AN ANALYSIS OF MOMENTUM AND MEAN REVERSION EFFECTS ON EQUITY INDICES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF APPLIED MATHEMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ARMAĞAN ÖZBİLGE

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
FINANCIAL MATHEMATICS

JUNE 2015

Approval of the thesis:

AN ANALYSIS OF MOMENTUM AND MEAN REVERSION EFFECTS ON EQUITY INDICES

submitted by **ARMAĞAN ÖZBİLGE** in partial fulfillment of the requirements for the degree of **Master of Science in Department of Financial Mathematics, Middle East Technical University** by,

Prof. Dr. Bülent Karasözen
Director, Graduate School of **Applied Mathematics** _____

Assoc. Prof. Dr. Ali Devin Sezer
Head of Department, **Financial Mathematics** _____

Assoc. Prof. Dr. Yeliz Yolcu Okur
Supervisor, **Financial Mathematics, METU** _____

Kamil Korhan Nazlıben
Co-supervisor, **School of International Studies, Avansas University of Applied Sciences** _____

Examining Committee Members:

Assist. Prof. Dr. Seza Danışoğlu
Business Administration, METU _____

Assoc. Prof. Dr. Yeliz Yolcu Okur
Financial Mathematics, METU _____

Prof. Dr. Aydın Ulucan
Business Administration, HACETTEPE UNIVERSITY _____

Date: _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: ARMAĞAN ÖZBİLGE

Signature :

ABSTRACT

AN ANALYSIS OF MOMENTUM AND MEAN REVERSION EFFECTS ON EQUITY INDICES

Özbilge, Armağan

M.S., Department of Financial Mathematics

Supervisor : Assoc. Prof. Dr. Yeliz Yolcu Okur

Co-Supervisor : Kamil Korhan Nazlıben

June 2015, 69 pages

Momentum and mean-reversion effects have become very popular in finance literature for the last two decades since their presence can generate abnormal profit patterns by applying either relative strength or contrarian trading strategy accordingly. Even though there are some common factor explanations for return reversals, they might not provide the full picture for return persistence. In our theoretical framework, we analyse some of the well-known discrete time momentum studies including the initial one and try to explain why a novel approach is needed.

Henceforth, in this work, we focus on a continuous time model that aims to capture both momentum and contrarian effects in stock returns. Our model nests the standard stochastic framework proposed by Koijen, Rodriguez and Sbuelz [24]. In our empirical analysis, we examine the term structure of return continuation (momentum) and mean reversion in Turkish stock market (BIST-100) using historical observation from 2004 to 2014. Further, the results of BIST-index are compared to both previous studies on it and other benchmark results in the literature in which US CRSP-index returns are investigated. Accordingly, we observe that, unlike US, Turkish stock market contains mean reversion, but not momentum effect, as Bildik and Gülay [5] states by analysing dozens of possible portfolio strategies. Thus, rather than constructing specified portfolios (decile, industry, size, etc.), presence of momentum and mean reversion effects in a stock market might be anticipated accurately by only analysing equity index of that market.

Keywords : stock momentum, mean reversion, contrarian effect

ÖZ

MOMENTUM VE ORTALAMAYA DÖNÜŞ ETKİLERİNE BORSA ENDEKSLERİ ÜZERİNDEN BİR İNCELEME

Özbilge, Armağan

Yüksek Lisans, Finansal Matematik Bölümü

Tez Yöneticisi : Doç. Dr. Yeliz Yolcu Okur

Ortak Tez Yöneticisi : Kamil Korhan Nazlıben

Haziran 2015, 69 sayfa

Momentum ve ortalamaya dönüş etkilerinin varlığı, göreceli güç veya karşıt alım satım stratejilerinin, duruma göre, uygulanmasıyla anormal getiri patikaları meydana getirdiği için, bu iki etki, son yirmi yılda finans literatüründe oldukça popüler hale gelmiştir. Getirilerdeki ters yönlü harekete bazı ortak faktör açıklamaları getirilse de bunlar, getirilerdeki devamlılığı açıklarken tüm tabloyu yansıtmayabiliyor. Teorik çerçevemiz içinde, ilki dahil, bazı bilinen ayrık zaman momentum çalışmalarını analiz ederek neden alışılmışın dışında bir yaklaşımın gerektiğini açıklamaya çalışıyoruz.

Bundan dolayı, bu çalışmada, hem momentum hem de ortalamaya dönüş etkilerini yakalamayı hedefleyen bir sürekli zaman modeli üzerine yoğunlaşıyoruz. Modelimiz, Koijen, Rodriguez and Sbuelz [24] tarafından önerilen standart stokastik çerçevenin üzerine kurulmaktadır. Yürüttüğmüz ampirik analizde, 2004-2014 yılları arası, getiri devamlılığının (momentum) ve ortalamaya dönüşün Türkiye borsasındaki (BİST-100) dönem yapısını inceledik. Ayrıca, BİST-100 endeksinden elde edilen sonuçlar, hem kendi literatürüyle, hem de Amerikan CRSP-endeks getirileri üzerine yapılan geçmiş literatür çalışmalarıyla kıyaslanıyor. Buna göre, Bildik and Gülay [5]'ın da düzinelerce portföyü analiz ettikten sonra belirttiği gibi, Türkiye borsası, Amerikan borsasının aksine, ortalamaya dönüş etkisi içermekte, ancak momentum etkisi içermemektedir. Bu yüzden, belki de bir borsadaki momentum ve ortalamaya dönüş etkileri, o borsadaki hisse senetlerinin büyüklüğüne, endüstrisine, vb. özelliklerine göre belirli portföyler oluşturmadan, o piyasanın borsa endeksini analiz ederek doğru bir şekilde öngörülebilir.

Anahtar Kelimeler : hisse senedi momentumu, ortalamaya dönüş, karşıt etki

To My Parents and
My Dear Nevcihan

ACKNOWLEDGMENTS

I would like to express my very great appreciation to my thesis supervisor Assoc. Prof. Dr. Yeliz Yolcu Okur and my thesis co-supervisor Kamil Korhan Nazlıben for their patient guidance, enthusiastic encouragement and valuable advices during the development and preparation of this thesis. Their willingness to give their time and to share their experiences has brightened my path. I would like to thank Kamil Korhan Nazlıben for introducing me this exciting topic as well.

I am also grateful to Assoc. Prof. Dr. Yeliz Yolcu Okur for her constant encouragements in the field of Financial Mathematics since my enrolment to IAM. She kept trusting and motivating me through her lectures and this thesis process in my tough times. Beside, her friendship and support even in my personal issues boost my commitment to this field and the institute.

I declare my thanks to the committee members Prof. Dr. Aydın Ulucan and Assist. Prof. Dr. Seza Danışoğlu for their precious attendance, comments and advices. Prof. Dr. Aydın Ulucan has given form to my academic studies by suggesting me this master program in METU and also inspired my enthusiasm since my undergrad years. Assist. Prof. Dr. Seza Danışoğlu gave great suggestions for this thesis and my further academic research.

Furthermore, I also thank all the institute members for providing such a friendly working environment. I received support from all academic and administrative staff one by one. However, I would like to have special thanks to Assoc. Prof. Dr. Ömür Uğur who kindly answered all my questions even if most of them were irrelevant to his lectures, Prof. Dr. Gerhard-Wilhelm Weber for his kindness and support, and, of course, Prof. Dr. Azize Hayfavi for her inspirational character, knowledge, and support in my thesis defence.

Many thanks to Soner Orhan, Selin Tekten, Meral Şimşek, and all my friends for their invaluable friendship and motivation. Their efforts broke my concerns and desperation and assisted me to move on. There is also an exceptional friend, Kaan Tanık, who helped me to initialize and maintain my graduate studies.

I would like to express my pleasure to having such a great family who always back me up in all my decisions, and show me great patience and trust. I am very appreciated for their endless, priceless, and unconditional love.

Finally, my lifetime fellow Nevcihan Karaosman who makes possible the existence of this thesis and all the nice things in my life deserves a great thank. Without her trust, patience, sacrifices, and smile, I would not have been anywhere close to here.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABBRV	Abbreviation
$\mathbb{R}$	Real Numbers
$\mathbb{E}$	Expectation operator
$\mathbb{V}$	Variance operator
$\mathbb{C}ov$	Covariance operator
EW	Equally Weighted Market Index
VW	Value Weighted Market Index
VAR	Vector Auto-regressive
MLE	Maximum Likelihood Estimator
SDE	Stochastic Differential Equation
B-S	Black-Scholes
OU	Ornstein Uhlenbeck
E-M	Euler-Maruyama
MSE	Mean Square Error
KRS	Koijen, Rodrígez, and Sbuelz
JT	Jegadeesh and Titman

CHAPTER 1

INTRODUCTION

Stock momentum has been examined extensively in the asset pricing literature. Simply, existence of stock momentum implies that stock returns exhibit a persistent characteristics in which velocity and direction of the return series are significantly preserved. In other words, stocks that performed well (poorly) in the past are expected to preserve their good (bad) performances in the future. On the other hand, mean-reversion behaviour of stock returns states that, stock returns fluctuate around their long run mean. For example, once a financial shock occurs, return of a stock eventually turns back to its long term fundamental level which could be either its average past performance or a specified portfolio mean like industry, size, equity index, etc.

The existence of return continuation and mean-reversion can generate abnormal profit patterns by applying either relative strength or contrarian trading strategy[1] accordingly. Indeed, these two effects may vary across the portfolios, countries, and investment horizons. For example, it is documented that, US market realizes return continuation on short term [23], 3 to 12 months, and mean-reversion on long term [7], 3 to 5 years, investment horizons whereas Turkish stock market preserves only short term mean-reversion [5]. Because of this discrepancy, associating their impact with a common factor is a challenge, in particular for the cases where return patterns exhibit momentum.

Analysing persistence and reversal of asset returns are theoretically crucial for recognizing market inefficiency. Because, a rational investor can establish an investment plan in which she aims to outperform the market by identifying these effects in return characteristics. Though Fama-French manages to explain long term return reversals under classical risk-return scheme [20](higher return can only be obtained in exchange for extra risk), momentum still violates the weak-form efficiency [16] by allowing a systematic trading system to beat the market without bearing additional risk.

However, it might be too costly to find out which of these two states dominate the market on which investment horizon. Conducting dozens of portfolio scenarios with corresponding significance tests is a non-trivial issue and it can even produce nothing significant[2]. In fact, in practice, momentum and relative strength indices are used as a part of the technical analysis [34, 29] for assessing the attainability of buy and sell

[1] Buying past winners and selling past losers or vice versa.

[2] See, for example, Japanese capital market[21]

signals for each stock, particularly in short run. But, computations are not sophisticated and not informative on their own. Beside, they are usually in consensus with simple arithmetic moving averages. So, momentum, in this sense, is very similar but not congruent to its theoretical meaning.

In this work, we focus on a continuous time model, proposed by Koijen, Rodríguez and Sbuelz [24], that aims to capture both momentum and contrarian effects in stock returns. In our empirical analysis, we estimate model parameters for Turkish stock market (BIST-100) equity index between years 2004-2014. Our theoretical framework also includes the detailed analysis of some of the well-known momentum studies in discrete time literature to both see the initial quantification of this phenomena and provide comparison. Therefore, rather than trying to explain its source by some common factors, our continuous time model takes momentum as a factor (because it is shown to exist [23, 25]) of return. Even though, this method is much easier than examining various portfolios, it might not be as informative as classical scheme since it only considers the time series predictability.

Stochastic parameter estimations are executed by the help of Euler-Maruyama(E-M) discretization and maximum likelihood function. All computations are performed on *MATLAB* and codes are readily available on appendices. We try to provide explanations, examples and connections at each stage to justify and even clarify what has been done and why. For example, E-M scheme is chosen for discretization, since it is shown to provide fair approximations with tolerable errors for lower order SDEs.

The usage of BIST-100 index data is not a coincidence. This is the first study that aims to examine momentum and mean-reversion effects in BIST-100. There is not any index level study about these issues for Turkish stock market either. So, providing evidences from an emerging market might develop some explanations for previous findings. Further, there is a stock level past study [5] for BIST in discrete time, which is also case for the US capital market. Therefore, relation between domestic studies can be verified by comparing these two distinct markets.

According to our empirical analysis, we observe negligibly small momentum and dominant mean reversion effects in BIST-100 index. Bildik and Gülay [5] reports same result for individual stocks by examining dozens of portfolios and even manages to exploit this opportunity by applying contrarian trading strategy in Turkish stock market. Same issue is also valid for US market, but in opposite direction. Koijen, Rodríguez and Sbuelz [24] finds stronger evidence for momentum in Equally Weighted(EW) index in which relative strength strategy is shown to be profitable. This consistency might indicate a possible relationship between index and corresponding stock level return patterns by means of momentum and mean-reversion. Of course, this claim needs so many additional examinations to be proven.

This study consists of five chapters. The first chapter introduces comprehensive knowledge about momentum and mean-reversion effects and a brief literature review. In the second chapter, mathematical intuition behind momentum and some of the seminal discrete time approaches are discussed. Thereafter, reasons behind passing to continuous time setting are explained and the structure of our model is investigated. The Chapter three demonstrates how the model is calibrated to stochastic parameter estimation by

means of Euler approximation and likelihood function. The fourth chapter implements the methods cited in chapter three to simulated and actual(BIST-100) data to figure out the role of variables inside our model and discusses the evidences of Turkish stock index. Finally, the last chapter concludes this master thesis with comments and further study suggestions.

1.1 A Brief Literature Review

Beside of some exceptional equity indices and international portfolio analysis, most of the past studies rely on individual stocks and/or specified portfolios, particularly in discrete time. De Bondt and Thaler [7] shows that US stock prices contain mean reversion effect over 3 to 5 years investment horizon according to their past 3 to 5 years performance. Lo and MacKinlay [26, 27] states that there are some patterns that stock prices follow and introduce their celebrated profit decomposition to express source of abnormal returns. Jegadeesh and Titman [23] concentrates on 3 to 12 months(relatively short term) investment horizon with the past 3 to 12 months return performances of financial securities and notice that, stocks that generate abnormal returns in the past continue to do so. They attribute this "momentum effect" to firm-specific factors. Moskowitz and Grinblatt [28] points out industry portfolios as a source of return continuation whereas Lewellen [25] claims excess cross covariation between stocks and not idiosyncratic factors to explain it, since his study analyses well-diversified portfolios not only by industry; but also by size and book-to-market value. Chen and Hong [12] asserts that Lewellen's findings are debated. They also show that assumption quantification might be tricky and biased. We will also go in deep on this issues by next chapter.

For Turkish stock market, Bildik and Gülay observes short term mean-reversion effect on BIST-100 by applying decile portfolios of Jegadeesh and Titman. They also examine the weak form efficiency of Turkish stock market and emphasize the unique characteristics of it. Accordingly, BIST is a capital market with record high inflation, high volatility, and low correlated returns which associate its return pattern behaviour with market specific factors.

While papers mentioned above generally focus on portfolio construction to exploit momentum and/or mean-reversion effects, there are some other significant studies aim to explain them. One of the most popular approaches is to explain these effects by imperfect investor behaviours by addressing either over- or under-reaction(See, for example, Barberis, Shleifer, and Vishny[4], Hong and Stein [22] and Chui, Titman, and Wei [14]).

Fama and French [19, 20] suggests three factor model which manages to catch big part of the mean-reversion effect, but fails to explain momentum. Carhart [10] adds momentum effect to three factor model and Fama and French [21] shows presence of momentum in international stock markets by using that four factor model. However, they again stress individual stocks and conduct numerous analysis to provide coefficients of factors and even fail to explain extreme cases. As a final discrete time study Chan, Hameed, and Tong [11] reports existence of momentum in international equity

markets by applying relative strength strategy to equity indices. The last two works are critical as they provide evidences for the presence of mean-reversion and momentum effects in international markets at index level as well as individual stock level.

Unlike discrete time models, there is not much study to capture momentum in continuous time setting. Wachter [33] investigates mean reversion anomaly in continuous time setting along with optimal portfolio selections. Rodríguez and Sbuelz [30] follows similar pattern for stock momentum. Both studies establish a stochastic framework to capture mean-reversion/momentum effect. Their aim is to identify optimal portfolio choices of an investor who considers mean-reversion/momentum effect.

Finally, Koijen, Rodríguez, and Sbuelz [24] suggests a novel continuous time model to feature both momentum and mean-reversion behaviour of equity index returns. They broadly introduce their stochastic framework and perform similar parameter estimations with us. They also examine optimal asset allocation for a portfolio, consists of an equity index and a riskless bond, by means of dynamic programming. Results are both compared with the case in which mean-reversion is the only state and some of the benchmark journals(Wachter [33] and Rodríguez and Sbuelz [30]).

CHAPTER 2

SOME OF THE MOMENTUM MODELS IN THE LITERATURE

This chapter consists analyses of early studies that were proposed to capture stock momentum. Dividing literature into two parts: 1) Discrete Time Models and 2) Continuous Time Models will be appropriate to clarify the intuition behind our study.

2.1 Discrete Time Models

There are numerous studies examining momentum in discrete time. While some of them examine it by specified portfolios like size, book-to-market value, industry, etc.[23, 28, 25], others advocate behavioural approaches[4, 14]. In addition to these, Fama and French[20, 21] argues the multi factor explanations to capture persistence and reversal in returns. Under the scope of this thesis, we will analyse the some studies from the first group which also includes the initial study on momentum. Because, like our model, the works we will identify also emphasize the presence and profitability of momentum.

2.1.1 Model Analysis of Jegadeesh and Titman [23]

De Bondt and Thaler [7] shows that stock prices contain mean reversion effect over 3 to 5 years investment horizon according to their past 3 to 5 years performances. Then, Jegadeesh and Titman [23] concentrates on 3 to 12 months (relatively short term) investment horizon with the past 3 to 12 months return performances of financial securities. Although, their base to explain momentum (idiosyncratic-risk) is debated, their study is the first one which claims the existence of stock returns momentum with evidences.

They started with ranking security returns in ascending order. Then, first ten securities named as losers portfolio while last ten named as winners. This procedure applied to stocks for their 3, 6, 9, and 12 months past performance. Portfolios are formed for the next 3, 6, 9, and 12 months for all past performances. In order to avoid bid-ask spread and increase the power of the test, parallel portfolios are processed just after a week of their formation.

Because Jegadeesh and Titman run the first serious study about stock momentum, their practice mostly emphasizes the return auto-covariances to back up their empirical findings. The theoretical scheme of them starts by a simple one factor model:

$$\begin{aligned} r_{i,t} &= \mu_i + \beta_i x_t + \varepsilon_{i,t}, \\ \mathbb{E}[x_t] &= 0, \\ \mathbb{E}[\varepsilon_{i,t}] &= 0, \\ \mathbb{C}ov[\varepsilon_{i,t}, x_t] &= 0, \ \forall \ i \\ \mathbb{C}ov[\varepsilon_{i,t}, \varepsilon_{j,t-1}] &= 0, \ \forall \ i \neq j \end{aligned} \tag{2.1}$$

where, $r_{i,t}$ is the observed return of security i at time t, μ_i is unconditional expected return of security i. Yet, x_t is unconditional and unexpected return on a factor-mimicking portfolio with the sensitivity β_i, and $\varepsilon_{i,t}$ is the firm specific error of security i. Under these assumptions, in accordance with their empirical analysis JT manages to quantify the momentum effect as follows:

$$\mathbb{E}[r_{i,t} - r_{m,t} | r_{i,t-1} - r_{m,t-1} > 0] > 0,$$

and

$$\mathbb{E}[r_{i,t} - r_{m,t} | r_{i,t-1} - r_{m,t-1} < 0] < 0,$$

where $r_{m,t}$ is equally weighted market return at time t. The above expectations are vital to understand the mathematical intuition behind momentum. Therefore, by combining these two expectations one can obtain the following expression

$$\mathbb{E}[(r_{i,t} - r_{m,t})(r_{i,t-1} - r_{m,t-1})] > 0. \tag{2.2}$$

In order to exploit the return continuation opportunity, JT provides such weights that equation (2.2) becomes the expected excess return of stock i with respect to the market.

$$w_{i,t} = \frac{1}{N}(r_{i,t-1} - r_{m,t-1}). \tag{2.3}$$

Notice that for an Equally Weighted index examination this weights setting leads to a zero cost portfolio since $\sum_i (X_i - \mathbb{E}[X]) = 0$. Therefore, profit equation could be written in the form

$$\pi_t = \sum_i w_{i,t} r_{i,t} = \frac{1}{N} \sum_i (r_{i,t-1} - r_{m,t-1}) r_{i,t},$$

and if we rather multiply weights expression with excess returns (could be either positive or negative in realization) $(r_{i,t} - r_{m,t})$ we obtain the equation (2.2), which is claimed to be positive indeed. Therefore, to convert covariance to something more

meaningful:

$$\begin{aligned}
\mathbb{E}\big[(r_{i,t}-r_{m,t})(r_{i,t-1}-r_{m,t-1})\big] &= \mathbb{E}\big[(\mu_i+\beta_i x_t+\varepsilon_{i,t}-\mu_m-\beta_m x_t-\varepsilon_{m,t}) \\
&\quad (\mu_i+\beta_i x_{t-1}+\varepsilon_{i,t-1}-\mu_m-\beta_m x_{t-1} \\
&\quad -\varepsilon_{m,t-1})\big], \\
&= \mathbb{E}[\mu_i^2-\mu_i\mu_m-\mu_m\mu_i+\mu_m^2] \\
&\quad +\mathbb{E}\big[x_t x_{t-1}(\beta_i^2-\beta_i\beta_m-\beta_m\beta_i+\beta_m^2)\big] \\
&\quad +\mathbb{E}[\varepsilon_{i,t}\varepsilon_{i,t-1}-\varepsilon_{i,t}\varepsilon_{m,t-1}-\varepsilon_{m,t}\varepsilon_{i,t-1} \\
&\quad +\varepsilon_{m,t}\varepsilon_{m,t-1}\big], \\
&= \mathbb{E}\big[(\mu_i-\mu_m)^2\big]+\mathbb{E}\big[x_t x_{t-1}(\beta_i-\beta_m)^2\big] \\
&\quad +\mathbb{E}\big[(\varepsilon_{i,t}-\varepsilon_{m,t})(\varepsilon_{i,t-1}-\varepsilon_{m,t-1})\big], \\
\Rightarrow \mathbb{E}[(r_{i,t}-r_{m,t})(r_{i,t-1}-r_{m,t-1})] &= \sigma_\mu^2+\sigma_\beta^2\mathbb{C}ov(x_t,x_{t-1})+ \\
&\quad \overline{\mathbb{C}ov}(\varepsilon_{i,t},\varepsilon_{i,t-1}). \qquad (2.4)
\end{aligned}$$

Under this setting, σ_μ^2 is cross-sectional variance of expected returns, and σ_β^2 is variance of factor sensitivities. The last term stands for average serial covariance.

This last expression decomposes three potential sources of excess momentum profits. The first term of it can be explained like a stock with high unconditional mean at one period is expected to carry its abnormal performance to next period. The second term indicates that, for instance, when factor mimicking portfolio generates positive returns with positive autocorrelation, choosing the stocks with high σ_β would be profitable. The last expression might be called as the idiosyncratic component of the stocks.

Jegadeesh and Titman also specify when momentum indicates market inefficiency. Since, the fist two terms of equation (2.4) generate profits in exchange of extra risk, they wouldn't be a signal of an inefficient market. On the other hand, idiosyncratic component would, since it could be diversified. In other words, only the idiosyncratic component is a part of avoidable unsystematic risk factors and if momentum profits are due to this term, their presence violates the traditional risk and return scheme (higher return can only be achieved in exchange of greater risk) and so indicates market inefficiency.

They also suggested a model to capture lead-lag effect. Their model somehow examines the behavioural concepts which are later employed to explain momentum in the literature[1].

$$r_{i,t}=\mu_i+\beta_{1,i}x_t+\beta_{2,i}x_{t-1}+\varepsilon_{i,t}, \qquad (2.5)$$

where $\beta_{1,i}$ and $\beta_{2,i}$ are sensitivities to lagged factor portfolio returns. When $\beta_{2,i}>0$, return of stock i reacts to a news and if $\beta_{2,i}<0$ as well, overreaction exists it is corrected in the following period. For this model, $\varepsilon_{i,t}$ has zero autocorrelation and

[1] See, for example, [4, 14].

factors are also serially *uncorrelated*. First, notice that

$$\begin{aligned}
\mathbb{C}ov(r_{m,t}, r_{m,t-1}) &= \mathbb{E}\big[(\mu_m + \beta_{1,m}x_t + \beta_{2,m}x_{t-1}\varepsilon_{m,t} - \mu_m) \\
&\quad (\mu_m + \beta_{1,m}x_{t-1} + \beta_{2,m}x_{t-2}\varepsilon_{m,t-1} - \mu_m)\big], \\
&= \beta_{1,m}\beta_{2,m}\mathbb{E}[x_{t-1}^2], \\
\Rightarrow \mathbb{C}ov(r_{m,t}, r_{m,t-1}) &= \beta_{1,m}\beta_{2,m}\sigma_x^2, \qquad (2.6)
\end{aligned}$$

now, set

$$\delta \equiv \frac{1}{N}\sum_{i=1}^{N}(\beta_{1,i} - \beta_{1,m})(\beta_{2,i} - \beta_{2,m})$$

Therefore,

$$\begin{aligned}
\mathbb{E}[(r_{i,t} - r_{m,t})(r_{i,t-1} - r_{m,t-1})] &= \mathbb{E}[\mu_i^2 - \mu_i\mu_m - \mu_m\mu_i + \mu_m^2] \\
&\quad +\mathbb{E}\big[x_{t-1}^2(\beta_{1,i}\beta_{2,i} - \beta_{1,m}\beta_{2,i} - \beta_{1,i}\beta_{2,m} \\
&\quad +\beta_{1,m}\beta_{2,m})\big], \\
&= \mathbb{E}\big[x_{t-1}^2(\beta_{1,i} - \beta_{1,m})(\beta_{2,i} - \beta_{2,m})\big] \\
&\quad +\mathbb{E}\big[(\mu_i - \mu_m)^2\big], \\
\Rightarrow \mathbb{E}\big[(r_{i,t} - r_{m,t})(r_{i,t-1} - r_{m,t-1})\big] &= \sigma_\mu^2 + \delta\sigma_x^2. \qquad (2.7)
\end{aligned}$$

Thus, when $\delta < 0$ lead-lag relation has a negative affect on momentum profits and vice versa. Further, if $\beta_2 > 0$ ($\beta_2 < 0$) returns are positively (negatively) correlated. JT conclude this model with their analysis. According to JT, retuns are negatively correlated and $\delta < 0$. So, they attributed momentum profits to idiosyncratic component. However, for the upcoming years, many well diversified portfolios are shown to contain momentum as well.

To sum up, though JT lacks while explaining the source of return continuation, their study is very important as they introduce the momentum concept in the literature. The persistence characteristic of return series is examined by many other researchers and it is shown to exist in various portfolios. Indeed, even latter works face with similar shortcomings while explaining the source of return continuation.

2.1.2 Model Analysis of Lewellen[25]

Jegadeesh and Titman[23] use decile portfolios and initializes the momentum concept. Moskowitz and Grinblatt[28] find similar return patterns by constructing industry portfolios. Lewellen[25] extends these results to size, book-to-market(B/M) value and double sorted size-B/M portfolios. Presence of return continuation into such well diversified portfolios shows the inadequacy of previous explanation(firm specific factor) and robustness of momentum. Unlike JT, Lewellen considers the cross serial correlation between stocks to explain momentum.

Lewellen also follows similar zero cost portfolio construction with JT. This time, he aims to provide explanatory models to identify the origin of momentum under distinctly sorted zeros cost portfolios. It is possible to write the equation (2.3) in a more

specific way

$$w_{i,t} = \frac{1}{N}(r^k_{i,t-1} - r^k_{m,t-1}), \tag{2.8}$$

where $r^k_{i,t-1}$ stands for the k month return of the stock which ends by month $t-1$. Similarly, $r^k_{m,t-1}$ indicates the equally weighted index return. It is, again, obvious that:

$$\sum_{i=1}^{N} w_{i,t} = \frac{1}{N}\sum_{i=1}^{N}(r^k_{i,t-1} - r^k_{m,t-1}) = 0. \tag{2.9}$$

since $\sum_i (X_i - \mathbb{E}[X]) = 0$. Then, by setting $\mu \equiv \mathbb{E}[r_t]$ and auto-covariance matrix $\Omega \equiv \mathbb{E}[(r_{t-1} - \mu)(r_t - \mu)']$, one period return will be:

$$\pi_t = \sum_i w_{i,t} r_{i,t} = \frac{1}{N}\sum_i (r_{i,t} - r_{m,t}) r_{i,t}. \tag{2.10}$$

From here, expected return becomes:

$$\begin{aligned} \mathbb{E}[\pi_t] &= \frac{1}{N}\mathbb{E}\left[\sum_i r_{i,t-1} r_{i,t}\right] - \frac{1}{N}\mathbb{E}\left[r_{m,t-1}\sum_i r_{i,t}\right], & (2.11)\\ &= \frac{1}{N}\sum_i (\rho_i + \mu_i^2) - (\rho_m + \mu_m^2). & (2.12) \end{aligned}$$

In order to be able to obtain equation (2.12), first part of the (2.11) is calculated like that:

$$\begin{aligned} (r_{i,t-1} - \mu_i)(r_{i,t} - \mu_i) &= r_{i,t-1} r_{i,t} + \mu_i^2 - \mu_i r_{i,t-1} - \mu_i r_{i,t},\\ \mathbb{E}[r_{i,t-1} r_{i,t}] &= \mu_i^2 + \mathbb{C}ov(r_{i,t-1}, r_{i,t}),\\ \frac{1}{N}\sum_i \mathbb{E}[r_{i,t-1} r_{i,t}] &= \frac{1}{N}\sum_i (\mu_i^2 + \mathbb{C}ov(r_{i,t-1}, r_{i,t})). \end{aligned} \tag{2.13}$$

which yields the first part of the equation (2.12) by setting $\mathbb{C}ov(r_{i,t-1}, r_{i,t}) = \rho_i$. Likewise, second part of the equation (2.11) becomes:

$$\mathbb{E}\left[r_{m,t-1}\frac{1}{N}\sum_i r_{i,t}\right] = \rho_m + \mu_m^2, \tag{2.14}$$

as $\frac{1}{N}\sum_i^N r_{i,t} = r_{m,t}$. Equations (2.13) and (2.14) together bring the expression (2.12).

One can also write momentum profits under Lo and MacKinlay [27] decomposition:

$$\begin{aligned} \mathbb{E}[\pi_t] &= \frac{1}{N} tr(\Omega) - \frac{1}{N^2}\mathbb{1}'\Omega\mathbb{1} + \sigma_\mu^2,\\ &= \frac{N-1}{N^2} tr(\Omega) - \frac{1}{N^2}[\mathbb{1}'\Omega\mathbb{1} - tr(\Omega)] + \sigma_\mu^2, \end{aligned} \tag{2.15}$$

where $tr(.)$ is trace operator and $\mathbb{1}$ is vector of ones. Then, by separating diagonal and off-diagonal of the auto-covariance matrix Ω, equation (2.15) leads to three sources of momentum. First term stands for positive auto-correlation of returns, high future return of a stock triggered by its own good past performance. Second term makes positive contribution to momentum profit only if stocks are negatively cross-correlated. In other words, today's high return for a stock predicts low future returns for other stocks. Third component claims that, momentum profit simply arises from investing stocks with higher unconditional mean than average.

Equations (2.12) and (2.15) shows the momentum profits under the zero-cost portfolio setting. These equations are very informative as they are constructed in the absence of an asset pricing model(just by considering the investment plan on EW index). Now, for any given model we can write profit equations under this decompositions to see which assumption attributes profits to which of those three sources.

Now consider a random walk model with log prices p_t:

$$\begin{aligned} p_t &= q_t + \varepsilon_t, && (2.16)\\ q_t &= \mu + q_{t-1} + \eta_t. && (2.17) \end{aligned}$$

which claims that, prices basically consist of a random walk, q_t (present value of dividends discounted at constant rate), and stationary component ε_t with zero mean. $\eta_t \backsim wn(0, \Sigma)$ is a white noise. Log returns can be calculated as follows

$$\begin{aligned} p_t &= \mu + q_{t-1} + \eta_t + \varepsilon_t, \\ p_t - p_{t-1} &= \mu + q_{t-1} + \eta_t + \varepsilon_t - q_{t-1} - \varepsilon_{t-1}, \\ r_t &= \mu + \eta_t + \Delta\varepsilon_t. && (2.18) \end{aligned}$$

He constructs his under and overreaction models by just changing the the stationary component ε_t accordingly. For underreaction he sets:

$$\varepsilon_t = -\rho\eta_t - \rho^2\eta_{t-1} - \rho^3\eta_{t-2} - ... \tag{2.19}$$

where $0 < \rho < 1$ and η is dividend news. In this setting:

$$\begin{aligned} \varepsilon_{t-1} &= -\rho\eta_{t-1} - \rho^2\eta_{t-2} - \rho^3\eta_{t-3} - ... \\ \rho\varepsilon_{t-1} &= -\rho^2\eta_{t-1} - \rho^3\eta_{t-2} - \rho^4\eta_{t-3} - ... \\ \varepsilon_t &= \rho\varepsilon_{t-1} - \rho\eta_t. && (2.20) \end{aligned}$$

Now plug equation (2.20) into log returns r_t (2.18):

$$\begin{aligned} r_t &= \mu + \eta_t + \rho\varepsilon_{t-1} - \rho\eta_t - \varepsilon_{t-1}, \\ r_t &= \mu + \eta_t(1-\rho) + \varepsilon_{t-1}(\rho - 1). && (2.21) \end{aligned}$$

So, for auto-covariance by using above equation (2.21)

$$\begin{aligned} \mathbb{E}[(r_t - \mu)(r_{t-1} - \mu)] &= \mathbb{E}[r_t r_{t-1}] - \mu^2 \\ &= (1-\rho)^2\mathbb{E}[\underbrace{(\eta_t - \varepsilon_{t-1})}_{(*)}\underbrace{(\eta_{t-1} - \varepsilon_{t-2})}_{(**)}], && (2.22) \end{aligned}$$

$$
\begin{aligned}
(*) &= \eta_t\rho^0 + \eta_{t-1}\rho^1 + \eta_{t-2}\rho^2 + ... \\
&= \sum_{j=0}^{\infty} \rho^j \eta_{t-j}. \qquad (2.23)
\end{aligned}
$$

$$
\begin{aligned}
(**) &= \eta_{t-1}\rho^0 + \eta_{t-2}\rho^1 + \eta_{t-3}\rho^2 + ... \\
&= \sum_{k=0}^{\infty} \rho^k \eta_{t-1-k}. \qquad (2.24)
\end{aligned}
$$

To simplify, put expressions (2.23) and (2.24) by indices change $j = 1+k$ that enables us to take expectation of equation (2.22) and get auto-covariance:

$$
\begin{aligned}
(1-\rho)^2 \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \rho^k\rho^j \mathbb{E}[\eta_{t-j}\eta_{t-1-k}] &= (1-\rho)^2 \sum_{j=0}^{\infty} \rho^j\rho^{j+1} \underbrace{\mathbb{E}[\eta_{t-j}\eta_{t-j}]}_{\Sigma}, \\
&= (1-\rho)^2\rho \sum_{j=0}^{\infty} (\rho^2)^j \Sigma, \\
&= (1-\rho)^2\rho \frac{1}{1-\rho^2}\Sigma, \\
&= \rho\frac{1-\rho}{1+\rho}\Sigma > 0. \qquad (2.25)
\end{aligned}
$$

This expression is obtained by using geometric series. Moreover, Σ is dividend covariance matrix. Model auto-covariance shows that under-reaction results with positive auto-correlation *under this setting*. Then, Lo and MacKinlay decomposition of momentum profits becomes(by equation (2.15))

$$
\mathbb{E}[\pi_t] = \rho\frac{1-\rho}{1+\rho}\left[\frac{1}{N}tr(\Sigma) - \frac{1}{N^2}\mathbb{1}'\Sigma\mathbb{1}\right] + \sigma_\mu^2. \qquad (2.26)
$$

Because $\sigma_\mu^2 \geq 0$, this profit equation is positive which means under-reaction might be a source of momentum.

On the other hand, for overreaction, Lewellen just sets $cov(\eta_t) = \sigma_\eta^2 I$ (I is an identity matrix) and changes ε_t such that:

$$
\varepsilon_t = B\eta_t + B\rho\eta_{t-1} + B\rho^2\eta_{t-2} + \dots \qquad (2.27)
$$

where $0 < \rho < 1$ and B is an $N \times N$ zero diagonal with positive off-diagonals matrix which necessarily claims that, each asset reacts correctly to news about itself, while it overreacts the news about other firms. So, by equation (2.27)

$$
\begin{aligned}
\varepsilon_{t-1} &= B\eta_{t-1} + B\rho\eta_{t-2} + B\rho^2\eta_{t-3} + \dots \\
\Rightarrow \varepsilon_t &= \rho\varepsilon_{t-1} + B\eta_t \\
\Rightarrow \Delta\varepsilon_t &= \varepsilon_{t-1}(\rho - 1) + B\eta_t \qquad (2.28)
\end{aligned}
$$

And by log-return equation (2.18):

$$\begin{aligned} r_t &= \mu + \eta_t + \Delta\varepsilon_t \\ &= \mu + \eta_t + \varepsilon_{t-1}(\rho - 1) + B\eta_t \\ &= \mu + \eta_t(I + B) + \varepsilon_{t-1}(\rho - 1) \end{aligned}$$

Then, return variance:

$$\begin{aligned} \mathbb{E}[(r_t - \mu)^2] &= \mathbb{E}[(\eta_t(I + B) + \varepsilon_{t-1}(\rho - 1))^2], \\ &= (I + B)(I + B)'\mathbb{E}[\eta_t^2] + (\rho - 1)^2\mathbb{E}(\varepsilon_{t-1}^2) + 2(I + B)(\rho - 1)Cov[\eta_t\varepsilon_{t-1}], \\ &= \sigma_\eta^2 I[I + B' + B + BB'] + (1 - \rho)^2 BB' \sum_{j=0}^{\infty} (\rho^2)^j \mathbb{E}[\eta_{t-j}^2], \end{aligned}$$

since ε_{t-1} doesn't depend on η_t and $\varepsilon_t = B\sum_{j=0}^{\infty}\rho^j\eta_{t-j}$ by equation (2.27). Thus,

$$\begin{aligned} \mathbb{E}[(r_t - \mu)^2] &= \sigma_\eta^2 I\left[I + B' + B + BB' + (1 - \rho)^2 BB'\frac{1}{1 - \rho^2}\right], \\ &= \sigma_\eta^2 I\left[I + B' + B + BB' + BB'(1 - \rho)\frac{1}{1 + \rho}\right], \\ &= \sigma_\eta^2 I\left[I + B' + B + BB'(1 - \rho + 1 + \rho)\frac{1}{1 + \rho}\right], \\ \Rightarrow \mathbb{V}[r_t] &= \sigma_\eta^2 I\left[I + B' + B + BB'\frac{2}{1 + \rho}\right], \end{aligned} \tag{2.29}$$

whose off-diagonals are greater than 0 stating excess covariance. Return auto-covariance can be calculated similarly:

$$\begin{aligned} \mathbb{E}[(r_t - \mu)(r_{t-1} - \mu)] &= \mathbb{E}\Big\{\big[(I + B)\eta_t + (\rho - 1)\varepsilon_{t-1}\big] \times \big[(I + B)\eta_{t-1} + (\rho - 1)\varepsilon_{t-2}\big]\Big\}, \\ &= (I + B)(I + B)'\mathbb{E}[\eta_t\eta_{t-1}] + (I + B)(\rho - 1)\mathbb{E}[\varepsilon_{t-2}\eta_t] \\ &\quad +(I + B)(\rho - 1)\mathbb{E}[\varepsilon_{t-1}\eta_{t-1}] + (\rho - 1)^2\mathbb{E}[\varepsilon_{t-1}\varepsilon_{t-2}], \\ &= 0 + (I + B)(\rho - 1)\mathbb{E}[B\eta_{t-1}^2 + B\rho\eta_{t-2}\eta_{t-1} + \ldots] \\ &\quad +(\rho - 1)^2\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} BB'\rho^j\rho^k\mathbb{E}[\eta_{t-1-j}\eta_{t-2-k}] \quad \textit{now, for } j = k + 1, \\ &= \sigma_\eta^2 I[(B + BB')(\rho - 1)] + (\rho - 1)^2 BB'\sum_{j=0}^{\infty}\rho^j\rho^{j+1}\sigma_\eta^2 I, \\ &= \sigma_\eta^2(\rho - 1)\left[B + BB' + (\rho - 1)BB'\rho\frac{1}{1 - \rho^2}\right], \\ &= \sigma_\eta^2(\rho - 1)\left[B + \frac{BB'\rho + BB' - \rho BB'}{\rho + 1}\right], \\ \Rightarrow \mathbb{C}ov(r_t, r_{t-1}) &= \sigma_\eta^2(\rho - 1)\left[B + \frac{1}{1 + \rho}BB'\right] < 0. \end{aligned} \tag{2.30}$$

Because $\rho < 1$, auto-correlation of returns is negative. Then, by defining $B = b[\mathbb{1}\mathbb{1}' - I]$ for $0 < b < 1$(actually he defines a small and simple stock covariation sensitivity) momentum profits are realized like that:

$$\mathbb{E}[\pi_t] = \frac{1}{N} \times tr\left(\sigma_\eta^2(\rho-1)\left[B + \frac{1}{1+\rho}BB'\right]\right) - \frac{1}{N} \times \mathbb{1}'\left[\sigma_\eta^2(\rho-1)\left[B + \frac{1}{1+\rho}BB'\right]\right]\mathbb{1} + \sigma_\mu^2.$$

We know that, B has zero diagonals and positive off-diagonals which are all b:

$$\begin{aligned}
\mathbb{E}[\pi_t] &= -\frac{1}{N}\sigma_\eta^2(\rho-1) \times \mathbb{1}'\left[b(\mathbb{1}\mathbb{1}' - I) + \frac{1}{1+\rho}b(\mathbb{1}\mathbb{1}' - I)[b(\mathbb{1}\mathbb{1}' - I)]\right]\mathbb{1} + \sigma_\mu^2, \\
&= -\frac{1}{N}N(N-1)b\sigma_\eta^2(\rho-1) - \frac{1}{N}\frac{1}{(1+\rho)}N(N-1)b^2\sigma_\eta^2(\rho-1) + \sigma_\mu^2, \\
&= \sigma_\eta^2(\rho-1)b\frac{(N-1)}{N}\left[\frac{b}{1+\rho} - 1\right] + \sigma_\mu^2 > 0. \qquad (2.31)
\end{aligned}$$

Notice that, this profit expression is also constructed from the equation (2.15) and it is positive as long as $0 < b < 1$ holds. In other words, if overreaction effect exceeds one ($b > 1$) momentum profits will be negative. Therefore, overreaction might be another source of momentum profits if it is not too large.

We examine two possible origins of momentum so far(under, over-reaction). Lewellen also points out that, aggregate risk premium variations across the time might cause excess covariance and so, momentum. Unlike, under-,over-reaction assumptions, stock prices oscillate around random walk in the absence of any behavioural irrationality. For instance, capital asset pricing model (CAPM) is one of the most famous representatives of this intuition. So, error term defined as:

$$\varepsilon_t = x_t\beta,$$

where x_t is a scalar with positive autocorrelation and zero mean, while β is an *N x 1* vector with positive elements. By following return equation (2.18), $r_t = \mu + \eta_t + \beta\Delta x_t$ where x_t is an AR(1) process, $x_t = \rho x_{t-1} + \nu_t$. Dividend yield and Δx_t have covariance:

$$\begin{aligned}
\delta \equiv \mathbb{C}ov(\eta_t, \Delta x_t) &= \mathbb{C}ov(\eta_t, x_{t-1}(\rho-1) + \nu_t), \\
&= \mathbb{C}ov(\eta_t, \nu_t), \qquad (2.32)
\end{aligned}$$

since η_t is uncorrelated with the past. Covariance between temporary price movements and dividends, $\delta > 0$ is another assumption. Return covariance:

$$\begin{aligned}
\mathbb{E}[(r_t - \mu)^2] &= \mathbb{E}[(\eta + \beta\Delta x_t)^2], \\
&= \mathbb{E}[\eta_t^2] + \beta\beta'\mathbb{E}[(\Delta x_t)^2] + 2\beta\underbrace{\mathbb{E}[\eta_t\Delta x_t]}_{\delta}, \\
\Rightarrow \mathbb{C}ov(r_t) &= \Sigma + \sigma_{\Delta x}^2\beta\beta' + \beta\delta' + \delta\beta'.
\end{aligned}$$

Notice that, covariance of returns are greater than dividend covariance which means time-varying risk premium increases the variances and covariances. The first-order

autocovariance is:

$$\begin{aligned}
\mathbb{E}[(r_t-\mu)(r_{t-1}-\mu)] &= \mathbb{E}\big[(\eta_t+\beta\Delta x_t)(\eta_{t-1}+\beta\Delta x_{t-1})\big],\\
&= 0+\beta\mathbb{E}[\eta_t\Delta x_{t-1}]+\beta\mathbb{E}[\Delta x_t\eta_{t-1}]+\beta\beta'\mathbb{E}[\Delta x_t\Delta x_{t-1}],\\
&= \mathbb{E}[(x_{t-1}(\rho-1)+\nu_t)\eta_{t-1}]\beta+\beta\beta'\rho_{\Delta x},\\
&= \mathbb{E}[\nu_{t-1}\eta_{t-1}](\rho-1)+\rho_{\Delta x}\beta\beta',\\
\Rightarrow \mathbb{C}ov(r_t,r_{t-1}) &= \rho_{\Delta x}\beta\beta'+(\rho-1)\beta\delta'<0, \qquad (2.33)
\end{aligned}$$

as the auto-covariance of Δx_t, $\rho_{\Delta x}<0$ and $\rho<1$. Thus, momentum profits

$$\mathbb{E}[\pi_t]=\rho_{\Delta x}\sigma_\beta^2+(\rho-1)\sigma_{\beta,\delta}+\sigma_\mu^2, \qquad (2.34)$$

where σ_β^2 denotes cross-sectional variance of β and $\sigma_{\beta,\delta}$ denotes cross-sectional covariance between β and δ. Since β stands for sensitivity of stock prices to risk premium changes, to be able to obtain positive momentum profits, relation between cash flow(dividend) covariances, and risk premium sensitivity of the stock prices has negative direction.

As mentioned, Lewellen also shows that momentum effect exists in industry, size, and book-to-market portfolios. Fama and French in [20] uses their three factor model to capture asset pricing anomalies. However, although they managed to explain big part of the mean-reversion effect, they lacked while capturing momentum. So, it might be claimed that, though industry, size, and book-to-market value factors are not able to explain (or at least they lack) momentum profits, they somehow contain that effect(see also Moskowitz and Grinblatt [28]).

Unlike under-reaction models suggested by Barberis et al. [4] and Hong and Stein [22], Lewellen [25] attributes momentum to over-reaction and excess covariance with these models since he detects negative autocorrelation of return patterns in his empirical research. On the other hand, Chen and Hong [12] publishes a discussion of Lewellen's paper and advocates different models in order to both explain momentum and inconsistency of him.

2.1.3 Model Analysis of Chen and Hong[12]

Lewellen[25] shows that returns are negatively auto- and cross-correlated and claims that, this would be only consistent with over-reaction and excess covariance based models. However, Chen and Hong[12] provides a different approach in which under-reaction can lead momentum in presence of negative auto and cross-correlation.

They consider a simple one-factor model in which momentum has a single source, under-reaction. Then, stock returns(log) are expressed as follows

$$r_{i,t}=\mu_i+\beta_i x_t+\varepsilon_{i,t}, \qquad (2.35)$$

where

$$x_t=\rho x_{t-1}+\nu_t.$$

This time x_t is demeaned market factor with variance σ_x^2 and serial auto-correlation ρ. It is also not a part of the return shock ε_t. Like previous setting, ν_t is uncorrelated shock to factor x_t and ε_t is positively correlated idiosyncratic shock with mean zero, variance σ_ε^2, and the first order auto-covariance $\mathbb{E}[\varepsilon_t\varepsilon_{t-1}] = \kappa\sigma_\varepsilon^2$ which is greater than zero. All other relations are disregarded. Furthermore, all stocks in the market are assumed to have same mean $\mu_i = \mu$ and same factor sensitivity $\beta_i = \beta = 1$. Each weight $w_{i,t}$ for stock i at time t is again defined as equation (2.8) which yields zero cost portfolios. So, profits are immediately realized by equation (2.10). We can also express weights in a more informative form:

$$\begin{aligned}
r_{i,t-1} - r_{m,t-1} &= \mu x_{t-1} + \varepsilon_{i,t-1} - \mu - x_{t-1} - \frac{1}{N}\sum_{j=1}^{N}\varepsilon_{j,t-1}, \\
\Rightarrow \frac{1}{N}(r_{i,t-1} - r_{m,t-1}) &= \frac{1}{N}\left[\varepsilon_{i,t-1} - \frac{1}{N}\sum_{j=1}^{N}\varepsilon_{j,t-1}\right], \\
&= \frac{1}{N}\left[\varepsilon_{i,t-1} - \varepsilon_{i,t-1}\frac{1}{N} - \frac{1}{N}\sum_{j\neq i}^{N}\varepsilon_{j,t-1}\right], \\
\Rightarrow w_{i,t} &= \frac{N-1}{N^2}\varepsilon_{i,t-1} - \frac{1}{N^2}\sum_{j\neq i}^{N}\varepsilon_{j,t-1}. \qquad (2.36)
\end{aligned}$$

Hence, by using these weights profit equation becomes

$$\pi_t = \sum_{i=1}^{N} w_{i,t}r_{i,t} = \sum_{i=1}^{N}\left[\frac{N-1}{N^2}\varepsilon_{i,t-1} - \frac{1}{N^2}\sum_{j\neq i}^{N}\varepsilon_{j,t-1}\right]r_{i,t}. \qquad (2.37)$$

Expected momentum profits can be found by taking the expectation of both sides.

$$\begin{aligned}
\mathbb{E}[\pi_t] &= \sum_{i=1}^{N}\left[\frac{N-1}{N^2}\mathbb{E}[\varepsilon_{i,t-1}r_{i,t}] - \frac{1}{N^2}\sum_{j\neq i}^{N}\cancel{\mathbb{E}[\varepsilon_{j,t-1}r_{i,t}]}\right], \\
&= \sum_{i=1}^{N}\left[\frac{N-1}{N^2}\big(\mathbb{E}[\varepsilon_{i,t-1}]\mu + \mathbb{E}[\varepsilon_{i,t-1}x_t] + \underbrace{\mathbb{E}[\varepsilon_{i,t-1}\varepsilon_{i,t}]}_{\kappa\sigma_\varepsilon^2}\big)\right], \\
&= \frac{N-1}{N^2}\sum_{i=1}^{N}\kappa\sigma_\varepsilon^2, \\
&= \kappa\sigma_\varepsilon^2\frac{N-1}{N} > 0. \qquad (2.38)
\end{aligned}$$

Notice that, last expression only depends on idiosyncratic return shocks, which are positively serially correlated. It is indeed guaranteed by model construction (constant unconditional means and betas).

Before expressing this profit equation under Lo and MacKinlay[27] decomposition, let us start from its generalized version.

$$\mathbb{E}[\pi_t] \equiv \sigma_\mu^2 + O - C. \qquad (2.39)$$

Accordingly, there are three possible sources of momentum profits (like in equation 2.15). First term stands for cross sectional variance of the unconditional means. So, our trading strategy might choose stocks with high unconditional mean in exchange for high risk without any irrational investor behaviour. In this one factor model, it is defined

$$\sigma_{\mu}^2 = \frac{1}{N}\sum_{i=1}^{N}(\mu_i - \mu_m)^2, \tag{2.40}$$

which is used while passing from one profit equation (2.12) to Lo and Mackinlay decomposition (2.15). Second term of equation (2.39 is average auto-covariance of the returns. If, today's good performance for a given stock foresees a good future performance for the same stock, this term makes positive contribution to momentum profits. Its role can be perceived as the time series predictability impact. For our model it becomes

$$O = \frac{N-1}{N^2}\sum_{i=1}^{N}\mathbb{E}[r_{i,t}r_{i,t-1} - \mu_i^2], \tag{2.41}$$

and the last term with minus coefficient is average cross-serial covariance. So, it has positive effect on momentum profits as long as today's bad performance for a given stock predicts a good future performance for another stock. So, abnormal returns could be generated by taking the advantage of negative correlation between return series. Under this one-factor model it is defined as follows

$$C = \mathbb{E}[r_{m,t}r_{m,t-1}] - \mu_m^2 - \frac{1}{N^2}\sum_{i=1}^{N}\mathbb{E}[r_{i,t}r_{i,t-1} - \mu_i^2]. \tag{2.42}$$

Therefore, applying equation (2.40) to this one factor model, $\sigma_{\mu}^2 = 0$ occurs because of constant mean assumption. For, O apply equation (2.41):

$$\mathbb{E}[r_{i,t}r_{i,t-1} - \mu_i^2] = \mathbb{E}[(\mu_i + \beta x_t + \varepsilon_{i,t})(\mu_i + \beta x_{t-1} + \varepsilon_{i,t-1}) - \mu_i^2].$$

Note that, $\mu_i = \mu$ is constant and $\beta = 1$ is assumed. Then,

$$\begin{aligned}\mathbb{E}[r_{i,t}r_{i,t-1} - \mu_i^2] &= \mathbb{E}[\mu^2 + \mu x_{t-1} + \mu\varepsilon_{i,t-1} + \mu\varepsilon x_t + x_t x_{t-1} + x_t\varepsilon_{i,t-1} + \varepsilon_{i,t}\mu \\ &\quad +\varepsilon_{i,t}x_{t-1} + \varepsilon_{i,t}\varepsilon_{i,t-1} - \mu^2], \\ &= \mu\mathbb{E}[x_{t-1} + x_t] + \mathbb{E}[x_t x_{t-1}] + \mathbb{E}[\varepsilon_{i,t}\varepsilon_{i,t-1}].\end{aligned}$$

Recalling that x_t is a demeaned process with autocorrelation ρ and taking summation of both sides while multiplying with $\frac{N-1}{N^2}$:

$$\begin{aligned}O &= \frac{N-1}{N^2}\sum_{i=1}^{N}(\sigma_x^2\rho + \kappa\sigma_{\varepsilon}^2), \\ &= \frac{N-1}{N}(\rho\sigma_x^2 + \kappa\sigma_{\varepsilon}^2).\end{aligned} \tag{2.43}$$

To find C, let us start from the first term of the equation (2.42):

$$\begin{aligned}
\mathbb{E}[r_{m,t}r_{m,t-1}] &= \mathbb{E}\bigg[\frac{1}{N}\sum_{i=1}^{N} r_{i,t}\frac{1}{N}\sum_{j=1}^{N} r_{j,t-1}\bigg], \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}[(\mu + x_t + \varepsilon_{i,t})(\mu + x_{t-1} + \varepsilon_{j,t-1})], \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\big\{(\mu^2 + \rho\sigma_x^2) + \mathbb{E}[\varepsilon_{i,t}\varepsilon_{j,t-1}]\big\}, \\
&= \mu^2 + \rho\sigma_x^2 + \frac{1}{N^2}\bigg[\sum_{i=1}^{N}\varepsilon_{i,t}\varepsilon_{i,t-1}\sum_{j=1}^{N}\varepsilon_{j,t-1}\bigg], \\
&= \mu^2 + \rho\sigma_x^2 + \frac{1}{N^2}\sum_{i=1}^{N}\kappa\sigma_\varepsilon^2, \\
&= \mu^2 + \rho\sigma_x^2 + \frac{1}{N}\kappa\sigma_\varepsilon^2. \qquad (2.44)
\end{aligned}$$

Hence, plugging this expression into the equation (2.42) brings C.

$$\begin{aligned}
C &= \mu^2 + \rho\sigma_x^2 + \frac{1}{N}\kappa\sigma_\varepsilon^2 - \mu_m^2 - \frac{1}{N^2}\sum_{i=1}^{N}\mathbb{E}[r_{i,t}r_{i,t-1} - \mu_i^2] \\
&= \rho\sigma_x^2 + \frac{1}{N}\kappa\sigma_\varepsilon^2 - \frac{1}{N}\rho\sigma_x^2 - \frac{1}{N}\kappa\sigma_\varepsilon^2 \\
&= \rho\sigma_x^2\bigg(\frac{N-1}{N}\bigg) \qquad (2.45)
\end{aligned}$$

Therefore, by combining equations (2.40),(2.43), and (2.45) results with momentum profit equation (2.38).

$$\begin{aligned}
\mathbb{E}[\pi_t] &= 0 + \bigg(\frac{N-1}{N}\bigg)(\rho\sigma_x^2 + \kappa\sigma_\varepsilon^2) - \rho\sigma_x^2\bigg(\frac{N-1}{N}\bigg). \\
&= \frac{N-1}{N}\kappa\sigma_\varepsilon^2. \qquad (2.46)
\end{aligned}$$

However, this result indicates two important interpretations mentioned by Chen and Hong[12]. First, though momentum profits seemingly doesn't depend on factor autocorrelation, in this model ρ, Lo and MacKinlay decomposition does. Therefore, it can be said that, this decomposition may lack while explaining the source of momentum profits. Another crucial observation, equations (2.43), and (2.45) can both be positive if the autocorrelation of the factor is necessarily positive or vice versa. Thereby, even under-reaction story for explaining momentum might be consistent with negative autocorrelation. Hong and Chen also supports this claim by empirical findings.

One can also interpret from the models analysed through subsections (2.1.1) - (2.1.3) that, neither of the arguments manage to provide full picture. Since models are constructed around some pre-assumptions, they might end up with different results for same data sets or vice versa.

There are also many other models including Fama and French three factor model[19]. However, under the scope of this study, we examine the similarity between two main momentum based models(Jegadeesh and Titman, and Lewellen) which are proposed in different years(Lewellen is roughly 9-10 years latter) and point out what happens when their assumptions are changed a little(Chen and Hong). Our analysis shows that, momentum really exists on data, and there is also descriptive decompositions to anticipate its source. On the other hand, none of them is enough to explain its source and it can be claimed that, these discrete time models, by construction, are in such a loop that they can't say more. By the way, one can attribute momentum under-reaction or over-reaction with positive or negative autocorrelation; but remember that, moving between these assumptions could be done just by changing the noise term.

Therefore, rather than trying to explain its source by some factors, our continuous time model takes momentum as a factor(because it exists) of return. Whenever a return series contain momentum and mean-reversion effects, their significance are determined by an estimated coefficient directly coming from the data. Even though, this method is much easier than examining various portfolios, it might not be as informative as classical scheme since it only considers the time series predictability.

2.2 A Fundamental Continuous Time Model

Unlike discrete time models, there is not much study to capture momentum in continuous time setting[2]. One reason for this can be stated like that, in equation (2.1) weights for the stocks are predetermined, in other words, momentum arises like a strategy in which an investor chooses stocks that generated abnormal returns in the past. However, to asses a continuous time model, one needs to define a momentum coefficient and/or seek for optimality conditions.

Our framework only encompasses the coefficient estimation part. This scheme has a significant advantage with compared the discrete time models, detecting momentum for a single time series. Rather than searching for the cross-serial correlations between stock, we try to investigate momentum and also mean reversion effect from a single return pattern. We followed the structure established by Koijen, Rodríguez, and Sbuelz [24]. So, it will be more logical to analyse this model setting and passing through our assumptions and parameter estimation methods.

2.2.1 Model Analysis of Koijen, Rodríguez, and Sbuelz[24]

This model combines the short term return persistence and long term return reversion in a intuitive way. Its structure actually, particularly in mean reversion, linked to Campbell and Viceira[9], and Wachter[33].

Since, momentum in financial markets means predictive power of short term perfor-

[2] For mean reversion anomaly see Campbell and Viceira[9], and Wachter[33]. And for continuous time momentum see Rodríguez and Sbuelz[30]

mance for the future, a state variable that summarizes the past performance would be able capture this impact accurately. Mean reversion variable is chosen as dividend yield by following the past studies mentioned above.

Equity index value at time t is defined by S_t in which dividends are reinvested. S_t would be constructed as a weighted average of its past performance and dividend yield.

$$\frac{dS_t}{S_t} = (\phi M_t + (1-\phi)\mu_t)dt + \sigma_S' dZ_t, \;\; 0 \leq \phi < 1, \tag{2.47}$$

where Z_t is a two-dimensional vector of independent Brownian Motions(like $[Z_t^1 \; Z_t^2]'$) and σ_S' is two dimensional volatility vector. According to equation (2.47) expected returns depend on two state variables μ_t and M_t.

M_t is called as performance variable and defined as weighted sum of past returns:

$$M_t = \int_0^t e^{-w(t-u)} \frac{dS_u}{S_u}, \tag{2.48}$$

where $w > 0$ and returns are weighted by $e^{-w(t-u)}$. Since, their estimation is close to 1, for the sake of simplicity $w = 1$ is assumed. Also, note that, S_t contains reinvested dividends which means M_t is calculated over cumulative returns.

Performance variable, by construction, follows the stock return dynamics. If M_t is totally differentiated, by applying Itó product rule[3]

$$\begin{aligned} M_t &= e^{-t} \int_0^t e^u \frac{dS_u}{S_u}, \\ \Rightarrow dM_t &= \underbrace{-e^{-t} \int_0^t e^u \frac{dS_u}{S_u} du}_{M_t} dt + e^{-t} e^t \frac{dS_t}{S_t}, \\ &= \frac{dS_t}{S_t} - M_t dt, \end{aligned} \tag{2.49}$$

relationship between M_t and S_t might be seen more clearly. Now, substituting equation (2.47) into (2.49):

$$\begin{aligned} dM_t &= (\phi M_t + (1-\phi)\mu_t)dt + \sigma_S' dZ_t - M_t dt, \\ &= (1-\phi)(\mu_t - M_t)dt + \sigma_S' dZ_t. \end{aligned} \tag{2.50}$$

This decomposition is informative as it shows the performance variable oscillates around a stochastic mean μ_t which is called as mean-reversion variable. This makes sense as equity index price is assumed to exhibit a mean reverting behaviour in long term which causes short term shocks to fluctuate around a long term fundamental level. μ_t is assumed to be stationary and follows Ornstein-Uhlenbeck process:

$$d\mu_t = \alpha(\mu_0 - \mu_1)dt + \sigma_\mu' dZ_t, \;\; \alpha > 0, \tag{2.51}$$

[3] Notice that quadratic variation between e^{-t} and $\int_0^t e^u \frac{dS_u}{S_u}$ is 0.

where μ_0 is long term mean and α is the convergence rate of μ_t to constant μ_0, and finally σ_μ is two-dimensional volatility vector. It is not too hard to solve this SDE(stochastic differential equation): Let,

$$\begin{aligned} L_t &= \mu_t - \mu_0 \\ \Rightarrow dL_t &= d\mu_t, \\ \Rightarrow d\mu_t &= -\alpha L_t dt + \sigma'_\mu dZ_t, \end{aligned} \tag{2.52}$$

Now, set:

$$\begin{aligned} e^{\alpha t}L_t &= K_t, \\ \Rightarrow K_o &= L_0 = L, \\ \Rightarrow dK_t &= \alpha e^{\alpha t}L_t dt + e^{\alpha t}dL_t, \\ &= \cancel{\alpha e^{\alpha t}L_t dt} + e^{\alpha t}(\cancel{-\alpha L_t dt} + \sigma'_\mu dZ_t), \\ \Rightarrow K_t - K_0 &= \int_0^t e^{\alpha u}\sigma'_\mu dZ_u, \\ \Rightarrow e^{\alpha t}L_t &= L + \int_0^t e^{\alpha u}\sigma'_\mu dZ_u, \\ \Rightarrow L_t &= e^{-\alpha t}L + \int_0^t e^{-\alpha(t-u)}\sigma'_\mu dZ_u, \end{aligned}$$

So, by equation (2.52) and assuming $L = \mu - \mu_0$[4]

$$\mu_t = e^{-\alpha t}(\mu - \mu_0) + \sigma'_\mu \int_0^t e^{-\alpha(t-u)} dZ_u. \tag{2.53}$$

Meanwhile, by equation (2.47), conditional expected returns are realized as follows

$$\begin{aligned} \mathbb{E}_t\Big[\frac{dS_t}{S_t}\Big] &= (\phi M_t + (1-\phi)\mu_t)dt, \\ &= (\mu_t + \phi(M_t - \mu_t))dt. \end{aligned} \tag{2.54}$$

This expression says that, if past performance have predictive power over future, $\phi > 0$ holds and according to value of $M_t - \mu_t$, upcoming returns are expected to be higher(if $M_t - \mu_t > 0$) or lower(if $M_t - \mu_t < 0$) than long term mean. Conversely, if $\phi = 0$ holds, then, past performance would have no predictive power on the return series and it will fluctuate around stochastic mean μ_t. These interpretations represent momentum and mean reversion effects on returns.

In order to calibrate their model, Koijen, Rodríguez, and Sbuelz use proxies for both mean-reversion variable μ_t and performance variable M_t. For, μ_t log dividend yields [5] are used with some adjustments as they are usually not easy to handle data series.

$$\mu_t = \mu_0 + \mu_1(D_t - \mu_D) = \mu_0 + \mu_1 X_t, \tag{2.55}$$

[4] Since μ_0 is constant long term mean, μ is used as the initial value of dividend yield series μ_t.

[5] For predictive power of dividend yields on mean reversals in US financial market see Fama and French [17, 18], Cochrane [15], Chen [13]. For emerging markets including Turkey see Aras and Yılmaz [2].

where D_t stands for log dividend yield, and $\mathbb{E}[D_t] = \mu_D$. Meanwhile, $X_t = D_t - \mu_D$ becomes demeaned dividend yield. M_t is discretized, by following Euler's scheme, over each observed time intervals between $[0, t]$ as a weighted summation. Monthly returns are used to approximate the integral 2.48.

$$M_t \approx \sum_{i=1}^{t} e^{-i} \left(\frac{S_{t-i+1} - S_{t-i}}{S_{t-i}} \right). \tag{2.56}$$

Therefore, general outlook of the model becomes:

$$\begin{aligned}
\frac{dS_t}{S_t} &= ((\mu_0 + \mu_1 X_t)(1-\phi) + \phi M_t)dt + \sigma_S' dZ_t, && (2.57)\\
dM_t &= (1-\phi)(\mu_0 + \mu_1 X_t - M_t)dt + \sigma_S' dZ_t, && (2.58)\\
dX_t &= -\alpha X_t dt + \sigma_X' dZ_t, && (2.59)
\end{aligned}$$

with $\sigma_X = \sigma_\mu / \mu_1$.

Notice that, equation(2.59) comes from (2.51):

$$\begin{aligned}
\mu_t &= \mu_0 + \mu_1 X_t,\\
\Rightarrow \mu_t &= \frac{\mu_t - \mu_0}{\mu_1} X_t \ \ and,\\
\Rightarrow d\mu_t &= \mu_1 dX_t.\\
\frac{d\mu_t}{\mu_1} &= \alpha \underbrace{\frac{\mu_0 - \mu_t}{\mu 1}}_{-X_t} dt + \sigma_\mu' / \mu_1 dZ_t,\\
\Rightarrow dX_t &= -\alpha X_t dt + \sigma_X' dZ_t.
\end{aligned}$$

They handle data by following Campbell at al.[8], where dividend yield is taken as natural logarithm of sum of the dividend payments over past year divided by current price index.

As mentioned before the randomness of the equations (2.58)-(2.59) is driven by two independent Brownian motions. The Cholesky decomposition of volatility matrix is:

$$\tilde{\Sigma} = \begin{pmatrix} \sigma_S' \\ \sigma_X' \end{pmatrix} = \begin{pmatrix} \sigma_{S(1)} & 0 \\ \sigma_{X(1)} & \sigma_{X(2)} \end{pmatrix}. \tag{2.60}$$

In line with Sangvinatsos and Wachter [31], and Binsbergen, Bandt, and Koijen [6], Z_t^1 is the return shock which is orthogonal to dividend yield shock Z_t^2.

Koijen, Rodríguez, and Sbuelz then, discretized the constructed model as a Vector Auto-regressive(VAR) model and used maximum likelihood function(MLE) to estimate parameters of equations (2.58)-(2.59) with $\phi > 0$ and $\phi = 0$ case for NYSE, NASDAQ, and AMEX markets. They also examine Value Weighted(VW) and Equally Weighted(EW) indices in their analysis. Our framework also nests on this methodology. On the next chapter, model will be handled with dicretizations and MLE outlook.[6]

[6] Equity index momentum also has previous studies. See, for instance, [11].

CHAPTER 3

MODEL CALIBRATION

In this chapter, methods used to execute parameter estimation processes will be discussed. We mainly examine two celebrated numeric discretization schemes for two well-known stochastic differential equations: Black-Scholes stock prices and Ornstein-Uhlenbeck process. Because they also have analytic solution, it is possible to make comparison between these methods by means of mean square error. In the light of our results, we also convert Koijen, Rodrí, and Sbuelz model into a vector autoregressive form. Moreover, both univariate and multivariate forms of maximum likelihood function is discussed and applied to a couple of converted SDEs inside our model.

3.1 Discretization of Stochastic Differential Equations

Two mainly important SDE discretization schemes will be discussed in this section. One of them is Euler's method and other is Milstein's method.

3.1.1 Euler-Maruyama Discretization Scheme

Consider a stochastic process $X(t,\omega)$, on the interval $[0,T]$, has the form:

$$X(t,\omega) = X(0,\omega) + \int_0^t f(s,X(s,\omega))ds + \int_0^t g(s,X(s,\omega))dW(s,\omega), \tag{3.1}$$

for $0 \leq t \leq T$,where $W(s,\omega)$ is a one-dimensional Brownian motion. $X(t,\omega)$ also has the differential form:

$$dX(t,\omega) = f(t,X(t,\omega))dt + g(t,X(t,\omega))dW(t,\omega) \tag{3.2}$$

Rather than going through exact solution or when it does not exist, one can approximate this stochastic differential form by:

$$\begin{cases} X_{k+1}(\omega) & = X_k(\omega) + f(t_k,X_k(\omega))\Delta t + g(t_k,X_k(\omega))\Delta W_k(\omega), \\ X_0(\omega) & = X(0,\omega), \end{cases} \tag{3.3}$$

for $k = 0,1,2,...,N-1$ where ω stands for sample path and,

$$\Delta t = T/N, \;\; t_k = k\Delta t,$$
$$X_k(\omega) \approx X(t_k, \omega),$$
$$\Delta W_t(\omega) = (W(t_{k+1}, \omega) - W(t_k, \omega)) \sim N(0, \Delta t).$$

For further information and Euler's method error see E. Allen [1].

Since they are two of the most celebrated stochastic processes with closed-form solution, and our model also uses one of them(Ornstein-Uhlenbeck), B-S stock prices OU process are really good candidates for this topic. We will first discretize them and then simulate paths for both closed-form solution and Euler-Maruyama approximations.

Example 3.1. Application of Euler-Maruyama method to Black-Scholes setting of stock prices:

Black-Scholes stock prices driven by geometric Brownian motion:

$$\frac{dS_t}{S_t} = \mu\, dt + \sigma\, dW_t, \;\; \mu, \sigma \in \mathbb{R}$$

where S_t is a stock price at time $0 \leq t \leq T$ with drift term μ and volatility σ. Let us define a partition over the interval $[0, t]$, $\mathbb{T} = \{t_k : 0 = t_0 < t_1 < \ldots < t_{n-1} < t_n = t\}$. By applying E-M

$$\begin{aligned} S_{t_{k+1}} - S_{t_k} &= S_{t_k}[\mu\, \Delta t + \sigma\, z_k \sqrt{\Delta t}], \\ S_{t_{k+1}} &= S_{t_k}[1 + \mu\, \Delta t + \sigma\, z_k \sqrt{\Delta t}], \end{aligned} \tag{3.4}$$

where $k = 0, \ldots, n-1$, $z_k \sim N(0,1)$. Since $z_k\sqrt{\Delta t} \sim N(0, \sqrt{\Delta t})$, this equations satisfies the conditions of 3.3. S_t has an exact solution by applying Itó lemma to $f(x) = log(x)$:

$$f'(x) = \frac{1}{x} \quad \text{and} \quad f''(x) = -\frac{1}{x^2},$$

$$\begin{aligned} \Rightarrow f(S_t) &= log(S_0) + \int_0^t \frac{1}{S_u} dS_u - \frac{1}{2}\int_0^t \frac{1}{S_u^2} d < S, S > u, \\ log(S_t) &= log(S_0) + \int_0^t \frac{1}{S_u} S_u[\mu du + \sigma dW_u] - \frac{1}{2}\int_0^t \frac{1}{S_u^2} S_u^2 \sigma^2 du, \\ log(S_t) &= log(S_0) + \int_0^t (\mu - \frac{1}{2}\sigma^2) du + \sigma \int_0^t dW_u, \\ \Rightarrow S_t &= S_0 \cdot \exp\{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t\}. \end{aligned} \tag{3.5}$$

For, $T = 1, N = 2^8, \mu = 0.1, \sigma = 0.5, S_0 = 1$ and discretized Brownian motion $W_t = z\sqrt{t}$, $z \sim N(0,1)$ a ***MATLAB*** simulation is performed [32]. It seems like a quite good approximation from figure (3.1). Mean square error between SDE and

solution is 1.7658×10^{-5} and squared error graph (3.2) is also shown[1]. Mean square errors are calculated as:

$$\mathbb{E}[(\hat{S} - S)^2] = \frac{1}{N}\sum_{i=1}^{N}(S_i^{EM} - S_i)^2$$

Example 3.2. Now, consider an Ornstein-Uhlenbeck process:

$$dX_t = -\gamma\, X_t\, dt + \sigma\, dW_t, \quad \gamma, \sigma \in \mathbb{R}$$

where X_t is a mean reverting process. Therefore, E-M approximation of it:

$$\begin{aligned} X_{t_{k+1}} - X_{t_k} &= -\gamma\, X_{t_k}\, \Delta t + \sigma\, z_k \sqrt{\Delta t}, \\ X_{t_{k+1}} &= X_{t_k}(1 - \gamma\, \Delta t) + \sigma\, z_k \sqrt{\Delta t}, \end{aligned} \tag{3.6}$$

where z_k is defined similar to equation (3.4). This process also has exact solution which is already solved in subsection (2.2.1)(see the equation (2.53)):

$$X_t = e^{-\gamma t} x_0 + \sigma \int_0^t e^{-\gamma(t-s)} dW_s, \quad X_0 = x_0. \tag{3.7}$$

Simulating[2] processes (3.6) and (3.7) for $T = 1, N = 2^8, \gamma = 0.1, \sigma = 0.5, X_0 = 1$ brings figure (3.3). It works far better with compared to B-S stock price example and its squared errors seem to be negligible but growing as time passes. OU estimation also has considerably smaller mean square error: $1.6095 \times 10^{-10} < 1.7658 \times 10^{-5}$.

3.1.2 Milstein Discretization Scheme

Euler scheme is simple and applicable. By the way, higher order approximations like Milstein is possible. Consider, once again, SDE (3.2):

$$\begin{aligned} X_{k+1}(\omega) &= X_k(\omega) + f(t_k, X_k(\omega))\Delta t + g(t_k, X_k(\omega))\Delta W_k(\omega) \\ &+ \frac{1}{2} g(t_k, X_k(\omega)) \frac{\partial g(t_k, X_k(\omega))}{\partial x} [(\Delta W_k(\omega))^2 - \Delta t], \\ X_0(\omega) &= X(0, \omega). \end{aligned} \tag{3.8}$$

It can also be written in the form:

$$\begin{aligned} X_{k+1}(\omega) &= X_k(\omega) + f(t_k, X_k(\omega))\Delta t + g(t_k, X_k(\omega)) z_k \sqrt{\Delta t} \\ &+ \frac{1}{2} g(t_k, X_k(\omega)) g_x(t_k, X_k(\omega)) \Delta t (z_k^2 - 1), \\ X_0(\omega) &= X(0, \omega), \end{aligned} \tag{3.9}$$

for $k = 0, 1, 2, ..., N-1$ is higher order approximation of it. Like in previous subsection, ω stands for sample path and:

[1] Even we have mean square error, squared errors are considered to detect if a sharp estimation errors and error path.

[2] Stochastic integral is discretized as a sum for simulation.

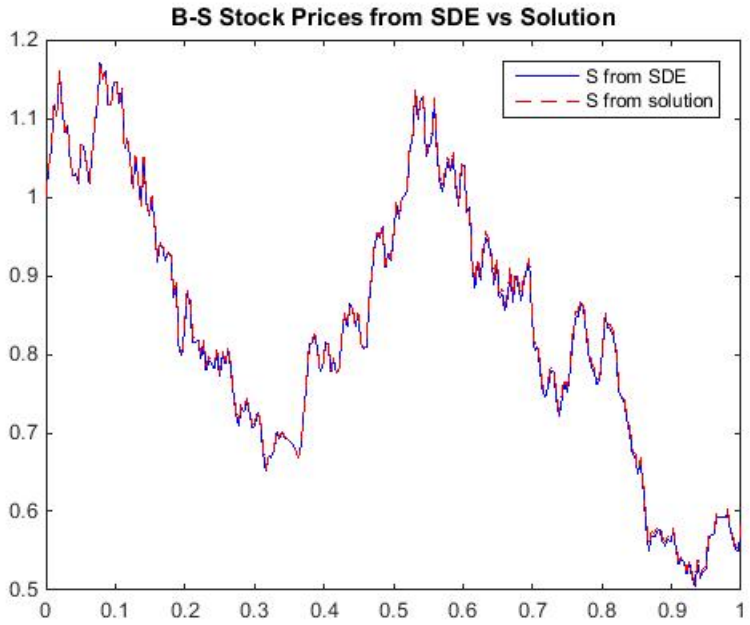


Figure 3.1: Comparision of E-M Approximation and Analytic Solution of B-S Stock Prices where $T = 1, N = 2^8, \mu = 0.1, \sigma = 0.5, S_0 = 1$.

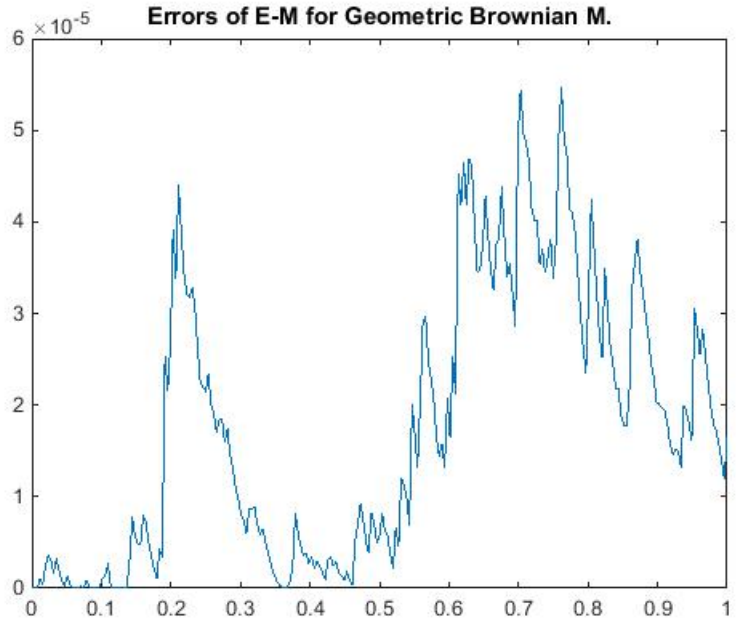


Figure 3.2: E-M Approximations Errors of B-S Stock Prices.

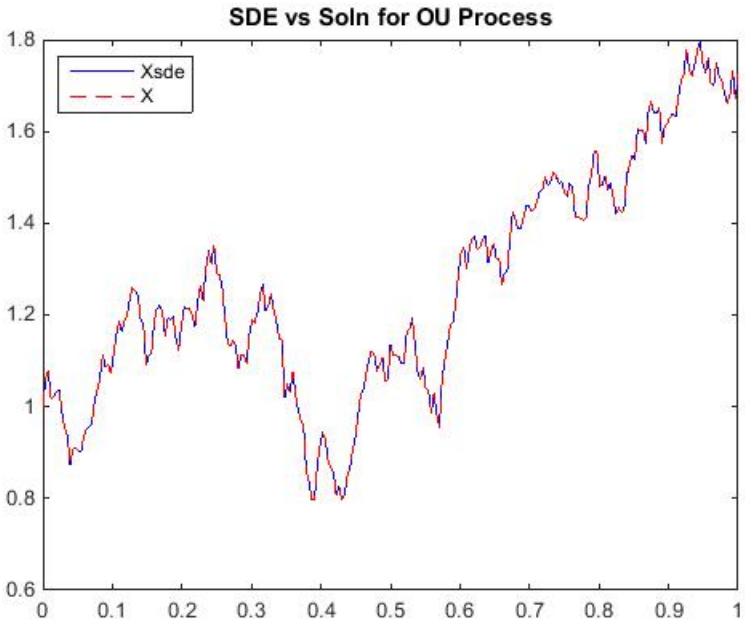


Figure 3.3: Comparision of E-M Approximation and Analytic Solution of OU Process where $T = 1, N = 2^8, \gamma = 0.1, \sigma = 0.5, X_0 = 1$.

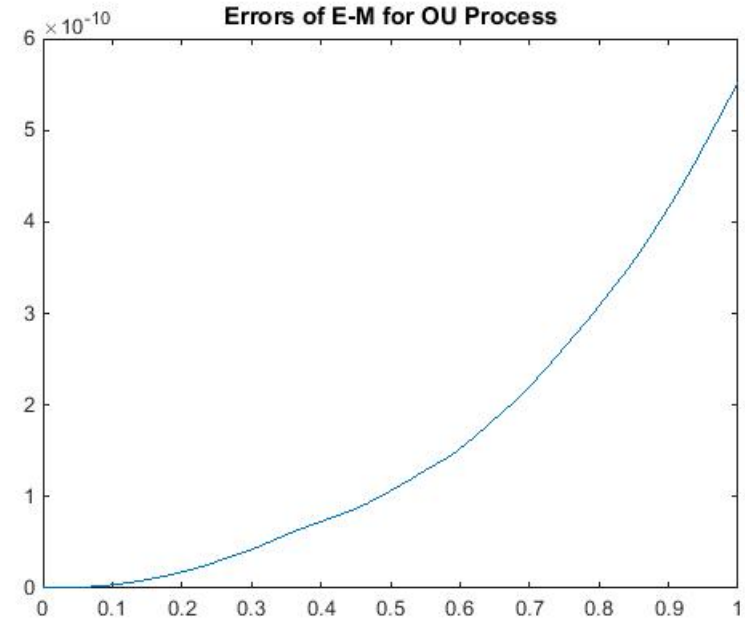


Figure 3.4: E-M Approximations Errors for OU Process.

$$\Delta t = T/N, \;\; t_k = k\Delta t,$$
$$X_k(\omega) \approx X(t_k, \omega),$$
$$\Delta W_t(\omega) = (W(t_{k+1}, \omega) - W(t_k, \omega)) \sim N(0, \Delta t).$$

Euler-Maruyama and Milstein are alike except and additional term. Because EM is of order $\mathcal{O}((\Delta t)^{1/2})$, this method converges to exact solution considerably faster with order $\mathcal{O}(\Delta t)$[3].

Example 3.3. Application of Milstein method to Black-Scholes setting of stock prices and Ornstein-Uhlenbeck processes:

For sake of simplicity recall 3.4. Now, just additional term is to be determined:

$$\begin{aligned} f(t_k, X_k(\omega)) &= S_{t_k}\,\mu \text{ and,} \\ g(t_k, X_k(\omega)) &= S_{t_k}\,\sigma, \\ \Rightarrow g_x(t_k, X_k(\omega)) &= \sigma. \end{aligned}$$

Therefore, equation (3.9) is noticed as:

$$S_{t_{k+1}} = S_{t_k}\{1 + \mu\Delta t + \sigma z_k\sqrt{\Delta t}\} + S_{t_k}\frac{1}{2}\sigma^2\Delta t(z_k^2 - 1). \tag{3.10}$$

Random number generator used to generate same random normal variates z_k. So, results will be comparable with figures (3.1) and (3.2). Parameters are also set same $T = 1, N = 2^8, \mu = 0.1, \sigma = 0.5, S_0 = 1$,.

Simulating equation (3.10) and exact solution leads to figures (**??**) and (**??**) which are slightly better than EM results. It has mean square error $2.7534 \times 10^{-08} < 1.7658 \times 10^{-5}$. For better understanding, simulations applied for different N and table 3.1 obtained. Milstein is converging in a faster manner as expected. However, each

Table 3.1: List of MSEs of E-M and Milstein

N	EM MSE	Milstein MSE
2^8	1.7658×10^{-5}	2.7534×10^{-8}
2^9	6.3991×10^{-6}	4.0630×10^{-9}
2^{10}	2.6481×10^{-6}	5.9169×10^{-10}
2^{11}	1.2312×10^{-6}	1.0630×10^{-10}
2^{12}	2.4894×10^{-7}	2.8171×10^{-11}
2^{13}	1.7596×10^{-7}	1.7735×10^{-11}

of the cases mean square error is truly small. For a parameter estimation from a single simple function, Milstein scheme would be preferable; but E.Allen [1] also shows that, predictions of these methods are quite close. Therefore, Milstein model's convergence might not create a necessary discrepancy for lower order models.

[3] For further information see E. Allen[1].

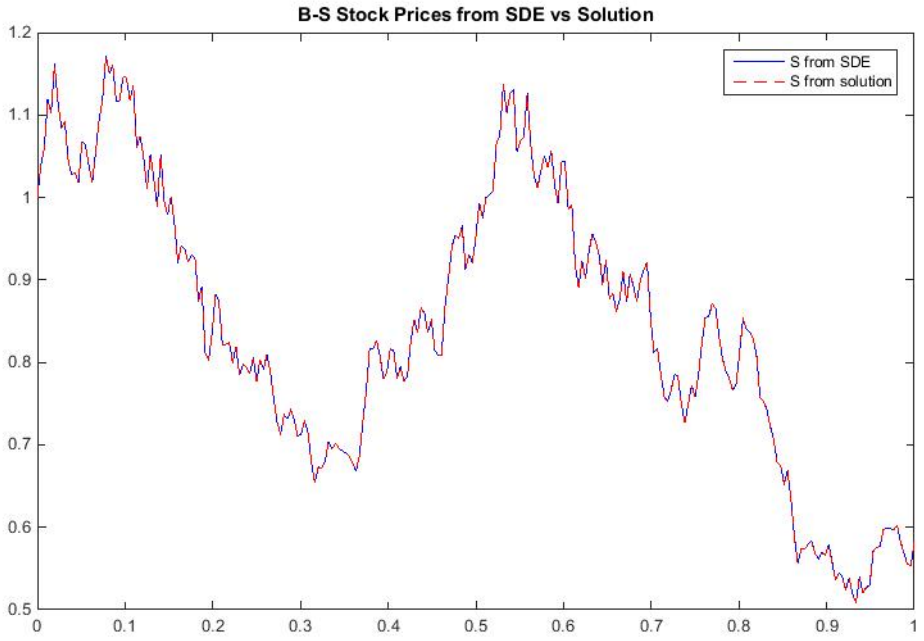


Figure 3.5: Comparision of Milstein Approximation and Analytic Solution of B-S Stock Prices where $T = 1, N = 2^8, \mu = 0.1, \sigma = 0.5, S_0 = 1$.

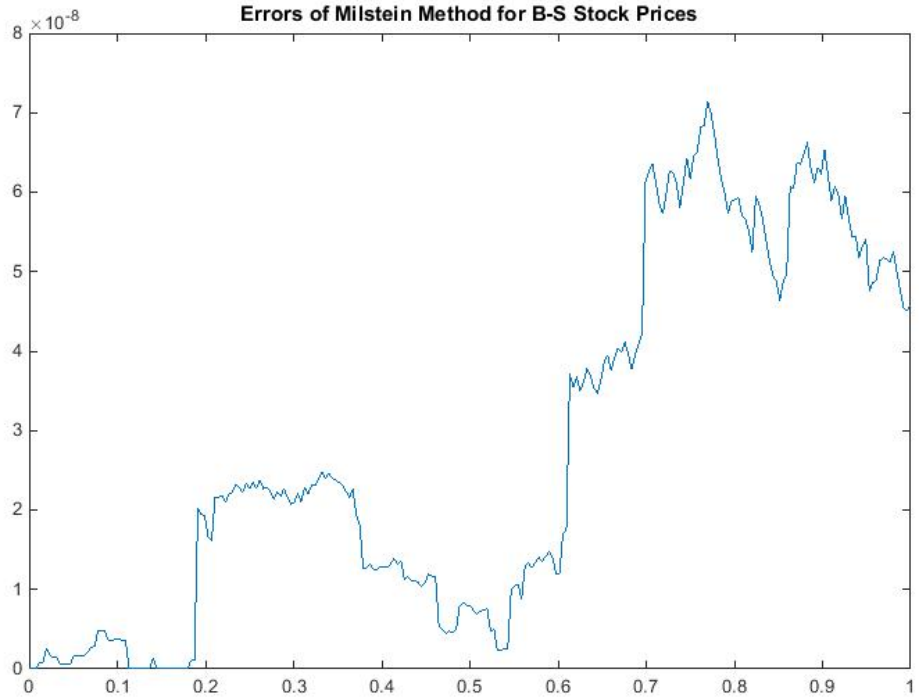


Figure 3.6: Milstein Approximations Errors for B-S Stock Prices.

On the other hand, for OU process, results are very appealing. Additional term calculations:

$$\begin{aligned} f(t_k, X_k(\omega)) &= -\gamma\, X_{t_k} \text{ and,} \\ g(t_k, X_k(\omega)) &= \sigma, \\ \Rightarrow g_x(t_k, X_k(\omega)) &= 0. \end{aligned} \tag{3.11}$$

Equation (3.11) demonstrates that, for an OU process, Milstein and Euler-Maruyama approximations coincide. This explains why we get stronger approximation for OU process with compared to B-S prices while applying E-M discretization.

These observations are very useful for converting SDEs to a vector auto-regressive structure, as our model contains a OU process and lower order SDEs. Hence, Euler's scheme is preferred for our framework.

3.1.3 Converting Koijen, Rodrígez, and Sbuelz [24] Setting to a VAR Model

Before initializing parameter estimations, the model needs to be adjusted to form a maximum likelihood function. Recall the equations (2.57)–(2.59):

$$\begin{aligned} \frac{dS_t}{S_t} &= ((\mu_0 + \mu_1 X_t)(1-\phi) + \phi M_t)dt + \sigma_S' dZ_t, \\ dM_t &= (1-\phi)(\mu_0 + \mu_1 X_t - M_t)dt + \sigma_S' dZ_t, \\ dX_t &= -\alpha X_t dt + \sigma_X' dZ_t, \end{aligned}$$

where $\sigma_X = \sigma_\mu/\mu_1$. Because of the reasons listed in previous subsections, Euler's method is chosen. It enables us to construct multivariate maximum likelihood easier indeed. Therefore,

$$\begin{aligned} \frac{S_{t_{k+1}} - S_{t_k}}{S_{t_k}} &\approx [(\mu_0 + \mu_1\, x_{t_k})(1-\phi) + \phi\, m_{t_k}]\Delta t \\ &\quad + \sigma_{S(1)}\sqrt{\Delta t}\,\eta_k + \sigma_{S(2)}\sqrt{\Delta t}\,\upsilon_k, \qquad (3.12) \\ M_{t_{k+1}} &\approx m_{t_k} + (1-\phi)(\mu_0 + \mu_1\, x_{t_k} - m_{t_k})\Delta t \\ &\quad + \sigma_{S(1)}\sqrt{\Delta t}\,\eta_k + \sigma_{S(2)}\sqrt{\Delta t}\,\upsilon_k, \qquad (3.13) \\ X_{t_{k+1}} &\approx x_{t_k} - \alpha\, x_{t_k}\,\Delta t + \sigma_{X(1)}\sqrt{\Delta t}\,\eta_k + \sigma_{X(2)}\sqrt{\Delta t}\,\upsilon_k, \qquad (3.14) \end{aligned}$$

where $(S_{t_{k+1}} - S_{t_k})/S_{t_k} \equiv R_{t_k}$, $X_{t_k} \equiv x_{t_k}$, $M_{t_k} \equiv m_{t_k}$, $\Delta t = T/N = 1$ (time intervals and data frequency coincide) and η_k, $\upsilon_k \sim N(0,1)$ and also $\eta_k \perp \upsilon_k$ stands for independent Brownian motions $Z_t^{(1)}$ and $Z_t^{(2)}$. From Cholesky decomposition of volatilities (2.60), $\sigma_S^{(2)} = 0$. Do we really need M_t in this form? It is already discretized

over its solution in equation (2.56). So, it becomes:

$$\begin{aligned} M_t \approx M_{t_n} &= \sum_{k=0}^{n-1} e^{-(t_n - t_k)} \left(\frac{S_{t_{k+1}} - S_{t_k}}{S_{t_k}} \right) = \sum_{k=0}^{n-1} e^{-(n-k)} r_{t_{k+1}}, \\ &= r_{t_1} e^{-n} + r_{t_2} e^{-(n-1)} + \ldots + r_{t_{n-1}} e^{-(n-(n-2))} + r_{t_n} e^{-1}, \\ &= \sum_{i=1}^{t} e^{-i} r_{t+1-i}. \end{aligned} \tag{3.15}$$

This means, estimating the parameters of R_t will enable us to get M_t. Moreover, X_k has exact solution as well. However, EM and Milstein schemes coincides in its setting which yields a pretty good approximation. Beside, this formation will enable us to take expectation and variance in a neat Cholesky decomposition (2.60).

Thus, equations (3.12)–(3.14) turns out to be a decomposition which will be very useful later on:

$$\begin{cases} R_{t_{k+1}} &= [(\mu_0 + \mu_1 x_{t_k})(1 - \phi) + \phi\, m_{t_k}]\Delta t + \sigma_{S(1)} \sqrt{\Delta t}\, \eta_k, \\ X_{k+1} &= x_{t_k} - \alpha\, x_{t_k} \Delta t + \sigma_{X(1)} \sqrt{\Delta t}\, \eta_k + \sigma_{X(2)} \sqrt{\Delta t}\, \upsilon_k. \end{cases} \tag{3.16}$$

3.2 Maximum Likelihood Estimation for SDEs

This section will be composed of a general approach that explains how MLE is used to carry out parameter estimation in SDEs [1] and its direct application for our system of equations.

3.2.1 General Approach

Once, a stochastic differential equation or set of stochastic differential equations either discretized by Euler-Maruyama or Milstein method, it is possible estimate their parameters using likelihood function for observed values. For univariate case consider the equation (3.2) in a way that:

$$dX(t) = f(t, X(t); \theta)dt + g(t, X(t); \theta)dW_t, \tag{3.17}$$

for $\theta \in \mathbb{R}^m$ is a vector of unknown parameters. $X(t)$ is assumed to be observed over values:

$$x_0, x_1, x_2, ..., x_N,$$

on equivalent time intervals $i\Delta t$, $i = 0, 1, 2, ..., N$ where $\Delta t = T/N$ like mentioned in previous section. We will try to estimate θ using these $N + 1$ values of X.

Suppose, $p(t_k, x_k | t_{k-1}, x_{k-1}; \theta)$ be the transition probability density of t_k, x_k for given the value x_{k-1} at time t_{k-1} for vector θ with initial state $p_0(x_0|\theta)$. Likelihood function

of theta is:

$$\begin{aligned} L(\theta) &= p_0(x_0|\theta) \cdot p(t_1, x_1|t_0, x_0; \theta) \cdot p(t_2, x_2|t_1, x_1; \theta) \cdots p(t_N, x_N|t_{N-1}, x_{N-1}; \theta), \\ &= p_0(x_0|\theta) \prod_{k=1}^{N} p(t_k, x_k|t_{k-1}, x_{k-1}; \theta). \end{aligned} \tag{3.18}$$

Then, the value of θ that maximizes this function, say θ^*, could be found by:

$$L(\theta^*) = \max_{\theta \in \mathbb{R}^m} \; p_0(x_0|\theta) \prod_{k=1}^{N} p(t_k, x_k|t_{k-1}, x_{k-1}; \theta). \tag{3.19}$$

Because any probability $p(.) \geq 0$ holds, maximizing the equation (3.18) will be equivalent to maximize (in order to simplify calculations) :

$$\begin{aligned} \ln L(\theta) &= \ln p_0(x_0|\theta) + \ln \left(\prod_{k=1}^{N} p(t_k, x_k|t_{k-1}, x_{k-1}; \theta) \right), \\ &= \ln p_0(x_0|\theta) + \sum_{k=1}^{N} \ln p(t_k, x_k|t_{k-1}, x_{k-1}; \theta), \end{aligned}$$

or minimize (which is usually used in computers for convenience):

$$-\ln L(\theta) = -\ln p_0(x_0|\theta) - \sum_{k=1}^{N} \ln p(t_k, x_k|t_{k-1}, x_{k-1}; \theta).$$

Knowing densities might not always be the case; but now consider the equation (3.17) with $X(t_{k-1}) \equiv x_{k-1}$ at time $t \equiv t_{k-1}$. Likewise the equation (3.3), applying Euler-Maruyama will transform this equation into the form:

$$X(t_k) \approx x_{k-1} + f(t_{k-1}, x_{k-1}; \theta)\Delta t + g(t_{k-1}, x_{k-1}; \theta)\eta_k \sqrt{\Delta t}, \tag{3.20}$$

where $\eta_k \sim N(0, 1)$. Now, it is possible to take conditional expectation of this expression. For notational convenience denote $\mathbb{E}[X(t_k)|x(t_{k-1})] \equiv \mathbb{E}_t[X(t_k)]$:

$$\begin{aligned} \mathbb{E}_t[X(t_k)] &= x_{k-1} + f(t_{k-1}, x_{k-1}; \theta)\Delta t + g(t_{k-1}, x_{k-1}; \theta)\sqrt{\Delta t}\, \mathbb{E}_t[\eta_k], \\ &= x_{k-1} + f(t_{k-1}, x_{k-1}; \theta)\Delta t, \\ &= \mu_k. \end{aligned}$$

Conditional variance can also be calculated:

$$\begin{aligned} \mathbb{V}_t[X(t_k)] &= \mathbb{E}_t[\big(X(t_k) - \mathbb{E}_t[X(t_k)]\big)^2], \\ &= \mathbb{E}_t[\big(x_{k-1} + f(t_{k-1}, x_{k-1}; \theta)\Delta t + g(t_{k-1}, x_{k-1}; \theta)\eta_k \sqrt{\Delta t} \\ &\quad -(x_{k-1} + f(t_{k-1}, x_{k-1}; \theta)\Delta t)\big)^2], \\ &= \mathbb{E}_t[\big(g(t_{k-1}, x_{k-1}; \theta)\eta_k \sqrt{\Delta t}\big)^2], \\ &= g^2(t_{k-1}, x_{k-1}; \theta)\Delta t\, \mathbb{E}_t[\eta_k^2], \\ &= g^2(t_{k-1}, x_{k-1}; \theta)\Delta t, \\ &= \sigma_k^2. \end{aligned}$$

Suddenly, corresponding probability density is:

$$p(t_k, x_k | t_{k-1}, x_{k-1}; \theta) \approx \frac{1}{\sigma_k \sqrt{2\pi}} \exp\left(\frac{-(x_k - \mu_k)^2}{2\sigma_k^2}\right),$$

which is to be maximized over vector $\theta \in \mathbb{R}^m$. This optimisation problem shouldn't be perceived as negligible. It might not be easy to calculate global maxima point of this transition density by changing θ, especially when there are too many observations of X are included to calculations and/or when θ is a very large vector. Nevertheless, it could be possible to calculate equation (3.19) by using partial derivative:

$$\frac{\partial}{\partial \theta} L(\theta) = 0. \tag{3.21}$$

One should be careful if equation (3.21) produces more than one solution. When it is the case, θ^* must be verified over each possible estimates of θ say $\hat{\theta}$ [3].

This will be very useful for upcoming sections, as it enables us to construct a benchmark θ values for relatively simple functions. So, both while determining θ_0 values for optimizers and having idea about the quality of final θ^* values produced by software, one can, at least, get a little help from the equation (3.21).

3.2.2 Construction of MLE for Koijen, Rodrígez, and Sbuelz [24] Model

In order to execute a numerical example for our model, likelihood function of discretized equation system (3.16) must be constructed.[4]

$$\begin{aligned}
\mathbb{E}_t[R_{k+1}] &= \mathbb{E}_t\big[[(\mu_0 + \mu_1 x_k)(1-\phi) + \phi\, m_k]\Delta t + \sigma_{S(1)}\sqrt{\Delta t}\,\eta_k\big], \\
&= [(\mu_0 + \mu_1 x_k)(1-\phi) + \phi\, m_k]\Delta t + \sigma_{S(1)}\sqrt{\Delta t}\,\mathbb{E}_t[\eta_k], \\
&= [(\mu_0 + \mu_1 x_k)(1-\phi) + \phi\, m_k]\Delta t, \\
&= \mu_R.
\end{aligned} \tag{3.22}$$

So, variance:

$$\begin{aligned}
\mathbb{V}_t[R_{k+1}] &= \mathbb{E}_t\Bigg[\Big([(\mu_0 + \mu_1 x_k)(1-\phi) + \phi\, m_k]\Delta t + \sigma_{S(1)}\sqrt{\Delta t}\,\eta_k \\
&\qquad -\big([(\mu_0 + \mu_1 x_k)(1-\phi) + \phi\, m_k]\Delta t\big)\Big)^2\Bigg], \\
&= \mathbb{E}_t\Big[\big(\sigma_{S(1)}\sqrt{\Delta t}\,\eta_k\big)^2\Big], \\
&= \sigma_{S(1)}^2 \Delta t\, \mathbb{E}_t[\eta_k^2], \\
&= \sigma_{S(1)}^2 \Delta t.
\end{aligned} \tag{3.23}$$

[4] For notational convenience $R(t_k) \equiv R_{k+1}$ and $X(t_k) \equiv X_{k+1}$ are used.

The conditional expectation of X becomes:

$$\begin{aligned}\mathbb{E}_t[X_{k+1}] &= \mathbb{E}_t[x_k - \alpha\, x_k \Delta t + \sigma_{X(1)}\sqrt{\Delta t}\,\eta_k + \sigma_{X(2)}\sqrt{\Delta t}\,\upsilon_k],\\ &= x_k - \alpha\, x_k \Delta t + \sigma_{X(1)}\sqrt{\Delta t}\,\mathbb{E}_t[\eta_k] + \sigma_{X(2)}\sqrt{\Delta t}\,\mathbb{E}_t[\upsilon_k],\\ &= x_k(1-\alpha\Delta t),\\ &= \mu_X. \end{aligned} \tag{3.24}$$

and variance:

$$\begin{aligned}\mathbb{V}_t[X_{k+1}] &= \mathbb{E}_t\Big[\big(x_k - \alpha\, x_k \Delta t + \sigma_{X(1)}\sqrt{\Delta t}\,\eta_k + \sigma_{X(2)}\sqrt{\Delta t}\,\upsilon_k\\ &\quad -(x_k(1-\alpha\,\Delta t)\big)^2\Big],\\ &= \mathbb{E}_t\Big[\big(\sigma_{X(1)}\sqrt{\Delta t}\,\eta_k + \sigma_{X(2)}\sqrt{\Delta t}\,\upsilon_k\big)^2\Big],\\ &= \sigma^2_{X(1)}\Delta t\,\mathbb{E}_t[\eta_k^2] + \sigma^2_{X(2)}\Delta t\,\mathbb{E}_t[\upsilon_k^2] + 2\sigma_{X(1)}\sigma_{X(2)}\Delta t\,\mathbb{E}_t[\eta_k\cdot\upsilon_k],\\ &= \sigma^2_{X(1)}\Delta t + \sigma^2_{X(2)}\Delta t,\\ &= \big(\sigma^2_{X(1)} + \sigma^2_{X(2)}\big)\Delta t. \end{aligned} \tag{3.25}$$

Notice that, $\mathbb{C}ov[\eta_k, \upsilon_k] = \mathbb{E}_t[\eta_k \cdot \upsilon_k] = 0$ comes from the fact that, $\eta_k \perp \upsilon_k$(since these randomness driven by independent Brownian motions in continuous time form). Further, following this model actually requires multivariate case of the transition probability density(in this case multivariate normal) which can be written in the form for a set of normal random variables $Y_1, ..., Y_k$ [3]:

$$f(y_1, ..., y_k) = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}}\exp\Big\{-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\Big\},$$

with $y' = (y_1, ..., y_k)$, $\mu' = (\mu_1, ..., \mu_k)$, and $\Sigma = \mathbb{C}ov[Y_i, Y_j]$, where $\mu_i = \mathbb{E}[Y_i]$ and Σ is a positive, semi-definite variance-covariance matrix.

Following similar approach for our bivariate case, to calculate variance-covariance matrix, covariance of R_{k+1} and X_{k+1} is needed:

$$\begin{aligned}\mathbb{C}ov[R_{k+1}, X_{k+1}] &= \mathbb{E}\Big[\big(R_{k+1} - \mathbb{E}[R_{k+1}]\big)\big(X_{k+1} - \mathbb{E}[X_{k+1}]\big)\Big],\\ &= \mathbb{E}\big[(\sigma_{S(1)}\,\eta_k\,\sqrt{\Delta t})\cdot(\sigma_{X(1)}\sqrt{\Delta t}\,\eta_k + \sigma_{X(2)}\sqrt{\Delta t}\,\upsilon_k)\big],\\ &= \sigma_{S(1)}\sigma_{X(1)}\,\Delta t\,\mathbb{E}[\eta_k^2] + \sigma_{S(1)}\sigma_{X(2)}\,\Delta t\,\mathbb{E}[\eta_k\cdot\upsilon_k],\\ &= \sigma_{S(1)}\sigma_{X(1)}\Delta t. \end{aligned} \tag{3.26}$$

KRS used monthly returns which makes $\Delta t = 1$. So, variance-covariance matrix of the model becomes (by equations (3.23) and (3.25)):

$$\begin{aligned}\Sigma &= \begin{pmatrix}\mathbb{V}ar[R_{k+1}] & \mathbb{C}ov[R_{k+1}, X_{k+1}]\\ \mathbb{C}ov[R_{k+1}, X_{k+1}] & \mathbb{V}ar[X_{k+1}]\end{pmatrix}\\ &= \begin{pmatrix}\sigma^2_{S(1)} & \sigma_{S(1)}\sigma_{X(1)}\\ \sigma_{S(1)}\sigma_{X(1)} & \sigma^2_{X(1)} + \sigma^2_{X(2)}\end{pmatrix}. \end{aligned} \tag{3.27}$$

This matrix should be familiar. Σ wasn't chosen arbitrarily, recall $\tilde{\Sigma}$ from equation (2.60):

$$\begin{aligned}
\tilde{\Sigma} &= \begin{pmatrix} \sigma_{S(1)} & 0 \\ \sigma_{X(1)} & \sigma_{X(2)} \end{pmatrix}, \\
\Rightarrow \tilde{\Sigma}\tilde{\Sigma}' &= \begin{pmatrix} \sigma_{S(1)} & 0 \\ \sigma_{X(1)} & \sigma_{X(2)} \end{pmatrix} \cdot \begin{pmatrix} \sigma_{S(1)} & \sigma_{X(1)} \\ 0 & \sigma_{X(2)} \end{pmatrix}, \\
&= \begin{pmatrix} \sigma^2_{S(1)} & \sigma_{S(1)}\sigma_{X(1)} \\ \sigma_{S(1)}\sigma_{X(1)} & \sigma^2_{X(1)} + \sigma^2_{X(2)} \end{pmatrix} = \Sigma.
\end{aligned}$$

So, it is showed that, how Cholesky decomposition of the system constructed. From here, finding bivariate likelihood function of the model is easy for parameter vector $\theta = (\sigma_{X(1)}, \sigma_{X(2)}, \alpha, \sigma_{S(1)}, \mu_0, \mu_1, \phi)'$. First set mean part using equations (3.22) and (3.24):

$$\begin{aligned}
y &= (r_{k+1}, x_{k+1})', \\
(y - \mu) &= (r_{k+1} - \mu_R, x_{k+1} - \mu_X)' = A, \qquad (3.28)
\end{aligned}$$

where transition density is:

$$\begin{aligned}
p(k+1, r_{k+1}, x_{k+1} | k, r_k, x_k; \theta) &\approx \frac{1}{\sqrt{(2\pi)^2|\Sigma|}} \exp\left\{ -\frac{1}{2} \begin{pmatrix} r_{k+1} - \mu_R \\ x_{k+1} - \mu_X \end{pmatrix}^T \cdots \right. \\
&\quad \left. \cdots \Sigma^{-1} \begin{pmatrix} r_{k+1} - \mu_R \\ x_{k+1} - \mu_X \end{pmatrix} \right\}, \\
&\approx \frac{1}{\sqrt{(2\pi)^2|\Sigma|}} \exp\left\{ -\frac{1}{2} A^T \Sigma^{-1} A \right\}. \qquad (3.29)
\end{aligned}$$

Then likelihood function becomes[5]

$$\begin{aligned}
L(\theta) &= p_0(r_0, x_0|\theta) \cdot p(r_1, x_1 | r_0, x_0; \theta) \cdots p(r_N, x_N | r_{N-1}, x_{N-1}; \theta), \\
&= p_0(r_0, x_0|\theta) \prod_{k=1}^{N} p(r_k, x_k | r_{k-1}, x_{k-1}; \theta) \\
&= p_0(r_0, x_0|\theta) \prod_{k=1}^{N} \frac{1}{\sqrt{(2\pi)^2|\Sigma|}} \exp\left\{ -\frac{1}{2} A^T \Sigma^{-1} A \right\} \\
\Rightarrow \ln L(\theta) &= \ln\left(p_0(r_0, x_0|\theta)\right) + \sum_{k=1}^{N} \left\{ \ln\left(\frac{1}{\sqrt{(2\pi)^2|\Sigma|}} \right) \right. \\
&\quad \left. -\frac{1}{2} \begin{pmatrix} r_k - \mu_R \\ x_k - \mu_X \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} r_k - \mu_R \\ x_k - \mu_X \end{pmatrix} \right\} \\
&= \ln\left(p_0(r_0, x_0|\theta)\right) - \frac{N}{2} \ln\left((2\pi)^2|\Sigma|\right) - \frac{1}{2} \sum_{k=1}^{N} \left\{ \cdots \right.
\end{aligned}$$

[5] Assume there are $N + 1$ observations for variables. There isn't any difference between going from $k + 1$ to $N - 1$ and from k to N.

$$\cdots \quad \begin{pmatrix} r_k - [(\mu_0 + \mu_1 x_{k-1})(1-\phi) + \phi\, m_{k-1}] \\ x_k - x_{k-1}(1-\alpha) \end{pmatrix}^T \Sigma^{-1} \cdots$$
$$\cdots \quad \begin{pmatrix} r_k - [(\mu_0 + \mu_1 x_{k-1})(1-\phi) + \phi\, m_{k-1}] \\ x_k - x_{k-1}(1-\alpha) \end{pmatrix} \Bigg\} \tag{3.30}$$

This is the function to be maximized ($-\ln L(\theta)$ used for optimization in software as mentioned before.). Realize that, how huge is this function with seven[6] unknown parameters. It is not possible to derive these unknowns from partial derivative. However, we can establish some benchmark values for some of the parameters. X is an OU process with only three unknown parameters α, $\sigma_{X(1)}$, $\sigma_{X(2)}$. It also doesn't have any interaction with R and M. Even if one might find a very different result for those three coefficients in multivariate case, it is worthwhile to estimate parameters of X by using MLE at least for determining θ_0 values for the software functions.

The mean and variance of X is already known by equations (3.24)–(3.25). Therefore probability density of X, say p^X becomes for $\theta^X = (\sigma_{X(1)}, \sigma_{X(2)}, \alpha)'$:

$$\begin{aligned} p^X(x_k|x_{k-1};\theta^X) &\approx \frac{1}{\sqrt{(\sigma^2_{X(1)} + \sigma^2_{X(2)})}\sqrt{2\pi}} \exp\left(\frac{-(x_k - \mu_X)^2}{2(\sigma^2_{X(1)} + \sigma^2_{X(2)})}\right), \\ &\approx \frac{1}{\sqrt{(\sigma^2_{X(1)} + \sigma^2_{X(2)})}\sqrt{2\pi}} \exp\left(\frac{-(x_k - x_{k-1}(1-\alpha))^2}{2(\sigma^2_{X(1)} + \sigma^2_{X(2)})}\right). \end{aligned}$$

Now, construct likelihood function for X:

$$\begin{aligned} L^X(\theta^X) &= p_0^X(x_0|\theta^X) \prod_{k=1}^{N} p^X(x_k|x_{k-1};\theta^X), \\ &= p_0^X(x_0|\theta^X) \prod_{k=1}^{N} \frac{1}{\sqrt{(\sigma^2_{X(1)} + \sigma^2_{X(2)})}\sqrt{2\pi}} \exp\left(\frac{-\big(x_k - x_{k-1}(1-\alpha)\big)^2}{2(\sigma^2_{X(1)} + \sigma^2_{X(2)})}\right), \\ \Rightarrow \ln L^X(\theta^X) &= \ln\big(p_0^X(x_0|\theta^X)\big) - \frac{N}{2}\ln\big((\sigma^2_{X(1)} + \sigma^2_{X(2)})2\pi\big) \\ &\quad - \sum_{k=1}^{N} \frac{\big(x_k - x_{k-1}(1-\alpha)\big)^2}{2(\sigma^2_{X(1)} + \sigma^2_{X(2)})}. \end{aligned} \tag{3.31}$$

Now, this is not a bad function to differentiate. Starting for α:

$$\begin{aligned} \frac{\partial}{\partial\alpha}\ln L^X(\theta^X) &= -\frac{2}{2(\sigma^2_{X(1)} + \sigma^2_{X(2)})} \sum_{k=1}^{N} \big(x_k - x_{k-1}(1-\alpha)\big)x_{k-1} = 0, \\ \Rightarrow \sum_{k=1}^{N} x_{k-1}^2\alpha &= -\sum_{k=1}^{N} x_{k-1}(x_k - x_{k-1}), \end{aligned}$$

[6] $\sigma_{X(1)}, \sigma_{X(2)}, \alpha, \sigma_{S(1)}, \mu_0, \mu_1, \phi$. Standard deviations are in the matrix Σ.

$$\Rightarrow \hat{\alpha} \;=\; -\sum_{k=1}^{N} x_{k-1}(x_k - x_{k-1}) \cdot \left(\sum_{k=1}^{N} x_{k-1}^2 \right)^{-1}. \tag{3.32}$$

Then, for standard deviations:

$$\begin{aligned}
\frac{\partial}{\partial \sigma_{X(1)}} \ln L^X(\theta^X) &= -\frac{N}{4\pi(\sigma_{X(1)}^2 + \sigma_{X(2)}^2)} 2\sigma_{X(1)} 2\pi \\
&\quad + \sum_{k=1}^{N} \frac{\big(x_k - x_{k-1}(1-\alpha)\big)^2}{4\big(\sigma_{X(1)}^2 + \sigma_{X(2)}^2\big)^2} 2\sigma_{X(1)} 2 = 0, \\
\Rightarrow \frac{N\sigma_{X(1)}}{\sigma_{X(1)}^2 + \sigma_{X(2)}^2} &= \frac{\sigma_{X(1)}}{\big(\sigma_{X(1)}^2 + \sigma_{X(2)}^2\big)^2} \sum_{k=1}^{N} \big(x_k - x_{k-1}(1-\alpha)\big)^2, \\
\Rightarrow \hat{\sigma}_{X(1)}^2 + \hat{\sigma}_{X(2)}^2 &= \sum_{k=1}^{N} \big(x_k - x_{k-1}(1-\hat{\alpha})\big)^2.
\end{aligned} \tag{3.33}$$

Notice that,

$$\frac{\partial}{\partial \sigma_{X(1)}} \ln L^X(\theta^X) = \frac{\partial}{\partial \sigma_{X(2)}} \ln L^X(\theta^X).$$

Though there is not a unique solution for $\sigma_{X(1)}$ and $\sigma_{X(2)}$, it exists for $(\sigma_{X(1)}^2 + \sigma_{X(2)}^2)$. On the next chapter, first robustness of these solutions will be tested, i.e. MLE of X will be tried to solved both by partial derivative and optimisation algorithms. After consistent results appeared, these solutions would be used as benchmark for vector-autoregressive(multivariate) negative log likelihood function (3.30).

So far, theoretical framework for the model has been constructed in cooperation with methodology. Thus, numerical implication could be executed by following this chapter.

CHAPTER 4

NUMERICAL IMPLEMENTATIONS

This chapter consists of three applications. The first one is actually a descriptive example from E.Allen [1], which has numerical results on hand, to make sure that methodology mentioned in the previous chapter works well.

After the techniques are verified, constructed model will be worked with simulated data to see what happens when the performance variable ϕ is taken back from the series.

Finally, last part of this section will be the execution of the model to BIST-100 with data analysis.

4.1 A Descriptive Example

This is example is taken from the book of E.Allen [1] (pp. 120) in which a square root stochastic process used for parameter estimations by means of Euler discretization and maximum likelihood. Since, E.Allen only gives the data set, model and results, that he finds, it would be a good exercise to estimate parameters for same data set with given approaches and see if findings exactly match with the book's results.

It is a data series of whooping crane population in Aransas-Wood Buffalo between the years 1939-1985 with stochastic structure:

$$dX_t = \theta_1 X_t dt + \sqrt{\theta_2 X_t} dW_t, \;\; X_0 = 18, \tag{4.1}$$

where $\theta = [\theta_1, \theta_2]'$ is the unknown parameter vector to be estimated and unlike our model, t is defined in terms of years(not months). Then, approximate X_t by:

$$X_{k+1} = x_k + \theta_1 x_k \Delta t + \sqrt{\theta_2 x_k}\sqrt{\Delta t}\,\eta_k, \tag{4.2}$$

with $x_k \equiv X_k$, and $\eta_k \sim N(0,1)$. Expectation and variance are calculated as follows to construct likelihood function:

$$\begin{aligned} \mathbb{E}[X_{k+1}] &= x_k + \theta_1 x_k + \sqrt{\theta_2 x_k}\sqrt{\Delta t}\,\mathbb{E}[\eta_k], \\ &= x_k + \theta_1 x_k, \end{aligned} \tag{4.3}$$

because yearly data used Δt can be taken as 1.

$$\begin{aligned}\mathbb{V}[X_{k+1}] &= \mathbb{E}\big[\big(x_k+\theta_1 x_k+\sqrt{\theta_2 x_k}\eta_k-(x_k+\theta_1 x_k)\big)^2\big],\\ &= \theta_2 x_k \mathbb{E}\big[(\eta_k)^2\big],\\ &= \theta_2 x_k.\end{aligned}$$

Transition density function of X_k becomes:

$$p(x_{k+1}|x_k;\theta) \approx \frac{1}{\sqrt{(2\pi)}\sqrt{\theta_2 x_k}}\exp\left\{-\frac{\big(x_{k+1}-x_k(\theta_1+1)\big)^2}{2\theta_2 x_k}\right\},$$

with likelihood function:

$$\begin{aligned}L(\theta) &= p_0(x_0|\theta)\prod_{k=0}^{N-1}p(x_{k+1}|x_k;\theta),\\ &= p_0(x_0|\theta)\prod_{k=0}^{N-1}\frac{1}{\sqrt{(2\pi)\theta_2 x_k}}\exp\left\{-\frac{\big(x_{k+1}-x_k(\theta_1+1)\big)^2}{2\theta_2 x_k}\right\},\\ \Rightarrow \ln L(\theta_1,\theta_2) &= \ln p_0(x_0|\theta)-\sum_{k=0}^{N-1}\left\{\frac{1}{2}\ln(2\pi\theta_2 x_k)+\frac{\big(x_{k+1}-x_k(\theta_1+1)\big)^2}{2\theta_2 x_k}\right\}.\end{aligned}$$

This equation is maximized over two decision variables θ_1 and θ_2 by using both software optimizer[1] and partial derivatives. Before going through the results, provide derivative estimates:

$$\begin{aligned}\frac{\partial}{\partial\theta_1}\ln L(\theta) &= \frac{1}{2\theta_2}2\sum_{k=0}^{N-1}\frac{x_{k+1}-x_k(\theta_1+1)}{x_k}x_k=0,\\ \Rightarrow\sum_{k=0}^{N-1}\big(x_{k+1}-x_k\big) &= \theta_1\sum_{k=0}^{N-1}x_k,\\ \therefore\hat{\theta}_1 &= \sum_{k=0}^{N-1}\big(x_{k+1}-x_k\big)\left(\sum_{k=0}^{N-1}x_k\right)^{-1},\end{aligned}\tag{4.4}$$

and for θ_2:

$$\begin{aligned}\frac{\partial}{\partial\theta_2}\ln L(\theta) &= -\sum_{k=0}^{N-1}\left\{\frac{1}{2}\frac{1}{2\pi\theta_2 x_k}2\pi x_k-\frac{\big(x_{k+1}-x_k(\theta_1+1)\big)^2}{2\theta_2^2 x_k}\right\}=0,\\ \Rightarrow\frac{N}{2\theta_2} &= \frac{1}{2\theta_2^2}\sum_{k=0}^{N-1}\left\{\frac{\big(x_{k+1}-x_k(\theta_1+1)\big)^2}{x_k}\right\},\\ \therefore\hat{\theta}_2 &= \sum_{k=0}^{N-1}\left\{\frac{\big(x_{k+1}-x_k(\theta_1+1)\big)^2}{x_k}\right\}.\end{aligned}\tag{4.5}$$

[1] Actually minus version is minimized in *MATLAB* part. Codes could be found in appendix.

In the book, estimations of θ_1 and θ_2 are stated as 0.0361 and 0.0609, respectively. Three optimizers used to avoid local maximum(minimum) issues. Whenever they all give same or very similar results, one can be sure that the values are at least around of the global maximum(minimum) point. Following table lists the results from both optimizers and partial derivatives:

Table 4.1: Stochastic Parameter Estimation Results of E.Allen[1] (Example 4.15)

Optimizer	$\hat{\theta}_1$	$\hat{\theta}_2$	$-\ln L(\theta)$
fminunc	0.0361	0.6094	136.7816
fminsearch	0.0361	0.6093	136.7816
patternsearch	0.0361	0.6093	136.7816
$\frac{\partial}{\partial\theta}L(\theta)$	0.0361	0.6093	136.7816

Estimations are consistent and they manage to capture the benchmark values. Once, parameters obtained it is possible to compare actual data with simulation results. Figure (4.1) shows the motion of the actual data and 4.2 shows two of the simulated paths with mean. Mean is taken by simulating one thousand paths and taking the arithmetic average for each time t. It seems like model shows plausible fit to the data.

This section is called descriptive, as it provides a good outlook of what is done in the previous chapter and why. Even though, we are working on a far different model with compared to this example, it is valuable for showing that we can implement the methods accurately. Partial derivative method is also said to be satisfactory. It is clear that, this technique can not be used in the case of seven parameters[2]. However, it might provide benchmarks or some intuitive results for some of the variables.

By the next section, all study will again focus on our stochastic framework. Similar simulations will be performed to be able to understand and appreciate the role of ϕ coefficient and thereafter, actual Borsa İstanbul (100) index data is going to be considered. While doing these, the interpretations of this section would be very useful for simulating data and providing initial values for unknown vector θ. In addition, once results are obtained, comparisons and comments would be cited without any hesitation about implementations.

4.2 Simulating Data for KRS Model to Understand the Role of ϕ

Koijen, Rodrígez, and Sbuelz [24] finds attractive results for their parameter estimations, under the setting (2.57)–(2.59). Two of them has vital importance for this study.

First, as it is already claimed as performance variable, ϕ has larger impact on equally weighted index(with compared to value weighted index) which means higher predictability by means of past performance. Because, almost all discrete time stock momentum literature has built on equally weighted index and its dominance in stock

[2] Since our model has seven unknown parameters.

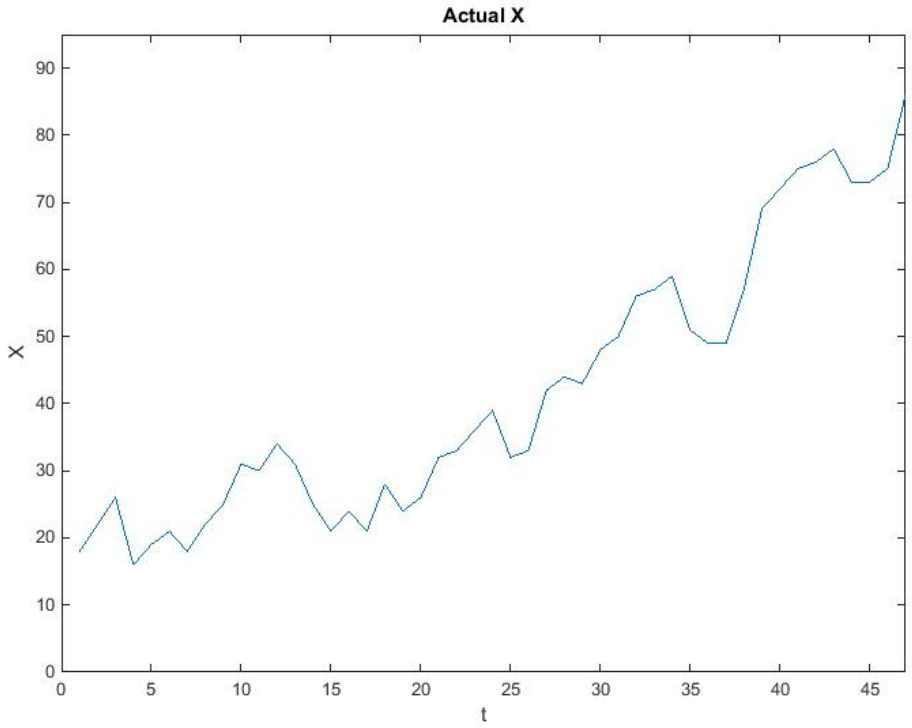


Figure 4.1: Actual X series from the data

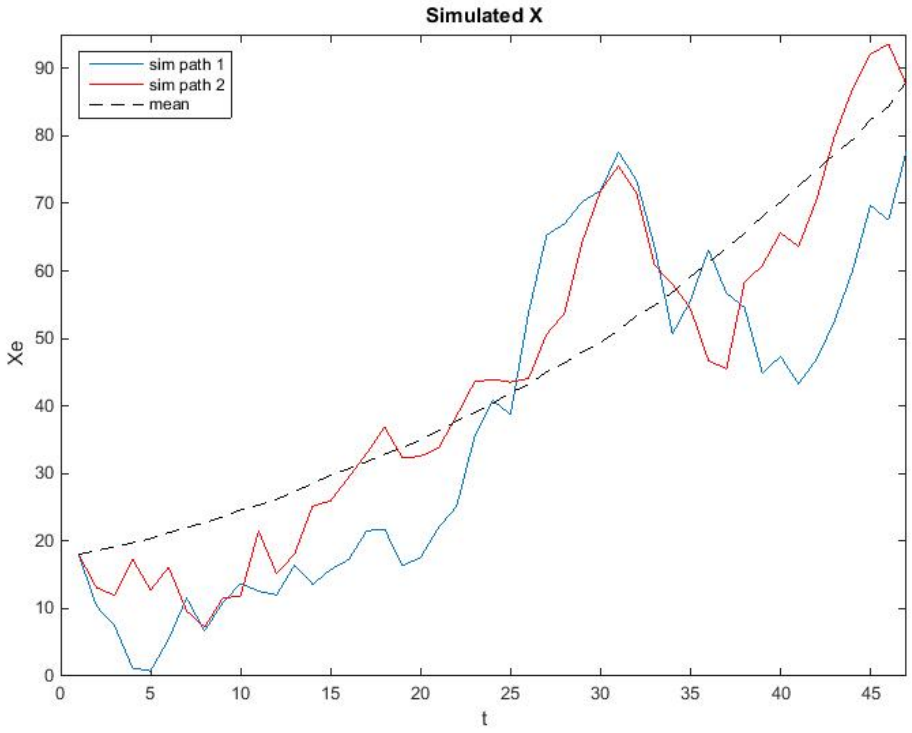


Figure 4.2: Simulated X series by parameters θ with its mean

momentum[3], this finding imposes confidence for the existence and qualitative power of ϕ.

Second, Koijen, Rodrígez, and Sbuelz showed that, autocorrelation of the returns R_t couldn't be captured in the absence of ϕ and it is captured when the performance variable is included to the model. Recall the chapter 2, the analysis of the models mainly stressed the importance of autocorrelation. Huge part of the previous study, even including behavioural models and specified portfolios(like size, industry, etc.), have made their main interpretations on the sample autocorrelation and/or have taken it as a main assumption while constructing their model.

Since, trustworthy of the methods are verified in previous section, our continuous time model could be simulated by using the parameters that are already found by Koijen, Rodrígez [24]. Simulations is going to be carried out for both equally- and value-weighted indices in the case of $\phi > 0$, i.e when performance variable exists. After that, given parameters will be estimated over the simulated data sets. Same estimation will also be done, by assuming $\phi = 0$, over same paths. It will provide another robustness check for the techniques used in this work indeed. However, the aim is actually different, observing the change of the other variables when the predictive power of the past actually exists and it is ignored. The reason why analysis is considered for all two indices is to realize if the magnitude of ϕ also has a noteworthy effect on other parameters when it is assumed to be zero while it isn't. These observation will be very useful for the upcoming section, while commenting on the findings obtained from BIST100.

Euler-Maruyama approximation is used for simulations, recall the equations (3.12)–(3.14) in the form:

$$\begin{aligned} R_{k+1} &= [\mu_k(1-\phi) + \phi\, m_k] + \sigma_{S(1)}\,\eta_k + \sigma_{S(2)}\,\upsilon_k, \\ M_{k+1} &= m_k + (1-\phi)(\mu_k - m_k) + \sigma_{S(1)}\,\eta_k + \sigma_{S(2)}\,\upsilon_k, \\ \mu_{k+1} &= \mu_0 + \mu_1\, x_{k+1}, \\ X_{k+1} &= x_k - \alpha\, x_k + \sigma_{X(1)}\,\eta_k + \sigma_{X(2)}\,\upsilon_k, \end{aligned}$$

where $\Delta t = 1$ inserted and again series placed right and side of the equations denoted with small scripts. By simply reverting the order of these processes one can obtain paths from X_t to R_t. It is obvious that, dividend yield series doesn't contain any of the other processes so, it naturally becomes initial simulation. Whenever X_t is obtained, μ_t could easily be found, it would be used to construct M_t and these two would be enough to get R_t.

As cited above, two different simulations are performed for two different, positive values of performance variable. Practically, these values stand for value- and equally-weighted indices. Parameter vector $\theta = [\phi, \mu_0, \mu_1, \sigma_{S(1)}, \alpha, \sigma_{X(1)}, \sigma_{X(1)}]'$ for each of the indexes is written in the table (4.2). Intuitively, taking back the performance variable from equally weighted index is expected to cause a larger deformation on other variables.

After the generation of these paths, each of the parameters estimated for both of the

[3] See for example, Jegadeesh and Titman[23], and Lewellen [25]

Table 4.2: Estimation Results of Koijen, Rodriguez and Sbuelz (2009)[24]

	Simulation Parameters	
	VW	EW
ϕ	0.15	0.39
μ_0 (%)	0.92	1.15
μ_1	0.016	0.012
$\sigma_{S(1)}$ (%)	5.24	6.28
α	0.011	0.0094
$\sigma_{X(1)}$ (%)	-5.77	-8.43
$\sigma_{X(2)}$ (%)	1.35	1.52

indexes by following the subsection (3.2.2).

Table 4.3: Estimation Results of Simulated Data

	Estimation Results of Simulated Data			
	Momentum and Mean-Reversion		Mean-Reversion only	
	VW	EW	VW	EW
ϕ	0.1509	0.39	$\emptyset$	$\emptyset$
μ_0 (%)	0.86	1.07	0.86	1.07
μ_1	0.0192	0.0162	0.0147	0.0037
$\sigma_{S(1)}$ (%)	5.31	6.37	5.43	7.09
α	0.0146	0.0131	0.0146	0.0131
$\sigma_{X(1)}$ (%)	-5.89	-8.58	-5.82	-7.92
$\sigma_{X(2)}$ (%)	1.32	1.48	1.58	3.62
$\hat{\alpha}$	0.0146	0.0131	0.0146	0.0131
$\hat{\sigma}^2_{X(1)} + \hat{\sigma}^2_{X(1)}$ (%)	0.36	0.76	0.36	0.76
$-\ln L(\theta)$	$-4.4271^{\times 10^3}$	$-4.1274^{\times 10^3}$	$-4.2253^{\times 10^3}$	$-3.1274^{\times 10^3}$

(4.3) where symbols denoted with hat shows the derivative check of estimations according to univariate likelihood function of X_t. Like mentioned before, these values might not be equal to estimates driven by multivariate likelihood function. It is already not possible to provide precise benchmarks for $\sigma_{X(1)}$ and $\sigma_{X(2)}$ because of the reasons mentioned in (3.2.2).

Before coming to the effects of performance variable, estimation power of MLE is considered. First of all, for each of the cases ϕ predictions seem quite fine. On the other hand, there are gaps between the real and estimated values of some variables

which might be considered negligibly small. The source of these gaps could be understood better by comparing real α and $\hat{\alpha}$. Because vector-autoregressive model based prediction of this parameter is exactly match with $\hat{\alpha}$, this error is most probably caused by MLE. For instance, $\sigma^2_{X(1)} + \sigma^2_{X(2)}$ of value weighted index in the table 4.3 equals to 0.00364345 consistent with $\hat{\sigma}^2_{X(1)} + \hat{\sigma}^2_{X(1)}$ value 0.0036. So, in this sense, accuracy could be preserved with the values $\sigma_{X(1)} = 0.0589$ and $\sigma_{X(2)} = 0.0132$ as well. This also explains why multiple optimization functions used for these implementations. Nevertheless, our model is able to tolerate those kind of errors because of its focus on state variables(M_t and μ_t). Further, a good analysis requires multiple descriptive tests to support findings which might also compensate small biases.

The impact of the performance variable is quite observable from table (4.3). Take $\sigma_{S(1)}$, for example, when actually existed ϕ value erased from the system, variation of return necessarily increases for each of the indices. This increase is even more prominent in equally weighted index which signals that, magnitude of change would be proportional to significance of momentum effect. Another interesting observation might be cited about μ_0 and μ_1. First, as one can easily notice, there is exactly no change on μ_0 which could be expected because of its role in the equation system(intercept point). On the other hand, μ_1 experiences a necessary decrease when performance variable is ignored which is closely related to calibration process of μ_t. Because μ_1 is the coefficient of demeaned dividend yields X_t, when $\phi = 0$, mean reversion effect automatically becomes only state of the return equation. This sudden domination is, in fact, synthetic and decreased coefficient of dividend yield is a reaction like increase in return variation to compensate ignorance of the return continuation. In other words, even though, mean reversion seems to have more remarkable role on returns in the absence of ϕ, its indicator's(X_t) value is decayed by μ_1. It is noteworthy that, proportion of this change is again closely related to momentum level.

Meanwhile, α is another unaffected variable. The hatted values inherently can't change; but it is seemingly the case for other variables that drive the X_t process(in multivariate estimation case) as well. It is not only valid for α; but also for $\sigma_{X(1)}$ and $\sigma_{X(1)}$. Even though, their individual values change, $\sigma^2_{X(1)} + \sigma^2_{X(2)}$ is around $3.64^{\times 10^{-3}}$ for value weighted index in the case of $\phi = 0$. So, most probably these dividend yield variations are adjusted according to change in $\sigma_{S(1)}$, such that value of their sum of squares preserved. Since, mean reversion process doesn't contain any of the other time series by construction, these situations should be explainable.

Finally, the objective function, which is minus log likelihood function for software issues in this case, shows tendency to move away from its minimum point. These loss of optimality is valid for each of the indices and it's much larger for equally weighted index in which momentum is more significant.

Though, parameter based analysis is informative, there is also another side of the medallion. What is the real risk exposure of ignoring an actually existed past performance? In fact, disregarding the momentum effect leads misunderstanding of the data behaviour.

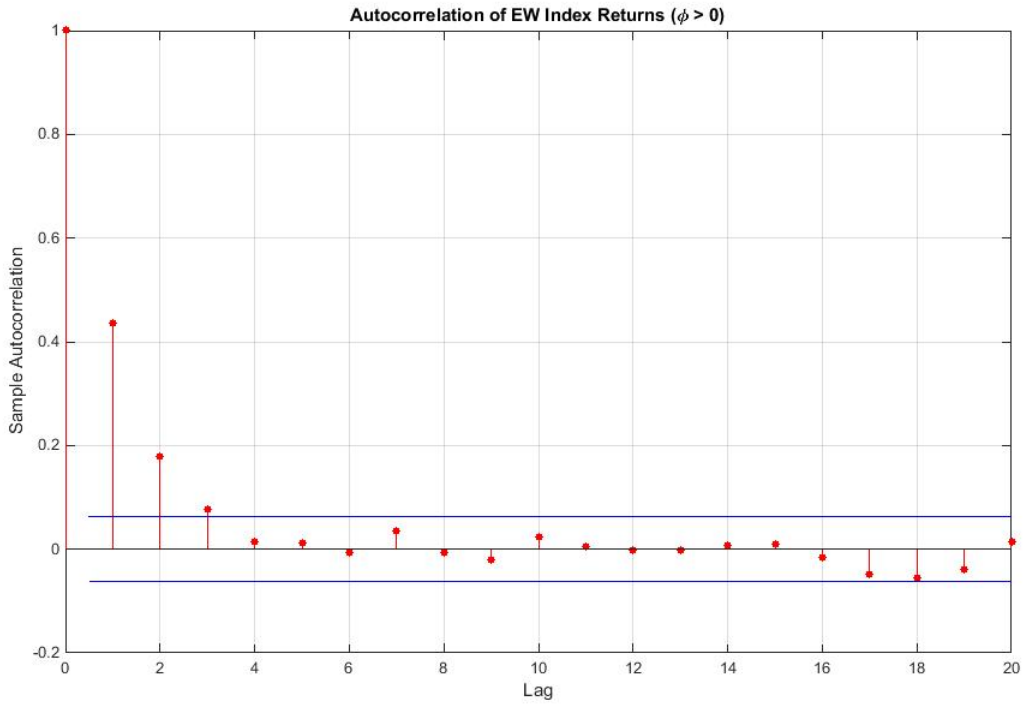


Figure 4.3: Autocorrelation Function of EW Index Returns ($\phi = 0.39$)

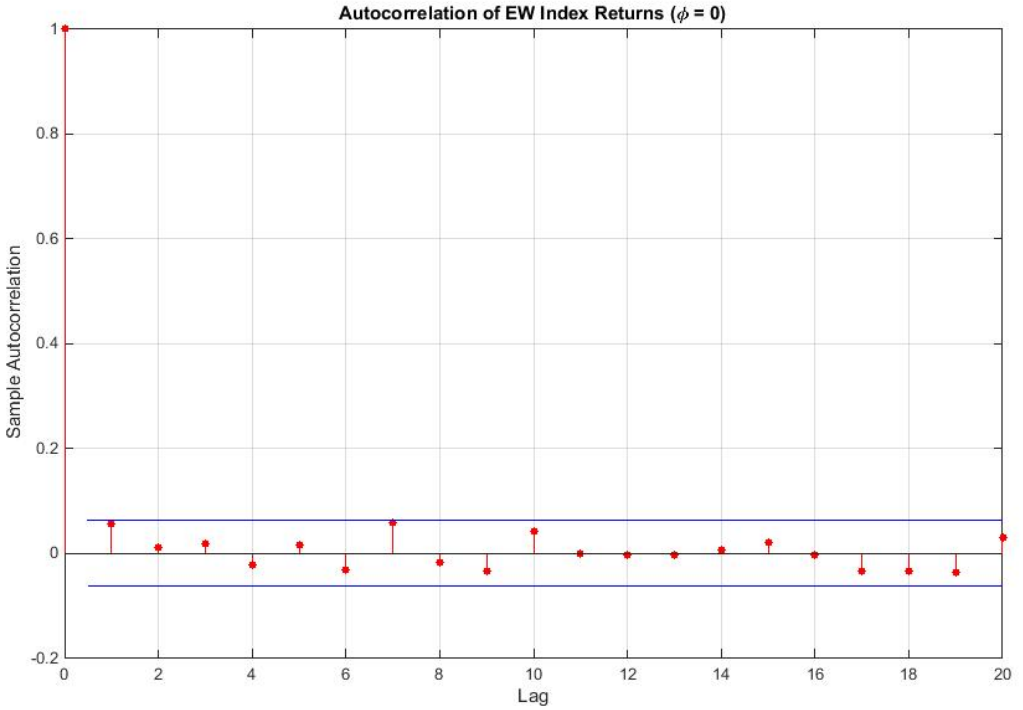


Figure 4.4: Autocorrelation Function of EW Index Returns ($\phi = 0$)

To see this, consider autocorrelation function of EW Index[4] when $\phi > 0$ and $\phi = 0$ in (4.3)–(4.4). Notice the vital difference between these autocorrelation functions. If momentum is not taken into account, returns are assumed to be uncorrelated which is not true. Figure (4.3) exhibits positive autocorrelation in returns until lag four whereas, figure (4.4) denies the return continuation completely.

Hence, simulations and estimations in this section showed that imposing ϕ to our system of equations generates memory to returns. Erasing the coefficient of performance variable takes the past predictability of returns away by making them uncorrelated. By the next section processes will be executed to a actual data and behaviour of BIST-100 is going to be analysed.

4.3 Empirical Analysis of BIST-100

The intuition behind momentum and mean reversion effects, how they are imposed into our model, and role of parameters have been discussed so far. For the empirical analysis BIST-100(XU100) data considered. Before going through the application results, description of constructed series will be explained.

Raw data is taken as monthly closing price and dividend payments of BIST-100 for Jan. 2004–Dec. 2014 from Bloomberg Data Center. Then, data calibrations are done in line with Koijen, Rodrígez, and Sbuelz as mentioned.

Accordingly, R_t was simply calculated as ordinary returns by following the left hand side of the equation (3.12). By using returns, M_t was calculated as (2.56) with a simple loop on *MATLAB*. For X_t, like mentioned in equation (2.55), D'_t was taken as index value of month t divided by corresponding year's total dividend payments. Rather than natural logarithm, Box-Cox transformation is used to normalize data and get D_t. Normality of these data series is verified by Q-Q plots, and variety of tests which could be found in appendices.

Figure 4.5 shows the series R_t and M_t. BIST-100 return seems to fluctuate around a value slightly higher than zero. It has mean 0.0149 (1.49%) which is a considerably high monthly return and standard deviation 0.0858 (8.58%). So, high return seems to occur in exchange of high risk. The patterns of return and momentum series are very similar probably because of the dominant recent past impact on M_t (by construction, see equation (2.56)). Momentum series is also smoother than return series with mean 0.0086 (0.86%) and standard deviation 0.0339 (3.39%). Recall the equation (3.16), we are interested with the lead lag relation between these series which makes their current pattern correlation irrelevant.

On the other hand, figure (4.6) shows the impact of applying Box-Cox transformation to demeaned dividend yield series. Even though their difference is visible, if one takes a closer look, she may notice the pattern similarity of these series. Box-Cox transformation causes a sharp increase in variation (check the y axes limits), but data behaviour

[4] Autocorrelations are taken for EW Index to be able to see the impact of ignoring ϕ better, since its value is greater for this index. See appendices for autocorrelation functions of VW Index.

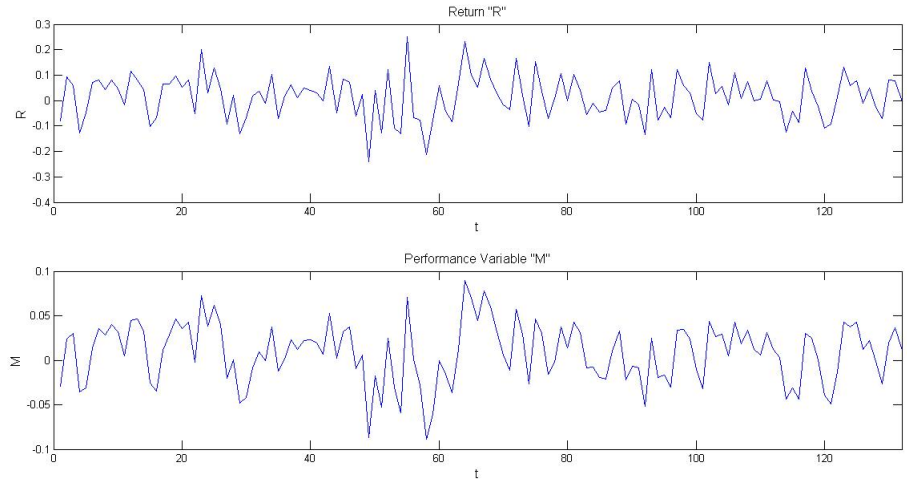


Figure 4.5: Plots of BIST-100, series R_t and M_t with μ_R=1.49%, σ_R=8.58% and μ_M=0.86%, σ_M=3.39%

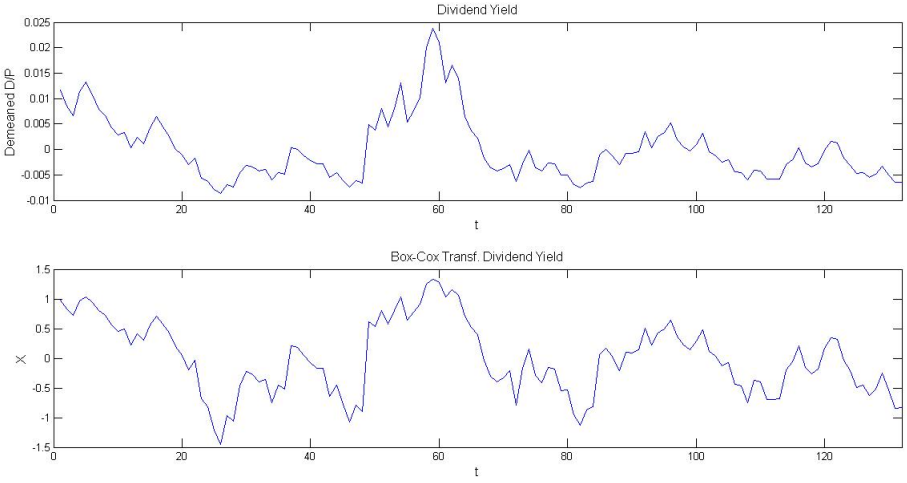


Figure 4.6: Plots of BIST-100, Demeaned Dividend Yields Comparison (Before and After Box-Cox Transformation)

is significantly preserved for even very small shocks. It is a reasonable price to satisfy normality assumption.

After generating R_t, M_t, and X_t, constructing maximum likelihood estimations is rather easy by following the previous section. Again, two different predictions are performed: for $\phi > 0$ and $\phi = 0$. However, this time Value Weighted Index isn't taken into account since it doesn't exist for BIST-100. So, each of the cases only includes Equally Weighted Index data. Table (4.4) shows the estimation results.

Table 4.4: Estimation Results of BIST-100 over the period Jan. 2004–Dec. 2014

	Estimation Results of BIST100	
	Momentum and Mean-Reversion	Mean-Reversion only
	VW	VW
ϕ	0.0006	$\emptyset$
μ_0 (%)	1.23	1.23
μ_1	0.0227	0.0227
$\sigma_{S(1)}$ (%)	8.4	8.4
α	0.1199	0.1199
$\sigma_{X(1)}$ (%)	-23.34	-23.34
$\sigma_{X(2)}$ (%)	16.22	16.22
$\hat{\alpha}$	0.1199	0.1199
$\hat{\sigma}^2_{X(1)} + \hat{\sigma}^2_{X(1)}$ (%)	8.08	8.08
$-\ln L(\theta)$	-192.4638	-192.4637

Results are very interesting; but not surprising indeed. The performance variable coefficient ϕ is negligible. It's insignificant by magnitude and its ignorance also does not change anything at all. This finding actually coincides with what Bildik and Gülay [5] claimed, short-term mean reversion effect. In fact, their study encompasses the years between 1991-2000; but their conclusion for this dominance is country-specific factors which might be the case and reason for endurance of this effect. So, both this study and paper of Koijen, Rodrígez, and Sbuelz [24] report consistent results with past literature for ϕ value. Beside the previous works, autocorrelation function of BIST-100 returns would be informative as well. Figure (4.7) is in line with expectations, just like mentioned in the previous section. Except an outlier in lag ten, returns of BIST-100 is uncorrelated, which means past performance doesn't have predictive power over future. Thus, collection of these findings imposes confidence about existence and descriptive power of ϕ.

Moreover, rather than constructing specified portfolios (decile, industry, size, etc.), presence of momentum and mean reversion effects in a stock market might be anticipated accurately by only analysing equity index of that market. So, one of the future studies might be cited as to provide this ϕ coefficient for some of the most popular

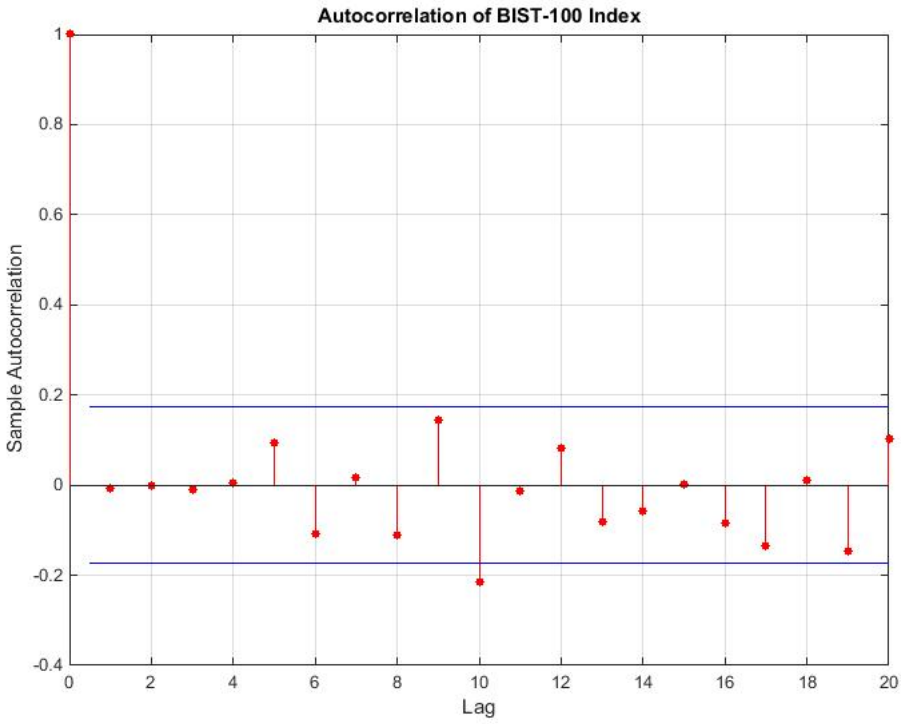


Figure 4.7: Autocorrelation Function of BIST-100 Index

indices in accordance with their literature and show more powerful evidences for existence and accuracy of it.

Another remarkable finding is the huge variation terms of BIST-100 with compared to US indices(Koijen, Rodrígez, and Sbuelz[24]). Bildik and Gülay also refers this high volatility and even attributed short-term mean reversion effect to it. However, source(s) of the differences on term structure of return reversals in these two distinct markets could be stated as another future work.

CHAPTER 5

CONCLUSION AND OUTLOOK

A vast number of studies provide evidences for existence of momentum and contrarian effects in international markets as well as US financial market. Since presence of momentum indicate return predictability, this effect challenges the efficient market hypothesis. Most of the previous studies examine investment strategies to recognize whether a capital market, an industry, or a set of national equity indices exhibit mean-reversion and return continuation, which might require an extensive research and data handling efforts. Also, considering momentum as a factor of return, rather than trying to explain it with common factors might provide a different outlook in this area.

This thesis adopts a recently developed continuous time model suggested by Koijen, Rodrígez, and Sbuelz[24] to the BIST-100 equity index between years 2004-2014. Our findings show that momentum effect is negligibly small($\phi = 0.0006$) while mean-reversion is considerably prominent(μ_t drives the changes in R_t) in BIST-100 index. There is no such evidence contradicts our assertion in the literature. In contrast, Bildik and Gülay[5] manages to generate abnormal returns by contrarian trading strategy(long past losers and short past winners) in Turkish stock market. This observation is in line with what Koijen, Rodrígez, and Sbuelz finds using CRSP based US data: existence of momentum in both VW(Value Weighted) and EW(Equally Weighted) indices, but more prominent in EW index where relative strength strategy(long past winners and short past losers) achieves abnormal returns. Hence, equity indices might be informative about return series behaviour of the stocks listed on that index, by means of momentum and mean-reversion. For example, a sufficiently large ϕ value for an index could signal the profitability of relative strength strategy at stock level.

Before going through our continuous time model, Chapter (2) draws the quantification of momentum as well as some of the fundamental discrete time models that are proposed on different dates. We observe substantial similarities between these models not only by construction but also by superiorities and weaknesses. They are flexible, for instance, a positively correlated return model can be easily converted into a negatively correlated one by adjusting the noise terms. However, they might not always provide the full picture since these models mostly result with a common profit decomposition.

Performing parameter estimation also requires knowledge of some methods which are considered in Chapter (3). Accordingly, the model is transformed from stochastic differential form to vector autoregressive(VAR) structure by means of Euler-Maruyama

discretization as it is shown that this scheme provides fair approximations for SDEs that are alike our model. Similarly, in Chapter (4), we exemplify stochastic parameter estimation with simulated data for the model and notice that positive ϕ values simply imply positive autocorrelation. Furthermore, if momentum exists and it is ignored, return shock($\sigma_{S(1)}$) reacts with an instant increase with respect to magnitude of ϕ.

This thesis can be extended for both theoretical and practical purposes. One possible future study can be cited as investor implications of the model on various equity indices to examine validity of the relation between return patterns of stocks and corresponding equity indices. Another issue is that, our model only relies on time series predictability, which might not always be the source of momentum(for example, negative cross serial correlation might be a source [25]). So, relaxing this assumption and making model applicable to individual stocks might be another possible future work. Finally, past predictability of the model could be supplied by fractional Brownian motions which might increase the complexity as well.

REFERENCES

[1] E. Allen, *Modeling with Itô stochastic differential equations*, volume 22, Springer Science & Business Media, 2007.

[2] G. Aras and M. K. Yilmaz, Price-earnings ratio, dividend yield, and market-to-book ratio to predict return on stock market: Evidence from the emerging markets, Journal of Global Business and Technology, 4(1), pp. 18–30, 2008.

[3] L. J. Bain and M. Engelhardt, *Introduction to probability and mathematical statistics*, volume 4, Duxbury Press Belmont, CA, 1992.

[4] N. Barberis, A. Shleifer, and R. Vishny, A model of investor sentiment, Journal of financial economics, 49(3), pp. 307–343, 1998.

[5] R. Bildik and G. Gülay, Profitability of contrarian strategies: Evidence from the Istanbul stock exchange*, International Review of Finance, 7(1-2), pp. 61–87, 2007.

[6] V. Binsbergen, H. Jules, M. W. Brandt, and R. S. Koijen, Optimal decentralized investment management, The Journal of Finance, 63(4), pp. 1849–1895, 2008.

[7] W. F. Bondt and R. Thaler, Does the stock market overreact?, The Journal of finance, 40(3), pp. 793–805, 1985.

[8] J. Y. Campbell, Y. L. Chan, and L. M. Viceira, A multivariate model of strategic asset allocation, Journal of financial economics, 67(1), pp. 41–80, 2003.

[9] J. Y. Campbell and L. Viceira, Consumption and portfolio decisions when expected returns are time varying, Quarterly Journal of Economics, 114(02), pp. 433–495, 1999.

[10] M. M. Carhart, On persistence in mutual fund performance, The Journal of finance, 52(1), pp. 57–82, 1997.

[11] K. Chan, A. Hameed, and W. Tong, Profitability of momentum stragegies in the international equity markets, Journal of Financial and Quantitative Analysis, 35(02), pp. 153–172, 2000.

[12] J. Chen and H. Hong, Discussion of" momentum and autocorrelation in stock returns", Review of Financial Studies, pp. 565–573, 2002.

[13] L. Chen, On the reversal of return and dividend growth predictability: A tale of two periods, Journal of Financial Economics, 92(1), pp. 128–151, 2009.

[14] A. C. Chui, S. Titman, and K. J. Wei, Individualism and momentum around the world, The Journal of Finance, 65(1), pp. 361–392, 2010.

[15] J. H. Cochrane, The dog that did not bark: A defense of return predictability, Review of Financial Studies, 21(4), pp. 1533–1575, 2008.

[16] E. F. Fama, Efficient capital markets: A review of theory and empirical work*, The journal of Finance, 25(2), pp. 383–417, 1970.

[17] E. F. Fama and K. R. French, Dividend yields and expected stock returns, Journal of financial economics, 22(1), pp. 3–25, 1988.

[18] E. F. Fama and K. R. French, Permanent and temporary components of stock prices, The Journal of Political Economy, pp. 246–273, 1988.

[19] E. F. Fama and K. R. French, Common risk factors in the returns on stocks and bonds, Journal of financial economics, 33(1), pp. 3–56, 1993.

[20] E. F. Fama and K. R. French, Multifactor explanations of asset pricing anomalies, The journal of finance, 51(1), pp. 55–84, 1996.

[21] E. F. Fama and K. R. French, Size, value, and momentum in international stock returns, Journal of financial economics, 105(3), pp. 457–472, 2012.

[22] H. Hong and J. C. Stein, A unified theory of underreaction, momentum trading, and overreaction in asset markets, The Journal of Finance, 54(6), pp. 2143–2184, 1999.

[23] N. Jegadeesh and S. Titman, Returns to buying winners and selling losers: Implications for stock market efficiency, The Journal of Finance, 48(1), pp. 65–91, 1993.

[24] R. S. Koijen, J. C. Rodriguez, and A. Sbuelz, Momentum and mean reversion in strategic asset allocation, Management Science, 55(7), pp. 1199–1213, 2009.

[25] J. Lewellen, Momentum and autocorrelation in stock returns, Review of Financial Studies, 15(2), pp. 533–564, 2002.

[26] A. W. Lo and A. C. MacKinlay, Stock market prices do not follow random walks: Evidence from a simple specification test, Review of financial studies, 1(1), pp. 41–66, 1988.

[27] A. W. Lo and A. C. MacKinlay, When are contrarian profits due to stock market overreaction?, Review of Financial studies, 3(2), pp. 175–205, 1990.

[28] T. J. Moskowitz and M. Grinblatt, Do industries explain momentum?, The Journal of Finance, 54(4), pp. 1249–1290, 1999.

[29] J. J. Murphy, *The visual investor: how to spot market trends*, volume 443, John Wiley & Sons, 2009.

[30] J. Rodriguez and A. Sbuelz, Understanding and exploiting momentum in stock returns, Advances in Corporate Finance and Asset Pricing. Elsevier Science, pp. 485–504, 2006.

[31] A. Sangvinatsos and J. A. Wachter, Does the failure of the expectations hypothesis matter for long-term investors?, The Journal of Finance, 60(1), pp. 179–230, 2005.

[32] Ö. Uğur, *An Introduction to Computational Finance*, volume 1, Imperial College Press, 2009.

[33] J. A. Wachter, Portfolio and consumption decisions under mean-reverting returns: An exact solution for complete markets, Journal of financial and quantitative analysis, 37(01), pp. 63–91, 2002.

[34] J. W. Wilder, *New concepts in technical trading systems*, Trend Research, 1978.

APPENDIX A

DESCRIPTIVE STATISTICS AND DATA CALIBRATION

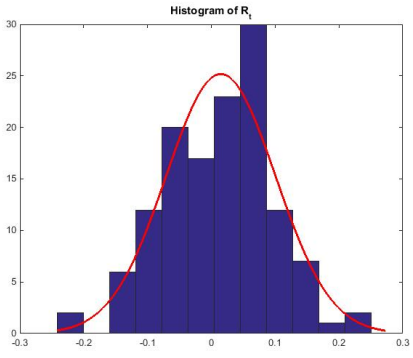
- Table of descriptive statistics of the series R_t, M_t, and X_t[1]: where Jarque-Bera

Table A.1: Table of Descriptive Statistics and Normality Tests

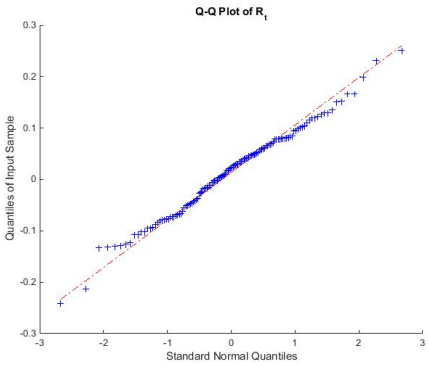
Series	Mean	Std. Dev.	Skewness	Kurtosis	JB-Test	AD-Test	KS-Test
R_t	0.0149	0.0858	-0.1019	3.1336	0	0	0
M_t	0.0086	0.0339	-0.3598	2.9074	0	1	0
X_t	0.0000	0.6130	0.0889	2.3038	0	0	0

and Anderson-Darling normality tests have null hypothesis that, sample data coming from normal distribution and One Sample Kolmogorov Simirnov has same hypothesis for standard normal distribution. The result "1" means rejecting null hypothesis with % 5 significance level for each of the tests. Because One Sample Kolmogorov Simirnov tests standard normality, $z = \frac{x-\mu}{\sigma}$ transformation applied to the data before performing the test.

Only M_t failed to pass AD-Test, probably because of its skewness which is high in absolute value with compared to R_t and X_t. Nevertheless, its normality verified by two other tests and rest of the data sets managed to pass all three tests. Also, histograms and Q-Q plots of these vectors are informative:

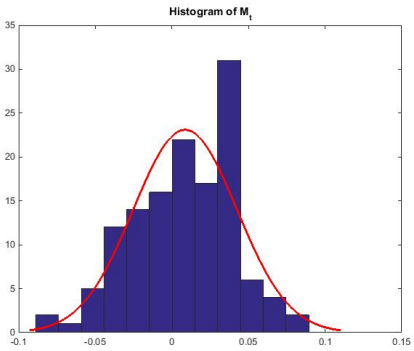


(a) Histogram of R_t

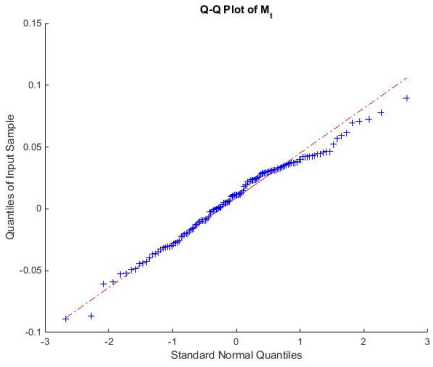


(b) Q-Q Plot of R_t

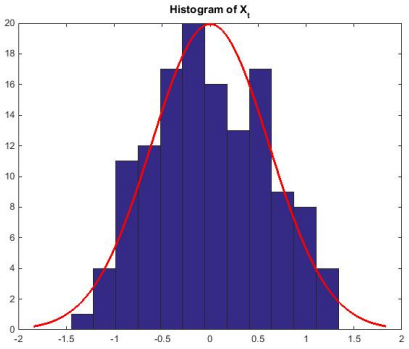
[1] JB-Test: Jarque-Bera Test, AD-Test: Anderson-Darling Test, KS-Test: One Sample Kolmogorov Simirnov Test



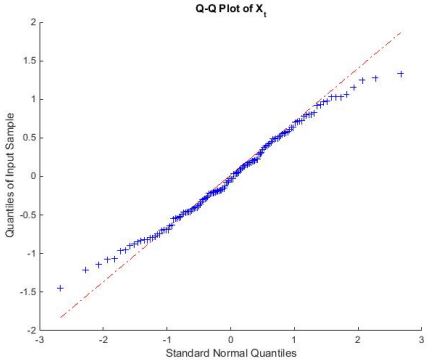
(a) Histogram of M_t



(b) Q-Q Plot of M_t



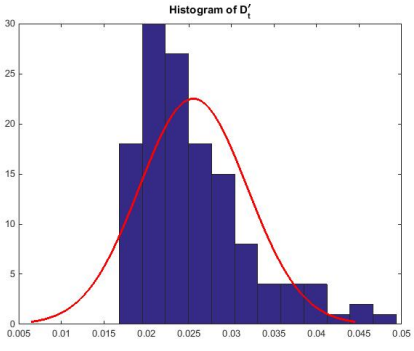
(a) Histogram of X_t



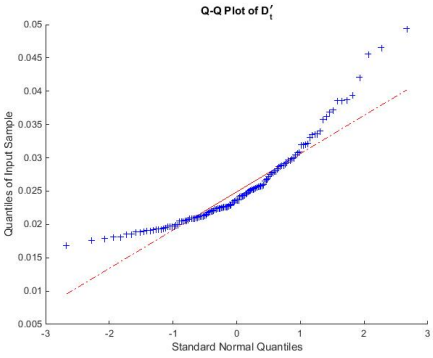
(b) Q-Q Plot of X_t

Therefore, it is seemingly safe to use Brownian Motions to define these variables.

- To see why Box-Cox transformation applied while generating X_t, consider the histograms and Q-Q plots of D'_t and its transformed version D_t[2]



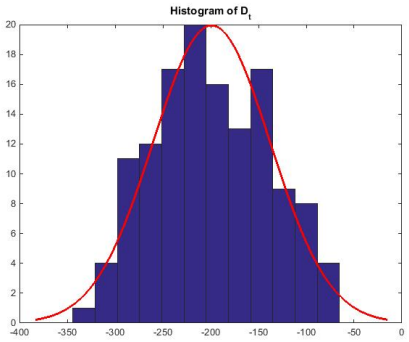
(a) Histogram of D'_t



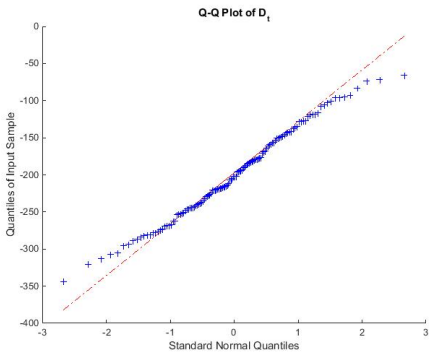
(b) Q-Q Plot of D'_t

D'_t also fails to pass any of the normality tests!

[2] Note that X_t is demeaned version of D_t, they are not same data sets.



(a) Histogram of D_t



(b) Q-Q Plot of D_t

As mentioned before, dividend yield series are usually really hard to handle data sets. Fortunately, this transformation method managed calibrate data successfully.

APPENDIX B

OTHER FIGURES AND TABLES

This chapter is organized for the tables and figures that are not shown inside of the thesis because of their similarity to shown tables.

- Autocorrelation functions of Value Weighted Index constructed by simulations in the section 4.2 has autocorrelation functions for different ϕ values (B.1–B.2).

  It is already shown that ϕ (presence of momentum) imposes return predictability into our model by means of positive autocorrelation. Because for the case $\phi = 0.39$ return series has positive autocorrelation vanishes by lag 4 (see the figure 4.3), we expect it to be smaller for this case and figure (B.1) satisfies this expectation. The more prominent momentum is the more predictable returns arise. Finally, even if the autocorrelation is only valid for the first lag (in this case), ignoring the return persistence again causes us to misunderstand the data behaviour.

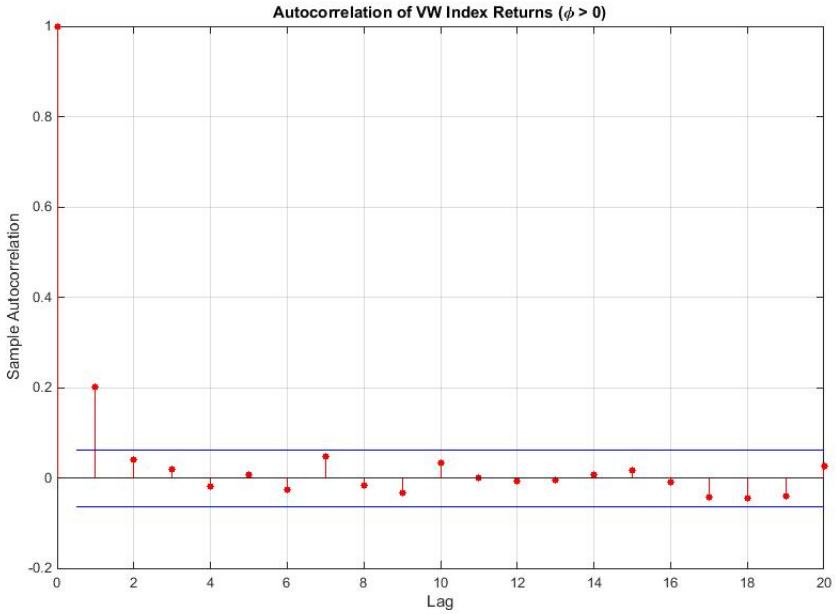


Figure B.1: Autocorrelation Function of VW Index Returns ($\phi = 0.15$)

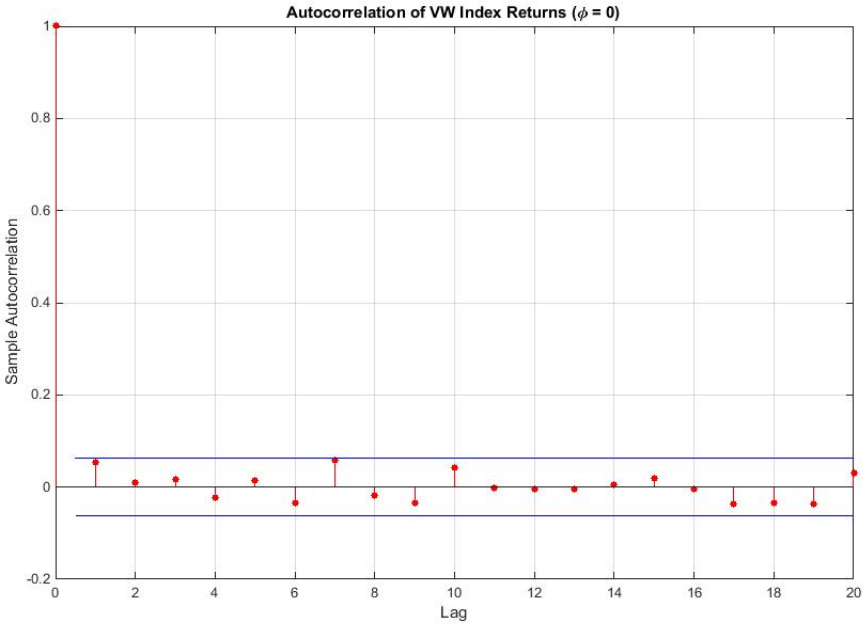


Figure B.2: Autocorrelation Function of VW Index Returns ($\phi = 0$)

APPENDIX C

MATLAB CODES

- To execute simulations in 3.1 and 3.3, following codes are used(Black-Scholes stock prices):

```
% dS = mu*S*dt + sigma*S*dW, S_0 = 1
rng(13,'v5normal');
T = 1;  mu = 0.1; sigma = 0.5; S0 = 1;
errorsM = zeros(6,2); Vrnd = randn(1,2^13);
for i = 8:13
    N = 2^i; dt = T/N;
    DW = zeros(1,N); S = zeros(1,N+1); S(1) = S0;
    Xm = [S0,zeros(1,N)]; Y = [S0,zeros(1,N)];
    for j = 1:N
        dW = sqrt(dt)*Vrnd(j);
        DW(j) = dW;
        % calculate coefficients
        a = mu*S(j); b = sigma*S(j);
        ap = mu*Y(j); bp = sigma*Y(j);
        %Euler-Maruyama
        S(j+1) = S(j) + a*dt + b*dW;
        %Milstein
        Y(j+1) = Y(j) + ap*dt + bp*dW + ...
         Y(j)*0.5*sigma^2*dt*((dW/sqrt(dt))^2-1);
        Xm(j+1) = S0*exp((mu-sigma^2/2)...
         *j*dt+sigma*sum(DW));
    end
    % for mean square errors
    errors = (Xm-Y).^2; errors2 = (Xm-S).^2;
    errorsM(i-7,1:2) = ...
     [mean(errors), mean(errors2)];
end
plot(0:dt:T,Y,'b'), hold on
plot(0:dt:T,Xm,'r--'),
legend('S from SDE','S from solution'),
title('B-S Stock Prices from SDE vs Solution'),
hold off;
display(Xm(end)); display(S(end)); display(Y(end));
```

- For OU prices:

```
% dX = -gamma*X*dt + sigma*dW, X_0 = 1
rng(10,'v5normal');
T = 1; N = 2^8; dt = T/N; gamma = 0.1; sigma = 0.5; X0 = 1;
X = [X0,zeros(1,N)]; Xsde = zeros(1,N+1); Xsde(1) = X0;
DW = zeros(1,N); stochInt = zeros(1,N);
for j = 1:N
    dW = sqrt(dt)*randn;
    W = cumsum(DW);
    a = -gamma*Xsde(j); b = sigma; % calculate coefficients
    Xsde(j+1) = Xsde(j) + a*dt + b*dW;
    stochInt(j) = exp(gamma*j*dt)*dW;
    X(j+1) = X0*exp(-gamma*j*dt) + ...
     sigma*exp(-gamma*j*dt)*sum(stochInt);
     % integral discretized as a sum
end
plot(0:dt:T,Xsde,'b'), hold on
plot(0:dt:T,X,'r--'), legend('Xsde','X'),
title('SDE vs Soln for OU Process'), hold off;
```

- For parameter estimations and calculations in 4.1 following codes are used:

```
data = xlsread('ornek'); % take data from an excel sheet
X = data(:,2); N = length(data); dt = 1;
objfun = @(theta) mlfornek(theta, X, N, dt);
theta0 = [-1; 2];
[theta, fv] = fminsearch(objfun, theta0),
[theta2, fv2] = fminunc(objfun, theta0),
[theta3, fv3] = patternsearch(objfun, theta0)
```

where mlfornek is the likelihood function of the form:

```
function f = mlfornek(theta, X, N, dt)
theta1 = theta(1);
theta2 = theta(2);
f = 0;
for j = 2:N % constructed -log Likelihood Function
    f = f + .5*log(2*pi*dt*theta2*X(j-1)) + ...
        (X(j)-X(j-1)-theta1*X(j-1)*dt)^2/(2*dt*theta2*X(j-1));
end
end
```

Partial derivative results and path simulations are performed by the following script:

```
% Use partial derivative to estimate data
% Simulate paths by obtained parameters
```

```
rng(13,'v5normal')
data = xlsread('ornek'); % take data from an excel sheet
X = data(:,2); N = length(data); dt = 1;
% sum partitions
t1 = zeros(1,N-1); t2 = zeros(1,N-1); t3 = zeros(1,N-1);

for j = 2:N % find optimal theta1 and theta2 by ML derivative
    t1(j-1) = X(j)-X(j-1);
    t2(j-1) = X(j-1)*dt;
    t3(j-1) = ((X(j)-X(j-1)-theta1hat*X(j-1)*dt)^2)/X(j-1);
end
theta1hat = 1/sum(t2)*sum(t1)
theta2hat = sum(t3)*(1/(N-1))
% Make simulation and compare with the actual data
Xe = zeros(N,1000); Xe(1,:) = 18;
for i = 1:N-1
    Xe(i+1,:) = Xe(i,:) + theta1hat*Xe(i,:) +...
        sqrt(theta2hat*Xe(i))*randn(1,1000);
end
m = mean(Xe,2);
figure(1), plot(X), xlim([0 47]), ylim([0 95]),
xlabel('t'), ylabel('X'), title('Actual X')
figure(2), plot(Xe(:,157)), xlim([0 47]), ylim([0 95]),
hold on, plot(Xe(:,80),'r'), plot(m,'k--'),
xlabel('t'), ylabel('Xe'), title('Simulated X')
legend('sim path 1','sim path 2','mean')
```

- For simulations and parameter estimations performed in the section 4.2:

```
% Simulate paths according to KRS model for both Indexes
% Lock or unlock related rows of values for simulations
% Estimate parameters by means of MLE
clear all, close all,
rng(13,'v5normal'); N = 1000;
% EW Index Returns where phi > 0
alpha = 0.0094; sigmaX1 = -0.0843; sigmaX2 = 0.0152;
phi = 0.39; mu0 = 0.0115; mu1 = 0.012; sigmaS1 = 0.0628;
% phi = 0
% VW Index Returns where phi > 0
% alpha = 0.011; sigmaX1 = -0.0577; sigmaX2 = 0.0135;
% phi = 0.15; mu0 = 0.0092; mu1 = 0.016; sigmaS1 = 0.0524;
% phi = 0
Z1 = randn(N,1); Z2 = randn(N,1);
X0 = 0; M0 = 0; R0 = 0;
mu = zeros(N,1); X = [X0; zeros(N-1,1)];
M = [M0; zeros(N-1,1)]; R = [R0; zeros(N-1,1)];
% Simulate X, mu, M, and R
for i = 1:N-1
```

```
    X(i+1) = (1-alpha)*X(i) + sigmaX1*Z1(i) + sigmaX2*Z2(i);
    mu(i) = mu0 + mu1*X(i);
    M(i+1) = M(i) + (1-phi)*(mu(i)-M(i)) + sigmaS1*Z1(i);
    R(i+1) = (1-phi)*mu(i) + phi*M(i) + sigmaS1*Z1(i);
end
% check if you can obtain parameters
objfun = @(theta) mlf3(theta, R, M, X, N, 1);
% for phi = 0 case:
% objfun = @(theta) mlf3_NoPhi(theta, R, M, X, N, 1);
theta0 = [-0.1104, 0.0283, -0.0672, 0.0747, ...
0.0105, 0.1004, 0.0165]';
% for phi = 0 case:
% theta0 = [-0.1104, 0.0283, -0.0672, ...
0.0747, 0.0105, 0.1004]';
[theta2, fv2] = fminunc(objfun,theta0),
[theta3, fv3] = patternsearch(objfun,theta0),
[theta4, fv4] = fminsearch(objfun,theta0)

% Optimal values of univariate likelihood function of X
t1 = zeros(N-1,1); t2 = zeros(N-1,1); t11 = zeros(N-1,1);
for j = 2:N % find optimal theta1 by MLE derivative
    t1(j-1) = X(j-1)*(X(j)-X(j-1));
    t2(j-1) = X(j-1)^2;
end
alphahat = -sum(t1)*(sum(t2))^(-1)
for j = 2:N % find optimal theta1 by MLE derivative
    t11(j-1) = (X(j) - X(j-1) + alphahat*X(j-1)*1)^2;
end
sqrOfSigma1PlusSigma2hat = 1/N * sum(t11)

% find autocorrelation of Returns
autocorr(R)
```

mlf3 and mlf3_ NoPhi are likelihood functions:

```
function f = mlf3(theta, RVW, MVW, d_VW, N, dt)
sigmaX1 = theta(1);
sigmaX2 = theta(2);
alpha = theta(3);
sigmaS1 = theta(4);
mu0 =  theta(5);
mu1 = theta(6);
phi = theta(7); % phi = 0 for mlf3\_ NoPhi
% Define var-covar matrix
K = [sigmaS1^2, sigmaS1*sigmaX1; ...
    sigmaS1*sigmaX1, (sigmaX1^2 + sigmaX2^2)];
[L, U] = lu(K); % for faster calculations
f0 = N*0.5*log((2*pi)^2*det(K));
```

```
f1 = 0;
for j = 2:N
    A = [RVW(j) - ((1-phi)*(mu0+mu1*d_VW(j-1))+...
     phi*MVW(j-1))*dt; d_VW(j)-d_VW(j-1)*(1-alpha*dt)];
    f1 = f1 + 0.5*A'*(U\(L\A));
end
f = f0+f1;
end
```

- Construction of performance variable M_t:

```
% Calculating Performance Variable
data = xlsread('BIST_2004_2014_Analysis');% take data
I = data(:,1); T = length(I); I_M = zeros(1,T);
for t = T:-1:1
    for i = 1:t
        I_M(t-i+1) = exp(-i)*I(t-i+1);
    end
    I_M(t) = sum(I_M(1:t));
end
% display(I_M);
```

- Construction of X_t:

```
% X_t calculations for BIST100
data = xlsread('BIST_2004_2014');
D = data(:,2); I = data(:,1);
N = length(D); Ds = zeros(N/12,1); Dy = zeros(N,1);
for i = 1:N/12
    Ds(i) = sum(D(12*(i-1)+1:12*i));
end
for j = 1:N/12
    for i = (j-1)*12+1:j*12
        Dy(i) = Ds(j)/I(i);
    end
end
L_Dy = log(Dy); DL_Dy = L_Dy - mean(L_Dy);
[DyBC, lambda] = boxcox(Dy);
% see the difference
figure(1), histfit(Dy), title('Histogram of D_t^{\prime}');
figure(2), qqplot(Dy), title('Q-Q Plot of D_t^{\prime}');
figure(3), histfit(DyBC), title('Histogram of D_t');
figure(4), qqplot(DyBC), title('Q-Q Plot of D_t');
```

- Descriptive statistics of all data sets:

```
% descriptive statistics
data = xlsread('BIST_2004_2014_Analysis');
X = data(:,3); M = data(:,2); R = data(:,1);
% for R
figure(1), histfit(R), title('Histogram of R_t');
figure(2), qqplot(R), title('Q-Q Plot of R_t');
mR = mean(R); stdR = std(R);
skwR = skewness(R); krtR = kurtosis(R);
jbtest(R); kstest((R-mean(R))./std(R)); adtest(R);
% for M
figure(3), histfit(M), title('Histogram of M_t');
figure(4), qqplot(M), title('Q-Q Plot of M_t');
mM = mean(M); stdM = std(M);
skwM = skewness(M); krtM = kurtosis(M);
jbtest(M); kstest((M-mean(M))./std(M)); adtest(M);
% for X
figure(5), histfit(X), title('Histogram of X_t');
figure(6), qqplot(X), title('Q-Q Plot of X_t');
mX = mean(X); stdX = std(X);
skwX = skewness(X); krtX = kurtosis(X);
jbtest(X); kstest((M-mean(M))./std(M)); adtest(X);
% for kstest z = (x-mu)/sigma used!
```

- Parameter estimations of BIST-100:

```
% MLE of BIST-100
data = xlsread('BIST_2004_2014_Analysis');
X = data(:,3); M = data(:,2); R = data(:,1); N=length(R);
% start optimisation for phi > 0
objfun = @(theta) mlf3(theta, R, M, X, N, 1);
opts3 = psoptimset('MaxIter',10000, ...
'MaxFunEvals',15000,'TolMesh',.0000001);
opts2 = optimset('MaxIter',10000,'MaxFunEvals',15000);
theta0 = [-0.167, 0.023, 0.1199, ...
 0.0747, 0.0105, 0.1004, 1]';
[theta, fv] = patternsearch(objfun,theta0, ...
[],[],[],[],[],[],opts3),
[theta2, fv2] = fminunc(objfun,theta0),
[theta4, fv4] = fminsearch(objfun,theta0,opts2),

% start optimisation for phi = 0
% objfun = @(theta) mlf3_NoPhi(theta, R, M, X, N, 1);
% opts3 = psoptimset('MaxIter',10000, ...
'MaxFunEvals',10000,'TolMesh',.0000001);
% opts2 = optimset('MaxIter',10000,'MaxFunEvals',25000);
% theta0 = [-0.167, 0.023, 0.1199, ...
0.0747, 0.0105, 0.1004 ]';
% [theta, fv] = patternsearch(objfun,theta0, ...
```

```
[],[],[],[],[],[],opts3),
% [theta2, fv2] = fminunc(objfun,theta0),
% [theta3, fv3] = fminsearch(objfun,theta0,opts2),

% Optimal values of univariate likelihood function of X
t1 = zeros(N-1,1); t2 = zeros(N-1,1); t11 = zeros(N-1,1);
for j = 2:N % find optimal theta1 by MLE derivative
    t1(j-1) = X(j-1)*(X(j)-X(j-1));
    t2(j-1) = X(j-1)^2;
end
alphahat = -sum(t1)*(sum(t2))^(-1)
for j = 2:N % find optimal theta1 by MLE derivative
    t11(j-1) = (X(j) - X(j-1) + alphahat*X(j-1)*1)^2;
end
sqrOfSigma1PlusSigma2hat = 1/N * sum(t11)

% find autocorrelation of Returns
autocorr(R)
```