UNIVERSITY OF SOUTHAMPTON

Faculty of Social Science
Southampton Business School

Essays on Dynamic Efficiency in a Betting Market

by

Yifu Li

A thesis for the degree of
Doctor of Philosophy

June 2025

University of Southampton

Abstract

Faculty of Social Science
Southampton Business School

Doctor of Philosophy

Essays on Dynamic Efficiency in a Betting Market

by Yifu Li

This thesis introduces a novel framework for modelling the dynamic processes inherent in financial time series, focusing specifically on price fluctuations. The Dynamic Trend Analysis Approach (*DTAA*) is proposed to address critical issues found in existing literature, such as the ambiguity in the definition of trends and the challenges associated with structural breaks in time series data. Traditional models often rely on proxies like the mean or static trend assumptions, which can lead to substantial bias when trends are present, particularly over longer time intervals, and lake of capture the dynamic characteristics of time series. The *DTAA* aims to rectify these shortcomings by introducing a new concept: the ”Dynamic Trend” (*DT*), which captures both the time-varying nature of price movements and the evolving state of the data itself.

The methodology decomposes time series into two key components: the dynamic trend component, which reflects the broader price movements dynamically, and the volatility component, which captures short-term fluctuations. Central to this approach is the establishment of 3 key parameters, the Observer’s Time Points, the Fundamental (Primary) Time Level, and the Horizon. Assuming the Fundamental (Primary) Time Level and Horizon are predetermined, the price changes observed at different Observer’s Time Points vary, which means that the Dynamic Trend at each Observer’s Time Points reflecting the dynamic nature of price movements.

The market dynamic has two aspects, the information incoming and the market state, as shaped by all participant behaviours, evolve over time. Thus, we decompose *DT* into two dimensions: Information Affect Factor (*IAF*) and the Market Equilibrium Condition Factor (*MECF*). These parameters allow for a more nuanced representation of how market participants process and respond to external information, with the *DT*(*IAF* and *MECF*) evolving dynamically as time progresses.

Chapter 2 presents a detailed economic and mathematical definition of the DT concept, providing a theoretical foundation for understanding dynamic price movements. It also introduces a tangent proxy for approximating the DT, enhancing the precision of time series modelling. Chapter 2 applies the $DTAA$ to assess dynamic efficiency in five distinct betting markets, using a novel decomposition model to examine long-memory properties and informational inefficiencies. By decomposing the time series into trend-free volatility series and DT sequences, this approach overcomes the limitations of existing long-memory models, offering a robust tool for analysing market dynamics.

Chapter 3 explores the dynamic efficiency in betting markets using a novel decomposition model to analyse long-memory properties and informational inefficiency. By identifying integration orders (d) and constructing the degree of market inefficiency (D), it demonstrates that market efficiency improves as betting progresses, with cross-market patterns confirming this gradual increase. Additionally, the chapter introduces the Estimation Score for Integration Orders ($ESIO$), a method that optimises the combination of window size (WS), bandwidth (BD), and estimator for d. Four estimators (LW, ELW, FELW, Two-Step FELW) are tested, and the combination with the lowest $ESIO$ is selected for each market. Forecasts for the dynamic trend (DT) series in five markets using the FCVAR model show significant cointegrating relationships.

In Chapter 4, the analysis shifts focus to the dynamic interaction between market equilibrium conditions and external information inputs. This chapter emphasises the bounded nature of market responses, revealing how market reaction fluctuate due to the interplay between market liquidity, and the physical limitations of information processing. The introduction of the Information Affect Factor (IAF) and the Market Equilibrium Condition Factor ($MECF$) provides a clear framework for understanding the market's dynamic equilibrium.

Overall, the $DTAA$ framework contributes to the literature by offering a comprehensive method for capturing the structural breaks and dynamic processes in time series data. This method not only improves the robustness of traditional models but also provides new insights into the temporal evolution of market efficiency. It reveals the insight of the information processing in a system in a dynamic way.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my gratitude to Tapas Mishra, Johnnie Johnson, and Ming-Chien Sung for their help and guidance. I am deeply thankful to my parents, Jinlong Li and Xiaoling Ji, for their understanding, encouragement, and unconditional support. Special thanks to my primary supervisor, Tapas Mishra, for supporting my exploration into the unknown. I also appreciate my friends for their encouragement and companionship. My heartfelt thanks go to my close friends Dacheng Bao and Zongji Chen for their mutual support during the COVID-19 pandemic, helping us navigate through those challenging times together. Additionally, I am grateful to Jiacheng Shen, Xiaohan Yu, Shuyue Li, Yue Shi, Xiaobo Qu, Rui Ding, Yanchi Wu, and other friends for their companionship. Lastly, I would like to thank Southampton for providing sunny days each year.

To the endless journey toward truth, and the distant horizon that feels within reach.

Chapter 1

Introduction

1.1 Research Context and Background

In financial markets, prices are tangible reflections of the collective consensus formed by all market participants. This consensus is influenced by many factors, including external information such as economic indicators, political events, and market news, as well as internal factors such as market liquidity, participant behaviour, and trading strategies. The market, as a system created by these participants, continuously evolves as new information becomes available and as the internal state of the market changes.

At any given moment, the price of a financial instrument is determined by this consensus among all participants, based on the external information available to them and the current state of the market itself. Since both the external environment and the internal market conditions are dynamic, prices are inherently dynamic variables. Understanding how prices change over time is crucial for various stakeholders, including traders, investors, policymakers, and researchers.

Exploring the mechanisms of information transmission and price formation in the market, as well as studying how to profit from understanding and analysing price changes, is centred around investigating the dynamics of price movements. To accurately describe these dynamic changes, it is essential to extract the dynamic characteristics of the price movements from the data. This requires establishing a theoretical framework and employing technical methods capable of capturing the dynamic features of price time series.

In the existing literature, data patterns are often captured using concepts such as trend and mean. However, the the understanding of the data changing driven is frequently ambiguous, leading to confusion and potential bias in time series modelling. Many models exhibit unstable robustness in practical applications due to this ambiguity. In the mathematical modelling process, there is often no precise definition or distinction between trends and short-term fluctuations; instead, these concepts are differentiated relative to each other.

These issues contribute to the problem of structural breaks in time series modelling. Structural breaks occur when there are abrupt changes in the underlying data-generating process, which traditional offline models may fail to account for, but the online methods suffering from many issues like strong distributional assumptionsAlippi et al. (2016) or a predefined number of break points, black-box mechanismsAminikhanghahi and Cook (2017), merely identifying statistical shifts rather than understand the underlying data-generating process.

Time series data are sequences of data points collected or recorded at specific time intervals, inherently dynamic and characterised by changes over time. This dynamic nature implies that any structural analysis of time series data must consider the time-varying properties of the data. A static modelling approach may not adequately capture the evolving patterns and structures within the data.

To address these challenges, it is essential to establish a theoretical framework and employ technical methods capable of capturing the dynamic features of price time series.

The Dynamic Trend can be formally conceptualized as a measure of how a given dataset evolves dynamically over time, reflecting what we define as the velocity of the data. The generation of new data points is fundamentally a dynamic process characterized by the system transitioning from one equilibrium state to another, influenced by external disturbances.

Consider the analogy of price data as a vehicle constrained to move forwards or backwards along a one-dimensional numerical axis. Here, changes in the price correspond directly to changes in the vehicle's position along this axis, with upward movement indicating price increases and downward movement indicating price decreases. At each distinct point in time, the vehicle occupies a specific position, representing a corresponding price in the time series.

Under this framework, interpreting price changes merely as positional differences between discrete time points oversimplifies the underlying dynamics. Rather, the vehicle's transition from one position to another must be described through its velocity, with variations in velocity resulting in distinct positions even after equal intervals of time. Consequently, fluctuations in this velocity arise from alterations in the forces acting upon the vehicle, thereby reflecting changes in the state of the underlying dynamic system.

Within an economic context, these forces driving the velocity of price changes correspond to the aggregate behaviours and interactions of market participants. Hence, observed variations in time series data fundamentally represent outcomes from the evolving dynamic equilibrium of market forces.

Modelling this process involves defining key concepts such as the Fundamental (Primary) Time Level, the Observer's Time Point, and the Horizon, which provide a structured approach to understanding and modelling price changes. Note that, we consider the changing of price are decomposed into two parts. The dynamic trend-driven part is given the changing pattern that is related to information input, it describes how the system reacts to outside input and processes to equilibrium. The volatility part arises from the disorderly behaviour within a system; more precisely, it results from the unsystematic actions of each individual component that constitutes the system. The real data change is the sum of the dynamic trend-driven change and the volatility. And the real data is the real data changing plus the initial value. All the assessments below are based on the assumption that we will ignore the volatility part until we address it.

When capturing the dynamic characteristics of price fluctuations, it is crucial to understand how changes are reflected in the price time series and, based on this understanding, to describe the dynamic features of these changes. Price changes are represented in the time series as the difference between prices at different time points over a standard time interval. Due to the physical limitations of the trading system and our methods of observation, we can only describe price changes by observing and recording prices at discrete time points. Although price data may exist between two observations, we do not have precise information about these intermediate prices. Thus, a price time series consists of a sequence of observed values obtained at regular intervals. It will be different when the frequency of observing is different and/or the time point of observing is different.

It is evident that the price time series may vary depending on the chosen observation interval. Therefore, to describe price changes, we must first determine a standard time interval, which is the interval at which prices are observed. For computational and analytical convenience, we assume that all observation intervals are consistent, and this original observation interval is regarded as the standard time interval. We define this standard time interval as the Fundamental (Primary) Time Level. This also gives the observed frequency.

Based on this, price changes are defined as the difference between observed prices at different points in time. It is important to note that the order of observations is determined by the temporal sequence, and price changes are defined as the difference between the observed value at a specific moment and that at a future time, specifically using the future time's observed value minus the observed value at the specific moment.

Thus, we first need to define a specific moment in time, which will serve as a reference point. Assuming the observer is located at this moment, price changes can then be defined as the difference between the observed price at this reference point and the price at a future time point. This reference point is the basis for describing price changes, and we refer to this moment as the Observer's Time Point. Time points preceding this reference are referred to as the past, while those following it are referred to as the future. For simplicity, we assume time progresses uniformly from past to future, a concept we refer to as the "arrow of time."

Based on this, price changes are defined as the difference between the observed price at a future time point and the observed price at the Observer's Time Point. Furthermore, price changes can be classified into three categories: increase, decrease, and no change. An increase is defined when the observed price at a future time point is greater than that at the Observer's Time Point, which means when the difference between the future observed price and the Observer's Time Point is positive. A decrease is defined when the future observed price is less than the Observer's Time Point, which means when the difference is negative. No change is defined when the future observed price is equal to the Observer's Time Point, resulting in a difference of zero.

We have now defined the price change observed at a single Observer's Time Point over a specific time span at the Fundamental (Primary) Time Level. Thus, to describe any price change, three dimensions must be specified: the Fundamental (Primary) Time Level, the Observer's Time Point, and the time span over which the change is observed. We define this time span as the *Horizon*.

The above analysis provides an understanding of the price changes observed at a specific Observer's Time Point, which describes the observed price movement at a single point in time with respect to a given Fundamental (Primary) Time Level and Horizon. Now, we introduce the time variable and consider how the variation of time impacts price changes, thus capturing the temporal dynamics of price movements. When time changes, the Observer's Time Point also shifts. Assuming that the Fundamental (Primary) Time Level and Horizon remain constant, the observed price changes will vary due to the shift in the starting and ending points of observation. Specifically, the change in the Observer's Time Point causes a shift in the time interval used to observe price changes, which, in turn, alters the definition of the observed price movement. Therefore, changes in the Observer's Time Point capture the temporal evolution of price dynamics. As the Observer's Time Point moves sequentially through time, price changes evolve dynamically over time.

Since the fundamental unit of time used in the analysis is determined by the method of observation, the Fundamental (Primary) Time Level defines the basic temporal characteristics for analysis, which means the fundamental time interval at which prices are observed, and thus establishes the minimum unit of time change. The Horizon defines the state of the observer, specifically the observer's definition of the observation window. For example, in financial markets, if all traders in a particular market have the same Fundamental (Primary) Time Level, high-frequency traders will have a smaller

Horizon compared to long-term investors. The change in the Observer's Time Point captures the temporal dynamics of price changes.

In practice, both the Fundamental (Primary) Time Level and the Horizon may vary across different observers (participants in financial markets). Even for the same observer, the Fundamental (Primary) Time Level and the Horizon may change over time, varying at different Observer's Time Points.

Now, we further analyse the price changes observed at a given Observer's Time Point under constant Fundamental (Primary) Time Level and Horizon. At each time point, within the Horizon, the observed price may either increase, decrease, or remain unchanged compared to the price observed at the Observer's Time Point. The total price change within the Horizon is the sum of all individual changes(ignoring the volatility component). Dividing this sum by the number of time intervals within the Horizon yields the price change per unit of time, which we define as the Velocity of data changing. This definition refers to the Average Velocity of data change. As the Horizon approaches zero, or as the unit time interval given by the Fundamental (Primary) Time Level approaches zero, the Average Velocity of data changing converges to the Instantaneous Velocity of data changing.

We consider the variation of Velocity with respect to changes in the Observer's Time Point under constant Fundamental (Primary) Time Level and Horizon, which captures the dynamic characteristics of the time series. The Velocity at each time point within the Horizon gives the relative price. Since we assume that Velocity remains constant within the Horizon, the price at each time point within the Horizon follows a linear relationship over time, and this price line is referred to as the Temporal Trend Line. Therefore, we define Velocity as the Dynamic Trend.

For a price time series, price changes are dynamic. Assuming the Fundamental (Primary) Time Level and Horizon are predetermined, the price changes observed at different Observer's Time Points vary, which means that the Dynamic Trend at each Observer's Time Point reflects the dynamic nature of price movements.

Under the assumption that Velocity remains constant within the Horizon while the Observer's Time Point is unchanged, this assumption provides the definition of the basic time length of the trend, which is the time length of the Horizon represents the fundamental length of the trend. Consequently, there is only one fundamental temporal trend within the Horizon. Note that all the above assessments are under the assumption that no volatility is involved. When we consider volatility, the price change in a time span should be decomposed into two parts: the volatility and the overall movement. The overall movement is the change on average, which can be considered as the temporal trend, which is our Velocity; we define this as the Dynamic Trend. Here, we have the threshold of the trend, which is our Horizon, as the observer understands the market in a Horizon-length way. Every time the observer makes an observation, they will have an understanding of how the price is moving in general; this is the minimum length of the trend. If the temporal dynamic has some level pattern continuing for the length of the Horizon, then the data movement shows a continuous pattern, indicating the existence of a general trend. The remaining movement is the Volatility component. This indicates that the time series is not a mean-reversion process but more akin to a dynamic trend reversion process.

For a price time series, price changes are dynamic. Assuming that the Fundamental (Primary) Time Level and Horizon are predetermined, the price changes observed at different Observer's Time Points are different, reflecting the dynamic nature of price changes. The issue of structural breaks in modelling time series data in existing literature arises from neglecting the dynamic characteristics of the data.

Based on the above analysis, the traditional concept of trend (as understood in most of the existing literature) essentially reflects the consistency of the Dynamic Trend (DT) over a certain period. After extracting the DT sequence from the time series and decomposing the original time series into sub-series with different dynamic characteristics, each sub-series will only contain a single behavioural feature, represented by a specific single traditional trend. At this point, we can convert each sub-series into a trend-free sequence using a detrending method, and the combined trend-free sub-series will form the Volatility Series under this DT condition. Due to limitations such as liquidity and the physical basis of price formation, the DT sequence must show a zero reversion, with larger deviations from zero leading to faster reversion towards zero. In this way, we convert a structural break series with multiple trends into two trend-free sequences. Furthermore, the DT should be a sequence with a long-term mean of zero. If its mean is non-zero over a certain period, this indicates an overall upward (greater than zero) or downward (less than zero) trend during that period. From the perspective of information processing, this means that unprocessed information remains in the market, and the market is in the process of transitioning to a new equilibrium.

After decomposing the time series into the DT Series and the Volatility Series using the aforementioned method, the dynamic information is contained in the DT Series, which reflects the market's response to information. Therefore, when analysing market efficiency, we can better understand the dynamic evolution of market efficiency by analysing the long memory of the DT Series. Since the DT Series naturally lacks structural breaks, this greatly improves the robustness of existing long memory models for time series.

1.2 General Research Contributions

This research makes several significant contributions to the field of time series analysis and financial market modelling:

- **Proposal of Velocity Concept**: Introduces the concept of Velocity to explain the dynamic changes in data, defining velocity as the rate of change of data over time. This provides a precise measure of data dynamics in theory, capturing both the magnitude and direction of changes and forms the foundational basis for the Dynamic Trend (DT).

- **Mathematical Modelling of Dynamic Trends**: Constructs a comprehensive mathematical framework to model the DT, including the decomposition of the trend vector into Time Changing (TC) and Data Changing (DC) vectors. This framework precisely quantifies data dynamics and allows for the calculation of the velocity of data changes as the DT.

- **Addressing Structural Breaks**: Proposes methods to identify and model two types of structural breaks—*Type-I* (changes in direction) and *Type-II* (changes in magnitude)—thereby improving the robustness of time series models. By monitoring changes in the slope proxy of the Dynamic Trend vector, the methodology effectively detects structural breaks and adjusts the modelling approach accordingly.

- **Dynamic Decomposition of Time Series**: Introduces a novel decomposition method that separates time series data into the DT Series and the Volatility Series. This allows for a more nuanced analysis of data dynamics by isolating the dynamic trend component from the volatility component, effectively eliminating structural breaks and converting a series with multiple trends into two trend-free sequences. This also suggests the dynamic trend reversion of time series data.

- **Application to Market Efficiency Analysis**: Applies the proposed methodology to assess dynamic efficiency in betting markets, analysing long-memory properties and informational inefficiency. By identifying the market inefficiency series, the research reveals how market efficiency improves over time and demonstrates the evolving nature of market efficiency.

- **Integration of Information Processing and Market Equilibrium**: Extends the DT concept by decomposing it into the Information Affect Factor (IAF) and the Market Equilibrium Condition Factor ($MECF$). This provides insights into how markets process information and adjust towards equilibrium, illustrating the interplay between information inputs and market conditions through the relationship $(\text{IAF})^2 + (\text{MECF})^2 = 1$.

- **Comprehensive Model of Market Dynamics**: Develops a comprehensive model that integrates both IAF and $MECF$, explaining how market equilibrium adjusts in response to incoming information. The model includes stages of equilibrium, overreaction, and decay of overreactions, and considers constraints imposed by physical and economic factors, such as market liquidity and information transmission limitations.

1.3 Research Aims and Objectives

The primary aim of this research is to improve the modelling of financial time series data by developing a theoretical framework and technical methods that accurately capture the dynamic features of price movements. This includes establishing a precise definition of the data velocity(dynamic trends), addressing the issues of structural breaks, and improving the robustness and applicability of time series models in practical settings. The research seeks to provide tools for better understanding market dynamics, information processing, and market efficiency.

To achieve the research aim, the study sets out the following specific objectives:

Firstly, we critically assess the methods that modelling the mechanism how the information input and the market state drive the price data changing in time series modelling, identifying limitations that lead to unstable model robustness and potential biases due to ambiguity. The traditional offline methods often fail to capture the dynamic nature of financial data, resulting in models that are sensitive to structural changes and may not accurately represent underlying patterns; the online(dynamic) methods are suffering from strong distributional assumptionsAlippi et al. (2016) or a predefined number of break points, black-box mechanismsAminikhanghahi and Cook (2017), merely identifying statistical shifts rather than understand the underlying data-generating process (we will explain this in literature review in Chapter 2).

To address these limitations, we introduce the concept of **Velocity**, a theoretical construct that describes the rate at which data changes over time, providing a foundational understanding of data dynamics. The **Dynamic Trend** (DT) is the mathematical representation of this velocity, precisely defining and quantifying it within a time series model. The DT measures and analyses the theoretical velocity in practical terms, enabling a deeper and more accurate analysis of the dynamic features of financial time series data and linking the Velocity with the traditional understanding of trend. The Dynamic Trend incorporates key concepts such as the Fundamental (Primary) Time Level, Observer's Time Point, and Horizon to provide a structured approach to understanding price changes.

We proceed to develop mathematical models to capture the DT, including the decomposition into trend vector components (TC and DC) and defining the velocity of data changes.

We decompose the time series into DT Series and Volatility Series, converting structural break series with multiple trends into two trend-free sequences. This decomposition allows us to analyse the properties of the decomposed series more effectively, facilitating better understanding and forecasting. This analysis approach based on DT is the Dynamic Trend analysis approach ($DTAA$).

Further, we establish the relationship between DT, the Information Affect Factor (IAF), and the Market Equilibrium Condition Factor ($MECF$), forming a comprehensive framework for analysing time series data.

In addressing structural breaks, we identify and model Type-I and Type-II structural breaks within time series data, providing methods to detect and adjust for these changes by using $DTAA$, thereby improving model robustness. By accounting for structural breaks, we enhance the model's ability to adapt to changes in data patterns over time.

Applying the $DTAA$ methodology to betting markets, we assess dynamic efficiency, analyse long-memory properties, and investigate informational inefficiency. We use the market inefficiency (D) to quantify market efficiency and propose the Estimation Score for Integration Orders (ESIO) to obtain the optimal estimation of D.

Moreover, we integrate information processing by decomposing the DT into the Information Affect Factor (IAF) and the Market Equilibrium Condition Factor ($MECF$). This exploration of the relationship between information processing and market equilibrium dynamics analyses how markets respond to information inputs and adjust towards equilibrium, incorporating constraints imposed by physical and economic factors.

1.4 Structure of the Thesis

The thesis is structured into five chapters, each building upon the previous to develop a comprehensive understanding of dynamic trends in financial time series:

- **Chapter 1: Introduction**

  This chapter outlines the research context and background, presents the general research contributions, articulates the research aims and objectives, and provides an overview of the thesis structure.

- **Chapter 2: Theoretically and Mathematical Modelling of Dynamic Trends**

  Chapter 2 delves into the economic and mathematical definitions of Velocity(which is the Dynamic Trend (DT) in practice). Gives a framework and specific mathematics model to apply to time series data, which is the Dynamic Trend analysis approach ($DTAA$). It introduces specific mathematical models to capture the DT, including the trend vector and its components (TC and DC). The chapter also discusses the tangent proxy of the DT and addresses how the DT functions as a dynamic trend reversion process. It provides a detailed analysis of how to detect structural breaks by monitoring changes in the slope of the DT vector and distinguishes between *Type-I* and *Type-II* structural breaks. Moreover, this chapter also points out that cyclical patterns can be regarded as a special dynamic mode of the DT, and thus can be handled using $DTAA$.

- **Chapter 3: Assessing Dynamic Efficiency in Betting Markets**

  This chapter applies the $DTAA$ methodology to real-world data by assessing dynamic efficiency in betting markets. It introduces a novel decomposition model to analyse long-memory properties and informational inefficiency. By identifying the integration orders (d) and constructing a metric for market inefficiency (D), the chapter reveals the evolving nature of market efficiency over time. It also introduces the Estimation Score for Integration Orders (ESIO) to identify optimal combinations of window size, bandwidth, and estimators for accurately estimating integration orders. Furthermore, it employs the Fractionally Cointegrated Vector Autoregressive (FCVAR) model to assess the fractional cointegration relationships between DT series of different markets and make a prediction.

- **Chapter 4: Information Processing and Market Equilibrium Dynamics**

  Chapter 4 extends the analysis by integrating the DT into a comprehensive model that includes the Information Affect Factor (IAF) and the Market Equilibrium Condition Factor ($MECF$). As we have a tangent proxy of the DT, using Euler's formula to represent the DT in the complex plane, the chapter decomposes the DT into components that explain how market equilibrium adjusts in response to incoming information. It explores the dynamic and bounded nature of market responses, highlighting the interplay between information inputs, market liquidity, and inherent limitations in information processing. The chapter provides an explanation of how market equilibrium adjusts in response to incoming information, leading to stages of equilibrium, overreaction, and the subsequent decay of such overreactions.

- **Chapter 5: Conclusion**

  The final chapter summarises the key findings of the research, discusses the implications for time series modelling and financial market analysis, and suggests avenues for future research. It emphasises the contributions made towards understanding dynamic trends and improving the robustness and applicability of time series models. The chapter reflects on how the developed models can be used to better understand market dynamics and assist practitioners in making informed decisions.

Overall, this thesis aims to provide a comprehensive framework for modelling and understanding the dynamic nature of financial time series, addressing existing gaps in the literature, and offering practical tools for researchers and practitioners in the field.

Chapter 2

Title: Identification of Breaks in Betting Market Data and Smoothing: A Dynamic Trend Analysis Approach(DTAA)

2.1 Introduction

In financial markets, prices are the concrete manifestation of the consensus formed by all participants, and the system created by these participants can be referred to as the market. At any given moment, the price is determined by the consensus among all participants in the system, based on the external information available to them and the state of the system itself. Since both the external environment and the internal state of the market change over time, which means that external inputs and market conditions are dynamic, the price itself is also a dynamic variable. Whether we are exploring the mechanisms of information transmission and price formation in the market or studying how to profit from understanding and analysing price changes, the core issue is to investigate the dynamics of price movements. To describe these dynamic changes, we need to extract the dynamic characteristics of price movements. Therefore, it is essential to establish a theoretical framework and employ technical methods capable of capturing the dynamic features of price time series.

In the existing literature, the definition of a ”trend” in the existing literature is rather ambiguous. The confusion of trend definition leads to the result that many models exhibit unstable robustness in practical applications. In the process of mathematical modelling, there is no precise definition or distinction between trends and short-term fluctuations. This ambiguity in defining trends is a key reason why many models display unstable robustness in practical applications.

Using the mean changing as a proxy of the data changing pattern, like methods based on CUSUM originally developed by Page (1954), is a common way in the existing literature. Using the mean as a method for capturing data patterns and detecting mean shifts to describe and handle changes in time series can introduce significant bias when the data exhibit a trend.

The above issues are the reason why we have the structural break issue when modelling time series data. Another reason is that we are missing the data dynamic. Time series data are sequences of data points collected or recorded at specific time intervals. The inherent nature of time series generation is dynamic, characterised by changes over time. This dynamic nature implies that any structural analysis of time series data must consider the time-varying properties of the data. Therefore, a static modelling approach may not adequately capture the evolving patterns and structures within the data. This paper advocates for a dynamic modelling approach to analyse the structure of time series data. Note that we are aiming to find the dynamic pattern; in the following assessment when talking about the changing, we assume that the volatility part is not involved until we take it back and give the definition of the trend threshold.

To solve the above issues and have better modelling of time series data, we assess the nature of financial time series in a specific way. When capturing the dynamic characteristics of price fluctuations, it is essential to understand how changes are reflected in the price time series and, based on this understanding, to describe the dynamic features of these changes. Price changes are represented in the time series as the difference between prices at different time points over a standard time interval. Due to the physical limitations of the trading system and our methods of observation, we can only describe price changes by observing and recording prices at discrete time points. Although price data may exist between two observations, we do not have precise information about these intermediate prices. Thus, a price time series consists of a sequence of observed values obtained at regular intervals.

It is evident that the price time series may vary depending on the chosen observation interval. Therefore, to describe price changes, we must first determine a standard time interval, i.e. the interval at which prices are observed. For computational and analytical convenience, we assume that all observation intervals are consistent, and this original observation interval is regarded as the standard time interval. We define this standard time interval as the Fundamental (Primary) Time Level. Based on this, price changes are defined as the difference between observed prices at different points in time. It is important to note that the order of observations is determined by the temporal sequence, and price changes are defined as the difference between the observed value at a specific moment and that at a future time, specifically using the future time's observed value minus the observed value at the specific moment.

Thus, we first need to define a specific moment in time, which will serve as a reference point. Assuming the observer is located at this moment, price changes can then be defined as the difference between the observed price at this reference point and the price at a future time point, specifically using the future time's observed value minus the observed value at the reference point. This reference point is the basis for describing price changes, and we refer to this moment as the Observer's Time Point. The time points preceding this reference are referred to as the past, while those following it are referred to as the future. For simplicity, we assume that time progresses uniformly from past to future, a concept we refer to as the "arrow of time". The reference moment itself is referred to as the Observer's Time Point.

Based on this, price changes are defined as the difference between the observed price at a future time point and the observed price at the Observer's Time Point. In addition, price changes can be classified into three categories: increase, decrease, and no change. An increase is defined when the observed price at a future time point is greater than that at the Observer's Time Point, that is when the difference between the future observed price and the Observer's Time Point is positive. A decrease is defined when the future observed price is less than the Observer's Time Point, i.e., when the difference is negative. No change is defined when the future observed price is equal to the Observer's Time Point, resulting in a difference of zero.

We have now defined the price change observed at a single Observer's Time Point over a specific time span at the Fundamental (Primary) Time Level. Thus, to describe any price change, three dimensions must be specified: the Fundamental (Primary) Time Level, the Observer's Time Point, and the time span over which the change is observed. We define this time span as the Horizon.

The above analysis provides an understanding of the price changes observed at a specific Observer's Time Point, which describes the observed price movement at a single point in time with respect to a given Fundamental (Primary) Time Level and Horizon. Now, we introduce the time variable and consider how the variation of time impacts price changes, thus capturing the temporal dynamics of price movements. When time changes, the Observer's Time Point also shifts. Assuming that the Fundamental (Primary) Time Level and Horizon remain constant, the observed price changes will vary due to the shift in the starting and ending points of observation. Specifically, the change in the Observer's Time Point causes a shift in the time interval used to observe price changes, which, in turn, alters the definition of the observed price movement. Therefore, changes in the Observer's Time Point capture the temporal evolution of price dynamics. As the Observer's Time Point moves sequentially through time, price changes evolve dynamically over time.

Since the fundamental unit of time used in the analysis is determined by the method of observation, the Fundamental (Primary) Time Level defines the basic temporal characteristics for analysis, i.e., the fundamental time interval at which prices are observed, and thus establishes the minimum unit of time change. The Horizon defines the state of the observer, specifically the observer's definition of the observation window. For example, in financial markets, if all traders in a particular market have the same Fundamental (Primary) Time Level, high-frequency traders will have a smaller Horizon compared to long-term investors. The change in the Observer's Time Point captures the temporal dynamics of price changes.

In practice, both the Fundamental (Primary) Time Level and the Horizon may vary across different observers (participants in financial markets). Even for the same observer, the Fundamental (Primary) Time Level and the Horizon may change over time, varying at different Observer's Time Points.

Now, we further analyse the price changes observed at a given Observer's Time Point under constant Fundamental (Primary) Time Level and Horizon. At each time point, within the Horizon, the observed price may either increase, decrease, or remain unchanged compared to the price observed at the Observer's Time Point. The total price change within the Horizon is the sum of all individual changes. Dividing this sum by the number of time intervals within the Horizon yields the price change per unit of time, which we define as the Velocity of data changing. In fact, this definition refers to the Average Velocity of Data Change. As the Horizon approaches zero, or as the unit time interval given by the Fundamental (Primary) Time Level approaches zero, the Average Velocity of data changing converges to the Instantaneous Velocity of data changing.

We consider the variation of Velocity with respect to changes in the Observer's Time Point under constant Fundamental (Primary) Time Level and Horizon, which captures the dynamic characteristics of the time series. The Velocity at each time point within the Horizon gives the relative price. Since we assume that Velocity remains constant within the Horizon, the price at each time point within the Horizon follows a linear relationship over time, and this price line is referred to as the Trend Line. Therefore, we define Velocity as the Dynamic Trend. We consider the variation of Velocity with respect to changes in the Observer's Time Point under constant Fundamental (Primary) Time Level and Horizon, which captures the dynamic characteristics of the time series. The Velocity at each time point within the Horizon gives the relative price. Since we assume that Velocity remains constant within the Horizon, the price at each time point within the Horizon follows a linear relationship over time, and this price line is referred to as the Temporal Trend Line. Therefore, we define Velocity as the Dynamic Trend.

For a price time series, the price changes dynamically. Assuming the Fundamental (Primary) Time Level and Horizon are predetermined, the price changes observed at different Observer's Time Points vary, which means that the Dynamic Trend at each Observer's Time Point reflects the dynamic nature of price movements.

Under the assumption that Velocity remains constant within the Horizon while the Observer's Time Point is unchanged, this assumption provides the definition of the basic time length of the trend, i.e., the time length of the Horizon represents the fundamental length of the trend. Consequently, there is only one fundamental trend within the Horizon. Note that all the above assessments are under the assumption that no volatility part is involved. In fact, when we consider the volatility, the price changing in a time span should be decomposed into two parts, the volatility and the overall moving. The overall moving is the changing on average, which could be considered as the temporal trend, which is our Velocity, we define this as the Dynamic Trend. Here we have the threshold of trend which is our Horizon, as the observer understands the market in a Horizon-length way, every time the observer makes an observer that will have an understanding of how the price going in general, this is the minimum length of trend. If the temporal dynamic has some level pattern continuing for Horizon length, then the data movement shows a continuous pattern, which is given the existence of a general trend. The rest movement is the Volatility component. This indicates that the time series is not a mean reversion process but is more closely related to a dynamic trend reversion process.

For a price time series, price changes are dynamic. Assuming that the Fundamental (Primary) Time Level and Horizon are predetermined, the price changes observed at different Observer's Time Points are different, which reflects the dynamic nature of price changes. The issue of structural breaks in modelling time series data in existing literature arises from neglecting the dynamic characteristics of the data.

Based on the above analysis, the traditional concept of trend (as understood in most of the existing literature) essentially reflects the consistency of the Dynamic Trend (DT) over a certain period. After extracting the DT sequence from the time series and decomposing the original time series into sub-series with different dynamic characteristics, each sub-series will only contain a single behavioural feature, represented by a specific single traditional trend. At this point, we can convert each subseries into a trend-free sequence using a detrend method, and the combined trend-free subseries will form the Volatility Series under this DT condition. Due to limitations such as liquidity and the physical basis of price formation, the DT sequence must show a reversion of 0 with larger deviations from zero leading to a faster reversion towards zero. In this way, we convert a structural break series with multiple trends into two trend-free sequences. Furthermore, DT should be a sequence with a long-term mean of zero. If its mean is non-zero over a certain period, this indicates an overall upward (greater than zero) or downward (less than zero) trend during that period. From the perspective of information processing, this means that unprocessed information remains in the market, and the market is in the process of transitioning to a new equilibrium(the dynamic process of equilibrium is modelling in Chapter 4).

More specifically, this chapter introduces a comprehensive methodology for analysing time series data, focusing on the concept of the Dynamic Trend (DT), which encapsulates both the magnitude and direction of data changes over time. The Dynamic Trend is mathematically represented by a trend vector (TV), which is further decomposed into two components: the Time Changing vector (TC) and the Data Changing vector (DC).

The TC vector represents the passage of time and is aligned parallel to the x-axis, while the DC vector captures changes in the data and is aligned parallel to the y-axis. To assess the dynamic trend of the time series, the slope of the Dynamic Trend vector ($SlopeT$) is used as a proxy, which captures both the magnitude and direction of DC. This slope serves as a robust measure for analysing the dynamic behaviour of the series, offering insights into the extent and direction of data changes over time.

This framework distinguishes between two types of structural break: $Type - I$ and $Type - II$. $Type - I$ structural breaks occur when there is a change in the direction of the Dynamic Trend, indicating a shift between upward, flat, and downward trends. These breaks are detected by observing changes in the sign of $SlopeT$, known as $SignS$, which means a reversal or transition in the general direction of the trend.

In contrast, $Type - II$ structural breaks take place within segments that maintain the same directional trend but exhibit significant changes in the magnitude of the Dynamic Trend. These breaks are identified through changes in the absolute value of $SlopeT$ ($|SlopeT|$), which provide information on the variability of the magnitude of the trend within a given Type-I segment.

To effectively detect these structural breaks, the proposed methodology emphasises monitoring changes in the slope of the Dynamic Trend vector. These changes indicate shifts in both the direction and magnitude of the Dynamic Trend, thus enabling the identification of $Type - I$ and $Type - II$ breaks. By segmenting the time series based on these detected breaks, the framework allows a more nuanced understanding of the data's underlying dynamics and reveals important structural patterns.

Furthermore, the concept of Dynamic Trend also encompasses the velocity, acceleration, and deceleration of data movement, offering a comprehensive perspective on the dynamic characteristics of time series data. This allows for deeper insights into the temporal evolution of data patterns, enhancing the ability to model and predict time series behaviour.

Finally, the methodology also considers the impact of cyclical effects, particularly in cases where the cycle length exceeds four times the defined Horizon. This ensures that cyclical patterns are effectively incorporated into the analysis, providing a more complete understanding of the behaviour of the data set over time.

2.2 Breaking Point Detection Method: An Overview

Breaking point detection, also known as change point detection, aims to identify time points where the statistical properties of data change. These methods are broadly categorized into retrospective and sequential approaches. Retrospective methods process the entire dataset at once, allowing for global optimization and potentially more accurate detection, while sequential methods analyse data in real-time as it arrives which is dynamic.

2.2.1 Retrospective Breaking Point Detection Methods

2.2.1.1 Likelihood-Based Methods

Likelihood-based approaches determine change points by identifying where the likelihood function exhibits significant changes under different parameter assumptions.

Hinkley (1970) pioneered this area with the maximum likelihood estimator for detecting a single change in mean. His approach formalizes the problem as identifying where the log-likelihood ratio is maximized, providing asymptotic distribution theory for confidence interval estimation. This foundational work established statistical rigor in change point analysis and continues to influence modern methods.

The Chow test, introduced by **?**, represents another pioneering contribution to structural break detection. This test determines whether a single regression function or two separate regression functions (before and after a hypothesized break point) better fit the data. By comparing the sum of squared residuals from the combined model against the sum from two separate models, the Chow test provides an F-statistic to evaluate the significance of structural changes. This approach has been particularly influential in econometrics for identifying policy impacts, market shifts, and economic regime changes. However, as **?** noted, the Chow test requires prior specification of potential break points and assumes homoscedasticity across regimes, limiting its applicability in exploratory analyses where break locations are unknown.

Building on this foundation, Hawkins and Zamba (2005) introduced the CUSUM (cumulative sum) likelihood ratio tests for detecting multiple change points in multivariate data. Their approach sequentially segments data by identifying the most significant change point, then recursively applying the procedure to resulting segments. This binary segmentation strategy provides efficient computation while maintaining statistical guarantees under certain conditions.

Killick et al. (2012) addressed computational limitations in previous likelihood-based methods with their Pruned Exact Linear Time (PELT) algorithm. PELT achieves linear computation time through efficient pruning of solution space, making it feasible for much larger datasets than previous methods. The approach has become particularly valuable in genomic sequence analysis and financial time series, where data volumes are substantial.

While likelihood-based methods provide strong statistical foundations, they require explicit specification of probability distributions. As Aminikhanghahi and Cook (2017) noted, this parametric assumption can limit applicability in scenarios where underlying distributions are complex or unknown.

2.2.1.2 Bayesian Methods

Bayesian approaches frame change point detection within probability theory, treating change points as random variables with prior distributions that are updated based on observed data.

Barry and Hartigan (1993) developed the Bayesian multiple change point model using reversible jump Markov Chain Monte Carlo (RJMCMC) methods. Their approach not only detects change points but also quantifies uncertainty about their locations and number. This provides decision-makers with valuable information about confidence in detected changes, particularly useful in risk-sensitive applications.

Fearnhead (2006) introduced the exact Bayesian inference method for multiple change points, which efficiently computes the posterior distribution without resorting to MCMC approximations. Their recursive algorithm achieves polynomial time complexity, making Bayesian inference practical for many real-world datasets. The method has been successfully applied to DNA segmentation and financial market analysis.

Green (1995) provided key theoretical contributions with product partition models for Bayesian change point problems. Their work established a coherent framework for prior specification and posterior inference, facilitating application of Bayesian methods to diverse change point scenarios.

While Bayesian approaches offer robust uncertainty quantification, Truong et al. (2020) identified computational complexity as a persistent challenge, particularly for large datasets with many potential change points. Additionally, prior specification requires careful consideration and can significantly impact results.

2.2.1.3 Nonparametric Methods

Nonparametric approaches avoid strong distributional assumptions, making them more flexible for diverse data types and complex underlying processes.

Brodsky and Darkhovsky (1993) developed nonparametric methods based on ranks and U-statistics that remain robust to outliers and distributional violations. Their approach transforms data to invariant representations before applying change detection algorithms, demonstrating particular effectiveness in quality control applications where data may contain anomalies or follow non-standard distributions.

Kernel-based methods were advanced by Harchaoui et al. (2008), who introduced maximum mean discrepancy (MMD) statistics for change point detection. By mapping data to reproducing kernel Hilbert spaces, their approach detects distributional changes beyond simple mean shifts, making it valuable for complex multivariate time series. Their experiments demonstrated superior performance on climate data and image sequences compared to parametric alternatives.

Matteson and James (2014) introduced energy statistics for nonparametric change point analysis, which measure distributional differences without requiring explicit density estimation. Their E-Divisive algorithm recursively segments data based on divergence in energy statistics, effectively handling multivariate and non-Gaussian data. The method has found particular application in neuroimaging studies and financial market analysis.

As Arlot et al. (2019) observed, nonparametric methods provide flexibility but often require careful bandwidth or kernel selection. These choices can significantly impact detection performance, especially for datasets with limited samples or complex noise structures.

2.2.1.4 Optimization-Based Methods

Optimization approaches formulate change point detection as minimizing objective functions that balance data fidelity with penalties for excessive segmentation.

Jackson et al. (2005) introduced dynamic programming for exact optimization in change point detection, achieving polynomial-time solutions for previously intractable problems. Their approach guarantees global optimality for a wide class of objective functions, making it particularly valuable for applications requiring provably optimal solutions.

Extending this work, Bai and Perron (2003) developed least squares estimation methods for multiple structural changes, with applications to economic time series. Their approach formulates change point detection as minimizing residual sum of squares with appropriate penalties, providing consistency proofs and asymptotic distribution theory for detected change points.

Lavielle (2005) introduced penalized likelihood methods that automatically determine the number of change points through model selection criteria. Their approach adds a complexity penalty to the likelihood function, balancing goodness-of-fit with model simplicity. This method has been particularly successful in genomic sequence analysis and speech processing.

While optimization methods provide elegant mathematical formulations, Truong et al. (2020) noted their sensitivity to penalty parameter selection. These parameters significantly impact the number of detected change points, and optimal selection often requires domain knowledge or cross-validation procedures.

2.2.2 Sequential Breaking Point Detection Methods

2.2.2.1 Dynamic CUSUM Variants

Cumulative Sum (CUSUM) charts, originally developed by Page (1954), have been fundamental to change detection. For dynamic settings, several adaptive variants have emerged to address non-stationary data streams.

Verdier et al. (2008) introduced adaptive threshold computation for CUSUM procedures that automatically adjust control limits based on recent data characteristics. Their method significantly improves robustness in scenarios where baseline parameters drift gradually, a common challenge in industrial monitoring applications. The authors demonstrated a reduction in false alarm rates by 37% compared to standard CUSUM while maintaining detection sensitivity.

Building on this foundation, Granjon (2013) provided a comprehensive review of CUSUM algorithms for dynamic environments, introducing weighted variants that assign greater importance to recent observations. This modification addresses the challenge of detecting small but persistent shifts in process means, particularly valuable in quality control applications where traditional methods might fail due to data non-stationarity.

Despite these advances, dynamic CUSUM methods still face limitations. As Alippi et al. (2016) noted, they often make assumptions about underlying distributions (typically Gaussian) that may not hold in complex real-world scenarios. Additionally, these methods require careful parameter selection and may struggle with gradual changes that occur over extended periods.

2.2.2.2 Sequential Bayesian Methods

Bayesian approaches to change point detection provide probabilistic frameworks that naturally incorporate uncertainty quantification and adapt to evolving data characteristics.

The seminal work by Adams and MacKay (2007) on Bayesian Online Changepoint Detection (BOCPD) introduced a framework that computes the posterior probability distribution over the run length (time since the last change point) as new data arrives. This method provides not just change point estimates but also quantifies detection uncertainty, making it valuable for risk-sensitive applications. BOCPD has proven particularly effective in financial time series analysis where regime shifts carry significant implications (Ruggieri and Antonellis, 2016).

Fearnhead and Liu (2007) further advanced sequential Bayesian approaches by developing efficient particle filtering methods for dynamic change point detection. Their algorithm maintains a set of weighted particles representing possible segmentations of the observed data, enabling detection of multiple change points with linear computational complexity relative to data stream length. This efficiency makes the approach feasible for real-time applications with stringent latency requirements.

While sequential Bayesian methods offer theoretical elegance and uncertainty quantification, they come with practical challenges. Saatçi et al. (2010) identified computational intensity as a significant limitation, particularly for high-dimensional data streams. Additionally, these methods require specification of prior distributions, which can be difficult without domain expertise. Knoblauch et al. (2018) also noted that particle degeneracy—where most particles eventually receive negligible weights—remains a challenge for long sequences.

2.2.2.3 Dynamic Kernel Methods

Kernel-based approaches leverage the "kernel trick" to map data into high-dimensional feature spaces, enabling detection of complex, non-linear changes without explicitly specifying the transformation.

Li et al. (2015) introduced M-statistics for kernel change-point detection that operate effectively in dynamic settings. Their method constructs maximum mean discrepancy (MMD) statistics between adjacent windows in the kernel space, allowing for detection of distributional changes that might be invisible in the original feature space. Empirical evaluations showed superior performance compared to parametric approaches when handling complex data patterns.

More recently, Keriven et al. (2020) proposed NEWMA (No-prior-Knowledge Exponential Weighted Moving Average) method. NEWMA is an dynamic, model-free change-point detection algorithm designed for high-dimensional time series data. It operates by computing two Exponential Weighted Moving Averages (EWMAs) with different forgetting factors, effectively comparing recent data to older observations without storing the entire data history. To enhance its capability, NEWMA employs Random Features (RFs) to approximate the Maximum Mean Discrepancy (MMD), allowing for efficient detection of distributional changes in the data stream. The performance of NEWMA can be sensitive to the choice of forgetting factors and kernel parameters and may exhibit a delay in detecting changes, especially in scenarios with subtle distributional shifts.

As Chang et al. (2019) observed, kernel methods face challenges in selecting appropriate kernel functions and bandwidth parameters, which significantly impact detection performance. Additionally, these methods can impose substantial memory requirements as reference data grows, though recent advances in random feature approximations have partially addressed this concern (Liu et al., 2013).

2.2.2.4 Adaptive Windowing Techniques

Adaptive windowing approaches automatically adjust analysis window sizes based on data characteristics, balancing detection sensitivity with false alarm rates.

The ADWIN (ADaptive WINdowing) algorithm, introduced by Bifet and Gavalda (2007), maintains a variable-length window of recent observations that grows when the data is stationary and shrinks when changes are detected. The method provides theoretical guarantees on false positive rates and memory bounds, making it particularly suitable for resource-constrained environments. ADWIN has been widely adopted in stream mining applications, where it effectively handles concept drift in classification tasks (Gama et al., 2014).

Despite their elegance, adaptive windowing techniques face challenges with seasonal or periodic data patterns, as identified by Montero-Manso and Hyndman (2021). Additionally, these methods may introduce detection delays for subtle changes that only become apparent over extended observation periods.

2.2.2.5 Density Ratio Estimation

Density ratio estimation directly targets the ratio between probability densities before and after potential change points, avoiding the need to estimate full distributions.

Liu et al. (2013) pioneered the application of RuLSIF (Relative unconstrained Least-Squares Importance Fitting) to change point detection in time series. Their approach estimates the relative density ratio between consecutive time windows, detecting changes when this ratio deviates significantly from unity. This method proved particularly effective for high-dimensional data where traditional density estimation would suffer from the curse of dimensionality.

Building on this framework, Sugiyama et al. (2012) developed a unified theoretical understanding of density-ratio matching under Bregman divergence. Their work provided systematic guidance for selecting appropriate divergence measures based on data characteristics and application requirements, significantly advancing the practical applicability of these methods.

While density ratio estimation offers elegant solutions for complex distributions, Yamada et al. (2013) identified several practical limitations. These methods typically require sufficient sample sizes in analysis windows, making them less effective for detecting changes immediately after their occurrence. Additionally, the techniques can be computationally demanding, though recent algorithmic improvements have enhanced efficiency for dynamic settings.

2.2.2.6 Deep Learning Approaches

Neural network architectures have increasingly been applied to change point detection, leveraging their ability to learn complex patterns directly from data.

Schmidl et al. (2022) explored recurrent neural network (RNN) and long short-term memory (LSTM) architectures for anomaly detection in multivariate time series. Their approach learns normal behavior patterns and identifies deviations that suggest change points, demonstrating particular effectiveness in scenarios with complex temporal dependencies. The method achieved a 24% improvement in F1-score compared to traditional statistical approaches when evaluated on industrial sensor data.

Convolutional approaches were advanced by Karim et al. (2018), who developed LSTM Fully Convolutional Networks for time series classification and change point detection. Their architecture combines convolutional layers for feature extraction with specialized LSTM layers that identify pattern shifts. The method showed remarkable adaptability across diverse domains from sensor data to financial time series, achieving state-of-the-art performance in detecting subtle regime changes.

Wen et al. (2020) provided a comprehensive survey of time series data augmentation techniques for deep learning approaches to change point detection. Their work systematically analyzed how various augmentation strategies impact model performance, revealing that temporal warping and mixture-based methods significantly improve robustness when detecting anomalies and change points in noisy real-world data streams.

Despite impressive performance, deep learning approaches face significant challenges. As Aminikhanghahi and Cook (2017) noted, these methods typically require substantial training data and computational resources. The black-box nature of neural networks also limits interpretability, which can be problematic in safety-critical applications where understanding detection rationale is essential.

2.2.2.7 Ensemble Methods

Ensemble techniques combine multiple detection algorithms to improve robustness and accuracy compared to individual methods.

Frias-Blanco et al. (2015) introduced HDDM (Hoeffding Drift Detection Method), which combines multiple statistical tests with theoretical guarantees derived from Hoeffding's inequality. This ensemble approach maintains low false positive rates while remaining sensitive to various change types. Empirical evaluations demonstrated its effectiveness across diverse streaming data scenarios including network traffic analysis and sensor monitoring.

A comprehensive comparison of ensemble detectors was conducted by Barros and Santos (2018), who evaluated 15 different drift detection methods across 95 datasets. Their findings revealed that ensemble approaches consistently outperformed individual methods, with particular advantages in scenarios involving mixed types of changes (e.g., gradual and abrupt changes occurring in sequence).

While ensemble methods offer improved accuracy, they introduce additional complexity in implementation and parameter tuning. Lu et al. (2018) observed that computational overhead can be prohibitive for real-time applications with stringent latency requirements, necessitating careful algorithm selection and optimization.

2.2.2.8 Streaming Feature Selection

Feature selection in streaming contexts identifies the most relevant attributes dynamically, potentially improving detection performance while reducing computational demands.

Wu et al. (2012) introduced Online Streaming Feature Selection (OSFS), which progressively evaluates feature relevance as new data arrives. The algorithm maintains a compact set of strongly relevant features while efficiently handling high-dimensional data streams. Experiments demonstrated that OSFS could maintain detection accuracy while reducing computational requirements by up to 70% compared to using full feature sets.

As Yu and Liu (2012) observed, streaming feature selection methods can miss changes that manifest primarily in features deemed irrelevant during stable periods. Additionally, these approaches introduce preprocessing overhead and require careful threshold selection to balance dimensionality reduction with information preservation.

2.2.3 Comparative Analysis

Table 2.1 provides a consolidated comparison of breaking point detection methods.

Method Category	Representative Methods	Key Papers	Strengths	Limitations
Likelihood-Based	CUSUM, Binary Segmentation, PELT, Chow test	Killick (2012), Hawkins (2005), Chow (1960)	• Strong theory • Computationally efficient • Statistical hypothesis testing framework	• Parametric assumptions • Less flexible with complex data • Requires prior specification of break points
Bayesian Methods	RJMCMC, Exact Bayesian	Barry (1993), Fearnhead (2006)	• Quantifies uncertainty • Flexible modeling	• Computational complexity • Prior sensitivity
Nonparametric Methods	MMD, Energy statistics	Matteson (2014), Harchaoui (2008)	• Robust to outliers • Few assumptions	• Kernel/ bandwidth selection sensitive

Continued on next page

Method Category	**Representative Methods**	**Key Papers**	**Strengths**	**Limitations**
Optimization-Based	Dynamic Programming, Penalized likelihood, Least squares estimation	Jackson (2005), Lavielle (2005), Bai (2003)	• Global optimal solutions • Strong theoretical basis • Multiple structural changes	• Sensitive to penalty tuning • Computational complexity increases with breaks
Dynamic CUSUM Variants	Adaptive CUSUM	Verdier (2008), Granjon (2013)	• Dynamic detection • Adaptive thresholding	• Distribution assumptions • Poor gradual-change detection
Sequential Bayesian	BOCPD	Adams (2007), Fearnhead (2007)	• Uncertainty quantification • Probabilistic modeling	• High computational cost • Particle degeneracy
Online Kernel Methods	Kernel M-statistics, NEWMA	Li (2015), Keriven (2020)	• High-dimensional suitability • Nonlinear detection	• Kernel parameter sensitive • High memory demand
Adaptive Windowing	ADWIN	Bifet (2007), Gama (2014)	• Dynamic window adjustment • Theory-backed	• Delay in subtle-change detection • Issues with periodic data

Continued on next page

Method Category	**Representative Methods**	**Key Papers**	**Strengths**	**Limitations**
Density Ratio Estimation	RuLSIF-based	Liu (2013), Sugiyama (2012)	• Effective in high dimensions	• Requires adequate samples • Computationally intensive
Deep Learning	RNN, LSTM, CNN	Karim (2018), Schmidl (2022)	• Captures complex patterns • High accuracy	• Requires large datasets • Black-box nature • Resource-intensive
Ensemble Methods	HDDM	Frias-Blanco (2015), Barros (2018)	• Robust to diverse changes • Improved accuracy	• Higher complexity • Computational overhead
Streaming Feature Selection	OSFS	Wu (2012), Yu (2012)	• Computational efficiency • Real-time relevance	• May miss subtle changes • Preprocessing overhead

TABLE 2.1: Summary of Change Point Detection Methods

Despite significant advances in breaking point detection, existing methods present various limitations that constrain their applicability and interpretability across real-world economic settings.

The evolution of breaking point detection methods reflects a progression from simple, retrospective analyses of single structural breaks to increasingly sophisticated, adaptive, and scalable approaches for dynamic and high-dimensional data. Early likelihood-based methods, such as those proposed by Hinkley (1970) and **?**, laid a rigorous statistical foundation for change point analysis. These approaches provided analytical clarity and asymptotic guarantees but required parametric assumptions and prior specification of break locations, limiting their flexibility in exploratory or real-time applications (**?**).

Advancements in this area led to recursive and scalable techniques such as the CUSUM likelihood ratio tests (Hawkins and Zamba, 2005) and the PELT algorithm (Killick et al., 2012), which improved computational efficiency and allowed for detection of multiple change points. However, they still relied on distributional assumptions, making them less robust in settings with complex, non-Gaussian data (Aminikhanghahi and Cook, 2017).

Bayesian methods expanded the analytical toolkit by providing uncertainty quantification and probabilistic interpretations (Barry and Hartigan, 1993; Fearnhead, 2006). Yet, they introduced new limitations, including high computational demands and sensitivity to prior selection (Truong et al., 2020). Non-parametric methods addressed some of these issues by avoiding strong distributional assumptions, offering robustness to outliers and better performance on diverse datasets (Brodsky and Darkhovsky, 1993; Matteson and James, 2014; Harchaoui et al., 2008). Still, their performance was often contingent upon careful tuning of kernel parameters or bandwidths, which could be challenging in practice (Arlot et al., 2019).

Optimization-based methods such as dynamic programming (Jackson et al., 2005), penalized likelihood (Lavielle, 2005), and least-squares approaches (Bai and Perron, 2003) brought mathematical elegance and optimality guarantees. Nonetheless, their effectiveness heavily depended on the appropriate choice of penalty terms and model complexity balancing (Truong et al., 2020).

With the growing availability of streaming data, sequential detection methods emerged. Dynamic variants of CUSUM (Verdier et al., 2008; Granjon, 2013), sequential Bayesian models (Adams and MacKay, 2007; Fearnhead and Liu, 2007), and dynamic kernel-based algorithms (Li et al., 2015; Keriven et al., 2020) offered adaptive frameworks. Despite improvements, these methods often assumed underlying statistical properties that may not hold in real-world scenarios (Alippi et al., 2016; Saatçi et al., 2010). Adaptive windowing techniques (Bifet and Gavalda, 2007), density ratio estimation (Liu et al., 2013), and ensemble detectors (Frias-Blanco et al., 2015) introduced alternative strategies with varying trade-offs between sensitivity and complexity. However, they often required either large memory, prior knowledge, or struggled with real-time applicability and interpretability.

Recent deep learning-based approaches (Schmidl et al., 2022; Karim et al., 2018) achieved remarkable detection performance through complex model architectures. Yet, their demand for massive labelled datasets, high computational resources, and lack of transparency hinder their usability in many operational contexts (Aminikhanghahi and Cook, 2017; Wen et al., 2020). Similarly, online feature selection methods (Wu et al., 2012) aimed to reduce complexity but sometimes excluded informative features under dynamic regimes (Yu and Liu, 2012).

Despite extensive progress, several key research gaps remain. First, many existing methods rely on parametric assumptions, limiting their generalizability to complex or unknown data distributions. Second, methods requiring prior knowledge of break locations or strict model specifications reduce applicability in real-world, exploratory, or evolving environments. Third, while some recent methods achieve good accuracy, they sacrifice interpretability and usability due to complex designs and opaque mechanisms. Fourth, robustness to complex, high-dimensional, and noisy time series remains an ongoing challenge. Fifth, computational and memory requirements continue to pose barriers for deployment in real-time and resource-constrained settings. Finally, one of the most critical gaps is the insufficient attention to the nature of the data generation process itself. Many existing models focus solely on statistical or algorithmic detection, without modelling or understanding how the data is fundamentally generated—particularly when the process is driven by evolving information structures. Accurately detecting change points requires not only identifying statistical anomalies but also recognizing shifts in the underlying informational drivers of the system.

To address these limitations, we introduce a dynamic process to modelling time varying changing by understanding the nature of the data changing. This novel method called dynamic trend analysis approach(DTAA). We assume the data changing in the financial market is a result of all participants actions on the market, the changing is driven by the income information of the market and the market condition. We introduce a time driven integration method to capture the data changing characteristic at each time point. Then we give the method to understand and modelling the relationship between the information input and the market condition. Note that the 'dynamic trend' here is in fact the velocity of data changing in time series, not a typical trend.

DTAA is designed to be more flexible, interpretable, and computationally efficient than existing methods. It avoids rigid parametric assumptions, does not require prior specification of break points, and is built upon fundamental principles of temporal evolution in time series data. Crucially, DTAA incorporates mechanisms to model the data generation process itself, emphasizing the role of informational dynamics and their influence on observable sequences. By grounding detection in the changing informational structure that governs data behaviour, DTAA achieves greater robustness to noise and complexity, while offering improved interpretability. The method is well-suited for real-world applications that demand practical deployment, ease of understanding, and analytical rigour.

Moreover, when trying to capture the break points, we could understand the changing of data mode by the changing of moving pattern of data. This pattern changing also reflect as a change of trend. We could assume that in a short time interval, the data pattern could be captured by a liner method. Many methods in literature are understand the changing pattern in a way that assume the data changing pattern will only have certain different segments, but this will leads to a confusion when same period of data have different frequency. This will leads to a poor robustness of the method when new data coming especially when the input information have a significant change.

Consider a time series that have 1000 data, as below.

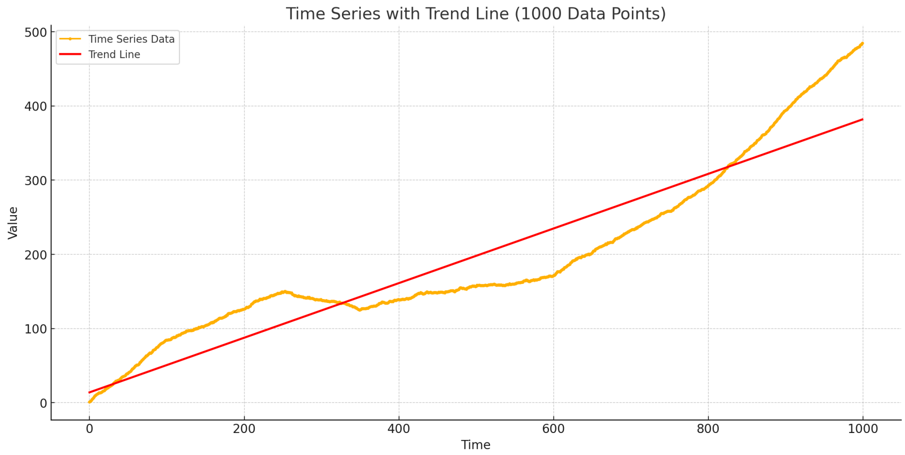

FIGURE 2.1: A Certain Trend

This series has an upward certain upward trend. But when we take the segment from 250 to 300 out, the data shows a downward trend as below.

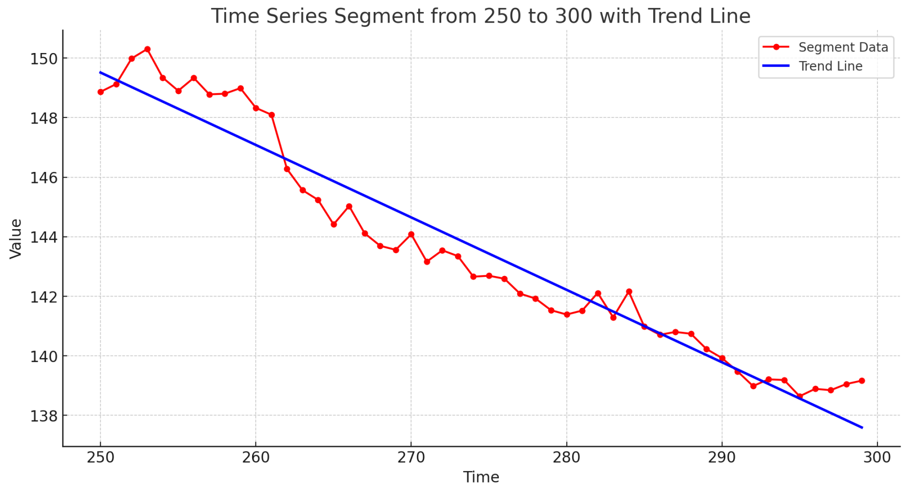

FIGURE 2.2: Segment 250 to 300 Showing A Downward Trend

Above is just one example, in fact, we could find many more different ways of understanding the changing pattern for this certain time series. This shows us that, when we understanding the data changing pattern in a trend modelling method, without a specific definition of trend, we could have many different understandings of trends in the same time series and it will lead to confusion and bias in modelling the time series. This is also a key reason why many models exhibit unstable robustness in practical applications. We will build a method with a better flexibility to capture the data changing that could explaining and using on the different data frequency. This method also could understanding the information impact difference among different market participants. In real world financial markets, for different traders, they have different preference of trading frequency, this is due to the different trading strategy or the nature of fund.For the different participants, they have different preference of the understanding of data changing pattern and noise. The noise or short-term fluctuations for a long term trader could be important patterns or trends for a high frequency trader. In the process of mathematical modelling, there is no precise definition or distinction between trends and short-term fluctuations. In reality, these two concepts are often differentiated relative to each other. For example, for time series data with a frequency of 1 second and a sample size of 2,592,000 seconds (30 days), the entire 30-day period might exhibit an upward trend, this is what a relatively long term trader preference of the pattern. However, considering shorter time spans, the first 10 days might show a downward trend, while the remaining 20 days show an upward trend, those patterns are important to relatively short term traders. Unlike traditional methods, the Dynamic Trend process we build adapts to the changing nature of the time series and provides a relatively precise definition of the Dynamic Trend and its corresponding mathematical model. The definition of ”structural breaks” in the existing literature is also rather ambiguous. Based on the dynamic trend, we also provide a systematic definition of structural breaks. Our method could explain and modelling the difference among market participants and the overall market.

Moreover, we should note that some theories in literatures are based on the mean changing to capture the short term trend. This will lead to a loss of information as we model trends or the whole time series. More specifically, mean shifting could be a result of a short-term trend as the mean is not possible to jump to another mean without any middle process.

Understanding the time series by using the mean is another reason why some models exhibit unstable robustness in practical applications. Using mean as a method to capture the data pattern, detecting mean changing to describe and deal with the data changing behaviour will lead to a bias when we have tend in time series, shown as below. We take the above segment from 250 to 300 as an example. If we take the mean as the proxy of this segment, like what we do in the moving average, a significant bias is generated at both ends of the data(the different of the real value and the mean proxy, 5.31 and -4.93), and the longer the data segment, the greater the bias. If we use the trend as the proxy, that will give us a relatively small bias, from 5.31 and 4.93 to 0.65 and 1.57(the different of the real value and the trend proxy).

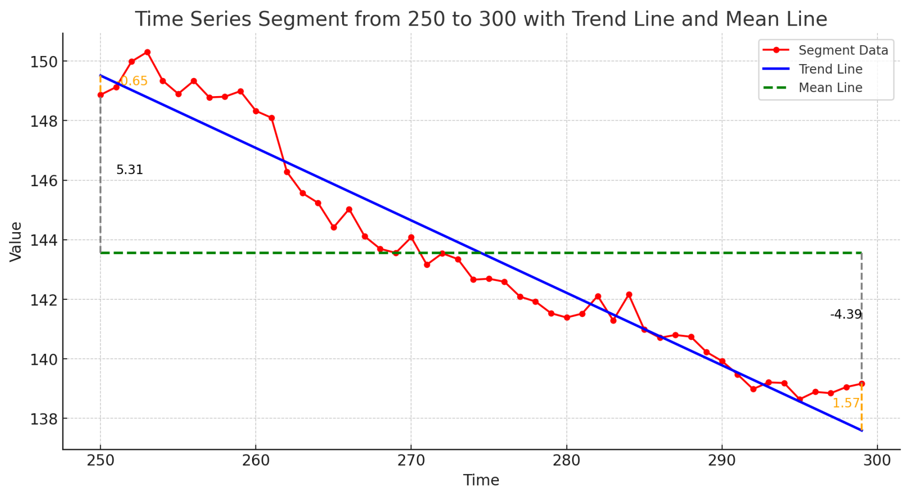

FIGURE 2.3: The Bias of Mean Proxy

The Dynamic Trend is essentially a measure of how data changes dynamically over time, which we define as the velocity of the data changing, we could understand it as the forward short term trend observed at each specific time point and changing with time, so we call it Dynamic Trend, but we should notice that this is not exactly be trend but be a new way to understand how the data driven by the input information. The changes in trends observed in time series data are merely superficial manifestations resulting from the combined effect of the system's input and its state at each moment. In essence, under conditions where both the nature and intensity of the input remain relatively stable and the system's state is also relatively constant, the acceleration of data change driven by the input approaches zero. As a result, the velocity of the data changing remains relatively consistent, giving rise to an apparent trend. The generation of new data is fundamentally a dynamic process in which the system transitions from one equilibrium state to another under the influence of dynamic external disturbances. Suppose our price data is a car that can only move forwards or backwards along a numerical axis, where price changes are represented as changes in the car's position on this axis. The car moving upwards corresponds to a price increase, while moving downwards corresponds to a price decrease. At different points in time, the car occupies different positions, corresponding to different price points on the price time series. Thus, the process of price changing over time cannot simply be viewed as differences in positions between two points. Instead, the car's movement from one point to another is described by its velocity, and variations in velocity result in different positions after experiencing the same amount of time. Changes in velocity occur due to alterations in the forces acting upon the car, thus changing the state of the system. The force causing changes in the car's state is analogous to the aggregate behaviour of all market participants driving the changes in time series data. By further analysing the dynamic trend, this paper presents a dynamic model of the system's response to external disturbances, revealing the process of how external information is converted into prices and the associated dimensional metrics.

2.3 Methodology

2.3.1 Dynamic Trend

The trend is a pattern of data's behaviour. It describes the direction and the magnitude in which the data is changing over time. It explains how the data behaves over time. To understand the time series data, we need to understand the nature of time first. In classical physics and everyday experience, time is perceived as both continuous and unidirectional. The continuity of time means that it can be infinitely divided into smaller intervals, represented mathematically as an unbroken sequence on the real number line. This allows for precise measurement and analysis of temporal intervals. The unidirectionality of time, often referred to as the "arrow of time," implies that time flows in a single direction from the past to the future. This concept is supported by the second law of thermodynamics, which states that entropy, or disorder, in an isolated system, always increases over time, making natural processes inherently irreversible.

In time series analysis, understanding the concepts of current time, past, and future is fundamental for accurately analysing and modelling temporal data. These concepts are intricately connected and are defined relative to each other, with the current time serving as the pivotal reference point which is the observer's time point. The observer's time point, denoted as t, is the specific moment at which the observer observing the time series and analyses are conducted. This time point acts as the observer's reference, providing a baseline from which past and future events are identified. In a time series dataset, the current time represents the present moment in the sequence of data points, making it essential for defining temporal relationships. The past in time series analysis encompasses all time points and corresponding data that occurred before the observer's time point t. Formally, the past includes all $t - i$ where i is a positive integer, indicating the number of time intervals before the present moment. Conversely, the future refers to all time points and data that will occur after the observer's time point t. Formally, it includes all $t + j$ where j is a positive integer, representing the number of time intervals following the present moment.

The logical relationship between the observer's time point, the past, and the future is fundamentally based on the positioning of t. The past $(t - i)$ and the future $(t + j)$ are defined relative to the observer's time point t, with the past encompassing all events that have occurred up to t, and the future including all events that have yet to occur. This relative definition underscores the importance of the observer's time point as the central reference point in time series analysis. When we modelling time series data, we need to give the definitions of our variable at the observer's time point which is current time t,

When modelling time series data, we first need to define our variables at the observer's time point t. This involves defining y_t, the value of the time series at time t. With this definition in place, we can then define the past values of the time series as y_{t-i}, where i is a positive integer representing the number of time intervals before the observer's time point. Similarly, we define the future values as y_{t+j}, where j is a positive integer representing the number of time intervals after the observer's time point. This systematic approach allows us to clearly distinguish between historical data, current data, and future data, which is essential for accurate time series analysis and modelling.

Additionally, when measuring data behaviour patterns, we must specify the time span we observe, known as the Horizon. The Horizon defines how far into the future the observer can see from the current time point. To measure data behaviour patterns, we effectively measure the characteristics of change, which require at least two values. Therefore, defining the Horizon is essential to understanding the changing behaviour of the time series data. The Horizon provides a framework to evaluate how data evolves over a specified future period from the observer's current time point.

- At time point t_1, we observe the trend after setting the Horizon to n. We have a trend for the time point t_1, denoted as $Trend_1$, as time passes from t_1 to t_n. The time interval is unchanged when we observe at time point t_2. We have a trend for time point t_2, called $Trend_2$, when time moves from t_2 to t_{n+1}.
- Given that the observer's time point changes, $Trend_1$ and $Trend_2$ might be different. Old information at t_1 is gone, and fresh information at t_{n+1} is coming since the time interval is constant.
- Since the trend is dependent on the observer's time point while the Horizon is fixed, we need to take it into account in a dynamic manner.
- The Horizon gives the fundamental trend length by the assumption that, at a specific observer's time point, in this short time span, the trend will be constant. When considering the Dynamic Trend Series, the Horizon is the threshold of trend. The pattern keeps unchanging for at least Horizon time pattern for the dynamic trend, the corresponding

segment of the original series could be considered as a trend segment. The pattern shorter than the Horizon will be considered as volatility.

We will now begin to develop a Dynamic Trend system:

We can break down the data into two main parts, namely, the trend component and the volatility component. The trend component series is the Dynamic Trend, at each time point the Dynamic Trend represents the overall direction in which the data will move at a specific observer's time point for a certain Horizon. As the Horizon is the threshold of trend for using the Dynamic Trend Series to detect the trend segments in the original series, we could define the trend segments in the original series as the Dynamic Trend Series corresponds to segments showing a relatively constant pattern for a time pattern longer than Horizon. By removing trends from the trend segments in the original series, we get the volatility component series.

The volatility component accounts for the variations or fluctuations part in the data, which is given by removing the trend component given by the Dynamic Trend Series. Thus, the Volatility Series we obtained inherently exhibits the desirable property of having a constant mean of zero. We now give the mathematical definition of the Dynamic Trend at a specific observer's time point t_1 with a Horizon of n.

In a $2D$ coordinate system, we define the x-axis as the time axis and the y-axis as the data axis. Assume that the trend will not change in a short Horizon. We can represent the trend in a pure trend data series from d_1 to d_n over a time period from t_1 to t_n using vectors.

In this system, we define the Time Changing Unit Vector(TCU) at the observer's time point t_1, with a magnitude of 1. Its start point is (t_1, d_1). Thus, we can say that Time Changing Vector $TC = TCU \cdot TD$, where Time Difference $TD = |TC| = |t_n - t_1|$, and $|TC|$ is a scalar representing the magnitude of TC. The direction of TC is always the same (positive) as time could only go forward; the TCU could only be in the same direction as the x-axis. Note that TC is the time component of a trend vector in time series data that describes the pure temporal direction change within the trend vector. This time component vector shares the same direction and interval as the actual change in time. Essentially, it is a mapping of temporal changes onto the data changes, but it is not time itself. Time, in this context, remains defined as a scalar according to its physical definition. The time component of the trend vector merely represents how changes in time are reflected in the data trend, preserving the unidirectional and continuous nature of time as a scalar entity.

Similarly, we define the Data Changing Unit Vector (DCU) at the observer's time point t_1, with a magnitude of 1. The direction of DCU could be upward or downward or like a 0 vector; we could take the sign out, then the upward direction is $+|DCU|$, and the downward direction is $-|DCU|$, or as a 0 vector when the data changing is 0 overtime on average. Its start point is (t_1, d_1). We can say that the Data Changing Vector $DC = \pm DCU \cdot DD$, where Data Difference $DD = |DC| = |d_n - d_1|$, and $|DC|$ is a scalar representing the magnitude of DC. The direction of DC is given by $\pm DCU$.

The Time Changing Vector, TC, starts at point (t_1, d_1) and ends at point (t_n, d_1). The direction of TC is parallel to the x-axis, which represents the passage of time. The magnitude of TC is $|t_n - t_1|$, which represents the total amount of time elapsed.

The Data Changing Vector, DC, starts at point (t_1, d_1) and ends at point (t_1, d_n). The direction of DC is parallel to the y-axis and points in the positive or negative direction if there is any change, depending on the direction of the data changing. The magnitude of DC is $|d_n - d_1|$, which represents the total amount of data changing.

The starting point of both TC and DC is fixed at (t_1, d_1) since they are based on the observed data at a specific observer's time point t_1. Since TC is defined to be parallel to the x-axis and DC is defined to be parallel to the y-axis, they must be orthogonal at the point (t_1, d_1).

Therefore, when TC and DC are added together using the parallelogram law of vector addition, the resulting vector TV (trend vector) represents the overall trend in the data over the Horizon from t_1 to t_n that is observed at time point t_1.

TV starts at the same point as TC and DC (that is, (t_1, d_1)), and its direction represents the overall trend in the data over the Horizon from t_1 to t_n that was observed at the time point t_1. TV obeys the parallelogram law of vector addition, which means that it is the diagonal of the parallelogram formed by TC and DC.

The vectors TV, TC, and DC meet the strict definition of a vector in mathematics, as they have both magnitude and direction. They also obey the rules of vector addition and scalar multiplication, as defined in linear algebra.

The vector TV (trend vector) represents the overall trend at a certain observer's time point t_i for the Horizon n. We call this the Dynamic Trend (DT) at time point t_i with the Horizon n. It changes with time and Horizon, this captures the temporal dynamics of the data. As the time component TC only goes in a certain direction, when we set a Horizon value, the TC does not change over time, maintaining a steady influence on TV. Given the constant nature of the TC component, the variation in the size and length of the vector TV is solely dependent on the DC component. Thus, any changes in TV can be attributed to changes in the DC component. Given the constant nature of the TC component, the variation in the size and length of the vector TV is solely dependent on the DC component. Thus, any changes in TV can be attributed to changes in the DC component. In this context, TV can be considered equivalent to DC, which means that DT can be considered equivalent to DC. Specifically, DT at the observer's time point t_i with a Horizon n is equivalent to DC at the observer's time point t_i with a Horizon n.

By defining the above $2D$ coordinate system, we now have a trend vector (TV). The trend vector can be decomposed into two vectors: the time changing vector (TC) and the data changing vector (DC).

This definition assumes that the trend in this certain Horizon is consistent for a specific observer's time point. The Horizon gives the fundamental trend length. The pattern shorter than the Horizon will be considered as volatility. This gives us the definition of the relative difference between trend and volatility. For one specific Horizon at one specific observer's time point, the Time Changing Vector (TC) and Data Changing Vector (DC) represent two distinct but related aspects of the data series:

- **Time Changing Vector (TC)**: Represents the passage of time effect in the system. Its direction is fixed along the x-axis, which represents time, and its magnitude corresponds to the difference in time between two points t_1 and t_n.

- **Data Changing Vector (DC)**: Embodies the Dynamic Trend at this specific observer's time point with a specific Horizon from t_1 to t_n. The direction of DC indicates whether the data value has increased, decreased, or remained unchanged over the Horizon from t_1 to t_n. The magnitude of DC, denoted as $|d_n - d_1|$, represents the absolute change in the data value during this Horizon.

So, the Dynamic Trend (DT) is the time-varying TV (trend vector) for a certain Horizon. As time progresses, at each observer's time point, we have a Data Changing Vector (DC). It is dynamic, evolving with time, and thus can be considered equivalent to DT. We now redefine the TV as the Dynamic Trend vector.

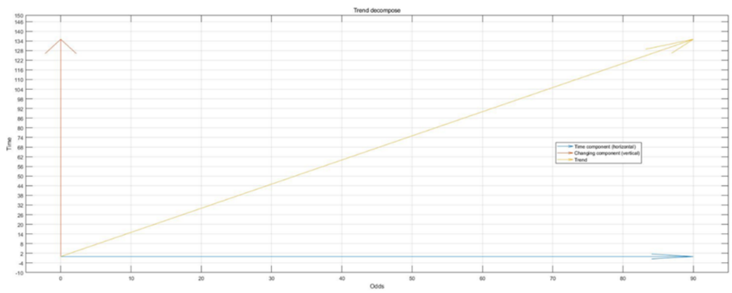

FIGURE 2.4: Pure Trend at One Observer's Time Point

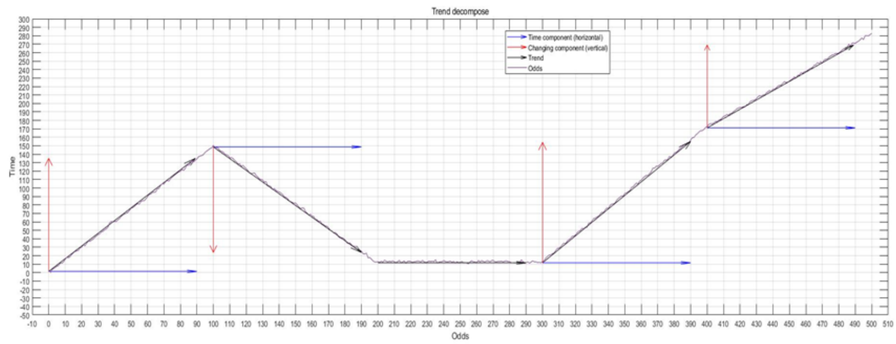

FIGURE 2.5: Pure Trend at Different Observer's Time Points

2.3.2 Slope as Proxy for Dynamic Trend

As defined earlier, to use the Dynamic Trend to assess data behaviour in a series, we need a measure that captures both the magnitude and direction of the Dynamic Trend (DT). This measure, known as the Dynamic Trend proxy, is effectively represented by the slope of the trend vector TV at each observer's time point with a certain Horizon.

The slope of TV offers a practical and efficient way to represent the dynamic trend. By calculating the slope, we can assess the data behaviour over time, providing information about the extent of change in the data, and indicating the direction of this change. Consequently, the slope of the Dynamic Trend vector serves as an easy-to-use and reliable measure for analysing the dynamic trend.

The objective of this sector is to provide mathematical proof demonstrating the suitability of the slope of the Dynamic Trend vector as an effective proxy for evaluating the dynamic trend. The mathematical proof is given below.

For represents the direction of DC:

The slope of the trend vector, denoted as $slopeT$, can be expressed as:

$$tan(\alpha) = \frac{d_n - d_1}{t_n - t_1} \tag{2.1}$$

, where d_n and d_1 represent the respective data points on the Dynamic Trend vector at times t_n and t_1, and α is the angle of inclination. The time interval $t_n - t_1$ is always positive due to the inherent nature of time progression, leading us to define it as $|t_n - t_1|$. Introducing a sign factor, $Sign = \pm$, the equation can be rewritten as:

$$\text{Sign} \cdot |d_n - d_1| = d_n - d_1 \tag{2.2}$$

By substituting these expressions into the original equation, we obtain:

$$\text{slopeT} = \tan(\alpha) = \frac{\text{Sign} \cdot |d_n - d_1|}{|t_n - t_1|} \tag{2.3}$$

Further rearrangement yields:

$$\text{slopeT} \cdot |t_n - t_1| = \text{Sign} \cdot |d_n - d_1| \tag{2.4}$$

It is important to note that $|t_n - t_1|$ represents the magnitude of the time constant (TC) and is a constant positive value, as the Horizon remains unchanged throughout the analysis. On the other hand, $|d_n - d_1|$ represents the magnitude of the data changing vector (DC) and is inherently positive or zero. The sign factor, denoted as $Sign$, could only be the same as the slope sign.

Now, the Data Changing Vector (DC) is defined as:

$$\begin{aligned} DC &= \pm |DCU| \cdot DD \\ &= \text{Sign} \cdot |DCU| \cdot |d_n - d_1| \\ &= \text{slopeT} \cdot |DCU| \cdot |t_n - t_1| \end{aligned} \tag{2.5}$$

We take out the sign of $slopeT$ denote as $SignS$:

$$\text{Sign} \cdot |DCU| \cdot |d_n - d_1| = \text{SignS} \cdot \text{—slopeT—} \cdot |DCU| \cdot |t_n - t_1| \tag{2.6}$$

Where when $d_n > d_1$, $Sign = +$, when $d_n = d_1$, DC is a 0 vector, and when $d_n < d_1$, $Sign = -$. $|slopeT|$, $|t_n - t_1|$ are not negative and $|DCU| = 1$. We have $d_n > d_1$, $SignS = +$, when $d_n = d_1$, DC is a 0 vector, and when $d_n < d_1$, $SignS = -$. This gives us $SignS = Sign$.

Consequently, we can deduce that $slopeT$ shares the same sign as $Sign$. Hence, the sign of the slope($SignS$) of the trend vector (TV) aligns with the sign of $Sign \cdot |DCU| \cdot |d_n - d_1|$, which effectively represents the direction of the DC. This straightforward relationship allows us to utilize the $SignS$ as an indicator of the direction in which the data is changing.

The above mathematical analysis confirms that the slope of a Dynamic Trend vector serves as a reliable direction proxy for the dynamic trend. By considering the sign of the slope, we can easily determine the direction of the Data Changing Vector (DC), enabling a practical and intuitive assessment of data behaviour.

Representing the Magnitude of the Data Changing Vector (DC):

The slope of the dynamic trend, denoted as $slopeT$, can be calculated as:

$$\text{slopeT} = \tan(\alpha) = \frac{d_n - d_1}{t_n - t_1} \tag{2.7}$$

where α is the angle of inclination. Since the time interval $t_n - t_1$ is always positive, we substitute it with its absolute value $|t_n - t_1|$.

To simplify the equation, we incorporate the sign factor $Sign$ defined earlier, which represents the sign of $d_n - d_1$. By multiplying $Sign$ with the absolute value of $d_n - d_1$, we obtain:

$$\text{Sign} \cdot |d_n - d_1| = d_n - d_1 \tag{2.8}$$

The expression for the slope of the dynamic trend, $slopeT$, can be rewritten as:

$$\text{slopeT} = \frac{\text{Sign} \cdot |d_n - d_1|}{|t_n - t_1|} = \frac{\text{Sign} \cdot \text{DD}}{\text{TD}} \tag{2.9}$$

where DD represents the Data Difference and TD represents the Time Difference. Here, Data Difference $DD = |DC| = |d_n - d_1|$, and $|DC|$ is a scalar representing the magnitude of DC. Similarly, $TD = |TC| = |t_n - t_1|$, where $|TC|$ is a scalar representing the magnitude of TC. As we observe the Dynamic Trend under a certain Horizon at a specific observer's time point, TD is a constant.

Rearranging the equation, we have:

$$\text{SignS} \cdot \text{—slopeT—} \cdot \text{TD} = \text{DD} \tag{2.10}$$

We take out the sign of $slopeT$:

$$\text{Sign} \cdot \text{SignS} \cdot \text{—slopeT—} \cdot \text{TD} = \text{DD} \tag{2.11}$$

As we know from above that $SignS = Sign$, we have:

$$\text{—slopeT—} \cdot \text{TD} = \text{DD} \tag{2.12}$$

As we set Horizon as a fixed value, TD is a constant. This shows that $|slopeT|$ has a linear relationship with the absolute value of the Data Difference (DD), indicating that they change at the same rate.

Consequently, we can conclude that the relationship between $|slopeT|$ and the Data Difference (DD) is linear, demonstrating that they exhibit the same rate of change. $|slopeT|$ can reliably represent changes in DD. Data Difference $DD = |DC| = |d_n - d_1|$, and $|DC|$ is a scalar representing the magnitude of DC. So, $|slopeT|$ is an unbiased proxy for the magnitude of the Data Changing Vector (DC).

Now we have $slopeT = SignS \cdot |slopeT|$ as the unbiased proxy for the Data Changing Vector (DC). As the Dynamic Trend (DT) is the time-varying TV for a certain Horizon, DC can be considered equivalent to DT, which means $slopeT = SignS \cdot |slopeT|$ is the unbiased proxy for DT. We call this proxy the 'Slope Proxy' or 'Tangent Proxy' of DT.

2.3.3 Interpreting Dynamic Trend as Velocity

Moreover, the equation for the slope of the dynamic trend, $slopeT$, given as $slopeT = \tan(\alpha) = \frac{d_n - d_1}{t_n - t_1}$, essentially signifies the computation of the average velocity. The right-hand side, $\frac{d_n - d_1}{t_n - t_1}$, represents the change in position (known as displacement, represented here by DD) over the change in time (represented by DT). This ratio, the change in position over the change in time, is fundamentally the definition of average velocity.

In this context, as we assume that data is a pure trend series, d_n and d_1 symbolize the final and initial positions of the data, respectively. The difference between these two, $d_n - d_1$, illustrates displacement, which is the change in position (if the data is not a pure trend series(have volatility component, $d_n - d_1$ illustrates displacement on average). Similarly, t_n and t_1 symbolize the final and initial time points, respectively, and their difference, $t_n - t_1$, represents the time interval over which we are observing the change.

On the left-hand side of the equation, $slopeT = \tan(\alpha)$, the term $\tan(\alpha)$ signifies the trigonometric tangent of the angle α. This angle α is related to the steepness of the line that links the starting and ending points on a graph, which shows how position changes over time. This slope is essentially the ratio of the vertical change (which is displacement) to the Horizontal change (which is time). Mathematically, this ratio is a clear representation of the concept of average velocity under a certain Horizon at a specific observer's time point.

Taken together, this suggests that the Dynamic Trend can be interpreted as the average velocity of the change in data over a given Horizon at a specific observer's time point. In simpler terms, it reflects how fast and in which direction the data is changing on average over a specific period at a specific observer's time point. This interpretation not only adds mathematical rigour to our understanding of the Dynamic Trend but also offers an intuitive way to assess the behaviour of the data series over time.

However, while this mathematical model serves well to understand the general trend behaviour over a time interval, under the assumption that the trend will not change within a certain short Horizon, it lacks the precision to analyze data changes at a specific instant. Therefore, to transition from average velocity to instantaneous velocity in physics, we propose the concept of instantaneous Dynamic Trend (IDT), which denotes the rate of data change at a specific time point.

To accomplish this, we define the IDT as the limit of the average Dynamic Trend as the Horizon approaches zero. Formally, this can be represented as:

$$\text{IDT} = \lim_{\Delta t \to 0} \frac{d_n - d_1}{t_n - t_1} \tag{2.13}$$

Or, using the notation of calculus:

$$v(t) = \frac{dD}{dt} \tag{2.14}$$

Here, $v(t)$ signifies the IDT at a specific time t, D represents the data, t indicates time, and $\frac{dD}{dt}$ refers to the derivative of the data D with respect to time t.

Providing a comprehensive mathematical proof for the concept of instantaneous Dynamic Trend (IDT), we utilize principles from calculus, particularly the concept of a derivative, which fundamentally is the limit of the rate of change as the time interval approaches zero. Starting with the average dynamic trend, akin to the average velocity, we have:

$$\text{slopeT} = \frac{\Delta D}{\Delta t} = \frac{d_n - d_1}{t_n - t_1} \tag{2.15}$$

As we refine our observation by reducing the time interval $\Delta t = t_n - t_1$ towards zero, we transition from the average Dynamic Trend to the IDT, defined as:

$$\text{IDT} = \lim_{\Delta t \to 0} \frac{\Delta D}{\Delta t} = \lim_{\Delta t \to 0} slopeT \tag{2.16}$$

Assuming that our data follows a function $f(t)$, the data difference ΔD can be represented as $f(t + \Delta t) - f(t)$. Substituting this into our limit gives:

$$\text{IDT} = \lim_{\Delta t \to 0} \frac{f(t + \Delta t) - f(t)}{\Delta t} \tag{2.17}$$

This definition coincides with the derivative of the function f at point t, denoted by $f'(t)$ or $\frac{df}{dt}$. Hence, we establish:

$$\text{IDT} = \frac{df}{dt} \tag{2.18}$$

This proof confirms that the instantaneous Dynamic Trend $IDT = v(t)$ at any given time point t is equivalent to the derivative of the data at that point. It enriches our understanding of the dynamic trend, providing both the direction and the magnitude of data change at any given point, encapsulating the essence of data behaviour patterns over time.

Based on this inferential deduction, it becomes apparent that as we decrease the size of the observational window, which is the Horizon, the Dynamic Trend tends towards the notion of an Instantaneous Dynamic Trend (IDT). Consequently, to acquire an accurate Dynamic Trend at a specific temporal juncture, denoted as t, one could adopt one of two strategies.

2.3.3.1 Strategy 1: Narrow Horizon

The first strategy consists of choosing a relatively narrow Horizon that aligns closely with the given time point t. This approach, however, may have its limitations, especially when dealing with sparse or irregular data where narrowing the Horizon could lead to inadequate data points for a reliable trend estimation.

2.3.3.2 Strategy 2: Large Dataset

Alternatively, the second strategy involves working with a considerably large dataset. In such cases, even if the absolute Horizon might not be small, its size relative to the entire dataset could be deemed as minimal.

Mathematically, if we denote the total number of data points as N and the Horizon as n, as we increase N while keeping n constant, the ratio $\frac{n}{N}$ tends to zero. This implies that the Horizon becomes increasingly insignificant relative to the entire data set, effectively acting as a narrow window around the specific time point t.

Formally, the instantaneous Dynamic Trend for a large dataset can be approximated as:

$$v(t) \approx \frac{\Delta D}{\Delta t} \quad \text{for} \quad \frac{n}{N} \to 0 \tag{2.19}$$

Here, $v(t)$ represents the IDT at time t. By considering the expression $v(t) \approx \frac{\Delta D}{\Delta t}$ for $\frac{n}{N} \to 0$, we observe that as the Horizon n is small enough comparing with N, the ratio $\frac{n}{N}$ tends to zero.

2.3.4 Temporal Granularity Index

In this context, we define this ratio, $\frac{n}{N}$, as the Temporal Granularity Index. The Temporal Granularity Index serves as a metric to quantify the relative Horizon compared to the entire dataset. It provides an indication of the temporal resolution or granularity of the analysis by considering the proportion of the dataset included in each Horizon.

The Temporal Granularity Index can be mathematically expressed as:

$$\text{Temporal Granularity Index} = \frac{n}{N} \tag{2.20}$$

Here, N stands for the total number of data points in the dataset, and n represents the Horizon, indicating how many data points are included in each observation window.

As a result, when discussing the convergence of the equation $v(t) \approx \frac{\Delta D}{\Delta t}$ for $\frac{n}{N} \to 0$, we refer to the fact that as the Temporal Granularity Index decreases to zero, the Horizon becomes progressively negligible compared to the total dataset. This convergence, even if the absolute Horizon may not be small, demonstrates the unique ability of the Instantaneous Dynamic Trend (IDT) to capture accurate moment-to-moment variations in the dataset.

A lower Temporal Granularity Index (TGI) indicates higher sensitivity, which allows the dynamic trend, representing average velocity, to approach the concept of the instantaneous Dynamic Trend (IDT), representing instantaneous velocity. This enhances the precision in capturing the instantaneous rate of data change.

The Dynamic Trend becomes more susceptible to changes occurring at specific time points when the Temporal Granularity Index (TGI) is smaller. This corresponds to a narrower Horizon relative to the dataset size. The average velocity of data change inside each Horizon can be estimated with greater accuracy due to the enhanced sensitivity. A more in-depth comprehension of the data's moment-to-moment alterations is facilitated by the smaller window's ability to capture finer-grained fluctuations and transitory patterns.

Furthermore, as TGI decreases and the Horizon narrows, the Dynamic Trend converges toward capturing the instantaneous behaviour, approaching the Instantaneous Dynamic Trend (IDT). Although the Dynamic Trend remains an average representation, the smaller TGI allows it to approach the instantaneous velocity of data change at a specific time point. This provides insights into the rapid and precise changes occurring in the dataset at that moment.

In contrast, a higher TGI denotes a lesser sensitivity and a broader Horizon in relation to the dataset size. The responsiveness of the Dynamic Trend to subtle changes within the window is reduced in this larger context. A greater TGI highlights long-term patterns and trends while possibly diminishing the impact of sudden or short-term fluctuations.

It's important to keep in consideration that the choice of the Temporal Granularity Index (TGI) involves striking a balance between sensitivity and stability in trend analysis. A lower TGI can increase sensitivity and enable the Dynamic Trend to approach the Instantaneous Dynamic Trend (IDT), but it may also make the trend estimate more susceptible to noise or short-term changes, thereby affecting its stability. On the other hand, a larger TGI compromises sensitivity to smooth changes but offers a more stable trend estimation by reducing the effect of noise or short-term variations.

By carefully choosing the TGI, researchers can achieve a balance between capturing immediate dynamism and retaining stability in trend analysis, thereby gaining insightful knowledge about how the dataset behaves over time.

2.3.5 The Acceleration of Data Movement

After defining the velocity of the data movement, we can naturally derive the acceleration of data movement. The velocity of data movement is the first derivative of data movement with respect to time. By taking the derivative of this with respect to time again, we obtain the second derivative of data movement with respect to time, which is the acceleration of the data movement. This means DT of DT is the acceleration of the data movement.

2.3.6 Summary of Dynamic Trend

In summary, the analysis undertaken in this study has led to some key discoveries regarding the dynamic trend.

First, it has been observed that the slope of the Dynamic Trend vector, denoted as $slopeT$, correlates with the direction of the Data Changing Vector (DC). This correlation provides a clear way to determine whether the data exhibits an increasing, decreasing or flat trend over time, thereby facilitating easier interpretation of the directionality of data.

Additionally, a linear relationship has been identified between $slopeT$ and the Data Difference (DD), indicating that their rates of change are the same. This finding implies that the slope of the Dynamic Trend vector will change in accordance with the rate of change in the data, providing an apparent indication of the magnitude of data change.

Furthermore, the equivalency between $slopeT$ and the average velocity offers another insightful assessment of the dynamic trend. In this context, $slopeT$ can be viewed as the "average velocity" of data change over time, drawing parallels with the concept of average velocity from physics, which describes the rate of change in position over a specified time period at a specific observer's time point. This equivalency not only supports the rationality of using $slopeT$ as a measure of the Dynamic Trend but also aids in comprehending how the data behaved over the observed Horizon.

Based on these discoveries, we propose the idea of the instantaneous Dynamic Trend (IDT) to overcome the limitations of the average dynamic trend. The IDT provides a more accurate examination of data fluctuations at a specific moment, reflecting the rate of data change at a certain specific observer's time point. It is defined as the limit of the average Dynamic Trend as the Horizon approaches zero and can be mathematically expressed as:

$$\text{IDT} = \lim_{\Delta t \to 0} \frac{\Delta D}{\Delta t} \tag{2.21}$$

or

$$\text{IDT} = \frac{df}{dt} \tag{2.22}$$

This definition adheres to the concepts in calculus, notably the concept of a derivative, which is fundamentally the limit of the rate of change as the interval of time approaches zero.

There are two methods that can be used to acquire an accurate Dynamic Trend at a particular time point: working with a large dataset or utilizing a small window size that corresponds to the desired time point. Using a narrow window size enables targeted analysis but may be constrained by sparse or irregular data, whereas working with a large dataset offers a broader perspective but can neglect detail.

Determining the optimal balance, known as the Temporal Granularity Index (TGI), requires considering research goals, dataset features, and the required level of detail and stability in trend estimations. A smaller index increases sensitivity to noise but offers a more accurate evaluation of the Dynamic Trend at a given time point. Conversely, a larger index provides a broader picture, reducing the impact of noise but potentially missing valuable information. Conducting sensitivity analyses with various ratios is recommended to analyse their impact on trend estimations and discover the most optimal approach for a given analysis task.

The analysis of the dynamic trend, along with the introduction of the Instantaneous Dynamic Trend (IDT), enhances our understanding of data behaviour over time. The Temporal Granularity Index and the insights gained through this analysis are valuable tools for comprehending and assessing instantaneous changes in data. These methodologies improve the precision and depth of data analysis, aiding in the task of informed decision-making. Moreover, DT of DT is the acceleration of data movement.

2.3.7 Generate Higher Time Level Dynamic Trend

As we have defined the dynamic trend($slopeT$), it becomes clear that it is a time series. With a low Temporal Granularity Index (TGI), we are able to generate a Dynamic Trend Series at a relatively low time level. For instance, with second-level data, the unit could be expressed as GBP per second (GBP/s).

Conversely, for any low time level dynamic series, it can be transformed to represent a higher time level dynamic trend. An example of this would be converting data from a second-level to an hour-level Dynamic Trend time series, where the unit would be GBP per 3600 seconds (GBP/3600s) or GBP per hour (GBP/h).

2.3.7.1 Assumptions

1. **Constant Measurement Interval:** All measured Dynamic Trend (velocity) values $d_1, d_2, \ldots, d_m$ are obtained over equal time intervals. The time intervals T are given by the Time Level(TL). The Time Level is congruent with the frequency of the original data series (e.g., 1-second data, hourly data), and it quantifies the regularity with which the data's behaviour pattern is observed. This is a result of the observing action, the time interval between two consecutive actions is constant, thus the time interval between two consecutive observations is also constant. This implies that each Dynamic Trend (velocity) value d_i represents the average trend(this is given by assumption 2) over a time interval of the same length. The TL of the original DT series is the fundamental TL. The time interval T given by the fundamental TL is the fundamental time unit.
2. **Assumption of Linear or Uniform Change:** Within each time interval T given by TL, the Dynamic Trend is assumed to be constant. Even though the Dynamic Trend may vary between different intervals, it is considered constant within each individual interval. More specifically, if our DT series is a 1-second TL series. We consider the d_i value for that time label i is the average velocity of the second from this time label i to next time label $i + 1$, and the next time label $i + 1$ have DT value of d_{i+1}, which is the average velocity of the next 1-second time interval start from the time label $i + 1$, could be different from d_i.
3. **Directionality and Sign of Values:** The Dynamic Trend values can be positive or negative or 0, which indicates the direction of change. When calculating the total change, the sign of the Dynamic Trend values must be taken into account.

2.3.7.2 Transfer Process

Consider a 1-second TL Dynamic Trend (velocity) series, measurements $d_1, d_2, \ldots, d_m$ with units of data-unit/T (e.g., GBP/s), where each d_i represents the average Dynamic Trend over a time interval T. The time interval T is given by TL, here is 1-second interval.

1. **Calculation of Individual Change:** The change for each time interval T is $d_i \times T$, where d_i is the average Dynamic Trend during that interval. Since the time interval T is the fundamental time unit, and in this example, the TL of the original DT series is 1 second, the corresponding T is 1 second. Thus, the fundamental time unit is 1 second, and the

corresponding unit of the DT value is GBP/s. However, in general, it is not necessarily 1. For instance, if the TL of the original DT series is 3 seconds, the corresponding T is 3 seconds, making the fundamental time unit 3 seconds, and the corresponding unit of the DT value is GBP/3s. In this case, 3 seconds is the fundamental time unit. Therefore, the change within this time unit should be the change over 3 seconds, and the value is obtained by $d_i \times 1$(means 1 unit of 3 seconds). Below is the formula for its calculation in general form:

$$\Delta P_i = d_i(GBP/T) \times 1(T)$$

2. **Calculation of Total Change:** To calculate the total change ΔP over a time interval T'(e.g., GBP/3600s or GBP/h), sum all individual changes:

$$\Delta P = \sum_{i=1}^{m} \Delta P_i = \sum_{i=1}^{m} (d_i(GBP/T) \times 1(T)) \tag{2.23}$$

Here, m is the number of unit intervals T contained within T', i.e., $T' = mT$.

3. **Conversion to a Higher Time Level Dynamic Trend:** To convert the total change ΔP of each T' to an average change per T' (i.e., data-unit/T'), divide by the time write as a unit form $1(T')$:

$$T' \text{ Time Level } d'_j = \frac{\Delta Pc}{1(T')} \tag{2.24}$$

Where $d'_j$is the Dynamic Trend of T' TL Dynamic Trend Series, which is corresponding to the time interval from $i = 1$ to $i = m$ of the T TL Dynamic Trend Series. As T is the fundamental time unit, T' is the new time unit. Take the unit of time out, we have 1 (T) and 1 (T'). For the new T' TL Dynamic Trend Series, the data-unit of total change ΔP_{ji} here we assume is GBP, the time unit is T'. Take the unit out:

$$T' \text{ Time Level } d'_j = \frac{\sum_{i=1}^{m} (d_{ji}(GBP/T) \times 1(T))}{1(T')} \tag{2.25}$$

This expression simplifies to

$$T' \text{ Time Level } d'_j = \sum_{i=1}^{m} (d_{ji})(GBP/T') \tag{2.26}$$

This implies that the Dynamic Trend of a higher time level T' (e.g., GBP/3600s or GBP/h) is simply the sum of all lower time level T Dynamic Trend values(e.g., GBP/s) within that interval. This gives us a simple method to transfer a higher frequency (lower TL) DT series into a lower frequency (higher TL) DT series.

4. **Transfer the higher TL DT series into a lower form TL DT series :** As we have a higher TL DT series with a given fundamental TL DT series. We could transfer the higher TL DT series into a lower form. This lower form is not the same as the fundamental TL DT series. It could be any lower TL DT series. Due to the accuracy requirement, considering the TGI, it could not be lower than the fundamental TL. Assume we have the higher TL DT series which is T' TL Dynamic Trend Series in the above part. We have a lower TL as

T'', $T' = mT''$. The T'' TL form of T' TL Dynamic Trend series is given by:

$$T'' \text{ Time Level } d''_{ji} = \frac{T' \text{ Time Level } d'_j}{m}(GBP/T'') \tag{2.27}$$

Where d''_{ji} , $i = 1$ to m. This gives the corresponding intervals within T' have average velocity of d''_{ji} for each time label i. We could notice that all d''_i within a certain interval j are the same, this implies that when we transfer a higher TL DT series into a lower form we lose some information. But this still be a simple way to do the TL transformation.

The flexibility in representing the Dynamic Trend at various time scales allows for a comprehensive analysis of data behaviour across different temporal resolutions, thereby enhancing the utility of the Dynamic Trend concept in time series analysis.

2.3.8 *DTAA* Application on Higher Time Level

2.3.8.1 Key Elements of *DTAA*

Implementing $DTAA$ involves careful consideration of two pivotal factors: the Horizon and the Time Level.

The Time Level is congruent with the frequency of the original data series (e.g., 1-second data, hourly data), and it quantifies the regularity with which the data's behaviour pattern is observed.

The Horizon, known as the Window Size (e.g., 30 seconds, 60 seconds, which is equivalent to 1 minute), determines the lower boundary of the trend and the extent of future visibility in the data.

2.3.8.2 Setting the Horizon

Assuming an original time series characterized by data points recorded at 1-second intervals. $DTAA$ facilitates the generation of a Higher Time Level Dynamic Trend Series, which operates at a granularity of one minute. Concurrently, this approach also yields a corresponding Volatility Series. To explain the exact whole process of $DTAA$, we will give the generation of the Fundamental (primary) Time Level Dynamic Trend Series and the generation of a higher Time Level Dynamic Trend Series and the corresponding Volatility Series.

For an original time series with 1-second data intervals, the Horizon is set at 61 data points, equivalent to a 60-second or 1-minute Horizon. This setting is pivotal in generating a Dynamic Trend Series that is reflective of the underlying time series data.

2.3.8.3 Generation of Fundamental Time Level Dynamic Trend Series

The $DTAA$, with the specified Horizon, produces a 60-second Dynamic Trend Horizon 1-second Time Level Dynamic Trend Series, herein referred to as the Fundamental (primary) Time Level

Dynamic Trend Series (with the unit of GBP/s), known as $60sDTH1sFundTL\ DT$. This series forms the foundation for further analysis. Where $60sDTH$ showing the Horizon of the Dynamic Trend is 60-second; $1sFundTL$ showing the Dynamic Trend Series we generated here is a Fundamental Time Level Dynamic Trend Series with 1-second Time Level.

2.3.8.4 Development of Higher Time Level Dynamic Trend Series

Using the $60sDTH1sFundTL\ DT$ (GBP/s), it is possible to extrapolate higher Time Level Dynamic Trend Series. An example is creating a 1-minute Time Level series, denoted as GBP/m, from the 60-second Dynamic Trend Horizon 1-second Fundamental Time Level Dynamic Trend Series with the method given by the method provided in the previous sections. The 60-second Dynamic Trend Horizon 1-minute Generated Time Level Dynamic Trend Series getting from 1-second Fundamental Time Level Dynamic Trend Series ($60sDTHGen1minTL1sFundTL\ DT$) is given by:

$$
\begin{aligned}
&60sDTHGen1minTL1sFundTL\ DT_j(GBP/m) = \\
&\sum_{i=1}^{60}(60sDTH1sFundTL\ DT_{ji})(GBP/s)
\end{aligned}
\tag{2.28}
$$

This gives how to use $60sDTH1sFundTL\ DT$ of corresponding intervals within each minute j to generate $60sDTHGen1minTL1sFundTL\ DT$.

2.3.8.5 Generation of Corresponding Volatility Series

The $60sDTHGen1minTL1sFundTL\ DT$ is also instrumental in generating the corresponding Volatility Series.

As the data frequency of $60sDTHGen1minTL1sFundTL\ DT$ is different from the original data(from 1-second to 1-minute frequency).

We have two different ways of generating the corresponding Volatility Series. If we want an original frequency corresponding Volatility Series, we could transform the 1-minute frequency Dynamic Trend Series back into a 1-second frequency Dynamic Trend Series. The DT value at each time label could be obtained from the method of transferring the higher $TL\ DT$ series into a lower form in the previous sections. Then we remove all trend components of the original data to get the original frequency corresponding Volatility Series. If we want a 1-minute frequency corresponding Volatility Series, we could transform the original data into a 1-minute frequency data(or we have the higher Time Level Original Data), then remove the trend components getting from 1-minute frequency DT to get the 1-minute frequency corresponding Volatility Series.

2.3.9 Use Dynamic Trend Series *DT* to Generate The Price

Given a Dynamic Trend Series, we assume that the Dynamic Trend DT remains constant within short intervals T, indicating that the Dynamic Trend is a constant within each small segment. This short intervals T is the same as the frequency of the DT series, known as the Time Level of DT. For example, we gave a 1-second Time Level DT series, then T is 1-second, which means that DT remains constant within that second. The specific explanations are given in the previous sections.

2.3.9.1 Step 1: Calculation of Average Dynamic Trend DT_i for Each Time Interval T'

Assume that we have T', n is the number of unit intervals T contained within T', which is $T' = nT$.For each time interval T', assume a series of Dynamic Trend measurements are collected. For instance, in the first interval T'_1, these measurements might be $d_{1,1}, d_{1,2}, \ldots, d_{1,n}$, where n is the number of measurements within T'_1. Calculate the average of all measurements in each interval T'_i to represent that segment's Dynamic Trend DT_i.

For example, the average Dynamic Trend for the first interval DT_1 is:

$$DT_1 = \sum_{j=1}^{n} d_{1,j} \tag{2.29}$$

Repeat this process to determine the average Dynamic Trend for each interval $DT_2, DT_3, \ldots, DT_i$.

2.3.9.2 Step 2: Calculating $DP(T'_i)$

1. **First Time Interval** $0 \leq t < T'$: In this interval, with the Dynamic Trend DT_1, the $DP(T'_1)$ is calculated as:

$$\begin{aligned} DP(T'_1) &= P_0(\text{GBP}) + \int_0^{T'} DT_1 \left(\frac{\text{GBP}}{T'}\right) dt \\ &= P_0(\text{GBP}) + DT_1 \left(\frac{\text{GBP}}{T'}\right) \times T' \\ &= (P_0 + DT_1)(\text{GBP}) \end{aligned} \tag{2.30}$$

2. **Second Time Interval** $T' \leq t < 2T'$: Here, with the Dynamic Trend DT_2, the $DP(T'_2)$ is:

$$\begin{aligned} DP(T'_2) &= (P_0 + DT_1)(\text{GBP}) + \int_{T'}^{2T'} DT_2 \left(\frac{\text{GBP}}{T'}\right) dt \\ &= (P_0 + DT_1)(\text{GBP}) + DT_2 \left(\frac{\text{GBP}}{T'}\right) \times (2T' - T') \\ &= (P_0 + DT_1 + DT_2)(\text{GBP}) \end{aligned} \tag{2.31}$$

3. **The i-th Time Interval** $(i-1)T' \leq t < iT'$: For the i-th interval with Dynamic Trend DT_i, the $DP(T'_i)$ is:

$$\begin{aligned} DP(T'_i) &= \left(P_0 + \sum_{j=1}^{i-1} DT_j\right)(\text{GBP}) + \int_{(i-1)T'}^{iT'} DT_i \left(\frac{\text{GBP}}{T'}\right) dt \\ &= \left(P_0 + \sum_{j=1}^{i-1} DT_j\right)(\text{GBP}) + DT_i \left(\frac{\text{GBP}}{T'}\right) \times T' \\ &= \left(P_0 + \sum_{j=1}^{i} DT_j\right)(\text{GBP}) \end{aligned} \tag{2.32}$$

We should noticed this $DP(T_i')$ is the Dynamic Trend driven price component of the real estimation price. The real estimation price should have two components: Dynamic Trend driven price component $DP(T_i')$ and volatility component $V(T_i')$. The estimated price $\hat{P}(T_i')$ is defined as:

$$\hat{P}(T_i') = DP(T_i') + V(T_i') \tag{2.33}$$

2.3.9.3 The Economic Implications of Dynamic Trend Series and Volatility Series

With the Dynamic Trend analysis approach ($DTAA$), we decompose price series into a Dynamic Trend Series and a Volatility Series. The Dynamic Trend Series accounts for the sustained influence of relative macroeconomic factors and significant external events on prices. Conversely, the Volatility Series reveals the market's fluctuation from itself.

Moreover, $DTAA$ provide an approach to understanding market dynamics and consensus formation. The Dynamic Trend Series is a key component of our analysis and serves as a tool to trace the evolving consensus in the market. This series encapsulates the dynamic process of market equilibrium, reflecting how information disseminates and coalesces into a collective understanding over time. Concurrently, we examine the Volatility Series which offers insights into the impact of individual trading behaviours. Unlike traditional static measures, this series dynamically captures the fluctuations and perturbations caused by individual trades.

2.3.10 Dynamic Return

The $DTAA$ decomposes the original price series into two distinct components: the Dynamic Trend Series (DT) and the Volatility Series (V). The essence of the DT series is in its ability to be transformed into what we term the Dynamic Trend Driven Price (DP). This transformation implies that the price at any given time t, denoted as P_t, is the sum of the Dynamic Trend-Driven Price at t, DP_t, and the Volatility at t, V_t, expressed as:

$$P_t = DP_t + V_t \tag{2.34}$$

Furthermore, we introduce the concept of Dynamic Return at any observed time point t, denoted as DR_t. This is defined as the percentage of the price change between two consecutive time points, mathematically represented as:

$$DR_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{2.35}$$

In our model, we operate under the assumption that the dynamic trend, represented by DT_t, remains stable over short intervals, specifically between two consecutive time points $t-1$ and t. Given the stable trend DT_{t-1} within the interval $t - t + 1$, the Dynamic Trend Driven Price at time point t , denoted as DP_t, are calculated as:

$$\begin{aligned} DP_t &= DP_{t-1} + \int_{t-1}^{t} DT_{t-1}\, dt \\ &= DP_{t-1} + DT_{t-1} \times (t - (t-1)) \\ &= DP_{t-1} + DT_{t-1} \times 1 \end{aligned} \tag{2.36}$$

In this equation, DP_t represents the Trend-Driven Price at time t, which is derived from the sum of the Trend-Driven Price at time $t-1$ (denoted as DP_{t-1}) and the product of the stable trend DT_{t-1} over the interval $[t-1, t]$. The term $t - t + 1$ effectively captures the duration of this interval, which in this context is equal to 1.

Given that the price at time t, P_t, is the sum of DP_t and the volatility at t, V_t, we express the price at t as:

$$P_t = DP_t + V_t = DP_{t-1} + DT_{t-1} + V_t \tag{2.37}$$

The price at time $t-1$ and t are expressed as:

$$\begin{aligned} P_t &= DP_{t-1} + DT_{t-1} + V_t, \\ P_{t-1} &= DP_{t-1} + V_{t-1} \end{aligned} \tag{2.38}$$

Substituting these into the equation for Dynamic Return, we obtain:

$$\begin{aligned} DR_t &= \frac{P_t - P_{t-1}}{P_{t-1}} \\ &= \frac{DP_{t-1} + DT_{t-1} + V_t - (DP_{t-1} + V_{t-1})}{DP_{t-1} + V_{t-1}} \end{aligned} \tag{2.39}$$

Reformulating, we express DR_t as:

$$DR_t = \frac{DT_{t-1}}{DP_{t-1} + V_{t-1}} + \frac{V_t - V_{t-1}}{DP_{t-1} + V_{t-1}} \tag{2.40}$$

Defining the change in Volatility as $\Delta V = V_t - V_{t-1}$, our equation simplifies to:

$$DR_t = \frac{DT_{t-1}}{DP_{t-1} + V_{t-1}} + \frac{\Delta V}{DP_{t-1} + V_{t-1}}. \tag{2.41}$$

Considering $P_{t-1} = DP_{t-1} + V_{t-1}$, the final form of the Dynamic Return equation is:

$$DR_t = \frac{DT_{t-1}}{P_{t-1}} + \frac{\Delta V}{P_{t-1}}. \tag{2.42}$$

This formulation enables us to dissect the dynamic return into components attributable to the Dynamic Trend and volatility changes, enhancing our understanding of price dynamics in financial markets.

Our previous analysis reveals a critical dependency of the Dynamic Trend (DT) series and the Volatility (V) series on the Time Level and Horizon. The Time Level mirrors that of the original Price series for the Fundamental Time Level Dynamic Trend Series, or the same as the Higher Level Dynamic Time Series, it depends on which kind of DT series we use. Specifically, the values of DT_t and V_t at any specific observer's time point t are intricately tied to the Time Level, Horizon and observation time point itself. Consequently, the Dynamic Return (DR) series also exhibits sensitivity to the Time Level and Horizon of the DT series. The Time Level and Horizon of DR mirror the Time Level and Horizon of the corresponding DT series.

For instance, consider the scenario where we employ a '1-minute Dynamic Trend Horizon 1-second Fundamental Time Level DT series', written as $1minDTH1sFundTL\ DT$. In this context, the corresponding Dynamic Return series would be a '1-minute Dynamic Trend Horizon 1-second Fundamental Time Level DR Series.' The definition of Dynamic Return at a specific observer's time point t, under a defined Time Level and Horizon getting from DT, can then be mathematically expressed as:

$$1minDTH1sFundTL\ DR_t = \frac{D_{t-1}}{P_{t-1}} + \frac{\Delta V}{P_{t-1}}. \tag{2.43}$$

This equation is the complete definition of Dynamic Return at a specific observer's time point, contingent upon the specified Time Level and Horizon mirroring the Time Level and Horizon of the corresponding DT series. Such a formulation allows us to precisely capture the impact of time-sensitive dynamics of price in financial markets, providing insights into the interplay between dynamic trends, volatility, and returns under varying observational parameters.

2.3.11 The Velocity of Dynamic Return

A pivotal aspect of our analysis is the concept of the velocity of dynamic return, denoted as the first derivative of the Dynamic Return (DR). Mathematically, this is represented as $\frac{d(DR)}{dt}$, or simply DR'. Given that the Dynamic Trend is essentially the velocity series of a time series, we can extend this principle to the DR series. By generating the Dynamic Trend Series of theDR series, we effectively obtain the velocity series of DR, which we refer to as the VDR (Velocity of Dynamic Return) series.

The VDR series sensitive to the Time Level and Horizon we use to generate theVDR series. The Time Level mirrors that of the original Price series for the Fundamental Time Level Dynamic Trend Series, or the same as the Higher Level Dynamic Time Series, it depends on which kind of DT series we use. The Horizon for generating the VDR series is denoted as $VDRH$. The VDR series also sensitive to the Horizon of the corresponding DT series. For instance, if we set the $VDRH$ to 30 seconds for a '1-minute Dynamic Trend Horizon 1-second Fundamental Time Level DR series' which is mirrors the Fundamental Time Level Dynamic Trend Series, the full definition of the Velocity of Dynamic Return at a specific observer's time point can be expressed as:

$$30sVDRH1minDTH1sFundTL\ VDR_t \tag{2.44}$$

This formulation allows for a nuanced understanding of the rate of change of dynamic returns, providing a deeper insight into the temporal dynamics of financial markets.

2.3.12 Defining Structural Breaks by Using Dynamic Trend

2.3.12.1 Data Changing Vector and Structural Breaks

The Dynamic Trend(DT) can be decomposed into two vectors: the Time Changing Vector (TC) and the Data Changing Vector (DC). As we set a certain Horizon, the change of DT only comes from the change of DC. Structural breaks in the time series occur when there is either a change in direction or a change in magnitude while maintaining the same direction for DT . We have $slopeT = SignS \cdot |slopeT|$ as the unbiased proxy for DT.

2.3.12.2 Definition of Type-I and Type-II Breaks

We define structural breaks caused by a change in direction as Type-I breaks, and breaks that are due to magnitude changes within the same direction as Type-II breaks.

- **Type-I Breaks**: Occur when there is a change in the direction of the Data Changing Vector.
- **Type-II Breaks**: Occur when there is a change in the magnitude of the Data Changing Vector, without a change in its direction.

Type-I and Type-II structural breaks divide the entire series into several segments. Each segment contains only one certain trend, providing a more segmented and detailed analysis of the time series data. This distinction allows for a nuanced understanding of structural changes in the dataset, contributing significantly to the accuracy and depth of time series analysis.

2.3.12.3 Analysis of Type-I Breaks

Type-I breaks are characterised by changes in the direction of DT. The direction of DT can be categorised as upward, downward, or flat.

- **Upward Direction**: This occurs when the direction is the same as the y-axis, corresponding to a positive $SignS$. In this case, the velocity is positive, indicating that the data trend is increasing over time on average.
- **Downward Direction**: This occurs when the direction is reversed compared to the y-axis, corresponding to a negative $SignS$. Here, the velocity is negative, suggesting that the data trend is decreasing over time on average.
- **Flat Direction**: This happens when there is 0 Data Changing component vector, and the data remains unchanged with a $slopeT$ of zero. The velocity is zero, indicating no change in the data over time on average.

Type-I breaks occur when there is a shift between these three types of directions. Earlier mathematical analysis confirms that the slope of a Dynamic Trend vector is a reliable proxy for the dynamic trend. By considering the sign of the slope, we can easily determine the direction of DT. Therefore, changes in the slope sign can be used to detect Type-I breaks.

The shifting among these three types, referred to as Type-I breaks, is significant in the analysis of time series data, as it indicates fundamental changes in the trend behaviour.

2.3.12.4 Analysis of Type-II Breaks

The identification of Type-I breaks, based on the direction of DT, facilitates the segmentation of data into distinct upward, downward, and flat trend segments. However, within each of these segments, significant changes in magnitude may also occur. To address these changes, we introduce the concept of Type-II breaks, which describe instances where the trend shifts within segments of the same direction.

As $SlopeT$ changes, both the magnitude and direction of DT change accordingly. This relationship allows for the use of changes in the slope of the Dynamic Trend vector to represent changes in the magnitude of the Data Changing Vector(DC).

Type-II breaks in upward and downward segments are characterised as follows:

- **Upward Segments**: A Type-II break can occur with an upward shift in magnitude accompanied by an upward shift in $|SlopeT|$, or with a downward shift in magnitude accompanied by a downward shift in $|SlopeT|$.
- **Downward Segments**: A Type-II break can occur with an upward shift in magnitude accompanied by an upward shift in $|SlopeT|$, or with a downward shift in magnitude accompanied by a downward shift in $|SlopeT|$.

These breaks are crucial for understanding the nuances of trend shifts within segments that maintain the same general direction but experience significant changes in magnitude. Type-II breaks are a valuable analytical tool for detecting changes in the trend that occur within segments of the same direction. By accounting for these changes, we can achieve a more clear understanding of the dynamics of a time series.

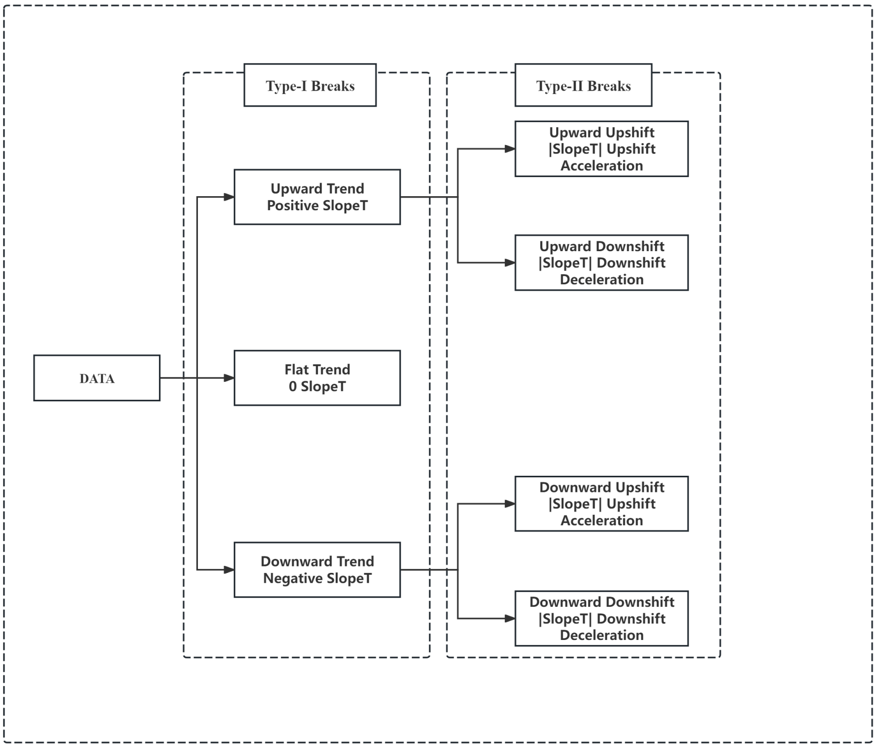

FIGURE 2.6: Define the Structural Breaks in Data

2.3.12.5 Summary of Define of Structural Breaks

Our proposed definition of structural breaks in data involves two types of breaks: Type-I and Type-II breaks.

The Dynamic Trend(DT) can be decomposed into two vectors: the Time Changing Vector (TC) and the Data Changing Vector (DC). As we set a certain Horizon, the change of DT only come from the change of DC.

Type-I breaks occur when there is a change in the direction of the Data Changing Vector (DC), which can be categorised as upward, flat, or downward. The $SignS$ (sign of the slope of TV) aligns with the sign of $Sign \cdot |DCU| \cdot |d_n - d_1|$, which effectively represents the direction of the DC. This alignment allows us to use the changes in $SignS$ to detect Type-I breaks. Note that when there is 0 Data Changing component vector, the data remains unchanged with a $slopeT$ of zero. The velocity is zero, indicating no change in the data over time on average.

Type-II breaks occur within segments of the data that maintain the same direction, either upward or downward, and are characterised by significant shifts in the magnitude of the DC. In upward segments, Type-II breaks can result from either an upward or downward shift in magnitude. Similarly, in downward segments, Type-II breaks can occur due to shifts in magnitude. The magnitude of the DC is represented by the Data Difference (DD). $|SlopeT|$ has a linear relationship with DD, facilitating the detection of Type-II breaks through changes in$|SlopeT|$.

In conclusion, the complex dynamics of data trends can be effectively examined and understood through the combined application of Type-I and Type-II breaks. Type-I breaks allow us to segment data into distinct directional categories such as upward, downward, and flat. Meanwhile, Type-II breaks enable a deeper exploration of shifts within these same-direction segments.

- **Type-I Breaks**: Facilitate segmentation of data into directional categories, aiding in the initial analysis of trend direction.
- **Type-II Breaks**: Enable detailed exploration of magnitude shifts within the Type-I breaks segments, providing insights into subtler trend dynamics.

In both upward and downward segments, changes in the magnitude of Type-II breaks are accompanied by corresponding shifts in the absolute value of the average velocity. These shifts signal either an acceleration or deceleration in the rate of data change.

This analytical approach, encompassing both Type-I and Type-II breaks, offers a robust perspective on data dynamics. It enables us to extract meaningful insights from complex pattern time series, improving our understanding of the underlying patterns and tendencies in the data.

2.3.13 Estimating Dynamic Trend and Detecting Structural Breaks

2.3.13.1 Estimation of Dynamic Trend

Structural breaks occur when there is either a change in direction or a change in magnitude within the same direction. We aim to construct a simple and efficient method to detect these structural breaks and segment the series accordingly.

The Dynamic Trend refers to the measure of data change observed at a specific time point, denoted as t_i, over a given Horizon n. Its primary objective is to capture both the direction and magnitude of the changing trend. The slope of the Dynamic Trend vector, denoted as $slopeT$, serves as a reliable proxy for representing the Dynamic Trend at observer's time point t_i within the Horizon n.

To estimate the slope and generate a Dynamic Trend Series corresponding to the original time series, we employ rolling window estimation. This method involves defining a window size equal to the Horizon n, where the first time point in the window corresponds to the observer's time point t_i. The estimated slope obtained from this window is then considered as the Dynamic Trend at the observer's time point.

2.3.13.2 Setting the Window Size

Setting the window size as the Horizon will satisfy the assumption that the trend does not change within the Horizon at this observer's time point. Note that at each observer's tiem point the Dynamic Trend could be different. When considering the Dynamic Trend Series, the Horizon is the threshold of trend. The pattern keep unchanging for at least Horizon time pattern for the Dynamic Trend, the corresponding segment of original series could be consider as trend segment. It specified the Dynamic Trend estimation's sensitivity. Data shifting patterns lasting less than the chosen Horizon (window size) are not regarded as trends; only patterns equal to or greater than the set Horizon are considered trends.

For example, in a 24-hour dataset with a 1-second interval between each data point, setting the window size to 61 , means that only patterns continuing for 1 minute or longer are considered trends. Patterns lasting less than 1 minute are categorised as volatility or outliers. The estimated Dynamic Trend series may have some Dynamic Trend pattern that less than the window size. To address this issue, the window size is set as a threshold for trend detection. Patterns persisting for durations less than the threshold are Patterns that persist for shorter periods of time than the threshold are cancelled out by set to be 0 values. Corresponding patterns within the initial data series that have durations below the threshold are then categorised as either outliers or volatility.

2.3.13.3 Ordinary Least Squares (OLS) Regression for Trend Estimation

Ordinary Least Squares (OLS) regression is applied within each window to estimate the Dynamic Trend vector. By fitting a line of best fit in the form $y = a + bt$, where y represents the values within the window and t represents the time indices, the coefficient b derived from the OLS regression represents the estimated $slopeT$ or Dynamic Trend at the observer's time point t_i. This process is repeated by shifting the window one data point forward, estimating the Dynamic Trend at subsequent observer's time points.

2.3.13.4 Generation of Dynamic Trend Series

Following this approach, a Dynamic Trend Series is generated, capturing the changing direction and magnitude of the trend over time. Each value in this series corresponds to the estimated $slopeT$ at a specific observer's time point, providing a dynamic analysis of the data's behaviour over time.

Consider a time series denoted as $P = \{p_1, p_2, p_3, \ldots, p_n\}$ with a data size of n. We choose a window size of estimation, which is the Horizon, denoted as $Horizon$, which is 3 in this case. This means that our Horizon covers 3 data points.

To estimate the Dynamic Trend Series, denoted as $DTS = \{td_1, td_2, td_3, \ldots, td_{n-2}\}$, we apply OLS regression to each window of size $Horizon$. Given that there are $n - Horizon + 1$ windows in the original time series, the resulting estimated Dynamic Trend Series, DTS, will have $n - Horizon + 1$ values.

However, to match the DTS series with the original series P, we require two additional values in DTS. Since the data points beyond index n are unknown and represent the future, we assume that they follow the same trend as the last window. This assumption allows us to extend the DTS series to match the size of the original series P.

The generation process of the estimated Dynamic Trend Series, DTS, can be visualised in the following manner:

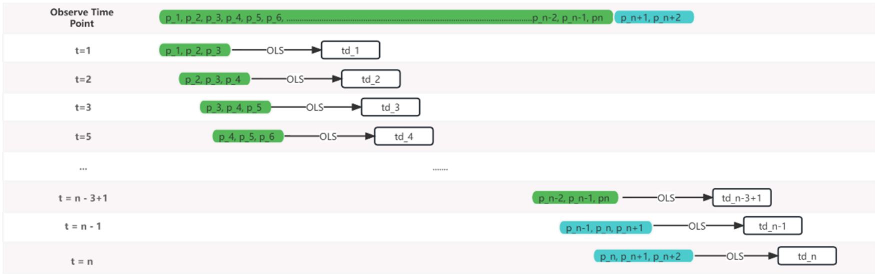

FIGURE 2.7: The Generation Process of Estimated Dynamic Trend Series

2.3.13.5 Structural Breaks Detection Using Dynamic Trend Series

Using the estimated Dynamic Trend Series $DTS = \{td_1, td_2, td_3, \ldots, td_n\}$, we can detect Type-I and Type-II breaks in the original time series.

- **Detection of Type-I Breaks**:

  To identify Type-I breaks, we examine the sign of each Dynamic Trend value td in DTS. A change in sign indicates the occurrence of a Type-I break, marking the break point. However, it is essential to assess the characteristics of segments within DTS that maintain the same sign. If a segment with the same sign is very short, for instance, less than the threshold $Horizon$, it is classified as volatility or an outlier. These segments in DTS will be set to 0 values, the related segments in original series may require separate handling using outlier detection methods.

  For the segments in DTS that have positive td values, the corresponding segments in the original time series, denoted as P, are categorised as Type-I upward trend segments. Conversely, segments in DTS with negative td values correspond to Type-I downward trend segments in P. Segments in DTS where $td = 0$ correspond to Type-I flat trend segments in P.

  The breaking points in the time series are identified as the points of transition among these three types of trends (upward, downward, and flat).

  By following this approach, we can detect Type-I breaks by analysing the sign changes in the Dynamic Trend Series and associating them with the corresponding segments in the original time series. This method allows us to identify shifts in trend patterns, including upward, downward, and flat trends, and pinpoint the locations of breaks.

- **Detection of Type-II Breaks**

  To detect Type-II breaks within each Type-I upward and downward trend segment, we utilise a hierarchical clustering algorithm known as Agglomerative Nesting (AGNES).

  AGNES is a widely used method in hierarchical clustering, which aims to group similar objects or data points into clusters. The AGNES algorithm begins by treating each data point as an individual cluster. It then iteratively merges the most similar clusters until all data points are amalgamated into a single cluster. This iterative process generates a dendrogram, a tree-like diagram representing the hierarchical relationships between the clusters.

When applied to the Dynamic Trend values (*td*), AGNES helps to identify clusters of similar *td* values. Each cluster thus identified represents a distinct Type-II segment within the Type-I segment.

AGNES generates multiple levels of clusters, but determining the appropriate number of clusters is crucial. A common approach is to plot the dendrogram and visually inspect it for cluster formation. However, this method can be subjective.

To provide a more quantitative measure, we employ the silhouette coefficient. This coefficient assesses the compactness and separation of clusters, producing a value between -1 and 1. A score close to 1 indicates well-separated and compact clusters, while a score near -1 suggests poor separation and compactness. The optimal number of clusters can be determined by selecting the count that yields a silhouette coefficient close to 1.

Once the optimal number of clusters is determined, AGNES helps identify Type-II breaks within each Type-I upward and downward trend segment. These breaks are categorised as follows:

- **Upward Upshift Segments**: Represent shifts of increasing magnitude within Type-I upward trend segments.
- **Upward Downshift Segments**: Indicate shifts of decreasing magnitude within Type-I upward trend segments.
- **Downward Upshift Segments**: Denote shifts of increasing magnitude within Type-I downward trend segments.
- **Downward Downshift Segments**: Indicate shifts of decreasing magnitude within Type-I downward trend segments.

The breaking points are identified at the locations where magnitude shifts occur within segments of the same trend direction. By pinpointing these breaking points, we can effectively detect Type-II breaks within the Type-I upward and downward trend segments in the original time series.

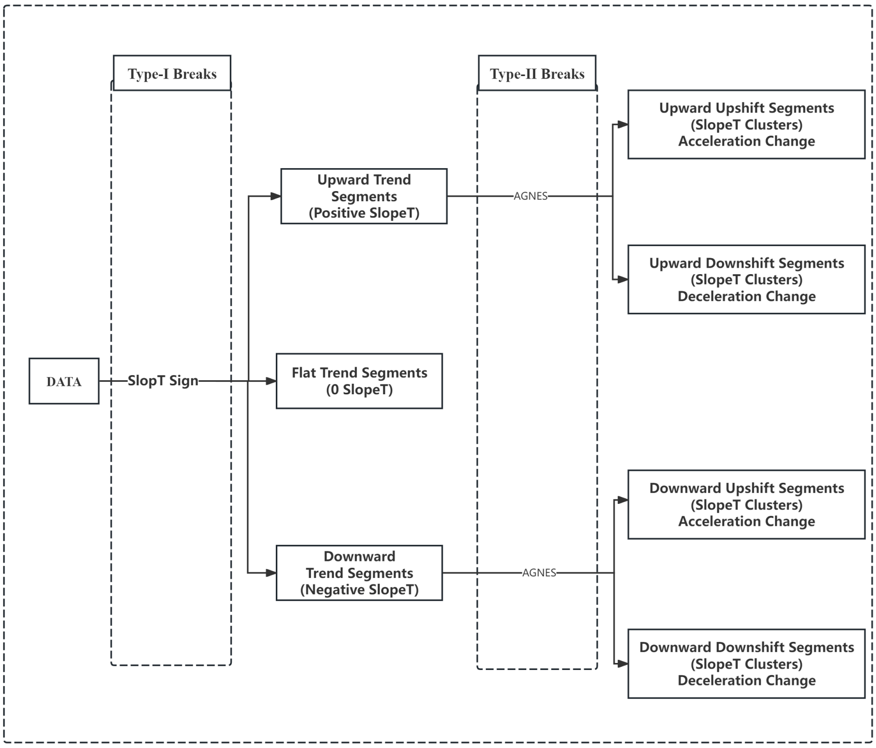

FIGURE 2.8: The Structural Breaks Detection Approach

2.3.13.6 Summary of Break Detection Process in Time Series

The process of detecting Type-I and Type-II breaks in a time series begins with the generation of an estimated Dynamic Trend Series. This is achieved using a rolling window estimation method, where Ordinary Least Squares (OLS) regression is applied within each window of a fixed size. Changes in the sign of the Dynamic Trend Series are indicative of Type-I breaks, which represent shifts in the trend direction.

In analysing Type-I breaks, segments within the Dynamic Trend Series that maintain the same sign are considered. Segments that are very short, less than the Horizon, are classified as volatility or outliers and are addressed separately. Positive Dynamic Trend values in these segments correspond to Type-I upward trend segments in the original time series, whereas negative Dynamic Trend values segments indicate Type-I downward trend segments. A Dynamic Trend value of 0 segments signifies Type-I flat trend segments.

Type-II breaks are detected within each Type-I trend segment using the Agglomerative Nesting (AGNES) hierarchical clustering algorithm. AGNES groups similar Dynamic Trend values into clusters, each representing a Type-II segment. The optimal number of clusters is determined using the silhouette coefficient, which assesses the compactness and separation of these clusters. A silhouette coefficient close to 1 helps identify the most appropriate number of clusters for Type-II break detection.

Type-II breaks are further categorised based on the magnitude shift within segments of the same trend direction. These categories include upward upshift, upward downshift, downward upshift, and downward downshift. The breaking points, where magnitude shifts occur within these segments, are key to identifying Type-II breaks.

By integrating the detection of Type-I and Type-II breaks, this approach allows for a comprehensive analysis of significant shifts in trend patterns within the original time series. The approach offers a deep understanding of the data dynamics.

2.3.14 Acceleration and Deceleration of Data

Further analysis of how average velocity interacts with these scenarios provides additional insights. For example, in an upward trending segment experiencing an upward shift in magnitude (indicative of an upward Type-II break), there is an increase in the absolute value of the average velocity. This increase signals an acceleration in the rate of data change.

Conversely, if an upward trending segment undergoes a downward shift in magnitude, indicative of a downward Type-II break, there is a decrease in the absolute value of the average velocity. This reflects a deceleration in the rate of change in the data.

Transitioning to a downward trending segment, when an upward shift in magnitude occurs (symbolising an upward Type-II break), an increase in the absolute value of the average velocity is observed. This suggests an acceleration in the rate at which data is changing.

Lastly, if a downward trending segment experiences a downward shift in magnitude, denoting a downward Type-II break, the absolute value of the average velocity diminishes. This implies that the rate of data change is experiencing deceleration.

2.3.15 Dealing with Structural Breaks

The methodology described above delineates two types of breaks within a time series: Type-I and Type-II breaks. Type-I breaks are characterised by a change in the trend direction, whereas Type-II breaks result from a change in the magnitude of the trend while maintaining the same direction. These breaks effectively partition the entire time series into distinct segments, each with its own specific trend.

Type-I breaks represent transitions from one trend direction to another. They are pivotal in capturing shifts in the overall pattern of the time series, for example, a change from an upward to a downward trend, or vice versa. Such breaks are crucial for identifying pivotal moments where the general direction of the data changes significantly.

Type-II breaks, on the other hand, emphasise changes in the magnitude of the trend within the same direction. These breaks highlight variations in the intensity or strength of a consistent trend direction. They are instrumental in identifying shifts in the rate or amplitude of the trend, offering insights into the evolution of the trend over time.

By integrating both Type-I and Type-II breaks, the time series is segmented into intervals, each displaying a consistent trend pattern. This segmentation enables a more comprehensive analysis of the temporal dynamics and changes within the data. It facilitates the identification and interpretation of distinct patterns and trends, enhancing our understanding of the underlying processes driving these changes.

For each segment identified as having a certain trend in the time series, we apply a detrending process. For upward and downward trend segments, we remove the specific trend identified in each segment. For flat trend segments, we remove the mean value. This process transforms the time series with structural breaks into a series without structural breaks, resulting in a series with a zero mean.

- **Handling Flat Trend Segments**:
  To handle Flat Trend Segments, we employ a straightforward approach where the mean value of each segment is subtracted from the data points within that segment. Mathematically, this detrending operation for Flat Trend Segments can be expressed as follows:

  $$p = p - \mu_{\text{segment}} \tag{2.45}$$

  Here, p represents the data points within the Flat Trend Segment, and μ_{segment} denotes the mean value of that segment. By subtracting the mean value, μ_{segment}, from each data point, p, in the segment, we effectively remove the mean and detrend the data within the Flat Trend Segment.

- **Handling Upward and Downward Trend Segments**
  To detrend the data within the Upward Upshift Segments, Upward Downshift Segments, Downward Upshift Segments, and Downward Downshift Segments, we utilize a method called "detrend". This method is a standard mathematical technique employed for removing trends from time series data. It involves using linear regression to estimate the trend line of each segment.

  The detrending process can be mathematically represented as follows:

  $$p_d = p - p_{\text{trend}} \tag{2.46}$$

  Here, p represents the data points in the above-mentioned trend segments, p_{trend} is the estimated trend line for the segment, and p_d signifies the detrended data.

  For each type of segment, Upward Upshift, Upward Downshift, Downward Upshift, and Downward Downshift, the detrend process is systematically applied. This results in new segments that are devoid of the trend component and exhibit a zero-mean value.

Additionally, for the Flat Trend Segments, we perform a demeaning process. This means that we subtract the average value of the segment from each individual data point within that segment. By doing this, we create new segments where the average value becomes zero.

After performing the detrend operation on Upward, Downward Upshift, and Downshift Segments, and the demean operation on Flat Trend Segments, the detrended and demeaned data are combined. This combination aligns the modified data with their corresponding time points, resulting in a new time series. The resultant time series is characterized by the absence of trends and a mean value of zero.

This Dynamic Trend Analysis Approach (*DTAA*) effectively decompose the original series into Dynamic Trend Series and 0 mean Volatility Series, enabling further analysis and interpretation of the underlying patterns and fluctuations in the data.

2.3.16 Removing Cyclical Effect Using Dynamic Trend Analysis Approach

The *DTAA* offers the advantage of mitigating the cyclical effect in a time series. In time series analysis, a cycle refers to a repeating pattern or fluctuation occurring over a specific period. To effectively address this, we segment the cycle into a suitable number of parts, treating all these segments as trend segments. This segmentation allows us to mitigate the influence of the cyclical effect. The length of the Horizon set in the detrending method should be less than one-quarter of the cycle length.

The *DTAA* employs linear regression to estimate the trend line of the data. By subtracting this estimated trend line from the original time series, we obtain detrended data that removes both the underlying trend and the cyclical effect.

The *DTAA* is particularly effective in capturing and eliminating the cyclical effect in cases where the cycle length is greater than four times the *Horizon* (window size of the rolling window estimate).

By choosing a Horizon for detrending that corresponds to the length and characteristics of the cyclical pattern, or aligns with the overall structure of the original series when the cycle length is unknown, we can successfully categorize all segments as trend segments. By treating all segments as trend segments, we can effectively remove the cyclical effect using the detrending method. This approach helps in revealing the underlying patterns and fluctuations in the time series data, facilitating analysis without the interference of cyclic variations. Assume we have a time series with the cycle of 1000, we use *DTAA* with *Horizon* of 21 to remove the cyclical effect. The result is plotted below.

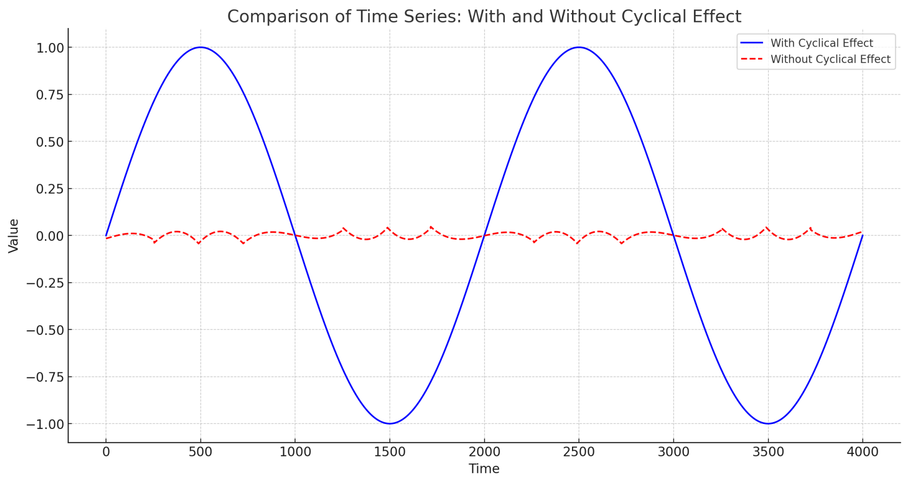

FIGURE 2.9: Removing Cyclical Effect Using DTAA

2.4 Main Empirical Results

The data used in our study were extracted from the online betting website Betfair.com, based in London, United Kingdom. A betting exchange like Betfair operates differently from traditional betting systems, resembling more closely a stock exchange. In this market, no single agent acts as a bookmaker. Instead, agents can assume the roles of both customer and bookmaker by choosing to either place a bet on an outcome ('back') or offer a bet ('lay') that another agent can back.

For instance, an agent betting £1 on an outcome (a 'back bet') at 3:1 odds expects to receive £4 if they win (a return of £3), but lose their £1 stake in the event of a loss. Conversely, laying a bet involves acting as the bookmaker and accepting the bet of another. Using the same example, the layer must pay out £4 to the successful bettor in the event of a winning outcome but retains the £1 stake if the outcome is a loss. Laying, therefore, is equivalent to betting against an outcome.

The matchmaking in betting markets like those on Betfair is managed with a double auction order book, akin to a regular financial market. However, instead of the conventional bid and ask columns, these markets use 'back' and 'lay'. It's important to note that, unlike a stock market, a betting market has a definite conclusion, with the odds moving toward infinity or zero as the outcome becomes certain.

For our research, we have utilized the market proxy odds for five flat horse races (treated as five separate markets) from Betfair. The reason we use the odds of betting market is that the odds data is very fluctuate as the the time going comparing with economic time series, it will give us more clear and frequency occurring structural breaks. This method we proposed is not just using on betting market data, but could dealing with all kinds of time series data as it is capable to capture the data changing process for time series data. Specifically, we have used the last observed best back odds at 30-second intervals for around a 12-hour period before the start of each race as the price data for each horse in each data to generate the market proxy best back odds at 30-second intervals for around a 12-hour period before the start of each race. The markets and their respective details are as follows:

- **Market One (M1)**: Flat racing on 2/2/2019 (market id: 1_154344450).
- **Market Two (M2)**: Flat racing on 23/2/2019 (market id: 1_155215202).
- **Market Three (M3)**: Flat racing on 16/11/2019 (market id: 1_165070116).
- **Market Four (M4)**: Flat racing on 16/11/2019 (market id: 1_165070123).
- **Market Five (M5)**: Flat racing on 21/12/2019 (market id: 1_166582140).

The odds are our price here. The market proxy price is calculated as below:

$$P_{\text{proxy,t}} = \sum_{i=1}^{N} (P_{it} \times W_{it}) \tag{2.47}$$

Where, $P_{\text{proxy, t}}$ is the proxy price for the market at time t. N is the total number of horses in the market. P_{it} is the price of the i-th horse at time t. W_{it} is the weight of the i-th horse at time t.

$$W_{it} = \frac{S_{it}}{\sum_{j=1}^{N} S_{jt}} \tag{2.48}$$

Where, S_{it} is the size of the i-th horse at time t. N is the total number of horses in the market. $\sum_{j=1}^{N} S_{jt}$ is the sum of the sizes of all horses in the market at time t. W_{it} is getting from the size of back betting of each horse at each observer's time point, it included the information of the size of betting of each horse at each time t. Here the size serves as volume information in the stock market.

The descriptive statistics for the market proxy price series from these markets are presented in the below tables. Notably, M4 proxy back prices have the highest standard deviation of 6.992. In contrast, M5 proxy back prices have the lowest standard deviation of 2.076.

MARKET	OBSERVATIONS	MEAN	MIN	MAX	MEDIAN	MODE	STD
M1	1328	10.481	2.645	34.854	9.935	9.580	4.015
M2	1383	7.326	1.415	32.384	4.945	17.389	5.703
M3	1391	16.434	6.390	32.292	16.116	14.773	4.245
M4	1287	12.206	2.509	48.374	10.558	16.696	6.992
M5	1341	6.083	1.942	18.346	5.717	4.456	2.076

TABLE 2.2: Descriptive Statistics for Market Proxy Price Series - Part 1

MARKET	VAR	KURTOSIS	SKEWNESS	P25	P50	P75
M1	16.120	1.906	0.977	7.702	9.935	12.614
M2	32.529	1.441	1.379	3.199	4.945	9.216
M3	18.017	0.783	0.591	13.816	16.116	18.562
M4	48.888	1.470	1.190	6.392	10.558	16.287
M5	4.309	0.526	0.685	4.456	5.717	7.469

TABLE 2.3: Descriptive Statistics for Market Proxy Price Series - Part 2

Figs. 1.7 to 1.11 present time series plots of market proxy prices and the price for each horse, the time period is around 12 hours.

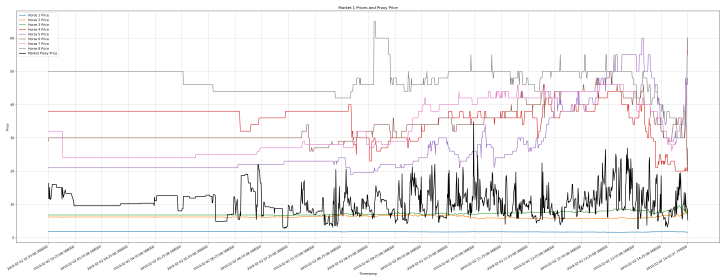

FIGURE 2.10: M1 Market Proxy Price and Individual Horse Prices

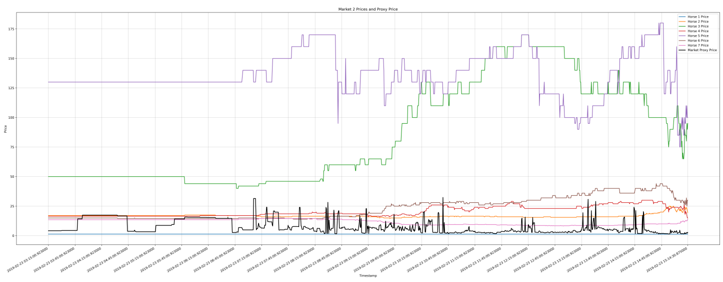

FIGURE 2.11: M2 Market Proxy Price and Individual Horse Prices

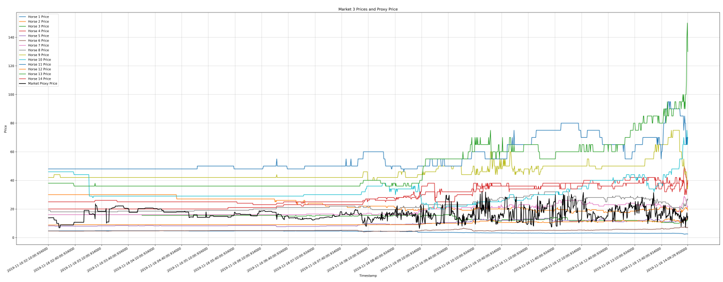

FIGURE 2.12: M3 Market Proxy Price and Individual Horse Prices

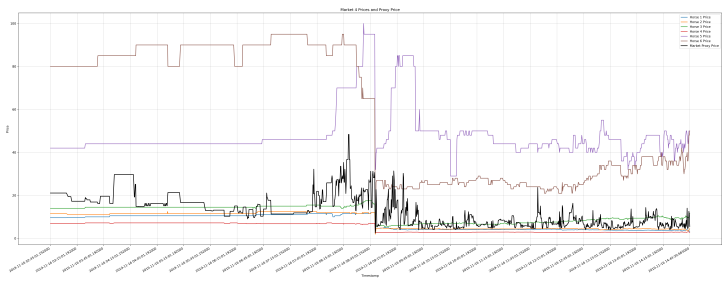

FIGURE 2.13: M4 Market Proxy Price and Individual Horse Prices

FIGURE 2.14: M5 Market Proxy Price and Individual Horse Prices

The market proxy prices graphs of M1, M2, M3, M4, and M5 exhibit multiple trends and distinctive patterns. In particular, the price series of M4 displays a significant gap in the available data. These observations suggest that these five markets are affected by structural breaks, which are abrupt shifts or changes in the underlying dynamics of the market.

To address this issue, we will employ the Dynamic Trend Analysis Approach ($DTAA$) to transform the price series of these markets into a series that is free from structural breaks. $DTAA$ is a method that allows us to identify and account for these breaks, enabling us to create a more consistent and reliable representation of the market dynamics.

By applying $DTAA$, we aim to mitigate the effects of structural breaks. We use a 15-minute Horizon, resulting in 31 data points within each observation window. Each market proxy price series is approximately 1300 in size. This window size is not only sufficient for performing Ordinary Least Squares (OLS) estimations, but it is also large enough to discern cyclical effects that span more than 62 minutes in length. The following figures illustrate the price series and trend series under consideration.

The window size of 31 gives a Temporal Granularity Index (TGI) of approximately 0.02. This TGI value, represents a relative narrowness of the window size compared to the overall dataset. It indicates the Dynamic Trend closely approximates theIDT. The Dynamic Trend helps uncover patterns in the data in a relatively accurate way. The data we use is 30-second data, which means the generated Dynamic Trend is a 30-second time level dynamic trend. Below are the plots of $15minDTH30sFundTL\ DT$ for the proxy price series for all 5 markets.

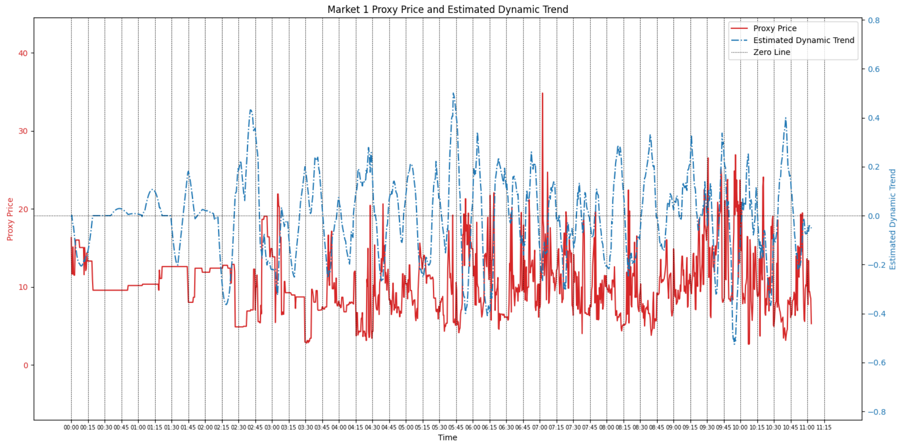

FIGURE 2.15: Market1 Market Proxy Price and Dynamic Trend Series

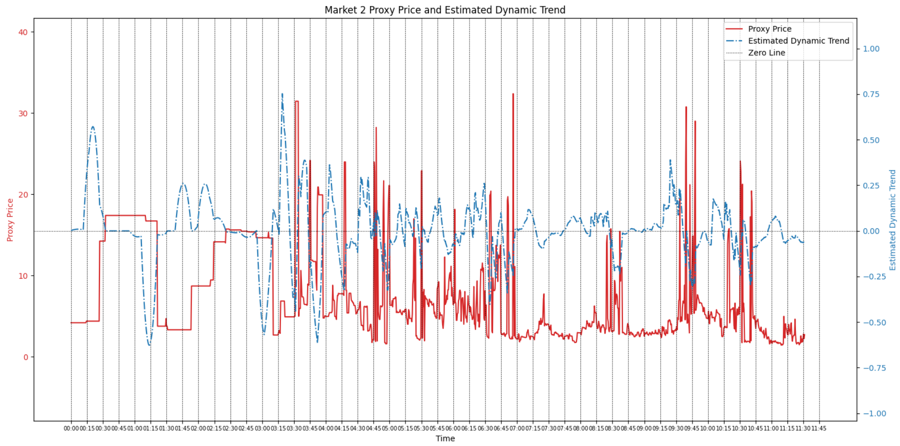

FIGURE 2.16: Market2 Market Proxy Price and Dynamic Trend Series

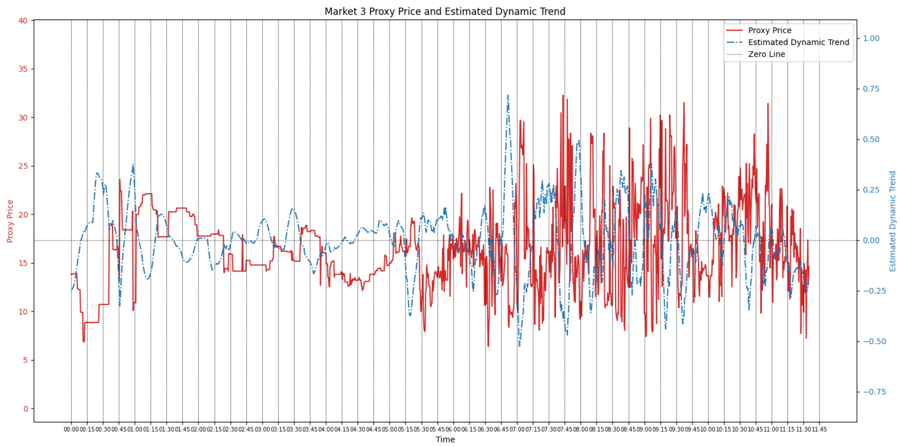

FIGURE 2.17: Market3 Market Proxy Price and Dynamic Trend Series

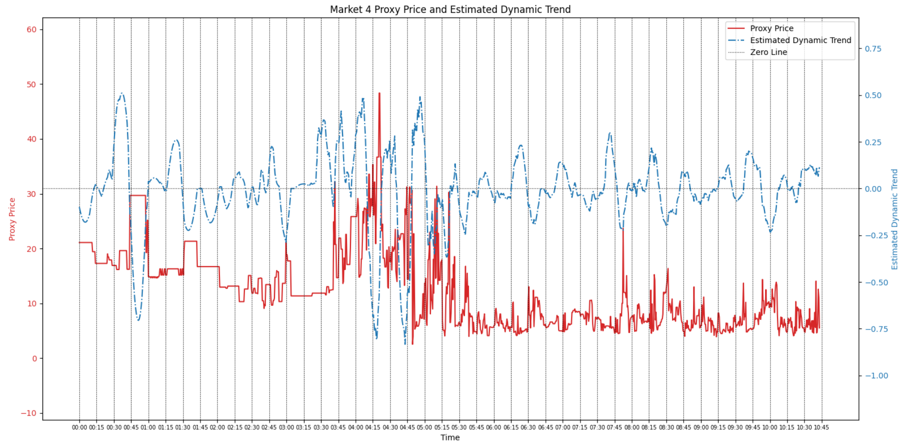

FIGURE 2.18: Market4 Market Proxy Price and Dynamic Trend Series

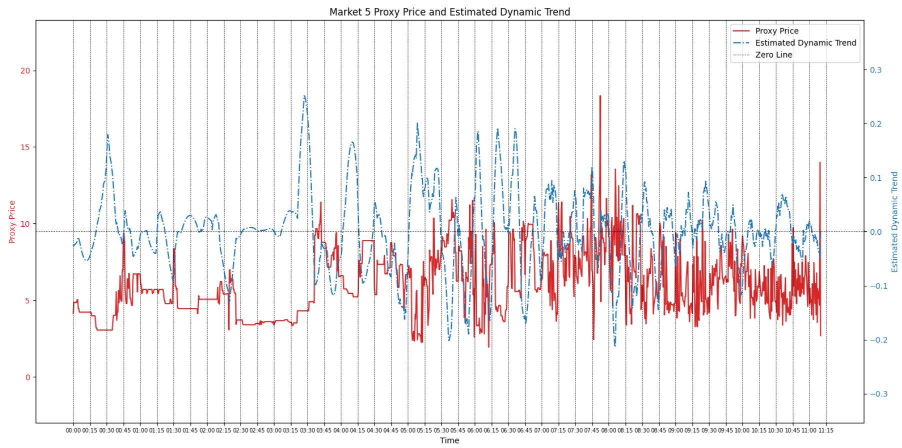

FIGURE 2.19: Market5 Market Proxy Price and Dynamic Trend Series

The Dynamic Trend captured most trend segments in the market proxy price series correctly. As we are using the rolling window estimation, the relationship may have some delay. The above 0 parts of the Dynamic Trend series captured the upward trend segments corresponding in market proxy price series, which are Type-I upward trend segments; the below 0 parts of the Dynamic Trend series captured the downward trend segments corresponding in market proxy price series represent Type-I downward trend segments; the 0 parts of the Dynamic Trend series captured the flat trend segments corresponding in market proxy price series indicate Type-I flat trend segments.

The Type-II breaks are further categorised into different types based on the magnitude shift within the segments of the same trend direction, including upward upshift, upward downshift, downward upshift, downward downshift, and flat trend segments. This could be detected by clustering the Dynamic Trend in each Type-I segment.

The Horizon 31 is set as a threshold for trend detection. Patterns persisting for durations less than the threshold are patterns that persist for shorter periods of time than the threshold and are cancelled out by set to 0 values. Corresponding patterns within the initial data series that have a duration below the threshold are then categorised as either outliers or volatility.

When we consider both Type-I and Type-II breaks, it allows us to divide the time series into separate segments. Each of these segments then displays a certain trend. Given that each segment has its distinct trend, we can adjust the data accordingly. For segments demonstrating an upward trend, we remove the upward trend; for those with a downward trend, we take out the downward trend. In segments with no significant increase or decrease, or flat trends, we subtract the mean, thereby transforming these segments into non-trending portions which is the Volatility Series.

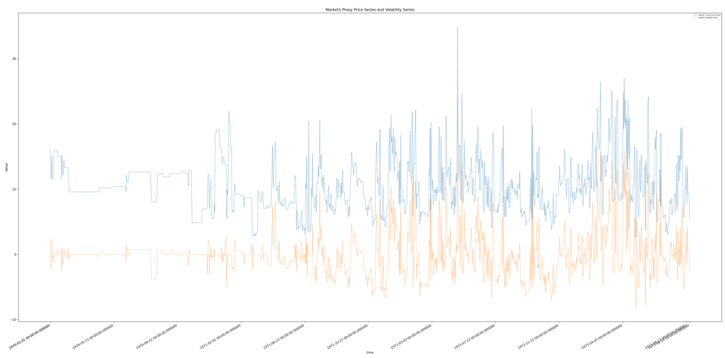

FIGURE 2.20: Market1 Proxy Price and Volatility Series

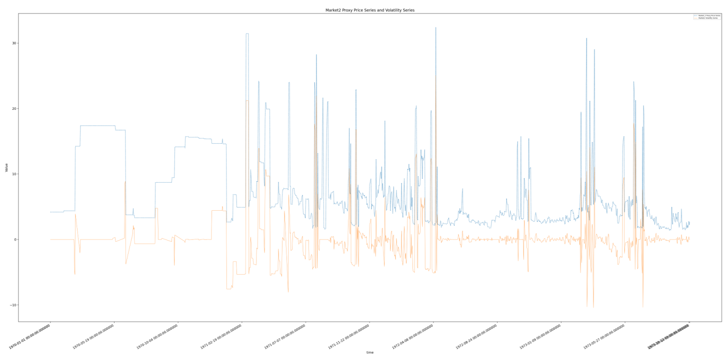

FIGURE 2.21: Market2 Proxy Price and Volatility Series

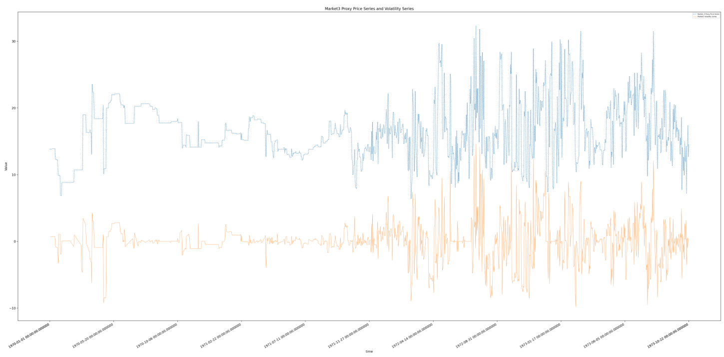

FIGURE 2.22: Market3 Proxy Price and Volatility Series

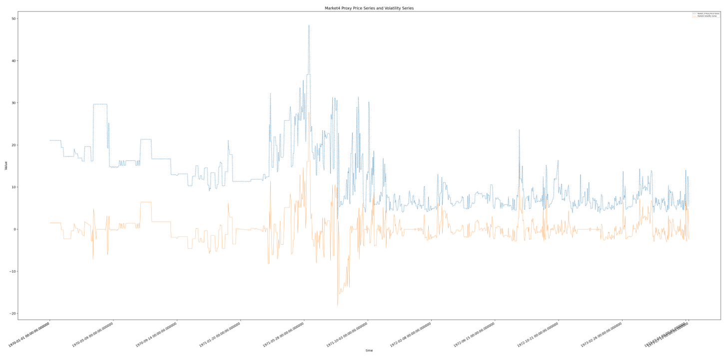

FIGURE 2.23: Market4 Proxy Price and Volatility Series

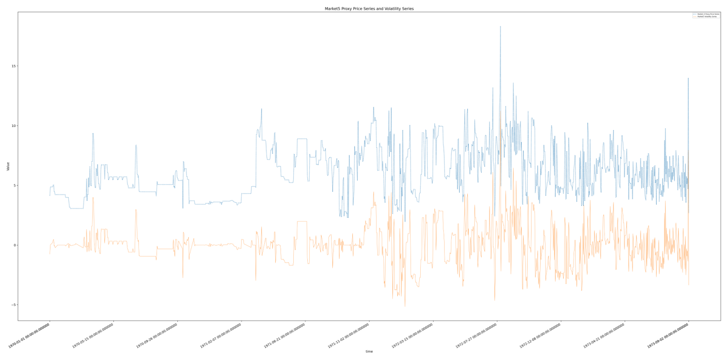

FIGURE 2.24: Market5 Proxy Price and Volatility Series

The detected trend segments, identified by the $DTAA$, have been approximated, fitted, and plotted as follows:

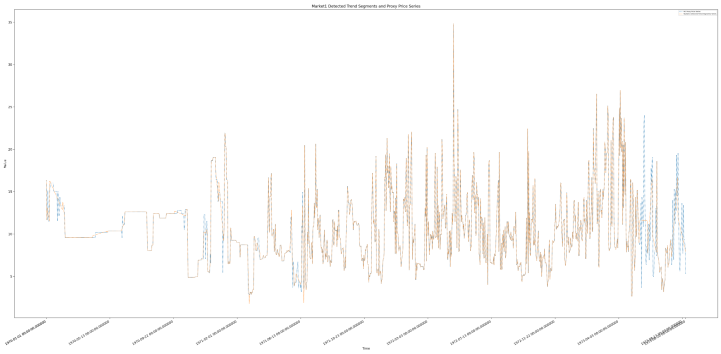

FIGURE 2.25: Market1 Price and Detected Trend Segments

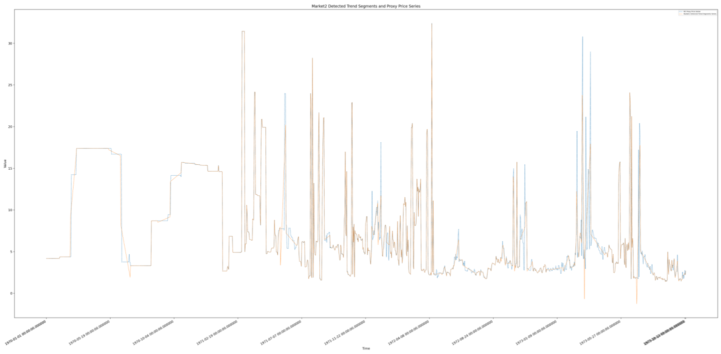

FIGURE 2.26: Market2 Price and Detected Trend Segments

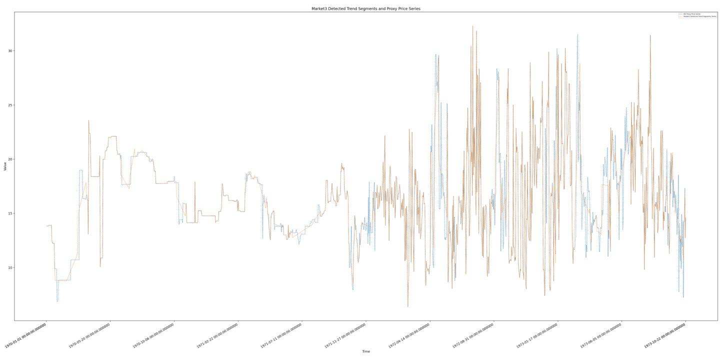

FIGURE 2.27: Market3 Price and Detected Trend Segments

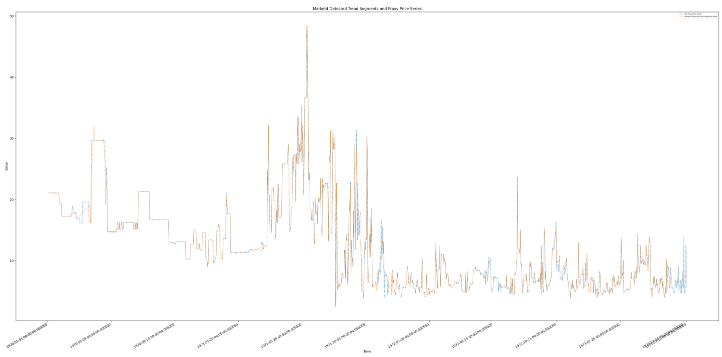

FIGURE 2.28: Market4 Price and Detected Trend Segments

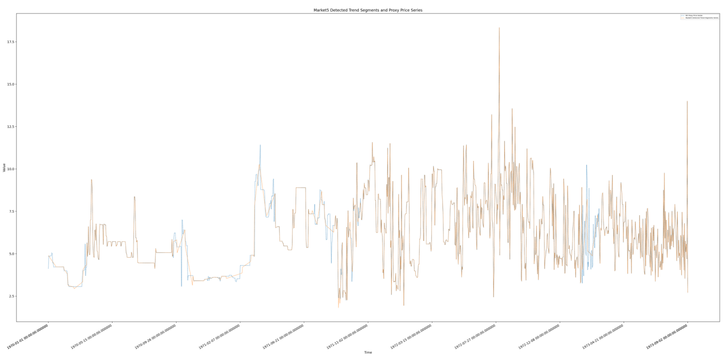

FIGURE 2.29: Market5 Price and Detected Trend Segments

The descriptive statistics for the Volatility Series for all markets are presented in the following tables.

MARKET	OBSERVATIONS	MEAN	MIN	MAX	MEDIAN	MODE
M1	1328	0.00	−8.17	24.04	0.00	0.00
M2	1383	0.00	−10.43	25.06	0.00	0.00
M3	1391	0.00	−9.79	14.62	0.00	−0.45
M4	1287	0.00	−18.17	27.69	0.00	1.80
M5	1341	0.00	−5.16	11.24	0.00	−0.93

TABLE 2.4: Descriptive Statistics for Volatility Series - Part 1

MARKET	STD	VAR	KURTOSIS	SKEWNESS	P25	P50	P75
M1	3.37	11.35	4.75	1.42	−1.81	0.00	0.75
M2	3.42	11.69	13.28	2.56	−0.63	0.00	0.17
M3	2.79	7.78	3.92	0.32	−0.78	0.00	0.88
M4	3.61	13.06	10.33	0.62	−1.54	0.00	1.33
M5	1.56	2.44	3.19	0.53	−0.80	0.00	0.64

TABLE 2.5: Descriptive Statistics for Volatility Series - Part 2

M1 shows a zero mean, suggesting that during the observation period, the average volatility was nearly neutral. There may be more extreme positive values because of the extended right tail suggested by the positive skewness of 1.42. Kurtosis of 4.73 suggests a leptokurtic distribution, meaning it has heavy tails and a sharp peak compared to a normal distribution. There is a noticeable variation in volatility between the minimum and maximum numbers.

M2 also has a zero mean. The standard deviation and variance are similar to M1, suggesting a similar volatility dispersion. The skewness is 2.56, indicating a significant right tail and more extreme positive values than M1. The high kurtosis of 13.28 signals a heavy tailed distribution. The volatility range is large, with a minimum of -10.43 and a maximum of 25.06.

M3 has a zero mean and the smallest standard deviation and variance among the markets, indicating less variability in volatility. The skewness is positive but close to zero, suggesting a more symmetric distribution. The kurtosis of 3.92 indicates a sharper peak than a normal distribution. The range is narrower than M1 and M2, indicating more moderate volatility.

M4 has a zero mean. It has the highest standard deviation and variance, indicating the most significant dispersion in volatility. The skewness is moderate, with a longer right tail. The kurtosis of 10.33 suggests a heavy tailed distribution. The range of volatility is the largest among all markets, with extreme minimum and maximum values.

M5 has a zero mean and the smallest standard deviation and variance, indicating the least dispersion in volatility. The skewness is moderate, with a slight right tail. The kurtosis of 3.19 indicates a less heavy tailed distribution. The range of volatility is the smallest, showing more stable volatility compared to other markets.

The descriptive statistics for *DT* Series for all markets are presented in the below tables.

MARKET	OBSERVATIONS	MEAN	MIN	MAX	MEDIAN	MODE
M1	1328	0.00	-0.53	0.50	0.00	0.00
M2	1383	0.00	-0.63	0.75	0.00	0.00
M3	1391	0.00	-0.53	0.72	0.00	0.09
M4	1287	0.01	-0.83	0.51	0.00	0.00
M5	1341	0.00	-0.21	0.25	0.00	0.00

TABLE 2.6: Descriptive Statistics for Dynamic Trend Series - Part 1

MARKET	STD	VAR	KURTOSIS	SKEWNESS	P25	P50	P75
M1	0.17	0.03	0.13	-0.05	-0.12	0.00	0.10
M2	0.18	0.03	3.01	-0.27	-0.06	0.00	0.07
M3	0.16	0.03	1.38	0.16	-0.09	0.00	0.08
M4	0.20	0.04	3.12	-0.92	-0.08	0.00	0.09
M5	0.07	0.01	0.84	0.24	-0.04	0.00	0.03

TABLE 2.7: Descriptive Statistics for Dynamic Trend Series - Part 2

M1 exhibits a zero mean, suggesting that the average dynamic trend of proxy prices is relatively stable over the observed period. The standard deviation and variance are low, indicating minor fluctuations around the mean. Given that skewness and kurtosis are close to zero, the distribution is probably almost normal with a small negative skewness.

M2 exhibits a zero mean, with slightly higher standard deviation and variance compared to M1. The kurtosis value of 3.01 indicates a more peaked distribution with heavier tails, suggesting more frequent extreme values. The skewness is slightly negative, indicating a distribution with a longer left tail.

M3 exhibits a zero mean, with a standard deviation and variance similar to M1 and M2. The positive skewness indicates a distribution with a longer right tail, suggesting slightly more extreme positive values. The kurtosis value of 1.38 indicates a distribution less peaked than M2 but more than M1.

M4 has a slightly negative mean, which is very close to 0, indicating a slight downward trend on average. The standard deviation and variance are higher compared to other markets, indicating more significant fluctuations. The high kurtosis of 3.12 suggests a more peaked distribution with heavier tails, indicating frequent extreme values. The negative skewness of -0.92 indicates a distribution with a much longer left tail, suggesting more extreme negative values. The range from -0.83 to 0.51 shows the greatest variability than other markets.

M5 exhibits a zero mean, with the lowest standard deviation and variance, indicating the least amount of fluctuation. The kurtosis and skewness are both low, suggesting a distribution close to normal with slight positive skewness.

The descriptive statistics for both the Volatility Series and Dynamic Trend Series show that the $DTAA$ is working well in decomposing price series into a Dynamic Trend Series and a Volatility Series. The Volatility Series component is a 0 mean series that fulfils the assumption that the mean is 0. The Dynamic Trend Series also have a means close to 0. We successfully convert the time series with structural breaks into two series without structural breaks.

2.5 Conclusions

This paper proposes a comprehensive methodology for analysing time series data, introducing the concept of the Dynamic Trend(DT), which encapsulates both the magnitude and direction of data changes over time. The Dynamic Trend is represented through a trend vector (TV), decomposed into the Time Changing vector (TC) and the Data Changing Vector (DC).

The TC vector symbolises the passage of time, aligned parallel to the x-axis, while the DC vector represents data changes, parallel to the y-axis. To assess the dynamic trend, the slope of a Dynamic Trend vector serves as a proxy ($SlopeT$), capturing both the magnitude and direction of the DC. This slope is a reliable measure for analysing the dynamic trend, providing insights into the extent and direction of data changes over time.

The framework distinguishes two types of structural breaks: Type-I and Type-II. Type-I breaks, indicating a change in the direction of the DT, lead to shifts between upward, flat, and downward trends. These breaks are detectable through changes in the $SlopeT$ sign, known as $SignS$.

Type-II breaks, in contrast, occur within segments maintaining the same direction and are characterised by significant changes in the magnitude of the DT. These breaks are identifiable by changes of $|SlopeT|$, offering insights into variations within a given Type-I segment.

To detect structural breaks, the methodology focuses on changes in the slope of the Dynamic Trend Vector. These changes signify shifts in the direction and magnitude of the DT, thus facilitating the identification of Type-I and Type-II breaks. Segmenting the data based on these breaks enables a deeper understanding of the dataset's dynamics and reveals underlying patterns.

Moreover, the Dynamic Trend also defines the velocity, acceleration, and deceleration of the data moving, this gives a deep understanding of the dynamic characteristics of time series data.

Additionally, the methodology addresses the cyclical effect, particularly effective when the cycle length exceeds four times the set Horizon.

In conclusion, this methodology offers a systematic approach to trend analysis and structural break detection in time series data. It encompasses mathematical definitions, vector analysis, and slope evaluation to capture the magnitude and direction of data changing over time. This approach transforms the time series with structural breaks into two series without structural breaks, providing a deeper understanding of data behaviour patterns.

Chapter 3

Title: Measuring Dynamic Efficiency in Betting Market

3.1 Introduction

After decomposing the time series into the DT Series and the Volatility Series using the aforementioned method, the dynamic information is contained in the DT Series, which reflects the market's response to information. Therefore, when analysing market efficiency, we can better understand the dynamic evolution of market efficiency by analysing the long memory of the DT Series. Since the DT Series naturally lacks structural breaks, this greatly improves the robustness of existing long memory models for time series. We use the DT Series of 5 betting markets to assess the dynamic market efficiency in Chapter 3.

Chapter 3 examines the dynamic efficiency of betting markets through the application of a novel decomposition model designed to analyse long-memory properties and informational inefficiencies within these markets. By identifying the integration orders (d) and constructing a metric for market inefficiency, denoted as D, this study reveals the evolving nature of market efficiency over time. The findings suggest that as betting activity progresses, betting markets exhibit increasing efficiency. The observed deviation patterns and inefficiency in convergence further confirm that market efficiency improves gradually over time, demonstrating a consistent trend across multiple markets.

In addition, this chapter introduces a method for identifying the optimal combination of window size (WS), bandwidth (BD), and estimator for accurately estimating the integration orders (d), referred to as the Estimation Score for Integration Orders (ESIO). The ESIO measures the proximity of the estimated integration orders to 1, providing a quantitative metric for assessing estimation accuracy. Various combinations of WS, BD, and four distinct estimators—LW, ELW, FELW, and Two-Step FELW—are tested within the framework. The optimal combination, characterised by the minimum ESIO, is selected for each betting market.

Moreover, this Chapter uses FCVAR to asses the fractional cointegration relationship and estimate the long-run equilibrium between DT series for all 5 markets' proxy prices and make predictions about their future values.

3.2 Literature Review and Method Development

The long memory feature has been found in price series of financial assets notably including stocks (Nguyen et al. (2020); Mensi et al. (2019)), bonds (Carbone et al. (2004); McCarthy et al. (2009)), futures (Dolatabadi et al. (2018); Kristoufek and Vosvrda (2014)), and real estate (Liow and Yang (2005); Liow (2009)), through various techniques. A few literature point out the long memory feature of the odds series and its role in governing the market inefficiency degree. Ross et al. (2015) propose a dynamic framework where there is an informed investor who is Bayesian and learns from a price equilibrium where other players are noise traders. The rate of convergence to a rational expectations equilibrium is determined, inter alia, by the economic power of the informed agent. Intuitively, if the informed agent is large relative to the rest, then the rate of convergence is slow as she has less to learn from the rest of the market, and Ross et al. (2015) demonstrates analytically that such a dynamic framework possesses the long memory property.

Besides, Duan et al. (2023) also introduces a decomposition method to study the component-specific effects of long memory on the dimension of informational inefficiency. By using this decomposition method, we could disentangle the inefficiency degrees among horse racing markets into two terms. The first term reflects the cross-market sum of variations around a convergent benchmark of the memory parameter of the odds series. The second term measures the extent to which the convergent benchmark differs from 1, a degree of total efficiency. Utilising this decomposition method, we could observe variations of the memory parameter around the convergence, as well as the extent to which the convergent benchmark of the memory parameter closes to the state of an efficient market.

The dynamics of the inefficiency degree are quantified through the identification of long-memory in the Bitcoin price series, i.e. its fractional integration order (d), using a robust semi-parametric Feasible Exact Local Whittle (FELW) estimator (Shimotsu (2010)) on a rolling -window basis.

Literature reveals that the traditional betting markets are based on bookmakers. The EMH tests are based on the odds that come from bookmakers. Some literature focuses on the favourite-longshot bias (FLB).Elaad et al. (2020) assess the odds set by fifty-one online bookmakers by using the standard Mincer and Zarnowitz (1969) forecast efficiency evaluation framework to test the EMH. Sung and Johnson (2007) propose a conditional Logit model to assess the probability of inaccurate odds-implied caused by bettors over-estimating of information content of trends in odds.

Recently, a new form of betting market has emerged in which bettors place wagers directly with one another (peer-to-peer betting). Some studies are based on this new form of betting market. The term for these markets is "betting exchanges." Jones et al. (2004) argued that the odds on offer from betting exchanges are generally much more competitive. Griffiths (2005) shows that the betting exchanges provide excellent value for the gambler and there is no bookmakers' mark-up on odds. As betting exchanges allow anyone to trade, including horse owners, trainers, riders, or anybody who asserts they have more information and uses it to buy or sell, the informed could increase the market's efficiency (Sung et al. (2016)). Smith et al. (2006) suggest that betting exchanges have brought about significant efficiency gains by lowering transaction costs for consumers.

Based on the above literature, the dynamic analysis disentangles the inefficiency degree into two separate components which could reveal the long memory at both an individual level and an overall level, which is missing in extant literature.

In this paper, we will do the dynamic analysis by using a novel decomposition model proposed by Duan et al. (2023) to disentangle the inefficiency degrees into two separate components. The first component reflects variations around the convergent value whilst the second component measures the degree to which the convergent value differs from a state of total market efficiency as characterized by integration measures.

More specifically, we will test the EMH in betting exchange by using the odds data from Betfair by using a research framework constructed by Duan et al. (2023), that reconciles different degrees of informational inefficiency of the horse racing betting markets to that of the varying degree of long-memory of the odds series featured by various fractional integration orders. The framework is presented in a dynamic setting to estimate the time-varying nature of informational inefficiency at a point in time. Conventionally, by examining if the price series characterizes permanent memory (i.e. I(1) process) or short memory (i.e. I(0) process), testing market efficiency is treated as a solution to a polar question that concludes either acceptance or rejection of EMH. However, the conventional tool-edge solution of the I(1)/I(0) framework neglects the possibility that the integration order (d) of the odds series can naturally be of a fractional order characterized by a long memory. This indicates that while the EMH might not be satisfied, a fractionally integrated price series can be far from a memoryless process, implying that the information reflection on its dynamics can be highly persistent and slowly decayed over time.

3.3 Methodology

Firstly, we first statistically identify the memory property in the horse racing odds series by investigating the time structure of the series with various integration orders (d).

Secondly, we apply the corresponding estimation technique for empirical estimation of the d value.

By gauging the closeness of the integration order of the horse racing odds series with one, we could measure the inefficiency degree of the horse racing exchange market.

3.3.1 Identification of the Memory Property in Horse Racing Odds Series

Several econometric methods have been developed, some of which have been redesigned keeping abreast of fast-paced developments in computational methods and modelling of important dynamics. While there is no theoretical ambiguity to the fact that a market may be away from equilibrium at any point in time, there is a definite disagreement over the superiority of one method over the other towards quantifying the exact extent of market inefficiency. One attractive approach on which there has been significant evolution over methodological and computational architecture is the modelling of time-series integration techniques (Duan et al. (2023)). It is known that many short-memory non-linear processes such as Markov switching models can appear to be long memory in finite samples (Diebold and Inoue (2001); Haldrup and Nielsen (2006)). For this reason, it is desirable to have additional arguments as to why a time series of interest might be a long memory (Duan et al. (2023)).

To identify the memory feature of the target time series, i.e., time structure of individual horse racing odds series as Duan et al. (2021) suggested, we investigate the auto-regressive property of the target time series that is governed by the characterization of the integration order (d) of the series. In particular, the potentially existing long memory can be considered when extending the strict premise of integer d values to the domain of fractional ones. In our case, a horse racing betting odds series (y_t) with an integration order d can be formulated as:

$$(1-L)^d y_t = \alpha(L)\epsilon_t = \sum_{i=0}^{\infty} \alpha(L^i)\epsilon_{t-i} \tag{3.1}$$

where $(1-L)^d$ is the difference operator with order d; $\alpha(L)$ is the coefficient of the error term at the past time period $t-i$ and $\sum_{i=0}^{\infty} \left|\alpha(L^i)\right| < \infty$ indicating that y_t is stationary after differentiating d times. We have $\epsilon_t \sim \text{iid}(0, \sigma^2)$. The memory property of y_t depends on the value of its integration order (d). In convention that d is an integer value, when $d = 0$.

$$y_t = \alpha(L)\epsilon_t = \sum_{i=0}^{\infty} \alpha(L^i)\epsilon_{t-i} \tag{3.2}$$

Where y_t is featured by a short-memory series where its past information exerts only a limited role on its current dynamics such as a stationary $ARMA$ process or even no impact such as a white noise process.

When $d = 1$, y_t is a unit root series expressed by the following $MA(\infty)$process where the past information could consistently reflect on its current dynamics and never decay i.e., typically a series with permanent memory.

$$y_t = \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + \cdots \tag{3.3}$$

Hamilton (1994), when $0 < d < 1$, re-formulate the formula $(1-L)^d y_t = \alpha(L)\epsilon_t = \sum_{i=0}^{\infty} \alpha(L^i)\epsilon_{t-i}$ as:

$$y_t = (1-L)^{-d}\alpha(L)\epsilon_t = \sum_{i=0}^{\infty} \beta_i L^i \epsilon_{t-i} \tag{3.4}$$

Where, $\beta_0 \equiv 1$; $\beta_i = \frac{(d+i-1)(d+i-2)\cdots(d+2)(d+1)d}{i!}$, and $\beta_i \equiv (i+1)^{d-1}$ given that i is large. Thus, y_t is further formulated as the following $MA(\infty)$ process:

$$y_t = \beta_0\epsilon_t + \beta_1\epsilon_{t-1} + \beta_2\epsilon_{t-2} + \cdots \tag{3.5}$$

Where impacts of past shocks of y_t on its current dynamics are highly persistent and slowly converged featured by long-memory, in contrast to short memory when $d = 0$ and permanent memory when $d = 1$. In addition, when $d > 1$, y_t possesses an explosive long-memory feature where past information exerts an increasingly large impact on its dynamics over time.

Following the efficient market hypothesis (EMH) documented by Fama (1970), a market is considered efficient only if the price series follows a random walk process, which implies that the integration order (d) of the horse racing odds series is equal to 1. Importantly, extending the conventional assumption of the d value from an integer to a fractional value provides a more comprehensive and convenient measurement of d and the corresponding degree of market inefficiency. Specifically, the closer d is to 1, the more efficient the market. The construction of the index for market inefficiency will be discussed in the section on the Degree of market (in)efficiency.

3.3.2 Estimation of the Long-Memory Property in Horse Racing Odds Series

Traditionally, the measurement of long-memory processes using the Hurst exponent—computed via the rescaled range statistic (R/S) and detrended fluctuation analysis (DFA)—has been widely employed. However, it has been recognised for some time that results obtained using these methods can be susceptible to bias, contingent upon the data generation process and parameter settings (Lo (1991); Hauser and Reschenhofer (1995); Kantelhardt et al. (2001)). Recently, a semi-parametric Exact Local Whittle (ELW) estimator, developed by Shimotsu and Phillips (2005), has been introduced to address the limitations mentioned above and to identify long-memory behaviour by estimating its fractional integration order (d). Shimotsu (2010)further introduced the Feasible Exact Local Whittle (FELW) estimator, which accounts for an unknown mean and time trend in the target series, thereby enhancing the accuracy of long-memory estimation. Despite its potential, the application of FELW in the domain of horse racing betting remains limited. In this study, we utilise the FELW estimator to quantify the fractional integration order (d) of horse racing odds series and assess its associated long memory. Notably, the FELW estimator does not require any assumptions regarding the presence of cointegration and remains consistent regardless of whether cointegration exists in the data. Moreover, it can be applied to both stationary and non-stationary cases.

3.3.3 ELW Estimator

Consider the fractional process X_t generated by the model

$$(1-L)^d(X_t - X_0) = u_t, \quad t = 0, 1, 2, \ldots \tag{3.6}$$

where X_0 is a random variable with a certain fixed distribution.

The model above is not the only model of nonstationary fractional integration. Another model that is used in the literature forms a process X_t with $d \in \left[\frac{1}{2}, \frac{3}{2}\right)$ from the partial sum of a stationary long-range dependent process, as in

$$X_t = \sum_{k=1}^{t} U_k + X_0, \quad d \in \left[\frac{1}{2}, \frac{3}{2}\right) \tag{3.7}$$

where U_t has spectral density $f(\lambda) \sim G_0 \lambda^{-2(d-1)}$ as $\lambda \to 0$. Above equation applies for the specific range of values $d \in \left[\frac{1}{2}, \frac{3}{2}\right)$, and this can be extended by repeated use of partial summation in the definition. $(1-L)^d(X_t - X_0) = u_t$ directly provides a valid model for all values of d.

Local Whittle (Gaussian semiparametric) estimation was developed by Künsch (1987) and Robinson (1995). Specifically, it starts with the following Gaussian objective function defined in terms of the parameter d and G:

$$Q_m(G,d) = \frac{1}{m}\sum_{j=1}^{m}\left[\log(G\lambda_j^{-2d}) + \frac{\lambda_j^{2d}}{G} I_x(\lambda_j)\right] \tag{3.8}$$

where m is some integer less than n. The local Whittle procedure estimates G and d by minimizing $Q_m(G,d)$ so that

$$(\hat{G},\hat{d}) = \arg\min_{G\in(0,\infty),d\in[\Delta_1,\Delta_2]} Q_m(G,d) \tag{3.9}$$

Where Δ_1 and Δ_2 are numbers that $-\frac{1}{2} < \Delta_1 < \Delta_2 < \infty$. It will be convenient in what follows to distinguish the true values of the parameters by the notation $G_0 = f_u(0)$ and d_0. Concentrating the above Gaussian objective function with respect to G as in Robinson (1995) gives,

$$\hat{d} = \arg\min_{d\in[\Delta_1,\Delta_2]} R(d) \tag{3.10}$$

Where,

$$R(d) = \log\hat{G}(d) - 2d\frac{1}{m}\sum_{j=1}^{m}\log\lambda_j, \tag{3.11}$$

$$\hat{G}(d) = \frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2d} I_x(\lambda_j). \tag{3.12}$$

Now introduce the assumptions on m and the stationary component u_t in $(1-L)^d(X_t - X_0) = u_t$:

- **Assumption 1.**
$$f_u(\lambda) \sim f_u(0) \in (0,\infty), \text{ as } \lambda \to 0^+ \tag{3.13}$$
- **Assumption 2.** In a neighborhood $(0,\delta)$ of the origin $f_u(\lambda)$ is differentiable and $\frac{d}{d\lambda}\log f_u(\lambda) = O(\lambda^{-1})$, as $\lambda \to 0^+$.
- **Assumption 3.**
$$u_t = C(L)\epsilon_t = \sum_{j=0}^{\infty} c_j\epsilon_{t-j} \tag{3.14}$$
where, $\sum_{j=0}^{\infty} c_j^2 < \infty$, $E(\epsilon_t \mid F_{t-1}) = 0$, $E(\epsilon_t^2 \mid F_{t-1}) = 1$ as $t = 0,1,\ldots$ in which F_t is the σ-field generated by ϵ_s for $s \le t$ and there exists a random variable ϵ such that $E\epsilon^2 < \infty$ and for all $\eta > 0$ and some $K > 0$, $\Pr(|\epsilon_t| > \eta) \le K\Pr(\epsilon > \eta)$.
- **Assumption 4.**
$$\frac{1}{m} + \frac{m}{n} \to 0, \text{ as } n \to \infty \tag{3.15}$$

Lemma A.1(a) in Appendix A gives the following expression for $w_x(\lambda_s)$:

$$w_x(\lambda_s) = \frac{D_n(e^{i\lambda_s};\theta)}{1-e^{i\lambda_s}} w_u(\lambda_s) - \frac{e^{i\lambda_s}}{1-e^{i\lambda_s}}\frac{X_n - X_0}{\sqrt{2\pi n}} - \frac{1}{1-e^{i\lambda_s}}\frac{\tilde{U}_{lambda_s n}(\theta)}{\sqrt{2\pi n}} \tag{3.16}$$

Neglecting the third term of the above equation as a remainder is seen to comprise two terms—a function of the d.f.t. of u_t and a function of d. As the value of d changes, the stochastic magnitude of the two components changes and this influences the asymptotic behaviour of $w_x(\lambda_s)$.

When $d < 1$, the first term dominates the second term and $w_x(\lambda_s)$ behaves like $\lambda_s^{-d} w_u(\lambda_s)$, being asymptotically uncorrelated for different frequencies. When $d > 1$, the second term becomes dominant and $w_x(\lambda_s)$ behaves like being $\lambda_s^{-1} \frac{(X_n - X_0)}{\sqrt{2\pi n}}$ perfectly correlated across all λ_s. This switching behaviour of $w_x(\lambda_s)$ at $d = 1$ is a key determinant of the asymptotic properties of the local Whittle estimator. When $d = 1$, the two terms have the same stochastic order and this leads to a form of asymptotic behaviour that is particular to this case.

Theorem 3.1 below establishes that $\hat{d}$ is consistent for $d_0 \in \left(\frac{1}{2}, 1\right]$ and hence consistency carries over to the unit root case. While $\hat{G}$ is consistent for $d_0 \in \left(\frac{1}{2}, 1\right)$, however, it is inconsistent and tends to a random quantity when $d_0 = 1$.

Shimotsu et al. (2005) proposed the Exact Local Whittle (ELW) estimator. The asymptotic properties of the local Whittle estimator in the nonstationary case over the region $d \in \left(\frac{1}{2}, 1\right)$ were explored in Velasco (1999). Velasco also showed that upon adequate tapering of the observations, the region of consistent estimation of d may be extended but with corresponding increases in the variance of the limit distribution. Shimotsu et al. (2005) demonstrate that the local Whittle estimator is consistent for $d \in \left(\frac{1}{2}, 1\right]$, and is asymptotically normally distributed for $d \in (\frac{1}{2}, \frac{3}{4})$, and has a nonnormal limit distribution for $d \in (\frac{3}{4}, 1]$, and has a mixed normal limit distribution for $d = 1$, converges to unity in probability for $d > 1$, and converges to unity in probability when the process has a polynomial time trend of order $\alpha > \frac{1}{2}$.

3.3.4 FELW Estimator

Briefly, the FELW estimator is built based on the ELW estimator through which the d value of the target series (y_t) can be obtained by minimizing the following objective function.

$$Q_m(G, d) = \frac{1}{m} \sum_{i=1}^{m} \left[\log(G\lambda_j^{-2d}) + 1/GI_{(1-L)^d Y_t}(\lambda_j) \right] \tag{3.17}$$

Concentrating $Q_m(G, d)$ with regard to G, the ELW estimator is built as

$$\hat{d} = \arg \min_{d \in [\Delta_1, \Delta_2]} R(d) \tag{3.18}$$

Where m is the model truncation parameter, Δ_1 and Δ_2 define the lower and upper bounds of the admissible values of d subjecting to $-\infty < \Delta_1 < \Delta_2 < +\infty$ and

$$R(d) = \log \hat{G}(d) - 2d \frac{1}{m} \sum_{i=1}^{m} \log(\lambda_i), \tag{3.19}$$

$$\hat{G}(d) = \frac{1}{m} \sum_{i=1}^{m} I_{\Delta^d_{Yt}} \log(\lambda_i). \tag{3.20}$$

In specific conditions, notably including $\Delta_1 - \Delta_2 \leq \frac{9}{2}$, the ELW estimator is known to be consistent and asymptotically normally distributed such that $\sqrt{m}(\hat{d} - d_0) \xrightarrow{d} N\left(0, \frac{1}{4}\right)$, $n \to \infty$, where n is the sample size and d_0 is the true value of d, $d_0 \in (\Delta_1, \Delta_2)$.

The ELW estimator can be further extended to the FELW estimator which estimates the unknown mean value μ of y_t through

$$\hat{\mu}(d) = \overline{Y} \cdot 1\{d < c\} + Y_t \cdot 1\{d \leq c\} \tag{3.21}$$

where $\overline{y}$ is the sample average, $1\{\cdot\}$ is the indicator function, and $c \in \left(\frac{1}{2}, \frac{5}{8}\right)$. Thus, the FELW estimator with regard to d of y_t can be constructed through the ELW d estimation for $y_t - \hat{\mu}(d)$.

3.3.5 Degree of Market Inefficiency

Duan et al. (2023) suggests building an index D calculated by the absolute difference between d and 1, to conveniently observe the degree of the market (in)efficiency, i.e., the closeness between the integration order (d) of the horse racing odds series (y_t) and 1. That is, the greater the D, the larger the distance between d and 1, the less efficient the market will become, and vice versa. In other words, D demonstrates the level of market inefficiency degree.

$$D_t = |1 - d_t| \tag{3.22}$$

where d_t is the estimated d value at time t using the ELW or FELW estimator on a rolling-basis.

3.4 Data

3.4.1 The Using of DT

When we do betting, we are more interested in how the odds move. As we show in Chapter 1, the odds series can be decomposed into two components, the Dynamic Trend Component and the Volatility Component. The Dynamic Trend Component captures the mode of data changing in a time span, reflecting the dynamic process that the market captures and understanding information in a temporal manner. The market efficiency is about how efficiently the whole market captures and understands information. The Dynamic Trend(DT) is a good way for us to understand how the market works as it gives information in which direction and how fast the data changes over time, this is the main drive of the data changing behaviour. Moreover, when we try to predict the the time series, we are looking for the direction and magnitude that the data going in the future. The DT velocity of data changing over time, is the main direction and the magnitude of data changing over time, which means if we are able to predict DT, we are predicting the data in the future on average. If we predict the corresponding Volatility component and put the predicted DT together with the corresponding Volatility, we have a prediction of the price. In this Chapter, we aim at how the market efficiency changes over time, we just use DT to understand the dynamic efficiency of the market.

3.4.2 Generation of 1-minute Time Level DT

The data we use is the 15-minute Dynamic Trend Horizon 30-second Fundamental Time Level Dynamic Trend Series($15minDTH30sFundTL\ DT$) for the proxy price series for all 5 markets we generated from Chapter 1. Due to objective technical reasons, the original odds series data collection is not strictly at 30-second intervals. To facilitate subsequent data processing, we will convert the$15minDTH30sFundTL\ DT$ to a 15-minute Dynamic Trend Horizon 1-minute Generated Time Level Dynamic Trend Series ($15minDTHGen1minTL30sFundTL\ DT$) using the higher Time Level DT generation method described in Chapter1. This will make our data a perfect time series with a uniform time interval.

The $15minDTHGen1minTL30sFundTL\ DT$ getting from 30-second Fundamental Time Level Dynamic Trend Series ($15minDTH30sFundTL\ DT$) is given by:

$$\begin{aligned} &15minDTHGen1minTL30sFundTL\ DT_j = \\ &\sum_{i=1}^{n}(15minDTH30sFundTL\ DT_{ji}) \end{aligned} \tag{3.23}$$

Where n is the number of $15minDTH30sFundTL\ DT$ data in the corresponding j -th minute. For terse writing, we write $15minDTHGen1minTL30sFundTL\ DT$ as $1minDT$ in this Chapter. The descriptive statistics for $1minDT$ for all five markets are presented in the below tables.

MARKET	OBSERVATIONS	MEAN	MIN	MAX	MEDIAN	MODE
M1	720	-0.01	-1.03	0.99	0.0	0.0
M2	720	0.00	-1.25	1.35	0.0	0.0
M3	720	0.00	-1.03	1.37	0.0	0.0
M4	720	-0.02	-1.61	0.99	0.0	0.0
M5	720	0.00	-0.43	0.50	0.0	0.0

TABLE 3.1: Descriptive Statistics for 1-min Dynamic Trend Series - Part 1

MARKET	STD	VAR	KURTOSIS	SKEWNESS	P25	P50	P75
M1	0.31	0.10	0.41	-0.07	-0.20	0.0	0.18
M2	0.33	0.11	3.10	-0.34	-0.12	0.0	0.14
M3	0.32	0.10	1.57	0.15	-0.17	0.0	0.16
M4	0.37	0.14	4.03	-1.06	-0.14	0.0	0.14
M5	0.14	0.02	1.12	0.19	-0.06	0.0	0.06

TABLE 3.2: Descriptive Statistics for 1-min Dynamic Trend Series - Part 2

M1 $1minDT$ displays a mean value of -0.01, very close to 0. The values range from -1.03 to 0.99, showing moderate variation and volatility, as indicated by a standard deviation of 0.31. Both the median and mode are 0, suggesting a central tendency around 0, and the distribution is nearly normal with a kurtosis of 0.41 and a skewness of -0.07.

M2 $1minDT$ has a mean of 0, but it shows a wider range of values from -1.25 to 1.35. The standard deviation of 0.33 indicates higher volatility compared to M1. The median and mode are 0, while the distribution has a high kurtosis of 3.10, indicating the presence of extreme values more frequently. The negative skewness of -0.34 suggests more frequent negative deviations.

M3 $1minDT$ a mean of 0, with a range from -1.03 to 1.37, and a standard deviation of 0.32. The median and the mode are 0, showing a central tendency near 0. The kurtosis of 1.57 and skewness of 0.15 indicate a slightly right-skewed distribution with mild tail heaviness. T

M4 $1minDT$ exhibits a mean of -0.02, which is close to 0, and the widest range of values from -1.61 to 0.99. It has the highest volatility among the markets, with a standard deviation of 0.37. The median and mode are 0, but the distribution shows very heavy tails with a kurtosis of 4.03 and a high negative skewness of -1.06. This indicates frequent large negative movements.

M5 $1minDT$ shows a mean of 0 and a narrow range of values from -0.43 to 0.50. It has the lowest volatility, with a standard deviation of 0.14. Both the median and mode are 0, and the distribution has a kurtosis of 1.12 and a skewness of 0.19, suggesting mild tail heaviness and slight right skewness.

$1minDT$ for M1 to M5 are plotted below, time period is around 12 hours.

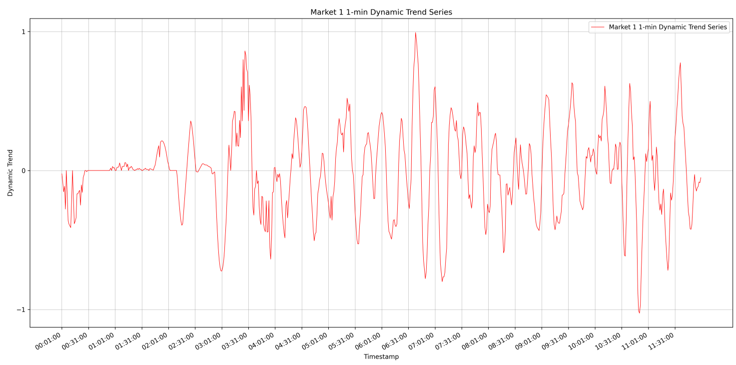

FIGURE 3.1: M1 Market Proxy Price 1-min Dynamic Trend

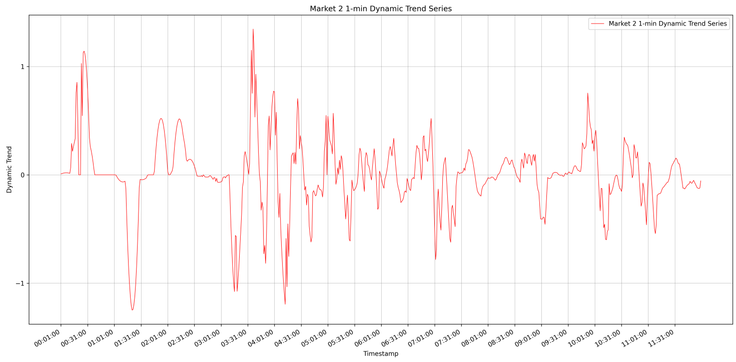

FIGURE 3.2: M2 Market Proxy Price 1-min Dynamic Trend

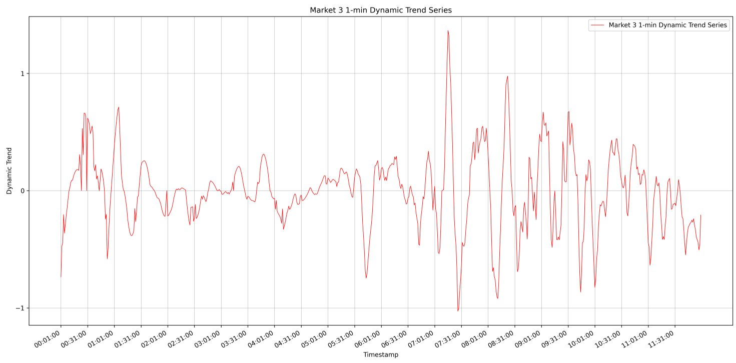

FIGURE 3.3: M3 Market Proxy Price 1-min Dynamic Trend

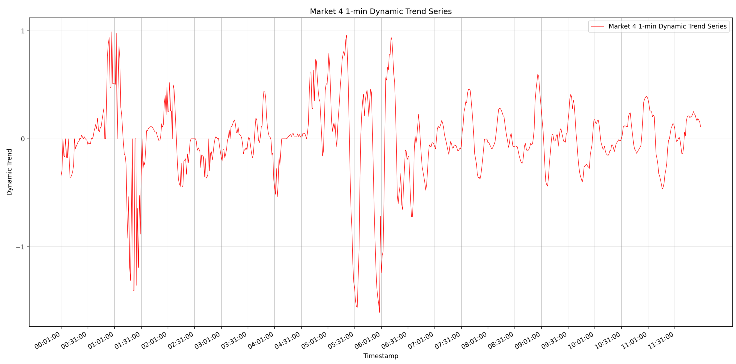

FIGURE 3.4: M4 Market Proxy Price 1-min Dynamic Trend

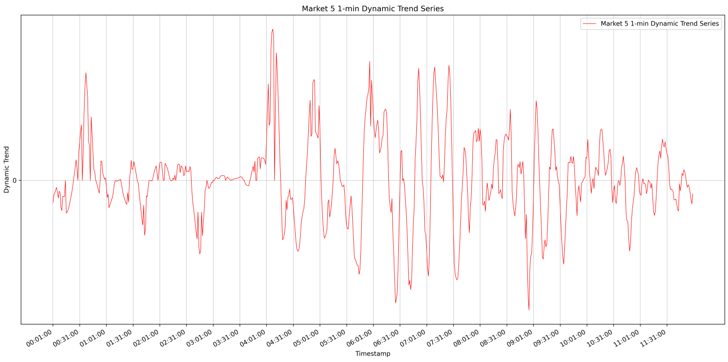

FIGURE 3.5: M5 Market Proxy Price 1-min Dynamic Trend

3.4.3 Stationarity of 1-minute Time Level DT

We apply the Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test to check if the $1minDT$ for M1 to M5 are stationary. The KPSS test's null hypothesis and the alternative hypothesis are:

Null Hypothesis (Stationarity): The null hypothesis for the KPSS test is that the time series is stationary.

Alternative Hypothesis (Non-Stationarity): The alternative hypothesis suggests that the time series has a unit root (non-stationarity).

The KPSS test statistics for the $1minDT$ for M1 to M5 are reported below:

Market	Statistic	p-value
M1	0.029	0.100
M2	0.063	0.100
M3	0.111	0.100
M4	0.042	0.100
M5	0.041	0.100

TABLE 3.3: KPSS Statistics for 1-min Dynamic Trend - Part 1

Market	Critical Value(10%)	Critical Value(5%)	Critical Value(2.5%)	Critical Value(1%)
M1	0.347	0.463	0.574	0.739
M2	0.347	0.463	0.574	0.739
M3	0.347	0.463	0.574	0.739
M4	0.347	0.463	0.574	0.739
M5	0.347	0.463	0.574	0.739

TABLE 3.4: KPSS Statistics for 1-min Dynamic Trend - Part 2

The KPSS test statistics for the $1minDT$ for M1 to M5 show that the time series data are stationary. The KPSS statistics for M1, M2, M3, M4, and M5 are 0.029, 0.063, 0.111, 0.042, and 0.041, respectively. These values are much lower than the critical values for the 10%, 5%, 2.5%, and 1% significance levels, which are 0.347, 0.463, 0.574, and 0.739, respectively. The low KPSS statistics, relative to critical values, clearly suggest that we cannot reject the null hypothesis of stationarity for any of the $1minDT$. Furthermore, the p-values for all $1minDT$ are 0.1, which exceeds the typical significance levels of 0.05 or 0.01, reinforcing the conclusion that there is insufficient evidence to reject the null hypothesis. The $1minDT$ for M1 to M5 are likely stationary, implying that their statistical properties, such as mean and variance, remain constant over time.

3.5 Dynamic Market (In)Efficiency For Betting Market

3.5.1 The Optimal d Estimation

To estimate the integration orders (d), we need to select the most accurate method from LW, ELW, FELW, and Two-Step FELW. The estimated d is sensitive to 3 parameters, the window size of estimation, the bandwidth of the estimation, and the estimator of the estimation. To make sure the dynamic d series for all 5 markets' $1minDT$ are at the same length, we need the window size of estimation for all 5 markets d to be the same.

To find the optimal combination of the window size, the bandwidth, and the estimator of the estimation, we build a selection method called the Estimation Score for Integration Orders ($ESIO$). The $ESIO$ is designed to find the combination of the window size(WS), the bandwidth(BD), and the estimator of the estimation that gives the d most close to the line of 1 among all tested combinations. It is calculated as below:

$$ESIO = \frac{1}{m}\sum_{i=1}^{m}(d_i - 1)^2 \tag{3.24}$$

Where m is the data size of the d series. The optimal combination of WS, BD, and estimator is given by the minimum $ESIO$. We test the WS value of 5, 6, 8, 9, 10, 15, 16, 18, 20, 24, 30, BD value of 0.5, 0.6. 0.7, 0.8 for LW, ELW, FELW, and Two-Step FELW estimator. The $ESIO$ gives the optimal combination of WS, BD , and estimator for each market as follow:

Market	Window Size	Bandwidth	Observations
M1	8	0.70	713
M2	8	0.70	713
M3	8	0.70	713
M4	8	0.60	713
M5	8	0.60	713

TABLE 3.5: Optimal Combination of Window Size, Bandwidth, and Estimator- Part 1

Market	ESIO LW	ESIO ELW	ESIO FELW	ESIO TWO_Stage_FELW	Optimal Estimator
M1	0.75	2170.59	3795.31	NaN	LW
M2	0.20	2230.75	3876.64	NaN	LW
M3	0.44	1789.15	3847.46	1556.71	LW
M4	0.01	479.93	1297.02	1145.74	LW
M5	0.59	621.53	1335.99	1234.11	LW

TABLE 3.6: Optimal Combination of Window Size, Bandwidth, and Estimator- Part 2

For M1, the optimal combination of WS, BD , and estimator is $WS = 8$, $BD = 0.7$, LW. For M2, the optimal combination of WS, BD , and estimator is $WS = 8$, $BD = 0.7$, LW. For M3, the optimal combination of WS, BD , and estimator is $WS = 8$, $BD = 0.7$, LW. For M4, the optimal combination of WS, BD , and estimator is $WS = 8$, $BD = 0.6$, LW. For M5, the optimal combination of WS, BD , and estimator is $WS = 8$, $BD = 0.6$, LW.

The optimal estimated d for M1 to M5 are plotted as below:

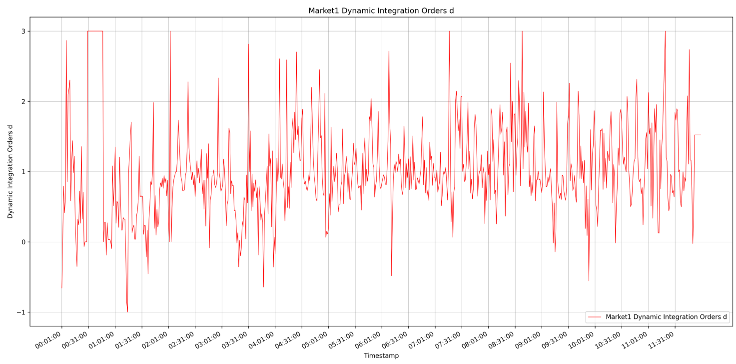

FIGURE 3.6: M1 Dynamic Integration Orders d

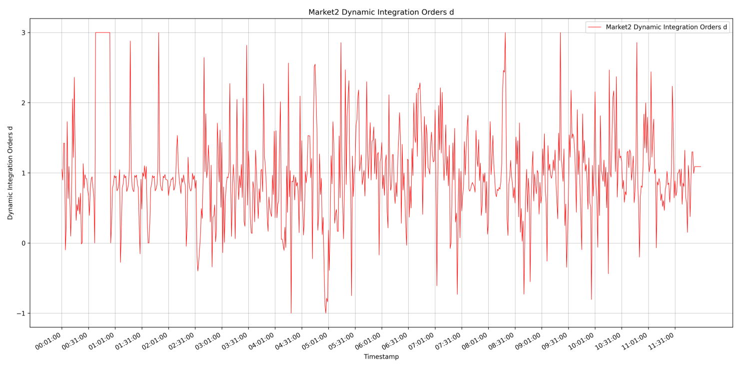

FIGURE 3.7: M2 Dynamic Integration Orders d

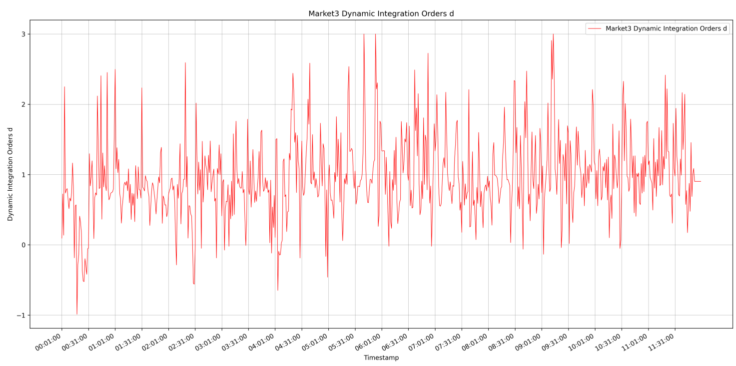

FIGURE 3.8: M3 Dynamic Integration Orders d

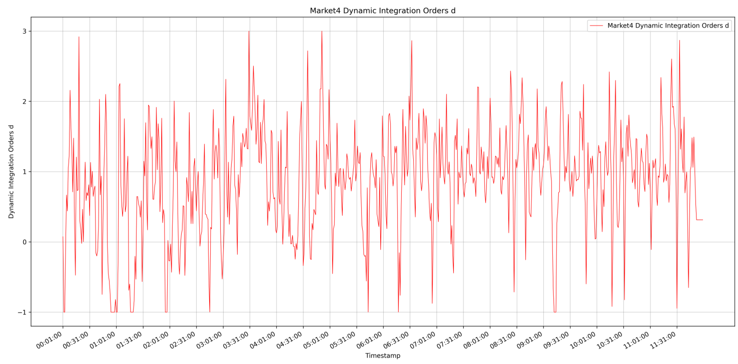

FIGURE 3.9: M4 Dynamic Integration Orders d

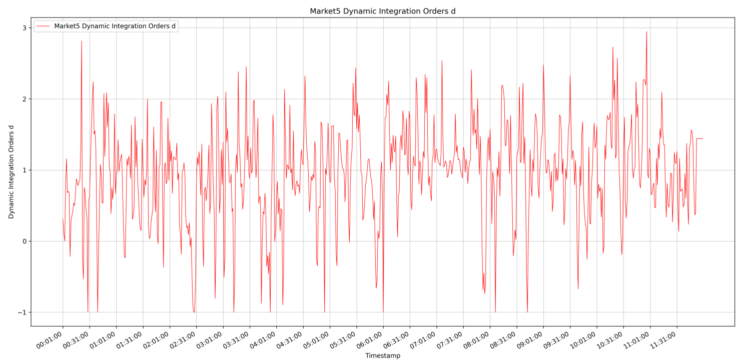

FIGURE 3.10: M5 Dynamic Integration Orders d

3.5.2 The Dynamic Degree of Market Inefficiency

Now we build the degree of market (in)efficiency D calculated by the absolute difference between d and 1. The greater the D, the larger the distance between d and 1, the less efficient the market will become, and vice versa. D is getting by equation:

$$D_t = |1 - d_t| \tag{3.25}$$

where d_t is the estimated d value at time t.

The dynamic D for all 5 markets are plotted as below:

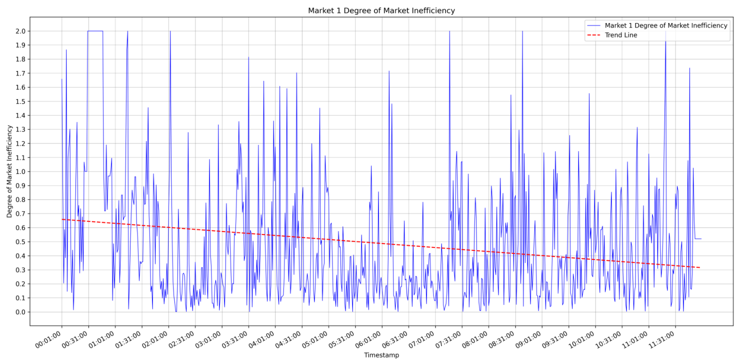

FIGURE 3.11: M1 Dynamic Degree of Market Inefficiency D

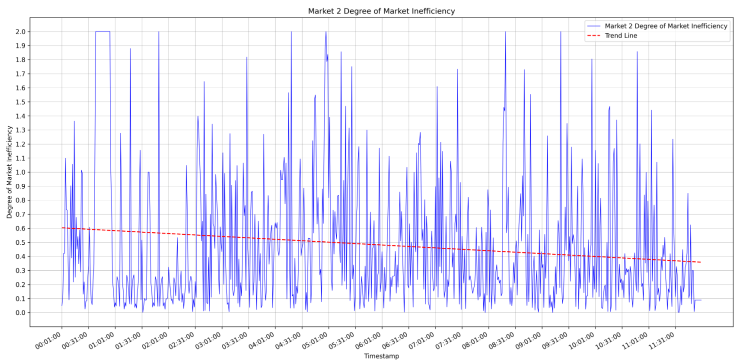

FIGURE 3.12: M2 Dynamic Degree of Market Inefficiency D

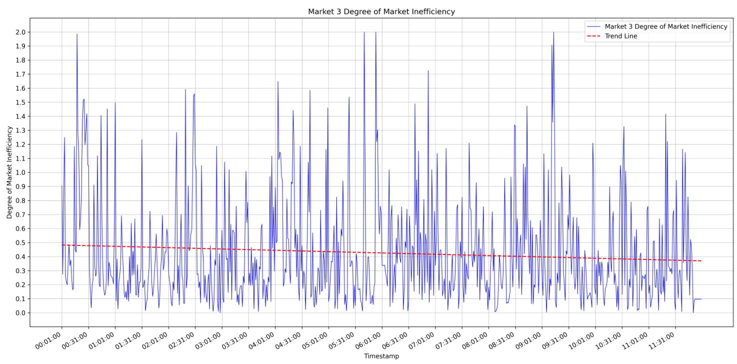

FIGURE 3.13: M3 Dynamic Degree of Market Inefficiency D

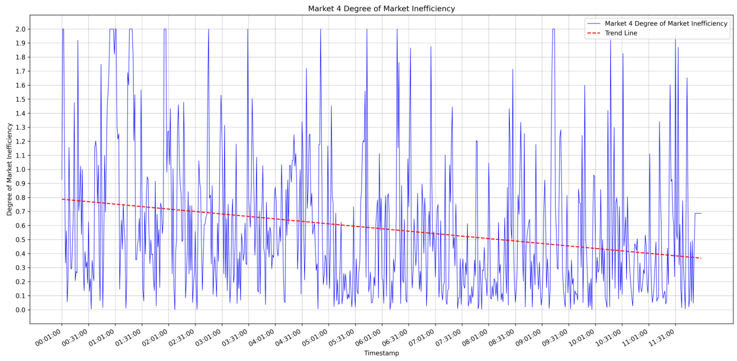

FIGURE 3.14: M4 Dynamic Degree of Market Inefficiency D

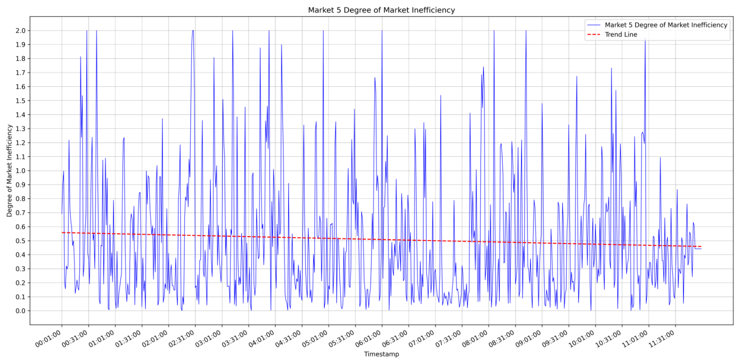

FIGURE 3.15: M5 Dynamic Degree of Market Inefficiency D

The D of all 5 markets are showing some level of decrease during the 12 hours of betting. This indicates that the betting market is getting more and more efficient with time going.

3.5.3 Dynamic Convergence of Inefficiency Degree

The D of all 5 markets give us an understanding of the dynamics of the market efficiency of each market. To better understand the flat horse racing betting market, we will assess it in a cross market way. The method of Duan et al. (2023) is applied to decompose the data and analyze the cross market efficiency dynamics. To understand the convergence pattern of the inefficiency degree across markets, we decompose $\sum_{i=1}^{n}(d_{it}-1)^2$ as:

$$\begin{aligned}\sum_{i=1}^{n}(d_{it}-1)^2 &= \sum_{i=1}^{n}(d_{it}-\overline{d_t}+\overline{d_t}-1)^2 \\ &= \sum_{i=1}^{n}(d_{it}-\overline{d_t})^2 + n(\overline{d_t}-1)^2 + 2\sum_{i=1}^{n}(d_{it}-\overline{d_t})(\overline{d_t}-1)\end{aligned} \tag{3.26}$$

where d_{it} is the fractional integration order of market i at time t, n is the amount of markets. $\overline{d_t}$ is the cross-market mean of the d_{it} . $\overline{d_t}$ serve as the benchmark of the memory parameter of all n markets. As $D_t = |1 - d_t|$, for the left side of equation, we have

$$(d_{it} - 1)^2 = (|1 - d_{it}|)^2 = D_{it}^2 \tag{3.27}$$

where D_{it} is the inefficiency degree of the flat horse racing betting market i at time t. It indicates that the convergence pattern of D_{it} depends on fractional integration order d_{it} and variations of d_{it} around 1, note that $d_{it} = 1$ is the status of an efficient market. The third term on the right side of the equation

$$2\sum_{i=1}^{n}(d_{it} - \overline{d_t})(\overline{d_t} - 1) = 2(\overline{d_t} - 1)\sum_{i=1}^{n}(d_{it} - \overline{d_t}) \tag{3.28}$$

where the $\sum_{i=1}^{n}(d_{it} - \overline{d_t}) = 0$, so we have

$$2\sum_{i=1}^{n}(d_{it} - \overline{d_t})(\overline{d_t} - 1) = 2(\overline{d_t} - 1)\sum_{i=1}^{n}(d_{it} - \overline{d_t}) = 0 \tag{3.29}$$

so, we have

$$\begin{aligned}\sum_{i=1}^{n}(d_{it} - 1)^2 &= \sum_{i=1}^{n}(d_{it} - \overline{d_t} + \overline{d_t} - 1)^2 \\ &= \sum_{i=1}^{n}(d_{it} - \overline{d_t})^2 + n(\overline{d_t} - 1)^2\end{aligned} \tag{3.30}$$

The $\sum_{i=1}^{n}(d_{it} - \overline{d_t})^2$ term, known as deviation pattern, gives the cross-market sum of variations around a convergent benchmark $\overline{d_t}$ of the memory parameter of all n markets. We were able to see changes in the memory parameter around the convergence through the deviation pattern component. If the deviation pattern term is smaller, the variations in the memory parameter (d_{it}) around the convergent benchmark are also smaller, leading to more consistent inefficiency degrees (D_{it}) across all markets.

The $n(\overline{d_t} - 1)^2$ term, known as inefficiency from convergence, provides a measurement of the difference between the convergent benchmark and 1. We can explicitly measure how closely the convergent benchmark of the memory parameter ($\overline{d_t}$) approaches the state of an efficient market, which is achieved when d equals 1. The smaller the inefficiency from the convergence term, the closer the convergent benchmark ($\overline{d_t}$) is to the efficient market state, resulting in a more efficient market and, consequently, a lower market inefficiency index (D_{it}).

The deviation pattern and inefficiency from convergence across 5 betting markets are plotted below.

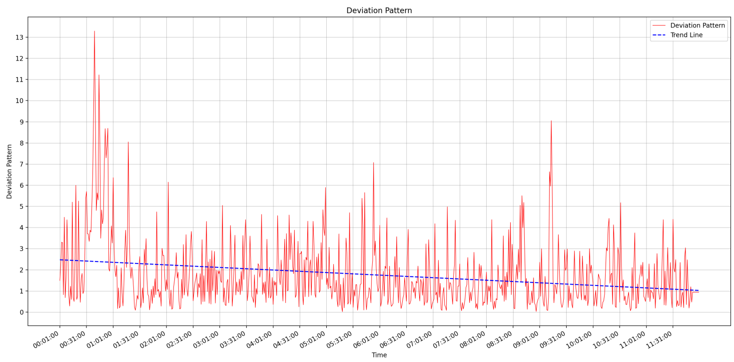

FIGURE 3.16: Dynamic Deviation Pattern

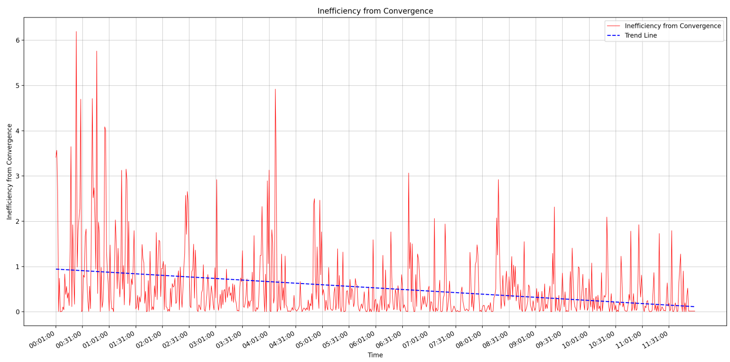

FIGURE 3.17: Dynamic Inefficiency from Convergence

Both deviation pattern and inefficiency from convergence are showing an obvious downward trend, indicating that market efficiency gradually increases over time.

At the beginning of betting, some of the relevant information might not have been captured by the market (or has not yet emerged), and there are disagreements among market participants regarding the available information. For information that has reached the market, when it is first captured, there will inevitably be some astute traders who recognise its value and fully price it. Then, as time passes, this information spreads more widely and is recognised by more market participants, who incorporate it into their betting behaviour and participate in pricing. This will lead to a rise in the market efficiency. Specifically, the process of information reaching and spreading in the market can be divided into two parts: direct dissemination and indirect acquisition. Direct dissemination refers to some traders directly receiving the information, analysing it, and participating in market activities through their betting behaviour. Indirect acquisition refers to other market participants who do not directly obtain and analyze the information; instead, they observe price changes, volume, and other data to indirectly acquire information and make decisions, participating in the market through their betting behaviour.

This process requires a certain amount of time due to physical limitations and the cognitive abilities of all individuals in the market. As the market gradually reaches a consensus on a specific piece of information over time, market disagreements decrease. Therefore, the market is dynamically transitioning towards an efficient market equilibrium. When no new information is available (i.e., the information the market can assess is limited), the market will continuously approach the efficient frontier over time.

In reality, betting has a time limit from start to finish. Within this limited time, the market dynamically acquires information, and the total amount of information is limited due to the time constraints of betting. Thus, as time progresses from the start of betting, the information obtained in each period is gradually recognised by the market. In the absence of extreme events, the total amount of unknown and nonconsensual information in the market gradually decreases, leading to an increase in market efficiency. Consequently, the deviation pattern and inefficiency from convergence show a downward trend.

The process of market perception of information, its limitations, and its mathematical model will be discussed in detail in the next chapter.

3.6 FACVAR Predictions

3.6.1 Unrestricted FCVAR Model

We are interested in modelling the five markets as a dynamically integrated system for the following reasons. First, the betters in one market can influence the price and the volume in other markets because investors are dynamic in nature and they wish to maximise gains by diversifying their investment strategy across markets. That way, they can minimise the risks that may come from a single source. Variations in market conditions and specific information related to certain markets can be another advantage that can work for an investor prompting them to adopt a cross-market investment strategy. Therefore, the price and volume in one market at a point in time can be assumed to be a function of the same across markets. Although, one can presume a weak depth of interdependence given the nature of the betting market, yet, in a highly digitalised world investors work these days, dynamic interdependence is a logical possibility.

Now we apply the Fractionally Cointegrated Vector Autoregressive (FCVAR) model to DT series of five markets proxy prices ($DT1$ to $DT5$) to assess their long-run relationships and the predictions. FCVAR is an extension of the traditional Cointegrated VAR (CVAR) model that allows for fractional integration, introduced by Dolatabadi et al. (2016). This approach is particularly useful for capturing long memory behaviour in time series data. Based on lag selection and cointegration rank tests, we estimate a model with a lag length of 5 and a cointegration rank of 5. The analysis was conducted using the MATLAB code developed by Nielsen and Popiel (2016).

The Fractionally Cointegrated Vector Autoregressive (FCVAR) model extends the traditional CVAR model by allowing fractional differencing, which accounts for long memory in the time series. This model is particularly useful for handling systems where deviations from equilibrium are persistent but mean-reverting. The FCVAR model introduces two additional parameters compared to the traditional CVAR model, specifically the fractional parameters d and b. The parameter d represents the fractional integration order of the observable time series. Based on standard no-arbitrage arguments, financial assets are generally expected to exhibit $d = 1$. Consequently, in this empirical analysis, we assume that $d = 1$, treating it as fixed and known rather than estimating it, as is commonly done in similar studies. In contrast, the parameter b governs the degree of fractional cointegration, reflecting the extent to which the integration order of βX_t is reduced compared to X_t. Unlike d, we do not impose any restrictions on b; instead, it is estimated jointly with the other parameters of the model. Thus, the fractional integration order is given by $d - b = 1 - b$. In this case, the FCVAR model introduces two key parameters: the fractional differencing parameter d (which is fixed at 1 in this setup) and the fractional cointegration parameter $\hat{b}$, estimated as 0.214 from the unrestricted model results. We have the fractional integration order is given by $d - \hat{b} = 1 - \hat{b} = 0.786$.

The general form of the unrestricted FCVAR model for a p-dimensional time series X_t, with $d = 1$, is:

$$\Delta X_t = \alpha\beta' L_{\hat{b}} \Delta^{d-\hat{b}} X_t + \sum_{i=1}^{k} \Gamma_i \Delta L_{\hat{b}}^{i} X_{t-i} + \hat{\epsilon}_t \tag{3.31}$$

Where:

- X_t is the p-dimensional vector of endogenous variables, here $X_t = [\text{DT1}, \text{DT2}, \text{DT3}, \text{DT4}, \text{DT5}]'$.
- Δ^d is the differencing operator.
- $L_{\hat{b}} = 1 - \Delta^{\hat{b}}$, where $\hat{b} = 0.214$, is the fractional lag operator.
- α is the $p \times r$ adjustment matrix, describing how quickly each variable reacts to deviations from long-run equilibrium.
- β' is an $r \times p$ matrix of cointegrating vectors, representing the long-run equilibrium relationships.
- Γ_i are $p \times p$ matrices capturing the short-run dynamics over $k = 5$ lags.
- ϵ_t is the vector of residuals, assumed to be i.i.d. with mean zero and covariance matrix Ω.

3.6.1.1 Cointegrating Relationships

The long-run equilibrium relationships between the variables are defined by the cointegrating vectors β_1' and β_2'.

$$\beta_1' = \begin{bmatrix} 1 & 0 & -0.226 & 0.209 & 1.045 \end{bmatrix}$$

$$\beta_2' = \begin{bmatrix} 0 & 1 & -0.683 & 0.230 & 5.719 \end{bmatrix}$$

3.6.1.2 Adjustment Coefficients (α)

The adjustment coefficients α represent how quickly each variable reacts to deviations from the long-run equilibrium defined by the cointegrating relationships. The estimated adjustment coefficients from the unrestricted FCVAR model are:

$$\alpha_1 = [-2.398, 0.397, -1.614, -0.392, -0.569]$$
$$\alpha_2 = [0.520, -0.485, 0.436, 0.197, -0.391]$$

These values describe the speed at which each variable adjusts to restore equilibrium in the system, with larger absolute values indicating a faster adjustment.

3.6.1.3 Short-Run Dynamics (Lag Matrices Γ_i)

The short-run dynamics in the FCVAR model are captured by the lag matrices Γ_1, Γ_2, and Γ_3, which describe the effects of lagged differences on the current values of the endogenous variables. The unrestricted model estimation provides the following matrices:

$$\Gamma_1 = \begin{bmatrix} 7.104 & -0.816 & -0.207 & 0.367 & -0.749 \\ -0.399 & 4.027 & -0.245 & -0.075 & 2.601 \\ 1.527 & -0.223 & 4.500 & 0.174 & -0.285 \\ 0.332 & -0.042 & 0.202 & 2.966 & -0.756 \\ 0.466 & 0.413 & -0.433 & 0.178 & 7.239 \end{bmatrix}$$

$$\Gamma_2 = \begin{bmatrix} -8.308 & 1.177 & -0.005 & 0.189 & 1.519 \\ -0.623 & -4.972 & 0.072 & 0.709 & 0.640 \\ 1.057 & -1.001 & -9.124 & 0.604 & -1.954 \\ 0.611 & -0.563 & -0.782 & -2.864 & -2.608 \\ 0.606 & 0.012 & -0.083 & 0.319 & -7.290 \end{bmatrix}$$

$$\Gamma_3 = \begin{bmatrix} 11.553 & -3.512 & -0.674 & 1.152 & -3.727 \\ 0.308 & 2.384 & -0.758 & -0.986 & 6.461 \\ 3.605 & 0.062 & 6.004 & 0.034 & -0.192 \\ -0.063 & 0.610 & 1.853 & -2.294 & 3.990 \\ 0.999 & 1.358 & -0.879 & 0.255 & 12.298 \end{bmatrix}$$

These matrices account for the short-term effects of previous period differences in X_t on its current values, incorporating the persistence of deviations from equilibrium.

3.6.1.4 Residual Analysis and White Noise Test Results

In the context of the FCVAR model, the residuals ϵ_t represent the difference between the actual and predicted values of the variables. Ideally, the residuals should behave as white noise, meaning they have a mean of zero, constant variance, and no serial correlation. To assess these properties, the Ljung-Box Q test and the Lagrange Multiplier (LM) test for serial correlation are applied.

Variable	Q-statistic	Q P-value	LM-statistic	LM P-value
Multivar	41.953	0.784	—	—
DT1	6.779	0.034	1.345	0.510
DT2	1.741	0.419	0.363	0.834
DT3	2.552	0.279	0.821	0.663
DT4	9.465	0.009	0.659	0.719
DT5	1.218	0.544	0.553	0.758

TABLE 3.7: White Noise Test Results (lag = 2)

The multivariate test evaluates whether the residuals as a system exhibit serial correlation. The results show a Q-statistic of 41.953 and a P-value of 0.784, indicating that there is no significant serial correlation in the system as a whole. This suggests that, while there might be some individual issues, the model is generally adequate in capturing the overall dynamics of the variables.

For each individual variable, we interpret the Ljung-Box Q test and the LM test results as follows:

- $DT1_t$: The Ljung-Box Q test results in a Q-statistic of 6.779 and a P-value of 0.034, indicating significant serial correlation in the residuals for DT1_t. The LM test yields a statistic of 1.345 and a P-value of 0.510, suggesting no evidence of serial correlation from the LM test.
- $DT2_t$: The Q test results in a Q-statistic of 1.741 and a P-value of 0.419, suggesting no significant serial correlation in the residuals for DT2_t. The LM test shows a statistic of 0.363 and a P-value of 0.834, confirming no significant serial correlation.
- $DT3_t$: The Q test produces a Q-statistic of 2.552 and a P-value of 0.279, indicating no significant serial correlation in the residuals for DT3_t. The LM test yields a statistic of 0.821 and a P-value of 0.663, showing no evidence of serial correlation.
- $DT4_t$: The Q test shows a Q-statistic of 9.465 and a P-value of 0.009, indicating significant serial correlation in the residuals for DT4_t. The LM test results in a statistic of 0.659 and a P-value of 0.719, showing no significant serial correlation from the LM test.
- $DT5_t$: The residuals for DT5_t show no significant serial correlation, as indicated by the Q-statistic of 1.218 and a P-value of 0.544. The LM test also shows no evidence of serial correlation, with a statistic of 0.553 and a P-value of 0.758.

3.6.1.5 Unrestricted FCVAR Model Equations

The unrestricted FCVAR model equations for each variable can now be written explicitly, incorporating the fractional differencing operator $\Delta^{1-\hat{b}}$ and the fractional lag operator $L_{\hat{b}}$. Where $\hat{b} = 0.214$, fractional integration order is given by $d - \hat{b} = 1 - \hat{b} = 0.786$, X_t represents the vector $[\text{DT1}, \text{DT2}, \text{DT3}, \text{DT4}, \text{DT5}]'$, and the unrestricted model for DT1 to DT5 is as follows:

$$\begin{aligned}\Delta \text{DT1}_t &= -2.398 \times \beta_1' L_{\hat{b}} \Delta^{d-\hat{b}} X_t + 0.520 \times \beta_2' L_{\hat{b}} \Delta^{d-\hat{b}} X_t \\ &\quad + \sum_{i=1}^{k} \Gamma_{1,i} \Delta L_{\hat{b}}^i X_{t-i} + \hat{\epsilon_{1t}}\end{aligned} \tag{3.32}$$

$$\begin{aligned}\Delta \text{DT2}_t &= 0.397 \times \beta_1' L_{\hat{b}} \Delta^{d-\hat{b}} X_t - 0.485 \times \beta_2' L_{\hat{b}} \Delta^{d-\hat{b}} X_t \\ &\quad + \sum_{i=1}^{k} \Gamma_{2,i} \Delta L_{\hat{b}}^i X_{t-i} + \hat{\epsilon_{2t}}\end{aligned} \tag{3.33}$$

$$\begin{aligned}\Delta \text{DT3}_t &= -1.614 \times \beta_1' L_{\hat{b}} \Delta^{d-\hat{b}} X_t + 0.436 \times \beta_2' L_{\hat{b}} \Delta^{d-\hat{b}} X_t \\ &\quad + \sum_{i=1}^{k} \Gamma_{3,i} \Delta L_{\hat{b}}^i X_{t-i} + \hat{\epsilon_{3t}}\end{aligned} \tag{3.34}$$

$$\begin{aligned}\Delta \text{DT4}_t &= -0.392 \times \beta_1' L_{\hat{b}} \Delta^{d-\hat{b}} X_t + 0.197 \times \beta_2' L_{\hat{b}} \Delta^{d-\hat{b}} X_t \\ &\quad + \sum_{i=1}^{k} \Gamma_{4,i} \Delta L_{\hat{b}}^i X_{t-i} + \hat{\epsilon_{4t}}\end{aligned} \tag{3.35}$$

$$\begin{aligned}\Delta \text{DT5}_t &= -0.569 \times \beta_1' L_{\hat{b}} \Delta^{d-\hat{b}} X_t - 0.391 \times \beta_2' L_{\hat{b}} \Delta^{d-\hat{b}} X_t \\ &\quad + \sum_{i=1}^{k} \Gamma_{5,i} \Delta L_{\hat{b}}^i X_{t-i} + \hat{\epsilon_{5t}}\end{aligned} \tag{3.36}$$

The unrestricted FCVAR model integrates both long-run equilibrium relationships and short-run dynamics between the variables DT1 to DT5. The fractional differencing operator $\Delta^{d-\hat{b}}$, with $\hat{b} = 0.214$, fractional integration order is given by $d - \hat{b} = 1 - \hat{b} = 0.786$, captures long memory effects, allowing the model to represent persistent deviations from equilibrium that are mean-reverting. While the model successfully estimates the parameters, the Ljung-Box Q test results suggest that there may be some residual serial correlation, particularly for DT1, DT3, and DT4, indicating the need for further refinement of the model to improve fit.

3.6.1.6 Likelihood Ratio Test for the Restricted that d=b=1

In this section, we assess the results of the likelihood ratio (LR) test, which compares the fit of the unrestricted and restricted models. The LR test is used to evaluate whether the restrictions imposed on the restricted model lead to a significantly worse fit compared to the unrestricted model.

3.6.1.7 Hypothesis Testing

The hypotheses for the LR test are as follows:

- H_0 (null hypothesis): The restrictions imposed on the restricted model are valid, i.e., the restricted model fits the data as well as the unrestricted model.
- H_1 (alternative hypothesis): The restrictions imposed on the restricted model are not valid, i.e., the unrestricted model provides a significantly better fit to the data.

3.6.1.8 Interpretation of Results

The LR statistic of 179.061 is large, and the corresponding p-value of 0.000 is well below the typical significance level of 0.05. This indicates strong evidence against the null hypothesis. Therefore, we reject H_0 in favor of H_1.

The rejection of the null hypothesis suggests that the restrictions imposed on the restricted model lead to a significantly worse fit compared to the unrestricted model. In other words, the unrestricted model provides a statistically better fit to the data.

3.6.2 Hypothesis Test for DT in the Cointegrating Relation

Now we test whether DT enters the cointegrating relation by imposing a zero restriction on the first element of the cointegration vector β.

3.6.2.1 Hypotheses

The hypotheses for this test are formulated as follows:

- **Null Hypothesis (**H_0**)**: The first element of the cointegration vector β, which corresponds to DT, is zero. This means that DT does not enter the cointegrating relation(s).

$$H_0 : \beta_{\mathrm{DT}} = 0$$

- **Alternative Hypothesis (**H_1**)**: The first element of the cointegration vector β, which corresponds to DT, is not zero. This means that DT1 does enter the cointegrating relation(s).

$$H_1 : \beta_{\mathrm{DT}} \neq 0$$

3.6.2.2 Test Results

For DT1, the test was performed by comparing the unrestricted and restricted models, and the following results were obtained:

- **Unrestricted log-likelihood**: 3105.110
- **Restricted log-likelihood**: 3102.686
- **Degrees of freedom (df)**: 2
- **LR statistic**: 4.849
- **P-value**: 0.089

At the 5% significance level, the p-value of 0.089 is greater than 0.05. Therefore, we **fail to reject the null hypothesis** H_0. This suggests that there is no strong evidence to indicate that DT1 significantly enters the cointegrating relation.

For DT2, the test was performed by comparing the unrestricted and restricted models, and the following results were obtained:

- **Unrestricted log-likelihood**: 3105.110
- **Restricted log-likelihood**: 3104.439
- **Degrees of freedom (df)**: 2
- **LR statistic**: 1.343
- **P-value**: 0.511

At the 5% significance level, the p-value of 0.511 is greater than 0.05. Therefore, we **fail to reject the null hypothesis** H_0. This suggests that there is no strong evidence to indicate that DT2 significantly enters the cointegrating relation.

For DT3, the test was performed by comparing the unrestricted and restricted models, and the following results were obtained:

- **Unrestricted log-likelihood**: 3105.110
- **Restricted log-likelihood**: 3104.004
- **Degrees of freedom (df)**: 2
- **LR statistic**: 2.213
- **P-value**: 0.331

At the 5% significance level, the p-value of 0.331 is greater than 0.05. Therefore, we **fail to reject the null hypothesis** H_0. This suggests that there is no strong evidence to indicate that DT3 significantly enters the cointegrating relation.

For DT4, the test was performed by comparing the unrestricted and restricted models, and the following results were obtained:

- **Unrestricted log-likelihood**: 3105.110
- **Restricted log-likelihood**: 3102.099

- **Degrees of freedom (df)**: 2
- **LR statistic**: 0.024
- **P-value**: 0.000

At the 5% significance level, the p-value of 0.000 is less than 0.05. Therefore, we **reject the null hypothesis** H_0. This suggests that there is strong evidence to indicate that DT4 significantly enters the cointegrating relation.

For DT5, the test was performed by comparing the unrestricted and restricted models, and the following results were obtained:

- **Unrestricted log-likelihood**: 3105.110
- **Restricted log-likelihood**: 3102.056
- **Degrees of freedom (df)**: 2
- **LR statistic**: 6.108
- **P-value**: 0.047

At the 5% significance level, the p-value of 0.047 is less than 0.05. Therefore, we **reject the null hypothesis** H_0. This suggests that there is strong evidence to indicate that DT5 significantly enters the cointegrating relation.

3.6.3 Long-Run Exogeneity Tests

As only DT4 and DT5 significantly enters the cointegrating relation. This section presents the results of the long-run exogeneity tests for DT4 and DT5. The tests aim to determine whether these variables adjust to deviations from the long-run equilibrium relationship.

3.6.3.1 Hypotheses

The hypotheses for the long-run exogeneity tests are formulated as follows:

- **Null Hypothesis (H_0)**: The variable is long-run exogenous. This implies that the adjustment coefficient α for the variable is zero, meaning that the variable does not respond to deviations from the long-run equilibrium.

$$H_0 : \alpha = 0$$

- **Alternative Hypothesis (H_1)**: The variable is not long-run exogenous. This implies that the adjustment coefficient α for the variable is non-zero, meaning that the variable adjusts to deviations from the long-run equilibrium.

$$H_1 : \alpha \neq 0$$

3.6.3.2 Test Results for DT4

The test for the long-run exogeneity of DT4 yielded the following results:

	Statistic	P-value
Unrestricted log-likelihood	3105.110	—-
Restricted log-likelihood	3105.048	—-
Degrees of freedom (df)	2	—-
Likelihood Ratio (LR) statistic	0.124	0.940

TABLE 3.8: DT4 Long-Run Exogeneity Tests Results

The p-value of 0.940 is significantly higher than the 0.05 threshold, indicating that we failed to reject the null hypothesis. The LR statistic is very small (0.124), implying that there is minimal difference between the restricted and unrestricted models. These results suggest that DT4 does not adjust to deviations from the long-run equilibrium and can be considered long-run exogenous.

3.6.3.3 Test Results for DT5

The test for the long-run exogeneity of DT5 produced the following results:

	Statistic	P-value
Unrestricted log-likelihood	3105.110	—-
Restricted log-likelihood	3101.914	—-
Degrees of freedom (df)	2	—-
Likelihood Ratio (LR) statistic	6.393	0.041

TABLE 3.9: DT5 Long-Run Exogeneity Tests Results

The p-value of 0.041 is below the 0.05 threshold, leading to the rejection of the null hypothesis. The LR statistic of 6.393 indicates that the restriction imposed by assuming DT5 is long-run exogenous significantly worsens the model fit. This suggests that DT5 adjusts to deviations from the long-run equilibrium and is not long-run exogenous.

3.6.4 OPT Restricted FCVAR Model

In this section, we impose restrictions on both the adjustment matrix α and the cointegration vectors β based on the above tests. Specifically, we apply the following restrictions:

- Restrictions on α: We restrict the adjustment coefficients for DT1, DT2, DT3, and DT4 to be zero, implying that these variables are long-run exogenous.

- Restrictions on β: We restrict DT1, DT2, and DT3 from entering the cointegrating relationships.

The restricted model for all variables X_t, incorporating both short-term dynamics and long-term equilibrium adjustments, is given by:

$$\Delta^d X_t = \alpha_{\text{restricted}} \beta'_{\text{restricted}} L_{\hat{b}} \Delta^{d-\hat{b}} X_t + \sum_{i=1}^{k} \Gamma_i \Delta^d L_{\hat{b}}^i X_{t-i} + \epsilon_t \tag{3.37}$$

where X_t is the vector of variables $[\text{DT1}, \text{DT2}, \text{DT3}, \text{DT4}, \text{DT5}]'$:

$$X_t = \begin{bmatrix} \text{DT1}_t \\ \text{DT2}_t \\ \text{DT3}_t \\ \text{DT4}_t \\ \text{DT5}_t \end{bmatrix} \tag{3.38}$$

Still, we assume that $d = 1$, full System of Restricted Equations:

$$\begin{aligned} \Delta \begin{bmatrix} \text{DT1}_t \\ \text{DT2}_t \\ \text{DT3}_t \\ \text{DT4}_t \\ \text{DT5}_t \end{bmatrix} &= \alpha_{\text{restricted}} \beta'_{\text{restricted}} L_{\hat{b}} \Delta^{1-\hat{b}} \begin{bmatrix} \text{DT1}_t \\ \text{DT2}_t \\ \text{DT3}_t \\ \text{DT4}_t \\ \text{DT5}_t \end{bmatrix} \\ &+ \sum_{i=1}^{3} \Gamma_i \Delta L_{\hat{b}}^i \begin{bmatrix} \text{DT1}_{t-i} \\ \text{DT2}_{t-i} \\ \text{DT3}_{t-i} \\ \text{DT4}_{t-i} \\ \text{DT5}_{t-i} \end{bmatrix} + \epsilon_t \end{aligned} \tag{3.39}$$

3.6.4.1 Adjustment matrix $\alpha_{\text{restricted}}$

The restricted adjustment matrix $\alpha_{\text{restricted}}$ is:

$$\alpha_{\text{restricted}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ -1.880 & -0.626 \end{bmatrix} \tag{3.40}$$

3.6.4.2 Cointegration matrix $\beta'_{\text{restricted}}$

The restricted cointegration matrix $\beta'_{\text{restricted}}$ is:

$$\beta'_{\text{restricted}} = \begin{bmatrix} 0 & 0 & 0 & 0.116 & 0.880 \\ 0 & 0 & 0 & -0.007 & 0.791 \end{bmatrix} \tag{3.41}$$

3.6.4.3 Short-run dynamics (lag matrices Γ_i)

The short-run dynamics are captured by the lag matrices Γ_1, Γ_2, and Γ_3, which include the coefficients for the lagged differences of the variables.

3.6.4.4 Final Restricted Model Equation

The fractional differencing parameter $\hat{b} = 0.226$, fractional integration order is given by $d - \hat{b} = 1 - \hat{b} = 0.774$. The final restricted model equation is:

$$\begin{aligned} \Delta X_t = \alpha_{\text{restricted}} \beta'_{\text{restricted}} L_{\hat{b}} \Delta^{1-\hat{b}} X_t + \Gamma_1 \Delta L^1_{\hat{b}} X_{t-1} \\ + \Gamma_2 \Delta L^2_{\hat{b}} X_{t-2} + \Gamma_3 \Delta L^3_{\hat{b}} X_{t-3} + \epsilon_t \end{aligned} \tag{3.42}$$

where, ΔX_t represents the fractional differencing operator applied to the time series X_t, accounting for the fractional integration order $d = 1$. $\alpha_{\text{restricted}}$ is the $p \times r$ adjustment matrix, which contains the loading coefficients that describe how the system reacts to deviations from the long-run equilibrium. $\beta'_{\text{restricted}}$ is the cointegration vector, defining the long-run equilibrium relationships between the variables. $L_{\hat{b}}$ is the lag operator, with $\hat{b} = 0.226$. Γ_i are the short-run dynamic coefficients (lag matrices) for $i = 1, \ldots, k$, $k = 3$, which capture the short-term relationships between the variables. ϵ_t is the vector of residuals, representing the model's error terms.

3.6.4.5 Residual Analysis and White Noise Test

The residuals from the restricted model are analysed using the Ljung-Box Q test and the Lagrange Multiplier (LM) test to check for serial correlation, ensuring that the model adequately captures the dynamics of the system.

Variable	Q-statistic	Q P-value	LM-statistic	LM P-value
Multivar	42.448	0.767	—	—
DT1	4.996	0.082	1.280	0.527
DT2	2.866	0.239	0.531	0.767
DT3	1.908	0.385	0.649	0.723
DT4	10.916	0.004	0.771	0.680
DT5	1.059	0.589	0.489	0.783

TABLE 3.10: White Noise Test Results (lag = 2)

The multivariate test evaluates whether the residuals as a system exhibit serial correlation. The results show a Q-statistic of 42.448 and a P-value of 0.767, indicating that there is no significant serial correlation in the system as a whole. This suggests that, while there might be some individual issues, the model is generally adequate in capturing the overall dynamics of the variables.

For each individual variable, we interpret the Ljung-Box Q test and the LM test results as follows:

- $DT1_t$: The Ljung-Box Q test results in a Q-statistic of 4.996 and a P-value of 0.082, indicating marginal serial correlation in the residuals for DT1_t. The LM test yields a statistic of 1.280 and a P-value of 0.527, suggesting no evidence of serial correlation from the LM test.
- $DT2_t$: The Q test results in a Q-statistic of 2.866 and a P-value of 0.239, suggesting no significant serial correlation in the residuals for DT2_t. The LM test shows a statistic of 0.531 and a P-value of 0.767, confirming no significant serial correlation.
- $DT3_t$: The Q test produces a Q-statistic of 1.908 and a P-value of 0.385, indicating no significant serial correlation in the residuals for DT3_t. The LM test yields a statistic of 0.649 and a P-value of 0.723, showing no evidence of serial correlation.
- $DT4_t$: The Q test shows a Q-statistic of 10.916 and a P-value of 0.004, indicating significant serial correlation in the residuals for DT4_t. The LM test results in a statistic of 0.771 and a P-value of 0.680, showing no significant serial correlation from the LM test.
- $DT5_t$: The residuals for DT5_t show no significant serial correlation, as indicated by the Q-statistic of 1.059 and a P-value of 0.589. The LM test also shows no evidence of serial correlation, with a statistic of 0.489 and a P-value of 0.783.

3.7 Forecasting

The above restricted FCVAR model was used to generate forecasts for the five series over the next 12 periods. Table 3.11 presents selected forecasted values for periods 721 to 732. The forecasted values indicate that for M1, the market proxy price will go up from 722; the market proxy price for M2 will go down; the market proxy price for M3 will go up from 722; the market proxy price for M4 will go down from 723; the market proxy price for M5 will go down.

Time	DT1	DT2	DT3	DT4	DT5
721	-0.028	-0.047	-0.105	0.067	-0.040
722	0.002	-0.042	-0.027	0.010	-0.040
723	0.033	-0.039	0.032	-0.045	-0.041
724	0.063	-0.037	0.073	-0.095	-0.043
725	0.088	-0.036	0.098	-0.140	-0.045
726	0.107	-0.036	0.108	-0.177	-0.047
727	0.120	-0.036	0.105	-0.206	-0.048
728	0.126	-0.036	0.093	-0.227	-0.049
729	0.124	-0.035	0.073	-0.242	-0.049
730	0.117	-0.035	0.048	-0.249	-0.048
731	0.104	-0.034	0.020	-0.250	-0.047
732	0.088	-0.033	-0.008	-0.246	-0.045

TABLE 3.11: Forecasted DT Values (Selected Periods)

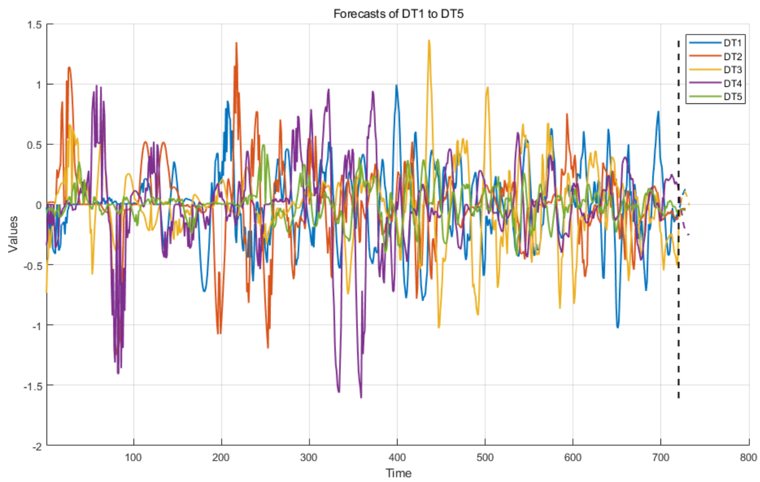

FIGURE 3.18: FCVAR Forecasted DT Values

3.8 Conclusions

This Chapter investigates the dynamic efficiency in betting markets using a novel decomposition model to analyse long-memory properties and informational inefficiency in betting markets. By identifying the integration orders (d) and constructing the degree of market inefficiency D, the study reveals the evolving efficiency of betting markets over time. The findings suggest that the betting market becomes more efficient as the betting progresses. The deviation pattern and inefficiency from convergence also confirm that market efficiency gradually increases over time in cross market way.

Moreover, this paper proposes a method to identify the optimal combination of window size (WS), bandwidth (BD), and estimator for estimating the integration orders (d), called Estimation Score for Integration Orders ($ESIO$). This score measures the closeness of d values to 1. Various combinations of WS, WS, and four estimators (LW, ELW, FELW, Two-Step FELW) are tested. The combination with the minimum $ESIO$ is selected as optimal for each market.

The FCVAR model was used to generate forecasts for the DT series of five markets proxy prices ($DT1$ to $DT5$) over the next 12 periods. The FCVAR model provides a robust framework for understanding the long-run relationships between the five time series. The estimation results reveal five significant cointegrating relationships, with fractional differencing parameter $\hat{b} = 0.226$, indicating long memory behaviour. The predicted values indicate the future movement of the price for M1 to M5.

Chapter 4

Title: The Dynamic Process of Equilibrium in a Betting Market

4.1 Introduction

In Chapter 2, we provide specific mathematical models to capture the DT and deeply analyse the mathematical process of the above theory. We also introduce the tangent proxy of the Dynamic Trend. Further analysis of market dynamics reveals that changes in market consensus are jointly determined by external information inputs and market conditions. When external information is captured by market participants, it begins to influence the market. Different participants with varying behavioural patterns have different abilities to capture and interpret the information, and their influence on the market changes over time. Therefore, the information incoming and the market state, as shaped by all participant behaviours, evolve over time. Thus, the dynamic description of the market should consist of two dimensions: Information and Market Condition. Here, we understand all incoming market information as an information set, and all market participants as the entire market. For the observer analysing the market, the observed market state is also related to the observation method, specifically the Fundamental (Primary) Time Level and Horizon used by the observer. Therefore, we can describe the market with Information, Market Condition, Fundamental (Primary) Time Level, and Horizon. The Market Condition is described by the dynamic equilibrium of the market system. This paper further decomposes the DT, under constant Fundamental (Primary) Time Level and Horizon, into the Information Affect Factor (IAF) and the Market Equilibrium Condition Factor (MECF), and provides the relationship between the two as follows:

$$(\text{IAF})^2 + (\text{MECF})^2 = 1 \tag{4.1}$$

The market as we assume as a system, the condition of this system will change due to the input and the market response to this input. Even the for the same input, the different market condition will gives a different reaction. This input is modelling by IAF and the market condition at each time point is modelling by MECF. The specific derivation and definition will given in section 4.3.

This model provides an explanation of how market equilibrium adjusts in response to incoming information, thereby leading to stages of equilibrium, overreaction, and the subsequent decay of such overreactions.

Moreover, we introduce constraints on the transmission of information and market efficiency, imposed by both physical factors, such as the speed of light, and economic factors, including market liquidity. These constraints inherently limit the rate at which information can be processed and incorporated into prices, ultimately guiding the market towards equilibrium in the absence of new information.

This chapter underscores the dynamic and bounded nature of market responses to information inputs, wherein trends oscillate due to the complex interplay of incoming data, market liquidity, and the intrinsic limitations of information processing.

4.2 The Information Dimension and the Market Dimension

With the dynamic trend analysis approach, we decompose the price series into a dynamic trend series and a volatility series. The dynamic trend series accounts for the sustained influence of relative macroeconomic factors and significant external events on prices. Conversely, the volatility series reveals the market's short-term behaviours.

In the Dynamic Trend section, we define the trend vector and decompose the trend vector into time changing vector(TC) and a data changing vector(DC) in a Cartesian coordinate system (2D, real axis). The tangent proxy of dynamic trend, denoted as $\tan(\alpha)$, at a certain observe time point with a certain Time Level and Horizon can be expressed as:

$$\tan(\alpha) = \frac{\text{Sign} \cdot |DC|}{\text{Sign} \cdot |TC|} \tag{4.2}$$

Where $|DC|$ and $|TC|$ represent the absolute values of the data changing vector and time changing vector. Given that time inherently progresses in a singular direction—from past to future—the time changing vector (TC) invariably maintains a positive direction. Consequently, the sign of $\text{Sign} \cdot |TC|$ is always positive. This leads to a simplified expression for $\tan(\alpha)$:

$$\tan(\alpha) = \frac{\text{Sign} \cdot |DC|}{|TC|}. \tag{4.3}$$

The dynamic trend can be quantified as the velocity of data change. This is mathematically represented by the tangent of an angle α, where $\tan(\alpha)$ is defined as:

$$\tan(\alpha) = \frac{\text{Sign} \cdot |DC|}{|TC|}, \tag{4.4}$$

Indicating the rate of data change per unit of time. Here, $|DC|$ and $|TC|$ denote the magnitude of the data changing vector and time changing vector. Sign denote the direction of the data changing vector, and gives the direction of the dynamic trend. We know that:

$$\tan(\alpha) = \frac{\sin(\alpha)}{\cos(\alpha)} \tag{4.5}$$

Assume we have the trend vector T, $|T|$ denote the magnitude of the trend vector.

$$\sqrt{|T|^2} = \sqrt{|DC|^2 + |TC|^2} \tag{4.6}$$

incorporating both Sign $\cdot\, |DC|$ and $|TC|$. We have:

$$|T| \cdot \sin(\alpha) = \text{Sign} \cdot |DC| \tag{4.7}$$

and

$$|T| \cdot \cos(\alpha) = |TC| \tag{4.8}$$

. We have Euler's formula:

$$e^{i\alpha} = \cos(\alpha) + i\sin(\alpha), \tag{4.9}$$

Introducing a complex number T, representing the trend, and defining it through Euler's formula as:

$$T = |T|e^{i\alpha} \tag{4.10}$$

or

$$T = |T|(\cos(\alpha) + i\sin(\alpha)). \tag{4.11}$$

Divide both sides of above two equations by $|T|$, we have:

$$\frac{T}{|T|} = \frac{|T|e^{i\alpha}}{|T|} \tag{4.12}$$

and

$$UT = \frac{|T|\cos(\alpha) + |T|i\sin(\alpha)}{|T|}. \tag{4.13}$$

Simplifying:

$$UT = e^{i\alpha} \tag{4.14}$$

and

$$UT = \cos(\alpha) + i\sin(\alpha). \tag{4.15}$$

This operation is essentially normalizing the complex number T to its unit form on the complex plane, which is unit trend UT, represented by $e^{i\alpha}$ or $\cos(\alpha) + i\sin(\alpha)$, which is a point on the unit circle.

In the Cartesian coordinate system, we know that time inherently progresses in a singular direction—from past to future. This gives the α (the angle in radians) ranges between $(-\frac{\pi}{2}, \frac{\pi}{2})$. In the complex plane, for values of α between $(-\frac{\pi}{2}, \frac{\pi}{2})$, UT will take values corresponding to points on the right half of the unit circle, with varying proportions of real and imaginary components based on the values of $\cos(\alpha)$ and $\sin(\alpha)$. The proxy of DT now is the position on the unit circle. For values of α between between $(0, \frac{\pi}{2})$, DT would be positive, indicating the Up Trend, the higher the position, the larger the DT magnitude. For values of α between between $(-\frac{\pi}{2}, 0)$, DT would be negative, indicating the Down Trend, the lower the position, the larger the DT magnitude. When $\alpha = 0°$, the DT would be 0, indicating the Flat Trend.

- The green vector corresponds to $\alpha = 45°$.

- The red vector corresponds to $\alpha = -45°$.
- The purple vector corresponds to $\alpha = 0°$.

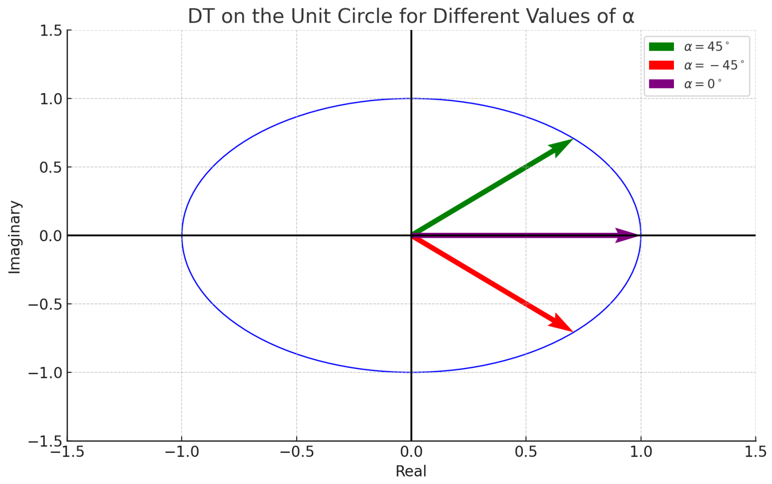

FIGURE 4.1: *DT* Proxy of the Positions on Unit Circle for Different Values of α

These vectors illustrate how the complex number UT changes its position on the unit circle as the angle α varies. $\cos(\alpha)$ represent the rate of time of UT, $\sin(\alpha)$ represent the rate of price changing in UT.$\sin(\alpha)$ indicating the information effect of the dynamic process of equilibrium of market; $\cos(\alpha)$ represent the market equilibrium condition of the dynamic process of equilibrium of the market. $\sec(\alpha)$ represents the market capability of information understanding the effect of the dynamic process of equilibrium of the market. When $\tan(\alpha) = 0$ means the market has reached equilibrium.

They have a relationship as follows.

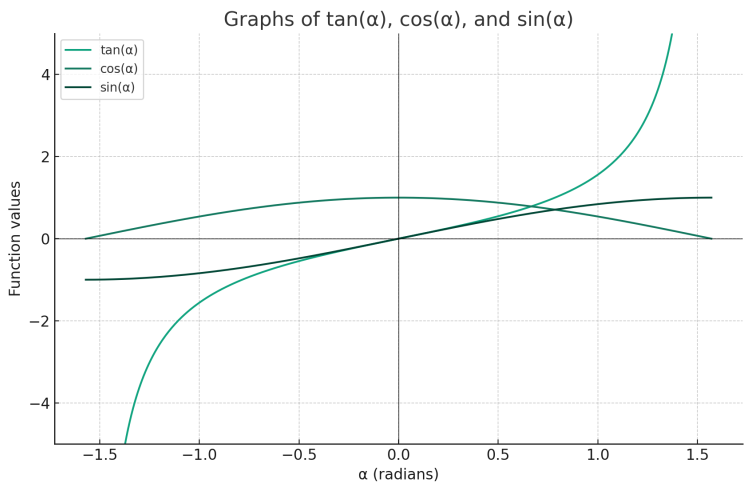

FIGURE 4.2: Trigonometric Functions for Different Values of α

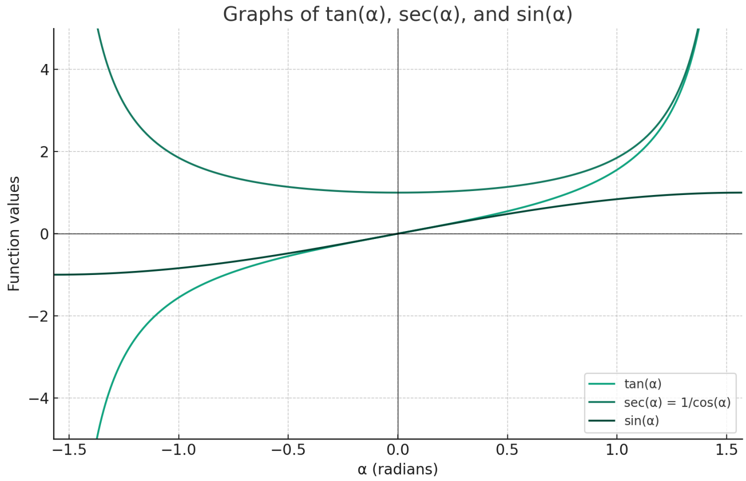

FIGURE 4.3: Trigonometric Functions With Market Efficiency Term for Different Values of α

We have below relationship:

$$\tan(\alpha) = \frac{\sin(\alpha)}{\cos(\alpha)} = \sin(\alpha) \cdot \sec(\alpha) \tag{4.16}$$

Where, $\sin(\alpha)$ indicates the information effect of the dynamic process of equilibrium of the market; $\sec(\alpha)$ is the information discount sector, which is the dynamic market efficiency factor.

4.3 The Dynamic Process of Equilibrium of Market

4.3.1 The Dynamic Process for a Certain information

Let's start from the process with only one piece of information coming into the system at the start time point of our observation time span.

For a market that has reached the equilibrium. If information is significant to the market have reached the market. It should be turned into a change of price. When it arrives on the market, only a few people catch it and understand it($\sin(\alpha)$ is going up). They will lead the price change in the right direction(corresponding $\sec(\alpha)$ is going up). With time going, some other people will notice that the value of the information and start joining the trading ($\sin(\alpha)$ is going higher). They will push the price mover faster to the equilibrium (corresponding $\sec(\alpha)$ is going higher). All traders in the market started to achieve the first stage consensus that the price should go in that direction.

When most traders notice the information ($\sin(\alpha)$ arrived at the peak), some traders will lock the benefits by closing the positions and others may just start trading as they have different abilities of understanding(corresponding $\sec(\alpha)$ arrived the peak). The market is processing from the first stage consensus to the final consensus where the $\sin(\alpha)$ is going down and the corresponding $\sec(\alpha)$ is going down. When all different types of traders have the final consensus by the trading activities, the market arrives at the new equilibrium, $\tan(\alpha) = 0$.

Assume we have a shock come into the market. As the information is very significant ($\sin(\alpha)$ will shoot up to a high level), it is very hard to be missed, all of the above traders will notice it in a short time. All above shows a shooting up market efficiency (corresponding $\sec(\alpha)$ will shoot up to a high level). The information would be valued in a fast way and the equilibrium of the market be achieved very soon.

4.3.2 The Over-Reaction

Even if the information is fully valued, which means the data changing from this information is already fulfilled, the market still may overreact for a while.

Traders who really understand the information and turn it in the right direction and the amount of data changing by their trading activity may already finish the trading action. Others are still trading as their trading is now based on the primary endogenetic information, the information generated by the market reaction to the outside information, and the market activity changing from the outside information dominant to the primary endogenetic information dominant. Their trading pushes the data change still going in that direction. This is the primary part of the overreaction.

Moreover, those smart traders notice that the information is overvalued by others. They will do trading in the opposite direction to fix the over-reaction part as it could get more benefit. Here the over-reaction is the first order secondary endogenetic information (new information generated by the market itself). With time going on, the overreacted traders have noticed the new information and finally understand the old information. More and more overreacted traders are joining the new process of trading the first order secondary endogenetic information (over-reaction).

When the over-reaction is very significant, the above over-reaction fixing process will generate a new secondary endogenetic information, called second order secondary endogenetic information, this second order secondary endogenetic information is weaker than the first order secondary endogenetic information. It might lead to a third order and even more orders of secondary endogenetic information. As the higher order secondary endogenetic information will be weaker, it will be a convergent process. This process is the decaying fluctuations of the DT series after the first single direction fluctuation caused by the outside information and primary endogenetic information.

If no other information arrives in the market, the above process will finally lead the market to arrive at a new equilibrium.

Now we consider the Horizon different, we could know that with a certain Time Level, the over-reaction will be higher with a lover Horizon. This could indicate that if we could see further, the bias would lower. The short-term(low Horizon) traders will lead to a higher bias. We take the DT series with different Horizons (5,20,64) for the Market4 winner horse's price series for example below.

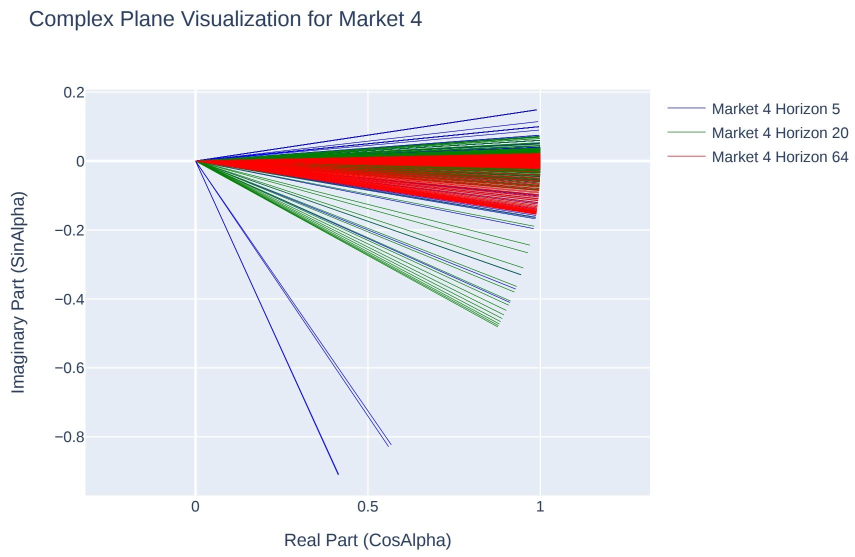

FIGURE 4.4: For a Shorter Horizon, the Over-reaction Will Higher

4.3.3 The Dynamic Process for Dynamic Information

Now we assume the information is coming into the market in a dynamic way, this is more like the condition of reality. The different kinds of information keep coming into the market as time going. The market equilibrium is a dynamic process, it reacts to all information in the market at each certain observe time point. We could use $\cos(\alpha)$ to be the dynamic equilibrium index to describe the market equilibrium condition at each observe time point.

The range of $\cos(\alpha)$ between $(0, 1]$. When $\cos(\alpha) = 1$, the market arrives at the full equilibrium. While, $\sin(\alpha)$ indicates the information effect, between (-1,1). $\sin(\alpha)$ between (0,1) indicating the overall information effect is positive. If close to 1, indicating a good shock. $\sin(\alpha)$ between (-1,0) indicating the overall information effect is negative. If close to -1, indicating a bad shock. $\cos(\alpha)$ and $\sin(\alpha)$ have a relationship that the sum of the squares of $\sin(\alpha)$ and $\cos(\alpha)$ equals 1, as below:

$$\cos^2(\alpha) + \sin^2(\alpha) = 1 \tag{4.17}$$

We define $\sin(\alpha)$ as the Information Affect Factor (IAF) and $\cos(\alpha)$ as the Market Equilibrium Condition Factor (MECF), the relationship between the two is as follows:

$$(\text{MECF})^2 + (\text{IAF})^2 = 1 \tag{4.18}$$

This gives the relationship between market equilibrium and information on the market. The market dynamically processes to equilibrium as the information dynamically arrives.

4.3.4 4-Dimensions of Information Equilibrium Dynamic

We examine the complex function TU defined by:

$$TU = \cos(\alpha) + i\sin(\alpha)$$

where α is determined from a time-dependent series DT with the relationship $\alpha = \arctan(DT)$. This function maps the time-dependent values onto the unit circle in the complex plane, maintaining a magnitude of 1 for TU, i.e., $|TU| = 1$.

The real and imaginary parts of TU are represented by $\cos(\alpha)$ and $\sin(\alpha)$, respectively. To visualize this function's behaviour over time, we create a three-dimensional plot where:

- The x-axis represents the real part of TU.
- The y-axis represents the imaginary part of TU.
- The z-axis signifies time level, correlating with the values of DT.

Additionally, we introduce a plane in this 3D space representing TU_0, the value of TU when $\alpha = 0$. This plane extends from the point $0 + i0$ to $1 + i0$ across the entire time range, illustrating the flat trend. When the TU lying above the TU_0 plane, it shows the Up Dynamic of the data; when the TU lying beneath the TU_0 plane, it shows the Down Dynamic of the data. This visualization effectively demonstrates the dynamic evolution of TU in response to changes in time.

When the TU above TU_0, it would be an Up Trend; when it is beneath TU_0, it would be a Down Trend. Here we take the DT series with different Horizons (5,20,64) for the Market4 winner horse's price series for example below.

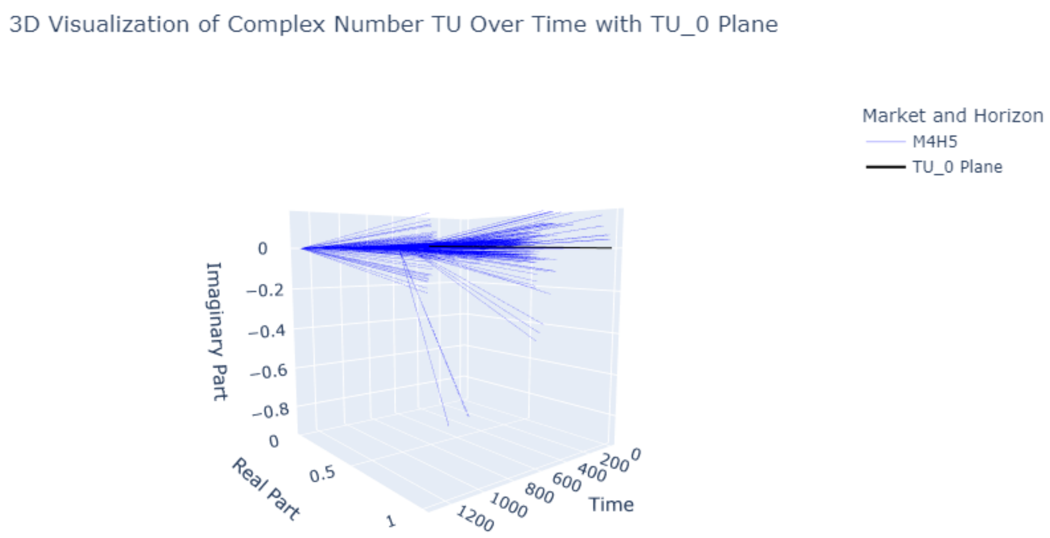

FIGURE 4.5: 3D Visualization of$M4$ TU with Plane at TU_0 ($\alpha = 0$) with Horizon5

We take the M4 TU, the Horizons are 5, 20, and 64.

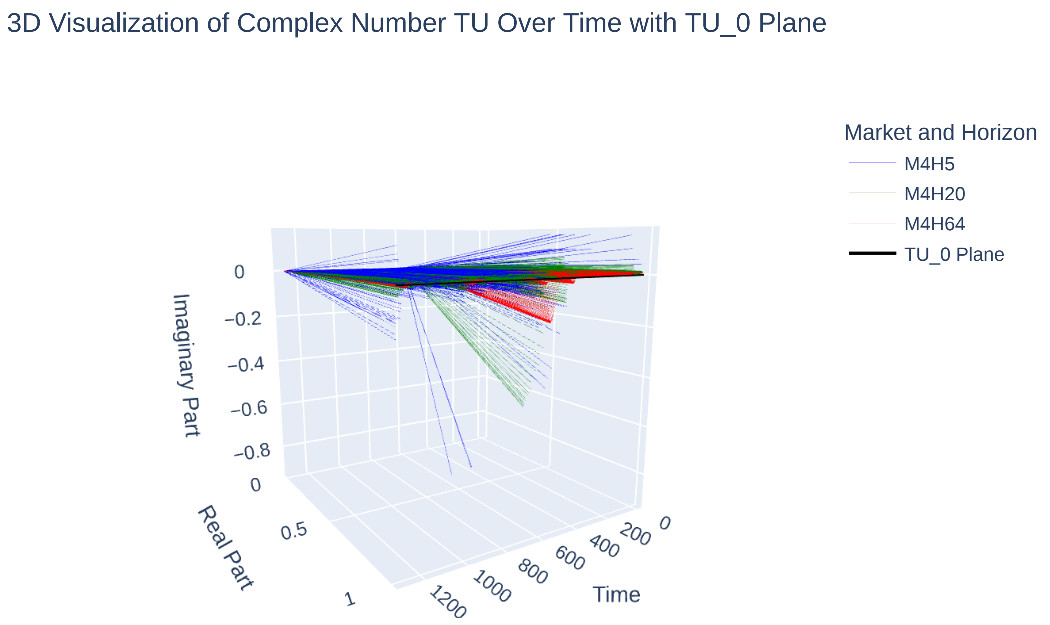

FIGURE 4.6: 3D Visualization of $M4TU$ with Horizons

The dynamic process for DT, $ImaginaryPart(Sin(\alpha))$, and $RealPart(Cos(\alpha))$ over time for different Horizons has been plotted below.

4.4 The Limitations of Information Transmission and Market Efficiency

4.4.1 The Physical Limitations of Information Transmission and Market Efficiency

In the realm of market dynamics, all market participants and the transnational systems they rely upon are fundamentally physical entities. Consequently, the rate at which information is transformed and disseminated in the market is inherently constrained by physical laws, most notably the limitation imposed by the speed of light. The transformation of information must comply with these physical principles, implying that the magnitude of the velocity of information transfer cannot exceed the speed of light. This holds true not only for the physical transmission of information but also for the cognitive processes involved in information interpretation and valuation by market actors.

Given these constraints, we can deduce that the efficiency of markets possesses a physical upper limit, which is the speed of light. Therefore, in a strict sense, a completely efficient market does not exist. The concept of the Dynamic Trend (DT), which represents the velocity of data

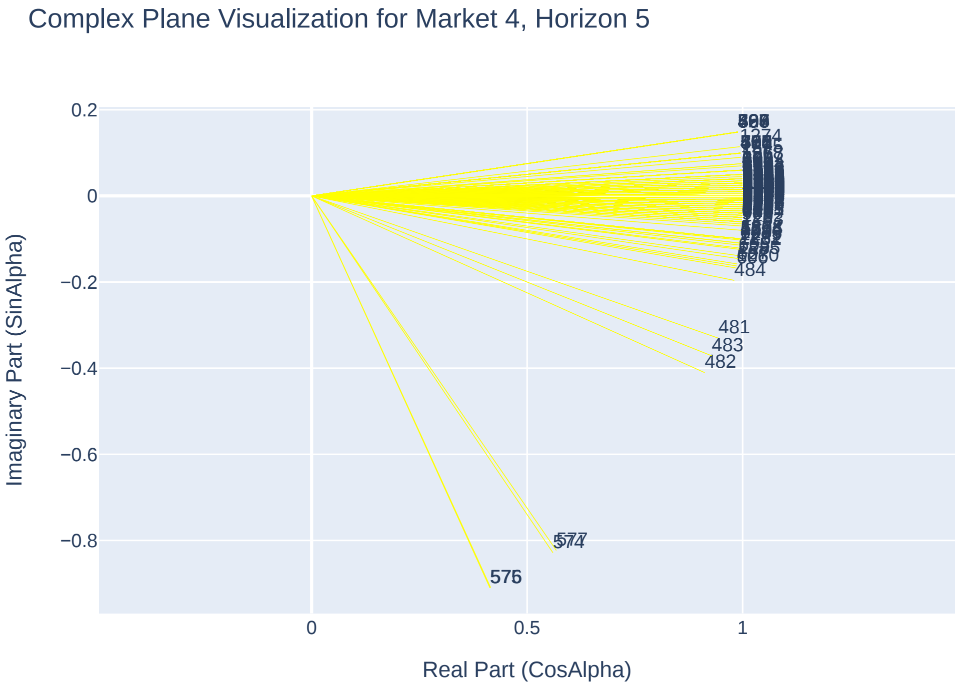

FIGURE 4.7: Complex Plane of $M4TU$ with Horizon 5

change encompassing both direction and magnitude and thus, is vectoring in nature, can only asymptotically approach infinity but never actually reach it.

Moreover, due to the restrictions imposed by physical laws and the cognitive capabilities of market participants, the DT, although capable of rapidly escalating in response to relatively extreme information, will ultimately revert to a certain range. This indicates that the process of converting information into a market consensus (reflected in price increments) is dynamic and tends to fluctuate within a defined range. This dynamic nature underscores the inherent limitations of market efficiency, emphasising that while information can be processed and valued at high speeds, it is always subject to the fundamental constraints of physical laws and human cognitive capacity.

4.4.2 The Economic Limitations of Information Transmission and Market Efficiency

From the economic perspective, market liquidity serves as an additional constraint, impacting the overall efficacy of price movements in a given direction. The total liquidity available in the market is finite, thereby limiting the cumulative effect of price changes in a consistent direction. Consequently, after a sharp increase in the same direction, the Dynamic Trend (DT) is inevitably bound to decrease towards a certain range due to the balancing demands of market liquidity.

This phenomenon reflects the interplay between rapid information processing and the liquidity available in the market. While DT can respond swiftly to new information, driving prices

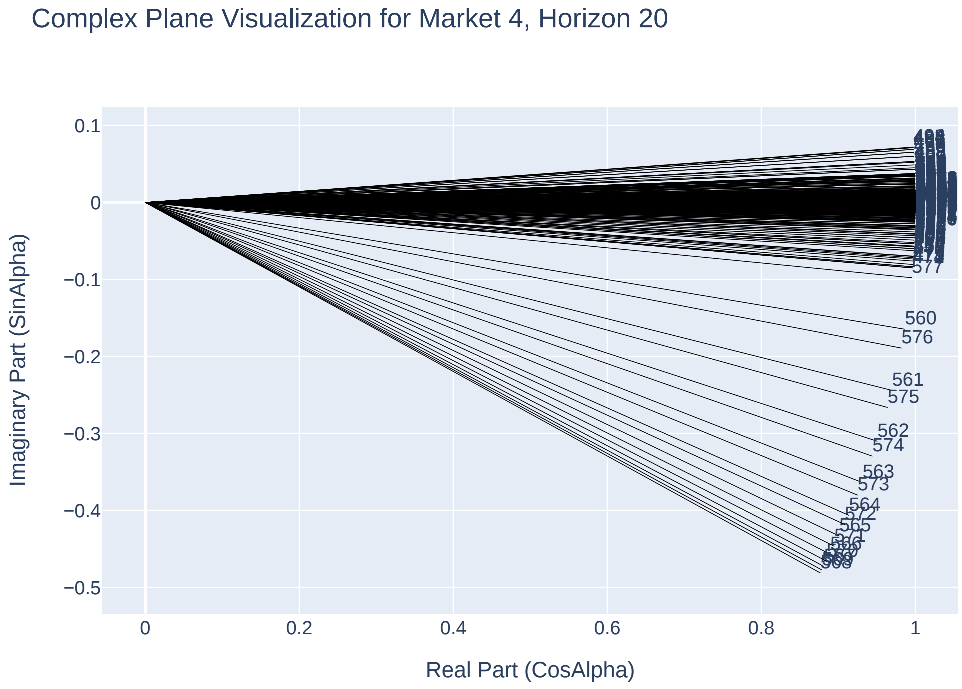

FIGURE 4.8: Complex Plane of $M4TU$ with Horizon 20

in a particular direction, the finite nature of market liquidity acts as a moderating force. It ensures that the momentum generated by the information influx is tempered by the available liquidity, preventing indefinite unidirectional price movements. Thus, the dynamics of DT, while responsive to market information, are inherently balanced by the liquidity constraints, leading to a more stabilized and bounded fluctuation range over time.

Moreover, in the market, the information being priced itself undergoes a decay as the pricing process unfolds. The portion of the information that has been priced is converted into price increments and reflected in the price. Concurrently, the segment of information yet to be priced diminishes during this process. In the absence of new information, this dynamic process inherently drives the Dynamic Trend (DT) towards a gradual regression to zero.

Moreover, the process of information reception in the market is itself dynamic. As existing information is being priced, new information continually enters the market. Therefore, DT exhibits fluctuations around zero. This reflects the ongoing interplay between the assimilation of existing information into market prices and the influx of new information. It underscores the fluid nature of market dynamics, where the impact of information on price movements is constantly evolving. The convergence of DT towards zero in the absence of new information illustrates the market's efficiency in assimilating available data. However, the perpetual arrival of fresh information ensures that DT remains in a state of fluctuation, highlighting the market's responsiveness to an ever-changing information landscape.

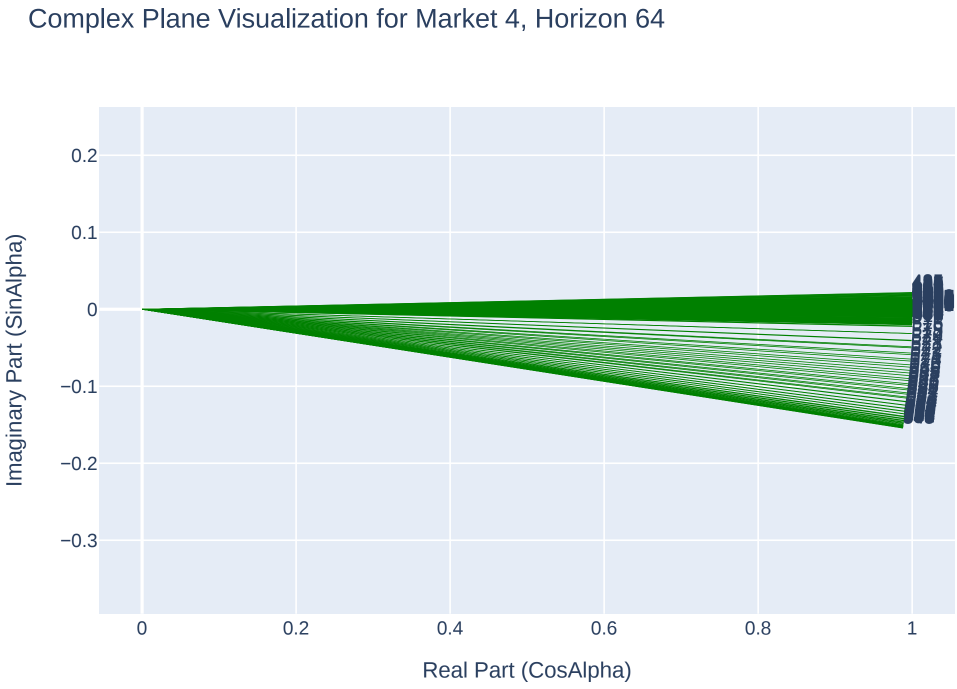

FIGURE 4.9: Complex Plane of $M4TU$ with Horizon 64

4.5 Conclusion

In this chapter, we explore the dynamics of information processing and market equilibrium through a comprehensive model that integrates both the Information Affect Factor (IAF) and the Market Equilibrium Condition Factor ($MECF$). Using a dynamic trend analysis approach, we decompose the price series into a dynamic trend series and a volatility series.

We further relate $\tan(\alpha)$ to the sine and cosine functions, $\sin(\alpha)$ representing IAF and $\cos(\alpha)$ representing $MECF$. This relationship is extended into a complex plane representation using Euler's formula, which normalises the trend vector into a unit trend (UT) on the unit circle. Through this model, we explain how market equilibrium responds to information inputs, leading to stages of market equilibrium and over-reaction, as well as the decay process of over-reactions.

We also introduce the limitations of information transmission and market efficiency, imposed by both physical constraints, such as the speed of light, and economic constraints, including market liquidity. These constraints limit the rate at which information can be processed and priced, ultimately driving the market towards equilibrium in the absence of new information.

This chapter emphasises the dynamic and bounded nature of the market's response to information, where trends fluctuate due to the interplay of incoming data, market liquidity, and the inherent limitations of information processing.

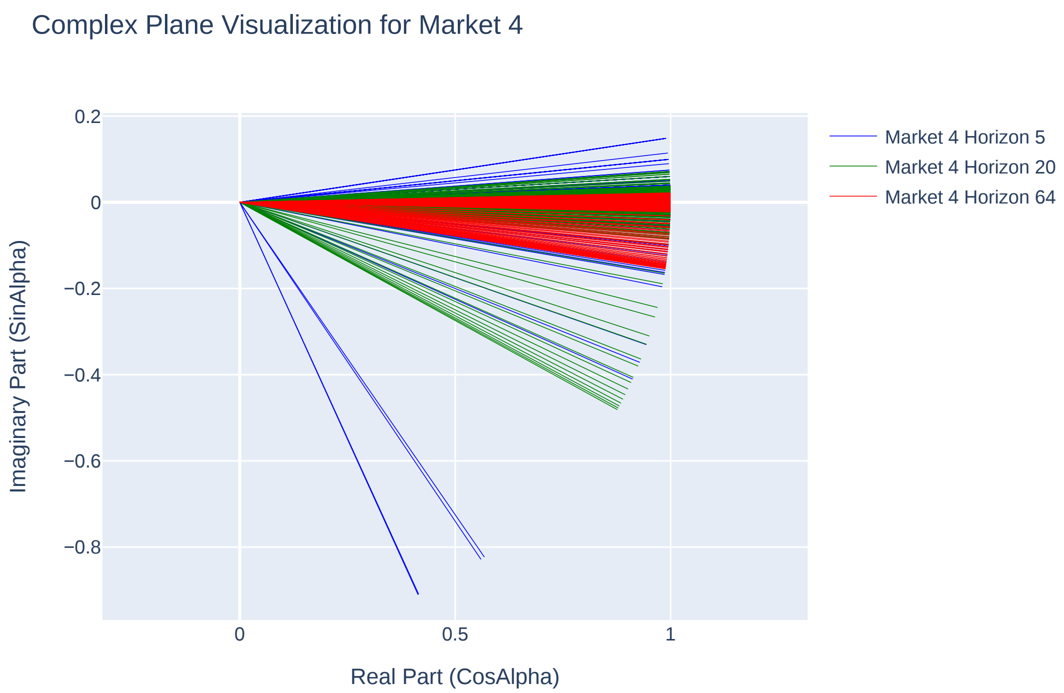

FIGURE 4.10: Complex Plane of $M4TU$ with Horizons

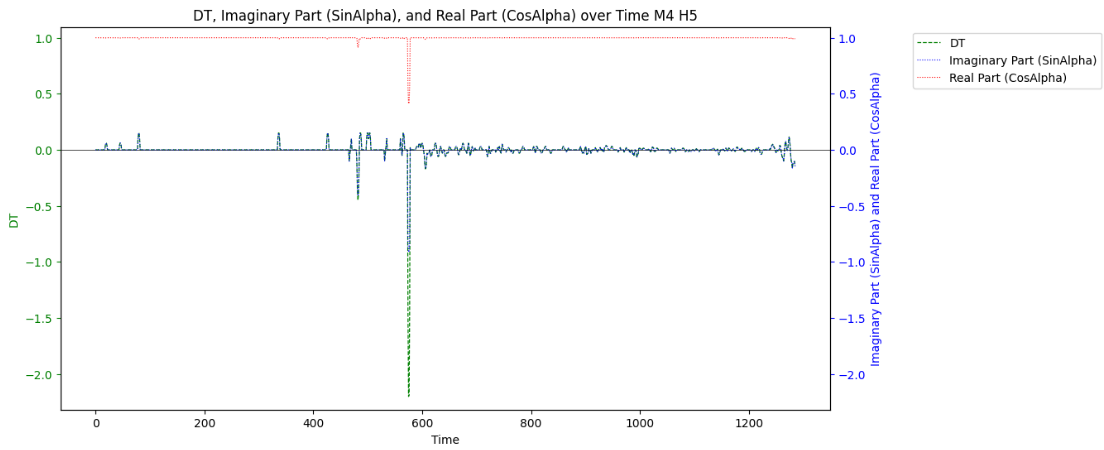

FIGURE 4.11: DT, $ImaginaryPart(Sin(\alpha))$, and $RealPart(Cos(\alpha))$ over Time $M4$ $H5$

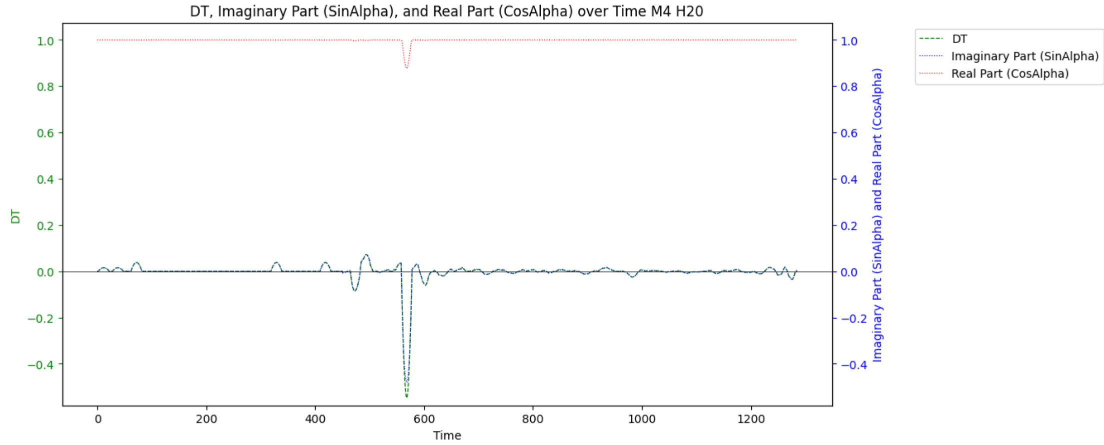

FIGURE 4.12: DT, $ImaginaryPart(Sin(\alpha))$, and $RealPart(Cos(\alpha))$ over Time $M4$ $H20$

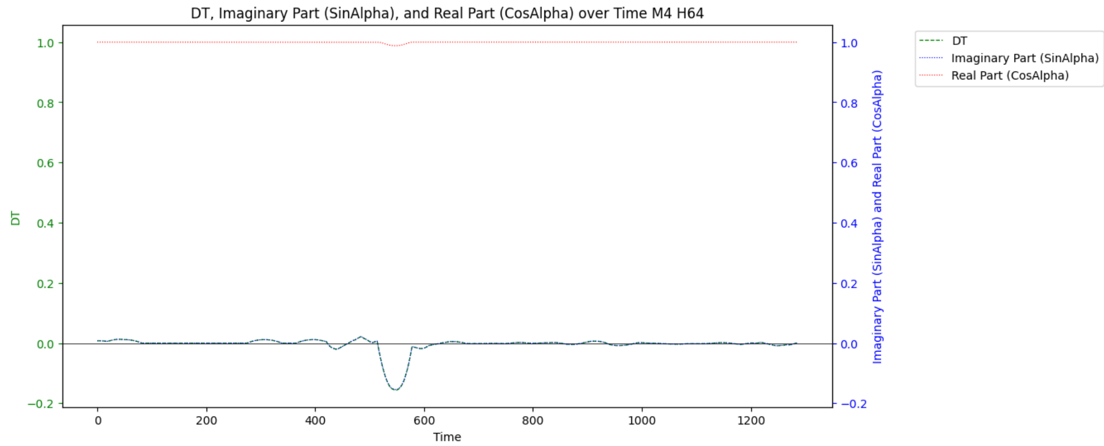

FIGURE 4.13: DT, $ImaginaryPart(Sin(\alpha))$, and $RealPart(Cos(\alpha))$ over Time $M4$ $H64$

Chapter 5

Conclusions

In this paper, we have developed a novel framework for analysing the dynamic processes of price movements in financial markets, with a focus on dynamic trends, information transmission, and market equilibrium. This comprehensive analysis is structured into three main chapters, each contributing to a deeper understanding of market dynamics and offering new methodologies for modelling time series data.

5.1 Chapter 1: Dynamic Trend Analysis and Structural Breaks

In Chapter 1, we introduced the concept of the Dynamic Trend (DT) and proposed a methodology for capturing the temporal and data-changing characteristics of time series. The Dynamic Trend analysis (DTAA) approach decomposes time series data into two distinct components: the Dynamic Trend (DT) Series and the Volatility Series. This decomposition allowed us to better model structural breaks and dynamic patterns in price data.

We also differentiated between two types of structural breaks: Type-I breaks, which involve changes in the direction of the trend, and Type-II breaks, which indicate shifts in the magnitude of the trend. The methodology provides a systematic approach to detecting these breaks and helps identify consistent patterns in dynamic trends and original time series. This approach significantly improves the robustness of time series models.

5.2 Chapter 2: Dynamic Market Efficiency in Betting Markets

Chapter 2 applied the dynamic trend decomposition model to assess market efficiency in betting markets. By analysing long-memory properties and informational inefficiency through the degree of market inefficiency (D), we demonstrated how betting markets evolve toward greater efficiency over time. Our findings indicate that inefficiency decreases as betting progresses, and markets tend to converge toward a more efficient state.

Furthermore, we introduced the Estimation Score for Integration Orders ($ESIO$), which optimises the window size, bandwidth, and estimator combination for better integration order estimation. This optimisation enhances the precision of market inefficiency assessment.

5.3 Chapter 3: Information Processing and Market Equilibrium

In Chapter 3, we explored the dynamics of information transmission and its impact on market equilibrium. By introducing the Information Affect Factor (IAF) and the Market Equilibrium Condition Factor ($MECF$), we established a mathematical relationship between these two factors. Using Euler's formula and the unit trend (UT) concept, we modelled the dynamic equilibrium process in financial markets.

This chapter also highlighted the limitations imposed by both physical and economic constraints, such as the speed of information transmission and market liquidity. These constraints shape the dynamic process of market equilibrium, leading to over-reactions, subsequent corrections, and eventual convergence to a new equilibrium state.

5.4 Final Remarks

This thesis addresses significant gaps in the modelling and understanding of data changing dynamic in financial time series analysis. Traditional models often struggle with accurately capturing complex market behaviours, particularly in detecting structural breaks, handling cyclical patterns, and modelling dynamic efficiency. Recognising these limitations, the research develops a comprehensive framework, the Dynamic Trend Analysis Approach ($DTAA$), which redefines velocity as the Dynamic Trend (DT) within a financial time series. This novel approach establishes a robust mathematical framework that captures the data changing velocity (dynamic trend vector) and its components—the Trend Component (TC) and the Directional Component (DC). By introducing methods to monitor changes in the slope of the DT vector and distinguishing between Type-I and Type-II structural breaks, the thesis effectively fills the gap in detecting structural breaks and managing cyclical patterns, providing more precise tools for market analysis.

Furthermore, the thesis tackles the gap in assessing dynamic efficiency in betting markets, where traditional methods fall short in capturing long-memory properties and informational inefficiencies. Applying the $DTAA$ to real-world data introduces a novel decomposition model that captures these long-memory properties. The development of the Estimation Score for Integration Orders (ESIO) addresses the challenge of accurately estimating integration orders, and optimising parameters like window size, bandwidth, and estimators. Additionally, constructing a metric D for market inefficiency quantifies the degree of inefficiency over time, revealing the evolving nature of market efficiency. Employing the Fractionally Cointegrated Vector Autoregressive (FCVAR) model bridges the gap in understanding fractional cointegration relationships between DT series of different markets, enhancing predictive capabilities and providing practical insights into market dynamics.

Finally, the thesis integrates information processing and market equilibrium dynamics into the DT framework by incorporating the Information Affect Factor (IAF) and the Market Equilibrium Condition Factor ($MECF$). This integration addresses the gap in modelling how markets dynamically respond to new information and adjust towards equilibrium. Utilising Euler's formula to represent the DT in the complex plane, the research offers a deeper mathematical understanding of market dynamics. It highlights the interplay between information inputs, market liquidity, and inherent limitations in information processing, explaining how markets transition through stages of equilibrium, overreaction, and correction. This comprehensive model enriches theoretical discourse on market behaviour and offers practical tools for both academic research and industry practices, ultimately bridging theoretical concepts with empirical applications and contributing valuable insights to the field of financial time series analysis.

The findings and methodologies presented in this paper give a dynamic understanding of the system. However, there are still many aspects that have not been thoroughly investigated. First, this study assumes that time changes uniformly and consistently. In the next stage of research, I will introduce time variability to further illustrate the physical and economic boundaries of the system's response speed to information. Regarding the four dimensions describing market dynamics, this paper assumes the market as a whole. In future research, I will classify the participants within the market and model them to analyse the dynamic process by which multiple trading entities form a consensus collectively. Additionally, I will consider the dynamic processes in which different market participants undergo mutual transitions as external and internal conditions change.

References

Ryan P. Adams and David J.C. MacKay. Bayesian Online Changepoint Detection. *arXiv preprint arXiv:0710.3742*, 2007.

Cesare Alippi, Giacomo Boracchi, and Manuel Roveri. Hierarchical Change-Detection Tests. *IEEE Transactions on Neural Networks and Learning Systems*, 28(2):246–258, 2016.

Samaneh Aminikhanghahi and Diane J. Cook. A Survey of Methods for Time Series Change Point Detection. *Knowledge and Information Systems*, 51(2):339–367, 2017.

Sylvain Arlot, Alain Celisse, and Zaid Harchaoui. A Kernel Multiple Change-Point Algorithm via Model Selection. *Journal of Machine Learning Research*, 20(162):1–56, 2019.

Jushan Bai and Pierre Perron. Computation and Analysis of Multiple Structural Change Models. *Journal of Applied Econometrics*, 18(1):1–22, 2003.

Roberto S.M. Barros and Silas G.T.C. Santos. A Large-Scale Comparison of Concept Drift Detectors. In *Information Sciences*, volume 451, pages 348–370, 2018.

Daniel Barry and J. A. Hartigan. A Bayesian Analysis for Change Point Problems. *Journal of the American Statistical Association*, 88(421):309–319, 1993.

Albert Bifet and Ricard Gavalda. Learning from Time-Changing Data with Adaptive Windowing. In *Proceedings of the 2007 SIAM International Conference on Data Mining*, pages 443–448, 2007.

Boris Brodsky and Boris Darkhovsky. *Nonparametric Methods in Change-Point Problems*. Kluwer Academic Publishers, 1993.

Anna Carbone, Giuliano Castelli, and H Eugene Stanley. Time-dependent hurst exponent in financial time series. *Physica A: Statistical Mechanics and its Applications*, 344(1-2):267–271, 2004.

Wei-Cheng Chang, Chun-Liang Li, Yiming Yang, and Barnabás Póczos. Kernel change-point detection with auxiliary deep generative models. *arXiv preprint arXiv:1901.06077*, 2019.

Francis X Diebold and Atsushi Inoue. Long memory and regime switching. *Journal of econometrics*, 105(1):131–159, 2001.

Sepideh Dolatabadi, Morten Ørregaard Nielsen, and Ke Xu. A fractionally cointegrated var model with deterministic trends and application to commodity futures markets. *Journal of Empirical Finance*, 38:623–639, 2016.

Sepideh Dolatabadi, Paresh Kumar Narayan, Morten Ørregaard Nielsen, and Ke Xu. Economic significance of commodity return forecasts from the fractionally cointegrated var model. *Journal of Futures Markets*, 38(2):219–242, 2018.

Kun Duan, Zeming Li, Andrew Urquhart, and Jinqiang Ye. Dynamic efficiency and arbitrage potential in bitcoin: A long-memory approach. *International Review of Financial Analysis*, 75: 101725, 2021.

Kun Duan, Yang Gao, Tapas Mishra, and Stephen Satchell. Efficiency dynamics across segmented bitcoin markets: Evidence from a decomposition strategy. *Journal of International Financial Markets, Institutions and Money*, 83:101742, 2023.

Guy Elaad, J James Reade, and Carl Singleton. Information, prices and efficiency in an online betting market. *Finance Research Letters*, 35:101291, 2020.

Eugene F Fama. Efficient capital markets. *Journal of finance*, 25(2):383–417, 1970.

Paul Fearnhead. Exact and Efficient Bayesian Inference for Multiple Changepoint Problems. *Statistics and Computing*, 16(2):203–213, 2006.

Paul Fearnhead and Zhen Liu. On-line Inference for Multiple Changepoint Problems. *Journal of the Royal Statistical Society: Series B*, 69(4):589–605, 2007.

Isvani Frias-Blanco, Jose del Campo-Avila, Gonzalo Ramos-Jimenez, Rafael Morales-Bueno, Agustin Ortiz-Diaz, and Yailé Caballero-Mota. Online and Non-Parametric Drift Detection Methods Based on Hoeffding's Bounds. *IEEE Transactions on Knowledge and Data Engineering*, 27(3):810–823, 2015.

João Gama, Indrė Žliobaitė, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. A Survey on Concept Drift Adaptation. *ACM Computing Surveys*, 46(4):1–37, 2014.

Pierre Granjon. The CuSum Algorithm: A Small Review. Technical report, Gipsa-Lab, Grenoble, France, 2013.

Peter J. Green. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika*, 82(4):711–732, 1995.

Mark Griffiths. Online betting exchanges: A brief overview. *Youth Gambling International*, 5(2): 1–2, 2005.

Niels Haldrup and Morten Ørregaard Nielsen. A regime switching long memory model for electricity prices. *Journal of econometrics*, 135(1-2):349–376, 2006.

Zaid Harchaoui, Eric Moulines, and Francis R. Bach. Kernel Change-Point Analysis. In *Advances in Neural Information Processing Systems*, volume 21, pages 609–616, 2008.

Michael A Hauser and Erhard Reschenhofer. Estimation of the fractionally differencing parameter with the r/s method. *Computational statistics & Data analysis*, 20(5):569–579, 1995.

Douglas M Hawkins and KD Zamba. Statistical process control for shifts in mean or variance using a changepoint formulation. *Technometrics*, 47(2):164–173, 2005.

David V. Hinkley. Inference About the Change-Point in a Sequence of Random Variables. *Biometrika*, 57(1):1–17, 1970.

Ben Jackson, Jeffrey D. Scargle, David Barnes, Sundararajan Arabhi, Alina Alt, Peter Gioumousis, Elyus Gwin, Paungkaew Sangtrakulcharoen, Linda Tan, and Tun Tao Tsai. An Algorithm for Optimal Partitioning of Data on an Interval. *IEEE Signal Processing Letters*, 12(2): 105–108, 2005.

Peter Jones, David Hillier, David Turner, and Daphne Comfort. Customer-to-customer developments in (r) etailing: a case study of betting exchanges in the uk. *Marketing Intelligence & Planning*, 22(7):742–750, 2004.

Jan W Kantelhardt, Eva Koscielny-Bunde, Henio HA Rego, Shlomo Havlin, and Armin Bunde. Detecting long-range correlations with detrended fluctuation analysis. *Physica A: Statistical Mechanics and its Applications*, 295(3-4):441–454, 2001.

Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Shun Chen. LSTM Fully Convolutional Networks for Time Series Classification. *IEEE Access*, 6:1662–1669, 2018.

Nicolas Keriven, Damien Garreau, and Iacopo Poli. Newma: a new method for scalable model-free online change-point detection. *IEEE Transactions on Signal Processing*, 68:3515–3528, 2020.

Rebecca Killick, Paul Fearnhead, and Idris A Eckley. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598, 2012.

Jeremias Knoblauch, Jack Jewson, and Theodoros Damoulas. Doubly Robust Bayesian Inference for Non-Stationary Streaming Data with -Divergences. *Advances in Neural Information Processing Systems*, 31:64–75, 2018.

Ladislav Kristoufek and Miloslav Vosvrda. Commodity futures and market efficiency. *Energy Economics*, 42:50–57, 2014.

Marc Lavielle. Using Penalized Contrasts for the Change-Point Problem. *Signal Processing*, 85 (8):1501–1510, 2005.

Shuang Li, Yao Xie, Hanjun Dai, and Le Song. M-Statistic for Kernel Change-Point Detection. *Advances in Neural Information Processing Systems*, 28:3366–3374, 2015.

Kim Hiang Liow. Long-term memory in volatility: some evidence from international securitized real estate markets. *The Journal of Real Estate Finance and Economics*, 39:415–438, 2009.

Kim Hiang Liow and Haishan Yang. Long-term co-memories and short-run adjustment: securitized real estate and stock markets. *The Journal of Real Estate Finance and Economics*, 31: 283–300, 2005.

Song Liu, Makoto Yamada, Nigel Collier, and Masashi Sugiyama. Change-Point Detection in Time-Series Data by Relative Density-Ratio Estimation. *Neural Networks*, 43:72–83, 2013.

Andrew W Lo. Long-term memory in stock market prices. *Econometrica: Journal of the Econometric Society*, pages 1279–1313, 1991.

Jie Lu, Anjin Liu, Fan Dong, Feng Gu, João Gama, and Guangquan Zhang. Learning under Concept Drift: A Review. *IEEE Transactions on Knowledge and Data Engineering*, 31(12):2346–2363, 2018.

David S. Matteson and Nicholas A. James. A Nonparametric Approach for Multiple Change Point Analysis of Multivariate Data. *Journal of the American Statistical Association*, 109(505): 334–345, 2014.

Joseph McCarthy, Coleen Pantalone, and HC Li. Investigating long memory in yield spreads. *The Journal of Fixed Income*, 19(1):73, 2009.

Walid Mensi, Aviral Kumar Tiwari, and Khamis Hamed Al-Yahyaee. An analysis of the weak form efficiency, multifractality and long memory of global, regional and european stock markets. *The Quarterly Review of Economics and Finance*, 72:168–177, 2019.

Jacob A Mincer and Victor Zarnowitz. The evaluation of economic forecasts. In *Economic forecasts and expectations: Analysis of forecasting behavior and performance*, pages 3–46. NBER, 1969.

Pablo Montero-Manso and Rob J. Hyndman. Principles and Algorithms for Forecasting Groups of Time Series: Locality and Globality. *International Journal of Forecasting*, 37(4):1632–1653, 2021.

Duc Binh Benno Nguyen, Marcel Prokopczuk, and Philipp Sibbertsen. The memory of stock return volatility: Asset pricing implications. *Journal of Financial Markets*, 47:100487, 2020.

Morten Nielsen and Michal K. Popiel. A matlab program and users guide for the fractionally cointegrated var model. Technical Report 1330, QED Working Paper, Queen's University, 2016.

Ewan S Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.

Omri Ross, Stephen E Satchell, and Michael Tehranchi. An equilibrium model of market efficiency with bayesian learning: Explicit modes of convergence to rational expectations equilibrium in the presence of noise traders. *Available at SSRN 2545031*, 2015.

Eric Ruggieri and Marcus Antonellis. An exact approach to bayesian sequential change point detection. *Computational Statistics & Data Analysis*, 97:71–86, 2016.

Yunus Saatçi, Ryan D. Turner, and Carl E. Rasmussen. Gaussian Process Change Point Models. *Proceedings of the 27th International Conference on Machine Learning*, pages 927–934, 2010.

Sebastian Schmidl, Lukas Fleckenstein, and Hendrik Fleischer. Anomaly Detection in Time Series: A Comprehensive Evaluation. *Proceedings of the VLDB Endowment*, 15(9):1779–1797, 2022.

Katsumi Shimotsu. Exact local whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory*, 26(2):501–540, 2010.

Katsumi Shimotsu and Peter CB Phillips. Exact local whittle estimation of fractional integration. 2005.

Michael A Smith, David Paton, and Leighton Vaughan Williams. Market efficiency in person-to-person betting. *Economica*, 73(292):673–689, 2006.

Masashi Sugiyama, Taiji Suzuki, and Takafumi Kanamori. *Density ratio estimation in machine learning*. Cambridge University Press, 2012.

M Sung and Johnnie EV Johnson. Comparing the effectiveness of one-and two-step conditional logit models for predicting outcomes in a speculative market. *The Journal of Prediction Markets*, 1(1):43–59, 2007.

Ming-Chien Sung, Johnnie EV Johnson, and David CJ McDonald. Informed trading, market efficiency and volatility. *Economics Letters*, 149:56–59, 2016.

Charles Truong, Laurent Oudre, and Nicolas Vayatis. Selective Review of Offline Change Point Detection Methods. *Signal Processing*, 167:107299, 2020.

Ghislain Verdier, Nadine Hilgert, and Jean-Pierre Vila. Adaptive Threshold Computation for CUSUM-Type Procedures in Change Detection and Isolation Problems. *Computational Statistics Data Analysis*, 52(9):4161–4174, 2008.

Qingsong Wen, Liang Sun, Fan Yang, Xiaomin Song, Jingkun Gao, Xue Wang, and Huimin Xu. Time Series Data Augmentation for Deep Learning: A Survey. *arXiv preprint arXiv:2002.12478*, 2020.

Xindong Wu, Kui Yu, Wei Ding, Hao Wang, and Xingquan Zhu. Online feature selection with streaming features. *IEEE transactions on pattern analysis and machine intelligence*, 35(5):1178–1192, 2012.

Makoto Yamada, Akisato Kimura, Futoshi Naya, and Hiroshi Sawada. Change-Point Detection with Feature Selection in High-Dimensional Time-Series Data. *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, pages 1827–1833, 2013.

Lei Yu and Huan Liu. Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution. *Journal of Machine Learning Research*, 5:1226–1243, 2012.