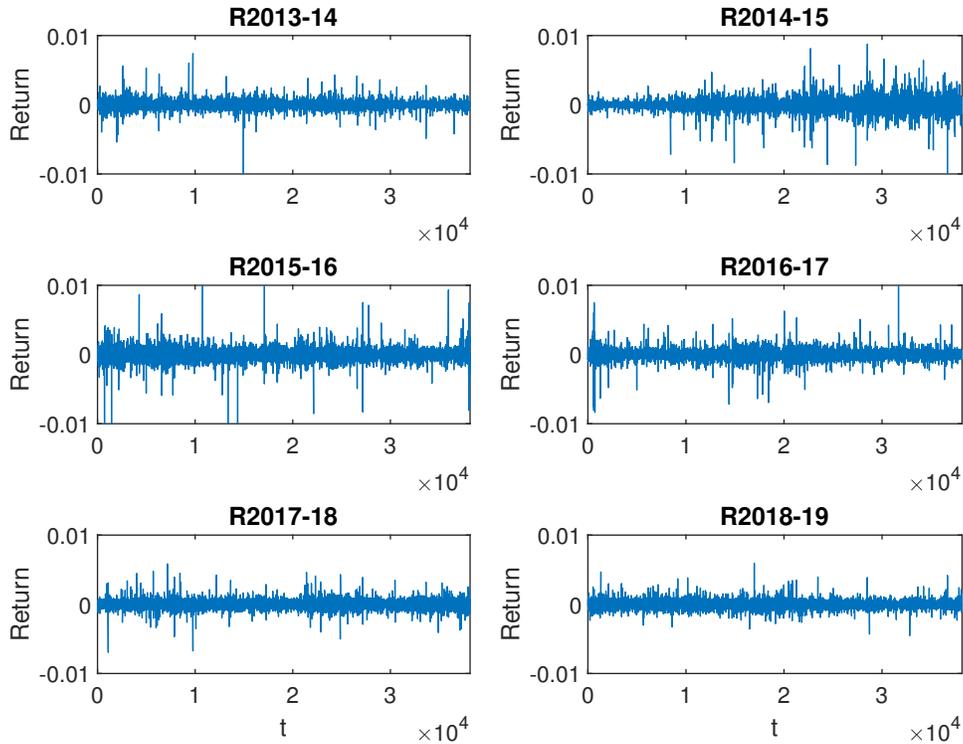


Figure 17: Plots of six non-overlapping 1-year period EUR/USD close bid 10-minutes returns.

caused by the ties in the constructed segments.

It is worth mentioning that [Bian et al. \(2012\)](#) have proposed a modified PE (mPE) for improving the symbolization of constructed segments with equal entries. They showed that the mPE better characterizes the heart rate systems and provides clearer discrimination between heart rate observed in young, elderly and congestive heart failure groups. An interesting future study would be applying mPE as an alternative to treat equal entries on our empirical financial time series under study. However, it is out of the scope of the current thesis.

### 5.3 Temporal dependence of 10-minute EUR/USD returns

This section applies the ordinal pattern based analysis on 10-minute EUR/USD returns to provide an illustrative example of empirical financial time series analysis.

The main tools I use is the PE and PD measures covered in sections 3.1 and 3.4. Additionally, the associated ordinal pattern distribution of fixed length segments partitioned in the observed time series is also informative in identifying the nature of the detected temporal dependence structures.

An overview of the specific procedures of the subsequent empirical analysis is as follows. I first examine the existence of dependence structures in the investigated return series by comparing the value of 1-PE and PD with their respective 99% C.I. under randomness and search for the dominant contributor to the detected temporal dependence structures in 10-minute returns. I find the discretization caused by the existence of the smallest increments and decrement units in financial prices is the dominant source of temporal dependence structures in the investigated return series. Since the discretization effect is the dominant source of dependence in returns, I make some adjustments to the original return series to break the discretion and search for the remaining structures in intraday return dynamics. For the adjusted return series, I decompose the return series into two separate components: sign of return; and the absolute magnitude of return. I investigate the temporal dependence structures exhibited in the binary return sign, in the magnitude of return as separating univariate time series and their bivariate dependence relations, in order to isolate their respective contributions to the temporal structures in return dynamics.

From the real-world application of ordinal pattern analysis on high-frequency exchange rate return, I show that the relative novel PE and PD measures and the associated ordinal pattern analysis not only can detect the existence of temporal dependence structures exhibited in the investigated time series, which is that overlooked by the conventional approaches, but can also indicate of the strength and nature of the detected temporal dependence structures.

### 5.3.1 Preliminary analysis to test against randomness

Both PE and PD have defined asymptotic distributions of zero mean under the assumption that the investigated process is purely random. By plotting  $1 - PE_{\tau}^{D=3}$  and  $PD_{\tau}^{D=3}$  over increasing lags/delays on a 10-minute intraday EUR/USD return series, one can construct randomness hypothesis tests to find evidence in the investigated data against pure randomness. The constructed hypothesis tests are one-tail tests, since both measures  $1 - PE_{\tau}^{D=3}$  and  $PD_{\tau}^{D=3}$  are strictly positive. The greater their values, the stronger the temporal dependence structure. The null hypothesis is

$$\mathcal{H}_0 : \text{time series } X = \{x_t : t = 1, \dots, N\} \text{ follows a purely random process}$$

versus the alternative hypothesis that

$$\mathcal{H}_1 : \text{time series } X = \{x_t : t = 1, \dots, N\} \text{ have temporal dependent structures}$$

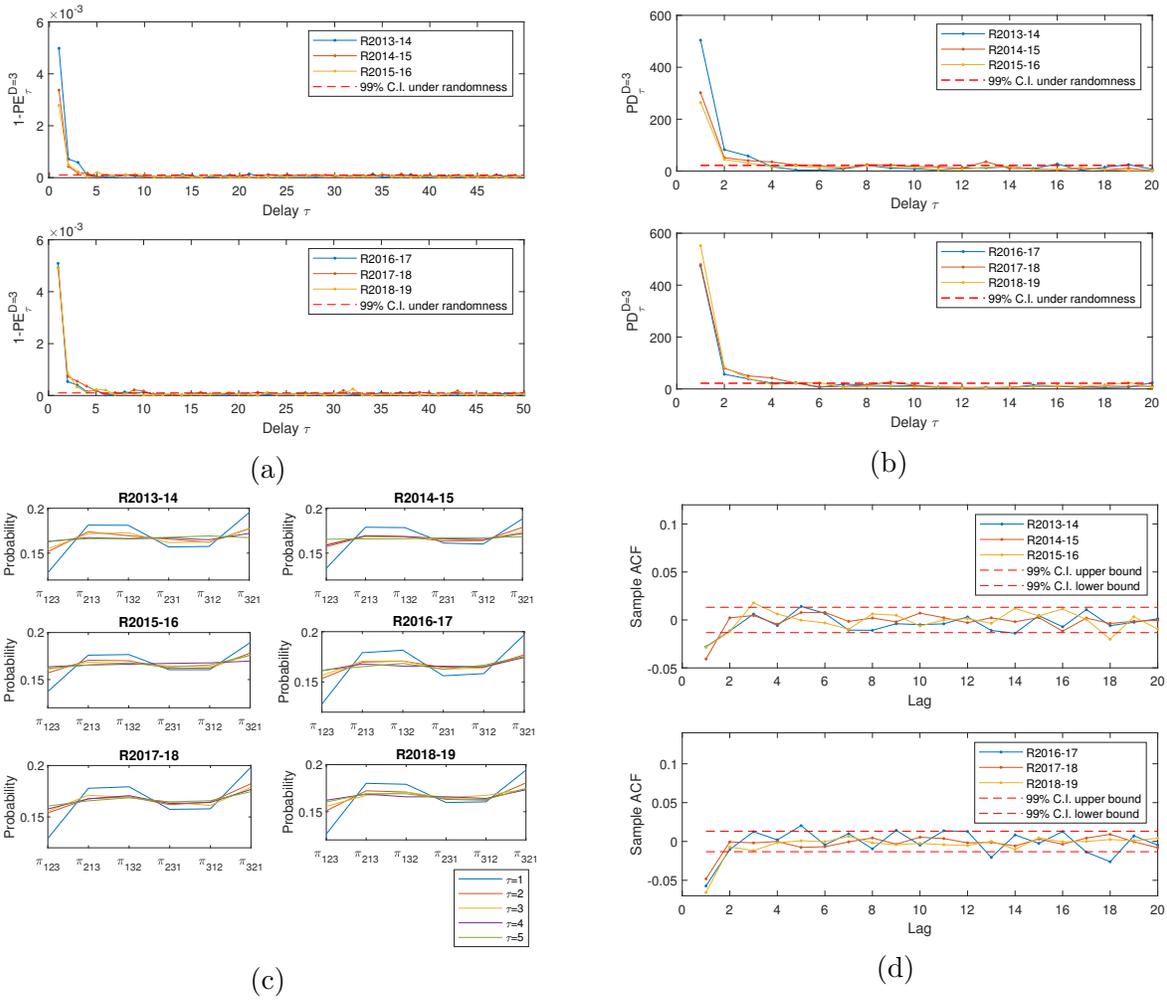


Figure 18: (a) Plot of  $1 - PE_{\tau}^{D=3}$  as a function of delay  $\tau$  in conjunction with its 99% C.I.. (b) Plot of  $PD_{\tau}^{D=3}$  as a function of delay  $\tau$  in conjunction with its 99% C.I.. (c) Probability of each ordinal pattern contributing to  $1 - PE_{\tau}^{D=3}$  that is significant. (d) Sample ACF plot in conjunction with its 99% C.I.. All measures are computed on six none-overlapping 1-year-long EUR/USD 10-minutes return series.

The randomness test can be constructed on both  $1 - \text{PE}_\tau^{D=3}$  and  $\text{PD}_\tau^{D=3}$ . Under  $\mathcal{H}_0$ , based on the result obtained by [Matilla-García & Marín \(2008\)](#)

$$2\{[N - (D - 1)\tau] \ln D!\} [1 - \text{PE}_\tau^D(X)] \longrightarrow \chi_{D!-1}^2$$

Accordingly, for our choice of pattern length  $D = 3$  and investigated data size  $N = 38160$ , the 99% C.I. of  $1 - \text{PE}_\tau^{D=3}$  under  $\mathcal{H}_0$  is around 1.1033e-04 for  $\tau = 1$ , and will increase by a negligible amount for larger delays since the data size  $N$  excessively exceeds  $(D - 1)\tau$  for our choice of  $D$  and the considered delays  $\tau$ . As for the 99% C.I. of  $\text{PD}_\tau^D$  under  $\mathcal{H}_0$ , since I have not derived its theoretical asymptotic distribution under randomness, we use Monte Carlo simulations to construct an estimate of the 99% C.I. of  $\text{PD}_\tau^{D=3}$  on random series. The simulation indicates the 99% C.I. of  $\text{PD}_\tau^{D=3}$  under randomness is around 22. Note that the ordinal based measures neglect and are invariant to the marginal distribution of the investigated series, hence the statistical property of  $\text{PD}_\tau^D$  or  $\text{PE}_\tau^D$  are identical for all random process with any form of univariate distribution. Therefore, an estimate of the 99% C.I. of  $\text{PD}_\tau^{D=3}$  can be obtained by taking the average of the 99th percentile of 500 generated iid series each of length 38160 following any arbitrary distribution. In our simulation, we choose the simplest standard uniform distribution. The plot of  $1 - \text{PE}_\tau^{D=3}$ ,  $\text{PD}_\tau^{D=3}$  and sample ACF over increasing delays for the six period' return series is shown in [Figure 18](#) in conjunction with the 99% confidence interval under randomness. In addition, the ordinal pattern probabilities associated with delays indicating non-randomness are also provided in [Figure 18](#).

The contrast between the plots of PE, PD and ACF as functions of delay/lag on the same investigated series can be quite informative since they all reflect the temporal dependence structures of the underlying dynamics under study, but place emphasis on different aspects. By observing the PE, PD and ACF plots and the ordinal patterns distributions on the 10-minute EUR/USD return series, a number of inferences can be drawn. First of all, the significant value of  $1 - \text{PE}_\tau^{D=3}$  and  $\text{PD}_\tau^{D=3}$  at short-term delays confirms the existence of the temporal dependence structure in the EUR/USD 10-minute return dynamics. In fact, it can be seen from [Figure 18](#) that values of  $1 - \text{PE}_\tau^{D=3}$  and  $\text{PD}_\tau^{D=3}$  drop rapidly from  $\tau = 1$  to  $\tau = 2$  and are only significant before  $\tau = 5$ , suggesting the temporal dependence structure in intraday return mainly exists over the short-term. In addition, the ACF plot only records significant negative correlations at delay 1, which coincides with the stylized fact that high-frequency financial returns normally display extremely high negative first-order autocorrelation. However, if the intraday returns under study only exhibit linear serial correlation, the ordinal pattern associated with the PE should have approximately equal probabilities of ordinal pattern  $\pi_{123}$  and  $\pi_{321}$ , as the result shown in [equation 13](#) proven by [Bandt & Shiha \(2007\)](#). Therefore

we conjecture that the detected temporal dependence structures have a rather significant non-linear component where the negative first order linear serial relation detected by the ACF is not the only, or the dominant structures in the intraday return dynamics. Lastly, despite the varied values of PE and PD for different investigated periods, the distributions of the ordinal pattern that correspond significant PE are surprisingly similar, indicating the nature of the temporal dependence structures in EUR/USD returns are time invariant.

### 5.3.2 Discretization effect

In the preliminary analysis of the intraday returns series, I confirm the existence of short-term dependence structures underlying the governing process. Further, I exclude the negative linear serial correlation as a possible main contributor to the detected structures. The previous preliminary analysis leaves us an unanswered question of what is the dominant source of the temporal dependence structures within intraday returns.

As guided by the summary of the properties of high-frequency financial returns given in section 2.2, other than the negative first-order correlation, the other three empirical properties of high-frequency returns are: extremely high kurtosis; intraday seasonality in volatility; and discretization. Since the former two properties are either only related to the empirical marginal distribution of the investigated data that is overlooked by ordinal analysis, or are expected to have a pronounced effect only on squared returns over long-term structures, this section focuses on investigating whether and how discretization brings temporal dependence relations into the return dynamics. If so, how strong is this structure compared to the overall temporal dependence structures exhibited in the observed return series? By simulating the discretization effect and comparing it with the original return series, the analysis leads to a striking conclusion that the discretization effect caused by limited decimal places price rounding conventions in financial asset price recording is the dominant contributor to the temporal dependence detected in the 10-minute EUR/USD returns under study.

The prices of all financial instruments are discretized. The discretization comes from the existence of the smallest possible price increment unit in the financial markets. The smallest possible price increment unit is often referred to as the tick-size in the stock and futures market or the pip in the exchange rate market. For instance, the tick-size for the stock and exchange rate is 1/1000th, except for the JPY related foreign exchange rates, which are normally quoted to three decimal places, and the US Government securities, which are quoted in 1/32nds of a dollar. Since the asset price can only jump in an integer multiple of a certain unit, the possible value of every corresponding return is subject to a pre-defined range that varies according to the price at the start of the return interval.

Discretization is an important feature of high-frequency (intraday) financial time series that is not generally an issue in lower frequency series (daily or longer than daily intervals). The discretization affects many fields in quantitative finance. Münnix et al. (2010) indicate that tick-size alters the distribution of financial return, especially for small return intervals, and distorts the calculation of correlations. Onnela et al. (2009) conclude that investors do not use all price fractions uniformly as allowed for by tick size, leading to a clustering of prices on certain fractions. Moreover, when the tick-to-price ratio is high, we expect more occurrence of zero returns. Chaos and the compass rose suggests price discreteness seriously distorts BDS test, thereby the evidence for chaos that has been found in stock returns might be due to price discreteness. Other than the above studies, existing studies about discretization on high-frequency returns are scarce and the issue has not been sufficiently addressed.

Discretization in price series imposes a non-linear temporal dependence structure in high-frequency return dynamics. Let  $X = \{x_t; t = 1, \dots, T\}$  denote the price series with tick size  $a$  and  $R = \{r_t; t = 1, \dots, T - 1\}$  represents the corresponding returns, that is,  $r_t = x_{t+1}/x_t - 1$ . Since price  $x_t$  can only jump in an integer multiple of a certain unit, the domain that restricts the possible values that the random variable  $r_t$  can take is not continuous, but restricted in the range  $R \in \{(x_t + I \times a)/x_t - 1 : I = \{0, 1, 2, \dots, \infty\}\}$ . I call the discretization in returns caused by the existence of tick size in price the discretization effect. Due to the discretization effect, the starting price  $x_t$  affects the distribution of its corresponding return  $r_t$ . Also, price  $x_t$  as the closing price of the previous period determines the value of return  $r_{t-1}$ . Since both current return and the lagged return are dependent on  $x_t$ ,  $r_t$  is dependent on  $r_{t-1}$  as well.

The strength of the dependence structure caused by the discretization effect differs for different return series even with the same tick size. Apart from the tick size, the strength of the discretization effect is also affected by the empirical distribution of the return series. Obviously, if the investigated returns are overall of relatively large values or there are more substantial returns (both positive or negative) in the time series compared to small ones, the effect from discretization will be more negligible. This is also the reason why discretization is less an issue in returns with longer-intervals such as daily returns, since in daily returns the ratio between the tick size and the price increments is too small to be influential.

Due to the varied level of discretization effect in each empirical return series, the discretization effect is best investigated through simulations so we can replicate the same level of discretization effect present in individual observed return series. I create a time series  $R^d = \{r_t^d; t = 1, \dots, T - 1\}$  that replicates the empirical distribution of the original return series. Simulated series  $R^d$  has the same level of discretization effect as in the original return series  $R$  but does not have any other temporal dependence structures, if these exist. The

time series  $R^d$  is generated by randomly shuffling the entries in the original return series  $R = \{r_t; t = 1, \dots, T - 1\}$  so that we destroy the temporal dependence structures in the return series (if existing), while conserving its empirical distribution. The shuffled return series is then converted back to a price series  $X^d = \{x_t^d; t = 1, \dots, T\}$  by following the relation  $x_{t+1}^d = (r_t^{\text{shuffled}} + 1) \times x_t^d$ , using the same initial price  $x_1^d = x_1$ . I round all the entries in the new price series  $X^d$  to four decimal places to simulate the discretization effect, and generate the simulated return series  $R^d = \{r_t^d; t = 1, \dots, T - 1\}$ , which replicates the discretization effect in the original data. To reduce the uncertainty and errors arising in the simulations, I repeat the process 500 times and generate 500  $R^d$  series for each year's return series. Figure 19(a) compares the average value of  $1 - \text{PE}_\tau^{D=3}$  of the 500 simulated return series and that of the original returns for each investigated period, both as functions of delays.

Figure 19(b) plots the average ordinal pattern probabilities for the simulated series at delays up to  $\tau = 5$ . From Figure 19(a), we can see that the value of  $1 - \text{PE}_\tau^{D=3}$  in the simulated discretized series  $R^d$  is significant at delays  $\tau = 1$  to  $\tau = 5$  and is very close to that in the original return series. This suggests the discretization effect cause a rather significant temporal dependence structure in the short-term dynamics of intraday return. Moreover, the striking close value of  $1 - \text{PE}_\tau^{D=3}$  and similar ordinal pattern distributions in  $R^d$  and that in the original returns suggests the discretization effect could be the primary contributor to the dependence structure detected in the return series, and this is true for all investigated periods.

One may wonder, if the average of bid and ask quotes is used to compute returns, would the discretization effect vanish? To resolve this question, I undertook the same analysis for the EUR/USD 10-minute returns using the average of bid and ask quote as the price series. The discretization effect is slightly weakened in this case, but is still significant and the dominant component of the temporal dependence structures in the resulting return series.

Our analysis suggests the discretization effect is the primary source of temporal dependence structures in the intraday return dynamics. An important implication of this discovery is that, in the study of high-frequency time series, more attention and caution is needed in treating and considering the existence of the discretization effect. More specifically, I believe there are at least two possible ways in which the discretization effect can distort or mislead standard analysis. First, discretization can lead to affirmative results in a non-linearity test and an independence test. Therefore, any model that eliminates the non-linearity or temporal dependence structure detected in the investigated high-frequency return series is not sufficient to prove that it is a good representation of the data under study. It may only be due to the fact that the fitting process distorts the discretization in the return series. Second, with

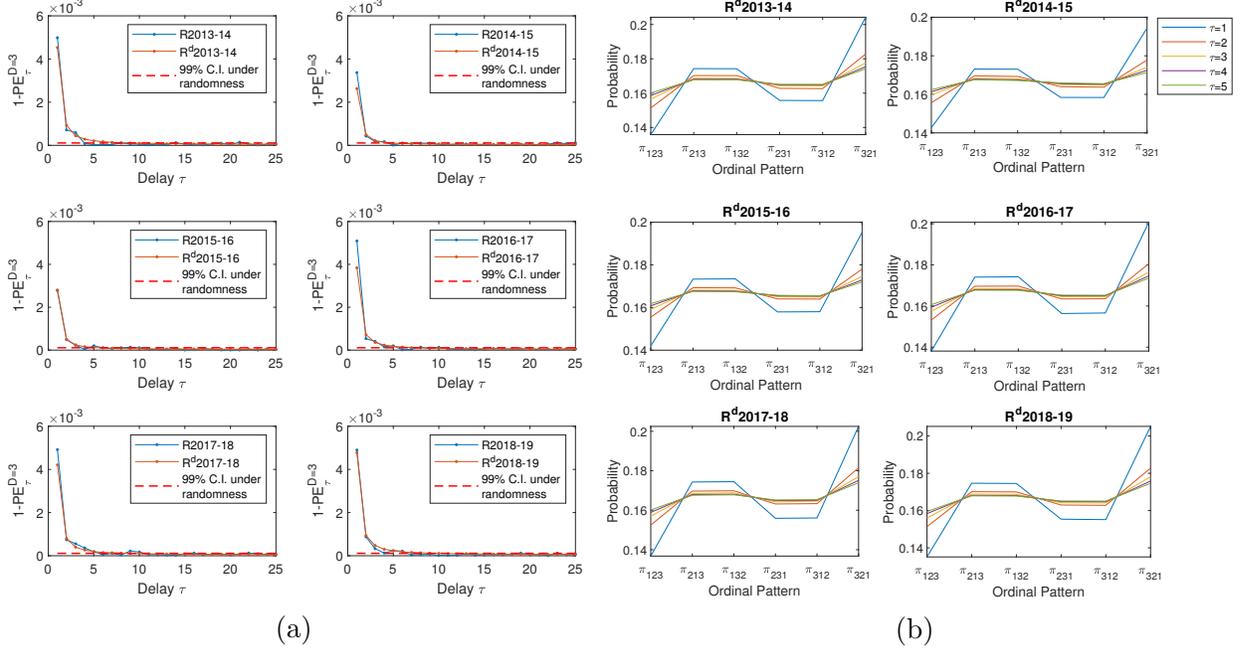


Figure 19: (a) Comparison between  $1 - PE_{\tau}^{D=3}$  as a function of delay on original EUR/USD 10-minutes return series and that on the simulated discrete return series  $R^d$ . (b) Probability of each ordinal pattern corresponding to  $1 - PE_{\tau}^{D=3}$  at delays where  $1 - PE_{\tau}^{D=3}$  is significant on simulated discrete return series  $R^d$ .

discretization as the dominating structure in the return series, other deterministic structures that may have exploitation potential for prediction can be hardly identified.

Since the discretization effect constitutes the primary component in the temporal dependence structures detected in the return series, the remaining question is whether there is any other determinism left in the intraday returns besides the discretization effect. One possible way to tackle this question is by converting the originally discrete return series to a continuously distributed one, and search for remaining structures. This can be done by adding a uniformly distributed drift that fluctuates within the range  $[-5 \times 10^{-5}, 5 \times 10^{-5}]$  to the original price series so that the price is no longer restricted to four decimal places. Then the new returns, **continuous-adjusted return (RC)**, can be computed by the continuously-adjusted price series.

Figure 20 compares the distribution of the original return series versus the continuous-adjusted returns for the six non-overlapping 1-year periods. It is clear in the histogram that after the conversion the originally discrete distributed intraday return series now has a continuous distribution. Figure 21 displays the plot of  $1 - PE_{\tau}^{D=3}$ ,  $PD_{\tau}^{D=3}$ , sample ACF and the ordinal pattern distribution at delay 1 for RC of each investigated period against increasing delays. The significant value of  $1 - PE$  and  $PD$  suggest that apart from the discretization effect, the investigated intraday return series has other temporal dependence structures. The

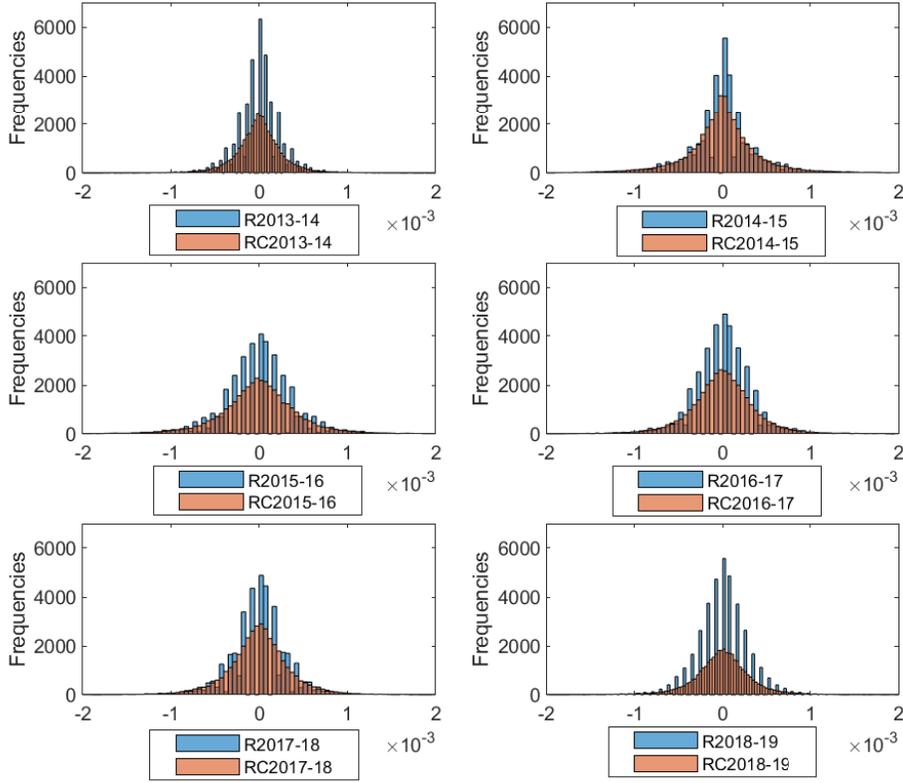


Figure 20: Comparison of distributions of original EUR/USD 10-minutes return versus after continuous-adjusted returns for the six none-overlapping 1 year periods.

plot of  $1 - PE_{\tau}^{D=3}$  indicates the existence of temporal dependence structures only at delay  $\tau = 1$ , whereas  $PD_{\tau}^{D=3}$  displays its superior capabilities to PE in unveiling the relatively weak but persistent temporal dependence. The PD plot on RC detected slowly diminishing dependence structures over delays up to  $\tau = 30$ . Additionally, based on the approximately equal probabilities of monotonic patterns and non-monotonic patterns shown in Figure 20 (c), the temporal structures within intraday returns after adjustments to continuous distribution can be characterized by the linear relation with negative autocorrelations, which seems to coincide with the first-order negative autocorrelation detected in the ACF plot. In the following sections, I will uncover the cause and the nature of the detected temporal structures within continuous-adjusted returns, and verify whether the detected structures can be sufficiently replicated by a linear process.

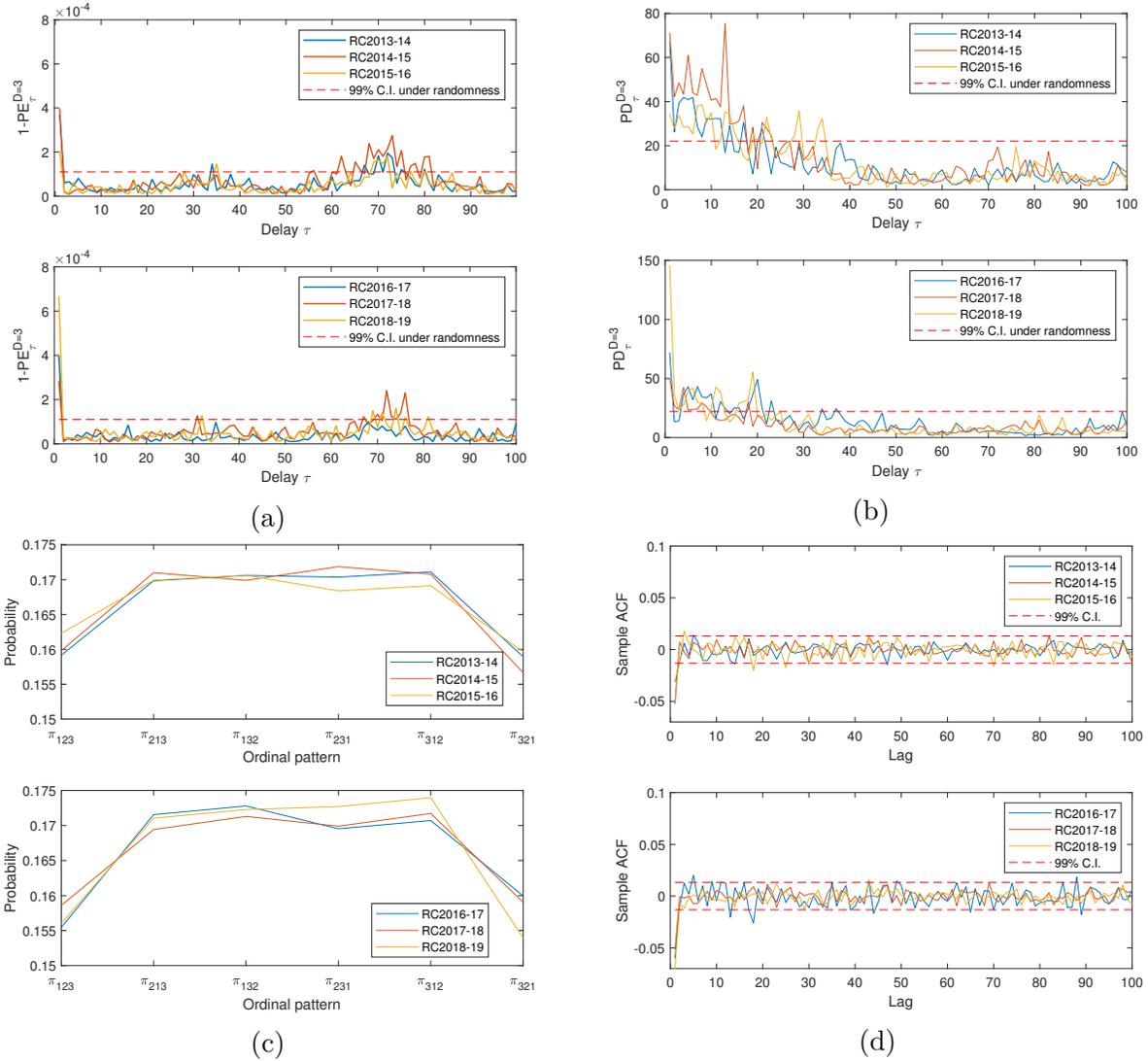


Figure 21: (a) Plot of  $1 - PE_{\tau}^{D=3}$  as a function of delay along with the 99% C.I. of  $1 - PE_{\tau}^{D=3}$  under randomness. (b) Plot of  $PD_{\tau}^{D=3}$  as a function of delay along with the 99% C.I. of  $PD_{\tau}^{D=3}$  under randomness. (c) Probability of each ordinal pattern in the computation of  $PE_{\tau}^{D=3}$  at  $\tau = 1$  (d) Plot of sample ACF as a function of lag along with the 99% C.I. of ACF under randomness. All measures are computed on six non-overlapping 1-year period continuous-adjusted return RC.

### 5.3.3 Temporal dependence in returns caused by structures in return magnitudes

The previous analysis indicates that while the discretization effect constitutes the primary component of the temporal dependence structures in the investigated intraday return dynamics, there are other remaining short-term and long-term dependence structures.

To search for the cause and nature of the detected structures within the intraday returns in addition to the aforementioned discretization effect, I adopt a novel approach that to the best of my knowledge has not previously been undertaken. In contrast to the conventional practice of analyzing return as an univariate series, I treat return as a product of its sign and its magnitude. I decompose the temporal dependence structures in return series into three separate components: structures in the sign of return; structures in the magnitude of return; and the bivariate dependence between sign and the magnitude of return. I investigate the contribution from each component respectively. Hopefully, by doing so we can uncover additional information about the dynamics underlying the complex process, and provide a more complete picture of the information by treating return as a whole.

This section briefly summarizes the dependence structures within return magnitude/volatility, and how they are reflected in return dynamics. The detailed analysis of the volatility dynamic is outlined in section 5.4. This section only briefly demonstrates the role of the temporal dependence structures within return magnitude(volatility) as a constituting component in return dynamics.

A simple simulation can indicate how temporal dependence structures within return magnitude/volatility are reflected within return dynamics. We generate a time series by assigning a random sign to the absolute magnitude of each RC series. The generated series denoted by `randsignRC` conserves the dynamics in the return absolute magnitudes/volatility and has the same number of negative and positive returns as the original observed returns, but loses any dependence structure within the sign of return or the interaction between the sign and the magnitude of returns. We generate 500 such time series for the six investigated periods, and compute the value of  $1 - PE_{\tau}^{D=3}$  and  $PD_{\tau}^{D=3}$  against increasing delays  $\tau$  on the simulated data. Figure 22 plots the average of  $1 - PE_{\tau}^{D=3}$  and  $PD_{\tau}^{D=3}$  of the 500 simulations for each investigated period, and compares it with that on the original RC series.

The plot given in Figure 22 shows that the simulated series `randsignRC` almost replicates the behaviour of  $1 - PE_{\tau}^{D=3}$  and  $PD_{\tau}^{D=3}$  on the original RC series for all considered delays except for  $\tau = 1$ . The similarity of  $1 - PE_{\tau}^{D=3}$  and  $PD_{\tau}^{D=3}$  plot on RC and `randsignRC` suggests the long-term dependence structure within time series RC is mainly attributed to the temporal

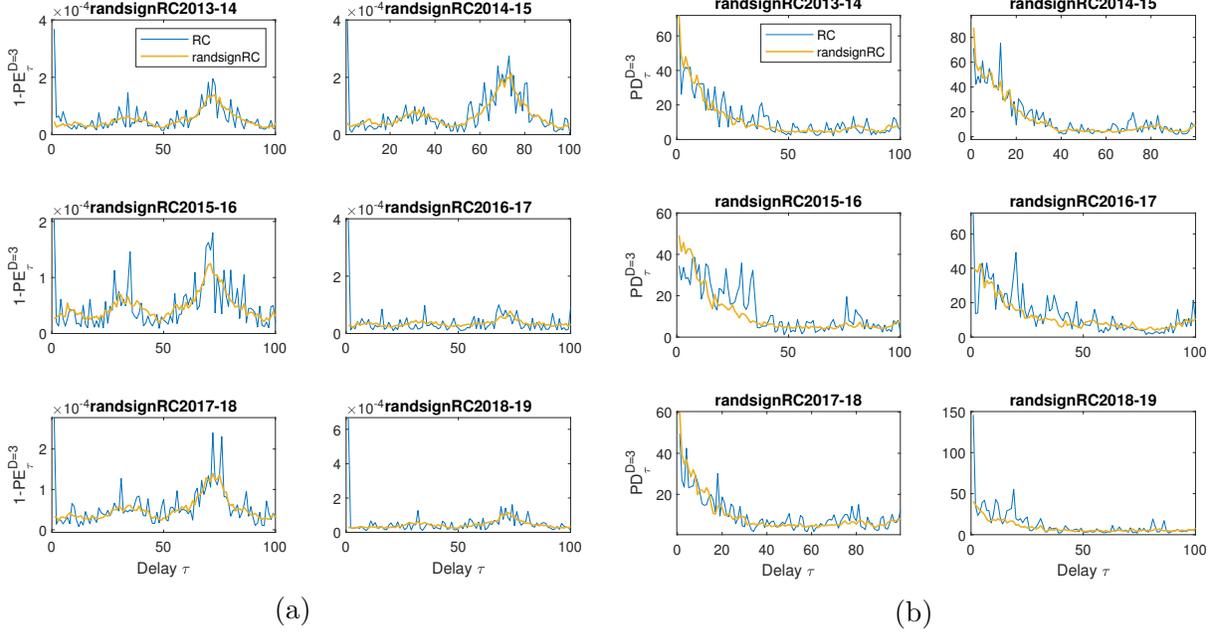


Figure 22: (a) Comparison between  $1 - PE_{\tau}^{D=3}$  on continuous-adjusted return RC and that on shuffled sign continuous-adjusted return randsignRC. (b) Comparison between  $PD_{\tau}^{D=3}$  on continuous-adjusted return RC and that on shuffled sign continuous-adjusted return randsignRC.

dependence structures present in the return magnitude/volatility dynamics. However, there is significant disparity in the  $PD_{\tau}^{D=3}$  and  $1 - PE_{\tau}^{D=3}$  measures at delay 1, especially prominent in  $1 - PE_{\tau=1}^{D=3}$ , between RC and that in randsignRC, indicating that the structure within return magnitude/volatility cannot explain the short-term temporal dependence structures in RC.

### 5.3.4 Signs of returns

This subsection studies temporal dependence solely in the signs of the return series. My main objective is to assess the predictability of the direction of price movements and the contribution from the determinism (if existing) within the sign of return into the temporal dependence structures in return as a whole. I will analyze the bivariate dependence relation between return signs and return magnitudes in the subsequent subsections.

In order to investigate the temporal dependence structures in the sign of returns, I construct a return sign series  $S = \{s_t\}$  by converting the continuously-adjusted return series RC into a binary series with two possible outcomes: - and +, where - indicates negative return and + indicates positive returns. Since the sign series is derived from continuous-adjusted return, there is no zero return.

In the original definition of PE, the investigated time series are partitioned into  $D$ -length segments and map into  $D!$  possible ordinal patterns so that we can obtain a discrete probability distribution. However, since the sign series is already a discrete time series, there is no need to compare the relative ranks of the entries in the constructed segments. Therefore we can compute its entropy directly. For instance, if the segment length is chosen to be 3, there are 8 possible patterns associated with the return sign series:

Pattern	Condition
$\pi_1^s$	$\{s_t = -, s_{t+\tau} = -, s_{t+2\tau} = -\},$
$\pi_2^s$	$\{s_t = +, s_{t+\tau} = +, s_{t+2\tau} = +\},$
$\pi_3^s$	$\{s_t = -, s_{t+\tau} = -, s_{t+2\tau} = +\},$
$\pi_4^s$	$\{s_t = +, s_{t+\tau} = +, s_{t+2\tau} = -\},$
$\pi_5^s$	$\{s_t = -, s_{t+\tau} = +, s_{t+2\tau} = +\},$
$\pi_6^s$	$\{s_t = +, s_{t+\tau} = -, s_{t+2\tau} = -\},$
$\pi_7^s$	$\{s_t = -, s_{t+\tau} = +, s_{t+2\tau} = -\},$
$\pi_8^s$	$\{s_t = +, s_{t+\tau} = -, s_{t+2\tau} = +\}.$

The normalized entropy for the return sign series  $S = \{s_t : t = 1, \dots, T\}$  with chosen vector length  $D$  and delay  $\tau$  is computed by

$$E_\tau^D(S) = \sum_{i=1}^{2^D} \frac{-p_\tau^D(\pi_i^s) \log(p_\tau^D(\pi_i^s))}{\ln 2^D},$$

where

$$p_\tau^D(\pi_i^s) = \frac{\#\{s_{t,\tau}^D | s_{t,\tau}^D \text{ has pattern } \pi_i^s\}}{N - (D-1)\tau}, \quad i = 1, \dots, 2^D,$$

and

$$s_{t,\tau}^D = (s_t, s_{t+\tau}, \dots, s_{t+(D-1)\tau}), \quad t = 1, 2, \dots, N - (D-1)\tau.$$

$E_\tau^D(S)$  share similar statistical properties with PE, with its value also bounded between 0 and 1, where 0 indicates a completely predictable dynamic, and 1 indicates completely stochastic dynamic. I use

$$1 - E_\tau^D(S)$$

to measure the level of departure from randomness of the dynamics underlying the observed sign series. Under the assumption that the time series  $S$  is iid,

$$2\{[N - (D-1)\tau] \ln 2^D\} [1 - E_\tau^D(S)]$$

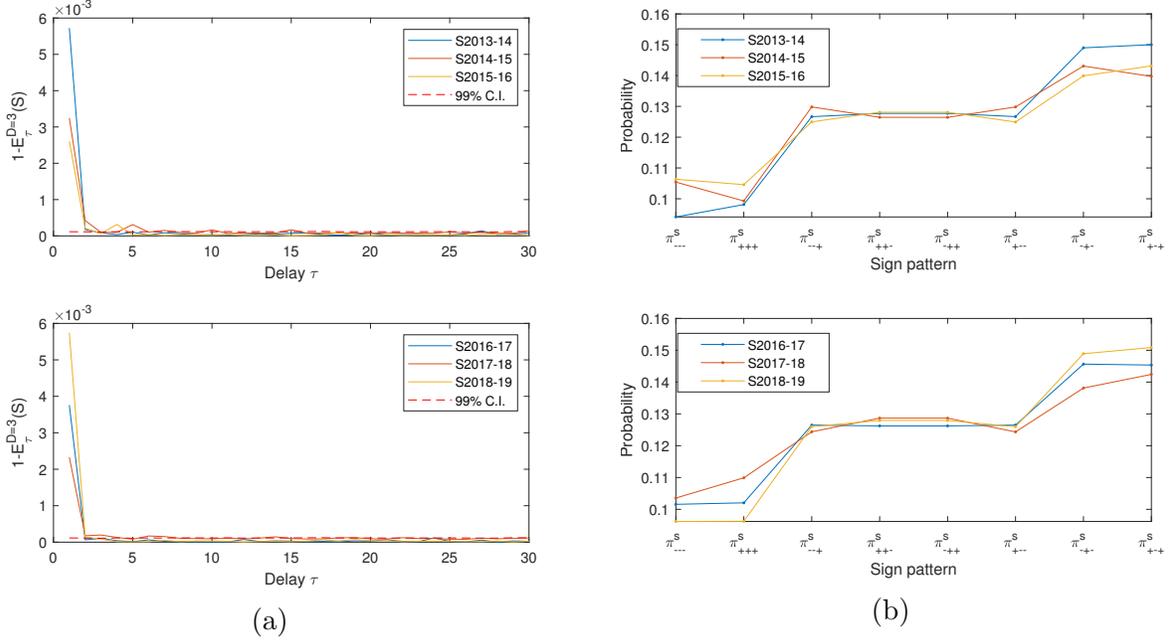


Figure 23: (a) Plot of  $1 - E_{\tau}^{D=3}(S)$  where  $S$  is the binary sign series of continuous-adjusted return of each six non-overlapping 1-year periods. (b) The plot of the probability distributions of the eight sign patterns corresponding to  $1 - E_{\tau=1}^{D=3}(S)$ .

is asymptotically  $\chi_{2D-1}^2$  distributed.

Figure 23(a) plots the standardized entropy on return sign series for increasing delays of each investigated period, along with the 99% C.I. under pure randomness. All the investigated periods display significant value of  $1 - E_{\tau}^{D=3}$  at and only at delay 1. Furthermore, Figure 23 (b) provides the probability distributions of the eight sign patterns for each investigated period that contribute to the value of  $E_{\tau=1}^{D=3}(S)$ . I find that the reason behind the significant  $1 - E_{\tau=1}^{D=3}(S)$  is the distinct disparities between the likelihood of persistent sign patterns ( $\pi_1^s = \{-, -, -\}, \pi_2^s = \{+, +, +\}$ ), patterns that alter sign once and the patterns that are constantly switching signs ( $\pi_1^s = \{-, +, -\}, \pi_2^s = \{+, -, +\}$ ). With the switching sign patterns being the most likely one and the persistent sign patterns being the least likely one, the result suggests the EUR/USD intraday returns tend to alter signs. In other words, positive returns are likely to be followed by negative returns, and vice versa. Even though the strength of sign dependence differs for different periods, the same rule applies to all investigated periods.

### 5.3.5 Bivariate dependence relation between signs of return and return magnitude

This section investigates the bivariate dependence relation between return signs and return magnitudes using the bivariate PE. I test the existence of the interaction between return sign and return magnitude. In addition, the positively dependent and negatively dependent combinations of sign and magnitude of returns reveal the likely and unlikely price movement habits of the high-frequency exchange rate.

To compute the bivariate PE between signs and the magnitude of returns we need to adjust the definition of original bivariate PE given in section 3.9. The reason is one of our investigated series returns, sign series  $S$ , is a categorical time series, but the original bivariate PE is designed for two continuous time series. In line with the definition of return sign's entropy given in the previous section, the joint patterns between return signs and return magnitude are adjusted to be the combination of sign patterns and ordinal patterns as follows if we choose  $D = 3$ :

Joint Pattern	$(a_s, b_s, c_s; a_y, b_y, c_y)$	Condition
$\pi_{1,1}^s$	$(-, -, -; 3, 2, 1)$	if $s_t = -, s_{t+\tau} = -, s_{t+2\tau} = -$ and $y_t > y_{t+\tau} > y_{t+2\tau}$ ,
$\pi_{2,1}^s$	$(+, +, +; 3, 2, 1)$	if $s_t = +, s_{t+2\tau} = +, s_{t+\tau} = +$ and $y_t > y_{t+\tau} > y_{t+2\tau}$ ,
$\pi_{3,1}^s$	$(-, -, +; 3, 2, 1)$	if $s_{t+\tau} = -, s_t = -, s_{t+2\tau} = +$ and $y_t > y_{t+\tau} > y_{t+2\tau}$ ,
$\vdots$	$\vdots$	$\vdots$
$\pi_{8,6}^s$	$(+, -, +; 1, 3, 2)$	if $s_{t+\tau} = +, s_t = -, s_{t+2\tau} = +$ and $y_{t+\tau} > y_{t+2\tau} > y_t$ .

There are total  $2^D \times D!$  joint patterns. With the changed number of possible joint patterns, the bivariate PE between return sign and return magnitude is modified into:

$$\text{BIPE}_\tau^s(S, Y) = \sum_{i=1}^{D!} \sum_{j=1}^{D!} \frac{-p_\tau^D(\pi_{i,j}^s) \log p_\tau^D(\pi_{i,j}^s)}{\log(2^D) \log(D!)},$$

where

$$p_\tau^D(\pi_{i,j}^s) = \frac{\#\left\{(s_{t,\tau}^D, y_{t,\tau}^D)' \mid (s_{t,\tau}^D, y_{t,\tau}^D)' \text{ has joint ordinal pattern } \pi_{i,j}^s\right\}}{N - (D-1)\tau}, \quad i = 1, \dots, 2^D; \quad j = 1, \dots, D!,$$

and

$$\begin{pmatrix} s_{t,\tau}^D \\ y_{t,\tau}^D \end{pmatrix} = \begin{pmatrix} s_t, s_{t+\tau}, \dots, s_{t+(D-1)\tau} \\ y_t, y_{t+\tau}, \dots, y_{t+(D-1)\tau} \end{pmatrix} : \quad t = 1, 2, \dots, N - (D-1)\tau$$

are the segment pairs derived from two equal length time series  $S = \{s_t; t = 1, \dots, N\}$  and

$Y = \{y_t; t = 1, \dots, N\}$ . The statistic measuring the dependence between return sign and squared return is

$$\text{HS}_\tau^D(S, Y) = \frac{[E_\tau^D(S) + \text{PE}_\tau^D(Y)] - 2\text{BIPE}_\tau^{sD}(S, Y)}{\min\{E_\tau^D(S), \text{PE}_\tau^D(Y)\}},$$

so that  $0 \leq \text{HS}_\tau^D(S, Y) \leq 1$ , with  $\text{HS}_\tau^D(S, Y) = 0$  iff time series  $S$  and  $Y$  are independent, and  $\text{HS}_\tau^D(S, Y) = 1$  iff  $Y$  is a fixed function of  $S$ . The individual joint pattern's contribution to dependence is measured by quantity  $d_\tau^D(\pi_{i,j}^s)$ , which is the ratio between the probability of any joint pattern  $\pi_{i,j}^s$  and the product of their respective marginal probabilities, that is,

$$d_\tau^D(\pi_{i,j}^s) = \frac{p_\tau^D(\pi_{i,j}^s)}{p_\tau^D(\pi_i^s) p_\tau^D(\pi_j)}, \quad i = 1, \dots, 2^D, \quad j = 1, \dots, D!.$$

Figure 24 plots the value of dependence measure  $\text{HS}_\tau^{D=3}$  between continuous-adjusted return signs  $S$  and return magnitude  $|\text{RC}|$  over increasing delays for the six investigated years, and the individual joint pattern correlation ratio quantified by  $d_\tau^{D=3}(\pi_{i,j}^s)$  for all 48 possible joint patterns at delays in which  $\text{HS}_\tau^{D=3}$  is most significant. Figure 24(a) indicates that there exists a bivariate dependence relation between return signs and magnitudes which is only significant at short-term dynamics, that is,  $\tau = 1$ . Moreover, the close resemblance of the plot of the correlation factor  $d_{\tau=1}^{D=3}(\pi_{i,j}^s)$  of each joint pattern for the returns recorded in the six different investigation periods suggests, despite the varied dependence strength, the nature and the form of the dependence relation between return sign and magnitude remains the same in time. Table 5 lists the top six positively and negatively dependent joint patterns contributing to the bivariate dependence relations.

To better demonstrate the implication of the bivariate dependence between return sign and magnitude, Figure 25 visually displays the most dependent joint return patterns and the associated resulting price movement patterns. By observing the plots given in Figure 25, the relation between return sign and magnitude reveals the micro-structure underlying the intraday exchange rate dynamics. The most positively dependent joint pattern  $\pi_{7,1}$ ,  $\pi_{5,1}$ ,  $\pi_{8,1}$  and  $\pi_{6,1}$  have a typical habit of price movement of the micro-structure of exchange rate market: after a relatively large price rise or drop, the exchange rate is likely to experience a reverse centre conversion movement over the next interval, followed by an even smaller adjustment. In addition, the subsequent smaller adjustment is more likely to be in the opposite direction to the first centre conversion movement as evidenced by the higher strength of dependence factor in joint pattern  $\pi_{7,1}$ ,  $\pi_{8,1}$ , compared to that in joint pattern  $\pi_{5,1}$  and

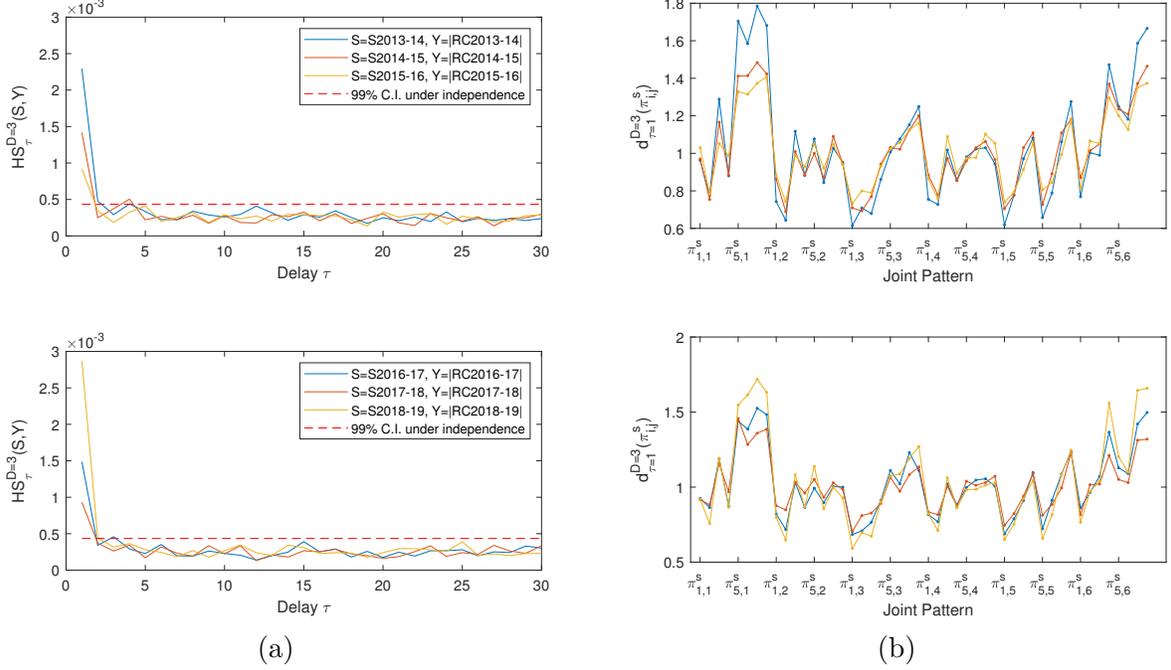


Figure 24: (a) Plot of dependence measure  $HS_{\tau}^{D=3}(S, Y)$  against delays, where  $S$  denotes the sign and  $Y$  is the magnitude of continuous-adjusted EUR/USD 10-minutes return for the six investigated 1-year periods. (b) Plot of the measure  $d_{\tau}^D(\pi_{i,j}^s)$  indicating the contribution to the overall dependence at delay 1 of each joint patterns between sign and magnitude of continuous-adjusted EUR/USD 10-minutes return for the six investigated 1-year periods.

$\pi_{6,1}$ . Moreover, such price patterns happen more often in price drop than in price rise as overall joint patterns  $\pi_{7,1}$ ,  $\pi_{5,1}$  are more positively dependent than  $\pi_{8,1}$  and  $\pi_{6,1}$ . The other two top positively dependent joint patterns  $\pi_{8,5}$  and  $\pi_{7,5}$  unveil another customary price movement pattern: before a relatively large price rise or drop, the exchange rate market is likely to prepare a big price movement with two prior reverse price movements of increasing magnitude and in opposite directions. In contrast, such market preparation happens more often before a large price increment than prior to a significant price decrement. From the most negatively dependent joint patterns plotted in the bottom two rows of Figure 25 (b), the least probable price movement pattern is to increase or decrease persistently for three consecutive intervals, especially if the significant price movement is in the middle.

Conclusively, I find the bivariate dependence relation between signs of return and return magnitude only exist over short-terms. Moreover, the detected dependence relations are persistent across different investigated periods. Some return sign and return magnitude ordinal pattern combinations are more likely to occur than others even after considering their marginal probabilities.

Table 5: The top six most positively dependent joint patterns and the top six most positively dependent joint patterns associated with the computation of the bivariate dependence measure  $HS_{\tau=1}^{D=3}(S, Y)$  where  $S$  is the sign and  $Y$  is the absolute magnitude of the continuously-adjusted EUR/USD 10-minutes return for six non-overlapping 1-year periods.

Top six most positively dependent joint patterns ( $d_{\tau=1}^{D=3}(\pi_{ij}^s)$ )					
RC2013-14	RC2014-15	RC2015-16	RC2016-17	RC2017-18	RC2018-19
$\pi_{7,1}^s$ (1.78)	$\pi_{7,1}^s$ (1.48)	$\pi_{8,1}^s$ (1.41)	$\pi_{7,1}^s$ (1.53)	$\pi_{5,1}^s$ (1.46)	$\pi_{7,1}^s$ (1.72)
$\pi_{5,1}^s$ (1.70)	$\pi_{5,1}^s$ (1.47)	$\pi_{8,5}^s$ (1.37)	$\pi_{8,5}^s$ (1.50)	$\pi_{8,1}^s$ (1.39)	$\pi_{8,5}^s$ (1.66)
$\pi_{8,1}^s$ (1.68)	$\pi_{6,1}^s$ (1.46)	$\pi_{7,1}^s$ (1.37)	$\pi_{8,1}^s$ (1.48)	$\pi_{7,1}^s$ (1.36)	$\pi_{7,5}^s$ (1.64)
$\pi_{8,5}^s$ (1.67)	$\pi_{7,5}^s$ (1.42)	$\pi_{7,5}^s$ (1.35)	$\pi_{5,1}^s$ (1.44)	$\pi_{8,5}^s$ (1.32)	$\pi_{8,1}^s$ (1.63)
$\pi_{7,5}^s$ (1.59)	$\pi_{8,5}^s$ (1.41)	$\pi_{5,1}^s$ (1.33)	$\pi_{7,5}^s$ (1.42)	$\pi_{7,5}^s$ (1.31)	$\pi_{6,1}^s$ (1.61)
$\pi_{6,1}^s$ (1.58)	$\pi_{8,1}^s$ (1.40)	$\pi_{6,1}^s$ (1.31)	$\pi_{6,1}^s$ (1.39)	$\pi_{6,1}^s$ (1.28)	$\pi_{4,5}^s$ (1.56)
Top six most negatively dependent joint patterns ( $d_{\tau=1}^{D=3}(\pi_{ij}^s)$ )					
RC2013-14	RC2014-15	RC2015-16	RC2016-17	RC2017-18	RC2018-19
$\pi_{1,3}^s$ (0.61)	$\pi_{2,3}^s$ (0.67)	$\pi_{1,3}^s$ (0.73)	$\pi_{1,3}^s$ (0.68)	$\pi_{1,3}^s$ (0.70)	$\pi_{1,3}^s$ (0.59)
$\pi_{1,6}^s$ (0.62)	$\pi_{1,3}^s$ (0.69)	$\pi_{1,6}^s$ (0.74)	$\pi_{1,6}^s$ (0.69)	$\pi_{1,6}^s$ (0.75)	$\pi_{2,2}^s$ (0.65)
$\pi_{2,2}^s$ (0.64)	$\pi_{2,2}^s$ (0.72)	$\pi_{2,2}^s$ (0.74)	$\pi_{2,3}^s$ (0.71)	$\pi_{2,3}^s$ (0.81)	$\pi_{1,6}^s$ (0.65)
$\pi_{5,6}^s$ (0.66)	$\pi_{5,6}^s$ (0.72)	$\pi_{2,4}^s$ (0.76)	$\pi_{2,2}^s$ (0.72)	$\pi_{5,6}^s$ (0.81)	$\pi_{5,6}^s$ (0.66)
$\pi_{3,3}^s$ (0.68)	$\pi_{1,6}^s$ (0.73)	$\pi_{2,1}^s$ (0.78)	$\pi_{5,6}^s$ (0.72)	$\pi_{1,5}^s$ (0.82)	$\pi_{3,3}^s$ (0.67)
$\pi_{2,3}^s$ (0.71)	$\pi_{2,1}^s$ (0.77)	$\pi_{3,3}^s$ (0.79)	$\pi_{3,3}^s$ (0.77)	$\pi_{2,4}^s$ (0.82)	$\pi_{2,3}^s$ (0.70)

### 5.3.6 Summary

This section investigated the temporal dependence exhibited in the 10-minute EUR/USD return as an illustrative example of the ordinal pattern based analysis on real-world financial time series.

I used the PE and PD measures introduced in chapter 3 to detect and decompose the temporal dependence structure in the EUR/USD 10-minute returns dynamics. I found the discretization effect is the primary source of short-term temporal dependence structures in 10-minute return dynamics. The existence of tick-size, the smallest unit of increments and decrements in a financial instrument, leads high-frequency returns to follow a discrete distribution. The discretization imposes a non-linear dependence structure between adjacent returns. I found that the temporal dependence caused by the discretization is the most significant source contributing to the non-randomness detected in the high-frequency EUR/USD return series over short-term dynamics.

Such non-linear structure cannot be detected by conventional tools, such as the ACF,

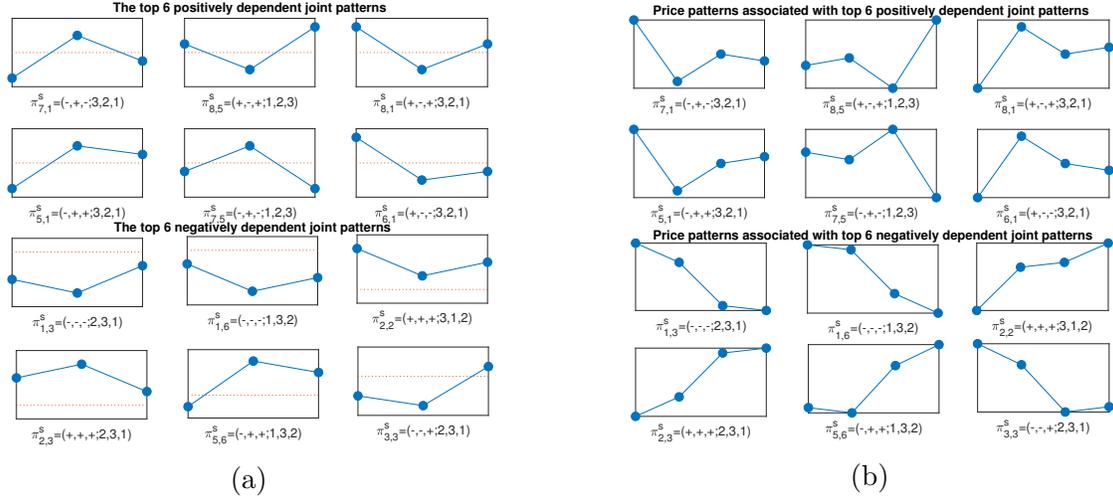


Figure 25: (a) Top 6 positive and negative dependent return sign and magnitude joint patterns. (b) Price movement patterns associated with the top 6 positive and negative dependent sign and magnitude joint patterns.

since the ACF is mainly aimed at capturing linear relations. However, measures based on ordinal patterns are not restricted by any pre-defined functional forms of structures and can successfully detect this nonlinear structure that ACF neglects. The superior robustness in detecting non-linear structures is the most important reason why PE and PD measures are important supplements to the existing tool kit in the analysis of financial time series.

Empirically, the discovery of the discretization effect as a dominant source of temporal dependence structure in high-frequency returns has significant implications. The analysis results stress the importance of considering discretization in studying the dynamic underlying high-frequency financial returns, and calls for extra caution when using high-frequency returns in many types of analysis, including but not limited to assessing market predictability, detecting the existence of chaos dynamics or investigating the sufficiency of modelling procedures. This is because some statistics and hypothesis tests used in the above areas might be seriously distorted by the discretization effect, or be erroneously treated as evidence of something that creates predictability.

Besides discretization, there are other sources contributing to the temporal dependence structures within our investigated intraday returns. However, compared to the structure caused by the discretization effect, the strength of the remaining structure in returns is relatively weak. This explains why intraday returns display a non-random and non-linear nature but have marginal predictability. By treating return as a compound product of collective structures within return sign, return magnitudes and the bivariate interaction between return sign and return magnitude, ordinal analysis reveal many interesting findings.

I found that the detected relatively weak long-term dependence present within the investigated returns are attributed to the structures exhibited in the volatility. Moreover, the high-frequency returns have a tendency to alter signs. In other words, positive return is likely be followed by negative return and vice versa. Consecutive positive returns and negative returns are unlikely to occur. The switching return sign tendency is the cause of the negative first-order autocorrelations that appeared in the ACF plot of investigated returns. Since positive returns are likely to be followed by negative returns and vice versa, ACF reflects these negative autocorrelations in returns. According to my findings, the switching return sign tendency is a persistent feature for every investigated period. However, the strength varies from time to time.

In addition, I also discovered a bivariate dependence relation between sign of return and return magnitude. Bivariate PE analysis reveals the interaction between return sign and return magnitude exists and only exists over the short-term. The various degrees of dependence between different return sign and return magnitude pattern combinations reveals that certain price movement patterns are more likely to occur than others. These results shed light on the micro-structures of financial market behaviours. The likely and unlikely price patterns reflect currency market customary trading habits. More specifically, after a relatively large price rise or drop, the exchange rate is likely to experience a reverse centre conversion movement over the next interval and then an even smaller adjustment. In addition, the smaller adjustment is more likely to be in the opposite direction to the first centre conversion movement, and such price pattern happens more often in price drop than in price rise. Before a relatively large price rise or drop, the exchange rate market is likely to prepare a major price movement by starting with two prior reverse price movements of increasing magnitude and opposite directions. This likely price pattern happens more often before a large price increment than a large price decrement. In contrast, the intraday exchange rate is very unlikely to increase or decrease persistently for three consecutive intervals, especially when there is a relative significant price movement in the middle.

#### **5.4 Temporal dependence of EUR/USD 10-minute squared returns and 1-hour realized volatilities**

The term volatility generally represents the level of variation of return for a given financial asset and is considered as a measure of risk for the associated asset. In the area of financial time series analysis, volatility is often termed to as the conditional variance of returns of the target asset, formally denoted as  $\text{Var}(r_t|F_{t-1})$  where  $F_t$  represents all historical information available prior to time  $t$ . Volatility is very important for portfolio selection, option pricing

and risk management. However  $\text{Var}(r_t|F_{t-1})$  is a latent variable that cannot be directly observed. Practitioners normally use the squared returns as the approximation of the true unobservable volatility. Alternatively, for return over a relatively long interval, such as the daily interval, the realized return that aggregates the squared returns with higher frequencies within the given intervals can also be employed as the proxy for volatility.

Since volatility is an unobservable quantity, I adopt the conventional practice of using squared returns and realized returns as its proxy. I conduct ordinal analysis on the 10-minute squared returns and 1-hour realized returns to investigate their respective temporal dependence structures. Similar to the ordinal analysis procedures carried out in the previous section on return series, I first detect the existence of temporal dependence structures in our investigated series by plotting  $1 - \text{PE}_\tau^{D=3}$  as a function of delay  $\tau$ . As seasonality in volatility is a well-documented property in high-frequency financial returns, as covered in section 2.2.7, as well as the most distinctive features exhibited in the PE plot and ACF plot on our investigated volatility series given in Figure 26, in order to search for other temporal dependence structures and avoid possible distortion of seasonal patterns on our analytical tools, I employ two commonly-used seasonal filters to remove the seasonality. I choose to consider more than one seasonal filters because seasonal filters may introduce artificial structures while removing periodicities. The usage of two completely different functional form seasonal filters helps to verify that the detected structures are not caused by the effect of deseasonalizing procedures but are intrinsic features underlying the dynamics of the original data. I compare the performance of the two seasonal filters and discover a possible reason for their imperfect removal of periodicities. After removing seasonality, I conduct ordinal analysis on deseasonalized 10-minute squared returns and deseasonalized 1-hour realized volatility to investigate the temporal dependence structures in squared returns besides seasonality. The PE and the associated ordinal analysis reveal the strength of temporal dependence in the deseasonalized squared returns and the deseasonalized realized volatility, how their temporal dependence structures diminish over increasing delays/lags and promising models to replicate their structure.

Note that in ordinal pattern based analysis such as PE computation and PD computation, squared returns and absolute returns are equivalent. Therefore in the subsequent analysis, the plots of PE and PD on squared returns are compared with the ACF plot of absolute returns in some instances. In fact, all monotonic transformation of the data set under study has no impact on the PE and PD measures.

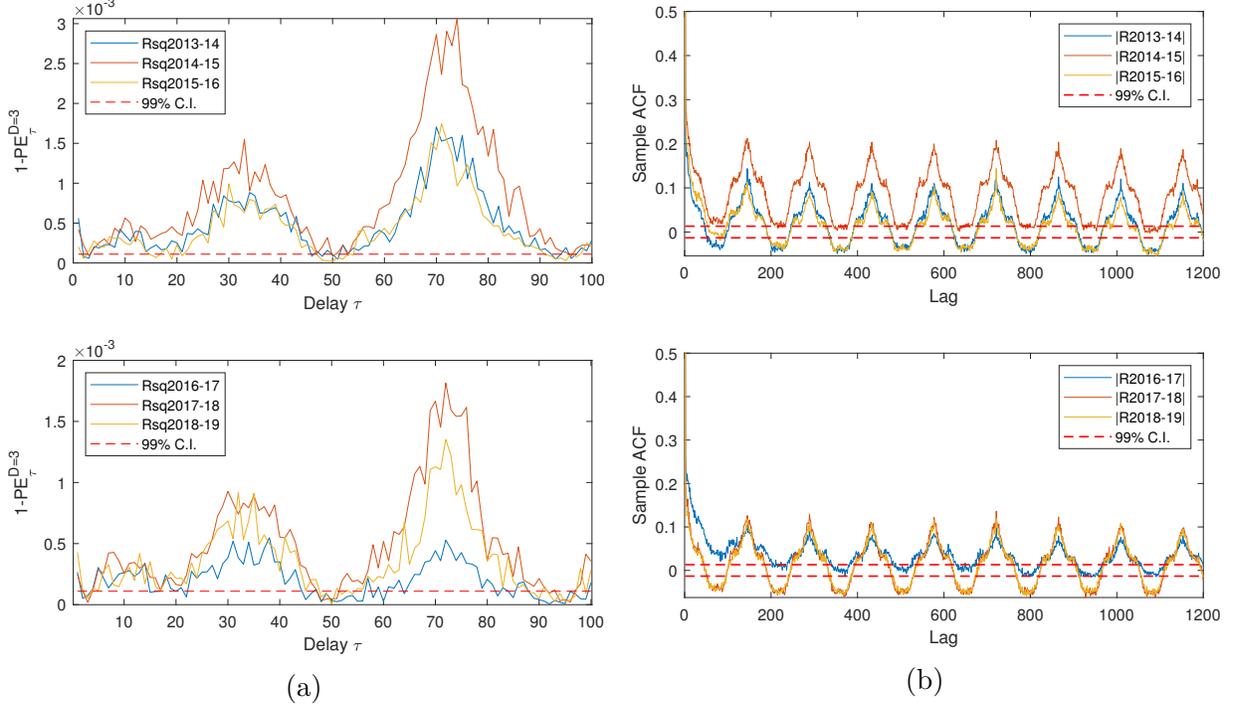


Figure 26: (a) Plot of  $1 - PE_{\tau}^{D=3}$  over delays  $\tau$  for six non-overlapping 1-year period 10-minutes EUR/USD squared returns. (b) Plot of sample ACF on six non-overlapping 1-year period 10-minutes EUR/USD absolute return series.

#### 5.4.1 Preliminary analysis on 10-minute EUR/USD squared returns

I conduct the preliminary analysis on 10-minute intraday EUR/USD squared returns by plotting  $1 - PE_{\tau}^{D=3}$  and sample ACF over increasing delays/lags on the investigated data, along with the 99% C.I. of statistic  $1 - PE_{\tau}^{D=3}$  and ACF under randomness. Unlike the return dynamic that mainly exhibits short-term structure, the temporal dependence structure in squared returns shows relative weak dependence structures in short delays but strong periodicities over long-term delays. For the six different considered periods, the strength of the periodicities varies significantly for different investigated periods.

The reason I did not compute PD measure on squared returns is shown in section 3.6, which demonstrate that the PD measure is susceptible to non-stationarity. The PD plot on squared returns exhibits strong periodicities and can therefore be very unreliable. But PE is more robust to the effect of non-stationarity, especially at short-term delay. Therefore, prior to the removal of the seasonality in the squared return, the preliminary analysis on 10-minute EUR/USD squared returns is mainly based on the PE measure.

### 5.4.2 Periodicity in squared returns

As mentioned in section 2.2.7, intraday periodicity in volatility is a commonly observed feature in high-frequency financial returns. The periodicity in intraday foreign exchange rate returns is a 24 hours pattern mainly attributed to the differences in trading times in the global foreign exchange markets. Also, certain financial news, such as interest rate announcements, are regularly released at the same time, which may cause some periodicity in the intraday volatility.

Since periodicity is the most distinct feature observed in our investigated squared return dynamics, without removing the periodicity, we cannot further investigate the existence and the nature of other deterministic structure in the squared returns besides seasonality. The most direct way of dealing with periodicities is to use seasonal adjustment filters. There are a number of candidate seasonal adjustment filters that are commonly used in the analysis of financial time series.

Andersen & Bollerslev (1998) provide a robust methodology, namely the **Flexible Fourier Form (FFF)** method, for capturing the distinct seasonal volatility components. This involves formulating a deterministic intraday volatility pattern to capture high-frequency volatility periodicity, and imposing a predetermined volatility response pattern associated with calendar and other effects. The FFF method is parsimonious and allows for smooth volatility dynamics. It is rigid in functional form, and imposes a smooth cyclical pattern in the characterization of intraday periodicity.

A second way to estimate the seasonal pattern is advocated by Taylor & Xu (1997). They use the averages of squared returns to estimate the seasonality patterns for different time of the week. Andersen & Bollerslev (1997) adjust their method by using the logarithm of the squared returns to help estimate seasonal patterns.

I use the above mentioned two deseasonalizing procedures: the **Averages of Squared Returns (ASR)** method and the FFF method of Andersen & Bollerslev (1997) to construct the seasonal adjustment filters and remove the periodicities exhibited in our investigated series.

Since our return series is recorded in **Greenwich Mean Time (GMT)** and does not reflect the daylight saving schemes followed by most of the major financial markets, I adjust the recorded time of the observed returns to adapt to the US daylight saving period in the estimation of seasonal adjustment filters. I choose to follow the US daylight saving period mainly because our investigated exchange rate pair EUR/USD is most heavily traded in the US and UK markets, and the daylight saving period in the UK almost overlaps with that in the US. Hence there is little difference between them.

The simplest and most intuitive way of removing periodicities is to use the averages of squared returns to estimate the seasonality patterns for different time of the week advocated by [Taylor & Xu \(1997\)](#). Let  $R_{d,n}$  denote the  $n$ th intraday return on day  $d$  in a week. The intraday seasonality for the  $n$ th intraday interval on day  $d$  ( $d = 1$  represents Monday,  $d = 2$  represents Tuesday...) is estimated by

$$(s_{d,n}^{ASR})^2 = \frac{1}{S_d} \sum_{s=1}^{S_d} (R_{s,n} - \bar{R})^2,$$

where  $S_d$  is the number of weeks in the investigated periods and  $\bar{R}$  is the sample mean of the intraday returns. Similarly, the deseasonalized return is obtained by

$$DR_{d,n}^{ASR} = \frac{R_{d,n}}{s_{d,n}^{ASR}}.$$

[Andersen & Bollerslev \(1997\)](#) suggest estimating the seasonalities in the intraday volatilities by using different frequencies of sine and cosine functions, namely, the Flexible Fourier Form (FFF) method. FFF method is the most commonly used approach for deseasonalizing periodic financial time series. They assume

$$R_{d,n} - \bar{R}_{d,n} = \frac{\sigma_d}{N_d^{1/2}} \times s_{d,n} \times Z_{d,n},$$

where  $R_{d,n}$  denotes the  $n$ th intraday return on day  $d$  in a week ( $d = 1$  represents Monday,  $d = 2$  represents Tuesday...),  $\bar{R}_{d,n}$  are the expected intraday return and  $Z_{d,n}$  is an iid random variable with mean zero and variance one,  $N_d$  is the number of intraday returns in one day,  $\sigma_d$  represents daily volatility and  $s_{d,n}$  is intraday seasonality. By squaring both sides of the above equation, taking logs, approximating  $\bar{R}_{d,n}$  by the sample mean  $\bar{R}$ , and replacing  $\sigma_t$  by a daily volatility estimate  $\hat{\sigma}_t$ , we have the following expression

$$2 \log \frac{|R_{d,n} - \bar{R}|}{\hat{\sigma}_d / N_d^{1/2}} = \log(s_{d,n}^2) + \log(Z_{d,n}^2).$$

Then [Andersen & Bollerslev \(1997\)](#) construct a FFF regression model to estimate the intraday seasonality factor:

$$f_{d,n} = \alpha + \delta_1 n + \delta_2 n^2 + \sum_{l=1}^L \lambda_l I_{l;d,n} + \sum_{p=1}^P \left[ \delta_{c,p} \cos \left( \frac{p2\pi}{N_d} n \right) + \delta_{s,p} \sin \left( \frac{p2\pi}{N_d} n \right) \right] + \varepsilon_{d,n},$$

where  $f_{d,n} = 2 \log\{|R_{d,n} - \bar{R}|\} + \log(N_d) - 2 \log(\hat{\sigma}_d)$ . The regression model includes the intercept  $\alpha$ , the linear and the quadratic function of the time of a day  $n$ , the indicator variables  $I_{l;d,n}$ , which are used to control the outliers, and an error term. The choice of  $P$  generally ranges from 3 to 6, I choose  $P$  to be 5, although all specifications lead to a similar smooth fit. To avoid using future information that is needed to construct estimation of daily volatility  $\hat{\sigma}_d$ , I assume there is no interaction between the daily volatility and the seasonal volatility pattern. Hence, the term  $2 \log(\hat{\sigma}_d)$  is removed from the definition of  $f_{d,n}$ . The estimate of intraday seasonality  $\hat{s}_{d,n}^{FFF}$  is obtained by  $\exp(\hat{f}_{d,n}/2)$ . To filter the periodicities of the intraday return, I simply divide the returns by the corresponding seasonality factor

$$DR_{d,n}^{FFF} = \frac{R_{d,n}}{s_{d,n}^{FFF}},$$

where  $DR_{d,n}^{FFF}$  denotes the deseasonalized return using the FFF filtering method.

Since I am going to investigate the forecasting performance of various methods in predicting squared returns and realized volatility in the subsequent sections, I only use the prior 43 weeks data for the estimation of seasonality in each investigation period to prevent using the data in the evaluation subsets.

Figure 27 records the estimated weekly seasonal pattern of the two seasonal adjustment filters, and the ACF plot of the original squared returns and the deseasonalized squared returns. It can be seen from the estimated seasonal patterns that the FFF approaches forms a smooth function of the seasonality, whereas the seasonal pattern constructed from ASR displays more irregular sharp peaks and troughs. Despite the differences in the functional form of the seasonal filter, it can be seen in the ACF plot that the performance of the two selected seasonal filters is very similar. They both remove most of the seasonality exhibited in the squared returns, but for some investigation periods there is still some remaining periodicity left in the deseasonalized squared returns, including in the year 2013-14, 2017-18 and 2018-19.

### 5.4.3 Discretization effect in squared return dynamics and on deseasonalization procedures

Section 5.3.2 demonstrated that the discretization effect is the dominant source of the temporal dependence structure detected in the return series. It makes us wonder whether the discretization effect also imposes a dependence structure in squared return dynamics. Prior to further analysis of the temporal dependence structures in squared returns in addition to seasonality, this section addresses the issues related to the discretization effect in the squared return dynamics. I find that price discretization introduces a significant but weaker dependence

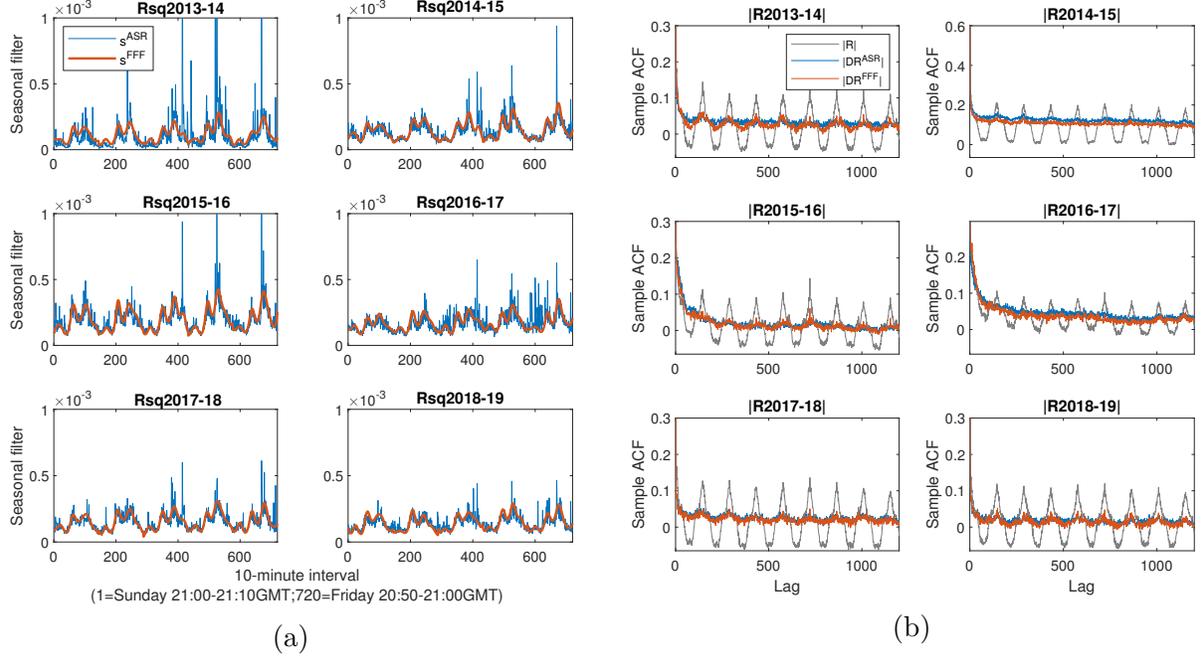


Figure 27: (a) Estimated weekly seasonal filters using ASR and FFF methods applied on six non-overlapping 1-year period 10-minute EUR/USD squared returns. (b) Comparison of sample ACF plot on original absolute return and deseasonalized returns after ASR and FFF seasonal filter respectively.

structure in squared return dynamics compared to return dynamics. In addition, the non-linear temporal dependence structure caused by the discretization creates an obstruction for the seasonal adjustment filter to work properly, especially when the time series have relatively strong temporal dependence structures. Moreover, despite seemingly weak strength of the discretization effect in squared returns, even after deseasonalizing procedures, the impact from it cannot be fully eliminated and may distort the PE and the other related ordinal pattern analysis. For that reason, I need to eliminate the discretization in the squared return dynamic before investigating the temporal dependence structures underlying the deseasonalized squared returns.

Figure 28 (a) plots the empirical distribution of the absolute value of investigated return  $|R|$  and the distribution of the absolute return after adjusting to follow a continuous distribution according to the procedures I proposed in section 5.3.2. The histogram clearly shows that the distribution of the return magnitude is discrete. In order to investigate the impact of discretization effect on squared returns, I adopt the approach proposed in section 5.3.2 to replicate the same level of discretization in the observed return. The simulated discretized return series is denoted by  $R^d$ . I compute the value of  $1 - \text{PE}_\tau^{D=3}$  on the squared of the  $R^d$  to reveal the contribution of discretization effect into the temporal dependence structure

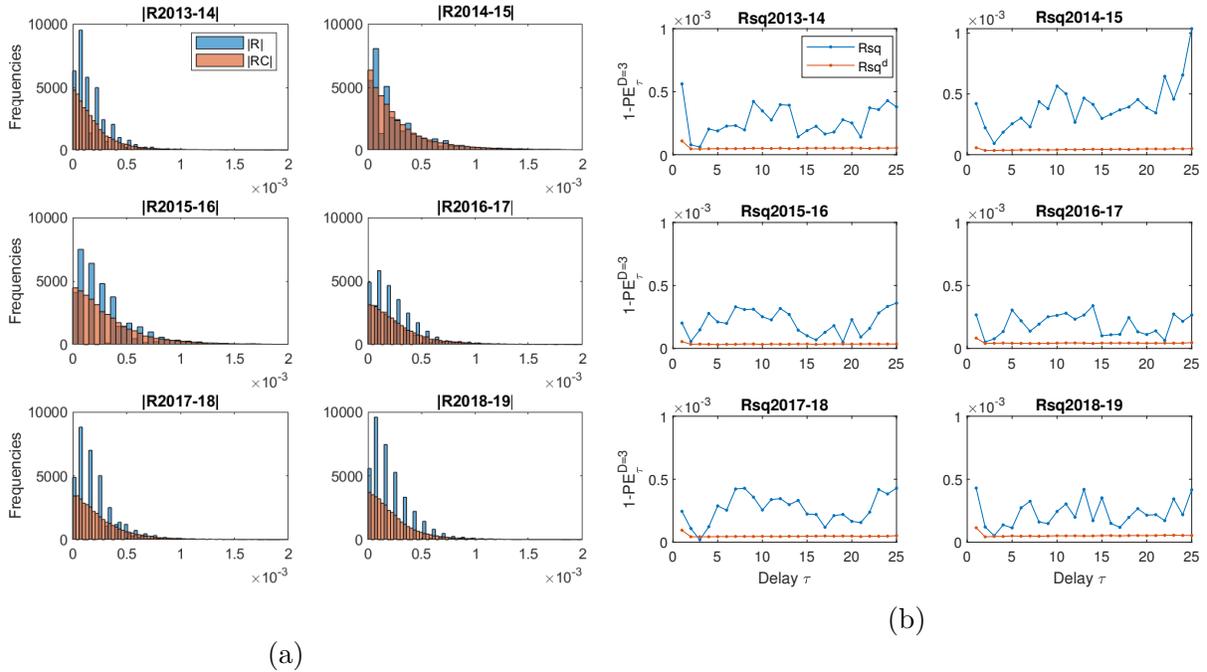


Figure 28: (a) Distribution of absolute Return versus continuous-adjusted absolute return. (b) Comparison between  $1 - PE_\tau^{D=3}$  on squared returns and simulated discretized squared returns. It worth noting that absolute return and squared return are equivalent to PE.

on squared return dynamics. Figure 28 (b) displays the PE plot on  $Rsq^d = (R^d)^2$  and that on original 10-minutes squared returns, and suggests the discretization effect introduces a significant but much weaker temporal dependence structure in squared returns dynamics compared to that in the return dynamics.

The discretization effect not only constitutes a small portion of the temporal dependence structures in the squared return dynamics, but can also hamper the deseasonalization procedures. In order to illustrate how the discretization effect hampers the proper performance of the deseasonalization procedures, I generate four periodic return time series  $Rsim1$ ,  $Rsim2$ ,  $Rsim3$  and  $Rsim4$ , where  $Rsim1$  and  $Rsim3$  follow continuous distributions and  $Rsim2$  and  $Rsim4$  exhibit discretization. The FFF seasonal filter is employed to deseasonalize the simulated series. The detailed specifications of the four simulated series are as follows.

- $Rsim1$  (random return  $\times$  seasonality):  $Rsim1$  is constructed by multiplying the formerly estimated seasonal pattern  $s^{FFF}$  on squared returns  $Rsq2013-14$ , with a random series  $r$  generated from the empirical distribution of the corresponding  $DR^{FFF}$ .
- $Rsim2$  (random return  $\times$  seasonality with discretization):  $Rsim2$  follows the same generating procedures as in  $Rsim1$ , except I incorporate the discretization into the simulated

series by converting Rsim1 back to a price series and round all prices to four decimal places, then convert back to simple return.

- Rsim3 (GARCH return×seasonality): Rsim3 is generated by multiplying the same seasonal pattern  $s^{\text{FFF}}$  with a GARCH(1,1) modelled return series  $r_{\text{GARCH}}$ .
- Rsim4 (GARCH return×seasonality with discretization): In Rsim4, discretization effect is incorporated to the generating procedures of series Rsim3.

The performance of the seasonal filters is assessed by (a) the existence of remaining periodicities in the ACF plots in the generated series after deseasonalization, and (b) how closely the estimated seasonal filters coincide with the “true” seasonal patterns I incorporate into the simulations. The distortion from the discretization effect on the dependence structures is measured by comparing the PE and PD plots on the temporal dependence structure on the deseasonalized series and those on the original simulations prior to the inclusion of seasonality.

Figure 29 (a) provides the ACF plot comparing simulated series and deseasonalized series of the above four designed scenarios. By comparing the performance of the seasonal filters on Rsim1 and Rsim2, as well as on Rsim3 and Rsim4, it is clear that seasonal filters perform worse in the simulated return series with discretization. Subplot (b) in Figure 29 indicates that the estimated seasonal filter on series with the discretization effect tends to overestimate the peaks and underestimate the local minimal of the periodic patterns. This distortion effect on the seasonal filter is especially severe in the time series with significant baseline serial dependence structures as shown in Rsim4. Moreover, from subplot (c) and (d) in Figure 29, we can see that despite the marginal remaining periodicity in the deseasonalized series in the ACF plot, the remaining periodicities are distinctly captured and reflected on the PE and PD plots. Even after the deseasonalization procedures, the discretization effect is still not fully diminished, evidenced from the significant disparity in  $1 - \text{PE}_\tau^{D=3}$  and  $\text{PD}_\tau^{D=3}$  between deseasonalized |Rsim1| and deseasonalized |Rsim2|, and between deseasonalized |Rsim3| and deseasonalized |Rsim4|, especially at delay 1.

#### 5.4.4 Temporal dependence structure in deseasonalized squared returns

In this section, I implement PE and ordinal analysis on the deseasonalized squared returns to uncover the temporal dependence structures in return magnitude/volatility in addition to seasonality.

I showed in the previous section that the discretization effect can lead to imperfect performance of seasonal filters and distortions on PE and PD analysis. To avoid unwanted

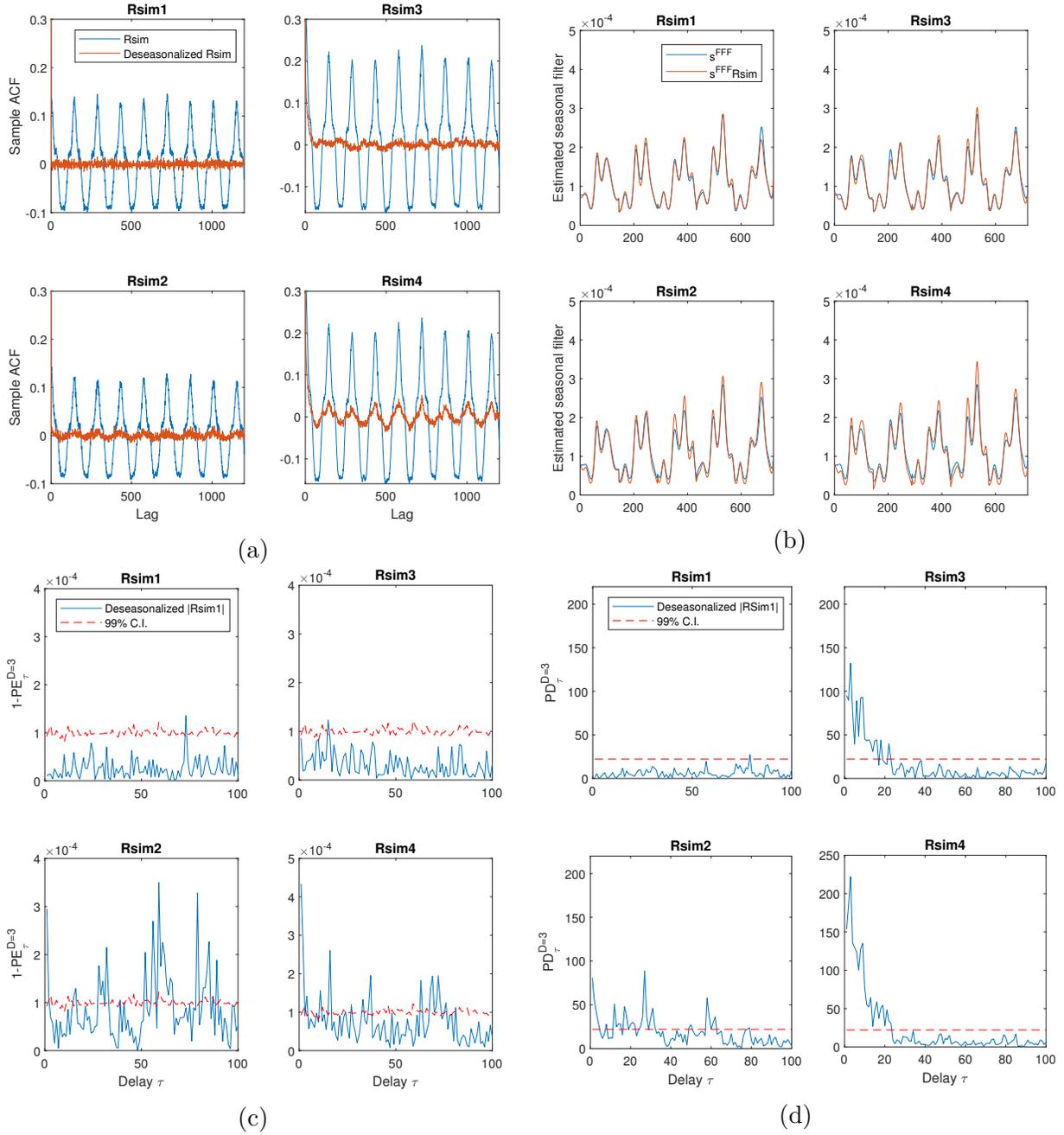


Figure 29: (a) Sample ACF on simulated periodic return series Rsim1, Rsim2, Rsim3 and Rsim4, and that on deseasonalized Rsim1, Rsim2, Rsim3 and Rsim4 after the FFF seasonal filtering procedure. (b) The comparison of estimated seasonal pattern and the “true” seasonal pattern incorporated into the simulation. (c) Plot of  $1 - PE_{\tau}^{D=3}$  over increasing delays on deseasonalized Rsim1, Rsim2, Rsim3 and Rsim4. (d) Plot of  $PD_{\tau}^{D=3}$  over increasing delays on deseasonalized Rsim1, Rsim2, Rsim3 and Rsim4.

influences from the discretization effect, I choose to conduct the temporal dependence investigation on deseasonalized continuous-adjusted squared returns, denoted by DRsqC. The deseasonalized continuous-adjusted squared returns are obtained by applying the seasonal filters on continuous-adjusted returns following the continuous conversion procedure provided in section 5.3.2. In fact, the deseasonalized continuous-adjusted squared returns are very close to those on deseasonalized original returns. On average, their difference is only around 0.1%.

In order to investigate the temporal dependence structures in continuous-adjusted deseasonalized squared returns, I compute the measures  $1 - \text{PE}_\tau^{D=3}$ ,  $\text{PD}_\tau^{D=3}$  and sample ACF as functions of increasing delays, and the ordinal pattern distribution at delay one on DRsqC for the six individual investigated periods. They are all plotted in Figure 30. Since the PE plots and PD plots of the deseasonalized squared returns after the two seasonal filters are very similar, the choice of the FFF and ASR seasonal adjustment methods makes little difference in the dynamics of deseasonalized intraday volatilities. The plots show that the value of  $1 - \text{PE}_\tau^{D=3}$  fluctuates around a marginal significant level from small delays to large delays, but  $\text{PD}_\tau^{D=3}$  is clearly far from the insignificant level and diminishes slowly from short-term to long-term. The slow diminishing rate of  $\text{PD}_\tau^{D=3}$  suggests there are long-range temporal dependence structures in the deseasonalized squared return. The seemingly contradictory inferences from the PE and PD plots is due to the fact that the temporal dependence structure of the deseasonalized squared returns are weak but gradually diminishing. As mentioned in section 3.7, the value of PE is determined by the strength of the temporal dependence at the selected delay and the rate at which the dependence changes. If the dynamics underlying the observed data have relatively weak but persistent temporal structures, this may not be detected by PE. That is why the PE plot suggests a random process, but the PD measure indicates the existence of long-range temporal dependence structures in the observed deseasonalized squared returns. According to the PD plot displayed in Figure 30 (b), the temporal dependence structures within the deseasonalized squared returns remain significant until  $\tau = 25$  to 30, suggesting that the impact from historical 10-minute squared returns recorded around 4 to 5 hours ago can still affect the volatility of current 10-minute returns. The delays at which the dependence structures fully diminished are determined by the delays at which the PD plot stopped declining, instead of the delays at which the PD measure is insignificant. This is because, as covered in section 3.6, non-stationarity can lead measure PD to stay significant for large delays. In particular, the steady level of significant value of measure PD is a typical indication of non-stationarity exhibited in the investigated series.

By contrasting the PD plot and the sample ACF plot for deseasonalized squared returns,

the slopes of diminishing rate of the newly-proposed PD measure and the measure ACF of increasing delays are similar, suggesting linear serial correlation might be the main form of the temporal dependence present in the intraday deseasonalized squared return series. Figure 30 (c) displays the ordinal pattern distribution of deseasonalized squared returns at delay one, with the solid line corresponding to the ordinal pattern probabilities associated with significant PE, whereas the ordinal pattern probabilities resulting in insignificant PE are plotted in the dashed line. Only the ordinal pattern probabilities associated with significant PE are informative. For all ordinal pattern distribution associated with significant PE, ordinal pattern  $\pi_{321}$  is more probable than pattern  $\pi_{123}$ . Referencing the ordinal pattern distribution features I extracted for various mainstream models in section 3.7, the ARMA model would not be an appropriate choice for modelling the investigated deseasonalized squared return series since realizations from any ARMA model are expected to display equal probability of ordinal pattern  $\pi_{123}$  and  $\pi_{321}$ . For the same reason, the GARCH model with Gaussian innovations can also be excluded for consideration as GARCH model with Gaussian innovations is expected to exhibit  $p(\pi_{123}) > p(\pi_{321})$ , which is opposite to the features observed in our investigated data. However, the GARCH model with student's  $t$  innovations can replicate the observed features of  $p(\pi_{123}) < p(\pi_{321})$ , and hence can be considered as a promising candidate for modelling intraday deseasonalized squared returns.

#### 5.4.5 Temporal dependence structure in 1-hour realized volatilities

Using the same deseasonalization procedures, I also deseasonalized the [realized volatility \(RV\)](#) of an interval of 1-hour.

$$RV1h_i = \sum_{t=1+6(i-1)}^{6i} r_t^2$$

I choose an interval of 1-hour because 1-hour consists of six sub-intervals of the original time series frequency of 10-minutes, so that the sum of an aggregate volatility quantity eliminates a considerable amount of noise exhibited in the original intraday time series, whilst we still have a acceptable length of data for the PE analysis that requires a certain amount of data size to produce reliable results. The deseasonalized 1-hour realized volatility series is denoted by DRV1h. Note that, for the realized volatilities, there is no need to consider the effect of discretization as for shorter-interval squared returns. The influence from discretization will become negligible when aggregating the squared returns that originally follow discrete distribution.

The purpose of analyzing the 1-hour interval volatility dynamics is first to investigate how

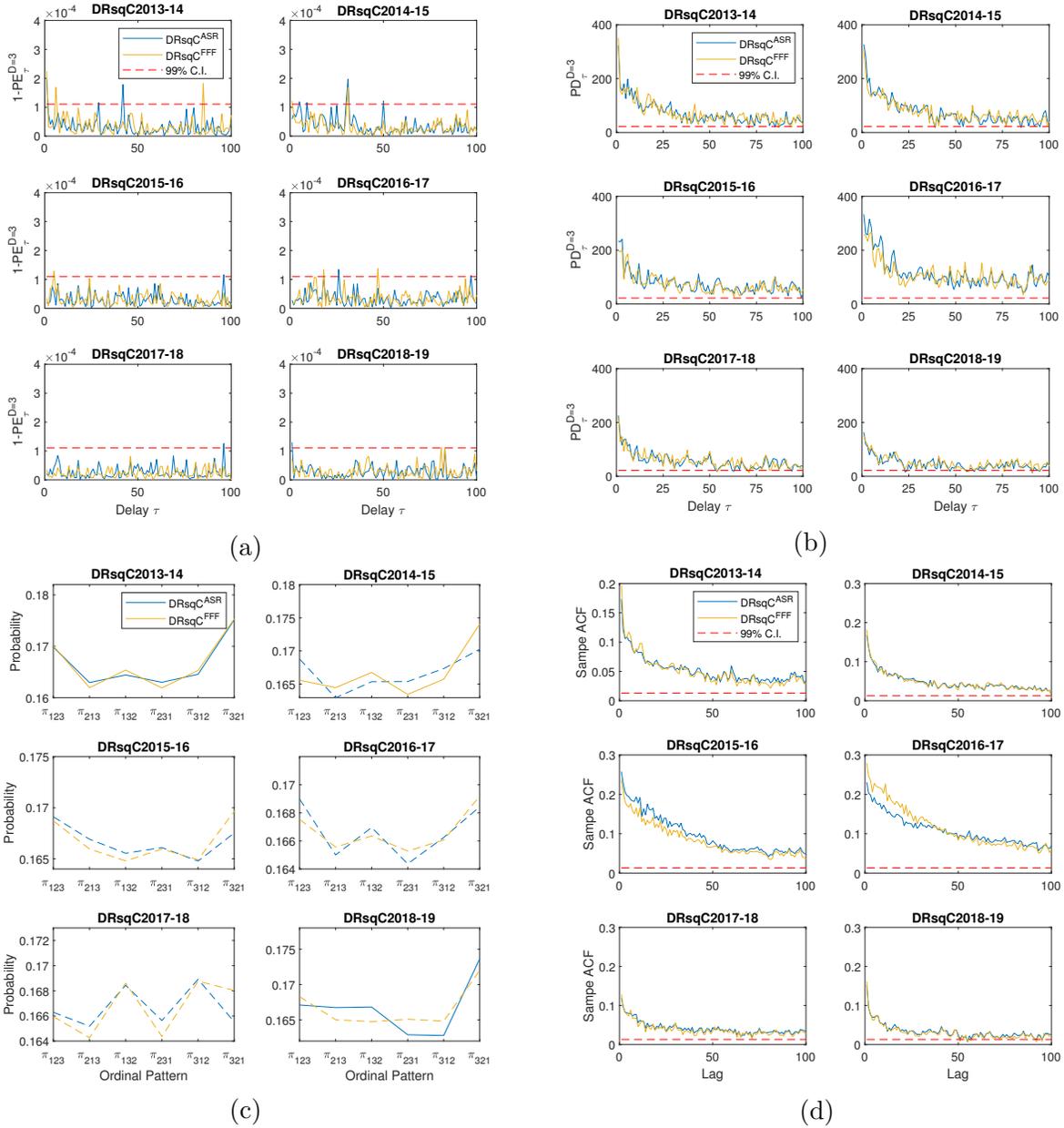


Figure 30: (a) Plot of  $1 - PE_{\tau}^{D=3}$  against delays on deseasonalized 10-minutes EUR/USD squared returns after continuous adjustment. (b) Plot of  $PD_{\tau}^{D=3}$  against delays on deseasonalized 10-minutes EUR/USD squared returns after continuous adjustment. (c) Distribution of ordinal patterns associated with  $PE_{\tau}^{D=3}$  at delay  $\tau = 1$ . (d) Plot of sample ACF on deseasonalized 10-minutes EUR/USD squared returns after continuous adjustment.

the dependence structures of volatilities change over different choices of time interval  $\Delta t$ . Additionally, empirical studies suggest that the mean absolute and mean squared returns are proportional to a power of the interval size, thus the distributions of return and volatility scale for a range of time intervals. The contrast between the dependence structures within 10-minutes interval volatilities and within 1-hour aggregated volatilities might help answer the question of how and whether the functional form of deterministic relations change for different choices of time interval  $\Delta t$ .

To investigate the temporal dependence structures underlying DRV1h, I plot the  $1 - PE_\tau^{D=3}$ ,  $PD_\tau^{D=3}$  and sample ACF as functions of delay and the ordinal pattern probability associated with PE at  $\tau = 1$  for the series under investigation in the six non-overlapping periods, after employing the two different seasonal filters. Similar to our observation in section 5.4.4 of the analysis on DRsqC, the PE plot of DRV1h displays significant value at delay 1 for all periods and occasionally marginal significant values at longer delays for some periods. In contrast, the PD plot indicates the significant but gradually diminishing dependence relations from delay  $\tau = 1$  to around  $\tau = 13$  suggesting the influence from historical aggregated volatilities within 1-hour intervals can last around 13 hours. From Figure 31 (c) we can see that the feature of the ordinal pattern distribution on DRV1h at delay 1 across the six investigated periods is very similar, with  $p(\pi_{321}) > p(\pi_{123})$ , and minimal difference between  $p(\pi_{213})$  and  $p(\pi_{231})$ . The invariant ordinal pattern probability rank across different choices of investigation periods suggests that, although the strength of the temporal dependence structures varies, the functional form of the dynamics governing the 1-hour realized volatility do not change over time.

By contrasting the PD plot and the ordinal pattern distributions on DRsqC and DRV1h, I find that the strength of temporal dependence structures in the 1-hour aggregated volatilities is stronger than that in DRsqC. I also find that despite the increased dispersion of ordinal pattern distributions in DRV1h, the rank of ordinal pattern probabilities in DRV1h is very similar to that in the 10-minute squared returns, indicating the main characteristics of volatility dynamics might be invariant regardless of the choice of time interval  $\Delta t$ . In other words, volatilities over various intervals can be represented and modelled through the same stream of parametric models.

Based on the results and insights from section 3.7, the rank of ordinal pattern probabilities can help identify the functional form of the time series under study. The observed feature of ordinal pattern probability rank in DRV1h and DRsqC series can be potentially replicated by a GARCH or GARCH variant model with leptokurtosis innovations. In the following sections, I further investigate how well GARCH models characterize the dynamics underlying the

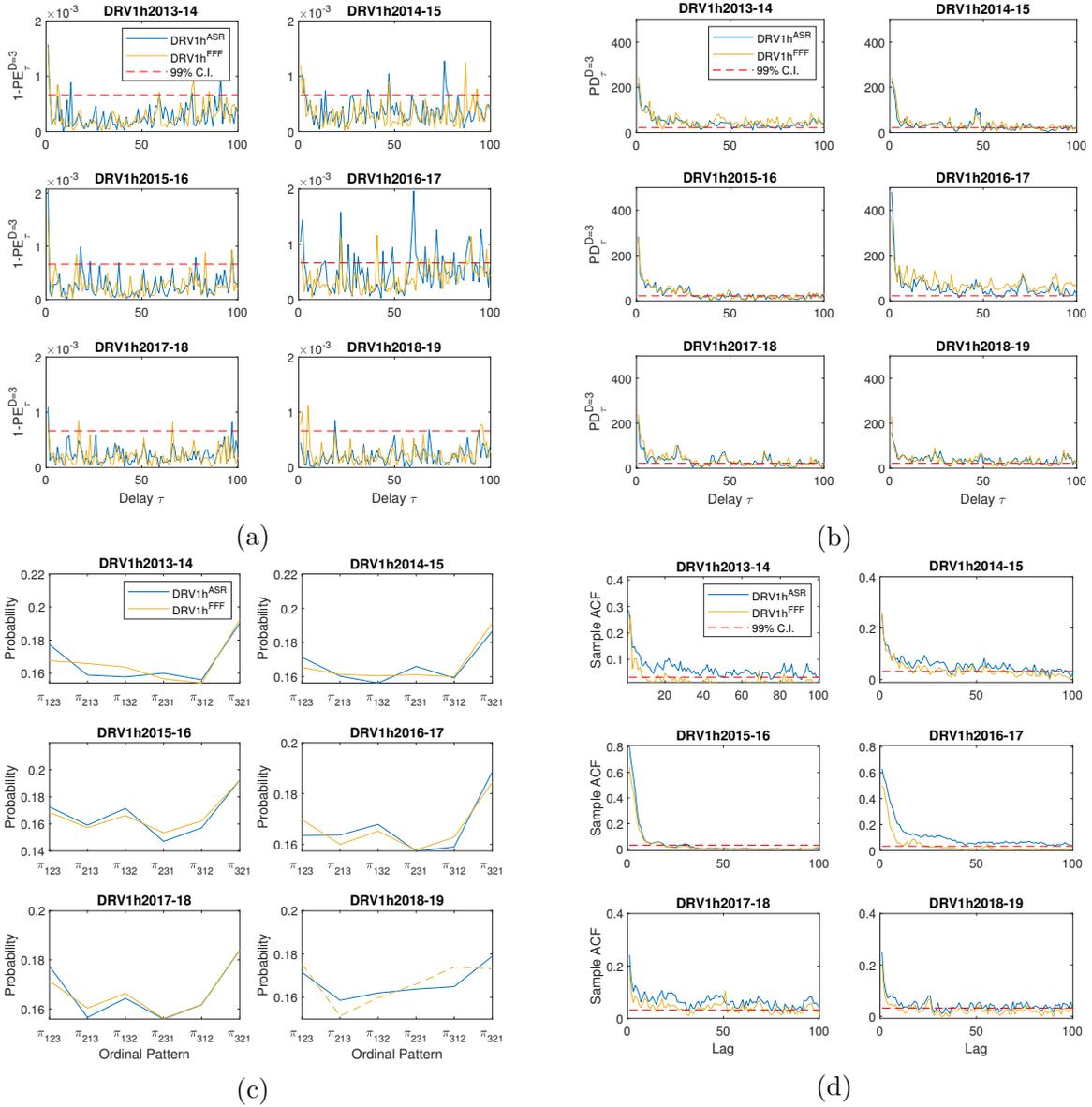


Figure 31: (a) Plot of  $1 - PE_{\tau}^{D=3}$  against delays on deseasonalized 1-hour realized volatilities. (b) Plot of  $PD_{\tau}^{D=3}$  against delays on deseasonalized 1-hour realized volatilities. (c) Distribution of ordinal patterns associated with PE at delay  $\tau = 1$ . (d) Plot of sample ACF on deseasonalized 1-hour realized volatilities.

10-minute squared returns and the 1-hour aggregated volatilities of the EUR/USD rate.

#### 5.4.6 Summary

This section analyzed the temporal dependence structures exhibited in EUR/USD 10-minutes squared returns and 1-hour realized volatilities.

Consistent with the literature, I found that high-frequency financial returns display pronounced intraday seasonality. Two seasonal adjustment filters Flexible Fourier Form (FFF) and average squared returns (ASR) were employed to remove seasonality in the squared returns and realized volatilities. Despite their completely different functional form, performance of the filters was shown to be very similar. Both can remove most of the periodicities but not exhaustively.

In addition to seasonality, discretization imposes a nonlinear structure on intraday squared return dynamics but to a much lesser extent when compared to that in the return itself. Moreover, even after applying seasonal filters, the temporal dependent structures in the squared returns introduced by discretization cannot be fully diminished. Discretization was also found to hamper the performance of the deseasonalization procedures, especially when the investigated time series have strong baseline deterministic structures.

Compared to the temporal dependence structure present in the high-frequency returns, which mainly exhibited determinism over the short-term, the PD measure detected slowly diminishing dependence in squared returns, consistent with the “long-memory” property of financial returns recorded in subsection 2.2.3. Our empirical analysis suggested the temporal dependence structure in deseasonalized squared returns does not fully diminish until delay  $\tau = 50$ , which is around eight hours.

By comparing the values of PE and PD in 10-minute squared returns and 1-hour realized volatilities, I found aggregated volatilities within longer intervals are more structural than squared returns recorded in shorter intervals. Additionally, realized volatilities also exhibited “long-memory” property. The PD plot, as a function of delay, suggests the temporal dependence structures in deseasonalized 1-hour aggregated volatilities gradually decay and vanish after delay  $\tau = 13$ , indicating the impact from a historical realized volatility can last 13 hours. Despite the varied strength and diminishing rate of the temporal structures exhibited in deseasonalized squared returns and aggregated volatilities, the similar rank in their ordinal pattern probabilities suggests the process governing their dynamics share similar functional form. Hence, they have the potential to be characterized through similar models. Based on the feature of their ordinal pattern distributions, I exclude the ARMA model and standard

GARCH class model for replicating the dynamics of the deseasonalized squared returns and deseasonalized realized volatilities, but identify the GARCH class model with leptokurtosis innovations as a promising candidate for replicating their underlying dynamics.

Lastly, by comparing the temporal dependence analysis results on different investigation periods, I found the overall strength and the diminishing rate of the deterministic structure underlying squared returns and aggregated volatilities vary over time. Nevertheless, the similarity of the ordinal pattern distributions indicates that the functional form governing the process does not vary regardless of which investigation period is chosen.

## 5.5 Modelling and forecasting of 10-minute returns

In section 5.3 I analyzed the temporal dependence structures underlying the EUR/USD 10-minute returns in detail and decomposed the detected dependence structures within observed returns into multiple component. In this section I investigate how well the mainstream models capture and exploit the temporal dependence structures underlying the intraday EUR/USD deseasonalized returns for the purpose of point prediction. The sufficiency of a predicting model will be evaluated based on a number of criterion. First, the newly proposed PD sufficiency test will be carried out to detect any remaining structures in the residuals and the dependence structures between residuals and past observations to assess the ability of the models to exploit the potential of point predicting one-step-ahead returns. Second, I expect a sufficient model to replicate the temporal dependence structures as in the deseasonalized return series. Therefore I compare the value of PE, PD, ACF and ordinal pattern distribution in the simulation generated from the estimated models and that in the observed deseasonalized returns to see whether they are alike.

I choose to consider the simplest MA(1) model to fit and forecast deseasonalized returns. This choice is motivated by the observation that all six years' investigated return series only exhibit negative first order autocorrelations, and the ordinal pattern rank in the continuous-adjusted return is similar to that of linear models. In fact, fitting the high-frequency returns to a MA(1) model first to eliminate the negative first-order autocorrelation, then fitting the squared residuals to a GARCH-class model is a standard procedure in modelling high-frequency returns. In most cases, after fitting to the MA(1) model, intraday return series no longer exhibit negative first ACF. However, whether the MA(1) model is able to replicate all the temporal dependence structures in the deseasonalized return series remains an open question. In the remainder of this section I aim to answer this question.

### 5.5.1 The sufficiency of MA(1) model

In the previous section, I showed that intraday squared returns exhibit strong periodicities. Therefore, the fitting process will be conducted on deseasonalized returns to eliminate the known periodicities exhibited in the return magnitude. In section 5.4.2, two seasonal filters, ASR and FFF, were employed to remove the seasonalities in the return volatilities. Despite the differences in their functional form, the two selected seasonal filters performed similarly. Hence in this analysis, I choose to only investigate the deseasonalized return/deseasonalized squared returns using the more regular functional form FFF seasonal filters.

The deseasonalized return used in this section for prediction is after the continuous adjustment, denoted by **deseasonalized continuous-adjusted return (DRC)**. The DRC is obtained by applying the seasonal filtering procedures on continuous adjusted returns. The continuous convention procedure is described in section 5.3.2. The reason we convert to the original returns into a continuously distributed series prior to the deseasonalization procedures it is that in section 5.3.2, I showed that the discreteness in intraday returns is the primary source contributing to the temporal dependence structures in the intraday return series. As evidenced in section 5.4.3, even the deseasonalizing procedures cannot entirely eliminate its effect. To avoid the distortion of the remaining discretization effect, I choose to predict DRC instead of the original deseasonalized returns. Despite the data adjustments, the magnitudes of DRC are very close to that without any adjustments. On average their difference is around 1.32%. Therefore the adjustments procedures will not introduce any significant change to the results of the following analysis.

Among the total length of 38160 entries of return series for each investigated period, the former 30960 observations are used for estimating the parameters and the latter 7200 observations are for out-of-sample evaluation purposes. In order to assess the sufficiency of the point prediction performance of the fitted MA(1) model, I apply the newly-proposed PD sufficiency test on both the in-sample subsets and out-of-sample subsets of the investigated data.

Table 6 summarizes the statistics used to evaluate the performance of the fitted MA(1) model in its attempt to exploit the predictability of EUR/USD 10-minute deseasonalized returns. Table 6 records the estimated parameter  $\theta_1$ , the in-sample and out-of-sample forecasting accuracy measure  $\text{adjMSEratio}$  and the  $\text{PD}_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$  and  $\text{PD}^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$  statistics required

by the newly proposed PD model sufficiency test.  $\text{adjMSEratio}$  is computed by

$$m = \frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t),$$

$$\text{adjMSEratio} = \frac{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t - m)^2}{\text{Var}(y_t)}.$$

I use  $\text{adjMSEratio}$  to indicate the forecasting accuracy because it quantifies the proportion of variations unexplained by the estimated model. Additionally, it omits the systematic bias in the estimation of the constant term of the proposed model, which our PD sufficiency test cannot identify. The statistic  $\text{PD}_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$  in the sufficiency test indicates the strength of dependence between past residuals and current residual after fitting the MA(1) model, whereas statistic  $\text{PD}^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$  quantifies the dependence between the lagged observations and current residual of the data under study. By comparing the values of the  $\text{PD}_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$  and  $\text{PD}^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$  with their 95% C.I. under independence, they all show insignificant or marginally significant values. The results suggest the fitted MA(1) models are able to exploit most of the point prediction potentials in the deseasonalized return dynamics, despite the fact that the predicted returns can only explain an average of 1% variations in the observed returns. In addition, Figure 32 plots the  $1 - \text{PE}_{\tau}^{D=3}$  plot,  $\text{PD}_{\tau}^{D=3}$  plot, sample ACF plot and ordinal pattern distribution on the deseasonalized returns and their residuals after fitting the MA(1) model. The comparison between before and after fitting the MA(1) model suggests the estimated MA(1) model manages to eliminate almost all short-term temporal dependence structures in the original data. The sufficiency evaluation of the simple MA(1) model on deseasonalized returns indicates the intraday returns have minimal point prediction potential where noise constitutes around 99% of the magnitude of the observed returns.

### 5.5.2 Estimated MA(1) dynamics versus real return dynamics

In this section, I compare the expected behaviour of the PE and PD measures applied to the simulated time series from the fitted MA(1) model with that applied to original data. The analysis carried out in this section aims to further verify the sufficiency of the MA(1) model in characterizing the dynamics underlying the intraday return.

I showed earlier that the distribution of the innovations can affect the ordinal pattern distributions, thus the values of PE and PD. For that reason I choose to simulate the time series from the estimated model with innovations following the empirical distribution of the residuals using kernel functions instead of the standard normal distribution with estimated standard deviations from fitting the MA(1) model. I generate 500 paths of simulations to

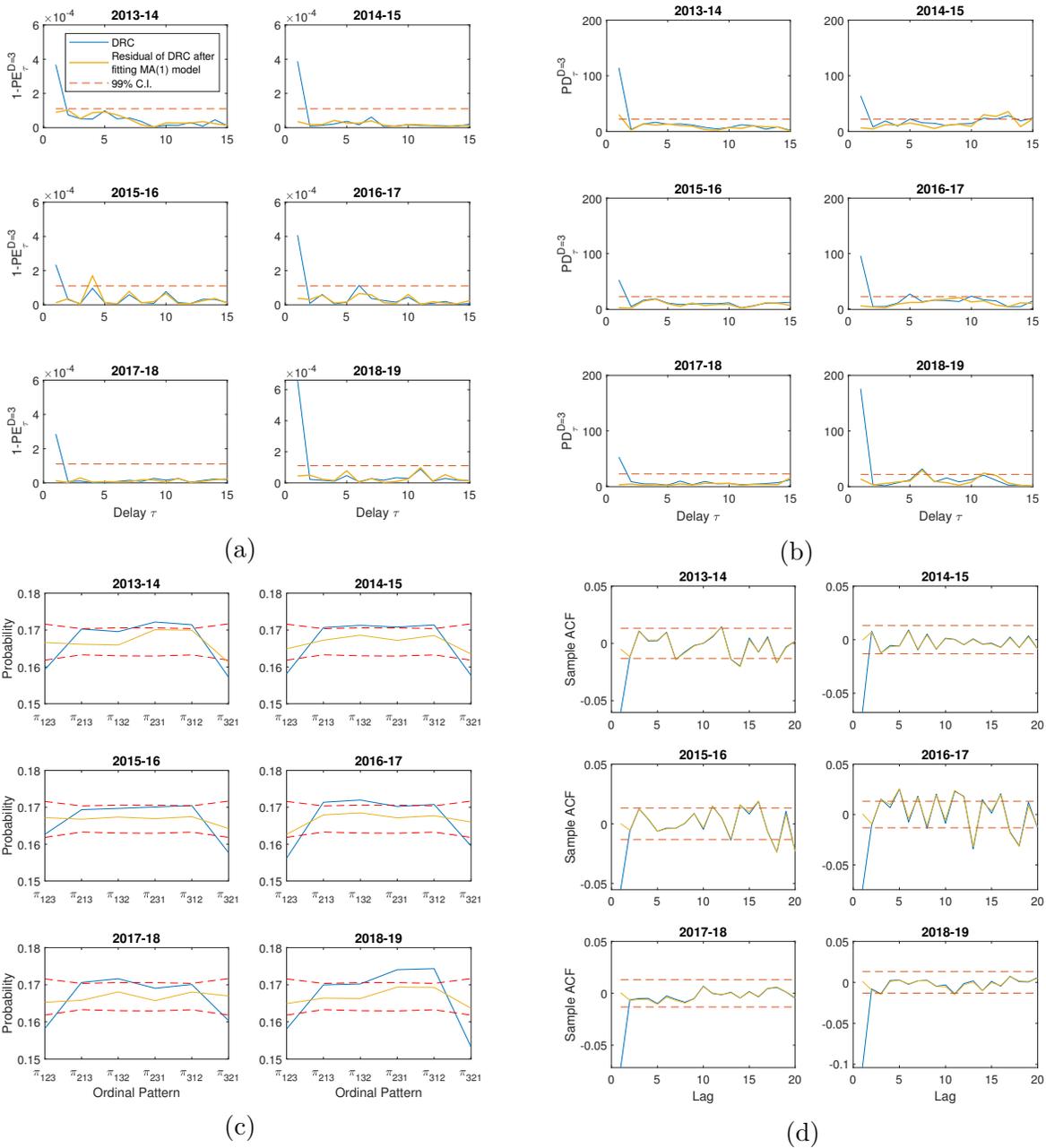


Figure 32: Comparison of (a) plot of  $1 - PE_{\tau}^{D=3}$  over delays (b) plot of  $PD_{\tau}^{D=3}$  over delays (c) probabilities of ordinal patterns associated with  $PE_{\tau=1}^{D=3}$  (d) plot of sample ACF on continuous adjusted deseasonalized returns and residuals after fitting MA(1) model.

Table 6: The estimated parameter  $\hat{\theta}_1$ , the in-sample and out-of-sample forecasting accuracy measure adjMSEratio and the statistics  $\text{PD}_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$  and  $\text{PD}^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$  required by the PD sufficiency test used to evaluate the one-step-ahead point prediction of the fitted MA(1) on 10-minutes EUR/USD continuous adjusted deseasonalized returns. The one-tail 95% C.I. of statistics  $\text{PD}_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$  and  $\text{PD}^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$  under randomness is 40.

Investigation period	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
Fitted MA(1) model on DRC						
$\hat{\theta}_1$	-0.056	-0.069	-0.057	-0.078	-0.074	-0.102
adjMSEratio (in sample)	99.25%	99.40%	99.10%	99.21%	98.89%	98.96%
adjMSEratio (out of sample)	98.98%	99.64%	99.73%	99.61%	99.46%	98.03%
$\text{PD}_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$ (in-sample)	65.66	45.23	12.44	37.93	21.17	44.14
$\text{PD}^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$ (in-sample)	60.23	54.45	18.88	42.99	20.73	36.05
$\text{PD}_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$ (out-of-sample)	27.81	30.57	16.20	11.52	25.03	34.02
$\text{PD}^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$ (out-of-sample)	25.02	31.39	19.03	7.51	22.95	32.05

construct an estimate of the 95% confidence interval of the expected level of PE and PD of the simulations. The comparisons of  $1 - \text{PE}_{\tau}^{D=3}$ ,  $\text{PD}_{\tau}^{D=3}$ , sample ACF and ordinal pattern distribution at delay 1 between the simulations and the original deseasonalized return are shown in Figure 33. The plots suggest all employed temporal dependence measures in the simulation are not distinctively different from that computed on the original data, indicating the simple linear correlation specified by the MA(1) model is sufficient or at least very close to sufficiency in capturing the deterministic structures exhibited in return dynamics.

### 5.5.3 Summary

In this section, I evaluated the sufficiency of MA(1) in replicating the dynamics underlying the EUR/USD intraday deseasonalized returns using a number of different techniques introduced earlier in this thesis.

I concluded that the MA(1) model is sufficient or at least very close to sufficiency in replicating the structure embedded in the 10-minute EUR/USD returns. The PD sufficiency test indicates that predictions constructed from the MA(1) model exploit most of the forecasting potential embedded in the intraday return dynamics, as evidenced by the insignificant or marginally significant values of  $\text{PD}_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$  and  $\text{PD}^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$ . In addition, the simulations generated from the estimated MA(1) model closely replicate the PE plot, PD plot and ordinal pattern distribution computed on the real intraday return series. Moreover, the estimated residuals after fitting the prediction model showed no significant remaining structures throughout all considered delays. All the above evidence supports the conclusion that the prediction generated from the MA(1) model is very close to the best prediction one can possibly construct

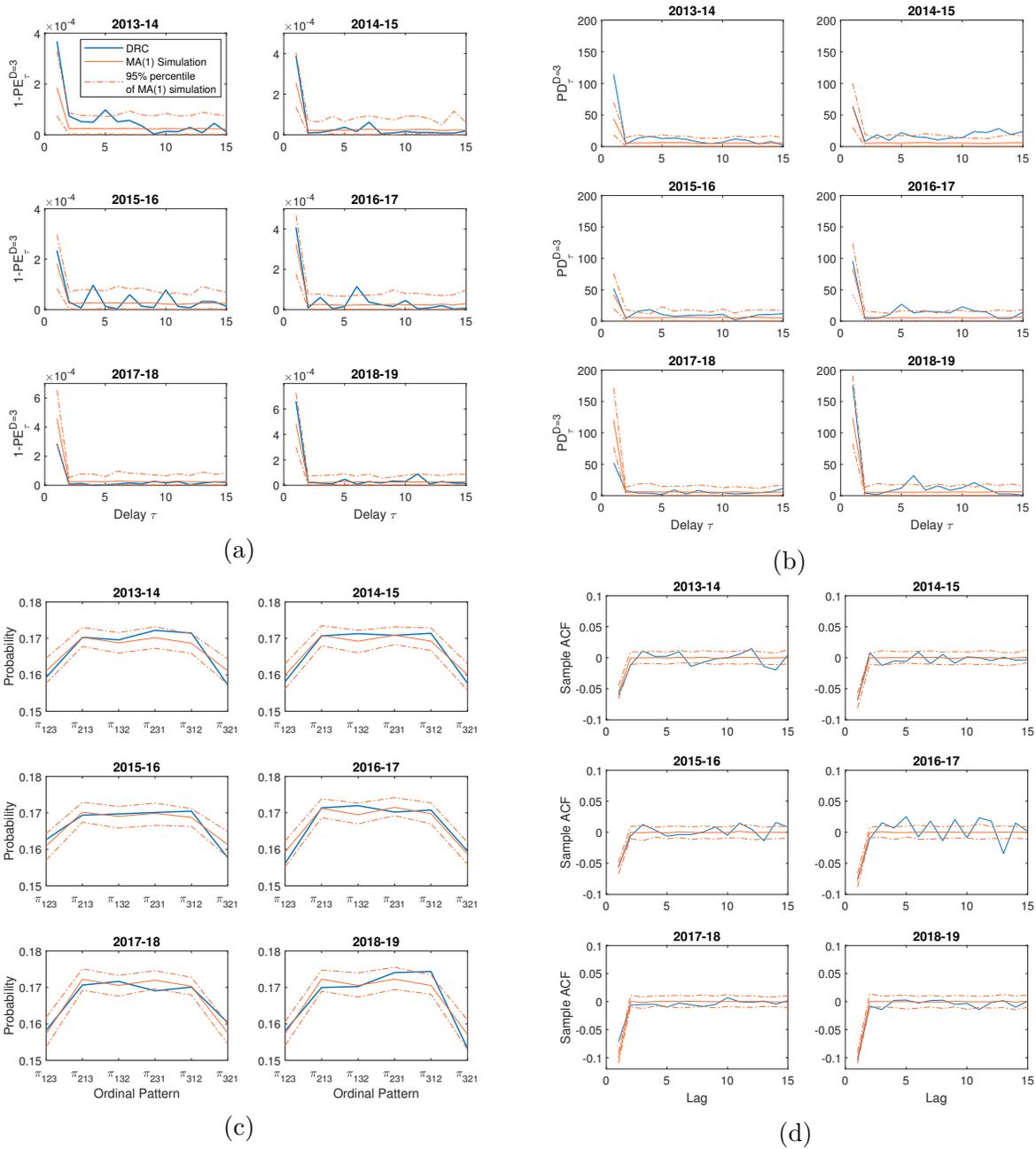


Figure 33: Comparison between (a) plot of  $1 - PE_{\tau}^{D=3}$  over delay, (b) plot of  $PD_{\tau}^{D=3}$  over delay, (c) probabilities of ordinal patterns associated with  $PE_{\tau=1}^{D=3}$ , (d) plot of sample ACF over lag, on investigated continuous adjusted deseasonalized returns and the average value of considered measures computed on MA(1) simulations, along with their respective 95% percentile computed in the generated MA(1) simulations.

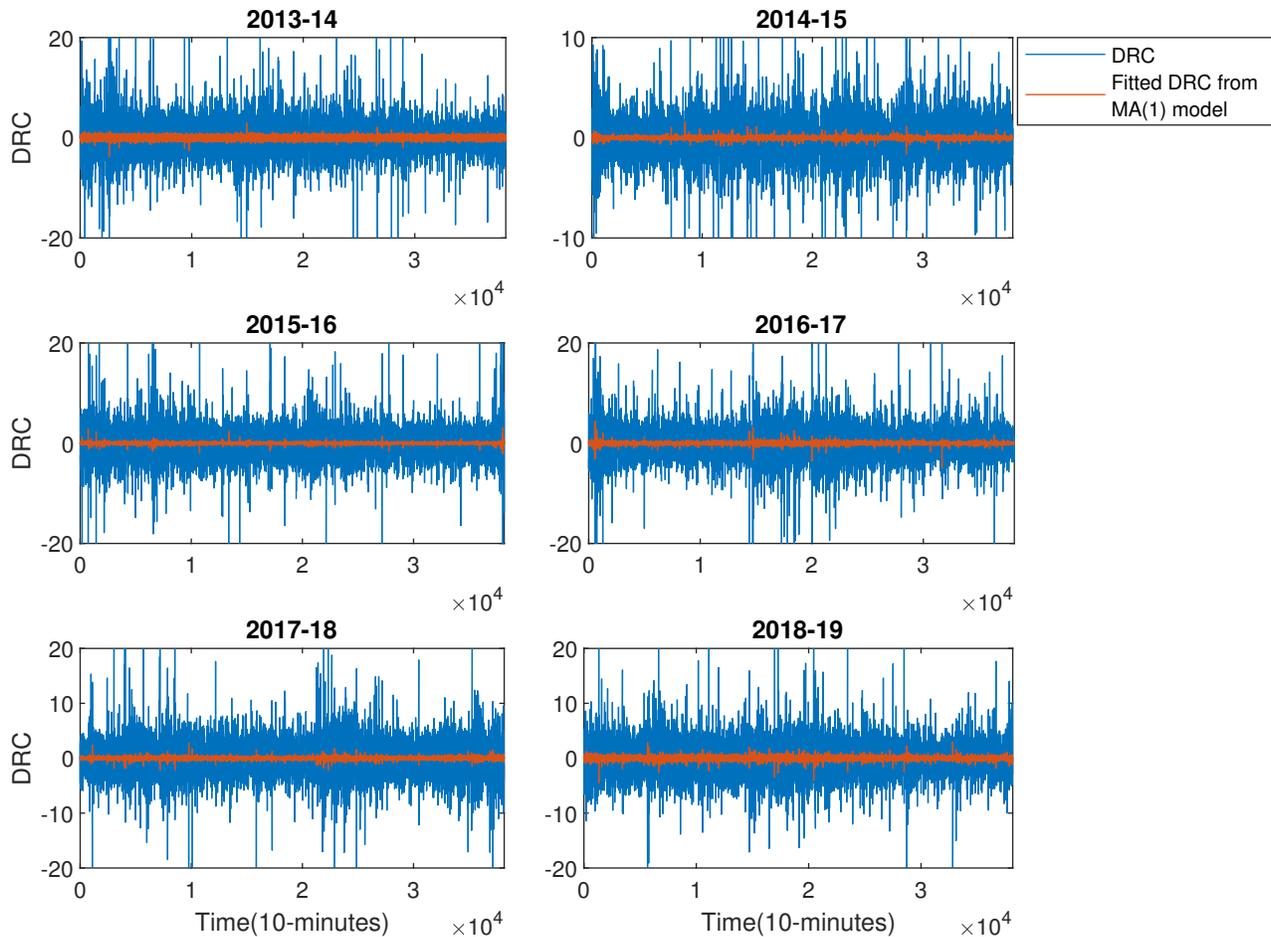


Figure 34: Plot of deseasonalized continuous-adjusted return and the fitted/foretasted value generated from the MA(1) model.

in predicting future intraday returns based on past returns, despite the fact it only improves the prediction accuracy by 0.3% to 2% over random guesses. The sufficient but marginal improvement of prediction accuracy implies that 10-minute EUR/USD returns are very noisy where error is the dominant component of the process, constituting 99% of the observed return. Therefore there is minimal potential for forecasting high-frequency return solely using its historical trajectories.

## 5.6 Modelling and forecasting of 10-minute squared returns and 1-hour aggregate volatilities

In this section, I evaluate the performance of various models in forecasting 10-minute squared returns and 1-hour aggregate volatility. More importantly, I investigate the reasons for their respective superior and suboptimal performances. My ultimate objective is to uncover the

main obstacles associated with each model in dealing with the properties of high-frequency financial data and provide promising directions for constructing better forecasting models. The considered models include three parametric models, ARMA, GARCH and FIGARCH, and two non-parametric approaches, SVR and GPR. The detailed description and specification of these models are provided in section 2.3. All five models are widely used in the area of financial time series analysis and are often used as the benchmark for comparison when practitioners evaluate the prediction performance of new models.

There are a number of reasons behind our choice of competing models. The ARMA and GARCH models presume linear relations between the object of interest and its past realizations. Even though the GARCH model is normally classified as non-linear, the deterministic relation postulated by it is still in linear form. GARCH model takes into account the volatility clustering feature in estimating the linear relation between squared returns. The comparison between the performance of the ARMA model and the GARCH model reveals the importance of incorporating the volatility clustering feature in predicting return volatilities. The FIGARCH model is a GARCH model variant. It relieves the fixed exponentially decaying constrain specified in the standard GARCH model by introducing an additional parameter  $d$ . From the comparison between the prediction accuracy of the standard GARCH model and the FIGARCH model, we can identify whether more flexibility in the form of the linear deterministic relationship enables the GARCH-class models to exploit the predictability underneath the intraday exchange rate volatility dynamics sufficiently. The non-parametric models GPR and SVR, compensate for the limitations of the above three parametric models by introducing the capability to capture non-linear deterministic relations. As described in section 2.3, that GPR and SVR models capture the behaviour of past movement of observed data through completely different mechanisms, which can result in quite different predictions on data with different properties. The comparison of the prediction accuracy between the popular non-parametric models and the classical parametric models reveals the usefulness of non-linear models in predicting intraday volatilities of the EUR/USD exchange rates.

### **5.6.1 Evaluation of various models in predicting 10-minute squared returns and 1-hour realized volatilities**

As for the previous section, in which I evaluate models' capacity to predict future returns, in this section I employ similar procedures to evaluate the performance of models in predicting squared returns. There are two criteria I am basing on to assess the sufficiency of the considered models.

1. The newly proposed PD model sufficiency test is used to assess the sufficiency of the

constructed predictions proposed by the considered models. The PD test is based on the comparison of the strength of temporal dependence within the estimated residuals and the strength of dependence between the estimated residuals and past observations.

2. When the PD model sufficiency test is unable to generate an affirmative conclusion, we will use the DM test with quadratic loss function (MSE) to compare the models with the one that generate least MSE or are considered inadequate. Any inferior model concluded by the DM test is also concluded insufficient.

Note that our evaluation studies focus on determining whether the considered models generate sufficient one-step-ahead point predictors of squared returns rather than the latent volatility. Namely, the competing models are evaluated in terms of their abilities to predict  $E(r_t^2|F_{t-1})$  instead of  $\text{Var}(r_t|F_{t-1})$ . It is important to distinguish the difference between these two to ensure the validity of the evaluation criteria listed above. The reasons are as follows. First, recall section 4.3, in which the design and construction of the PD sufficiency test aims to assess how the employed models capture and replicate the deterministic relation governing the first-order moment structures underlying the observed time series. Therefore the proposed test distinguishes whether the predictions are close or significantly depart from the oracle point forecast  $\hat{x}_t^{\text{oracle}} = E(x_t|F_{t-1})$ . With squared returns as our data, a sufficient prediction is defined in the PD sufficiency test as close to  $E(r_t^2|F_{t-1})$ . If  $E(r_t^2|F_{t-1})$  and  $\text{Var}(r_t|F_{t-1})$  are equal or equivalent, then the conclusion drawn from the PD sufficiency test can be extended to evaluate the performance of various models in prediction the latent volatility. However,  $E(r_t^2|F_{t-1})$  and  $\text{Var}(r_t|F_{t-1})$  are not always equal or equivalent especially in the presence of high kurtosis.

Under certain circumstances, for instance, if

$$r_t|F_{t-1} \sim \sigma_t \times \mathcal{N}(0, 1)$$

then

$$E(r_t^2|F_{t-1}) = \text{Var}(r_t|F_{t-1}).$$

However, the above equality does not necessarily always hold. For example, if

$$r_t|F_{t-1} \sim \sigma_t \times \text{Student's } t(0, \nu)$$

then

$$E(r_t^2|F_{t-1}) = \frac{\nu}{\nu - 2} \text{Var}(r_t|F_{t-1}). \quad (26)$$

In that case, the optimal point prediction of squared returns is not equivalent to that of the latent volatility, especially when the degree of freedom of the  $t$ -distribution that is negatively related to the level of leptokurtosis is of relative small value. As noted in section 2.2.1, a heavy tail is a widely documented empirical property of financial returns. Additionally, the distribution is increasingly heavy-tailed for higher frequency returns. In fact, all of our investigated 10-minute returns exhibit high levels of leptokurtosis. Compared to Gaussianity, the Student's  $t$  distribution with a relatively small degree of freedom parameter is a more realistic assumption of the distribution of the conditional return  $r_t|F_{t-1}$ , which corresponds to the latter case of the example listed above. Therefore the sufficient/insufficient conclusions drawn from the PD sufficiency test cannot be extended to evaluating the performance of various models in predicting the latent volatility for our data.

For the second evaluation criteria, if the quantity of interest is  $\text{Var}(r_t|F_{t-1})$ , the MSE metric and the DM test based on MSE functions can lead to erroneous ranking of the performance of competing models. Patton (2011) indicates that the MSE is a “robust” loss function for correctly ranking competing volatility forecasts. Therefore even the “true” volatility is a latent variable, the MSE is a robust function for comparing volatility forecasts when using squared returns as a noisy proxy. However, in order for the above argument to be true,  $E(r_t^2|F_{t-1})$  needs to be equal to  $\text{Var}(r_t|F_{t-1})$ . The proof is simple and is given below.

Let  $h_t^*$  denote the optimal forecasts for the MSE loss function using squared returns as a proxy for latent volatility:

$$h_t^* = \underset{h}{\text{argmin}} E[(r_t^2 - h)^2|F_{t-1}].$$

The search for the optimal forecast defined above is equivalent to the search for  $h$  that makes the derivative of  $E[(r_t^2 - h)^2|F_{t-1}]$  equal to zero. We have

$$\begin{aligned} \frac{\partial}{\partial h} E[(r_t^2 - h)^2|F_{t-1}] &= E[2(r_t^2 - h)|F_{t-1}] \\ &= 2E(r_t^2|F_{t-1}) - 2h \\ &= 0 \end{aligned}$$

Therefore,

$$h_t^* = E(r_t^2 | F_{t-1}).$$

I show in equation (26) that  $E(r_t^2 | F_{t-1})$  is not always equal to the latent volatility. As a result, the predictions that minimize the MSE of the observed squared returns are not necessarily the optimal prediction of the population latent volatilities of returns. Hence MSE or any prediction comparison test, such as the DM test based on MSE using squared returns as a volatility proxy, can be misleading if the objective is to compare the performance of competing models in predicting  $\text{Var}(r_t | F_{t-1})$ .

In addition to the PD sufficiency test and DM model comparison test, there are four other metrics I use to reflect the performance of the competing models. They are out-of-sample MSEratio, out-of-sample adjMSEratio, out-of-sample mean error ( $m$ ) and the out-of-sample mean error contribution in MSE. The four statistics are defined as follows:

$$\begin{aligned} \text{MSEratio} &= \frac{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2}{\frac{1}{N} \sum_{t=1}^N (y_t)^2}, \\ m &= \frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t), \\ \text{adjMSEratio} &= \frac{\frac{1}{N} \sum_{t=1}^N [y_t - \hat{y}_t - \frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)]^2}{\frac{1}{N} \sum_{t=1}^N (y_t - \bar{y}_t)^2}, \\ \text{m ratio in MSE} &= \frac{m^2}{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2}. \end{aligned}$$

MSEratio measures the mean squared error of the estimated prediction relative to that if we assume all  $\hat{y}_t = 0$ .  $\hat{y}_t = 0$  represents the prediction one can construct with no knowledge about the future or the underlying dynamics at all, and thus can act as a benchmark to be compared with the predictions estimated by the competing models.  $\text{MSEratio} > 1$  is a clear indication that the prediction estimated by the considered model is even worse than a random guess. Additionally, in order to separate the error in estimating the constant term and that in estimating the deterministic function, I split the most widely-used prediction accuracy metric MSE into two components. If the one-step-ahead predictor of  $x_t$  can be written in the following form

$$\hat{x}_t = \hat{c} + \hat{g}(x_{t-1}, x_{t-2}, \dots, x_{t-\tau_T})$$

where  $\tau_T$  is the furthest lag in which past entries affect the current entries, adjMSEratio

measures how well the predictor replicates the deterministic function  $g(\cdot)$  governing the investigated data.  $m$  accounts for the error in estimating the constant term  $c$ . More specifically,  $\text{adjMSEratio}$  measures the proportion of un-explained variations in the investigated data by each estimated predictor after adjusting for systematic bias. The  $m$  ratio in MSE indicates the proportion to which the misspecification of constant term contributes to the overall MSE.

The reason I modify the conventional MSE and isolate the contributions of constant estimation error is to ensure comparability of model performance across different periods in the presence of nonstationarity. It is clear from Figure 35 that the our investigated deseasonalized squared return series displays pronounced non-stationarity, especially inconstant mean. Since our selected models, both parametric and non-parametric, are designed for stationary or at least weakly stationary time series, they are unable to flexibly adjust themselves to adapt the varying baseline volatilities of the investigated volatility series. Even though a number of novel or hybrid models, such as MS-GARCH and so on are proposed to model and forecast non-stationary time series, assessing their performance is beyond the scope of this thesis. It is anticipated that the out-of-sample forecasting performance of our selected models will be better on the series where the baseline mean of the test set is closer to that of the training set. Since our primary objective is to investigate among all of our selected models, which mimics the function form of the volatility dynamic most closely, and try to explain why is so.

For parametric models ARMA and GARCH, around 80% of data (30960) is used for estimation and 20% (7200) is used to evaluate out-of-sample forecasting performance. For non-parametric models GPR and SVR, around 60% of data (24000) is used for the training set, 20% (6960) is used for optimal lag selection and hyper-parameter tuning, and 20% (7200) is used to evaluate out-of-sample forecasting performance. For non-parametric models, a number of kernel functions, including linear and rbf functions, have been considered, and a wide range of possible values of hyper-parameters has been compared based on the MSE in the 20% of hyper-parameter tuning subset.

Table 7 records the statistics I use to evaluate the performance of the models considered in this study in predicting 10-minute EUR/USD volatilities, including the estimated parameters in each employed model, the out-of-sample MSEratio,  $\text{adjMSEratio}$ , mean error out-of-sample mean error contribution in MSE defined above, and the statistics  $\text{PD}_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$  and  $\text{PD}^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$  required in the PD model sufficiency test. The orange cell corresponds to the models that are concluded to be insufficient according to the two criteria. Note that the models that are not orange in the table are not necessarily the oracle model for predicting the investigation data, rather we do not have enough evidence to reject their sufficiency. Additionally, I also highlight the model with the best prediction accuracy in each investigation

period by showing the smallest out-of-sample  $\text{adjMSEratio}$  in each  $\text{DRsq}$  series with red underscore.

From the table consisting of the out-of-sample forecasting evaluation merits of the five competing models for all investigated periods 10-minute EUR/USD deseasonalized squared returns, a number of conclusions can be drawn. First, I confirm the existence of prediction potential for intraday exchange rate volatility. Except for 2016-17, in which none of the considered models can beat random guesses, the best performing model of the rest of investigating periods' are able to provide significant reduction in uncertainty about future volatility compared to if we made no effort into forecasting. The increment of accuracy ranges from 6% to 19% in forecasting deseasonalized intraday squared returns. Second, the results suggest there is no single model that outperforms its competitors in terms of prediction accuracy across all periods of the study. For instance, the GARCH(1,1) model provides much better prediction accuracy compared to its competitors in predicting  $\text{DRsq}_{2015-16}$ , whereas predictions generated from the SVR model are superior to those made from the ARMA model in predicting  $\text{DRsq}_{2013-14}$ . In other words, the prediction performance of the employed models varies significantly across different investigated time periods. Additionally, almost all constructed predictors are not sufficient to exploit the maximum prediction potential.

Using the same set of competing models and evaluation criteria, I also assess the sufficiency of the one-step-ahead prediction performance of the deseasonalized realized volatility  $\text{DRV1h}$ , the temporal dependence structures of which are analyzed in section 5.4.5. The statistics summarizing the forecasting performance of the models in relation to  $\text{DRV1h}$  are given in Table 8. Overall 1-hour aggregated volatilities are more predictable compared to 10-minute squared returns, as the best performing model can increase the prediction accuracy of the investigated data by 10% on average over random guesses and up to 42% in  $\text{DRV1h}_{2015-16}$ . However, despite the vast potential of predictability of aggregated volatilities, similar to that found in the  $\text{DRsq}$  series, almost all of the constructed predictors are not sufficient to exploit the maximum prediction potential. Additionally, there is no single model that outperforms the rest of its competitors across all periods. The ranking of prediction accuracies of the models largely depends on the selected investigation period.

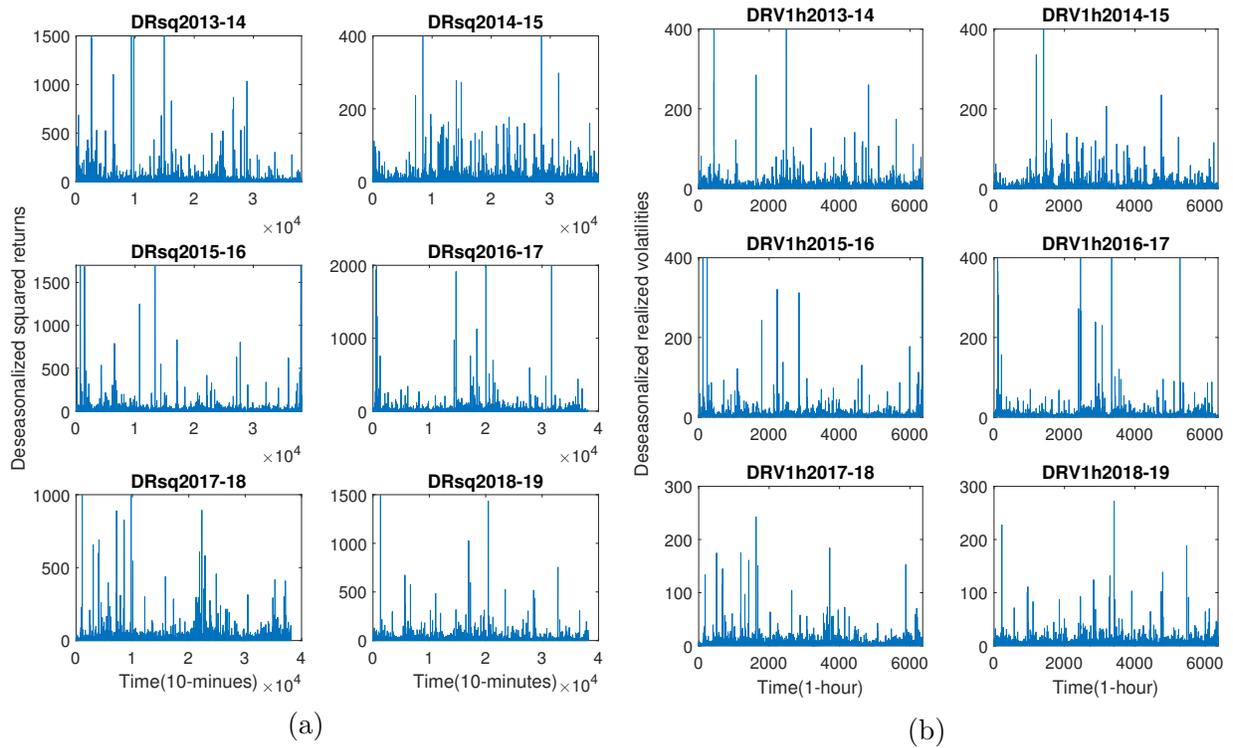


Figure 35: (a) Plot of the six non-overlapping 1-year period EUR/USD 10-minutes deseasonalized squared returns. (b) Plot of the six non-overlapping 1-year period EUR/USD 1-hour deseasonalized aggregated volatilities.

Table 7: Comparison of prediction performances of various models on deseasonalized EUR/USD 10-minutes squared returns recorded in six non-overlapping 1-year investigation periods. MSEratio measures the MSE of the estimated prediction relative to that if we assume all  $\hat{y}_t = 0$ :  $MSEratio = [\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2] / [\frac{1}{N} \sum_{t=1}^N (y_t)^2]$ . adjMSEratio and  $m$  measure how well the predictor replicates the deterministic function  $g(\cdot)$  governing the investigated data and the error in estimating the constant term  $c$ :  $adjMSEratio = [\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t - m)^2] / [\frac{1}{N} \sum_{t=1}^N (y_t - \bar{y}_t)^2]$ ;  $m = \frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)$ .  $m$  ratio in MSE records the proportion that the misspecification of constant term contributing to the overall MSE:  $m$  ratio in MSE =  $m / \frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2$ .  $PD_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$  and  $PD^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$  are the statistics required in the PD model sufficient test. The orange cells correspond to the models that are concluded to be insufficient. The model with smallest out-of-sample adjMSEratio is highlighted in red underscore, and termed “Function Estimation Winner”. The model with smallest out-of-sample MSEratio is termed “Overall Winner”. All the model evaluation metrics and statistics are computed in the out-of-sample test set that is not used for training the model.

Overall Winner	SVR	SVR	GARCH	GPR	ARMA	FIGARCH
Function Estimation Winner	SVR	SVR	GARCH	GPR	ARMA	ARMA
	DRsq2013-14	DRsq2014-15	DRsq2015-16	DRsq2016-17	DRsq2017-18	DRsq2018-19
ARMA						
$\hat{\phi}_1, \hat{\theta}_1$	0.62,0.42	0.92,0.84	0.88,0.80	0.98,0.87	0.89,0.84	0.84,0.78
MSEratio out-of-sample	99.95%	92.38%	83.62%	102.89%	85.28%	93.70%
adjMSEratio out-of-sample	<u>100.40%</u>	<u>100.40%</u>	84.38%	<u>103.47%</u>	<u>98.75%</u>	<u>98.61%</u>
$m$ out-of-sample	-4.25	-0.31	1.21	-0.58	0.32	-1.98
$m$ ratio in MSE out-of-sample	9.70%	0.37%	0.02%	0.01%	0.04%	1.57%
$PD_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$ out-of-sample	627.18	550.53	543.17	953.31	414.69	349.43
$PD^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$ out-of-sample	449.86	109.87	174.60	237.90	119.10	143.42
PD sufficiency test out-of-sample	Inconclusive	Inconclusive	Inconclusive	Inconclusive	Unsure	Unsure
GARCH						
$\hat{\alpha}_1, \hat{\beta}_1$	0.12,0.80	0.10,0.84	0.10,0.82	0.1,0.85	0.1,0.81	0.11,0.67
MSEratio out-of-sample	94.54%	93.44%	80.47%	102.03%	86.11%	93.57%
adjMSEratio out-of-sample	<u>99.53%</u>	<u>101.64%</u>	<u>81.20%</u>	<u>102.60%</u>	<u>99.74%</u>	<u>98.77%</u>
$m$ out-of-sample	-3.07	-0.27	0.76	-0.63	0.03	-1.78
$m$ ratio in MSE out-of-sample	5.36%	0.27%	0.01%	0.01%	0.00%	1.27%
$PD_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$ out-of-sample	920.13	770.05	829.35	841.23	659.82	421.84
$PD^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$ out-of-sample	293.17	149.71	167.74	225.24	201.77	236.87
PD sufficiency test out-of-sample	Inconclusive	Inconclusive	Inconclusive	Inconclusive	Inconclusive	Inconclusive
FIGARCH						
$\hat{\phi}_1, \hat{d}, \hat{\beta}_1$	0.06,0.15,1.4e-05	0.38,0.22,0.44	0.40,0.20,0.44	0.40,0.20,0.45	0.07,0.20,0.16	0.07,0.09,6.7e-04
MSEratio out-of-sample	91.99%	93.19%	82.67%	101.42%	86.14%	92.91%
adjMSEratio out-of-sample	<u>100.68%</u>	<u>101.50%</u>	83.41%	<u>102.00%</u>	<u>99.47%</u>	<u>99.07%</u>
$m$ out-of-sample	-1.66	0.20	1.71	0.41	0.94	-0.80
$m$ ratio in MSE out-of-sample	1.60%	0.15%	0.04%	0.00%	0.31%	0.26%
$PD_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$ out-of-sample	715.27	939.43	739.83	784.58	611.77	525.50
$PD^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$ out-of-sample	376.59	196.53	223.72	280.59	221.78	334.99
PD sufficiency test out-of-sample	Inconclusive	Inconclusive	Inconclusive	Inconclusive	Inconclusive	Inconclusive
GPR						
Selected lag for input	8	5	8	8	7	3
MSEratio out-of-sample	95.95%	100.16%	93.27%	99.78%	85.66%	112.80%
adjMSEratio out-of-sample	<u>103.34%</u>	<u>108.70%</u>	94.09%	<u>100.34%</u>	<u>99.21%</u>	<u>118.09%</u>
$m$ out-of-sample	-2.38	-0.38	1.83	-0.79	0.13	-2.50
$m$ ratio in MSE out-of-sample	3.18%	0.50%	0.04%	0.02%	0.01%	2.08%
$PD_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$ out-of-sample	947.07	259.31	498.59	<b>150.38</b>	426.90	370.88
$PD^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$ out-of-sample	349.95	145.78	413.19	<b>199.52</b>	181.27	345.96
PD sufficiency test out-of-sample	Inconclusive	Inconclusive	Inconclusive	<b>Reject</b>	Inconclusive	Inconclusive
SVR						
Selected lag for input	2	7	1	1	1	1
MSEratio out-of-sample	91.52%	92.26%	96.31%	100.00%	98.31%	93.13%
adjMSEratio out-of-sample	<u>99.36%</u>	<u>99.66%</u>	96.50%	<u>100.35%</u>	<u>105.81%</u>	<u>99.48%</u>
$m$ out-of-sample	2.02	-0.51	7.29	2.93	4.79	0.47
$m$ ratio in MSE out-of-sample	2.40%	0.96%	0.56%	0.22%	7.08%	0.09%
$PD_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$ out-of-sample	464.17	96.17	425.38	352.60	282.67	255.65
$PD^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$ out-of-sample	326.03	116.66	317.33	333.05	303.37	218.05
PD sufficiency test out-of-sample	Inconclusive	Inconclusive	Inconclusive	Inconclusive	Inconclusive	Inconclusive

Table 8: Comparison of prediction performances of various models on deseasonalized EUR/USD 1-hour aggregated volatilities recorded in six non-overlapping 1-year investigation periods. MSERatio measures the MSE of the estimated prediction relative to that if we assume all  $\hat{y}_t = 0$ :  $\text{MSERatio} = [\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2] / [\frac{1}{N} \sum_{t=1}^N (y_t)^2]$ . adjMSERatio and  $m$  measure how well the predictor replicates the deterministic function  $g(\cdot)$  governing the investigated data and the error in estimating the constant term  $c$ :  $\text{adjMSERatio} = [\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t - m)^2] / [\frac{1}{N} \sum_{t=1}^N (y_t - \bar{y}_t)^2]$ ;  $m = \frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)$ .  $m$  ratio in MSE records the proportion that the misspecification of constant term contributes to the overall MSE:  $m$  ratio in MSE =  $m / \frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2$ .  $\text{PD}_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$  and  $\text{PD}^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$  are the statistics required in the PD sufficiency model sufficient test. The orange cells correspond to the models that are concluded to be insufficient. The model with smallest out-of-sample adjMSERatio is highlighted with red underscore, and is termed “Function Estimation Winner”. The model with smallest out-of-sample MSERatio is termed “Overall Winner”. They are computed in the training set due to the inadequate data length of the investigated data in the test set.

Overall Winner	SVR	ARMA	GARCH	SVR	ARMA	SVR
Relation Estimation Winner	FIGARCH	SVR	GARCH	SVR	SVR	SVR
	DRV1h2013-14	DRV1h2014-15	DRV1h2015-16	DRV1h2016-17	DRV1h2017-18	DRV1h2018-19
ARMA						
$\hat{\phi}_1, \hat{\theta}_1$	0.79,0.63	0.76,0.55	0.87,0.75	0.86,0.41	0.88,0.77	0.57,0.39
MSERatio out-of-sample	81.70%	69.03%	63.71%	117.85%	56.40%	78.92%
adjMSERatio out-of-sample	<u>94.53%</u>	<u>100.95%</u>	<u>65.17%</u>	<u>120.61%</u>	<u>96.64%</u>	<u>99.11%</u>
$m$ out-of-sample	-0.61	-0.12	0.34	-0.19	0.11	-0.44
$m$ ratio in MSE out-of-sample	18.11%	0.50%	0.07%	0.04%	0.23%	5.91%
$\text{PD}_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$ in sample	161.44	<b>104.37</b>	123.46	<b>222.50</b>	119.96	122.23
$\text{PD}^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$ in sample	107.04	<b>150.36</b>	119.60	<b>289.53</b>	103.21	121.19
PD sufficiency test in sample	Inconclusive	<b>Reject</b>	Inconclusive	<b>Reject</b>	Inconclusive	Inconclusive
GARCH						
$\hat{\alpha}_1, \hat{\beta}_1$	0.23,0.43	0.26,0.55	0.16,0.74	0.35,0.49	0.05,0.94	0.20,0.29
MSERatio out-of-sample	83.13%	70.88%	56.77%	110.81%	56.45%	79.14%
adjMSERatio out-of-sample	<u>93.33%</u>	<u>103.67%</u>	<u>58.10%</u>	<u>113.39%</u>	<u>96.91%</u>	<u>99.17%</u>
$m$ out-of-sample	-0.66	-0.12	0.22	-0.20	0.05	-0.45
$m$ ratio in MSE out-of-sample	20.55%	0.50%	0.03%	0.05%	0.05%	6.12%
$\text{PD}_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$ in sample	141.19	<b>138.49</b>	160.25	181.49	223.13	140.33
$\text{PD}^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$ in sample	129.38	<b>204.21</b>	134.79	191.99	86.58	144.54
PD sufficiency test in sample	Inconclusive	<b>Reject</b>	Unsure	Inconclusive	Inconclusive	Inconclusive
FIGARCH						
$\hat{\phi}_1, d, \hat{\beta}_1$	0.46,0.08,0.32	0.37,0.19,0.28	1.36e-05,0.24,0.06	0.43,0.15,0.23	0.38,0.24,0.50	0.08,0.11,0.0012
MSERatio out-of-sample	71.76%	69.83%	57.69%	107.50%	56.93%	74.46%
adjMSERatio out-of-sample	<u>92.87%</u>	<u>102.44%</u>	<u>58.91%</u>	<u>110.05%</u>	<u>95.70%</u>	<u>98.88%</u>
$m$ out-of-sample	-0.39	0.07	0.61	0.08	0.32	-0.12
$m$ ratio in MSE out-of-sample	8.4%	0.18%	0.25%	0.01%	2.12%	0.50%
$\text{PD}_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$ in sample	125.60	199.57	140.84	237.69	108.24	120.80
$\text{PD}^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$ in sample	145.72	200.12	104.20	212.10	102.62	114.41
PD sufficiency test in sample	Inconclusive	Inconclusive	Inconclusive	Inconclusive	Inconclusive	Inconclusive
GPR						
Selected lag for input	6	1	1	3	4	10
MSERatio out-of-sample	84.38%	69.04%	86.62%	98.29%	56.48%	81.05%
adjMSERatio out-of-sample	<u>102.75%</u>	<u>100.68%</u>	<u>88.55%</u>	<u>100.61%</u>	<u>96.92%</u>	<u>99.14%</u>
$m$ out-of-sample	-0.54	-0.15	0.55	-0.1	0.06	-0.53
$m$ ratio in MSE out-of-sample	13.82%	0.77%	0.14%	0.01%	0.08%	8.36%
$\text{PD}_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$ in sample	323.98	366.58	<b>178.35</b>	135.92	132.87	138.54
$\text{PD}^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$ in sample	304.35	408.39	<b>243.73</b>	126.84	132.64	150.56
PD sufficiency test in sample	Inconclusive	Inconclusive	<b>Reject</b>	Inconclusive	Inconclusive	Inconclusive
SVR						
Selected lag for input	3	1	1	1	4	2
MSERatio out-of-sample	69.37%	69.72%	93.32%	97.70%	60.56%	73.09%
adjMSERatio out-of-sample	<u>96.24%</u>	<u>98.58%</u>	<u>94.85%</u>	<u>100.02%</u>	<u>95.54%</u>	<u>97.52%</u>
$m$ out-of-sample	-0.18	0.33	1.30	0.03	0.64	0.03
$m$ ratio in MSE out-of-sample	1.81%	3.79%	0.72%	0.00%	7.79%	0.03%
$\text{PD}_{\tau=1}^{D=4}(\hat{\varepsilon}_t)$ in sample	73.47	221.47	152.23	<b>258.70</b>	59.21	83.70
$\text{PD}^{*,D=4}(\hat{\varepsilon}_t, x_{t-1})$ in sample	62.34	272.96	177.22	<b>369.19</b>	81.11	74.98
PD sufficiency test in sample	Inconclusive	Inconclusive	Inconclusive	<b>Reject</b>	Inconclusive	Inconclusive

## 5.6.2 Reasons behind the insufficient predicting performance of the employed models

Why is the prediction performance of our selected models largely dependent on the choice of investigation period? And why are the considered models unable to construct a sufficient one-step-ahead predictor of future squared returns? The prediction evaluation conducted in the previous section leaves those questions unanswered.

To answer those questions, I deploy the PD visualization plot I proposed in section 3.8 to reveal the deterministic relation  $\hat{g}(\cdot)$  captured by each forecasting model. The PD visualization graphs of the five competing models used in predicting DRsq and DRV1h are given in Figures 36 and 37. In the PD visualization plots, the competing models that are found to be inadequate are plotted in dashed lines, and \* indicates the model that provides the best prediction accuracy amongst its competitors. Note that the models that are plotted in solid lines are not necessarily the oracle prediction model for the investigation data, rather we do not have enough evidence to reject its sufficiency.

The newly-proposed temporal dependence measure  $PD_{\tau}^{D=3}$  on the simulations generated from the fitted prediction model reflects the “weight” that the estimated function  $\hat{g}(\cdot)$  of the considered model assigns to the lagged input variable  $x_{t-\tau}$ . For instance, from Figure 36 we can see that, for DRsq2013-14 series, the deterministic relation estimated by the GPR model is much stronger than its competitors as it assigns heavier “weights” to all relevant lagged input variables. In other words, the prediction made from the GPR model is more reliant on the value of lagged squared returns compared to that generated from its competitors. On the contrary, the SVR model provides the most conservative predictions amongst all competing models where the estimated deterministic relation presumes minimal impact from historical observations onto the prediction of future realizations. As a result, the deterministic function estimated by the SVR model is close to a steady constant line that does not vary, regardless of the value of input variables. The parametric ARMA, GARCH and FIGARCH models estimate very close deterministic relations in predicting DRsq2014-15 as evidenced by their similar values of  $PD_{\tau}^{D=3}$  over various delays.

Among the three parametric models, the GARCH model postulates a more gradually decaying rate of the linear serial correlation in constructing its predictor. The remark from the PD visualization plot perfectly coincides with the estimated parameters in the fitted ARMA(1,1), GARCH(1,1) and FIGARCH(1,d,1) models. In section 2.3 the serial correlation decaying rate in GARCH(1,1) model is discussed, establishing that it is controlled by its parameter  $\alpha_1 + \beta_1 = 0.92$ , whereas the decaying rate of ACF is determined by  $\phi_1 = 0.62$  in ARMA(1,1).

FIGARCH(1,d,1) model does not assume a constant exponentially decaying rate of its linear serial correlation like the ARMA and GARCH models, following the formulation provided in equation (9), the average ACF decaying rate for the first five lags is around 0.75. As suggested by the estimated parameters, the fitted GARCH(1,1) of DRsq2014-15 series has the slowest diminishing rate of its linear serial relation, compared to that proposed by the fitted ARMA(1,1) and FIGARCH(1,d,1) models.

Since none of the above models is able to outperform its competitors in predicting DRsq or DRV1h across all considered investigation periods, additionally almost all of the constructed predictors from our selected models are not sufficient to exploit the maximum prediction potential, hence there must be some obstacles that prevent the making of optimal predictions. The simulation studies in section 4.4 show that the forecasting performance of the SVR and GPR models can be undetermined by dynamical innovations, especially when the investigated data has a relatively small signal to noise ratio. Unfortunately our investigated DRsq and DRV1h series happen to meet both characteristics. First, high-frequency exchange rate return is a prominent example of noisy data. In DRsq series across all investigated periods, even in the most successful predictions, noise constitutes 81.2% of the observed data. Second, past empirical studies summarized in section 2.2.3 suggest that volatility clustering is a widely observed feature of financial returns. Since large price variations are likely to be followed by large price variations, large past return shocks are likely to lead to large dispersion of future returns. This feature causes a secondary-moment dependence structure in the squared returns. As a result, by transforming the dynamics governing the behaviour of squared returns into the following form:

$$x_t = c + g(x_{t-1}, x_{t-2}, \dots, x_{t-\tau_T}) + \varepsilon_t, \quad (27)$$

the innovation term  $\varepsilon_t$  is dependent on each other in a way that large past  $\varepsilon_{t-\tau}$  tends to lead to more volatile future  $\varepsilon_t$ .

Further evidence supporting the existence of dependence structures within the innovations in time series DRsq and DRV1h is that the PD measure capturing the temporal dependence structures within the residuals after fitting the prediction models are all far in excess of the insignificant level for all considered models across all investigated periods. Therefore, there is a high likelihood that the innovations of the original dynamics of DRsq and DRV1h are dependent on each other. By combining the former simulation study results for the GPR and SVR models and the empirical results obtained here, I conclude that the SVR and GPR models generally produce poor predictions of intraday squared returns and volatilities because the existence of dependence structures presented in the innovation undermined their forecasting performance. The extent that the dependence structure in innovation affects the

prediction performance of the GPR and SVR models is determined by the strength of the innovations' structures and the magnitude of innovations relative to signals.

I have outlined the reasons for the suboptimal performance of the non-parametric SVR and GPR models, but what is the cause of the inadequate performance of the parametric ARMA, GARCH and FIGARCH models? As indicated by the simulation study carried out in section 4.4, the ARMA and GARCH models are robust to dynamical innovations. I suspect their inadequacy might be due to their incapacity to formulate non-linear deterministic relations. However, this suspicion appears to be unfounded because the deterministic relation captured by the non-parametric models are not very nonlinear, even in the investigation period in which non-parametric models outperform the parametric models that presume linear serial relations. The estimated  $\hat{g}(\cdot)$  from the SVR and GPR models with superior prediction accuracy is actually very close to linear form. By ruling out a number of potential factors that can lead to inadequate performance of the ARMA and GARCH models, I finally identify the primary cause behind the unsatisfactory performances of the ARMA, GARCH and FIGARCH models. That is non-stationarity. More specifically, the form of non-stationarity where the deterministic relations governing the process are time-varying.

The non-stationarity present in our investigated DRsq and DRV1h series can be illustrated by splitting the investigation data into several sub-periods. We plot the sample ACF of DRsq and DRV1h in each sub-period in Figures 38 and 39. The first five sub-periods correspond to the data used in the training set and the last sub-period in the testing set. In addition, I also plot the sample ACF of simulations generated from the estimated ARMA(1,1), GARCH(1,1) and FIGARCH(1,  $d$ , 1) models to display the departure of the deterministic structure postulated by the estimated model from that in the actual data in the testing set. Figures 38 and 39 indicate the deterministic structure underneath our targeted squared return and aggregated volatility series varies considerably over time. Also, the deterministic structure seems to change abruptly and in an unsystematic and irregular manner. From the sample ACF from the estimated model simulations we can see, under this type of non-stationarity, the ARMA and GARCH models try to estimate the average level of the linear dependence relation underlying the investigated data, thereby overestimating the relation in the sub-period of relatively weaker deterministic dependence structures, and underestimating the relation in the sub-period of stronger than average deterministic dependence structures. Even though the SVR and GPR models are also affected by non-stationarity, they (especially SVR) tend to construct a smoother function compared to the ARMA and GARCH model, thus are more conservative in approximating temporal dependence structures.

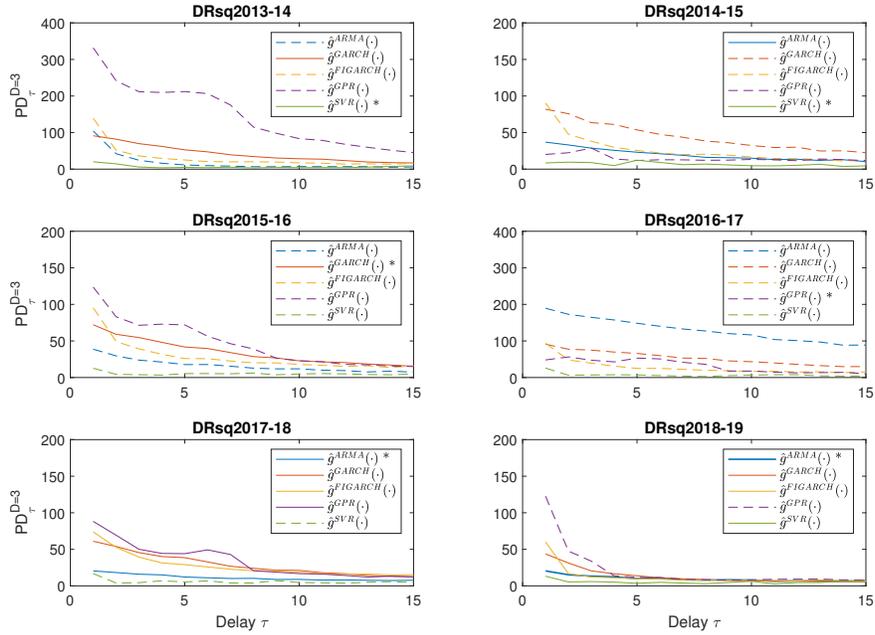


Figure 36: PD visualization plot of the estimated function  $\hat{g}(\cdot)$  postulated by each considered models fitted to the deseasonalized EUR/USD 10-minutes squared returns in every 1-year investigation period. The dashed line corresponds to the model that is concluded insufficient. \* indicates the model with the best prediction accuracy amongst its competitors.

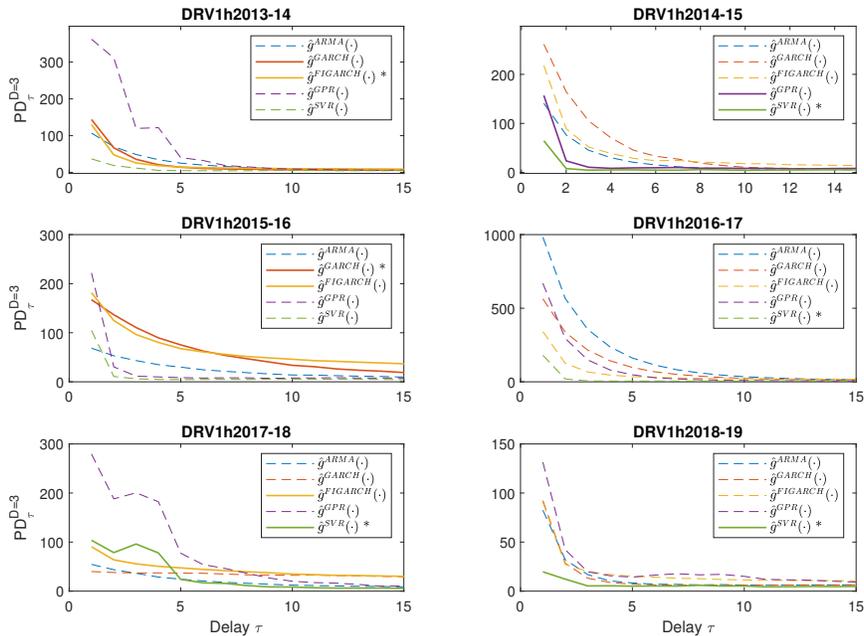


Figure 37: PD visualization plot of the estimated function  $\hat{g}(\cdot)$  postulated by each considered models fitted to the deseasonalized EUR/USD 1-hour aggregated volatilities in every 1-year investigation period. The dashed line corresponds to the model that is concluded insufficient. \* indicates the model with the best prediction accuracy amongst its competitors.

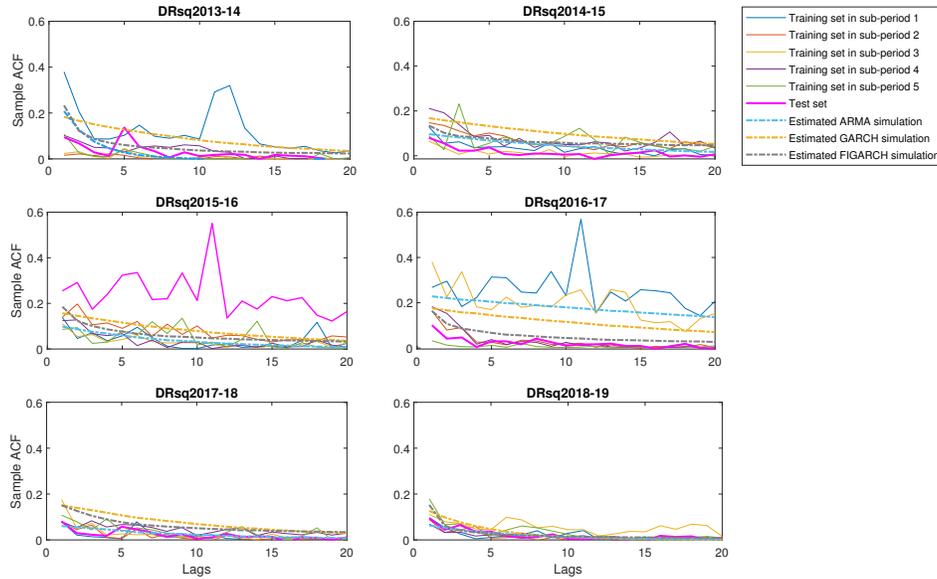


Figure 38: The plot of sample ACF of the deseasonalized EUR/USD 10-minutes squared returns in the six split sub-period of the original investigation period. The first five sub-periods correspond to the data used in the training set and the last sub-period in the testing set. The sample ACF of simulations generated from the estimated ARMA(1,1), GARCH(1,1) and FIGARCH(1,d,1) models are also included in conjunction to show the departure of the linear deterministic structure postulated by the estimated model to that in the actual data in the testing set.

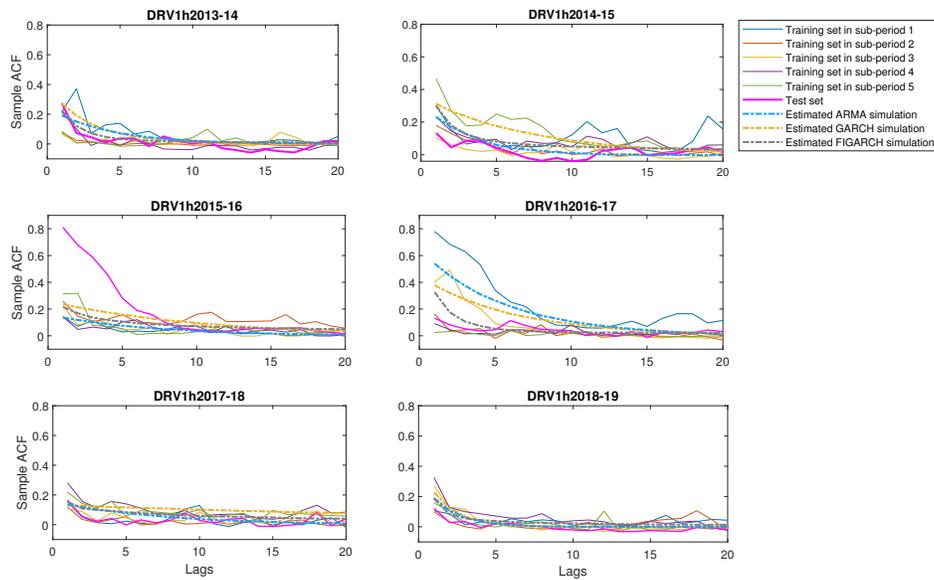


Figure 39: The plot of sample ACF of the deseasonalized EUR/USD 1-hour aggregated volatilities in the six split sub-periods of the original investigation period. The first five sub-periods correspond to the data used in the training set and the last sub-period in the testing set. The sample ACF of simulations generated from the estimated ARMA(1,1), GARCH(1,1) and FIGARCH(1,d,1) models are also included in conjunction to show the departure of the linear deterministic structure postulated by the estimated model to that in the actual data in the testing set.

### 5.6.3 Main obstacles to predicting return volatilities

By investigating the temporal dependence structures exhibited in intraday squared return series, and evaluating the performance of a number of mainstream models, I am able to identify the main obstacles in predicting and modelling high-frequency financial volatilities and provide promising directions for constructing and selecting better forecasting models. I believe there are three primary challenges in predicting financial asset volatilities:

- high level of noise contamination;
- structural innovation;
- non-stationarity.

The first two properties of asset returns severely undermine the proper estimation mechanism of more complex non-parametric models such as GPR and SVR. Even though simpler parametric models are robust to structural innovation, the sufficiency of the simple linear models such as ARMA and GARCH models in replicating the dynamics of real-world asset returns remains an open question. Among them, non-stationarity is the greatest challenge to constructing accurate predictions of future volatilities. Our analysis indicates that the form of non-stationarity exhibited in the financial volatility series is not a simple one, such as unit root non-stationarity which can be easily removed by taking the difference, or systematic time-varying mean or variance that can be eliminated by de-trending the systematic pattern in the mean or the variance of the investigated data as a function of time. The form of non-stationarity exhibited in the financial volatility series is more like a piecewise stationarity where although the time series as a whole exhibits non-stationarity, weak stationarity rules are satisfied in some local regimes and the transition between those regimes is instantaneous. To model piecewise stationarity, a Markov switching model can be employed to account for time-varying temporal dependence relations exhibited in the dynamics. However, it seems that the transition happens arbitrarily and abruptly rather than in a systematic manner. If so, even non-stationary models cannot construct an accurate forecast of such a process since no information can be indicative of when the transition is going to happen and how it is going to change. In that case, even though volatilities series are proven to have strong temporal dependence structures and prediction potential, the arbitrarily abrupt transition between different temporal dependence structures undermines the ergodicity prerequisite for most forecasting models to work properly. From the analysis conducted in this thesis, I cannot ascertain whether the investigated volatility series switches the regions in a systematic manner or arbitrarily and abruptly. There is another possibility that the volatility series switches the regions systematically but not according to its historical movement patterns

instead determined by some exogenous factors.

To overcome these obstacles, a promising prediction model of asset volatilities is expected to have the following properties. First, it needs to be robust to structural innovations or equivalently not misled by higher-moment structures, and has relatively low estimation error if the dynamic displays a high level of noise contamination. Favorably, this model has the flexibility to capture any form of deterministic relation, both linear and nonlinear structures. Most importantly, the desired model needs to detect and adapt swiftly to the abruptly varying dependence relations exhibited in the observed data and identify the right subset of historical observations to estimate the deterministic relation required to construct predictions.

## 5.7 Ordinal analysis on sea surface temperature data

Previous empirical studies have focused on the application of the ordinal analysis in high-frequency financial time series. In this section, I will provide an example of PE, PD and bivariate PE analysis on empirical time series from a completely different field, sampled at a relatively low frequency. Doing so is essential as it demonstrates the potential of the tools I have studied and proposed in this thesis to a much wider range of disciplines in addition to the financial area.

The empirical time series under study is monthly sea surface temperature (SST) data. The SST data is obtained from <https://www.cpc.ncep.noaa.gov/data/indices/ersst5.nino.mth.81-10.ascii> published by the National Oceanic and Atmospheric Administration (NOAA), a scientific agency within the U.S. Department of Commerce. The data records the level of the average monthly SSTs in four different regions: Niño 1+2, Niño 3, Niño 3.4 and Niño 4, from January 1982 to November 2020. Among the four univariate time series, each corresponds to one defined region, and all are of length 851. The Niño 1+2, Niño 3, Niño 3.4 and Niño 4 regions correspond to the ship tracks that crossed the ocean area specified as follows.

- Niño 1+2 (0-10°South)(90°West-80°West): The Niño 1+2 region is located at the Eastern Tropical Pacific, which is the smallest and most eastern of the Niño SST regions. The SST index in this region tends to have the largest variance among of all SST regions.
- Niño 3 (5°North-5°South)(150°West-90°West): The Niño 3 region is located at the eastern tropical Pacific. This region was once the primary focus for monitoring and predicting El Niño, but researchers later learned that the key region for coupled ocean-atmosphere interactions for ENSO lies further west (Trenberth 1997).

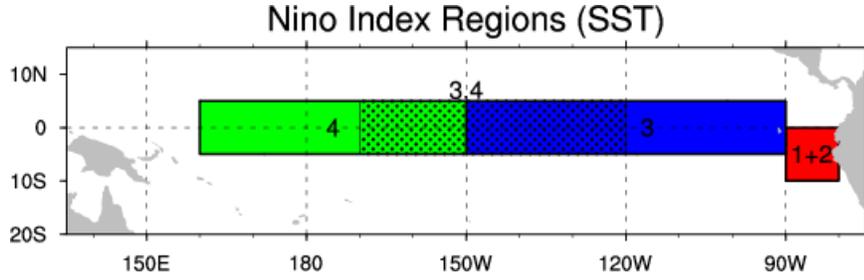


Figure 40: The ocean areas that belongs to each Niño Regions.

- Niño 3.4 (5°North-5°South)(170-120°West): The Niño 3.4 SST records the central tropical Pacific El Niño conditions. The Niño 3.4 SST represents the average equatorial SSTs across the Pacific from about the dateline to the South American coast.
- Niño 4 (5°North-5°South) (160°East-150°West): The Niño 4 region is located at the western tropical Pacific. This region tends to have less variance compared to other Niño regions.

Figure 40, created by [Trenberth \(2020\)](#), visually identifies the four regions that correspond to each SST data. The four univariate SST time series plots are provided in Figure 41.

SST is one of the key variables associated with the global ocean-atmosphere system. Many efforts have been made to predict SST, as the accurate prediction of SST enables the forecasting of extreme climate events ([Chowdary et al. 2010, 2011](#), [Kang & Shukla 2006](#)). Additionally, anomalies of SST are often used to identify and characterize El Niño or La Niña events. El Niño and La Niña are the coupled phase of the [El Niño-Southern Oscillation \(ENSO\)](#) phenomenon, where El Niño corresponds to the warming phase of the sea temperature and La Niña the cooling phase. The [Anomalies of Sea Surface Temperature Index \(SSTA\)](#) takes the difference between the observed SST and the climatological SST, where the climatological SST is created by fitting the time series of SST data to a cosine function, to capture the annual and semi-annual seasonality:

$$SST = A \cos(2\pi * t + B) + C \cos(4\pi * t + D) + E \quad (28)$$

The constants A and B capture the magnitude of the annual and semi-annual seasonality, and constants C and D correspond to the annual and semi-annual phase. The constant E denotes the long-term mean at the time of each observation. It is worth mentioning that the last constant term E can also be specified as  $E * t$  to account for the global warming effect. Since the resultant SSTA series doesn't exhibit prominent nonstationarity, there would be little difference in using either of the specifications.

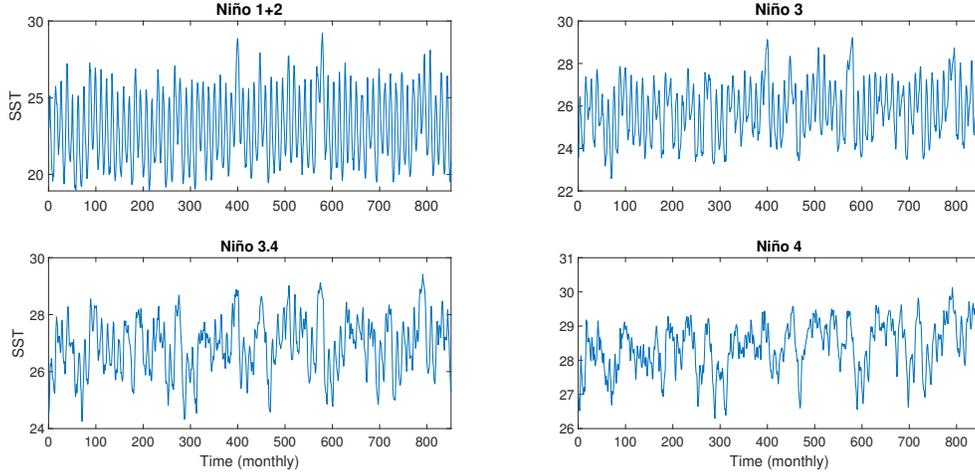


Figure 41: The plot of monthly average SST (in Celsius) in each Niño region.

Generally, an El Niño or La Niña event is identified when the SSTAs in one region or several regions exceed a defined threshold in one or in several consecutive months. For example, [Trenberth \(1997\)](#) suggests El Niño events can be declared if 5-month running means of SSTAs in the Niño 3.4 region exceed  $0.4^{\circ}\text{C}$  for six months or more. [Trenberth & Stepaniak \(2001\)](#) argue that the Niño 3.4 index should be used in combination with an index they introduce, called the [Trans-Niño Index \(TNI\)](#). The TNI is defined as the difference in normalized SST anomalies between the Niño 1+2 and Niño 4 regions. The Japan Meteorological Agency declares an El Niño event when the average five-month NINO 3 SSTA is over  $0.5^{\circ}\text{C}$  for six consecutive months or longer. However, there is no consensus on what constitutes an El Niño or La Niña event.

In the context of existing literature relating to SST and SSTA analysis, I apply the PE and PD measures on the SST data and SSTA data in the four specified regions, to measure their level of predictability and show the temporal dependence structures governing the dynamics of SST in each region. Our analysis uncovers the SST predictable limits in different ocean areas and indicates the maximum time frame that SST or SSTA can be predicted. Additionally, by making use of bivariate PE, I measure the bivariate dependence relation between the SSTA index in different regions. My analysis shows the similarity and difference between the dynamics of the SSTA index in different regions. With such information, insights can be provided regarding the informative and effective use of the SSTA index in various regions in characterizing the features of each El Niño or La Niña event.

The plots of  $1 - \text{PE}_{\tau}^{D=3}$ ,  $\text{PD}_{\tau}^{D=3}$  and sample ACF as functions of delay(lag) on each region's SST data are provided in [Figure 42](#). It shows clearly in the plots that the SST time series exhibits strong yearly and half-yearly seasonality, as evidenced by the regular pattern that

repeats itself every 12 delays in the plots of both measures. The seasonality in the SST data coincides with our expectations as the SST are affected by the seasonal weather in an annual cycle. What is more interesting is that the strength of the seasonality in the SST index varies significantly for different regions. From the magnitudes of the regular peaks of the PE and PD measures shown in the plots of different regions, the seasonality is most prominent in the SST of region Niño 1+2 and is weakened in Niño 3 and least displayed in Niño 3.4 and Niño 4.

The level of seasonality of the SST time series in each Niño regions suggested by the PE and PD plot is also verified by the monthly plot of SST data shown in Figure 43. In order to compute the SSTA, I fit the SST time series to a seasonal cosine function formulated as in (28), and take the difference between the original SST observations and the fitted smoothed periodic function. We can see from the plot that the seasonality constitutes the most dominating components in the SST in Niño 1+2 region, and is least evident and influential in the SST index of region Niño 3.4 and Niño 4.

After removing the seasonal components, the SSTA of each region is plotted in Figure 44. By computing the value of  $1 - PE_{\tau}^{D=3}$ ,  $PD_{\tau}^{D=3}$  and sample ACF on SSTA in the four Niño regions on various delays, we can see that the SSTA indexes are highly predictable. The value of  $1 - PE_{\tau}^{D=3}$  and  $PD_{\tau}^{D=3}$  are far from the insignificant level, especially in short delays. The PE and PD plots indicate the predictability of SSTA index are rather persistent where the temporal dependence between monthly SSTA index decays gradually and does not fully diminish until around delay 10. That means even though the accuracy and credibility deteriorate as the forecast horizon increases, the SSTA index can be predicted at most 10-months ahead. Additionally, the value of the PD measure shows that the SSTA in Niño 4 regions has the strongest temporal dependence structures among all Niño regions, whereas the SSTA index in region Niño 1+2 is least deterministic. The result indicates that the onset of El Niño and La Niña declared based on the SSTA index in Niño 4 or Niño 3 will be more regular and predictable compared to that based on the SSTA index in region Niño 1+2. An additional insight from the analysis is that the similar trend of PD and sample ACF as delay/lag increases suggests the form of temporal structure exhibited in SSTA is likely to be in linear form.

In addition to the investigation of the univariate dynamics of each SST and SSTA time series, we also employ the dependence measure  $H_{\tau}^{D=3}$  defined in section 3.9 to measure the dependence relation between the SSTA index in different regions. The dependence measure  $H_{\tau}^{D=3}$  stems from the multivariate PE and can detect any form of dependence relations between two or more individual time series. The dependence measure plots on the SSTA

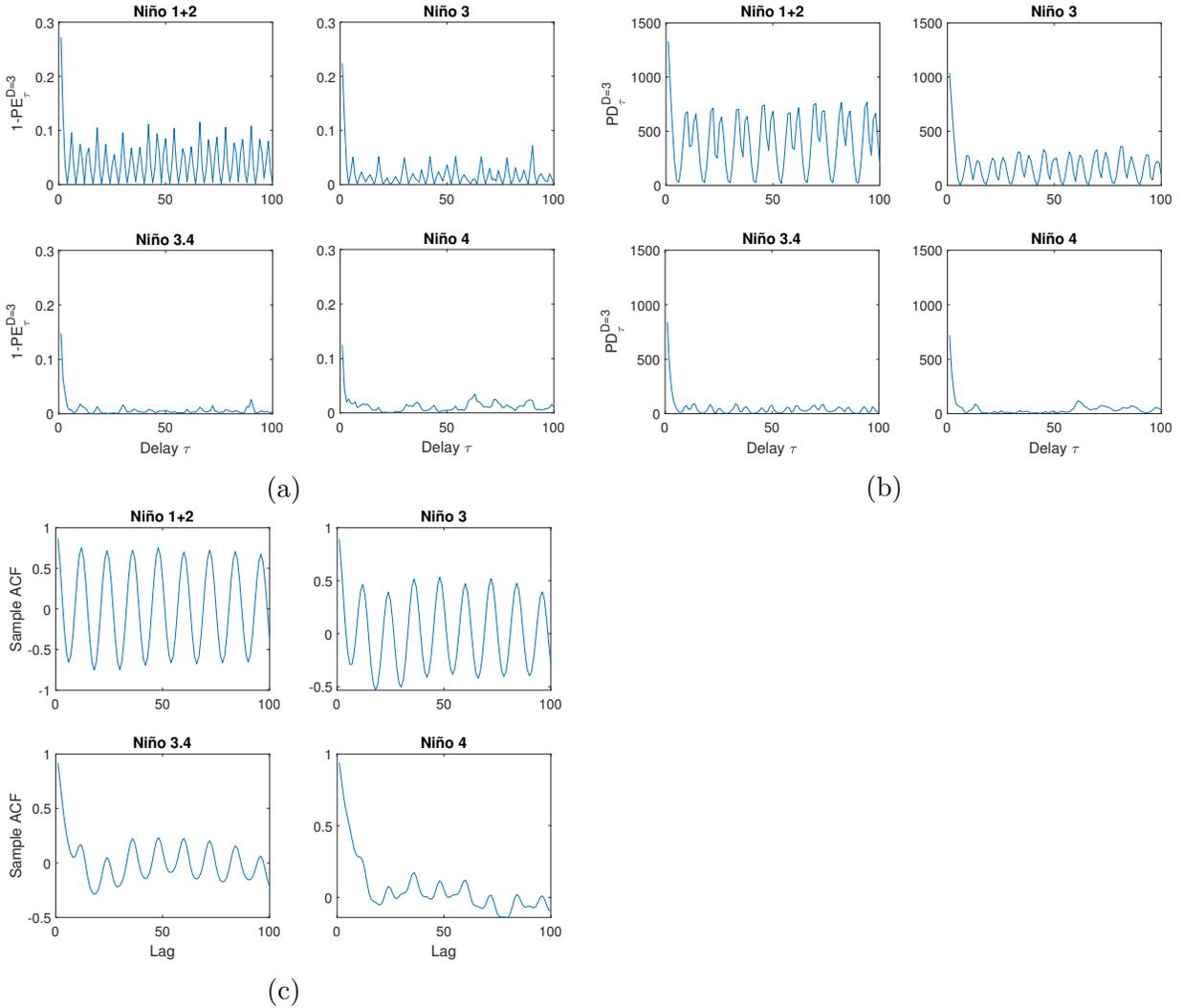


Figure 42: (a) Plot of  $1 - PE_{\tau}^{D=3}$  as a function of delay  $\tau$  on the monthly average SST time series in the four Niño regions. (b) Plot of  $PD_{\tau}^{D=3}$  as a function of delay  $\tau$  on the monthly average SST time series in the four Niño regions. (c) Plot of the sample ACF as a function of lag on the monthly average SST time series in the four Niño regions.

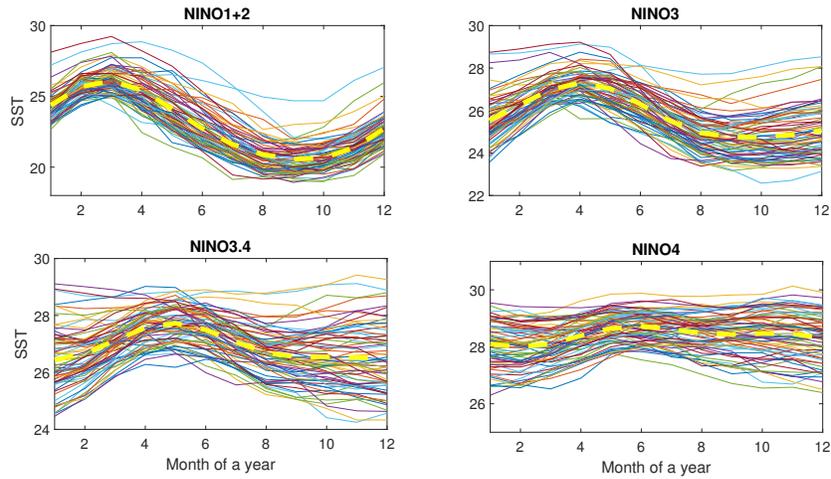


Figure 43: The plot of monthly average SST of each year in the four Niño regions. The yellow dashed line represents the fitted cosine function that captures the annual and the semi-annual seasonality exhibited in the SST index.

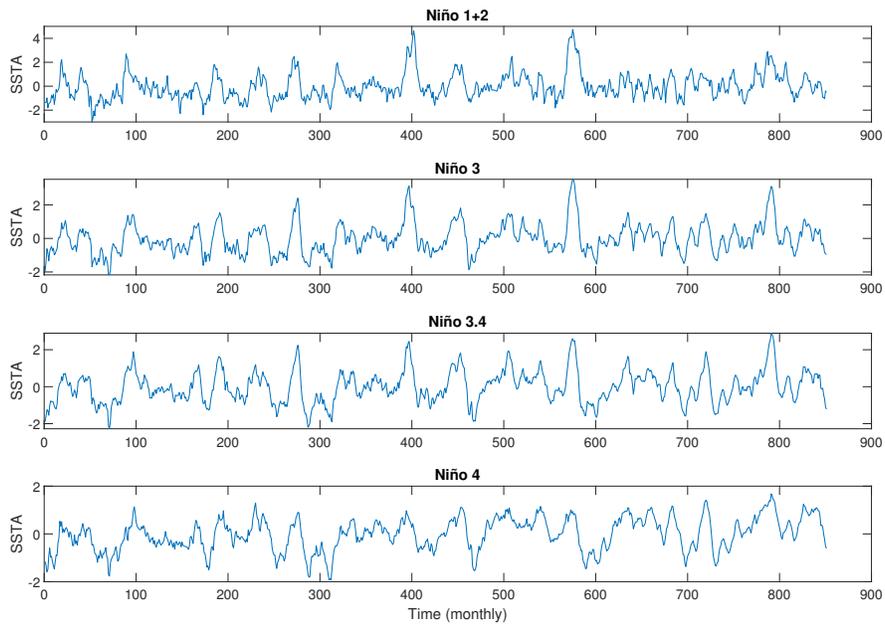


Figure 44: The plot of SSTA in each Niño region.

index from the six pairs of distinct ocean regions as functions of delay are shown in Figure 46. We also provide the cross-correlation plot of the pair of SSTA in different region in Figure 47 as a comparison. The plot of dependence measure indicates that the level of causality relation between the SSTA index from different regions is weaker at smaller delays and gradually increases to a steady level for longer delays. This observation implies that the overall trend of the SSTA indexes of the four regions are alike, but the short-term fluctuations of each region are distinct. Additionally, the strength of dependence between the SSTA in different regions is shown to accord with their distances. Among all combinations, the dependence relation between SSTA in Niño 1+2 with Niño 4 is the weakest as they are the furthest apart Niño areas and the dependence between SSTA in Niño 3 and Niño 3.4 regions are the strongest, and they share the largest common areas.

On the other hand, the cross-correlation plot reveals the dependence relation between the SSTA and the lagged or future SSTA in different regions. Since cross-correlation measures the correlation of two series as a function of the displacement of one relative to the other, the implication of cross-correlation is completely different from dependence measure  $H_7^D$ . The cross-correlation plot suggests the correlation between the SSTAs in different regions is strongest when there is no displacement of time. The level of correlation diminishes as the displacement of time increases. However, due to the specification of the cross-correlation, it is unable to reflect the evolution of the bivariate dependence between the investigated series from the short-term to long-term dynamics.

In summary, this section extends the application of PE and PD measures to SST time series. The empirical studies demonstrate the potential for the PE and PD measures to be used for a wide range of time series from different disciplines not only in applied finance. I showed that the PE and PD measures indicate that SST time series in different Niño regions exhibit different levels of seasonality and predictability, where SST in Niño 1+2 displays the most prominent seasonality and weakest predictability and SST in Niño 4 the opposite. This result is consistent with the general perception in the literature that the SST index in Niño 1+2 region tends to have the largest variation and the SST in Niño 4 tends to have more stable behaviour. Moreover, we found that all SSTA indexes are highly predictable and the predictability is rather persistent where the SST anomaly can be predicted up to around two seasons in advance. Additionally, the bivariate PE analysis suggests the long-term trends of SSTA index of all regions are alike, their distinctions reside in the short-term fluctuations. Their level of similarity is determined by the distance between the restricted areas in which they are recorded. Therefore in terms of efficiency, it is better to use the combination of SSTA index in further apart regions instead of adjacent regions to characterize the features

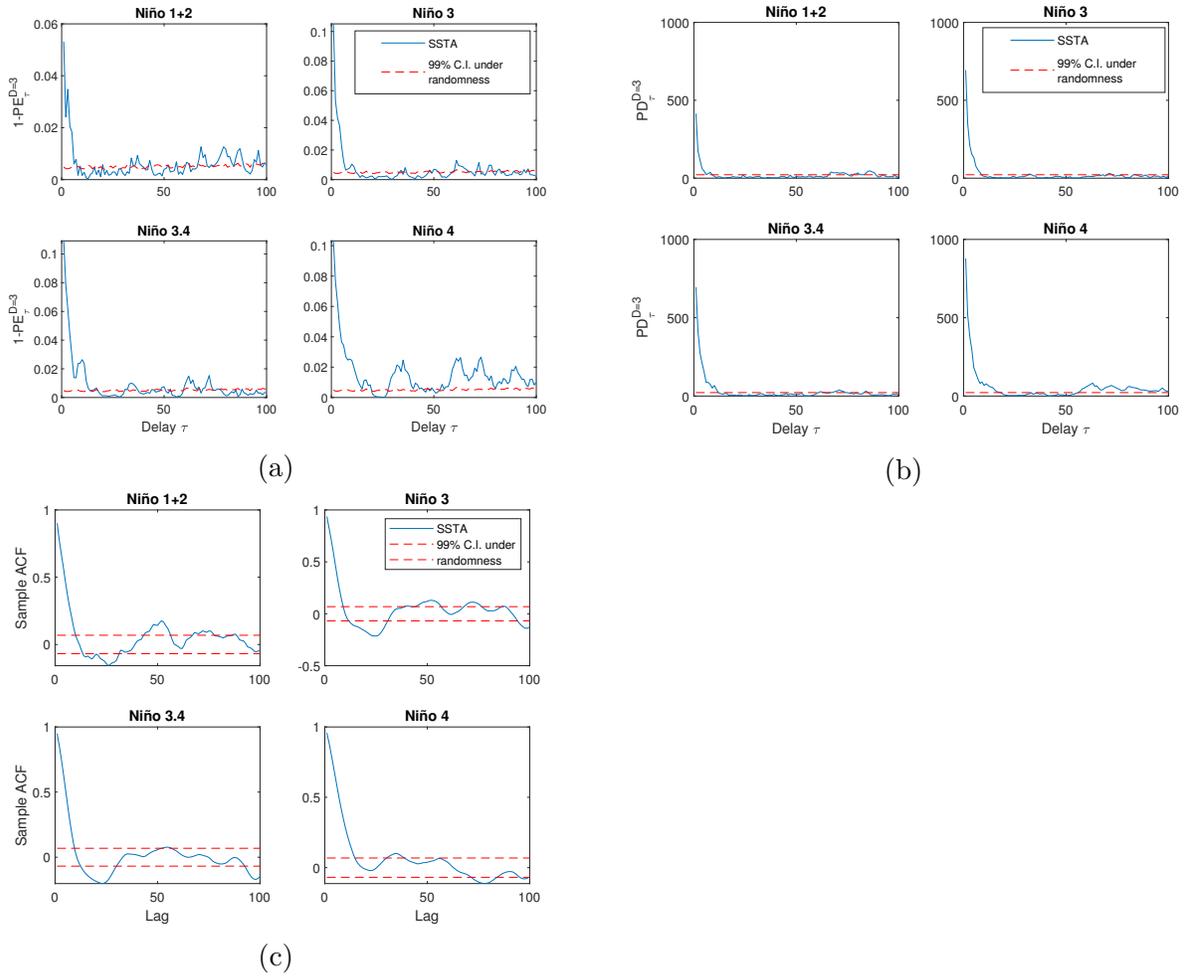


Figure 45: (a) Plot of  $1 - PE_{\tau}^{D=3}$  as a function of delay  $\tau$  on the SSTA time series in the four Niño regions. (b) Plot of  $PD_{\tau}^{D=3}$  as a function of delay  $\tau$  on the SSTA time series in the four Niño regions. (c) Plot of the sample ACF as a function of delay  $\tau$  on the SSTA time series in the four Niño regions.

of either El Niño or La Niña events.

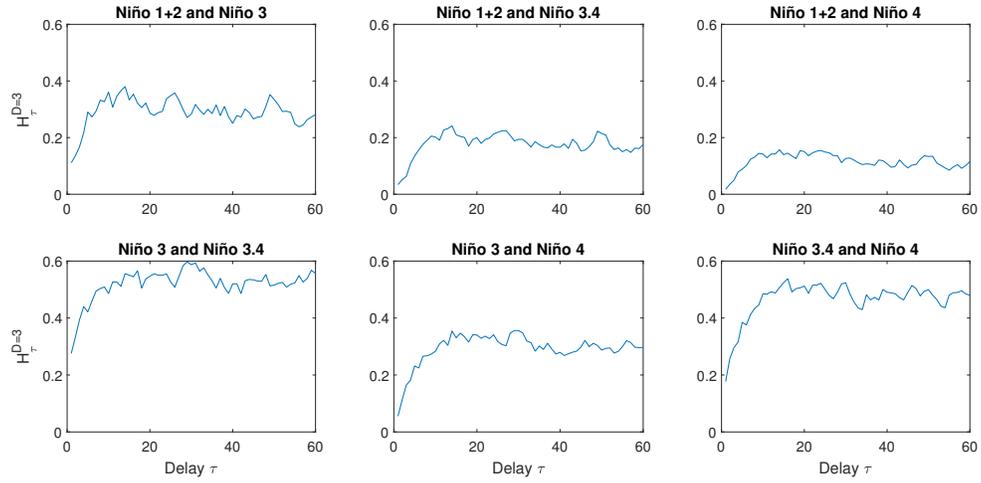


Figure 46: Plot of dependence measure  $H_{\tau}^{D=3}$  as function of delay on sea surface temperature anomaly (SSTA) recorded in two different regions.

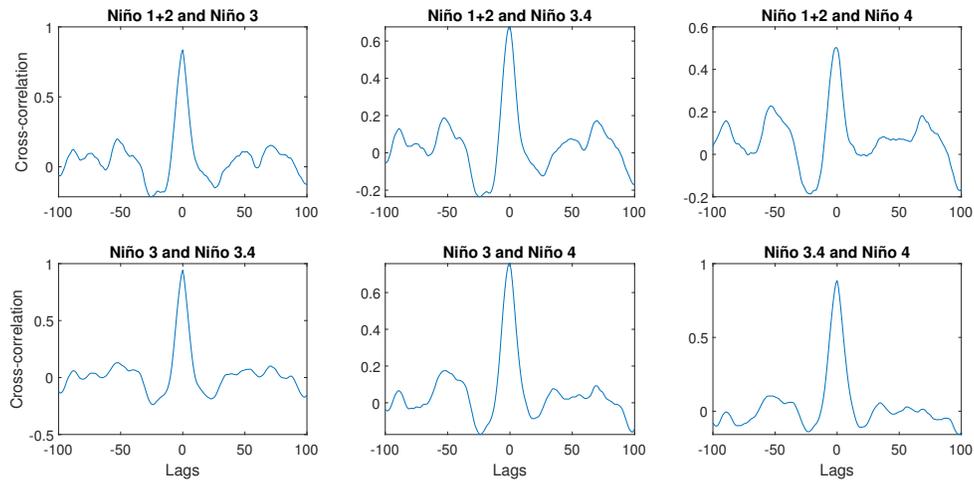


Figure 47: Plot of cross-correlation against lag on sea surface temperature anomaly (SSTA) recorded in two different regions.

## 6 Conclusion

The aim of this thesis is to extend the application of PE and the concept of ordinal pattern analysis in the area of financial time series analysis, and present an empirical example of the PE and ordinal pattern analysis on real-world financial time series.

### 6.1 Methodological contributions

Methodologically, I related the concept of PE with the ACF in a Gaussian process and the parametrization in the ARMA and GARCH models. In addition, I investigated the way that PE responds to the number of commonly observed properties of financial time series such as high kurtosis and non-stationarity. By doing so, I established a close connection between PE and the related ordinal pattern based analysis and the existing mainstream literature, and provided appropriate interpretations of the results obtained from the PE analysis tailored to the widely recognized characteristics of financial time series. In addition, the novel tools studies and developed in this thesis can be used as independent checks of the general proposition of the dynamics of high-frequency financial assets perceived by the existing literature. In particular, I extended these relatively novel techniques and ordinal pattern concepts in the following areas relating to financial time series analysis: temporal dependence detection; model selection; model interoperation; and prediction sufficiency evaluation.

For temporal dependence detection, a hypothesis test can be constructed based on the value of PE to distinguish between processes with temporal dependence structures against purely random processes. Moreover, statistic  $1 - PE_{\tau}^D$  is often plotted as a function of its pre-chosen parameter delay  $\tau$  to track the temporal dependence evolution between increasingly sparsely spaced out entries, thus providing insight into the underlying dynamics. However, I showed that the value of  $PE_{\tau}^D$  is not only determined by the strength of temporal dependence over the selected delay, but also the temporal dependence structures over  $D - 1$  multiples of the chosen delay, where  $D$  is the segment length parameter required in the computation of PE. The absence of one-to-one correspondence between PE and the temporal dependence at the selected delay complicates its interoperation and implication, and weakens its sensitivity in detecting determinism in a process that has significant but slowly varying temporal dependence structures for increasing lags. By adjusting the specification of PE, I created a new statistic PD so that PD is exempted from the influences of the temporal dependence relation over other delays besides the selected delay. Therefore, the new statistic PD exclusively measures the temporal dependence relation over the selected delay in an observed time series.

Statistic PD can be regarded as a non-linear alternative to the ACF. In fact, I showed that

the PD measure resembles the behaviour of the most widely used linear dependence measure ACF when the observed series follows a linear dynamic. Furthermore, it has an additional capability to reveal non-linear temporal dependence relations that ACF overlooks. I found that by plotting  $1 - PE_{\tau}^D$  and  $PD_{\tau}^D$  against increasing delays  $\tau$ , the PE plot and PD plot can be used together to detect the temporal dependence structures (both linear and non-linear) underlying the observed time series and reflect how temporal dependence relations diminish as the lag between the observations increases.

For model selection, I found that the probabilities of the ordinal patterns used in the computation of PE has the capability to reflect many fundamental features of the underlying dynamical system of a given time series, such as the functional form of the governing deterministic relation, the asymmetry, kurtosis and the existence of dynamic structures present in the innovations. As a result, systems with similar characteristics also share similar features in their ordinal pattern probability distribution. Therefore the ordinal pattern probability ranking can help suggest the promising parametric models to apply when investigating observed time series, and point out possible inadequacy in a selected model's specifications.

Additionally, I invented a visualization plot to reveal the deterministic structures captured by different models fitted to the investigated time series. The invented visualization technique does not require any knowledge of the functional form of the models employed and provides a universal framework to facilitate comparison between different deterministic relations postulated by both parametric and non-parametric models. By employing the statistic PD, the visualization graph tracks the level of relevance of the increasingly lagged observations in the prediction function postulated by different models. The visualization technique is particularly critical in understanding the prediction constructed by the non-parametric models that have no closed-form formula. Despite the wide-ranging success of non-parametric approaches in many applications, interpreting the results and comprehending the way they approximate the behaviour of the past observations dynamic are always the greatest challenge for many practitioners. Our invention helps to open the "Black Box", as many non-parametric models, such as machine learning approaches, are often described, and sheds new light on explaining their superior and suboptimal performance when applied to data with different properties. With the help of the newly proposed visualization technique, named the PD visualization plot, I found that non-parametric models SVR and GPR can be distorted by higher-moment temporal dependence structures present in the time series, especially when the observed data has relatively low signal to noise ratio. Additionally, in the presence of non-stationarity, SVR tends to generate conservative predictors and underestimate the "true"

deterministic relation.

This thesis also proposed a point forecast sufficiency test that fills the lack of conceptual cohesion between the two primary genres of model/prediction evaluation methods: the residual independence test and the prediction accuracy metrics. I showed that the new proposed test, named the PD model sufficiency test, can overcome the limitations of the mainstream model evaluation approaches. The test assesses a model's performance in fulfilling the point prediction task, which is the main objective for most forecasting practices.

Compared to the conventional model sufficiency test, my proposed test is not built on the assumption that a sufficient model is expected to eliminate all structures initially present in the observed data, thus resulting in independent and identically distributed residuals. I demonstrated that a sufficient point predicting model can generate non-white residuals if there exists higher than first-moment dependence relations in its underlying process. Therefore, the inferences drawn from the conventional sufficiency test, such as the well-known BDS test, can be misleading in evaluating a model's point prediction performance. With empirical evidence indicating the prevalence of higher-moment structures present in the financial return series ([Khademalomoom et al. 2019](#)), my proposed test provides a more reliable approach in establishing the point forecasting sufficiency of a given model in the area of financial time series analysis.

In contrast to the prediction accuracy metrics, such as the very commonly used MSE which aims to reflect the postulated predictors' point prediction accuracy, the PD model sufficiency test is more informative. It evaluates the performance of the considered model without the need to compare it with a benchmark. It can particularly reveal a model's point forecasting performance relative to the maximum prediction potential of its underlying dynamic.

To the best of our knowledge, my proposed test is the first attempt to assess the point forecasting sufficiency of the considered models . It is specially designed to remain valid for time series with non-white innovation when cast into an additive form.

## 6.2 Main results in empirical analysis

Empirically, the most critical contribution of this thesis is providing detailed demonstrations of how PE and the related ordinal pattern analysis reveal important information about the dynamics underlying the real-world financial series. I chose 10-minute interval EUR/USD exchange rate return series as the object of investigation. With the aid of PE and the newly proposed measure PD, I answered a number of fundamental questions about the investigated data.

- Do intraday EUR/USD return and volatility follow a purely random process or have temporal dependence structures?
- If temporal dependence structures are present in observed intraday return and volatilities, what is the nature of the detected structures and what contributes to them?
- How predictable (point and density) are intraday return and volatility without using any exogenous variables as inputs?
- Can conventional models, such as linear models and conditional heteroscedastic models, fully replicate the underlying dynamics in the intraday EUR/USD returns and volatilities?
- How well do mainstream models such as ARMA, GARCH, GPR and SVR models point predict EUR/USD intraday returns and volatilities? What are the primary reasons behind their sufficient/insufficient point forecasting performances?

The main empirical findings on intraday EUR/USD return series are as follows. The statistics PE and PD are shown to be significant only at delays less than 3, indicating the temporal dependence structure in the EUR/USD 10-minute returns mainly exhibits over the short-term. In other words, for high-frequency returns, their temporal dependence relations deteriorate rapidly and the impact from historical returns on future returns immediately vanishes after only two consecutive intervals. Moreover, with the aid of simulation studies, I was able to decompose the detected temporal dependence structures and identify their contributory sources. I found that the non-linear temporal dependent structure caused by the price rounding discretization convention is the dominant contributor to the temporal dependence structures within intraday returns. Further, the switching tendency in the sign of consecutive returns and the certain likely and unlikely price movement patterns that result in bivariate dependence relations between return sign and return magnitude constitute the remaining temporal dependence structures. However, compared to the discretization effect, which provides minimal prediction exploitation potential, the strength of the remaining temporal dependence structure caused by other factors is almost negligible, indicating marginal predictability of EUR/USD intraday returns.

In addition, I found that the simplest MA(1) model can exploit most of the prediction potential in predicting conditional mean of future returns, even though it can only increase the prediction accuracy by around 1% compared to using the current value. This implies that innovations constitute around 99% of the return dynamics. My results are consistent with previous literature that assesses the efficiency of foreign exchange markets using high-frequency returns and found that despite EUR/USD intraday returns exhibit significant

temporal dependence structures, the temporal dependence structures in returns are not able to be exploited for prediction purposes, especially in obtaining an accurate point prediction of future returns.

In contrast, intraday volatilities are more predictable compared to the return itself. I investigated 10-minute interval squared returns series and 1-hour realized volatility series constructed by aggregating six consecutive 10-minutes squared returns within a 1-hour interval. The most distinct feature detected in the dynamics of intraday volatilities is the seasonality mainly caused by the differences in trading times in the global foreign exchange markets. After removing the seasonality, gradually diminishing temporal dependence structures are detected in the deseasonalized volatility series. Unlike intraday returns that mainly exhibit short-term temporal dependence structures, the temporal dependence in intraday volatilities diminish gradually with increasing delays and remain significant up to  $\tau = 20$  to 30. The results are consistent the widely-documented perspective of many past studies that financial return volatility displays long-range dependence.

As for the underlying dynamics of EUR/USD intraday volatility, I cannot confirm the exact form of the governing process. However, a number of insights from our empirical analysis can enhance the understanding of the dynamics of the intraday volatility process. First, other than the deterministic positive correlation between lagged and current volatilities in their conditional mean, EUR/USD intraday volatility also exhibit temporal dependence relation in its second-moment. If the process is transformed into the form specified in (18), the temporal dependence structure exhibited in the innovation are very similar to that specified in the GARCH model where large lagged volatilities tend to lead to large dispersion of the distribution of future volatility. Second, by investigating the performance of the ARMA, GARCH, FIGARCH, GPR and SVR models in predicting squared returns and aggregated volatilities, I found their performances vary significantly across different investigation periods. Many of the employed models are able to improve prediction accuracy when compared to a random guess thus confirming the predictability of intraday volatilities. But there is no single model that outperforms the rest of its competitors across all considered periods. Almost all of the constructed predictors are not sufficient to exploit the maximum prediction potential. I found the inadequacy of the employed models was due to a number of reasons. First, the underlying dynamics of EUR/USD intraday volatility series are proven to have dependent innovations (or equivalently higher-order dependence structures). Our simulation study shows that dependent innovations can undermine the prediction ability of the non-parametric GPR and SVR models, especially when the observed series has a relatively low signal to noise ratio. Even though the simpler ARMA and GARCH models are robust to dynamical noise, under

non-stationarity, especially when the deterministic relations governing the process are time-varying, the ARMA and GARCH models try to estimate the average level of the linear serial relation underlying the investigated data, thereby overestimating the temporal structures in the sub-periods of relatively weaker deterministic dependence structures, and underestimating that in the sub-periods of stronger than average deterministic dependence structures. The SVR and GPR models are also affected by non-stationarity, but they (especially SVR) tend to construct a smoother function compared to the ARMA and GARCH models, thus are more conservative in approximating temporal dependence structures.

There are several other interesting findings from the empirical analysis. In order to remove the seasonality exhibited in the volatility dynamics, seasonal filters constructed by the FFF method and ASR method are employed for comparison. They are both capable of removing most of the periodicities, but neither performs perfectly. I found that discretization can be the reason behind the imperfect performance of seasonal filters. A simulation study indicates the discretization effect hampers the proper performance of the seasonal filters, especially when the investigated time series have strong baseline deterministic structures. Therefore, by eliminating the discreteness and converting the formerly discrete intraday squared return series into a continuous one, there is a chance that the performance of the previously imperfect seasonal filter can be further improved.

Additionally, by comparing the value of the newly proposed dependence measure PD and the associated ordinal pattern distributions on volatility series over different intervals, I found the underlying dynamics of the 1-hour aggregate volatilities are similar to that in 10-minutes volatilities, except the former is more structural than the latter, and the discretization effect is more influential in 10-minute squared return' dynamics but negligible in longer interval realized volatility series. Therefore, they can be characterized and predicted through similar models. Furthermore, by exploring six separate periods' data, I found that the strength of the temporal dependence structure underlying the intraday volatility and realized volatility varies across different selected periods, but their functional forms are similar in nature.

As for the empirical results on the SST index, monthly SSTs are highly predictable, and their predictability is rather persistent. Even though the SST indexes in different Pacific regions exhibit different levels of seasonality and determinism, their long-term trends are alike. Their level of distinction in the short-term fluctuations is determined by the distance between the regions in which they are recorded.



## References

- Abhyankar, A. N. (1995), ‘Return and volatility dynamics in the FTSE 100 stock index and stock index futures markets’, *The Journal of Futures Markets* **15**(4), 457–488.
- Altman, N. & Leger, C. (1995), ‘Bandwidth selection for kernel distribution function estimation’, *Journal of Statistical Planning and Inference* **46**(2), 195–214.
- Amigó, J. M., Kocarev, L. & Tomovski, I. (2007), ‘Discrete entropy’, *Physica D: Nonlinear Phenomena* **228**(1), 77–85.
- Ammermann, P. A. & Patterson, D. M. (2003), ‘The cross-sectional and cross-temporal universality of nonlinear serial dependencies: evidence from world stock indices and the taiwan stock exchange’, *Pacific-Basin Finance Journal* **11**(2), 175–195.
- Andersen, T. G. & Bollerslev, T. (1997), ‘Intraday periodicity and volatility persistence in financial markets’, *Journal of Empirical Finance* **4**(2-3), 115–158.
- Andersen, T. G. & Bollerslev, T. (1998), ‘Deutsche mark–dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies’, *The Journal of Finance* **53**(1), 219–265.
- Ashley, R. A., Patterson, D. M. & Hinich, M. J. (1986), ‘A diagnostic test for nonlinear serial dependence in time series fitting errors’, *Journal of Time Series Analysis* **7**(3), 165–178.
- Aziz, W. & Arif, M. (2005), Multiscale permutation entropy of physiological time series, in ‘2005 Pakistan Section Multitopic Conference’, IEEE, pp. 1–6.
- Bahraminasab, A., Ghasemi, F., Stefanovska, A., McClintock, P. V. & Kantz, H. (2008), ‘Direction of coupling from phases of interacting oscillators: a permutation information approach’, *Physical Review Letters* **100**(8), 084101.
- Baillie, R. T., Bollerslev, T. & Mikkelsen, H. O. (1996), ‘Fractionally integrated generalized autoregressive conditional heteroskedasticity’, *Journal of Econometrics* **74**(1), 3–30.
- Bandt, C. (2005), ‘Ordinal time series analysis’, *Ecological Modelling* **182**(3-4), 229–238.
- Bandt, C. & Pompe, B. (2002), ‘Permutation entropy: a natural complexity measure for time series’, *Physical Review Letters* **88**(17), 174102.
- Bandt, C. & Shiha, F. (2007), ‘Order patterns in time series’, *Journal of Time Series Analysis* **28**(5), 646–665.

- Belaire-Franch, J. & Opong, K. K. (2005), ‘Some evidence of random walk behavior of Euro exchange rates using ranks and signs’, *Journal of Banking & Finance* **29**(7), 1631–1643.
- Bian, C., Qin, C., Ma, Q. D. & Shen, Q. (2012), ‘Modified permutation-entropy analysis of heartbeat dynamics’, *Physical Review E* **85**(2), 021906.
- Blanco, S., Figliola, A., Quiroga, R. Q., Rosso, O. & Serrano, E. (1998), ‘Time-frequency analysis of electroencephalogram series. iii. wavelet packets and information cost function’, *Physical Review E* **57**(1), 932.
- Bollerslev, T. (1986), ‘Generalized autoregressive conditional heteroskedasticity’, *Journal of Econometrics* **31**(3), 307–327.
- Bollerslev, T. & Mikkelsen, H. O. (1996), ‘Modeling and pricing long memory in stock market volatility’, *Journal of Econometrics* **73**(1), 151–184.
- Brock, W. A., Hsieh, D. A., LeBaron, B. D., Brock, W. E. et al. (1991), *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence*, MIT Press, Cambridge, Massachusetts, London, England.
- Broock, W. A., Scheinkman, J. A., Dechert, W. D. & LeBaron, B. (1996), ‘A test for independence based on the correlation dimension’, *Econometric Reviews* **15**(3), 197–235.
- Bruzzo, A. A., Gesierich, B., Santi, M., Tassinari, C. A., Birbaumer, N. & Rubboli, G. (2008), ‘Permutation entropy to detect vigilance changes and preictal states from scalp eeg in epileptic patients. a preliminary study’, *Neurological Sciences* **29**(1), 3–9.
- Campbell, J. Y., Champbell, J. J., Campbell, J. W., Lo, A. W., Lo, A. W. & MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton University Press, Princeton, N.J.
- Cao, Y., Tung, W.-w., Gao, J., Protopopescu, V. A. & Hively, L. M. (2004), ‘Detecting dynamical changes in time series using the permutation entropy’, *Physical Review E* **70**(4), 046217.
- Caporin, M. (2003), ‘Identification of long memory in GARCH models’, *Statistical Methods and Applications* **12**(2), 133–151.
- Chordia, T., Roll, R. & Subrahmanyam, A. (2005), ‘Evidence on the speed of convergence to market efficiency’, *Journal of Financial Economics* **76**(2), 271–292.
- Chortareas, G., Jiang, Y. & Nankervis, J. C. (2011), ‘The random-walk behavior of the euro exchange rate’, *Finance Research Letters* **8**(3), 158–162.

- Chowdary, J. S., Xie, S.-P., Luo, J.-J., Hafner, J., Behera, S., Masumoto, Y. & Yamagata, T. (2011), ‘Predictability of northwest pacific climate during summer and the role of the tropical indian ocean’, *Climate Dynamics* **36**(3-4), 607–621.
- Chowdary, J., Xie, S.-P., Lee, J.-Y., Kosaka, Y. & Wang, B. (2010), ‘Predictability of summer northwest pacific climate in 11 coupled model hindcasts: Local and remote forcing’, *Journal of Geophysical Research: Atmospheres* **115**(D22).
- Cont, R. (2001), ‘Empirical properties of asset returns: Stylized facts and statistical issues’, *Quantitative Finance* **1**(2), 223–236.
- Costa, M., Goldberger, A. L. & Peng, C.-K. (2002), ‘Multiscale entropy analysis of complex physiologic time series’, *Physical Review Letters* **89**(6), 068102.
- de Lima, P. J. (1996), ‘Nuisance parameter free properties of correlation integral based statistics’, *Econometric Reviews* **15**(3), 237–259.
- De Micco, L., González, C., Larrondo, H., Martin, M., Plastino, A. & Rosso, O. (2008), ‘Randomizing nonlinear maps via symbolic dynamics’, *Physica A: Statistical Mechanics and its Applications* **387**(14), 3373–3383.
- Ebeling, W., Steuer, R. & Titchener, M. (2001), ‘Partition-based entropies of deterministic and stochastic maps’, *Stochastics and Dynamics* **1**(01), 45–61.
- Fadlallah, B., Chen, B., Keil, A. & Príncipe, J. (2013), ‘Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information’, *Physical Review E* **87**(2), 022911.
- Fernandes, M. & Preumont, P.-Y. (2012), ‘The finite-sample size of the BDS test for GARCH standardized residuals’, *Brazilian Review of Econometrics* **32**(2), 241–260.
- Gençay, R., Dacorogna, M., Muller, U. A., Pictet, O. & Olsen, R. (2001), *An Introduction to High-frequency Finance*, Princeton University Press, Princeton, N.J.
- Goodhart, C. A. & O’Hara, M. (1997), ‘High frequency data in financial markets: Issues and applications’, *Journal of Empirical Finance* **4**(2-3), 73–114.
- Grassberger, P. & Procaccia, I. (1983), ‘Characterization of strange attractors’, *Physical Review Letters* **50**(5), 346.
- Groth, A. (2005), ‘Visualization of coupling in time series by order recurrence plots’, *Physical Review E* **72**(4), 046220.

- Hanias, M. P. & Curtis, P. G. (2008), ‘Time series prediction of dollar\euro exchange rate index’, *International Research Journal of Finance and Economics* **15**(15), 232–239.
- Hinich, M. J. (1982), ‘Testing for Gaussianity and linearity of a stationary time series’, *Journal of Time Series Analysis* **3**(3), 169–176.
- Hsieh, D. A. (1989), ‘Testing for nonlinear dependence in daily foreign exchange rates’, *The Journal of Business* **62**(3), 339–368.
- Huang, S.-C., Chuang, P.-J., Wu, C.-F. & Lai, H.-J. (2010), ‘Chaos-based support vector regressions for exchange rate forecasting’, *Expert Systems with Applications* **37**(12), 8590–8598.
- Humeau-Heurtier, A., Wu, C.-W. & Wu, S.-D. (2015), ‘Refined composite multiscale permutation entropy to overcome multiscale permutation entropy length dependence’, *IEEE signal processing letters* **22**(12), 2364–2367.
- Jiang, J., Shang, P., Zhang, Z. & Li, X. (2017), ‘Permutation entropy analysis based on gini–simpson index for financial time series’, *Physica A: Statistical Mechanics and its Applications* **486**, 273–283.
- Joe, H. (1989), ‘Relative entropy measures of multivariate dependence’, *Journal of the American Statistical Association* **84**(405), 157–164.
- Jones, M. C. (1990), ‘The performance of kernel density functions in kernel distribution function estimation’, *Statistics & Probability Letters* **9**(2), 129–132.
- Jordan, D., Stockmanns, G., Kochs, E. F., Pilge, S. & Schneider, G. (2008), ‘Electroencephalographic order pattern analysis for the separation of consciousness and unconsciousness: an analysis of approximate entropy, permutation entropy, recurrence rate, and phase coupling of order recurrence plots’, *Anesthesiology: The Journal of the American Society of Anesthesiologists* **109**(6), 1014–1022.
- Kang, I.-S. & Shukla, J. (2006), Dynamic seasonal prediction and predictability of the monsoon, in ‘The Asian Monsoon’, Springer, pp. 585–612.
- Keller, K. & Wittfeld, K. (2004), ‘Distances of time series components by means of symbolic dynamics’, *International Journal of Bifurcation and Chaos* **14**(02), 693–703.
- Khademalomoom, S., Narayan, P. K. & Sharma, S. S. (2019), ‘Higher moments and exchange rate behavior’, *Financial Review* **54**(1), 201–229.

- Kowalski, A., Martín, M., Plastino, A. & Rosso, O. (2007), ‘Bandt–Pompe approach to the classical-quantum transition’, *Physica D: Nonlinear Phenomena* **233**(1), 21–31.
- Kreuzer, M., Kochs, E. F., Schneider, G. & Jordan, D. (2014), ‘Non-stationarity of EEG during wakefulness and anaesthesia: advantages of EEG permutation entropy monitoring’, *Journal of Clinical Monitoring and Computing* **28**(6), 573–580.
- Li, X., Cui, S. & Voss, L. J. (2008), ‘Using permutation entropy to measure the electroencephalographic effects of sevoflurane’, *The Journal of the American Society of Anesthesiologists* **109**(3), 448–456.
- Li, X., Ouyang, G. & Richards, D. A. (2007), ‘Predictability analysis of absence seizures with permutation entropy’, *Epilepsy Research* **77**(1), 70–74.
- Lien, D. & Xiang, J. (2010), ‘Convergence to efficiency in FTSE-100 futures market’, *International Journal of Financial Markets and Derivatives* **1**(3), 243–257.
- Lim, K.-P., Brooks, R. D. & Hinich, M. J. (2008), ‘Nonlinear serial dependence and the weak-form efficiency of asian emerging stock markets’, *Journal of International Financial Markets, Institutions and Money* **18**(5), 527–544.
- Liu, Y., Cizeau, P., Meyer, M., Peng, C.-K. & Stanley, H. E. (1997), ‘Correlations in economic time series’, *Physica A: Statistical Mechanics and its Applications* **245**(3-4), 437–440.
- Matilla-García, M. & Marín, M. R. (2008), ‘A non-parametric independence test using permutation entropy’, *Journal of Econometrics* **144**(1), 139–155.
- Matilla-García, M., Marín, M. R. & Dore, M. I. (2014), ‘A permutation entropy based test for causality: The volume–stock price relation’, *Physica A: Statistical Mechanics and its Applications* **398**, 280–288.
- Mikosch, T. & Stărică, C. (2004), ‘Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects’, *Review of Economics and Statistics* **86**(1), 378–390.
- Mischaikow, K., Mrozek, M., Reiss, J. & Szymczak, A. (1999), ‘Construction of symbolic dynamics from experimental time series’, *Physical Review Letters* **82**(6), 1144.
- Münnix, M. C., Schäfer, R. & Guhr, T. (2010), ‘Impact of the tick-size on financial returns and correlations’, *Physica A: Statistical Mechanics and its Applications* **389**(21), 4828–4843.
- Nelson, D. B. (1991), ‘Conditional heteroskedasticity in asset returns: A new approach’, *Econometrica: Journal of the Econometric Society* pp. 347–370.

- Olofsen, E., Sleight, J. & Dahan, A. (2008), ‘Permutation entropy of the electroencephalogram: a measure of anaesthetic drug effect’, *British Journal of Anaesthesia* **101**(6), 810–821.
- Onnela, J.-P., Töyli, J. & Kaski, K. (2009), ‘Tick size and stock returns’, *Physica A: Statistical Mechanics and its Applications* **388**(4), 441–454.
- Ouyang, G., Dang, C., Richards, D. A. & Li, X. (2010), ‘Ordinal pattern based similarity analysis for eeg recordings’, *Clinical Neurophysiology* **121**(5), 694–703.
- Packard, N. H., Crutchfield, J. P., Farmer, J. D. & Shaw, R. S. (1980), ‘Geometry from a time series’, *Physical Review Letters* **45**(9), 712.
- Parlitz, U., Berg, S., Luther, S., Schirdewan, A., Kurths, J. & Wessel, N. (2012), ‘Classifying cardiac biosignals using ordinal pattern statistics and symbolic dynamics’, *Computers in Biology and Medicine* **42**(3), 319–327.
- Patton, A. J. (2011), ‘Volatility forecast comparison using imperfect volatility proxies’, *Journal of Econometrics* **160**(1), 246–256.
- Perron, P. & Qu, Z. (2010), ‘Long-memory and level shifts in the volatility of stock market return indices’, *Journal of Business & Economic Statistics* **28**(2), 275–290.
- Peter D, H. (1985), ‘Kernel estimation of a distribution function’, *Communications in Statistics-Theory and Methods* **14**(3), 605–620.
- Powell, G. & Percival, I. (1979), ‘A spectral entropy method for distinguishing regular and irregular motion of Hamiltonian systems’, *Journal of Physics A: Mathematical and General* **12**(11), 2053.
- Reboredo, J. C., Rivera-Castro, M. A., Miranda, J. G. & García-Rubio, R. (2013), ‘How fast do stock prices adjust to market efficiency? Evidence from a detrended fluctuation analysis’, *Physica A: Statistical Mechanics and its Applications* **392**(7), 1631–1637.
- Robinson, P. (1977), ‘The estimation of a nonlinear moving average model’, *Stochastic Processes and Their Applications* **5**(1), 81–90.
- Rosso, O., Larrondo, H., Martin, M., Plastino, A. & Fuentes, M. (2007), ‘Distinguishing noise from chaos’, *Physical Review Letters* **99**(15), 154102.
- Rubinov, M. & Sporns, O. (2010), ‘Complex network measures of brain connectivity: Uses and interpretations’, *Neuroimage* **52**(3), 1059–1069.
- Sewell, M. (2011), ‘Characterization of financial time series’, *Rn* **11**(01), 01.

- Shang, P. et al. (2019), ‘Multiscale tsallis permutation entropy analysis for complex physiological time series’, *Physica A: Statistical Mechanics and Its Applications* **523**, 10–20.
- Silva, A., Campos, S., Monteiro, J., Venâncio, C., Costa, B., Guedes de Pinho, P. & Antunes, L. (2011), ‘Performance of anesthetic depth indexes in rabbits under propofol anesthesia: prediction probabilities and concentration-effect relations’, *The Journal of the American Society of Anesthesiologists* **115**(2), 303–314.
- Silva, A., Cardoso-Cruz, H., Silva, F., Galhardo, V. & Antunes, L. (2010), ‘Comparison of anesthetic depth indexes based on thalamocortical local field potentials in rats’, *Anesthesiology: The Journal of the American Society of Anesthesiologists* **112**(2), 355–363.
- Siokis, F. M. (2018), ‘Credit market jitters in the course of the financial crisis: A permutation entropy approach in measuring informational efficiency in financial assets’, *Physica A: Statistical Mechanics and its Applications* **499**, 266–275.
- Takens, F. (1981), Detecting strange attractors in turbulence, in ‘Dynamical Systems and Turbulence, Warwick 1980’, Springer, pp. 366–381.
- Taylor, S. J. & Xu, X. (1997), ‘The incremental volatility information in one million foreign exchange quotations’, *Journal of Empirical Finance* **4**(4), 317–340.
- Trenberth, K. (2020), ‘The climate data guide: Niño SST indices (Niño 1+2, 3, 3.4, 4; ONI and TNI)’.  
**URL:** <https://climatedataguide.ucar.edu/climate-data/nino-sst-indices-nino-12-3-34-4-oni-and-tni>
- Trenberth, K. E. (1997), ‘The definition of El Niño’, *Bulletin of the American Meteorological Society* **78**(12), 2771–2778.
- Trenberth, K. E. & Stepaniak, D. P. (2001), ‘Indices of El Niño evolution’, *Journal of climate* **14**(8), 1697–1701.
- Tzouras, S., Anagnostopoulos, C. & McCoy, E. (2015), ‘Financial time series modeling using the hurst exponent’, *Physica A: Statistical Mechanics and its Applications* **425**, 50–68.
- Vapnik, V. (1998), The support vector method of function estimation, in ‘Nonlinear Modeling’, Springer, pp. 55–85.
- Vapnik, V. (2013), *The Nature of Statistical Learning Theory*, Springer Science & Business Media, New York.
- Veisi, I., Pariz, N. & Karimpour, A. (2007), Fast and robust detection of epilepsy in noisy

- EEG signals using permutation entropy, *in* ‘2007 IEEE 7th International Symposium on Bioinformatics and Bioengineering’, IEEE, pp. 200–203.
- Williams, C. K. & Rasmussen, C. E. (2006), *Gaussian Processes for Machine Learning*, Vol. 2, MIT Press Cambridge, MA, Cambridge.
- Yan, R., Liu, Y. & Gao, R. X. (2012), ‘Permutation entropy: A nonlinear statistical measure for status characterization of rotary machines’, *Mechanical Systems and Signal Processing* **29**, 474–484.
- Yin, Y. & Shang, P. (2014), ‘Weighted multiscale permutation entropy of financial time series’, *Nonlinear Dynamics* **78**(4), 2921–2939.
- Zhao, X., Shang, P. & Huang, J. (2013), ‘Permutation complexity and dependence measures of time series’, *EPL (Europhysics Letters)* **102**(4), 40005.
- Zunino, L., Soriano, M. C., Fischer, I., Rosso, O. A. & Mirasso, C. R. (2010), ‘Permutation-information-theory approach to unveil delay dynamics from time-series analysis’, *Physical Review E* **82**(4), 046212.
- Zunino, L., Zanin, M., Tabak, B. M., Pérez, D. G. & Rosso, O. A. (2009), ‘Forbidden patterns, permutation entropy and stock market inefficiency’, *Physica A: Statistical Mechanics and its Applications* **388**(14), 2854–2864.