Modeling and structural break detection using nonparametric methods

विद्या वाचस्पति की
उपाधि की अपेक्षाओं की आंशिक पूर्ति में प्रस्तुत शोध प्रबंध

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

द्वारा / By
छात्र का नाम / Name: Archi Roy

पंजीकरण सं. / Registration No.: 20192027

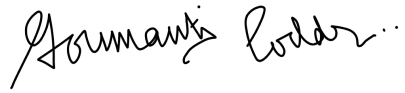
शोध प्रबंध पर्यवेक्षक / Thesis Supervisors: Dr. Moumanti Podder and Dr. Soudeep Deb

भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान पुणे
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE

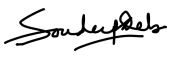
2024

Certificate

It is certified that the work incorporated in the thesis entitled **"*Modeling and structural break detection using nonparametric methods*"**, submitted by **Archi Roy**, was carried out by the candidate, under our supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or Institution.



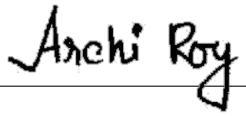
Moumanti Podder
(Supervisor)
Department of Mathematics
Indian Institute of Science Education and Research (IISER), Pune
Date: April 25, 2025



Soudeep Deb
(Co-supervisor)
Decision Sciences Area
Indian Institute of Management (IIM), Bangalore
Date: July 12, 2025

Declaration

I declare that this written submission represents my idea in my own words and where others' ideas have been included, I have adequately cited and referenced the original sources. I declare that I have acknowledged collaborative work and discussions wherever such work has been included. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea, data, fact or source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed. The work reported in this thesis is the original work done by me under the joint guidance of **Dr. Moumanti Podder** (IISER Pune) and **Dr. Soudeep Deb** (IIM Bangalore).

Signature of the Student: 

Date: July 13, 2025

To the women in my life-
the ones who stayed, the ones who did not,
the ones who fought, the ones who could not,
the ones who were in my shoes yesterday,
and the ones who will be tomorrow.

Acknowledgement

I would begin by expressing my deepest gratitude to my supervisors, Dr. Soudeep Deb and Dr. Moumanti Podder, whose unwavering encouragement and support have shaped not only the course of this thesis, but also my growth as a researcher. I am grateful to them for considering me potent to join their research group, and gradually guiding a naive M.Sc. graduate to discover how joyful and rewarding it is to be a statistician. They allowed me the freedom to explore, to stumble, and to live through the cycle of the frustration that numbers don't listen to the relief that numbers don't lie, and to love the process. As their student, I have never been without help whenever I needed it. This thesis is a result of their constant support and mentorship.

I would like to thank my doctoral advisory committee members for their invaluable support throughout my journey. I am grateful to Professor Apratim Guha for his consistent guidance and encouragement during these years of research. I am also especially thankful to Dr. Anindya Goswami, whose encouragement during my M.Sc. days was one of the key reasons I found the courage to pursue a Ph.D. Thanks to the Department of Mathematics at IISER Pune for their support. I am grateful to all my collaborators for generously sharing their knowledge and for the respect and trust they have shown me throughout our work together. I would especially like to thank Dr. Somabha Mukherjee for his contribution to my work on shape- constrained statistics, which forms one of the chapters of this thesis. I am deeply indebted to Professor Itai Dattner- not only for being a collaborator, but also for being an inspiring mentor whose guidance has been instrumental in shaping my understanding of the philosophy of nonparametric statistics. A significant chapter of this thesis is the result of our collaboration.

I would not have reached this point without the constant encouragement of my teachers throughout school and college. In particular, I am deeply grateful to my college professor, Dr. Durba Bhattacharya, whose words of assurance during one of the lowest points in my academic career gave me the strength to continue. This thesis would not have been possible without an exceptional group of peers who supported me at every step, both at IISER Pune and IIM Bangalore. I extend my heartfelt thanks to Anchal, Siddharth, Kapil, Kunal Rai, Amrutha, Reetha, Lizan, Shagun, Satyaki, and Anitha for being my family in Bangalore. I also wish to thank Kunal Arora, Kaustabh, and Dhruv at IISER Pune for their unwavering support and friendship. To my friends in Pune- Debesh Da, Debjyoti Da, Ranita Di, Saikat, and Shreehari- thank you for making this campus feel like home over the past six years. And to Supritha, thank you for being a sister in spirit, for laughing at my worst jokes, and

for joining me in every ridiculous idea without hesitation. In ways both big and small, knowingly or unknowingly, each of them has helped me gather myself during difficult times and return to this thesis with renewed focus and hope.

This thesis has come together as a team effort, made possible by the steadfast support of my family and the unwavering presence of a remarkable partner. I would thank my mother for teaching me to dream without boundaries and for encouraging me to focus on my goals unapologetically. Her presence in my life has possibly been one of my biggest inspirations. I thank my father for being a Superman and protecting me from all-things-unpleasant that would have distracted me from working on this thesis. I guess that is what all fathers are for, mine just is the best of them all. Thanks to Kakima (Rupa Chatterjee) and Kaku (Sahadeb Chatterjee) for their constant encouragement and support. I am forever thankful to Arijit, for not only being an extremely inspiring peer and my best friend, but for being my biggest cheerleader, for bringing me back to me whenever I lost myself to the chaos, for reassuring, for loving, for listening, and for believeing in me and this thesis when I could not.

Lastly, I extend my heartfelt gratitude to the women in statistics and mathematics- those whose names I have read in history books, those who remain unsung, those I have had the privilege to know, and those I have not- for continually paving the way and upholding the presence and power of women in science. Their perseverance and brilliance inspire me every day to contribute my own chapter to this ongoing story.

Abstract

Understanding the behavior of time-evolving systems is central to contemporary data science, with wide-ranging applications in domains such as finance, biology and environmental studies. These systems are often high-dimensional, nonlinear, and governed by complex internal dynamics as well as external shocks. Time series data offer a natural lens through which to understand and model such systems. However, conventional modeling approaches, especially parametric methods, are often inadequate in capturing the nuanced and structurally evolving nature of real-world data. This thesis develops a set of novel nonparametric tools aimed at advancing statistical inference in such dynamic and complex settings.

A primary focus of the thesis is the detection of structural breaks- abrupt changes in the underlying relationship between a pair of time-dependent variables. Classical break detection methods typically rely on strong parametric assumptions and are often limited to changes in mean or variance. In contrast, this work introduces a fully nonparametric, data-driven methodology capable of identifying changes in more intricate functional relationships, including nonlinear and time-varying patterns. The proposed approach employs adaptive smoothing techniques within a minimax optimal framework, ensuring robust performance even under model misspecification and irregular error structures. Simulation studies and real-world applications in economics and finance demonstrate the method's ability to uncover structural changes that remain undetected by conventional techniques.

The second component of the thesis explores the use of shape-constrained nonparametric regression, motivated by the observation that many real-world processes exhibit known structural properties, such as monotonicity, convexity, or quasiconvexity. Incorporating these constraints into a nonparametric framework allows for more interpretable and stable estimation while preserving model flexibility. The proposed estimators are shown to be statistically consistent and particularly effective in settings with limited data or high noise levels.

The thesis also addresses a critical but often overlooked problem: testing the adequacy of ordinary differential equation (ODE) models used to describe dynamic systems. Existing lack-of-fit tests are typically not well suited for ODEs, as they fail to consider the structural constraints imposed by the differential formulation. This work proposes a novel adaptive goodness-of-fit test for both fully and partially observed ODE systems. The test is constructed to directly assess deviations in the trajectories and derivative structures implied by the ODE, offering a rigorous and sensitive

diagnostic tool. The methodology is supported by theoretical results guaranteeing optimal detection rates, and its practical utility is confirmed through simulations and applications in biomedical and agricultural data.

Collectively, the contributions of this thesis advance the field of nonparametric inference for dynamic systems. By developing methodologies that are robust, adaptive, and theoretically grounded, this work addresses key limitations of existing approaches and provides practical tools for analyzing complex, real-world time series. The results have broad applicability across disciplines where understanding structural change, validating mechanistic models, and incorporating qualitative prior knowledge are essential for reliable statistical analysis.

Publications and preprints

1. **Roy, A.**, Soni, A., & Deb, S. (2023). A wavelet-based methodology to compare the impact of pandemic versus Russia–Ukraine conflict on crude oil sector and its interconnectedness with other energy and non-energy markets. Energy Economics, 124, 106830.

2. **Roy, A.**, Deb, S., & Chakarwarti, D. (2024). Impact of COVID-19 on public social life and mental health: A statistical study of google trends data from the USA. Journal of Applied Statistics, 51(3), 581-605.

3. **Roy, A.**, Moumanti Podder, & Soudeep Deb (2024). Nonparametric method of structural break detection in stochastic time series regression model. arXiv preprint arXiv:2410.15713.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In an increasingly data-rich world, many scientific and practical problems revolve around understanding the behavior of systems that evolve over time. These systems, whether economic markets, biological processes, climate phenomena, or engineered infrastructures, are rarely static. They change, adapt, and respond to internal dynamics and external influences. As a result, time series data, which capture observations recorded sequentially over time, provide an essential window into the underlying mechanisms of such dynamic systems. Working with time series data is not only central to forecasting future behavior but also critical to uncovering latent structures, causal pathways, and systemic vulnerabilities. The capacity to model, predict, and intervene in these systems depends on our ability to accurately characterize their temporal evolution.

Many of these systems are complex in nature. Complexity arises due to interdependencies among components, nonlinear feedback loops, emergent behavior, and the presence of noise or uncertainty. A financial market, for example, is influenced by investor sentiment, regulatory interventions, technological shifts, and geopolitical tensions, all interacting in ways that defy simplistic modeling. Similarly, a biological system-like gene regulatory networks or neural activity-features interactions that are often not directly observable, non-additive, and temporally varying. In such settings, the notion of "one-size-fits-all" models breaks down. Rigid frameworks that assume stable relationships or homogeneous effects can lead to misleading inferences and poor

predictions.

Traditional parametric methods, while mathematically elegant and computationally convenient, are often insufficient for analyzing such systems. Parametric models rely on a fixed functional form and a limited number of parameters to describe the data-generating process. While these assumptions make estimation and interpretation tractable, they impose strong constraints that may not align with the true behavior of complex, evolving systems. Misspecified models can lead to biased estimates, underestimation of uncertainty, and incorrect policy or scientific conclusions. Moreover, parametric models often struggle to adapt when the system undergoes regime changes or when new forms of dependence emerge over time.

This motivates the need for nonparametric methods. By allowing the data to inform the model structure with minimal prior assumptions, nonparametric approaches provide the flexibility to capture intricate, nonlinear, and time varying relationships. These methods are particularly well suited for exploratory analysis, model validation, and robust inference in environments where theoretical knowledge is limited or the system behavior is inherently unstable. For example, kernel smoothing and local regression methods have been widely used in analyzing volatility in financial time series, modeling growth curves in biology, and estimating functional relationships in environmental and climate data. Nonparametric techniques can also accommodate irregular sampling, missing data, and heterogeneous variances-common features of real-world time series.

Furthermore, nonparametric models play a critical role in the diagnostics and refinement of more structured models. They can serve as benchmarks against which

parametric models are evaluated, or as tools to reveal features-such as nonlinearity, thresholds, or discontinuities-that parametric forms may fail to capture. In many modern applications, hybrid approaches that combine the interpretability of parametric models with the flexibility of nonparametric methods are becoming increasingly valuable. For instance, in financial econometrics, semiparametric models help quantify risks while remaining responsive to changing market dynamics.

In addition to their modeling flexibility, nonparametric approaches align well with advances in computation and algorithm design. With the growth of high frequency and high dimensional data, there is a growing demand for scalable algorithms that can adapt to data complexity without overfitting. Nonparametric methods, when equipped with techniques such as cross validation, regularization, and adaptive smoothing, offer a principled way to navigate the bias-variance tradeoff and provide interpretable insights in complex settings.

In summary, the study of time series data and complex systems stands at the forefront of modern statistical research due to its relevance across diverse disciplines and its foundational role in understanding dynamic processes. Nonparametric methods provide the necessary tools to model these systems in a flexible, data driven manner, enabling robust inference even under structural shifts, nonlinearities, and unknown dependencies. As the complexity of available data continues to grow, and as systems become increasingly interconnected and dynamic, the importance of such methodological frameworks will only deepen-underscoring their centrality in the future of both theoretical statistics and applied domain research.

Building on this broader motivation, one pressing issue in the analysis of time evolv-

ing systems is the detection of structural breaks-sudden changes in the relationship between predictor and response variables. Classical methods for detecting such breaks typically rely on parametric assumptions, which can lead to misleading conclusions if the underlying relationship is nonlinear or nonstationary. This thesis advances the field by proposing a data driven, nonparametric methodology to identify structural breaks without imposing strong model assumptions. Unlike traditional approaches, which often focus on changes in mean or variance, the proposed framework is capable of detecting breaks in more complex functional relationships. By leveraging smoothing techniques and adaptive estimation methods, the methodology ensures robustness against misspecified models and varying noise structures. The theoretical underpinnings of this approach are developed in a minimax framework, ensuring optimal detection rates even in challenging data environments. Through extensive simulations and real-world applications, this work demonstrates that nonparametric structural break detection provides a more accurate and insightful understanding of dynamic processes in economic and financial systems.

In addition to regression-based models, many scientific phenomena are best described through ordinary differential equations (ODEs), which characterize dynamic processes governed by underlying physical, biological, or economic laws. While ODE models are widely used, their validity is rarely tested rigorously, leading to potential model misspecifications that undermine inference and prediction. Traditional lack-of-fit tests are insufficient in this setting because ODE models encode information not just about observed data points but also about the derivative structure that governs system evolution. This thesis introduces a novel, adaptive goodness-of-fit test for ODE models, bridging the gap between nonparametric hypothesis testing and differential equation

theory. The proposed test assesses whether observed data are consistent with a given ODE model while allowing for partially observed systems where some states remain unmeasured. Unlike previous approaches, which rely on indirect methods such as residual analysis, this framework directly evaluates discrepancies in trajectory, integration, and gradient representations of the system. The methodology is designed to be adaptive, ensuring optimal power across a wide range of alternatives, and is supported by rigorous theoretical results demonstrating its rate-optimality. Simulation studies highlight the test's ability to detect subtle model deviations that would go unnoticed using conventional techniques. Furthermore, real-data applications in biological and financial contexts illustrate its practical utility in diagnosing model misspecifications and improving ODE-based inference.

A complementary theme explored in this thesis is the role of shape constraints in nonparametric regression. While fully unconstrained nonparametric models offer flexibility, they can suffer from overfitting and lack interpretability, especially in cases where prior knowledge about the functional form exists. Many real-world processes exhibit inherent shape characteristics, such as monotonicity, convexity, or unimodality, which should be incorporated into the modeling process to enhance estimation efficiency and predictive accuracy. This thesis develops novel estimation techniques for shape-constrained nonparametric regression, ensuring that inferred relationships remain consistent with domain specific expectations. By integrating shape constraints within a nonparametric framework, the proposed methodology achieves a balance between flexibility and structure, leading to improved generalization in both low and high dimensional settings. Theoretical analysis establishes convergence properties and statistical consistency, while empirical studies confirm the advantages of shape con-

straints in financial risk assessment and macroeconomic trend analysis. By extending existing shape-constrained estimation techniques to dynamic environments, this work provides a valuable tool for modeling complex economic and scientific processes.

By addressing the limitations of existing approaches, our work introduces a novel methodology that unifies lack-of-fit testing across both fully and partially observed ODE systems. Unlike prior studies that focus on specific test statistics tailored to particular representations of the system, we develop a single adaptive test that is applicable to a broad class of ODE models, ensuring robustness against model misspecifications in both individual components and the system as a whole. Our test is constructed within a minimax nonparametric framework, allowing it to achieve optimal detection rates for local alternatives and to remain consistent even when the alternative model is close to the parametric form. Importantly, the method is not restricted to completely observed systems but can be effectively applied to cases where certain state variables are unobserved due to experimental constraints. This adaptability makes our approach particularly relevant for applications in fields like epidemiology, finance, and biological systems, where data incompleteness is a common challenge. Furthermore, our test overcomes the limitations of traditional nonparametric procedures, which are highly sensitive to the smoothness of alternatives, by designing an adaptive framework that automatically adjusts to the underlying complexity of the data generating process. To establish the efficacy of our methodology, we provide a rigorous theoretical analysis, demonstrating that our test achieves comparable or superior power performances relative to existing approaches while maintaining computational efficiency. Our results are further substantiated through an extensive simulation study and real-data applications across diverse domains, high-

lighting the practical significance of our framework. Overall, this thesis contributes to the broader field of statistical modeling by addressing fundamental challenges in detecting structural changes, assessing model adequacy, and incorporating domain knowledge through shape constraints. The methodologies developed herein are not only of theoretical interest but also offer practical insights into real-world problems across diverse disciplines. By pushing the boundaries of nonparametric inference and model validation, this work lays the foundation for more reliable and interpretable statistical modeling in dynamic systems.

Chapter 2

Detection of structural breaks

2.1 Background

Detection of structural breaks is a crucial part of analysis and forecasting of time series data, for which the topic has received much attention in diverse fields starting from finance [1], climate research [2], to cryptography [3] and hydrology [4]. The general problem of structural break detection concerns the inference of a change in distributional characteristics for a set of time-ordered observations. The detection can be sequential [see 5, 6, and relevant references therein] or retrospective [see 7, for a brief review]. The latter stream of literature can be broadly classified into parametric and nonparametric detection procedures. The parametric methodologies assume the underlying data generating process to follow a known distribution or a functional form. For example, [8] developed a method of detecting changes in time series data assuming it to be piece-wise autoregressive (AR) in nature. [9] also worked under the same assumption with weak dependence between the pieces, and developed a test based on the quasi-likelihood ratio of the process in a small scanning window before and after the potential break-point. [10] developed a test for structural breaks in a linear regression model with autoregressive moving average (ARMA) residuals using a Wald test statistic. On the other hand, sequential Monte Carlo methods have been developed [11, 12] for detecting breaks in generalized autoregressive conditional heteroskedastic (GARCH) models and stochastic volatility models, both of which are common techniques to deal with financial datasets. A major issue with such para-

metric procedures is their specific assumptions about the structure of the underlying process, which may be impractical in real-life applications. An interesting development in this context was recently published by [13], who investigated the theoretical properties of different extensions of moving sum procedures in detection of structural breaks. Although their test statistic is based on a given parametric model, the framework allows for model misspecification in the theoretical analysis. Interestingly, their work was only concerned with detection of changes in the mean levels of a piece-wise stationary time series.

Parallelly, there have been multiple attempts in the extant literature to detect structural breaks avoiding the imposition of any parametric structure on the concerned time series. For example, [14] detected breaks in the path of a time series by looking at the f-divergence measure between the likelihood of the data in consecutive scanning windows in either side of potential break-points. [15] extended the work of [16] to a time series setup by proposing a cost function based on an initial segmentation of the data where the position of breaks are determined as the solution of the optimal segmentation that maximizes the log-likelihood obtained from the empirical distribution function. Similar works in the same direction were done by [17, 18]. In parallel, [19] proposed a self-normalized test procedure based on a cumulative sum (CUSUM) type test statistic to detect breaks in mean or other distributional properties of the time series, and [20] proposed a test of detection of structural breaks in the covariance structure of multivariate time series utilizing the Euclidean difference in the spectral density matrices. More recently, [21] proposed a methodology for detecting structural breaks in distributional properties of a time series with the help of the empirical distribution functions, and [22] proposed a nonparametric algorithm for

detecting structural breaks in the spectral density of a locally stationary time series.

A common approach in this stream of literature is to detect changes in the mean response of a time-evolving variable by assuming the mean to be constant between consecutive breaks. These works often focus on optimizing a well-defined loss function obtained from evaluating a certain quantity of interest in two segments of the time series sample [see 23, 24, among others]. Another available approach is to identify structural breaks through the maximization of a loss function computed based on the regression curves in the different segments of the data. In this technique, rather than assuming the mean response to be constant between two consecutive breaks, it is considered to be an arbitrary smooth curve between breaks. Our proposed methodology aligns with this setup in particular. Many interesting developments in this direction have been done using a fixed design model $Y_i = f(x_i) + \epsilon_i$, where Y_i's are the observed response, x_i's are deterministic terms, ϵ_i's are independent model errors and f is an unknown smooth function. For example, [25] used local linear smoothing to estimate the piece-wise regression curve f assuming different numbers and locations of structural breaks. The optimal positions of the structural breaks are chosen as the ones which minimize the jump information criteria for the model. More recently, [26] considered a similar setup with stochastic covariates, but the theory was developed under independent data assumption. In the domain of time series regression, [27] worked with the model

$$Y_t = \mu(X_t) + \epsilon_t, \tag{2.1}$$

and tested whether the shape of the mean regression function $\mu(.)$ stays the same for

all time-points t in the sample space. The test statistic proposed in this chapter is based on the kernel-based Euclidean distance between all possible pairs $\mu(u)$ and $\mu(v)$, for $u, v \in \mathbb{R}$. [28] detected structural breaks under the same model using a CUSUM type statistic assuming the process $\{X_t\}$ to be α-mixing. In another pertinent work, [29] considered a modified version of (2.1), with the conditional mean being a smooth time-varying function $g_t(.)$. Their test was developed utilizing the Fourier transform of Y_t using an instrumental variable to infer if g_t is time invariant. More recently, [30] considered detection of breaks in a special case of the model (2.1) with $X_t = Y_{t-1}$, but with more general assumptions on the true properties of the $\mu(.)$ function.

There have been comparatively fewer works in the time series location-scale model

$$Y_t = \mu(X_t) + \sigma(X_t)\epsilon_t, \tag{2.2}$$

which is an extension of (2.1). In an early study, [31] proposed detecting structural breaks in terms of the squared differences of the right and left-hand limits of the $\mu(.)$ and $\sigma(.)$ functions estimated by local linear estimator. They assumed $\{X_t\}$ to be α-mixing. [32] considered the same model to estimate structural breaks in the first derivative i.e. the slope of the mean regression function $\mu(\cdot)$, by using an extension of the traditional zero-crossing technique developed by [33]. Their setup allowed for long-range dependence in $\{X_t\}$.

Our focus in this chapter is on a similar structure of stochastic regression model, under which we develop a novel structural break detection framework where a change in the response time series $\{Y_t\}$ is identified by a statistically significant change in the global functional behavior of the mean regression function $\mu(\cdot)$, or the conditional behavior

function $\sigma(\cdot)$, or both. Such approach allows us to detect large scale structural shifts in the data and is more robust to short-term anomalies. We make mild assumptions on the properties of these functions and allow for both short and long range dependence in the covariate. Further, our method does not require any knowledge on the number of breaks in the model. As illustrated in Section 2.4, despite making mild assumptions, our proposed procedure is not only computationally less extensive, but is also superior in terms of performance for heavy-tailed financial data.

The rest of the chapter is organized in the following way. In Section 2.2, we present the mathematical framework of the problem, while Section 2.3 contains the proposed methodology of structural break detection and the relevant asymptotic theory. We assess the empirical performance of the proposed methodology for various time series structures, along with moderate to heavy-tailed residual distributions, and present the findings in Section 2.4. Next, Section 4.4 contains an application of the proposed method to cryptocurrency data. We conclude with some necessary and important remarks in Section 4.5.

2.2 Mathematical framework

Let us start with setting some notations. Throughout this chapter, $E(\cdot)$ and $V(\cdot)$ are used to indicate the expectation and variance of a random variable. For any matrix A, $A_{i,j}$ represents the $(i,j)^{th}$ element. The p^{th} norm will be indicated by $\|.\|_p$, while $\mathcal{L}^p$ is used for the space of all random variables with finite p^{th} norm. We shall use $\xrightarrow{P}$ for convergence in probability and $\xrightarrow{d}$ for convergence in distribution. For any set $S \subset \mathbb{R}$, denote by $\mathcal{C}^p(S)$ the space of functions with the q^{th} derivative bounded on S for all integers $q \leqslant p$, and let $S(\delta) = \bigcup_{\omega \in S}\{x \mid |x - \omega| \leqslant \delta\}$ denote the δ-neighborhood of

S. Throughout the section we denote the filtration generating $\{X_t\}$ as $\{\mathcal{F}_t\}$.

For our main analysis, following many existing works in the structural breaks literature [e.g., 27], we scale the index set of the time domain and restrict it to the interval $[0, 1]$, henceforth denoted as $\mathcal{T}$. We consider the stochastic regression model

$$Y_t = \mu(X_t) + \sigma(X_t)\epsilon_t, \quad \text{for } t \in \mathcal{T}, \tag{2.3}$$

where $\{Y_t\}$ and $\{X_t\}$ are two real-valued time series, $\mu : \mathbb{R} \to \mathbb{R}$ is the conditional mean regression function, $\sigma^2 : \mathbb{R} \to \mathbb{R}^+$ is the conditional variance function and $\{\epsilon_t\}$ is a real-valued independently and identically distributed (iid) random noise process with unit variance. At this stage, it is important to highlight that in all our theoretical derivations, X_t is going to be assumed univariate for convenience, but the proposed methodology and the results will work even if X_t is a vector-valued process of finite dimension. We assume the following about the error and covariate processes.

Assumption 1. *Error process $\{\epsilon_t\}$ is independent of $\{X_t\}$, and has finite fourth moments.*

Assumption 2. *The time series $\{X_t\}$ is stationary and is generated by a sequence of iid random variables $\{\eta_t\}$ i.e. $X_t = m(\mathcal{F}_t)$ where $m(.)$ is a measurable function and $\{\eta_t\}$ is $\{\mathcal{F}_t\}$-adapted. Let f_X be the density of the $\{X_t\}$ process. The temporal dependence structure of $\{X_t\}$ is expressed through the quantity Ξ_k, which, for any $k \in \mathbb{N}$, is defined as*

$$\Xi_k = k\Theta_{2k}^2 + \sum_{r=k}^{\infty} (\Theta_{k+r} - \Theta_r)^2, \quad \Theta_k = \sum_{i=1}^{k} \theta_i,$$

where $\theta_i = \sup_{x\in\mathbb{R}} ||P_0\left(f_X(x \mid \mathcal{F}_{i-1})\right)|| + \sup_{x\in\mathbb{R}} ||P_0\left(f'_X(x \mid \mathcal{F}_{i-1})\right)||$, *with* P_ℓ *for any* $l \in \mathbb{Z}$ *being a projection operator defined as* $P_\ell(Z) = E\left(Z \mid \mathcal{H}_\ell\right) - E\left(Z \mid \mathcal{H}_{\ell-1}\right)$ *for any* $\{\mathcal{H}_\ell\}$*-adapted random variable* $Z \in \mathcal{L}^1$.

Hereafter, the joint filtration generated by $\{\eta_t, \epsilon_t\}$ is denoted by $\mathcal{G}_t$. It should be noted that the term θ_i is a rough quantification of the contribution of η_0 in predicting X_i. Smaller values of θ_i would denote lesser dependence of X_i on the previous observations of the process. Following this argument, for a sample size n, we say that $\{X_t\}$ is a short-range dependent (SRD) process if $\Theta_n < \infty$; otherwise we say that $\{X_t\}$ is a long-range dependent (LRD) process. Note that our proposed framework allows for both short and long–range dependence in the covariate X. It can be shown that for SRD processes, $\Xi_n = O(n)$, whereas for LRD processes the rate of decay of Ξ_n is much slower. For example, if θ_i is of the form $l(i)/i^\beta$, where $\beta > 1/2$ and $l(.)$ is a slowly varying function, then we can write $\Xi_n = O\left(n^{3-2\beta}l^2(n)\right)$ or $\Xi_n = O\left(n\bar{l}^2(n)\right)$ where $\bar{l}(n) = \sum_{i=1}^n |l(i)|/i$. This follows from a straightforward application of Karamata's theorem. Interested readers may refer to [34].

Assumption 3. *For some* $\delta > 0$, *each of* $f_X(\cdot)$, $\mu(\cdot)$ *and* $\sigma(\cdot)$ *is a four-times differentiable function in* $\mathcal{X}(\delta)$, *where* $\mathcal{X}(\delta)$ *denotes a* δ*-neighborhood of* $\mathcal{X}$. *Also,* $\inf_{x\in\mathcal{X}}\{f_X(x)\} > 0$ *and* $\inf_{x\in\mathcal{X}}\{\sigma(x)\} > 0$.

It is important to highlight that, under Assumption 3, the modeling framework in (2.3) covers a fairly large class of traditional time series models. For example, if $X_t = Y_{t-1}$ and $\sigma(\cdot)$ is a constant function, we get the class of nonlinear AR models. As a special case, if we further put $\mu(x) = ax$ for some constant $a \in \mathbb{R}$, we obtain

the linear AR process. Similarly, by setting $\mu(x) = a\max\{x, 0\} + b\min\{x, 0\}$ or $\mu(x) = a + be^{-cx^2}$, where $a, b, c \in \mathbb{R}$, we obtain the threshold AR and exponential AR processes, respectively. One can also obtain heavy-tailed data generating processes from (2.3), for example $\mu(x) = 0$ and $\sigma(x) = \sqrt{a + bx^2}$, gives the traditional ARCH process. On the other hand, letting $Y_t = X_{t+1} - X_t$ and assuming $\{\epsilon_t\}$ to be iid Gaussian, we obtain the continuous time stochastic diffusion model:

$$dX_t = \mu(X_t) + \sigma(X_t)dW_t,$$

where $\{W_t\}$ is a standard Brownian motion. As pointed out by [35], this structure covers a wide class of financial models. Further, Assumption 2 is in line with several linear and nonlinear time series processes [36, 37].

2.3 Theory

In the stochastic regression model specified by (2.3), we assume that the data is observed over the time span $\mathcal{T}$ with increasing sampling frequency. We use $\{t_1, t_2, \ldots, t_n\}$ for the time-points at which the sample observations are obtained, and for convenience of notations, this set will also be denoted as $\mathcal{T}$. The cardinality of the sample is denoted as n, and the asymptotic theory will be derived for $n \to \infty$. Further, denote by (Y_i, X_i) the paired observations recorded at time-point $t_i \in \mathcal{T}$. Hereafter, we shall use the notation $\sum_{\mathcal{T}} \pi_i$ to denote the sum of the values π_i recorded at time-points $\{t_i\}$.

The main theory proposed in this work is concerned with the detection of structural

break in the at-most-one-change setup. Specifically we consider the model

$$Y_t = \mu_t(X_t) + \sigma_t(X_t)\epsilon_t,$$

where the changepoint testing problem can be formally stated as

$$H_0 : g_t(x) = g(x) \quad \forall x \in \mathcal{X} \text{ vs } H_1 : \quad \exists \tau \text{ such that } g_t(x) = \begin{cases} g_1(x), & t \leq \tau \\ g_2(x), & t > \tau \end{cases}$$

where $g \equiv \mu$ or σ. As we will show in the subsequent sections, the localization for multiple structural breaks can be done using a simple binary segmentation type procedure. To that end, our main objective is to develop a test-based procedure to detect a break in the functional characteristic $g(x) \coloneqq g(Y_t \mid X_t = x)$ of the data. In particular, we shall focus on the case where g is either the conditional expectation or the conditional variance. As a first step to develop the test, we divide the time domain $\mathcal{T}$ into two disjoint parts, denoted by $\mathcal{T}_-$ and $\mathcal{T}_+$ respectively. Due to the assumption of increasing sampling frequency, the cardinality of both $\mathcal{T}_-$ and $\mathcal{T}_+$ will approach ∞. With a slight abuse of notation, we continue to denote the number of observations in each set by n. Assume the true functional characteristic $g(\cdot)$ to be $g_1(\cdot)$ on $\mathcal{T}_-$ and $g_2(\cdot)$ in $\mathcal{T}_+$. In the absence of a structural break, we should have $g_1(x) = g_2(x)$ for all $x \in \mathcal{X}$. However, if a structural break is indeed present, then $g_1(\cdot)$ and $g_2(\cdot)$ will exhibit significant difference over the range $\mathcal{X}$. For a fixed $x \in \mathcal{X}$, let

$$g_{\text{diff}}(x) = g_1(x) - g_2(x),$$

which, in the presence of a structural break, should take a large value for some

$x \in \mathcal{X}$. Therefore, under the assumption that the signal is not symmetric around the change, a significantly large value of $\sup_{x\in\mathcal{X}}\{|g_{\text{diff}}(x)|\}$ indicates the presence of a structural break. Our procedure relies on this phenomenon. In Sections 2.3.1 and 2.3.2, we provide the estimates of both $g_{\text{diff}}(\cdot)$ and $\sup_{x\in\mathcal{X}}\{|g_{\text{diff}}(x)|\}$ (taking g to be conditional mean function or the conditional variance function), and establish their asymptotic theory.

2.3.1 Asymptotic theory for point-wise estimates

In this subsection, we derive the asymptotic properties of $g_{\text{diff}}(x)$ for a fixed $x \in \mathcal{X}$, where g is the conditional mean function $\mu(\cdot)$, or the conditional variance function $\sigma^2(\cdot)$. Let us use $\mu_1(\cdot)$ and $\mu_2(\cdot)$ to denote the mean function in the segments $\mathcal{T}_-$ and $\mathcal{T}_+$, respectively. For a fixed $x \in \mathcal{X}$, the point-wise disparity between these two functions is given by

$$\mu_{\text{diff}}(x) = \mu_1(x) - \mu_2(x).$$

We estimate this with a Nadaraya-Watson type estimator $\widehat{\mu}_{\text{diff}}(x) = \widehat{\mu}_1(x) - \widehat{\mu}_2(x)$, with

$$\widehat{\mu}_1(x) = \frac{1}{nb_n\widehat{f}_X(x)}\sum_{\mathcal{T}_-} Y_t K\left(\frac{x - X_t}{b_n}\right), \widehat{\mu}_2(x) = \frac{1}{nb_n\widehat{f}_X(x)}\sum_{\mathcal{T}_+} Y_t K\left(\frac{x - X_t}{b_n}\right),$$

where $\widehat{f}_X(x) = (nb_n)^{-1}\sum_{\mathcal{T}} K\left((x - X_t)/b_n\right)$ is the estimated density of the covariate $\{X_t\}$, $K(\cdot)$ is an appropriately chosen kernel function, and $b_n = b(n)$ is a bandwidth sequence satisfying the following assumption.

Assumption 4. *The kernel function $K(.)$ is symmetric, bounded, has bounded deriva-*

tive and bounded support $[-1, 1]$. *The bandwidth sequence* $b_n = b(n)$ *satisfies* $b_n \to 0$ *and* $nb_n \to \infty$.

Akin to above, denote the conditional variance function $\sigma^2(\cdot)$ in the segments $\mathcal{T}_-$ and $\mathcal{T}_+$ as $\sigma_1^2(\cdot)$ and $\sigma_2^2(\cdot)$ respectively. For a fixed $x \in \mathcal{X}$, the point-wise difference $\sigma^2_{\text{diff}}(x) = \sigma_1^2(x) - \sigma_2^2(x)$ is estimated as

$$\widehat{\sigma}^2_{\text{diff}}(x) = \widehat{\sigma}_1^2(x) - \widehat{\sigma}_2^2(x), \tag{2.4}$$

where

$$\widehat{\sigma}_1^2(x) = \frac{1}{nb_n \widehat{f}_X(x)} \sum_{\mathcal{T}_-} (Y_t - \widehat{\mu}_1(X_t))^2 K\left(\frac{x - X_t}{b_n}\right),$$
$$\widehat{\sigma}_2^2(x) = \frac{1}{nb_n \widehat{f}_X(x)} \sum_{\mathcal{T}_+} (Y_t - \widehat{\mu}_2(X_t))^2 K\left(\frac{x - X_t}{b_n}\right).$$

Note that we have used the same bandwidth b_n for the estimation of both the conditional mean and variance functions. One can also use a different bandwidth sequence h_n for the estimation of variance. It does not affect the theoretical results provided that $\lim_{n\to\infty} |h_n/b_n|$ is bounded away from 0 and ∞. For the kernel function $K(\cdot)$, define $\phi(K) = \int_{\mathbb{R}} (K(u))^2 du$ and $\psi(K) = \int_{\mathbb{R}} (u^2/2) K(u) du$. The following results describe the asymptotic point-wise behavior of $\widehat{\mu}_{\text{diff}}(\cdot)$, under specific conditions on whether the variance function should be assumed to be same for the entire time horizon or not.

In order to establish the properties of the test statistic, we start by proving a few prerequisite lemmas. Hereafter, we use the notation $\sum_{\mathcal{R}}$ to indicate that the intended

result for the summation term holds for both $\mathcal{R} = \mathcal{T}_+$ and $\mathcal{R} = \mathcal{T}_-$. For notational convenience throughout this section, we shall use $K_{b_n}(u)$ to denote the term $K(b_n^{-1}u)$. The following theorem establishes the pointwise asymptotic normality of the test statistic for detecting changepoint in mean structure. The proofs will largely follow from few prerequisite lemmas, which we state and prove in the Appendix.

Theorem 2.1. *Along with the previously stated assumptions in Section 2.2, assume that the conditional variance function remains the same throughout the time domain $\mathcal{T}$, i.e., $\sigma_1(x) = \sigma_2(x) = \sigma(x)$ for all x, and suppose the bandwidth is chosen such that*

$$nb_n^9 + \frac{1}{nb_n} + \Xi_n\left(\frac{b_n^3}{n} + \frac{1}{n^2}\right) \xrightarrow{n\to\infty} 0.$$

Fix an $x \in \mathcal{X}$ such that $f_X(x) > 0, \sigma(x) > 0$ and $f_X, \mu \in \mathcal{C}^4(x-\delta, x+\delta)$ for some $\delta > 0$. Then, as $n \to \infty$,

$$\frac{\sqrt{nb_n\widehat{f}_X(x)}}{\sqrt{2\phi(K)\widehat{\sigma}^2(x)}}\left[\widehat{\mu}_{diff}(x) - (\mu_1(x) - \mu_2(x)) - \left(b_n^2\psi(K)\rho_{\mu_1}(x) - b_n^2\psi(K)\rho_{\mu_2}(x)\right)\right] \xrightarrow{d} \mathcal{N}(0,1),$$

where $\widehat{\sigma}^2(x)$ is a consistent nonparametric estimate of the common conditional variance function, and the asymptotic bias of the estimate is defined through

$$\rho_{\mu_1}(x) = \mu_1''(x) + 2\mu_1'(x)\frac{f_X'(x)}{f_X(x)}, \quad \rho_{\mu_2}(x) = \mu_2''(x) + 2\mu_2'(x)\frac{f_X'(x)}{f_X(x)}.$$

Proof. Define

$$\widehat{w}_n^{(1)}(x) = \frac{f_X(x)}{\widehat{f}_X^{(1)}(x)} \quad \text{and} \quad \widehat{w}_n^{(2)}(x) = \frac{f_X(x)}{\widehat{f}_X^{(2)}(x)}.$$

Lemma 2.1 can be utilized to prove that

$$\widehat{w}_n^{(i)}(x) = 1 + O_p\left(\frac{1}{\sqrt{nb_n}} + b_n^2 + \frac{\Xi_n^{\frac{1}{2}}}{n}\right), i = 1, 2.$$

Further, defining

$$\mathcal{U}_n^{(1)}(x) = \frac{1}{nb_n f_X(x)} \sum_{\mathcal{T}_-} (\mu_1(X_t) - \mu_1(x)) K_{b_n}(x - X_t),$$
$$\mathcal{U}_n^{(2)}(x) = \frac{1}{nb_n f_X(x)} \sum_{\mathcal{T}_+} (\mu_2(X_t) - \mu_2(x)) K_{b_n}(x - X_t),$$

it can be shown that, for $i = 1, 2$,

$$\mathcal{U}_n^{(i)}(x) = b_n^2 \psi(K) \rho_{\mu_1}(x) + O_p\left(\sqrt{\frac{b_n}{n}} + b_n^2 + \frac{\sqrt{\Xi_n} b_n}{n}\right).$$

Therefore, under bandwidth conditions similar to those specified in Lemma 2.2, we have

$$(\widehat{\mu}_1(x) - \widehat{\mu}_2(x)) - (\mu_1(x) - \mu_2(x)) - \left(b_n^2 \psi(K) \rho_{\mu_1}(x) - b_n^2 \psi(K) \rho_{\mu_2}(x)\right) = \mathcal{V}_{\mathcal{T}}(x)$$

where

$$\mathcal{V}_{\mathcal{T}}(x) = \frac{1}{nb_n f_X(x)} \sum_{\mathcal{T}_-} \sigma(X_t) \epsilon_t K_{b_n}(x - X_t) - \frac{1}{nb_n f_X(x)} \sum_{\mathcal{T}_+} \sigma(X_t) \epsilon_t K_{b_n}(x - X_t).$$

Theorem 2.1 can then be proved using Lemma 2.2 with specific forms of the functions f_1 and f_2. □

It is imperative to point out that, in practical applications, the true conditional mean functions $\mu_1(.)$ and $\mu_2(.)$ are not known. Hence, the bias terms involving $\rho_{\mu_1}(\cdot)$ and $\rho_{\mu_2}(\cdot)$ need to be estimated as well. Following [38], we avoid this by utilizing a jackknife correction, which is equivalent to estimating $\widehat{\mu}_1(\cdot)$ and $\widehat{\mu}_2(\cdot)$ using the kernel $K^*(u) = 2K(u) - K(u/\sqrt{2})/\sqrt{2}$. Note that K^* has support $[-\sqrt{2}, \sqrt{2}]$. Hereafter, the bias-corrected estimates will be denoted with $*$ above them, e.g., $\widehat{\mu}_1^*(\cdot)$.

Corollary 2.1. *Under the assumptions stated in Theorem 2.1,*

$$\frac{\sqrt{nb_n\widehat{f}_X(x)}}{\sqrt{2\phi(K)\widehat{\sigma}^2(x)}}\left[\widehat{\mu}^*_{diff}(x) - (\mu_1(x) - \mu_2(x))\right] \xrightarrow{d} \mathcal{N}(0,1)\,,$$

where $\widehat{\mu}^*_{diff}(x) = \widehat{\mu}_1^*(x) - \widehat{\mu}_2^*(x)$ is the estimate of $\mu_{diff}(x)$ using the kernel K^*.

Corollary 2.2. *Assume that the conditional variance function may have different behavior in $\mathcal{T}_-$ and $\mathcal{T}_+$, and is estimated separately in the two segments as $\widehat{\sigma}_1^2(x)$ and $\widehat{\sigma}_2^2(x)$. Then, letting $\widehat{S}(x) = \widehat{\sigma}_1^2(x) + \widehat{\sigma}_2^2(x)$, under the same conditions specified in Theorem 2.1, for a fixed $x \in \mathcal{X}$ as $n \to \infty$,*

$$\frac{\sqrt{nb_n\widehat{f}_X(x)}}{\sqrt{\phi(K)\widehat{S}(x)}}\left[\widehat{\mu}^*_{diff}(x) - (\mu_1(x) - \mu_2(x))\right] \xrightarrow{d} \mathcal{N}(0,1)\,.$$

The proof of Corollary 2.1 is straightforward by noting that $\psi(K^*) = 0$ for the modified kernel function. In the following corollary, we establish the point-wise behavior of the $\widehat{\mu}^*_{\text{diff}}(.)$ function under the setup where we allow for the possibility of a structural break in the conditional variance $\sigma(\cdot)$, i.e., if $\sigma_1(x) \neq \sigma_2(x)$ for at least some $x \in \mathcal{X}$. . The proof of Theorem 2.1 involves expressing $\widehat{\mu}_{\text{diff}}(x)$ as a martingale difference sequence with respect to the filtration $\{\mathcal{F}_t\}$, followed by a straightforward application

of martingale central limit theorem. The proof of Corollary 2.2 follows along similar lines. .

As previously discussed, we similarly evaluate the disparity between the conditional variance functions in $\mathcal{T}_-$ and $\mathcal{T}_+$ using the expression $\widehat{\sigma}^2_{\text{diff}}(x) = \widehat{\sigma}^2_1(x) - \widehat{\sigma}^2_2(x)$. In this formulation, each of the conditional variance estimates $\widehat{\sigma}^2_i(\cdot)$ depends on the respective conditional mean estimates $\widehat{\mu}_i(\cdot)$, for $i = 1, 2$. However, they inherently possess a bias of order $O(b_n^4)$. Although the asymptotic distribution of $\widehat{\sigma}^2_{\text{diff}}(x)$ can be derived using this bias, for improved performance of the test statistic, we incorporate the bias-corrected mean estimates $\widehat{\mu}^*_i(\cdot)$ in the estimation of the conditional variances. The revised estimates are therefore given as,

$$\widehat{\sigma}^2_1(x) = \frac{1}{nb_n\widehat{f}_X(x)} \sum_{\mathcal{T}_-} (Y_t - \widehat{\mu}^*_1(X_t))^2 K\left(\frac{x - X_t}{b_n}\right),$$
$$\widehat{\sigma}^2_2(x) = \frac{1}{nb_n\widehat{f}_X(x)} \sum_{\mathcal{T}_+} (Y_t - \widehat{\mu}^*_2(X_t))^2 K\left(\frac{x - X_t}{b_n}\right).$$

Furthermore, it can be shown that both the above estimates are biased, with the bias being of the order $O(b_n^2)$. We therefore follow the jackknife type correction once again and use the modified kernel K^* to obtain bias-corrected estimates $\widehat{\sigma}^{*^2}_i(x)$, for $i = 1, 2$, and define

$$\widehat{\sigma}^{*^2}_{\text{diff}}(x) = \widehat{\sigma}^{*^2}_1(x) - \widehat{\sigma}^{*^2}_2(x).$$

We next derive the point-wise asymptotic distribution of the $\widehat{\sigma}^{*^2}_{\text{diff}}(.)$ function.

Theorem 2.2. *Along with the previously stated assumptions in Section 2.2, consider*

the bandwidth condition

$$b_n^{\frac{3}{2}} \log n + \frac{1}{n^2 b_n^5} + \frac{\Xi_n}{n^2} \xrightarrow{n\to\infty} 0.$$

Fix an $x \in \mathcal{X}$ *such that* $f_X(x) > 0, \sigma(x) > 0$ *and* $f_X, \mu, \sigma \in \mathcal{C}^4(x-\delta, x+\delta)$ *for some* $\delta > 0$. *If* $\nu_\epsilon = E(\epsilon_0^4) - 1$ *and* $\widehat{S}(x)$ *is as defined in Corollary 2.2, then as* $n \to \infty$,

$$\frac{\sqrt{nb_n \widehat{f}_X(x)}}{\nu_\epsilon \sqrt{\phi(K^*)\widehat{S}(x)}} \left[\widehat{\sigma}^{*^2}_{diff}(x) - \left(\sigma_1^2(x) - \sigma_2^2(x)\right)\right] \xrightarrow{d} \mathcal{N}(0,1).$$

Proof. Define

$$W_n^{(1)}(x) = \frac{1}{nb_n f_X(x)} \sum_{\mathcal{T}_-} \sigma_1(X_t)\epsilon_t K_{b_n}(x - X_t),$$
$$W_n^{(2)}(x) = \frac{1}{nb_n f_X(x)} \sum_{\mathcal{T}_+} \sigma_2(X_t)\epsilon_t K_{b_n}(x - X_t).$$

One can similarly define $W_n^{*^{(1)}}(x)$ and $W_n^{*^{(2)}}(x)$ by replacing K by K^* in the above equations. Further, define

$$r_n = \sqrt{\frac{b_n \log n}{n}} + b_n^4 + \frac{\sqrt{b_n \Xi_n}}{n}, q_n = \sqrt{\frac{\log n}{nb_n}} + b_n^2 + \frac{\sqrt{\Xi_n}}{n}, \chi_n = \sqrt{\frac{\log n}{nb_n}} + \frac{\log n}{\sqrt{n^3 b_n^5}},$$

and denote $\Delta_n = r_n + q_n(b_n^2 + \chi_n)$. Simple calculations show that

$$\widehat{\mu}^*(x) - \mu(x) = \begin{cases} W_n^{*(1)}(x) + O_p(\Delta_n) & \text{in } \mathcal{T}_-, \\ W_n^{*(2)}(x) + O_p(\Delta_n) & \text{in } \mathcal{T}_+. \end{cases}$$

Assuming $\Delta_n = o\left((nb_n)^{-\frac{1}{2}}\right)$, we have

$$\Delta_n \sum_{\mathcal{T}_-} \left(Y_t - \widehat{\mu}^*(X_t)\right) K_{b_n}(x - X_t) \xrightarrow{P} 0.$$

Then, we can write

$$\widehat{\sigma}_1^2(x) = \frac{1}{nb_n \widehat{f}_X(x)} \sum_{\mathcal{T}_-} \left(\sigma_1(X_t)\epsilon_t - W_n^{*(1)}(X_t) + O_p(\Delta_n)\right)^2 K_{b_n}(x - X_t). \tag{2.5}$$

Since $\sup_{x \in \mathcal{T}_-} \left\{\left|W_n^{*(1)}\right|(x)\right\} = O_p(\chi_n)$, it can be shown that

$$\begin{aligned}\sum_{\mathcal{T}_-} \left(\sigma_1(X_t)\epsilon_t - W_n^{*(1)}(X_t) + O_p(\Delta_n)\right)^2 K_{b_n}(x - X_t) &= \sum_{\mathcal{T}_-} \left(\sigma_1(X_t)\epsilon_t\right)^2 K_{b_n}(x - X_t) \\ + O_p(\Delta_n) \sum_{\mathcal{T}_-} \sigma_1(X_t)\epsilon_t K_{b_n}(x - X_t) &- 2 \sum_{\mathcal{T}_-} \sigma_1(X_t)\epsilon_t W_n^{*(1)}(X_t) K_{b_n}(x - X_t).\end{aligned} \tag{2.6}$$

Combining (2.5) and (2.6) and taking absolute value on both sides, we can write

$$\widehat{\sigma}_1^2(x) \leqslant \frac{T_{\mathcal{T}_-}(x)}{nb_n \widehat{f}_X(x)} + \frac{2\left|L_{\mathcal{T}_-}(x)\right| + O_p(\Delta_n) J_{\mathcal{T}_-}(x)}{nb_n \widehat{f}_X(x)}.$$

where

$$\begin{aligned}T_{\mathcal{T}_-}(x) &= \sum_{\mathcal{T}_-} \left(\sigma_1(X_t)\epsilon_t\right)^2 K_{b_n}(x - X_t), \\ L_{\mathcal{T}_-}(x) &= \sum_{\mathcal{T}_-} \sigma_1(X_t)\epsilon_t W_n^{*(1)}(X_t) K_{b_n}(x - X_t),\end{aligned}$$

$$J_{\mathcal{T}_-}(x) = \sum_{\mathcal{T}_-} \sigma_1(X_t)\,|\epsilon_t|\,K_{b_n}(x - X_t).$$

Since $J_{\mathcal{T}_-}(x) = O_p\left(nb_n\left(1 + q_n + \chi_n\right)\right)$ and $\sup_{x\in\mathcal{T}_-}\left\{\left|L_{\mathcal{T}_-}(x)\right|\right\} = O_p\left(b_n^{-\frac{3}{2}}\right)$, under pre-specified bandwidth conditions we can write, $\widehat{\sigma}_1^2(x) = (nb_n\widehat{f}_X(x))^{-1}T_{\mathcal{T}_-}(x)$. Defining

$$D_{\mathcal{T}_-}(x) = \sum_{\mathcal{T}_-}\left(\sigma_1^2(X_t) - \sigma_1^2(x)\right)K_{b_n}(x - X_t),\ E_{\mathcal{T}_-}(x) = \sum_{\mathcal{T}_-}\sigma_1^2(X_t)\left(\epsilon_t^2 - 1\right)K_{b_n}(x - X_t),$$

we can write

$$\widehat{\sigma}_1^2(x) - \sigma_1^2(x) = \frac{D_{\mathcal{T}_-}(x) + E_{\mathcal{T}_-}(x)}{nb_n\widehat{f}_X(x)}. \tag{2.7}$$

On the other hand,

$$\sup_{x\in\mathcal{T}_-}\left\{\left|\frac{D_{\mathcal{T}_-}(x)}{nb_n\widehat{f}_X(x)} - b_n^2\psi(K)\rho_{\sigma_1}(x)\right|\right\} = O_p(r_n), \tag{2.8}$$

where $\rho_{\sigma_i}(x) = 2(\sigma_i'(x))^2 + 2\sigma_i(x)\sigma_i''(x) + 4(f_X(x))^{-1}\sigma_i(x)\sigma_i'(x)f_X'(x)$ for $i = 1, 2$. Now, using (2.8) in (2.7), we can write

$$\widehat{\sigma}_1^2(x) - \sigma_1^2(x) - b_n^2\psi(K)\rho_{\sigma_1}(x) = \frac{E_{\mathcal{T}_-}(x)}{nb_n\widehat{f}_X(x)}.$$

Performing a similar calculation in $\mathcal{T}_+$ and implementing the jackknife correction using the kernel K^*, we have

$$\left(\widehat{\sigma}_1^{*^2}(x) - \widehat{\sigma}_2^{*^2}(x)\right) - \left(\sigma_1^2(x) - \sigma_2^2(x)\right) = \frac{E^*_{\mathcal{T}_-}(x)}{nb_n\widehat{f}_X(x)} - \frac{E^*_{\mathcal{T}_+}(x)}{nb_n\widehat{f}_X(x)}.$$

Theorem 2.2 now follows from Lemma 2.2 by taking the appropriate forms of f_1 and f_2. □

Given that we do not impose any specific distributional assumptions on the random noise process, the moments of the distribution are not known and ν_ϵ needs to be estimated. Denoting the standardized residuals from the model (2.3) as $\{\widehat{r}_t\}$, one may replace the term ν_ϵ in Theorem 2.2 by

$$\widehat{\nu_\epsilon} = \frac{\sum_{\mathcal{T}_-} \widehat{r}_{1_t}^4 1_{\{X_t \in \mathcal{X}_-\}} + \sum_{\mathcal{T}_+} \widehat{r}_{2_t}^4 1_{\{X_t \in \mathcal{X}_+\}}}{\sum_{\mathcal{T}_-} 1_{\{X_t \in \mathcal{X}_-\}} + \sum_{\mathcal{T}_+} 1_{\{X_t \in \mathcal{X}_+\}}} - 1,$$

where $\widehat{r}_{1_t} = \frac{Y_t - \widehat{\mu}_1^*(X_t)}{\sqrt{\widehat{\sigma}_1^{*2}(X_t)}}$ and $\widehat{r}_{2_t} = \frac{Y_t - \widehat{\mu}_2^*(X_t)}{\sqrt{\widehat{\sigma}_2^{*2}(X_t)}}$ are the standardized residuals for (2.2) estimated in the segments $\mathcal{T}_-$ and $\mathcal{T}_+$ respectively, and $\mathcal{X}_-$ and $\mathcal{X}_+$ are the ranges of the covariate X in these two segments. It can be shown that under the previously mentioned conditions, $\widehat{\nu}_\epsilon$ is a consistent estimate for ν_ϵ.

2.3.2 Test for presence of structural break

The asymptotic theory developed for the estimates $\widehat{\mu}_{\text{diff}}(\cdot)$ and $\widehat{\sigma}^2_{\text{diff}}(\cdot)$ enable us to assess the disparity between the conditional mean and variance functions in two disjoint parts of the data at specific covariate profile $x \in \mathcal{X}$. However, they do not provide sufficient information to determine whether the overall behavior of the functions differ in the two segments. Consider the specific example of the conditional mean function. As we are interested in the equality of the functional characteristic in $\mathcal{T}_+$ and $\mathcal{T}_-$, a change in the global behavior of $\mu(\cdot)$ may be estimated through the test statistic

$$\sup_{x \in \mathcal{X}} \{|\widehat{\mu}^*_{\text{diff}}(x)|\} = \sup_{x \in \mathcal{X}} \{|\widehat{\mu}^*_1(x) - \widehat{\mu}^*_2(x)|\}\,. \tag{2.9}$$

Since we can nonparametrically estimate the $\widehat{\mu}_1^*(\cdot)$ and $\widehat{\mu}_2^*(\cdot)$ functions only point-wise, evaluating the above quantity over a continuous range $\mathcal{X}$ is practically not possible. In practice, we can evaluate the quantity over a fine enough grid of points in $\mathcal{X}$. We therefore define a sequence of partitions $\{\Pi_n\}$ of $\mathcal{X}$ as

$$\Pi_n = \left\{ x_{t_j} \mid x_{t_j} = \Lambda_1 + 2jb_n, j = 0, 1, 2, \ldots, m_n - 1 \right\} \text{ where } m_n = \left\lceil \frac{\Lambda_2 - \Lambda_1}{2b_n} \right\rceil.$$

for some $\Lambda_1, \Lambda_2 \in \mathcal{X}, \Lambda_1 < \Lambda_2$. Note that the partitions $\{\Pi_n\}$ become dense in this subsection of $\mathcal{X}$ as $n \to \infty$. It can be argued that under appropriate smoothness conditions on the true conditional mean function (as mentioned in Assumption 3), $\{\mu(x) \mid x \in \mathcal{X}\}$ can be well approximated by $\{\mu(x) \mid x \in \Pi_n\}$ for a sufficiently large value of n. We can therefore conclude that

$$\sup_{x \in \mathcal{X}} \{|\widehat{\mu}^*_{\text{diff}}(x)|\} \approx \lim_{n \to \infty} \left[\sup_{x \in \Pi_n} \{|\widehat{\mu}^*_{\text{diff}}(x)|\} \right].$$

Our interest is in detecting the presence of structural break in the conditional mean function, and we can treat this as a testing of hypothesis problem of the following form:

$$\begin{aligned} H_0 &: \text{ There is no structural break in the conditional mean,} \\ H_1 &: \text{ There is a structural break in the conditional mean.} \end{aligned} \tag{2.10}$$

In other words, the null hypothesis H_0 corresponds to the assumption $\mu_1(x) = \mu_2(x)$ for all $x \in \mathcal{X}$ whereas the alternative hypothesis H_1 points to the inequality of the

two functions for at least one x. As mentioned before, we can use the statistic given in (2.9), and use its null distribution to detect significant deviation from H_0. The following theorem provides the large sample behavior of the test statistic under the null hypothesis of structural stability when the conditional variance function stays the same throughout the time domain $\mathcal{T}$.

Theorem 2.3. *Along with the previously stated assumptions in Section 2.2, assume that the conditional variance function remains the same throughout the time domain $\mathcal{T}$, i.e., $\sigma_1(x) = \sigma_2(x) = \sigma(x)$ for all x. Define*

$$\mathcal{B}_r(p) = \sqrt{2\log p} - \frac{1}{\sqrt{2\log p}}\Big[\log\log(p) + \log(2\sqrt{\pi})\Big] + \frac{p}{\sqrt{2\log r}}.$$

Considering the bandwidth condition

$$nb_n^9 \log n + \frac{(\log n)^3}{nb_n^3} + \Xi_n \left\{ \frac{b_n^3 \log n}{n} + \frac{(\log n)^2}{n^2 b_n^{\frac{4}{3}}} \right\} \xrightarrow{n\to\infty} 0,$$

under the null hypothesis mentioned in (2.10), for any $z \in \mathbb{R}$,

$$\lim_{n\to\infty} P\left[\frac{\sqrt{nb_n}}{\sqrt{\phi(K^*)}} \sup_{x\in\Pi_n} \left\{ \left(\frac{\sqrt{\widehat{f}_X(x)}}{\sqrt{\widehat{\sigma}^2(x)}} \right) \left| \widehat{\mu}^*_{diff}(x) \right| \right\} \leqslant \mathcal{B}_{m_n}(z) \right] = e^{-2e^{-z}},$$

where $\widehat{\sigma}^2(x)$ is a nonparametric estimate of conditional variance, as mentioned in Theorem 2.1.

Proof. Fix n, pick any k real numbers from $\Pi_n \subset \mathcal{X}$ and denote them by $x_{t_{j_1}}, x_{t_{j_2}}, ..., x_{t_{j_k}}$. Let $\mathbf{x} = \left(x_{t_{j_1}}, x_{t_{j_2}}, ..., x_{t_{j_k}}\right)'$. Under the assumption that the conditional variance process remains the same throughout the time horizon, following an argument similar

to that used in the proof of Theorem 2.1, the null hypothesis indicates that for any $x \in \mathcal{X}$,

$$\widehat{\mu}_1^*(x) - \widehat{\mu}_2^*(x) = \frac{1}{nb_n f_X(x)} \sum_{\mathcal{T}_-} \sigma(X_t)\epsilon_t K_{b_n}^*(x - X_t) - \frac{1}{nb_n f_X(x)} \sum_{\mathcal{T}_+} \sigma(X_t)\epsilon_t K_{b_n}^*(x - X_t).$$

Define

$$\zeta_t^{(1)}(x) = \frac{\sigma(X_t)\epsilon_t K_{b_n}^*(x - X_t)}{\sigma(x)\sqrt{nb_n\phi(K^*)f_X(x)}}\mathbf{I}_{\{t\in\mathcal{T}_-\}}, \zeta_t^{(2)}(x) = \frac{\sigma(X_t)\epsilon_t K_{b_n}^*(x - X_t)}{\sigma(x)\sqrt{nb_n\phi(K^*)f_X(x)}}\mathbf{I}_{\{t\in\mathcal{T}_+\}}$$

Let

$$\begin{aligned}
\zeta_t(x) &= \zeta_t^{(1)}(x)\mathbf{I}_{\{t\in\mathcal{T}_-\}} - \zeta_t^{(2)}(x)\mathbf{I}_{\{t\in\mathcal{T}_+\}}, \\
S_n(x_{j_m}) &= \sum_{\Pi_n} \zeta_t(x_{j_m}), \\
\boldsymbol{\lambda}_t(x) &= \left(\zeta_t(x_{t_{j_1}}), \zeta_t(x_{t_{j_2}}), \ldots, \zeta_t(x_{t_{j_k}})\right)', \\
\mathbf{S}_{n,k} &= \left(\sum_{\Pi_n} S_n(x_{j_1}), \sum_{\Pi_n} S_n(x_{j_2}), \ldots, \sum_{\Pi_n} S_n(x_{j_k})\right)'.
\end{aligned}$$

Theorem 2.3 then follows by applying Lemma 2.3 to the quadratic characteristic matrix of $\mathbf{S}_{n,k}$. □

Following Theorem 2.3, the test for a structural break in the conditional mean can be implemented. We first construct the discrete grid of points Π_n and evaluate the test statistic

$$T(\mu) = \frac{\sqrt{nb_n}}{\sqrt{\phi(K^*)}} \max_{x\in\Pi_n} \left\{ \frac{\sqrt{\widehat{f}_X(x)}}{\sqrt{\widehat{\sigma}^2(x)}} \left|\widehat{\mu}_1^*(x) - \widehat{\mu}_2^*(x)\right| \right\}.$$

Then, we reject the null hypothesis of structural stability in conditional mean at level of significance η if the observed value of the statistic $T(\mu)$ exceeds the critical value $\mathcal{B}_{m_n}(z_\eta)$, $z_\eta = -\log(-2\log(1-\eta))$ being the $100(1-\eta)\%$ quantile of standard Gumbel distribution.

Along a similar line, the testing problem for the structural stability in the conditional variance function $\sigma^2(\cdot)$ is formulated as

$$\begin{aligned} H_0 &: \text{ There is no structural break in the conditional variance,} \\ H_1 &: \text{ There is a structural break in the conditional variance.} \end{aligned} \tag{2.11}$$

It can be tested with the statistic

$$\sup_{x\in\mathcal{X}} \left\{ \left| \widehat{\sigma}^{*^2}_{\text{diff}}(x) \right| \right\} = \sup_{x\in\mathcal{X}} \left\{ \left| \widehat{\sigma}^{*^2}_{1}(x) - \widehat{\sigma}^{*^2}_{1}(x) \right| \right\},$$

which, in practice, can be approximated over a dense grid of values of x as defined in Π_n before, i.e., we shall consider $\max_{x\in\Pi_n} \left\{ \left| \widehat{\sigma}^{*^2}_{1}(x) - \widehat{\sigma}^{*^2}_{1}(x) \right| \right\}$ in practical applications. The below-stated theorem illustrates the behavior of the supremum statistic for the conditional variance under the null hypothesis of structural stability. The proof is similar and straightforward.

Theorem 2.4. *Along with the previously stated assumptions in Section 2.2, let*

$$nb_n^9 \log n + \frac{\log n}{nb_n^4} + \Xi_n \left\{ \frac{b_n^3 \log n}{n} + \frac{(\log n)^2}{n^2 b_n^{\frac{4}{3}}} \right\} \xrightarrow{n\to\infty} 0.$$

Under the null hypothesis of (2.11)*, for any* $z \in \mathbb{R}$*,*

$$\lim_{n\to\infty} P\left[\frac{\sqrt{nb_n}}{\widehat{\nu}_\epsilon\sqrt{\phi(K^*)}}\sup_{x\in\Pi_n}\left\{\sqrt{\frac{\widehat{f}_X(x)}{\widehat{S}(x)}}\left|\widehat{\sigma}^{*^2}_{diff}(x)\right|\right\} \leqslant \mathcal{B}_{m_n}(z)\right] = e^{-2e^{-z}},$$

where $\mathcal{B}_{m_n}(z)$ *is as defined as Theorem 2.3,* $\widehat{S}(x)$ *and* $\widehat{\nu}_\epsilon$ *are as defined in Section 2.3.1.*

Following Theorem 2.4, the test for a structural break in the conditional variance can be implemented similarly as before. Upon constructing the discrete grid of points Π_n, we evaluate

$$T(\sigma) = \frac{\sqrt{nb_n}}{\widehat{\nu}_\epsilon\sqrt{\phi(K^*)}}\max_{x\in\Pi_n}\left\{\frac{\sqrt{\widehat{f}_X(x)}}{\sqrt{\widehat{S}(x)}}\left|\widehat{\sigma}^{*^2}_1(x) - \widehat{\sigma}^{*^2}_2(x)\right|\right\}.$$

and the decision is to reject the null hypothesis of structural stability in conditional variance at level of significance η if the observed value of $T(\sigma)$ is bigger than the critical value $\mathcal{B}_{m_n}(z_\eta)$.

Note that the key steps to proving the above theorems is to express $\widehat{\mu}^*_{\text{diff}}(x)$ as a martingale difference sequence and considering its quadratic characteristic matrix. A martingale maximum deviation theorem proposed by [39] is then applied to obtain the asymptotic distribution. The technical details of the proof are deferred to the Appendix.

Let us now move on to the test of a structural break in both the conditional mean and the conditional variance function. This can be formulated as the following testing

problem:

$$\begin{aligned} H_0 &: \text{ There is no structural break in } \mu(\cdot) \text{ or } \sigma^2(\cdot), \\ H_1 &: \text{ There is a structural break either in } \mu(\cdot) \text{ or in } \sigma^2(\cdot). \end{aligned} \tag{2.12}$$

The test for the above hypothesis can be performed similarly with the help of the following result, which can be derived easily as a straightforward consequence of Theorem 2.2.

Corollary 2.3. *Under the previously stated assumptions in Theorem 2.3 for any* $z \in \mathbb{R}$

$$\lim_{n\to\infty} P\left[\frac{\sqrt{nb_n}}{\sqrt{\phi(K^*)}} \sup_{x\in\Pi_n}\left\{\sqrt{\frac{\widehat{f}_X(x)}{\widehat{S}(x)}}\left|\widehat{\mu}^*_{diff}(x)\right|\right\} \leqslant \mathcal{B}_{m_n}(z)\right] = e^{-2e^{-z}},$$

where all symbols have the same meaning as before.

The implementation of Corollary 2.3 also follows along similar lines as that of Theorems 2.3 and 2.4, although our simulation experiments suggest that the test proposed in last corollary suffers from low power, especially in smaller samples. We thus suggest testing for a structural break separately for the conditional mean and for the conditional variance by simultaneously utilizing Theorems 2.3 and 2.4, along with a Holm-Bonferroni correction. In this approach, define $T_{\max} = \max\{|T(\mu)|, |T(\sigma)|\}$ and $T_{\min} = \min\{|T(\mu)|, |T(\sigma)|\}$. Then, both the null hypotheses in (2.10) and (2.11) are rejected at η level of significance if $T_{\min} \geqslant \mathcal{B}_{m_n}(z_{\frac{\eta}{2}})$ and $T_{\max} \geqslant \mathcal{B}_{m_n}(z_\eta)$. The performance of this test is illustrated in Section 2.4.

As a last point of discussion in this section, we want to highlight that the above results

also facilitate the construction of $100(1-\eta)\%$ simultaneous confidence bands for the differences in the conditional mean function or the same in the conditional variance function in the two parts of the data. These confidence bands are quite effective in obtaining valuable insights about how the covariate process impacts the response variable before and after a potential structural break. We state the expressions for these confidence bands below. The derivations of these bands are straightforward from the previous theorems.

Corollary 2.4. *Under the conditions stated in Theorem 2.3 the confidence band for* $\mu_{diff}(x)$ *is*

$$\widehat{\mu}^{*}_{diff}(x) \pm \frac{\sqrt{\phi(K^{*})\widehat{\sigma}^{2}(x)}}{\sqrt{nb_{n}\widehat{f}_{X}(x)}}\mathcal{B}_{m_n}(z_\eta);$$

and under the conditions stated in Theorem 2.4, the confidence band for $\sigma^{2}_{diff}(x)$ *is*

$$\widehat{\sigma}^{*^2}_{diff}(x) \pm \frac{\widehat{\nu}_{\epsilon}\sqrt{\phi(K^{*})\widehat{f}_{X}(x)}}{\sqrt{nb_{n}\widehat{S}(x)}}\mathcal{B}_{m_n}(z_\eta).$$

2.3.3 A note on the choice of the bandwidth

The asymptotic theory derived in the previous subsection emphasize the critical role of selecting an appropriate bandwidth for the estimation of the conditional mean and variance functions in order to get desired performance in the proposed tests. The conditions set forth in these results impose specific constraints on both the bandwidth b_n and the dependence range of the covariate $\{X_t\}$. For instance, the first part of the bandwidth condition in Theorem 2.3, $nb_n^9 \log n \to \infty$, ensures that the bandwidth is not excessively large. Meanwhile, the second part, $(\log n)^3/nb_n^3 \to 0$, prevents the

bandwidth from being too small. The third condition involving Ξ_n guarantees that the dependence range of the covariate process $\{X_t\}$ remains within an appropriate range. Notably, for short-range dependent (SRD) processes, $\Xi_n = O(n)$, ensuring that the bandwidth condition is met as long as the first two terms are $o(1)$. This aligns with choosing the bandwidth as

$$b_n = O(n^{-\beta}), \quad \beta \in \left(\frac{1}{9}, \frac{1}{3}\right).$$

The situation for long-range dependent (LRD) processes is more complex. Consider, for example, a zero-mean *iid* process $\{\eta_t\}$ with $\eta_0 \in \mathcal{L}^q, q \geqslant 2$, and define a process of the form

$$X_t = \sum_{j=0}^{\infty} a_j \eta_{t-j}, \quad \text{with } a_j = \frac{l(j)}{j^{\kappa}}, \quad \kappa \in \left(\frac{1}{2}, 1\right],$$

where $l(\cdot)$ is a slowly varying function. The structure defined above encompasses a broad range of linear and nonlinear processes. For example, if we set

$$a_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)},$$

where Γ denotes the incomplete gamma function, with $d \in (0, 0.5)$, we obtain the long-range dependent FARIMA$(0, d, 0)$ process. This specification can also produce various nonlinear processes, including the class of nonlinear AR and ARCH models.

Our setting also accommodates heavy-tailed $\{\eta_t\}$. It can be shown that the process

$\{X_t\}$ of this form exhibits LRD characteristics. Elementary calculations yield

$$\Xi_n = \begin{cases} O\left(n^{3-2\kappa} l^2(n)\right) & \text{if } \kappa \in \left(\frac{1}{2}, 1\right), \\ O\left(n\left(\sum_{i=1}^{n} \frac{|l(i)|}{i}\right)^2\right) & \text{if } \kappa = 1. \end{cases}$$

Now, to obtain a bandwidth condition similar to those in Theorem 2.1 and Theorem 2.3, we require $\kappa \in \left(\frac{17}{26}, 1\right]$, under which the bandwidth $b_n = O(n^{-\beta})$ should satisfy

$$\beta \in \left(\max\left\{\frac{1}{9}, \frac{2-2\kappa}{3}\right\}, \min\left\{\frac{1}{3}, \frac{3(2\kappa-1)}{4}\right\}\right).$$

In particular, the MSE-optimal bandwidth $n^{-0.2}$ nearly satisfies both the requirements of LRD and SRD processes. In parallel, a common approach to obtaining the optimal bandwidth in nonparametric regression is through a cross-validation procedure, which involves selecting the bandwidth that minimizes a specified loss criterion. One such criterion is the mean squared error (MSE) criterion [40], defined as

$$L_{\text{MSE}} = E\left[\left(\widehat{g}(x \mid b_n) - g(x)\right)^2\right], \quad g(\cdot) \equiv \mu(\cdot) \text{ or } \sigma(\cdot),$$

where $\widehat{g}(x \mid b_n)$ is the estimate of $g(x)$ using the Nadaraya-Watson kernel estimator with bandwidth b_n. More generally, the bandwidth can also be chosen by minimizing the conditional weighted mean squared error

$$L_{\text{WMSE}} = \int_{-\infty}^{\infty} E\left[\left(\widehat{g}(x \mid b_n) - g(x)\right)^2\right] \omega(x)\, dx,$$

where $\omega(\cdot)$ is a suitable weight function with compact support. Another pertinent

work was done by [41] who designed a feed forward neural network with one hidden layer that is trained to minimize the prediction error of the local linear regression estimates. While this approach may be an effective alternative, it is important to note that the authors worked with only a nonlinear autoregressive model structure.

2.3.4 Localization structural breaks

With the asymptotic theory developed for different types of tests to assess the similarity of functional characteristics in the two segments, we move on to leverage these tests for detecting structural breaks at unknown locations within a dataset. We use a two-stage localization where in the first stage, we recursively apply a binary segmentation type approach by splitting the time series at the midpoint and testing for structural breaks. If the test is not rejected for a particular segment, then the procedure stops for that part, whereas, if a break is indeed detected, then the algorithm continues to evaluate both segments before and after the midpoint in the same way. This process is repeated recursively, testing for breaks and splitting the data, until each resulting segment contains fewer than $L_{\min}$ number of data points (this is a user defined quantity). Note that this approach is efficient in identifying multiple structural breaks, as it simultaneously examines both sides of any detected break, ensuring comprehensive detection throughout the dataset.

In the second stage of the algorithm we perform a post-processing, where the partition generated by the series of tests in the first stage is examined further. The objective is to ensure that two consecutive tests in the first stage are not rejected because of the same break-point. For instance, if $b_1 \leqslant b_2 \leqslant \ldots \leqslant b_k$ are the points where the tests are rejected in the first stage, then to confirm that b_i is indeed a structural break, we

consider the subdata $\{(Y_t, X_t)\}$, $b_{i-1} \leqslant t \leqslant b_{i+1}$ and split it into two segments at b_i. Let $\widehat{\mu}_{b_i^-}(\cdot)$, $\widehat{\mu}_{b_i^+}(\cdot)$ and $\widehat{\sigma}^2_{b_i^-}(\cdot)$, $\widehat{\sigma}^2_{b_i^+}(\cdot)$ be the conditional mean and variance estimates in the two segments, using the kernel estimators defined in Section 2.3.1. Then, the disparity between two parts is assessed by the quantities

$$\widehat{\mu}_{\text{cp}}(b_i) = \sup_{x \in \mathcal{X}} \left\{ \left| \widehat{\mu}_{b_i^-}(x) - \widehat{\mu}_{b_i^+}(x) \right| \right\}, \quad \widehat{\sigma}^2_{\text{cp}}(b_i) = \sup_{x \in \mathcal{X}} \left\{ \left| \widehat{\sigma}^2_{b_i^-}(x) - \widehat{\sigma}^2_{b_i^+}(x) \right| \right\}, \tag{2.13}$$

and we declare b_i to be indeed a structural break if either $\widehat{\mu}_{\text{cp}}(b_i)$ or $\widehat{\sigma}^2_{\text{cp}}(b_i)$ is statistically significantly large. The asymptotic properties of $\widehat{\mu}_{\text{cp}}(b_i)$ and $\widehat{\sigma}^2_{\text{cp}}(b_i)$ follow similarly to those outlined in Section 2.3.2, and the corresponding tests can be designed in the same fashion. Note that the second stage of the algorithm serves as a confirmatory procedure, helping to prevent overestimation of the number of structural breaks in the data. The pseudo-code of the procedure is presented in Algorithm 1, and its consistency is proved in the theorem below.

Theorem 2.5. *Let $\{(Y_t, X_t)\}$ be a jointly observed time series of length n with a true structural break in the functional characteristic $g(\cdot)$ at the time-point τ and let $\widehat{\tau}$ be a structural break detected by the binary segmentation algorithm. If the power of the corresponding test used in the algorithm is $(1-\beta)$, then*

$$P\left(|\tau - \widehat{\tau}| \leqslant \frac{L_{min}}{2} \right) \geqslant 1 - \beta \text{ as } n \to \infty.$$

Proof. with power $(1-\beta)$ at each step. This ensures a probability of at least $(1-\beta)$ of identifying the break at each iteration. After several iterations, the interval reduces to a length $L_{\min}$, with a maximum error of $L_{\min}/2$. As $n \to \infty$, the procedure becomes more accurate, ensuring with probability $(1-\beta)$ that the estimated break-point $\widehat{\tau}$ is

Procedure 1: Localization of structural break in functional characteristic $g(\cdot)$

Input : Time series data $D = \{(Y_i, X_i)\}$ for i corresponding to time-point t_i.
Output: List containing locations of structural breaks.
Function `binary segmentation(`D`):`

```
L = length(D)
Initiate B, an empty list of structural breaks
if L < L_min then
    return B
end
Compute midpoint m = ⌊L/2⌋;
Test for structural break in the functional characteristic g(·) at m at 5%
 level of significance;
if No break detected at m then
    Proceed to the second stage.
end
if Break detected at m then
    Add m to the list B;
    Segment1 ← D[1 : m];
    Segment2 ← D[m + 1 : end];
    breaks1 ← binary segmentation(Segment1);
    breaks2 ← binary segmentation(Segment2);
end
Sort the time-points in B in increasing order to form a sequence
 {1 = b_0, b_1, b_2, ..., b_k, b_{k+1} = L}.
foreach b_j, j = 1, ..., k do
    Consider the sub-data D_sub = {(Y_i, X_i)}, b_{j-1} ⩽ i ⩽ b_{j+1};
    Test for structural break in the functional characteristic g(·) at b_j at
     5% level of significance;
    if the test is not rejected then
        Remove b_j from B
    end
end
return B
```

within $L_{\min}/2$ of the true break τ. Thus, $P\left(|\tau - \widehat{\tau}| \leqslant L_{\min}/2\right) \geqslant 1 - \beta$. □

At this stage, it is important to emphasize that an alternative approach is to imple-

ment the second stage directly to detect structural break, i.e. to split the data at a random time-point t_0 and assessing the disparity between the functional characteristics of the data before and after t_0 using test statistics $\widehat{\mu}_{\text{cp}}(t_0)$ or $\widehat{\sigma}^2_{\text{cp}}(t_0)$ of the form (2.13). Assuming the data contains at most one structural break, the test statistics can then be inverted to detect the location of this break. For practical implementation, we define a sequence of partitions of $\mathcal{T}$ as $\{\mathcal{T}_n\}$, where $\mathcal{T}_n = \{t'_1, t'_2, \ldots \mid t'_j = 2jb_n, j = 0, 1, 2, \ldots, k_n - 1\}$ with $k_n = \lceil 1/2b_n \rceil$. A structural break in the conditional mean or in the conditional variance can then be estimated as

$$\widehat{\tau}^{\mu} = \underset{t \in \mathcal{T}_n}{\operatorname{argmax}} \left\{ \widehat{\mu}_{\text{cp}}(t) \right\}, \quad \widehat{\tau}^{\sigma} = \underset{t \in \mathcal{T}_n}{\operatorname{argmax}} \left\{ \widehat{\sigma}^2_{\text{cp}}(t) \right\}.$$

The consistency of these estimated breaks is guaranteed by the following result.

Theorem 2.6. *Let τ^{μ} and τ^{σ} denote the true structural breaks in the conditional mean and the conditional variance, respectively. Then, as $n \to \infty$, $\widehat{\tau}^{\mu} \xrightarrow{P} \tau^{\mu}$ and $\widehat{\tau}^{\sigma} \xrightarrow{P} \tau^{\sigma}$.*

Proof. consider the number of data points on either side of t_0 as $n \to \infty$. Let $\widetilde{\mu}(x \mid t_0) = \mu_{t_0^-}(x) - \mu_{t_0^+}(x)$, where $\mu_{t_0^-}(\cdot)$ and $\mu_{t_0^+}(\cdot)$ respectively denote the true conditional mean function before and after the splitting point t_0. We denote by $\widehat{\widetilde{\mu}}(x \mid t_0)$ the Nadaraya Watson kernel estimate of $\widetilde{\mu}(x \mid t_0)$. Clearly, under relevant bandwidth conditions, as $n \to \infty$, $\widehat{\widetilde{\mu}}(x \mid t_0) \xrightarrow{P} \widetilde{\mu}(x \mid t_0)$ for any $x \in \mathcal{X}$ and $t \in \mathcal{T}$. Now, note that $\widetilde{\mu} : \mathcal{X} \times \mathcal{T} \to \mathbb{R}$. Define a map $C : \mathcal{T} \to \mathbf{C}(\mathcal{X})$, where $\mathbf{C}(A)$ for any set A denotes the set of non-null compact subsets of A, as $C(t_j) = x_{t_j}$ where $x_{t_j} \in \Pi_n, t_j \in \mathcal{T}_n$. Since C is a compact-valued map as $n \to \infty$, then using Berge's

maximum theorem we can write $\widehat{\mu}_{\text{cp}}(t_0) \to \mu_{\text{cp}}(t_0)$. The consistency of $\widehat{\tau}^{\mu}$ thus is a consequence of the uniqueness of the maximum of $\widehat{\widehat{\mu}}$ in its second argument. One can follow a similar approach for proving the consistency of $\widehat{\tau}^{\sigma}$. □

Detection of multiple structural breaks using this approach can be achieved through techniques such as wild binary segmentation [42] or linear segmentation [43], with appropriate modifications. We find it imperative to reiterate that this alternative approach requires substantially more number of iterations to detect a single structural break, whereas the binary segmentation algorithm works out much faster in practice.

2.4 Simulation study

To evaluate the effectiveness of the proposed method, we conduct a series of simulations designed to replicate real-world scenarios in the time series domain. The simulation setup involves several key steps, including data generation, parameter specification, and validation procedures. Let us first outline the details of these steps, providing a comprehensive overview of the experimental framework.

We start by generating the covariate series $\{X_t\}$ from different data generating processes (DGP), the response $\{Y_t\}$ is then obtained through certain forms of conditional mean and variance functions, subject to different forms of normal, moderate and heavy tailed noise. The considered DGPs, noise distributions and different forms of the conditional mean and variance functions (depending on the number of structural breaks in the data) is described in Table 2.1. To detect the power of the test and for assessing the accuracy of the proposed binary segmentation algorithm, structural breaks are randomly introduced in the entire horizon and the segments mentioned in

Table 2.1 refer to different parts generated because of the break-points.

Table 2.1: Specifications of various components of the simulation settings.

Segment	**Conditional mean $\mu(x)$**	**Conditional variance $\sigma^2(x)$**
Segment 1	$0.5 + 0.2x$	1
Segment 2	$0.1 + 0.3x^2 + 0.1x^3 + 0.2x^4$	x^2
Segment 3	$\log(0.4 + 0.1x^2)$	$0.1 + 0.4x^2$
Segment 4	$\exp(0.01x)$	$0.5 + (0.8 + x)^4$
Segment 5	$0.9\sin(x)$	$\log(1 + 0.4x^2)$
DGP	**Model Structure**	**True Parameter Values**
White noise	$y_t \sim \mathcal{N}(\mu, \sigma^2) + \epsilon_t$	$\mu = 0, \sigma = 1$
ARMA-GARCH	$y_t = \mu + \phi_1 y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1},$	$\mu = 0, \phi_1 = 0.5, \theta_1 = -0.4$
	$\epsilon_t \sim \mathcal{N}(0, \sigma_t^2), \sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$	$\omega = 0.1, \alpha_1 = 0.1, \beta_1 = 0.8$
TAR	$y_t = \begin{cases} \phi_{11} y_{t-1} + \phi_{12} y_{t-2} + \epsilon_t, y_{t-1} \leqslant 0 \\ \phi_{21} y_{t-1} + \phi_{22} y_{t-2} + \epsilon_t, y_{t-1} > 0 \end{cases}$	$\phi_{11} = 0.6, \phi_{12} = 0.3,$ $\phi_{21} = -0.6, \phi_{22} = 0.4$
Noise process	**Noise distribution**	**Parameter specification**
Normal	$\mathcal{N}(\mu, \sigma^2)$	$\mu = 1, \sigma = 1$
Student's t	t_ν	$\nu = 10$
Power law	$f(x) = x_0 \alpha x^{1-\alpha}$	$x_0 = 1, \alpha = 0.6$

It is pertinent to discuss the choices of the conditional mean and variance structures used in this simulation study. The mean function $\mu(x)$ incorporates a polynomial regression function of different degrees in segments 1 and 2. Segment 3 employs a log-linear regression, frequently used in econometrics to capture diminishing returns and stabilized volatility in financial data. Segment 4 features an exponential regression, widely adopted in finance for modeling compound interest and in biological contexts for growth rates. Finally, in segment 5 we utilize a trigonometric regression model,

suitable for modeling seasonal behaviors common in meteorology, economics, and biology. For the variance function $\sigma^2(x)$, the first segment assumes homoskedasticity, while the other specifications are well-established in the literature for addressing heteroskedasticity [44]. Particularly, segments 2 and 3 consider standard GARCH structures, segment 4 employs a power GARCH framework, and segment 5 utilizes an exponential GARCH approach.

The nonparametric estimations are done using the parabolic kernel $K(u) = 0.75\mathbf{I}_{\{|u|\leqslant 1\}}$ and the MSE-optimal bandwidth of $b_n = n^{-0.2}$ where n is the length of the data. One may alternatively select the bandwidth using cross-validation procedures, however, our observation has been that the performances of the test with the cross-validated bandwidth are more or less similar to those obtained with the MSE optimal bandwidth. We conduct the simulation study for sample sizes of $n = \{500, 1000, 2000\}$. All the simulations performed for evaluating the size and the power of the tests are conducted at 5% level of significance and under the presence of a single structural break. The conditional mean and variance functions before and after the structural break are taken as mentioned in segment 5 and 2 respectively. To evaluate the performance of the tests, we simulate 100 independent samples from a chosen combination of a data generating process (DGP) and noise distribution. In each sample, a single structural break is introduced at a random time point. For each test, the size is measured as the average empirical Type-I error rate, while the power is assessed by calculating the average proportion of correct rejections of the null hypothesis across all 100 samples.

We present the performance of the test in the presence of a single structural break in

the conditional mean and/or variance. For structural breaks in the conditional mean, the test shows good size control across all noise structures and DGPs. The power of the test improves significantly with increasing sample size, starting with moderate power at a sample size of 500 and reaching up to 67% at a sample size of 2000, with TAR models consistently demonstrating higher power compared to White Noise and ARMA-GARCH. When detecting breaks in the conditional variance, the test exhibits similarly good size control, with minimal deviations even at larger sample sizes. The power for variance breaks is notably higher than for mean breaks, achieving relatively strong power even at a sample size of 500 and further improving with sample size. Overall, the test is more effective at detecting variance breaks than mean breaks, with the power increasing consistently across all DGPs as the sample size grows, especially for more complex models like TAR. The simultaneous test for structural breaks in mean and variance has around 50% to 70% power in smaller sample sizes, working comparatively well in the case of heavy-tailed noises. As the sample size increases the power of the test gets better across all combinations, consistently reaching values in the range of 80% to 90%.

Turn attention to the detection performance of the binary segmentation algorithm for sample sizes of $n \in \{500, 1000, 2000\}$. To evaluate the performance of the structural break detection algorithm, we introduce a randomly generated number of breaks in the conditional mean and/or variance functions, placing these breaks at random locations within the data. We ensure that there is a minimum gap of 100 days between any two breaks to maintain distinct separations and the maximum number of breaks for sample sizes of up to 2000 is 4. The effectiveness of the detection process is assessed using several metric. At first we calculate a 'deviation' metric, which quantifies the

Table 2.2: Performance of the test when there is a single structural break in the conditional mean.

		Noise: $\mathcal{N}(0,1)$		Noise: t_{10}		Noise: Power law	
Sample size	DGP	Size	Power	Size	Power	Size	Power
500	White Noise	0.00	0.21	0.00	0.31	0.01	0.25
	ARMA-GARCH	0.01	0.23	0.00	0.29	0.00	0.26
	TAR	0.01	0.31	0.00	0.27	0.00	0.34
1000	White Noise	0.00	0.43	0.02	0.54	0.00	0.48
	ARMA-GARCH	0.00	0.37	0.00	0.51	0.00	0.48
	TAR	0.03	0.47	0.01	0.59	0.01	0.57
2000	White Noise	0.00	0.53	0.01	0.58	0.00	0.56
	ARMA-GARCH	0.04	0.55	0.02	0.58	0.00	0.62
	TAR	0.01	0.58	0.01	0.65	0.00	0.67

Table 2.3: Performance of the test when there is a single structural break in the conditional variance.

		Noise: $\mathcal{N}(0,1)$		Noise: t_{10}		Noise: Power law	
Sample size	DGP	Size	Power	Size	Power	Size	Power
500	White Noise	0.00	0.65	0.01	0.61	0.00	0.71
	ARMA-GARCH	0.00	0.56	0.01	0.61	0.00	0.68
	TAR	0.01	0.81	0.01	0.73	0.03	0.89
1000	White Noise	0.00	0.66	0.01	0.69	0.01	0.78
	ARMA-GARCH	0.00	0.60	0.01	0.75	0.00	0.82
	TAR	0.00	0.88	0.03	0.88	0.01	0.91
2000	White Noise	0.01	0.78	0.01	0.72	0.00	0.86
	ARMA-GARCH	0.00	0.74	0.02	0.73	0.03	0.84
	TAR	0.05	0.92	0.04	0.93	0.05	0.97

average distance between the detected breaks and the nearest actual structural break. Mathematically, if $\{\text{CP}_1, \text{CP}_2, \ldots, \text{CP}_m\}$ are the true structural breaks, then we define

Table 2.4: Performance of the test when there is a single structural break in conditional mean or variance.

		Noise: $\mathcal{N}(0,1)$		Noise: t_{10}		Noise: Power law	
Sample size	DGP	Size	Power	Size	Power	Size	Power
500	White Noise	0.00	0.66	0.01	0.52	0.01	0.72
	ARMA-GARCH	0.02	0.44	0.00	0.52	0.01	0.62
	TAR	0.00	0.72	0.01	0.68	0.00	0.74
1000	White Noise	0.00	0.78	0.00	0.70	0.01	0.74
	ARMA-GARCH	0.01	0.70	0.02	0.68	0.00	0.60
	TAR	0.01	0.80	0.00	0.64	0.00	0.80
2000	White Noise	0.01	0.84	0.01	0.80	0.00	0.82
	ARMA-GARCH	0.00	0.90	0.01	0.92	0.00	0.88
	TAR	0.02	0.85	0.00	0.87	0.01	0.91

the average minimum deviation measure (AMD) as the mean of

$$\mathrm{MD} = \sum_{r=1}^{m'} \min_{1 \leqslant i \leqslant m} \{|\widehat{\tau}_r - \mathrm{CP}_m|\},$$

taken over all repetitions, where $\{\widehat{\tau}_1, \widehat{\tau}_2, \ldots, \widehat{\tau}_{m'}\}$ are the detected structural breaks in the data. This average deviation measure, computed over 50 iterations, provides a quantitative assessment of the accuracy of our detection algorithm by comparing the detected breaks to the true breaks, highlighting the precision and reliability of the procedure. We further compute the average error in terms of number of detected structural breaks, denoted as 'ADN' (Average Deviation in Numbers) as the average of

$$\mathrm{DN} = |m - m'|\,.$$

taken over all repetitions. From Table 2.5, we observe that the nonparametric PELT [45] algorithm severely overestimates the number of structural breaks present in the data, especially for larger sample sizes and more complex data structures. While it helps the method achieve better accuracy in terms of the minimum deviation metric in few cases, it is practically not a suitable approach. Our method, in contrast, estimates at most one additional structural break on average while typically retaining a deviation of less than ± 70 data points from the true structural break. Only exception is when the covariate series comes from an ARMA-GARCH process and there are higher number of structural breaks in the main data. In that scenario, the error in deviation from the true break-point is in the range of 100 to 150. Overall, our simulation study suggests that the size of the test of structural stability in the conditional mean is consistently near zero across all scenarios, indicating that the test does not falsely identify structural breaks when none exist. However, the power of the test remains low for smaller sample sizes. As the sample size increases to 2000, the power improves across all models. This demonstrates that the test becomes more effective at detecting structural breaks with larger sample sizes and in the presence of more complex noise, such as t-distributed or power-law noise, which often occurs in financial time series data. On the other hand, the power performance of the test of structural stability in the conditional variance function is considerably high even for smaller sample sizes of 500, for higher sample sizes the power approaches 1. Further, the power of the test is typically high for heavy-tailed noises, suggesting that the test can be particularly useful for financial data. The size of this test also stays controlled across all cases.

Table 2.5: Detection performance of binary segmentation and nonparametric PELT algorithms where there exists a random number of structural breaks in either conditional mean or variance

Sample Size	DGP	Noise	AMD		ADN	
			CPFind	PELT	CPFind	PELT
500	White Noise	$\mathcal{N}(0,1)$	53.18	24.49	0.32	0.97
500	White Noise	t_{10}	57.88	21.96	0.28	0.96
500	White Noise	Power law	17.30	29.46	0.24	1.03
500	ARMA-GARCH	$\mathcal{N}(0,1)$	76.52	32.21	0.42	1.10
500	ARMA-GARCH	t_{10}	82.98	29.21	0.48	1.12
500	ARMA-GARCH	Power law	41.24	29.27	0.22	1.40
500	TAR	$\mathcal{N}(0,1)$	61.20	30.02	0.26	1.21
500	TAR	t_{10}	65.50	29.14	0.22	1.38
500	TAR	Power law	24.06	38.02	0.20	1.38
1000	White Noise	$\mathcal{N}(0,1)$	42.34	34.90	0.46	1.46
1000	White Noise	t_{10}	68.76	95.76	0.52	1.34
1000	White Noise	Power law	17.54	99.76	0.68	2.22
1000	ARMA-GARCH	$\mathcal{N}(0,1)$	104.46	55.40	0.60	2.98
1000	ARMA-GARCH	t_{10}	169.84	108.21	0.80	2.34
1000	ARMA-GARCH	Power law	48.28	132.91	0.38	3.60
1000	TAR	$\mathcal{N}(0,1)$	53.06	83.65	0.46	3.76
1000	TAR	t_{10}	64.28	85.65	0.44	3.38
1000	TAR	Power law	30.82	70.63	0.66	3.68
2000	White Noise	$\mathcal{N}(0,1)$	57.14	67.30	1.06	1.76
2000	White Noise	t_{10}	53.73	56.02	1.46	1.68
2000	White Noise	Power law	24.94	91.73	1.71	1.42
2000	ARMA-GARCH	$\mathcal{N}(0,1)$	146.88	62.57	1.10	3.58
2000	ARMA-GARCH	t_{10}	120.00	109.77	0.69	3.54
2000	ARMA-GARCH	Power law	115.65	113.92	0.82	3.46
2000	TAR	$\mathcal{N}(0,1)$	72.00	68.38	1.62	5.34
2000	TAR	t_{10}	47.31	76.23	1.88	5.26
2000	TAR	Power law	21.24	68.02	1.70	4.96

2.5 Application to Bitcoin data

News sentiment plays a significant role in shaping Bitcoin prices, much like it does with traditional financial assets. Several recent studies have explored how macroeconomic news and social media activity affect Bitcoin's price movements [46]. Public attention to Bitcoin has also been found to predict its price behavior [47]. In this section, we utilize our proposed method to analyze shifts in the relationship between public attention, measured by Google News Interest Score (GNIS) from Google Trends, and Bitcoin's price and volatility.

The GNIS data, denoted as $\{G_t\}$, spans January 1 2020 to September 4 2024, assigning a relative score (0-100) to the search volume on "Bitcoin", with 100 indicating peak popularity. This score, extracted using the 'pytrends' library, reflects search volume as a proportion of all global searches. The daily Bitcoin log-price series covering the same period, denoted as $\{P_t\}$, is sourced from FRED (link: https://fred.stlouisfed.org/series/CBBTCUSD).

We model the current day Bitcoin log-return as a function of the last day's GNIS, in order to incorporate for a lag in the impact. The model is,

$$r_t = \mu(G_{t-1}) + \sigma(G_{t-1})\epsilon_t,$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ are unknown smooth functions representing the mean and conditional variance of the log-price variable. Note that since we are working with the log-price, $\sigma(\cdot)$ can be thought of as a proxy for the volatility. For the detection procedure, we apply a rule-of-thumb bandwidth, $h_n = n^{-0.2} = 0.2258$. We also ran

the same analysis using a cross-validated bandwidth choice, however the detected structural breaks in these two procedures were reasonably close to one another. The algorithm requires a minimum of 200 days between consecutive structural breaks. We detect two structural breaks in the mean return level, no breaks were found in the conditional variance of the return series. The detected breaks are illustrated on the price series in Figure 2.1. We start with a brief exploratory analysis of the data, presented in Table 2.6. In the table, "JB-Test" denotes Jarque-Bera test of normality, "LB-Test" denotes the Ljung-Box test for autocorrelation, "LM-Test" denotes the Lagrange multiplier test for heteroskedasticity and "ADF-Test" denotes the augmented Dickey-Fuller test for stationarity. All tests are conducted at 5% level of significance. The descriptive statistics and these tests provide a clear picture of the market behavior across the three segments. The first one, i.e., 1 January 2020 to 7 March 2021, is characterized by relatively lower log-prices, high volatility in log-returns (as indicated by the high kurtosis), and the most dispersed GNIS values. The positive skewness of log-prices and negative skewness of log-returns highlight significant price increases and occasional sharp decreases during this period. In this segment, the JB test reveals significant deviations from normality, particularly in log-returns, which exhibit high kurtosis. The LB test indicates autocorrelation in log-returns, while the LM test suggests heteroskedasticity, though only in log-returns. The ADF test confirms the stationarity of log-returns, but log-prices and GNIS are borderline non-stationary. Segment 2 (8 March 2021 to 8 July 2023) shows a rise in the mean log-price, reduced volatility in log-returns (lower kurtosis), and increased GNIS, suggesting greater market stability and media attention compared to Segment 1. Log-prices have a nearly symmetrical distribution, indicating fewer extreme price movements. The JB test in this segment shows improvements in normality for log-prices and GNIS, though log-

returns still deviate. Autocorrelation weakens during this period, as seen in the LB test, and there is no evidence of heteroskedasticity in any variables, reflecting greater stability. However, the ADF test still indicates trends in log-prices and GNIS, while log-returns remain stationary. Finally, Segment 3 (8 July 2023 to 4 September 2024) marks the highest average log-prices with relatively lower GNIS compared to Segment 2, showing a moderate decrease in media attention. The volatility in log-returns remains low, while skewness and kurtosis suggest more stabilized market behavior, almost approaching the normal distribution. While log-returns continue to exhibit non-normality, the deviations are less pronounced, and there is no autocorrelation or heteroskedasticity detected. The volatility in GNIS diminishes, and stationarity is confirmed for log-returns, though trends persist in log-prices and GNIS.

Table 2.6: Descriptive statistics for log-price, returns and GNIS, for the three segments generated by the detected structural breaks. Segment 1 refers to 1 January 2020 to 7 March 2021, Segment 2 is 8 March 2021 to 8 July 2023, and Segment 3 is from 8 July 2023 until 4 September 2024.

Entire data	**Mean (SD)**	**Range**	**Quartiles**	**Skewness**	**Kurtosis**	**JB Test**	**LB Test**	**LM Test**	**ADF Test**
Log-Price	10.24 (0.64)	(8.50, 11.20)	(9.86, 10.32, 10.75)	−0.57	2.37	0.00	0.00	1.00	0.76
Log-Returns	0.00 (0.03)	(−0.32, 0.16)	(−0.02, 0.00, 0.02)	−0.60	10.60	0.00	0.41	0.00	0.01
GNIS	28.72 (16.32)	(10, 100)	(17.82, 22.71, 35.00)	1.79	6.51	0.00	0.00	1.00	0.02
Segment 1	**Mean (SD)**	**Range**	**Quartiles**	**Skewness**	**Kurtosis**	**JB Test**	**LB Test**	**LM Test**	**ADF Test**
Log-Price	9.46 (0.58)	(8.50, 10.95)	(9.12, 9.26, 9.78)	1.07	3.10	0.00	0.00	1.00	0.79
Log-Returns	0.00 (0.00)	(−0.36, 0.02)	(0.00, 0.00, 0.00)	−1.32	18.11	0.00	0.91	0.00	0.01
GNIS	27.67 (22.99)	(10, 98)	(14.00, 16.93, 28.96)	1.75	4.81	0.00	0.00	1.00	0.56
Segment 2	**Mean (SD)**	**Range**	**Quartiles**	**Skewness**	**Kurtosis**	**JB Test**	**LB Test**	**LM Test**	**ADF Test**
Log-Price	10.38 (0.40)	(9.66, 11.12)	(10.03, 10.36, 10.72)	−0.01	1.80	0.00	0.00	1.00	0.78
Log-Returns	0.00 (0.00)	(−0.02, 0.01)	(0.00, 0.00, 0.01)	−0.37	6.90	0.00	0.30	0.00	0.01
GNIS	32.16 (14.35)	(15, 100)	(20.57, 29.71, 39.00)	1.32	5.17	0.00	0.00	1.00	0.01
Segment 3	**Mean (SD)**	**Range**	**Quartiles**	**Skewness**	**Kurtosis**	**JB Test**	**LB Test**	**LM Test**	**ADF Test**
Log-Price	10.79 (0.35)	(10.13, 11.20)	(10.35, 10.72, 11.06)	−0.31	1.58	0.00	0.00	1.00	0.92
Log-Returns	0.00 (0.00)	(−0.01, 0.01)	(0.00, 0.00, 0.00)	0.06	4.20	0.00	0.41	0.00	0.01
GNIS	22.92 (8.28)	(12, 54)	(18.00, 20.71, 25.14)	1.64	5.83	0.00	0.00	1.00	0.67

Following the methodology proposed in the previous sections, we find two structural breaks in the conditional mean, presented in Figure 2.1. The first break, detected on 7 March 2021, aligns with Bitcoin's rapid rise to an all-time high in early 2021, when institutional interest in Bitcoin surged and the cryptocurrency gained widespread media attention. This period saw companies like Tesla investing in Bitcoin, which drove significant news coverage and heightened interest in the asset, contributing to the price surge. The second structural break on 8 July 2023, suggests another structural shift, possibly linked to market stabilization after major corrections. Before this date, the market was volatile and the news was largely negative (declining trend), reflecting uncertainty and regulatory pressures. However, by July, the tone of news coverage shifted to more positive stories, focusing on institutional adoption, technological advancements, and market stabilization. This change in sentiment likely contributed to Bitcoin's market recovery, with the news acting as a stabilizing factor, reinforcing the upward momentum in the market. We report no structural breaks in the conditional variance.

Below, in Figure 2.2, we present the confidence bands for the disparity in mean regression functions caused by the two structural breaks, as well as the conditional variance function in the entire time horizon. The top panel of Figure 2.2 indicates a slightly increasing trend in the effect of GNIS on the mean function. It further highlights that when news attention is low, there are greater fluctuations in the difference between the average log-price behavior on either side of the first structural break. However, as news attention increases, the confidence band narrows and the difference in the effect on mean log-price diminishes. It implies that lesser news attention has less impact on the average price in the first segment as compared to the second, whereas

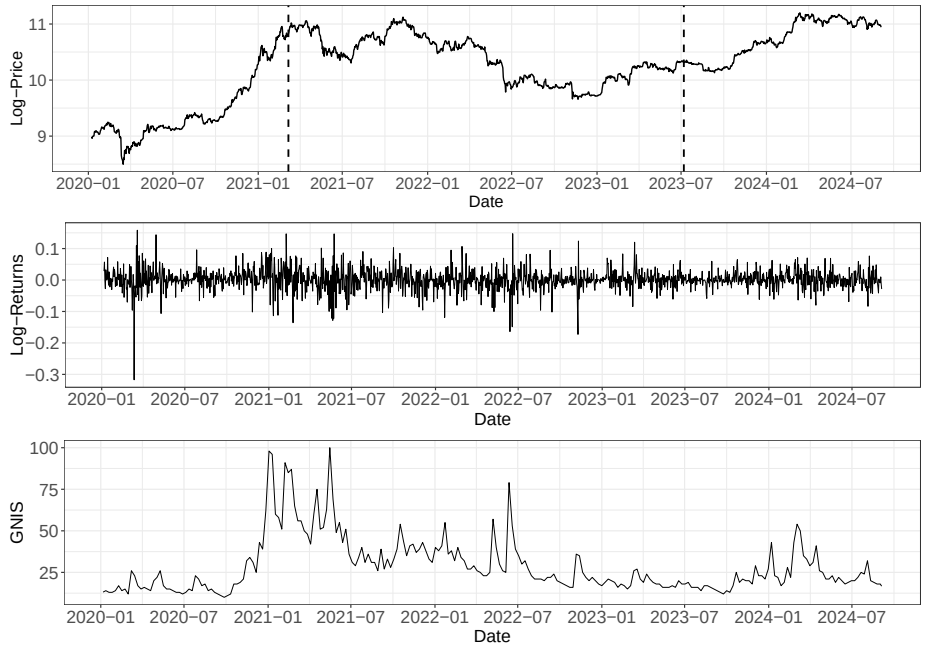


Figure 2.1: (Top) Bitcoin price (log-transformed) and the detected structural breaks in the impact of public attention on its average level; (middle) log-return of Bitcoin price; (bottom) GNIS series for the entire time period.

heightened news attention brings more certainty in the mean price movement surrounding the break. The middle panel of the same plot illustrates a uniformly wider confidence band of the disparity in the mean function, implying that the estimate of the difference has more variability around the second structural break.

The bottom panel of the same figure examines the conditional variance function. The confidence band for the conditional variance is uniformly much narrower than the bands for the disparity in mean level. This shows that the conditional variance has been more consistent throughout the time horizon, further justifying the absence of a structural break. Additionally, we observe a similar trend as in the conditional mean: lower news attention correlates with greater volatility, as seen by wider fluctuations in the variance. Conversely, with increased news coverage, the volatility stabilizes,

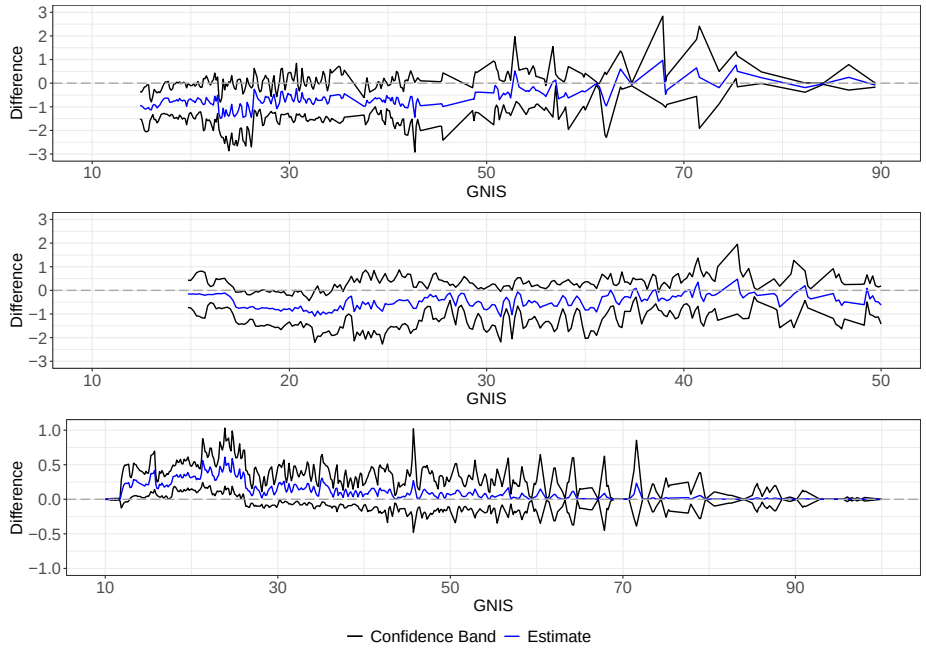


Figure 2.2: (Top) Confidence band of the disparity of mean regression functions before and after the first structural break. (Middle) Confidence band of the disparity of mean regression functions before and after the second structural break. (Bottom) Confidence interval of the conditional variance function.

demonstrating that as the media focuses more on Bitcoin, the volatility in its price becomes more predictable and less prone to sudden changes. Overall, the figure underscores how news sentiment impacts both the mean and variance of Bitcoin prices, with greater news attention leading to more stable price behavior.

2.6 Conclusion

In this chapter, we presented a test designed to detect structural breaks occurring within the middle of a dataset. The theoretical foundations of the test's asymptotic properties are rigorously established. Additionally, the test can be adapted to identify structural breaks in the conditional mean, conditional variance, or both,

at unknown points in time. The effectiveness of the proposed test is demonstrated through a comprehensive simulation study, which covers a wide array of linear and nonlinear data-generating processes, incorporating various contamination levels (thin-tailed, moderately heavy-tailed, and heavy-tailed distributions). The simulation results indicate that our algorithm performs robustly in detecting structural breaks at unknown locations.

Note that even though in the current work we only focus on detecting structural breaks in the conditional mean and variance only, the same theoretical idea can be utilized to extend the method to detect breaks in higher order conditional moment functions as well. We leave the detailed derivations to a future work. It is also possible to derive similar results using the local linear [48] or spline [49] estimator instead of a kernel estimator with slight modifications in the bandwidth conditions. Additionally, our current study focuses on univariate response series, but one may attempt to extent the proposed methodology to high-dimensional setting. This approach would enable the modeling of structural breaks as part of a broader stochastic process, potentially offering richer insights into the underlying dynamics of financial data.

2.7 Appendix: Prerequisite lemmas

Lemma 2.1. *Define*

$$\mathcal{I}_n(x) = \sum_{\mathcal{R}} \{f_X(x \mid \mathcal{F}_{t-1}) - E[f_X(x \mid \mathcal{F}_{t-1})]\}.$$

Recall the definition of Ξ_n *from Assumption 2. Then, for a fixed* $x \in \mathbb{R}$ *and* $\Delta > 0$

$$\left\| \sup_{|x| \leqslant \Delta} \{|\mathcal{I}_n(x)|\} \right\|_2 = O\left(\sqrt{\Xi_n}\right).$$

Proof. The proof follows from Theorem 1 of [50]. □

Lemma 2.2. *Let* f_1 *and* f_2 *be measurable functions such that* $f_2(\epsilon_0) \in \mathcal{L}^2$ *and each of* $f_1(x)$ *and* $f_X(x)$ *belongs to* $\mathcal{C}^0(x-\delta, x+\delta)$ *for some* $\delta > 0$*. Further, assume that* f_1 *and* f_X *do not vanish anywhere on* $\mathcal{X}$*. Define, for any* $K \in \mathcal{K}$*,*

$$\upsilon_t(x) = \frac{f_1(X_t)\left[f_2(\epsilon_t) - E(f_2(\epsilon_t))\right] K_{b_n}(x - X_t)}{f_1(x)\sqrt{nb_n V(f_2(\epsilon_0))\phi(K) f_X(x)}}.$$

Then, assuming $b_n \to 0, nb_n \to \infty$ *and* $n^{-2}\Xi_n \to 0$ *as* $n \to \infty$*, we have for a fixed* $x \in \mathcal{X}$*,*

$$S_n(x) = \sum_{\mathcal{R}} \upsilon_t(x) \xrightarrow{d} \mathcal{N}(0,1).$$

Proof. Due to the independence of $\{X_t\}$ and $\{\epsilon_t\}$, it is straightforward to show that $\{\upsilon_t\}$ forms a martingale difference sequence with respect to the filtration $\{\mathcal{G}_t\}$. Now, define

$$\gamma_t(x) = f_1^2(X_t) K_{b_n}(x - X_t).$$

Setting $U_t = \gamma_t(x) - E(\gamma_t(x) \mid \mathcal{F}_{t-1})$ and $V_t = E(\gamma_t(x) \mid \mathcal{F}_{t-1}) - E(\gamma_t)$, we can write

$$\sum_{\mathcal{R}} \{\gamma_t - E(\gamma_t)\} = \sum_{\mathcal{R}} U_t + \sum_{\mathcal{R}} V_t.$$

Since $\{U_t\}$ forms a martingale difference sequence with respect to the filtration $\{\mathcal{F}_t\}$ and $E(U_t^2) = O(b_n)$, we can write

$$\sum_{\mathcal{R}} U_t = O_p(\sqrt{nb_n}).$$

On the other hand, taking advantage of the fact that $f_1(x) \in \mathcal{C}^0(x-\delta, x+\delta)$ and $K \in \mathcal{K}$, Lemma 2.1 implies that

$$\left\| \sum_{\mathcal{R}} V_t \right\|_2 = O\left(b_n \sqrt{\Xi_n} \right).$$

Simple calculations show that since $b_n \to 0$, $nb_n \to \infty$ and $n^{-2}\Xi_n \to 0$,

$$\sum_{\mathcal{R}} E\left(\upsilon_t^2 \mid \mathcal{G}_{t-1} \right) = \frac{\sum_{\mathcal{R}} U_t + \sum_{\mathcal{R}} V_t + \sum_{\mathcal{R}} E(\gamma_t)}{nb_n \phi(K) f_X(x) f_1^2(x)} \xrightarrow{P} 1. \tag{2.14}$$

Next, since $f_1(x) \in \mathcal{C}^0(x-\delta, x+\delta)$ and K is bounded, we must have, for some $c > 0$ and sufficiently large n,

$$\sup_{u \in \mathbb{R}} \{ |f_1(u) K_{b_n}(x-u)| \} \leqslant c.$$

For any $s > 0$, define $d(x) = c^{-1} s f_1(x) \sqrt{nb_n V(f_2(\epsilon_0)) \phi(K) f_X(x)}$. Again, exploiting the independence of $\{X_t\}$ and $\{\epsilon_t\}$, we deduce

$$\begin{aligned} &\sum_{\mathcal{R}} E\left(\upsilon_t^2(x) \mathbf{1}_{\{|\upsilon_t(x)| \geqslant s\}} \right) \leqslant \\ &\qquad \frac{E\left(\gamma(X_t)\right) E\left(\left[f_2(\epsilon_0) - E(f_2(\epsilon_0)) \right]^2 \mathbf{1}_{\{|f_2(\epsilon_0) - E(f_2(\epsilon_0))| \geqslant d(x)\}} \right)}{b_n V(f_2(\epsilon_0)) f_1^2(x) \phi(K) f_X(x)} \to 0. \end{aligned} \tag{2.15}$$

Finally, combining (2.14) and (2.15), we can conclude that the sequence $\{\upsilon_t(x)\}$ satisfies the conditions for martingale central limit theorem, which implies the statement of the lemma. □

Lemma 2.3. *Let $K \in \mathcal{K}$ and $f_2(\epsilon_0) \in \mathcal{L}^3$. Assume that f_X and f_1 never vanish on $\mathcal{X}$, and each of f_X and f_1 belongs to $\mathcal{C}^4(\mathcal{X}(\delta))$ for some $\delta > 0$. Choose the bandwidth b_n such that*

$$b_n^{\frac{4}{3}} \log n + \frac{(\log n)^3}{n b_n^3} + \frac{\Xi_n (\log n)^2}{n^2 b_n^{\frac{4}{3}}} \xrightarrow{n\to\infty} 0. \tag{2.16}$$

Recall $S_n(x)$ from Lemma 2.2. Then

$$\lim_{n\to\infty} P\left(\sup_{x\in\Pi_n} \{|S_n(x)|\} \leqslant \mathcal{B}_{m_n}(z) \right) = e^{-2e^{-z}},$$

where $\{\Pi_n\}$ and $\mathcal{B}_n(\cdot)$ are as defined in Section 2.3.2.

Proof. For a fixed $x \in \mathcal{X}$, recall the definition of $\{\upsilon_t(x)\}$ from Lemma 2.2. Choose k random points from Π_n and denote them by $x_{t_{j_1}}, x_{t_{j_2}}, ..., x_{t_{j_k}}$. Set $\mathbf{x} = \left(x_{t_{j_1}}, x_{t_{j_2}}, ..., x_{t_{j_k}}\right)'$, and for each $l = 1, 2, \ldots, k$ define $S_n(x_{t_{j_l}}) = \sum_{\mathcal{R}} \upsilon_t(x_{j_l})$ and $\mathbf{S}_{n,k}(\mathbf{x}) = \left(S_n(x_{t_{j_1}}), S_n(x_{t_{j_2}}), ..., S_n(x_{t_{j_k}})\right)'$. Without loss of generality, we may assume that the first p elements of $\mathbf{x}$ are from $\mathcal{X}_-$ and rest are from $\mathcal{X}_+$, where $\mathcal{X}_-$ and $\mathcal{X}_+$ respectively denote the ranges of the covariate X in the segments $\mathcal{T}_-$ and $\mathcal{T}_+$. Let $\mathcal{Q}$ be the quadratic characteristic matrix of $\mathbf{S}_{n,k}$, i.e., for any $\mathbf{z} \in \mathbb{R}^k$,

$$\mathcal{Q}(\mathbf{z}) = \sum_{\mathcal{R}} E\left(\upsilon_t(\mathbf{z}) \upsilon_t^{\intercal}(\mathbf{z}) \middle| \mathcal{G}_{t-1} \right).$$

Since the support of the kernel K is $[-1, 1]$, it is clear that $\mathcal{Q}$ is a diagonal matrix. The diagonal terms of $\mathcal{Q}$ are given by

$$\mathcal{Q}_{r,r} = \frac{1}{f_1^2(x_{t_{j_r}}) n^{\alpha} b_n \phi(K) f_X(x_{t_{j_r}})} \sum_{\mathcal{R}} E\left[\gamma_t(x_{t_{j_r}}) \mid \mathcal{F}_{t-1}\right],$$

for each $r = 1, 2, \ldots, k$. Now, define,

$$\begin{aligned}
\vartheta_t(r) &= f_1^2(X_t) K_{b_n}^2\left(x_{t_{j_r}} - X_t\right) - E\left[f_1^2(X_t) K_{b_n}^2\left(x_{t_{j_r}} - X_t\right) \mid \mathcal{F}_{t-1}\right], \\
\varrho_t(r) &= E\left[f_1^2(X_t) K_{b_n}^2\left(x_{t_{j_r}} - X_t\right) \mid \mathcal{F}_{t-1}\right] - E\left[f_1^2(X_t) K_{b_n}^2\left(x_{t_{j_r}} - X_t\right)\right].
\end{aligned}$$

Using an argument similar to that used in Lemma 2.2, it is straightforward to show that

$$\left\|\sum_{\mathcal{R}} \vartheta_t(r)\right\|_2 = O\left(\sqrt{n b_n}\right) \text{ and } \left\|\sum_{\mathcal{R}} \varrho_t(r)\right\|_2 = O\left(b_n \sqrt{\Xi_n}\right). \tag{2.17}$$

On the other hand, using Taylor's expansion, we can write

$$\left|\sum_{\mathcal{R}} E\left[f_1^2(X_t) K_{b_n}^2\left(x_{t_{j_r}} - X_t\right)\right] - n b_n \mathcal{Q}_{r,r}\right| = O\left(n b_n^3\right). \tag{2.18}$$

Combining (2.17) and (2.18), we have,

$$E\left(|\mathcal{Q}_{r,r'} - \mathcal{I}_{r,r'}|^{\frac{3}{2}}\right) = O\left(\left(\frac{1}{\sqrt{n b_n}} + b_n^4 + \frac{\sqrt{\Xi_n}}{n}\right)^{\frac{3}{2}}\right) \forall\, r, r', \tag{2.19}$$

where $\mathcal{I}$ is the $k \times k$ identity matrix. Elementary calculations also show that

$$\sum_{\mathcal{R}} \left| \upsilon_t(x_{t_{j_r}}) \right|^3 = O\left(\frac{1}{\sqrt{nb_n}} \right), \tag{2.20}$$

so that (2.19) and (2.20) together imply

$$E\left(|\mathcal{Q}_{r,r'} - \mathcal{I}_{r,r'}|^{\frac{3}{2}} \right) + \sum_{\mathcal{R}} \left| \upsilon_t(x_{t_{j_r}}) \right|^3 = \frac{1}{\sqrt{nb_n}} + b_n^3 + \frac{\Xi_n^{\frac{3}{4}}}{n^{\frac{3}{2}}}.$$

Next, let $\mathcal{A}_{j_r}$ be the event $\left\{ \left| S_n(x_{t_{j_r}}) \right| \geqslant \mathcal{B}_{m_n}(z) \right\}$ and $\mathcal{E}_{m_n} = \bigcup_{j=0}^{m_n} \mathcal{A}_{j_r}$. Under (2.16), it can be easily verified that, for any fixed $x \in \mathcal{X}$,

$$\left(\frac{1}{\sqrt{nb_n}} + b_n^3 + \frac{\Xi_n^{\frac{3}{4}}}{n^{\frac{3}{2}}} \right) \{1 + \mathcal{B}_{m_n}(z)\}^4 \, e^{\frac{\mathcal{B}_{m_n}^2(z)}{2}} \xrightarrow{n \to \infty} 0.$$

Now, using Theorem 1 of [39], we can write

$$P\left(\bigcap_{r=1}^{k} \mathcal{A}_{j_r} \right) = \left(\frac{2e^{-z}}{m_n} \right)^k (1 + o(1)).$$

The lemma hence follows using the principle of inclusion and exclusion. □

Chapter 3

Nonparametric shape-constrained modeling

3.1 Background

Consider a time series dataset denoted as $\{(Y_t, X_t)\}_{t=1}^n$, where Y_t and X_t are real-valued time series. We model this dataset according to the same structure described in Chapter 2 as:

$$Y_t = \mu(X_t) + \sigma(X_t)\epsilon_t. \tag{3.1}$$

We assume that the error terms $\{\epsilon_t\}$ are independent and identically distributed with zero mean and finite variance and that they are independent of the covariates $\{X_t\}$. Note that in contrast to Chapter 2, we make no smoothness assumptions on the conditional mean and variance. A common approach is to estimate them by the usual Nadaraya-Watson kernel estimate as follows.

$$\widehat{\mu}(x) = \frac{1}{nb(n)\widehat{f}_{b(n)}(x)} \sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{b(n)}\right), x \in \mathbb{R}, \tag{3.2}$$

$$\widehat{\sigma}^2(x) = \frac{1}{nh(n)\widehat{f}_{h(n)}(x)} \sum_{i=1}^{n} \left(Y_i - \widehat{\mu}(X_i)\right)^2 K\left(\frac{x - X_i}{h(n)}\right), x \in \mathbb{R}, \tag{3.3}$$

where $b(.)$ and $h(.)$ are bandwidth parameters (dependent on the sample size n), $K(.)$ is the kernel function and $\widehat{f_a}(x) = \frac{1}{na}\sum_{i=1}^n K\left(\frac{x-X_i}{a}\right)$ is the estimated density of the covariate X at point $x \in \mathbb{R}$ calculated using bandwidth $a \in \mathbb{R}$. We assume that the covariate $\{X_t\}$ is bounded in $\mathcal{X} \subset \mathbb{R}$, and that $\widehat{\mu}$ and $\widehat{\sigma}$ are both bounded measurable lower-semicontinuous functions. Note that all our results hold true if we use the local

linear estimator (see [51]) instead of the kernel estimator.

Now, consider a scenario where we have prior knowledge about the structural form of the true conditional mean or conditional variance function. In many financial modeling applications, such prior knowledge may pertain to the shape of the function- specifically, properties such as monotonicity, quasiconvexity, or a combination of the two. However, the nonparametric estimators typically employed for these functions do not inherently satisfy these shape constraints unless they are explicitly enforced. A number of existing works- such as [52], [53], [54], and [55], have proposed approaches that impose shape restrictions directly within the estimation procedure. These methods typically design estimators to satisfy the desired shape properties by embedding explicit constraints into the estimation process. While effective in ensuring compliance with the assumed shape, this approach can sometimes compromise the estimator's statistical efficiency or interpretability, especially when the constraints are strict or misaligned with the underlying data structure. In contrast, we propse a methodology to incorporate shape information in an ex-ante fashion through well-defined shape-enforcing operators. Instead of constraining the estimation process from the outset, we begin with an initial unconstrained estimator and subsequently apply a shape-enforcing operator to project the estimator onto the space of functions satisfying the specified shape constraint. This two-step approach allows for greater flexibility and preserves much of the statistical richness of the original estimator while ensuring the final estimate adheres to the structural assumptions. Moreover, while the majority of prior literature has focused primarily on monotonicity and convexity constraints, our framework accommodates more general shape restrictions, such as quasiconvexity or hybrid constraints, thereby offering a broader and more adaptable modeling paradigm

for financial data.

Global estimation methods aim to approximate the entire functional form of the conditional mean function using a parametric or semi-parametric basis. For instance, the conditional mean function might be expressed as a weighted sum of sinusoidal basis functions, with the weights estimated via least squares. Common approaches in this category include polynomial regression, trigonometric expansions, and the Fourier Flexible Form (FFF). For a detailed overview of such global estimation techniques—both nonparametric and semiparametric—the reader is referred to [56] and [57]. On the other hand, local estimation methods focus on estimating the conditional mean function pointwise, without requiring specification of the global functional form. These include techniques such as kernel smoothing and local linear regression, which offer greater flexibility in capturing localized features of the data. A notable advantage of our proposed framework is its generality: both global and local estimators can be used as the initial input to the shape-enforcing operators. We also demonstrate that these operators can yield substantial improvements in adherence to the desired shape constraints without degrading the underlying statistical properties of the original estimator.

The rest of the chapter is organized as follows. In Section 3.2, we define some notations to be commonly used throughout the chapter and revisit some well known definitions required to develop the theory. In Section 3.3, we define the shape-enforcing operators and theoretically establish their superiority over common nonparametric estimates under shape information available for the corresponding population functions. As specified before, our main focus would be monotonicity and/or quasiconvexity. Finally

in Section 3.4 we demonstrate the performance of the shape-enforced estimates in synthetic data, and conclude with some necessary remarks in Section 3.5.

3.2 Notations and definitions

We denote the range of the true function $\theta(\cdot)$ by $\mathcal{B}$, where $\mathcal{B} \subset \mathbb{R}$. For any set $\mathcal{A} \subset \mathbb{R}$, the Lebesgue measure of $\mathcal{A}$ is denoted by $\mathcal{M}(\mathcal{A})$. The set of all bounded, measurable, real-valued functions defined on χ is denoted by $\mathcal{L}^\infty$. Within this, we denote by $\mathcal{L}^\infty_{\mathbb{M}}$ the collection of all bounded, measurable, real-valued, and weakly increasing functions on χ. Similarly, $\mathcal{L}^\infty_{\mathbb{S}}$ denotes the set of all bounded, measurable, real-valued, lower-semicontinuous functions on χ, and $\mathcal{L}^\infty_{\mathbb{Q}}$ denotes the set of all bounded, measurable, real-valued, lower-semicontinuous quasiconvex functions on χ. The intersection of these two classes, denoted by $\mathcal{L}^\infty_{\mathbb{QM}}$, is given by $\mathcal{L}^\infty_{\mathbb{QM}} = \mathcal{L}^\infty_{\mathbb{Q}} \cap \mathcal{L}^\infty_{\mathbb{M}}$. For a function $f : \chi \to \mathbb{R}$, we define the p-th order Euclidean norm as $||f||p = (\int \chi |f(x)|^p, dx)^{1/p}$ for $p \in [1, \infty)$. The supremum norm is given by $||f||\infty = \sup_{x \in \chi} |f(x)|$. Correspondingly, the p-th order Euclidean distance between two functions f and g mapping χ to $\mathbb{R}$ is defined as $d_p(f, g) = ||f - g||_p$, for $p > 1$. For any function $f : \chi \to \mathbb{R}$, the lower-contour set at level $y \in \mathbb{R}$ is defined as $\mathcal{I}_f(y) = x \in \chi \mid f(x) \leq y$. Finally, for any set $\mathcal{A}$, we use $\text{conv}(\mathcal{A})$ to denote its convex hull, which is the smallest convex set that contains $\mathcal{A}$.

Definition 3.1 (Shape enforcing operator)**.** *Let $\mathcal{L}^\infty_0$ and $\mathcal{L}^\infty_1$ be two subsets of the space $\mathcal{L}^\infty$. An operator O is said to be $\mathcal{L}^\infty_1$-enforcing on $\mathcal{L}^\infty_0$ with respect to a given distance or semi-distance function ρ defined on $\mathcal{L}^\infty$, if it satisfies the following four properties:*

1. ***Reshaping***: *For every function f in $\mathcal{L}_0^\infty$, the transformed function $O(f)$ belongs to $\mathcal{L}_1^\infty$. In other words, the operator O maps functions in $\mathcal{L}_0^\infty$ into the shape-constrained class $\mathcal{L}_1^\infty$.*

2. ***Invariance***: *If a function f already belongs to the shape-constrained class $\mathcal{L}_1^\infty$, then applying the operator O does not alter it. That is, $O(f) = f$ for any f in $\mathcal{L}_1^\infty$.*

3. ***Order preservation***: *The operator O respects the pointwise ordering between functions. Specifically, if two functions f and g in $\mathcal{L}_0^\infty$ satisfy $f \leq g$ (meaning $f(x) \leq g(x)$ for all x in the domain), then the transformed functions must also satisfy $O(f) \leq O(g)$.*

4. ***Distance reduction***: *The application of the operator O does not increase the distance between any two functions. That is, for any f, g in $\mathcal{L}_0^\infty$, the distance $\rho(O(f), O(g))$ is less than or equal to the original distance $\rho(f, g)$.*

Definition 3.2 (Equimeasurable rearrangement of a function)**.** *Define for any measurable function $f : E \to \mathbb{R}$,*

$$\lambda_f(y) = \mathcal{M}\left(\left\{x \in E \middle| |f(x)| > y\right\}\right), y > 0.$$

Note that if $\mathcal{M}(E) = \infty$ then we additionally assume that $\lambda_f(y) < \infty$ for any non-negative real value of y.

Definition 3.3 (Non-increasing rearrangement)**.** *A function f^* is said to be the non-increasing rearrangement of a function f if it satisfies the following two conditions.*

1. *The function f^* is non-increasing on the interval $(0, \mathcal{M}(E))$, where $\mathcal{M}(E)$ denotes the Lebesgue measure of the set E.*

2. *The function f^* is equimeasurable with the absolute value of f, which means that for every $y > 0$, the measure of the set where f^* exceeds y equals the measure of the set where $|f|$ exceeds y. Formally, this can be written as:*

$$\lambda_{f^*}(y) = \mathcal{M}\left(\left\{t \in (0, \mathcal{M}(E)) \,\middle|\, f^*(t) > y\right\}\right) = \mathcal{M}\left(\left\{x \in E \,\middle|\, |f(x)| > y\right\}\right) = \lambda_f(y).$$

Condition (2) above implies that the two distribution functions of the original and rearranged function are pointwise same. Note that the above definition does not define the non-increasing rearrangement uniquely, it can take different values at the points of discontinuities. For definiteness we assume in addition that f^* is left-continuous everywhere in $(0, \mathcal{M}(E))$. Then the relation between the distribution function and the non-increasing rearrangement is given by

$$f^*(t) = \inf_{y>0} \{\lambda_f(y) < t\}, t \in (0, \mathcal{M}(E)).$$

This shows that in a certain sense, the non-increasing rearrangement is the inverse of the distribution function.

The following equivalent definitions of equimeasurable non-increasing rearrangement is taken from [58] and [59].

$$f^*(t) = \sup_{\{F \subset E | \mathcal{M}(F) = t\}} \left\{ \inf_{x \in F} \{|f(x)|\} \right\}, t \in (0, \mathcal{M}(E)). \tag{3.4}$$

Similarly, a non-decreasing equimeasurable rearrangement is defined as

$$f_*(t) = \inf_{\{F \subset E | \mathcal{M}(F)=t\}} \left\{ \sup_{x \in F} \{|f(x)|\} \right\}. \tag{3.5}$$

Clearly, f_* is non-negative, equimeasurable with $|f|$ on E and non-decreasing on $(0, \mathcal{M}(E))$. It can be shown that both the supremum and infimum in Equation (3.4) and Equation (3.5) are attained. Note that the effectiveness of equimeasurable rearrangement of functions is the fact that in certain cases they preserve the properties of the original functions and often has a simpler form.

Another non-increasing equimeasurable rearrangement of f is given by

$$f_d(t) = \sup_{\{F \subset E | \mathcal{M}(F)=t\}} \left\{ \inf_{x \in F} f(x) \right\}, t \in (0, \mathcal{M}(E)). \tag{3.6}$$

Note that f_d is equimeasurable with f while f^* was equimeasurable with $|f|$, all other properties of f^* and f_d (non-increasing, left-continuous on $(0, \mathcal{M}(E))$ etc.) remain same. Clearly, if f is a non-negative function then f^* and f_d coincide.

Definition 3.4 (Submodular (discrepancy) functions)**.** *For two sets A and B of the same sample space $\mathbf{S}$ such that $A \subset B$ and for any element $x \in \mathbf{S} \setminus (A \cup B)$, a submodular function f must satisfy*

$$f(A \cup \{x\}) - f(A) \geq f(B \cup \{x\}) - f(B).$$

A submodular function in general is defined as a set function (a function whose domain is a family of subsets of a given set that usually takes values from the extended real

line) whose value, informally, has the property that the difference in the incremental value of the function that a single element makes when added to an input set decreases as the size of the input set increases.

We finally state the following equivalent definition formally proposed in the works on [60].

Definition 3.5. *Let $L : \mathbb{R}^2 \to \mathbb{R}_+$ be a measurable function. L is called a submodular function if for each pair (v, t) and (v', t') in $\mathbb{R}^2$ we have,*

$$L\left(\min\{v, v'\}, \min\{t, t'\}\right) + L\left(\max\{v, v'\}, \max\{t, t'\}\right) \leq L(v, t) + L(v', t'). \tag{3.7}$$

Equation (3.7) means that a function L measuring the discrepancy between pairs of vector is submodular if co-monotonization of the pair reduces the discrepancy. A popular example of such function is the power function i.e. $L(v, t) = |v - t|^p$ for some $p > 0$. Many other loss functions can also be shown to be submodular.

Definition 3.6 (Quasiconvex functions)**.** *A function $f : \chi \to \mathbb{R}$ is called quasiconvex if and only if all its lower-contour sets are convex.*

There are many other equivalent definitions of quasiconvexity, we in this document would mainly work with the above definition only. Note that quasiconvexity is a global property of a function which is weaker than convexity. A convex function is always quasiconvex but a quasiconvex function may not always be convex. Quasiconvex functions are widely used in economics as utility and production functions, see [61] and [62] for some example applications. They are also preferred in many applications

because of their good optimization properties.

Definition 3.7 (Compact sets (sequential definition))**.** *A set $\mathcal{A}$ is called compact if and only if every sequence in the set $\mathcal{A}$ has a convergent subsequence with respect to a well-defined metric.*

3.3 Mathematical framework

3.3.1 Monotonicity enforcement

Suppose that the true parameter function $\theta(.)$ is monotonic and the corresponding estimate $T(.)$ is not monotonic. Rearrangement is a popular method to transform the original estimate to a monotonic estimate that is potentially better both as a point and interval estimate with respect to a common metric.

We define,

$$T^*(x) = \inf_{y \in \mathbb{R}} \left\{ \int_{\chi} 1_{\{T(u) \leq y\}} du \geq x \right\} \tag{3.8}$$

We can write

$$T^*(x) = \inf_{y \in \mathbb{R}} \left\{ \int_{\chi} 1_{\{T(u) \leq y\}} du \geq x \right\} = \inf_{y \in \mathbb{R}} \left\{ \mathcal{M}\left(\left\{ z \in \mathcal{B} \middle| z \leq y \right\} \right) \geq x \right\}, \tag{3.9}$$

Equation (3.9) shows that $T^*(x)$ can be thought of the x-th quantile of $T(x)$'s, which is by definition a weakly increasing function of x.

Definition 3.8 ($\mathbb{M}$-operator)**.** *The monotonicity enforcing operator ($\mathbb{M}$-operator) is defined as $\mathbb{M} : T \to T^*$.*

In fact, when $T(.)$ is known to be continuous (in case of a global estimation), we can think of the $\mathbb{M}$-operator as a sorting operator in the following way. Given the values of the function $T(.)$ at x in a fine enough net of equidistant points, we simply sort the values in an increasing order to create the rearranged function T^*. We begin by showing that the $\mathbb{M}$-operator is distance contracting.

Theorem 3.1. *$\mathbb{M}$-operator is $\mathcal{L}^{\infty}_{\mathbb{M}}$ enforcing on $\mathcal{L}^{\infty}$ with respect to d_p-metric where $p > 1$.*

Proof.

Let the target function $\theta : \chi \to \mathcal{B}$ be a weakly increasing measurable function and let $T : \chi \to \mathcal{B}_T$ be estimate of θ. We first show that for any $p \geq 1$,

$$||T^* - \theta||_p \leq ||T - \theta||_p \tag{3.10}$$

To prove the Equation (3.10), first assume that both T and θ are step functions taking constant values in the intervals $\left(0, \frac{1}{r}\right]$, $\left(\frac{1}{r}, \frac{2}{r}\right]$,...,$\left(\frac{r-1}{r}, 1\right]$, for some fixed integer r. We denote by θ_s and T_s the values of θ and T respectively in the interval $\left(\frac{s-1}{r}, \frac{s}{r}\right]$, $s = 1(1)r$. We write a step function f as $f = (f_1, f_2, ..., f_n)$ where f_i is the value of the function f at the i-th step. Clearly if in such case we integrate f over its domain $\mathcal{D}_f \subset \mathbb{R}$ with respect to Lesbegue measure we would have $\int_{\mathcal{D}_f} f d\mathcal{M} = \sum_{i=1}^{n} f_i$.

We define the operator $\mathcal{S}$ as follows. Let $k \in \{1, 2, ..., r\}$ be such that $T_k > T_m$ for some $m > k$. If such k does not exist (i.e. T_k is the step with the minimum value)

then we write $\mathcal{S}(T) = T$. Otherwise we define

$$\mathcal{S}(T) = (T_1, T_2, ..., T_{k-1}, T_m, T_{k+1}, ..., T_{m-1}, T_k, T_{m+1}, ..., T_r).$$

Since the true function θ is known to be weakly increasing so we must have $\theta_k \leq \theta_m$. Now suppose we have $T_k \geq T_m$. Then for any submodular function L we must have (using Equation (3.7)),

$$\begin{aligned} &L\left(\min\{T_k, T_m\}, \min\{\theta_k, \theta_m\}\right) + L\left(\max\{T_k, T_m\}, \max\{\theta_k, \theta_m\}\right) \\ &= L\left(T_m, \theta_k\right) + L\left(T_k, \theta_m\right) \leq L\left(T_k, \theta_k\right) + L\left(T_m, \theta_m\right) \end{aligned} \tag{3.11}$$

Now note that

$$\mathcal{S}(T) = (T_1(x), T_2(x), ..., T_{k-1}(x), T_m(x), T_{k+1}(x)..., T_{m-1}(x), T_k(x), T_{m+1}(x)..., T_r(x))$$

and $\theta = (\theta_1(x), \theta_2(x), ..., \theta_{k-1}(x), \theta_k(x), \theta_{k+1}(x), ..., \theta_{m-1}(x), \theta_m(x), \theta_{m+1}(x), ..., \theta_r(x))$ (since θ is weakly increasing). Therefore using Equation (3.11) we must have,

$$\begin{aligned} &\int_{\chi} L(\mathcal{S}(T(x)), \theta(x) dx \\ &= L(T_1, \theta_1)(x) + ... + L(T_{k-1}, \theta_{k-1})(x) + L(T_m, \theta_k)(x) + L(T_{k+1}, \theta_{k+1})(x) + ...+ \\ &L(T_{m-1}, \theta_{m-1})(x) + L(T_k, \theta_m)(x) + L(T_{m+1}, \theta_{m+1})(x) + ...L(T_r, \theta_r)(x) \\ &\leq L(T_1, \theta_1)(x) + ... + L(T_{k-1}, \theta_{k-1})(x) + L(T_k, \theta_k)(x) + L(T_{k+1}, \theta_{k+1})(x) + ... \end{aligned}$$

$$+L(T_{m-1},\theta_{m-1})(x)+L(T_m,\theta_m)(x)+L(T_{m+1},\theta_{m+1})(x)+...L(T_r,\theta_r)(x)$$
$$=\int_\chi L\left((T_1,...T_r),(\theta_1,...,\theta_r)\right)(x)dx=\int_\chi L\left(T,\theta\right)(x)dx.$$

Therefore we must have

$$\int_\chi L(\mathcal{S}(T),\theta)(x)dx\leq\int_\chi L(T,\theta)(x)dx. \tag{3.12}$$

Now note that the complete rearrangement of T (denoted as T^*) can be achieved through repeated (but finite number of times) application of $\mathcal{S}$ on T, i.e.

$$T^*=\mathcal{S}\circ...\circ\mathcal{S}(T). \tag{3.13}$$

Therefore repeating the argument in Equation (3.12) we would get

$$\int_\chi L\left(T^*(x),\theta(x)\right)dx=\int_\chi L\left(\mathcal{S}\circ...\circ\mathcal{S}(T)(x),\theta(x)\right)dx\leq\int_\chi L\left(T(x),\theta(x)\right)dx. \tag{3.14}$$

Using the fact that $\mathbb{M}\equiv\mathcal{S}\circ...\circ\mathcal{S}$ and taking L to be the power function given by $L(f,g)=|f-g|^p$ for any two functions f, g on the same domain with $p\in[1,\infty)$ we get the desired result for the step-function case.

Now we prove the result for a general measurable functions T and θ. As stated previously, we assume the domain of both the functions to be $[0,1]$, so that we can think of θ to be a quantile function. Using Lemma 3.4 let $\{T^{(q)}\}_q$ and $\{\theta^{(q)}\}_q$ be two sequences of bounded step functions converging to T and θ almost everywhere respectively. Note that $\{\theta^{(q)}\}_q$ is a sequence of quantile functions. Since $\{T^{(q)}\}_q$ con-

verges to T almost everywhere then the quantile functions of $\{T^{(q)}\}_q$ would converge to the quantile function of T almost everywhere. Now note that by definition the quantile function of $T(q)$ is $T^{*(q)}$, where $T^{*(q)} = \mathbb{M}(T^{(q)})$. Thus using Equation (3.14) and taking L to be the power function we would have,

$$\int_\chi L\left(T^{*(q)}(x), \theta(x)\right) dx \leq \int_\chi L\left(T^{(q)}(x), \theta(x)\right) dx, q \geq 1$$
$$\implies ||T^{*(q)}(x) - \theta(x)||_p \leq ||T^{(q)}(x) - \theta(x)||_p, p \in [1, \infty)$$

Using dominated convergence theorem,

$$||T^*(x) - \theta(x)||_p \leq ||T(x) - \theta(x)||_p, p \in [1, \infty). \tag{3.15}$$

This proves our claim in Equation (3.10). Now note that

1. (Reshaping) By construction, $\mathbb{M}(T)$ is weakly increasing. Additionally if T is bounded and measurable then its increasing rearrangement $\mathbb{M}(T)$ should also be bounded and measurable.

2. (Invariance) Since a montonically non-decreasing rearrangement of a weakly increasing function is nothing but itself, therefore we must have $\mathbb{M}(T) = T$ if the initial estimate T is already monotonically weakly increasing.

3. (Order preservation) Clearly, for two (bounded, measurable) estimates of the true function θ given by T_1 and T_2 with $T_1(x) \leq T_2(x) \forall x \in \chi$, their corresponding montonically non-decreasing rearrangements $T_1^* = \mathbb{M}(T_1)$ and $T_2^* = \mathbb{M}(T_2)$

respectively will also satisfy $T_1^*(x) \leq T_2^*(x) \forall x \in \chi$.

All the above points and Lemma 3.3 together prove that the $\mathbb{M}$-operator is $\mathcal{L}_{\mathbb{M}}^{\infty}$-enforcing on $\mathcal{L}^{\infty}$ with respect to $d_p, p > 1$. □

We now prove that $T^* = \mathbb{M}(T)$ works better as a point and interval estimate, provided that the true function θ is weakly increasing.

Suppose the confidence interval for T is $[T_L, T_U]$. This requires us to prove the following statements.

$$\mathbf{1}_{\{\theta \in [\mathbb{M}(T_L), \mathbb{M}(T_U)]\}} \geq \mathbf{1}_{\{\theta \in [T_L, T_U]\}} \tag{3.16}$$

$$||\mathbb{M}(T) - \theta||_p \leq ||T - \theta||_p, p \in [1, \infty) \tag{3.17}$$

$$||\mathbb{M}(T_L) - \mathbb{M}(T_U)||_p \leq ||T_L - T_U||_p, p \in [1, \infty) \tag{3.18}$$

To show that Equation (3.16) is true, we first show that

$$\{T_L \leq \theta \leq T_U\} \implies \{\mathbb{M}(T_L) \leq \theta \leq \mathbb{M}(T_U)\} \tag{3.19}$$

Since θ is known to be weakly increasing therefore $\mathbb{M}(\theta) = \theta$. Thus by order preservation,

$$\{T_L \leq \theta \leq T_U\} \implies \{\mathbb{M}(T_L) \leq \mathbb{M}(\theta) \leq \mathbb{M}(T_U)\} = \{\mathbb{M}(T_L) \leq \theta \leq \mathbb{M}(T_U)\}.$$

This implies that $\{T_L \leq \theta \leq T_U\} \subset \{\mathbb{M}(T_L) \leq \theta \leq \mathbb{M}(T_U)\}$, which directly implies Equation (3.16).

Equation (3.17) is a direct consequence of Theorem 1 and Equation (3.18) follows directly from that. This proves that subjecting the original estimate T to the $\mathbb{M}$-operator makes it substantially better both as point and interval estimate.

3.3.2 Quasiconvexity enforcement

Define

$$T^{\dagger}(x) = \min\left\{y \in \mathbb{R} \middle| x \in \text{conv}(I_T(y))\right\}. \tag{3.20}$$

Definition 3.9 ($\mathbb{Q}$-operator)**.** *The quasiconvexity enforcing operator ($\mathbb{Q}$-operator) is defined as $\mathbb{Q} : T \to T^{\dagger}$.*

Theorem 3.2. *The $\mathbb{Q}$-operator exists.*

Proof. Define

$$\mathcal{Q}_T(x) = \left\{y \in \mathbb{R} \middle| x \in \text{conv}\left(\mathcal{I}_T(\,y)\right)\right\} = \left\{y \in \mathbb{R} \middle| x \in \text{conv}\left(\left\{x \in \chi \middle| T(x) \leq y\right\}\right)\right\}.$$

To show that $T^{\dagger}(x)$ exists, we have to show that the infimum of $\mathcal{Q}_T(x)$ exists and is attained. We start by stating the following lemma which we will frequently use throughout the proof. We have,

$$\mathcal{Q}_T(x) = \left\{y \in \mathbb{R} \middle| T(x) \leq y \text{ and some other elements to make the set convex}\right\} \tag{3.21}$$

This implies that $\mathcal{Q}_T(x) \subset \left\{y \in \mathbb{R} \middle| T(x) \leq y\right\} = \mathcal{I}_T(y)$. Using Lemma 3.4 in above equation we must have

$$\inf\left(\mathcal{Q}_T(x)\right) \geq \inf\left(\mathcal{I}_T(y)\right). \tag{3.22}$$

Again,

$$\mathcal{I}_T(y) = \left\{x \in \chi \middle| T(x) \leq y\right\} \subset \left\{T(x) \middle| x \in \chi\right\}.$$

Using Lemma 3.4 in above equation we again have

$$\inf\left(\mathcal{I}_T(y)\right) \geq \inf\{T(x) \middle| x \in \chi\}. \tag{3.23}$$

Using Equation (3.22) and Equation (3.23) we have, $\inf\left(\mathcal{Q}_T(x)\right) \geq \inf\{T(x) \middle| x \in \chi\}$. Now since we assume T is a bounded function i.e. $\inf\left\{T(x) \middle| x \in \chi\right\} > -\infty$, therefore we must have $\inf\left(\mathcal{Q}_T(x)\right) > -\infty$.

Let $y^0 = \inf\left(\mathcal{Q}_T(x)\right)$. To show that $y^0 = \min\left(\mathcal{Q}_T(x)\right)$ we have to show that there exists a sequence of numbers $\{y^{(k)}\}_k$ in $\mathcal{Q}_T(x)$ such that $y^{(k)} \to y^0$ as $k \to \infty$.

Now for each k,

$$y^{(k)} \in \mathcal{Q}_T(x) \implies x \in \text{conv}\left(\left\{\mathcal{I}_T\left(y^{(k)}\right)\right\}\right) \implies x \in \text{conv}\left(\left\{x \in \chi \middle| T(x) \leq y^{(k)}\right\}\right). \tag{3.24}$$

According to Carathéodory's theorem, we will then find $x_1^{(k)}, x_2^{(k)} \in \left\{x \in \chi \middle| T(x) \leq y^{(k)}\right\}$ and $\alpha_1^{(k)}, \alpha_2^{(k)} \in [0,1]$ such that $x = \alpha_1^{(k)} x_1^{(k)} + \alpha_2^{(k)} x_2^{(k)}$. Define, $X^{(k)} = \left(x_1^{(k)}, x_2^{(k)}\right)^\intercal$ and $\alpha^{(k)} = \left(\alpha_1^{(k)} \alpha_2^{(k)}\right)^\intercal$. Consider the sequence $\left\{\left(X^{(k)}, \alpha^{(k)}\right)\right\}_k$ in $\chi \times [0,1]$. Since both χ and $[0,1]$ are compact sets therefore we will find $X^0 = (x_1^0 x_2^0)^\intercal$ and $\alpha^0 = (\alpha_1^0 \alpha_2^0)^\intercal$ such that $X^{(k)} \to X^0$ and $\alpha^{(k)} \to \alpha^0$ as $k \to \infty$. Therefore we have,

$$x = \alpha_1^{(k)} x_1^{(k)} + \alpha_2^{(k)} x_2^{(k)} = lim_{k\to\infty}\left[\alpha_1^{(k)} x_1^{(k)} + \alpha_2^{(k)} x_2^{(k)}\right] = \alpha_1^0 x_1^0 + \alpha_2^0 + x_2^0. \tag{3.25}$$

Now assuming T is a lower semicontinuous function we must have

$$T(x_1^0) \leq \liminf_{k\to\infty} \left[T(x_1^{(k)})\right] \tag{3.26}$$

$$\leq \liminf_{k\to\infty} \left[T(y^{(k)})\right] = y^0. \tag{3.27}$$

Therefore we have $T(x^0) \leq y^0$ and similarly $T(x^0) \leq y^0$, which in turn means that $x_1^0, x_2^0 \in \left\{x \in \chi \middle| T(x) \leq y^0\right\} = \mathcal{I}_T(y^0)$.

So for any $\alpha_1^0, \alpha_2^0 \in [0,1]$ we must have $\alpha_1^0 x_1^0 + \alpha_2^0 x_2^0 \in \text{conv}\left(\mathcal{I}_T(y^0)\right)$. This further means that $y^0 \in \mathcal{Q}_T(x)$ i.e. the infimum is attained for any arbitrarily chosen value of x. Therefore the minimum exists, which proves the existence of the $\mathbb{Q}$-operator.

□

Theorem 3.3. *$\mathbb{Q}$-operator is $\mathcal{L}_{\mathbb{Q}}^{\infty}$ enforcing on $\mathcal{L}_{\mathbb{S}}^{\infty}$ with respect to d_∞ metric.*

Proof. We first argue that $\mathbb{Q}(T)$ is quasiconvex. Note that $x \in \text{conv}\left(\mathcal{I}_T(y)\right)$ if and only if $\mathbb{Q}(T(x)) \leq y$. The lower-contour sets of $\mathbb{Q}(T)$ can be written as,

$$\mathcal{I}_{\mathbb{Q}(T)}(t) = \left\{u \middle| \mathbb{Q}(T(u)) \leq t\right\} = min\left\{u \middle| min\{y | u \in \text{conv}(\mathcal{I}_T(y))\} \leq t\right\} \tag{3.28}$$

$$= \left\{u \middle| u \leq t \text{ and } u \in \text{conv}(\mathcal{I}_T(y))\right\} \tag{3.29}$$

$$= \left\{u \middle| T(u) \leq t \text{ and } u \in \text{conv}\left(\{T(u) \leq t\}\right)\right\} = \text{conv}\left(\mathcal{I}_T(t)\right). \tag{3.30}$$

Since $\text{conv}\left(\mathcal{I}_T(t)\right)$ is convex for all values of t, therefore $\mathcal{I}_{\mathbb{Q}(T)}(t)$ is convex for all values of t. Therefore $\mathbb{Q}(T)$ must be quasiconvex.

Next we show that $\mathbb{Q}(T)$ is bounded and lower semicontinuous. Note that

$$\mathbb{Q}(T(x)) = \min\left\{y\middle|x \in \text{conv}\left(\mathcal{I}_T(y)\right)\right\} = \min\left\{y\middle|x \in \text{conv}\left(\{y|T(x) \leq y\}\right)\right\}.$$

Since $\{y|T(x) \leq y\} \subset \mathcal{B}_T$ (where $\mathcal{B}_T$ is the range of T) therefore we must have $\inf\left(\{T(x)|x \in \chi\}\right) \leq \mathbb{Q}(T(x)) \leq sup\left(\{T(x)|x \in \chi]\}\right)$.This implies that $\mathbb{Q}(T)$ is bounded.

Now let us assume that $\mathbb{Q}(T)$ is not lower semicontinuous. Then there exists a sequence $\{x^{(n)}\}$ in χ converging to x^0 (the convergence is guaranteed by the fact that χ is a a compact set) with $\mathbb{Q}(T(x^0)) = y_0$ but due to non lower semicontinuity,

$$\liminf_{n\to\infty} \mathbb{Q}(T(x^{(n)})) < y_0 - \epsilon,$$

for some $\epsilon > 0$. We denote by

$$y_k = \mathbb{Q}(T(x^{(k)})) = \min\left\{y\middle|x \in \text{conv}\left(\mathcal{I}_T(y_k)\right)\right\} \tag{3.31}$$

Now for each k, $x^{(k)}$ can be written as $x^{(k)} = \alpha_1^{(k)}x_1^{(k)} + \alpha_2^{(k)}x_2^{(k)}$ where $x_1^{(k)}, x_2^{(k)} \in \chi$ and $\alpha_1^{(k)}, \alpha_2^{(k)} \in [0,1]$. Consider the sequneces $\{x_1^{(k)}\}_k$, $\{x_2^{(k)}\}_k$, $\{\alpha_1^{(k)}\}_k$ and $\{\alpha_2^{(k)}\}_k$. Again, since χ and $[0,1]$ are compact sets we will find subsequences $\{x_{1_n}^{(k)}\}$ and $\{x_{2_n}^{(k)}\}$ converging to x_1^0 and x_2^0 ($x_1^0, x_2^0 \in \chi$) respectively. We can similarly find subsequneces $\{\alpha_{1_n}^{(k)}\}$ and $\{\alpha_{2_n}^{(k)}\}$ converging to α_1^0 and α_2^0 ($\alpha_1^0, \alpha_2^0 \in [0,1]$) respectively. Therefore we have, $x^{(k)} \xrightarrow{k\to\infty} x^0$ where $x^0 = \alpha_1^0 x_1^0 + \alpha_2^0 x_2^0$.

Now assuming T is lower semicontinuous we must have,

$$T(x_1^0) \leq \liminf_{k\to\infty} T(x_1^{(k)}) \tag{3.32}$$

Now recall that $x_1^{(k)} \in \text{conv}\left(\mathcal{I}_T(y_k)\right) \implies T(x_1^{(k)}) \leq y_k$. Combining Equation (3.31) and Equation (3.32) we have,

$$T(x_1^0) \leq \liminf_{k\to\infty} T(x_1^{(k)}) \leq \liminf_{k\to\infty} y_k = \liminf_{k\to\infty} \mathbb{Q}(T(x^{(k)})) < y_0 - \epsilon. \tag{3.33}$$

Similarly we would have $T(x_2^{(0)}) < y_0 - \epsilon$. From this we can write,

$$x^0 = \alpha_1^0 x_1^0 + \alpha_2^0 x_2^0 \implies T(x^0) = T(\alpha_1^0 x_1^0 + \alpha_2^0 x_2^0) < y_0 - \epsilon \implies x^0 \in \mathcal{I}_T(y_0 - \epsilon) \tag{3.34}$$

Note that the second inequality is taken directly from [63]. Now, since $x^0 \in \mathcal{I}_T(y_0 - \epsilon)$ therefore we must have,

$$\mathbb{Q}(T(x^0)) = \min\left\{y \middle| x^0 \in \text{conv}\left(\mathcal{I}_T(y)\right)\right\} \leq y_0 - \epsilon < y_0. \tag{3.35}$$

Equation (3.35) implies that $\mathbb{Q}(T)$ is lower-semicontinuous, which is a contradiction to our initial assumption. Since x^0 was arbitrarily chosen therefore we must have $\mathbb{Q}(T)$ is lower-semicontinuous.

Next we want to show that if $T \in \mathcal{L}_{\mathbb{Q}}^{\infty}$ then $\mathbb{Q}(T) = T$. Now if T is quasiconvex this means that $\mathcal{I}_T(y)$ is convex for all values of y. Then the lower contour set of $\mathbb{Q}(T)$ is given by,

$$\begin{aligned} \mathcal{I}_{\mathbb{Q}(T)}(y) &= \left\{x \in \chi \middle| \mathbb{Q}(T(x)) \leq y\right\} \\ &= \left\{x \in \chi \middle| \min\{y^* \middle| x \in \text{conv}(\mathcal{I}_T(y^*))\} \leq y\right\} = \text{conv}\left(\mathcal{I}_T(y)\right), \end{aligned}$$

which is a convex set. Therefore $\mathcal{I}_{\mathbb{Q}(T)}(y)$ is convex for all values of y, implying that

$\mathbb{Q}(T)$ is quasiconvex. Therefore we must have $\mathbb{Q}(T) \in \mathcal{L}_{\mathbb{Q}}^{\infty}$.

To prove the order preservation property, consider two estimates T_1 and T_2 of the parameter function θ such that $T_1, T_2 \in \mathcal{L}_{\mathbb{S}}^{\infty}$ and $T_1(x) \leq T_2(x) \forall x \in \chi$. We want to show that $\mathbb{Q}(T_1(x)) \leq \mathbb{Q}(T_2(x)) \forall x \in \chi$.

Now since $T_1(x) \leq T_2(x) \forall x \in \chi$ we must have,

$$\begin{aligned} \left\{x \in \chi \middle| T_1(x) \leq y\right\} \subset \left\{x \in \chi \middle| T_2(x) \leq y\right\} \forall y \\ \implies \mathcal{I}_{T_1}(y) \subset \mathcal{I}_{T_2}(y) \forall y \\ \implies \text{conv}\left(\mathcal{I}_{T_1}(y)\right) \subset \text{conv}\left(\mathcal{I}_{T_2}(y)\right) \forall y. \end{aligned} \tag{3.36}$$

Therefore we must have,

$$\mathcal{I}_{\mathbb{Q}(T_1)}(y) = \left\{x \in \chi \middle| \mathbb{Q}(T_1(x)) \leq y\right\} = \{x| \text{ all elements of conv}\left(\mathcal{I}_{T_1}(y)\right) \text{ which are} \leq y\} \tag{3.37}$$

$$\subset \{x| \text{ all elements of conv}\left(\mathcal{I}_{T_2}(y)\right) \text{ which are} \leq y\} = \mathcal{I}_{\mathbb{Q}(T_2)}(y). \tag{3.38}$$

Since $\mathcal{I}_{\mathbb{Q}(T_1)}(y) \subset \mathcal{I}_{\mathbb{Q}(T_2)}(y) \forall y$ therefore we must have $\mathbb{Q}(T_1)(y) \leq \mathbb{Q}(T_2)(y) \forall y$. This completes the proof.

Next, to prove the distance contraction, we want to show that the $\mathbb{Q}$-operator is distance contracting with respect to d_{∞} metric. Assume that the true conditional mean function $\mu \in \mathcal{L}_{\mathbb{Q}}^{\infty}$. We want to show that $||\mathbb{Q}(T) - \theta||_{\infty} \leq ||T - \theta||_{\infty}$.

Define, $\delta = ||T - \theta||_\infty = sup_{x \in \chi}\{|T(x) - \theta(x)|\}$. This means,

$$|T(x) - \theta(x)| \leq \delta \forall x \in \chi \implies \theta(x) - \delta \leq T(x) \leq \theta(x) + \delta \forall x \in \chi. \tag{3.39}$$

Now following [63] we must have $\mathbb{Q}(f + c) = \mathbb{Q}(f) + c$ for any $f \in \mathcal{L}_{\mathbb{S}}^{\infty}$ and for any constant c. Therefore due to order preservation we have,

$$\mathbb{Q}(\theta)(x) - \delta \leq \mathbb{Q}(T)(x) \leq \mathbb{Q}(\theta)(x) + \delta \forall x \in \chi \implies |\mathbb{Q}(T)(x) - \mathbb{Q}(\theta)(x)| \leq \delta \forall x \in \chi \tag{3.40}$$

$$\implies sup_{x \in \chi}\{|\mathbb{Q}(T)(x) - \mathbb{Q}(\theta)(x)|\} < \delta = ||T - \theta||_\infty. \tag{3.41}$$

Putting $\mathbb{Q}(\theta) = \theta$ (due to invariance property) in Equation (3.41) we have $||\mathbb{Q}(T) - \theta||_\infty \leq ||T - \theta||_\infty$. □

The proof of $\mathbb{Q}(T)$ being a better point and interval estimate as compared to T can be done in similar lines with the same proof done in case of the monotonicity constraint. Details are skipped as no technical difficulties are involved. The implementation of the operator for practical applications has been illustrated in Algorithm 2.

3.3.3 Joint enforcement of quasiconvexity and monotonicity operator

Define,

$$T^{\Delta}(x) = \mathbb{Q} \circ \mathbb{M}(T). \tag{3.42}$$

Definition 3.10 ($\mathbb{QM}$-operator)**.** *The joint quasiconvexity and monotonicity enforcing operator ($\mathbb{QM}$-operator) is given by $\mathbb{QM} : T \to T^{\Delta}$.*

Theorem 3.4. *$\mathbb{QM}$-operator is $\mathcal{L}_{\mathbb{QM}}^{\infty}$ enforcing on $\mathcal{L}_{\mathbb{S}}^{\infty}$ with respect to the d_∞ metric.*

Procedure 2: Enforcement of $\mathbb{Q}$-operator

Input : A point $x \in \mathcal{X}^n$, and an ordered dataset $\mathcal{Y}^n = \{y_{(1)}, y_{(2)}, \ldots, y_{(n)}\}$
Output: Estimated quantile value $\mathbb{Q}_f(x)$
Function `QOperator(`x, $\mathcal{Y}^n$`):`
  Initialize $y_L \leftarrow y_{(1)}$;
  Initialize $y_U \leftarrow y_{(n)}$;
  while $y_U \neq y_L$
    Define $\mathcal{Y}_{LU} = \{y \in \mathcal{Y}^n : y_L \leq y \leq y_U\}$;
    Compute median $y^* \leftarrow \text{median}(\mathcal{Y}_{LU})$;
    Compute lower contour set $\mathcal{I}_{f,n}(y^*) = \{z \in \mathcal{X}^n : f_n(z) \leq y^*\}$;
    if $x \in \text{conv}[\mathcal{I}_{f,n}(y^*)]$ **then**
      Set $y_U \leftarrow y^*$
    else
      Set $y_L \leftarrow y^*$
    end
  end
  return $\mathbb{Q}_f(x) \leftarrow y_U$

Proof. We first deal with the reshaping property of the $\mathbb{QM}$-operator, i.e. we have to show that $\mathbb{QM}(T) \in \mathcal{L}^{\infty}_{\mathbb{QM}}$ where $T \in \mathcal{L}^{\infty}_{\mathbb{S}}$. Let $g = \mathbb{M}(T)$. Then clearly g is weakly increasing, and hence also quasiconvex. Due to invariance property of $\mathbb{Q}$-operator, $\mathbb{Q}(g)$ would also be quasiconvex. The other properties of $\mathbb{QM}$-operator (invariance, order preservation and distance reduction) can be similarly proved using the fact that $\mathbb{Q}(g) = g$. □

The proof of $\mathbb{QM}(T)$ being a better point and interval estimate as compared to T can be done in similar lines with the same proofs done in the last two sections. We skip the details as no technical difficulties are involved.

3.4 Simulation study

We evaluate the performance of the proposed shape-enforcing operators through extensive Monte Carlo simulations under various data generating processes (DGPs), noise structures, and sample sizes. The population mean function is taken as $\mu(x) = e^x$, which is monotonic and quasiconvex. Specifically, we consider three sample sizes: $n \in \{300, 500, 1000\}$. For the DGP, we use three types of processes: White Noise (where the time series is generated as standard Gaussian), ARMA(1,1) to capture short-range dependence, SETAR (Self-Exciting Threshold Autoregressive) process to introduce regime-switching dynamics based on the value of past observations. Each DGP is simulated under three different noise distributions: standard normal (denoted as *norm* in the simulation tables), t-distribution with 3 degrees of freedom (to model heavy tails), and a power-law noise generated using a Pareto distribution with shape parameter 2.5. For each combination of sample size, DGP, and noise type, we compute the percentage gain in both Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) obtained using the $\mathbb{M}$-operator relative to a usual nonparametric Nadaraya Watson estimator. The results are averaged over 1000 simulation replications.

Table 3.1 reveals several consistent patterns regarding the effectiveness of the proposed $\mathbb{M}$-operator across varying sample sizes, data generating processes (DGPs), and noise distributions. The performance gains—measured in terms of improvements in Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE)—are most notable when the sample size is small. For a sample size of 300, the $\mathbb{M}$-operator yields substantial improvements, particularly under the ARMA and SETAR processes, and

Table 3.1: Gain in accuracy upon enforcing $\mathbb{M}$-operator

Sample size	**DGP**	**Noise**	**% Gain in MAE**	**% Gain in RMSE**
300	White Noise	norm	2.5686	3.1174
300	ARMA	norm	10.2954	11.4295
300	SETAR	norm	8.3861	9.1793
300	White Noise	t	6.5038	7.0081
300	ARMA	t	11.4550	12.1464
300	SETAR	t	5.9463	7.0827
300	White Noise	power-law	−0.0003	0.0959
300	ARMA	power-law	0.7757	1.6569
300	SETAR	power-law	0.1924	0.4223
500	White Noise	norm	1.9400	2.7310
500	ARMA	norm	0.5107	0.6138
500	SETAR	norm	0.5596	0.7257
500	White Noise	t	2.2581	2.9949
500	ARMA	t	2.9566	3.4881
500	SETAR	t	5.2912	6.1396
500	White Noise	power-law	0.0000	0.0152
500	ARMA	power-law	0.0005	0.0240
500	SETAR	power-law	0.0112	0.0447
1000	White Noise	norm	0.3149	0.6505
1000	ARMA	norm	0.6956	0.9443
1000	SETAR	norm	2.5490	3.2455
1000	White Noise	t	6.2784	6.9865
1000	ARMA	t	0.0828	0.1935
1000	SETAR	t	6.0066	6.7279
1000	White Noise	power-law	0.0000	0.0055
1000	ARMA	power-law	0.0470	0.1024
1000	SETAR	power-law	0.0006	0.0091

Table 3.2: Gain in accuracy upon enforcing $\mathbb{Q}$-operator

Sample size	**DGP**	**Noise**	**% Gain in MAE**	**% Gain in RMSE**
300	White Noise	norm	30.4182	23.0704
300	ARMA	norm	73.0874	60.1088
300	SETAR	norm	42.7238	35.3949
300	White Noise	t	56.5914	44.7615
300	ARMA	t	85.9528	67.5336
300	SETAR	t	31.9044	25.7731
300	White Noise	power-law	-2.4885	−3.0008
300	ARMA	power-law	-34.4652	−32.2777
300	SETAR	power-law	-9.7912	−10.0112
500	White Noise	norm	47.1233	33.4629
500	ARMA	norm	-5.5351	−3.1960
500	SETAR	norm	-19.4027	−12.4332
500	White Noise	t	59.6190	42.2530
500	ARMA	t	73.7085	49.6222
500	SETAR	t	47.6229	36.2088
500	White Noise	power-law	−0.9615	−1.1505
500	ARMA	power-law	−1.4650	−1.5278
500	SETAR	power-law	−3.2876	−3.4731
1000	White Noise	norm	45.0405	28.8704
1000	ARMA	norm	55.1561	34.7556
1000	SETAR	norm	105.4194	76.7823
1000	White Noise	t	247.0137	172.2439
1000	ARMA	t	3.4014	3.4634
1000	SETAR	t	206.9370	144.7681
1000	White Noise	power-law	−0.8870	−1.1591
1000	ARMA	power-law	−14.3507	−15.1223
1000	SETAR	power-law	−2.0085	−2.2033

Table 3.3: Gain in accuracy upon enforcing $\mathbb{QM}$-operator

Sample size	**DGP**	**Noise**	**% Gain in MAE**	**% Gain in RMSE**
300	White Noise	norm	0.3044	0.2321
300	ARMA	norm	0.7415	0.6158
300	SETAR	norm	0.4315	0.3599
300	White Noise	t	0.5721	0.4542
300	ARMA	t	0.8711	0.6900
300	SETAR	t	0.3200	0.2606
300	White Noise	power-law	−0.0249	−0.0300
300	ARMA	power-law	−0.3414	−0.3196
300	SETAR	power-law	−0.0979	−0.1000
500	White Noise	norm	0.4717	0.3363
500	ARMA	norm	−0.0554	−0.0320
500	SETAR	norm	−0.1940	−0.1243
500	White Noise	t	0.5966	0.4253
500	ARMA	t	0.7373	0.4974
500	SETAR	t	0.4773	0.3645
500	White Noise	power-law	−0.0096	−0.0115
500	ARMA	power-law	−0.0147	−0.0153
500	SETAR	power-law	−0.0329	−0.0347
1000	White Noise	norm	0.4504	0.2891
1000	ARMA	norm	0.5516	0.3478
1000	SETAR	norm	1.0546	0.7691
1000	White Noise	t	2.4732	1.7278
1000	ARMA	t	0.0340	0.0346
1000	SETAR	t	2.0714	1.4509
1000	White Noise	power-law	−0.0089	−0.0116
1000	ARMA	power-law	−0.1434	−0.1511
1000	SETAR	power-law	−0.0201	−0.0220

these gains are most pronounced when the noise follows a normal or t-distribution. In contrast, under heavy-tailed power-law noise, the performance improvements are negligible or marginal, especially for white noise DGPs. As the sample size increases to 500 and then 1000, the gains in MAE and RMSE become progressively smaller across all settings. This indicates that the advantage offered by the $\mathbb{M}$-operator diminishes with more data, likely due to the consistency of baseline estimators in larger samples. However, even in larger samples, modest improvements persist under t-distributed noise and nonlinear DGPs such as SETAR, suggesting that the operator continues to offer robustness benefits in challenging scenarios. Overall, the $\mathbb{M}$-operator demonstrates the greatest utility in small-sample and moderately non-Gaussian environments, particularly when the underlying process exhibits temporal dependence or regime switching. Its relative advantage is less evident in settings dominated by heavy-tailed power-law noise or when the data volume is sufficiently large to offset the complexity of the DGP.

Similarly, the performance of the $\mathbb{Q}$-operator, as presented in Table 3.2, demonstrates a sharp contrast depending on the noise distribution and the nature of the data generating process (DGP). Under both normal and t-distributed noise, the operator leads to substantial improvements in estimation accuracy, especially for smaller sample sizes. These improvements are particularly prominent for ARMA and SETAR models, with gains being stronger in settings with heavier-tailed t noise, where the operator appears to enhance robustness considerably. Even for white noise DGPs, the $\mathbb{Q}$-operator consistently offers strong gains under these two noise distributions. However, the advantages of the $\mathbb{Q}$-operator deteriorate notably when the data is contaminated with power-law noise. In such cases, the operator not only fails to offer improvement but

actually leads to performance deterioration across all DGPs and sample sizes. This negative impact is especially stark for ARMA processes and remains consistent even as the sample size increases. Interestingly, unlike the $\mathbb{M}$-operator, the performance of the $\mathbb{Q}$-operator does not exhibit a clear diminishing trend with increasing sample size in the favorable noise settings; instead, it continues to yield strong and even increasing gains in many of those scenarios, particularly in nonlinear DGPs like SETAR. The $\mathbb{Q}$-operator is highly effective in structured or autocorrelated environments under light to moderately heavy-tailed noise, but its susceptibility to performance degradation in the presence of heavy-tailed power-law noise highlights a limitation in its robustness to extreme distributional deviations.

Finally, we present the relative performance of the $\mathbb{QM}$-operator in Table 3.3. The $\mathbb{QM}$-operator exhibits a generally stable but modest performance profile. Across all sample sizes and DGPs, it yields consistently small but positive gains in estimation accuracy under normal and t-distributed noise. The gains are slightly higher under t noise, especially for nonlinear DGPs like SETAR, suggesting that the operator retains some degree of robustness to moderate deviations from normality. However, the improvements are relatively marginal compared to those observed for the individual $\mathbb{M}$ and $\mathbb{Q}$ operators. A key strength of the $\mathbb{QM}$-operator lies in its stability. Even under power-law noise, where both $\mathbb{M}$ and $\mathbb{Q}$ operators faced more substantial deterioration in performance, the degradation observed with the $\mathbb{QM}$-operator is minimal. The percentage losses under this heavy-tailed regime are consistently close to zero, indicating that the operator, while not particularly effective in boosting accuracy, avoids making the performance worse in challenging scenarios. Unlike the $\mathbb{Q}$-operator, the $\mathbb{QM}$-operator does not display significant sensitivity to the choice of sample size, with

its performance largely holding steady as the data volume increases. This consistency makes it a conservative and reliable choice in scenarios where the underlying data characteristics are uncertain or where robustness across varied noise types is preferred over aggressive improvement. Overall, the $\mathbb{QM}$-operator provides a balanced trade-off between accuracy and robustness, though its limited gain magnitude may reduce its appeal in scenarios prioritizing maximal performance improvement.

3.5 Conclusion

This work introduces a flexible and broadly applicable framework for estimating conditional mean and variance functions in time series models while incorporating shape-based prior information. Traditional nonparametric estimators, such as kernel smoothing or local linear regression, offer flexibility and robustness but do not inherently satisfy structural shape constraints that are often motivated by domain knowledge, particularly in financial applications. These constraints might include monotonicity, quasiconvexity, or combinations thereof, which can reflect important economic theory or empirical regularities. Rather than embedding such constraints directly into the estimation procedure—which can reduce statistical efficiency or introduce rigidity—our approach proposes a two-step method. We first compute an unconstrained estimator and then apply a shape-enforcing operator that projects this estimate onto the space of functions satisfying the desired structural assumptions. This strategy allows us to preserve the richness of the initial estimator while ensuring that the final result aligns with known or expected functional behavior. Importantly, our methodology is compatible with both local and global estimation techniques. This includes estimators that work pointwise as well as those that approximate the entire

function using parametric or semi-parametric bases. Through empirical studies, we demonstrate that the shape-enforcing operators can significantly improve the compliance of the estimates with prior shape knowledge without compromising predictive accuracy or interpretability. By supporting a wider class of shape constraints than typically addressed in the literature, our framework offers a powerful and generalizable tool for nonparametric modeling in time series and financial econometrics.

Building on the contributions made in this work, several exciting directions for future research can be explored to further enhance the understanding and application of nonparametric models under shape restrictions. One key direction is the extension of the setup to incorporate more complex shape restrictions, particularly those that involve higher-order derivatives or partial derivatives of the regression functions. This would allow for a deeper understanding of the curvature properties of the functions, which are crucial in many economic applications. For instance, the ability to model higher-order sensitivities like gamma and vega would provide more accurate pricing of derivatives and better risk management in financial markets. This extension would improve economic models by capturing marginal effects in a more precise manner, offering a more nuanced understanding of economic and financial systems.

Another important area for future research lies in the detection of structural breaks within models that enforce shape restrictions. Existing methods for nonparametric changepoint detection often overlook the shape-related information inherent in regression functions, such as monotonicity and convexity. Bridging this gap by incorporating shape constraints directly into changepoint detection frameworks could significantly enhance the precision and robustness of structural break detection. This

would be particularly valuable in financial data, where structural shifts in market behavior or regime changes are often subtle yet impactful. My future work will focus on developing methods that integrate shape-enforcing operators into these detection frameworks, potentially offering improved accuracy in identifying structural breaks and better understanding shifts in underlying data patterns.

Lastly, it is also an interesting problem to devise methodologies for the shape-enforced estimation of nonparametric structural functions, an area that, to the best of my knowledge, has not been fully explored in the literature. This involves developing robust estimation techniques that respect theoretical shape constraints, such as monotonicity or concavity, while estimating key structural functions in economic models. These functions, which could represent concepts like demand, supply, or cost curves, often have clear theoretical shape restrictions that are not typically incorporated into standard estimation methods. By enforcing these shape restrictions, the resulting estimates would be more consistent with economic theory, offering improved accuracy and interpretability. This line of research would not only fill an important gap in the literature but also provide valuable tools for applied economists and financial analysts seeking to model real-world data more effectively.

3.6 Appendix: Prerequisite lemmas

Lemma 3.1. *1.* $f_*(t) = f^*\left(\mathcal{M}(E) - t\right)$ *holds at every point of continuity i.e. almost everywhere on* $(0, \mathcal{M}(E))$.

2. For any function ψ that is monotonic on $[0, \infty)$,

$$\int_0^{\mathcal{M}(E)} \psi(f^*(x))dx = \int_0^{\mathcal{M}(E)} \psi(f_*(x))dx = \int_E \psi(|f(x)|)dx.$$

3. $\sup_{\{F \subset E | \mathcal{F}=t\}} \{\int_F |f(x)|dx\} = \int_0^t f^*(u)du \forall t \in (0, \mathcal{M}(E))$.

4. $\inf_{\{F \subset E | \mathcal{F}=t\}} \{\int_F |f(x)|dx\} = \int_0^t f_*(u)du \forall t \in (0, \mathcal{M}(E))$.

Proof. Interested readers may refer to [59] for detailed proof. □

Lemma 3.2. *Let $\{f_k\}$ be a sequence of positive bounded measurable functions on $(0, 1)$ with corresponding equimeasurable decreasing rearrangement $\{f_k^*\}$. The necessary and sufficient conditions for a continuous function $\phi(x; u_1, u_2, ..., u_n)$ defined for $0 < x < 1$ and $u_k \geq 0 \forall k = 1(1)n$ to satisfy*

$$\int_0^1 \phi(x; f_1(x), f_2(x), ..., f_n(x))dx \leq \int_0^1 \phi(x; f_1^*(x), f_2^*(x), ..., f_n^*(x))dx \tag{3.43}$$

for any subset of functions $f_1, f_2, ..., f_n$ from $\{f_k\}$ are-

$$\phi(u_i + h, u_j + h) - \phi(u_i + h, u_j) - \phi(u_i, u_j + h) + \phi(u_i, u_j) \geq 0 \tag{3.44}$$

$$\int_0^\delta [\phi(x - t; u_i + h) + \phi(x + t; u_i) - \phi(x + t; u_i + h) - \phi(x - t; u_i)]\, dt \geq 0 \tag{3.45}$$

for all $h > 0$, $0 < \delta < x$, $\delta < 1 - x$ and $i \neq j$. In particular, if ϕ has continuous second partial derivatives with respect to all variables then Equation (3.44) and

Equation (3.45) *respectively become equivalent to*

$$\frac{\delta^2\phi}{\delta u_i \delta u_j} \geq 0, \quad \frac{\delta^2\phi}{\delta x \delta u_i} \leq 0.$$

Proof. The proof of the above can be found in [64]. Note that the direction of the inequality in Equation (3.43) alters when we work with increasing rearrangements. □

[65] discuss that when a function $L : \mathbb{R}^2 \to \mathbb{R}_+$ (where $\mathbb{R}_+$ denotes the set of positive real numbers) is smooth then submodularity is equivalent to $\frac{\delta^2 L}{\delta v \delta t} \leq 0 \forall (v,t) \in \mathbb{R}^2$. It can also be shown that Equation (3.44) and Equation (3.45) hold for un-smooth submodular functions. Therefore for any submodular function $L : \mathbb{R}^2 \to \mathbb{R}_+$ we must have

$$\int_\chi L(q^*(x), g^*(x))dx \leq \int_\chi L(q(x), g(x))dx, \tag{3.46}$$

where q and g are two functions on χ and q^* and g^* are their corresponding increasing rearrangements.

Lemma 3.3. *For any sequence of cumulative distribution functions* $\{F_n\}_n$, $F_n^{-1} \hookrightarrow F^{-1}$ *if and only if* $F_n \hookrightarrow F$, *where '*$\hookrightarrow$*' denotes weak convergence.*

Proof. See [66] (Chapter 21) for complete proof. □

Lemma 3.4. *Given a measurable function* f *on* $E \subset \mathbb{R}$ *there exists a sequence of step functions* $\{f_k\}$ *which converges to* f *pointwise almost everywhere.*

Proof. See [67] for complete proof. □

Chapter 4

Test of lack-of-fit in ODE models

4.1 Background

The problem of testing whether data originate from a parametric regression function, where the alternative hypothesis is a nonparametric model, has been extensively studied in the statistical literature [see 68, 69, for comprehensive reviews]. However, the specific case in which the regression function is defined as the solution of a system of ordinary differential equations (ODEs) has received comparatively less attention. To the best of our knowledge, this problem has been investigated in only a few studies, including [70], [71] and more recently, [72] and [73]. In this chapter, we develop an adaptive nonparametric testing procedure to assess whether the underlying regression function satisfies a fully or partially specified system of ODEs, considering a nonparametric alternative. Many early contributions in the domain of lack-of-fit included evaluating the validity of a hypothesized model by examining the discrepancy between the least squares estimate (LSE) and the nonparametric kernel estimate [74, 75], similar ideas were extended by many future chapters like [76],[77] and [78]. Other parallel approaches to lack-of-fit testing broadly emerged from empirical process theory [79], likelihood-ratio based tests [80] or properties of the residuals [81]. In contrast, the field of lack-of-fit testing for ordinary differential equation (ODE) models remains largely unexplored despite its wide applicability in a wide range of scientific fields [82]. Existing methodologies often fall short here because, unlike static models or regression-based frameworks, ODE systems represent dynamic processes driven

by latent temporal dynamics, making their underlying data-generating mechanisms more complex to evaluate. This challenge has also been acknowledged in the recent review by [83], which underscores the limited attention given to statistical methods for diagnosing potential model misspecifications in dynamic systems. In this context, [70] and [71] considered the scenario where the evolution of the system is driven by a certain ODE with unknown parameter and some external influence modeled as a nonparametric forcing function. The lack-of-fit test was then done through testing whether the forcing function is identically zero. However, the authors acknowledged that their proposed test might be conservative for chaotic alternatives, which could be mitigated by observing the system over a longer time span with more frequent measurements. Notably, these studies lacked a rigorous theoretical investigation of the size and power of the test, particularly against both global and local alternatives. Some related works, such as [84] and [85], focused on regularization methods for parameter estimation in ODEs. These studies also introduced forcing function perturbations at the derivative level to account for model misspecification. While their primary goal was parameter estimation, the authors suggested that such an approach could be adapted for hypothesis testing, emphasizing that a statistical test at the derivative level may be more relevant for detecting model misspecifications than one based on residuals. A more recent nonparametric approach to lack-of-fit testing in dynamic systems was introduced by [72] and [73]. The authors proposed three different test statistics, each capturing the discrepancy between the observed data and different representations of the hypothesized system: one based on trajectory representation, another on integration representation, and the third on gradient representation. They highlighted that the trajectory-based statistic fails to test individual components of the system due to mixed component and parameter effect, while the latter two can-

not handle partially observed systems. The test's asymptotic power was theoretically established for local alternatives, and its performance was shown to be comparable to that of [70] in terms of power across a wide range of ODE systems. Finally, it is worth noting the recent contribution by [86], who developed an efficient algorithm for selecting the best model from a set of competing ODE models. Our work addresses the same problem of lack-of-fit testing in ODE systems, but takes a fundamentally different approach from the above-cited chapters. We propose a single test statistic that is applicable to both fully and partially observed ODE systems, accommodating model misspecifications in either the entire system or in specific components, as long as the concerned system is statistically identifiable. To that end, we introduce an omnibus test of lack-of-fit for a broad class of ODE models. Our contribution in this chapter is threefold. First, to the best of our knowledge, this study is the first to develop theory and methods for testing lack-of-fit for nonlinear ODEs with a minimax nonparametric framework. Second, our framework is capable of handling ODEs that describe a single evolving state, as well as models with multiple states, where one or more states may remain unobserved due to practical limitations or experimental costs. This makes our framework a versatile tool for assessing the adequacy of ODE models in evolving systems. Finally, the proposed test is designed to be adaptive and rate-optimal, meaning it adjusts to the unknown smoothness of the alternative model and is uniformly consistent against alternatives whose distance from the parametric model converges at the fastest possible rate. By doing this we overcome the limitation of the traditional nonparametric lack-of-fit tests, the performance of which are heavily dependent on the regularity of the alternatives [87]. We establish the theoretical consistency of our test through a minimax approach [88, 89] for a wide class of alternatives. Our test achieves power performances comparable to [72]. The

performance of the test is further validated through an extensive simulation study and applied to a range of real datasets from fields like biology, agriculture, and drug research.

4.2 Mathematical framework

4.2.1 Assumptions

In this section, we outline the assumptions necessary for ensuring the consistency of the test. These assumptions primarily concern the properties of the parametric family $\mathcal{F}_\Theta$, the behavior of the parametric estimators, the nature of noise heteroscedasticity, the spacing of design points, and other relevant factors. Most of these assumptions are standard in the literature on model misspecification testing (see [89], [90] among others).

Assumption 1. *Let $\{\theta_n^0\}$ be a sequence in Θ whose limit, if exists, is in Θ and $\widehat{\sigma}_i^2$'s be the heteroscedasticity estimates. Let $\widehat{\theta_n^*}$ be the estimate of θ_n^0 obtained from the data generated from the model $y^*(t_i) = z(t_i, \theta_n^0, \zeta) + \epsilon_i^*$ where $\epsilon_i^* \sim \mathcal{N}(0, \widehat{\sigma}_i^2)$ independently. Then*

$$\lim_{n\to\infty} P\left(\sqrt{n}\left|\widehat{\theta_n^*} - \theta_n^0\right| > \upsilon\right) < \delta,$$

for some $\delta > 0$ and sufficiently large $\upsilon > 0$.

This assumption establishes a stability property of the parametric estimator that is used to justify the simulation procedure for obtaining the critical values of the test statistic.

Assumption 2. *$h_{max} > h_{min} \geqslant n^{-\gamma}$ for some $\gamma \in \left(0, \frac{1}{3}\right)$ and $h_{max} = C_H((\log\log n)^{-1})$*

for some $C_H > 0$. Note that imposing this assumption on the structure of $\mathcal{H}_n$ gives $J_n \leqslant O(\log n)$ as $n \to \infty$.

Assumption 3. *There exists constants C_1 and C_2 such that $C_1 nh \leqslant M_h(t_i) \leqslant C_2 nh \forall i = 1(1)n$ where $M_h(t)$ denotes the number of observed time points falling in the interval $[t-h, t+h]$.*

It can be shown that this assumption holds with a probability approaching 1 as $n \to \infty$ if $\mathcal{H}_n$ meets the conditions specified in assumption 4, and the sampling time points t_i's are drawn from a distribution that is absolutely continuous with respect to the Lebesgue measure, has bounded support, and possesses a density that is bounded away from 0. However, it is not strictly necessary for the t_i's to be generated from such a distribution.

Assumption 4. *The measurement errors ϵ_i's have up to $4+p$ moments finite for some $p > 0$.*

Assumption 5. *The heteroscedasticity and fourth-moment function of the random noise are Lipschitz continuous. This means if we consider $\sigma_i^2 \coloneqq \sigma^2(t_i)$ and $S_i \coloneqq E[\epsilon_i^4] \coloneqq S(t_i)$ then*

$$\left|\sigma_i^2 - \sigma_j^2\right| \leqslant \nu \left|t_i - t_j\right| \text{ and } \left|S_i - S_j\right| \leqslant \upsilon \left|t_i - t_j\right| \forall i, j,$$

for some finite constants ν and υ. This assumption ensures the consistency of the heteroscedasticty estimates.

Assumption 6. *K is non negative, symmetric about origin with support $[-1, 1]$. In particular we assume $K(u) \geqslant \kappa$ when $|u| \leqslant \delta_K$ for some known κ,$\delta_K > 0$ and*

$K(u) \leqslant 1 \forall u$.

4.2.2 Problem setup

Consider a r-dimensional system observed with noise over discrete time points $t \in \mathcal{T} \subseteq [0,1], \mathcal{T} = \{t_1, t_2, \ldots, t_n\}, t_1 = 0, t_n = 1$ as

$$y(t_i) = x(t_i) + \epsilon_i, j = 1, 2, \ldots, r; i = 1, 2, \ldots, n, \tag{4.1}$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ for $\sigma_i > 0, i = 1, 2, \ldots, n$ independently. Note that we assume only $d(\leqslant r)$ states of the system are observed. We are interested in testing whether the noisy observations in (4.1) are generated by the system

$$\begin{cases} x'(t) = F(x(t), \theta), \\ x(0) = \zeta, \end{cases} \tag{4.2}$$

where the functional form of $F(., \theta) \in \mathcal{F}_\Theta$ is known for a given $\theta \in \Theta \subseteq \mathbb{R}^q, q \geqslant 1$. Typically, the initial condition ζ is known in applications; however, if unavailable, they can estimated alongside the system parameter θ using the approach outlined in the following sections. The testing problem can then be formally stated as

$$H_0 : x(t) = z(t, \theta, \zeta) \text{ for some } \theta \in \Theta,$$

where $z(., \theta, \zeta)$ is the solution of (4.2) either numerically or analytically evaluated. We assume that, in cases where the system is not exactly solvable, the solution can be numerically approximated with arbitrarily high accuracy. Therefore, for simplicity, we

do not account for the numerical error introduced during this evaluation, in addition to the random noise already present in the model. Nevertheless, we believe extending the analysis to incorporate such numerical errors is straightforward and leave this aspect for future work. Note that the hypothesis testing problem described above is a classic model specification testing problem. However, the main difficulty in this case is two-fold. One, the misspecification is tested at the derivative level, which we do not observe and second, even at the derivative level the system is not completely observed.

4.2.3 Proposed test

To construct the test statistic for (4.2.2) we need to estimate the unknown parameter θ in the ODE model (4.2) under the null hypothesis. The estimation of unknown parameters in dynamical systems is well-studied, with popular methods including nonlinear least squares [91] and two-step collocation [90, 92, 93]. Theoretical results on the asymptotic properties of these estimates are also established; see [94] and [83] for detailed reviews.

We assume there exists an estimator $\widehat{\theta}$ of θ that is $\sqrt{n}-$consistent for θ_0, where θ_0 is the true value of θ under H_0. We also assume that $\widehat{\theta}$ is a stable estimator. Under reasonably mild restrictions (see [95]) one can show that most of the usual estimators, such as least-squares of M-estimators, satisfy these properties. The construction of the test statistic also requires the estimation of the heteroscedasticity σ_i^2's. Since we require an estimator that is consistent regardless of whether H_0 is true we cannot use an estimator based on the estimated model residuals from (4.1). We adopt the heteroscedasticity estimator $\widehat{\sigma}_i^2$ motivated from the works of [96]. The asymptotic

properties of this estimate is rigorously established in [89]. In what follows the notations are defined assuming only one of the states in the system (4.1) is observed, i.e. $d = 1$. The case for multiple observed states $d > 1$ can be achieved by a simple Bonferroni correction. We leave the development of a stronger test to a future work. For notational simplicity in the $d = 1$ case, we write (4.1) as $y_i = x(t_i) + \epsilon_i$ where $y_i \coloneqq y(t_i)$.

We now describe the kernel smoothing procedure that is used in our test. Given an appropriately chosen kernel function $K(.)$ and bandwidth $h = h(n)$, for each $i, j = 1, 2, \ldots, n$ define the weight matrix as

$$W_h(i,j) = \int_0^1 w(t,t_i)w(t,t_j)dt, \quad w(t,t_i) = \frac{K\left(\frac{t-t_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{t-t_i}{h}\right)}.$$

The kernel smoother of nonparametric estimator of the observed process and the parametric estimator are respectively given as

$$\widehat{x}(t) = \sum_{i=1}^n y(t_i)w(t,t_i), \quad \widehat{z}(t,\theta,\zeta) = \sum_{i=1}^n z(t_i,\theta,\zeta)w(t,t_i).$$

Denote $\widehat{x} = (\widehat{x}(t_1), \widehat{x}(t_2), \ldots, \widehat{x}(t_n))^\intercal$ and $\widehat{z}(\theta,\zeta) = (\widehat{z}(t_1,\theta,\zeta), \widehat{z}(t_2,\theta,\zeta), \ldots, \widehat{z}(t_n,\theta,\zeta))^\intercal$. Our proposed test statistic is based on the integrated $\mathcal{L}^2$-distance between the nonparametric and smoothed parametric estimator, defined as,

$$S_h(\theta) = \widehat{e}^\intercal W_h \widehat{e}, \quad \widehat{e} = \widehat{x} - \widehat{z}(\theta,\zeta), \quad \theta \in \Theta.$$

Define $N_h = E\left[S_h(\theta_0)\right]$ and $V_h^2 = \mathrm{Var}\left[S_h(\theta_0)\right]$. Consider the standardized version of

$S_h(\theta)$ as,

$$T_h(\theta) = \frac{S_h(\theta) - N_h}{V_h}.$$

Since the values of $T_h(\theta)$ are extremely sensitive to the selected bandwidth h, we evaluate them over a sequence of possible bandwidths defined as

$$\mathcal{H}_n = \left\{ h_i \middle| h_{\min} \leqslant h_i \leqslant h_{\max}, h_i = h_{\max} a^i, i = 0, 1 \ldots (J_n - 1) \right\}$$

where $J_n = \log_{\frac{1}{a}} \left(\frac{h_{\max}}{h_{\min}} \right)$, for some $0 \leqslant h_{\min} < h_{\max} \leqslant 1$ and $a \in (0, 1)$. The final test statistic is then defined as,

$$T^* = \max_{h \in \mathcal{H}_n} \left\{ T_h(\widehat{\theta}) \right\} = \frac{S_h(\widehat{\theta}) - N_h}{V_h}.$$

This approach, first proposed by [89], has been widely adopted in the literature on lack-of-fit testing to ensure test adaptivity [97, 98]. Note that in the above equations, N_h and V_h^2 can be replaced by their respective consistent estimates. Since the true value of θ under H_0, denoted by θ_0, is not known, hence evaluating the critical values of T^* analytically is not possible. Alternatively, we suggest a bootstrap algorithm to evaluate the critical values. Let t_α be the $\alpha-$level critical value which is obtained as the $(1-\alpha)-$th quantile of the bootstrap distribution of the test statistic. The proposed test rejects H_0 with asymptotic level α is $T^* > t_\alpha$. As highlighted by [75], the choice of bootstrap procedure is critical in this context. The naive and adjusted residual bootstrap methods are noted to often produce overly conservative tests, which can fail even under homoscedastic error structures. To address these issues, the authors advocate for the use of the wild bootstrap as a more robust alternative. However, our simulation studies indicate that the naive bootstrap performs adequately and is

sufficient in practice.

4.2.4 Properties of the test

We assume that the function $F(.,\theta)$ in (4.2) is smooth and the system (4.2) is structurally identifiable i.e. the solution z uniquely determines the initial value ζ and the parameter vector θ, and subsequently the entire system (4.2). Verifying the identifiability of a system in the presence of statistical noise presents significant challenges. [99] employed differential algebra to assess global identifiability of dynamic models, whereas [100] developed an eigenvalue-based test for checking the identifiability of ODE systems. More recently, [93] provided necessary and sufficient conditions for the structural identifiability of a specific class of separable ODEs. We also require some assumptions pertaining to the characteristics of the parametric family $\mathcal{F}_\Theta$, the behavior of the parametric estimators, the nature of noise heteroscedasticity, the spacing of design points, and other related factors. Most of these assumptions are standard in the literature on model misspecification testing (see [89], [90], among others). All the proofs in this section will follow from few prerequisite lemmas, which have been detailed in the Appendix.

We now establish the asymptotic size and power of the test. In what follows, t_α denotes the $(1-\alpha)-$level critical value of the bootstrapped distribution of the test statistic T^*.

Theorem 4.1. *Under the previosuly stated, if H_0 is true then* $\lim_{n\to\infty} P(T^* > t_\alpha) = \alpha$.

Proof. The theorem is a straightforward consequence of Lemma 4.4, Lemma 4.5 and

Lemma 4.6, which are detailed in the Appendix. We skip the details since there are no technical difficulties involved. □

This theorem ensures the asymptotic size of the test. The proof follows directly from a probabilistic equivalence established between the behavior of the statistic $S_h(\theta)$ constructed with bootstrapped data and its behavior under both the null and alternative hypotheses.

For any function f, we empirically measure its distance from the parametric class $\mathcal{F}_\Theta$ as,

$$\rho_n^2(f, \mathcal{F}_\Theta) = \inf_{\theta \in \Theta} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(f(t_i) - \zeta - z(t_i, \theta) \right)^2 \right\}. \tag{4.3}$$

An asymptotically consistent test should reject H_0 if the empirical distance between the true (unknown) dynamics x and the parametric class $\mathcal{F}_\Theta$ is large enough for sufficient sample size, say if $\rho_n^2(x, \mathcal{F}_\Theta) \geqslant c_\rho$ for some $c_\rho > 0$.

Theorem 4.2. *Under the previously stated assumptions, if there exists a* $n_0 = n_0(c_\rho)$ *for which* $\rho_n^2(x, \mathcal{F}_\Theta) \geqslant c_\rho$ *for all* $n \geqslant n_0$ *then* $\lim_{n\to\infty} P(T^* > t_\alpha) = 1$.

Proof. Define,

$$\vartheta = \underset{\theta \in \Theta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \left(x(t_i) - \zeta - z(t_i, \theta) \right)^2 \right\}.$$

Then according to Lemma 4.9, it would be enough to show that if there exists a $n_0 = n_0(c_\rho)$ for which $\rho^2(x, \mathcal{F}_\Theta) \geq c_\rho$ for all $n \geq n_0$ then

$$\int_0^1 \sum_{i=1}^{n} \left[\left(x(t_i) - \zeta - z(t_i, \vartheta) \right) w(t, t_i) \right]^2 dt \geq \frac{4 V_h \widetilde{t}_\alpha^{(0)}}{n\sqrt{h}}. \tag{4.4}$$

Since $\int_0^1 \sum_{i=1}^n \left[\left(x(t_i) - \zeta - z(t_i, \vartheta)\right) w(t, t_i)\right]^2 dt$ can be seen as the smoothed distance between the true data generating process and the null hypothesis therefore assuming the functions x and z to be smooth and the bandwidth h to be small enough we can write

$$\int_0^1 \sum_{i=1}^n \left[\left(x(t_i) - \zeta - z(t_i, \vartheta)\right) w(t, t_i)\right]^2 dt \approx \sum_{i=1}^n \left(x(t_i) - \zeta - z(t_i, \vartheta)\right)^2$$

Therefore it would suffice to show that $4V_h \widetilde{t}_\alpha^{(0)} \leqslant n^2 \sqrt{h} c_\rho$, which follows from assumption 4 and Lemma 4.8 in the Appendix. □

Next we focus on the power of the test against a sequence of local alternatives. Formally, the hypothesis testing problem can be formulated as

$$H_0 : F \in \mathcal{F}_\Theta$$

vs

$$H_1^{(n)} : F \in \mathcal{F}_\Theta^n \text{ where } \mathcal{F}_\Theta^n = \left\{x_n | x_n(t) = z(t, \theta) + \rho_n g(t), x \in \mathcal{X}_\Theta, \theta \in \Theta\right\}, \tag{4.5}$$

where $\mathcal{X}_\Theta$ denotes the parametric class of the dynamics of x under null hypothesis, for some continuous function g and $\{\rho_n\} \xrightarrow{n\to\infty} 0$ where $\rho_n \geqslant k\sqrt{\frac{\log\log n}{n}}$ for some $k > 0$. We assume that $g(\cdot)$ is a continuous function that is normalized such that $\sum_{i=1}^n |g(t_i)|^2 \geqslant n$. We further assume

$$\sum_{i=1}^n \left[g(t_i)\left(1 - \frac{z_{\theta'}^2(t_i, \theta')}{B_{\theta'}}\right) + \frac{z_{\theta'}(t_i, \theta')}{B_{\theta'}} \sum_{j=1, j\neq i}^n g(t_j) z_{\theta'}(t_j, \theta')\right]^2 \geqslant \delta_g \sum_{i=1}^n |g(t_i)|^2,$$

for some $\delta_g > 0, \theta' \in \Theta$ where $B_{\theta'} = \sum_{i=1}^n \left|\frac{\delta}{\delta\theta'} z(t_i, \theta')\right|$ and $z_{\theta'}(t, \theta') = \frac{\delta}{\delta\theta'} z(t, \theta')$. Note

that these two assumptions only allow us to take functions $g(\cdot)$'s for which

$$\sum_{i=1}^{n}\left(x_n(t_i)-\zeta-z(t_i,\theta_n^0)\right)^2 > cn\rho_n \text{ for some } c>0.,$$

where $\{\theta_n^0\}$ is a sequence of points in θ, whose limit point, if exists, is in θ. This further ensures that the rate of convergence of the sequence of alternatives to the null hypothesis depends solely on how fast the sequence $\{\rho_n\}$ converges to 0. This, along with the previously stated assumptions imply that

$$\inf_{\theta\in\Theta}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(x_n(t_i)-\zeta-z(t_i,\theta)\right)^2\right\} \geqslant \delta_g\rho_n(1-o(1)).$$

Theorem 4.3. *Along with the previously stated assumptions, if g satisfies the above described restrictions, then $\lim_{n\to\infty} P(T^* > t_\alpha) = 1$ for each of the alternatives of the form (4.5).*

Proof. Recall the sequence of alternatives defined in (4.5). According to Lemma 4.9 we have to show that if $H_1^{(n)}$ is true then for large enough value of n and for all $h \in \mathcal{H}_n$,

$$\int_0^1\left[\sum_{i=1}^{n}\left\{\left(x(t_i)-\zeta-z(t_i,\theta_n^0)\right)w(t,t_i)\right\}^2\right]dt \geq \frac{4V_h\widetilde{t}_\alpha^{(0)}}{n\sqrt{h}}, \tag{4.6}$$

where $\{\theta_n^0\}, \theta_n^0 = \text{argmin}_{\theta\in\Theta}\left\{\sum_{i=1}^{n}\left(x_n(t_i)-\zeta-z(t_i,\theta)\right)^2\right\}$ is a sequence of points whose limit, if exists, is in θ. Fix t and define:

$$a^2 = \sum_{i=1}^{n}\left\{\left(x(t_i)-\zeta-z(t_i,\theta_n^0,\xi)\right)w(t,t_i)\right\}^2,$$

$$b^2 = \rho_n^2\sum_{i=1}^{n}k^2(t_i)w^2(t,t_i),$$

where

$$k(t_i) = B_{\theta'}^{-1}\left[g(t_i)\left(1 - z_{\theta'}(t_i, \theta', \xi)\right) - z_{\theta'}(t_i, \theta', \xi)\sum_{j=1, j\neq i}^{n} g(t_j) z_{\theta'}(t_j, \theta', \xi)\right],$$

and all symbols are as previously defined in (4.2.4). Then using reverse Minkowski inequality,

$$(b-a)^2 \leq \sum_{i=1}^{n}\left[\left\{x_n(t_i) - \zeta - z(t_i, \theta_n^0) - \rho_n k(t_i))\right\} w(t, t_i)\right]^2.$$

Therefore under $H_1^{(n)}$,

$$\int_0^1 \left[\sum_{i=1}^{n}\left\{\left(x_n(t_i) - \zeta - z(t_i, \theta_n^0)\right) w(t, t_i)\right\}^2\right] dt \geqslant \frac{\delta_g}{2}.\frac{n\log\log n}{n}.$$

Establishing (4.14) is therefore equivalent to demonstrating that

$$\left(\frac{\delta_g k}{2}\right)\cdot \log\log n \geqslant \frac{4V_h \widetilde{t}_\alpha^{(0)}}{n\sqrt{h}},$$

which follows directly from Lemma 4.2 and Lemma 4.8. □

This theorem establishes the rate-optimality of the test, which means that the test maintains maximum power even when the class of alternatives approach the null at the fastest rate possible. Interested readers are referred to [88] for more details on the concept of rate-optimality for nonparametric tests. However, Theorem 4.3 does not uniformly hold for any choice of continuous function g. The issue of uniform consistency is what we delve into next.

Consider the class of Hölder continuous functions with smoothness order $s \in \mathbb{R}, s \geqslant 2$ and finite Hölder norm, denoted as $\mathbb{H}(s)$. Now for any function $f \in \mathbb{H}(s)$ we denote

$$\mathrm{B}(\mathbb{H}(s), n) = \left\{ f \in \mathbb{H}(s) \middle| \rho_n(f, \mathcal{F}_\Theta) \geqslant C_\alpha \left(\frac{\sqrt{\log \log n}}{n} \right)^{\frac{2s}{4s+1}} \right\}$$

for some $C_\alpha > 0$, where the distance ρ_n is as defined in (4.3). Then for any $s \geqslant 2$, our testing problem can be formulated as

$$H_0 : F \in \mathcal{F}_\Theta \quad \text{vs} \quad H_1^{(n)} : F \in \mathrm{B}(\mathbb{H}(s), n). \tag{4.7}$$

Theorem 4.4. *Under the alternatives of the form (4.7), when C_α is sufficiently large then*

$$\lim_{n\to\infty} \left[\inf_{x \in \mathrm{B}(\mathbb{H}(s),n)} \{P(T^* > t_\alpha)\} \right] = 1.$$

Proof. Take

$$\vartheta = \operatorname*{argmin}_{\theta\in\Theta} \left\{ \sum_{i=1}^{n} (x(t_i) - \zeta - z(t_i, \theta))^2 \right\}$$

and write $d(t_i) = x(t_i) - \zeta - z(t_i, \vartheta)$. By Lemma 4.9 it suffices to show that

$$\int_0^1 \left\{ \sum_{i=1}^{n} d^2(t_i) w^2(t, t_i) \right\}^2 dt \geq \frac{4V_h \widetilde{t}_\alpha^{(0)}}{n\sqrt{h}} \text{ for some } h \in \mathcal{H}_n,$$

which we will establish by approximating $d(.)$ by a piece-wise polynomial and proving that each piece satisfies the required condition. Now let m denote the greatest integer less than s. Since $d \in \mathbb{H}(s)$ using a Taylor expansion in $\mathcal{T}$ with center t_C and scale h

we can write,

$$P(u) = \beta_0 + \beta_1 \left(\frac{u - t_C}{h}\right) + \beta_2 \left(\frac{u - t_C}{h}\right)^2 + \ldots + \beta_m \left(\frac{u - t_C}{h}\right)^m.$$

The approximation error for the above approximation is given by

$$|d(u) - P(u)| = \frac{\left|d^{(m+1)}(c)\right|}{(m+1)!} |u - t_C|^{m+1} h^{m+1} \leq Ch^s, \tag{4.8}$$

for some constant C. This means that $d(.)$ can be approximated as $P(.)$ in $\mathcal{T}$ with error rate $O(h^s)$. Define $\widehat{d}(u) = \sum_{i=1}^n d(t_i)w(u, t_i)$ and $\widehat{P}(u) = \sum_{i=1}^n P(t_i)w(u, t_i)$. Then clearly $\left|\widehat{d}(u) - \widehat{P}(u)\right| \leqslant Ch^s$. Elementary calculations show

$$\sum_{\{i|t_i \in \mathcal{T}\}} \left|\widehat{d}(t_i)\right|^2 \geq \frac{1}{2} \sum_{\{i|t_i \in \mathcal{T}\}} \left|\widehat{P}(t_i)\right|^2 - n_{\mathcal{T}} C^2 h^{2s},$$

where $n_{\mathcal{T}}$ is the no. of elements in the set $\{i|t_i \in \mathcal{T}\}$. Now let V be a $(m+1)\times(m+1)$ matrix with elements

$$V(i, j) = \sum_{\{k|t_k \in \mathcal{T}\}} \left(\frac{t_k - t_C}{h}\right)^{i+j}.$$

Define an array of numbers $\{\psi(i, j)\}, i = 1, 2, \ldots, n, j = 0, 1, \ldots, m$ which satisfy the equation

$$\sum_{j=1}^n \left(\frac{t_j - t_C}{h}\right)^k w(t_i, t_j) = \left(\frac{\psi(i, k) - t_C}{h}\right)^k \text{ for all } k.$$

Let $\widetilde{V}$ be another $(m + 1) \times (m + 1)$ matrix whose $(p, q)-$th element is given by

$$\widetilde{V}(p, q) = \sum_{\{i|t_i \in \mathcal{T}\}} \left(\frac{\psi(i, p) - t_C}{h}\right)^r \left(\frac{\psi(i, q) - t_C}{h}\right)^s.$$

It can be shown that there exists a constant R, depending on C_1 and C_2 from assump-

tion 5, for which $\left\|\widetilde{V}\right\|_\infty \leq R$ and $\left\|\widetilde{V}^{-1}\right\|_\infty \leq R$. This is a straightforward application of the regular design case of [101] and lemma 6.6 of [102]. We can then write,

$$\sum_{\{i|t_i\in\mathcal{T}\}} \left|\widehat{d}(t_i)\right|^2 \geqslant \frac{1}{4R^2} \sum_{\{i|t_i\in\mathcal{T}\}} |d(t_i)|^2 - \left(1+\frac{1}{2R^2}\right) n_{\mathcal{T}} C^2 h^{2s}. \tag{4.9}$$

Now split $[0,1]$ into L non-overlapping intervals $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_L\}$ and apply (4.9) in each of them. Due to assumption 4, for suitable constants C_1^* and C_2^* we can write

$$\sum_{i=1}^{n} \left|\widehat{d}(t_i)\right|^2 \geqslant C_1^* n^{\frac{2s+1}{2s}} (\log\log n)^{\frac{s}{4s+1}} + C_2^* n (\log\log n)^{-2s}. \tag{4.10}$$

Using approximation of an integral as a sum,

$$\int_0^1 \left\{\sum_{i=1}^{n} d^2(t_i) w^2(t,t_i)\right\} dt \approx \frac{1}{2n} \sum_{i=1}^{n}\sum_{j=1}^{n} d^2(t_j) w^2(t_j,t_i) = O\left(\sum_{i=1}^{n} \left|\widehat{d}(t_i)\right|^2\right).$$

The remainder of the proof follows in a straightforward manner. □

This result establishes the uniform consistency of the proposed test against the class of Hölder continuous alternatives. Extending this consistency to broader function classes, such as Besov or Sobolev spaces, can be achieved through similar reasoning, with minor adjustments to the technical assumptions regarding the spacing of observation time points. We skip the details as there are no significant difficulties involved.

4.3 Simulation study

4.3.1 Empirical size and power of the test

This section presents the results of Monte Carlo experiments conducted to evaluate the numerical performance of the adaptive, rate-optimal test. Each experiment involves specifying a parametric null hypothesis model and an alternative models. The alternative is formed by introducing a small sinusoidal perturbation to the null. Monte Carlo simulations are then used to estimate both the test's rejection probability under the null hypothesis and its power against the alternative model. The null hypothesis is defined as

$$\begin{cases} x'(t) = -(k_b + k_e)x(t) + k_t r(t), \\ r'(t) = k_b x(t) - k_t r(t), \\ x(0) = x_0,\ r(0) = r_0, \end{cases} \tag{4.11}$$

where x is the observed variable (observed under noise as in (4.1)) and r is the unobserved/latent variable. Note that (4.11) is widely used to model drug diffusion in plasma and tissue [103]. Under H_0, the true values of the parameters are taken as $k_b = 0.98, k_e = 0.22, k_t = 0.33$. This choice is motivated from the real dataset recording the drug diffusion pattern in different animal subjects, as we will elaborate in (4.4). In the alternative, we introduce a sinusoidal perturbation of the form $A\sin(\omega\pi t)$ with $A = 5$ and $\omega = 20$. The initial conditions are assumed to be pre-specified as $x_0 = 500, r_0 = 0$. All experiments are done for a 5% level of significance. As mentioned previously, all time points are scaled to be in $[0, 1]$. The null and the alternative is presented in (4.1).

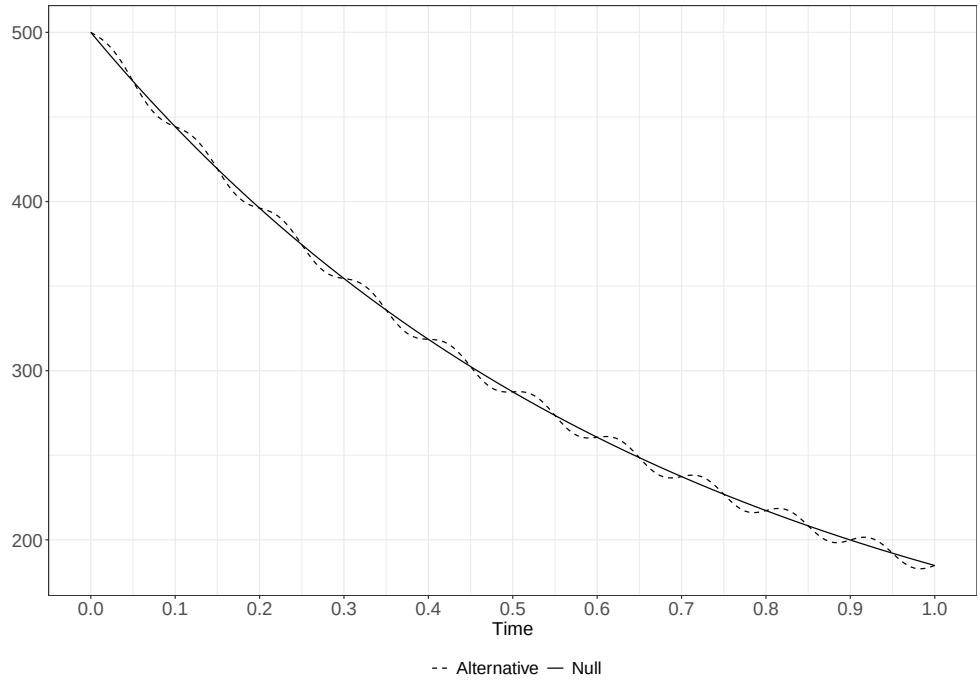


Figure 4.1: Null and alternative hypothesis for MCMC experiments

The simulation procedure is as follows. Data is first generated under the null hypothesis (4.11) (to evaluate size) and under the alternative hypothesis (to evaluate power) for n equispaced time points $\{t_1, t_2, \ldots, t_n\}$, with homoscedastic random noise introduced corresponding to signal-to-noise ratios of 10% and 30%, respectively. Consistent with the earlier notation, the observed noisy system output at time t_i is denoted as y_i for $i = 1, 2, \ldots, n$. Parameter estimation for the null hypothesis model is performed using a nonlinear least squares algorithm, while the error variance is estimated with the Rice estimator [104]. The estimated parameter and the estimated variances are denoted as $\widehat{\theta} = (\widehat{k}_b, \widehat{k}_e, \widehat{k}_t)$ and $\widehat{\sigma}^2$'s respectively. Nonparametric smoothing of the data and the parametric solution of (4.11) are implemented using the Epanechnikov kernel, with bandwidth configurations deliberately chosen to slightly overfit the data for optimal test performance. The experiments are conducted for sample sizes of 50,

100, and 250, the corresponding bandwidth grid were chosen to be 10 equispaced points in the intervals $(0.025, 0.030)$, $(0.015, 0.020)$ and $(0.008, 0.015)$ respectively. Upon evaluating the test statistic, the critical values were obtained using a naive bootstrap procedure. The steps of the bootstrap are as follows.

1. For each $i = 1, 2, \ldots, n$, generate $y_i^* = \zeta + z(t_i, \widehat{\theta}) + \epsilon_i^*$, where ϵ_i^*'s are sampled from $\mathcal{N}(0, \widehat{\sigma}^2)$.

2. Use the dataset $\{y_1^*, y_2^*, \ldots, y_n^*\}$ to estimate the parameter θ in (4.11) and the error variance. Denote these as $\widehat{\theta}^*$ and $\widehat{\sigma}^{2^*}$ respectively.

3. Compute the test statistic using the generated dataset and estimated parameters in step 2. Denote the bootstrapped test statistic as $\mathbf{T}^*$.

4. Repeat steps 1–3 many times to evaluate the $(1-\alpha)-$th quantile of $\mathbf{T}^*$, denote it as t_α. This is taken as the critical value for a test of $100(1-\alpha)\%$ level of significance.

All experiments were conducted in RStudio, using R version 3.6.3. For size computation, 1000 Monte Carlo replications were performed, while 500 replications were used for power estimation. This approach ensured sufficient precision in size calculation while optimizing computational efficiency to maintain an adequate power for the test. The results for the simulation for different sample sizes and signal-noise ratios (SNR) are presented in (4.1). The size of the test remains close to the nominal level across all configurations, demonstrating its reliability. The power of the test increases with larger sample sizes and lower signal-to-noise ratios. For instance, at $n = 50$, the power is 0.396 for an SNR of 10%, while at $n = 250$, the power rises significantly to

0.993. Higher SNRs (30%) result in reduced power, as expected, but the test still exhibits robust performance, particularly with larger sample sizes.

Sample size (n)	**SNR**	**Size**	**Power**
50	10%	0.064	0.396
	30%	0.052	0.174
100	10%	0.062	0.844
	30%	0.064	0.305
250	10%	0.052	0.993
	30%	0.050	0.576

Table 4.1: Size and power of the proposed test for different sample sizes and signal-noise ratios (only x is observed)

We further examined whether the test's performance improves significantly when all components of the system are observable, i.e. both the dynamics of x and r are observed. The test procedure remains unchanged, with the only difference being that the parameter θ is estimated by minimizing the loss function with respect to both variables x and r. Simulation results indicate a marginal improvement in both size and power when the system is completely observed. This underscores the robustness and efficacy of our test for both fully and partially observed systems. We consider the same null hypothesis as specified in (4.3):

$$\begin{cases} x'(t) = -(k_b + k_e)x(t) + k_t r(t), \\ r'(t) = k_b x(t) - k_t r(t), \\ x(0) = x_0,\ r(0) = r_0, \end{cases}$$

where both variables x and r are observed with noise. Similar to the setup described in (4.3), the true values of the parameters under the null hypothesis are taken as $k_b = 0.98$, $k_e = 0.22$, $k_t = 0.33$, with the initial conditions specified as $x_0 = 500$ and $r_0 = 0$.

Under the alternative hypothesis, we introduce a sinusoidal perturbation of the form $A\sin(\omega\pi t)$ to the x component, where $A = 5$ and $\omega = 20$. Note that although both components of the system are observed under noise, the testing is performed only on the x component; hence, the perturbation is introduced solely to the x component. Simultaneous testing for both x and r components, particularly when the system may not be structurally identifiable, is also a feasible extension. However, this lies beyond the scope of the current paper and is left as a direction for future research.

The setup of the simulation is the same, except the least-squares objective function for estimating the unknown parameter θ is modified to incorporate the loss in both the x and r components. The variance of both components are estimated using the estimator proposed in [96]. The power and size for finite sample size in different noise intensities is described in (4.2).

The results indicate a slight advantage in terms of both size and power when both components are observed. Specifically, the size of the test remains below the nominal level even for small sample sizes (n), demonstrating its robustness under limited data. Additionally, the power is comparable to the scenario where only one component is observed (as shown in (4.1)), with minor gains observed for smaller sample sizes. However, as the sample size increases, the performance of the test becomes nearly equivalent regardless of whether all components are observed, suggesting that

Sample size (n)	SNR	Size	Power
50	10%	0.053	0.401
	30%	0.051	0.188
100	10%	0.050	0.857
	30%	0.049	0.332
250	10%	0.049	0.994
	30%	0.042	0.580

Table 4.2: Size and power of the proposed test for different sample sizes and signal-noise ratios (both x and r are observed)

the additional information provided by observing both components has equivalent performance in large-sample settings. This highlights the test's adaptability to different data availability scenarios, performing well even when some components are unobserved.

4.3.2 Comparison of the test with other exiting tests in the literature

We now compare the performance of the test statistic with that of [70] and [72]. We consider the null hypothesis of a linear form as follows.

$$H_0 : \begin{cases} x_1'(t) = a\tau x_1(t) \\ x_2'(t) = \tau\left(x_1(t) + bx_2(t)\right), \end{cases}$$

where (a, b, τ) are unknown parameters. We consider different forms of oscillating and low-frequent perturbations in the alternative as follows.

$$H_1^{(1)} : \begin{cases} x_1'(t) = \tau \left(a x_1(t) + 0.2 \cos \left(a x_1(t) \right) \right) \\ x_2'(t) = \tau \left(a x_1(t) + b x_2(t) + 0.2 \cos \left(a x_1(t) + b x_2(t) \right) \right), \end{cases}$$

$$H_1^{(2)} : \begin{cases} x_1'(t) = \tau \left(a x_1(t) + 0.05 (a x_1(t))^3 \right) \\ x_2'(t) = \tau \left(a x_1(t) + b x_2(t) + 0.05 (a x_1(t) + b x_2(t))^3 \right), \end{cases}$$

$$H_1^{(3)} : \begin{cases} x_1'(t) = \tau \left(a x_1(t) + \exp \left(a x_1(t) \right) \right) \\ x_2'(t) = \tau \left(a x_1(t) + b x_2(t) + 2.5 \exp \left(a x_1(t) + b x_2(t) \right) \right). \end{cases}$$

Under the null hypothesis the true parameter values are taken as $(\tau, a, b) = (10, -0.06, -0.24)$. The data is generated from the above systems with independent random noise following a $\mathcal{N}(0, 0.05^2)$ distribution. The sample size is set as $n = 300$ and all tests are done at 5% level of significance.

As shown in Table 4.3, our proposed test exhibits comparable power to existing methods in both fully and partially observed settings. When the entire ODE system is observed, all tests perform similarly in terms of power. However, in scenarios where only one component of the system is observed, our test often outperforms the gradient-matching-based test statistic proposed by [72]. Notably, our test achieves power levels close to those of [70]. The results remain robust across all choices of alternatives considered.

Both Observed				
	Size	**Power**		
		$H_1^{(1)}$	$H_1^{(2)}$	$H_1^{(3)}$
[70]	0.047	1.000	1.000	1.000
[72] (TM)	0.045	1.000	1.000	1.000
Our test	0.052	1.000	1.000	1.000
Only x_1 Observed				
	Size	**Power**		
		$H_1^{(1)}$	$H_1^{(2)}$	$H_1^{(3)}$
[70]	0.047	1.000	1.000	1.000
[72] (GM)	0.038	0.270	0.919	0.296
[72] (IM)	0.036	0.499	1.000	0.134
Our test	0.061	1.000	0.934	1.000
Only x_2 Observed				
	Size	**Power**		
		$H_1^{(1)}$	$H_1^{(2)}$	$H_1^{(3)}$
[70]	0.047	1.000	1.000	1.000
[72] (GM)	0.034	0.193	0.606	0.062
[72] (IM)	0.058	0.998	1.000	0.721
Our test	0.054	1.000	1.000	1.000

Table 4.3: Comparison of size and power for different tests under various completely and partially observed setting. Here TM, IM and GM respectively denote the trajectory-matching, integral -matching and gradient-matching based test statistic.

4.4 Applications

Although the theory presented in this chapter primarily focuses on partially observed systems of differential equations, we would like to highlight that it is applicable to a wider range of systems. To demonstrate this, we apply the theory to three real-world examples. The first example involves testing the dynamics of drug diffusion between the blood and tissue compartments over time. Here, the concentration of the drug in tissue is difficult and costly to measure, making it an unobserved variable. This scenario directly aligns with the partially observed systems discussed in this chapter. The second example examines the relationship between crop yield and soil fertilizer, where "crop yield" serves as a proxy for"time" in the system's evolution. This is a completely observed system, as data on both the response (crop yield) and the predictor (soil fertilizer) are available. Finally, we explore predator-prey dynamics between two pairs of bacterial species. This is another completely observed system, though it involves multiple responses, all of which are observed. As detailed in this section, the test yields interpretable results in all three cases. Notably, despite having a limited number of sample points in each case, the underlying dynamics are sufficiently captured, allowing for reliable inference. This was achieved through careful experimental design setups inspired by [105].

4.4.1 Mathematical modeling of drug-diffusion data

The mathematical modeling of drug diffusion in complex systems combines biology, physics, and computational mathematics to understand the absorption, distribution, metabolism, and elimination of drugs within biological systems. This research is crucial for optimizing drug delivery, improving efficacy, and minimizing side effects.

Models integrate diffusion theory, reaction kinetics, and fluid dynamics to simulate drug transport, accounting for factors like molecular size, solubility, and biological barriers, ranging from simple compartment models to more complex partial differential equations. For a comprehensive introduction, readers may refer to [106] and [107].

A widely used mathematical framework in this context, known as "compartment modeling" [108], simplifies biological systems by dividing them into functional compartments representing specific physiological regions or tissue groups. This framework assumes that, after administration, the drug is absorbed into these compartments through diffusion and elimination. In particular, the two-compartment model [103, 109] considers two compartments: plasma and tissue. The drug is transferred from the plasma to the tissue at a rate k_b, diffuses back from tissue to plasma at a rate k_t, and is eliminated from the system at a rate k_e. The dynamics of this system resemble that described in (4.11), where $x(t)$ represents the drug concentration in plasma and $r(t)$ in the tissue. As noted in (4.3), due to the high cost of measuring tissue concentration, we can only observe the drug concentration in the plasma (x).

We test the null hypothesis of (4.11) using experimental data on the decay of a chemical's concentration in the plasma over time in a specific species of laboratory rat. The data is provided by US Department of Health and Human Services, and can be found here. In this experiment, the rat was administered a dose of 300 μg/kg of the drug, and its plasma concentration was measured over a 48-hour period. The measurements were taken at 12 irregularly spaced time points during this period. The decay of the plasma concentration is illustrated in the below plot.

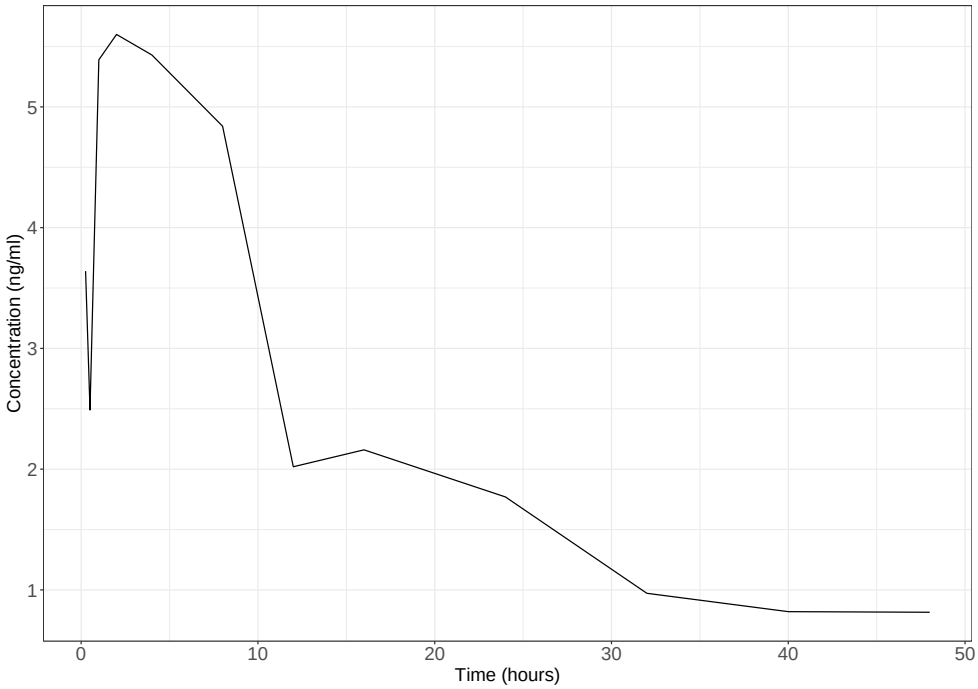


Figure 4.2: Decay of drug concentration in plasma post-intake

Note that the concentration of the drug in plasma at $t = 0$ was not measured but back extrapolated, typically it was set to be the first observed. The test was conducted using the Epanechnikov kernel and the bandwidth grid was taken to be 10 equispaced points in the interval $(0.05, 0.07)$. The test was accepted at 5% level of significance based on the observed data. The same experiment with 10 other subjects of the same species. Each of them were injected with the same dose of 300 μg/kg at $t = 0$, and were observed over different time spans. The test was conducted for each of the subjects individually. However, since the number of observation time points was approximately the same across all subjects, a common bandwidth grid was used for consistency: 10 equally spaced points within the interval $(0.05, 0.07)$. The results of the test are presented in (4.3).

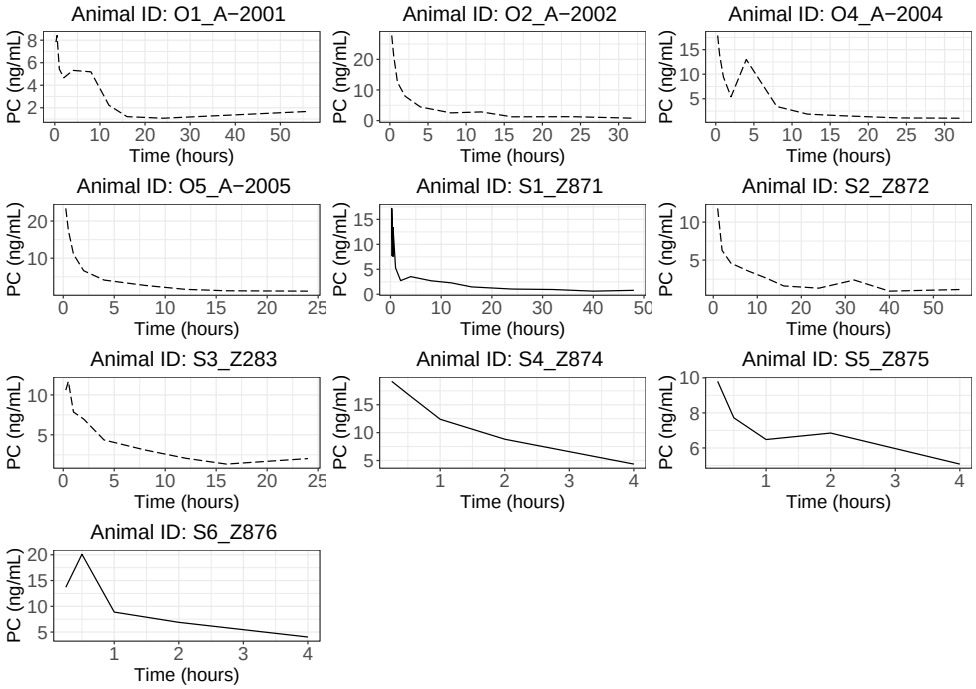


Figure 4.3: Decay of drug concentration in plasma post-intake in different subjects. 'PC' denotes plasma-concentration. The subjects with dashed-line **rejected** the null hypothesis.

Our test results indicate that 5 out of the 11 subjects accepted the null hypothesis, while the remaining 6 rejected it. However, it is important to note that limited data was available for some subjects, which may impact the reliability of the test results for those cases. Each subject was tested individually, as outlined in the methodology. Nevertheless, simultaneous testing across all subjects could potentially enhance the accuracy of the test, particularly in scenarios with limited data. Exploring this approach is beyond the scope of the current study and is left for future research.

4.4.2 Mathematical modeling of agricultural trial data

Understanding the dynamics between crop yield and fertilizer levels is critical for optimizing agricultural productivity and sustainability. Properly managing fertilization

not only maximizes yields but also minimizes environmental impacts, such as soil degradation and water pollution caused by excessive fertilizer use. A deeper understanding of this relationship helps inform data-driven agricultural practices, ensuring efficient resource allocation and promoting long-term soil health. It also supports policymakers in designing strategies that balance food security with ecological sustainability, especially in regions with constrained agricultural resources.

We analyze agricultural trial data from the Rothamsted Research field trials to study the relationship between crop yields and soil fertilization. Specifically, we use the data provided by [110], which includes the yields of spring barley from 20 fields in relation to soil phosphorus levels. Let y denote the soil fertilizer level and x the crop yield. As the null hypothesis, we consider the Mitscherlich model, an exponential form widely used in the literature to model crop growth dynamics [111, 112]. The model is given by:

$$\begin{cases} y(x) = \alpha_1(\alpha_2 - y(x)) \\ y(0) = \alpha_0 \end{cases}$$

where $\alpha_0, \alpha_1, \alpha_2$ are model parameters estimated using nonlinear least squares as described before. The fitted model is illustrated in (4.4). The test was conducted using the Epanechnikov kernel the bandwidth grid was taken to be 10 equispaced points in the interval $(0.90, 1.00)$. The data accepts the null hypothesis at 5% level of significance. This aligns with the findings of [86] which also speaks in favor of the exponential model sufficiently capturing the yield-fertilizer dynamics.

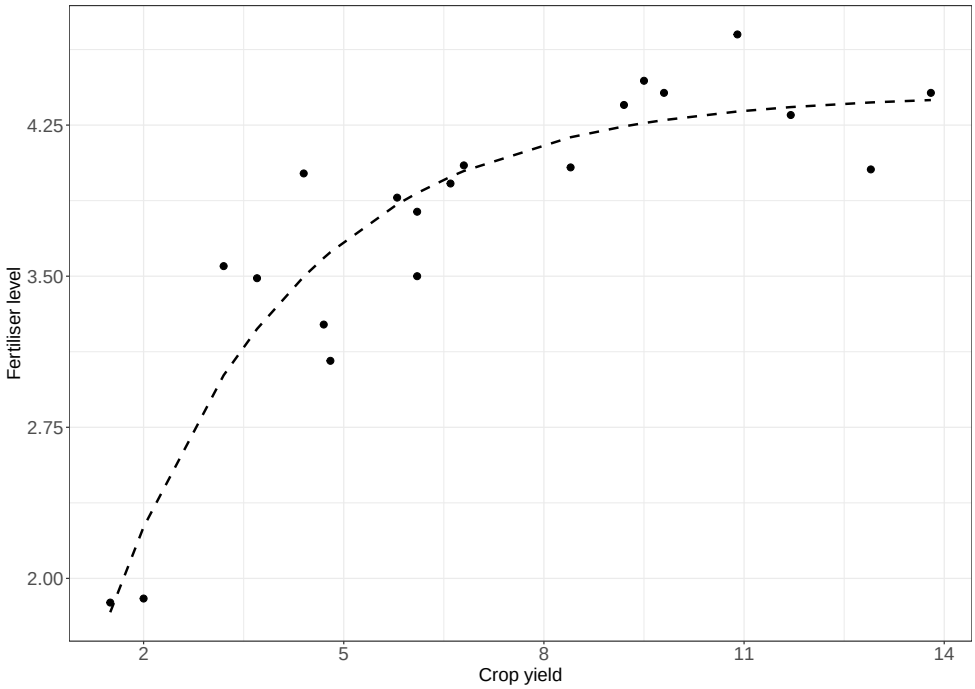


Figure 4.4: Dynamics between crop yield and phosphorus fertilizer level

4.4.3 Mathematical modeling of predator-prey dynamics data

Understanding predator-prey dynamics is crucial for studying ecological systems, as it provides insights into how populations interact and stabilize over time. These dynamics play a key role in maintaining biodiversity, regulating ecosystem functions, and informing conservation strategies. Accurate models of predator-prey interactions allow researchers to predict changes in population sizes under various environmental conditions, assess the impacts of external factors such as habitat loss or climate change, and design effective management interventions for endangered or invasive species. These models also serve as valuable tools in laboratory experiments, helping to replicate and understand ecological behavior in controlled settings.

We test the predator-prey dynamics data using two different predator-prey pairs

of microorganisms. The data originates from the lab experiments of [113], which observed the population counts of the predatory species and the prey species over time in a controlled environment. As the null hypothesis, we consider the widely used Lotka-Volterra model, which has been a cornerstone of predator-prey dynamics research in the existing literature [114], despite its well-documented limitations [115, 116]. The model is given by:

$$\begin{cases} y_1(t) = \alpha_2\alpha_3 y_1(t)y_2(t) - \alpha_4 y_1(t) \\ y_2(t) = \alpha_1 y_2(t) - \alpha_2 y_1(t)y_2(t) \\ y_1(0) = \alpha_{1_0}, y_2(0) = \alpha_{2_0} \end{cases}$$

where y_1 and y_2 respectively denote the numbers of predators and prey respectively and $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_{1_0}, \alpha_{2_0}$ are unknown parameters. The parameters are estimated using nonlinear least squares as before by minimizing the data loss due to both observed variables. Note that in this setup both y_1 and y_2 are to be tested jointly, which is implemented by a Bonferroni correction. While a more sophisticated way of simultaneous testing might be possible we leave it for future work. All tests was conducted using the Epanechnikov kernel the bandwidth grid was taken to be 10 equispaced points in the interval $(0.07, 0.08)$. We find that both the concerned predator-prey microorganism pairs *Paramecium aurelia* (predator) vs *Saccharomyces exiguus* (prey), and *Paramecium bursaria* (predator) vs *Schizosaccharomyces pombe* (prey) reject the Lotka-Volterra model at 5% level of significance. This further aligns with the findings of [86], which favours more sophisticated models over the basic Lotka-Volterra model in capturing the predator-prey dynamics between microorganisms. The fitted models for the respective pairs are shown in (4.5) and (4.6) respectively.

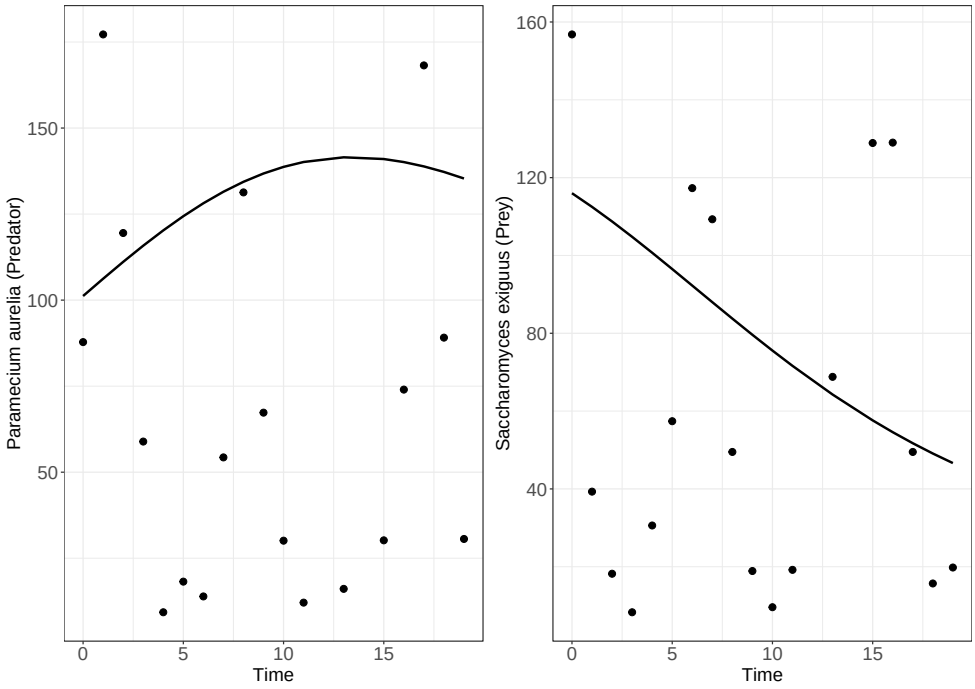


Figure 4.5: Predator-prey dynamics between Paramecium aurelia (predator) and Saccharomyces exiguus (prey)

4.5 Conclusion

In this chapter, we introduce a novel lack-of-fit test for systems of ordinary differential equations (ODEs) that may not have full observability of all their components. Our proposed test is designed to be adaptive to the smoothness of the alternative model, making it highly flexible and rate-optimal, ensuring that it remains robust across a wide range of potential system dynamics. The theoretical framework we provide guarantees both the size and asymptotic consistency of the test, addressing different forms of model misspecifications in a comprehensive manner. Importantly, if the system is found to be misspecified, this framework can pinpoint the component of the system potentially responsible for the misspecification, provided that the system is structurally identifiable. This capability marks a significant advancement in the

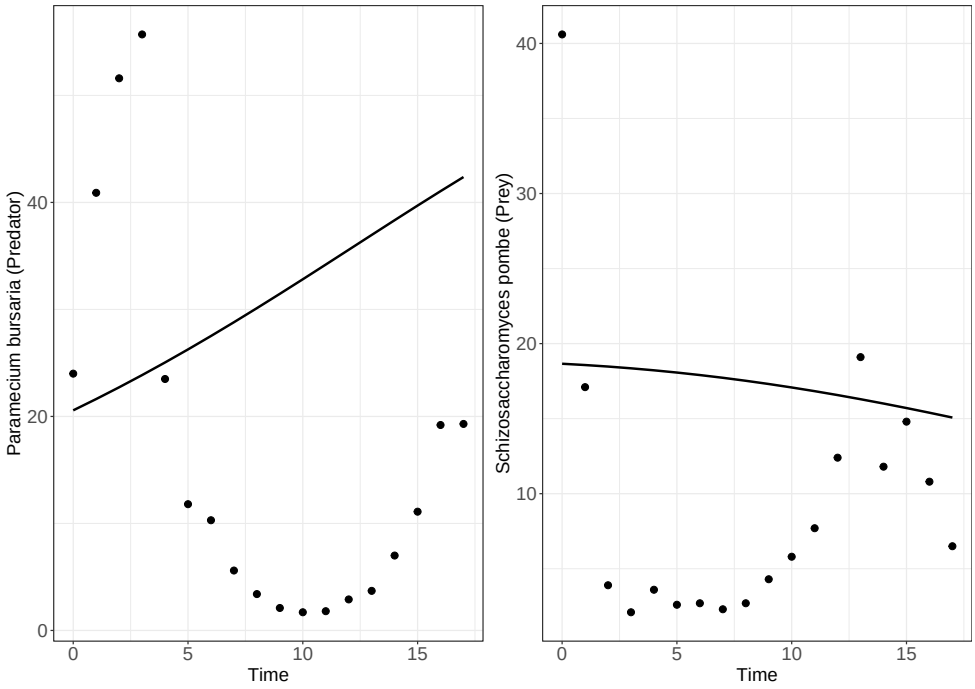


Figure 4.6: Predator-prey dynamics between Paramecium bursaria (predator) and the prey Schizosaccharomyces pombe (prey)

field, as it allows for more targeted diagnostics of model fit.

To validate the practical applicability of the test, we conduct an extensive simulation study that demonstrates its finite-sample performance. Furthermore, we apply the proposed test to a vast variety of experimental datasets. In each application, we perform the test under the null hypothesis that each observation is generated from the same dynamics as specified by H_0. A more complex scenario arises when we hypothesize that each observation arises from a dynamical system of the same parametric form but with the parameters subject to random fluctuations around a baseline unknown value. In this case, direct parameter estimation is not feasible due to the limited availability of a single output per subject at each time point. This essentially leads to the random design problem, which is more intricate and lies beyond

the scope of the current chapter. We leave it to future research.

Another promising direction for future research involves extending our approach to handle systems with multiple observed variables, such as systems of partial differential equations or ODEs with multiple components observable. Additionally, we aim to adapt our methodology to stochastic differential equations, which would open new avenues for applications in fields like financial econometrics, where modeling uncertainty in dynamical systems is of paramount importance. These extensions would significantly enhance the versatility of our method, broadening its impact across various domains that require sophisticated model testing and misspecification analysis.

4.6 Appendix: Prerequisite lemmas

Lemma 4.1. *For some finite constants C_{N_1} and C_{N_2}, $\frac{C_{N_1}}{\sqrt{h}} \leqslant N_h \leqslant \frac{C_{N_2}}{\sqrt{h}}$ for all $h \in \mathcal{H}_n$.*

Proof. We have,

$$N_h = n\sqrt{h}\sum_{i=1}^{n}\sigma_i^2\int_0^1 w^2(t,t_i)dt.$$

Recall the constants C_1 and C_2 defined in assumption 5, and δ_K, κ defined in assumption 8 of (4.2.4). We have,

$$N_h \leqslant n^2\sqrt{h}\max_{i=1(1)n}\left\{\sigma_i^2\right\}\max_{i=1(1)n}\left\{\int_0^1 w^2(t,t_i)dt\right\} \leqslant n^2\sqrt{h}\max_{i=1(1)n}\left\{\sigma_i^2\right\}\frac{2\delta_K^2}{\kappa^2 C_1^2 n^2 h} = \frac{C_{N_2}}{\sqrt{h}}$$

Again,

$$N_h \geqslant n^2\sqrt{h}\min_{i=1(1)n}\{\sigma_i^2\}\int_0^1 w^2(t,t_i)dt \geqslant n^2\sqrt{h}\min_{i=1(1)n}\{\sigma_i^2\}.\frac{2\kappa^2\delta_K}{C_2^2 n^2 h} = \frac{2\kappa^2\delta_K \min_{i=1(1)n}\{\sigma_i^2\}}{C_2^2\sqrt{h}} = \frac{C_{N_1}}{\sqrt{h}}$$

Combining these two equations the proof is complete. □

Lemma 4.2. *For some finite constants C_{V_1} and C_{V_2}, $\frac{C_{V_1}}{h} \leqslant V_h^2 \leqslant \frac{C_{V_2}}{\sqrt{h}}$ for all $h \in \mathcal{H}_n$.*

Proof. We have,

$$V_h^2 = n^2 h \left[2 \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_i^2 \sigma_j^2 \left(\int_0^1 w(t,t_i) w(t,t_j) dt \right)^2 + \sum_{i=1}^{n} S_i \left(\int_0^1 w^2(t,t_i) dt \right)^2 \right],$$

where $S_i = E[\epsilon_i^4], i = 1, 2, \ldots, n$. Recall the constants C_1 and C_2 defined in assumption 5, and δ_K, κ defined in assumption 8 of (4.2.4). Denote,

$$V_1(h) = 2n^2 h \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_i^2 \sigma_j^2 \left(\int_0^1 w(t,t_i) w(t,t_j) dt \right)^2,$$

$$V_2(h) = n^2 h \sum_{i=1}^{n} S_i \left(\int_0^1 w^2(t,t_i) dt \right)^2.$$

Elementary calculations show that

$$V_2(h) \leqslant \frac{2\delta_K^2 \max_{i=1(1)n}\{S_i\}}{\kappa^2 C_1^2}.$$

Therefore we can conclude $V_2(h) = o(1)$, implying $V_h \approx V_1(h)$ asymptotically. On the other hand we have,

$$V_1(h) \leqslant \frac{4\delta_K^2 C_N \max_{i=1(1)n}\{\sigma_i^4\}}{\kappa^2 C_1^2 \sqrt{h}},$$

where $\|W_h\|_\infty \leqslant C_N$ for some finite positive constant C_N. The other part of the inequality similarly follows. □

Lemma 4.3. *Denote $b_h(\theta) = n\sqrt{h} \int_0^1 \left\{ \sum_{i=1}^{n} \left(z(t_i,\theta_0) - z(t_i,\theta) \right) w(t,t_i) \right\}^2 dt$. Then*

for every $\delta > 0$

$$\max_{h \in \mathcal{H}_n} \left\{ \sup_{\{\theta \in \Theta || \theta - \theta_0| \leqslant \delta\}} \{b_h(\theta)\} \right\} = o(n)$$

Proof. The proof of this lemma is straightforward. We skip the details as no technical difficulties are involved. □

Lemma 4.4. $\mathbf{T}_h(\widehat{\theta}^*) = T_h(\theta_0) + o_p(1)$ *uniformly for all* $h \in \mathcal{H}_n$.

Proof. Let $\widehat{\sigma}_i^{*^2}$ be the heteroscedasticity estimate from the bootstrap data and $\epsilon_i^*, i = 1, 2, \ldots, n$ be independently sampled from $\mathcal{N}(0, \widehat{\sigma}_i^{*^2})$. Define,

$$A_1 = \frac{nh}{\sqrt{C_{V_1}}} \left| \int_0^1 \left\{ \sum_{i=1}^n \sum_{j=1, j \neq i}^n \epsilon_i^* \left(z(t_j, \widehat{\theta}) - z(t_j, \widehat{\theta}^*) \right) w(t, t_i) w(t, t_j) \right\} dt \right|$$

$$A_2 = \frac{nh}{\sqrt{C_{V_1}}} \left| \int_0^1 \left\{ \sum_{i=1}^n \left(z(t_i, \widehat{\theta}) - z(t_i, \widehat{\theta}^*) \right) w(t, t_i) \right\}^2 dt \right|$$

Using Lemma 4.3 we can write,

$$\left| \mathbf{T}_h(\widehat{\theta}^*) - T_h(\theta_0) \right| \leqslant A_1 + A_2$$

Since $\widehat{\theta}^*$ is a $\sqrt{n}-$consistent estimate of $\widehat{\theta}$ from the bootstrapped data, under the assumptions on the kernel function and the observation points, elementary calculations show that $A_1 = o_p\left(\frac{1}{\sqrt{n}}\right)$ and $A_2 = o_p(1)$. This proves the statement of the lemma. □

Lemma 4.5. $T_h(\widehat{\theta}) = T_h(\vartheta) + o_p(1)$ *uniformly for all* $h \in \mathcal{H}_n$.

Proof. We can write,

$$S_h(\widehat{\theta}) = S_h(\vartheta) + n\sqrt{h}\int_0^1 \left\{\sum_{i=1}^{n}\left(z(t_i,\vartheta) - z(t_i,\widehat{\theta})\right)w(t,t_i)\right\}^2 dt + o_p(1)$$

Now if H_0 is false then $\widehat{\theta}$ is a $\sqrt{n}-$consistent estimate of ϑ. Then using a similar argument as in Lemma 4.4 we can write,

$$n\sqrt{h}\int_0^1 \left\{\sum_{i=1}^{n}\left(z(t_i,\vartheta) - z(t_i,\widehat{\theta})\right)w(t,t_i)\right\}^2 dt = o_p(1)$$

Now consider the case where H_0 is true. Then $\theta_0 = \vartheta$, therefore the statement of the lemma trivially holds true. □

Lemma 4.6. *If H_0 is true then $T_h(\vartheta) = T_h(\theta_0) + o_p(1)$ uniformly for all $h \in \mathcal{H}_n$.*

Proof. Since $\theta_0 = \vartheta$ under H_0 therefore the statement of the lemma trivially holds true. □

Lemma 4.7. *For any $z \geqslant 1$ and sufficiently large n, $P\left(T_h(\theta_0) > z\right) \leqslant \exp\left(-\frac{z^2}{4}\right)$ uniformly for all $h \in \mathcal{H}_n$.*

Proof. Let $\Sigma = \text{diag}\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2\right)$ be the variance-covariance matrix of $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^\intercal$. Recall W_h from (**??**). Then

$$S_h(\theta_0) = \epsilon^\intercal W_h \epsilon. \tag{4.12}$$

We can write $\epsilon = \Pi\Upsilon$, where Υ is a vector of i.i.d $\mathcal{N}(0,1)$ random variables, and $\Pi = \Sigma^{-1}$. Then, (4.12) simplifies to:

$$S_h(\theta_0) = \Upsilon^\intercal \Pi^\intercal W_h \Pi \Upsilon.$$

Now let $\Lambda = \text{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_n)$ be a diagonal matrix containing the eigenvalues of $\Pi^\intercal W_h \Pi$ and let P be an orthogonal matrix such that $\Pi^\intercal W_h \Pi = \Pi^\intercal \Lambda P$. Define $Z = P\Upsilon$. Note that the elements of Z are independent $\mathcal{N}(0,1)$ variables. Then we can write,

$$\epsilon^\intercal W_h \epsilon = \sum_{i=1}^{n} \lambda_i Z_i^2, E\left[\epsilon^\intercal W_h \epsilon\right] = \sum_{i=1}^{n} \lambda_i \text{ and } \text{Var}\left[\epsilon^\intercal W_h \epsilon\right] = 2\sum_{i=1}^{n} \lambda_i^2.$$

Denote, $\nu^2 = 2\sum_{i=1}^{n} \lambda_i^2$. The MGF of $T_h(\theta_0)$ can be written as the following.

$$E\left[\exp\left(\frac{r}{\nu}\sum_{i=1}^{n}\lambda_i\left(Z_i^2-1\right)\right)\right] = \prod_{i=1}^{n}\exp\left(-\frac{r\lambda_i}{\nu} - \frac{1}{2}\log\left(1-\frac{2r\lambda_i}{\nu}\right)\right), \frac{r\lambda_i}{\nu} < 1 \quad (4.13)$$

Now for large enough n, $\lambda_i < \nu\delta$ for an arbitrarily chosen $\delta > 0$. Applying the inequality $-\log(1-u) \leqslant u + u^2$ (for small enough u) to (4.13) we can write

$$E\left[\exp\left(\frac{r}{\nu}\sum_{i=1}^{n}\lambda_i\left(Z_i^2-1\right)\right)\right] \leqslant \exp(r^2)$$

The statement of the lemma then follows due to exponential Chebyshev's inequality taking $r = \frac{z}{2}$. □

Lemma 4.8. *Let J_n be the length of the user specified bandwidth grid $\mathcal{H}_n$ and $t_\alpha^{(0)}$ be the $(1-\alpha)-th$ quantile of $T_0 = \max_{h\in\mathcal{H}_n}\{T_h(\theta_0)\}$. Then $t_\alpha^{(0)} \leqslant 2\sqrt{\log J_n - \log\alpha}$ for sufficiently large n.*

Proof. Using Lemma 4.7 we can write

$$P\left(T_0 > z\right) \leqslant P\left(\bigcup_{h\in\mathcal{H}_n}\{T_h(\theta_0) > z\}\right) \leqslant J_n \exp\left(-\frac{z^2}{4}\right)$$

Now if $z = t_\alpha^{(0)}$ is the $(1-\alpha)-$th quantile of T_0 then $P\left(T_0 > t_\alpha^{(0)}\right) = \alpha$, consequently the statement of the lemma holds. □

Lemma 4.9. *Define* $\widetilde{t}_\alpha^{(0)} = \max\left\{t_\alpha^{(0)}, \sqrt{2\log J_n + \sqrt{2\log J_n}}\right\}$. *If*

$$\int_0^1 \sum_{i=1}^n \left[\left(x(t_i) - \zeta - z(t_i,\vartheta)\right) w(t,t_i)\right]^2 dt \geq \frac{4V_h \widetilde{t}_\alpha^{(0)}}{n\sqrt{h}} \tag{4.14}$$

for some $h \in \mathcal{H}_n$ *then* $\lim_{n\to\infty} P(T^* > t_\alpha) = 1$.

Proof. Due to the equivalence lemmas Lemma 4.4, Lemma 4.5 and Lemma 4.6 it would be enough to prove that $\lim_{n\to\infty} P(T_h(\vartheta) > t_\alpha^{(0)}) = 1$ holds for some $h \in \mathcal{H}_n$. Now since

$$T_h(\vartheta) = T_h(\theta_0) + \frac{n\sqrt{h}}{V_h}\int_0^1 \left\{\sum_{i=1}^n \left(x(t_i) - \zeta - z(t_i,\vartheta)\right) w(t,t_i)\right\}^2 dt$$

holds true, (4.14) implies

$$\lim_{n\to\infty} P(T_h(\vartheta) > t_\alpha^{(0)}) \geq P\left(T_h(\theta_0) > t_\alpha^{(0)} - 4\widetilde{t}_\alpha^{(0)}\right).$$

If $\widetilde{t}_\alpha^{(0)} = t_\alpha^{(0)}$, the statement of the lemma follows directly from Lemma 4.8. Alternatively, if $\widetilde{t}_\alpha^{(0)} = \sqrt{2\log J_n + \sqrt{2\log J_n}}$, the statement of the lemma holds because $2\sqrt{\log J_n - \log\alpha}$ grows asymptotically slower than $4\sqrt{2\log J_n + \sqrt{2\log J_n}}$ as $n \to \infty$.

□

Bibliographic references

1. Andreou, E. & Ghysels, E. Structural breaks in financial time series. *Handbook of financial time series,* 839–870 (2009).

2. Beaulieu, C., Chen, J. & Sarmiento, J. L. Change-point analysis as a tool to detect abrupt climate variations. *Philosophical transactions of the royal society a: mathematical, physical and engineering sciences* **370,** 1228–1249 (2012).

3. Fragkiadakis, A., Tragos, E., Kovacevic, L. & Charalampidis, P. *A practical implementation of an adaptive compressive sensing encryption scheme* in *2016 ieee 17th international symposium on a world of wireless, mobile and multimedia networks (wowmom)* (2016), 1–6.

4. Morabbi, A. *et al.* A multiple changepoint approach to hydrological regions delineation. *Journal of hydrology* **604,** 127118 (2022).

5. Dette, H. & Gösmann, J. A likelihood ratio approach to sequential change point detection for a general class of parameters. *Journal of the american statistical association* **115,** 1361–1377 (2020).

6. Aue, A. & Kirch, C. The state of cumulative sum sequential changepoint testing 70 years after Page. *Biometrika* **111,** 367–391 (2024).

7. Truong, C., Oudre, L. & Vayatis, N. Selective review of offline change point detection methods. *Signal processing* **167,** 107299 (2020).

8. Davis, R. A., Lee, T. C. M. & Rodriguez-Yam, G. A. Structural break estimation for nonstationary time series models. *Journal of the american statistical association* **101,** 223–239 (2006).

9. Yau, C. Y. & Zhao, Z. Inference for multiple change points in time series via likelihood ratio scan statistics. *Journal of the royal statistical society series b: statistical methodology* **78,** 895–916 (2016).

10. Robbins, M. W., Gallagher, C. M. & Lund, R. B. A general regression changepoint test for time series data. *Journal of the american statistical association* **111,** 670–683 (2016).

11. He, Z. & Maheu, J. M. Real time detection of structural breaks in GARCH models. *Computational statistics & data analysis* **54,** 2628–2640 (2010).

12. Chen, C. W., Gerlach, R. & Liu, F.-C. Detection of structural breaks in a time-varying heteroskedastic regression model. *Journal of statistical planning and inference* **141,** 3367–3381 (2011).

13. Kirch, C. & Reckruehm, K. Data segmentation for time series based on a general moving sum approach. *Annals of the institute of statistical mathematics* **76,** 393–421 (2024).

14. Liu, S., Yamada, M., Collier, N. & Sugiyama, M. Change-point detection in time-series data by relative density-ratio estimation. *Neural networks* **43,** 72–83 (2013).

15. Haynes, K., Fearnhead, P. & Eckley, I. A. A computationally efficient nonparametric approach for changepoint detection. *Statistics and computing* **27,** 1293–1305 (2017).

16. Zou, C., Yin, G., Feng, L. & Wang, Z. Nonparametric maximum likelihood approach to multiple. *The annals of statistics* **42,** 970–1002 (2014).

17. Matteson, D. S. & James, N. A. A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the american statistical association* **109,** 334–345 (2014).

18. Diop, M. L. & Kengne, W. A general procedure for change-point detection in multivariate time series. *Test* **32,** 1–33 (2023).

19. Zhang, T. & Lavitas, L. Unsupervised self-normalized change-point testing for time series. *Journal of the american statistical association* **113,** 637–648 (2018).

20. Sundararajan, R. R. & Pourahmadi, M. Nonparametric change point detection in multivariate piecewise stationary time series. *Journal of nonparametric statistics* **30,** 926–956 (2018).

21. Fu, Z., Hong, Y. & Wang, X. On multiple structural breaks in distribution: an empirical characteristic function approach. *Econometric theory* **39,** 534–581 (2023).

22. Casini, A. & Perron, P. Change-point analysis of time series with evolutionary spectra. *Journal of econometrics* **242,** 105811 (2024).

23. Gösmann, J., Kley, T. & Dette, H. A new approach for open-end sequential change point monitoring. *Journal of time series analysis* **42,** 63–84 (2021).

24. Gösmann, J., Stoehr, C., Heiny, J. & Dette, H. Sequential change point detection in high dimensional time series. *Electronic journal of statistics* **16,** 3608–3671 (2022).

25. Xia, Z. & Qiu, P. Jump information criterion for statistical inference in estimating discontinuous curves. *Biometrika* **102,** 397–408 (2015).

26. Wang, Y. Multiple change-point detection for regression curves. *Canadian journal of statistics,* e11816 (2024).

27. Vogt, M. Testing for structural change in time-varying nonparametric regression models. *Econometric theory* **31,** 811–859 (2015).

28. Yang, Q., Li, Y.-N. & Zhang, Y. Change point detection for nonparametric regression under strongly mixing process. *Statistical papers* **61,** 1465–1506 (2020).

29. Fu, Z. & Hong, Y. A model-free consistent test for structural change in regression possibly with endogeneity. *Journal of econometrics* **211,** 206–242 (2019).

30. Cui, Y., Yang, J. & Zhou, Z. State-domain change point detection for nonlinear time series regression. *Journal of econometrics* **234,** 3–27 (2023).

31. Gao, J., Gijbels, I. & Van Bellegem, S. Nonparametric simultaneous testing for structural breaks. *Journal of econometrics* **143,** 123–142 (2008).

32. Wishart, J. & Kulik, R. Kink estimation in stochastic regression with dependent errors and predictors. *Electronic journal of statistics* (2010).

33. Goldenshluger, A., Tsybakov, A. & Zeevi, A. Optimal change-point estimation from indirect observations. *The annals of statistics,* 350–372 (2006).

34. Wu, W. B. Empirical processes of long-memory sequences. *Bernoulli* **9,** 809–831 (2003).

35. Fan, J. A selective overview of nonparametric methods in financial econometrics. *Statistical science,* 317–337 (2005).

36. Tong, H. *Non-linear time series: a dynamical system approach* (Oxford university press, 1990).

37. Wu, W. B. Nonlinear system theory: Another look at dependence. *Proceedings of the national academy of sciences* **102,** 14150–14154 (2005).

38. Wu, W. B. & Zhao, Z. Inference of trends in time series. *Journal of the royal statistical society: series b (statistical methodology)* **69,** 391–410 (2007).

39. Grama, I. & Haeusler, E. An asymptotic expansion for probabilities of moderate deviations for multivariate martingales. *Journal of theoretical probability* **19,** 1–44 (2006).

40. Hall, P., Sheather, S. J., Jones, M. & Marron, J. S. On optimal data-based bandwidth selection in kernel density estimation. *Biometrika* **78,** 263–269 (1991).

41. Giordano, F. & Parrella, M. L. Neural networks for bandwidth selection in local linear regression of time series. *Computational statistics & data analysis* **52,** 2435–2450 (2008).

42. Korkas, K. K. & Fryzlewicz, P. Multiple change-point detection for non-stationary time series using wild binary segmentation. *Statistica sinica* **27,** 000–000 (2017).

43. Dufays, A., Houndetoungan, E. A. & Coën, A. Selective linear segmentation for detecting relevant parameter changes. *Journal of financial econometrics* **20,** 762–805 (2022).

44. Palm, F. C. 7 garch models of volatility. *Handbook of statistics* **14,** 209–240 (1996).

45. Killick, R., Fearnhead, P. & Eckley, I. A. Optimal detection of changepoints with a linear computational cost. *Journal of the american statistical association* **107,** 1590–1598 (2012).

46. Critien, J. V., Gatt, A. & Ellul, J. Bitcoin price change and trend prediction through twitter sentiment and data volume. *Financial innovation* **8,** 1–20 (2022).

47. Figà-Talamanca, G. & Patacca, M. Disentangling the relationship between bitcoin and market attention measures. *Journal of industrial and business economics* **47,** 71–91 (2020).

48. Fan, J. & Gijbels, I. *Local polynomial modelling and its applications: monographs on statistics and applied probability 66* (CRC Press, 1996).

49. Yang, Y., Xu, Y. & Song, Q. Spline confidence bands for variance functions in nonparametric time series regressive models. *Journal of nonparametric statistics* **24,** 699–714 (2012).

50. Wu, W. B. Strong invariance principles for dependent random variables. *The annals of probability* (2007).

51. Fan, J. *Local polynomial modelling and its applications: monographs on statistics and applied probability 66* (Routledge, 2018).

52. Matzkin, R. L. Restrictions of economic theory in nonparametric methods. *Handbook of econometrics* **4,** 2523–2558 (1994).

53. Silvapulle, M. J. & Sen, P. K. *Constrained statistical inference: Inequality, order and shape restrictions* (John Wiley & Sons, 2005).

54. Seijo, E. & Sen, B. Nonparametric least squares estimation of a multivariate convex regression function (2011).

55. Koenker, R. & Ng, P. Inequality constrained quantile regression. *Sankhyā: the indian journal of statistics,* 418–440 (2005).

56. Gallant, A. R. & Souza, G. On the asymptotic normality of Fourier flexible form estimates. *Journal of econometrics* **50,** 329–353 (1991).

57. Andrews, D. W. Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica: journal of the econometric society,* 307–345 (1991).

58. Ryff, J. V. Measure preserving transformations and rearrangements. *Journal of mathematical analysis and applications* **31,** 449–458 (1970).

59. Korenovskii, A. *Mean oscillations and equimeasurable rearrangements of functions* (Springer, 2007).

60. Chernozhukov, V., Fernandez-Val, I. & Galichon, A. Improving point and interval estimators of monotone functions by rearrangement. *Biometrika* **96,** 559–575 (2009).

61. Crouzeix, J. P., Eatwell, M. J. & Newman, P. in *The new palgrave dictionary of economics* (Palgrave Macmillan UK, 1987).

62. Guerraggio, A. & Molho, E. The origins of quasi-concavity: a development between mathematics and economics. *Historia mathematica* **31,** 62–75 (2004).

63. Chen, X., Chernozhukov, V., Fernández-Val, I., Kostyshak, S. & Luo, Y. Shape-enforcing operators for generic point and interval estimators of functions. *The journal of machine learning research* **22,** 10034–10075 (2021).

64. Lorentz, G. G. An inequality for rearrangements. *The american mathematical monthly* **60,** 176–179 (1953).

65. Chang, C.-S. & Yao, D. D. Rearrangement, majorization and stochastic scheduling. *Mathematics of operations research* **18,** 658–684 (1993).

66. Van der Vaart, A. W. *Asymptotic statistics* (Cambridge university press, 2000).

67. Stein, E. M. & Shakarchi, R. *Real analysis: measure theory, integration, and Hilbert spaces* (Princeton University Press, 2009).

68. Hart, J. *Nonparametric smoothing and lack-of-fit tests* (Springer Science & Business Media, 2013).

69. González-Manteiga, W. & Crujeiras, R. M. An updated review of goodness-of-fit tests for regression models. *Test* **22,** 361–411 (2013).

70. Hooker, G. Forcing function diagnostics for nonlinear dynamics. *Biometrics* **65,** 928–936 (2009).

71. Hooker, G. & Ellner, S. P. Goodness of fit in nonlinear dynamics: misspecified rates or misspecified states? *The annals of applied statistics* **9,** 754–776 (2015).

72. Liu, R., Fang, Y. & Zhu, L. Model checking for parametric ordinary differential equations systems. *Statistica sinica* **33,** 1853–1878 (2023).

73. Liu, R. & Zhu, L. Specification testing for ordinary differential equation models with fixed design and applications to covid-19 epidemic models. *Computational statistics & data analysis* **180,** 107616 (2023).

74. Kozek, A. S. A nonparametric test of fit of a parametric model. *Journal of multivariate analysis* **37,** 66–75 (1991).

75. Hardle, W. & Mammen, E. Comparing nonparametric versus parametric regression fits. *The annals of statistics,* 1926–1947 (1993).

76. Liu, Z., Stengos, T. & Li, Q. Nonparametric model check based on local polynomial fitting. *Statistics & probability letters* **48,** 327–334 (2000).

77. Li, C.-S. Using local linear kernel smoothers to test the lack of fit of nonlinear regression models. *Statistical methodology* **2,** 267–284 (2005).

78. Gharaibeh, M. M. & Wang, H. A new nonparametric lack-of-fit test of nonlinear regression in presence of heteroscedastic variances. *Communications in statistics-theory and methods* **52,** 7886–7914 (2023).

79. Diebolt, J. & Zuber, J. Goodness-of-fit tests for nonlinear heteroscedastic regression models. *Statistics & probability letters* **42,** 53–60 (1999).

80. Hjort, N. L., McKeague, I. W. & Van Keilegom, I. Extending the scope of empirical likelihood. *The annals of statistics* (2009).

81. Heuchenne, C. & Van Keilegom, I. Goodness-of-fit tests for the error distribution in nonparametric regression. *Computational statistics & data analysis* **54,** 1942–1951 (2010).

82. Sapienza, F. *et al.* Differentiable programming for differential equations: a review. *Arxiv preprint arxiv:2406.09699* (2024).

83. Dattner, I. Differential equations in data analysis. *Wiley interdisciplinary reviews: computational statistics* **13,** e1534 (2021).

84. Brunel, N. J. & Clairon, Q. A tracking approach to parameter estimation in linear ordinary differential equations. *Electronic journal of statistics* **9,** 2903–2949 (2015).

85. Clairon, Q. A regularization method for the parameter estimation problem in ordinary differential equations via discrete optimal control theory. *Journal of statistical planning and inference* **210,** 1–19 (2021).

86. Dattner, I., Gugushvili, S. & Laskorunskyi, O. Model selection for ordinary differential equations: a statistical testing approach. *Biometrical journal* **66,** e70013 (2024).

87. Fan, Y. & Li, Q. Consistent model specification tests: kernel-based tests versus bierens' icm tests. *Econometric theory* **16,** 1016–1041 (2000).

88. Ingster, Y. I. On the minimax nonparametric detection of signals in white gaussian noise. *Problemy peredachi informatsii* **18,** 61–73 (1982).

89. Horowitz, J. L. & Spokoiny, V. G. An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69,** 599–631 (2001).

90. Gugushvili, S. & Klaassen, C. A. J. $\sqrt{n}$-consistent parameter estimation for systems of ordinary differential equations: bypassing numerical integration via smoothing. *Bernoulli* **18,** 1061–1098 (2012).

91. Chen, S., Shojaie, A. & Witten, D. M. Network reconstruction from high-dimensional ordinary differential equations. *Journal of the american statistical association* **112,** 1697–1707 (2017).

92. Ramsay, J. & Hooker, G. *Dynamic data analysis* (Springer, 2017).

93. Dattner, I. & Klaassen, C. A. Optimal rate of direct estimators in systems of ordinary differential equations linear in functions of the parameters. *Electronic journal of statistics* (2015).

94. McGoff, K., Mukherjee, S., Pillai, N., *et al.* Statistical inference for dynamical systems: a review. *Statistics surveys* **9,** 209–252 (2015).

95. Amemiya, T. *Advanced econometrics* (Harvard university press, 1985).

96. Yatchew, A. *Some tests of nonparametric regression models* in *Dynamic econometric modelling, proceedings of the third international symposium in economic theory and econometrics, cambridge university press, cambridge* (1988), 121–135.

97. Lopez, O. & Patilea, V. Nonparametric lack-of-fit tests for parametric mean-regression models with censored data. *Journal of multivariate analysis* **100,** 210–230 (2009).

98. Zhao, Y., Zou, C. & Wang, Z. An adaptive lack of fit test for big data. *Statistical theory and related fields* **1,** 59–68 (2017).

99. Audoly, S., Bellu, G., D'Angio, L., Saccomani, M. P. & Cobelli, C. Global identifiability of nonlinear models of biological systems. *Biomedical engineering, ieee transactions on* **48,** 55–65 (2001).

100. Quaiser, T. & Mönnigmann, M. Systematic identifiability testing for unambiguous mechanistic modeling–application to jak-stat, map kinase, and nf-κb signaling pathway models. *Bmc systems biology* **3,** 50 (2009).

101. Ingster, Y. I. Asymptotically minimax hypothesis testing for nonparametric alternatives. I, II, III. *Math. methods statist* **2,** 85–114 (1993).

102. Härdle, W., Sperlich, S. & Spokoiny, V. G. *Component analysis for additive models* tech. rep. (SFB 373 Discussion Paper, 1997).

103. Khanday, M., Rafiq, A. & Nazir, K. Mathematical models for drug diffusion through the compartments of blood and tissue medium. *Alexandria journal of medicine* **53,** 245–249 (2017).

104. Rice, J. Bandwidth choice for nonparametric regression. *The annals of statistics,* 1215–1230 (1984).

105. Dattner, I., Miller, E., Petrenko, M., Kadouri, D. E., Jurkevitch, E. & Huppert, A. Modelling and parameter inference of predator–prey dynamics in heterogeneous environments using the direct integral approach. *Journal of the royal society interface* **14,** 20160525 (2017).

106. Nestorov, I. Whole body pharmacokinetic models. *Clinical pharmacokinetics* **42,** 883–908 (2003).

107. Siepmann, J. & Siepmann, F. Mathematical modeling of drug dissolution. *International journal of pharmaceutics* **453,** 12–24 (2013).

108. Holz, M. & Fahr, A. Compartment modeling. *Advanced drug delivery reviews* **48,** 249–264 (2001).

109. Feizabadi, M. S., Volk, C. & Hirschbeck, S. A two-compartment model interacting with dynamic drugs. *Applied mathematics letters* **22,** 1205–1209 (2009).

110. Welham, S. J., Gezan, S. A., Clark, S. J. & Mead, A. *Statistical methods in biology: design and analysis of experiments and regression* (CRC press, 2014).

111. Bélanger, G., Walsh, J. R., Richards, J. E., Milburn, P. H. & Ziadi, N. Comparison of three statistical models describing potato yield response to nitrogen fertilizer. *Agronomy journal* **92,** 902–908 (2000).

112. Afzal, S., Islam, M. & Obaid-Ur-Rehman. Application of mitscherlich–bray equation for fertilizer use on groundnut. *Communications in soil science and plant analysis* **45,** 1636–1645 (2014).

113. Gause, G. Experimental demonstration of volterra's periodic oscillations in the numbers of animals. *Journal of experimental biology* **12,** 44–48 (1935).

114. Dedrick, S., Warrier, V., Lemon, K. P. & Momeni, B. When does a lotka-volterra model represent microbial interactions? insights from in vitro nasal bacterial communities. *Msystems* **8,** e00757–22 (2023).

115. Summers, J. K. & Kreft, J.-U. The role of mathematical modelling in understanding prokaryotic predation. *Frontiers in microbiology* **13,** 1037407 (2022).

116. Davis, J. D., Olivença, D. V., Brown, S. P. & Voit, E. O. Methods of quantifying interactions among populations using lotka-volterra models. *Frontiers in systems biology* **2,** 1021897 (2022).