9-11-2017

Structure Learning and Break Detection in High-Frequency Data

Yaohua Zhang
University of Connecticut - Storrs, zyh.yaohua@gmail.com

STRUCTURE LEARNING AND BREAK DETECTION IN HIGH-FREQUENCY DATA

Yaohua Zhang, Ph.D.
University of Connecticut, 2017

ABSTRACT

The accurate learning of the underlying structure in high-frequency data has become critical in the analysis of time series for capturing valuable information that facilitates decision-making. The time series data in finance often is large, dynamic, heterogeneous and even structural unstable. Each aspect of these characteristics will add a degree of difficulty in efficient analysis. The goal of this dissertation is to discover the latent structure of dynamic high-frequency data that may have structural breaks, from both univariate and network perspective. We focus our analysis on durations between user-defined events in transaction-by-transaction stock prices from the Trade and Quotes (TAQ) data base at Wharton Research Data Services (WRDS). Our proposed approach can be easily adapted to other models.

The dissertation has three main contributions. First, we propose a fast and accurate distribution-free approach using penalized martingale estimating functions on logarithmic autoregressive conditional duration (Log ACD) models. We discuss three approaches for parameter estimation. Our approach employs effective starting values from an approximating time series model and provides investigators accurate fits and predictions that can assist in trading decisions. Second, we propose a sequential monitoring scheme to detect structural breaks in the estimated parameters of a univariate piecewise Log

ACD model. Based on martingale estimating function, this scheme does not require any distributional assumption. This monitoring scheme can detect structural breaks and choose model orders at the same time. Assuming data is given, we compare the performance of our scheme with that of a state-of-the-art offline scheme via simulation studies. Third, we propose a framework for detecting structural breaks in dynamic networks of a large number of stocks. In particular, we discover unobserved dynamic network structure from nodal observations governed by both the latent network and time. Our empirical analysis on the 30 most liquid stocks in S&P100 is an exploratory study. Such an analysis would be useful to economists studying the structural breaks in financial networks.

STRUCTURE LEARNING AND BREAK DETECTION IN HIGH-FREQUENCY DATA

Yaohua Zhang

B.S., Applied Mathematics, Henan Normal University, Henan, China, 2012

M.S., Statistics, University of Connecticut, CT, USA, 2014

A Dissertation
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
at the
University of Connecticut

2017

APPROVAL PAGE

Doctor of Philosophy Dissertation

STRUCTURE LEARNING AND BREAK DETECTION IN HIGH-FREQUENCY DATA

Presented by

Yaohua Zhang, B.S. Applied Mathematics, M.S. Statistics

Major Advisor _____

Nalini Ravishanker

Associate Advisor _____

Jun Yan

Associate Advisor _____

Kun Chen

University of Connecticut

2017

Acknowledgements

When I started in this doctoral program, I supposed that academic work would be a largely solitary pursuit. This assumption turned out to be false. I could not have done this without the support of the great people who helped me along the way.

Firstly, I would like to express my sincere gratitude to my advisor Professor Nalini Ravishanker for your continuous support of my graduate study, for your patience, motivation, and immense knowledge. Thank you for teaching me how to be a researcher. It would be impossible to list everything I have learned from you across five year, as you taught me everything I know about this field. I could not have imagined having a better advisor and mentor for my Ph.D. study. It has been a privilege to work with you, and I look forward to learning from you and working with you in the future.

Besides my advisor, I would like to thank the rest of my thesis committees: Professor Jun Yan and Professor Kun Chen, for all their invaluable contribution. Their comments and guidance helped me a great deal in the writing of this work and enhanced my dissertation in various ways. I have always admired your ability to see substantial issues in research.

I would like to offer my special thanks to all faculty members in the Department of Statistics at University of Connecticut. Especially, I want to express my sincere appreciation to Professor Zhiyi Chi and Dr. Naitee Ting. Our frank discussions and your advice encouraged me throughout my doctoral work. Your experience is of great help and influence to my career development. I also want to thank Tracy Burke and Megan Petsa for the immense work that they put behind my five years of study.

My sincere thanks also go to Professor Jian Zou for providing me the financial data and great comments in my research. I would also like to thank the faculty and graduate students in the University of Connecticut. I want to thank Chongliang Luo, Daoyuan

Shi, Gao Niu, Gregory Vaughan, Qian Meng, Yan Zhuang, Yeongjin Gwon. It has been wonderful to spend five years among such intelligent and beautiful people.

I would like to thank my family, all of who have supported me in various ways throughout my doctoral work. I am grateful to my father Youqing Zhang, my mother Baodi Li, and my brother Xinghua Zhang for all their love and understanding in my endeavors.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Structural Learning in High-Frequency Data

Recent years have seen a surge in the volume and size of the data being processed and analyzed and the corresponding availability of big data statistical techniques. Such a "big data movement" is driven by the fact that massive amounts of very high dimensional and/or high-frequency structured or unstructured data are continuously produced and stored in an inexpensive way. Many of these multi-dimensional data may also have an underlying network structure that may evolve dynamically.

Accurate learning of the underlying structure has become an important issue in the analysis of time series for capturing valuable information that facilitates decision-making. It is often the case that time series data is large, dynamic, heterogeneous, and even unobservable. Each aspect of these characteristics will add a degree to the difficulty of carrying out efficient analysis.

One of the fundamental questions in this dissertation is how can we discover the latent structure which is dynamic and may have structural breaks. The problem can be formulated in the following way: We denote the observed time series by $\{\boldsymbol{Y_t}\}, t =$

$1, 2, \cdots, n$, indexed by time. Here $\boldsymbol{Y_t}$'s can be univariate, or multivariate. Our first goal is to uncover the underlying structure. The second goal is to detect where observations undergo changes in the structure viz, the structural break points. For simplicity, suppose there is only one break point. The sequence is identically distributed as F_0 until a time τ the distribution changes abruptly to F_1 :

$$\boldsymbol{Y_t} \sim F_0, \quad t = 1, \cdots, \tau - 1$$

$$\boldsymbol{Y_t} \sim F_1, \quad t = \tau, \tau + 1, \cdots, n$$

Here F_0 and F_1 are two probability measures that differ on a set of nonzero measure.

In this dissertation, we focus our analysis on durations between user-defined events in transaction-by-transaction stock prices from the Trade and Quotes (TAQ) data base at Wharton Research Data Services (WRDS). Our proposed approaches can be easily adapted to other models or data sets.

1.2 Thesis Overview

In Chapter 2 Section 2.1, we introduce the background of modeling high-frequency time series and give definition of a financial duration. Section 2.2 describes the financial transactions database and shows how we obtain raw inter-event durations. We discuss how these raw durations may be subject to diurnal effect due to trading behavior of

stocks, and how this behavior may vary between stocks depending on characteristics such as liquidity. We present plots to show the patterns in the diurnally adjusted durations process for each day for four different stocks. Section 2.3 presents a brief review of statistical models that we fit to the diurnally adjusted durations and makes an argument for using an approach for estimation and prediction that relies on minimal distributional assumptions.

In Chapter 3, our objective is to describe analysis of intra-day financial durations and provide various numerical and visual aids that would serve as a useful guide for practitioners. In Section 3.2, we discuss a novel penalized estimation function (EF) framework for the class of Log Autoregressive Conditional Duration (Log ACD) models for the adjusted durations. In Section 3.2.2, we implement the penalized approach via a smoothly clipped absolute deviation (SCAD) penalty under a local linear approximation, and describe the details of fitting and prediction. Section 3.2.3 describes three computational approaches for parameter estimation, solution of a system of nonlinear equations, and two online recursive approaches. Section 3.2.4 discusses an efficient approach to obtain initial values from simple approximating time series models to start the nonlinear equations solver or the online recursions. A simulation study in Section 3.3 shows the relative accuracies of the three approaches. In Section 3.4, we return to our primary objective of modeling long sequences of intra-day adjusted financial durations. We discuss results from fitting Log ACD(p, q) models to data from four stocks. We use durations from the trading days in the first three weeks of June 2013 as training data,

holding out data from the last week in June for predictive cross-validation. For each day in the training set, we allow the data to select the order of the Log ACD(p, q) model. Section 3.5 provides a discussion and summary.

We start to discuss structural break detection in univariate time series of financial durations in Chapter 4. In Section 4.1, we give a detailed description of the structural break detection problem in univariate time series. We introduce the definition of a piecewise Log ACD model and illustrate different kinds of structural breaks via visual plots. In Section 4.2, we review the lines of research focusing on structural break detection using offline methods. We give detailed review for two state-of-the-art offline methods in structural break detection, Automatic Piecewise Autoregressive Modeling (Auto-PARM) and Group LASSO for autoregressive (AR) Models in Section 4.2.1 and 4.2.2 respectively. Both methods focus on structural break detection in piecewise autoregressive models. In the following, we outline our monitoring procedure in two ways. In Section 4.3, we assume that all data points are available. We describe an automatic algorithm that helps investigators to identify the locations of observed structural breaks. Extensive simulation studies in Section 4.4 show the relative accuracies of our retrospective procedures, and a comparison between penalized EF and Group LASSO approaches is given. In Section 4.5, we return to our primary objective to detect structural break detection in univariate time series of durations between events in intra-day high-frequency transaction level financial data. Another lines of research is sequential approaches. Section 4.6 gives an introduction to the background of sequential test of structural stability.

In Section 4.7, we outline the ideas of our monitoring procedure. Simulation studies in Section 4.8 show that our online monitoring procedure can control type I errors as well as maintain a relatively high power for detecting structural breaks. Real data analysis and a brief summary are given in Section 4.9 and Section 4.10 respectively.

In Chapter 5, we consider structural break detection in dynamic networks. In Section 5.1, we introduce the background of dynamic networks and illustrate different kinds of structural breaks in networks. In Section 5.2, we give some necessary preliminaries to networks. In Section 5.3, we provide detailed review for two state-of-the-art methods that focus on structural learning in time-varying networks. We outline a framework for detecting structural breaks in dynamic networks of dependent time series in Section 5.4. In particular, we discover unobserved dynamic network structure from nodal observations that are governed by both the latent structure and time. In Section 5.5, we apply our proposed method to 30 the most liquid stocks in S&P 100.

Chapter 2

Financial Durations

2.1 Introduction

Modeling high-frequency time series data is becoming increasingly relevant in multiple fields due to increased availability of long, complex data stemming from technological advancements in data collection and data storage. In the medical field, devices such as the electrocardiogram (ECG) provide data in real time at frequencies exceeding 100 cycles per second (Hejjel and Roth, 2004). In transportation analysis, data on traffic volume and congestion is collected every 90 seconds (Vlahogianni et al., 2011). Financial databases, which earlier only recorded daily data on the opening or closing characteristics of financial markets, now increasingly contain real-time information on characteristics such as price and volume of every transaction.

Transaction-by-transaction financial data fit the definition (Yan and Zivot, 2003) of high-frequency data where observations such as stock prices or volumes are obtained at a time scale finer than once per day. Such data are widely available nowadays for a host of different financial instruments on markets at several locations worldwide and at various scales, from individual bids to buy and/or sell, to the full distribution of such

bids. With the availability of high-frequency financial data, there is a natural interest in and an increasing demand for better and faster statistical modeling of the duration dynamics of the assets, including flexible nonparametric or semiparametric methodology with minimal assumptions, thereby posing a set of new challenges for researchers and practitioners. Typically, high-frequency financial data arrive over irregularly spaced time intervals, so that the time duration between consecutive data points is not uniform; data may arrive rapidly, separated by short durations, or arrive slowly, with longer durations between arrival times. Empirical evidence shows for instance, that stock market activity tends to peak around market opening and closing times (Yan and Zivot, 2003), and exhibit varying patterns of intra-day and intra-week behavior. Recognizing that traditional discrete-time models which bin the data into equally spaced time intervals are inadequate, Engle and Russell (1998) proposed a nonlinear model for irregularly spaced inter-event durations, called the Autoregressive Conditional Duration (ACD) model.

In financial applications, an event may be defined as a single transaction, as a price change exceeding a certain amount, or as some other characteristic of interest. Engle and Russell (1998) argued that clusters of high activity versus low activity, indicated by short durations versus long durations between events, can reveal very useful information about the market microstructure, and may be predictable. By modeling the time taken by the market until stock prices change beyond or below a certain threshold, a financial trader can gain knowledge which influences the speed with which she/he places quotes. A duration until a price reaction can potentially be understood as a trader's delayed

reaction or response to an informative event. Easley and O'hara (1992) pointed out that while traders may generally have little or no knowledge on the occurrence of an informative event, a cluster of short durations between transactions could often be a data based indicator of some new information in the market. The time between subsequent stock price changes of a stated magnitude can often provide a structural interpretation in terms of the rate of change by which information is released and the rate at which the market incorporates this information into prices. Formally, durations are defined as the time between two consecutive events. Let t_i be the time of occurrence for the i^{th} event. The i^{th} duration is a nonnegative random variable defined as the time interval between the events occurring at times t_i and t_{i-1}:

$$D_i = t_i - t_{i-1} \tag{2.1}$$

Several classes of statistical models for durations have been proposed in the literature since the ACD(p, q) models of Engle and Russell (1998) and the more flexible Log ACD models (Bauwens and Giot, 2000). These models are described in Section 2.3. In high-frequency financial data analysis, one can model financial durations based on all transactions available (Hafner, 2005), or model via a point process defined by a subset of the transaction arrival times with specific characteristics(Bacry and Muzy, 2014; Bauwens and Hautsch, 2009).

2.2 Durations Between Events in Transaction Level Stock Prices

Modeling inter-event durations is of significant interest in financial applications as mentioned in Section 2.1. A duration is defined as the time interval between successive transactions in financial markets. A trade duration is one of the most popular measures in financial econometrics for quantifying trading speed and intensity. It is closely related to the liquidity of the underlying asset. Successful statistical modeling and inference of the duration process will give insight into the market's buyer and seller trading activity patterns which is also a topic of interest in financial microstructure theory. The concept of stock liquidity refers to financial transaction frequency and is closely related to a measure of trading activity intensity, relation among trades, trading volume and stock price dynamics. Kyle (1985) proposed three market liquidity measures: tightness, depth, and resiliency, which evaluate bid-ask spread and transaction costs for traders, the market's ability to absorb and execute large orders with minimal price impact, and the speed for the prices to recover from a fluctuation. In this thesis, we have identified some new patterns between stock liquidity and durations based on ultra high-frequency financial data. Our findings could shed light on how practitioners design their trading strategies according to the stability and liquidity of the different asset classes in their portfolio.

High-frequency data were obtained from the Trade and Quotes (TAQ) database at Wharton Research Data Services (WRDS) from the Wharton School at the University of

Pennsylvania. Since the high-frequency data possess unique features such as microstructure noise, we first processed the data using the same method as in Zou, An, and Yan (2015). We considered the prices from 9:30am to 4:00pm each day in June 2013, and took the average price if the stocks were traded at different exchange centers at the same time. We retained the observations if the price was greater than zero. Only regular trade records were used, i.e., we deleted observations if the trade correction indicator was not zero. In terms of the conditions of sale, we deleted observations with special trade sale conditions, i.e., O, Z, B, T, L, G, W, J and K, since these records may be abnormal and cause inconsistency issues in the subsequent data analysis. Here, "O" stands for an opening trade that occurs in sequence but is reported to the tape at a later time, "Z" represents sold out of sequence, "B" represents average price trade, "T" represents extended hours trade, etc.

In Chapter 3 and 4, we considered transaction-by-transaction stock price data from the trading days in June 2013 of four assets: Bank of America (BAC), General Electric (GE), IBM (IBM) and 3M (MMM). The observations are tick-by-tick transactions for stocks traded in leading stock exchanges including the New York Stock Exchange (NYSE) and NASDAQ. GE and BAC are liquid stocks with high trading volumes (100,000 to 200,000 transactions per day on average) and a relatively low price spread of about \$0.5 per day on average. This is considered to be stable behavior. IBM is also a liquid stock with a smaller number of average transactions per day (around 20,000) and an absolute price spread between \$3 and \$4. MMM has variable numbers of transactions

(between 4000 and 15,000 per day) and an absolute price spread of about $1.50.

2.2.1 Transactions to Raw Durations

As mentioned in Section 2.1, the raw durations were computed from the raw transactions data as the elapsed time between two successive occurrences of an event using (2.1), i,e., $D_i = t_i - t_{i-1}$. We say that an event occurs when the change in price between two successive transactions exceeds a certain threshold δ. Suppose the open price is P_0, say. The first event occurs when the stock price first changes to $P_1 > P_0+\delta$ or $P_1 < P_0-\delta$; the second event is recorded when the price next changes to $P_2 > P_1+\delta$ or $P_2 < P_1-\delta$, etc. The value of δ for different stocks is determined through a combination of the empirical behavior of the stock in some previous time period and available financial information about the stock behavior. Following an approach that is commonly used by financial professionals, we used data from the previous month (May 2013) to calculate average turnover ratio as average daily volume divided by total shares outstanding. Specifically, δ equals 0.00377, 0.004, 0.00376 and 0.00408 for GE, BAC, IBM and MMM respectively. By taking the differences of successive transaction times and filtering out any zero-valued durations, we obtained a time series of raw durations. In Figure 1, we illustrate how to calculate raw durations from the intra-day transactions data during the first three seconds of the trading day on June 3, 2013 for BAC. The x-axis shows the time stamps of each transaction (in milliseconds) while the y-axis shows the corresponding stock price. Each vertical dashed line represents the time at which an event occurs, i.e., there is a

price change exceeding δ ($\delta = 0.004$ for BAC). The raw durations were computed as the elapsed time between two price changes. As expected, the lengths of the duration time series vary for different stocks in proportion to the liquidity of the stock.

It is reasonable to ask why one would not simply compute raw durations as the gap between each transaction, instead of employing a user defined δ. First, defining an event via a δ seems to be a meaningful financial practice. By modeling the time taken by the market until stock prices change beyond or below a certain threshold, a financial trader can gain knowledge which influences the speed with which she/he places quotes. A duration until a price reaction can potentially be understood as a trader's delayed reaction or response to an informative event. Second, computing raw durations as the gap between each transaction is likely to lead to a large number of zero or near-zero durations that must then be excluded before the statistical modeling. If we model the duration between subsequent transactions, we are reckoning every single movement in the price as a price change (this change is zero for most cases). Peculiar portfolio adjustment from individual bank decisions should not be considered as a movement in the fundamental price at which individuals are exchanging. Choosing a threshold δ that is greater than zero can help us capture movements in the price at which transactions are occurring. Setting $\delta > 0$ has been widely used in the literature. Engle and Russell (1998) set δ as half of the largest observed spread. Bauwens and Giot (2000) chose \$0.125 as a threshold. Yuan et al. (2015) summarizes the threshold that people usually use when modeling volume duration (change in volume). They found that people use the

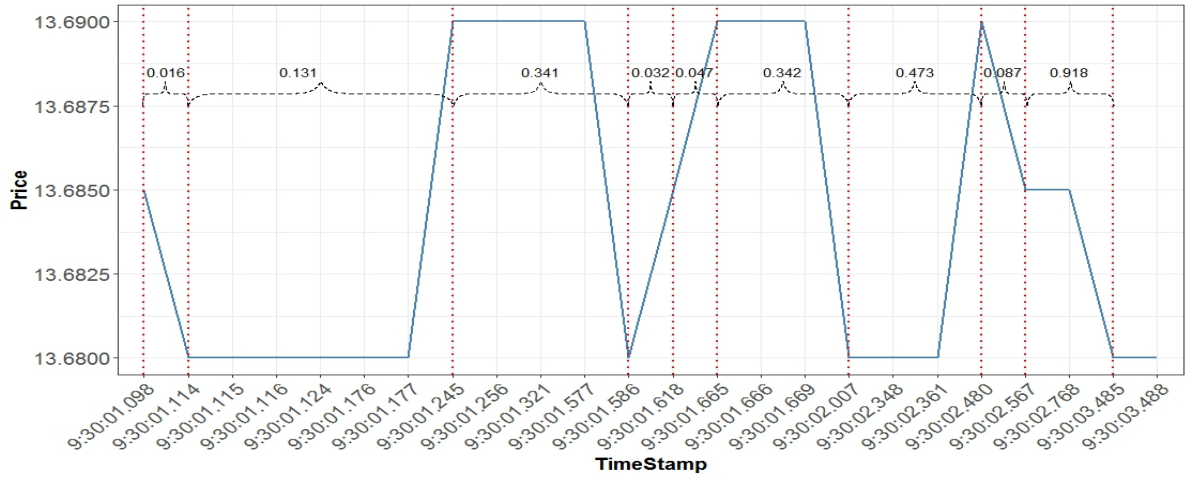

Figure 1: Transactions to raw durations for BAC during the first 3 seconds on June 3, 2013

multiple of the mean volume per transaction as the threshold. In this chapter, we used data from the previous month (May 2013) to calculate δ. But in practice, a financial analyst may define an event in different ways, such as a directional change in price, or a change in volume, etc.

2.2.2 Filtering the Diurnal Effect

Intra-day periodic trading patterns is a stylized feature commonly seen in high-frequency financial data. This refers to the high trading intensity during the opening and closing periods of the trading day, with relatively lower trading activity around noon. The effect thus corresponds to shorter durations in the opening and closing periods of each trading day, and longer durations around noon. This phenomenon is usually referred to as the "diurnal effect", see Engle and Russell (1998) and Tsay (2005) for more details.

Financial professionals expect that intra-day trading volume patterns are likely to vary for different stocks due to differences in their characteristics such as liquidity. They are also expected to exhibit some dissimilarities on different days for the same stock, for reasons including dependence on particular news items (e.g., Federal Open Market Committee (FOMC) minutes, jobs reports), shares held by institutional investors and average holding period. For a given stock, with the exception of a few days in a year (earnings announcements and other major news, more often than not known well in advance) intra-day trading volume patterns are likely to be similar from one day to another. In particular, one would expect to observe persistence in diurnal patterns, level, and temporal stochastic patterns from day to day. We investigate these properties in this section using data on four different stocks over trading days in a month.

The diurnal effect is illustrated in Figure 2 for BAC, again using data from June 3, 2013. The raw durations in Figure 2(a) reveal the typical reverse U-shaped trading intensity behavior (diurnal effect). This effect is more clearly seen in the solid line in Figure 2(b) which depicts the smoothed mean raw duration computed over a 15-minute window.

We used an approach proposed by Tsay (2005) to estimate this diurnal pattern and obtain diurnally adjusted durations. Specifically, we defined two functions $O(t_i)$ and $C(t_i)$ as

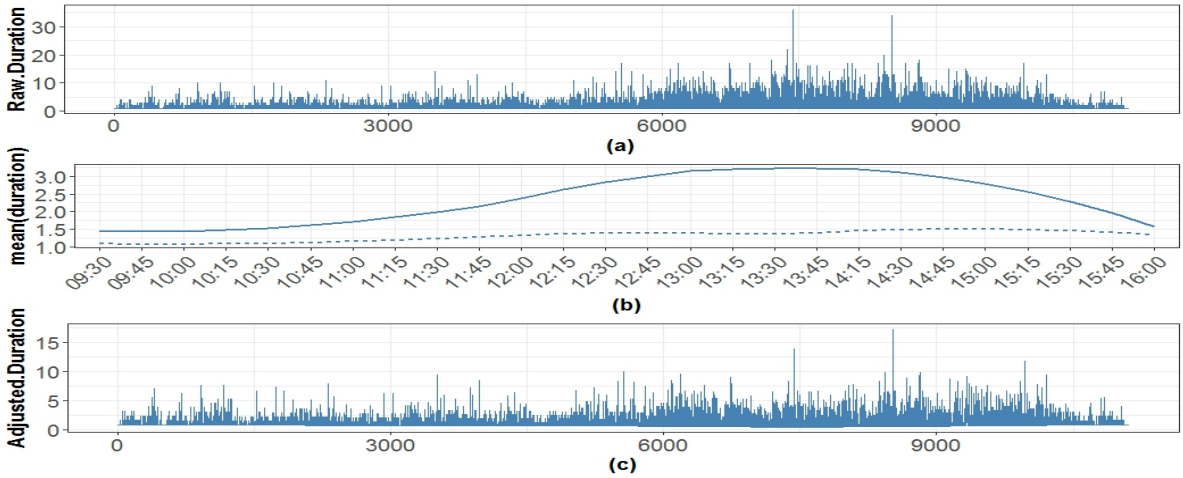

Figure 2: Information for BAC on June 3, 2013. (a) Raw durations; (b) Smoothed mean of raw durations (solid line) and smoothed mean of adjusted durations (dashed line) over each 15 minute window; (c) Durations adjusted for diurnal effect

$$O(t_i) = \begin{cases} t_i - 34200 & \text{if } t_i < 46800 \\ 0 & \text{otherwise} \end{cases}$$

and

$$C(t_i) = \begin{cases} 57600 - t_i & \text{if } t_i \geq 46800 \\ 0 & \text{otherwise} \end{cases}$$

where t_i is the time in seconds when the ith transaction occurs, and 34200, 46800 and 57600 correspond to 9:30AM, 1:00PM and 4:00PM respectively. Raw durations were then adjusted through a linear regression to yield adjusted durations x_i:

$$\begin{aligned} \log(D_i) &= \gamma_0 + \gamma_1 o(t_i) + \gamma_2 c(t_i) + \zeta_i \\ x_i &= \exp(\log(D_i) - \hat{\gamma}_0 - \hat{\gamma}_1 o(t_i) - \hat{\gamma}_2 c(t_i)), \end{aligned} \tag{2.2}$$

where $o(t_i) = O(t_i)/10000$, $c(t_i) = C(t_i)/10000$ and $\hat{\gamma}_i$ denote the least squares estimates of the coefficients. The plot of adjusted durations for June 3, 2013 for BAC in Figure 2(c) show durations adjusted for the diurnal effect while preserving the stochastic temporal dependence. The reverse U-shaped pattern is clearly absent in the dotted line in Figure 2(b) which shows the smoothed mean diurnally adjusted adjusted duration computed over a 15-minute window; the overall mean is close to 1.0.

Due to day-by-day variations in the data, it is possible that the diurnal effect coefficients in (2.2) are different for different days. These differential effects are illustrated in Figures 3(a) - 6(a) which show histograms of trading intensities for five trading days of the first week for the four stocks. In general, the plots show a U-shaped behavior which corresponds to shorter durations in the opening and closing periods of each trading day, and longer durations around noon. Nevertheless, there are some differences. For instance in Figure 3(a), we observe a typical U-shaped behavior for GE on the first, fourth and fifth day, while on the second and third days, we observe a reverse L-shaped trading behavior which corresponds to fewer transactions (flat pattern) occurring from the beginning of the day until late afternoon followed by more transactions near the end of the day. Similarly, for BAC, we observe a U-shaped behavior on the third and fifth days, while on the first day, a slow decay in the morning gives a skewed U-shaped pattern. Also, we observe a cluster of long durations in the afternoon and around noon for the second and fourth days respectively. For lower liquidity stocks like IBM (Figure 5) and MMM (Figure 6), the typical U-shaped behavior seems to occur less frequently;

in general the number of durations for each day and for each stock is small, so that frequent transactions in any small time period will make a fraction of the day prominent, leading to a non-U-shaped behavior (see the second and third days for IBM and the fourth day for MMM). A reverse L-shaped behavior seems to be common overall in these low liquidity stocks. Figures 3(b) - 6(b) display the diurnally adjusted durations where the time of day effect is largely removed.

Even for a high liquidity stock like BAC, the trading patterns on certain days may differ from the general pattern. Figure 7 shows trading intensities (frequencies of diurnally adjusted durations in 15-minute intervals during the trading day) for BAC for June 10, June 11, June 14, June 17, June 18 and June 19. The left panel shows the histograms while the right panel shows the corresponding smoothed mean raw durations computed over a 15-minute window. We see for example that on Mon. June 10, the histogram exhibits a deep U-shaped behavior, with higher frequencies at the opening and closing periods relative to smaller frequencies in the middle of the day. By contrast, histograms on Tue. June 11 and Mon. June 17 display a wider U-shape, i.e., a small number of transactions that start early in the morning and end late in the afternoon. The histograms for Fri. June 14 and Tue. June 18 are flatter.The histogram for Wed. June 19 is very different from any U-shaped behavior, and rather exhibits a reverse Z-shape. There do not appear to be any day-of-the-week effect, but the second week in June appears to exhibit the most disparate behavior.

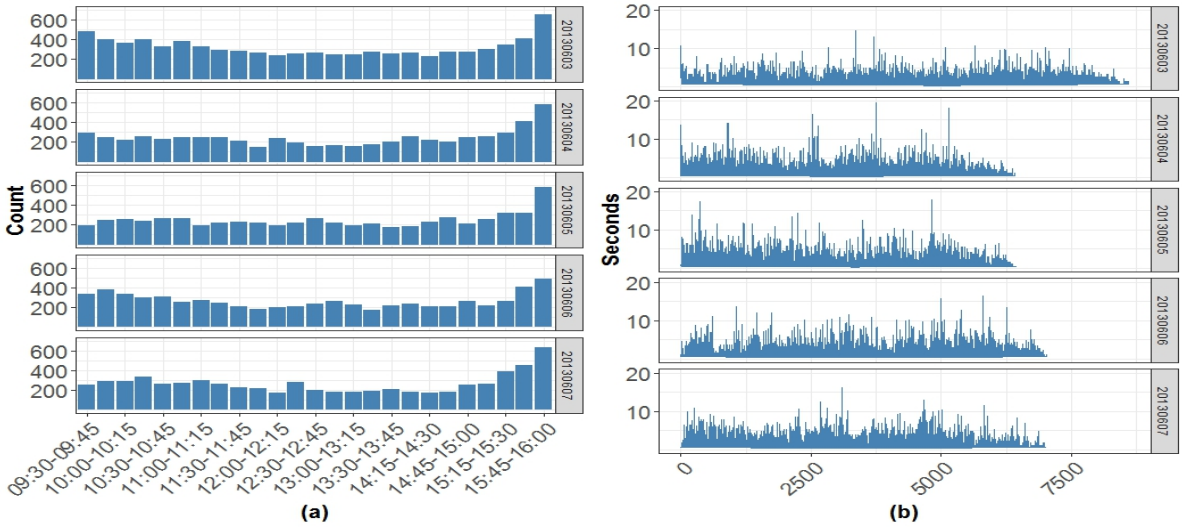

Figure 3: Durations for GE in the first week of June 2013. (a) Histograms of trading intensities over each 15 minute window; (b) Durations adjusted for diurnal effect.

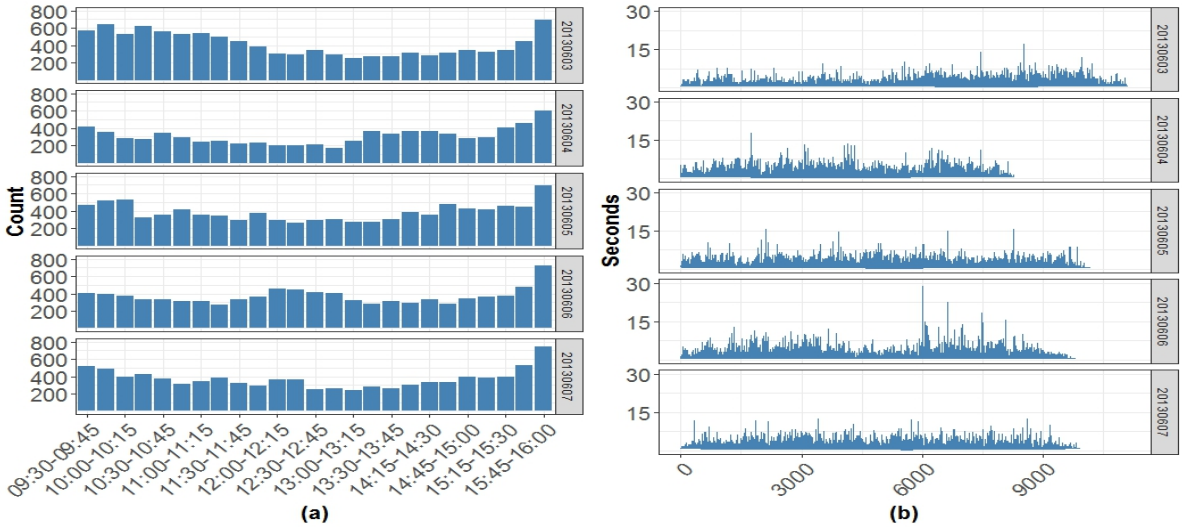

Figure 4: Durations for BAC in the first week of June 2013. (a) Histograms of trading intensities over each 15 minute window; (b) Durations adjusted for diurnal effect.

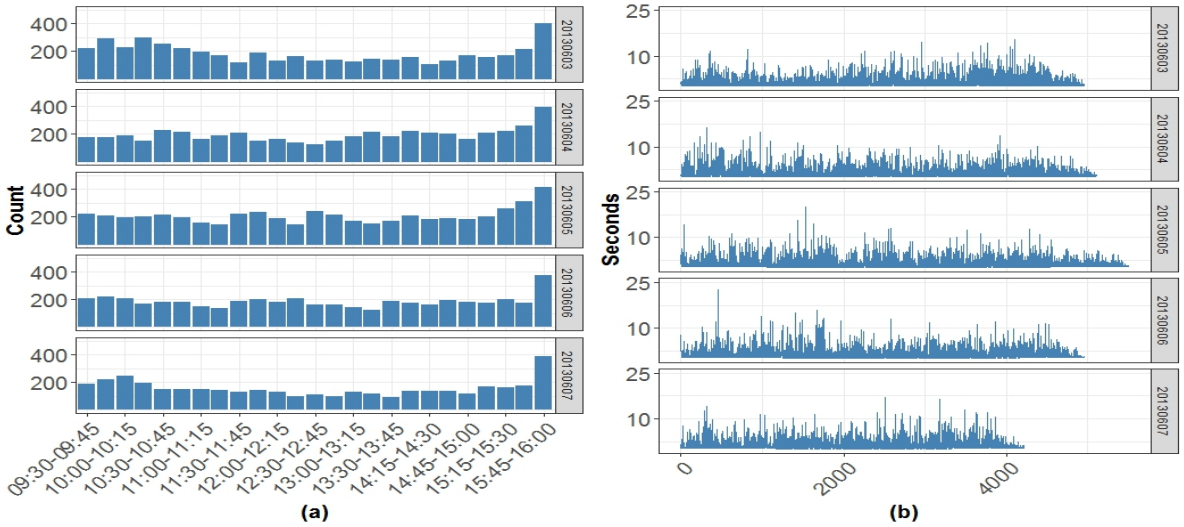

Figure 5: Durations for IBM in the first week of June 2013. (a) Histograms of trading intensities over each 15 minute window; (b) Durations adjusted for diurnal effect.

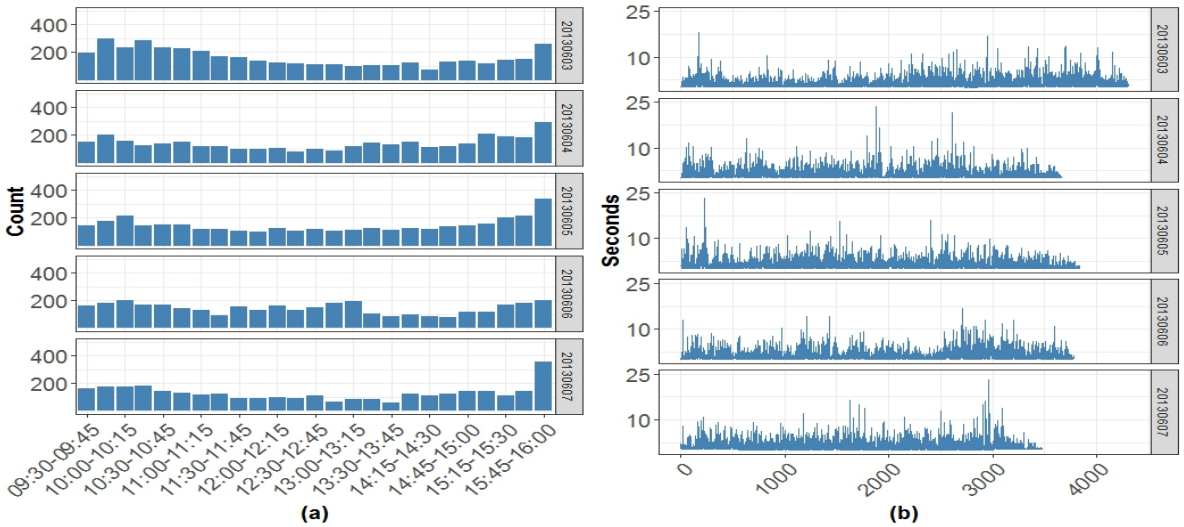

Figure 6: Durations for MMM in the first week of June 2013. (a) Histograms of trading intensities over each 15 minute window; (b) Durations adjusted for diurnal effect.

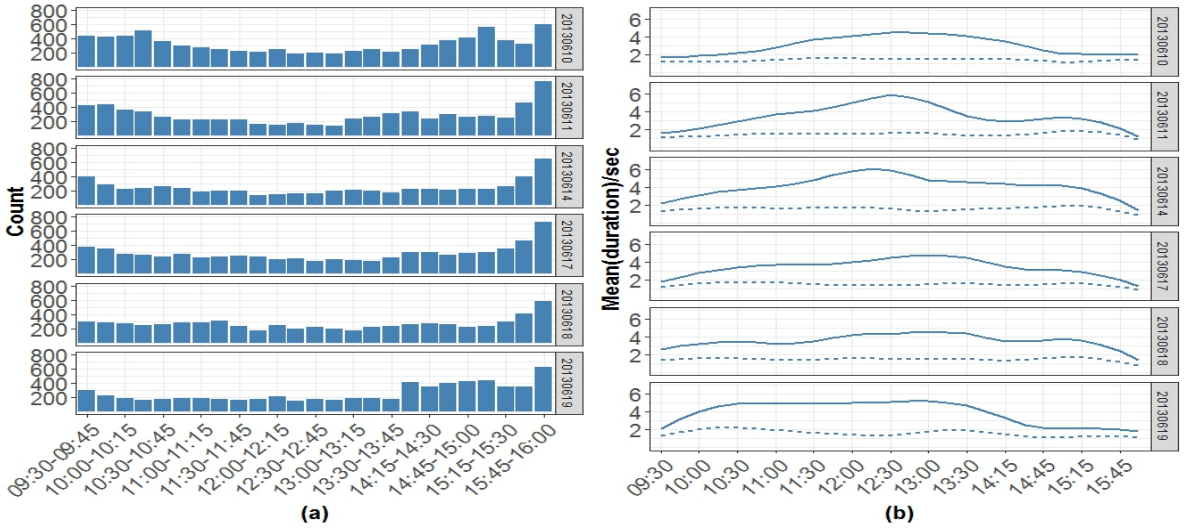

Figure 7: BAC for June 10, June 11, June 14, June 17, June 18, and June 19. (a) Histograms of trading intensities over each 15 minute window; (b) Smoothed means of raw (solid line) and adjusted (dashed line) durations over each 15 minute window.

2.3 A Class of Duration Models

Let x_i, $i = 1, 2, \ldots$, denote a time series of adjusted durations defined as equation (2.2), and let $\mathcal{F}_{i-1}^x = \{x_1, x_2, ..., x_{i-1}\}$ denote the information associated with previous durations. Engle and Russell (1998) introduced the exponential autoregressive conditional duration (EACD) model which assumes that

$$\begin{aligned} x_i &= \psi_i \varepsilon_i \\ \psi_i &= E(x_i|\mathcal{F}_{i-1}^x) = \omega + \sum_{j=1}^{p} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j}, \end{aligned} \tag{2.3}$$

where ε_i are assumed to be independent and identically distributed (i.i.d.) exponential random variables with $E(\varepsilon_i) = 1$. Variations of the EACD(p, q) model include the

Weibull or Gamma ACD(p, q) models by assuming alternate error distributions. Verification of the conditions $\omega > 0$, $\alpha_j \geq 0$ for $j = 1, \ldots, p$, $\beta_j \geq 0$ for $j = 1, \ldots, q$ and $\sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j < 1$ is necessary in order to ensure nonnegativity and weak stationarity of the duration process, and may be tedious in practice, especially for large p or q.

Note that the ACD and GARCH models (Bollerslev, 1986) have similar setups. In fact, the ACD model is commonly viewed as another version of the GARCH model for financial durations rather than for financial returns. Both models are often used to capture market changes induced by a cluster of news or events. A GARCH model could be used for any time series that has intervals of changing variance. Similarly, an ACD model is used to model clusters of short and long durations observed in high-frequency financial data.

The Log ACD(p, q) model (Bauwens and Giot, 2000) relaxes the parameter constraints that ensure nonnegativity and thus provides greater flexibility than the ACD(p, q) model. Let ε_i be i.i.d. nonnegative errors which are independent of $\mathcal{F}_{i-1}^{x}$ and with $E(\varepsilon_i) = \mu_\varepsilon$. The logarithmic ACD, or the Log ACD(p, q) model is defined as

$$x_i = \exp(\Phi_i)\varepsilon_i, \tag{2.4}$$

so that the conditional mean of the ith duration given the history $\mathcal{F}_{i-1}^{x}$ is

$$E(x_i|\mathcal{F}_{i-1}^{x}) = \exp(\Phi_i)\mu_\varepsilon \tag{2.5}$$

Suppose that

$$\begin{aligned} \exp(\psi_i) &= \exp(\Phi_i)\mu_\varepsilon, \text{where,} \\ \psi_i &= \omega + \sum_{j=1}^{p} \alpha_j \log x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} \end{aligned} \tag{2.6}$$

Then, we can rewrite (2.4) as

$$x_i = \frac{\exp(\psi_i)}{\mu_\varepsilon}\varepsilon_i \tag{2.7}$$

The condition $\sum_{j=1}^{\max(p,\ q)}(\alpha_j + \beta_j) < 1$ ensures weak stationarity.

Zhang, Russell, and Tsay (2001) introduce a Threshold ACD models (TACD) to allow a more flexible specification rather than allowing the conditional means to depend linearly on the past information. The TACD model is commonly viewed as the generalization of the threshold-GARCH model (Rabemananjara and Zakoian, 1993; Zakoian, 1994). A series of durations $\{x_i\}$ fit a $J-$regime TACD(p, q) model if

$$\begin{aligned} x_i &= \psi_i \varepsilon_i^{(j)}, \\ \psi_i &= \omega^{(j)} + \sum_{k=1}^{p} \alpha_k^{(j)} x_{i-k} + \sum_{k=1}^{q} \beta_k^{(j)} \psi_{i-k}, \quad \text{if } x_{i-1} \in R_j. \end{aligned} \tag{2.8}$$

where $R_j = [r_{j-1}, r_j], j = 1, 2, ..., J$ are J regimes and $0 = r_0 < r_1 < ... < r_J = \infty$ are the threshold values. $(\omega^{(j)}, \alpha_k^{(j)}, \beta_k^{(j)})$ denote the parameters of ACD models for each regime and should be nonnegative. In addition, the i.i.d. nonnegative error terms $\varepsilon_i^{(j)}$ are regime-specific, independent of past information and with positive expectations.

In this setup, each regime has a different ACD models with different duration persistence, conditional means and error distributions. Therefore a TACD model greatly increases the flexibility when modeling durations compared to the standard ACD model. However, the estimates are given by maximizing the likelihood functions for each pair of threshold values and the choice of threshold values are based on a grid search across possible permutations of duration deciles. When the value of J increases, the model estimation process can be computationally intensive.

Chapter 3

Modeling Financial Durations Using Penalized Estimating Functions

3.1 Motivation

The use of estimating functions (EFs) in statistical inference has a long history, starting with Fisher (1924) who showed that maximum likelihood and minimum chi-squared methods are asymptotically equivalent by comparing the first order conditions of the two estimation procedures, i.e., analyzing properties of estimators by focusing on the corresponding EFs rather than on the objective functions or estimators themselves. Godambe (1960) and Durbin (1960) gave a fundamental optimality result for EFs for the scalar parameter case, while Godambe (1985) studied inference based on the EF approach for discrete-time stochastic processes. Bera et al. (2006) gave an excellent survey on the historical development of this topic. This approach for time series problems had been developed by Hansen and Singleton (1982), Merkouris et al. (2007), Ghahramani and Thavaneswaran (2012), and Liang et al. (2011), among others. Thavaneswaran,

Ravishanker, and Liang (2015) introduced a combined (linear and quadratic) estimation approach for a class of generalized duration models. They discussed the theory and methodology, but did not provide details on specific models, implementation or real data applications.

In this chapter, we introduce a penalized estimating functions (PEF) approach to model financial durations. We assume that observed adjusted durations $\{x_i\}_{i=1}^n$ follow the Log ACD(p,q) model. Let $\boldsymbol{x}_i = (x_1, x_2, \ldots, x_i)'$, $\boldsymbol{\theta} = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)'$ and $g = \max(p,q)$. Let $P = 1 + p + q$. If we assume a known form of the density $f_\varepsilon(.)$ of the errors, the distributions $f(x_i|\boldsymbol{x}_{i-1}, \boldsymbol{\theta})$ can be derived directly, and the maximum likelihood estimates (MLEs) $\widehat{\boldsymbol{\theta}}$ can be obtained by maximizing the conditional likelihood function Tsay (2009):

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{x}_n) = \prod_{i=g+1}^{n} f(x_i|\boldsymbol{x}_{i-1}, \boldsymbol{\theta})$$

However, $f_\varepsilon(.)$ is usually unknown in practice, and this imposes restrictive and unrealistic assumptions in empirical model building in practice. We describe an alternate distribution-free approach based on martingale estimating functions (EFs) which does not require any distributional assumption (such as Exponential, Weibull, etc.) on the errors. A general framework for parameter estimation in the class of generalized duration models based on combined martingale EFs was discussed in Thavaneswaran et al. (2015). Our distribution-free approach is similar to the Generalized Method of Moments (Hansen, 1982), only requiring specification of the first four conditional (on the history)

moments of the duration process, and is described in Section 3.2. The EF approach relies only on a specification of the first few moments of the random variable at each time conditional on its history, and does not require specification of the form of the conditional probability distribution. It provides a flexible approach for modeling financial durations. We believe that financial analysts would prefer to employ predictive statistical models which are not hinged to distributional assumptions. We introduce a penalized estimating functions approach in order to achieve sparsity in model building and show that these are more informative than non-penalized EFs.

3.2 Penalized Combined Martingale EF for Log ACD Models

3.2.1 Martingale Estimating Function

The martingale estimating function (EF) approach (Godambe, 1985) was introduced for duration processes by Thavaneswaran et al. (2015). The main advantage of EF framework lies in its flexibility with no restrictions on the error distributions of the underlying process. In this chapter, we have incorporated effective estimation under the EF framework that can enable big data applications such as online streaming capability.

Let x_i, $i = 1, \ldots, n$ be the adjusted duration process parametrized by $\boldsymbol{\theta} \in \mathbb{R}^P$, and let $\mathcal{F}_{i-1}^{x}$ denote the information associated with $\boldsymbol{x}_{i-1} = \{x_1, \ldots, x_{i-1}\}$. The combined

EF approach only requires specification of the first four conditional moments of x_i given its history. The estimating function is then constructed via suitable martingale differences that arise naturally from the conditional moments. Suppose the conditional mean, variance, third central moment, and fourth central moment of $\{x_i\}$ given $\mathcal{F}_{i-1}^x$ are respectively $\mu_i(\boldsymbol{\theta})$, $\sigma_i^2(\boldsymbol{\theta})$, $\gamma_i(\boldsymbol{\theta})$, and $\kappa_i(\boldsymbol{\theta})$. Define the following linear and quadratic martingale differences:

$$m_i(\boldsymbol{\theta}) = x_i - \mu_i(\boldsymbol{\theta}) \text{ and } M_i(\boldsymbol{\theta}) = m_i^2(\boldsymbol{\theta}) - \sigma_i^2(\boldsymbol{\theta}) \tag{3.1}$$

Their quadratic variations and covariation are

$$\begin{aligned} \langle m \rangle_i &= E[m_i^2(\boldsymbol{\theta})|\mathcal{F}_{i-1}^x] = \sigma_i^2(\boldsymbol{\theta}), \\ \langle M \rangle_i &= E[m_i^4(\boldsymbol{\theta})|\mathcal{F}_{i-1}^x] - \left(E[m_i^2(\boldsymbol{\theta})|\mathcal{F}_{i-1}^x]\right)^2 = \kappa_i(\boldsymbol{\theta}) - \sigma_i^4(\boldsymbol{\theta}), \\ \langle m, M \rangle_i &= E[m_i^3(\boldsymbol{\theta})|\mathcal{F}_{i-1}^x] = \gamma_i(\boldsymbol{\theta}) \end{aligned} \tag{3.2}$$

Consider the class $\mathcal{M}$ of zero-mean, square integrable P-dimensional combined (i.e., combining the linear and quadratic) martingale EFs,

$$\mathcal{M} = \left\{ \mathbf{g}_C(\boldsymbol{\theta}) : \ \mathbf{g}_C(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[\mathbf{a}_{i-1}(\boldsymbol{\theta}) m_i(\boldsymbol{\theta}) + \mathbf{b}_{i-1}(\boldsymbol{\theta}) M_i(\boldsymbol{\theta}) \right] \right\}, \tag{3.3}$$

where $\mathbf{a}_{i-1}(\boldsymbol{\theta})$ and $\mathbf{b}_{i-1}(\boldsymbol{\theta})$ are matrix-valued functions of $\boldsymbol{\theta}$ and $\boldsymbol{x}_{i-1}$, $1 \leq i \leq n$. We assume that $\mathbf{g}_C(\boldsymbol{\theta})$ are almost surely differentiable with respect to the components of

$\boldsymbol{\theta}$, and are such that for each $n \geq 1$, $E\left(\partial \mathbf{g}_C(\boldsymbol{\theta})/\partial \boldsymbol{\theta}|\ \mathcal{F}_{n-1}^x\right)$ and $E(\mathbf{g}_C(\boldsymbol{\theta})\mathbf{g}_C(\boldsymbol{\theta})'|\ \mathcal{F}_{n-1}^x)$ are nonsingular for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where all expectations are taken with respect to the underlying probability measure $P_{\boldsymbol{\theta}}$. The optimal combined EF based on the linear and quadratic martingale differences in (3.1) is the one which maximizes, in the partial order of nonnegative definite matrices, the Godambe information matrix

$$\begin{aligned}\mathbf{I}_{\mathbf{g}_C}(\boldsymbol{\theta}) = &\left[E\left(\frac{\partial \mathbf{g}_C(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\middle|\ \mathcal{F}_{n-1}^x\right)\right]' \left[E(\mathbf{g}_C(\boldsymbol{\theta})\mathbf{g}_C(\boldsymbol{\theta})'|\ \mathcal{F}_{n-1}^x)\right]^{-1} \\ &\times \left[E\left(\frac{\partial \mathbf{g}_C(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\middle|\ \mathcal{F}_{n-1}^x\right)\right],\end{aligned}$$

and was derived in Thavaneswaran et al. (2015) as

$$g_C^*(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left(\mathbf{a}_{i-1}^*(\boldsymbol{\theta}) m_i(\boldsymbol{\theta}) + \mathbf{b}_{i-1}^*(\boldsymbol{\theta}) M_i(\boldsymbol{\theta})\right), \tag{3.4}$$

where

$$\begin{aligned}\mathbf{a}_{i-1}^*(\boldsymbol{\theta}) = \rho_i^2 \Bigg(&-\frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{1}{\langle m \rangle_i} - \Bigg(E\left[\frac{\partial (m_i^2(\boldsymbol{\theta}) - \sigma_i^2(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}}\middle| \mathcal{F}_{i-1}^x\right] \\ &- \frac{\partial E\left[(m_i^2(\boldsymbol{\theta}) - \sigma_i^2(\boldsymbol{\theta}))\middle| \mathcal{F}_{i-1}^x\right]}{\partial \boldsymbol{\theta}}\Bigg) \eta_i\Bigg),\end{aligned} \tag{3.5}$$

and

$$\mathbf{b}^*_{i-1}(\boldsymbol{\theta}) = \rho_i^2 \left(\frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \eta_i + \left(E\left[\frac{\partial (m_i^2(\boldsymbol{\theta}) - \sigma_i^2(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} \middle| \mathcal{F}^x_{i-1} \right] - \frac{\partial E\left[(m_i^2(\boldsymbol{\theta}) - \sigma_i^2(\boldsymbol{\theta})) \middle| \mathcal{F}^x_{i-1} \right]}{\partial \boldsymbol{\theta}} \right) \frac{1}{\langle M \rangle_i} \right), \tag{3.6}$$

with

$$\rho_i^2 = \left(1 - \frac{\langle m, M \rangle_i^2}{\langle m \rangle_i \langle M \rangle_i} \right)^{-1} \text{ and } \quad \eta_i = \frac{\langle m, M \rangle_i}{\langle m \rangle_i \langle M \rangle_i} \tag{3.7}$$

The optimal Godambe information matrix corresponding to $g^*_C(\boldsymbol{\theta})$ is then

$$\begin{aligned} \mathbf{I}_{\mathbf{g}^*_C}(\boldsymbol{\theta}) &= E(\mathbf{g}^*_C(\boldsymbol{\theta}) \mathbf{g}^*_C(\boldsymbol{\theta})' \mid \mathcal{F}^x_{n-1}) \\ &= \sum_{i=1}^{n} \rho_i^2 \left[\frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \frac{1}{\langle m \rangle_i} + \frac{\partial \sigma_i^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_i^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \frac{1}{\langle M \rangle_i} \right. \\ &\quad \left. - \left(\frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_i^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} + \frac{\partial \sigma_i^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right) \eta_i \right] \end{aligned} \tag{3.8}$$

Lindsay (1985) showed that the asymptotic variance of the resulting optimal estimate $\widehat{\boldsymbol{\theta}}$ is $\mathbf{I}^{-1}_{\mathbf{g}^*_C}(\boldsymbol{\theta})$ evaluated at $\widehat{\boldsymbol{\theta}}$, and if the estimator $\widehat{\boldsymbol{\theta}}$ is obtained from the "most" informative EF, then it will be asymptotically most efficient. General consistency and asymptotic distributional results were discussed by Heyde (2008, chap. 12). The standard errors (SEs) of $\widehat{\theta}_j$, $j = 1, \ldots, P$, are thus obtained as the positive square roots of the diagonals of the inverse optimal information matrix.

3.2.2 Penalized Estimating Functions

We introduce model fitting via penalized estimating functions. While penalized methods have been widely used in regression settings, the literature on using a penalized approach for variable selection in the context of estimating functions is rather sparse; see Wang et al. (2012) and references therein for penalized generalized estimating equations (GEE) in the longitudinal data setup. To our knowledge, this is among the first attempts of this approach for martingale estimating functions.

We include a penalty in (3.4) in order to obtain penalized estimates of $\boldsymbol{\theta}$. Starting with $g_C^*(\boldsymbol{\theta})$ defined in (3.4), we first define an objective function with penalty as

$$g_{C,\lambda}^* := g_C^*(\boldsymbol{\theta}) - np_\lambda'(|\boldsymbol{\theta}|), \tag{3.9}$$

where, for some $a > 2$ and penalty (tuning) parameter λ, $p_\lambda'(|\boldsymbol{\theta}|)$ is the first derivative of the SCAD penalty (Fan and Li, 2001) defined by

$$p_\lambda'(|\boldsymbol{\theta}|) = \lambda\{I(|\boldsymbol{\theta}| \leq \lambda) + \frac{(a\lambda - |\boldsymbol{\theta}|)_+}{(a-1)\lambda} I(|\boldsymbol{\theta}| > \lambda)\} \tag{3.10}$$

It is well known that while $p_\lambda'(|\boldsymbol{\theta}|)$ is differentiable on $(-\infty, 0) \cup (0, \infty)$, it has a singularity at zero. To remedy this, we use its local linear approximation

$$p_\lambda'(|\boldsymbol{\theta}|) \approx \{p_\lambda'(|\boldsymbol{\theta}_0|)/|\boldsymbol{\theta}_0|\}\boldsymbol{\theta}, \tag{3.11}$$

where $\boldsymbol{\theta}_0$ denotes the vector of initial values, see Section 3.4. The SCAD penalty allows for a

continuous solution by not excessively penalizing larger values of components of $\boldsymbol{\theta}$. Since the numerical approximation does not give exactly zero estimates for the penalized coefficients, we set these to zero if they are estimated at magnitude less than a small cut-off point, say 10^{-5}.

As mentioned in Section 3.2.1, The optimal information matrix of $\widehat{\boldsymbol{\theta}}$ for the penalized combined EF is derived as follows

$$\begin{aligned}
\mathbf{I}_{\mathbf{g}^*_{C,\lambda}}(\boldsymbol{\theta}) &= E(\mathbf{g}^*_{C,\lambda}(\boldsymbol{\theta})\mathbf{g}^*_{C,\lambda}(\boldsymbol{\theta})'|\ \mathcal{F}^x_{n-1}) \\
&= \sum_{i=1}^{n} \rho_i^2 \left[\frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \frac{1}{\langle m \rangle_i} + \frac{\partial \sigma_i^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_i^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \frac{1}{\langle M \rangle_i} \right. \\
&\quad \left. - \left(\frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_i^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} + \frac{\partial \sigma_i^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right) \eta_i \right] - \sum_{i=1}^{n} W_i \\
&= \mathbf{I}_{\mathbf{g}^*_C}(\boldsymbol{\theta}) + \mathbf{I}_{\mathbf{g}^*_\lambda}(\boldsymbol{\theta}) \qquad (3.12)
\end{aligned}$$

evaluated at $\widehat{\boldsymbol{\theta}}$, where

$$W_i = \left(\frac{p'_\lambda(|\boldsymbol{\theta}_0|)}{|\boldsymbol{\theta}_0|} \right)^2 \boldsymbol{\theta}\boldsymbol{\theta}'$$

From (3.12) we may compute the asymptotic standard errors of the estimates as the square root of the diagonal elements. The penalized combined EF is therefore more informative than the combined EF.

For the Log ACD(p, q) model, recall that $\boldsymbol{\theta} = (\omega, \boldsymbol{\alpha}', \boldsymbol{\beta}')'$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)'$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)'$, so that $P = p + q + 1$. The conditional mean, variance, third, and fourth

central moments of $\{x_i\}$ given $\mathcal{F}_{i-1}^x$ are respectively

$$\begin{aligned}\mu_i(\boldsymbol{\theta}) &= \exp(\psi_i),\ \sigma_i^2(\boldsymbol{\theta}) = \sigma_\varepsilon^2 \exp(2\psi_i)/\mu_\varepsilon^2, \\ \gamma_i(\boldsymbol{\theta}) &= \gamma_\varepsilon \exp(3\psi_i)/\mu_\varepsilon^3, \text{ and } \kappa_i(\boldsymbol{\theta}) = \kappa_\varepsilon \exp(4\psi_i)/\mu_\varepsilon^4,\end{aligned}$$

from which, we get the forms of the linear and quadratic martingale differences from (3.1) as

$$\begin{aligned}m_i(\boldsymbol{\theta}) &= x_i - \exp(\psi_i), \text{ and} \\ M_i(\boldsymbol{\theta}) &= (x_i - \exp(\psi_i))^2 - \sigma_\varepsilon^2 \exp(2\psi_i)/\mu_\varepsilon^2\end{aligned}$$

From (3.2), their quadratic variations and covariation are respectively

$$\begin{aligned}\langle m \rangle_i &= \sigma_\varepsilon^2 \exp(2\psi_i)/\mu_\varepsilon^2, \\ \langle M \rangle_i &= (\kappa_\varepsilon - \sigma_\varepsilon^4) \exp(4\psi_i)/\mu_\varepsilon^4, \text{ and} \\ \langle m, M \rangle_i &= \gamma_\varepsilon \exp(3\psi_i)/\mu_\varepsilon^3\end{aligned}$$

so that from (3.7),

$$\rho_i^2 = \frac{\sigma_\varepsilon^2(\kappa_\varepsilon - \sigma_\varepsilon^4)}{\sigma_\varepsilon^2(\kappa_\varepsilon - \sigma_\varepsilon^4) - \gamma_\varepsilon^2} \text{ and } \eta_i = \frac{\gamma_\varepsilon \mu_\varepsilon^3}{\sigma_\varepsilon^2(\kappa_\varepsilon - \sigma_\varepsilon^4)\exp(3\psi_i)}$$

These quantities enable us to compute the forms of $\mathbf{a}_{i-1}^*(\boldsymbol{\theta})$ and $\mathbf{b}_{i-1}^*(\boldsymbol{\theta})$ in (3.5) and (3.6), and thus derive the optimal EF $g_{C,\lambda}^*(\boldsymbol{\theta})$ in (3.9). Solving the set of nonlinear equations $g_{C,\lambda}^*(\boldsymbol{\theta}) = \mathbf{0}$ gives the EF estimates of $\boldsymbol{\theta} = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$.

3.2.3 Parameter Estimation

We describe three computational approaches that enable fast and accurate estimation of the vector of model parameters $\boldsymbol{\theta}$ in practical applications. The most direct approach involves the solution of the system of nonlinear equations $g^*_{C,\lambda}(\boldsymbol{\theta}) = \mathbf{0}$ or $g^*_C(\boldsymbol{\theta}) = \mathbf{0}$, for penalized EF and non-penalized EF approach, respectively. This is discussed in Section 3.2.3. Section 3.2.3 describes recursive formulas shown in (3.13)-(3.14), and enables online estimation of $\boldsymbol{\theta}$ as the i^{th} duration is observed. Section 3.2.3 discusses the use of approximating scalar recursions for estimating the components of $\boldsymbol{\theta}$, and iterating to convergence. Especially in big data situations involving streaming data, the recursive methods will provide computationally attractive choices.

Method 1: Nonlinear Equations Solver Estimation

The direct method of estimating $\boldsymbol{\theta}$ given an observed set of durations $x_i, i = 1, \ldots, n$ consists of solving the system of nonlinear equations $g^*_{C,\lambda}(\boldsymbol{\theta}) = \mathbf{0}$ for $\widehat{\boldsymbol{\theta}}$, where the optimal EF was given by (3.9). For best numerical accuracy, we use the MATLAB function `lsqnonlin`, which is called from an R wrapper. `lsqnonlin` is a nonlinear least squares optimization routine that calls the trust-region reflective algorithm with some boundary conditions on the solutions (Coleman and Li, 1996). Solving $g^*_C(\boldsymbol{\theta}) = 0$ gives the estimates based on the non-penalized EF.

Method 2: Approximate Vector Recursive Estimation

The second method for estimating $\boldsymbol{\theta}$ consists of deriving recursive formulas through a first-order Taylor approximation of $g^*_{C,\lambda}(\boldsymbol{\theta})$ about $\widehat{\boldsymbol{\theta}}_{i-1}$. From (3.9), we see that $g^*_{C,\lambda}(\boldsymbol{\theta}) = \sum_{i=1}^{n} g^*_{i,\lambda}(\boldsymbol{\theta})$, where $g^*_{i,\lambda}(\boldsymbol{\theta}) = \mathbf{a}^*_{i-1}(\boldsymbol{\theta})m_i(\boldsymbol{\theta}) + \mathbf{b}^*_{i-1}(\boldsymbol{\theta})M_i(\boldsymbol{\theta}) - p'_\lambda(|\boldsymbol{\theta}|)$. For $i = 1, \ldots, n$, let $\widehat{\boldsymbol{\theta}}_i$ denote the estimate of $\boldsymbol{\theta}$ given data $x_1, \ldots, x_i$. We can write $g^*_{i,\lambda}(\boldsymbol{\theta}) = \sum_{s=1}^{i} \Lambda_s(\boldsymbol{\theta})$. Let $g^{*(1)}_{i,\lambda}(\widehat{\boldsymbol{\theta}}_{i-1}) \equiv \partial g^*_{i,\lambda}(\boldsymbol{\theta}_{i-1})/\partial \boldsymbol{\theta}_{i-1}$, evaluated at $\widehat{\boldsymbol{\theta}}_{i-1}$. Then,

$$g^*_{i,\lambda}(\widehat{\boldsymbol{\theta}}_i) \simeq g^*_{i,\lambda}(\widehat{\boldsymbol{\theta}}_{i-1}) + g^{*(1)}_{i,\lambda}(\widehat{\boldsymbol{\theta}}_{i-1})(\widehat{\boldsymbol{\theta}}_i - \widehat{\boldsymbol{\theta}}_{i-1})$$

provided $\widehat{\boldsymbol{\theta}}_i - \widehat{\boldsymbol{\theta}}_{i-1}$ is "small". Since $g^*_{i,\lambda}(\boldsymbol{\theta}) = g^*_{i-1,\lambda}(\boldsymbol{\theta}) + \Lambda_i(\boldsymbol{\theta})$, we have

$$g^*_{i,\lambda}(\widehat{\boldsymbol{\theta}}_i) - g^*_{i-1,\lambda}(\widehat{\boldsymbol{\theta}}_{i-1}) \simeq g^{*(1)}_{i,\lambda}(\widehat{\boldsymbol{\theta}}_{i-1})(\widehat{\boldsymbol{\theta}}_i - \widehat{\boldsymbol{\theta}}_{i-1}) + \Lambda_i(\widehat{\boldsymbol{\theta}}_{i-1}),$$

or

$$0 \simeq g^{*(1)}_{i,\lambda}(\widehat{\boldsymbol{\theta}}_{i-1})(\widehat{\boldsymbol{\theta}}_i - \widehat{\boldsymbol{\theta}}_{i-1}) + \Lambda_i(\widehat{\boldsymbol{\theta}}_{i-1})$$

Therefore,

$$(\widehat{\boldsymbol{\theta}}_i - \widehat{\boldsymbol{\theta}}_{i-1}) \simeq -[g^{*(1)}_{i,\lambda}(\widehat{\boldsymbol{\theta}}_{i-1})]^{-1}\Lambda_i(\widehat{\boldsymbol{\theta}}_{i-1})$$

Further,

$$\widehat{\boldsymbol{\theta}}_i \simeq \widehat{\boldsymbol{\theta}}_{i-1} - [g^{*(1)}_{i,\lambda}(\widehat{\boldsymbol{\theta}}_{i-1})]^{-1}\Lambda_i(\widehat{\boldsymbol{\theta}}_{i-1}),$$

where $\Lambda_i(\widehat{\boldsymbol{\theta}}_{i-1})$ denotes $g^*_{i,\lambda}(\boldsymbol{\theta})$, evaluated at $\widehat{\boldsymbol{\theta}}_{i-1}$. Let $\boldsymbol{K_i} = -[g^{*(1)}_{i,\lambda}(\widehat{\boldsymbol{\theta}}_{i-1})]^{-1}$. The optimal recursive estimate for $\boldsymbol{\theta}$ based on the penalized EF is therefore

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_i = \widehat{\boldsymbol{\theta}}_{i-1} + \mathbf{K}_i \Big(& \mathbf{a}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1}) m_i(\widehat{\boldsymbol{\theta}}_{i-1}) \\ & + \mathbf{b}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1}) M_i(\widehat{\boldsymbol{\theta}}_{i-1}) - p'_\lambda(|\widehat{\boldsymbol{\theta}}_{i-1}|) \Big), \text{where}\end{aligned} \tag{3.13}$$

$$\begin{aligned}\mathbf{K}_i = \mathbf{K}_{i-1} \Bigg(\mathbf{I}_p - \Bigg(& \mathbf{a}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1}) \frac{\partial m_i(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial \boldsymbol{\theta}'} + \frac{\partial \mathbf{a}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial \boldsymbol{\theta}} m_i(\widehat{\boldsymbol{\theta}}_{i-1}) + \mathbf{b}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1}) \\ & \times \frac{\partial M_i(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial \boldsymbol{\theta}'} + \frac{\partial \mathbf{b}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial \boldsymbol{\theta}} M_i(\widehat{\boldsymbol{\theta}}_{i-1}) - p''_\lambda(|\widehat{\boldsymbol{\theta}}_{i-1}|) \Bigg) \mathbf{K}_{i-1} \Bigg)^{-1}\end{aligned} \tag{3.14}$$

for $i = 1, \ldots, n$, and $\mathbf{I}_P$ is the P-dimensional identity matrix. Use of the recursions (3.13)-(3.14) yields fast, online estimates of the model parameters. That is, for each $i = 1, \ldots, n$, we construct $\widehat{\boldsymbol{\theta}}_i$ in terms of $\widehat{\boldsymbol{\theta}}_{i-1}$ and $\boldsymbol{K}_i$. We start the recursions at initial parameter values (see Section 3.2.4). The estimate $\widehat{\boldsymbol{\theta}}_n$ gives the final estimate of $\boldsymbol{\theta}$. We can obtain the optimal recursive estimate for $\boldsymbol{\theta}$ based on the non-penalized EF by dropping $p'_\lambda(|\widehat{\boldsymbol{\theta}}_{i-1}|$ and $p''_\lambda(|\widehat{\boldsymbol{\theta}}_{i-1}|)$ from (3.13) and (3.14) respectively.

Note that for the Log ACD(p, q) model, all partial derivatives in the recursive estimation procedure may be written in terms of $\partial\psi_i/\partial\boldsymbol{\theta}$, where

$$\frac{\partial \psi_i}{\partial \boldsymbol{\theta}} = (1, \log x_{i-1}, \ldots, \log x_{i-p}, \psi_{i-1}, \ldots, \psi_{i-q})$$

Such expressions can also be derived for different classes of duration models discussed in Pacurar (2008) or Thavaneswaran et al. (2015). In some cases, especially when P is large and/or the

coefficients are too close to the boundary of the stationarity region, the recursions in (3.13)-(3.14) may face some issues due to singularity of the $\mathbf{K}_i$ matrix for some values of i. Method 3 overcomes this issue.

Method 3: Approximate Iterated Scalar Recursive Estimation

We may implement the formulas in (3.13)-(3.14) as a sequence of scalar recursions for each component of $\boldsymbol{\theta}$, and iterate these to convergence. For instance, suppose $\boldsymbol{\theta} = (\omega, \alpha, \beta)$ so that $P = 3$ for the Log ACD$(1, 1)$ model. We start with initial parameter values and run the recursion on each of these components singly, holding the other two parameters fixed at the previous iteration value. Once all three parameters get updated for a given i, we move to the next duration. We iterate the entire procedure until the estimation procedure has converged.

3.2.4 Choice of Initial Values in Log ACD Models

Starting the nonlinear equation solver under Method 1 or the vector or scalar recursions under Method 2 and Method 3 from reasonably good initial values will lead to faster convergence to accurate final estimates. Initial parameter values for Log ACD models can be based on properties of approximating non-Gaussian autoregressive moving average (ARMA) models for log durations.

Suppose $\{x_i\}$ follows the Log ACD(p, q) model given in (2.6)-(2.7). Let $y_i = \log x_i$, the

natural logarithm of x_i, and let $\delta = E(\log \varepsilon_i)$. Then,

$$\begin{aligned} y_i &= \psi_i + \log \varepsilon_i - \log \mu_\varepsilon \\ &= \omega + \sum_{j=1}^{p} \alpha_j y_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} + \log \varepsilon_i - \log \mu_\varepsilon \\ &= \omega + \sum_{j=1}^{p} \alpha_j y_{i-j} + \sum_{j=1}^{q} \beta_j (y_{i-j} - \log \varepsilon_{i-j} + \log \mu_\varepsilon) + \log \varepsilon_i - \log \mu_\varepsilon \\ &= \omega^\star + \sum_{j=1}^{p} \alpha_j y_{i-j} + \sum_{j=1}^{q} \beta_j y_{i-j} - \sum_{j=1}^{q} \beta_j \nu_{i-j} + \nu_i \end{aligned} \tag{3.15}$$

where $\omega^\star = \omega - (1 - \sum_{j=1}^{q} \beta_j)(\log \mu_\varepsilon - \delta) = \omega - \tau$, say, and $\nu_i = \log \varepsilon_i - \delta$, with $E(\nu_i) = 0$. From (3.15), we see that y_i follows an ARMA($\max(p, q), q$) model with $p \geq q$ and zero mean non-Gaussian errors ν_i , i.e.,

$$(1 - \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) B^j) y_i = \omega^\star + (1 - \sum_{j=1}^{q} \beta_j B^j) \nu_i, \tag{3.16}$$

with intercept $\omega^\star$, AR coefficients given by $\phi_j \equiv \alpha_j + \beta_j$ for $j = 1, \ldots, \max(p, q)$, and MA coefficients given by $\theta_j \equiv -\beta_j$ for $j = 1, \ldots, q$. Equation (3.16) gives us a way to obtain initial values of Log ACD(p, q) model parameters. We use the *arima* function in **R**, with method = "*css*" to estimate $\mathbf{\Omega} = (\omega^\star, \phi_1, \cdots, \phi_{\max(p,q)}, \theta_1, \cdots, \theta_q)$.

Then, $\tilde{\beta}_j = -\tilde{\theta}_j$, $j = 1, \ldots, q$ and $\tilde{\alpha}_j = \tilde{\phi}_j - \tilde{\beta}_j$, $j = 1, \ldots, p$. Since we are unable to estimate μ_ε and δ, we cannot directly obtain $\tau = (1 - \sum_{j=1}^{q} \tilde{\beta}_j)(\log \mu_\varepsilon - \delta)$, and thence an initial estimate of ω. We therefore obtain $\tilde{\tau}$ and hence $\tilde{\omega}$ as follows. Given the initial values

$(\tilde{\omega^\star}, \tilde{\alpha}_1, \ldots, \tilde{\alpha}_p, \tilde{\beta}_1, \ldots, \tilde{\beta}_q)$, we obtain residuals

$$\tilde{\varepsilon}_i = \frac{x_i}{\exp(\tilde{\psi}_i)} = \frac{x_i}{\exp(\tilde{\omega^\star} + \sum_{j=1}^{p} \tilde{\alpha}_j y_{i-j} + \sum_{j=1}^{q} \tilde{\beta}_j \tilde{\psi}_{i-j})},$$

and estimate τ by $\log(\bar{\tilde{\varepsilon}_i})$ (since $\exp(\omega^\star)\exp(\tau) = \exp(\omega)$, and $\tilde{\varepsilon}_i = \varepsilon_i \exp(\tau)/\mu_\varepsilon$, we can show after some algebra that this quantity estimates τ).

We also need good initial values for $\mathbf{K}$ (see (3.14)) to start off the recursions, for which we use the "delta method". Let $\tilde{\mathbf{V}} = \text{Var}(\tilde{\mathbf{\Omega}})$ denote the variance-covariance matrix of initial estimates from the ARMA$(\max(p,q), q)$ model and let $\tilde{\mathbf{P}} = d\mathbf{\Theta}/d\mathbf{\Omega}$ be the matrix of first derivatives of the Log ACD(p,q) model parameters $\mathbf{\Theta}$ with respect to $\mathbf{\Omega}$, evaluated at the initial values. Then, $\tilde{\mathbf{K}} = (\tilde{\mathbf{P}}\tilde{\mathbf{V}}\tilde{\mathbf{P}}')^{-1}$.

3.3 Simulation Study

We first assess the accuracy of the three methods described in Section 3.2.3 for Log ACD(p,q) model parameter estimation using non-penalized estimating functions. We describe five scenarios. For each scenario, we generated $L = 200$ sets of durations data, each of length $n = 7500$. For Scenario 1, we generated durations from a Log ACD$(1,0)$ model with $\omega = 0.25, \alpha = 0.05$, and $\varepsilon_i \sim$ gamma$(0.6, 0.7)$ distribution. For Scenario 2, durations were generated from a Log ACD$(1,1)$ model with $\omega = 0.04, \alpha = 0.05, \beta = 0.75$, and $\varepsilon_i \sim$ exponential(1) distribution. In Scenario 3, we generated data from a Log ACD$(1,1)$ model with $\omega = 1.0, \alpha = 0.05, \beta = 0.60$, and $\varepsilon_i \sim$ Weibull$(0.4, 0.5)$ distribution. For Scenario 4, we generated durations from a Log ACD$(2,1)$ model with $\omega = 0.5, \alpha_1 = 0.05, \alpha_2 = 0.10, \beta_1 = 0.60$, and errors

$\varepsilon_i \sim$ Weibull$(0.9, 0.9)$ distribution. In Scenario 5, we generated data from a Log ACD$(2, 2)$ model with $\omega = 0.15, \alpha_1 = 0.05, \alpha_2 = 0.10, \beta_1 = -0.05, \beta_2 = 0.70$, and $\varepsilon_i \sim$ gamma$(0.5, 0.8)$ distribution. The true parameters were selected to reflect values we observed with the real data . Specifically, all the parameters are close to what we observed in our real data analysis (to be discussed in Section 3.4). We rearranged them and got different scenarios. Although we generated errors from different parametric distributions, the EF estimation approach itself does not rely on distributional assumptions as we have seen in the previous sections.

As described in Section 3.2.3, final model parameter estimates were obtained using Methods 1, 2, and 3, starting with initial values obtained by fitting an ARMA$(\max(p, q), q)$ model to each log-transformed duration series as discussed in Section 3.2.4. Recall that Method 1 corresponds to NESE, which solves a system of P nonlinear equations, Method 2 or AVRE is based on recursive formulas for the entire vector $\boldsymbol{\theta}$, while Method 3 is denoted by AISRE, and iterates the recursive formulas for each scalar parameter until convergence is achieved. Figure 8 shows the boxplot of the final parameter estimates. All three methods provide final estimates that are close to the true values of the model parameters, which verifies the usefulness of these approaches in data analysis. Among the three methods, Method 2 (AVRE) is the fastest, and additionally allows us to trace variation of parameters. Hence we use Method 2 in the following sections for penalized EF and non-penalized EF approaches.

To implement the penalized estimating functions approach, we must select the two unknown tuning parameters a and λ. Following Fan and Li (2001), we take $a = 3.7$. Unlike their cross validation method, we use a different approach to select λ. We fit a Log ACD$(p, 0)$ model over a set of λ values, and select λ based on a trade-off between accuracy and sparsity. Specifically,

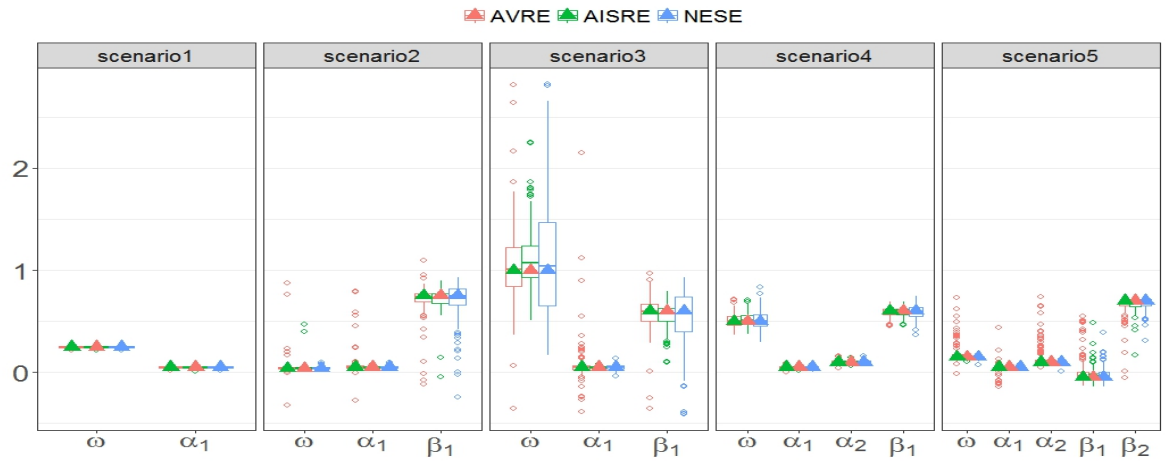

Figure 8: Boxplots of parameter estimates for different Log ACD(p, q) models for $L = 200$ simulated durations of length $n = 7500$. the triangle represents the true value.

we select the λ value that gives us the smallest MAD among the models with a certain specified level of sparsity, i.e., a certain number of user-specified nonzero estimates, say five.

To study the penalized model fit, we simulated data from two different Log ACD models, and fit a Log ACD$(20, 0)$ model in each case, thereby approximating a Log ACD(p, q) model with a Log ACD$(p, 0)$ model for large p. This is similar to the practice in the time series literature of building penalized autoregressive AR(p) rather than penalized autoregressive moving average ARMA(p, q) models. For Scenario 1, we generated durations of length $n = 7500$ from a Log ACD$(2, 0)$ model with $\omega = 0.25, \alpha_1 = 0.20, \alpha_2 = 0.10$, and $\varepsilon_i \sim$ gamma$(0.5, 0.6)$. Figure 9 shows the solution path for this simulated data set. The vertical line marks the optimal λ selected by using the MAD criterion we mentioned earlier. Since the α coefficients of the Log ACD models are usually small, we used a grid of small values for the search.

For Scenario 2, $n = 7500$ durations were generated from a Log ACD$(4, 0)$ model with $\omega = 0.10, \alpha_1 = 0.10, \alpha_2 = 0.05, \alpha_3 = 0.05, \alpha_4 = 0.10$, and $\varepsilon_i \sim$ gamma$(0.5, 0.6)$ distribution.

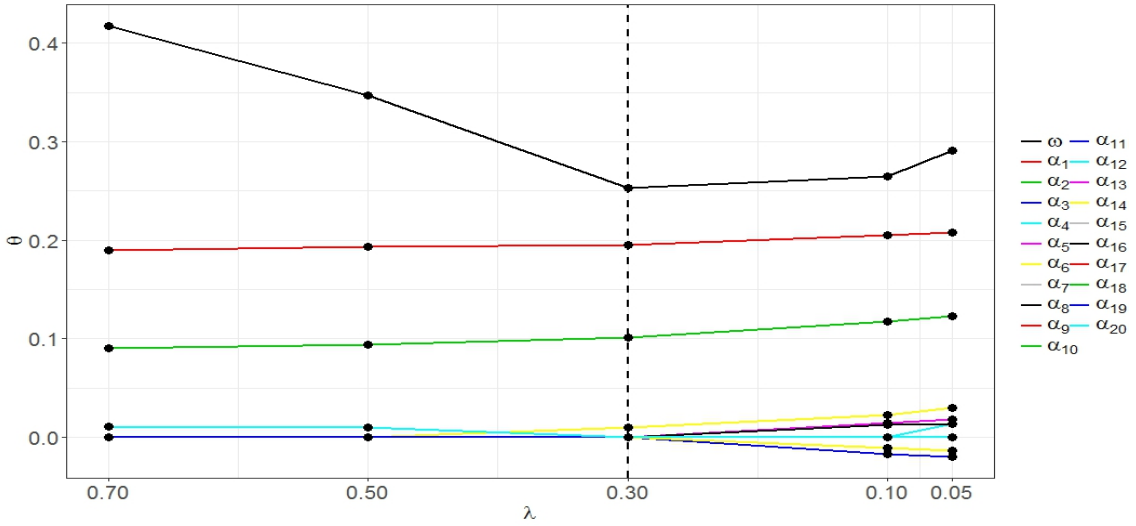

Figure 9: Solution path for a simulated set of durations from the Log ACD$(2, 0)$ model. The vertical bar represents the optimal λ.

Both scenarios were replicated 500 times. Table 1 displays the number of times that nonzero parameters were correctively identified and the percentiles of the parameter estimates. Clearly, our penalized estimating function approach is able to correctly identify the model order and estimate the parameter values well.

3.4 Modeling Daily Adjusted Financial Durations

We analyzed adjusted durations for trading days in June 2013 for four assets: BAC, GE, IBM and MMM, which were described in Section 2.2. Recall that the adjusted durations showed a considerable amount of clustering, suggesting that periods of rapid trading are likely to be followed by rapid trading, while periods of slow trading are likely to be followed by slow trading. We used data from the first $T_C = 15$ trading days in June 2013 as the training (or calibration) data, and observations from the last $T_H = 5$ days as holdout data. We let $x_i(t)$ denote the

Table 1: Number of times the nonzero parameters are correctively identified and percentiles of parameter estimates for the Log ACD$(p, 0)$ models for $L = 500$ simulated durations of length $n = 7500$.

$f_\varepsilon(.)$	Param	True	Number of Times Selected	5^{th}	50^{th}	95^{th}
Gamma$(0.5, 0.6)$	ω	0.25	500	0.22	0.25	0.27
	α_1	0.20	500	0.19	0.20	0.20
	α_2	0.10	500	0.09	0.10	0.10
Gamma$(0.5, 0.6)$	ω	0.10	480	0.06	0.09	0.14
	α_1	0.10	500	0.09	0.10	0.10
	α_2	0.05	500	0.04	0.05	0.05
	α_3	0.05	498	0.04	0.05	0.05
	α_4	0.10	500	0.09	0.10	0.10

ith adjusted duration for the tth day; $i = 1, \ldots, n(t)$, and $t = 1, \ldots, T_C$. In section 3.4.1, we fit candidate models of different orders to data from each day, and selected the best model for each day based on the Mean Absolute Deviation (MAD) criterion on the predictions. In section 3.4.2, we describe penalized estimating functions approach in which we fit the same Log ACD model with large model order and allow the penalty function to determine the estimated coefficients for each day. We then computed the prediction coverage for the holdout days.

3.4.1 Estimating Functions Approach

We fit Log ACD(p, q) models to the daily adjusted durations using the optimal recursions described in (3.13)-(3.14), which enables online parameter estimation of $\boldsymbol{\theta}(t) = (\omega(t), \boldsymbol{\alpha}'(t), \boldsymbol{\beta}'(t))'$ (we use the notation $\boldsymbol{\theta}(t)$ to denote the vector of parameters $\boldsymbol{\theta}$ for the tth day). For each day, we allowed the data to select the model orders p and q. Let $\widehat{\boldsymbol{\theta}}(t) = (\hat{\omega}(t), \hat{\boldsymbol{\alpha}}'(t), \hat{\boldsymbol{\beta}}'(t))'$ denote the estimated model parameter vector on day t, and let $\hat{x}_i(t) = \exp(\hat{\psi}_i(t))$, $i = 1, \ldots, n(t)$ be the estimated adjusted durations for that day. For each day $t = 1, \ldots, T_C$, we chose the best

Table 2: Parameter estimates for adjusted BAC durations in June, 2013.

Date	$\hat{\omega}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	d
20130603	0.227	0.047						0.001
20130604	0.319	0.103						0.005
20130605	0.258	0.074						0.001
20130606	0.289	0.093						0.006
20130607	0.261	0.034						0.000
20130610	0.299	0.011						0.005
20130611	−0.009	−1.146						0.054
20130612	0.330	0.051						0.001
20130613	0.293	0.025						0.001
20130614	0.137	0.046	0.048		−0.057	0.753		0.111
20130617	−0.015	0.555						0.305
20130618	0.045	0.014	0.041	−0.009	0.167	0.394	0.341	0.162
20130619	−0.501	−0.625						0.767
20130620	0.240	0.066						0.002
20130621	0.213	0.059						0.002

model as the one which minimized the Mean Absolute Deviation (MAD) defined under each fitted model as MAD $= \sum_{i=2}^{n(t)} |\hat{x}_i(t) - x_i(t)|/(n(t) - 1)$. That is, for different days, we could choose different model orders p and q, with $p \geq q$. If two models for the same day had very similar MAD values, the model for which the absolute difference between the average of the empirical durations and fitted durations, $d = |\bar{x}_i - \bar{\hat{x}}_i|$, is closest to zero, was selected. The optimal recursive estimates from modified (3.13) - (3.14) are shown in Table 2 - 5 for the four stocks. The estimates are reasonably stable across the different days in the calibration set. Although we do not show the SEs of the estimates due to space limitations, these are easily obtained as discussed at the end of Section 3.2.1.

For BAC, our algorithm selected the Log ACD$(1, 0)$ model for 13 out of 15 trading days in the calibration set, the Log ACD$(2, 2)$ model for Fri. June 14 and the Log ACD$(3, 3)$ model for

Table 3: Parameter estimates for adjusted GE durations in June, 2013.

Date	$\hat{\omega}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	d
20130603	0.281	0.064						0.001
20130604	0.129	0.067	0.048	0.010	0.311	−0.204	0.552	0.025
20130605	0.277	0.056						0.144
20130606	0.018	-0.021			0.960			0.127
20130607	0.203	0.060	−0.013		0.870	−0.400		0.053
20130610	0.526	0.652						0.537
20130611	0.124	0.010	0.058		−0.150	0.813		0.001
20130612	0.167	0.003	−0.080	−0.103	0.739	0.201	−0.150	0.818
20130613	−0.034	0.415						0.431
20130614	0.215	0.036	0.026		0.612	0.046		0.376
20130617	0.340	0.072						0.005
20130618	1.041	0.123	0.142		−1.392	−0.465		0.042
20130619	0.295	0.056						0.002
20130620	0.279	0.024						0.000
20130621	0.274	0.040						0.001

Table 4: Parameter estimates for adjusted IBM durations in June, 2013.

Date	$\hat{\omega}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	d
20130603	1.967	0.030						5.596
20130604	0.003	−0.341	0.103	−0.069	0.017	0.905		0.226
20130605	0.123	0.127	0.177	−0.038	−0.452	0.407	0.737	0.035
20130606	0.482	0.082	−0.039		0.656	−0.677		0.049
20130607	−1.718	−0.211	−1.637	−0.496	2.543	0.708		1.585
20130610	−0.037	0.544						0.475
20130611	0.105	0.048	−0.207		0.739	0.169		1.990
20130612	0.459	−0.549						0.223
20130613	0.411	−0.461						0.088
20130614	0.385	0.066						0.129
20130617	0.557	−0.031	0.109	0.050	−0.511	0.042	0.595	0.331
20130618	0.461	0.066						0.047
20130619	0.238	0.215						0.349
20130620	0.358	−0.068						0.011
20130621	0.339	0.148						0.007

Table 5: Parameter estimates for MMM adjusted durations in June, 2013.

Date	$\hat{\omega}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	d
20130603	1.765	0.324						4.561
20130604	0.186	0.077	0.007	0.029	0.349	−0.508	0.747	0.009
20130605	−0.100	0.045						0.705
20130606	0.210	−0.304						0.315
20130607	0.383	−0.173						0.125
20130610	0.351	−0.093	0.105		−0.142	0.646		0.513
20130611	0.443	−0.356						0.017
20130612	0.133	0.234	0.102	−0.048	0.109	0.581		0.042
20130613	0.157	0.086	0.003		−0.054	0.712		0.028
20130614	0.497	0.099						0.033
20130617	0.417	0.123						0.088
20130618	0.506	0.011						0.052
20130619	−0.497	0.613						0.984
20130620	2.126	−0.117						6.914
20130621	0.155	0.018	−0.094	−0.031	0.134	0.679		0.748

Tue. June 18. The estimates of ω and α_1 in Table 3 were very similar under the Log ACD$(1,0)$ model except on three days, Tue. June 11, Mon. June 17 and Wed. June 19.

The differences between the fitted models on different days may be at least partly explained by differences in their trading intensities (frequencies of diurnally adjusted durations in 15-minute intervals during the trading day) as seen from Figure 7 for June 10, June 11, June 14, June 17, June 18 and June 19. On days where the histogram exhibited a deep U-shaped behavior, with higher frequencies at the opening and closing periods relative to smaller frequencies in the middle of the day, as for example on Mon. June 10, a Log ACD(1,0) model with $\hat{\alpha}_1$ ranging in the interval $(0.01, 0.1)$ was selected.

By contrast, histograms on Tue. June 11 and Mon. June 17 displayed a wider U-shape, and although the Log ACD$(1,0)$ was the best model for these two days, the estimates of α_1 were

-1.146 and 0.555 respectively, very different from those on the typical days. The histograms for Fri. June 14 and Tue. June 18 were flatter, and higher order Log ACD models were selected to capture this pattern. The histogram for Wed. June 19 was very different from any U-shaped behavior, and had a reverse Z-shape. Although the Log ACD$(1, 0)$ was selected as the best model for that day, we observed that the fits were not good, and the value of d was 0.769, compared to d values that ranged between 10^{-4} and 0.3 for the other days. There did not appear to be any day of the week effect, but the second week in June appeared to exhibit the most disparate behavior. Our overall assessment is that the distribution free and online recursive EF approach is very attractive since it adapts very well to the daily behavior of the data and fits the best model, and provides reliable estimates. We observed that the best model fits were obtained when the histograms of adjusted duration counts followed a U-shaped behavior.

We carried out a detailed analysis for the other three stocks as well, but only provide brief summaries here. For GE, the best model was again the Log ACD$(1, 0)$ model for 8 of the 15 trading days. However, on Tue. June 4 and Wed. June 12, the Log ACD$(3, 3)$ model was selected; on Fri. June 7, Tue. June 11, Fri. June 14 and Tue. June 18 the Log ACD$(2, 2)$ model was the best, and on Thur. June 6, the Log ACD$(1, 1)$ model was selected. The absolute difference d between the average of the empirical durations and fitted durations ranged between 10^{-4} and 0.4 for most of the trading days except for Mon. June 10 ($d = 0.54$) and Wed. June 12 ($d = 0.82$).

For IBM, the Log ACD$(1, 0)$ model was selected on 9 of the 15 days. On Wed. June 5 and Mon. June 17, the Log ACD$(3, 3)$ model was the best; on Thur. June 6 and Tue. June 11, the

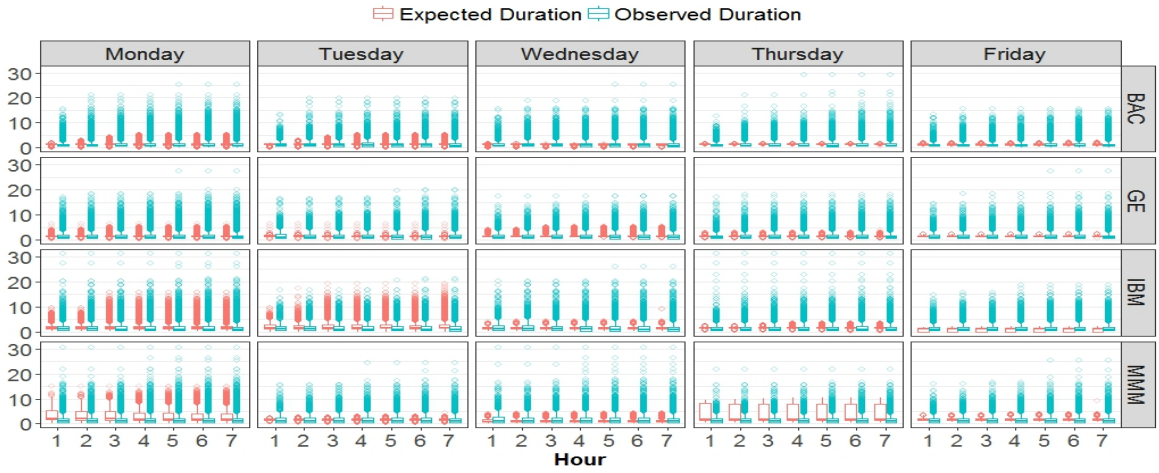

Figure 10: Boxplot of expected and observed durations (calibration days) for different days of the week across the four stocks. Both durations are cumulated on a 1 hour running window.

Log ACD$(2,2)$ model was selected; and the Log ACD$(3,2)$ was selected on Tue. June 4 and Fri. June 7. The values of d ranged between 10^{-4} and 0.5 for most of the trading days except for Mon. June 3 ($d = 5.6$), Fri. June 7 ($d = 1.6$) and Tue. June 11 ($d = 2.0$). Since IBM is a relatively low liquidity stock (average length of adjusted durations was about $n = 4800$) with a relatively large price spread (\$3 to \$4 on average), outliers (very long durations) may occur. This added another dimension of difficulty in fitting the Log ACD models to this stock's durations.

For MMM, the best model was again the Log ACD$(1,0)$ model for 10 of the 15 days. On Tue. June 4, the Log ACD$(3,3)$ model was selected; on Mon. June 10 and Thur. June 13, the Log ACD$(2,2)$ model was selected; and on Wed. June 12 and Fri. June 21, the Log ACD$(3,2)$ model was selected. The value of d ranged between 10^{-4} and 0.5 for most of the days except for Mon. June 3 ($d = 4.6$), Wed. June 5 ($d = 0.71$), Wed. June 19 ($d = 0.98$), Thur. June 20 ($d = 6.91$), and Fri. June 21 ($d = 0.748$).

Figure 10 displays both expected durations and observed durations (calibration days) for different days of the week across the four stocks. Both durations were cumulated on a one-hour running window. Specifically, we collected the expected/observed durations for all three Mondays in the 15 calibration days, and produced the boxplot for each running time window. Hour 1 refers to 9:30am to 10:30am, Hour 2 refers to 9:30am to 11:30am, etc. The side-by-side boxplots reveal that the expected durations have different variability for different days of the week and for the different stocks. In general, the variations seem to be higher on Mondays and also higher for lower liquidity stocks. The plot also enables us to assess the goodness of fit of the modeling. For most days, we can see the observed durations generally have higher variabilities than the expected durations. One violation of this pattern are Tuesday's boxplots for IBM, where the variation in the expected duration is almost the same as that in the observed durations, likely due to an inadequate model fit for the second Tuesday in June. Figure 11 is created in the same fashion as Figure 10, by showing boxplots expected durations for the calibration days against the mean observed durations for the $T_H = 5$ holdout days. The patterns in these plots motivate us to study the prediction coverage for out-of-sample model evaluation, as discussed below.

Prediction coverage for the $T_H = 5$ holdout days can give useful information in practice. Based on the parameter estimates $\widehat{\boldsymbol{\theta}}(t) = (\hat{\omega}(t), \hat{\boldsymbol{\alpha}}'(t), \hat{\boldsymbol{\beta}}'(t))'$ for each calibration day $t = 1, \ldots, T_C$, we computed $\hat{x}_i(t) = \exp(\hat{\psi}_i(t))$, $i = 1, \ldots, n(t)$ as the estimated adjusted durations for day t, and then obtained the average estimated adjusted duration $\overline{\hat{x}}(t) = \sum_{i=1}^{n(t)} \hat{x}_i(t)/n(t)$. This yielded the values $\overline{\hat{x}}(1), \ldots, \overline{\hat{x}}(T_C)$. We computed $\overline{\hat{x}}^{(.025)}$ and $\overline{\hat{x}}^{(.975)}$ which were respectively the 2.5th and 97.5th percentiles of $\overline{\hat{x}}(1), \ldots, \overline{\hat{x}}(T_C)$. We checked whether the observed

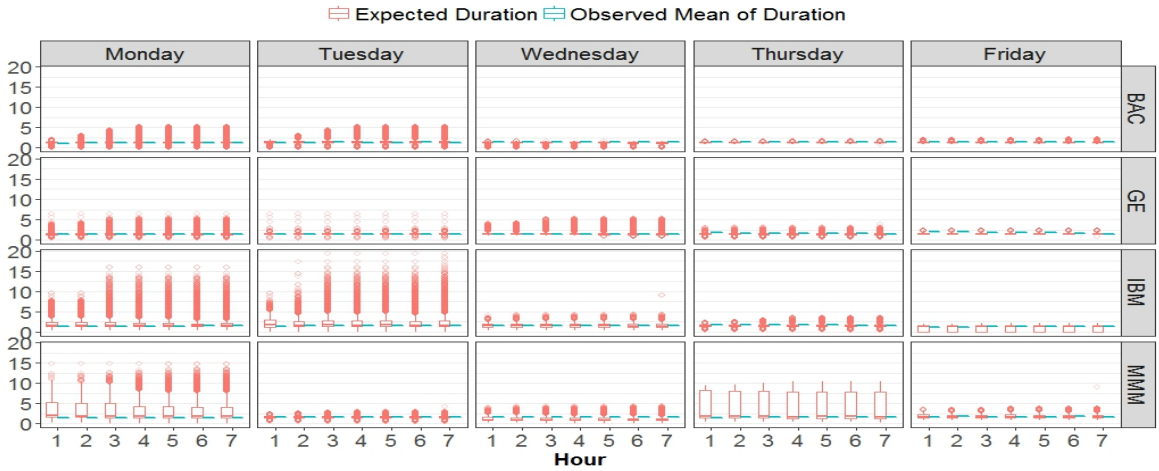

Figure 11: Boxplot of expected durations (calibration days) for different days of the week across the four stocks. A bar on the side of each boxplot represents the mean of observed durations (holdout days). Both durations are cumulated on a 1 hour running window.

average adjusted durations for any hold-out day t^* given by $\overline{x}(t^*) = \sum_{i=1}^{n(t^*)} x_i(t^*)/n(t^*)$, fell within $(\overline{\hat{x}}^{(.025)}, \overline{\hat{x}}^{(.975)})$, i.e., within the 95% limits that we obtained from the calibration data. For each of the four stocks, $\overline{x}(t^*)$ was within the 95% limits for all of the T_H hold-out days, leading to the conclusion that the EF modeling approach gave good out-of-sample predictive validity. A similar calculation may be easily done over cumulated one-hour windows and are shown in Figure 12.

Figure 12 is an alternate representation of Figure 11, where instead of the boxplots of expected durations, we have displayed the 2.5% and 97.5% bounds, together with the mean observed durations for the holdout days; these lie entirely within the bounds in all cases, showing that the prediction coverage of the EF approach is very high. These visual aids are easily constructed from the estimation results and can enable practitioners to observe and evaluate patterns leading to trading decisions.

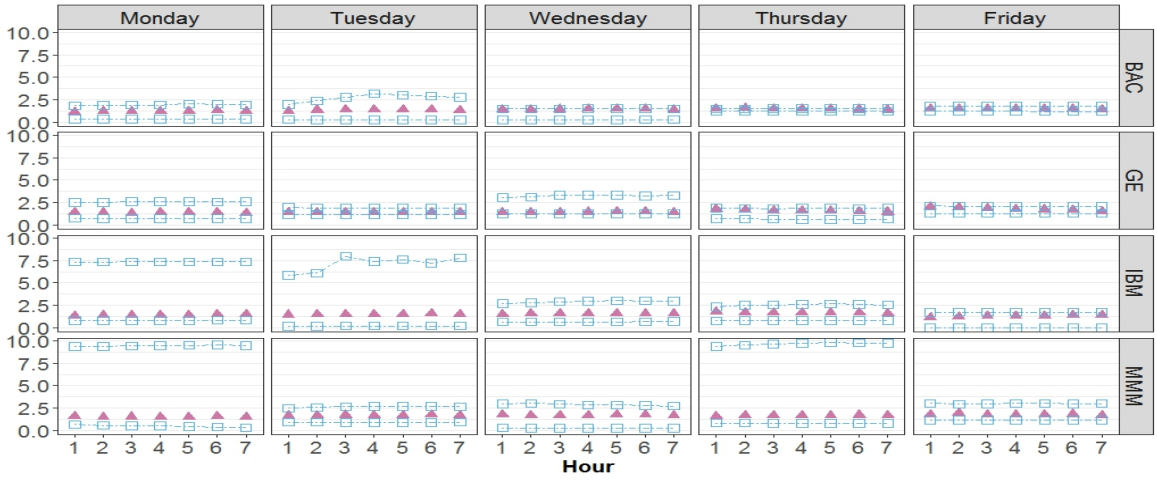

Figure 12: Quantiles of expected durations (calibration days) for different days of the week across the four stocks. Squares represent the 2.5% and 97.5% quantiles and solid triangles represent the means of observed durations (holdout days). Both durations are cumulated on a 1 hour running window.

3.4.2 Penalized Estimating Functions Approach

We ran the penalized approach for daily durations for the four stocks. We fit a Log ACD$(20, 0)$ model to data for each day, selected the best λ based on the criterion we mentioned above, and obtained the optimal recursive estimates from (3.13) - (3.14). Due to space limitations, only the estimates of ω are shown in Table 6; these are more stable across stocks than under the method without penalty. In general, we observe better prediction accuracy (d) for IBM and MMM and for a few days of BAC and GE. The choice of λ seems to depend on the liquidity of a stock which is proportional to n. Since λ controls the step size of the gradient of the penalty, a more highly liquid stock, such as BAC, will require a smaller value of λ for convergence than a less liquid stock, like MMM.

For the penalized estimating function approach, Figures 13 - 15 are created in the same fashion as Figures 10 - 12. A comparison of these figures with their counterparts in Section

Table 6: Parameter estimates for adjusted BAC, GE, IBM and MMM durations in June, 2013.

Date	BAC			GE			IBM			MMM		
	λ	$\hat{\omega}$	d	λ	$\hat{\omega}$	d	λ	$\hat{\omega}$	d	λ	$\hat{\omega}$	d
20130603	0.7	0.220	0.010	0.5	0.28	0.009	10	0.425	0.023	14	0.428	0.025
20130604	0.7	0.306	0.018	0.7	0.36	0.012	10	0.421	0.029	20	0.447	0.029
20130605	0.5	0.251	0.012	1.0	0.36	0.026	10	0.405	0.041	41	0.463	0.022
20130606	0.5	0.255	0.007	1.0	0.34	0.021	10	0.426	0.031	50	0.437	0.055
20130607	1.0	0.289	0.020	1.0	0.34	0.019	10	0.442	0.030	34	0.467	0.024
20130610	0.7	0.315	0.019	1.0	0.34	0.016	10	0.437	0.036	24	0.430	0.027
20130611	0.5	0.324	0.009	1.0	0.37	0.013	10	0.437	0.032	50	0.475	0.037
20130612	0.5	0.285	0.010	1.0	0.40	0.010	10	0.417	0.037	2	0.458	0.025
20130613	0.7	0.316	0.019	1.0	0.36	0.027	10	0.427	0.021	8	0.419	0.048
20130614	0.7	0.341	0.013	1.0	0.39	0.035	10	0.451	0.032	88	0.505	0.034
20130617	0.7	0.316	0.019	0.7	0.33	0.008	10	0.443	0.028	11	0.401	0.020
20130618	0.7	0.341	0.013	1.0	0.32	0.035	10	0.464	0.040	24	0.445	0.016
20130619	1.0	0.339	0.040	0.7	0.29	0.013	40	0.454	0.072	90	0.499	0.085
20130620	1.0	0.231	0.013	0.5	0.27	0.011	1	0.359	0.010	11	0.401	0.020
20130621	1.0	0.203	0.015	0.5	0.27	0.007	1	0.323	0.025	24	0.445	0.016

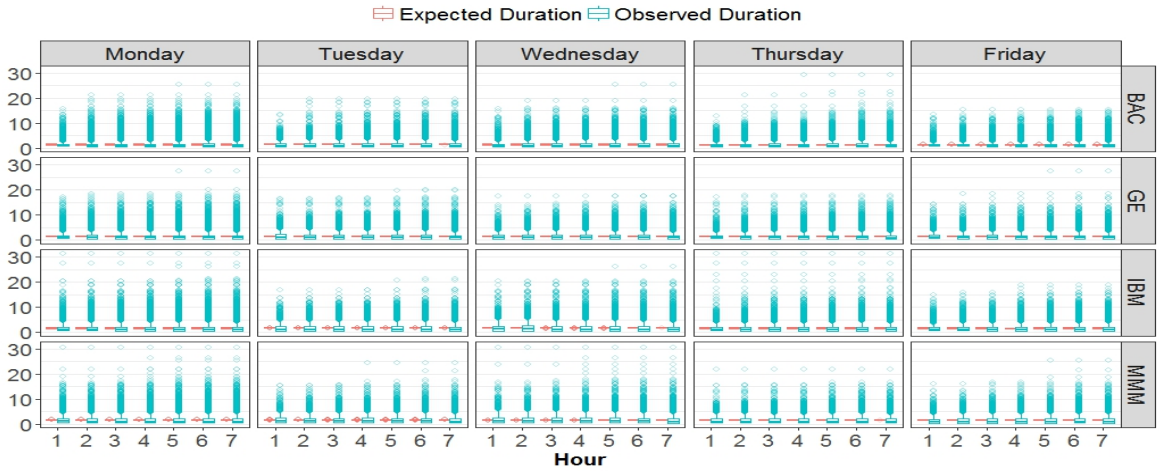

Figure 13: Boxplot of expected and observed durations (calibration days) for different days of the week across the four stocks. Both durations are cumulated on a 1 hour running window.

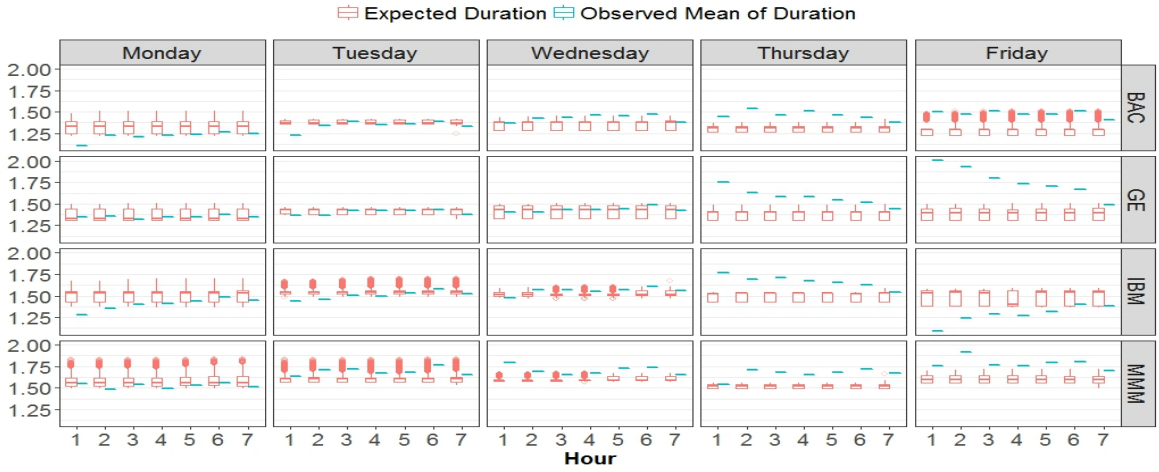

Figure 14: Boxplot of expected durations (calibration days) for different days of the week across the four stocks. A bar on the side of each boxplot represents the mean of observed durations (holdout days). Both durations are cumulated on a 1 hour running window.

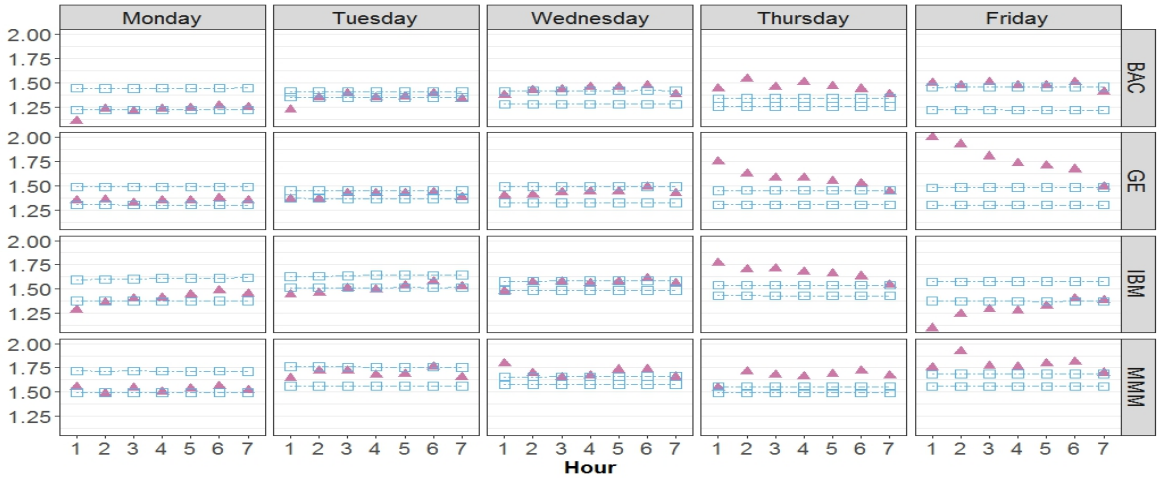

Figure 15: Quantiles of expected durations (calibration days) for different days of the week across the four stocks. Squares represent the 2.5% and 97.5% quantiles and solid triangles represent the means of observed durations (holdout days). Both durations are cumulated on a 1 hour running window.

3.4.1 is instructive in understanding the differences between the penalized and unpenalized estimating function approach. In short, there is a bias-variance trade-off. There are a few reasons that explain the differences between Figure 12 and Figure 15 (and similar reasons hold for differences between Figures 10 and 13 and Figures 11 and 14 respectively). First, expected durations calculated under the non-penalized method tend to have have larger variances than those computed using the penalized method. The distance between upper and lower bounds of the expected durations in Figure 12 is therefore larger than its counterpart in Figure 15. Second, the penalized approach generally has high precision (less variance) in terms of point estimates. Errors in terms of prediction accuracy (d) shown in Table 6 are smaller than those in Table 2. For example, the precision (d) estimate for the non-penalized approach is fairly large, being equal to 0.767 on June 19th, 2013 for BAC, while it is only 0.040 for the penalized approach. However, higher order models tend to result in underestimation of variations of expected durations. Thus, the coverages in Figure 15 are relatively lower compared to those in Figure 12.

3.5 Summary and Discussion

The important and necessary contribution of this chapter is to take the theory and method described in the literature to implementation in the context of financial durations modeling. Considerable novelty has gone into consolidating and implementing the three ways in which an analyst can now carry out the estimation and prediction. The implementation of this in the recursive estimation context will be especially useful for streaming big data. The benefits of our approach for practical application are three-fold. First, unlike the traditional

parametric methods, the EF approach only requires assumptions on the first few conditional moments of the durations and does not require specification of the probability distribution of the process. Second, the models and methods are data driven and are very flexible in accommodating various stochastic patterns. Third, we have developed a comprehensive computational framework via three approaches for parameter estimation, including solution of nonlinear estimating equations, recursive formulas for the vector-valued parameter estimates, and iterated component-wise scalar recursions. In addition, we have also introduced and implemented a penalized estimation functions approach both for the nonlinear equations solver and the online recursions. Our R code provides an easy framework of estimation, model selection and prediction of dynamic patterns in financial durations. It will be useful to investigate the approach used in Chen and Chan (2011) for model fitting using the PEF approach.

Chapter 4

Structural Break Detection in Univariate Durations

4.1 Problem Description

Structural breaks detection, or change points detection, has its origins in quality control (Page, 1955, 1954), but has since then become a widely used statistical analysis approach in various fields. Jassby and Powell (1990) conducted analysis of changing seasonal pattern, trend, oscillatory behavior, and unusual events in ecological time series data. Dias and Embrechts (2002) explored changes in dependent time series data in finance and insurance. In general, there are two main issues, one is whether or not a change or several changes might have occurred; one is identifying the times of any such changes. Much of the methodology was first proposed to tackle independent observations. However, detecting structural breaks of temporally correlated observations can be appealing for several reasons: piecewise models are usually more flexible than corresponding non-piecewise models; detected break points can be linked to the behavior of external variables of interests or historical events, thus providing more information to researchers; the last segment, viewed as current trend, is also helpful in forecasting the

future process.

Conceivably the simplest model with structural break points is defined as

$$Y_t = f_t + \varepsilon_t, \quad t \in \mathcal{T} \tag{4.1}$$

where $\mathcal{T}$ denotes the set of times, and $\{f_t : t \in \mathcal{T}\}$ is a one-dimensional time series model with m unknown structural break points. The locations $\tau_1, \tau_2, ..., \tau_m$ are usually unknown. $\{\varepsilon_t : t \in \mathcal{T}\}$ follows a parametric distribution with mean 0 and variance σ^2. Our objective is to detect number of break points m and their corresponding locations $\tau_1, \tau_2, ..., \tau_m$.

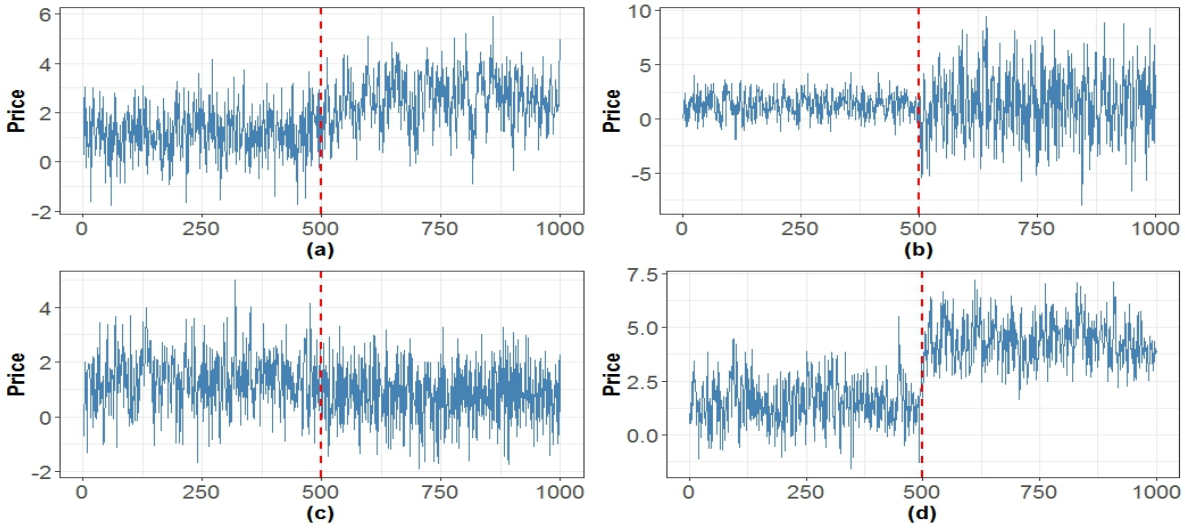

Figure 16: Some examples of structural breaks in univariate time series. (a): changing autoregressive (AR) model order; (b): changing noise variance; (c): changing parameter sign; (d): changing overall means

Many kinds of structural breaks problems exist in various areas. Figure 16 illustrates commonly occurring structural breaks in the case of such univariate time series. A structural break could be induced for instance, by a change of mean, or variance, or correlation structure, or model order.

In practice, what we observe at each time t or some other indexes could be numerically valued (structured data) or be networks or graphs (unstructured data). In this chapter, we mainly discuss structural break detection in univariate time series of durations between events in intra-day high-frequency transaction level financial data.

To describe the setup, let x_i, $i = 1, 2, \ldots, n$, denote a time series of adjusted durations, and let $\mathcal{F}_{i-1}^x$ denote the information associated with previous durations. Denote the structural break points between the j^{th} and $(j+1)^{th}$ observations following a Log ACD model as τ_j, $j = 1, ..., m$. Set $\tau_0 = 1$ and $\tau_{m+1} = n$. Then $\{x_i, \tau_j < i < \tau_{j+1}\}$ fits the j^{th} piece of Log ACD(p_j, q_j) process if

$$x_i = \exp(\psi_i^{(j)}) \frac{\varepsilon_i^{(j)}}{\mu_\varepsilon^{(j)}}, \tag{4.2}$$

where

$$\psi_i^{(j)} = \omega^{(j)} + \sum_{k=1}^{p_j} \alpha_k^{(j)} \log x_{i-k} + \sum_{k=1}^{q_j} \beta_k^{(j)} \psi_{i-k}^{(j)}, \quad \text{if } \tau_{j-1} < i < \tau_j. \tag{4.3}$$

The i.i.d. nonnegative errors $\varepsilon_i^{(j)}$ are segment-specific, independent of $\mathcal{F}_{i-1}^x$ and with $E(\varepsilon_i^{(j)}) = \mu_\varepsilon^{(j)}$.

In Figure 16, we display only one structural break for each case. It is possible that in real applications, there exist multiple structural break points. Also, a combination of structural breaks due to different causes may exist in the same time series. Figure 17 illustrates two such time series of financial durations. In Figure 17(a), 5 structural breaks evenly divide a length of 10000 simulated durations from equation (4.2) - (4.3) into 6 segments. The first, the third and the fourth segments are simulated from three different Log ACD$(1, 1)$ models (e.g. $\omega^{(j)}, \alpha^{(j)}, \beta^{(j)}, j = 1, 3, 4$ are different). The second, the fifth and the sixth segments are

generated from three different Log ACD$(1, 0)$ models. The error terms follow a Gamma$(1, 2)$ distribution and are the same for each segment. In this case, we show a combination of mean shift, model order change and existence of multiple structural break points. Figure 17(b) shows an example of a cluster of 3 structural break points. 4 segments are generated from 4 different LogACD$(1, 1)$ models but with the same ω. In addition, the first and fourth error terms follow a Weibull$(1.5, 2)$ distribution while the second and third error terms follow a Gamma$(2, 1)$ distribution. In this case, we show a combination of change in model parameter, change in error distributions and cluster of multiple structural break points.

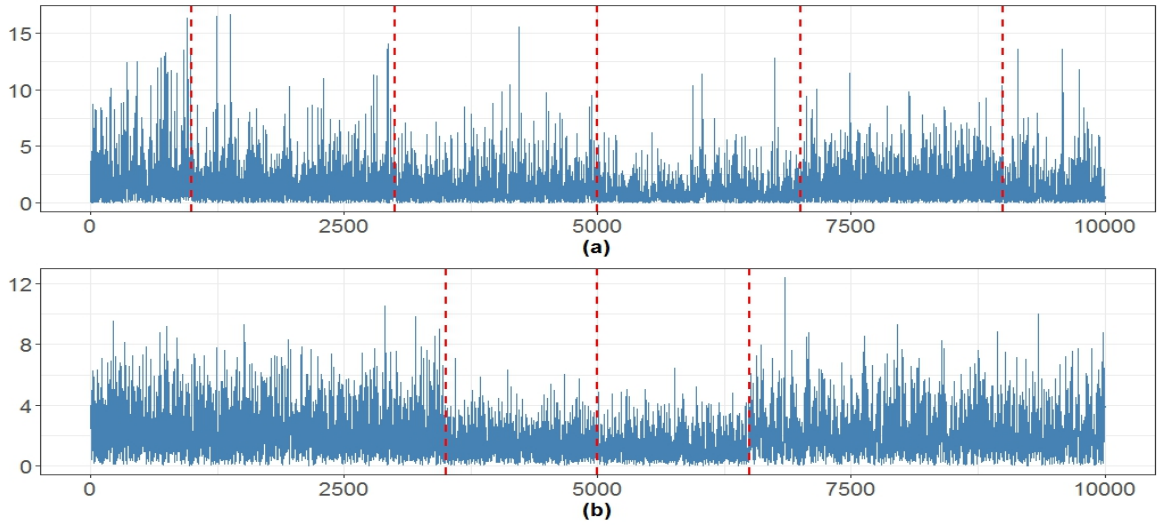

Figure 17: Two examples of structural breaks in univariate time series of financial durations. (a): A combination of mean shift, change in model order and existence of multiple structural break points. (b): A combination of change in model parameter, change in error distributions and cluster of multiple structural break points. Red dashed lines represent the structural break points.

In this chapter, we investigate structural breaks in terms of different combinations of mean shift, change in model order, change in model parameter and change in error distributions in high-frequency financial durations. We implement both offline and online method to detect

the changes of parameters estimated from the penalized estimating functions approach that was described in detail in Chapter 3.

4.2 A Review of Offline Methods

Structural break detection has been extensively studied for univariate time series $\{Y_t\}$, that is, for data where the observations are continuously valued scalars at each time point. Most of the methodologies focus on a posteriori or offline detection in structural break points. Traditionally, researchers viewed structural break detection as a hypothesis testing problem. The test statistic, usually one of the CUSUM-type family statistics, was used to investigate whether the null hypothesis of structural stability holds. The hypothesis testing procedures were non-parametric and needed not to fit any time series models (Bai, 1999; Banerjee and Urga, 2005; Csörgö and Horváth, 1997; Döring, 2011; Robbins, Gallagher, and Lund, 2016). Recently, several studies started to consider structural break detection problem as a model selection procedure (Aue, Cheung, Lee, Zhong, et al., 2017; Chan, Yau, and Zhang, 2014; Davis, Lee, and Rodriguez-Yam, 2006; Lu, Lund, Lee, et al., 2010; Robbins, Gallagher, Lund, and Aue, 2011). Section 4.2.1 - 4.2.2 are detailed reviews for two state-of-the-art approaches. Unlike the hypothesis testing procedures, likelihood ratio procedures took into account parametric time series assumptions, which described explicitly the dependence structure of time series data. In this way, researchers (Davis, Huang, and Yao, 1995; Hansen and Yu, 2000; Robbins, Gallagher, and Lund, 2016) easily implemented forecasting procedures. The relevant lines of research are summarized in the monograph by Csörgö and Horváth (1997), and in the more recent review paper by Aue and Horváth (2013).

The literature for structural break detection in high-frequency financial analysis is sparse. To our best knowledge, threshold autoregressive conditional duration model proposed by Zhang, Russell, and Tsay (2001) is the first and only one to detect structural break points. A brief introduction of TACD model has been already given in Section 2.3. A $J-$regime TACD model essentially allows the conditional means to depend nonlinearly on past information. Zhang et al. (2001) (referred to as ZRT) found unknown structural break points by sequentially using Lagrange Multiplier Tests. Then ZRT fitted a 3-regime TACD$(1,1)$ model for each segment and argued that a TACD model gave a good fit in terms of modeling nonlinearity. The procedure we proposed differs from a TACD model in that a TACD model has to specify the order of the model and the error distribution at first, and is applied as an offline model.

4.2.1 Automatic Piecewise Autoregressive Modeling (Auto-PARM)

Lately, there have been attempts to view structural break detection as a model selection problem. Davis et al. (2006) considered the problem of modeling a class of non-stationary time series using piecewise autoregressive (AR) processes. To be specific, a random process $\{Y_t\}$ that fits a piecewise autogregressive model (PARM) of order p (or PARM(p)) is defined as follows. Let $j = 1, \cdots, m$, denote the break point between the j^{th} and $(j+1)^{th}$ segments as τ_j, and set $\tau_0 = 1$ and $\tau_{m+1} = n$. Then the j^{th} piece of the series is modeled as

$$Y_t = X_{t,j}, \quad \tau_{j-1} \leq t < \tau_j \tag{4.4}$$

where $\{X_{t,j}\}$ denote a piecewise AR(p_j) models with white noise ε_t .

$$X_{t,j} = \gamma_j + \phi_{j1} X_{t-1,j} + \cdots + \phi_{j,p_j} X_{t-p_j,j} + \sigma_j \varepsilon_t \tag{4.5}$$

Given an observed time series $\{y_t\}_{t=1}^n$, the objective then is to obtain a "best"-fitting model from this class. This is equivalent to finding the "best" combination of the number of pieces $m+1$, the breakpoint locations $\tau_1, \cdots, \tau_m$ and the parameters for piecewise models (e.g., orders p_j's for Auto-PARM).

To solve the model selection problem, Davis et al. (2006) utilized the minimum description length (MDL) principle of Rissanen (1985) to define a best-fitting model. It begins with splitting observations $Y = (y_1, y_2, \cdots, y_n)$ into two parts. The first part, denoted by $\hat{\mathcal{F}}$, represents the fitted piecewise stationary models, and the second part, denoted by $\hat{\varepsilon} = Y - \hat{Y}$, represents the residuals, where $\hat{Y}$ is the fitted value for Y. Notice that once $\hat{\mathcal{F}}$ and $\hat{\varepsilon}$ are known, Y can be completely retrieved. The idea of MDL principle is to find the best pair $\hat{\mathcal{F}}$ and $\hat{\varepsilon}$ so that via encoding (or compressing) them, Y can be transmitted with the least amount of code length.

Practical minimization of the MDL criterion is not a trivial task. Genetic Algorithm (Holland, 1992) was used to handle this minimization problem. Genetic algorithms (GA) are a class of stochastic optimization techniques which are based on the idea of Darwin's theory of natural selection. Typically a GA begins with a random population of possible solutions to the optimization problems. These solutions are known as chromosomes and are often represented in vector form. These chromosomes are allowed to evolve over time through the so-called crossover

and mutation operations. The evolution process would ultimately lead to a chromosome which represents a good answer to the optimization problem.

A chromosome should contain information of all the break points τ_j as well as the piecewise model orders p_j's for any $\mathcal{F} \in \mathcal{M}$, where $\mathcal{M}$ denotes the whole class of piecewise models. A chromosome $\delta = (\delta_1, \cdots, \delta_n)$ is of length n with gene values δ_t defined as

$$\delta_t = \begin{cases} -1 & \text{if no break point at time } t; \\ p_j & \text{if } t = \tau_{j-1} \text{ and the parameter for the jth piece model is } p_j. \end{cases}$$

In fact, GA is essentially a large scale adaptive trial-and-error procedure which depends on a number of tuning parameters including an upper bound P_0 on the order of p_j, a minimum length m_B for each piecewise process, crossover rate, and mutation rate, etc. Davis et al. (2006) only established a consistent result for known number of structural break points. Hancock (2008) proved that estimated number of structural break and piecewise AR model orders could be weakly consistent using conditional maximum (Gaussian) likelihood variance estimates.

4.2.2 Group LASSO for Structural Break Autoregressive Models

Chan, Yau, and Zhang (2014) handled piecewise AR models from a new perspective. The modeling procedure embedded the structural break detection problem into a LASSO (Yuan and Lin, 2006) framework. To implement Group LASSO procedure, Chan et al. (2014) reformulated the piecewise AR model into a matrix format. Let $\boldsymbol{Y_n^0} = (Y_1, Y_2, ..., Y_n)'$, $\boldsymbol{\eta}(n) = (\sigma_1\varepsilon_1, \sigma_2\varepsilon_2, ..., \sigma_n\varepsilon_n)'$, $\boldsymbol{\beta_j} = (\gamma_j, \phi_{j,1}, \phi_{j,2}, ..., \phi_{j,p_j})$ for $j = 1, ..., m$,and $\boldsymbol{\theta}(n) = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2}, ..., \boldsymbol{\theta_n})'$.

They rewrote equations 4.4 and 4.5 as follows

$$\boldsymbol{Y_n^0} = \boldsymbol{X_n}\boldsymbol{\theta}(n) + \boldsymbol{\eta}(n) \tag{4.6}$$

where $\boldsymbol{X_n}$ is an $n \times np$ matrix and was defined by

$$\boldsymbol{X_n} = \begin{pmatrix} \boldsymbol{Y_0'} & 0 & 0 & \cdots & 0 \\ \boldsymbol{Y_1'} & \boldsymbol{Y_1'} & 0 & \cdots & 0 \\ \boldsymbol{Y_3'} & \boldsymbol{Y_2'} & \boldsymbol{Y_2'} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{Y_{n-1}'} & \boldsymbol{Y_{n-1}'} & \boldsymbol{Y_{n-1}'} & \cdots & \boldsymbol{Y_{n-1}'} \end{pmatrix}$$

, $\boldsymbol{Y_k'} = (Y_k, Y_{k-1}, ..., Y_{k-p+1})$, and

$$\boldsymbol{\theta_i} = \begin{cases} \boldsymbol{\beta_1}, & \text{when } i = 1 \\ \boldsymbol{\beta_{j+1}} - \boldsymbol{\beta_j} & \text{when } i = \tau_j \\ \mathbf{0}, & \text{otherwise} \end{cases}$$

for $i = 1, 2, ..., n$. By utilizing Group LASSO procedure, $\boldsymbol{\theta}(n)$ can be estimated as

$$\widehat{\boldsymbol{\theta}}(n) = \underset{\boldsymbol{\theta}(n)}{\operatorname{argmin}} \frac{1}{n} ||\boldsymbol{Y_n^0} - \boldsymbol{X_n}\boldsymbol{\theta}(n)||^2 + \lambda_n \sum_{i=1}^{n} ||\boldsymbol{\theta_i}||, \tag{4.7}$$

where λ_n is the tuning parameter which controls the sparsity of the estimates and $||\cdot||$ is the l_2 norm. Through the construction of $\boldsymbol{\theta_i}$, the AR parameters in each regime can be obtained by

$$\hat{\boldsymbol{\beta}}_{\mathbf{1}} = \hat{\boldsymbol{\theta}}_{\mathbf{1}}, \quad \text{and} \quad \hat{\boldsymbol{\beta}}_{\boldsymbol{j}} = \sum_{i=1}^{\tau_j} \hat{\boldsymbol{\theta}}_{\boldsymbol{i}}, \quad j = 1, ..., \hat{m} \tag{4.8}$$

Chan et al. (2014) proved that the number of structural break points are overestimated if the number of break points is unknown. However, the probability of all the true break points fall in the break points set tends to be 1. Therefore the authors proposed a two-step procedure. In step one, the candidate set of true break points, $\mathcal{A} = \{\tau_1^0, \tau_2^0, ..., \tau_{m_0}^0\}$, was estimated via Group LASSO procedure. The best possible subset of break points were obtained via prescribed information criteria in step two.

4.3 A Retrospective Monitoring Approach

We outline our monitoring procedure for parameter change in two ways. In this section, we assume that all data points are available. Based on visual plots, we would know the existence of structural break points, see Figure 18 - 19. We describe an automatic algorithm that helps investigators to identify the locations of observed structural breaks. We will discuss the details of an online monitoring scheme in Section 4.7.

Before outlining the central ideas of our retrospective procedure, we illustrate the online feature of approximate vector recursive estimation (AVRE) approach described in Section 3.2.3. To illustrate this feature, we generate two sets of durations. In scenario one, durations of length $n = 7500$ are generated from a Log ACD$(2, 0)$ model with parameters $\omega = 0.25, \alpha_1 = 0.2, \alpha_2 = 0.1$ and $\varepsilon_i \sim$ Gamma$(0.5, 0.6)$. In scenario two, durations of length $n = 7500$ are generated from a 3-segment piecewise Log ACD model defined in (4.2) and (4.3):

$$\psi_i = \begin{cases} 0.1 + 0.3 \log x_{i-1}, & \text{if } 1 \leq i < 3000 \\ 0.1 - 0.2 \log x_{i-1}, & \text{if } 3001 \leq i < 5250 \\ 0.1 + 0.3 \log x_{i-1}, & \text{if } 5251 \leq i \leq 7500 \end{cases}$$

and error terms $\varepsilon_i^{(j)} \sim \text{Weibull}(3, 4)$ for 3 segments. Figure 18 shows plots of both simulated durations. It is clear that two structural break points divide the simulated durations into three segments in Figure 18(b), while no structural break appears in Figure 18(a).

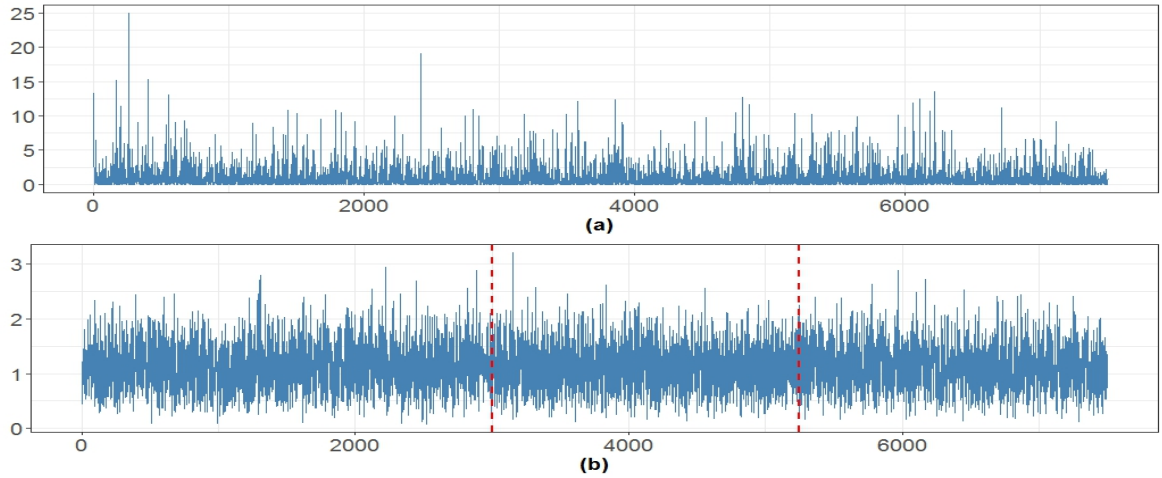

Figure 18: Plots of simulated durations generated from scenario one (a) and scenario two (b) respectively. Red dashed lines represent the structural break points.

We fit a Log ACD$(5, 0)$ model to the first set of simulated durations using penalized estimating functions (PEF) with $\lambda = 0.3$ and non-penalized estimating functions (EF) approaches. Figure 19(a) shows the online trace of parameters for two methods. It is clear that the true non-zero components (ω, α_1 and α_2) of parameters are not excessively penalized while zero valued components of parameters get penalized considerably. This phenomena is in line with the benefit of using SCAD penalty as discussed in Section 3.2.2. If the underlying true model

is a non-piecewise Log ACD model, the SCAD penalty monotonically shrinks non-zero values to zero. However, if the true underlying model is a piecewise Log ACD model, we are able to observe turning points in the online trace plots where the model structure changes.

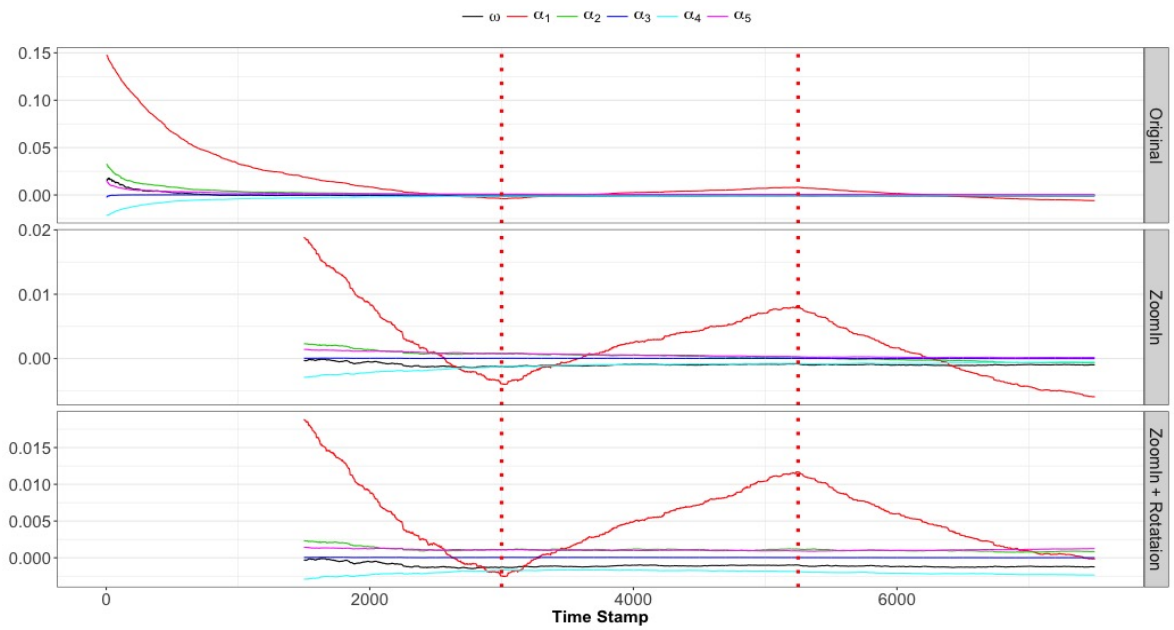

Figure 19: Trace plots of parameters estimated from both PEF and non-penalized EF approaches. (a) estimates using two methods for the first set of simulated durations. (b) estimates using two methods for the second set of simulated durations. Red dashed lines represent the true structural break points.

We fit a Log ACD$(5, 0)$ to the second set of simulated durations using both PEF ($\lambda = 2$) and EF. Figure 19(b) displays the online trace plots of parameters from two approaches separately. The red dashed lines represent the two true structural break points. To show the structural break points clearly, we censored out the first 1500 estimates of each parameter. For PEF, we can clearly find two turning points which are the same as the true locations of structural break points. While for EF, there are no such turning points. In addition, the two turning points appear only in the trace plot of α_1 estimated from PEF approach. This is in line with

the construction of the 3-segments piecewise Log ACD model where only α_1 changes across 3 segments. The model orders can then be determined by checking the number of parameter components that have turning point(s).

In Figure 19(b), the two turning points may be visually identified in the trace plot of $\widehat{\alpha}_1$ estimated using PEF approach. Can we provide and validate an algorithm to find the exact locations of such turning points? An efficient way is to find the peaks and valleys of the corresponding trace plot. Before the introduction of our FindPeaks algorithm, it is worthwhile to check the characteristics of trace plots of estimated parameters. Obviously, the trace plot of estimated parameter from PEF is not smooth. If we implement an algorithm that can locate turning points or peaks and valleys, we may end up with many false local minimum/maximum. To avoid getting local peaks/valleys, we fit a locally scatterplots smoothing (LOESS) (Cleveland, Grosse, and Shyu, 1992) with degree 1. A virtue of LOESS is that it builds up a function that only contains the deterministic part of the trace plot by fitting simple models. In Figure 20, we demonstrate a segment of the trace plot of an estimated parameter calculated from PEF. The red line represents a fitted line derived from LOESS function in R. A peak is always above the smoothed line while a valley is always below the smoothed line. The blue arrow is the position of the true structural break point. Our goal then is to find the point on the black line that has the largest distance with its corresponding position on the red line.

Algorithm 1, referred to as FindPeaks procedure, summarizes our method of finding the structural break points. First, the FindPreaks procedure locates all the points that are above their corresponding fitted curve. Next, the threshold c (the unit of estimated standard variance of errors) increases until the size of the set of possible peak points (referred to as peak

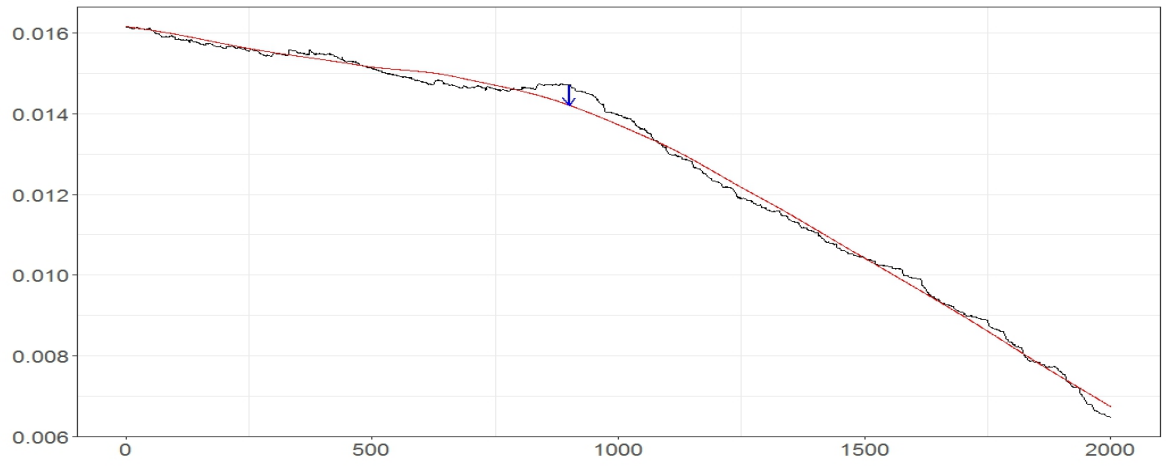

Figure 20: A segment of the trace plot of an estimated parameter calculated from PEF. Red line represent the fitted line derived from LOESS function in R. Blue arrow represents the position of true turning point.

Algorithm 1 FindPeaks

procedure FINDPEAKS($\boldsymbol{x}, b, \xi, \kappa, \eta$)
- $\boldsymbol{x} \leftarrow \boldsymbol{x}[-c(1:b)]$
- Initial $n = length(x), \kappa = nl$
- $\tilde{\boldsymbol{x}} := \{\tilde{x}_i = 1/3 \sum_{t=i}^{i+2} x_t, i = 1, 2, \cdots, n-2\}$
- $\boldsymbol{x}^* = \text{loess}(\text{x} \sim 1 : n\ , \xi)$
- $\mathcal{W} := \{i, \quad \text{such that} \quad \tilde{x}_i < \tilde{x}_{i+1}, \quad 2\kappa < i < n - 2\kappa\}$
- $m \leftarrow$ number of clusters in $\mathcal{W}$
- **for** i in $1 : m$ **do**
  - $c \leftarrow 0.1$, $d \leftarrow 1000$
  - **while** $d > \eta$ **do**
    - $\mathcal{T} \leftarrow NULL$
    - **for** j in $\mathcal{W}$ **do**
      - $\mu_1 \leftarrow mean(\sum_{j-\kappa}^{j+\kappa} x_i^*)$
      - $\mu_2 \leftarrow mean(\sum_{j-2\kappa}^{j-\kappa} x_i^* + \sum_{j+\kappa}^{j+2\kappa} x_i^*)$
      - **if** $\mu_1 - \mu_2 > c\hat{\sigma}$ **then**
        - $\mathcal{T} := augment(\mathcal{T}, j)$
    - $\mathcal{B} := bin(\mathcal{T})$
    - Update $\mathcal{W}$ by dropping the least frequent bin in $\mathcal{B}$
    - $c \leftarrow c + 0.1$, $d \leftarrow min(diff(\mathcal{B}))$
  - $\tau_i = argmax_t\{\hat{\varepsilon}_t, t \in \mathcal{T}\}$
- **return** $\{\tau_i, i = 1, 2, ..., m\}$

candidates) is within a predefined value η. Finally, the point is found such that it has the furthest distance with its corresponding fitted value (as showed in Figure 20) within the set it belongs to. A valley can be obtained the same way by multiplying -1 to all the estimated values.

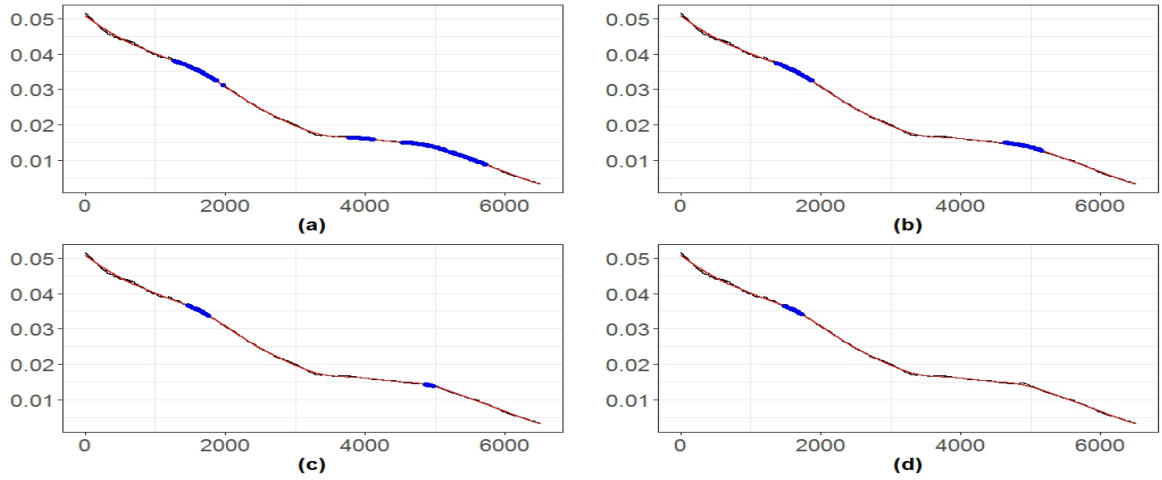

Figure 21: An illustration of our FindPeaks procedure. Black line is the trace plot of a parameter estimated using PEF. Red line represents the corresponding smoothed curves using LOESS. Blue points are the peak candidates. (a) - (d) are four snapshots during the iteration. $c = 0.1, 0.5, 1, 1.1$ for four steps respectively

Figure 21 are four snapshots of the iteration of FindPeaks procedure. Durations of length $n = 8000$ are simulated from (4.12). We fit a Log ACD$(5, 0)$ model with $\lambda = 2$ to the simulated durations. The black lines are the censored (first 1500 estimates are truncated) trace plot of $\widehat{\alpha}_1$ and the red lines are the corresponding smoothed curves obtained by using LOESS. Figure 21(a)-(d) are the four steps described above. Threshold values c are 0.1, 0.5, 1, 1.1 respectively in each step. At the first step, we identify almost all the points (see blue points in Figure 21(a)) that are above the red line. As seen in Figure 21, the number of clusters and the size of each cluster shrink when c increases. At the third snapshot, the size of the second cluster of peak

candidates is less than our predefined neighborhood size $\eta = 100$. All the points in the second cluster are saved and c continues to increase. At the fourth snapshot, the size of the last cluster of peak candidates is less than η, thus all the peak candidates are saved. Two estimated peaks are then obtained by identifying the furthest point away from the corresponding smoothed value in each candidate set.

We now elaborate on the choice of the control parameters ξ, κ, η, number of burn-in b and penalty parameter λ.

LOESS function in R requires parameter ξ which controls the degree of smoothing. If $0 < \xi < 1$, 100ξ percent of the total number of points will be used in the calculation of local fitted values. If $\xi > 1$, all points will be used. Since our objective is not to minimize the sum squared errors, a smaller value of ξ (e.g. 0.25) and a smoother curve will be sufficient. To avoid local minimum/maximum, a window size κ is needed. A peak is defined as the center of a cluster size 2κ such that its mean is higher than the mean of its two adjacent neighbors of size κ by $c\hat{\sigma}$, where c is the unit of $\hat{\sigma}$ and $\hat{\sigma}$ is the estimated standard errors of residuals between estimates and fitted values. Based on our simulation studies, we suggest $\kappa = 0.05$, i.e. 5% of the total length of estimates. η is the parameter that controls the purity of peak candidates. A too large η may give false alert while a too small η may introduce bias. Therefore, we suggest $\eta = 100$.

We have observed that the PEF approach takes a certain number of iterations to reach to a relative stable stage in Figure 19(a). Thus, censoring out some points will amplify the deterministic pattern of a trace plot and reduce false alert rate. Based on extensive simulation studies, we suggest that b should be within 10% and 20% of the input series.

Penalty parameter λ controls the step size of each iteration. A larger value of λ generates a smoother curve, while a smaller value of λ creates a jagged curves and thus can amplify the pattern of turning points. In practice, investigators can always start with a relatively large number, say 2. If the turning points are not discernible, smaller values of λ can then be used, e.g. 1 or 0.5.

4.4 Simulation Study

In this section, we access the accuracy of our retrospective monitoring (here referred to as RM) procedure via the simulation results. Four sets of simulations are conducted. Different combinations of structural breaks are considered in each setup. As we described in Section 3.2.4, a time series which follows a Log ACD model can be transformed into an AR model with Non-Gaussian errors through logarithmic operation. In this way, the structural break detection in a piecewise Log ACD model can be viewed as the structural break detection in a piecewise AR model. Assuming the data is given, the four scenarios show that the RM procedure is comparable to the results of Chan et al. (2014) (referred to as TSGL) in terms of bias and standard errors of estimates. Before the comparison of RM and TSGL, we set up their tuning parameters. For TSGL, $K = 10$ (maximum number of possible structural break points) is used for all four setups. For RM, λ is set to be 2 for all scenarios. ξ, κ, η are $0.25, 0.05, 100$ respectively for all cases. Also, we set b as 1500, which indicates that we censored out the first 20% data points for each series.

4.4.1 Change in α_1 Only

In scenario one, we look for a change in α_1 of a piecewise Log ACD$(1,0)$ model and the existence of a single structural break point. The time series studied in this example is generated from the following 2-segment piecewise Log ACD model

$$\begin{aligned} x_i &= \exp(\psi_i^{(j)})\varepsilon_i^{(j)}/\mu_\varepsilon^{(j)}, \\ \psi_i^{(j)} &= \begin{cases} 0.3 + 0.3\log x_{i-1}, & \text{if } 1 \le i \le 3750, \quad j = 1 \\ 0.3 - 0.2\log x_{i-1}, & \text{if } 3751 \le i \le 7500, \quad j = 2 \end{cases}, \end{aligned} \tag{4.9}$$

and error terms $\varepsilon_i^{(j)} \sim$ Weibull$(1.5, 2)$ for both segments. In the example, 200 realizations are generated from (4.9). One of the realizations are shown in Figure 22(a). We fit a Log ACD$(5,0)$ model to each realization and the trace plots of estimates for one realization is shown in Figure 22(b). We also apply TSGL procedure reported in Chan et al. (2014) on the logarithmic simulated data and obtain the estimated number of structural breaks and their corresponding locations.

It is clearly seen from Figure 22(b) that there is a turning point on the trace plot of $\widehat{\alpha}_1$ but no turning points are noticeable on other trace plots. Then we identify the structural break point via our FindPeaks algorithm. The percentage (%) of estimated number of structural breaks, the mean and standard errors of estimates of each location are recorded. Note that both procedures successfully identify all the structural breaks. Our RM procedure is comparable to TSGL in terms of mean (0.496 vs 0.499) and standard errors (0.017 vs 0.002).

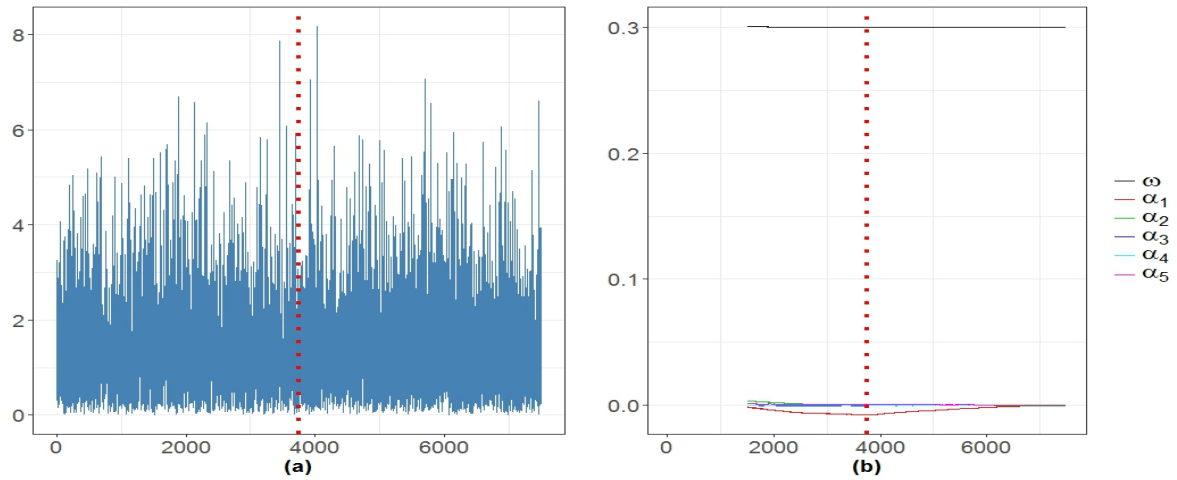

Figure 22: (a) One realization generated from a piecewise Log ACD model (4.9). (b) The trace plots of all estimates from PEF. Red dashed line is the true break point.

4.4.2 Change in Error Distribution Only

This example compares the performance between RM and TSGL procedure for the following piecewise Log ACD model

$$\begin{aligned} x_i &= \exp(\psi_i^{(j)})\varepsilon_i^{(j)}/\mu_\varepsilon^{(j)}, \\ \psi_i^{(j)} &= \begin{cases} 0.2 + 0.5\log x_{i-1}, & \text{if } 1 \le i \le 3750, \quad j = 1 \\ 0.2 + 0.5\log x_{i-1}, & \text{if } 3751 \le i \le 7500, \quad j = 2 \end{cases}, \end{aligned} \tag{4.10}$$

error terms $\varepsilon_i^{(1)} \sim \text{Weibull}(1, 0.5)$ and $\varepsilon_i^{(2)} \sim \text{Weibull}(2, 0.6)$. Similar to Section 4.4.1, 200 realizations are generated from (4.10). We fit a Log ACD$(5, 0)$ model to each realization and apply TSGL procedure on the logarithmic simulated data. One of the realizations and trace plots of estimates calculated from PEF are shown in Figure 23. Obviously, one turning point occurs on the trace plots of $\widehat{\omega}, \widehat{\alpha}_1$, and $\widehat{\alpha}_5$. That is to say, structural break in error distributions are reflected in the change of model parameters. Although changing only in error

distributions may be impractical in real application, this extreme example still confirms the power of our monitoring procedure. We then detect the location of structural break using FindPeaks procedure. Both procedures correctly estimate the number and the location of structural breaks. Our RM procedure performs equivalent to TSGL in terms of mean (0.479 vs 0.499) and standard errors (0.050 vs 0.002).

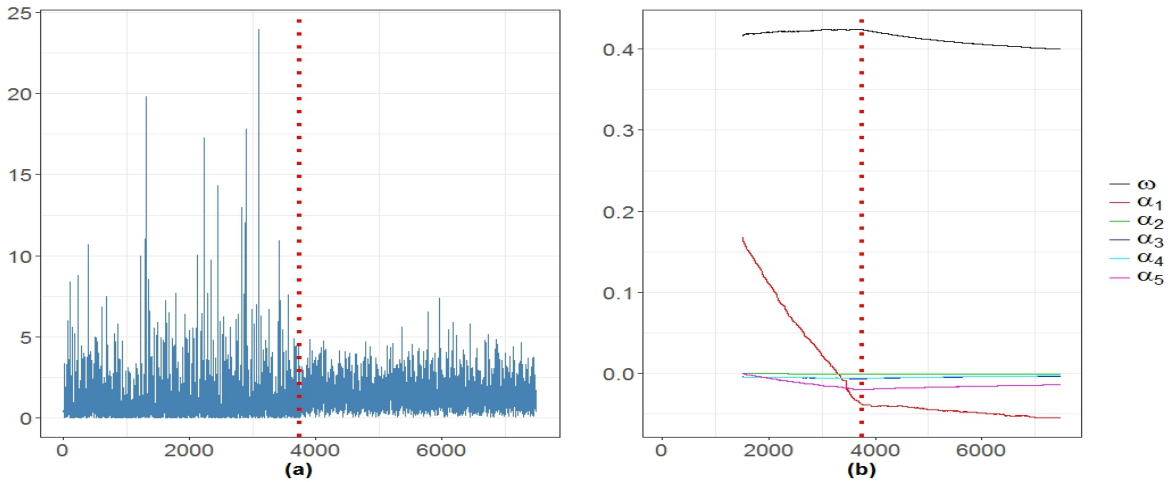

Figure 23: (a) One realization generated from a piecewise Log ACD model (4.10). (b) The trace plots of all estimates from PEF. Red dashed line is the true break point.

4.4.3 Change in Model Order and Mean Shift

In this example, we consider a combination of model order change and mean shift. The time series we study in this section is simulated from the following 2-segment piecewise Log ACD model

$$\begin{aligned} x_i &= \exp(\psi_i^{(j)})\varepsilon_i^{(j)}/\mu_\varepsilon^{(j)}, \\ \psi_i^{(j)} &= \begin{cases} 0.1 + 0.3\log x_{i-1} + 0.2\log x_{i-2}, & \text{if } 1 \le i \le 3750, \quad j = 1 \\ 0.5 - 0.2\log x_{i-1}, & \text{if } 3751 \le i \le 7500, \quad j = 2 \end{cases}, \end{aligned} \tag{4.11}$$

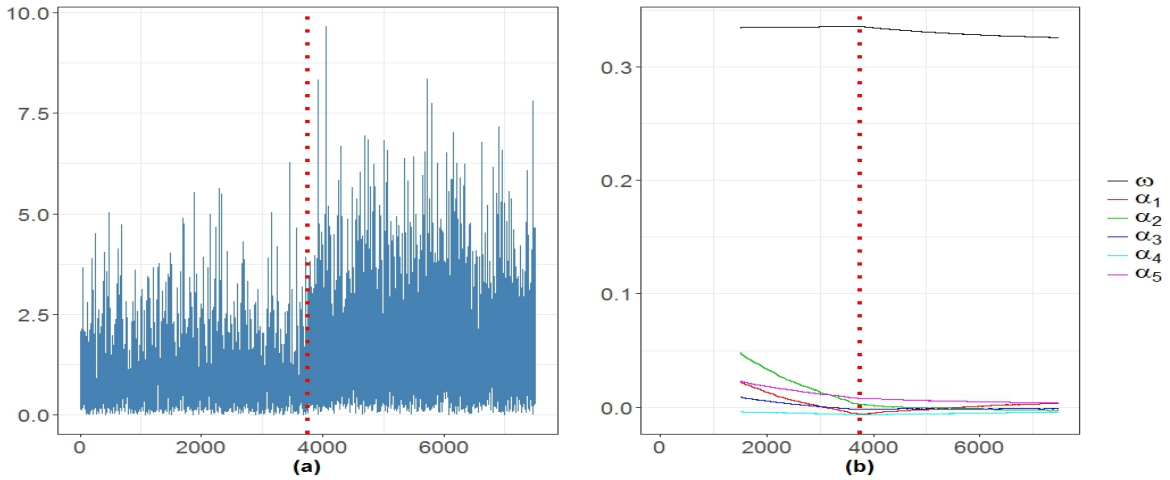

Figure 24: (a) One realization generated from a piecewise Log ACD model (4.11). (b) The trace plots of all estimates from PEF. Red dashed line is the true break point.

and error terms $\varepsilon_i^{(j)} \sim$ Weibull$(1.5, 2)$ for both segments. Similar to previous two examples, 200 realizations are generated from (4.11) and one of them is illustrated in Figure 24(a). Figure 24(b) displays the trace plots of estimates derived from a Log ACD$(5, 0)$ model using PEF. For three trace plots (e.g. $\widehat{\omega}_1, \widehat{\alpha}_1, \widehat{\alpha}_2$) in Figure 24(b), each of them has one turning point. Therefore, the change in model order and mean shift are captured by our PEF approach. To compare the performance of our RM procedure with the performance of an offline method, we use TSGL approach to locate structural breaks for each simulated data set. Again, the accuracy rates for both procedures are 100%.

4.4.4 A Combination of Structural Breaks

We consider a complex yet practical situation in this example. A time series of length 8000 are simulated from the following 4-segment piecewise LogACD model

$$
\begin{aligned}
x_i &= \exp(\psi_i^{(j)})\varepsilon_i^{(j)}/\mu_\varepsilon^{(j)}, \\
\psi_i^{(j)} &= \begin{cases} 0.5 - 0.3\log x_{i-1}, & \text{if } 1 \le i \le 3200, \quad j = 1 \\ 0.3 + 0.3\log x_{i-1}, & \text{if } 3201 \le i \le 4800, \quad j = 2 \\ 0.5 - 0.2\log x_{i-1}, & \text{if } 4801 \le i \le 6400, \quad j = 3 \\ 0.2 + 0.4\log x_{i-1}, & \text{if } 6401 \le i \le 8000, \quad j = 4 \end{cases},
\end{aligned} \tag{4.12}
$$

and error terms for the first and third segments are Gamma$(3, 4)$ while error terms follow a Gamma$(2, 5)$ distribution for the second and fourth segment. We fit a Log ACD$(5, 0)$ model to

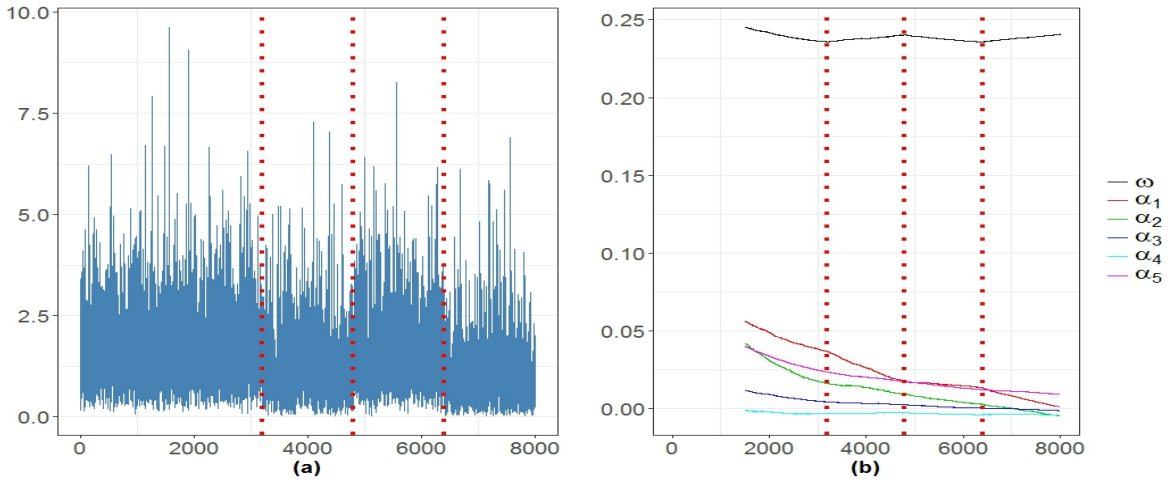

Figure 25: (a) One realization generated from a piecewise Log ACD model (4.12). (b) The trace plots of all estimates from PEF. Red dashed line is the true break point.

the simulated data to obtain the estimates of parameters and we repeat this process 200 times. Figure 25(a) shows one of 200 realizations of simulated data and Figure 25(b) shows the trace plots of estimates using PEF. From the trace plots, we can observe three turning points on $\widehat{\omega}$ and $\widehat{\alpha}_1$.

The estimation results for RM are summarized in Table 7, in which the performance of TSGL are compared. Note that RM procedure correctly identifies all three structural breaks

and TSGL procedure returns 2 structural breaks 6% of the time and 3 structural breaks 94% of the time. The TSGL procedure performs slightly better in terms of mean and standard errors though.

Table 7: Estimated structural break points using RM and TSGL procedures.

Estimated Number	RM			TSGL		
of Break Points	%	Mean	SE	%	Mean	SE
2	0			6	0.398	0.004
					0.802	0.004
3	100	0.401	0.015	94	0.400	0.001
		0.579	0.060		0.598	0.005
		0.780	0.060		0.800	0.001

Assuming all the online data is given in one time, we have demonstrated that RM procedure is comparable to the exceptional offline procedure proposed by Chan et al. (2014).

4.5 Real Application

We validate our retrospective monitoring procedure via three real examples in this section. For each example, we fit a Log ACD$(10, 0)$ model using PEF approach with $\lambda = 2$.

4.5.1 BAC for A Week

As we discussed in Section 3.4.1, estimates from non-penalized estimating functions for BAC during the week of June 10, 2013 vary a lot. For Mon. June 10, Wed. June 12 and Thur. June 13, Log ACD$(1, 0)$ were selected as the best model. Although Log ACD$(1, 0)$ was the best fit for Tue. June 11, the estimates are different from the other three days. For Fri. June 14, a

Log ACD$(2, 2)$ was chosen by the non-penalized estimating functions approach. Due to this discrepancy, we want to investigate the existence of structural breaks. Assuming durations for the whole week is given, Figure 26 displays the combined durations for BAC during the week of June 10, 2013. Black dashed lines represent the boundaries between different days. Red dashed

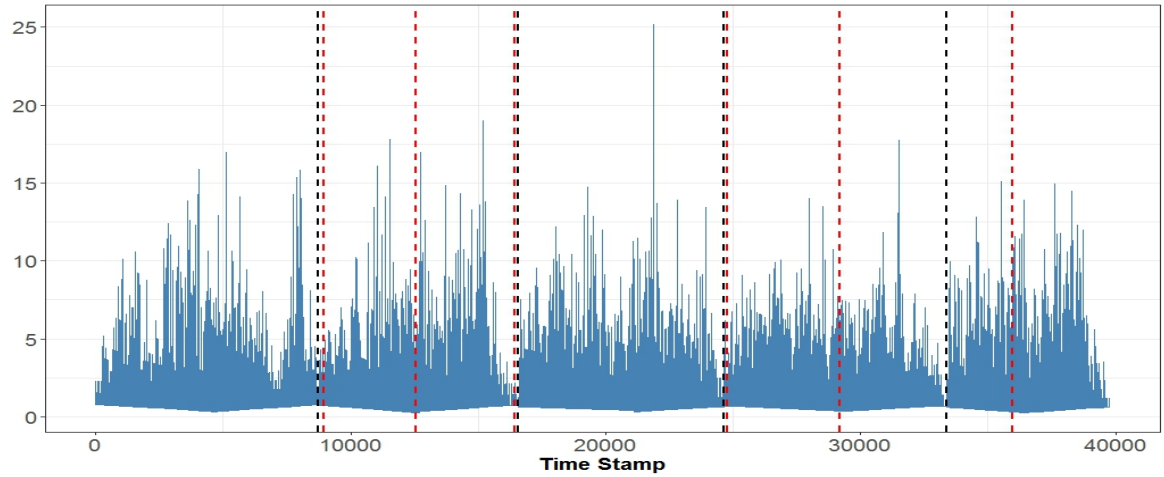

Figure 26: Combined durations for BAC during the week of June 10, 2013. Black dashed lines represent the boundaries between different days. Red dashed lines represent the estimated locations of structural breaks

lines represent the estimated locations of structural breaks using PEF approach. Noted that the first three boundaries are correctly identified while the fourth estimated boundary departs from the real value. Also, two more estimated structural breaks (the second and fifth red dashed lines) are found. The two structural breaks are within Tue. June 11 and Thur. June 13 respectively. Instead of being two false alarms, news or analytical reports were announced on both days. On Tue. June 11 2013, PRNewswire (2013c) reported the news below

> *Company Director, R. David Yost, purchased his share at price of US $11.51, which was almost 15.55% return by the end of last trading session. Investors may want to find out how Bank of America insiders like CEOs, CFOs and Directors*

are thinking about the future of the company.

And on Thur. June 13 2013, PRNewswire (2013a) reported the news below

> *By the end of last trading session, the share rose about 12.5% this year. As the US banking industry experienced a strong first quarter with record profits, the market will consider it as a recovery signal after 2008 financial crisis.*

Given the above news reports, the two extra estimated structural breaks may indicate the underlying change of the market. A further investigation may be needed.

4.5.2 BAC for Three Days

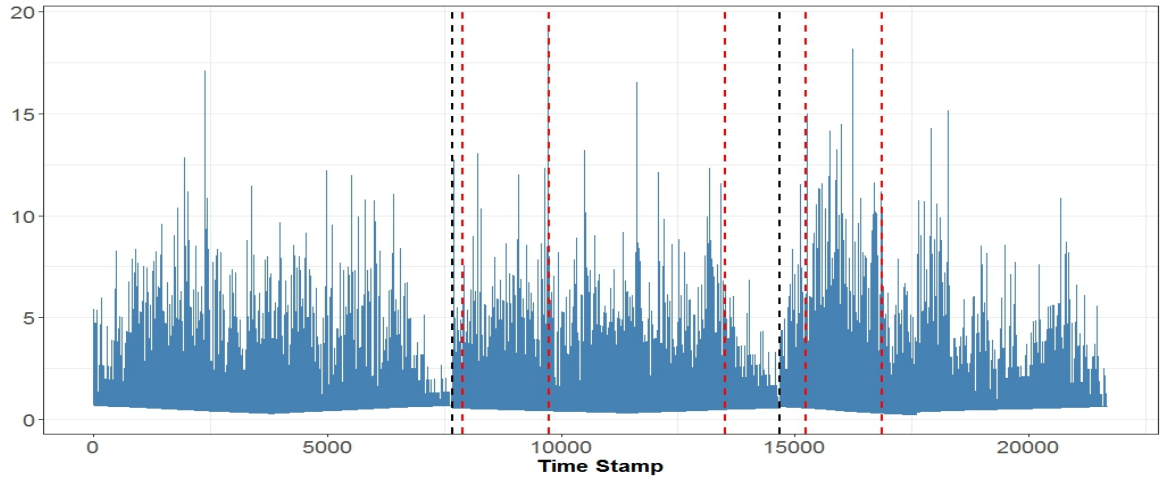

Figure 27: Combined durations for BAC during the first three days of week June 17, 2013. Black dashed lines represent the boundaries between different days. Red dashed lines represent the estimated locations of structural breaks

In this example, we examine the changes occurred for BAC in three consecutive days (Mon. Jun 17, Tue. Jun 18 and Wed. Jun 19). Abnormal Log ACD$(1,0)$ were selected for the first and third days while Log ACD$(3,3)$ were the best fit for the second day. Figure 27 displays

the combined durations for three days. Boundaries for different days and estimated locations are marked as black dashed lines and red dashed lines respectively. The two actual boundaries are correctly identified with a reasonable difference. There are three extra estimated locations of structural breaks. The second and third dashed lines occur in Tue. June 18. Since Log ACD$(3, 3)$ was the best model for that day, two structural breaks explains the underlying change of the stock behavior. We found that on Tue. June 18, PRNewswire (2013b) reported the following news.

> *Company Director, Charles K. Gifford, sold his shares at price of US$13.07 for about US$7.59 million on June 14. Investors may want to find out how Bank of America insiders like CEOs, CFOs and Directors are thinking about the future of the company.*

Since stock market changes rapidly, any kinds of insider information can greatly influence the pattern of stock behavior. We didn't find any news regarding Bank of America published on Wed. June 19. But the fifth estimated location of structural breaks divides Wed. June 19 into two halves. Long durations (frequent transactions) occurred in the morning while relatively short durations (less frequent transactions) occurred in the afternoon.

4.5.3 BAC for Two Days

The third example examines the performance of retrospective monitoring procedure when there is only one change in models between two consecutive days. In Table 2, we observe approximately the same estimates for Wed. June 5 and Thur. June 6. Similarly, we use black dashed lines to represent the boundaries between days and red dashed lines represent the

estimated locations of structural breaks. As clearly seen in Figure 28, the estimated boundary is not close to the actual boundary. But it is still a reasonable structural change since no

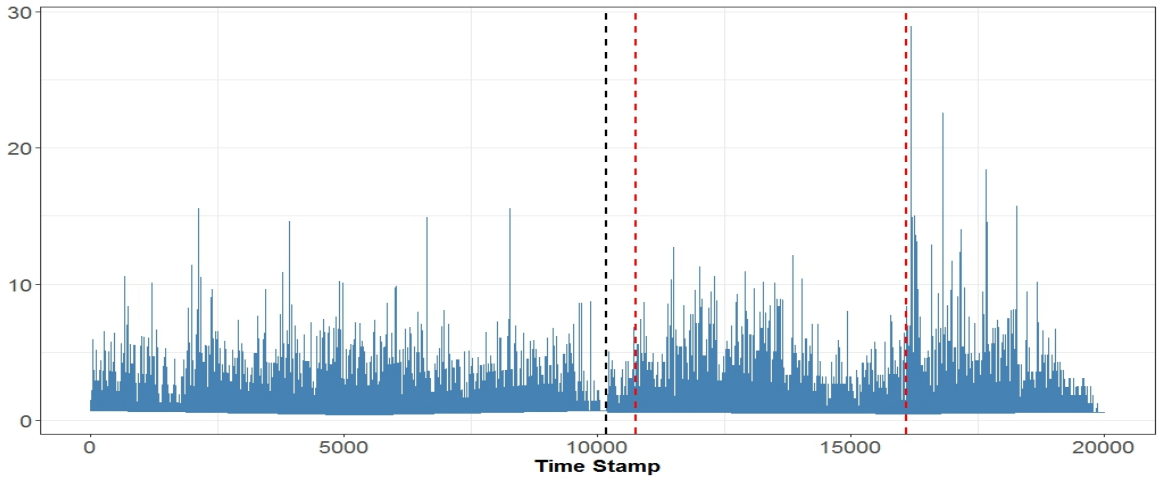

Figure 28: Combined durations for BAC during Wed. June 5 and Thur. June 6, 2013. Black dashed lines represent the boundaries between different days. Red dashed lines represent the estimated locations of structural breaks. Note that the first red dashed line overlaps with the first black dashed line.

seemingly change in pattern occurs before the first estimated boundary. The extra estimated structural breaks is within Thur. June 6. Although no any related news were reported on that day, dividing the durations of the second day into two clusters (short vs long) are reasonable since the mean of durations before the second red line is relatively smaller than the mean of durations after red line (as seen in Figure 28).

4.6 A Review of Online Methods

Another line of research is the sequential or online process. This is in line with the original interest in problems of statistical control (Page, 1954). For example, when manufacturing a product, one monitors and analyzes the quality of products and the goal is to detect a change

as soon as possible. Researchers, in recent years, have been trying to apply these ideas to other areas, notably economics and finance.

Chu, Stinchcombe, and White (1996) first developed two sequential tests of structural stability for online monitoring of economic behavior modeled using *linear* regression models. A fluctuation test (FL) based on recursive estimates of parameters from a linear regression model and a CUSUM test based on recursive estimated residuals were employed. The general paradigm requires an initial time period of length m and choice of a stopping rule, and the goal of structural break detection is achieved via monitoring for parameter changes online. Since then, several studies had extended and refined this procedure in other directions. Berkes, Gombay, Horváth, and Kokoszka (2004) proposed a different test of structural stability to monitor for change of parameter from a GARCH(p, q) process. Since squared residuals did not satisfy one of the conditions given by Chu et al. (1996), their procedure relied more directly on the quasi-likelihood function. Hušková, Prášková, and Steinebach (2007) studied the online monitoring procedure that were based on partial sums of weighted residuals for parameter changes in autoregressive models. Aue, Horváth, and Reimherr (2009) proposed a cumulative sum (CUSUM)-type statistic to monitor change in parameters in multiple time series *linear* models. The authors verified the properties of their methods via extensive simulation studies and on a real macroeconomic data set. Aue et al. (2012) applied similar monitoring approach to detect change in parameters in a functional capital asset pricing model. The aforementioned literature list is not meant to be complete but is capable of showing a tendency to the problem of online monitoring for structural breaks in financial time series. For a recent review of sequential test of structural breaks and some extensions, see, e.g. Horváth and Rice (2014).

Next, we review the work of Chu et al. (1996) and Berkes et al. (2004) in details since both methods are the most relevant ones to our work. Chu et al. (1996) first proposed the fundamental framework and Berkes et al. (2004) extended Chu et al. (1996)'s work to GARCH models.

To illustrate Chu et al. (1996)'s method, we consider the linear model

$$Y_i = \boldsymbol{X_i'}\boldsymbol{\beta_i} + \varepsilon_i, \quad i \geq 1, \tag{4.13}$$

where $\boldsymbol{X_i}$ is a $k-$dimensional random row vector and $\boldsymbol{\beta_i}$ is the a $k-$dimensional nonstochastic parameter row vector. The training sample size of m is assumed to be "noncontaminated", i.e. $\boldsymbol{\beta_1} = \boldsymbol{\beta_2} \cdots = \boldsymbol{\beta_m}$. The null hypothesis describes structural stability of $\boldsymbol{\beta}$, while the alternative contains one structural break point τ that is greater than m. A monitoring scheme is a stopping rule which depends on a detecting statistic (detector) Γ_τ and a threshold $g(m, \tau)$, according to $\tau_g(\Gamma_\tau) \equiv min\{\tau \geq m, \Gamma_\tau > g(m, \tau)\}$.

A CUSUM detector which is based on the recursive residuals is defined as follows. Let $\widehat{\boldsymbol{\beta}}_\tau$ denote the least square estimates calculated from (4.13) at time τ. Define recursive residuals as $w_k = 0$ and $w_\tau = \hat{\varepsilon}_\tau / \nu_\tau^{1/2}$, $\hat{\varepsilon}_\tau = Y_\tau - \boldsymbol{X_\tau'}\widehat{\boldsymbol{\beta}}_\tau$, $\quad m+1 \leq \tau < \infty$, and $\nu_\tau^{1/2}$ is the standard error of $\hat{\varepsilon}_\tau$. The τth cumulated sum of recursive residuals is then given as $Q_m(\tau) = \hat{\sigma}^{-1/2} \sum_{t=m}^{m+\tau} w_t$, where $\widehat{\sigma}$ is the estimator for the common variance of the ε_m's computed from the training sample.

The idea of Chu et al. (1996)'s procedure can be summarized as follows. Let the standard

Wiener process be $W(.)$. If it can be shown that, as $m \to \infty$,

$$m^{-1/2} Q_m(\tau) \to W(\tau), \quad \tau \in (0, \infty) \tag{4.14}$$

and a boundary function can be found such that

$$\lim_{m\to\infty} P\left(|Q_m(\tau)| \geq m^{1/2} g(m,\tau) \quad \text{for some} \quad \tau \geq m\right) = P\left(|W(\tau)| \geq g(\tau) \quad \text{for some} \quad \tau \geq 1\right), \tag{4.15}$$

provided the sequence $Q_m(\tau)$ satisfies a functional central limit theorem. The probability on the right-hand side of (4.15) can be computed analytically, once several sensible choices of $g(\cdot)$ were made. As a result, a monitoring procedure can be developed such that as $m \to \infty$, the decision function $|Q_\tau^m|$ crosses the boundary $m^{1/2} g(m,\tau)$ at some future time τ with a predefined probability α, under the null hypothesis.

However, Berkes et al. (2004) pointed out that due to the presence of extra terms in the covariance structure, the squares of residuals from a GARCHp,q) process do not meet a functional limit theorem with the Weiner process, making it impossible to apply Chu et al. (1996)'s procedure in GARCH process. Instead, they put forward a different approach that relies on the quasi-likelihood functions, rather than that uses model residuals. Denote $\widehat{l}_i'(\boldsymbol{\beta})$ the $(p+q+1)-$dimensional row vector of partial derivatives of the conditional quasi-likelihood of $Y_i \sim$GARCH(p,q) with respect to the $\boldsymbol{\beta}$ and consider the matrix

$$\widehat{\boldsymbol{D}}_m = \frac{1}{m} \sum_{1<i\leq m} \left(\widehat{l}_i'(\widehat{\boldsymbol{\beta}}_m)\right)' \left(\widehat{l}_i'(\widehat{\boldsymbol{\beta}}_m)\right), \tag{4.16}$$

where $\widehat{\boldsymbol{\beta}}_m$ is the quasi-maximum likelihood parameter estimate. Then they constructed a $(p+q+1)-$dimensional detector

$$G_m(\tau) = \sum_{m<i\leq m+\tau} \left(\widehat{l}_i'(\widehat{\boldsymbol{\beta}})\right) \widehat{\boldsymbol{D}}_m^{-1/2}, \quad \tau \in [0,\infty), \tag{4.17}$$

and a stopping time τ_m

$$\tau_m = min\{\tau : |G_m(\tau)| > m^{1/2}(1+\tau/m)g(m,\tau)\} \tag{4.18}$$

4.7 Monitoring for Change of Parameter from Penalized EFs

In high-frequency financial data analysis, online procedures seem to be more useful since investigators have to make decisions promptly. The aforementioned online methods either require independent and identically distributed regressors in *linear* regression models or predefined model orders of a *nonlinear* model. The procedure we proposed overcomes these disadvantages by monitoring change in parameters estimated from PEF. That is, our procedure can detect structural breaks and choose model orders simultaneously.

Motivated by the work of Berkes et al. (2004), we propose a monitoring scheme for Log ACD models. Suppose we have observed a set of durations $x_1, x_2, \cdots, x_n$. We assume that

$$x_i = \exp(\psi_i)\varepsilon_i/\mu_\varepsilon, \quad i = 1,2,...,n \tag{4.19}$$

and under structural stability null hypothesis

$$H_0: \quad \psi_i = \omega + \sum_{k=1}^{p} \alpha_k \log x_{i-k} + \sum_{k=1}^{q} \beta_k \psi_{i-k}, \quad 1 \leq i < n \tag{4.20}$$

Let $\boldsymbol{\theta} = (\omega, \alpha_1, ...\alpha_p, \beta_1, ...\beta_q)$ be the parameter of the Log ACD(p, q) model. Under the alternative that one structural break point occurs at time τ.

$$H_A: \quad \psi_i \quad = \quad \begin{cases} \omega + \sum_{k=1}^{p} \alpha_k \log x_{i-k} + \sum_{k=1}^{q} \beta_k \psi_{i-k}, & 1 \leq i < \tau \\ \omega^* + \sum_{k=1}^{p^*} \alpha_k^* \log x_{i-k} + \sum_{k=1}^{q^*} \beta_k^* \psi_{i-k}, & \tau \leq i \leq n \end{cases} \tag{4.21}$$

That is, after time τ, the model order and parameters change to (p^*, q^*) and $\boldsymbol{\theta}^* = (\omega^*, \alpha_1^*, ...\alpha_{p^*}^*, \beta_1^*, ...\beta_{q^*}^*)$ respectively. As we described in Section 3.2.3, the penalized estimates for the unknown parameter based on these data can be obtained by solving the penalized estimating function

$$g_{C,\lambda}^*(\boldsymbol{\theta}) = \sum_{i=1}^{n} g_{i,\lambda}^*(\boldsymbol{\theta}) = \sum_{i=1}^{n} (\mathbf{a}_{i-1}^*(\boldsymbol{\theta}) m_i(\boldsymbol{\theta}) + \mathbf{b}_{i-1}^*(\boldsymbol{\theta}) M_i(\boldsymbol{\theta}) - p_\lambda'(|\boldsymbol{\theta}|)),$$

where each term of $g_{i,\lambda}^*(\boldsymbol{\theta})$ is given in Section 3.2.1. Godambe (1985) shown that the derivative of the quasi-(log)likelihood function is the optimal combination of elementary orthogonal estimating functions in the case of discrete time stochastic process. Thus, we can rewrite (4.16) as

$$\widehat{\boldsymbol{D}}^{(m_1,m_2)} = \frac{1}{m_2 - m_1} \sum_{m_1 < i \leq m_2} \left(g_{i,\lambda}^*(\widehat{\boldsymbol{\theta}}_{m_2}) \right) \left(g_{i,\lambda}^*(\widehat{\boldsymbol{\theta}}_{m_2}) \right)', \tag{4.22}$$

where m_1 is the number of burn-in observations, m_2 is the size of training samples and $\widehat{\boldsymbol{\theta}}_{m_2}$

is the penalized estimates at iteration m_2. Recall that PEF approach takes a certain number of iterations to reach to a relative stable stage (described in Section 4.3). Since the essential idea of our monitoring procedure is the deviation of the updated parameter $\widehat{\boldsymbol{\theta}}_\tau$ at step τ from the historical estimate $\widehat{\boldsymbol{\theta}}_{m_2}$, truncating the fluctuation stage of recursions will remove the discrepancy between sample/historical observations and future data. Other definitions for $\widehat{\boldsymbol{D}}^{(m_1,m_2)}$ will be investigated in future work.

As the monitoring starts at m_2+1, we define our detector for each parameter $\theta_j, j = 1,2,...,d = p+q+1$

$$G_j^{m_2}(\tau) = \sum_{m_2<i\leq m_2+\tau} \left(g^*_{i,\lambda}(\widehat{\boldsymbol{\theta}}_{m_2})_j\right) / \sqrt{\widehat{\boldsymbol{D}}_{j,j}^{(m_1,m_2)}}, \quad \tau \in [0,\infty). \tag{4.23}$$

Based on the construction of $g^*_{i,\lambda}(\boldsymbol{\theta})$, $G_j^{m_2}(\tau)$ is essentially the sum of increment in the standardized $\widehat{\theta}_j$. Due to the consistency and asymptotic normality properties of estimating functions (Heyde, 2008, Ch. 12, p. 180 -196), $(m_2-m_1)^{-1/2}\boldsymbol{G}_j^{m_2}(\tau)$ can be approximated by $\left((1+\tau)W_j\left(\frac{\tau}{1+\tau}\right)\right), \quad \tau \in [0,\infty)$ under null hypothesis, where $W_j(\cdot)$, $j = 1,2,...,d$ are independent standard Wiener process. We now define the stopping time τ_{j,m_2} for each θ_j, $j = 1,...,d$ as

$$\tau_{j,m_2} = min\{\tau_j : |G_j^{m_2}(\tau)| > (m_2-m_1)^{1/2}(1+\tau_j/(m_2-m_1))g(\tau_j,m_2)\}. \tag{4.24}$$

If $\tau_{j,m_2} < \infty$, we consider that a structural break occurs for parameter θ_j. Boundary $g(\tau_j, m_2)$ is chosen such that

$$\lim_{m_2\to\infty} P_{H_0}\{\tau_{j,m_2} < \infty\} = \alpha$$

where $0 < \alpha < 1$ is the controlled type I error and

$$\lim_{m_2 \to \infty} P_{H_A}\{\tau_{j,m_2} < \infty\} = 1$$

We follow Berkes et al. (2004)'s proposal that a constant boundary function $g(\tau_j, m_2) = c$ is sufficient. Based on the following well-know formula for the distribution function of $\sup_{0 \le t \le 1} |W(t)|$ (Csörgo and Révész, 2014), we can establish the relation between critical value c and type I error α.

$$\begin{aligned} \lim_{m_2 \to \infty} P_{H_0}\{\tau_{j,m_2} < \infty\} &= 1 - P\{\sup_{0 \le t \le 1} |W(t)| \le c\} \\ &= 1 - \left\{ \frac{4}{\pi} \sum_{0 \le k \le \infty} \frac{(-1)^k}{2k+1} \exp\left(-\frac{\pi^2 (2k+1)^2}{8c^2} \right) \right\} \end{aligned} \tag{4.25}$$

Equation (4.25) calculates the critical value c for different significance level α.

4.8 Simulation Studies

In this section, we report the results of simulation studies, which aim to assess the performance of our monitoring process. For each model (as described in (4.26) and (4.27) - (4.29)), we simulate data of length $n = 7500$ and replication time is 500. Then we fit a Log ACD$(5, 0)$ model to each realization. Estimates for each realization are obtained by using PEF with $\lambda = 2$. In Section 4.8.1, we monitor the change in parameter and record the probabilities of rejecting the null hypothesis for each parameter at different future time τ. In Section 4.8.2, we conduct the power analysis for detecting structural breaks.

4.8.1 Empirical Type I Errors

In tis section, we consider the following four scenarios.

$$\begin{aligned}
&\text{Scenario 1} \quad : \quad \omega = 0.3, \alpha = 0.1, \varepsilon \sim \text{Weibull}(1.5, 2), m_1 = 1500, m_2 = 2000 \\
&\text{Scenario 2} \quad : \quad \omega = 0.3, \alpha = 0.1, \varepsilon \sim \text{Weibull}(1.5, 2), m_1 = 1500, m_2 = 2500 \\
&\text{Scenario 3} \quad : \quad \omega = 0.1, \alpha_1 = 0.3, \alpha_2 = 0.1, \varepsilon \sim \text{Gamma}(0.8, 1), m_1 = 1500, m_2 = 2000 \\
&\text{Scenario 4} \quad : \quad \omega = 0.1, \alpha_1 = 0.3, \alpha_2 = 0.1, \varepsilon \sim \text{Gamma}(0.8, 1), m_1 = 1500, m_2 = 2500
\end{aligned} \tag{4.26}$$

In Table 8 we report the empirical rejection probabilities of the null hypothesis of structural stability in the model parameters assuming the hypothesis is true (i.e. the empirical type I error). It is seen that our procedure can identify the correct model order since no significant rejecting probabilities for non-zero parameters occur. The empirical type I errors for parameter α_3 to α_5 are almost all 0. There are few cases that the rejection probabilities are larger than 0.05 (when τ is fairly large though). We note that the theory we proposed in this chapter shows that the the empirical type I errors tends to be less than 0.05 as $m_2 \to \infty$. Thus for any finite m_2, distortion will be present. This is particularly visible if the Log ACD parameters are difficult to estimate, as in Scenario 3 and Scenario 4 (Gamma$(0.8, 1)$ is skewer than Weibull$(1.5, 2)$, thus add more fluctuated noise); the monitoring procedure has a high probability of Type I error. We suggest that in such situations a m_2 larger than 1000 would be required to obtain empirical type I error close to the nominal value.

Table 8: Empirical rejecting probabilities for Scenario 1 - 4.

Parm/ τ	100	200	300	400	500	600	700	800	900	1000
Scenario 1: $\omega = 0.3, \alpha = 0.1, \varepsilon \sim$ Weibull$(1.5, 2)$ $\boldsymbol{m_1 = 1500, m_2 = 2000}$										
ω	0	0	0	0	0	0	0	0	0	0
α_1	0	0	0	0	0	0	0	0	0	0
α_2	0	0	0	0	0	0	0	0	0	0
α_3	0	0	0	0	0	0	0	0	0	0
α_4	0	0	0	0	0	0	0	0	0	0
α_5	0	0	0	0	0	0	0	0	0	0
Scenario 2: $\omega = 0.3, \alpha = 0.1, \varepsilon \sim$ Weibull$(1.5, 2)$ $\boldsymbol{m_1 = 1500, m_2 = 2500}$										
ω	0	0	0	0	0	0	0	0	0	0
α_1	0	0	0	0	0	0	0	0	0	0
α_2	0	0	0	0	0	0	0	0	0	0
α_3	0	0	0	0	0	0	0	0	0	0
α_4	0	0	0	0	0	0	0	0	0	0
α_5	0	0	0	0	0	0	0	0	0	0
Scenario 3: $\omega = 0.1, \alpha_1 = 0.3, \alpha_2 = 0.1, \varepsilon \sim$ Gamma$(0.8, 1)$ $\boldsymbol{m_1 = 1500, m_2 = 2000}$										
ω	0	0	0	0.016	0.04	0.076	0.112	0.178	0.21	0.256
α_1	0	0	0	0	0	0	0	0	0	0
α_2	0	0	0	0	0	0	0	0	0	0
α_3	0	0	0	0	0	0	0	0	0	0
α_4	0	0	0	0	0	0	0	0	0	0
α_5	0	0	0	0	0	0	0	0.004	0.006	0.008
Scenario 4: $\omega = 0.1, \alpha_1 = 0.3, \alpha_2 = 0.1, \varepsilon \sim$ Gamma$(0.8, 1)$ $\boldsymbol{m_1 = 1500, m_2 = 2500}$										
ω	0	0	0	0.006	0.008	0.026	0.056	0.106	0.172	0.236
α_1	0	0	0	0	0	0	0	0	0	0
α_2	0	0	0	0	0	0	0	0	0	0
α_3	0	0	0	0	0	0	0	0	0	0
α_4	0	0	0	0	0	0	0	0	0	0
α_5	0	0	0	0	0	0	0.002	0.002	0.002	0.002

4.8.2 Power Analysis

We proceed with investigation in the power of our monitoring procedure. We consider 3 piecewise Log ACD models in this section. The stochastic parts for each model are described in (4.27) - (4.29). Based on the results described in Section 4.8.1, we use $m_1 = 1500$ and $m_2 = 2500$ in this simulation study. That is, the size of training sample is 1000.

$$\begin{aligned} \text{Scenario 5} : \psi_i &= \begin{cases} 0.2 + 0.1 \log x_{i-1} + 0.2 \log x_{i-2}, & 1 \le i < 3000 \\ 0.2 + 0.1 \log x_{i-1} + 0.5 \log x_{i-2}, & 3001 \le i \le 7500 \end{cases} \\ \varepsilon &\sim \text{Gamma}(2,1) \quad \text{for both segments} \end{aligned} \tag{4.27}$$

$$\begin{aligned} \text{Scenario 6} : \psi_i &= \begin{cases} 0.3 + 0.1 \log x_{i-1} + 0.4 \log x_{i-2}, & 1 \le i < 3000 \\ 0.3 + 0.3 \log x_{i-1} + 0.1 \log x_{i-2}, & 3001 \le i \le 7500 \end{cases} \\ \varepsilon &\sim \text{Weibull}(1.5,2) \quad \text{for both segments} \end{aligned} \tag{4.28}$$

$$\begin{aligned} \text{Scenario 7} : \psi_i &= \begin{cases} 0.2 + 0.1 \log x_{i-1} + 0.2 \log x_{i-2}, & 1 \le i < 3000 \\ 0.2 + 0.1 \log x_{i-1} + 0.3 \log x_{i-2}, & 3001 \le i \le 7500 \end{cases} \\ \varepsilon &\sim \text{Gamma}(2,1) \quad \text{for both segments} \end{aligned} \tag{4.29}$$

The empirical powers are summarized in Table 9, in which powers for each parameter at different future time τ are recorded. As can be expected, large changes in parameters are detected more reliably (Scenario 5 versus Scenario 7 and α_1 versus α_2 in Scenario 6). Except for a few false alarm (when τ is very large), our procedure can correctly identify the structure break for non-zero valued parameter.

Table 9: Empirical power for Scenario 5 - 7.

Parm/ τ	350	400	450	500	550	600	650	700	750	800
Scenario 5: model changes at $\tau = 500$										
ω	0	0	0	0	0	0	0	0	0	0
α_1	0	0	0	0	0	0.002	0.006	0.034	0.080	0.158
α_2	0	0	0	0	0.022	0.236	0.630	0.898	0.976	1
α_3	0	0	0	0	0	0	0	0	0	0.002
α_4	0	0	0	0	0	0	0	0	0	0
α_5	0	0	0	0	0	0.002	0.002	0.002	0.004	0.004
Scenario 6: model changes at $\tau = 500$										
ω	0	0	0	0	0.002	0.002	0.004	0.014	0.018	0.038
α_1	0	0	0.002	0.002	0.002	0.010	0.046	0.098	0.172	0.280
α_2	0.002	0.018	0.066	0.216	0.346	0.482	0.628	0.716	0.844	0.902
α_3	0	0	0	0	0	0	0	0	0	0
α_4	0	0	0	0	0	0	0	0	0	0
α_5	0	0	0	0	0.004	0.010	0.016	0.028	0.034	0.046
Scenario 7: model changes at $\tau = 500$										
ω	0	0	0	0	0	0	0	0	0	0
α_1	0	0	0	0	0	0	0	0	0	0.004
α_2	0	0	0	0	0	0.022	0.080	0.240	0.474	0.702
α_3	0	0	0	0	0	0	0	0	0	0
α_4	0	0	0	0	0	0	0	0	0	0
α_5	0	0	0	0	0	0	0	0	0	0

Understanding the distribution of the detection delay time when the decision function first exceeds a critical level can provide to practitioners insights in making prompt decisions over online data. In Table 10, we report the summary statistics for the delayed time for each scenario. As expected again, the more significant change a parameter experiences, the smaller their corresponding means of delayed time. In general, the delayed time range from around 100 to 400, depending on the effect size.

Table 10: Distribution of delayed time for Scenario 5 - 7.

Parm/ τ	Mean	SE	Min	Q1	Median	Q3	Max
Scenario 5:							
α_3	136.4	49.0	22.0	101.0	132.5	168.0	294.0
Scenario 6:							
α_2	382.9	136.8	-53.0	289.8	380.0	483.5	799.0
α_3	122.9	141.6	-180.0	13.8	104.5	212.2	816.0
Scenario 7:							
α_3	263.1	87.4	53.0	203.0	255.0	313.2	517.0

4.9 Monitoring Structural Break in Stock BAC

Before we finish this chapter, we describe two real examples. In example one, we combine two days of adjusted durations for BAC (Wed. June 5, 2013 and Thur. June 6, 2013) together. Note that from Table 2, $\widehat{\omega}$ and $\widehat{\alpha}_1$ on Wed. June 5 changes from 0.258 and 0.074 to 0.289 and 0.093 respectively. In example two, we combine Mon. June 10, 2013 and Tue. June 11, 2013. Although Log ACD$(1, 0)$ was selected as the best model for both dates, we know that Tue. June 11 has an abnormal pattern since the estimates of ω and α_1 are different from the other days. For each example, we fit a Log ACD$(20, 0)$ model with $\lambda = 0.5$. The estimates for each parameter are obtained via PEF. (m_1, m_2) are set to be (5000,6000) and (4000, 5000) respectively. Since the number of observed adjusted durations in the first example is larger than the the one in example two, we set a higher value of m_1 so that we can get rid of the fluctuation part of the recursion. m_2 is set according to the suggestion we made in previous section.

Figure 29 and 30 display the plots of sequences of the calculated detector for each parameter. In both figures, we use black dashed and red dashed lines to represent the true boundary

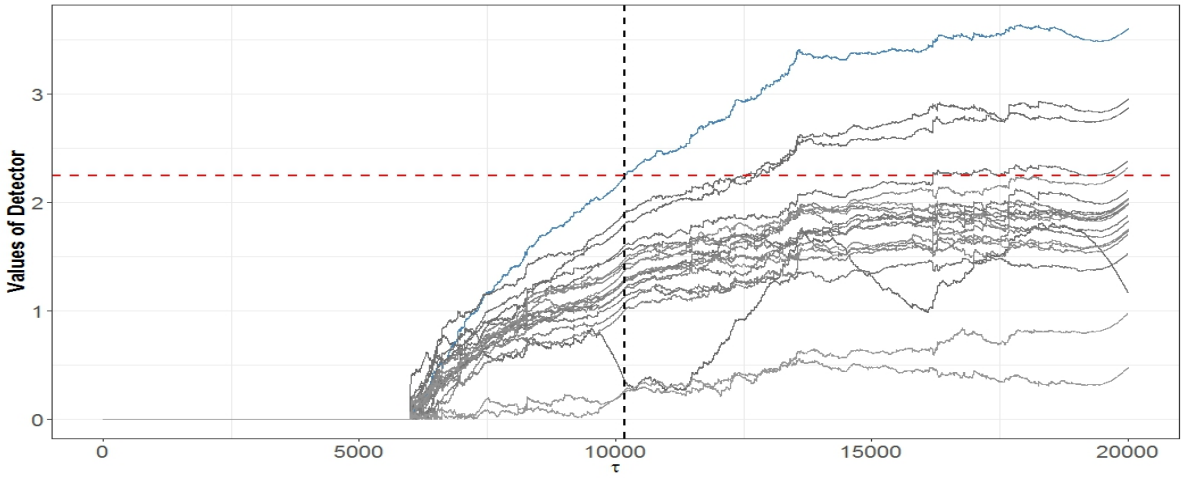

Figure 29: Plot of the sequences of detector for each parameter. Blue line represents the monitoring sequence for parameter α_1. Red dashed line is the true structural break point while red dashed line is the critical level.

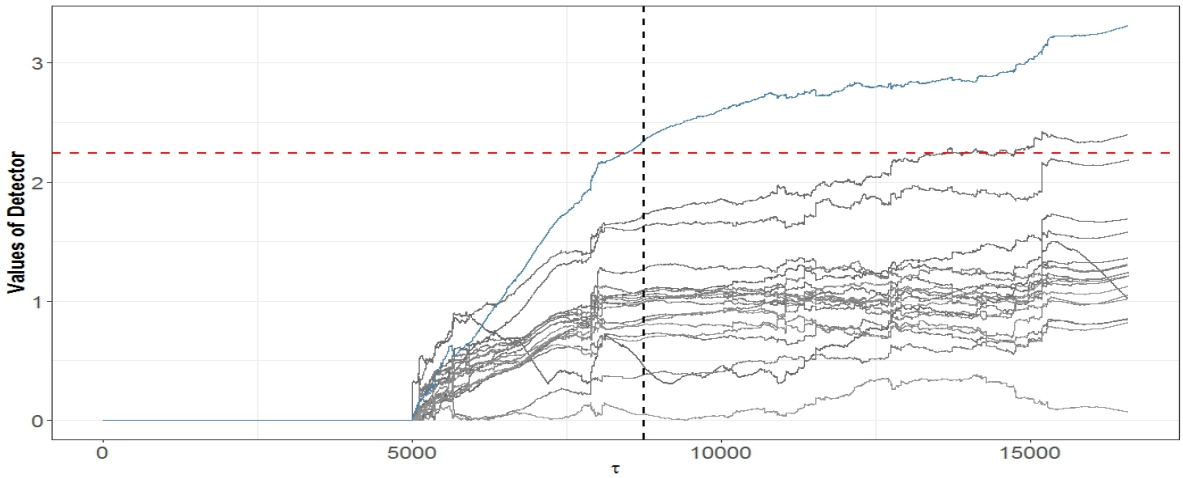

Figure 30: Plot of the sequences of detector for each parameter. Blue line represents the monitoring sequence for parameter α_{20}. Red dashed line is the true structural break point while red dashed line is the critical level.

between days and the critical level. Note that in the first example, $\widehat{\alpha}_1$ (labeled as blue line in Figure 29) first cross the critical level at $\tau = 10182$ (the true boundary is 10183). However, most of other parameters do not cross the boundary, indicating that our monitoring procedure

can successfully detect structural break points with a correct model order. In the second example, $\widehat{\alpha}_{20}$ first cross the critical level (as indicated as blue line in Figure 30) at $\tau = 8435$ (the true boundary is 8745). A possible explanation is that our procedure fits a Log ACD$(20, 0)$ model to the observed durations. That is, it uses a large number of p. However, the stochastic pattern is largely affected by βs, which truly reflect the change of $\widehat{\alpha}_{20}$.

4.10 Summary and Discussion

In this chapter, we have addressed the structural break detection problem in univariate time series of durations. We combine the recursive feature of our penalized estimating functions approach and our newly proposed stopping rule to construct a monitoring scheme. Our monitoring scheme allows us to detect number of structural breaks, locations of those structural breaks as well as model order simultaneously. A nice feature of our monitoring scheme is that economic analysts can detect structural breaks in both offline and online ways.

Chapter 5

Structural Break Detection in Dynamic Networks

5.1 Introduction

In this big data era, people have experienced fast advancement of large data collection and storage. As a result, new statistical techniques are badly needed to understand the structure under complex data sets. Networks can naturally capture structures and provide an insight into dependencies among observed nodal attributes. In many examples, such as in financial, social, or information sciences, it is often necessary to analyze a large collection of random variables interconnected by a complex network. A network is usually denoted as $G(V, E)$, where V is a set of vertices or nodes and $E \subseteq V \times V$ is a set of edges connecting the vertices. A network is called undirected if for all $v, w \in V$, $(v, w) \in E$ is equivalent to $(w, v) \in E$, otherwise directed. Depending on the application, each node can represent a stock or a gene while each edge represents a linkage or connection between the corresponding nodes. A structural break may occur in the network at some points in time due to a sudden event, possibly resulting in a totally different network structure. Apart from dynamic changes where a few edges (the

number is bounded) may be added or removed frequently over time, structural breaks are sudden changes. That is, the stochastic structure of the network remains unchanged until a time τ, when a change in number of edges may occur, thus altering the network structure. Figure 31 illustrates some examples of structural breaks in a network evolving over time. At the time point τ_1, four red nodes are connected. At time point τ_2, these four red nodes are disconnected. At time point τ_3, three other red nodes get connected.

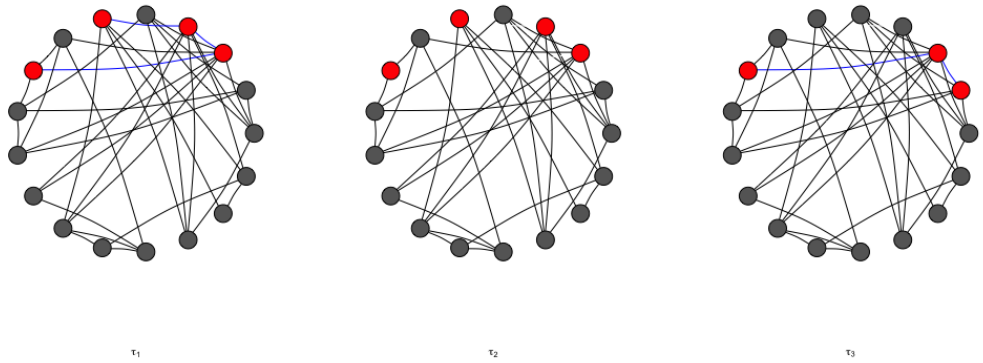

Figure 31: An example of structural breaks in a network. At time point τ_1, four red nodes are connected. At time point τ_2, four red nodes are disconnected. At time point τ_3, three other red nodes get connected

In this chapter, we propose a framework for detecting structural breaks in dynamic networks. In particular, we discover unobserved dynamic network structure from nodal observations that are governed both by the latent network and time. We develop methods for simultaneously determining number of structural breaks and locations of break points. We demonstrate our method in networks of a large number of stocks that evolve over time.

5.2 Preliminaries on Networks

Let $G = (V, E)$ be a graph with vertices set V and edge set E. A node $u \in V$ represents an entity (e.g., a stock) and an edge $(u, v) \in E$ represents a relationship between them (e.g., correlation). At each time t, each node in the vertex set $V = (1, \cdots, p)$ corresponds to an element of a $p-$dimensional random vector $X = (X_1, \cdots, X_p)'$ of nodal characteristics following a probability distribution indexed by $\theta \in \Theta$. X can be either discrete or continuous. When the elements of the random vector X are discrete, the model is referred to as a discrete Markov random field (MRF), whereas when X is continuous-valued, the model is referred to as a Gaussian graphical model (GGM). Under MRF, the edge set $E \subseteq V \times V$ encodes certain conditional independence assumptions among nodes X. For example, the random variable X_a is conditionally independent of X_b given the rest of the variables if $(a, b) \notin E$.

A special case of a MRF is the Ising model where each nodal attribute can only assume binary values. That is, $X_a \in \{-1, 1\}$, for all $a \in V$ and the joint probability distribution over X can be expressed as:

$$P_\theta(X = x) = \frac{1}{Z(\theta)} exp\left(\sum_{(a,b)\in E} \theta_{ab} x_a x_b\right) \tag{5.1}$$

where $Z(\theta) = \sum_{x\in\{-1,1\}^p} exp(\sum_{(a,b)\in V} \theta_{ab} x_a x_b)$ is the partition functions that ensures the integration of (5.1) to be 1. The model (5.1) is completely determined by $\{\theta_{ab}\}_{(a,b)\in E}$. Furthermore, the parameters specify the graph structure, that is, we have that $\theta_{ab} = 0$ for all $(a, b) \notin E$. Therefore, knowing $\{\theta_{ab}\}_{(a,b)\in V\times V}$ can help uncover the structure.

Under the Gaussian graphical model (GGM), the probability distribution can be fully

specified by the first two moments. Let

$$X = (X_1, X_2, \cdots, X_p)' \sim \mathcal{N}_p(0, \Sigma) \tag{5.2}$$

where $\mathcal{N}_p(0, \Sigma)$ denote a p-dimensional multivariate normal distribution with mean 0 and covariance matrix $\Sigma = \{\sigma_{ab}\}$ where $(a, b) \in V$. $\Sigma^{-1} = \{\omega_{ab}\}$ is called precision matrix which encodes the conditional independence under MRF. If $\omega_{ab} = 0$, X_a and X_b are conditionally independent given the rest of the nodes, while if $\omega_{ab} \neq 0$, then X_a and X_b are dependent. Therefore, the graph $G(V, E)$ is completely determined by the precision matrix Σ^{-1}.

5.3 A Review of Estimating Time-Varying Networks

Much of the existing methodology for structure learning in networks assumes that the data display time-invariance property and that the relational structure is not dynamic (Friedman, Hastie, and Tibshirani, 2008; Peng, Wang, Zhou, and Zhu, 2009; Ravikumar, Wainwright, Lafferty, et al., 2010; Yuan and Lin, 2006). To our knowledge, Kolar et al. (2010), Zhou, Lafferty, and Wasserman (2010), Kolar and Xing (2012) and Roy, Atchadé, and Michailidis (2016) were the first to work on dynamic relational structure.

Kolar et al. (2010) considered data collected at different time points as time-varying networks. They focused on estimating a sequence of graph $\{G_t\}_{t \in \mathcal{T}}$ corresponding to observations $\boldsymbol{X}_t \sim \mathbb{P}_{\theta_t}$, where $X_t \in \{0, 1\}^p$. Further, they assumed that the probability distributions change smoothly over time, or there exists a partition of the interval [0,1] into segments where the graph structure within each segment is invariant. Zhou et al. (2010) developed a nonparametric

method for estimating a time-varying Gaussian graphical model, under the assumption that $\boldsymbol{X}_t \sim \mathcal{N}_p(0, \Sigma_t)$ are independent, but not identically distributed, realizations of a multivariate distribution whose covariance matrix changes smoothly over time. Although both papers took the time-varying component into account, their focus was not structural break detection. Roy et al. (2016) studied single structural break detection on MRFs. They proposed a profile pseudo-likelihood approach which allows the number of nodes to change with time.

Kolar and Xing (2012) considered structural break detection in time-varying networks. Since their work is the most related one to our goal, we explain their approach in detail. To describe their setup, let $\{\boldsymbol{X}_t\}_{t\in n} \in \mathcal{R}^p$ be a sequence of n independent observations from some $p-$dimensional multivariate Gaussian distributions. Let $\{\mathcal{B}^j\}_{j\in[B]}$ be a disjoint partition of set n where each block of the partition consists of a certain length of observations. Let $\mathcal{T} := \{T_0 = 1 < T_1 < \cdots < T_B = n + 1\}$ denote the set of partition boundaries. They considered the following model:

$$\boldsymbol{X}_t \sim \mathcal{N}_p(0, \Sigma^j), \quad t \in \mathcal{B}^j \tag{5.3}$$

such that each block consists of a piecewise constantly evolving network, and the boundary set $\mathcal{T}$ consists of the locations of break points. To uncover the latent network structure, one needs to solve the precision matrix $(\Sigma^j)^{-1} = \{\omega^j_{ab}\}$, where $(\omega^j_{ab}) = 0$ indicates a conditional independence between $\boldsymbol{X}_a$ and $\boldsymbol{X}_b$ at segment j. Kolar and Xing (2012) proposed a time-varying covariance selection procedure based on the time-coupled neighborhood selection using the fused-type penalty. They called this procedure Temporal-Different Lasso (TD-Lasso).

With the neighborhood selection procedure, one can recover the underlying structure by combining local structures for each node $a \in V$. Let S_a^j denote the neighborhood of the node a on the block $\mathcal{B}^j$ and N_a^j denote nodes that do not belong to the neighborhood of node a on the jth block, $N_a^j = V \setminus S_a^j$. The neighborhood structure is derived by minimizing the following objective function

$$\widehat{\boldsymbol{\beta}^a} = \underset{\boldsymbol{\beta} \in \mathcal{R}^{(p-1)\times n}}{\operatorname{argmin}} \; L(\boldsymbol{\beta}) + pen_{\lambda_1,\lambda_2}(\boldsymbol{\beta}) \tag{5.4}$$

where the loss function of $\boldsymbol{\beta} = (\beta_{b,t})_{b\in[p-1],t\in n}$ is

$$L(\boldsymbol{\beta}) = \sum_{t\in n} \left(x_{t,a} - \sum_{b\in\setminus a} x_{t,b}\beta_{b,t} \right)^2 \tag{5.5}$$

and the penalty term is defined as

$$pen_{\lambda_1,\lambda_2}(\boldsymbol{\beta}) = 2\lambda_1 \sum_{t=2}^{n} ||\boldsymbol{\beta}_{.,t} - \boldsymbol{\beta}_{.,t-1}||_2 + 2\lambda_2 \sum_{t=1}^{n} \sum_{b\in\setminus a} ||\beta_{b,t}||_1 \tag{5.6}$$

where λ_1 controls the number of structural breaks while λ_2 controls the sparsity of the precision matrix. Then the estimated set of boundaries is given as

$$\hat{\mathcal{T}} = \{\hat{\mathcal{T}}_0 = 1\} \cup \{\hat{\mathcal{T}}_j : \hat{\boldsymbol{\beta}}^a_{.,\hat{T}_j} \neq \hat{\boldsymbol{\beta}}^a_{.,\hat{T}_j-1}\} \cup \{\hat{\mathcal{T}}_{\hat{B}} = n+1\} \tag{5.7}$$

Based on the estimated boundaries, one can define the neighborhood estimation for node a at block j is

$$\hat{S}_a^j = \{\hat{\boldsymbol{\beta}}^a_{.,t} : \hat{\boldsymbol{\beta}}^a_{b,t} \neq 0, b \in \setminus a, t \in [\hat{\mathcal{T}}_{j-1} : \hat{\mathcal{T}}_j]\} \tag{5.8}$$

Then the global structure is revealed by combining different neighborhoods for each node $a \in V$.

5.4 Monitoring Change in Financial Networks

The aforementioned work are mostly applied in bioinformatics or social sciences, where the data are usually independent and identically distributed. The observations in our financial application are dependent time series. It is insufficient to only consider the dependent structure between nodes at each time point. The temporal structure contained in the data at each node should also be taken into account when modeling such networks. Therefore, we need to adjust the work just described to meet the requirement of both dependence (column-wise and row-wise). To our best knowledge, no one has considered structural break detection in financial networks from a statistical modeling point of view. In this section, we address this problem.

To illustrate our procedure, let us consider a network $G = (V, E)$ with vertices set $V = (1, 2, ..., p)$ and edge set E. A node $u \in V$ denotes a stock in our case and an edge $(u, v) \in E$ represents a connection between them.

As we described in Chapter 2, we obtained transactional level data for S&P100 stocks from Trade and Quotes database at Wharton Research Data Services and pre-cleaned the data by deleting observations with special trade sale conditions. We focus on 30 the most liquid stocks in the S&P100 from 9:30am to 4:00pm in June 2013. In order to construct a financial network, we process the transactional level data in the following steps.

Step 1. In step one, we calculate the number of transactions in every one-minute interval for each stock over the month of June, 2013. Since the counts are sufficiently large, we view

counts as continuous-valued time series. Step 2. In step two, we consider time series of log counts each of length n within each of T windows. The windows are chosen to be either one-day and three-hours. Let t be the intra-window time index and τ index the set of windows, so $t = 1, 2, ..., n$, $\tau = 1, ..., T$. For windows of length one-day, n and T are 360 and 20 respectively. Likewise for a three-hour window, n and $\mathcal{T}$ are 180 and 40 respectively. Within each window, let $\boldsymbol{X}_\tau$ denote a p dimensional time series ($p = 30$ corresponding to 30 stocks) observed for $\tau = 1, 2, ..., T$.

Step 3. In step three, we establish the correlation between nodes pairwise within each window. For example, we look at $(\boldsymbol{X}_u, \boldsymbol{X}_v)$, where $u, v \in V$. The squared coherence (Shumway and Stoffer, 2010, Sec. 4.8) between $(\boldsymbol{X}_u, \boldsymbol{X}_v)$ is defined as

$$\rho^2_{u \cdot v}(\omega_l) = \frac{|f_{uv}(\omega_l)|^2}{f_{uu}(\omega_l) f_{vv}(\omega_l)},$$

where $\omega_l = l/n$ are the Fourier frequencies, $f_{uv}(\omega_l)$ is the cross-spectrum between $(\boldsymbol{X}_u, \boldsymbol{X}_v)$, while $f_{uu}(\omega_l)$ and $f_{vv}(\omega_l)$ are the direct spectra of $\boldsymbol{X}_u$ and $\boldsymbol{X}_v$ respectively.

Step 4. Let

$$s_{uv} = \max_{w_l} \{\rho^2_{u \cdot v}(\omega_l)\}.$$

We construct the matrix of maximum squared coherences, $\boldsymbol{S}_\tau = \{s_{uv}\}_{u,v \in V}$. Our structural break detection is based on the coherence matrix $\boldsymbol{S}_\tau$ of max squared coherences.

The network or graph $G_{\boldsymbol{S}_\tau}$ of conditional dependences can be estimated by determining the zero entries of the inverse matrix $\boldsymbol{K}_\tau = \boldsymbol{S}_\tau^{-1}$. We can use the graphical lasso (glasso) estimator proposed by Yuan and Lin (2007) and Banerjee et al. (2008) to estimate $\boldsymbol{K}_\tau$. Specifically, at

every window τ, we calculate

$$\widehat{\boldsymbol{K}}_\tau = \underset{\boldsymbol{K}_\tau}{\operatorname{argmin}} \{-\log det\{\boldsymbol{K}_\tau\} + tr(\boldsymbol{S}_\tau \boldsymbol{K}_\tau) + \lambda ||\boldsymbol{K}_\tau||_1\} \tag{5.9}$$

where λ is a tuning parameter. We monitor the change in the estimated precision matrix $\widehat{\boldsymbol{K}}_\tau$. Structural break(s) are signaled when any two adjacent $\widehat{\boldsymbol{K}}_\tau$ and $\widehat{\boldsymbol{K}}_{\tau+1}$ have element-wise changes.

How does our setup relate to a GGM? We first transform our observed time series of log transaction counts within each window indexed by τ to its Discrete Fourier Transform (DFT) (Shumway and Stoffer, 2010, Sec. 4.4) of $\{x_{u,t}\}$ is defined as

$$d_u(\omega_l) = n^{-1/2} \sum_{t=1}^{n} x_{u,t} \exp\{-2\pi i \omega_l t\} \quad l = 0, 1, ..., n-1.$$

It is known that $\{d_u(\omega_l)\}_{u \in V}$ is asymptotically complex Gaussian with mean 0 and variance given by the spectrum, $f_u(\omega_l)$. Let $\boldsymbol{d}(w_l) = (d_1(w_l), ..., d_p(w_l))$. The joint distribution of $\boldsymbol{d}(w_l)$ at frequency ω_l is asymptotically multivariate complex Gaussian with mean $\mathbf{0}$ and variance given by the $p \times p$ spectral density matrix, $\boldsymbol{f}(\omega_l)$. Notice that every element in $\boldsymbol{S}_\tau$ is a function of the corresponding element in $\boldsymbol{f}(\omega_l)$, and our setup can be regarded as a "Fourier transformed" GGM.

5.5 An Application to the Network of 30 Stocks

In this section, we apply our method to a time-varying network of real financial data to investigate structural changes. The data process was described in Section 5.4. Table 11

summarizes the stocks names, their corresponding sectors and number of transactions in June, 2013 for these 30 stocks. We monitor the network among stocks across two windows (daily and every three hours).

Table 11: Summary of the 30 most liquid stocks in S&P 100.

Stock	Sector	Number of transactions
AAPL	Information Technology	3,687,407
PFE	Healthcare	3,559,605
BAC	Financials	3,019,195
MSFT	Information Technology	2,802,762
C	Financials	2,738,675
INTC	Information Technology	2,395,144
ORCL	Information Technology	2,168,090
CSCO	Information Technology	1,930,667
GE	Industrials	1,928,661
JPM	Financials	1,916,533
WFC	Financials	1,766,857
F	Consumer Discretionary	1,698,967
MS	Financials	1,656,932
T	Telecommunication Service	1,654,819
MRK	Healthcare	1,616,907
QCOM	Information Technology	1,482,100
EMC	Consumer Discretionary	1,429,295
XOM	Energy	1,418,870
ABT	Healthcare	1,351,462
FCX	Materials	1,334,192
KO	Consumer Staples	1,258,577
GILD	Healthcare	1,139,049
EBAY	Information Technology	1,127,442
JNJ	Healthcare	1,126,607
HPQ	Information Technology	1,098,411
VZ	Telecommunication Service	1,093,048
ABBV	Healthcare	1,064,180
BMY	Healthcare	1,058,203
PG	Consumer Staples	947,543
MDLZ	Consumer Staples	916,643

We first report our analyses for a one-day window. That is, we consider the end of day one is the first time stamp (e.g. $\tau = 1$) and the end of day two as the second time stamp (e.g. $\tau = 2$) etc.. Thus for each stock, there are $n = 360$ log counts within each window. We then calculate the sample squared coherence matrix $\boldsymbol{S}_\tau$, $\tau = 1, 2, ..., 20$. The first quartile, median and third quartile of $\boldsymbol{S}_\tau$ are roughly 0.6, 0.7 and 0.75 respectively, indicating high correlations among stocks. The precision matrix $\boldsymbol{K}_\tau$ is obtained via group lasso (5.9) with $\lambda = 0.9$. Note that λ controls the sparsity of the precision matrix $\boldsymbol{K}_\tau$. Since a small value of λ would imply many connected stocks, we recommend that λ should exceed the median of sample coherences by 0.2. In this way, we can obtain a relatively sparse network which indicates the strongest connections among stocks.

Figure 32 displays the detected structural breaks, which occur on 5 different days (the 9th, 11th, 13th, 14th and 17th). For the remaining 15 days, there is no network among the 30 stocks. We have color coded the different sectors, Information Technology, Healthcare, Financial, Industrials, Consumer Discretionary, Telecommunication Service, Energy, Materials and Consumer Staples are labeled as red, blue, green, grey, dark red, light blue, purple, dark green and pink respectively. For the 9th day, JPM and BAC are connected, suggesting a highly similar trading pattern between both stocks. Figure 33 confirms that the JPM and BAC have similar trading patterns on the 9th day. Such information could deepen investigators' understanding of the market. It is seen that Financial stocks are more likely to have strong connections with other stocks within or outside the financial sector. On the 13th day, 10 stocks which belong to 6 sectors have strong connections, possibly indicating intensive trading. We searched the historical S&P100 index in the month of June 2013 and found that the index on

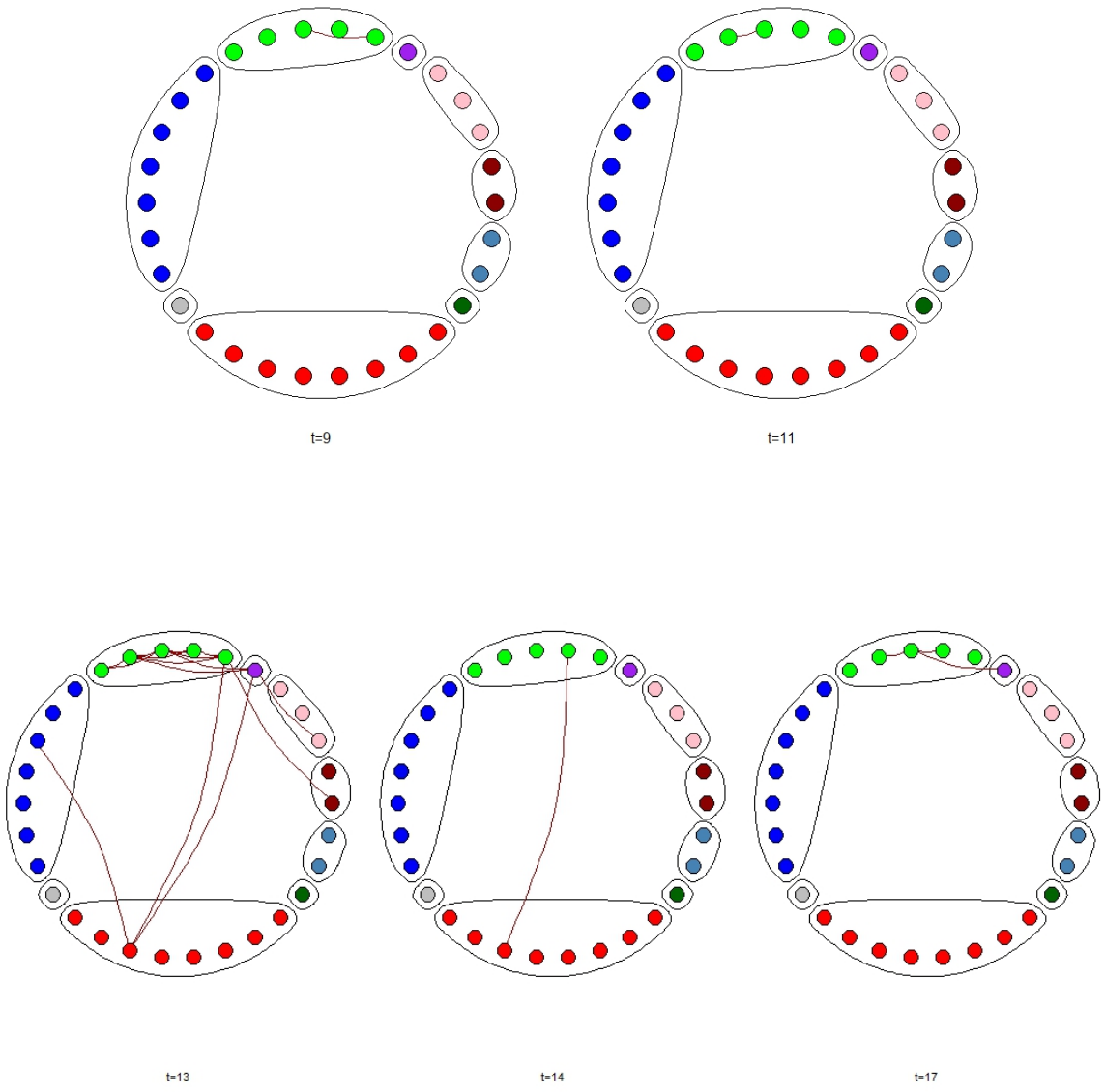

Figure 32: Detected structural breaks at five different days.

June 19, 2013 (13th day) was the highest in that month.

We also conduct similar analyses on a three-hour window. Due to the decreased length of observations (from $n = 360$ to $n = 180$) for each stock within each window, it is possibly

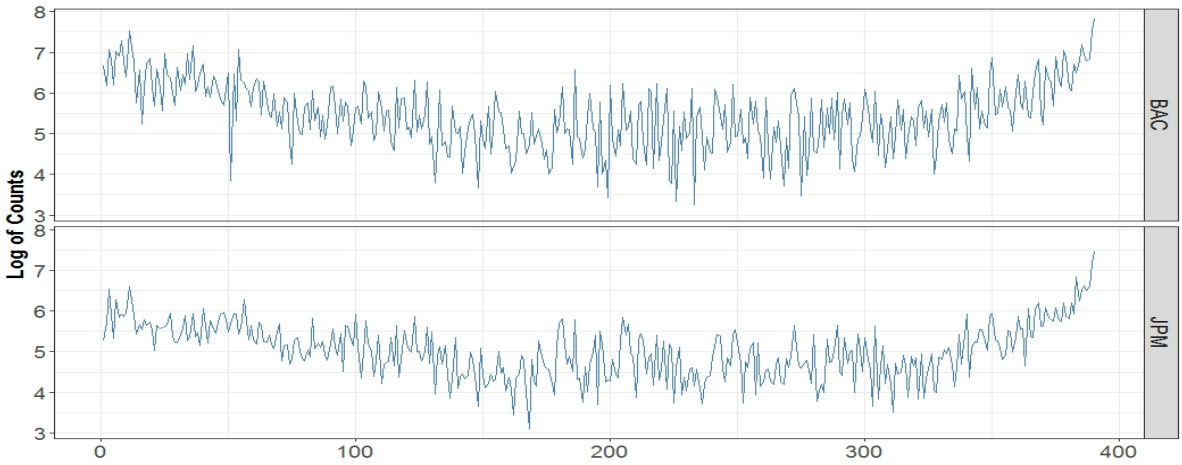

Figure 33: Plots of log counts for BAC and JPM on Day 9.

more difficult for two stocks to have very similar patterns, as indicated by lower values for the maximum squared coherences (values in S_τ). We suggest that a smaller penalty λ be used to detect structural breaks in networks. To avoid redundancy, we only report findings that are substantially different than that obtained from analyses on one-day window. In general, morning sessions have weaker connections among stocks; the median of maximum squared coherences ranges from 0.2 to 0.3, while in the afternoon session, connections among stocks are slightly stronger; the median of the maximum squared coherences ranges from 0.3 to 0.4 approximately. Therefore, more structural breaks in networks occur during the afternoon session. This phenomenon is in line with the market behavior where more transactions occur in the afternoon perhaps because more information are gathered till the afternoon sessions. In addition, the sets of stocks that are analyzed on a three-hour window generally cover the sets of stocks detected on one-day window, providing evidence that our analyses are consistent on different scales.

5.6 Summary and Discussion

In this chapter, we propose a framework for detecting structural breaks in dynamic financial networks. Unlike existing methodologies, we focus on temporal structure and dependence relation between nodes. The use of matrix of maximum squared coherences allows us to detect structural breaks in a Fourier transformed GGM setup. This is a preliminary analysis for structural break detection in dynamic networks and future work can investigate this more in-depth analysis.

Chapter 6

Discussion and Future Directions

High-frequency financial data are of direct interest for issues dealing with the market infrastructure, structural instability and modeling of real-time market dynamics. Accurate structure learning and break detection are of considerable interest in high-frequency financial data analysis. The main contributions of this dissertation are three-fold. First, we introduce a penalized estimating functions framework for modeling the class of Log ACD models for financial durations under minimal distributional assumptions. We believe that financial analysts would prefer to employ predictive statistical models which are not hinged to distributional assumptions, and in this regard that we have chosen to employ the estimating functions approach to carry out the predictive modeling and introduced penalized EFs to induce sparsity. Second, we combine the recursive feature of our penalized estimating functions approach and our newly proposed stopping rule to monitor structural instabilities. Our procedure can help financial analysts to detect structural breaks in both offline and online approaches. Third, we introduce structural break detection in dynamic financial networks. Our exploratory study should be helpful to financial analysts understanding the underlying infrastructure of dynamic financial networks.

In addition, the estimating function approach presented in this dissertation is appealing for use with a wide range of durations models (Pacurar, 2008; Thavaneswaran et al., 2015) that

are useful for financial modeling, since the estimating functions approach is very general and enables users to fit models with least restrictive assumptions which are very useful in practice.

Our proposed approach regarding structural break detection in dynamic financial networks is essentially a two-step method. First, we transform each time series to its Discrete Fourier Transform (DFT). Second, assuming the transformed data follow a multivariate complex Gaussian distribution, we obtain the precision matrix at different time stamps using graphical lasso. A possible future research direction could be to detect structural breaks as well as temporal patterns in a single step. Let $\{\boldsymbol{y_t}\}_{t=1}^{T}$ denote a k dimensional vector time series and the errors $\{\boldsymbol{e_t}\}_{t=1}^{T}$ follows a multivariate distribution with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. The objective could be the following

$$\min_{\boldsymbol{\nu},\boldsymbol{\Phi},\boldsymbol{\Sigma}} = \sum_{t=1}^{T} ||\boldsymbol{y_t} - \nu - \sum_{l=1}^{p} \Phi^{(l)} \boldsymbol{y_{t-l}}||^2 + \lambda \left(\mathcal{P}_1(\boldsymbol{\Phi}) + \mathcal{P}_2(\boldsymbol{\Sigma})\right), \tag{6.1}$$

where $\boldsymbol{\nu}$ denotes a $k \times 1$ intercept vector, $\boldsymbol{\Phi}^{(l)}$ denotes a time-invariant $k \times k$ matrix. $\mathcal{P}_i(\cdot), i = 1, 2$ is a penalty function which controls the sparsity of the model. The difficulty of implementing setup in 6.1 lies in an fast and efficient algorithm that minimizes this objective function.

Bibliography

Aue, A., R. C. Cheung, T. C. Lee, M. Zhong, et al. (2017). Piecewise quantile autoregressive modeling for nonstationary time series. *Bernoulli 23*(1), 1–22.

Aue, A., S. Hörmann, L. Horváth, M. Hušková, and J. G. Steinebach (2012). Sequential testing for the stability of high-frequency portfolio betas. *Econometric Theory 28*(4), 804–837.

Aue, A. and L. Horváth (2013). Structural breaks in time series. *Journal of Time Series Analysis 34*(1), 1–16.

Aue, A., L. Horváth, and M. L. Reimherr (2009). Delay times of sequential procedures for multiple time series regression models. *Journal of Econometrics 149*(2), 174–190.

Bacry, E. and J.-F. Muzy (2014). Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance 14*(7), 1147–1166.

Bai, J. (1999). Likelihood ratio tests for multiple structural changes. *Journal of Econometrics 91*(2), 299–323.

Banerjee, A. and G. Urga (2005). Modelling structural breaks, long memory and stock market volatility: an overview. *Journal of Econometrics 129*(1), 1–34.

Banerjee, O., L. E. Ghaoui, and A. dAspremont (2008). Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *Journal of Machine learning research 9*(Mar), 485–516.

Bauwens, L. and P. Giot (2000). The logarithmic acd model: an application to the bid-ask quote process of three nyse stocks. *Annales d'Economie et de Statistique*, 117–149.

Bauwens, L. and N. Hautsch (2009). Modelling financial high frequency data using point processes. *Handbook of financial time series*, 953–979.

Bera, A. K., Y. Bilias, P. Simlai, et al. (2006). Estimating functions and equations: An essay on historical developments with applications to econometrics. *Palgrave Handbook of Econometrics 1*, 427–476.

Berkes, I., E. Gombay, L. Horváth, and P. Kokoszka (2004). Sequential change-point detection in garch (p, q) models. *Econometric Theory 20*(6), 1140–1167.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics 31*(3), 307–327.

Chan, N. H., C. Y. Yau, and R.-M. Zhang (2014). Group lasso for structural break time series. *Journal of the American Statistical Association 109*(506), 590–599.

Chen, K. and K.-S. Chan (2011). Subset arma selection via the adaptive lasso. *Statistics and its Interface 4*(2), 197–205.

Chu, C.-S. J., M. Stinchcombe, and H. White (1996). Monitoring structural change. *Econometrica: Journal of the Econometric Society*, 1045–1065.

Cleveland, W. S., E. Grosse, and W. M. Shyu (1992). Local regression models. *Statistical models in S 2*, 309–376.

Coleman, T. F. and Y. Li (1996). An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on optimization 6*(2), 418–445.

Csörgö, M. and L. Horváth (1997). *Limit theorems in change-point analysis*, Volume 18. John Wiley & Sons Inc.

Csörgo, M. and P. Révész (2014). *Strong approximations in probability and statistics*. Academic Press.

Davis, R. A., D. Huang, and Y.-C. Yao (1995). Testing for a change in the parameter values and order of an autoregressive model. *The Annals of Statistics*, 282–304.

Davis, R. A., T. C. M. Lee, and G. A. Rodriguez-Yam (2006). Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association 101*(473), 223–239.

Dias, A. and P. Embrechts (2002). Change-point analysis for dependence structures in nance and insurance. *Novos Rumos em Estatstica(Ed. C. Carvalho, F. Brilhante and F. Rosado), Sociedade Portuguesa de Estatstica, Lisbon*, 69–86.

Döring, M. (2011). Convergence in distribution of multiple change point estimators. *Journal of Statistical Planning and Inference 141*(7), 2238–2248.

Durbin, J. (1960). Estimation of parameters in time-series regression models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 139–153.

Easley, D. and M. O'hara (1992). Time and the process of security price adjustment. *The Journal of finance 47*(2), 577–605.

Engle, R. F. and J. R. Russell (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica 66*, 1127–1162.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association 96*(456), 1348–1360.

Fisher, R. A. (1924). The conditions under which $\chi 2$ measures the discrepancey between observation and hypothesis. *Journal of the Royal Statistical Society 87*, 442–450.

Friedman, J., T. Hastie, and R. Tibshirani (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics 9*(3), 432–441.

Ghahramani, M. and A. Thavaneswaran (2012). Nonlinear recursive estimation of volatility via estimating functions. *Journal of Statistical Planning and Inference 142*(1), 171–180.

Godambe, V. (1985). The foundations of finite sample estimation in stochastic processes. *Biometrika 72*(2), 419–428.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics 31*(4), 1208–1211.

Hafner, C. M. (2005). Durations, volume and the prediction of financial returns in transaction time. *Quantitative Finance 5*(2), 145–152.

Hancock, S. (2008). *Estimation of structural breaks in nonstationary time series.* Colorado State University.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, 1029–1054.

Hansen, L. P. and K. J. Singleton (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica: Journal of the Econometric Society 50*(5), 1269–1286.

Hansen, M. and B. Yu (2000). Wavelet thresholding via mdl for natural images. *IEEE Transactions on Information Theory 46*(5), 1778–1788.

Hejjel, L. and E. Roth (2004). What is the adequate sampling interval of the ecg signal for heart rate variability analysis in the time domain? *Physiological measurement 25*(6), 1405.

Heyde, C. C. (2008). *Quasi-likelihood and its application: a general approach to optimal parameter estimation.* New York, NY: Springer Science & Business Media.

Holland, J. H. (1992). Adaptation in natural and artificial systems. 1975. *Ann Arbor, MI: University of Michigan Press and*.

Horváth, L. and G. Rice (2014). Extensions of some classical methods in change point analysis. *Test 23*(2), 219–255.

Hušková, M., Z. Prášková, and J. Steinebach (2007). On the detection of changes in autoregressive time series i. asymptotics. *Journal of Statistical Planning and Inference 137*(4), 1243–1259.

Jassby, A. D. and T. M. Powell (1990). Detecting changes in ecological time series. *Ecology 71*(6), 2044–2052.

Kolar, M., L. Song, A. Ahmed, and E. P. Xing (2010). Estimating time-varying networks. *The Annals of Applied Statistics*, 94–123.

Kolar, M. and E. P. Xing (2012). Estimating networks with jumps. *Electronic journal of statistics 6*, 2069.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society 53*(6), 1315–1335.

Liang, Y., A. Thavaneswaran, and B. Abraham (2011). Joint estimation using quadratic estimating function. *Journal of Probability and Statistics 2011*, 1–14. Article ID 372512.

Lindsay, B. G. (1985). Using empirical partially bayes inference for increased efficiency. *The Annals of Statistics 13*(3), 914–931.

Lu, Q., R. Lund, T. C. Lee, et al. (2010). An mdl approach to the climate segmentation problem. *The Annals of Applied Statistics 4*(1), 299–319.

Merkouris, T. et al. (2007). Transform martingale estimating functions. *The Annals of Statistics 35*(5), 1975–2000.

Pacurar, M. (2008). Autoregressive conditional duration models in finance: a survey of the theoretical and empirical literature. *Journal of Economic Surveys 22*(4), 711–751.

Page, E. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika 42*(3/4), 523–527.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika 41*(1/2), 100–115.

Peng, J., P. Wang, N. Zhou, and J. Zhu (2009). Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association 104*(486), 735–746.

PRNewswire (2013a). *Earning Forecast Highlight for Pfizer, Bank of America, Lululemon Athletica, Facebook, Elan, and QUALCOMM.* `http://en.prnasia.com/story/81203-0.shtml` [Published: June 13, 2013].

PRNewswire (2013b). *Insider Trading Investment for Bank of America, Hewlett-Packard, Bristol Myers Squibb, Annaly Capital Management, Walt Disney, and American Capital Agency.* `http://en.prnasia.com/story/81456-0.shtml` [Published: June 18, 2013].

PRNewswire (2013c). *Stock News for Bank of America, ACADIA Pharmaceuticals, American Capital Agency, Advanced Micro Devices, Annaly Capital Management, and Research In Motion*. `http://en.prnasia.com/story/81083-0.shtml` [Published: June 11, 2013].

Rabemananjara, R. and J.-M. Zakoian (1993). Threshold arch models and asymmetries in volatility. *Journal of Applied Econometrics 8*(1), 31–49.

Ravikumar, P., M. J. Wainwright, J. D. Lafferty, et al. (2010). High-dimensional ising model selection using 1-regularized logistic regression. *The Annals of Statistics 38*(3), 1287–1319.

Rissanen, J. (1985). *Minimum description length principle*. Wiley Online Library.

Robbins, M., C. Gallagher, R. Lund, and A. Aue (2011). Mean shift testing in correlated data. *Journal of Time Series Analysis 32*(5), 498–511.

Robbins, M. W., C. M. Gallagher, and R. B. Lund (2016). A general regression changepoint test for time series data. *Journal of the American Statistical Association 111*(514), 670–683.

Roy, S., Y. Atchadé, and G. Michailidis (2016). Change point estimation in high dimensional markov random-field models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Shumway, R. H. and D. S. Stoffer (2010). *Time series analysis and its applications: with R examples* (Third Edition ed.). New York, NY: Springer Science & Business Media.

Thavaneswaran, A., N. Ravishanker, and Y. Liang (2015). Generalized duration models and optimal estimation using estimating functions. *Annals of the Institute of Statistical Mathematics 67*(1), 129–156.

Tsay, R. S. (2005). *Analysis of financial time series*, Volume 543. Hoboken, New Jersey: John Wiley & Sons.

Tsay, R. S. (2009). Autoregressive conditional duration models. In T. C. Mills and K. Patterson (Eds.), *Palgrave handbook of econometrics*, pp. 1004–1024. London, UK: Palgrave Macmillan.

Vlahogianni, E. I., M. G. Karlaftis, and K. Kepaptsoglou (2011). Nonlinear autoregressive conditional duration models for traffic congestion estimation. *Journal of Probability and Statistics 2011*.

Wang, L., J. Zhou, and A. Qu (2012). Penalized generalized estimating equations for high-dimensional longitudinal data analysis. *Biometrics 68*(2), 353–360.

Yan, B. and E. Zivot (2003). Analysis of high-frequency financial data with s-plus. Technical report, UWEC-2005-03 [online]. Available at http://ideas. repec. org/p/udb/wpaper/uwec-2005-03. html.

Yuan, J., Y. Pan, and X. Zhang (2015). Ultrahigh frequency data liquidity duration estimation: a case study of chinese a shares. *Mathematical Problems in Engineering 2015*.

Yuan, M. and Y. Lin (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(1), 49–67.

Yuan, M. and Y. Lin (2007). Model selection and estimation in the gaussian graphical model. *Biometrika 94*(1), 19–35.

Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and control 18*(5), 931–955.

Zhang, M. Y., J. R. Russell, and R. S. Tsay (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics 104*(1), 179–207.

Zhou, S., J. Lafferty, and L. Wasserman (2010). Time varying undirected graphs. *Machine Learning 80*(2), 295–319.

Zou, J., Y. An, and H. Yan (2015). Volatility matrix inference in high-frequency finance with regularization and efficient computations. In *Big Data (Big Data), 2015 IEEE International Conference on*, pp. 2437–2444. IEEE.