PARAMETER ESTIMATION OF A NOVEL STOCK PRICE MODEL

by

Nihal Bahtiyar

A Thesis Submitted to the
Graduate School of Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Industrial Engineering

Koç University

August, 2012

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Nihal Bahtiyar

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

____
Assoc. Prof. Mine Çaglar

____
Prof. Ayse Humeyra Bilge

____
Assoc. Prof. Lerzan Ormeci

Date: ______________________

To my parents...

ABSTRACT

In this study, a stock price process is considered as an integral with respect to a Poisson random measure which governs several parameters of the trading agents. This model is powerful since it reflects two important properties of high frequency financial data, long-range dependence and self-similarity. We estimate parameters of this model using real data. The estimation procedure is demonstrated on log-returns of a particular stock in banking industry from Istanbul Stock Exchange between February 2007 and December 2009. We estimate the Hurst parameter describing long-range dependence. The numerical values found here verify the long-range dependence assumption. We also estimate the order duration parameter which follows a Pareto distribution. Interarrival rate of orders are calculated under the assumption that they arrive according to a Poisson Process. Effect rate and effect function are numerically fitted to the price data.

ÖZETÇE

Bu çalışmada, hisse senedi işlemleri yapan yatırımcılar ile bağlantılı parametreleri yöneten bir rassal Poisson ölçüsüne göre integralden oluşan bir hisse senedi süreci dikkate alınmıştır. Bu model, yüksek sıklıktaki finansal verilerin iki önemli özelliği olan öz benzerlik ve uzun dönemli bağımlılığı yansıttığı için güçlüdür. Gerçek veriler kullanarak bu modelin parametrelerini tahminledik. Kestirim için, ubat 2007 ve Aralık 2009 arasında stanbul Menkul Kıymetler Borsasında işlem gören bankacılık sektörüne ait bir hisse kullanılmıştır. Uzun dönemli bağımlılığı gösteren Hurst parametresini tahminledik. Bulunan sayısal değerler uzun dönemli benzerlik varsayımını destekler nitelikteydi. Pareto dağılımına uyan emir süresi parametresinin kestirimi yapıldı. Bir Poisson süreci ile geldiği varsayılan emirlerin geliş hızı hesaplandı. Etki hızı ve etki fonksiyonu da gerçek verilerden yola çıkılarak numerik olarak hesaplandı.

ACKNOWLEDGMENTS

First, I would like to thank my advisor Assoc. Prof. Mine Caglar for her supervision and endless support.

I am thankful to Prof. Ayse Humeyra Bilge and Assoc. Prof Lerzan Ormeci for kindly accepting my last minute call for attending my committee and for their critical suggestions and comments.

I am also thankful to all my instructors and all my professors in my BSc and MSc period, I learnt so much from them.

I also thank TÜBİTAK (The Scientific and Technological Research Council of Turkey) for their financial support during my thesis studies. I also thank Istanbul Stock Exchange for sharing stock price data with us.

I thank to all my classmates Buşra Urek, Filiz Sayin, Umut Ariturk, Müge Guclu, Kübra Yaman. I am thankful especially to Görkem Sariyer and Abdullah Bilgin for their endless support for being life-long good friends and making times of trouble enjoyable. I am also thankful to my collegues Sezin Orucu, Zeynep Furtun, Emine Ekici and Omer Gozum and to my managers Evrim Yetismis and Canan Kalay for their insightful approach.

Support, love and considerate approach of my dear friends Duygu Sap, Duygu Balci, Esma Baytok, Nazife Bedel, Meltem Oz Poyraz, Meral Sonmez, Seyda Onursal, Eda Gungor, Hande Akkan Dikmen, Caner Ulusoy, Oktay Balaman and again Görkem Sariyer strengthen and empowered me. I thank them for their fellowship.

I am indebted to all my family especially to my parents Zubeyde and Aydin Bahtiyar for their worthless, eternal love, endless patience and devotion. I am proud to be their daugther. This thesis is dedicated to them because of the endless support that they provide me during my education. I also thank to all members of our family:

my grandmothers, grandfathers, aunt, uncle and cousins especially to Sevgi Gurler since they encourage me in each step of my life. I am also thankful to Isa Engiz as newest member of our family. His valuable support, cheering and encouraging approach and love will be lifelong precious for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

The impact of the behavior and the strategies of trading agents on financial markets has been widely studied in the literature. Different academic fields show a growing interest to agents' behavior as physical units determining the prices of stocks in the market and to the analysis of their interactions. The agents are most commonly categorized as "chartist" and "fundamentalist" [23], [24], [22]. The strategy of fundamentalists is based on buying more stocks as the price of stocks they own decrease and thus making profit. On the other hand, the investor group called chartist adopts the principle of buying more stocks as the prices of the ones they own increase while selling these when the prices decrease since they rely on the behavior of their competitors as well as on actual price movements. There might be also other agent behaviors and different categorizations. Our approach does not rely on these categorizations as explained in [5]. Each behavior is specified either by a specific function defining its characteristic or it is represented by a probability distribution. We will prefer the probability distribution approach which allows modeling of individual differences in the group. Agent behavior has also been studied by physicists which use network techniques [10], [11], [12]. Alfarano and Milakovic showed that network structure which was expressed by a matrix representing the interaction of agents[9]. Alfarano et al., introduce a stochastic model using this interaction matrix[14].

Long-range dependence and self-similarity properties, which are mostly observed together, are related to the correlation structure of stochastic process. From a general point of view, long-range correlations observed in a process is called "long-range dependence", while sudden fluctuations in almost every time scale are called "self-similarity". They refer to the conditions which are mathematically expressed in terms of correlation functions and probability distributions [15]. It is shown that these properties are also observed in financial data [16].

Stochastic analysts concentrate on two stochastic processes where long-range dependence and self-similarity are exactly modeled. These are the "Levy processes" and "fractional Brownian motion" whose theoretical characteristics were intensively investigated in the last decade. The references [17] and [18] can be shown as examples for the use of these processes as price process. The majority of recent studies concentrated on whether the condition of not allowing arbitrage in fair markets is satisfied in cases where these two processes are used. Arbitrage refers to making profit without taking any risks. Since fractional Brownian motion allows arbitrage, it is not appropriate for stock prices [19]. Levy process which does not allow arbitrage is suitable for stock models; We should note however it is not the only method to prevent arbitrage [25]. Since these two processes, which attract the attention of mathematicians in theoretical terms, require stochastic calculus, their simulations in computer media require special numerical methods.

Inspired by the above mentioned studies, Bayraktar et al. [16] constructed a stochastic process which is appropriate for simulation or numerical trials and allows for drawing theoretical conclusions. They established a semi-Markov process and included finite number of agents, however envisaged that this number goes to infinity in the limit [16]. As another simplification, they reflected agent behavior in the model only at stagnation periods. On the other hand, Kluppelberg and Kuhn, established a model on change of price in time by speculative information coming to the market according to a Poisson process [20]. After arriving at the market, this information will slowly affect the price, and this effect will decrease in time.

In this thesis, all analysis are based on a new price process proposed by Akcay [19] and presented in Akcay and Caglar [21]. The superiority of this stochastic process analyzed is that it can give fractional Brownian motion and stable Levy motion by different limiting processes. In this thesis, physical parameters of the considered model are predicted using real stock exchange data. We assume that the demand for a stock and the change in its price are directly proportional, and furthermore each buy order increases the stock price and each sell order decreases the price. Variables such as order time, order quantity and the duration a transaction are considered to be regulated by a Poisson random measure. The model also assumes that the trade duration of an agent follows a heavy tailed distribution and shows a long-range dependence as a result. This process being a semi-martingale does not allow arbitrage in the market.

In Chapter 2, we give the basic definitions for stochastic processes and statistical methods that we use in other chapters. In Chapter 3 of this study, we discuss the nature of high frequency financial data using both the stylized facts of high frequent data and the available models. In Chapter 4, we introduce the model introduced in [19] and describe the dataset used in parameter estimation. Then, we describe the parameters of the involved distributions and estimate those parameters using data retrieved from Istanbul Stock Exchange in Chapter 5. Finally, the conclusions are given in Chapter 6.

Chapter 2

PRELIMINARIES

In this chapter, we state certain definitions that we will be use in the sequel. Section 2.1 and Section 2.2 contains definitions respectively related to stochastic processes and statistics.

2.1 Stochastic Process Preliminaries

In this section, we give fundamental definitions from stochastic processes cited in [31]. We assume that the reader is familiar with axiomatic probability theory, martingale theory and stochastic processes. In particular, the notions of probability space Ω, σ-algebra, expected values, local martingale and finite variation process will be used below.

2.1.1 Fractional Brownian motion and Hurst parameter

Definition A real process X is called a *semi-martingale* with respect to a given filtration $\mathcal{F}_t$ if X can be decomposed as $X = M + V$ where M is a local martingale and V is a finite variation process [30].

A fractional Brownian motion $(B_H^t, t \geq 0)$ with Hurst parameter $0 < H < 1$ is a continuous Gaussian process such that

$$E[B_t^H] = 0, \;\; E[B_t^H B_s^H] = 1/2(\mid t \mid^{2H} + \mid s \mid^{2H} - \mid t - s \mid^{2H}) \qquad t, s \in \mathbb{R}_+ \quad (2.1)$$

The characteristic of this process dependent crucially on the Hurst parameter, H.

- If $H = 1/2$, (B_t^H) becomes the standard Brownian motion with independent increments which is a semi-martingale. This process is also well-known as the Wiener Process[27].

- If $0 < H < 1/2$, the process is not affected by long-past and fluctuates around mean.

- If $1/2 < H < 1$, the process represents long-range dependence property with hyperbolically decaying autocorrelation coefficient. Definition of this property will be given in Subsection 2.1.2.

2.1.2 Long-range dependence and self-similarity

Stock price data have two fundamental properties, long-range dependence and self-similarity. In this sub-section, we give definition of these properties for any process Z. We also give special case for fractional Brownian motion cited from [26].

Definition For any stochastic process Z, let $r(n) = E[Z_1(Z_{n+1} - Z_n)]$. If $\sum_{n-1}^{\infty} r(n) = \infty$, then Z has *long-range dependence* property.

In particular, for fractional Brownian motion, long-range dependence property is observed when following condition is satisfied

$$\sum_{n-1}^{\infty} r(n) = \infty \text{ where } r(n) = E[B_1^H(B_{n+1}^H - B_n^H)] \tag{2.2}$$

In long-range dependent processes, the data are correlated across arbitrarily large time lag.

Definition Let Z be any stochastic process, Z is called *self-similar* with Hurst parameter H if $Z_{\alpha t}$ and $\alpha^H Z_t$ has the same probability law. Particularly, a fractional Brownian motion is self-similar with Hurst parameter H when $(B_{\alpha t}^H, t \geq 0)$ has the same probability law as $(\alpha^H B_t^H, t \geq 0)$ [26]. If the data represent self-similarity, it has bursty paths over a wide range of time scales.

2.2 Stastistics Preliminaries

In this section, we define two statistical method that we will be used in the following chapters. First one is periodogram which is used to detect seasonal effect. Second is two-way analysis of variance that we use to distinguish the effect of months and week-days.

2.2.1 Periodogram

Periodogram is the empirical version of power spectral density [29]. The definition of periodogram is given below.

Definition Let X be a process. $\hat{S}_{X,n}(f) = \mid n^{-1/2} \sum_{k=1}^{n} X_k \exp^{2i\pi kf} \mid$ is called the order n periodogram of X.

In practice, the periodogram is generally computed by means of a fast Fourier transform that provides a sampled version of $\hat{S}_{X,n}(f)$ [29]. Peak points of the periodogram corresponds to periods observed in dataset and they allow to detect seasonal component of the data.

2.2.2 Two-way analysis of variance (ANOVA)

Analysis of variance(ANOVA) is used to detect the source of the variation within several treatments. In two-way ANOVA, the data are arranged a matrix x_{ij} refering observation for i^{th} block and j^{th} treatment where the term *block* refers to a matched group of observations from each population.

Let $\bar{x}_i$ be the mean of the observations in i^{th} block $(i = 1, 2, ..., b)$ and $\bar{x}_j$ be mean of the observations in j^{th} treatment $(j = 1, 2, ..., k)$ where b and k be number of treatments and number of blocks respectively. Let $\bar{\bar{x}}$ be general mean of data. These means are calculated as follows:

$$\begin{aligned}
\bar{x}_i &= \sum_{j=1}^{k} \frac{x_{ij}}{k} \\
\bar{x}_j &= \sum_{i=1}^{b} \frac{x_{ij}}{b} \\
\bar{\bar{x}} &= \sum_{j=1}^{k} \sum_{i=1}^{b} \frac{x_{ij}}{bk}
\end{aligned} \tag{2.3}$$

The purpose is to reduce the within-treatment variation to more easily detect differences between the treatment means. In this analysis, total sum of squares (SS_{Total}) is defined to be the sum of squares for error (SSE), the sum of squares for treatments (SST) and the sum of squares for blocks (SSB).

$$SS_{Total} = SSE + SST + SSB \tag{2.4}$$

Sum of squares are calculated by the

$$\begin{aligned}
SS_{Total} &= \sum_{j=1}^{k} \sum_{i=1}^{b} (x_{ij} - \bar{\bar{x}})^2 \\
SST &= \sum_{j=1}^{k} b(\bar{x}_j - \bar{\bar{x}})^2 \\
SSE &= \sum_{j=1}^{k} \sum_{i=1}^{b} (x_{ij} - \bar{x}_j - \bar{x}_i + \bar{\bar{x}})^2
\end{aligned} \tag{2.5}$$

MST and MSE stands for mean squares of treatments and mean squares of errors respectively. They are computed by dividing the sums of squares by their respective degree of freedom.

$$\begin{aligned}
MST &= \frac{SST}{k-1} \\
MSB &= \frac{SSB}{b-1} \\
MSE &= \frac{SSE}{n-k-b+1}
\end{aligned} \tag{2.6}$$

Test statistic is calculated using mean squares as follows:

$$F = \frac{MST}{MSE} \tag{2.7}$$

which is F distributed with $\nu_1 = k - 1$ and $\nu_2 = n - k - b + 1$ degrees of freedom.

If we want to test to determine whether the block means differ, test statistics becomes

$$F = \frac{MSB}{MSE} \tag{2.8}$$

where F is distributed with $\nu_1 = b - 1$ and $\nu_2 = n - k - b + 1$ degrees of freedom.

Chapter 3

HIGH FREQUENCY FINANCIAL DATA

In this chapter, we describe the high frequency data retrieved from Istanbul Stock Exchange by introducing the stylized facts observed and review the available models used for such data.

3.1 Stylized Facts

In this section, we state stylized facts by using high frequency stock price data. This dataset is retrieved from Istanbul Stock Exchange and consists of price of a single stock between February 2007 and December 2009 with granularity of second. Details about the data are given in Section 4.2. In this section, we analyze seasonal and periodical components of *logarithmic return* series of this stock since we continue with a logarithmic return model in the sequel. Logarithmic return series is denoted as X_t and defined by

$$X_t = \log(Y_t) - \log(Y_{t-1}), \tag{3.1}$$

where $Y(t)$ represents stock price at time t. First, we obtain daily log-return i.e. the differences of the logarithms of day-end closing prices. We use periodogram as defined in Section 2.2.1 to determine the frequency of seasonal components. Figure 3.1 represents periodogram of 3 years data. Although the most dominant three peak points could be selected, a significant period could not be identified since they did not correspond to periods such as day, week or months.

Since our dataset consists of high frequency data for 3 years, we analyze seasonality of daily data on weekly, monthly and annual time scale. In this part of analysis, we do not consider intra-day seasonality because intra-day characteristics will be analyzed

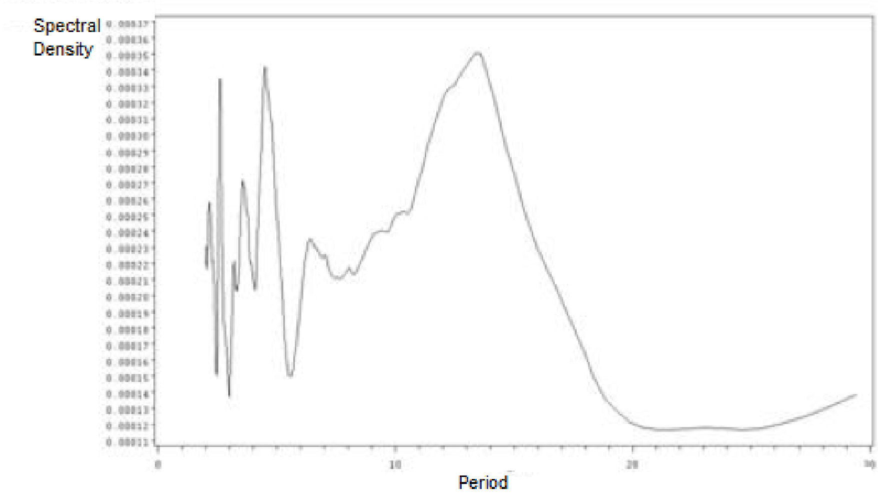

Figure 3.1: Smoothed Periodogram (Estimated spectral density function) obtained by using SAS 9.1 statistical software. "Smoothed" version refers to the periodogram obtained after removal of measurement noise. This method is applied to daily log-return series of 3 years of a single stock. No significant period is identified.

in estimation of buy-sell orders' physical parameters. For this reason, in this part also, we use daily log-return series that we have obtained using day-end price.

Daily, monthly and yearly averages obtained from the processed series are presented in Figure 3.2.

When the averages are subtracted from time series data, seasonality are removed and time series shown in Figure 3.3 is obtained. These steps are introduced as follows.

First, log-return series $X(t)$ is re-indexed as $X_{i,j,k,l}$ where the day t corresponds to weekday i, month of the year j and the year k it belongs and $l = 1, ..., L$ is in unit of week in a month. In this setting, $i \in \{1, 2, 3, 4, 5\}$ where $i = 1$ refers to Monday, $i = 2$ refers to Tuesday and so forth. $j \in \{1, 2, ..., 12\}$ where $j = 1$ corresponds to January, $j = 2$ corresponds to February and so forth. Similarly, $k \in \{1, 2, 3\}$ where $k = 1$, $k = 2$ and $k = 3$ refer to 2007, 2008 and 2009 respectively. l is used to distinguish first Monday and second Monday within a month. For this reason, $i \in \{1, 2, 3, 4, 5\}$. For example, let $X(t)$ be log-return in Thursday, 20^{th} of March, 2008. When we re-index, we obtain $X_{4,3,2,3}$ since $i = 4$, $j = 3$, $k = 2$ corresponds to Thursday, March and

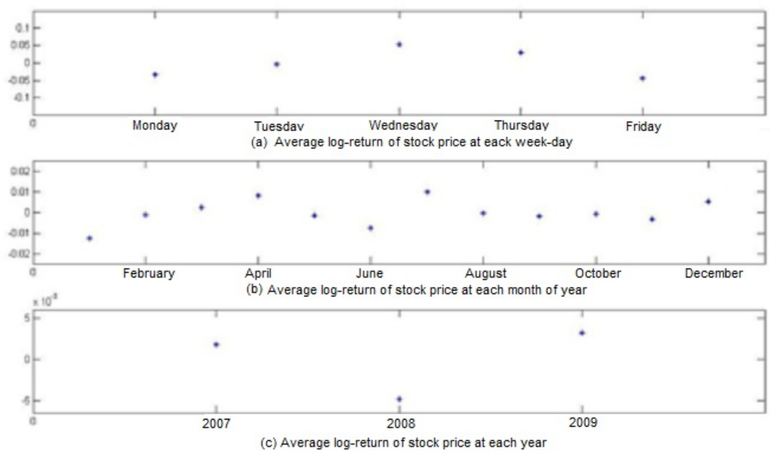

Figure 3.2: (a) Average of log-return price at each week-day. (b) Average of log-return price at each month of year. (c) Average of log-return price at each year.

2008 respectively. In March 2008, we have 4 different Thursday's corresponding to 6^{th}, 13^{th}, 20^{th} and 27^{th} of March. To identify 20^{th} of March, we use the last index as $l = 3$ since it is 3^{rd} Thursday in March.

After re-indexing procedure, we find general mean, daily, monthly and yearly zero-mean averages m, α_i, β_j, γ_k values are subtracted respectively and the following time series is obtained:

$$X_{(i,j,k,l)} - m - \alpha_i - \beta_j - \gamma_k \qquad i = 1, ..., 5,\ j = 1, ..., 12,\ k = 1, ..., 3,\ l = 1, ..., 5. \quad (3.2)$$

thus $X_{i,j,k,l}$ is purified from averages according to weekday i, month of the year j and the year k it belongs. Remaining oscillations are presented in Figure 3.3. The decrease towards the middle of the series corresponds to financial crisis occurred in 2008. The difference between general average m and 0 was not found to be statistically significant. Therefore, it will be assumed as 0 below.

Whether daily or monthly effects indicated in Equation 3.2 and their interactions have a statistically significant effect is analyzed with analysis of variance(ANOVA) method. Since the number of years is low here, annual averages will not be considered.

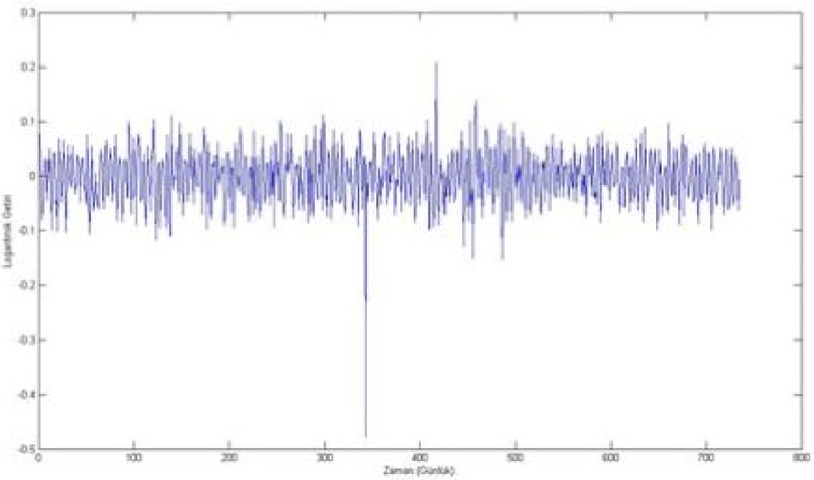

Figure 3.3: Time series obtained after removal of trend(general mean, m) and seasonal means(mean at weekday i, α_i; mean at month of the year j, β_j; mean in year k, γ_k.

The results of SAS GLM operation are presented in Figure 3.4. According to these results, general linear model is not significant with 0.1033 P-value. When the effects are analyzed individually, the effect months is slightly significant. Due to high P-values, days alone or days and months collectively do not have an effect. Since the effect of the months is not very strong, no seasonality is assumed for our data. If the effect of months is desired to be included in the model, by fitting a curve to the monthly averages in Figure 3.3, this effect can be estimated.

Since Hurst parameter is a shape parameter, it is independent from extracted averages. For this reason, to predict the Hurst parameter, logarithmic return series will be used without even subtracting monthly averages in Section 5.2.1. In addition to seasonality, how long the Hurst parameter remains constant is also related to the stationarity of the series and will be analyzed.

3.2 Available Models

In this section, we first introduce two fundamental models in stock pricing literature: fractional Bachelier Model with its fractional version and Black and Scholes Model

The GLM Procedure

Dependent Variable: Log-Returns

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	59	0.10333506	0.00175144	45,658.00	0.1033
Error	675	0.94380263	0.00139823		
Corrected Tot:	734	10,471,377.00			

R-Square	Coeff Var	Root MSE	Y- Mean
0.098683	22050.08	0.037393	0.00017

Source	DF	Type I-SS	Mean Square	F Value	Pr > F
Month	11	0.02789387	0.00253581	29,587.00	0.0483
Day	4	0.00405437	0.00101359	0.72	0.5751
Month*Day	44	0.07138682	0.00162243	42,370.00	0.2254

Source	DF	Type III-SS	Mean Square	F Value	Pr > F
Month	11	0.02667818	0.00242529	26,665.00	0.0622
Day	4	0.00312037	0.00078009	0.56	0.6933
Month*Day	44	0.07138682	0.00162243	42,370.00	0.2254

Figure 3.4: SAS output for variance analysis showing that month of year, day of week and their interaction do not have statistically significant effect on log-return.

with its fractional version. Then, we continue with more recent models having agent based approach in stock price modeling.

Bachelier model is an additive model as follows

$$Y_t = Y_0 + \nu t + \sigma B_t \qquad t \in [0, T]. \tag{3.3}$$

In equation 3.3, Y represents the stock price process while B represents a Brownian motion, $\nu \in \mathbb{R}$ and $\sigma, T \in \mathbb{R}_+$ where σ is the volatility and ν is the drift of the model. In fractional version of Brownian motion, B is replaced by fractional Brownian motion, as follows

$$Y_t = Y_0 + \nu t + \sigma B_t^H \qquad t \in [0, T]. \tag{3.4}$$

The Black-Scholes model is another well-analyzed model for stock prices given by

$$Y_t = Y_0 e^{(r+\nu)t+\sigma B_t} \qquad t \in [0,T]. \tag{3.5}$$

In that model, we also see the impact of interest rate, $r \in \mathbb{R}_+$, different from Bachelier model. Similarly, in fractional version of this model, fractional Brownian motion, B^H is used instead of Brownian motion B as given below.

$$Y_t = Y_0 e^{(r+\nu)t+\sigma B_t^H} \qquad t \in [0,T]. \tag{3.6}$$

Besides these well-known models, there are several approaches in stock price modeling. We continue with agent-based approaches.

In [24], Lux considers that all agents belong to two different groups according to their trading behavior: fundamentalists or chartists. Fundamentalists buy when price is below and sell when price is above of a fundamental value which is assumed to be known with certainty and to be constant. On the other hand, chartists rely on the behavior of their competitors as well as on actual price movements and they have also two subgroup as optimist chartists and pessimist chartists.

In [23], agents are able to compare strategies and change their behavior accordingly. It is assumed that birth rate of the process is equal to its death rate, i.e. the model assumes that a constant portion of agents is replaced by new entrants. The probability of exit trading is equal for chartists and fundamentalists while new entrants are assumed to first to act as chartists. In this setting, price changes are modeled as endogenous responses of the market to imbalances between demand and supply. Imbalances in the market occur according to transition probability from fundamentalists to chartists, optimist chartists to pessimist chartist and vice et versa since these transitions change supply and demand to the market. Change in fundamental value of fundamentalists is also another factor that governs change of demand and supply balance and consequently stock price. Further analysis of this model exist in the literature. Analysis of its time variation of higher moments is given in [14] while volatility clustering analysis is stated in [22].

In [11], deterministic functions are used to define strategies of different types of agents. According to their approach, each trading strategy is a signal processing element using past price information and current net order then provides a price formation process by inserting white noise. In this model, Farmer and Joshi consider two type of traders existing in the market and assumed that only market makers have an impact on the price formation. By modeling this type of agents impact, they assumed that positions, orders, and strategies are anonymous and also market maker must be risk neutral. They also assumed that both buy and sell orders have not a priori difference in their impact on price. The price process is constructed by the following dynamic system

$$p_{t+1} = p_t + \frac{1}{\lambda}\sum_{i=1}^{N}\omega^{(i)}(p_t, p_{t-1}, ..., I_t) + \xi_{t+1} \tag{3.7}$$

where p_t, ξ_{t+1}, $I(t)$, $\omega^{(i)}$, λ represent log-price at time t, white noise, any additional information at time t, order quantity of agent i, a scale factor that normalizes the order size respectively.

In [10], an agent based model is constructed by agents' decisions among three states: buying, selling or staying idle. In their decisions, all agents take benefits of public information ($\epsilon_t, t = 0, 1, 2, ...$) characterized by a sequence of independent and identically distributed Gaussian random variables where $\epsilon_t \sim N(0, D^2)$. The decision of each agent occurs sequentially by following these steps at each time period:

i) They receive a public information signal ϵ_t.

ii) Each of them compares the received signal with her own threshold value, $\theta_i(t)$. At each time period, any agent updates her threshold value with probability $0 < s < 1$.

iii) Each agent i compares public information signal with her own threshold values. If $|\epsilon_t| > \theta_i(t)$ then they give an order.

Excess of demand affects stock price and form a price process. As the result of simulation given in [10], it is illustrated that the model replicates self-similarity and long-range dependence of stock price data.

In [16], agents are divided into two groups: inert and active agents. For each agent, a semi-Markov process is associated with the trading activity where a Markov chain is used to determine the mood of an agent including sell, buy or inactive states. Active agents trade the stock frequently whereas the inert agents remain in the inactive state for a time period with a heavy-tailed distribution. When the effect of these two types of agents are combined, the stock price is approximated in law by a superposition of stochastic integrals with respect to a fractional Brownian motion and a Wiener process. The aggregate process does not allow arbitrage when the Hurst parameter of fractional Brownian motion which is inherited from the sojourn time of the inert agents in the inactive state is between 1/2 and 3/4 .

Chapter 4

STOCK PRICES: MODEL AND DATA

In this chapter, we first describe the essentials of the model that we use with its asymptotic limits. The model is an agent based model studied by Caglar [5]. It is based on a stochastic process modeling log-return of a single stock. We also describe the dataset that we use for parameter estimation in this chapter.

4.1 An Agent Based Model

The model used in the following of this work consists of a stochastic price process given by $Y_t = Y_0 e^{(mt+cZ_t)}$ where $Z(t)$ is the log-price process. The log price $Z(t)$ is modeled as aggregation of the effects of orders placed by all active agents in $[0,t]$ where $Z_0 = 0$ [5]. Effect of each order is proportional to its duration and volume. Other assumption of the model is positive correlation between the total net demand and the price change. Thereby, we expect that a buy order of an agent increases the price while a sell order decreases it. Since the process is a semi martingale, it does not allow arbitrage [19].

We assume that each agent's arrival is independent and identically distributed and occurs according to the underlying Poisson process and its duration is also random. We use the triplet (S_j, U_j, R_j) to represent a single order from an agent where S_j is the arrival time of the order, U_j is the duration of its effect on the price, and R_j is its rate. The rate is convertor to monetary units having positive values in case of a buy order and having negative values in case of a sell order. These triplets compose the atoms of a Poisson random measure N with mean measure μ.

Let $(\Omega, F, \mathbb{P})$ be a probability space. Let $\mathcal{B}_{\mathbb{R}}$ denote Borel-σ algebra on $\mathbb{R}$. Let N be a Poisson random measure on $\left(\mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}, \mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\mathbb{R}}\right)$ with mean measure

$$\mu(ds, du, dr) = \lambda ds\, \nu(du)\, \gamma(dr) \tag{4.1}$$

where $\lambda > 0$ is the arrival rate of the underlying Poisson Process, γ is the distribution of a random variable R and ν is a probability measure satisfying

$$\int_u^\infty \nu(dy) \sim h(u)\frac{u^{-\delta}}{\delta} \quad \text{as} \quad u \to \infty \tag{4.2}$$

where $1 < \delta < 2$ and h is a slowly varying function at infinity. The log-price process is constructed as

$$Z(t) = \int_{-\infty}^{\infty}\int_0^{\infty}\int_{-\infty}^{\infty} ru \left[f\left(\frac{t-s}{u}\right) - f\left(\frac{-s}{u}\right)\right] N\left(ds, du, dr\right). \tag{4.3}$$

In this setting, $f : \mathbb{R} \longrightarrow \mathbb{R}$ is a Lipschitz continuous function representing the effect of the order. In the limit, the log price process $Z(t)$ converges to a fractional Brownian motion (fBm) or a Lévy Process under different scalings [5].

4.2 Available Data

The data used in this paper are retrieved from Istanbul Stock Exchange including high frequent trading data in granularity of second between February 2007 and December 2009. We select most liquid stock to analyze which belongs to a company from banking industry in Turkey.

The data set that Istanbul Stock Exchange shares only for academic purposes contains not only price information but also order type, order quantity, transaction quantity, order time, order date and transaction time. Using the available information, we obtain log-price and log-second price processes in granularity of second, the duration and the interarrival distribution of orders and effects of these orders. Therefore, these data sets provide required information to estimate all parameters.

To obtain a unit interval time series, tick-by-tick interpolation method is used based on the assumption that the price remains constant if there is no occurring transaction [2]. SAS 9.1 statistical software is used for this purpose. Time series obtained for the selected stock price is presented in Figure 4.1. We use MATLAB in estimation of Hurst parameter while we use both SAS 9.1 and MATLAB in estimation of other parameters.

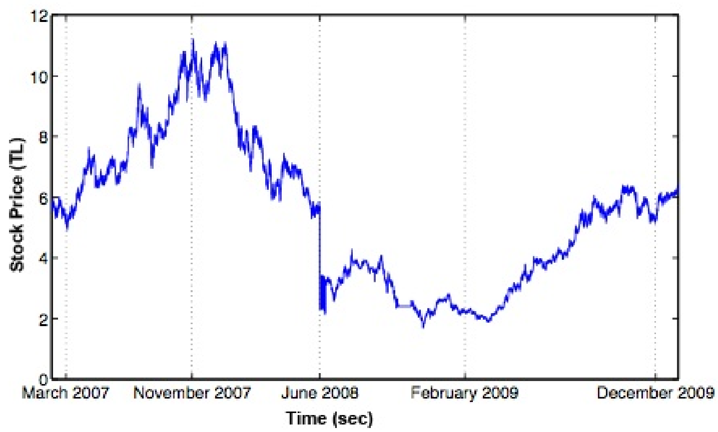

Figure 4.1: Time series plot of selected stock's price (between 02.02.2007-31.12.2009)

In Subsection 5.2.1, we consider entire dataset of 3 years to estimate Hurst parameter. We use six days data between 01.08.2008 and 08.08.2008 to estimate the interarrival rate of orders in Section 5.1, duration distribution in Subsection 5.2.2 and effect rate in 5.4 since these six days data have correct form of ID values enabling to distinguish different orders in the month of August of 2008. These first six days were selected due to technical obligation. During these six days, the orders were enumerated with 16-digit identification numbers and enable to identify different orders' beginning and ending times and other information stored. After the first six days, since order ID's exceeded 16 digits, Istanbul Stock Exchange system is not appropriate to distinguish different orders by always keeping the same number. Since the system is reset at the beginning of each month, it is possible to use the first days of other months but we choose August as representative.

Chapter 5

PARAMETER ESTIMATION

In this chapter, we estimate all three parameters of the model introduced in Section 4.1. In each section, we introduce the estimation methods and give estimation results of each parameter. Estimation results are also interpreted in the modeling context.

5.1 Interarrival of Orders

In this section, we analyze interarrival times of orders. Interarrival time is defined as the time difference between two consecutive orders. In stock markets, buy and sell orders given over night accumulate and they are transacted immediately or throughout the day. These over night orders will be excluded from our analysis of interarrival times.

In our analysis, we begin with classification of orders occurred in August 2008 according to their buy/sell order characteristic and their validity as day/session type of orders. Frequencies according to these classifications are presented in Table 5.1.

According to model presented in Equation 4.3, arrivals of orders occur accord-

Table 5.1: Observed numbers of orders used in interarrival estimation

	Sell Orders	Buy Orders	Buy and Sell Orders (Total)
Day type valid orders	27 862	36 370	64 232
Session type valid orders	76 037	71 517	147 554
Total number of valid orders	103 899	107 887	211 786
Canceled orders			60 485
Total number of orders			272 271

Table 5.2: Estimation of Interarrival of Buy and Sell Orders

	Observed Mean	Observed Standard Deviation	Estimated Mean	$\hat{\lambda}$
Buy Orders	3.295762	3.295855	3.295762	0.303420
Sell Orders	4.184979	4.185126	4.184979	0.238950

ing to a Poisson Process where interarrival times represents by random variable S_j. Therefore,the interarrivals are driven by independent and follow an exponential distribution. Interarrival rates of buy and sell orders are predicted in the light of this assumption of the model.

We fit independent exponential distributions to buy orders' and sell orders' interarrival times separately by using maximum likelihood estimator $\hat{\lambda}$ where it is given by

$$\hat{\lambda} = \frac{1}{\bar{x}}. \tag{5.1}$$

We obtain different λ for each distribution with 1/sec unit where mean of interarrival times is given by $1/\lambda$. Prediction results of both buy and sell orders are given in Table 5.2. As indicated in the table, an order occurs every 3-4 seconds.

Averages and standard deviations observed in buy and sell orders are almost equal. Although this result strengthened our exponential distribution hypothesis, Kolmogorov-Smirnov test is not significant due to "curse of dimension" with 211,786 orders in total.

In Figure 5.1 and Figure 5.2, probability density functions of day type buy and sell orders are respectively shown with their fitted distributions. These figures show a visual fit. Although statistical significance does not exist, this visual fit and nearly equal mean and standard deviations given in Table 5.2 support our conclusion that orders arrive according to a Poisson process.

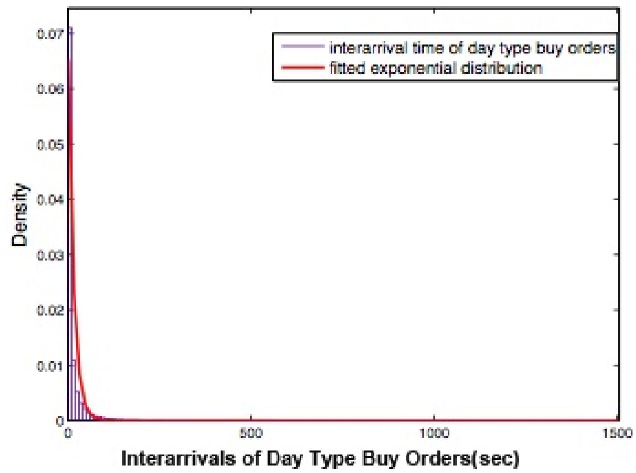

Figure 5.1: Estimated and empirical probability density function of interarrival times of day type buy orders occurs within first 6 days of August 2008

5.2 Order durations and Hurst parameter

In this section, we analyze both distribution of order durations and Hurst parameter of fBm characterizing long range dependence of the model. First, we estimate the Hurst parameter which is expected to be related to the tail parameter of order durations. Then, we continue by fitting a heavy tailed distribution to order durations and estimate its tail parameter.

5.2.1 Estimation of Hurst parameter

In this section, we estimate Hurst parameter of log-price process using data of 3 years as explained in Section 4.2. In estimation procedure, we use wavelet based estimator introduced by Vietch and Abry [3]. We begin with description of the model, then we state the results. Since wavelet estimation method is robust to linear or non-linear trend and seasonality effects, in estimation of Hurst parameter, logarithmic return series is used without removing trend and seasonality.

Estimator introduced in [3] is based on discrete wavelet transform of data. Let $d(j,k), k = 1, ..., n_j, j = 1, ..., J$ be discrete wavelet coefficients of data set $x(t)$. $d(j,k)$

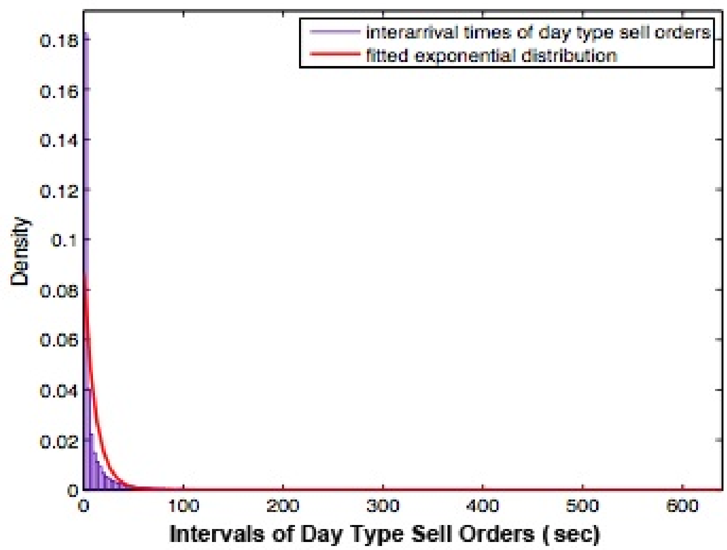

Figure 5.2: Estimated and empirical probability density function of interarrival times of day type sell orders occurs within first 6 days of August 2008

is found as follows:

$$d(j,k) = \int_{-\infty}^{\infty} x(t)\psi_{j,k}(t)dt, \quad j,k \in \mathbb{Z}, \tag{5.2}$$

where $\psi_{j,k}(t)$ are wavelet functions. These functions are generated by mother wavelet $\psi_0(t)$ according to following equation:

$$\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k), \quad j,k \in \mathbb{Z}. \tag{5.3}$$

Using these coefficients, expected value of squares of wavelet coefficients are denoted as μ_j and calculated as follows:

$$\mu_j = \frac{1}{n_j}\sum_{k=1}^{n_j} d_x^2(j,k) \tag{5.4}$$

where n_j is number of coefficients at octave j and $n_j \approx n2^{-j}$.

In the equation below, using μ_j and $g_j \sim \frac{-1}{n_j ln2}$, another function is calculated to guarantee unbiasdness.

$$y_j = \log_2(\mu_j) - g_j \tag{5.5}$$

At the end of these procedures, linear regression of y_j on j is found. The slope of regression line α equals to $2H - 1$. Thus, the estimator of $\widehat{H} = (1 + \alpha)/2$ is unbiased and consistent [3]. Log-log scale diagram in Figure 5.3 represents plot of y_j versus j and regression line where we use log-return data of 3 years with granularity of 1min.

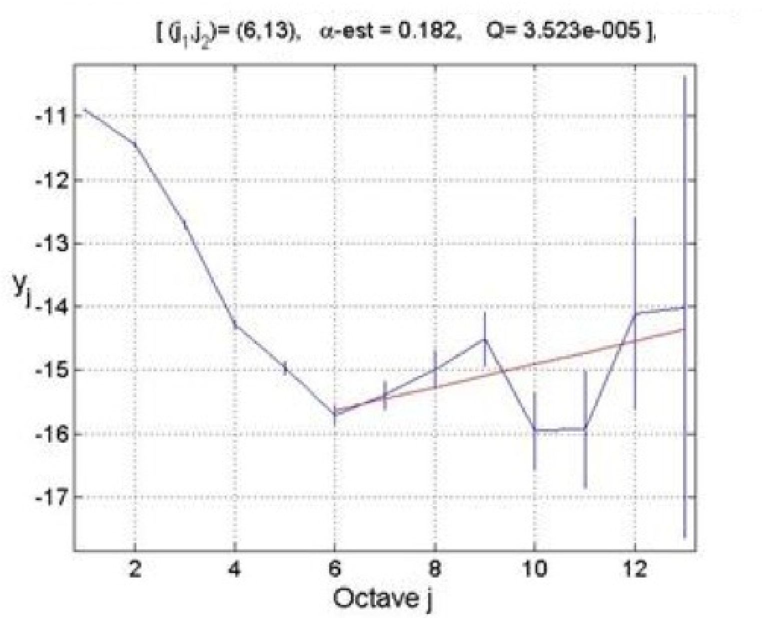

Figure 5.3: Scaling diagram in log-log scale where $y_j = \log_2(\mu_j) - g_j$ is plotted against octave j. Regression line gives estimate of Hurst parameter as the slope is equal to $2\hat{H} - 1$.

In calculation of estimated Hurst parameter, we use MATLAB codes shared by Vietch and Abry in [8]. We begin our estimations by keeping highest granularity that we have in the data. We obtain Hurst parameter estimates within $(0, 1/2)$ indicating negative correlations when a granularity smaller than 1 minute is used. This negative correlation reflects the reactional transactions of buyer and seller agents against each other as also explained in [2], [1].

Table 5.3 shows estimated Hurst parameter for the entire data set and its confidence interval calculated using log-returns in 1 minute, 5 minute and 10 minute granularity. Considering 1 minute, 5 minute and 10 minute log-returns, we still preserve the high frequency nature of the data [2]. The parameter is not estimated for the data in seconds to avoid the myopia effect that we observe in log-returns in seconds.

Table 5.3: Hurst Parameter Estimation for Entire Data Set

Properties of the Data			Estimation Results			
Aggregation Interval	Length of Aggregate Data	Maximum Available Octave	Initial Octave (j_1)	Final Octave (j_2)	Hurst Parameter (H)	Confidence Interval for H
1min	236539	15	7	15	0.546	[0.521, 0.572]
5min	47567	13	6	13	0.591	[0.548, 0.634]
10min	23838	12	5	12	0.596	[0.553, 0.639]
15min	15830	11	4	11	0.513	[0.476, 0.549]
20min	11954	11	6	11	0.627	[0.522, 0.732]
30min	7961	10	4	10	0.538	[0.482, 0.594]
60min	4299	9	4	9	0.524	[0.448, 0.619]

Results stated in Table 5.3 show that we observe Hurst parameter greater than 0.5 even in the lower bound of the confidence interval if we consider log-returns in 1 min, 5 min or 10 min. These results are consistent with the results found by Bayraktar in [16]. When we observe log-returns in time intervals longer than 10 minutes, the confidence intervals are wide since we lose the high frequency nature of the data and have shorter sequences of aggregate data. Therefore, rest of the work related to Hurst parameter estimation is performed for 1 min, 5 min and 10 min intervals.

To check whether the Hurst parameter changes by year or not, we analyze three years separately. Results of this analysis are given in Table 5.4.

In the results of log-returns in 1 minute and 5 minutes, the estimate Hurst parameter for 2009 is significantly less than those for 2007 and 2008. Even the results of 2008 and 2007 do not lie on the confidence interval of 2009. In the analysis of 10 min data, the estimates of Hurst parameter for 2008 and 2009 are nearly equal while they are less than that for 2007. This is considered as a signal of non-stationary data and the analysis is repeated for 1/2, 1/4, 1/8 and 1/16 of the data [4] to detect the

Table 5.4: Yearly Estimation of Hurst Parameter

	Results(2007)		Results(2008)		Results(2009)	
	Hurst Parameter	Confidence Interval	Hurst Parameter	Confidence Interval	Hurst Parameter	Confidence Interval
1min	0.622	[0.544, 0.701]	0.689	[0.507, 0.872]	0.537	[0.492, 0.582]
5min	0.622	[0.528, 0.717]	0.629	[0.423, 0.834]	0.531	[0.454, 0.609]
10min	0.574	[0.479, 0.669]	0.515	[0.430, 0.601]	0.517	[0.439, 0.594]

non-stationary part and the reason of this difference.

Table 5.5 shows the analysis results when we divide the data into 1/2. In this analysis, we keep initial octave (j_1) the same as that used in the analysis of the entire data stated in Table 5.3. First half of the data consist of log-returns in given intervals between 02.02.2007 and 16.08.2008 while second half consists of the data between 17.08.2008 and 31.12.2009. The difference between these two subsets of main data set is more apparent in 5 and 10 minute result.

Table 5.5: Estimation of Hurst Parameter for Each Half of the Data

	Results(First Half)		Results(Second Half)	
	Hurst Parameter	Confidence Interval	Hurst Parameter	Confidence Interval
1min	0.558	[0.520, 0.595]	0.534	[0.497, 0.572]
5min	0.626	[0.560, 0.693]	0.530	[0.464, 0.596]
10min	0.617	[0.551, 0.683]	0.524	[0.458, 0.590]

The above result evokes the necessity of the analysis for 1/4 of the dataset. These results are given in Table 5.6. In the analysis of each quarter of the data, estimated Hurst parameter of second quarter is notably lower than the rest of the data. It is even lower than the lower bounds of confidence intervals of the other parts of data. In order to detect the non-stationary parts of the data, we increase the number of splits in data set.

Table 5.7 gives the estimation results of Hurst Parameter for each piece corresponding to one eighth of the dataset. As indicated in Table 5.7, the estimated Hurst parameter in part 4 is the lowest one. This result is parallel to the decrease observed in the second quarter in Table 5.6. Moreover, it shows that part 4 is the main cause of the decrease observed in second quarter rather than part 3. Part 4 where we have lower Hurst parameter stands for the data between 07.04.2008 and 08.10.2008 where Late-2000's Global Financial Crisis begins. This fact illuminates the cause of the decrease in Hurst parameter in 5.8.

Table 5.6: Estimation of Hurst Parameter for Each Quarter of the Data

	1min (j_1=7, j_2=13)		5min (j_1=6, j_2=11)		10min (j_1=5,j_2=10)	
	Hurst Parameter	Confidence Interval	Hurst Parameter	Confidence Interval	Hurst Parameter	Confidence Interval
Part 1	0.550	[0.494, 0.607]	0.643	[0.538, 0.749]	0.567	[0.462, 0.673]
Part 2	0.558	[0.501, 0.614]	0.514	[0.409, 0.620]	0.333	[0.228, 0.439]
Part 3	0.561	[0.504, 0.617]	0.513	[0.408, 0.618]	0.542	[0.436, 0.648]
Part 4	0.467	[0.410, 0.523]	0.564	[0.458, 0.669]	0.510	[0.404, 0.615]

Table 5.7: Estimation of Hurst Parameter for Each Eighth Part of the Data

	1min (j_1=7, j_2=12)		5min (j_1=6, j_2=10)		10min (j_1=5,j_2=9)	
	Hurst Parameter	Confidence Interval	Hurst Parameter	Confidence Interval	Hurst Parameter	Confidence Interval
Part 1	0.478	[0.390, 0.567]	0.618	[0.442, 0.795]	0.559	[0.277, 0.640]
Part 2	0.609	[0.521, 0.698]	0.692	[0.515, 0.869]	0.595	[0.413, 0.777]
Part 3	0.499	[0.411, 0.588]	0.712	[0.535, 0.889]	0.653	[0.471, 0.834]
Part 4	0.332	[0.244, 0.421]	0.370	[0.193, 0.547]	0.312	[0.131, 0.494]
Part 5	0.585	[0.497, 0.674]	0.494	[0.317, 0.671]	0.673	[0.492, 0.855]
Part 6	0.563	[0.475, 0.652]	0.448	[0.272, 0.625]	0.544	[0.362, 0.725]
Part 7	0.461	[0.373, 0.550]	0.515	[0.339, 0.692]	0.505	[0.324, 0.687]
Part 8	0.455	[0.367, 0.543]	0.581	[0.405, 0.758]	0.574	[0.392, 0.755]

For further analysis of this effect, we estimate Hurst parameter for each sixteenth part of the data as detailed in Table 5.8. As indicated in Table 5.7, estimated Hurst parameter in part 4 is the lowest one. This result is parallel to decrease observed in second quarter in Table 5.6. Moreover, it shows that part 4 is the main cause of the decrease observed in second quarter rather than part 3. Part 4 where we have lower Hurst parameter stands for the data between 07.04.2008 and 08.10.2008 where Late-2000's Global Financial Crisis begins. This fact illuminates the cause of the decrease in Hurst parameter. For further analysis of this effect, we estimate Hurst parameter for each sixteenth part of the data as detailed in Table 5.8.

Table 5.8: Estimation of Hurst Parameter for Each Sixteenth Part of the Data

	1min (j_1=7, j_2=11)		5min (j_1=6, j_2=9)	
	Hurst Parameter	Confidence Interval	Hurst Parameter	Confidence Interval
Part 1	0.548	[0.404, 0.692]	0.626	[0.296, 0.955]
Part 2	0.542	[0.398, 0.686]	0.259	[-0.070, 0.589]
Part 3	0.631	[0.487, 0.775]	0.773	[0.443, 1.103]
Part 4	0.607	[0.463, 0.752]	0.614	[0.285, 0.944]
Part 5	0.382	[0.237, 0.526]	0.280	[-0.050, 0.610]
Part 6	0.625	[0.481, 0.769]	0.681	[0.351, 1.011]
Part 7	0.318	[0.174, 0.462]	-0.159	[-0.489, 0.171]
Part 8	0.581	[0.437, 0.725]	0.792	[0.463, 1.122]
Part 9	0.595	[0.451, 0.739]	0.520	[0.190, 0.850]
Part 10	0.631	[0.487, 0.775]	0.628	[0.298, 0.958]
Part 11	0.628	[0.484, 0.772]	0.484	[0.154, 0.813]
Part 12	0.569	[0.424, 0.713]	0.640	[0.310, 0.970]
Part 13	0.513	[0.369, 0.657]	0.435	[0.105, 0.764]
Part 14	0.500	[0.356, 0.644]	0.776	[0.446, 1.106]
Part 15	0.426	[0.282, 0.571]	0.566	[0.236, 0.896]
Part 16	0.491	[0.347, 0.636]	0.564	[0.234, 0.894]

In Table 5.8, we give analysis for 5 min and 10 min log-returns since we have not enough number of octaves for the analysis of 15min data. Non-stationarity is observed in Part 2 and Part 7 for especially in 5 min data. The decrease in Hurst parameter occurred in Part 2 is not observed in previous divisions while the decrease in Part 7 reflects the decrease occurred in previous divisions. In order to have a stationary data set, we exclude these two parts.

5.2.2 Estimation of Order Duration

In the price model, the order duration is associated with random variable U_j assuming that it follows a heavy-tailed distribution. To predict order duration parameter, we use data of first six days in August 2008 that we choose as we explained in Section 4.2. Table 5.9 shows frequency of analyzed orders according to their type, buy or sell, day or session type.

Table 5.9: Observed numbers of orders used in duration estimation

	Buy Orders	Sell Orders	Total
Transacted Day Type Orders	6851	6851	12515
Transacted Session Type Orders	21006	21006	35112
Total Transacted Orders	27857	27857	47627
Untransacted Orders	7819	7819	20450
Total Number of Orders	35676	35676	68077

Order durations are calculated by taking the difference between initial registration time of the order and last trade time of the order by combining the two datasets using order ID number. Figure 5.4 shows the histogram of orders given during the day while Figure 5.5 shows the histogram of the orders given during the session. Since session type orders are valid only during one session, the times of these types of orders involve a shorter period of time compared to day type orders. The distributions are heavy tailed as assumed by the model. In the estimation, we used generalized Pareto distribution as a well-known heavy-tailed distribution using MATLAB. Density function of generalized Pareto distribution is defined as

$$f(x|k,\sigma,\theta) = \left[1 + k(\frac{x-\theta}{\sigma})\right]^{-1-\frac{1}{k}} \tag{5.6}$$

where σ is the scale parameter, θ is the threshold parameter and k is the shape parameter effective on the tail in particular. The parameter δ that we have used in the model coincides with $1/k$ here. Hence, we can find the Hurst index to be $(3-1/k)/2$. Table 5.10 gives the parameters of the generalized Pareto distribution fitted on the order durations. Here, the threshold parameter is not constrained to be positive for the sake of fitting the tail better. As the model predicts, the important factor in self-similarity and long-range dependence is the tail parameter, while there is no special requirement on the initial part of the duration distribution.

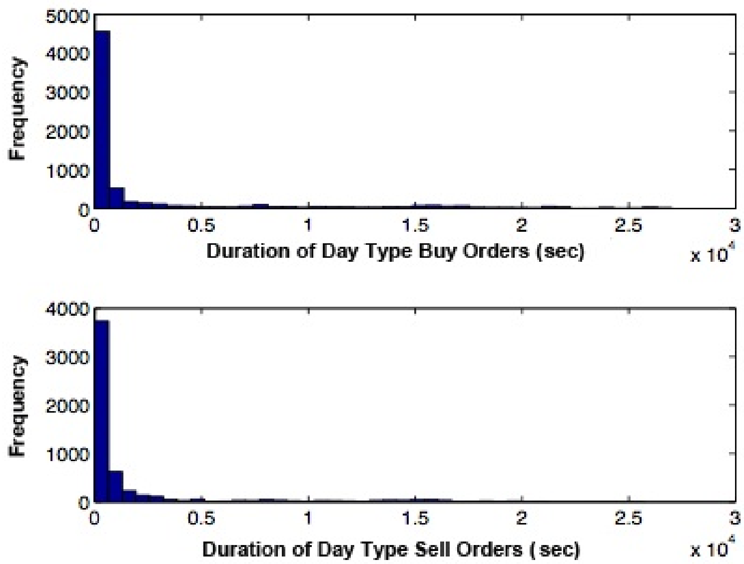

Figure 5.4: Histogram of day type orders occurred within first 6 days of August 2008 which follows heavy-tailed distribution.

Table 5.10: Estimation result of generalized Pareto distribution fit to order durations

	tail parameter (k)	scale parameter (σ)	threshold parameter (θ)	Hurst parameter (H)
Day type buy orders	0.695319	1359.58	-600	0.780905599
Day type sell orders	0.572437	1103.84	-480	0.626541436
Session type buy orders	0.643222	266.13	-120	0.722663404
Session type sell orders	0.615648	215.81	-90	0.687847601

Table 5.10 shows that the distributions of order durations give greater Hurst parameter estimates than direct estimation of Hurst parameters from price time series

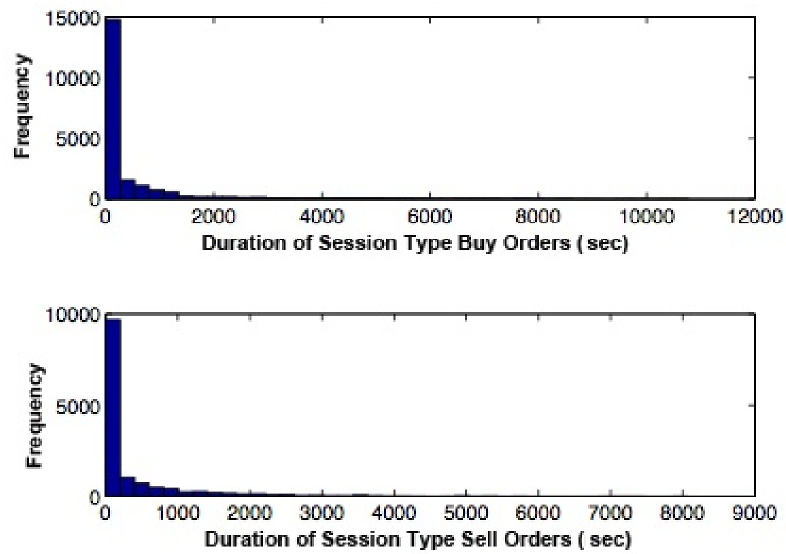

Figure 5.5: Histogram of session type orders occurred within first 6 days of August 2008 which follows heavy-tailed distribution.

using wavelet estimation method. However this difference is considered acceptable since models are idealized estimates of real data. In spite of this difference, order realization times have heavy tailed distribution as assumed by the model. Since the orders that are traded immediately upon arrival are not separately considered in the model, we plot also histogram of non-zero durations as seen in Figure 5.6 where only 52% of orders take at least 1 sec to transact completely. If we ignore zero duration orders, mean of non-zero durations is found as 1939sec and shape parameter is 0.7. On the other hand, if we take both zero and non-zero durations into account, mean of duration is found as 1010sec.

5.3 Effect function estimation

In this section, we begin with construction of effect function and estimation of its parameters. Then we use our dataset to estimate effect rate for this selected effect function by using assumptions of the model given in Section 4.1.

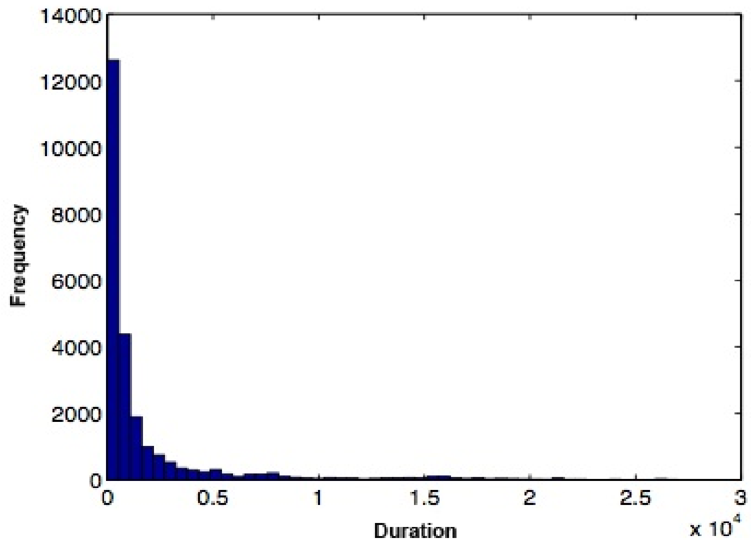

Figure 5.6: Histogram of all nonzero durations; excluded are those transactions which take place immediately.

In estimation of effect function and effect rate, we use same dataset that we have used in duration estimation. First, we estimate effect function f as defined in Section 4.1. Then we estimate effect rate r that we have considered as the monetary conversion parameter.

As explained in Section 5.2, our model does not consider the effect of orders that are traded instantaneously upon arrival. When we analyze realization rate of remaining orders, we note that 21,639 (76%) of all 28,357 orders were traded in one piece after a certain duration while only 24% were traded in several pieces in time. The distribution of these according to different types of orders is presented in Figure 5.7.

Since our price model did not consider orders that are traded instantaneously, we propose an exponential function which most fits to this dynamic, of the form

$$f(x) = \frac{\exp(ax) - 1}{\exp(a) - 1} \tag{5.7}$$

This function constructs a link between transacted quantity of each order and its

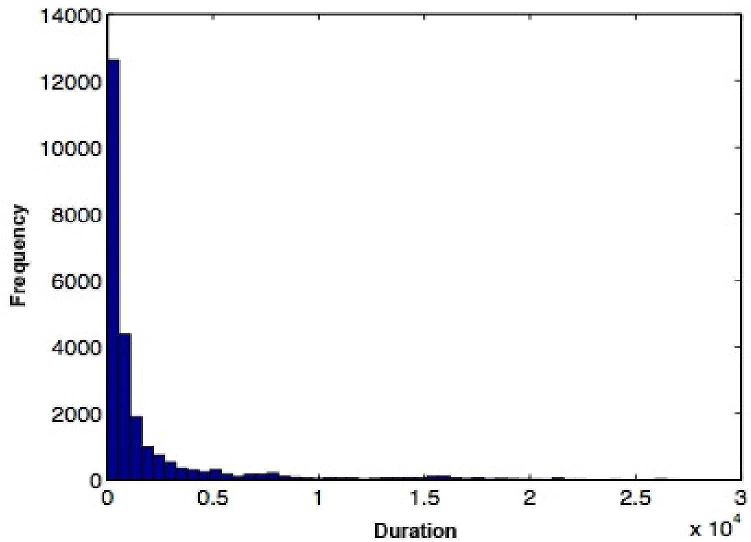

Figure 5.7: Histogram of all orders occurred within first 6 days of August 2008 that trades throughout the order duration or at the last minute

corresponding duration. The parameter $a > 0$, which we chose as parameter, models effect rate in a non-linear form. Thus, this continuous function represents orders as much as possible that transact in one piece after a certain time. The change of traded quantity in time for a given order is represented with this function. Standard f function is calculated for different values of a in Figure 5.8, it will be shifted according to order arrival time and duration as required by the model and will be scaled with effect parameter r. Here, we will use a constant r parameter to convert unit of quantity into monetary unit. At the end it will be multiplied by the amount of each order and will get a separate value for each order.

It is clear that the method we adopted changed the real model to represent the behavior in real data. We expressed difference between the orders with random variable a instead of r which is a parameter which linearly entered the model. Using *lsqcurvefit* function in MATLAB, we fitted f to additive trade realization function that each order formed while realization. Dividing observational function by total trade quantity, we obtained image in [0,1] interval; We scale also the domain of f into

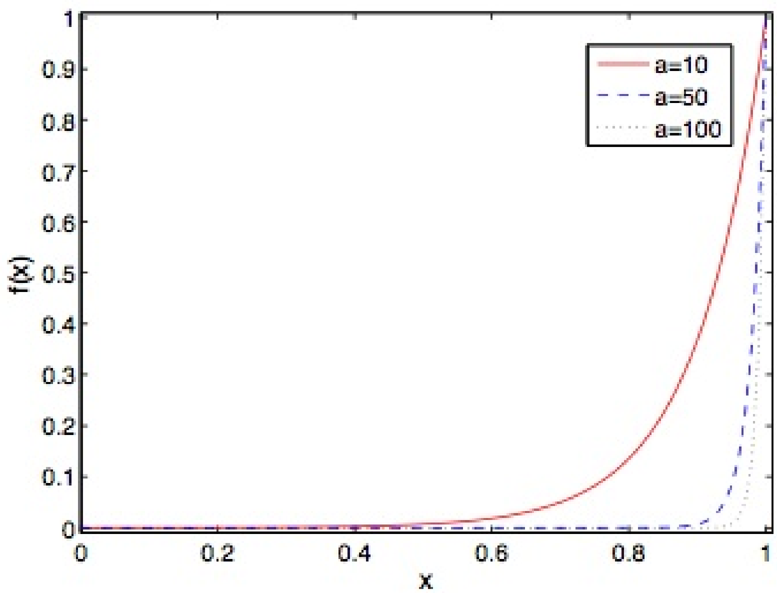

Figure 5.8: The effect function for various values of a

[0,1]. As a result, domain and codomain of observational and theoretical functions became equal. Figure 5.9 shows an order's dynamic where a was predicted as 1.2542 and used in the effect function f.

Figure 5.10 shows the frequency distribution of a estimated from the data. The accumulation observed around the value 700 matches with the orders that are all transacted at once, since the upper limit that we use in the least squares method is 700. This is even steeper than $a = 100$ curve of Figure 5.9. As the numeric precision in MATLAB limits us to this magnitude, 700 is chosen and a better precision is not looked for. The accumulation here is in accordance with the fact that 76% of 28357 orders are orders that are transacted all at once rather than piecewise.

Since distribution of a given in Figure 5.10 is bimodal, a single distribution cannot be fitted. Keeping orders that were realized at the last minute, a heavy tailed distribution like Pareto can be appropriate for the first section. Another option at the stage of simulation involves drawing random number using Bootstrap method from observed distributions, instead of fitting a parametric distribution.

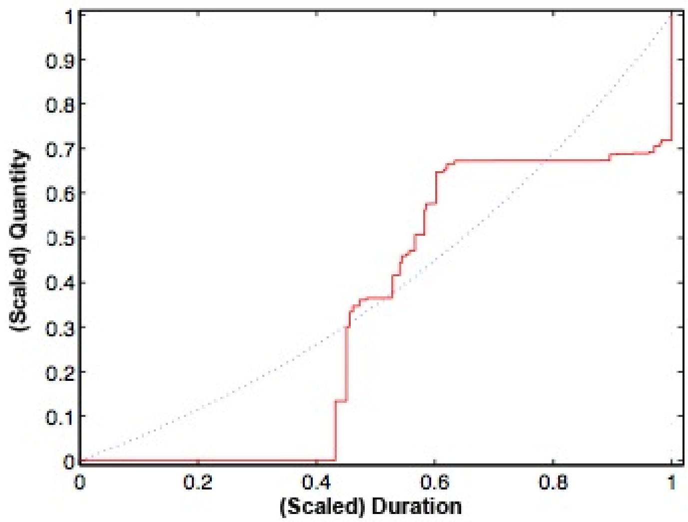

Figure 5.9: An order realization curve with estimated $a = 1.2542$

5.4 Estimation of effect rate

As expressed in Section 4.1, effect rate is represented by a random variable R_j and has a monetary conversion role. Assuming that the effect of any order i is directly proportional with trade duration in our model, expressing with

$$r_i u_i f\left(\frac{t - s_i}{u_i}\right) \tag{5.8}$$

We fit function f which determines the shape of effect in Section 5.3 after its normalization by dividing into transaction quantity. For this reason, r_i is directly proportional to quantity. According to expression 5.8, r_i also plays the role at converting time unit of trade duration into monetary unit of stock price change. To join these two facts, we define effect rate as $r_i = rQ_i$ where r is a conversion parameter. Accordingly, when we convert the total effect of an order i into money unit,

$$r_i u_i = rQ_i u_i$$

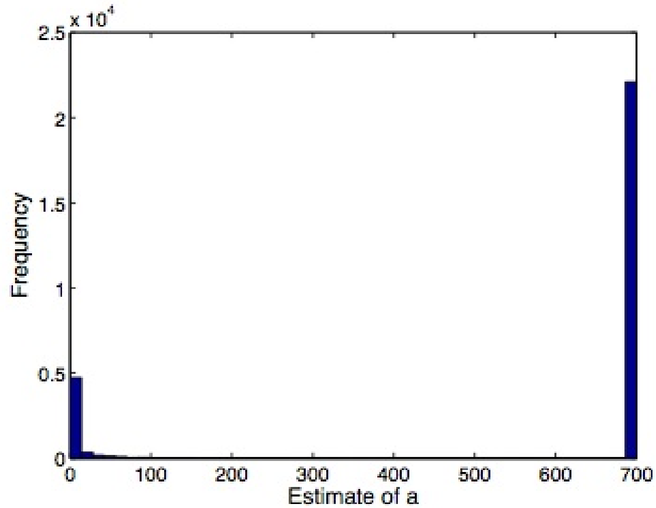

Figure 5.10: Distribution of parameter a of effect function

For this reason, to estimate r, we use data of six days used in prediction of duration distribution, defined in Section 4.2. Considering that buy orders have positive while sell orders have negative effects, r_i becomes a real valued random variable. Making use of this random variable and effect function f, we are able to calculate the effect of an order at any time until its last transaction. At this stage, it is appropriate to test one of the model assumptions stating the independence of effect rate and order duration. We need to apply the test into two random variables R and U defining effect rate and order duration respectively. According to our effect rate definition, we obtain $R = rQ_i$. For these reasons, we test independence of order quantity Q and order duration U. Scatter diagram of these variables is presented in Figure 5.11. According to this figure, no dependence is observed in visual sense even if a certain accumulation is observed on smaller times and quantities. Since correlation coefficient is -0.0732, we conclude the verification of model assumption by the fact that effect rate and order duration are independent random variables.

All trade quantities are presented in Figure 5.12 with positive values representing buy orders and negative values representing sell orders. When normal distribution is

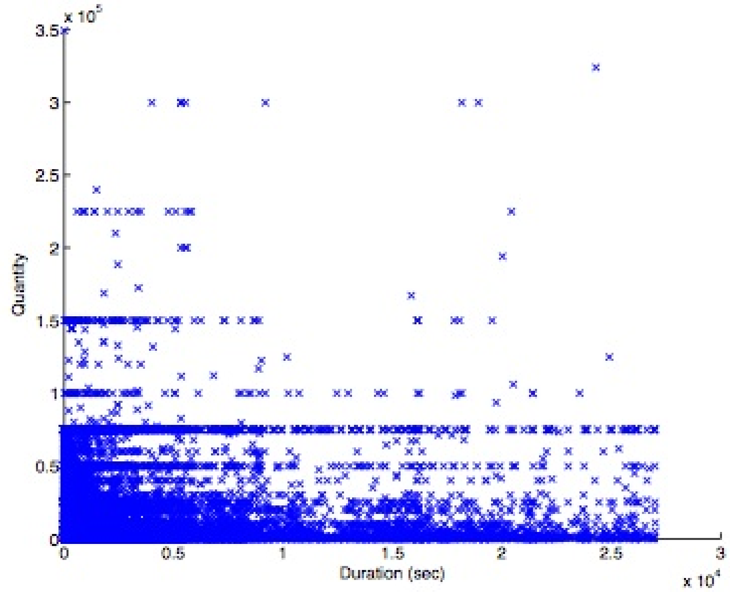

Figure 5.11: Scatter diagram of quantity versus duration

fitted into this histogram, mean is found as 1039.7 and standard deviation is found as 35442. In the following calculations, if we make 1/10000 scaling we will reach smaller numbers; it is convenient to continue with a distribution with 0.1 average and 3.54 standard deviation. Let the indexes of buy orders realized throughout a certain day be collected in set B, and indexes of sell orders be collected at set S. Let the transaction quantity of any order i is indicated with Q_i . Since all transactions will be finalized at the end of the day, for simplicity, we consider a common r value both for buy and sell orders. Considering that buy orders increase prices while sell orders decreases, the following relationship appears by the end of the day:

$$\sum_{i \in A} rQ_i u_i - \sum_{i \in S} rQ_i u_i = r \left[\sum_{i \in A} Q_i u_i - \sum_{i \in S} Q_i u_i \right] \propto \log P_1 - \log P_0. \tag{5.9}$$

where P_0 and P_1 indicates opening and closing prices of one day respectively. Left side of Equation 5.9 can be interpreted as the net effect of buy and sell on price. Since total effect of orders is taken into account due to day-long calculation, intermediate values

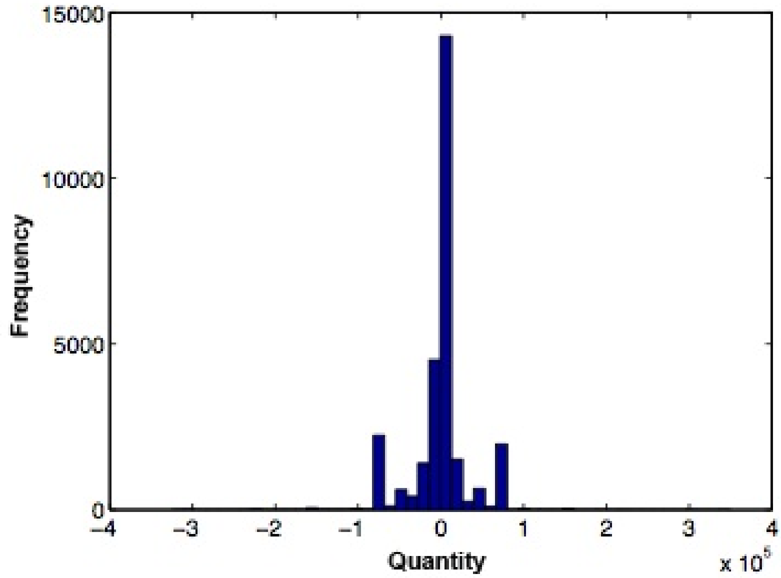

Figure 5.12: The distribution of order quantity; buy (positive) and sell (negative) orders

of effect function are not included in this calculation. For 6 days that the data are involved, we continue by calculating the terms in right and left side of Equation 5.9 in terms of r and fitting least squares line. The slope of the line gives scaled estimate of r, $\hat{r}$ where calculations are made by scaling trade quantity as 1/10000 and duration as 1/3600 to maintain the numerical sensitivity used by the computer. After calculating slope of regression line, we turn back to previous scale to obtain an estimate of the effect rate. In Figure 5.13, for 6-days period, $\sum_{i\in A} Q_i u_i - \sum_{i\in S} Q_i u_i$ values are plotted in x-axis while daily logarithmic return are plotted in y-axis. According to the figure the effect of trades is not related to logarithmic return. Moreover, the gradient of regression line is not significantly different than zero, correlation coefficient is extremely low. As a result of this observation, thinking that this effect is reflected on the prices after a certain period of time, we plot the right hand side of (2) with a delay of 1 and 2 days. The 2-day shifted state of the price changes is shown in Figure 5.14. Here, a positive linear relationship with a high correlation coefficient of 0.9258 is observed. The conversion factor r is found to be $1.93e - 05$. Although it is a very small number, it is significant with a P-value of 0.008. As a result of the regression, the y-intercept is found to be -0.013. The associated P-value is 0.03. Even though it is statistically meaningful to some degree, this parameter will be taken to be 0 in order for r to act as a direct factor. Although it is different from the original model, a shift in time does not fundamentally change the nature of the model. It just means that the price sequence will emerge as a result of the transactions only after a certain time delay.

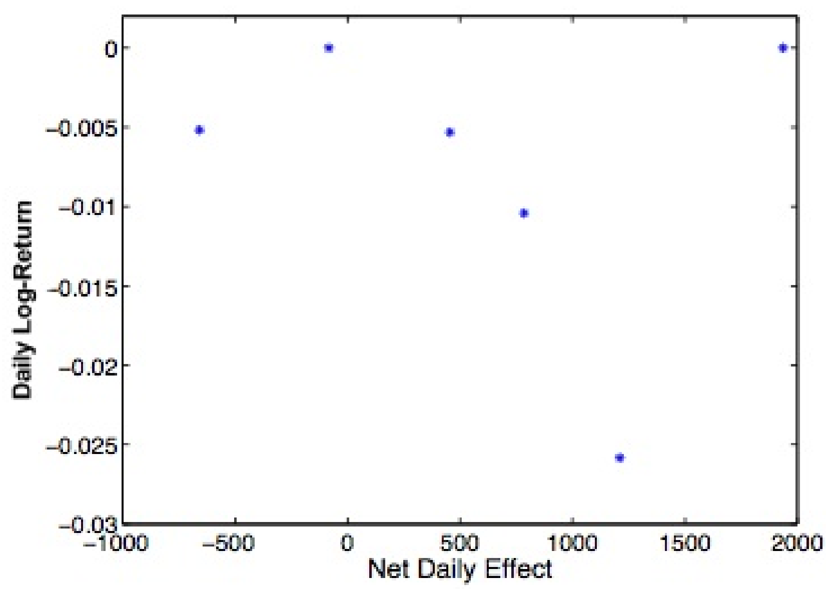

Figure 5.13: Logarithmic return versus net daily effect

To be used at the fitting stage, we assume that the factor r_i has a normal distribution since we have previously assumed that the amount distribution is normal. We will work back from the scalings above and recover the correct parameters. In view of the scaling for the amount, we had calculated the mean to be 0.1 and the standard deviation to be 3.54. Since the scaling for time amounts to converting seconds into hours, we can assume that the coefficients r_i come from a normal distribution with mean $(1.93e{-}05)(0.1) = 1.93e{-}06$ and standard deviation $(1.93e{-}05)(3.54) = 6.8322e{-}05$,

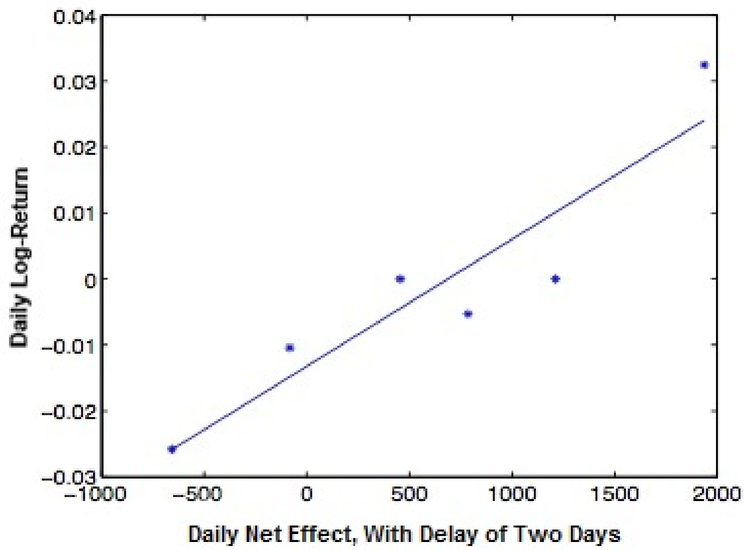

Figure 5.14: Daily log-return vs daily net contribution with delay of two days

and that time is measured in units of hours. If one measures time in units of seconds, one needs to make the mean and the standard deviation even smaller. The reason for obtaining numbers as small as above is that the effect at the end is proportional to the logarithm of the price.

Chapter 6

CONCLUSIONS

In this thesis, we have estimated all parameters of a stock price model using real data retrieved from Istanbul Stock Exchange. Since this stochastic model considers effects of agent-level behavior on stock price, we estimate agent-based parameters. Moreover, the model converges asymptotically to a fractional Brownian motion when high frequency trading occurs. For this reason, we express and estimate the long-range dependence parameter of the model as the Hurst parameter corresponding to the limiting fractional Brownian motion.

First, we have estimated interarrival of orders after classification of orders into two groups: buy and sell orders. We have fit independent exponential distributions to interarrivals of both buy and sell orders by using maximum likelihood estimation method. Although results of Kolmogorov-Smirnov test are not significant, we have found that estimated mean, observed mean and observed standard deviation are close to each other which support our assumption that arrivals occur according to a Poisson process.

We have estimated Hurst parameter governing long-range dependence property using wavelet based estimation method. Estimate of Hurst parameter is found around 0.6 for time scales greater than 1 minute. This results justifies long range dependence nature of high frequency stock price data. Then, we have estimated the order duration parameter. According to the model, order duration follows a heavy-tailed distribution. Histogram of order duration visually supports this assumption. We fit generalized Pareto distribution as a well-known heavy-tailed distribution after grouping orders according to their duration type as day/session type orders and buy/sell type. We have also calculated Hurst parameter this time by using its relationship with the tail parameter of generalized Pareto distribution. Even if the two estimates of Hurst parameter are not equal to each other, this difference is acceptable.

Finally, we have constructed an effect function that represents the local dynamics of trading of each order over its duration . Then, parameter of this function is estimated. We end up with estimation of effect rate. Effect rate is assumed as a multiple of order quantity and a monetary conversion parameter. Since all orders are finalized at the end of day, we consider daily price change to estimate the conversion parameter. In this procedure, we could not find a linear proportion between daily net contribution and daily log return. Since we observe a linear relation after two days of delay, we estimate conversion parameter with delayed data.

Our analysis of high frequency data have confirmed that mild long-range dependence is observed in stock prices at 1min granularity. While smaller scales involve reactional movements of buyers and sellers, the Hurst parameter is estimated to be around 0.6 using scales of 1min and larger. Our overall conclusion is that the stochastic price model fits the data well The Poisson arrivals and heavy tailed duration distribution assumptions hold and the effect of the buy and sell orders on the price can be numerically approximated by a deterministic function with a random parameter. Furthermore, all the estimations of this thesis are used in validation of the model by simulations reported elsewhere [6].

BIBLIOGRAPHY

[1] Tsay, R. S., Analysis of Financial Time Series, 2002, Wiley.

[2] Dacorogna, M. M., Gencay, R., Muller, U., Olsen, R. B. and Pictet O. V., ”An Introduction to High Frequency Finance” Academic Press, 2001.

[3] Vietch, D. and Abry, P., ”A Wavelet-Based Joint Estimator of the Parameters of Long-Range Dependence”, IEEE Transactions on Information Theory, 45:878-897, April 1999.

[4] Willinger, W., Taqqu, M. S., Teverovsky, V., ”Stock market prices and long-range dependence”,Finance and Stochastics, 3:1-13, 1999.

[5] Çaglar, M., ”Stock Price Processes with Infinite Source Poisson Agents”, Mathematics of Operations Research, accepted and to be appeared, 2010.

[6] Çaglar, M., Bahtiyar N., Altintas I., ”Parameter Estimation of an Agent Based Stock Price Model”, XXXXX, submitted, 2012.

[7] Lau, M. S., Erramilli, A., Wang, J.L., Willinger, W.,”Self-similar Traffic Parameter Estimation: A Semi-parametric Periodogram-Based Algorithm”, Global Telecommunications Conference IEEE, 3:2225-2231, 1995.

[8] http://www.cubinlab.ee.unimelb.edu.au/ darryl/

[9] Alfarano, S., Lux, T., Wagner, F., ”Time variation of higher moments in a financial market with heterogeneous agents: An analytical approach”, Journal of Economic Dynamics and Control, 32: 101-136, 2008.

[10] Ghoulmie, F., Cont, R. and Nadal, J.-P., "Heterogeneity and feedback in an agent-based market model", Journal of Physics: Condensed Matter, 17: 1259-1268, 2005.

[11] Farmer, J. D., Joshi, S., "The Price dynamics of common trading strategies", Journal of Economic Behavior and Organization, 49:149-171, 2002.

[12] Keskin, M., Deviren B., Kocakaplan Y., "Topology of the correlation networks among major currencies using hierarchical structure methods", Physica A, 390:719-730, 2011.

[13] Alptekin N., "Long Memory Analysis of USD/TRL Exchange Rate", International Journal of Human and Social Sciences, 1: 2: 111-116, 2006.

[14] Alfarano, S., Lux, T., Wagner, F., "Time variation of higher moments in a financial market with heterogeneous agents: An analytical approach", Journal of Economic Dynamics and Control 32: 101-136, 2008.

[15] Çaglar, M., "Simulation of Fractional Brownian Motion with Micropulses", Advances in Performance Analysis, 3: 43-69, 2004.

[16] Bayraktar, E., U. Horst, R. Sircar, "A limit Theorem for Financial Markets with Inert Investors", Mathematics of Operations Research, 31: 789-810, 2006.

[17] Geman, H., "Pure Jump Levy Processes for Asset Price Modelling", Journal of Banking and Finance, 26: 1297-1316, 2002.

[18] Hu, Y., B. Oksendal, "Fractional white noise calculus and applications to finance, Infinite Dimensional Anal.", Quant. Probab. Rel. Topics 6: 1-32, 2003.

[19] Akcay, Z., "Fractional Brownian motion in finance from arbitrage point of view". MSc Thesis, August 2007, Koc University.

[20] Kluppelberg, C., C. Kuhn, "Fractional Brownian motion as a weak limit of Poisson shot noise processes-with applications to finance" , Stochastic Processes and their Applications, 113: 333-351, 2004.

[21] Akcay, Z., M. Çaglar, "An Agent Based Stock Price Model", The 32nd Conference on Stochastic Processes and their Applications, August 6-10, 2007, Urbana-Champaign, Illinois.

[22] Lux,T., "Volatility slustering in financial markets: A microsimulation of interacting agents", International Journal of Theoretical and Applied Finance", 675-702, 2000.

[23] Lux, T., "The socio-economic dynamics of speculative markets: Interacting agents, chaos, and the fat tails of return distribution", Journal of Economic Behavior and Organization, 143-165, 1998.

[24] Lux, T. and Marchesi, M., "Scaling and criticality in a stochastic multi-agent model of financial market", Nature, 498-500, 1999.

[25] Cont, R., and Tankov, P., "Calibration of jump-diffusion option pricing models: a robust non-parametric approach", Journal of Computational Finance, Vol. 7, No. 3, 1-49, 2004.

[26] Duncan T. E., Hu Y., Pasik-Duncan B., Stochastic Calculus for Fractional Brownian Motion I. Theory, Preceedings of the 39th IEEE Conference on Decision and Control, 2000.

[27] Samoradnitsky G., Taqqu M., "Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance (Stochastic Modeling)", Chapman and Hall, 1994.

[28] Keller G., Warrack B., Statistics for Management and Economics, Thomson Brooks Cole, 2003.

[29] Chonavel T., Statistical Signal Processing: Modelling and Estimation, Springer, 2002.

[30] Metivier M., Semi-martingales: A Course on Stochastic Processes, Cloth DM, 1982.

[31] Cinlar E., Lecture Notes on Stochastic Differential Equations, Princeton University, 1993.

VITA

Nihal Bahtiyar was born in Istanbul, Turkey on January 5,1986. She received her B.Sc. degrees in Mathematics Engineering in 2009 and Management Engineering as double major in 2010 from Istanbul Technical University,Istanbul. From August 2009 to August 2011, she worked as a teaching and research assistant in Koc University, Turkey and had studied for the 109T665 project which was sponsored by TUBITAK, since October 2009. She has presented her work about *Parameter Estimation of Stock Price* in the following conferences: 24^{th} Mini EURO Conference on Continuous Optimization and Information-Based Technologies in The Financial Sector, 2010 (Izmir, Turkey), 35^{th} Conference on Stochastic Processes and their Applications, 2011 (Oaxaca, Mexico), International Conference on Mathematical Finance and Economics (Istanbul, Turkey).