AUTOREGRESSIVE CONDITIONAL DURATION AND LIQUIDITY

ÜMİT ALTAY BARAN

BOĞAZİÇİ UNIVERSITY

2022

AUTOREGRESSIVE CONDITIONAL DURATION AND LIQUIDITY

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Management

by

Ümit Altay Baran

Boğaziçi University

2022

Autoregressive Conditional Duration and Liquidity

The thesis of Ümit Altay Baran

has been approved by:

Assist. Prof. Cenk C. Karahan
(Thesis Advisor) ____________________

Assoc. Prof. Neslihan Yılmaz ____________________

Assoc. Prof. Tolga Umut Kuzubaş ____________________

Assoc. Prof. Cumhur Ekinci
(External Member) ____________________

Assoc. Prof. Mehmet Oğuz Karahan
(External Member) ____________________

January 2022

DECLARATION OF ORIGINALITY

I, Ümit Altay Baran, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature…………………………………………….......

Date …………………………………………….............

ABSTRACT

Autoregressive Conditional Duration and Liquidity

This thesis provides an in-depth Autoregressive Conditional Duration (ACD) application and examines the relationship between the consecutive transaction, price, and volume durations in Borsa Istanbul (BIST) stock exchange and investigates the explanatory power of models' coefficients on return and liquidity measures. ACD models enable the usage of intraday high-frequency order book data to model the duration between consecutive transactions and predict the next meaningful duration. Durations are modeled in two parts, past and conditional, and the power of dependencies between successive trades is investigated in this way. In this study, by utilizing a subset of the ten most traded stocks in BIST and applying a framework for investigating the most suitable error term specification, various ACD models and extensions are employed to understand the intraday duration behaviors of different stocks. First, a brief explanation is specified about why the widely used low-frequency liquidity measures are inadequate at capturing the appropriate intraday liquidity. Afterwards, the durations between the transactions, prices, and volumes are modeled by using various types of ACD models, and a framework is established. Finally, new coefficients taken from the ACD applications are comparatively studied using regressions across different trading windows. Building on the framework of in-depth applications and analysis of ACD, a panel regression is used to understand the explanatory power of the new model. The findings of this study demonstrate the effectiveness of using ACD applications for modeling intraday duration effects on equity returns and liquidity.

ÖZET

Otoregresif Koşullu Süre ve Likidite

Bu tez bir Otoregresif Koşullu Süre (ACD) uygulaması sunmakta ve Borsa İstanbul'da (BİST) ardışık işlem, fiyat ve hacim süreleri arasındaki ilişkiyi incelemektedir. ACD modelleri gün içi yüksek frekanslı emir defteri verilerini kullanarak, işlemler arasındaki sürenin modellenmesini ve bir sonraki anlamlı işlem süresinin tahmin edilmesini sağlar. Süreler geçmiş ve koşullu olmak üzere iki parçada modellenir ve ardışık işlemler arasındaki bağımlılıkların gücü bu şekilde araştırılır. Bu çalışmada ilk olarak, yaygın şekilde kullanılan düşük frekanslı likidite ölçütlerinin, gün içi yüksek frekanslı işlemlerden meydana gelen likiditeyi ölçmedeki eksikliklerine dair kısa bir açıklama yapılmıştır. İkinci olarak, işlemler arasındaki süreler, fiyatlar ve hacimler, farklı ACD metotları kullanılarak modellenmiş ve karşılaştırılmıştır. En uygun hata terimi spesifikasyonunun ve ACD uzantısının seçimi için bir çerçeve çizilmiştir. Son olarak, ACD uygulamalarının yeni çıktıları, gün içi alım satım faaliyetleri ile karşılaştırmalı olarak incelenmiş ve getiriler açıklanmıştır. Çeşitli ACD analizleri üzerine inşa edilen yeni değişkenlerin açıklayıcı gücünü anlamak için bir panel regresyon uygulanmış ve piyasa koşulları, model çıktıları ile test edilmiş, açıklama gücü rapor edilmiştir. Bu tezin bulguları, hisse senedi getirileri üzerindeki gün içi likidite etkilerini modellemek için ACD uygulamalarının etkinliğini ortaya koymaktadır.

CURRICULUM VITAE

NAME: Ümit Altay Baran

DEGREES AWARDED

Ph.D. in Management, 2022, Boğaziçi University
Master's Degree in Financial Engineering, 2014, Boğaziçi University
BSc in Industrial Engineering, 2012, Galatasaray University

AREAS OF SPECIAL INTEREST

Capital market, asset pricing, derivatives, option pricing, portfolio management and modeling.

PUBLICATIONS

Journal Articles

Baran, Ü. A. & Karahan, C. C., (2021). Beta Anomaly in Turkish Equity Market. International Journal of Disciplines in Economics & Administrative Sciences Studies, Vol:7, Issue:36, 1006-1013.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ACD	Autoregressive Conditional Durations
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
BIST	Borsa Istanbul
PS	Pastor and Stambaugh
LACAPM	Liquidity Adjusted Capital Asset Pricing Model
HFD	High-Frequency Data
HFT	High-Frequency Trading
EACD	Exponential ACD
GACD	Gamma ACD
GGACD	Generalized Gamma
WACD	Weibull ACD
BCACD	Box-Cox ACD
LACD	Logarithmic ACD
TACD	Threshold ACD
AACD	Augmented ACD
AMACD	Additive and Multiplicative ACD
ABACD	Augmented Box-Cox ACD
QMLE	Quasi-maximum likelihood estimate

CHAPTER 1

INTRODUCTION

Stock liquidity has been an essential issue in financial markets for several years. It can be briefly described as the ability of buying or selling a significant amount of a security quickly, anonymously, and with minimal or no price impact.

The stock liquidity and the trading activity have been investigated from a couple of different perspectives and have drawn much attention in the financial market and academia. In the 1970s first researchers to examine the concept of liquidity were Demsetz (1968) and Tinic (1972). Demsetz manages to list the determinants of liquidity as trading volume, number of trades, volatility, firm size, and price. Four years later, Tinic demonstrates a positive relationship between trading activity and liquidity and a negative association between trading activity and volatility.

Following these new studies, the researchers in the upcoming years started to explore whether the assets with high spread and/or price impact have higher average returns from the cross-sectional point of view. Amiduh and Mendelson (1986), studying the bid-ask spreads, and Brennan and Subrahmanyam (1996), focusing on price impacts, both pieces of research find a positive relationship between expected stock returns and alternative proxies for individual illiquidity levels.

Building on the previous research, new studies investigating the different liquidity perspectives emerge. The presence of predictability and commonality in liquidity is revealed in various studies, such as Subrahmanyam (2001), Amihud (2002), among many others. Asset pricing based on liquidity risks empirically aids in explaining the cross-section of stock returns and how the decreased liquidity is increasing expected stock returns. Liquidity plays an essential role in solving numerous asset pricing puzzles and anomalies. Amihud, Mendelson, and Pederson (2005), in their survey study on liquidity and assets prices, review the principles on how liquidity impacts the required returns of securities. They show that the effects of liquidity risks in empirical studies are statistically significant and economically vital.

The liquidity risk is a priced source of risk when the particular models are fitted into the financial data. The scale of the liquidity premium is significantly present, but quantity differs among the studies and the proxies. For example, Pastor and Stambaugh (2003) report a very high 7.5% annual premium even though Acharya and Pederson (2005) find it 1.1% annually.

Across the financial literature, the main difficulty with liquidity is that the liquidity itself is unobservable, and it is needed to use proxies to integrate into any model. Since 2002, the illiquidity measure developed by Amihud has been the most widely used liquidity proxy in the financial market and literature. In recent years, more than hundreds of researches published in several most famous and attended journals like the Journal of Financial Economics, Journal of Finance, etc., used Amihud Illiquidity measure as their liquidity proxy for their empirical analyses. Amihud's measure is widely used because of briefly two main reasons; first is the ease of use and the second one is that it is demonstrated in many analyses that the Amihud Illiquidity has a solid positive relation with the expected stock return. In his study, Amihud essentially describes the price impact by the proportion of absolute daily return to the trading volume. However, in financial theory, intraday trading activities are crucial to liquidity and volatility. So that while facilitating the measurement of liquidity, we are missing or ignoring some essential daily information of the related underlying securities by using Amihud (2002). It is discussed in Chordia, Huh, and Subrahmanyam (2009) that although many microstructure theories have been developed, extant economic models are unable to map precisely onto Amihud's construct of the ratio of absolute return to volume.

The excess returns captured by the Amihud measure are generally called the liquidity premium that should be taken into consideration while modeling any securities. Academics use it to analyze liquidity premium, build liquidity factors, or liquidity tests, but we cannot be sure if the pricing of the Amihud is undoubtedly due to the price impact or other reasons. Excluding the intraday trading activity like intraday volatility, excessive price changes, or volume differences between price-changing transactions, Amihud's Illiquidity is leaving some questions marks in the asset pricing models.

This study proposes a new way of measuring liquidity that takes intraday trading information into consideration to solve this missing information problem. It is considered that due to the improved automation of financial markets and the advancements in computing power, including intraday high-frequency data into the liquidity and volatility models have become a necessity. Modeling the duration between the transactions, forecasting the following significant price change and price impact becomes available with Autoregressive Conditional Duration (ACD) models by using intraday high-frequency data.

The ACD model was developed by Engle and Russell (1998), whose specific purpose is the modeling of times between events. Both Engle (2000) and Engle (2002) demonstrate that the ACD model shares many characteristics with the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models and that durations can be modeled like volatility. As it is in the GARCH model of Bollerslev (1986), ACD models also rely on economic motivation following from the clustering of news and financial events in the market. With all these characteristics of ACD models, it becomes viable to incorporate and model intraday data for liquidity measurement. From our point of view, by generating such intraday duration estimations by using corresponding properties, it becomes possible to address most of the concerns about the usage of Amihud's liquidity measure.

In this study, there are mainly three research questions that are focused on. Firstly, it is investigated whether it is possible to use ACD specifications on BIST intraday orderbook data to get statistically significant and meaningful results? Secondly, can a framework for selecting appropriate error terms specification and ACD extension be constructed for different durations like transaction, price, and volume? Finally, is it viable to use these results from ACD models on HFD for capturing the relationships between liquidity and returns? Moreover, what are the explanatory power of these coefficients from estimated models?

CHAPTER 2

LITERATURE REVIEW

The objective of this chapter is to present a review of the prior literature on ACD and liquidity. The first part analyses the most important liquidity studies in literature, and the second part is on both theoretical as well as empirical studies on ACD models.

2.1 Liquidity literature

Traditionally and practically, a market with a low transaction cost is defined as a liquid, while one with a high level of cost is illiquid. Since these costs affecting the transactions rely on many circumstances, such as the trade's size, timing, trading spot, counterparts, it is not easy to make the measurements. On the other hand, it can be said that the total information required for calculating the exact transaction cost is often unobtainable. Therefore, a wide range of measures utilized to assess liquidity appropriately should be identified.

Trading volume: It is described as the total number of shares or contracts traded for particular equity during a specified time horizon. It can be measured on any kind of security traded during a trading period. This measure is simple to calculate and available worldwide.

Trading frequency: It is the number of trades completed within a specific time window. The high trading frequency may be associated with higher liquidity, but on the other hand, it can also be related to increased volatility. The trade size is not taken into consideration while measuring the frequency.

Bid-Ask spread: Direct way to measure the cost of a small trade is generally computed as the difference between the bid and ask price as the bid-ask midpoint.

The spread determines the cost of performing a single exchange of a particular size. Since the data is widely available in real-time, it is easy to calculate straightforward. Given that there is a restricted number of waiting orders in the bid and ask spread, the bid-ask spread calculation cannot be utilized in large-scale transaction cost or liquidity calculations.

Quote size: This can be defined as the number of securities tradable at the bid and offer prices. It is precisely linked to market depth and accompanies the bid-ask spread. Measuring the quote size can be complicated for the majority of the markets since the market maker may have a regulation that makes it probable to hide some of the order book depth. (Iceberg orders etc.).

Trade size: Amount of securities traded at the corresponding bid and ask price points. Quantity traded will be different from the quantity quoted most of the time since transactions are not directly linked with the quoted quantity. This measure can alternate the usage of quote size.

Regarding all these measures used for calculating liquidity, different liquidity models have been developed by many researchers in the financial literature. In this study, we will mainly focus on comparing the models and theories with the popular Amihud's Illiquidity measure from 2002. However, first, it is relevant to review some of the most famous liquidity studies from the financial literature as well.

The capability of trading a large number of stocks with minimum price impact, cost, and delay can be labeled as liquidity. As Keynes explained in the 1930s, we can call an asset liquid if it is possible to trade it immediately without any loss in realization. All the characteristics that are briefly explained in the previous paragraph cannot be described in a unique measure. Therefore, an adequate measure of liquidity that denotes all these qualities continues to be a field of research.

We can categorize the stock market liquidity measurement literature into four different groups. The former literature suggests four primary characteristics of liquidity as trading volume, realization time, price impact, and trading(transaction) cost. These four different main characteristics can be relatively captured by various liquidity measures such as depth as volume and quantity measure, breadth as price impact measure, immediacy as time and speed measure, and finally, transaction costs as a cost measure. These measures differ from one another by the frequency of the data that is being used in the studies. While some measures are calculated on intraday high-frequency data, some others are measured in low frequency, for instance, monthly or yearly data.

Depending on the objectives and the calculation time windows of the corresponding works, measuring methodology displays numerous amount of variations. Various liquidity approaches have been used and proposed in the literature. As Chai et al. (2010) believe, it is impossible to find the best measure that can calculate every type of liquidity for various market systems and conditions because each kind of measure captures distinct characteristics of liquidity. As liquidity has multidimensional features, it cannot be captured by a single measure. The outcomes from separate measures of liquidity can denote different decisions. (Benic and Franic, 2008)

Many studies in the literature compare the liquidity measures with one another. Hallin et al. (2011) use the spread and volume-related liquidity, compare them with each other and inform that both procedures are negatively correlated and provide almost identical information about market liquidity. In another study of comparing liquidity measures, Marshall et al. (2013) found that three different measures of Gibbs, Amivest, and Amihud are all effective at measuring liquidity in the emerging markets. Moreover, as an additional study of assessing the measures and results, Bedowska-Sojka (2018) reports that the Amihud Illiquidity is the most accurate transactional cost-based measure in the literature, according to their findings.

From the point of view of market impact cost, Kyle (1985) and Breen, Hodrick, Korajczyk (2002) are both very important studies on explaining the illiquidity deriving from private information. They explain the illiquidity by generating a competitive and risk-neutral market maker modeling. They manage to show that the illiquidity will be present when two agents with different levels of information are against each other in any auction. The informed trader will be positioning himself according to the information he possesses so that this action will create an un-equilibrium between informed and uninformed agents.

Stoll (1978) and Amihud, Mendelson (1980) are essential papers concentrating on the bid-ask spread side of the liquidity. They all show that the fundamental source of illiquidity is the fragmentation of investors and markets, which means that not all market participants are present at the same and all time. For example, a seller may possibly arrive at the market when a genuine buyer is not present. This gap caused by the timing of the presence of the participants is filled by the market maker for the purpose of market continuity. However, this results in bearing additional risks for the market maker, and it should be compensated for this risk. Amihud and Mendelson (1980) and Stoll (1981) solve this problem by having the bid-ask prices depending on the market maker's inventory and modeling the illiquidity from inventory risk.

In the finance literature, there are also a collection of papers modeling liquidity from the point of view of search and delay costs, price and execution risks, and limits on trading. In both studies, Duffie et al. (2003) and (2005) manage to model the liquidity risk deriving from the search, bargain, and limits of trading, especially in the OTC markets. They conclude their study by explaining the positive relationship between search frictions and liquidity premium.

Datar, Naik, and Radcliffe (1998) measure the liquidity by share turnover, which is volume divided by the number of shares. They show that higher turnover is correlated with greater liquidity and find that liquidity (higher turnover) is linked with lower expected returns.

Brennan and Subrahmanyam (1996) estimate the liquidity from a Kyle-inspired model; Stoll and Whaley (1983) derive the liquidity as spread plus commissions from the direct and observable data, and Glosten and Harris (1988) approximate the mechanisms of bid-ask spreads as asymmetric information and inventory costs. Pastor and Stambaugh (2002) generate a liquidity proxy from their return reversal measures, and Lesmond, Ogden, and Trzcinka (1997) use the zero-return measure (proportion of zero daily returns). Pastor and Stambaugh (PS) measure built by Pastor and Stambaugh (2003) is acquired by the regression of daily returns in excess of daily market index returns with daily dollar volume.

Price impact size is assessed by the slope of the price function in Hasbrouck (2009). Goyenko et al. (2009) introduce the five-minute price impact by capturing the permanent price deviations in a 5-minute trade window. Bid-Ask spread's midpoints are measured from the beginning of the trade window until the close. Trzcinka et al. (2009) introduce a high correlated proxy with spread-related measures called FHT. They compare percentage cost and cost per volume proxies from intraday equity market data.

Liquidity is in different types of combination in related studies such as Vu et al. (2015), Lee (2011), and Bradrania and Peat (2013). In addition to stock market liquidity and calculating it with several different proxies, market liquidity risk with expected returns is also an essential issue for researchers to base their studies on. The market liquidity risk has been well calculated as the association between stock return, liquidity, and market return.

Vu et al. (2015) use the Liquidity Adjusted Capital Asset Pricing Model (LACAPM) and find out that liquidity risks are higher in bearish markets and positively correlated with market returns. Bradrania and Peat (2013) also reveal the existence of market liquidity risks' effects on stock returns. The occurrences of the liquidity risks are well impacting the expected returns, and it is already priced in the assets. Dang and Nguyen (2020) show the harmful effects of liquidity risks on stock returns across various global markets. Similarly, Nneji (2015) studies the strong relationship between stock market crashes and the hikes in stock market liquidity risks. Guloglu and Ekinci (2021) review the measures that are used for quantifying liquidity in empirical studies. Focusing on the high-frequency research, they show the differences and similarities according to different dimensions of liquidity and that market liquidity measures are categorizable according to their distinct characteristics.

The first liquidity proxy that has been used in the literature can be assessed as Keynes (1930), where liquidity is linked to the cost of realization of an asset's market order. The Sum of the buying premium and the selling discount, which is the spread between the bid and ask price, is accepted as the measure of illiquidity. As explained in previous paragraphs, building on both market and academic knowledge of the mid-1900s, the most famous studies on the market microstructure of liquidity and estimations of market liquidity proxies are shown in Table 1 below.

Table 1. Most-used Liquidity Proxies and their Remarks

Author/Year	Data Frequency	Proxy Dimension	Summary
Cooper et al. (1985)	Day	Price Impact	Amivest measurement
Chordia et al. (2001)	Minute	Transaction Cost	Bid-Ask Spread
Datar (1998)	Day or Month	Price Impact	Elasticity of Trades
Amihud (2002)	Day	Price Impact	Illiquidity Measurement, Sensitivity to dollar volumes
Pastor and Stambaugh (2003)	Day	Price Impact	Volume of stocks vs stock price impacts
Uddin (2009)	Month	Relative Measures	Systematic Liquidity vs Stock liquidity
Trzcinka et al. (2009)	Day and Minute	Comparative Analysis	High-Frequency Proxies Comparisons

These various liquidity proxies formed mainly by low-frequency data are commonly used in asset pricing research papers, but none of them are as popular as Amihud’s liquidity ratio. At this point, it is logical for our study to focus on Amihud (2002) and try to explain why the low-frequency liquidity proxies are insufficient for today’s financial markets.

The return premium linked with the universally used Amihud (2002) illiquidity measure is primarily accepted as a liquidity premium that offsets the price impact or transaction cost. A basic absolute value of the return to dollar volume ratio is utilized as a measure to calculate the liquidity premium. It is so simple to use and apply to any kind of trading data and uses only publicly available data. The main reason behind Amihud's (2002) popularity comes from this simplicity of application.

Even if the Amihud illiquidity is widely used by researchers for examining liquidity premium, building liquidity factors, or testing for liquidity, it is not clear whether the valuation of the Amihud is due to the price impact (stock liquidity) or other reasons. As it can be understood in the formula, the model excludes intraday trading activity like intraday volatility and price changes. The Amihud illiquidity is positively correlated with the "end of the day" return of the underlying asset. So, no matter what the intraday volume or the volatility of the related security is, if the closing price is equal or close enough to the opening price, this model will measure the illiquidity lower than it should be.

2.2 Autoregressive conditional duration literature

In recent years, with the introduced accessibility of the low-cost intraday transaction datasets, both the finance literature and market-wide financial analysts started to embrace the studies of High-Frequency Data (HFD). Most empirical studies using day-to-day data acquired by keeping either first or final observation of the time series are totally neglecting intraday events. They are being replaced by models using intraday databases, taking every single transaction into calculations. Using HFD and modeling it with econometrics, finance, and time-series statistics, a more profound knowledge of market activity and microstructure emerge. These new approaches make it possible to keep up with the constantly increasing speed of trading worldwide.

Recent studies in the market microstructure show that apart from traditional trade variables like price, volume, number of trades, etc., time is also carrying information and should, therefore, be modeled. Inspired by these arguments, the ACD model developed by Engle and Russell (1998), whose main objective is modeling the irregularly spaced transaction data, is used in this study.

The main concept of ACD is the trade duration which is essentially the waiting time between successive transactions of equity. (Engle and Russell, 1998) One of the essential features of this high-frequency trade data is that it is irregularly spaced in time. (Engle, 2000) Different kinds of distributions such as Exponential, Weibull, or Gamma are used to model these ACD structures. The flexibility and the versatility of the Weibull distribution become handy for this study of modeling the irregularly spaced duration data between the consecutive trades. (Meitz and Teräsvirta, 2006)

Time occurred between trades is informative, as proposed in both Easley and O'Hara (1992) and Easley et al. (1997). The empirical and theoretical research conducted on ACD models, which has drawn much attention on microstructure, trade intensity, and liquidity, are covered in research studies like (Bauwens and Giot, 2001; Engle, 2002; Hautsch, 2004; Tsay, 2002).

Building on the baseline model of Engle and Russell (1998), many researchers contribute to the literature by generating different types of ACD applications depending on their datasets and purposes. In the literature, there is a shared acknowledgment that the ACD models can be classified into three different generations. Exponential ACD (EACD), Gamma/Generalized Gamma ACD (GACD/GGACD), and Weibull ACD(WACD) can be considered as the most prevalent versions of ACD models that make use of different types of error term distributions since the model's inception. Models using Engle and Russell's baseline ACD with various error term specifications are classified as first-generation ACD models with the assumption of having the innovations following a distribution with nonnegative support.

There are also a significant amount of studies concentrating on the additional flexibility of parametric and nonparametric extensions. Box-Cox ACD (BCACD) model of Dufour and Engle (2000), the logarithmic version (LACD) of Bauwens and Giot (2000), threshold implementation (TACD) of Zhang et al. (2001), and augmented modeling (AACD) of Fernandes and Grammig (2006) can be referred as these second generation of ACD models.

Developing on the first-generation models and addressing the needs for additional flexibility of parametric and nonparametric extensions, there are significant amounts of studies expanding the standard ACD model for improved versatility. Box-Cox ACD (BCACD) model of Dufour and Engle (2000), the logarithmic version (LACD) of Bauwens and Giot (2000), threshold implementation (TACD) of Zhang et al. (2001), and augmented modeling (AACD) of Fernandes and Grammig (2006) can be referred as these second-generation of ACD models. Pacurar (2008) delivers a comprehensive review of these first and second generations of models that were developed prior to the global financial crisis. She concludes by declaring that the group of ACD-GARCH models, along with the connection between price durations and volatility, may help develop risk measures based on transaction data that are beneficial for active intraday traders. While some studies have already been introduced, she believes this subject would be given further attention.

Furthermore, Hautsch (2012) provides an interesting set of models to further explain recent developments and areas of usages of duration models. Motivated by the growing popularity of HFTs, he manages to present distinct styles of ACD models and illustrates their effectiveness and capabilities in different types of applications in his book, Econometrics of Financial High-Frequency Data. Starting with the standard ACD model of Engle and Russel (1998) and extending the model for specific purposes, the special cases are presented as; the Additive and Multiplicative Model (AMACD) for incorporating news impact curves, Box-Cox Model (BACD) for concave and convex news impact functions, Exponential Model (EXACD) of Dufour and Engle (2000) for capturing piece-wise linear impacts, Hentschel Model (HACD) by Fernandes and Grammig (2006) for introducing multiplicative stochastic component and Spline News Impact Model (SNIACD) by modeling the news response.

Bhogal and Variyam (2019) review the latest literature on conditional durations models in 2019. As they stated in their study, their paper can be noticed as an extension to Pacurar's study, where she studied the majority of "the first- and second-generation" models of conditional durations. They review the theoretical and empirical studies developed after 2008 and add the third-generation conditional duration models to the ACD literature review as an extension.

There are also some remarkable works on how to forecast some aspects of the monetary markets by using ACD models. Tse and Yang (2012) predict the intraday volatility of an equity, and Pyrlik (2013) delivers predictions on market crashes by using the same methodology. Huptas (2014) develops an empirical study on durations of selected stocks in the Polish stock exchange by using Bayesian inference on ACD models. Applying a logarithmic extension of ACD models, Gurgul and Syrek (2016) investigate and compares the microstructure of Polish and German stock exchanges.

The recent popularity of cryptocurrencies motivated a considerable effort to understand crypto markets' properties and microstructural behavior. The tech-savvy investor base and advent of HFT in this particular market require that particular attention be paid to high-frequency models to understand the sources of volatility and duration clustering. Dimpfl and Odelli (2020) use this duration perspective of ACD to estimate the probability of a significant price event in Bitcoin and investigate how Bitcoin features are associated with durations and volatility.

CHAPTER 3

METHODOLOGY AND MODEL

3.1 Standard ACD model and diurnal adjustments

As mentioned before, the HFD data are irregularly time-spaced and statistically considered as point processes. As explained in Bauwens and Giot (2001), an HFD contains a collection of information on financial trades, and “the times” of these events serve as the arrival times of the point process. Engle and Russell (1998) use this data as a basis for their study of the trading activity and market microstructure investigation.

A point process of the durations between market transactions is modeled by Engle and Russell (1998), and the model is identified as the ACD model. Since price or volume events are associated with the arriving time, the mechanism can be labeled as a marked point process. The stochastic process of the ACD model adopts a series of time points $\{t_0, t_1, \dots, t_n, \dots\}, t_0 < t_1 < \cdots < t_n$. If we call N(t), the quantity of events that occurred before the time t, the conditional intensity process can be written as:

$$\lambda\left(t \mid N(t), t_1, \dots, t_{N(t)}\right) = \lim_{\Delta t \to 0} P\left(N(t+\Delta t) > N(t) \mid N(t), t_1, \dots, t_{N(t)}\right)/\Delta t \qquad (1)$$

This function (1) can be specified as a hazard function, and as it is assumed in Engle and Russell (1998), it links the duration ahead with the likelihood of the next event occurring. The parameterized duration function is assumed as the conditional probability of transaction i, occurred at time t provided the duration among t_i and t_{i-1} where $x_i = t_i - t_{i-1}$ is the marginal duration between two consecutive transactions. As it is suggested by Engle and Russell, if we let ψ_i be the conditional expectation of i th duration, it can be written as:

$$E(x_i \mid x_{i-1}, \dots, x_1) = \psi_i(x_{i-1}, \dots, x_1; \theta) \equiv \psi_i \tag{2}$$

Furthermore, the standardized durations of the ACD model can be assessed as:

$$\varepsilon_i = \frac{x_i}{\psi_i} \tag{3}$$

are independent and identically distributed, which is iid. This indicates that the conditional expected duration can capture all the temporal dependencies and that ACD is characterized by the description of the conditional duration ψ_i and the distribution of ε_i. Engle and Russell (1998) show that the recent p durations characterize the conditional duration and that a conventional model of ACD(p,q) can be described as:

$$\psi_i = \omega + \sum_{j=0}^{p} \alpha_j x_{i-j} + \sum_{j=0}^{q} \beta_j \psi_{i-j} \tag{4}$$

p and q are lag orders.

The specification of error term ε_i arranges the ACD models into different categories. Exponential and Weibull distributions are the two most utilized distributions of the error term ε_i. The preference of distribution in (4) influences the conditional intensity or the hazard function of the ACD model. The flat conditional hazard function implied by Exponential distribution leads to the refusal of duration clustering. As it is indicated in many empirical studies like; Engle and Russell (1998), Dufour and Engle (2000), Feng et al. (2004), Lin and Tamvakis (2004), and various others, using a flat hazard function implied by Exponential distribution is limiting and can be simply omitted in empirical applications. The shape of the

Weibull distribution leads to more flexible analysis and fits the idea of duration clustering, which can even be seen in raw data.

The Weibull distribution with parameters (κ γ), the hazard function can be described as $h(x) = \kappa^{\gamma}\gamma^{\gamma-1}\gamma$ and the conditional intensity take the form:

$$\lambda\left(t \mid x_{N(t)}, \ldots, x_1\right) = \left(\Gamma(1+1/\gamma)\psi_{N(t)+1}^{-1}\right)^{\gamma}\left(t - t_{N(t)}\right)^{\gamma-1}\gamma \tag{5}$$

The two-parameter family is obtained by Γ being the gamma function and γ being the Weibull parameter, which specification indicates an increasing or decreasing hazard function. Depending on the Weibull parameter being greater or less than one, the hazard function is upward or downward sloping. The log-likelihood of the Weibull function decreases to the log-likelihood of exponential ACD if the parameter is equal to one. These parameters are acquired by using maximum-likelihood estimation. Finally, the log-likelihood function for Weibull ACD can be assessed as:

$$\log\ L = \sum_{i=1}^{N(T)} \ln\left(\gamma/x_i\right) + \gamma \ln\left(\Gamma(1+1/\gamma)x_i\psi_i^{-1}\right) - \left(\Gamma(1+1/\gamma)x_i\psi_i^{-1}\right)^{\gamma} \tag{6}$$

The strong intraday seasonality is a well-known characteristic of the HFD. Many researchers like Bollerslev and Domowitz (1993), Beltratti and Morana (1999), and several others report the same high-frequency trading data characteristics in their studies. Engle and Russell (1998) state that greater trading activity at the opening and closing of the trading sessions leads to smaller durations between transactions. It can be linked to the investor’s behaviors. Before proceeding with modeling of the durations, the intraday durations are decomposed into two parts: deterministic and stochastic. This method of removing the intraday seasonality effects is initially employed by Engle and Russell (1998). The diurnally adjusted durations $\tilde{x_i}$, of the durations x_i and the expected duration θ_ψ is given by (7) and (8), respectively:

$$\tilde{x}_i = x_i/\phi(t_{i-1};\theta_\phi) \tag{7}$$

$$E_{i-1}(x_i) = \phi(t_{i-1};\theta_\phi)\psi_i(\tilde{x}_{i-1},\ldots,\tilde{x}_1;\theta_\psi) \tag{8}$$

Two sets of parameters stated in function (5) are estimated using maximum likelihood with a two-stage estimation method. First, a cubic spline function is used to remove the intraday effects of duration, and then these diurnally adjusted durations are utilized in the ACD model.

3.2 Distributional assumptions for error term specifications

Because of the ease of use and application, the exponential distribution is a natural option very appropriate for ACD estimations. The standard exponential distribution with a shape parameter equal to one is used by Engle and Russell (1998) and is called the EACD model. A significant advantage of employing this distribution is that it provides a quasi-maximum likelihood (QML) estimator for the ACD parameters (Engle and Russell, 1998). It has been shown by Drost and Werker (2004) that a satisfactory level of consistency of the estimates is achieved when the QML estimation is built on the standard gamma family. The quasi-likelihood function can be written as:

$$L(\theta) = -\sum_{i=1}^{N(T)} \left[\frac{x_i}{\psi_i} + \log\ \psi_i\right] \tag{9}$$

Engle and Russell (1998) validate the EACD (1,1) by the resemblance between the ACD and GARCH models by applying the same results on the QMLE properties of the GARCH (1,1). Under the preliminary conditions of their model, the estimates of θ are achieved by maximizing the likelihood function given above (9).

The QML estimation produces steady estimations, and the processes in the case of exponential distributions are simple to execute, but this happens at the cost of efficiency. In practical applications, fully efficient maximum likelihood estimates might be best to use. However, in contrast, the selection of the distribution of the error term in the methodology affects the conditional intensity and hazard function of the ACD model. The exponential specification entails a fixed conditional hazard function that is relatively restricting and might be effortlessly rejected in practical economic applications. For better versatility and adaptability, Engle and Russell (1998) employ a standard Weibull distribution with a scale parameter being equal to one, and the following model is called WACD. The log-likelihood of the Weibull function decreases to the log-likelihood of exponential ACD if the parameter is equal to one. The Weibull parameter dictates the allowance of increasing or decreasing hazard function by being greater or less than one. As shown in function (6), these parameters are acquired using maximum-likelihood estimation.

Encouraged by the explanatory evaluation of observed price and volume durations, Grammig and Maurer (2000) cast doubt on the notion of monotonicity of the hazard function in Engle and Russell's standard ACD models. They support using a Burr distribution, which can be derived as a Gamma mixture of Weibull distributions (Lancaster, 1992), containing the Exponential, Weibull, and log-logistic as special cases. This specification of ACD using the Burr distribution is called Burr-ACD.

Having the same doubts on the standard ACD models parallel to Gramming and Maurer, Lunde (1999) suggests the usage of the generalized gamma distribution that leads into the G-ACD model. Both distributions used by the researchers permit similar shaped hazard functions, equally dependent on two parameters, for referring to circumstances where, for minor durations, the hazard function is escalating and, for lengthy durations, the hazard function is declining. Moreover, Hautsh (2002; 2004, 2012) compares the possible error term specifications and suggests Generalized F distributions that include Weibull and Log-logistic as special cases.

The ACD model has been expanded in numerous aspects, focusing primarily on increasing the conventional details of durations. The solid connection amongst the ACD and GARCH models encouraged the rapid development of complementary measurements of conditional durations. All these possible error term distributions and the ones used in this study can be found in Table 2. The most popular generalizations of the ACD methodology are reviewed in the following subsection of Standard ACD Model Extensions.

Table 2. Error Term Distributions

Weibull:	Burr (as in Grammig and Maurer, 2000):
$\varepsilon \sim W(\theta,\gamma), \qquad \theta,\gamma > 0$ $f_\varepsilon(\varepsilon) = \theta\gamma\varepsilon^{\gamma-1}e^{-\theta\varepsilon^{\gamma}}$ $E(\varepsilon) = \theta^{-\gamma^{-1}}\Gamma(\gamma^{-1}+1)$ Forcing $E(\varepsilon) = 1$; $\theta = [\Gamma(\gamma^{-1}+1)]^{\gamma}$ with Hazard Function: $h(\varepsilon) = \theta\gamma\varepsilon^{\gamma-1}$ (10)	$\varepsilon \sim Burr(\theta,\kappa,\sigma^2);\ \theta,\kappa,\sigma^2 > 0, \kappa > \sigma^2$ $f_\varepsilon(\varepsilon) = \frac{\theta\kappa\varepsilon^{\kappa-1}}{(1+\sigma^2\theta\varepsilon^{\kappa})^{\frac{1}{\sigma^2}+1}}$ $E(\varepsilon^s) = \theta^{-\frac{s}{\kappa}} \times \frac{\Gamma\left(1+\frac{S}{\kappa}\right)\Gamma\left(\frac{1}{\sigma^2}-\frac{S}{\kappa}\right)}{\sigma^2\left(1+\frac{S}{\kappa}\right)\Gamma\left(\frac{1}{\sigma^2}+1\right)}$ Forcing $E(\varepsilon) = 1$; $\theta = \left(\frac{\Gamma\left(1+\frac{1}{\kappa}\right)\Gamma\left(\frac{1}{\sigma^2}-\frac{1}{\kappa}\right)}{\sigma^2\left(1+\frac{1}{\kappa}\right)\Gamma\left(\frac{1}{\sigma^2}+1\right)}\right)^{\kappa}$ (11)
Generalized Gamma:	**Generalized F (Hautsch,2012):**
$\varepsilon \sim GG(\gamma,\kappa,\lambda), \qquad \gamma,\kappa,\lambda,\varepsilon > 0$ $f_\varepsilon(\varepsilon) = \frac{\gamma\varepsilon^{k\gamma-1}}{\lambda^{k\gamma}\Gamma(\kappa)}\exp\left\{-\left(\frac{\varepsilon}{\lambda}\right)^{\gamma}\right\}$ $E(\varepsilon^s) = \lambda^s\frac{\Gamma(\kappa+s/\gamma)}{\Gamma(\kappa)}$ Forcing $E(\varepsilon) = 1$; $\lambda = \frac{\Gamma(\kappa)}{\Gamma\left(\kappa+\frac{1}{\gamma}\right)}$ (12)	$\varepsilon \sim GF(\eta,\gamma,\kappa,\lambda), \qquad \eta,\gamma,\kappa,\lambda > 0$ $f_\varepsilon(\varepsilon) = \frac{\gamma\varepsilon^{\kappa\gamma-1}[\eta+(\varepsilon/\lambda)^{\gamma}]^{-\eta-\kappa}\eta^{\eta}}{\lambda^{\kappa\gamma}B(\kappa,\eta)}$ where $B(\kappa,\eta) = \frac{\Gamma(\kappa)\Gamma(\eta)}{\Gamma(\kappa+\eta)}$ $E(\varepsilon^s) = \lambda^s\eta^{s/\gamma}\frac{\Gamma(\kappa+s/\gamma)\Gamma(\eta-s/\gamma)}{\Gamma(\kappa)\Gamma(\eta)}$ Forcing $E(\varepsilon) = 1$; $\lambda = \frac{\Gamma(\kappa)\Gamma(\eta)}{\eta^{1/\gamma}\Gamma(\kappa+1/\gamma)\Gamma(\eta-1/\gamma)}$ (13)

3.3 Standard ACD model extensions

As it is deeply explained in Hautsch (2012), the class of generalized polynomial random coefficient ACD models includes a variety of extensions of the standard ACD model permitting for additive and multiplicative stochastic components known as specifications. This time, differing from the standard ACD model, the conditional mean function includes lagged innovations additively and/or multiplicatively. Furthermore, it includes parameterizations describing not just the linear but also further adaptable news impact curves. For ease of demonstration, the following methodology will be based on the models with a lag order of p and q equal to one. In the following sections, we will be discussing and demonstrating the methodology for Logarithmic ACD (LACD) of Bauwens and Giot (2000) and Lunde (1999), Additive and Multiplicative ACD (AMACD) of Hautsch (2012), and Augmented Box-Cox ACD (ABACD) of Hautsch (2012).

3.3.1 Logarithmic ACD

As an advanced and more complex form of the ACD model, Bauwens and Giot (2000) present the LACD model. In the standard ACD (m, q) model, appropriate conditions are required for the parameters to confirm the positivity of durations. When it is needed to add some variable taken from the market microstructure and have a possibility of negative coefficients, the durations may become negative. Bauwens and Giot (2000) propose the more adaptable LACD (m, q) model in which the logarithm of the conditional duration is used to prevent this situation. The logarithmic design permit introduction of supplementary variables into the methodology, deprived of sign limitations on their coefficients.

Both Bauwens and Giot (2000) and Lunde (2000) suggest the LACD model that guarantees the non-negativity of durations without any restrictions on parameters. The model is obtained by $x_i = \psi_i \varepsilon_i$ with (14),

$$\begin{aligned} \ln \Psi_i &= \omega + \alpha \ln \varepsilon_{i-1} + \beta \ln \Psi_{i-1} \\ &= \omega + \alpha \ln x_{i-1} + (\beta - \alpha) \ln \Psi_{i-1} \end{aligned} \tag{14}$$

where ε_i is i.i.d with mean equals to one. Differing from the standard ACD model, the LACD model entails a concave relationship between ε_{i-1} and x_i. Negative surprises from the news impact, which is ε_i <1, on x_i is greater than the positive surprises with ε_i >1. Similar to the linear ACD model introduced in the first part of the methodology, in the case of distribution with a mean non-equal to one, the model can be written as:

$$\begin{aligned} x_i &= \Psi_i \tilde{\varepsilon}_i / \zeta =: \Phi_i \tilde{\varepsilon}_i \\ \ln \Phi_i &= \tilde{\omega} + \alpha \ln \varepsilon_{i-1} + \beta \ln \Phi_{i-1} \end{aligned} \tag{15}$$

where $\tilde{\omega} := \omega + (\beta - 1)\ln \zeta$ and $\varepsilon_i := \tilde{\varepsilon}_i / \zeta$ with $E[\tilde{\varepsilon}_i] := \zeta \neq 1$

Finally, the general form of LACD (p, q) of Bauwens and Giot can be stated as:

$$\ln \mu_i = \omega + \sum_{j=1}^{p} \alpha_j \ln \varepsilon_{i-j} + \sum_{j=1}^{q} \beta_j \ln \mu_{i-j} \tag{16}$$

Inside the ACD literature, it is discussed that the standard ACD model given in the first section of methodology cannot completely capture the possible nonlinear dependencies between conditional durations and the past durations from the financial market data. To be specific, the authors state that linear models overpredict the conditional durations after extremely short or long durations. This promotes and encourages new methods looking for more versatile functional forms that might permit characteristic responses to small and large shocks in durations. Following the concerns in the finance literature on the versatility of ACD models, we will also be discussing AMACD and ABACD methodology in the following sections for our study.

3.3.2 Augmented ACD models

In this section, the methodology for two different types of generalized polynomial random coefficient ACD models is published. First is the Additive and Multiplicative ACD known as AMACD(p,r,q), and the second is Augmented Box-Cox ACD, known as ABACD(p,q). As explained in 3.3, with the additive and multiplicative form of the model, more flexible news impacts can be captured with parametrizations.

3.3.2.1 Additive and Multiplicative ACD (AMACD) Model

The ACD specification with additive and multiplicative extension for lag levels of P=Q=1 can be given by:

$$\Psi_i = \omega + (\alpha\Psi_{i-1} + v)\varepsilon_{i-1} + \beta\Psi_{i-1} \tag{17}$$

Where v is the parameter letting the model imply a news impact curve with the slope of $\alpha\Psi_{i-1} + v$. Therefore, the lagged enhancements come into the conditional mean function additively, along with multiplicatively. In this perception, the Additive and Multiplicative ACD (AMACD) model is more adaptable and includes the standard ACD model in the case of v =0.

A more general form of AMACD (p,r,q) of Hautsch (2012) can also be written as:

$$\mu_i = \omega + \sum_{j=1}^{p} \alpha_j x_{i-j} + \sum_{j=1}^{r} v_j \varepsilon_{i-j} + \sum_{j=1}^{q} \beta_j \mu_{i-j} \tag{18}$$

In which, as for the standard model, we assume that the duration $x_i = \mu_i \varepsilon_i$, where the error term ε_i is i.i.d and distributed with a mean E (ε_i) =1.

3.3.2.2 Augmented Box-Cox ACD

By applying the methodology for developing the asymmetric GARCH presented by Hentschel (1995), an extension of the ACD model, which is called ABACD, is obtained by Hautsch (2012):

$$\Psi_i^{\delta_1} = \omega + \alpha\big(|\varepsilon_{i-1} - b| + c(\varepsilon_{i-1} - b)\big)^{\delta_2} + \beta \Psi_{i-1}^{\delta_1} \tag{19}$$

In this design, the parameter b is related to the location of the kink and the δ_2 is the determinant of the shape of function around b. For $\delta_2 > 1$, the shape is convex, and for $\delta_2 < 1$ it is concave. The ABACD model is specified on an additive stochastic component. The more generalized for ABACD(p,q) can be provided by:

$${\mu_i}^{\delta_1} = \omega + \sum_{j=1}^{p} \alpha_j \big(|\varepsilon_{i-j} - v| + c_j|\varepsilon_{i-j} - b|\big)^{\delta_2} + \sum_{j=1}^{q} \beta_j {\mu_{i-j}}^{\delta_1} \tag{20}$$

Although this design of the news impact function makes it available for more elasticity, it holds one main disadvantage since the parameter restriction |c|<=1 must be imposed to be able to bypass complicated values when $\delta_2 \neq 1$.

CHAPTER 4

DATA AND MARKET DESCRIPTION

The data used in this study and the BIST stock exchange is described in this chapter.

4.1 BIST market information

Full intraday order book data that has been purchased from BIST, which consists of order book information of every single Borsa Istanbul stock and derivatives from the end of the year 2019, is used in this study. The total data that is gathered is approximately around one tb. The study sample is selected as the 10 most traded stocks in BIST. A stock must satisfy certain conditions not to be excluded from our sub-sample. The stock is not supposed to be under "special treatment" by BIST Administration during the research interval since it would artificially manipulate the pricing and duration activities. In their study, Zhang (2001) shows that the different price limits and trading rules would affect the trading activities in a biased sense.

According to our selection methodology, our sample stocks and their tickers are as follows; GARAN.E, THYAO.E, KCHOL.E, EREGL.E, TUPRS.E, ISCTR.E, BIMAS.E, PETKM.E, EKGYO.E, SAHOL.E.

4.2 Data and filtering

The duration data for the study are extracted from BIST orderbook dataset from September 2019 to November 2019. Each trade is recorded with label, datetime, price, volume, bid price, and ask price information. The total dataset of BIST equities consists of 23.3 million observations. To escape wrong calculations, the data is categorized in a sense that only the durations from continuous trading sessions are kept. All the data before 10:00 a.m. and after 6:00 p.m. are removed. The trading days in which the stock price would reach its trading limits of 10% are also omitted from the database. (None of the 10 most traded stocks reached this limit in this study). After the selection and filtering process, the subset of 10 most traded stocks in BIST, which is around 3 million observations, is used in our models. Table 3 demonstrates the summary statistics and the raw data information of the selected stocks.

Table 3. Descriptive Statistics for 10 most Traded Stocks in our Sample

Stock	Stats	Price	Duration	Volume	Ask	Bid	Stock	Stats	Price	Duration	Volume	Ask	Bid
BIMAS	N	124580	124580	124580	124580	124580	KCHOL	N	133892	133892	133892	133892	133892
	Mean	47.79	4.85	13330.73	47.80	47.76		Mean	18.42	4.51	12236.23	18.43	18.42
	Max	59.55	831	127000000	49.86	49.84		Max	21.54	469	11400000	19.44	19.43
	Min	36.8	0	38.96	46.06	46.02		Min	14.37	0	15.08	16.95	16.88
	Sd	0.87	14.84	361243.3	0.83	0.83		Sd	0.54	14.73	43915.07	0.54	0.54
EKGYO	N	157016	157016	157016	157016	157016	PETKM	N	256741	256741	256741	256741	256741
	Mean	1.24	3.84	13701.16	1.24	1.23		Mean	3.48	2.35	17429.52	3.49	3.48
	Max	1.57	321	2480000	1.42	1.41		Max	4.33	321	3406362	3.66	3.65
	Min	1.04	0	1.08	1.09	1.08		Min	2.68	0	2.8	3.33	3.32
	Sd	0.07	9.85	69564.78	0.07	0.07		Sd	0.07	5.53	88310.88	0.078	0.079
EREGL	N	295448	295448	295448	295448	295448	SAHOL	N	169238	169238	169238	169238	169238
	Mean	6.64	2.04	13287.22	6.65	6.64		Mean	8.82	3.57	12287.63	8.83	8.81
	Max	8.28	324	2835849	6.99	6.98		Max	11.1	339	6438474	9.68	9.67
	Min	5.17	0	5.26	6.43	6.42		Min	6.63	0	8.18	8.18	8.17
	Sd	0.11	5.48	41513.47	0.11	0.11		Sd	0.34	10.28	35133.93	0.34	0.34
GARAN	N	483272	483272	483272	483272	483272	THYAO	N	896539	896539	896539	896539	896539
	Mean	9.38	1.25	38156.45	9.38	9.37		Mean	11.60	0.67	31125.56	11.61	11.60
	Max	12.02	323	28000000	10.28	10.27		Max	14.85	321	20200000	12.4	12.39
	Min	6.68	0	6.68	8.25	8.24		Min	8.6	0	8.6	10.6	10.59
	Sd	0.50	3.94	132348.3	0.50	0.50		Sd	0.39	2.09	106848.5	0.39	0.39
ISCTR	N	176486	176486	176486	176486	176486	TUPRS	N	284780	284780	284780	284780	284780
	Mean	5.95	3.42	17960.79	5.96	5.94		Mean	127.96	2.12129	18243.99	128.02	127.90
	Max	7.38	395	8760000	6.34	6.33		Max	169.6	322	14900000	143.8	143.7
	Min	4.48	0	4.48	5.49	5.47		Min	94.9	0	97	117.4	117.3
	Sd	0.17	10.03	48191.59	0.16	0.16		Sd	7.49	6.20	70526.22	7.45	7.45

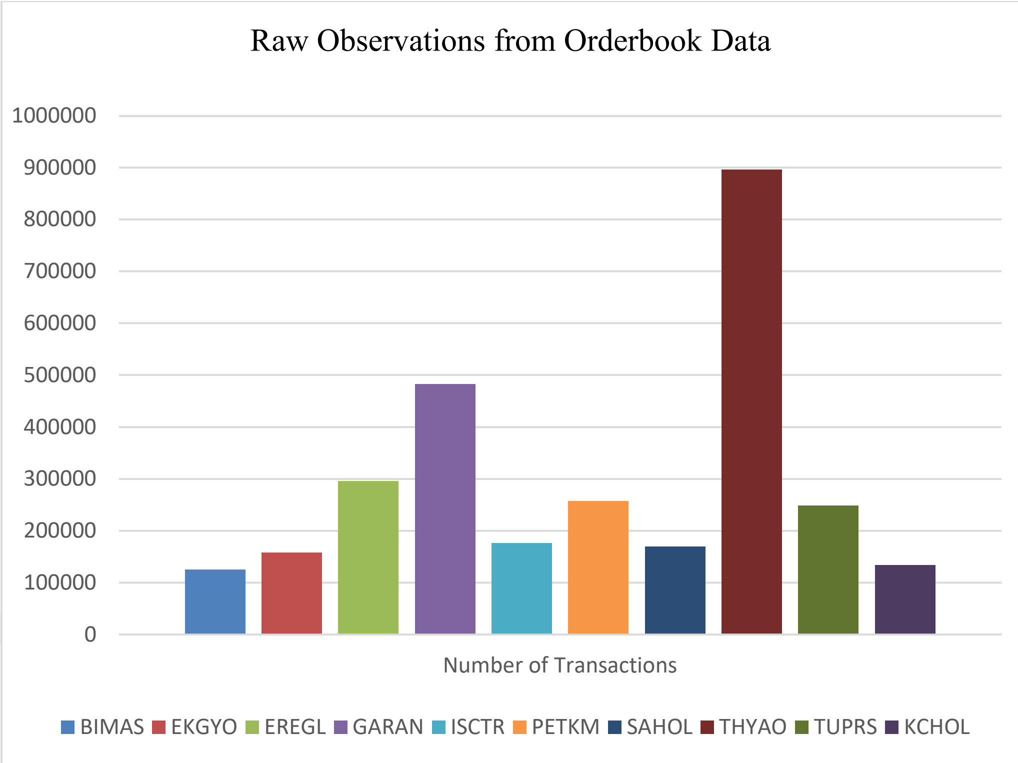


Figure 1. Number of observations for each equity separately

CHAPTER 5

EMPIRICAL RESULTS

5.1 ACD applications to transaction, price and volume durations

A Framework for various ACD models and extensions is conducted in this section.

5.1.1 Diurnal adjustments

As explained in Methodology with the functions (7) and (8), a cubic spline method is applied for removing the intraday effect from all our selected stocks separately. For demonstrating the removal of the seasonality from the data, most traded equity in the sub-sample THYAO.E is chosen, and the results of the diurnal adjustments for transaction durations are reported.

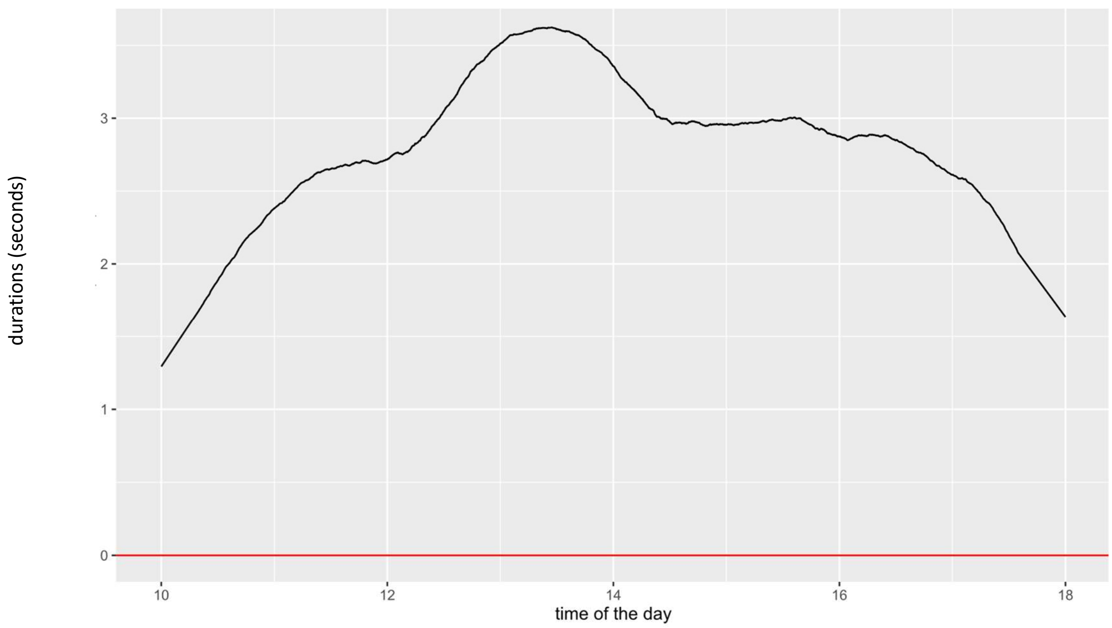


Figure 2. Diurnally adjusted transaction durations estimated by cubic spline function without day of the week aggregation.

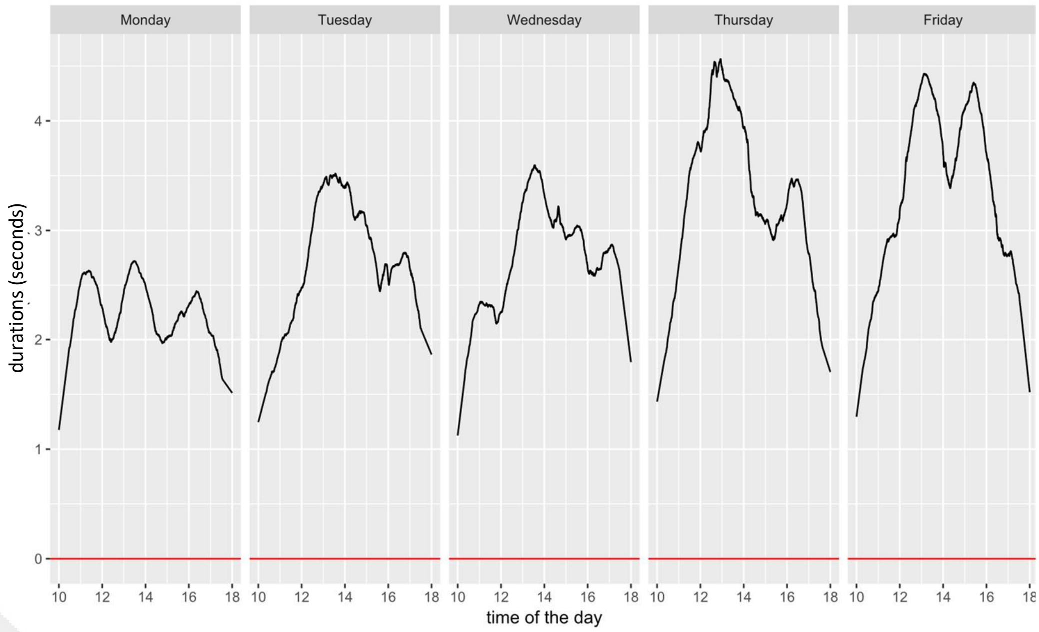


Figure 3. Diurnally adjusted transaction durations estimated by cubic spline function with day of the week aggregation

896539 transactions of THYAO are used for estimation of diurnal factors both for daily and day of the week specifications. A few interesting things can be observed from the outputs in Figure 2 and Figure 3. It can be clearly observed that durations are longer in the middle of the day or "lunch time" and shorter at the end of the day when traders are more involved. It can also be seen that the durations tend to get shorter and shorter after 15:30 local time, which is the opening of U.S stock markets.

The pattern does not seem to differ broadly between the durations occurring on the different days of the week. That is why it is logical not to apply a weekday aggregation and force different weekdays to have different diurnal adjustments. We continue with the intraday diurnal aggregation since we observe the same trading pattern with every equity in our sample. The results from diurnal adjustments in Figure 4 and via the correlogram of the ACD fitted BIST equities in the next chapter.

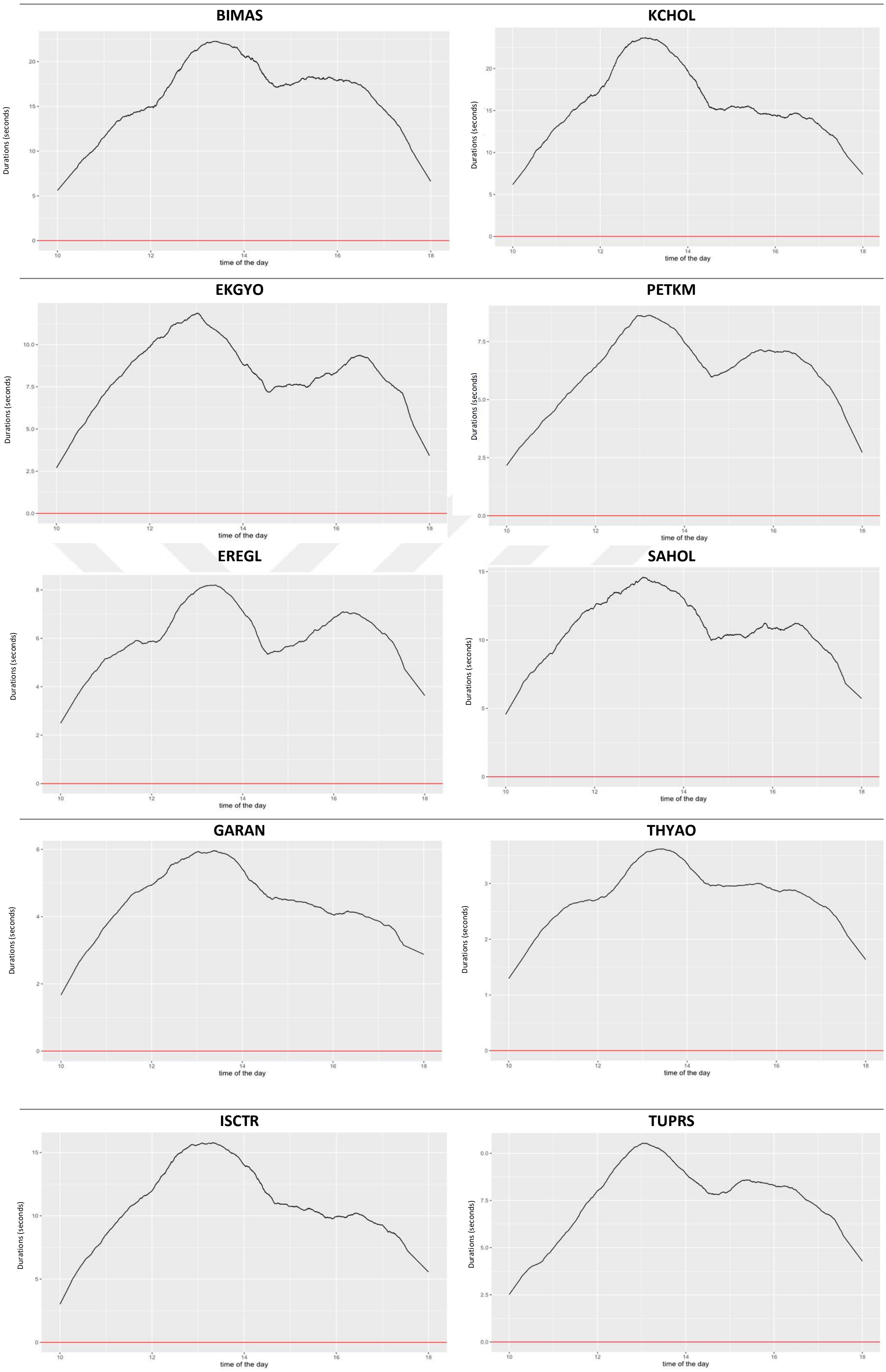


Figure 4. Diurnally adjusted transaction durations estimated by cubic spline function with daily aggregation for all stocks in our sample.

Diurnally adjusted durations are almost equal across whole equities in our sample from BIST. This similarity might be explained by the market-wide trading activity being parallel and correlated between different stocks. High-frequency trading algorithms of the institutional traders and analogous financial behaviors (bandwagon effect) of local traders clearly add up to these clustering effects that we observe in durations. This clustering of time events should be paid attention to in every single intraday high-frequency analysis to be able to understand the microstructure of the money markets.

5.1.2 Autoregressive conditional duration model for transaction durations

For the second step of the maximum likelihood method, diurnally adjusted transaction durations of all stocks are fitted by using an ACD(p,q) as in equation (8). As parallel to the literature, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are used to verify the optimal lag length and the lag length of 1 for both p and q, which is WACD (1,1) fits the data best. As it has been done in the previous chapter, THYAO has been chosen to further clarify why it is decided to use Weibull distribution for the durations and to understand the differences between different choices of the distribution of the error terms. Before getting into the results, an advanced explanation of this reasoning behind the distribution selection can be made by using comparative tables, quantile-quantile plots, hazard function graphics and scatter plots of conditional durations versus residuals.

Both Pacurar in 2008 and Bhogal, Variyam in 2019, well-rounded reviews in the ACD literature, report that Exponential, Weibull, Burr, and Generalized Gamma distributions are the most used error specification for capturing information from different types of financial duration data. Because of this reasoning, an ACD (1,1) with four different error term specifications is applied, and the results are reported in Table 4 and Figure 5. (Bhogal & Variyam, 2019; Pacurar, 2008)

Table 4. ACD (1,1) application for four different distributions; Weibull, Exponential, Gen.Gamma, and Burr

Weibull	Coef	SE	PV	Goodness	Value	Exponential	Coef	SE	PV	Goodness	Value
ω_1	0.003	0.000	0	LLH	-163661	ω_1	0.003	0.000	0	LLH	-154261
α_1	0.080	0.001	0	AIC	327330	α_1	0.080	0.001	0	AIC	328582.1
β_1	0.919	0.001	0	BIC	327370.7	β_1	0.918	0.001	0	BIC	328836.9
γ_1	1.383	0.002	0	MSE	1.05					MSE	1.079841
Gen.Gamma	Coef	SE	PV	Goodness	Value	Burr	Coef	SE	PV	Goodness	Value
ω_1	0.028	0.001	0	LLH	-99103	ω_1	0.021	0.001	0	LLH	-152064.9
α_1	0.106	0.002	0	AIC	298305.8	α_1	0.118	0.005	0	AIC	304209.9
β_1	0.868	0.003	0	BIC	298730.6	β_1	0.866	0.005	0	BIC	304549.6
κ_1	22.637	3.106	0	MSE	0.97	γ_1	1.170	0.009	0	MSE	1.00
γ_1	0.179	0.013	0								

Note: All p-values are statistically significant. The goodness of fit values of loglikelihood, AIC, BIC, and MSE are all reported.

When the results are compared in Table 4, it is observed that all four specifications can be used according to their goodness of fit values with Weibull and Exponential distribution fitting the data slightly better than the other two. Another standard method of assessing the closeness between theoretical distributions and the empirical samples is studying the quantile-quantile plots visually. The quantiles of samples are plotted against the theoretical quantiles, and the deviation from each other is examined, with lower deviation being a better fit. In Figure 6, we report Q-Q plots for four different error distributions.

Parallel to Table 4, it is seen that residual quantiles are closer to theoretical quantiles for Weibull and Exponential distribution. The deviation from the assumed error term distribution is more extensive for both Generalized Gamma and Burr distribution compared with the other two.

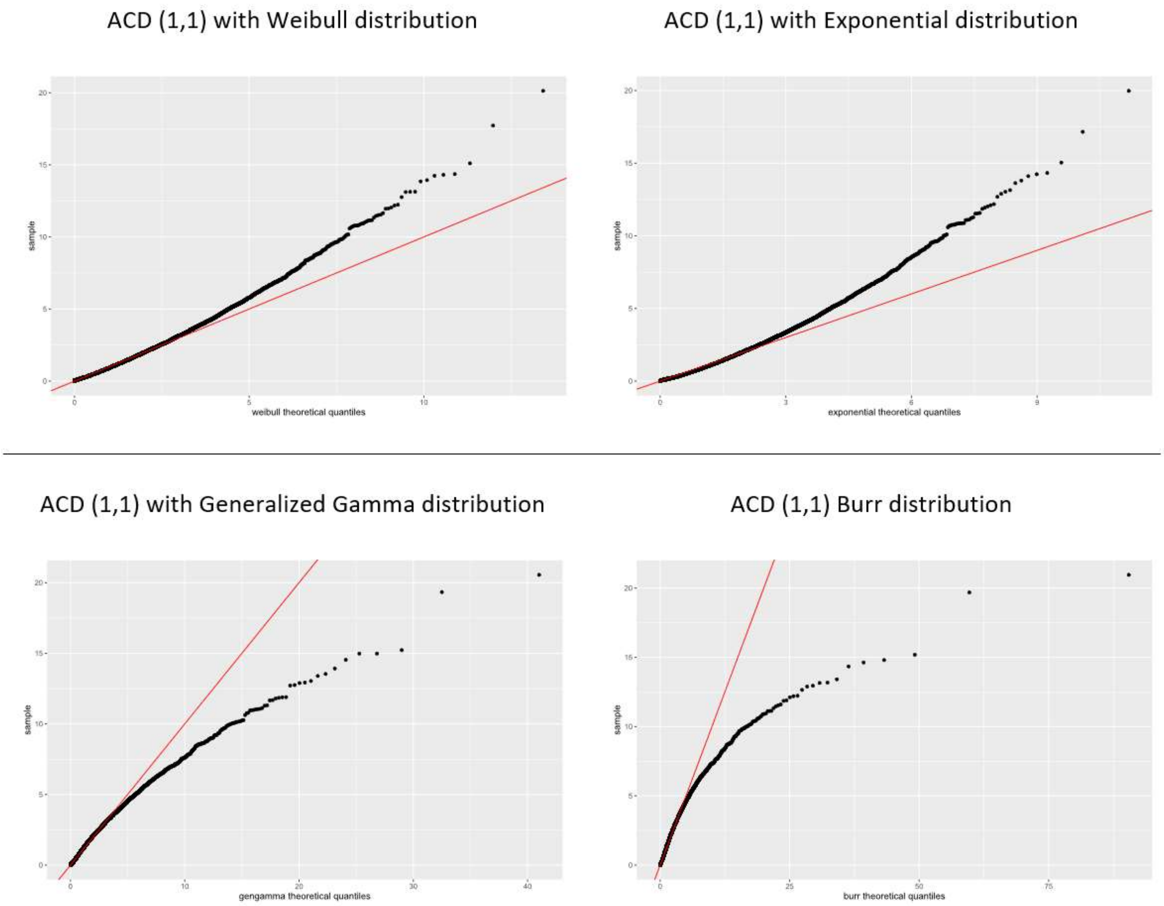


Figure 5. Q-Q plots for four different error term distributions; Weibull, Exponential, Generalized Gamma and Burr

The fact that the sample quantiles lie below the red line imposed by the distribution function indicates that the empirical distribution has a different tail than the applied distribution. At this point, the hazard functions of Weibull and Exponential distribution are plotted to compare the results graphically. Both the non-parametric estimate of the hazard function from the residuals and the hazard function implied from the model estimate are plotted. The empirical hazard function of the residuals from the non-parametric estimator is from Engle and Russell. When the residuals are compared with the estimated parameters of the selected error distribution, it can be clearly found out that Weibull is a better fit for the sample. (Engle and Russell, 1998)

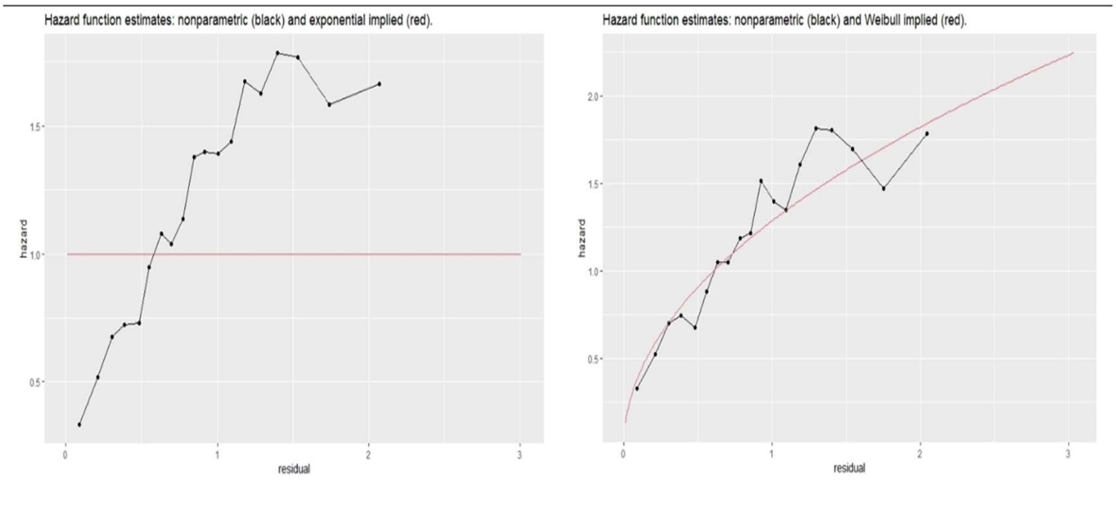


Figure 6. Hazard functions from the residuals estimated from ACD (1,1) with Exponential (left) and Weibull(right) error term specifications

For the exponential distribution, the hazard function is flat and equal to one as the distribution is forced to have a unit expectation which is explained in the methodology chapter. For the Weibull error term specification, it is also investigated how the estimated conditional means are predicting the mean size of the upcoming durations. In figure 7, the residuals and their means are plotted against the estimated conditional means. As expected, the mean of the residuals is constant across each level of conditional durations.

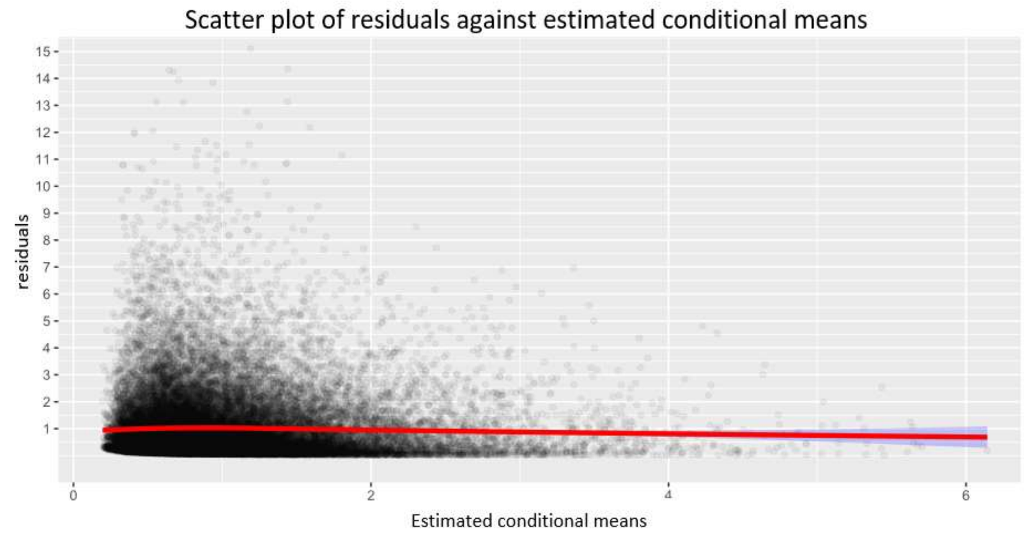


Figure 7. Plot of residuals versus conditional means

Weibull error term specification is used for the ACD estimation of 10 most traded stocks, and the results are reported in following Tables 5 and 6. The computational results confirm the autoregressive component of the durations data.

Table 6 shows the decreasing autocorrelation of durations due to the removal of diurnal components and ACD estimations. The autocorrelation behavior is similar across all the equities. The Weibull term is statistically significant at resolving the shape of the hazard function.

Table 5. Estimates from the WACD (1,1) model applied to each equity in sub-sample separately

BIMAS	Coef	SE	PV	Goodness	Value	KCHOL	Coef	SE	PV	Goodness	Value
ω_1	0.023	0.002	0	LLH	-33206.49	ω_1	0.017	0.001	0	LLH	-32281.28
α_1	0.104	0.004	0	AIC	66420.99	α_1	0.092	0.004	0	AIC	64570.56
β_1	0.875	0.005	0	BIC	66454.96	β_1	0.892	0.005	0	BIC	64604.54
γ_1	0.919	0.004	0	MSE	1.777	γ_1	0.916	0.004	0	MSE	1.89
EKGYO	Coef	SE	PV	Goodness	Value	PETKM	Coef	SE	PV	Goodness	Value
ω_1	0.014	0.001	0	LLH	-58006.61	ω_1	0.006	0.000	0	LLH	-82597.62
α_1	0.129	0.003	0	AIC	116021.23	α_1	0.071	0.002	0	AIC	165203.24
β_1	0.860	0.003	0	BIC	116057.77	β_1	0.924	0.002	0	BIC	165240.94
γ_1	1.020	0.003	0	MSE	1.86	γ_1	1.158	0.003	0	MSE	1.23
EREGL	Coef	SE	PV	Goodness	Value	SAHOL	Coef	SE	PV	Goodness	Value
ω_1	0.009	0.001	0	LLH	-81717.06	ω_1	0.008	0.001	0	LLH	-45872.82
α_1	0.083	0.002	0	AIC	163442.12	α_1	0.097	0.003	0	AIC	91753.65
β_1	0.909	0.002	0	BIC	163479.74	β_1	0.897	0.003	0	BIC	91789.02
γ_1	1.112	0.003	0	MSE	1.357095	γ_1	0.995	0.003	0	MSE	1.65
EREGL	Coef	SE	PV	Goodness	Value	THYAO	Coef	SE	PV	Goodness	Value
ω_1	0.009	0.001	0	LLH	-81717.06	ω_1	0.003	0.000	0	LLH	-163661
α_1	0.083	0.002	0	AIC	163442.12	α_1	0.080	0.001	0	AIC	327330
β_1	0.909	0.002	0	BIC	163479.74	β_1	0.919	0.001	0	BIC	327370.7
γ_1	1.112	0.003	0	MSE	1.357095	γ_1	1.383	0.002	0	MSE	1.052
ISCTR	Coef	SE	PV	Goodness	Value	TUPRS	Coef	SE	PV	Goodness	Value
ω_1	0.015	0.001	0	LLH	-48308.65	ω_1	0.007	0.001	0	LLH	-69066.55
α_1	0.094	0.003	0	AIC	96625.31	α_1	0.064	0.002	0	AIC	138141.11
β_1	0.893	0.004	0	BIC	96660.85	β_1	0.930	0.002	0	BIC	138178.05
γ_1	1.001	0.003	0	MSE	1.575	γ_1	1.112	0.003	0	MSE	1.31

Note: The p-values for every coefficient are statistically significant for a 5% level for each stock in the sample.

In Table 5, the estimation outcomes are presented. As it can be noticed, α_1 and β_1 coefficients are statistically significant, showing us a strong interrelationship in the transaction duration series. Since both coefficients are significantly greater than 0, it can be said that consecutive trades carry a high degree of dependence. The magnitude of the last duration can determine the timing of the next, which leads to duration clustering. The persistence of price durations can be seen from the sum of $\alpha_1$and β_1 coefficients which are very close to 1. The current transaction duration is helpful in defining the subsequent duration, which may validate the duration clusters at a particular time of the day. The Weibull parameter $\gamma_1$for all stocks is also statistically significant.

As it's explained in the methodology part of the study, conditional on the Weibull parameter being greater or less than one, the hazard function is upward or downward sloping. The log-likelihood decreases to the log-likelihood of ACD with exponential specification when the Weibull parameter is equal to one. In Engle and Russell's ACD analysis on the U.S stock market, it is reported that all γ_1 are less than one, and this implies a downward slopping hazard function. (Engle and Russell, 1998) In the finding, different results of Weibull parameters for individual stocks are observed. These differences in levels of coefficients can be explained by trading activity. In U.S markets, most traded stocks have a higher sum of coefficients $\alpha_1$and β_1 and lower level of Weibull since the institutional investors are favoring the high liquidity and that they are a lot more active compared to the Turkish stock market.

As opposed to U.S stock markets, 70 percent of the intraday volume of BIST is realized by local individual investors. Comparing the two markets, it can be noticed that the Turkish stock market has less trading activity and longer trade durations. These divergences from developed financial markets tend to get less evident as High-Frequency Trading becomes more utilized market-wide.

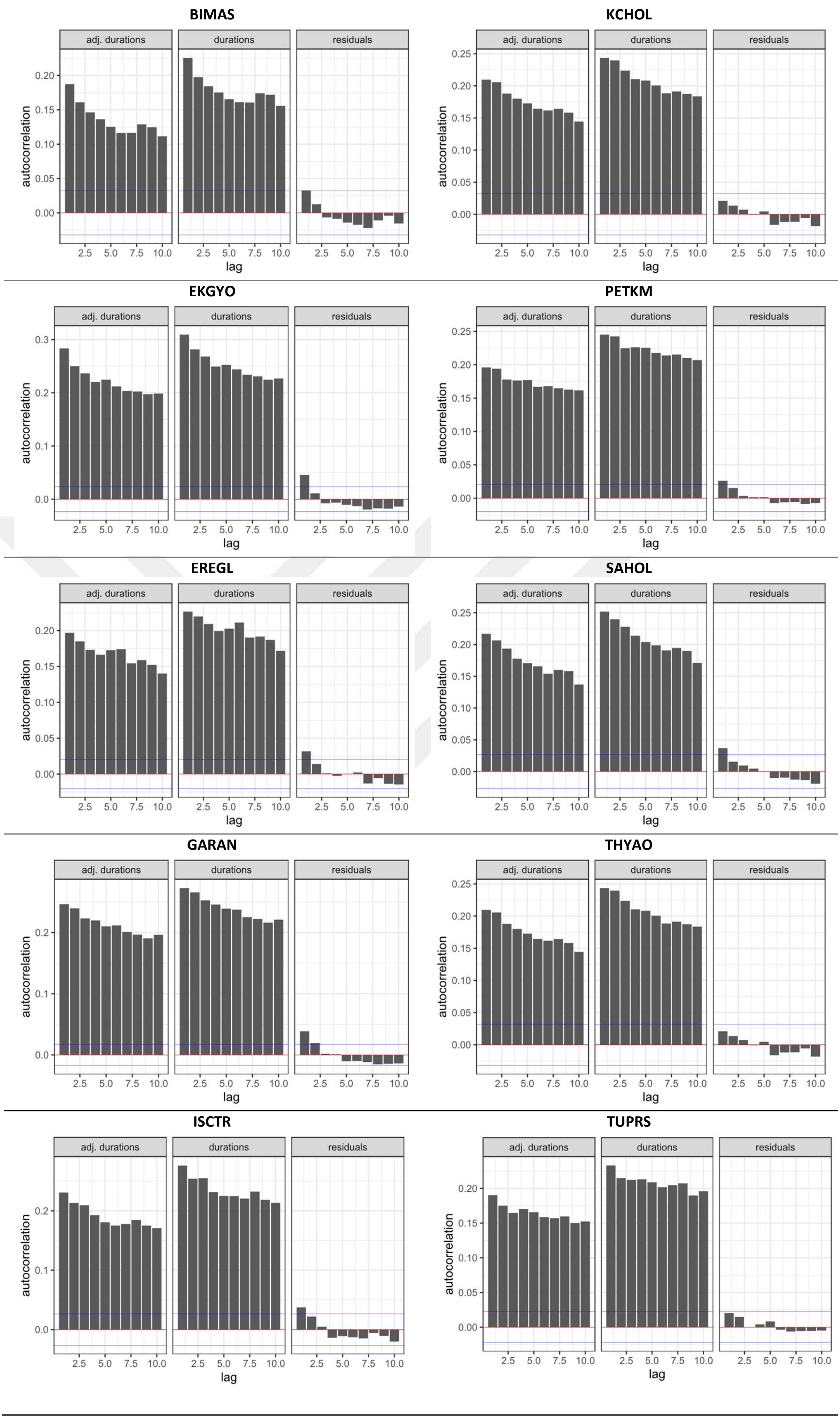


Figure 8. Auto-correlogram of durations, diurnally adjusted durations, and residuals for each stock in our sample.

To investigate the validity of the ACD framework and test the possible autocorrelation of residuals, Ljung-Box Q-statistics is used, the correlograms, and Meitz and Teräsvirta's Lagrange Multiplier tests for ACD models are conducted. (Meitz and Teräsvirta, 2006) Test results are almost indistinguishable across all stocks in our sub-sample. If the model is correctly specified, it is expected to have independent error terms, and the residuals should not show any further ACD structured dependency.

Table 6. Lagrange Multiplier Test of Meitz and Teräsvirta (2006) for no remaining ACD

Stock	L-M Stats	Degree of Freedom	P-value	# of durations
THYAO	455	5	0	896539
SAHOL	128	5	0	169238
EKGYO	212	5	0	157016
PETKM	228	5	0	256741
BIMAS	159	5	0	124580
ISCTR	92.5	5	0	176486
TUPRS	125	5	0	284780
EREGL	261	5	0	295448
KCHOL	124	5	0	133892
GARAN	423	5	0	483272

Note: LM test of Meitz and Teräsvirta (2006) for ACD models to investigate possible remaining ACD structures.

WACD (1,1) model successfully removes the autocorrelation in the data. The null hypothesis that the residuals are autocorrelated is rejected, indicating that our WACD (1,1) model is accurately identified and does a good job of fitting the data. It can be seen from Table 5 and Figure 8 that adjusted durations exhibit a strong autocorrelation structure, and the time of the day effect does not account fully for the dependence of the durations. The residuals are not significantly autocorrelated. ACF is slowly decreasing to zero in all cases. After removing the diurnal component, the correlations are declining somewhat, but the considerable reduction in autocorrelations comes for the residuals after removing the estimated WACD (1,1) dependency. It can also be understood that there are not any remaining ACD structures in the residuals from the M&T Lagrange Multiplier test, which can be seen in Table 6.

5.1.3 Autoregressive conditional duration models for price durations

To be able to further investigate bid-ask spread behaviors of the intraday transactions, the correct ACD model parallel to literature is conducted. The main idea behind this ACD fitting process is to fully understand the durations between price impacts and find an explication between durations, volume, and trade intensity. We want to have a closer look at the behaviors of price durations, the realized volume between the bid-ask spread changes, and the trade intensity according to the number of trades for each price duration.

First, the bid-ask spread durations are constructed by calculating the raw durations between “mid-prices” of the spread. It is assumed that the mid-price is the average of the bid and ask price value at the corresponding trade window. We find the most suitable duration calculation method by investigating different trade windows and comparing correlograms between different durations.

As it is done in chapter 5.1.1, for demonstrating the removal of the seasonality from the data, investigating the most suited smoothing method, and for applying various types of ACD methodology step by step, the most traded equity in the subsample, which is THYAO.E, is selected. The results of the diurnal adjustments and ACD fits for price durations are reported.

As seen above in Figure 9, even if the duration means windows are changed, the week of the day pattern for price durations, as it has been already seen in transaction durations, can be observed. According to these findings, it is reasonable to use a day of the week diurnal adjustment to clean our data from seasonality.

According to the literature, three different most used smoothing methods for the durations are applied, and then the results are tested and decided which one fits the data the best. Super Smoother, Flexible Fourier Form, and Cubic Spline Function methods are all applied separately to raw durations for the calculation of diurnal adjustments. The results of the smoothing methods can be seen in Figure 10.

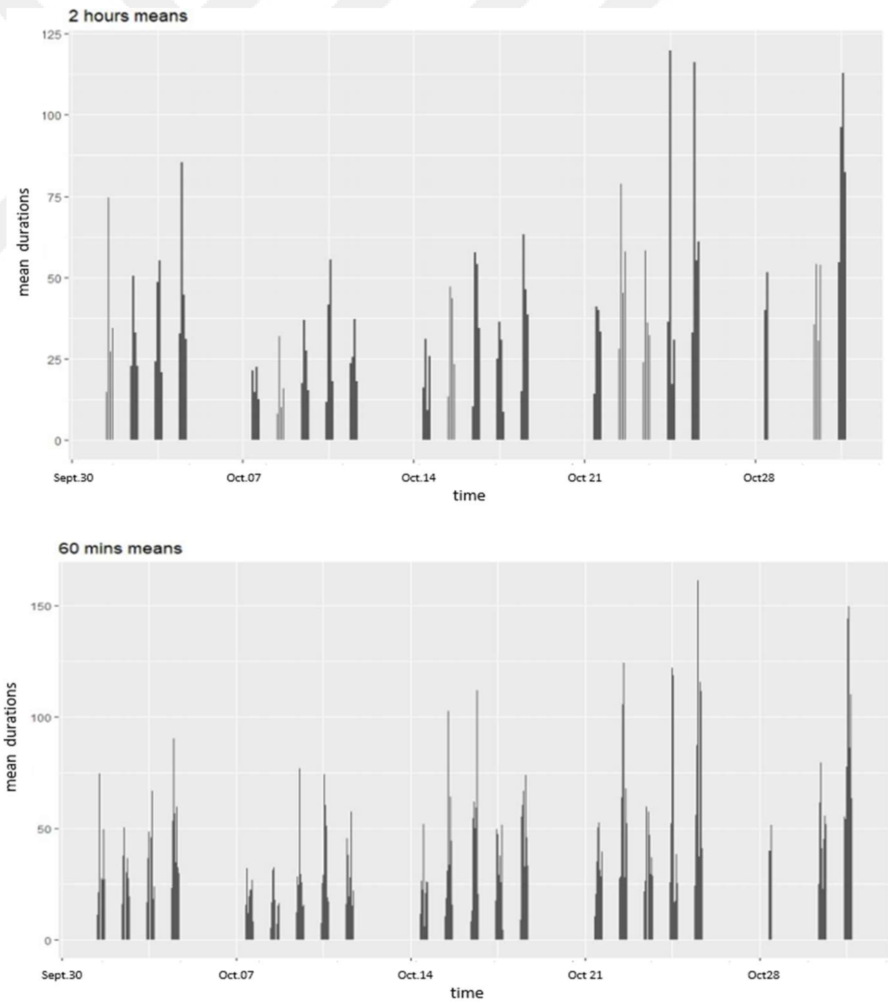


Figure 9. Intraday raw mean duration patterns for price durations

While converting from raw durations to adjusted durations by using the appropriate smoothing methodology, it is crucial not to get negative duration returns. In this instance, the cubic spline function fails to get positive results for all cases. The negative fitted line produced by the cubic spline technique can be explained by the sudden shift in the mean durations in the opening hours of that trading day. This estimation would yield negative adjusted durations and would be contradicting the assumptions of ACD methodology and prevent further estimations. Because of this reasoning, the cubic spline technique is eliminated for this case. On the other hand, when the differences between the other two methods are evaluated, one can see that FFF is more versatile and superior at capturing intraday duration instability. Parallel to raw durations, duration shifts and clusters are further noticeable.

The patterns and correlograms are evaluated for different calculations, and it is decided that the Flexible Fourier Form (FFF) fits the data the best. The improvements in the durations can be seen with the comparisons of autocorrelation between raw and adjusted durations generated by FFF. The correlogram for the Flexible Fourier Smoothing methodology is given in Figure 11.

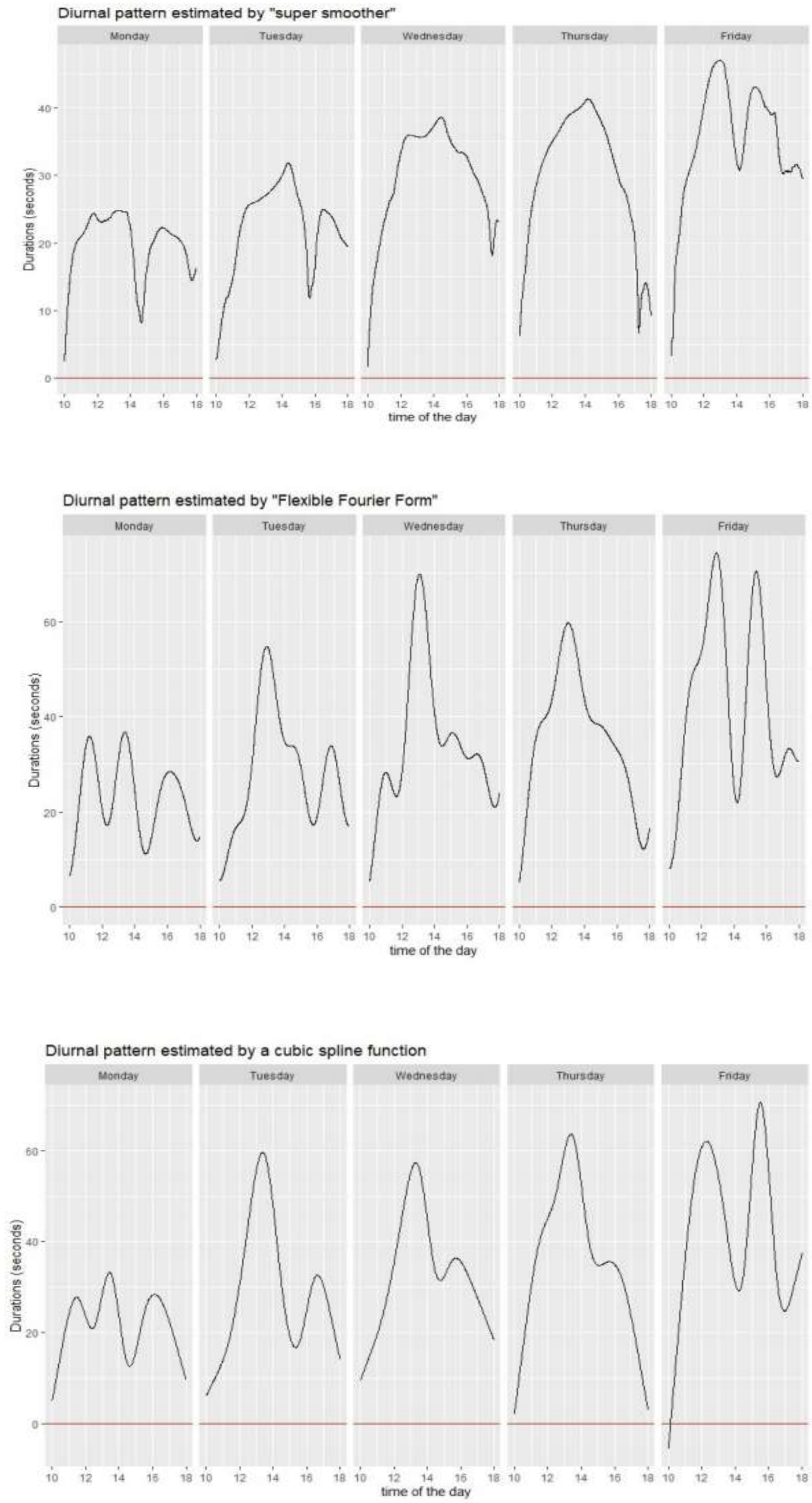


Figure 10. Diurnal patterns estimated by three different most used Smoothing Methods in literature

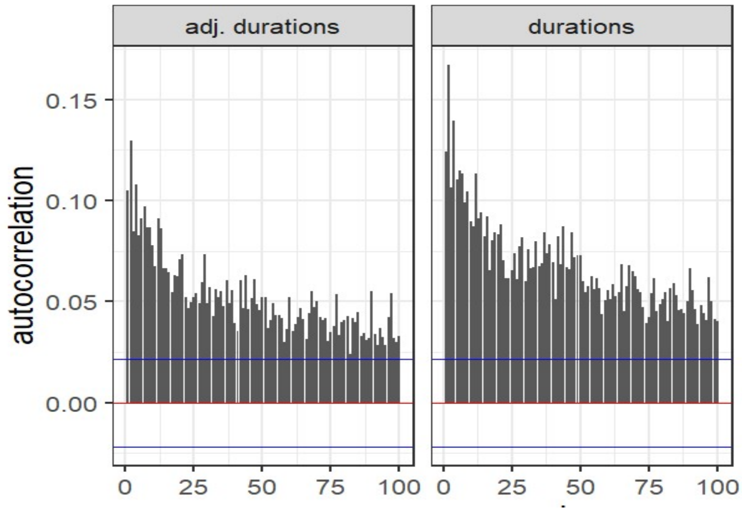


Figure 11. Correlogram for diurnal adjustments applied to raw durations with FF smoothing and day of the week seasonality removal

Completing the method selection process of diurnal durations makes it possible to apply various ACD and Log-ACD methods. We apply both ACD and Log-ACD estimations using Weibull error term distributions (We showed the reasoning behind the Weibull error term specification decision in the previous chapters.). The table below shows that the Weibull distribution estimations with four different ACD methodologies fit our price duration data. For the sake of having more flexibility with our coefficients, we use the Weibull distribution method for our study in the further parts. AMACD, ABACD, Log-ACD, and ACD methods, explained in the methodology part of our study, are applied to adjusted price durations estimated by our Flexible Fourier smoothing and day of the week diurnal seasonality removal methods. The descriptive statistics of adjusted durations and results of the ACD fits can be seen in Tables 7 and 8 below.

Since the price duration is the duration between each meaningful (price impact) transaction, the descriptive statistics should be understood from the point of view of price-changing intervals. The total number of transactions that changed the bid and ask spread in the orderbook is 7875, with a mean of 93.9, meaning that it takes approximately 94 transactions to change the price floor of the corresponding orderbook. All the details in descriptive statistics should be read from this perspective in Table 7.

Table 7. Descriptive Statistics of Adjusted Price Durations

Stock	Stats	Price	Duration	Volume (TL)	# Of Transaction
THYAO	N	7875	7875	7875	7875
	Mean	11.54	65.34	2930158	93.9
	Max	12.38	4332	50581230	1821
	Min	10.60	0	10	1
	Median	11.62	12	1340986	38

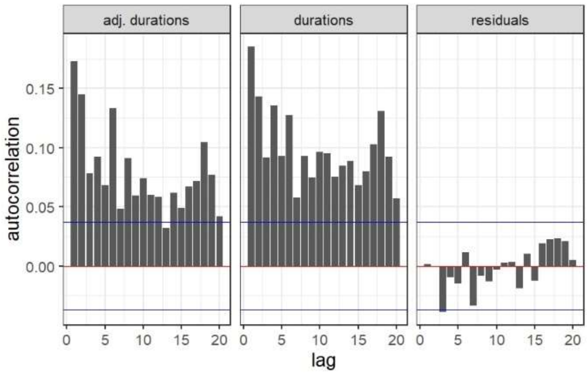
Table 8. Estimates from four Different ACD Specifications for Price Durations with Weibull Error Term

Log-ACD	Coef Value	SE	PV	Goodness	Value	ACD	Coef Value	SE	PV	Goodness	Value
ω_1	0.155	0.013	0	LLH	-4562.8	ω_1	0.041	0.007	0	LLH	-4499.19
α_1	0.143	0.012	0	AIC	9133.6	α_1	0.164	0.014	0	AIC	9006.386
β_1	0.895	0.021	0	BIC	9161.5	β_1	0.796	0.018	0	BIC	9034.272
γ_1	0.612	0.004	0	MSE	5.1879	γ_1	0.623	0.005	0	MSE	5.257101

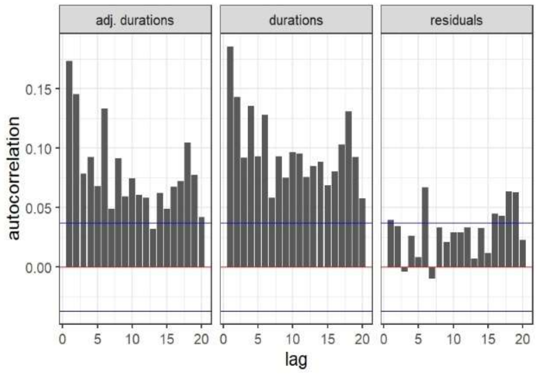
AMACD	Coef Value	SE	PV	Goodness	Value	ABACD	Coef Value	SE	PV	Goodness	Value
ω_1	0.0064	0.008	0.042	LLH	-4486.8	ω_1	-0.908	0.714	0.203	LLH	-4460.29
α_1	0.0725	0.018	0	AIC	8983.7	α_1	0.823	0.615	0.181	AIC	8936.579
ν_1	0.0613	0.013	0	BIC	9018.6	C_1	1.316	0.368	0	BIC	8992.351
β_1	0.8526	0.018	0	MSE	5.171	β_1	0.933	0.008	0	MSE	5.115013
γ_1	0.624	0.005	0			ν_1	-1.706	0.480	0		
						δ_1	0.199	0.112	0.076		
						δ_2	0.096	0.017	0		
						γ_1	0.625	0.004	0		

When we analyze the estimates from our ACD specifications in Table 8, since the values are within very close distance, it is challenging to decide which model better fits the price duration data. For that reason, at this point, as we did with transaction durations, it is logical to compare correlograms of each model separately and select our model depending on it.

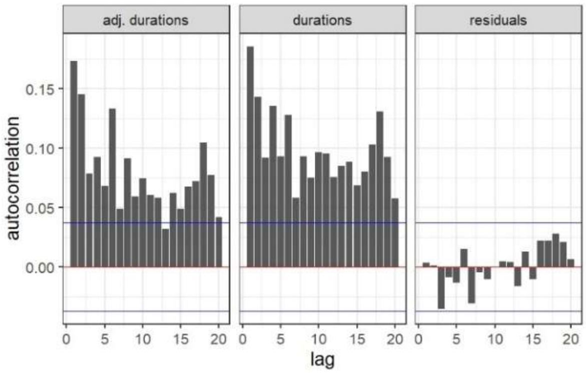
As it can be noticed, α_1 and β_1 coefficients are statistically significant, showing us a strong interrelationship in the price duration series for each model. Given that both coefficients are considerably greater than 0, it can be assumed that consecutive trades carry a high degree of dependence. The scale of the last duration can define the timing of the next, which leads to duration clustering. The persistence of transaction durations can be seen from the sum of α_1 and β_1 coefficients which are very close to 1 for all specifications. The current price duration is valuable in defining the following duration, and this may possibly confirm the duration clusters at a particular time of the day. The Weibull parameter γ_1 for all stocks is also statistically significant. We can justify this significance of Weibull parameters with hazard function plots in the figure below.



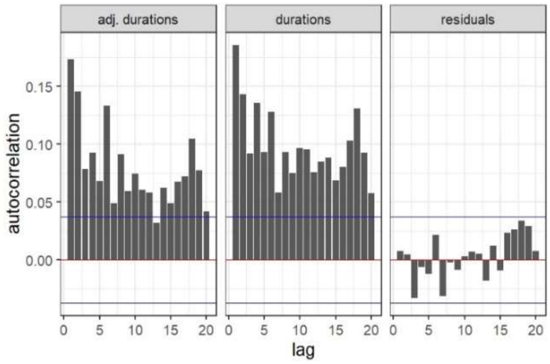
Correlogram for ACD



Correlogram for LACD

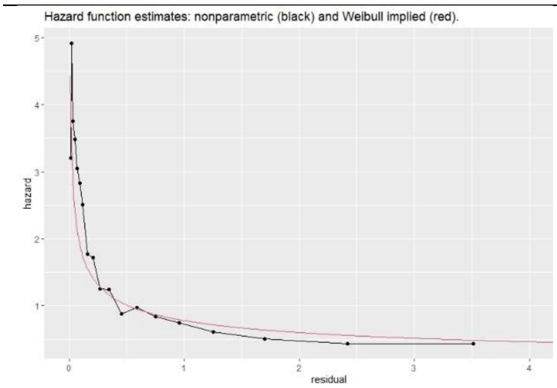


Correlogram for AMACD

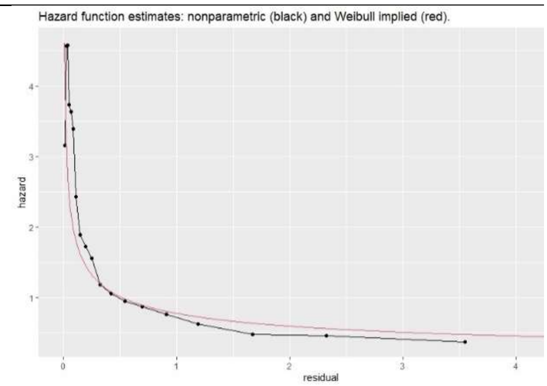


Correlogram for ABACD

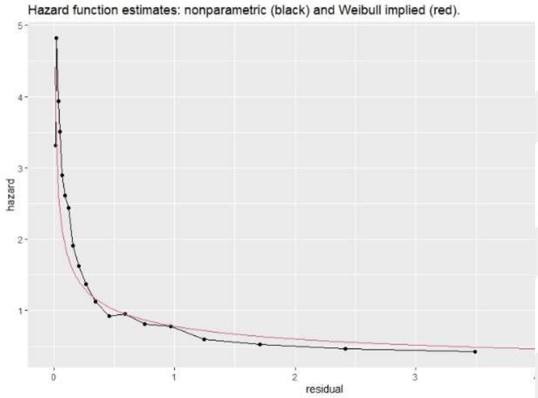
Figure 12. Correlogram for ACD, LACD, AMACD and ABACD separately



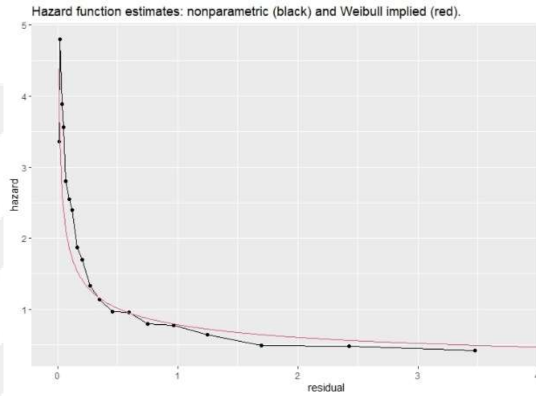
Hazar function for ACD



Hazard function for LACD



Hazard function for AMACD



Hazard function for ABACD

Figure 13. Plots for hazard function estimates for ACD, LACD, AMACD and ABACD separately

As clarified in the previous section 5.1.2, the hazard function is upward or downward sloping depending on the Weibull parameter being larger or less than one. The log-likelihood decreases to the log-likelihood of ACD with exponential specification when the Weibull parameter is equal to one. In Engle and Russell's ACD analysis on the U.S stock market, it is reported that all γ_1 are less than one, and this implies a downward slopping hazard function. (Engle and Russell, 1998) In our finding, we find the same results of Weibull parameters for price durations parallel to U.S stock markets. These levels of coefficients can be explained by trading activity.

With a statistically significant Weibull parameter less than 1, the downward slopping hazard function implies price duration clustering can be justified as expected.

Stocks with a higher sum of coefficients $\alpha_1$and β_1 and lower level of Weibull since the institutional investors are favoring the high liquidity, and this intuitional trading activity results in a “bandwagon” effect in transaction intensity. Price impacting transactions are clearly clustering when the activity rises in the market. When we compare the results from Figures 12 and 13, we can declare that both standard ACD and AMACD can explain the price durations. Since the results are significant and close to each other, we can choose to use AMACD because of its greater possibility of capturing news impacts. In the methodology part 3.3.2.1, it can be seen that with an additive and multiplicative form of the model, more flexible news impacts can be captured with parametrizations. We also report the scatter plot for the AMACD model in Figure 14 to understand the residuals' behavior better. As expected, the mean of the residuals is at the same level for all values of estimated conditional means.

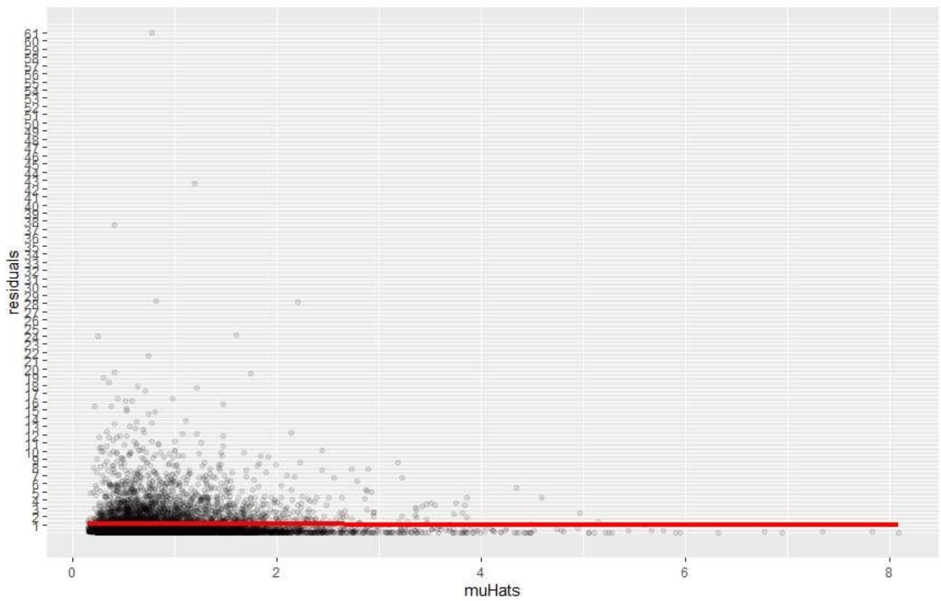


Figure 14. Scatter plot for estimated conditional means versus residuals for AMACD model

5.1.4 Autoregressive conditional duration models for volume durations

After investigating trade and price durations, we proceed to further analyses of volume durations for threshold volume behavior of intraday transactions. We construct the correct ACD model parallel to the literature. The main objective behind this volume process is to fully understand the durations between pre-determined Turkish Lira levels of volume that occur consecutively in the stock market. We intend to have a more decisive view of the behaviors of volume durations, the realized stalling time between the volume level changes, and the trade intensity according to the number of trades for each volume duration.

First, we construct the volume durations by calculating the raw durations between pre-determined levels of volumes. Since there is not a universally accepted volume level for volume threshold duration models, as an addition to the literature, instead of selecting only a pre-determined volume level for each equity (which is 1 million dollars of volume in most of the studies, around 10 million in Turkish liras), we assume that 0.001, 0.005 and 0.01 of the total Turkish liras of volume at the corresponding trade window are more satisfying for the study. We find the most suitable duration calculation method by investigating different volume windows and comparing correlograms between different durations. We report the descriptive statistics for different levels of volumes in Table 9 below.

As we have accomplished in chapter 5.1.1, for representing the elimination of the seasonality from the data, investigating the most appropriate smoothing technique, and for applying various types of ACD methodology step by step, we selected the most traded equity in our subsample, which is THYAO.E and describe the results of the diurnal adjustments and ACD fits for volume durations.

As it can be seen, even if duration means windows are adjusted, the week of the day pattern for volume durations, as previously found in transaction and price durations, is undoubtedly realized. Corresponding to these discoveries, a day of the week diurnal adjustment is chosen to cleanse the data from seasonality.

Matching the literature, three different smoothing methods for the durations are employed, tested the results, and decided that the methodology that fits the data the finest is Flexible Fourier Form. Methods are all applied separately to raw durations to calculate diurnal adjustments.

Table 9. Descriptive Statistics of Adjusted Volume Durations of Three Different Threshold Levels

Threshold	Stats	Price	Duration	# Of Transaction
2.3 million TL (0,01%)	N	9164	9164	9164
	Mean	11.62	56.55	80.93
	Max	12.38	1668	506
	Min	10.60	1	4
	Median	11.69	27	71
4.6 million TL (0.02%)	N	4838	4838	4838
	Mean	11.63	107.1	153.1
	Max	12.38	2783	771
	Min	10.60	1	7
	Median	11.70	60	138
11.5 million TL (0.05%)	N	1983	1983	1983
	Mean	11.63	261	373
	Max	12.38	3775	1438
	Min	10.60	1	40
	Median	11.63	168	338

The total number of transactions that occurred for the respective realized volume threshold in the orderbook can be seen in Table 9. All the details in descriptive statistics should be read from this perspective of “filling” the volume needed for the pre-determined threshold. The number of transactions and duration statistics show us how many transactions needed to be realized to reach the respective threshold level and how long it took to realize them.

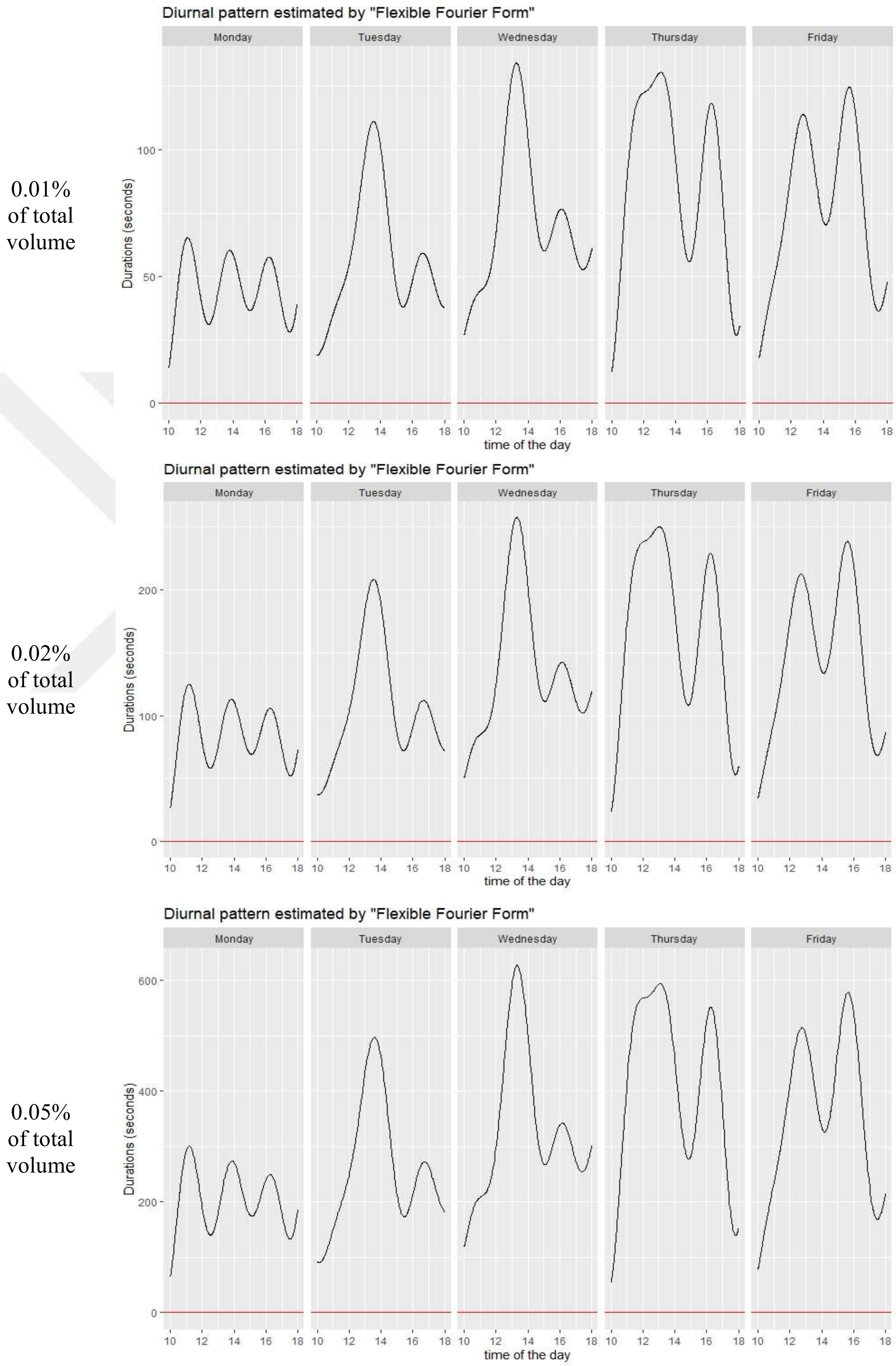


Figure 15. Diurnally adjusted duration patterns estimated by FFF for three different threshold levels

The LACD (1,1) specification estimates are analyzed in Table 10 for three threshold levels. As it can be observed, $\alpha_1$and $\beta_1$coefficients are statistically significant, showing us a strong interrelationship in the volume duration series for each threshold level. Given that both coefficients are significantly larger than 0, it can be expected that successive trades carry a high degree of dependence. The size of the final duration can explain the scheduling of the next, which advances into duration clustering for volume durations, as seen in both price and transaction durations. The persistence of volume durations can be seen from the sum of $\alpha_1$and β_1 coefficients being close to 1 for all specifications. The current volume duration is significant in defining the following duration, which confirms the duration clusters at a specific time of the day. The Weibull parameter $\gamma_1$for all specifications is also statistically significant. Just by analyzing the estimates from the LACD, we cannot make any decision between different threshold levels. That is why it is logical to check how the residuals and auto correlograms are at this point.

Table 10. Estimates for Log-ACD (1,1) Model for Three Different Volume Thresholds

Threshold	Log-ACD	Coeff Value	SE	PV	Goodness	Value
2.3 million TL	ω_1	0.0874	0.00404	0	Loglikelihood	-7594.47
(0.01%)	α_1	0.1807	0.00662	0	AIC	15196.94
	β_1	0.9366	0.00542	0	BIC	15225.44
	γ_1	1.0228	0.00777	0.003	MSE	1.280409
4.6 million TL	ω_1	0.0788	0.00508	0	Loglikelihood	-3866.56
(0.02%)	α_1	0.2562	0.00964	0	AIC	7741.112
	β_1	0.8877	0.00936	0	BIC	7767.049
	γ_1	1.236	0.01276	0	MSE	0.864163
11.5 million TL	ω_1	0.0757	0.00978	0	Loglikelihood	-1444
(0.05%)	α_1	0.4123	0.01938	0	AIC	2895.9
	β_1	0.8028	0.01764	0	BIC	2918.27
	γ_1	1.5433	0.02748	0	MSE	0.54656

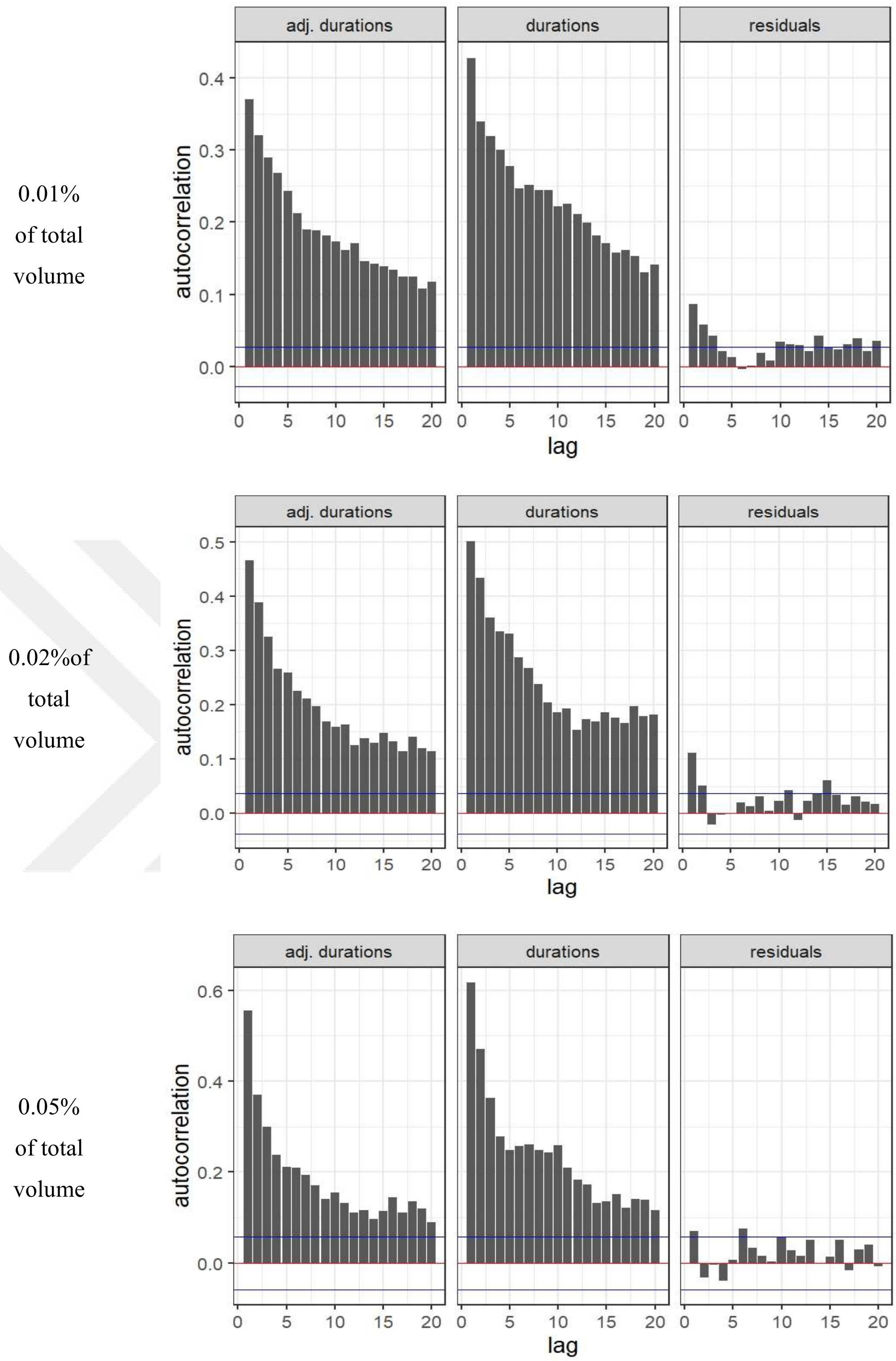


Figure 16. Correlograms for three different threshold levels

When we check the correlograms in Figure 16 for the threshold levels, we can immediately see the reduction in the correlations first with diurnal adjustment then an even more significant reduction with LACD modeling. In lower volume threshold levels, it is found that first and second lagged values are still autocorrelated even though LACD implies a considerable reduction for volume durations and adjusted durations. By analyzing the LACD estimates and correlogram, it becomes evident that the selection of volume threshold at 0,05 percent is better than other options. One can say that when the volume threshold is kept smaller than it should be, around 230.000 dollars, in this case, the autocorrelation is inevitable in residuals. Since there are no specific rules in the ACD literature for determining the volume levels for durations, it is logical to compare the results of different volume percentages and decide. At this point, the scatter plots and hazard functions are published to deeper understand the residuals with a 0.05 percent volume threshold.

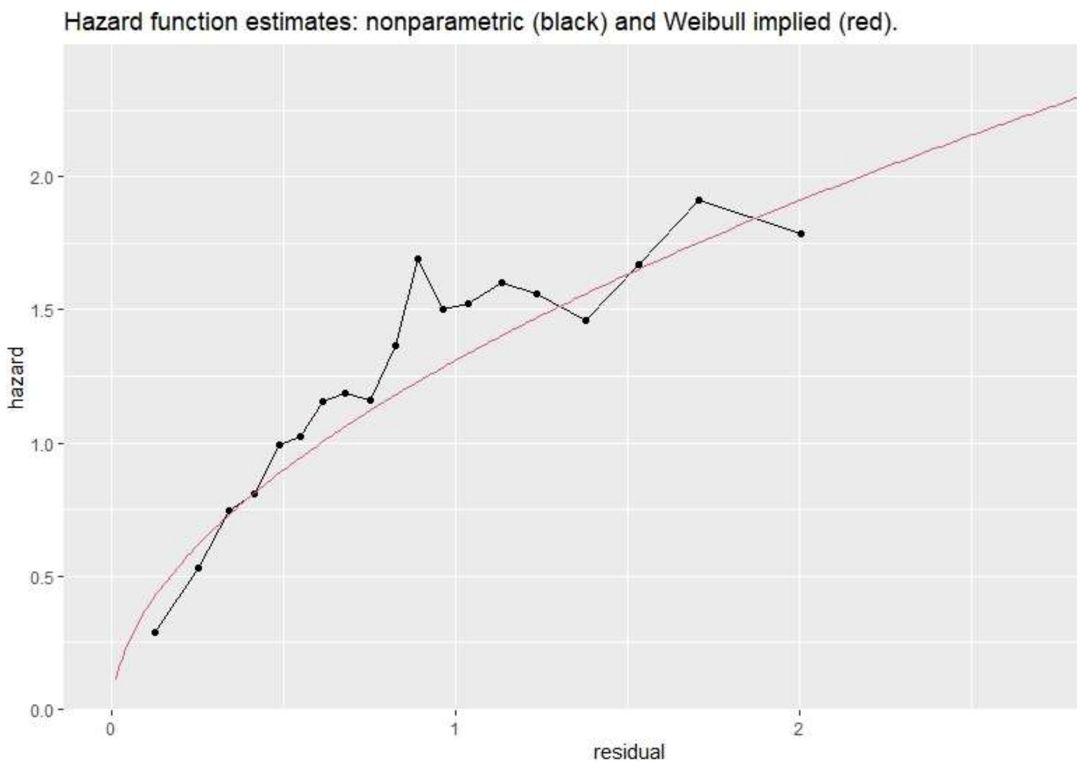


Figure 17. Hazard function estimates for LACD (1,1) with Weibull error distribution for 0,05% volume threshold specification

Figures 17,18, and 19 are helpful for verifying our LACD (1,1) model and understanding the residuals from the model. Nonparametric results from the model match closely with the Weibull implied function. In Figure 17, since the Weibull coefficient is around 1.5, the hazard function is upward sloping. When the residual quantiles are plotted against the Weibull distribution in Figure 18, the same matching the theoretical values are also reflected.

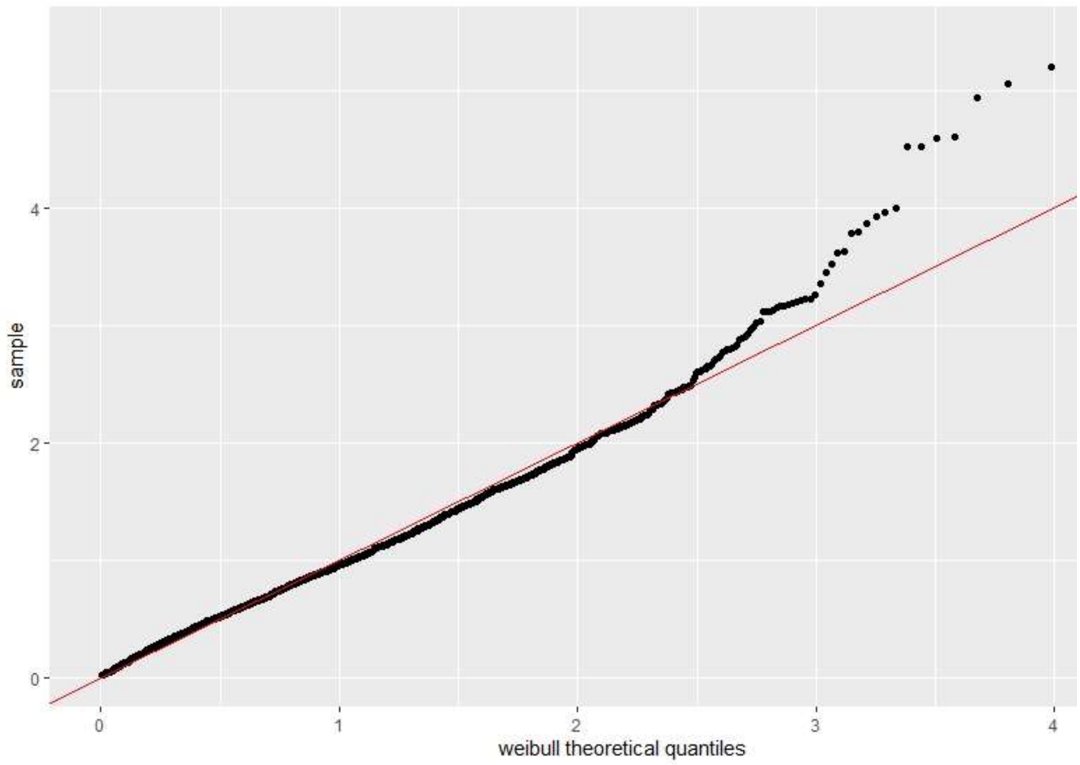


Figure 18. Quantile-quantile plot for residuals versus theoretical quantiles for LACD (1,1) with 0.05% threshold specification

In Figure 19, the mean of the residuals is at the same level for all values of estimated conditional means as theoretically expected. By exploring these three figures, successfully fitted LACD (1,1) on volume durations for 0.05% specification can be justified.

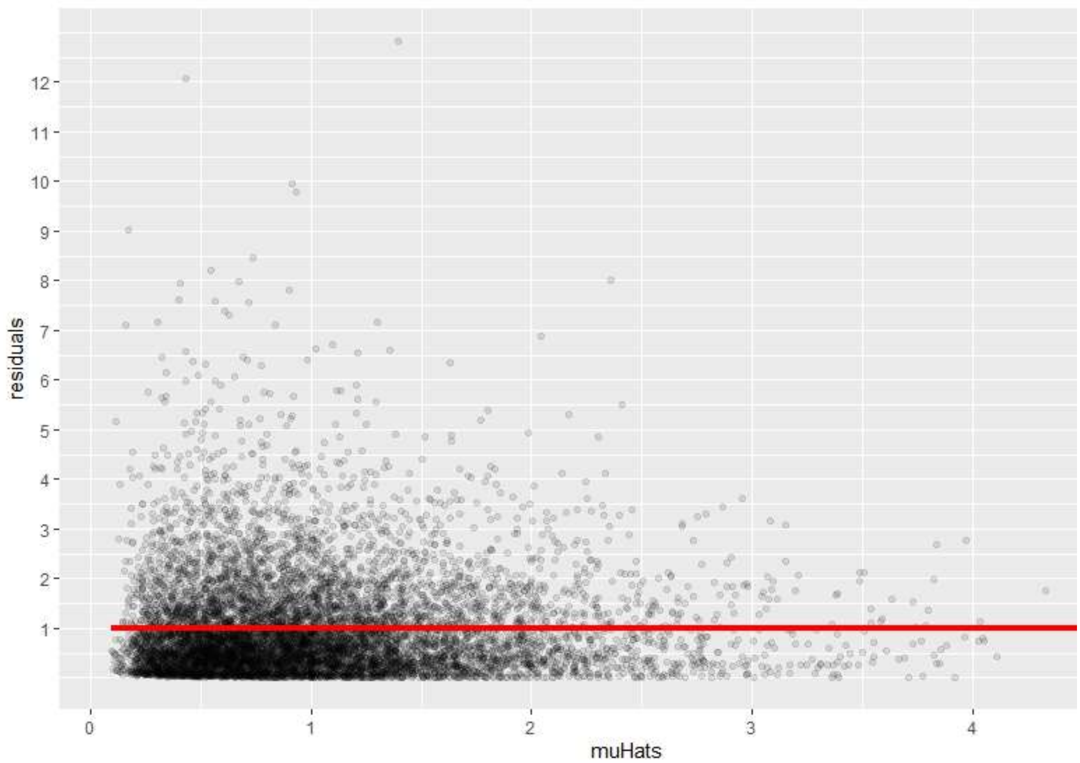


Figure 19. Scatter plot for estimated conditional means versus residuals for LACD (1,1) with 0.05% threshold specification

5.2 Modelling liquidity with autoregressive conditional duration models

As an extension to what has been done in previous chapter 5.1, in this part of the empirical findings, the link between the durations modeling and liquidity is further analyzed by using the same dataset of the 10 most traded equities in BIST. A Weibull ACD model is conducted on the dataset, but this time for daily and one-hour trading windows. The ACD methodology is applied individually for each stock and each hour of the trading session for every day.

Table 11. Summary Statistics for 10 Most Traded Stocks in BIST

Stock	Stats	Price	Duration	Volume	Ask	Bid	Stock	Stats	Price	Duration	Volume	Ask	Bid
BIMAS	N	124580	124580	124580	124580	124580	KCHOL	N	133892	133892	133892	133892	133892
	Mean	47.79	4.85	13331	47.8	47.76		Mean	18.43	4.51	12236	18.44	18.42
	Max	59.55	831	1.27E+08	49.86	49.84		Max	21.54	469	1.14E+07	19.44	19.43
	Min	36.8	0	39	46.06	46.02		Min	14.37	0	15	16.95	16.88
	Sd	0.87	14.85	361243	0.83	0.83		Sd	0.55	14.74	43915	0.54	0.54
EKGYO	N	157016	157016	157016	157016	157016	PETKM	N	256741	256741	256741	256741	256741
	Mean	1.24	3.84	13701	1.24	1.23		Mean	3.48	2.35	17430	3.49	3.48
	Max	1.57	321	2.48E+06	1.42	1.41		Max	4.33	321	3.41E+06	3.66	3.65
	Min	1.04	0	1	1.09	1.08		Min	2.68	0	3	3.33	3.32
	Sd	0.07	9.85	69565	0.07	0.07		Sd	0.07	5.53	88311	0.07	0.07
EREGL	N	295448	295448	295448	295448	295448	SAHOL	N	169238	169238	169238	169238	169238
	Mean	6.65	2.05	13287	6.65	6.64		Mean	8.82	3.57	12288	8.83	8.82
	Max	8.28	324	2.84E+06	6.99	6.98		Max	11.1	339	6.44E+06	9.68	9.67
	Min	5.17	0	5	6.43	6.42		Min	6.63	0	8	8.18	8.17
	Sd	0.12	5.49	41513	0.11	0.11		Sd	0.34	10.29	35134	0.35	0.35
GARAN	N	483272	483272	483272	483272	483272	THYAO	N	896539	896539	896539	896539	896539
	Mean	9.38	1.25	38156	9.39	9.38		Mean	11.61	0.67	31126	11.61	11.6
	Max	12.02	323	2.80E+07	10.28	10.27		Max	14.85	321	2.02E+07	12.4	12.39
	Min	6.68	0	7	8.25	8.24		Min	8.6	0	9	10.6	10.59
	Sd	0.51	3.94	132348	0.51	0.51		Sd	0.4	2.09	106849	0.39	0.39
ISCTR	N	176486	176486	176486	176486	176486	TUPRS	N	284780	284780	284780	284780	284780
	Mean	5.95	3.42	17961	5.96	5.95		Mean	127.96	2.12	18244	128.03	127.9
	Max	7.38	395	8.76E+06	6.34	6.33		Max	169.6	322	1.49E+07	143.8	143.7
	Min	4.48	0	4	5.49	5.47		Min	94.9	0	97	117.4	117.3
	Sd	0.17	10.04	48192	0.17	0.17		Sd	7.49	6.21	70526	7.46	7.46

Table 12. WACD (1,1) Estimation Results for Each Stock Separately

BIMAS	Coef	SE	PV	Goodness	Value	KCHOL	Coef	SE	PV	Goodness	Value
ω_1	0.023	0.002	0	Loglike.	-33206	ω_1	0.017	0.001	0	Loglike.	-32281
α_1	0.104	0.004	0	AIC	66421	α_1	0.092	0.004	0	AIC	64571
β_1	0.875	0.005	0	BIC	66455	β_1	0.892	0.005	0	BIC	64605
γ_1	0.919	0.004	0	MSE	1.778	γ_1	0.916	0.004	0	MSE	1.896
EKGYO	Coef	SE	PV	Goodness	Value	PETKM	Coef	SE	PV	Goodness	Value
ω_1	0.014	0.001	0	Loglike.	-58007	ω_1	0.006	0.000	0	Loglike.	-82598
α_1	0.129	0.003	0	AIC	116021	α_1	0.071	0.002	0	AIC	165203
β_1	0.860	0.003	0	BIC	116058	β_1	0.924	0.002	0	BIC	165241
γ_1	1.020	0.003	0	MSE	1.869	γ_1	1.158	0.003	0	MSE	1.231
EREGL	Coef	SE	PV	Goodness	Value	SAHOL	Coef	SE	PV	Goodness	Value
ω_1	0.0095	0.0006	0	Loglike.	-81717	ω_1	0.008	0.001	0	Loglike.	-45873
α_1	0.0832	0.0019	0	AIC	163442	α_1	0.097	0.003	0	AIC	91754
β_1	0.9091	0.0021	0	BIC	163480	β_1	0.897	0.003	0	BIC	91789
γ_1	1.1120	0.0027	0	MSE	1.357	γ_1	0.995	0.003	0	MSE	1.656
GARAN	Coef	SE	PV	Goodness	Value	THYAO	Coef	SE	PV	Goodness	Value
ω_1	0.008	0.000	0	Loglike.	-108000	ω_1	0.003	0.000	0	Loglike.	-163661
α_1	0.093	0.002	0	AIC	216299	α_1	0.080	0.001	0	AIC	327330
β_1	0.902	0.002	0	BIC	216338	β_1	0.919	0.001	0	BIC	327371
γ_1	1.166	0.002	0	MSE	1.526	γ_1	1.383	0.002	0	MSE	1.053
ISCTR	Coef	SE	PV	Goodness	Value	TUPRS	Coef	SE	PV	Goodness	Value
ω_1	0.015	0.001	0	Loglike.	-48309	ω_1	0.007	0.001	0	Loglike.	-69067
α_1	0.094	0.003	0	AIC	96625	α_1	0.064	0.002	0	AIC	138141
β_1	0.893	0.004	0	BIC	96661	β_1	0.930	0.002	0	BIC	138178
γ_1	1.001	0.003	0	MSE	1.575	γ_1	1.112	0.003	0	MSE	1.319

The primary purpose of this chapter is to investigate the explanatory power of the model and compare it with market-wide used proxies. The coefficient from the WACD model and related Amihud Illiquidity measures are used for building several correlation matrices and multivariate regressions to understand the relationship of our model with stock returns, volume, and liquidity measures. Stock-wise descriptive stats and WACD (1,1) estimates for raw data can be found in Table 11 and Table 12 above.

5.2.1 Estimation results for daily data

The dataset is split into daily trading intervals for further investigations of Amihud, return, and ACD coefficient relations. Tables 13 and 14 below show summary statistics and correlation matrix for all variables of daily estimations can be seen. As expected in previous chapters, a negative relationship between Weibull's slop coefficient γ and Amihud can be observed. Parallel to the economic interpretation, the Amihud Illiquidity is expected to increase when the durations are clustering.

Table 13. Summary Statistics for Daily Panel Data

Variable	Obs	Mean	Std. Dev.	Min	Max
Return	180	0.006	0.024	-0.036	0.988
ω	180	0.319	0.158	0.035	0.748
α	180	0.086	0.050	-0.045	0.266
β	180	0.605	0.156	0.095	0.910
Amihud	180	2E-09	3E-09	1E-10	2E-08
γ	180	1.094	0.137	0.884	1.673

Table 14. Correlation Matrix for Daily Estimations

	α	β	γ	Return	Amihud
α	1				
β	-0.1292	1			
γ	0.0643	0.3821	1		
Return	0.0628	0.0628	0.148	1	
Amihud	0.0578	-0.4054	-0.421	0.1576	1

Another result parallel to the market microstructure interpretations can be assessed as the relationship between β and Amihud Illiquidity. Since β is the coefficient of the past conditional duration estimated by ACD, one may expect higher β values when the market activity and trade intensity are elevated. It is possible to investigate these results in Tables 15 and 16 below.

Table 15. Pairwise Pooled Regressions for Daily Return

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
γ	0.026***				0.041
	(0.013)				(0.027)
β		0.010			0.006
		(0.012)			(0.014)
α			0.029		0.017
			(0.023)		(0.030)
Amihud				1.1E+06	2.E+06
				-1.2E+06	-1.5E+06
Constant	-0.022	0.000	0.004	0.003	-0.049
	(0.014)	(0.008)	(0.003)	(0.003)	(0.025)
Observations	180	180	180	180	180
R-squared	0.022	0.004	0.004	0.025	0.083

Robust standard errors in parentheses
*** P<0.01, ** p<0.05, p<0.1 *

Table 16. Pairwise Pooled Regressions for Daily Amihud

Variables	Model 1	Model 2	Model 3	Model 4
γ	-0.099***			-0.068***
	(0.013)			(0.019)
β		-0.083***		-0.047**
		(0.015)		(0.019)
α			0.037	0.030
			(0.062)	(0.060)
Constant	0.137	0.079	0.026	0.129
	(0.015)	(0.010)	(0.005)	(0.013)
Observations	180	180	180	180
R-squared	0.177	0.164	0.003	0.248

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Pairwise pooled regression results for return and Amihud Illiquidity versus all the coefficients are reported in the tables above. Weibull's slope parameter γ is significant in both cases. Both γ and β are significant for estimating Amihud's proxy. In Table 16, model 4 has an explanatory power of 0.248 as R-square, which indicates significant variables from WACD (1,1) are capable of explaining the corresponding proxy to an extent. Having negative and significant coefficient signs for β and γ demonstrates the connection between the persistence of durations and the upward slopping shape of the hazard function. With a hazard function leaning toward such a shape, one can expect to see “thicker” bid-ask spreads, as it’s explained in Engle and Russell (1998). This characteristic of the hazard’s shapes indicates a greater failure rate as durations get longer, making it possible to accumulate a greater amount of passive orders waiting at the bid and ask levels. By this perception, one can expect to see a less illiquid order book since a “thicker” bid-ask spread would increase the amount of equity buyable with minimum transaction cost.

The same perspective for understanding the economic interpretation of the negative relationship between γ and Amihud can also be applied to γ and return. As the bid-ask passive orders get “thicker, " larger transactions and volumes are required to generate returns. At this point, Fixed and Random Effect Panel Regressions are conducted for return versus Amihud and ACD coefficients in Tables 17 and 18.

Table 17. Fixed-Effects Regression for Daily Returns

Variables	FE Regression
ω	0.011
	(0.22)
α	0.002
	(0.023)
β	0.010
	(0.022)
Amihud	-1.9E+06
	(2,4E+06)
γ	-0.014**
	(0.006)
Constant	0.006
	(0.023)
Observations	180
R-squared	0.16
No of company	10

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 18. Random-Effects Regression for Daily Returns

Variables	RE Regression
ω	0.003
	(0.040)
α	-0.002
	(0.039)
β	0.002
	(0.040)
Amihud	-2.1E+06***
	(6.4E+05)
γ	-0.008***
	(0.002)
Constant	0.007
	(0.041)
Observations	180
R-squared	0.189
No of company	10

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Both regressions are statistically significant, as it is reported with F tests. Results from the corresponding regressions are superior to pooled regression in Table 15. For RE regression, both the Amihud and γ are significant. A Hausman test is deployed to be able to understand which regression is preferable, and the test results are given in Table 19. By using the chi-square test results, we fail to reject the null hypothesis that the RE model is suitable. As expected from the significance of the variables and the R square results, it is preferred to continue with RE regressions.

Table 19. Hausman Test for Daily FE and RE Regressions

	FE Panel	RE Panel	Difference	Std.E
ω	0.200	0.192	0.008	0
α	0.122	0.152	-0.030	0
β	0.190	0.186	0.004	0
Amihud	-1.9E+06	-2.0E+06	1.3E+05	-1.9E+05
γ_1	-0.115	-0.061	-0.053	-0.016
Chi square	0.51			
Prob>chi2	0.47			

Table 20. Friedman's Test of Cross-Sectional Independence for RE

	Correlation matrix of residuals									
	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10
r1	1									
r2	0.578	1								
r3	-0.100	0.237	1							
r4	0.612	0.709	0.127	1						
r5	0.406	0.488	0.446	0.443	1					
r6	0.619	0.594	-0.106	0.738	0.403	1				
r7	0.679	0.703	-0.014	0.574	0.571	0.747	1			
r8	-0.202	0.127	0.591	-0.137	0.140	-0.344	-0.243	1		
r9	0.704	0.608	-0.316	0.539	0.075	0.538	0.589	-0.016	1	
r10	0.540	0.394	-0.105	0.350	0.415	0.487	0.671	-0.369	0.290	1

Friedman's test of cross-sectional independence = 69.84, Pr = 0.00

Considering the Friedman Test results, the null hypothesis that no differences between the variables is rejected. RE regression is suitable for the data. In RE estimation results, the coefficient for the Weibull's γ is negative, and that leads to negative relations between γ and return. These results are parallel to microstructure interpretations. Since Weibull's γ directly accounts for the Hazard function's slope, duration clustering enforced by the slope is associated with the return. From the market behavior perspective, it can be assumed that the possibility of having inclined returns arises because of clustering imposed by "bandwagon effects" and institutional intraday trader impacts on the market.

5.2.2 Estimation results for hourly data

The estimations from Table 12 are divided into one-hour trade windows to be able to compare WACD (1,1) coefficients with Amihud Illiquidity measures, corresponding stock returns, and various variables. The resulting summary statistics for stock-wise WACD coefficients can be seen in Table 21. As can be noticed from Table 12, all coefficients are statistically significant. $\alpha+\beta$ values are close to one, demonstrating the persistence of lagged durations and conditional durations on determining the duration of t+1. Weibull parameters are equal or greater than one in most cases, indicating flat/upward slopping hazard function estimates. For further and in-depth analysis, Random-Effects (RE) and Fixed-Effects (FE) panel regressions are conducted for understanding variable relationships.

Table 21. Stock-wise Summary Statistics for WACD (1,1) Estimations on Hourly Adjusted Trading Windows

BIMAS	Obs	Mean	Std. Dv.	Min	Max	KCHOL	Obs	Mean	Std. Dv.	Min	Max
ω_1	144	0.460	0.414	-0.025	1.917	ω_1	144	0.404	0.395	-0.059	1.807
α_1	144	0.064	0.145	-0.354	0.490	α_1	144	0.043	0.148	-0.401	0.593
β_1	144	0.481	0.476	-1.086	1.123	β_1	144	0.561	0.437	-0.852	1.126
γ_1	144	0.962	0.092	0.704	1.304	γ_1	144	0.957	0.106	0.688	1.436
Return	144	-2.2E-04	0.005	-0.020	0.025	Return	144	-1.5E-04	0.006	-0.023	0.028
Amihud	144	4.70E-10	3.97E-10	0	1.84E-09	Amihud	144	5.08E-10	3.92E-10	0	2.00E-09
EKGYO	Obs	Mean	Std. Dv.	Min	Max	PETKM	Obs	Mean	Std. Dv.	Min	Max
ω_1	144	0.323	0.324	0.009	1.929	ω_1	144	0.332	0.425	0.015	2.021
α_1	144	0.139	0.131	-0.370	0.640	α_1	144	0.065	0.064	-0.101	0.229
β_1	144	0.547	0.352	-0.955	1.106	β_1	144	0.607	0.418	-1.025	1.019
γ_1	144	1.046	0.148	0.735	1.554	γ_1	144	1.156	0.106	0.929	1.559
Return	144	-0.001	0.008	-0.026	0.036	Return	144	-2.2E-04	0.005	-0.017	0.018
Amihud	144	9.05E-10	1.83E-09	0	1.48E-08	Amihud	144	1.87E-10	2.45E-10	0	1.59E-09
EREGL	Obs	Mean	Std. Dv.	Min	Max	SAHOL	Obs	Mean	Std. Dv.	Min	Max
ω_1	144	0.290	0.317	0.013	1.709	ω_1	144	0.417	0.400	0.025	1.935
α_1	144	0.088	0.074	-0.119	0.271	α_1	144	0.083	0.107	-0.239	0.349
β_1	144	0.627	0.310	-0.744	1.021	β_1	144	0.505	0.411	-0.956	1.073
γ_1	144	1.117	0.112	0.864	1.531	γ_1	144	1.017	0.091	0.809	1.318
Return	144	-9.1E-05	0.006	-0.019	0.014	Return	144	-4.4E-05	0.012	-0.025	0.123
Amihud	144	2.09E-10	1.83E-10	0	8.98E-10	Amihud	144	5.9E-10	9.0E-10	0	1.0E-08
GARAN	Obs	Mean	Std. Dv	Min	Max	THYAO	Obs	Mean	Std. Dv.	Min	Max
ω_1	144	0.202	0.218	0.008	1.581	ω_1	144	0.112	0.152	-0.014	1.151
α_1	144	0.114	0.075	-0.119	0.397	α_1	144	0.097	0.040	0.032	0.239
β_1	144	0.689	0.227	-0.606	1.033	β_1	144	0.794	0.165	-0.238	0.963
γ_1	144	1.154	0.161	0.859	1.709	γ_1	144	1.368	0.252	0.972	2.431
Return	144	6.57E-05	0.012	-0.036	0.113	Return	144	-4.0E-04	0.007	-0.027	0.0293
Amihud	144	1.73E-10	9.97E-10	0	1.19E-08	Amihud	144	3.24E-11	2.21E-11	0	1.03E-10
ISCTR	Obs	Mean	Std. Dv.	Min	Max	TUPRS	Obs	Mean	Std. Dv.	Min	Max
ω_1	144	0.382	0.358	0.039	1.807	ω_1	144	0.309	0.362	0.004	1.990
α_1	144	0.085	0.120	-0.244	0.601	α_1	144	0.055	0.060	-0.147	0.192
β_1	144	0.540	0.389	-0.737	1.116	β_1	144	0.641	0.366	-0.955	1.066
γ_1	144	1.031	0.113	0.788	1.380	γ_1	144	1.106	0.107	0.902	1.568
Return	144	-3.3E-04	0.007	-0.022	0.038	Return	144	-0.002	0.018	-0.208	0.026
Amihud	144	4.59E-10	8.61E-10	0	7.62E-09	Amihud	144	2.15E-10	9.83E-10	0	1.19E-08

In Table 22, the correlation matrices for WACD estimates from the previous summary statistics table can be examined. The link between the duration and liquidity is investigated by comparing the Amihud Illiquidity measure and return values with the WACD estimates. The goal here is to understand the inter-relationships from a cross-sectional point of view.

Table 22. Correlation Matrices for WACD (1,1) Estimations vs Amihud and Return

BIMAS	ω_1	α_1	β_1	γ_1	Return	Amihud	KCHOL	ω_1	α_1	β_1	γ_1	Return	Amihud
ω_1	1.00						ω_1	1.00					
α_1	0.24	1.00					α_1	0.14	1.00				
β_1	-0.96	-0.51	1.00				β_1	-0.95	-0.45	1.00			
γ_1	-0.11	-0.24	0.17	1.00			γ_1	-0.10	-0.13	0.14	1.00		
Return	-0.13	-0.05	0.13	0.01	1.00		Return	0.07	-0.03	-0.05	-0.17	1.00	
Amihud	0.11	0.14	-0.15	-0.09	-0.12	1.00	Amihud	-0.07	-0.02	0.06	0.10	-0.14	1.00
EKGYO	ω_1	α_1	β_1	γ_1	Return	Amihud	PETKM	ω_1	α_1	β_1	γ_1	Return	Amihud
ω_1	1.00						ω_1	1.00					
α_1	0.06	1.00					α_1	-0.13	1.00				
β_1	-0.94	-0.40	1.00				β_1	-0.99	-0.02	1.00			
γ_1	-0.21	-0.43	0.35	1.00			γ_1	-0.15	-0.13	0.17	1.00		
Return	0.14	0.07	-0.15	-0.10	1.00		Return	0.12	0.05	-0.13	-0.26	1.00	
Amihud	0.10	-0.06	-0.07	-0.12	0.05	1.00	Amihud	-0.14	-0.05	0.15	-0.15	-0.08	1.00
EREGL	ω_1	α_1	β_1	γ_1	Return	Amihud	SAHOL	ω_1	α_1	β_1	γ_1	Return	Amihud
ω_1	1.00						ω_1	1.00					
α_1	-0.17	1.00					α_1	-0.04	1.00				
β_1	-0.97	-0.06	1.00				β_1	-0.97	-0.22	1.00			
γ_1	-0.19	-0.22	0.24	1.00			γ_1	-0.21	-0.04	0.22	1.00		
Return	0.08	0.09	-0.10	-0.15	1.00		Return	0.06	-0.14	-0.02	-0.04	1.00	
Amihud	0.03	-0.10	-0.01	-0.24	-0.06	1.00	Amihud	0.03	-0.11	0.00	0.05	0.69	1.00
GARAN	ω_1	α_1	β_1	γ_1	Return	Amihud	THYAO	ω_1	α_1	β_1	γ_1	Return	Amihud
ω_1	1.00						ω_1	1.00					
α_1	0.00	1.00					α_1	0.28	1.00				
β_1	-0.95	-0.31	1.00				β_1	-0.98	-0.48	1.00			
γ_1	-0.18	-0.08	0.20	1.00			γ_1	-0.25	-0.37	0.31	1.00		
Return	0.11	-0.15	-0.05	-0.11	1.00		Return	0.06	0.04	-0.06	-0.19	1.00	
Amihud	0.10	-0.15	-0.05	0.03	0.79	1.00	Amihud	0.06	0.02	-0.06	-0.10	-0.15	1.00
ISCTR	ω_1	α_1	β_1	γ_1	Return	Amihud	TUPRS	ω_1	α_1	β_1	γ_1	Return	Amihud
ω_1	1.00						ω_1	1.00					
α_1	0.16	1.00					α_1	0.09	1.00				
β_1	-0.96	-0.44	1.00				β_1	-0.99	-0.25	1.00			
γ_1	-0.16	-0.42	0.27	1.00			γ_1	-0.13	0.15	0.11	1.00		
Return	0.01	0.01	-0.01	-0.01	1.00		Return	-0.32	0.04	0.31	-0.12	1.00	
Amihud	0.03	-0.08	-0.01	0.00	0.16	1.00	Amihud	0.36	-0.03	-0.35	0.02	-0.94	1.00

Fixed-Effect and Random-Effects regressions can be observed in Table 23 and Table 24 below. Both regressions are statistically significant, as reported with F and Wald Chi-Square tests. By using the chi-square test results, we fail to reject the null hypothesis that RE model is suitable. A Hausman test is deployed to be able to understand which regression is preferable, and the test results are given in Table 25.

Table 23. Fixed-Effect Regression Results for WACD (1,1) Coefficients with Amihud and Return

Variables	FE Regression
ω	0.002
	(0.012)
α	-0.001
	(0.011)
β	0.003
	(0.011)
Amihud	1.28E+05
	(-3.0E+06)
γ	-0.008***
	(0.002)
Constant	0.006
	(0.014)
Observations	1,440
No. of companies	10
R-squared	0.013

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 24. Random-Effects Regression Results for WACD (1,1) Coefficients with Amihud and Return

Variables	RE Regression
ω	0.000
	(0.017)
α	-0.001
	(0.017)
β	0.001
	(0.017)
Amihud	4.26E+04
	-2.95E+05
γ	-0.006***
	(0.002)
Constant	0.005
	(0.017)
Observations	1,440
No. of companies	10
R-squared	0.113

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Regarding the Hausmann test, we rejected the null hypothesis of the RE model being appropriate, and we decided to continue with RE panel regression. Friedman test is used for checking cross-sectional independencies in Table 26.

Table 25. Hausmann Test for RE and FE Panel Regressions

	FE Panel	RE Panel	Difference	Std.E
ω	0.002	0.000	0.002	0.001
α	-0.001	-0.001	0.001	0.001
β	0.003	0.001	0.002	0.001
Amihud	1.3E+05	4.3E+04	8.6E+04	5.8E+04
γ_1	-0.008	-0.006	-0.002	0.001
Chi-square	2.23			
Prob>chi2	0.1353			

Table 26. Friedman's Test for Cross-Sectional Independencies

	Correlation matrix of residuals									
	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10
r1	1									
r2	0.419	1								
r3	0.098	0.156	1							
r4	0.211	0.468	0.365	1						
r5	0.315	0.421	0.197	0.357	1					
r6	0.157	0.327	0.068	0.260	0.233	1				
r7	0.408	0.672	0.173	0.449	0.442	0.320	1			
r8	-0.016	0.205	0.085	0.262	0.280	0.076	0.093	1		
r9	0.468	0.730	0.008	0.436	0.359	0.272	0.665	0.186	1	
r10	0.102	0.475	0.037	0.217	0.163	0.150	0.491	-0.089	0.348	1

Friedman's test of cross-sectional independence = 547.845, Pr = 0.00

The null hypothesis that there are no differences between the variables is rejected regarding the Friedman Test results. RE regression is appropriate for the data. As it can be noticed from the regression results, the γ coefficient from the ACD's Weibull distribution error term specification is the only statistically significant variable. The coefficient for the Weibull's γ is negative, leading to a negative correlation between γ and return. These results are parallel to economic interpretations. Weibull's γ directly explains the slope of the Hazard function, duration clustering imposed by the slope is directly linked with return. As the γ of Weibull declines and that slope of the hazard function gets below one, the duration clustering begins. In Tables 27,28,29,30, and 31, the results for pair-wise pooled regressions are reported.

Table 27. Pairwise Pooled Regression for Hourly Amihud

Variables	Model 1	Model 2	Model 3	Model 4
γ	-0.007***			-0.006***
	(0.001)			(0.001)
β		-0.002**		-0.002*
		(0.001)		(0.001)
α			-0.002	-0.005
			(0.003)	(0.003)
Constant	0.011*	0.005*	0.004	0.012
	(0.001)	(0.001)	(0.003)	(0.011)
Observations	1,440	1,440	1,440	1,440
R-squared	0.021	0.011	0.001	0.028

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 28. Pair Regression for Return and Gamma

Variables	Return/Gamma
σ	-0.004**
	(0.002)
Constant	0.004**
	(0.002)
Observations	1440
R-squared	0.006

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 29. Pair Regression for Return and Amihud

Variables	Return/Amihud
Amihud	0.000
	(0.000)
Constant	-0.001
	(0.001)
Observations	1440
R-squared	0.000

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 30. Pair Regression for Return and Alpha

Variables	Return/Alpha
α	-0.002
	(0.002)
Constant	-0.000
	(0.000)
Observations	1,440
R-squared	0.000

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 31. Pair Regression for Return and Beta

Variables	Return/Beta
β	0.001
	(0.002)
Constant	-0.001
	(0.001)
Observations	1,440
R-squared	0.001

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

CHAPTER 6

CONCLUSION

This thesis investigates and models the transaction, price, and volume durations in Borsa Istanbul and shows the explanatory power of various ACD specifications and extensions on the stock market activities. Durations from intraday orderbook data display strong autocorrelation; therefore, a general framework from conventional literature and an ACD model is used to inspect market microstructure.

The analysis demonstrates that BIST durations can be successfully modeled using these methods by removing any autocorrelated dependencies parallel to developed financial markets. In addition, a framework for choosing the error term specification is provided by conducting analysis with various types of distributions and ACD model selections that are most commonly used in finance literature.

Statistically significant and positive coefficients from the models imply that the durations are strongly interrelated and can be estimated using past values of conditional and unconditional durations. The magnitude of the last duration can determine the next's timing, leading to duration clustering. Having a sum of coefficients of past durations close to one reveals the strong persistence of lagged durations in all cases. It is also found out that the cross-sectional differences within the most traded ten stocks reveal differing degrees of clustering in trade durations. Turkish duration dynamics appear to follow developing market examples rather than the developed ones like the US. As trading speed increases, intraday dynamics become ever more essential, and a more significant number of investors adopt HFT, one can expect to see a convergence in durations between worldwide financial markets.

The primary motivation behind the first part of this thesis was to generate a framework for modeling every type of market durations in BIST by comparing various models with distinctive specifications and extensions. In the second part, the aim was to use the outcomes of the respective framework to show the explanatory power of the intraday duration modeling, which cannot be clarified by traditional low-frequency liquidity models and proxies in the literature.

The results from the first part of the study show that the serial correlation between the durations can be eliminated and modeled by ACD models for different market scenarios. Distinctive error specifications and ACD extension can be used and compared by this framework, and every type of duration can be modeled in BIST. Moreover, the outcomes of the second part of the study reveal that intraday liquidity and duration modeled by the framework from the first part helps explaining equity market movements and the traditional Amihud Illiquidity model.

REFERENCES

Acharya, V. V., & Pedersen, L. H. (2005). Asset pricing with liquidity risk. *Journal of Financial Economics*, *77*(2), 375–410. https://doi.org/10.1016/j.jfineco.2004.06.007

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, *5*(1), 31–56. https://doi.org/10.1016/S1386-4181(01)00024-6

Amihud, Y., & Mendelson, H. (1980). Dealership market: Market-making with inventory. *Journal of Financial Economics*, *8*(1), 31–53.

Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics*, *17*(2), 223–249. https://doi.org/10.1016/0304-405X(86)90065-6

Amihud, Y., Mendelson, H., & Pedersen, L. (2005). Liquidity and Asset Prices. *Foundations and Trends in Finance,* 1(4), 269-364. http://doi.org/10.1561/0500000003

Bauwens, L., & Giot, P. (2000). The Logarithmic ACD Model: An Application to the Bid-Ask Quote Process of Three NYSE Stocks. *Annales d'Economie et de Statistique*, *60*. https://doi.org/10.2307/20076257

Bauwens, L., & Giot, P. (2001). *Econometric Modelling of Stock Market Intraday Activity.* Boston, Springer. https://doi.org/10.1007/978-1-4757-3381-5

Będowska-Sójka, B. (2018). Emerging and Mature Markets – Behavior of Low-Frequency Liquidity Measures. The Case of the German and Polish Stock Markets. *Problemy Zarzdania Management Issues*, (76), 24–36.

Beltratti, A., & Morana, C. (1999). Computing value at risk with high frequency data. *Journal of Empirical Finance*, *6*(5), 431–455.

Benic, V., & Franic, I. (2008). Stock Market Liquidity: Comparative Analysis of Croatian and Regional Markets. *Financial Theory and Practice*, *32*(4), 477–498.

Bhogal, S. K., & Variyam, R. T. (2019). Conditional Duration Models for High-Frequency Data: A Review on Recent Developments. *Journal of Economic Surveys*, *33*(1), 252–273. https://doi.org/10.1111/joes.12261

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Bollerslev, T., & Domowitz, I. (1993). Trading Patterns and Prices in the Interbank Foreign Exchange Market. *Journal of Finance*, *48*(4), 1421–1443.

Bradrania, R., & Peat, M. (2013). Characteristic Liquidity, Systematic Liquidity and Expected Returns. *Journal of International Financial Markets, Institutions and Money* https://doi.org/10.2139/ssrn.1913152

Breen, W. J., Hodrick, L. S., & Korajczyk, R. A. (2002). Predicting equity liquidity. *Management Science*, *48*(4), 470–483. https://doi.org/10.1287/mnsc.48.4.470.210

Brennan, M. J., & Subrahmanyam, A. (1996). Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. *Journal of Financial Economics*, *41*(3), 441–464. https://doi.org/10.1016/0304-405X(95)00870-K

Chai, D., Faff, R., & Gharghori, P. (2010). New evidence on the relation between stock liquidity and measures of trading activity. *International Review of Financial Analysis*, *19*(3), 181–192. https://doi.org/10.1016/j.irfa.2010.02.005

Chordia, T., Huh, S.-W., & Subrahmanyam, A. (2009). Theory-Based Illiquidity and Asset Pricing. *The Review of Financial Studies*, 22(9), 3629-3668 https://doi.org/10.1093/rfs/hhn121

Chordia, T., Roll, R., & Subrahmanyam, A. (2001). Market Liquidity and Trading Activity. *The Journal of Finance*, *56*(2), 501–530. https://doi.org/10.1111/0022-1082.00335

Dang, T. L., & Nguyen, T. M. H. (2020). Liquidity risk and stock performance during the financial crisis. *Research in International Business and Finance*, *52*, 101165. https://doi.org/10.1016/j.ribaf.2019.101165

Datar, V. T., Y. Naik, N., & Radcliffe, R. (1998). Liquidity and stock returns: An alternative test. *Journal of Financial Markets*, *1*(2), 203–219.

Demsetz, H. (1968). The Cost of Transacting. *The Quarterly Journal of Economics*, *82*(1), 33. https://doi.org/10.2307/1882244

Dimpfl, T., & Odelli, S. (2020). Bitcoin Price Risk—A Durations Perspective. *Journal of Risk and Financial Management*, *13*(7), 1–18.

Drost, F. C., & Werker, B. (2004). Semiparametric Duration Models. *Journal of Business & Economic Statistics*, *22*(1), 40–50.

Duffie, D., Gârleanu, N., & Pedersen, L. H. (2005). Over-the-Counter Markets. *Econometrica*, *73*(6), 1815–1847. https://doi.org/10.1111/j.1468-0262.2005.00639.x

Duffie, D., & Singleton, K. (2003). Credit Risk: Pricing Measurement and Management. In *Credit Risk: Pricing, Measurement, and Management*.

Dufour, A., & Engle, R. F. (2000). Time and the Price Impact of a Trade. *The Journal of Finance*, *55*(6), 2467–2498. https://doi.org/10.1111/0022-1082.00297

Easley, D., Kiefer, N., & O'Hara, M. (1997). One Day in the Life of a Very Common Stock. *Review of Financial Studies*, *10*(3), 805–835.

Easley, D., & O'Hara, M. (1992). Time and Process of Security Adjustment. *The Journal of Finance*, 47, 577-604. https://doi.org/10.1111/j.1540-6261.1992.tb04402.x

Engle, R. (2002). Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business & Economic Statistics*, *20*(3), 339–350.

Engle, R. F. (2000). The Econometrics of Ultra-High-Frequency Data. *Econometrica*, *68*(1), 1–22.

Engle, R. F., & Russell, J. R. (1998). Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica*, *66*(5), 1127–1162. https://doi.org/10.2307/2999632

Feng, D., Jiang, G. J., & Song, P. X.-K. (2004). Stochastic Conditional Duration Models with "Leverage Effect" for Financial Transaction Data. *Journal of Financial Econometrics*, *2*(3), 390–421. https://doi.org/10.1093/jjfinec/nbh016

Fernandes, M., & Grammig, J. (2006). A family of autoregressive conditional duration models. *Journal of Econometrics*, *130*(1), 1–23.

Gerhard, F., & Hautsch, N. (2002). Volatility estimation on the basis of price intensities. *Journal of Empirical Finance*, *9*(1), 57–89.

Glosten, L. R., & Harris, L. E. (1988). Estimating the components of the bid/ask spread. *Journal of Financial Economics*, *21*(1), 123–142. https://doi.org/10.1016/0304-405X(88)90034-7

Goyenko, R. Y., Holden, C. W., & Trzcinka, C. (2009). Do liquidity measures measure liquidity? *Journal of Financial Economics*, *92*(2), 153–181.

Grammig, J., & Maurer, K.O. (2000). Non-monotonic hazard functions and the autoregressive conditional duration model. *Econometrics Journal*, *3*, 16–38. https://doi.org/10.1111/1368-423X.00037

Groth, J., Cooper, S., & Avera, W. (1985). Liquidity, exchange listing, and common stock performance. *Journal of Economics and Business*, *37*, 19–33. https://doi.org/10.1016/0148-6195(85)90003-7

Guloglu, C. Z., & Ekinci, C. (2021). Liquidity measurement: A comparative review of the literature with a focus on high frequency. *Journal of Economic Surveys*, 36(1), 41-74. https://doi.org/10.1111/joes.12440

Gurgul, H., & Syrek, R. (2016). The logarithmic ACD model: The microstructure of the German and Polish stock markets. *Managerial Economics*, *17*(1), 77–92.

Hallin, M., Mathias, C., Pirotte, H., & Veredas, D. (2011). Market liquidity as dynamic factors. *Journal of Econometrics*, *163*(1), 42–50. https://doi.org/10.1016/j.jeconom.2010.11.005

Hasbrouck, J. (2009). Trading Costs and Returns for U.S. Equities: Estimating Effective Costs from Daily Data. *The Journal of Finance*, *64*, 1445–1477. https://doi.org/10.1111/j.1540-6261.2009.01469.x

Hautsch, N. (2004). *Modelling Irregularly Spaced Financial Data: Theory and Practice of Dynamic Duration Models*. Springer-Verlag. https://doi.org/10.1007/978-3-642-17015-7

Hautsch, N. (2012). *Econometrics of Financial High-Frequency Data*. Berlin, Springer-Verlag. https://doi.org/10.1007/978-3-642-21925-2

Ho, T., & Stoll, H. (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics*, *9*(1), 47–73.

Huptas, R. (2014). Bayesian Estimation and Prediction for ACD Models in the Analysis of Trade Durations from the Polish Stock Market. *Central European Journal of Economic Modelling and Econometrics*, *6*(4), 237–273.

Kyle, A. S. (1985). Continuous Auctions and Insider Trading. *Econometrica*, *53*(6), 1315–1335. https://doi.org/10.2307/1913210

Lancaster, T. (1992). *The Econometric Analysis of Transition Data.* Rhode Island, Cambridge University Press, https://doi.org/10.1017/CCOL0521265967

Lee, K.-H. (2011). The world price of liquidity risk. *Journal of Financial Economics*, *99*(1), 136–161. https://doi.org/10.1016/j.jfineco.2010.08.003

Lesmond, D., Trzcinka, C., & Ogden, J. (1997). A New Measure of Transaction Costs. *The Review of Financial Studies*, 12(5), 1113-1141

Lin, S. X., & Tamvakis, M. N. (2004). Effects of NYMEX trading on IPE Brent Crude futures markets: A duration analysis. *Energy Policy*, *32*(1), 77–82.

Lunde, A. (1999). A Generalized Gamma Autoregressive Conditional Duration Model. *Department of Economics, Politics and Public Administration, Aalborg University*

Marshall, B. R., Nguyen, N. H., & Visaltanachoti, N. (2013). Frontier Market Transaction Costs and Diversification. *Journal of Financial Markets*, 24, 1-24. https://doi.org/10.1016/j.finmar.2015.04.002

Meitz, M., & Teräsvirta, T. (2006). Evaluating Models of Autoregressive Conditional Duration. *Journal of Business & Economic Statistics*, *24*(1), 104–124.

Nneji, O. (2015). Liquidity shocks and stock bubbles. *Journal of International Financial Markets, Institutions and Money*, *35*, 132–146. https://doi.org/10.1016/j.intfin.2014.12.010

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. https://doi.org/10.1111/j.1467-6419.2007.00547.x

Pastor, L., & Stambaugh, R. (2002). Mutual fund performance and seemingly unrelated assets. *Journal of Financial Economics*, *63*(3), 315–349.

Pastor, L., & Stambaugh, R. (2003). Liquidity Risk and Expected Stock Returns. *Journal of Political Economy*, *111*(3), 642–685.

Pyrlik, V. (2013). Autoregressive conditional duration as a model for financial market crashes prediction. *Physica A: Statistical Mechanics and Its Applications*, *392*(23), 6041–6051. https://doi.org/10.1016/j.physa.2013.07.072

Stoll, H. R. (1978). The Pricing of Security Dealer Services: An Empirical Study of Nasdaq Stocks. *The Journal of Finance*, *33*(4), 1153–1172. https://doi.org/10.1111/j.1540-6261.1978.tb02054.x

Stoll, H., & Whaley, R. E. (1983). Transaction costs and the small firm effect. *Journal of Financial Economics*, *12*(1), 57–79.

Tinic, S. M. (1972). The Economics of Liquidity Services. *The Quarterly Journal of Economics*, *86*(1), 79–93.

Tsay, R. S. (2002). *Analysis of Financial Time Series*. John Wiley & Sons, Inc. https://doi.org/10.1002/0471264105

Tse, Y. K., & Yang, T. T. (2012). Estimation of high-frequency volatility: An autoregressive conditional duration approach. *Journal of Business and Economics Statistics*, *30*(4), 533–545. https://doi.org/10.1080/07350015.2012.707582

Uddin, M. H. (2009). Reexamination of stock liquidity risk with a relative measure. *Studies in Economics and Finance*, *26*, 24–35. https://doi.org/10.1108/10867370910946306

Vu, V., Chai, D., & Do, V. (2015). Empirical tests on the liquidity-adjusted capital asset pricing model. *Pacific-Basin Finance Journal*, *35*, 73–89. https://doi.org/10.1016/j.pacfin.2014.10.007

Zhang, M. Y., Russell, J. R., & Tsay, R. S. (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics*, *104*(1), 179–207. https://doi.org/10.1016/S0304-4076(01)00063-X