TIME SERIES ANALYSIS OF IMKB-30 EQUITY MARKET INDEX RETURNS AND THE EFFECT OF VOLUME AND VOLATILITY ON RETURNS

By

Çetin Ali Dönmez

Submitted to

The Institute for Graduate Studies in Social Sciences

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Bogazici University

2006

ABSTRACT

This thesis is done mainly to explore the time series dynamics and lead lag relationships among the Istanbul Stock Exchange (ISE) Equity Market Index, called ISE30, session to session and daily returns, volume and volatility. In addition to the well known classical definition of the returns, a new definition of return is made, namely, the returns are also calculated by using the average values. Moreover, many variables, some requiring detailed information on individual stock basis were also calculated and included in the analysis.

An expectation survey aimed at answering the question of how the market trade variables affect the expectations of brokers was conducted. This survey was found to provide very interesting hints about how the expectations of the market people form in case of different combinations of return, volume and other trade data variables. A very detailed analysis of the survey results are provided in this thesis. Additionally, distributional properties of return series are analysed for the whole period spanning 1997-2005. The period is divided into three sub-periods, namely the pre-crisis period, crisis period and post-crisis period and all the analyses are repeated to see whether the distribution and the sample moments of session to session and daily returns change between different data windows.

Return series were mainly modeled by using Autoregressive (AR), Moving Average (MA) and Autoregressive Moving Average (ARMA) techniques. The return series were found to possess the so called “long memory” or “persistency” problem. The long term memory property was explored in detail and the series are transformed by using the fractional integration method (ARFIMA). After an univariate time series analysis of

the returns, a multivariate analysis of the returns with the trade variables were conducted by using Vector Autoregressive Model (VAR).

In summary AR,MA and ARMA models were found to have little explanatory power for close to close returns. On the other hand, the returns calculated by the average values were found to have significant serial correlations, a fact that makes the AR, MA and ARMA models more useful. ARFIMA method proved to be useful in some cases, while it did not help in some others. Although the inclusion of other variables in the VAR models contributed to the explanatory power, the improvement is generally regarded to be not so prominent. Thus it can be said that changes in volume and volatility were found to have limited explanatory power with regard to the mean return for the next period, a result that is contradictory to what was implied by the expectation survey.

ÖZET

Bu tezde temel olarak İstanbul Menkul Kıymetler Borsası'nın İMKB30 olarak bilinen endeksinin seanslık ve günlük getirilerinin zaman serisi özellikleri ile hacim ve oynaklık gibi işlem bilgilerinden elde edilen veriler arasındaki öncül ardıl ilişkileri incelenmiştir. Klasik olarak kapanış değerleri dikkate alınarak hesaplanan getiri serileri yanında, ortalama değerlerden hesaplanan getiri serileri de analize dahil edilmiştir. Ayrıca işlem verilerinden elde edilen ve bazıları tek tek hisse bazında işlem verilerinden hesaplanan değişkenler de analize dahil edilmiştir.

Piyasada işlem yapan üye temsilcilerin beklentilerinin nasıl oluştuğunun tespit edilmesi amacına yönelik olarak bir beklenti anketi yapılmıştır. Bu anket piyasa oyuncularının değişik getiri hacim oynaklık vb. kombinasyonlarında beklentilerinin nasıl oluştuğuna ilişkin ilginç ipuçları sağlamıştır. Anketin sonuçlarının detaylı analizleri bu tez içinde yer almaktadır.

Ayrıca, 1997-2005 yılları arasındaki verilerin yer aldığı zaman dilimi, kriz öncesi kriz sırası ve kriz sonrası periyodlara ayrılarak ayrı ayrı incelenmiş ve olasılık dağılımlarındaki ve ilgili istatistiklerdeki değişimlerin analizi yapılmıştır.

Getiri serileri temel olarak AR, MA ve ARMA teknikleri kullanılarak modellenmeye çalışılmıştır. Getiri serilerinde uzun dönemli hafıza problemi tespit edilmiş olup, bu husus detaylı olarak ele alınmış ve getiri serileri ondalıklı entegrasyon yöntemi ile dönüştürülerek yeni seriler elde edilmiştir. Getiri serilerinin kendi aralarındaki zaman serisi analizi yanında diğer değişkenlerin de dahil edildiği çok değişkenli VAR modeli kullanılarak değişkenler arasındaki öncül ardıl ilişkileri ortaya çıkarılmaya çalışılmıştır

Özet olarak, AR,MA ve ARMA modellerinin kapanışlardan hesaplanan getiri serilerini açıklamada genel olarak yetersiz kaldığı, ancak ortalamalardan hesaplanan getiri serilerinin modellenmesinde daha çok işe yaradığı tespit edilmiştir. Ondalıklı entegrasyon metodu olarak adlandırılan ARFIMA olarak bilinen metodun bazı serilerde uzun dönemli hafıza problemini hallettiği, ancak bazı serilerde fazla işe yaramadığı tespit edilmiştir. Diğer değişkenlerin de dahil edilmesi ile yapılan VAR analizinin genel olarak getiri dinamiklerinin açıklanmasına katkıda bulunmakla birlikte bu katkının sınırlı olduğu sonucuna ulaşılmıştır. Genel olarak, hacim ve oynaklık verilerinin bir sonraki dönem getirilerinin tahmin edilmesinde sınırlı katkısı olduğu tespit eidlmiş olup, bu bulgunun beklenti anketinde ortaya çıkan sonuçlarla kısmen çelişkli olduğu sonucuna ulaşılmıştır.

Curriculum Vitae

Mr. Çetin Ali DÖNMEZ

Place and Date of Birth: Istanbul, September 27,1967

Education:

Doctoral Student in Finance at Bogazici University (2002-)

MBA Degree in Finance (Bogazici University, 1992)

B Sc. Degree in Industrial Engineering (Bogazici University, 1989)

Experience:

Istanbul Stock Exchange(ISE) Derivatives Market Department (1996-)

ISE Settlement and Custody Bank Credits Department (1995)

Istanbul Stock Exchange Stock Market Department (1990-1995)

Publications

An alternative Instrument of Borrowing for the Treasury: Tax Bonds, *Active Bankacılık ve Finans Dergisi, 2004, No:36*

An Empirical Analysis of Foreign Traders and Their Portfolio Performances and the Effects of Foreign Capital Flow on Market Dynamics in the Istanbul Stock Exchange Equity Market. (With Fatih Kiraz), *ERC/METU International Conference in Economics VII, September,2003*

A new Instrument for Turkish Capital markets: ETFs, *ISE Journal, Sept. 2002*.

A practical proposal for developing the mutual funds industry: Exchange Traded Funds, *Active Bankacılık ve Finans Dergisi, 2002, No:24*

Inroduction to Financial Derivaties, (Donmez, Basaran, Yılmaz,Ugur, Ugan, Kartalli, Dogru) *ISE Press, 2002*

Table of Contents

List of Tables

List of Figures

PREFACE

The exploration of the time series properties of equity market index returns and their relationships with volume is not actually a very interesting topic at the first sight since quite an extensive amount of research has been done on this subject. However this thesis is expected to make some contributions to the current finance literature, mostly in terms of the inclusion of new variables into the time series analysis that up to now generally have been ignored and moreover the readers will have the results of an expectation survey conducted among brokers that will probably propagate some further research in this field.

The main theme of this thesis is to explore the return volume and other trade data dynamics. The first question raised was related to the variable “return”. How should return be defined? Generally return is defined as the logarithm of the ratio of the closing value of the index at time “t” to the closing value of the index at time “t-1”. Why should we define return as such? Starting from this question, a new definition of return was made, namely, it was also calculated by using the average values in addition to the closing values. There are also some other trade variables which are not taken into consideration by almost any researchers, those variables proved to bring very interesting conclusions.

Another important point to note is that, the method known as technical analysis which is not usually taught in finance schools does have quite large popularity among traders. Technical analysts do claim that the prices have some patterns, the volume increase or decrease have important implications for the future price changes. Why is this method so popular? Is it because it is simple? Or is there something really magic behind it? Is it really possible to have a better guess for the next period by taking into account past values of return, volume volatility and other trade data?

During the literature survey phase which took long, various papers was examined on the discipline of behavioral finance as well as the papers directly related to the return volume dynamics. These readings further added to our curiosity for the following question: How do the variables related to trades such as return, volume, affect the expectation of market people? To answer this question, a survey was conducted and very interesting hints about how the expectations of the market people form in case of different combinations of return, volume and other trade data variables were documented.

For preparing this thesis I read many statistics books as well as academic papers in various journals. The most influential books during my work are the books titled "Analysis of Financial Time Series" written by Ruey S. Tsay and "Time Series Analysis", written by Hamilton, J. D. both of which were suggested by my thesis advisor whom I do thank for his valuable suggestions and confidence on me. The paper written by Karpoff, J.M. in 1987 titled "The relation between price changes and trading volume" published in *Journal of Financial and Quantitative Analysis* 22: 109-126 was also very helpful to initiate further research on my part.

I was really astonished by the fact that the Istanbul Stock Exchange does not have a handy and easy to use database which poses great problems for researchers. Many problems in the data were also discovered, thus the academicians should be very careful before using the ISE data in their researches. On the other hand, the prices and the volumes of each stock in the index were separately analysed and variables such as total volume, return dispersion and volume dispersion for the IMKB30 index were calculated for each session and each day, taking into account the trade data of each of the 30 stocks.

The stocks in the index do also change from period to period and there is no list of index stocks for each period readily available in the ISE. Therefore, collecting all of this information and transforming the data to make it ready for statistical analysis was really a tedious task. This is also the reason behind why the IMKB30 index was selected as the main market indicator. It is clear that dealing with 30 stocks is much easier than dealing with 100 stocks in an environment where many problems do exist as to the availability and reliability of the data.

A final note for the reader is about the structure of this thesis. The thesis starts with a discussion of previous research on return volume dynamics, touches upon studies of some behavioral finance scientists. It continues with an explanation of the ISE Equity Market and the related trade data. The second and third chapters explain the survey methodology and the evaluation and analysis of the results of survey with the help of empirical data from the market. Interesting conclusions have been reached, a fact that is believed to initiate further research in this field. Then, in chapter 4 the distributional properties of the returns are analysed, special importance is given to separately analyse the data by using the crisis year of 2001 as a benchmark. In chapter 5 time series analysis of the return series are conducted and the stationarity of the series is studied in detail. The return series calculated from the closing values and average values were first modeled by using Autoregressive (AR), Moving Average (MA) and Autoregressive Moving Average (ARMA) techniques. It was found that, although the close to close returns have very small, even non-significant autocorrelations, the serial correlations in average returns are larger in magnitude and statistically significant. Moreover, both the close to returns and the average return series were discovered to have long memory and significant heteroskedasticity. To overcome the “long memory problem” the ARFIMA methodology is applied to see whether the series can be redefined. And finally VAR model is applied to see the lead lag effects of trade variables on return generating process.

This thesis is the result of a very rigorous and long lasting study and I do believe that the reader will find it relatively easy to progress while reading through the chapters, since I did my best to gain knowledge of all the subjects that are included in this thesis. Some new variables which are generally ignored in most of the previous studies are included in the analyses. Therefore market professionals will find some interesting hints with regard to portfolio management strategies; this thesis is also believed to pave way to new academic studies in the fields of technical analysis and behavioral finance.

CHAPTER I

INTRODUCTION TO THE RETURN DYNAMICS AND THE ISE EQUITY MARKET

Introduction

The predictability of stock index returns based on their own past values and the trading volume of previous periods is an important topic that has been extensively researched in both empirical and theoretical finance. The presence of some form of relationship (linear or non-linear) between index return and its lagged values is often examined and quite different results are reported. If the time series do exhibit a linear relation, this means that the lagged values are autocorrelated and Autoregressive (AR) and/or Moving Average (MA) models can be used do define the functional form of the relationship. Whether the existence of autocorrelation implies a violation of market efficiency has been discussed extensively and various theoretical explanations have been suggested to explain the non-zero autocorrelation phenomenon. On the other hand, even if the consecutive returns are found to be uncorrelated this does not necessarily mean that they are independent. Consecutive returns may for example, be uncorrelated, but their squared values may well be correlated across time. However, there is another complication, namely, the relationship between the returns or squared returns and their past values across time has been found to be unstable in a number of studies and this finding further complicates the theoretical framework. As a rule of thumb, any time series should be plotted to see whether the sample moments and the shape of distribution vary over time or not. If, for example, the variance of returns change over the course of time and this can be modeled by autoregressive conditional heteroskedasticity (ARCH, GARCH) models.

The buying and selling decisions in the stock market are taken everyday by a huge number of market agents and the result is the market price and volume that are observed on various media. In the Turkish case, quite an extensive use of technical analysis is known to be used, a method that is usually overlooked by efficient market proponents. The main objective of this thesis is to explore the lead-lag relationship between prices and trading volumes and thus a sound empirical explanation supporting the use of technical analysis is sought. A common notion in handbooks of technical analysis is that increasing trading volume strengthens trend, i.e., increases the probability that the ongoing trend will continue. Conversely, falling trading volume signals that the current trend is going to reverse, i.e., the probability that a trend reversal is going to happen increases. Increasing number of financial research is focusing on time series properties of not only price but also volume. Thus, as noted before, from the point of view of the technical analysis method, lagged volume could be useful to predict price movements since market participants can not obtain a full information signal from the price alone. If traders frequently use volume data as an additional statistic to observe some sort of signal with regard to possible future path of assets, the supporting empirical evidence should be found to validate this belief. In our effort towards finding some sort of supporting evidence for the wide spread use of technical analysis and volume return relation a survey was conducted among the stock brokers. The results, which will be documented in detail in coming sections, indicate that brokers do heavily rely on technical analysis and they also give special importance to volume increases or decreases along with the magnitude of returns in forming their expectation for future returns.

The variables used in this research are mainly the trading volume (in YTL terms), net volume (YTL value of equities changing hands during a trading day after netting, a measure that can only take the attention of a market professional), close to close return, average return, return dispersion, max-min range (intra-day price range), direction of the market measured by the difference between the closing price and the weighted average price. A large majority of prior research failed to find any conclusive evidence on the nature of the relationship between return and volume. In addition to the net volume data, at least for the last three year-time period, the daily foreign investor share in total free float was also used to see whether the changes in the foreign holdings of Turkish shares produce any signal as to the direction of the whole market.

The very hot field in the recent financial studies, namely behavioral finance is also touched upon and various articles are summarised to gain insights into the return process. The fact that the maximum, minimum and the average prices along with the closing prices are available in the media, this may well mean that investors might also be using those statistics to form their expectations. In addition to measuring the differences between the closing prices, the difference between the maximum prices and the minimum prices may help to explain the seemingly random behaviour of returns.

A Brief Summary of the ISE Equity Market Operational Structure

The ISE Stock Market operates on a fully computerized trading system based on a multiple price-continuous auction method in which buy and sell orders match on price and time priority. Stock trading activities are carried out in two separate sessions, the first being held between 09:30-12:00 and the second between 14:00-16:30. There is also an "Accumulated Order Processing" (AOP) period at the beginning of each session. The AOP period is between 09:30-09:45 for the first session and 14:00-14:10 for the second session. During AOP, only limit orders are received via electronic interface from the internal systems of the brokerage firms, or alternatively, accumulated orders are downloaded from floppy discs through trading terminals. The main distinction of this period is that the brokers are not allowed to use the keyboard of trading terminals, therefore, direct manual order entry to the system is not allowed during AOP.

From the perspective of this study there is an important detail in the ISE Stock market trading mechanism. In the trading mechanism, there is a price limit of 10% for each stock at each session and the base price on which the price limit is calculated is the volume weighted average price of the stocks in the previous session. If any dividend payment or stock split occurs, the base price is adjusted accordingly by the ISE. This fact is quite important, because some of the exchanges do not have price limits and still many others calculate the price limits based on the closing prices.

ISE Equity Market transactions are settled at (T+2) (Two days after the trade date). At the end of each trading day, Takasbank (Clearing & Custody Bank) multilaterally nets transactions made during the day and calculates the settlement position on member basis. The Istanbul Stock Exchange Equity Market transactions were settled on a T+1 basis in the past, but for the period to be analysed in this thesis, namely (1997-2005) the settlement period was T+2 and did not change during the period.

IMKB Indices

The IMKB30 index is an important benchmark since it contains very liquid stocks and foreign investors usually invest in İMKB30 stocks. The original idea behind calculating and publishing this index actually traces back to the second half of the 1990s. At that time the İMKB was very keen to start index derivatives trading and the İMKB30 index was the main underlying index that was planned to be used.

In this study the main focus is on the Istanbul Stock Exchange 30 Share Index (IMKB30), which is composed of the 30 largest stocks in terms of equity capitalization, the index represents around 75% of the total equity capitalization of all stocks. The trading volume of these 30 stocks consists of approximately 65% of all trading volume. On the other hand the other well known index namely the IMKB100, contains 100 largest stocks and represents around 90% of the total market capitalization and trading volume. Both the IMKB30 and IMKB100 are value-weighted arithmetic mean of the constituent stocks based on equity capitalization. Capitalization weighing is done by using the free float rate as a multiplier so that the publicly held portion of the total capitalization is taken into account.

Both the IMKB30 and IMKB100 are price indices meaning the dividend payments of index stocks are not assumed to be reinvested. Corporate actions such as stock splits and equity offerings are taken into account, necessary adjustments are done in order to maintain the continuity of the index. In the calculation method, actual transaction prices are used to determine the level of the index. This point is quite important, because not all the indices are calculated in this manner. For example the very well known FTSE index is calculated by using prices that are based on the midpoint of the best (inside) bid-and-ask quotes (the touch) being displayed on the trading screen for each of the constituent stocks. This difference is quite important when one deals with intraday volatility and returns. This is because if there is no trade during a time interval and if in this case prior transaction price is taken into account then this may create artificial serial correlations. However, since the main focus of this research is the session to session returns and the stocks included in the İMKB30 index do have very liquid markets this issue can be ruled out. However, researchers should be careful when they concentrate on intraday return process, especially for very short time intervals.

Another important point that should be noted is the fact that the ISE30 index constituent stocks are revised quarterly by the exchange and the necessary adjustments are done accordingly. However, there is still another issue that should be taken into account. It is the fact that, in addition to periodic assessments, non-periodic changes in the list of index stocks are also made. For example, if a stock is not traded or is closed to trading for more than five consecutive trading days it is excluded from the index and another stock is added. It should be noted that there were such cases within the period of our study and consequently, necessary adjustments were made

For the whole sample period, both periodic and non-periodic changes to the IMKB30 index were all taken into account. This is important, because to calculate the variables, volume, volume dispersion and return dispersion, individual stock information was needed. The list of the IMKB30 constituent stocks for each period is provided in Appendix A. As seen, the İMKB30 index stocks are revised quarterly; however due to the problems related to the financial well-being of the constituent companies, some non-periodic changes occurred. For example in November 2000 when there should have been no change in the index composition Medya Holding (a large conglomerate with severe financial problems during that time) was excluded from the index and (MIGROS) company replaced it. The stocks excluded from and included in the index and the relevant dates are all provided in Appendix A. The construction of this table took quite a lot of time. It should also be noted that the index in this sense has survivorship bias, because the investors holding the stocks to mimic the performance of the IMKB30 face the problem of having those stocks with no value.

Theoretical Explanations Of The Return- Volume –Volatility Relationship

There are mainly two theories on the lead lag relations among return volume and volatility. The Sequential Information Arrival Hypothesis (SIAH) of Copeland (1976), Jennings, Starks & Fellingham (1981) assumes that traders receive new information in a sequential, random fashion. From an initial position of equilibrium where all traders possess the same set of information, new information arrives in the market and traders revise their expectations accordingly. However, traders do not receive the information signals simultaneously. Once all traders have reacted to the information signal, a final equilibrium is reached. The sequential reaction to information in the SIAH suggests that lagged values of volatility may have the ability to predict current trading volume, and vice versa.

On the other hand, the Mixture of Distribution Hypothesis (MDH) hypothesis (Clark (1973), Eps and Eps (1976)) is based on the assumption that all traders simultaneously receive the new price signals. As such, the shift to a new equilibrium is immediate and there is no intermediate partial equilibrium. Thus, under the MDH, there should be no information content in past volatility data that can be used to forecast volume (or vice versa) since these variables contemporaneously change in response to the arrival of new information. In other words MDH hypothesis states that volatility and volume are driven by the same information flow simultaneously.

The autocorrelation of returns phenomenon found in empirical studies of equity markets is also tried to be explained by researchers on several grounds. One of the explanations on the meaning of return autocorrelations is that, correlations arise due to market frictions (nonsynchronous trading, price discreteness etc.) Nonsynchronous trading may especially be important in case, all the stocks included in the index are not traded simultaneously. This may cause the prices of some stocks to lag behind others. Another explanation is that autocorrelations are observed because the economic risk premium is time varying. The proponents of this view also argue that in an efficient and even frictionless market the returns can be autocorrelated. Yet another group of researchers attribute the nonzero autocorrelation to the irrationality of market participants. The existence of irrational investors, irrational trading behaviour or psychological factors may produce profitable trading strategies for rational or astute investors. Especially this approach gained considerable support among the financial community during the last two decades and lead to the new discipline called behavioral finance.

The behavioral finance perspective seeks to explain the time dependency of returns and volatilities by different theoretical approaches. For a true believer of market efficiency there is little reason to believe that these past statistics are helpful in making investment decisions. However, the statistics published in the media may be satisfying some kind of a demand from the investors' point of view. The direction and the magnitude of the changes in variables such as changes in volume, changes the i.e., the highest and lowest of the last period, last one week, one month one year period etc. may be serving as important benchmark points rather than the actual transaction price to the investors especially in cases where the investor hold the stock for so long that he/she no longer recalls the purchase price or the current trading price is far from the purchase price.

The availability bias put forward by Tversky and Kahneman (1982) causes people to base their decisions on the most recent events which in turn causes investors to over-react to market conditions whether they are "positive" or "negative". This implies that the return series should exhibit reversals. However it is quite crucial to find out on what time horizon this is valid. Is it valid for very short periods like intra day or session or day or is it valid for the weekly periods or monthly periods. Therefore one should be very careful in interpreting the availability bias theory.

Contrary to the availability bias explanation some researchers like Daniel at al.(1998) Barberis, Nicholas & Thaler (2003) focus on the overconfidence bias. The overconfidence bias implies that people are too slow to change an established view, as opposed to being too willing to change. This theory is based on the assumption that *the human mind is conservative.* The implication is that people are slow to recognize the importance of an information arrival especially when it is contrary to market wide expectations. That is, they underweight evidence that disconfirms their prior views and overweigh confirming evidence.

Consequently, both analysts and investors interpret a permanent change as if it were temporary; thus the price is slow to adjust. This means that the return process may be positively correlated, the effect of an information arrival may show itself slowly in the market. The fact that investors gradually realize the effect of the new information, the market will probably underreact which implies that there is a profit chance for momentum traders or trend chasers.

Still another explanation from behavioral finance theory is termed as disposition effect. The disposition effect which was first coined by Shefrin and Statman (1985) is actually an extension of the prospect theory put forward by Kahneman and Tversky (1979). According to the Prospect Theory people value gains and losses relative to a reference point like the purchase price, and that they are risk-seeking in the domain of possible losses and risk-averse when faced with gain outcomes. Shefrin and Statman extended prospect to theory to investment decisions and claimed that investors have "the disposition to sell winners too early and ride losers too long". They labeled this behavioral phenomenon as "the disposition effect". An implication of disposition effect is that current volume should be negatively correlated with the volume on previous days if the current price is below the previous price(s) and positively correlated with the volume on previous days if the current price is above the previous price(s).

Additionally, research on learning and memory suggests that individuals are also likely to remember extreme observations (Fredrickson and Kahneman 1993; Fiske and Taylor 1991). As a result, individuals may focus on extreme observations when making investment decisions. Thus for example, an investor trying to decide when to sell a stock may view a trade price surpassing a prior high as an opportunity to sell. Similarly, an investor considering potential investments may view the stock trading below its historic low as a favorable time to buy.

General Findings of Previous Researches

The main path to investigating the lead-lag relations in the return series starts with checking whether there is serial correlation between returns. The shape of the distribution is also important in the analysis of return series. The existence of non-normality might be due to serial correlation and/or heteroskedasticity in return series.

The empirical investigation of equity returns were initially done by Fama (1965) and Mandelbrot(1963). Studies usually have shown that returns, especially in the short run are not normal. The return distributions do show positive skewness and a high kurtosis value. A kurtosis value larger than three implies of course the distribution has a fat tail problem. Efforts have been made to solve the fat tail problem by using the models such as ARCH and GARCH which are based on volatility clustering assumption. (Bollerslev, Chou & Kroner 1992), Akgiray (1989), Akgiray, Booth, Loistl (1989), Aparicio, Estrada (2001).

The non-normality property has very important consequences. One is that if a stochastic process is not normally distributed, the non- existence of serial correlation does not imply statistical independence Akgiray (1989). Another important consequence of the return distributions being non-normal is that the conclusions to be drawn from Box-Pierce Q test and Dickey Fuller unit root test should be evaluated carefully when the distribution is not normal. Lo and McKinlay (1998) showed that the variance ratio test is a better measure to test the random walk hypothesis than the Box Pierce Q test and Dickey Fuller test.

Akgiray (1989) also showed that the daily return series of the US equity indices are not normally distributed and sample moments differ from period to period and also concluded that daily returns are not independent of each other. He further showed that return series display high first lag autocorrelations. The significance was found to be even higher in absolute and squared return series. He then applies the AR(1) transformation of returns and finds that although the resulting residuals of this transformation are uncorrelated, they are not independent. He then proceeds to utilize the conditional heteroskedastic models namely the ARCH model of Engle (1982) and the GARCH model of Bollerslev (1987) to account for the dependence of the squared error terms and finds that GARCH (1,1) model fits quite well to the daily return data and the hypothesis that standardized residuals of this model are normally distributed could not be rejected. A similar study was conducted by Mougoue and Whyte (1996) for the German and French Equity Markets and they found that stock returns in both countries are best described by GARCH(1,1) model. They also documented changes in the mean variance relationship before and after the US stock market crash of 1987. This conclusion is consistent with the work of Akgiray (1989).

On the other hand, it is quite vital then to break up the sample to find out regime shifts if any, in return generating process. For example, Masulis and Victor (1995) studied the FTSE index series for the eight year period between 1984 and 1991 and in order to study the possibility of regime shifts occurring in the behavior of stock returns over the observation period, they divided the observation period into three subperiods and then compare the statistical properties of the return series across subperiods. They also excluded certain periods such as before and after market crash in order to avoid the negative effects of transition periods and to reduce the effects of several large outliers that occur at this potential structural breakpoint. Their model distinguishes between overnight and daytime return dynamics, permits overnight and daytime return dynamics, to follow different leptokurtic conditional distributions. They found that the distribution of overnight returns is more leptokurtic than that of daytime returns. In fact, daytime returns are found to conform much more closely to a normal distribution than do overnight returns. They also found that the probability of more extreme returns is higher in the post-crash period, overnight returns are strongly positively correlated with the most recent daytime return, mildly negatively correlated with the prior daytime return. They examined the serial correlations of squared returns and observed that overnight squared returns are strongly positively correlated with the most recent daytime squared return and mildly positively correlated with the prior two daytime squared returns.

On the other hand, Chowdury (1999) analysed the weekly stock returns in eight Asian and Pacific Markets and he found that although there is some first order autocorrelation, this does not indicate the availability of profitable short run investment opportunities. Taylor(2000) investigated daily returns of FTA(Financial Times All Share) index, FTSE100, twelve frequently traded UK stocks, and Dow-Jones Industrial Average and S&P500. As a result of his study he rejects the hypothesis that returns are uncorrelated and thus finds significant dependence between consecutive returns especially for the indices. However whether this finding may result in significant opportunities for trading to beat the market is unclear. Another approach is to find out whether the stock market overshoots during especially crisis period. For example Basci and Muradoglu (2001) by using weekly national index returns for 21 world markets documents international evidence that stock market rebounds after extreme falls. They used a third order polynomial model on lagged returns, coupled with GARCH residuals and found that the return forecasts from this model are better than the linear alternatives in weeks following extreme falls.

In addition to forecasting the magnitude of returns and volatility some researchers concentrated on forecasting the direction of the market. Because profitable trading strategies may result if one successful at forecasting market direction, quite apart from whether one is successful at forecasting returns themselves.

For example, Christioffersen and Diebold (2003) used the daily weekly, monthly and annual values of S&P100 index and found that if the expected returns are non-zero volatility dependence produces sign dependence especially at intermediate horizons of two or three months and this fact can not be captured by the widely used techniques such as analysis of sign autocorrelations, runs tests or traditional tests of market timing due to the fact that the nature of sign dependence is highly nonlinear.

They also found that the link between volatility forecastability and sign forecastability still holds in conditionally non-Gaussian environments, as for example with time-varying conditional skewness and/or kurtosis.

Some academics have tried to explain the price behaviour of assets by including the range which is defined as the difference between the highest and lowest prices throughout the day. Alizadeh, Brandt and Diebold (2002), found substantial gains in estimating volatility from using the range. This research and a number of previous studies Parkinson (1980), Garman and Klass (1980) showed that range is more efficient volatility proxy. These findings led us to include the daily range of returns as predictor variable in our analysis.

In a paper written by Huddart,Lang & Yetman very interesting results have been reached as to the relation between the trading volume and aspects of the stock's past prices. Their research suggests that investors focus on past stock price behaviour in making their trading decisions. More specifically, the authors document substantial increase in volume when a stock is trading above the highest price attained during the year ending 20 trading days before the current week. They also found that the effect is more visible the longer the time since the prior maximum is attained. They also find that the effect is stronger for NASDAQ stocks where the individual ownership is greater, than for NYSE and AMEX stocks, which implies that there is a negative relation between investor sophistication and reliance on reference points.

In addition to analyzing the price changes there is also extensive research focusing on the relationship between daily trading volume and stock price movements. Karpoff (1987) provides a very good overview of earlier research on the relationship between returns and volume. He classifies the studies into two groups; first, those that examine the relationship between absolute price change and trading volume, and second, those that examine the relationship between price change *per se* and trading volume, and finds that the majority of them report a positive relationship between price change (*per se* or absolute) and trading volume. Karpoff (1987) cites four reasons why price-volume relation is important. First, empirical evidence on price-volume relation is helpful to analyze how information is disseminated to financial markets and whether price by itself contains how much of that information. The second reason mentioned is that, understanding the joint distribution of returns and volume is important and this will probably increase the power of statistical tests. Third, joint dynamics between returns and volume are also important to examine the distribution of returns and changes in variances. And fourth, price volume relations may help to explain whether speculation is stabilizing or de-stabilizing effect on prices.

Lamourex and Lastrapes (1990) for example used daily trading volume as a proxy for information arrival and showed that volume has a significant explanatory power regarding the variance of daily returns. Contrary to this finding however, Lee, Chen and Rui (2001) used daily trading volume as a proxy for information arrival and found that volume has no significant explanatory power for the conditional volatility of daily returns. They further note that variance ratio tests reject the hypothesis that stock returns follow random walk. Silvapulle and Choi (1999) use linear and nonlinear causality tests to investigate causality between returns and trading volume on the Korean stock exchange. They find significant bilateral linear and nonlinear causality between returns and volume. Chen, Firth and Rui (2001) examine the dynamic relation between returns, volume, and volatility of nine national stock index for the period from 1973 to 2000. Their results show a positive correlation between trading volume and the absolute value of the stock price change. Granger causality tests demonstrated that for some countries, returns cause volume and volume causes returns. In general, they concluded that return cause volume but not vise versa.

Several recent theoretical papers also examine the role played by trading volume in asset markets. Lee and Swaminathan (2000) analyzed monthly returns of NYSE and AMEX stocks for the period January 1965 through December 1995 and showed that past trading volume predicts both the magnitude and the persistence of future price momentum. They found that high(low) volume winners(losers) experience faster momentum reversals. They further found that low volume stock generally outperform high volume stocks. Lo and Wang (2000) conclude that trading activity is fundamental to a deeper understanding of economic interactions. In a recent study by Connolly and Stivers (2003) examined the relationship among turnover shocks and price dispersion shocks returns. Their data set is the weekly (Wednesday to Wednesday) large firm portfolio returns of US UK and Japanese equity markets. They find that turnover shocks and return dispersion shocks as they define them are positively correlated. They also find that the first order autoregressive coefficient of return series is reliably positive and increasing when the trading volume increases. They further find that return in time t is positively correlated to the volume in time t-1, but the magnitude is found to be smaller than the contemporaneous relation of return and volume. They also find that consecutive equity index returns tend to display substantial momentum when there is unexpectedly high turnover in the latter period and reversals when there is unexpectedly low turnover or return dispersion. Fan, Groemewold and Wu (2003) examined the relation between trading volume and stock returns in Chinese equity market by using daily return and volume data and found that volume has low predictive power on future returns but a strong and predictable effect on absolute returns and they further documented stronger evidence of return causing volume. They also found that equity returns are not independent and GARCH (1,1) model fits well to the data. Salman (2000) recently investigated daily return volume and risk dynamics of Istanbul Stock Exchange by using the GARCH method. He finds that lagged volume has a statistically significant positive effect on returns and he also find a positive contemporaneous relationship between risk and return.

Another interesting study conducted by Gervals, Kaniel, Mingelgrin (1999) documents that stocks (NYSE stocks) experiencing high(low) trading volume over a period of one day to a week tend to appreciate(depreciate) in the following month. This effect is found to be stronger when the rise in volume is not accompanied by an abnormal rise in returns. They further show that profitable trading strategies can be implemented by using the information content of the volume data. Blume, Easley and O'Hara (1994) show that lagged volume could be useful to predict price movements. Basci, Özyildirim and Aydogan investigated the weekly price and volume series of 29 individual stocks traded in the Istanbul Stock Exchange for the period between January 1988 and March 1991, by regressing the price levels with the volume series, and using the lagged values of the residuals from this regression in an error correction model they found that it seems possible to forecast the future price changes of some stocks by using the current price and trading volume. Although this finding casts significant doubt on the efficiency of the Istanbul Stock Exchange market, data belongs to a period where the exchange was in the early stage and the study was concentrated on individual stock basis which is subject to the actions of insiders and large portfolio holders especially in the period examined. Ciner (2000) also investigates whether trading volume contains information to predict both the magnitude and direction of price changes on the Toronto Stock Exchange (TSE). He finds that linear causality tests show no predictive power for lagged volume for returns per se, although this conclusion is reversed by nonlinear causality tests which suggest nonlinear predictive power for lagged volume. Saatcioglu and Starks (1998) investigate emerging markets in Latin America and they report that volume leads returns in these markets.

The importance of trading volume in forecasting returns is also examined at individual stock level. For example Chordia and Swaminathan(2000) investigated the stocks in n NYSE/AMEX during the period from 1963 to 1996 found that holding the firm size constant, trading volume is a significant determinant of the cross-autocorrelation patterns in stock returns. More specifically they found that daily or weekly (measured as Wednesday close to Wednesday close) returns of stocks with high trading volume lead daily or weekly returns of stocks with low trading volume. Additional tests indicate that this effect is related to the tendency of high volume stocks to respond rapidly and low volume stocks to respond slowly to market wide information. On the other hand Richardson and Peterson (1999) used daily returns of New York Stock Exchange and American Stock Exchange stocks and found that lagged large firm returns predict current small firm returns after controlling for auto correlation in small firm returns.

Darrat, Rahman, Zhong (2003) used intraday volume and return volatility of Dow Jones Industrial Average Stocks and found that contemporaneous correlations are positive and statistically significant in only three of the 30 DJIA stocks and all the 27 remaining stocks of the DJIA exhibit no significant positive correlation between trading volume and return volatility. They concluded that such weak evidence of contemporaneous correlations contradicts the prediction of the MDH in intraday data. Contrary to the non existence of contemporaneous correlations trading volume and return volatility are found to follow a clear lead lag pattern in a large number of the DJIA stocks which means that the result support the Sequential Information Arrival Hypothesis.

Interaction of autocorrelation and volatility and volume is another interesting research area. Yanxiang Gu (2004) provides a brief overview of previous research about this issue. LeBaron(1992), by using daily and weekly data of the S&P composite index from January 1928 through May 1990 finds that first order autocorrelation is larger during periods of lower volatility and smaller during periods of high volatility for both daily and weekly returns. Sentana and Wadhwani(1992) reported that when volatility is low stock returns at short horizons exhibit positive serial correlation, and in case of high volatility they exhibit negative autocorrelation. Campbell Grossman and Wang(1993) examined the relation between autocorrelation and volume. They found that the first order autocorrelation tends to decline as volume increases.

From the market efficiency perspective Fama(1998) summarizes the previous research on return dynamics and Asserts that the concept of market efficiency is still valid. He bases this conclusion on two reasons. One is that in an efficient market, apparent underreaction will be about as frequent as overreaction. If anomalies split randomly between underreaction and overreaction, they are consistent with market efficiency. He documents several prior research supporting his view. The other reason that Fama cites is the his finding about the sensitivity of results on the methodology used in prior researches. He claims that the findings of previous research tend to become marginal or disappear when exposed to different models for expected (normal) returns or when different statistical approaches are used to measure them.

CHAPTER II

DATA, VARIABLES AND SURVEY DESIGN

Data

The data used in this research were the IMKB30 index minimum, maximum and closing values and the minimum maximum, closing , weighted average prices and the trading volume of IMKB30 stocks. Data were available on each trading session basis. The price and volume data of the individual stocks included in the index were also used for some statistical calculations which will be explained in the coming pages. There was a huge amount of data collecting effort for the preparation of this thesis. To give an example, one of the main variables of this study is the volume and the volume of the İMKB30 index was found to be not separately kept in the IMKB database, therefore the volume of the IMKB30 index was exclusively calculated by using the individual volume of the İMKB30 stocks.

On the other hand, due to the fact that İMKB30 index has been calculated since the beginning of 1997, the data period starts from the beginning of 1997 and ends in April 2005. Although the index data are available since the beginning of 1997, a major problem in the series was encountered, for the first quarter in one of the constituent stocks and therefore the analysis were started from the beginning of April 1997. The whole sample was also divided into three sub-samples to see the effect of the financial crisis which occurred during 2000 and 2001 in Turkey, on the mean, variance and the shape of the distribution of return series.

For each of the periods, on a trading session basis, minimum, maximum, closing and weighted average prices and trading volumes of all 30 stocks of the IMKB30 are provided. The adjusted prices of the stocks are supplied by the ISE. For the readers to get the logic behind the adjustment, the adjustment is done simply by multiplying all the past values of the stocks by the ratio of new opening price/old opening price. The opening price is simply the weighted average price of the last session rounded to the nearest price tick. The price data before 2005 also is adjusted to account for the adoption of the new currency namely the New Turkish Lira (*Yeni Türk Lirası*). The adjustment on each stock basis is necessary because some of the explanatory variables to be used in this study depend on values in TL or YTL, such as the so-called return dispersion, which is calculated by the square of the difference between the return of each of the index stocks from the index return. Thus, in order to calculate the return of individual stocks, adjusted prices are needed.

An exhaustive list of the data and the variables used in this research is provided below:

CI_t =	Closing level of the IMKB30 index during period t
Min_t =	Minimum Level of the IMKB30 during period t
Max_t =	Maximum Level of the IMKB30 during period t
AI_t =	Average Level of the IMKB30 during period t
CR_t =	Return calculated from the closing values
AR_t =	Return calculated from average values

$Range_t$ =	The difference between the minimum and maximum values of the index divided by the Average value of the index
DIR_t =	Direction of the index as of period t
Vol_t =	TL value of the stocks traded during the time period t
$NetVol_t$ =	The amount of money actually changing hands during a trading day.
N(up) =	Number ISE30 stocks went up during time t
N(down) =	Number ISE30 of stocks went down during time t
N(nochg) =	Number of ISE30 stocks that did not change
N(strongup) =	Number of ISE30 stocks which experienced both a price increase and volume increase
N(Strongdown) =	Number of ISE30 stocks which experienced a price decline and increase in volume
Artaz =	(N(up)-N(down)) / 30
RDt=	Return dispersion of the ISE30 stocks
VDt=	Volume dispersion of the ISE30 stocks
RV_t =	Average of the individual returns multiplied by volume. As the name implies, this variable is calculated by using all the return and trading volume of individual stocks.
Ret30vol =	Return of the index multiplied by the percent change in total trading volume of the index.

The explanations of some of the variables mentioned above are as follows:

Return is defined as the natural logarithm of the index level at time t divided by the index level at time t-1. In functional form close to close return, for instance, return is defined as:

$$CR_t = Ln\ (CI_{\ t} / CI_{\ t-1})$$

The Istanbul Stock Exchange does not calculate and publish any average index, in other words, an index calculated from the average prices of common stocks. For this reason a proxy for the average index defined as follows is used in the analysis

For Each Session:

AI (Session)= Average index = (Maximum Value+ Minimum Value + Closing Value)/3

For each day

AI (Day) = (The first session average + second session average)/2

Consequently, the average return is calculated by the following formula

$$AR_t = Ln\ (AI_{\ t} / AI_{\ t-1})$$

In addition to the calculation of index returns, the return series of individual stocks in the index is also calculated as the logarithm of the ratio of adjusted closing price of stock at time "t" to the adjusted closing price of stock i at time "t-1". This is done to find the value of return dispersion to be explained ahead.

In addition to calculating the returns, the volatility is approximated by the squared value of the returns. Additionally, the intra period volatility also is approximated by the difference between the minimum and maximum values divided by the closing index value. This measure is named as range and calculated as follows:

$$Range_t = (Max_t - Min_t) / CI_t$$

The direction of the index during a session or a trading day is also included in analysis. The variable to show the direction of the index is defined as follows:

$$DIR_t = (CI_t - AI_t) / AI_t$$

A positive value for this variable indicates that the index closing level is above the average value and this might imply the index is heading up, and for negative values vice versa.

In addition to the variables related to price levels, trading volume figures (Vol_t) are also taken into account. Volume is defined as the TL value of index constituent stocks traded during a trading session or trading day. Total volume of İMKB30 stocks is not really available, therefore the total volume of İMKB30 is calculated by summing up the individual volume figures of each stock in the index. In order to satisfy the stationarity condition in time series analysis changes in volume which is defined below is taken into account.

$$\Delta Vol_{t,t-1} = \text{Logarithm (Volume at time t / the Volume at time t-1)}$$

In addition to volume, another variable which is defined as Daily Net Volume of IMKB30 index ($NetVol_t$) was used. This variable represents the amount of money actually changing hands during a trading day. In other words this value is supposed to show the portfolio movements in brokerage houses. Similar to the reasoning applied to the volume figures, changes in net volume are calculated and taken into account in order to satisfy the stationarity condition.

The calculation of net volume figures needs some explanation. For example, assume for simplicity that there are four brokerage firms (A,B,C,D) trading in the market and there is only one stock, say it is S. Further assume that at time t the following trades occurred.

Firm A buys 3 million YTL worth of Stock S

Firm B sells 1 million YTL worth of stock S

Firm C sells 1 million YTL worth of Stock S

Firm D sells 1 million YTL worth of Stock S

After the above trades trading volume and net trading volume can easily be calculated as 3 million YTL. Firm A brings 3 million YTL cash and takes the stocks bought. Firm B, C, and D deposit the stocks they sold and take 1 million YTL each. The trading volume is 3 million YTL and 3 million YTL changes hands in this case.

Now assume that at time t+1 the following trades occurred.

Firm A buys 2 million YTL worth of Stock S

Firm B sells 1 million YTL worth of stock S

Firm C sells 1 million YTL worth of Stock S

Firm A sells 1 million YTL worth of Stock S (Firm A, for some reason, sells half of the 2 million YTL worth of stocks it bought)

Firm B buys 1 million YTL worth of Stock S

After the above trades the trading volume is 3 million YTL. However this time the net trading volume is only 1 million YTL. This is because Firm A sells back 1 million YTL of Stock S it bought and is obligated to bring only 1 million YTL cash to the clearing center. Firm B first makes a sale of stock S and buys back the exact amount and thus the net obligation is zero for firm B. Firm C makes just one trade and takes 1 million YTL cash against the delivery of 1 million YTL worth of Stock S. In this case although the trading volume is 3 million YTL, only one million YTL worth of stock S and 1 million YTL cash changes hands, thus the net volume is calculated to be one million YTL. The net volume is only available for each trading day and not for each session. Coming back to the definitions changes in net volume was used in order to get rid of non-stationarity problem. In other words, $\Delta NetVol_{t,t-1}$ is defined as logarithm of the ratio of Net Volume at time t to the Net Volume at time t-1.

Another important variable used in this study is the so called return dispersion. This variable measures whether the constituent stocks in index move in accordance with each other or not. Return dispersion (RD_t) of the stocks in the index is defined in functional form as follows:

$$RD_t = \sqrt{\left[1/(n-1)\sum_{i=1}^{n}(R_{i,t} - (\sum_{i=1}^{n} R_{i,t}/30))^2 \right]}$$

where n = 30 since there are 30 stocks in the İMKB30 index. This measure uses the individual firm returns included in the index and shows how dispersed are the returns of constituent stocks. If all the stocks move together in one direction this variable takes small values, on the other hand if there are significant differences among the returns of the index constituent stocks this variables takes a large value.

Another measure called Volume Dispersion VD_t was used to explain return series, this variable is defined as follows:

$$VD_t = \sqrt{\left[1/(n-1)\sum_{i=1}^{n}(V_{i,t} - (\sum_{i=1}^{n} V_{i,t}/30))^2 \right]}$$

here again n equals to 30 since there are 30 stocks in the İMKB30 index. Volume dispersion is normalized by dividing the above value to average volume of the period. This variable measures whether the stocks in the index do have volume figures close to each other or not.

RV_t = (Return multiplied by Volume) To find this variable, the volume of each stock in a session or in a day is multiplied by return of each of the stocks in the period and all these figures are summed up. As has been done for the return dispersion, the resulting figure is divided by the average daily volume of the index to normalize the series. More specifically the following formula will be used.

$$RV_t = \sum_{i=1}^{n} (R_{i,t} * Vol_{i,t}) \,/\, \mathrm{Vol}_t$$

where n = 30 since there are 30 stocks in the İMKB30 index.

The last variable used in this research is the variable called ret30vol which is calculated by the following formula:

$$\mathrm{Re}t30Vol = Ln(CI_t / CI_{t-1}) * Vol_t / Vol_{t-1}$$

In other words per cent change in the index is multiplied by the per cent change in total volume of the index stocks.

Survey Of Expectations Of Stock Market Brokers

In order to find out the effect of return, volume, price variables on the expectations of market professionals a survey was conducted on the Istanbul Stock Exchange Trading Floor. A questionnaire is prepared for this purpose.

In the questionnaire categorical variables are used to get information as to whether the respondent has investment in the Istanbul Stock Exchange Stock Market and which investment techniques or methodologies are used. The survey was conducted in one shot during the period between session 1 and session 2.

In order to find out the profiles of respondents in terms of demographic characteristics, four questions with categorical answers were asked. These are namely, questions about the gender, age group educational level and graduate major of the respondents.

There are also twenty questions mainly aimed at finding the expectations for the next day and/or next session of the brokers regarding the direction of the ISE Stock Market. The choices of these expectation questions are designed as an ordinal scale (likert type scale). The answers can be one of type “strongly up”, “up”, “horizontal”, “down” and “strongly down”. To give example, the first of these twenty questions is provided below:

“What is your expectation of the market direction for the next session or the next day if there is an increased in the index level accompanied with an increase in volume. Please write down your expectations as provided below.”

Strongly up (SU), Up(U), Horizontal (H), Down (D), Strongly Down (SD)

For all the twenty questions aimed at finding out the expectations, the subjects were also allowed to choose a "no idea" option, if they do not really have any idea as to the direction of the market.

General Findings From The Survey

Some interesting findings from the survey are provided below:

- A total of 500 questionnaires were distributed to the brokers and traders on the ISE Floor, 191 of them were returned, 107 of them being male and 73 of them being female while 11 of them did not check either male or female.
- The age of the brokers is heavily concentrated between 26-40, almost half of the brokers are between 31-35.
- A surprisingly large percentage of brokers (40%) answered the question of whether they had any investments in the stock market as "no". This ratio is approximately the same for male and female brokers. It seems however, quite interesting to get this answer since the market is actually on the finger tips of these people. This finding might be due to the fact that a considerable number of brokerage houses do restrict their brokers from investing in the market. Another interpretation might be the fact that brokers do not believe that they can earn extra profits by investing themselves. It is also interesting to note that six of the respondents did not check either yes or no

for the question of whether they have some investments in the stock market or not, and that they are all male respondents.

- The number of subjects who responded to all twenty questions about market expectations was 175. The remaining 16 respondents did not provide their answers for some of the 20 questions.
- Some subjects checked more than one expectation. For example, there are cases where they wrote both down (D) and horizontal (H). In such cases the first of the multiple expectations are taken as the answer. In some cases the respondents answer the questions as D/H/U, meaning down/horizontal/up, this kind of answer is regarded as a "no idea" answer.
- 14 out of 191 respondents replied that they do not use any method to make their investments. Of these 14 subjects only one had investment in the stock market while others said that they have no investments in the stock market. Therefore only one respondent (female) who had some investments in the stock market was inclined to indicate her method for investment.

The following tables summarize the demographic profile of the brokers who replied our "survey.

Table 1 - Ages of Participants

Age-group	Percentage
26-30	%14
31-35	%52
36-40	%27
41-45	%5

Two out of 191 respondents did not give any information about their ages.

Table 2 - Educational Levels of Participants

Level	Percentage
Graduate	%11
Undergraduate	%82
High school	%7

Four out of 191 respondents did not answer this question

Table 3 - School Major of Respondents

Major	Percentage
Social and Administrative Sciences	%74
Engineering	%11
Sciences	%6
Other	%9

Brokers that had high school degrees high school degree did not check any of the above majors. Three respondents out of 191 did not provide any information about their educational levels or majors.

The Analysis of Investment Methods From The Survey

The first important finding of this survey with regard to the investment methods used by brokers is that the most popular method for stock market investment among brokers is found to be the technical analysis method, more specifically, 73 % of subjects declared that they use technical analysis as an investment tool. Fundamental analysis method also is found to be heavily used by the brokers, namely, 62% of the respondents checked this method. Since respondents were allowed to check more than one method the sum of the percentages adds up to more than 100%.

A striking result of this question is actually the fact that the third most frequently cited method among brokers is intuition or their feelings. This is quite an important result since it provides some insights into the hypothesis that the emotional aspects of investment behaviour should not be neglected when evaluating the responses of investors and portfolio managers in case certain market conditions. The fact that the behavioral finance discipline is gaining importance seems to have a very solid base, and will gain more importance in the near future also is seemingly being supported by the results of our survey. The following table gives the overall results of the survey regarding the methods of stock market investment. As seen, nearly half of the respondents rely on their feelings while they might be using other methods.

Table 4 - Investment Methods

Method	Number of respondents	% of Respondents
Fundamental Analysis	119	%62
Technical Analysis	139	%73
Rumours	56	%29
Intuition, feelings	87	%46
Following big investors	64	%34
Other	15	%8

Since the respondents are discovered to making use of more than one method for investment, the following table, which shows the number and percentage of the respondents and the total number of methods used for managing their equity portfolios is also useful.

Table 5 - Number of Methods Used By Brokers

Number of Methods Used	Number of Subjects	Percentage
0	14	%7
1	25	%13
2	51	%27
3	67	%35
4	21	%11
5	10	%5
6	3	%2

From the table above, it can be seen that brokers are generally very cautious in making their investment decisions. More specifically, 35% of the subjects rely on three methods and 27% rely on two methods, and actually 80% of respondents use at least two methods.

The percentage of methods used by male and female respondents are also shown in the table below. Although the most commonly used three methods (technical analysis, fundamental analysis and relying on pure intuition) and their rankings do not change across gender, the fourth most commonly used method used by male respondents is “following the big investors” while it is “Rumour” for women.

Table 6 - Usage of methods By Males and Females

Method	% use by Male Respondents	% use by Female Respondents
Fundamental Analysis	%72	%51
Technical Analysis	%78	%63
Rumours	%30	%27
Intuition, feelings	%47	%41
Following big investors	%38	%23
Other	%8	%8

The percentages do not add up to one (or 100%), because the subjects were allowed to tick more than one method in this question. As seen, the sum of the percentages are greater for men than women. This might be an indication of males using more methods on the average than females. In our survey the average number of methods used by men in their investment decisions is 2.73, while it is 2.14 for women. On the other hand, the median of the number of methods used by males is three, while it is two for females. A simple t-test also shows that the mean number of methods used by male and female respondents are significantly different from each other.

Another important table is provided below which shows that although the most commonly used method is technical analysis, the number of respondents who use solely fundamental analysis is slightly greater than that of technical analysis and the other methods. However the numbers are small and very close to each other, therefore the figures can not be used to reach any conclusion on this subject.

Table 7 - Exclusive Usage of Methods

Method	Number of People Using only this method
Fundamental Analysis	10
Technical Analysis	8
Rumours	2
Intuition, feelings	3
Following big investors	1
Other	1

This table reassures the fact that, invetors generally rely on more than one method while investing in the market. However this table might also be an indication of self confidence of the people using a specific portfolio management method. Fundamental analysts seem to be more confident of their method than those using other methods.

Analysis Of The 20 Expectations Questions In The Survey

In the questionnaire, the brokers were asked to answer the question in the following format:

Please write your expectation for the next day or the next session for the following cases.

1) Equity market index increases and the volume or turnover also increases.
2) Equity market index increases but the volume remains the same
3) Market is down, turnover decreases
4) Market is down turnover increases
5) Index reaches a new high of the last twelve month period
6) Index reaches a new high of the last one month period
7) Index drops to new low of the last twelve month period
8) Index drops to a new low of the last one month period
9) Index increases for the last two or more consecutive session and/or days
10) Index decreases for the last two or more consecutive session and/or days
11) Index closes lower after a volatile session or day
12) Index closes higher after a volatile session or day
13) Index drops sharply after a volatile session or day
14) Index increases sharply after a volatile session or day
15) Index increases smoothly (with low volatility) during the day/session
16) Index decreases smoothly (with low volatility) during the day/session

17) Index close higher as of the end of day or session but there is a fall towards the end of trading period

18) Index close lower as of the end of day or session but there is a rise towards the end of trading period

19) Index rises while all the stocks in the index rises accordingly (Return dispersion).

20) Index rises but some of the stocks in the index experience large increases while some of them decreases (Return dispersion)

The choices of the above questions are given as a likert type scale as follows:

a) Strongly down
b) Down
c) Horizontal
d) Up
e) Strongly Up

Values of 1 to 5 is assigned to the choices "a" through "e", for example strongly down takes a value of 1, while strongly up takes a value of 5. The respondents are also allowed to write "no idea" option. In the following table average scores and the total number of respondents who checked each option and also checked the "no idea" option are provided.

While interpreting the results of the survey, it should be kept in mind, the notion of "up" , "strongly up", "down", "strongly down" or "horizontal" expectations, the notion of high and low volatility may differ among brokers. A more detailed survey asking the brokers about their expectations by providing cases with numerical values of variables might probably provide more insights into the expectation formation process. However the researcher in this field should be informed that, the brokers do not generally have much spare time to fill in such a "detailed" survey.

Table 8 - The Summary Of Expectations

QUEST.	AVRG. SCORE	STR. DOWN	DOWN	HRZNTL	UP	STR. UP	NO IDEA	TOTAL	RANK (*)
QUEST1	4,337	5	2	2	94	84	1	188	2
QUEST2	2,984	3	61	61	50	7	4	186	11
QUEST3	2,800	13	51	87	28	6	1	186	13
QUEST4	1,640	117	43	6	16	4	2	188	20
QUEST5	3,545	11	38	8	85	36	9	187	7
QUEST6	3,580	3	34	16	104	19	9	185	6
QUEST7	2,469	40	73	19	36	11	6	185	16
QUEST8	2,486	21	95	17	37	5	10	185	15
QUEST9	3,287	4	44	37	83	10	4	182	9
QUEST10	2,750	12	72	42	48	2	6	182	14
QUEST11	2,358	8	123	20	24	1	6	182	18
QUEST12	3,678	2	17	23	129	6	6	183	5
QUEST13	1,806	82	69	11	18	0	3	183	19
QUEST14	4,181	1	11	14	80	71	5	182	3
QUEST15	3,706	1	5	58	94	19	5	182	4
QUEST16	2,418	17	86	59	13	2	4	181	17
QUEST17	2,934	6	64	32	65	0	11	178	12
QUEST18	3,256	1	45	38	78	6	11	179	10
QUEST19	4,412	1	3	11	69	93	3	180	1
QUEST20	3,329	0	6	97	57	1	14	175	8

(*) Rank According to the averages score

The most answered question is the first question and the least answered questions are the last two questions. This might be due to the fact that the respondents may get tired or bored of answering all the questions as he/she proceeds. The questions that include a volume increase or volume decrease term with an increase/decrease in the index level namely the first and the third questions, have the least number of "no idea" answer. This might be interpreted as volume change accompanied by the index change have some clear implications for the brokers about the next session or the day. The largest score is calculated as approximately 4.4 for the 19th question. More specifically, the respondents expects a bull market after a rise in all of the index constituent stocks. The overall score for this question is somewhere in between up and strongly up choices. The lowest score is taken for the 4th question, which asks the expectations after a fall in the index level accompanied by a rise in the volume. In other words, the respondents do expect a fall in the index for the next day when the index falls with a rising volume. Remembering the scaling methodology, average scores close to the number 3 which corresponds to the horizontal expectation means that no up or down expectation is formed.

The average score of the questions numbered 19, 1, and 14 is above four, this means that the expectation of the respondents for next period is "up" in these cases. The last two expectations in the above table are also regarded as important since the average score of them is below two which corresponds to the down expectation. The other 15 expectation questions are somewhere around three.

In order to find out whether the mean score for a question is different from the value of three which corresponds to the horizontal expectation, a t-test is performed for each of the twenty questions . The null hypothesis is that the mean of the average score for any question is equal to three. If the null hypothesis is rejected, then this will mean that the sample on the average does not expect a horizontal market for . An example of the t-test is given below. As seen the average score of question 19 is significantly greater than three.

Table 9 - Sample output (t-test) for Question 19

Hypothesis Testing for CEV19		
Date: 02/15/06 Time: 14:28		
Sample(adjusted): 1 182		
Included observations: 180		
Excluded observations: 2 after adjusting endpoints		
Test of Hypothesis: Mean = 3.000000		
Sample Mean = 4.438889		
Sample Std. Dev. = 0.756084		
Method	Value	Probability
t-statistic	25.53250	0.0000

A similar t-test is conducted for all the twenty questions asked to the brokers. Except for questions, 2,3, 10 and 17, all the mean of all the other questions are found to be different from zero at 0.05 significance level. More specifically the questions can be classified to three groups, namely, Up Expectation, Down Expectation and Horizontal Expectation cases as follows:

Table 10 - Cases Classification with respect to expectations

Expectation	Cases
Up	1, 5, 6, 9, 12, 14, 15, 18, 19, 20
Down	4, 7 ,8, 11, 13, 16
Horizontal	2, 3, 10, 17

In addition to calculating the average score, the total percentage of "up" and "strongly up" answers together and "down" and "strongly down" answers together is regarded as more explanatory and a better indicator of the feel of respondents about the expectation of the direction of the market.

For the first three questions with the largest average expectation score in the above table the ratio of “up” and “strongly up” answers to the total is 90%, 95% and 83% respectively. In other words, 90 % of brokers expects the index to go either up or strongly up in the next day or session if the index rises while all the stocks in the index rise (question 19). On the other hand, approximately 95% of brokers expectations are either “up” or “strongly up” for the next session or day when the market index rises with rising volume (Question1) And finally 83% of respondents checked their expectations as either “up” or “strongly up” for the question asking their expectations when “index increases sharply after a volatile session or day”.

For the two questions with the lowest expectation score, it can be seen that the percentage of the total of “down” and “strongly down” answers are 85% and 83% respectively. More specifically, 85% of the brokers expects a fall in the index after a sharp fall in the index with volatile trading period (Question 13) and approximately 83 % of brokers expects a fall in the index after a fall in the index with large turnover. The total percentage of “up” and “strongly up” and “down” and “strongly down” expectations for each of the questions are provided in the following table. The percentages in each of the columns provide insights into the relative importance of the effects of different variables related to price, volume, volatility etc on the expectations of respondents. A declaration of a strongly up or a strongly down expectation can be regarded as an indication on how confident is the respondent in each of the twenty different cases. It is quite interesting to observe for example the fact that the expectations of respondents can be very diverse in some cases. For example, in question two, the respondents were asked to tell their expectations in case of an up move with no increase in volume, almost equal percentages of brokers have found to posses "strongly up" and "strongly down" expectations. Therefore the same information seemingly leads to completely different expectations, a result that really needs to be studied further by the researchers in this field.

Table 11 - Percentage Of "Up" And "Strongly Up" Expectations

QUESTION	Total % of Up and Strongly Up Expectations	Total % of Down and Strongly Down Expectations
QUEST19	%90	%2
QUEST1	%95	%4
QUEST14	%83	%7
QUEST15	%62	%3
QUEST12	%74	%10
QUEST6	%66	%20
QUEST5	%65	%26
QUEST20	%33	%3
QUEST9	%51	%26
QUEST18	%47	%26
QUEST2	%31	%34
QUEST17	%37	%39
QUEST3	%18	%34
QUEST10	%27	%46
QUEST8	%23	%63
QUEST7	%25	%61
QUEST16	%8	%57
QUEST11	%14	%72
QUEST13	%10	%83
QUEST4	%11	%85

The survey results also are evaluated to see whether there is any difference in the expectations of male and female respondents in each of the twenty cases. The following table provide the summary results of the average expectation scores of male and female respondents. A quick look at the table reveals the fact that although there is some change in their order, the scores for each of the twenty questions are very close to each other for male and female respondents. In order to find out whether the average scores are different between males and females a t-test is performed for each of the twenty questions. Only for the question number two was the average scores found to be different. For the all other 19 questions the average scores do not significantly differ from each other. To repeat, question two was as “Equity market index increases but the volume remains the same”. The overall average for this question has been found to be 2,984 and interpreted as a “horizontal” expectation since it is close to the Number 3. Males and Females differ however, in their responses to this question. Females expect a down market (2,710) while males expect a slightly up market (3,173). The null hypothesis of the average being equal to three is not rejected for males while it is rejected for females.

Table 12 - Scores Of Female and Male Subjects Sorted Acc. To Average Score

Quest	Avrg Score (Female)	Total(female)	Avrg Score (Male)	Total (male)
1	4,347	72	4,317	105
2	2,71	72	3,173	105
3	2,736	72	2,798	105
4	1,69	72	1,567	105
5	3,672	71	3,416	105
6	3,631	70	3,52	105
7	2,358	70	2,539	105
8	2,446	70	2,57	105
9	3,414	70	3,17	104
10	2,721	70	2,802	104
11	2,362	70	2,388	103
12	3,714	70	3,639	103
13	1,757	70	1,85	103
14	4,159	70	4,182	103
15	3,58	70	3,788	103
16	2,571	70	2,337	102
17	2,924	69	2,871	101
18	3,418	69	3,174	101
19	4,435	69	4,398	101
20	3,246	70	3,333	99

Additionally, all the twenty expectations questions asked in the survey are also evaluated to see whether there is any difference between the group of respondents who use technical analysis as one of their investment tools and those who use methods other than technical analysis. The below table shows the average scores of the two groups.

Table 13 - Average Scores of Technical Analysis Users versus others

Question	Technical Analysis Users	Others
QUEST1	4,360	4,271
QUEST2	2,985	2,979
QUEST3	2,831	2,714
QUEST4	1,594	1,771
QUEST5	3,594	3,400
QUEST6	3,621	3,454
QUEST7	2,448	2,533
QUEST8	2,439	2,628
QUEST9	3,328	3,170
QUEST10	2,664	3,000
QUEST11	2,391	2,256
QUEST12	3,619	3,860
QUEST13	1,761	1,935
QUEST14	4,198	4,130
QUEST15	3,765	3,533
QUEST16	2,382	2,522
QUEST17	2,967	2,841
QUEST18	3,194	3,432
QUEST19	4,462	4,267
QUEST20	3,308	3,386

The above table indicates that the average scores of technical analysis users and the average scores of the respondents who do not use technical analysis do not significantly differ from each other. A test is performed for each of the twenty questions and the null hypothesis of equal means between the two groups can not be rejected at 5 % significance level.

CHAPTER III

EMPIRICAL ANALYSIS OF SURVEY RESULTS

Introduction

In order to ascertain whether the expectations of the brokers taken from the survey have empirical support the mean return of sessions whose previous session have different properties with regard to return, volume, volatility and return dispersion have been analysed. The analysis is done by tabulating the values of variables for each of the cases.

Session to Session Returns

Lead Lag Relations Between The Returns

The first point to consider is to compare the mean returns after a session with positive return and after a session with negative return. As seen from the following table, the mean return after sessions with positive returns is greater than zero and the mean return after sessions with negative returns is less than zero and they are significantly different from zero and from each other. The fact that the absolute value of the t statistics is lower for negative returns than that of the positive returns implies that the market expectation for positive returns is stronger after positive returns than the expectation for negative returns after negative returns.

Table 14 - Comparison of the mean returns after up and down sessions

Variable	Mean	t value	probability	Anova F Value	F-Prob
posretson	0,0024	4,9366	0,0000	24,7524	0,0000
negretson	-0,0011	-2,1469	0,0319		

Another important statistics with regard to the expected return for the next session after a positive and negative return is the ratio of positive returns to the total number of returns after an "up session" and "down session". It has been calculated that 56 % percent of the returns after an "up session" is positive and 52 % of the returns after a "down session" is negative.

Next period returns are also analysed by differentiating the first and second sessions, normal returns and returns that are high in magnitude. The following table shows the results of our analysis for all these samples.

Table 15 - Comparison of returns after up sessions, known session on session basis

variable	Count	mean	Std. Dev.	Std Err. Of Mean	F-value	F-prob.
Pretson1	1031	0,00319	0,02048	0,00064	2,3940	0,1220
Pretson2	1059	0,00167	0,02414	0,00074		
Nretson1	633	-0,00055	0,02279	0,00091	0,6874	0,4072
Nretson2	632	-0,00170	0,02651	0,00106		
Phretson	611	0,00505	0,02708	0,00110	5,8674	0,0155
Pretson	2090	0,00242	0,02242	0,00049		
Nretson	1923	-0,00109	0,02219	0,00051	0,0989	0,7532
Hnegretson	506	0,00072	0,02790	0,00012		

The positive returns are grouped as session 1 positive returns and session 2 positive returns and similarly negative returns are grouped as session 1 negative returns and session 2 negative returns. This is done to see whether it makes any difference to have a positive return in session one or session two with regard to the mean expectation of the next session return. As can be seen from the table above, the mean expected return for the next session following a session with a positive return does not differ across session number. Therefore, the mean expectation can be said to be positive after an increase in the index regardless of the session number. Note, however, that the mean return after session 1 is higher than the mean return after session 2.

The mean return expectation after a negative return also is found to be not statistically different across sessions. After a negative return comes another negative return regardless of the session number. Note however, that, the mean expected return after a negative session is smaller(more negative), after session 2, compared to session 1.

The mean return for the next session after a large increase in returns i.e. returns which are approximately one standard deviation greater than zero also is calculated to see whether it makes any difference with regard to the mean return of the next session. There are 612 such cases which are shown by the variable “phretson” standing for consecutive return after a positive high return. As seen from the above table, the mean return after a large increase is higher than the mean return after just an increase in a session. The probability value is very small (0.0155), which implies that the difference is statistically significant. The ratio of positive returns after a large increase in returns becomes equal to 59 % that supports the above conclusion.

The same conclusion can not be drawn, however, for the large negative returns. Large negative return is defined as the returns whose magnitude is one standard deviation greater than zero. As it can again be seen from the table, the mean return after a negative return and after a large negative return are statistically indifferent from each other. Contrary to what has been found for the large positive returns, the mean return after large falls is even higher than the mean return after negative returns.

The expected return for the next session after two consecutive sessions are also analysed, because there was a question asking about the expectations of broker, the general expectation from the survey was favoring the up market. The mean return after two consecutive up movements is positive and significantly greater than zero. On the other hand, the mean return after three consecutive up sessions are also found to be greater than zero but it does not seem statistically significant. See the table below:

Table 16 - The mean return after two ups and three ups

Variable	Mean	t value	probability
ikiupson	0,002294	3,4386	0,0006
ucupson	0,001582	1,85746	0,0637

Additionally the mean return after an up session and the mean return after two up sessions are compared to see whether there is any significant difference in the between them. As it can be seen from the table below, the mean return after two up sessions and after an up session are quite close to each other they are not significantly different from each other. There is no sign of a strengthening trend and there is also no no sign of any reversal.

Table 17 - Comparison of Next period returns after two ups and after positive returns

Variable	Count	mean	Std. Dev.	Std Err. Of Mean	F-value	F-prob.
Ikiupson	1170	0,0023	0,0229	0,0007	0,0236	0,8778
Pretson	2090	0,0024	0,0224	0,0005		

.

On the other hand the null hypothesis of equal means for the cases of positive returns, two consecutive positive returns and three consecutive positive returns can not be rejected. However, the mean return after a positive session is greater that the mean return after two consecutive positive sessions and this in turn is greater than the mean expected return after three consecutive positive movements, which implies some kind of a reversal.

For the negative returns the same analysis is repeated. The following output shows that the mean expected return is negative after sessions with negative returns, the mean expected return is also negative after two consecutive negative returns, but the mean gets a bit closer to zero, and the mean expected return after three consecutive negative sessions becomes a bit larger than zero, this is regarded as a sign of reversal. On the other hand, the hypothesis that all the three means are equal to each other can not be rejected as shown below.

Table 18 - Comparison of down returns with two and three consecutive down returns

variable	Count	mean	Std. Dev.	Std Err. Of Mean	F-value	F-prob.
Ucdownson	501	0.000976	0.024626	0.001100	1.603.481	0.2013
Ikidownson	1003	-0.000500	0.023959	0.000757		
Nretson	1923	-0.001086	0.022190	0.000506		

Another sample is constructed where the index reaches a new high of the last one month period. There are 301 such observations and it has been found that if the index reaches a new high of the last one month period, the mean return after such a session is greater than zero with a probability value of 0,0007. This means that the index trespassing the maximum of the last one period can be regarded an indication of a further up move. However when this sample is compared to that with positive returns only, the result is that, although the mean return after sessions where the prices closes higher than the highest of the last month is higher than the mean returns after positive sessions, the two samples are not significantly different from each other (shown below)

Table 19 - Comparison of Returns to returns after index passes the highest of the last one period.

variable	Count	mean	Std. Dev.	Std Err. Of Mean	F-value	F-prob.
PSTRETSON	2090	0.002421	0.022421	0.000490	1,5367	0,1245
BAYMAXSON	301	0.004551	0.022914	0.001321		

The mean returns after sessions where the highest of the last one year period is attained, and there have been found 109 such cases and the mean return after such sessions are found to be not significantly greater than zero, i.e. t value is found to be small with probability of 0.1570.

The mean returns after sessions where the index drops to a new minimum of the last month period the following output is obtained. The mean return after such sessions is not significantly different from zero. It is also interesting to note that, contrary to the sign of the mean return after negative sessions the mean return after sessions with the new low of the month is even found to be positive.

In addition to the above findings, there are only eight cases where the session reaches a new low of the last one month period, thus the sample is very small, however the mean return after such sessions are found to very close to zero. In fact four out eight cases are positive and the remaining four are negative, which means that reaching to a new low for the last one year period does not say anything meaningful in terms of the direction of the market for the next session.

The Lead Lag Relation Between Return And Volume

Since the brokers responding the survey seem to give special importance to changes in volume, the return series are also analysed by taking the variable volume into account. There are four cases with regard to return volume relationship, namely:

i) Return is positive volume is up

ii) Return is positive volume is down

iii) Return is negative volume is up

iv) Return is negative volume is down

The positive and negative return series are each divided into two samples where the first sample is the returns with increasing volume and the second is the return series with decreasing volume. The following table depicts the fact that the mean expected return for the next session after an up session accompanied by a rising volume and the mean return after an up session accompanied by a falling volume are found to be not statistically different from each other. This means that an increase in return accompanied by an increase in volume can not be regarded as a sign of an up market and similarly, an increase in return which is not supported by an increasing volume can not be regarded as a sign of down market for the next period. Note, however, the fact that mean return for the next session is higher if the volume is also higher. Thus there is some evidence favoring the up expectation for the next session is the increase in the index comes with an increase in volume, but it is not statistically significant. A similar conclusion is reached for the last two of the four cases above. In other words, a fall in the index accompanied by a rise in volume can not be taken as a signal of a further fall and besides, a fall in the index level with a concurrent fall in the volume is not a sign of a recovery at least for the next session.

Table 20 - Comparison of positive negative returns accompanied by a negative and positive volume change

variable	Count	mean	Std. Dev.	Std Err. Of Mean	F-value	F-prob.
PRETNVOLSON	825	0.001584	0.020947	0.000729	1,9023	0,1680
PRETPVOLSON	1265	0.002967	0.023324	0.000656		
NRETNVOLSON	1195	-0.001479	0.021902	0.000634	0,9647	0,3261
NRETPVOLSON	727	-0.000454	0.022668	0.000841		

The volume return relationship is further analysed across sessions to see if there is any difference. There has been found to be no difference across sessions in terms of the volume return lead lag relationship. Therefore, it can be concluded that, a rising or a falling volume with a rising or a falling index can not be suggested to use in forming expectations about the mean return for the next session, this conclusion is valid regardless of the session number.

It is interesting to see that the above findings about the volume return relationship is clearly contrary to the expectations of brokers in the İMKB as taken from the survey. Note that the first four questions in the survey was asking the expectations of the brokers in he following cases.

21) Equity market index increases and the volume or turnover also increases.
22) Equity market index increases but the volume remains the same
23) Market is down, turnover decreases
24) Market is down turnover increases

The average scores were 4,337, 2,984,2800 and 1,640 respectively on a 1-5 scale where one stands for strongly down expectation and five stands for strongly up expectation. The scores of the first and the fourth questions clearly say that the expectations favours a bull market if an index rise is accompanied by a rising volume, and people generally expects a down market after a session with negative return and a rising volume. The above analysis implies however that, these expectations are unfounded. More specifically it has been found that the returns after sessions with a rise in the index level and a drop in volume are not statistically different from the returns after sessions with a rise in the index level and a rise in volume and similarly a drop in the index level with a falling volume does not cause the return of the next session to be different than the case where a fall in the index is accompanied by an increase in volume.

Since the average score of the first and fourth cases indicate a clear bias of brokers, the volume issue seems to deserve to be dwelled upon longer. For this reason, as a first step, the sample with means returns greater than zero is filtered according to the volume change criteria as follows:

- The days with positive return
- The days with positive return and an at least a 25% rise in volume
- The days with positive return and at least a 50% rise in volume

As shown below, the mean return for the next session increases as one moves from a mere rise in return, to at least a 25% increase in volume accompanying the return and further to at least a 50% rise in volume and they seem to have means which are statistically significantly different from each other. This result is in accordance with the average score of the first question from the survey which implies that the market expects an up market after a rise in the index and a rise in the volume.

Note that although the returns after sessions with positive returns and at least a 50% rise in volume is greater than that of the sessions with positive returns and at least 25 % rise in volume, the difference is not statistically significant. The ratio of positive returns after a positive return is 56 %, the ratio of positive returns after a positive return and at least 25 % rise in volume is 60 %, in other words, the returns are greater if the volume rise is more visible with a positive index.

A similar test is conducted for large positive returns, i.e. returns which are approximately one standard deviation greater than zero. There are 612 such cases. The sample is also filtered according to the volume criteria as has been done, namely, the returns accompanied by a 25 % rise in volume are filtered, and then the returns accompanied by a 50 % rise in volume are filtered. As the table shows the expected mean returns for the next session are a bit higher when accompanied by a large increase in volume, but the three samples namely, large returns, large returns with 25 % increases in volume and large returns with 50% increases in volume are found to be not statistically different from each other. Thus it can be concluded that large increases accompanied by large volume increases can not be regarded as a sign of a bull market.

Similarly negative returns are also categorized as negative returns, negative returns with at least 25% rise in volume and negative returns with at least 50% rises in volume. The mean returns for the next session for each of the three cases are calculated. As shown in the output below, the mean return for the three cases are not statistically different from each other. Thus one can not conclude that the next session will be lower if the index falls and the volume rises. Note that although the largest negative mean return for the next session is obtained in case of a negative return accompanied by more than a 50% rise in volume this result does not seem to have sound statistical support. A crosscheck of the ratio of positive and negative returns also assures the same conclusion. The ratio of negative changes after a fall in the index and a rise in the volume is approximately 50% in all the three cases above, which means that volume increase brings no new information for predicting the direction of the market.

To summarize, a positive return with a high volume (a visible volume change such as at least 25 %) might be an indication of a further rise in the index. This finding is in accordance with the average score of the first question of the survey. A negative return with a rise in volume is not however, an indication of a further fall in the index. This result is clearly contrary to the findings of the fourth question in the survey.

In addition to the increases in volume, the sessions with a decrease in volume are also analysed. For example, a negative return accompanied by large volume drop (25 %) case is analysed and it has been found that this can not be regarded as recovery sign for the next period as shown below.

Table 21 - Comparison of mean returns after session with different return volume combinations

variable	Count	mean	Std. Dev.	Std Err. Of Mean	F-value	F-prob.
PSTRET	2090	0.002421	0.022421	0.000490	3,1409	0,0434
HAC_50ARTIS01	311	0.005233	0.023080	0.001309		
HAC_25ARTIS01	700	0.004204	0.023266	0.000879		
HPOSRET	611	0.005004	0.025903	0.001048	0,2167	0,8052
HPOSRET25_VOLCHG01	324	0.006057	0.025690	0.001427		
HPOSRET50_VOLCHG01	169	0.005990	0.026026	0.002002		
NEGRTE	1923	-0.001086	0.022190	0.000506	0,7419	0,4763
NEGRET_50HAC01	100	-0.003215	0.028530	0.002853		
NEGRTE_25HAC01	330	-0.000109	0.023462	0.001292		
NEGRET25_FALLINVOL01	697	-0.001953	0.021216	0.000804	0,7986	0,3716
NRETSON	1923	-0.001086	0.022190	0.000506		

This finding is contrary to the widespread belief among technical analysts who generally claim that volume drop during negative sessions should be regarded as a sign of recovery. In our survey, however, the average scores of the brokers are very close to three in case of a drop in the index accompanied by a high volume. And this result is in accordance with our empirical findings.

Additionally session to session negative returns are filtered as large negative returns, this series is also filtered according to rise, fall, large rise and large fall in volume to see the effect of volume in case of large falls in the index. As was the case for negative returns, any change in volume accompanied by a large fall does not imply any direction for the return of the next session.

Lead Lag Relation Between Return And Volatility

Since the brokers were asked to write their expectation for the next session or the next day in the case of volatile sessions, the session to session returns were also analysed to see whether volatility makes any difference for the mean return of the next session. To do that, as a first step, volatility should be defined. In our case the best proxy for volatility is assumed to be the variable "range", which is defined as the difference between maximum and minimum values attained during a trading period divided by the closing value of the period. All the positive returns are then sorted according to the magnitude of "range". Then the sample is split to half and the returns with large range values are compared with the returns with the small range values.

As seen from the following output, the mean return following the sessions with a positive return and with high volatility is found to be significantly higher than the mean return after the sessions with a positive return and low volatility.

On the other hand, by applying the same line of reasoning it has been found that the same conclusion can not be drawn for the negative returns when the negative returns with low volatility and the negative returns with high volatility are compared. The mean return for the session following a session with negative return and accompanied by high volatility is not significantly different from that of a session with a negative return and low volatility as shown below:

Table 22 - Comparison of returns after different return and volatility combinations

variable	Count	mean	Std. Dev.	Std Err. Of Mean	F-value	F-prob.
HIGHVLT	1045	0.003398	0.026181	0.000810	3,9772	0,0463
LOWVLT	1045	0.001444	0.017847	0.000552		
NRETHIGHVLT	961	-0.000832	0.026824	0.000865	0,2528	0,6152
NRETLOWVLTLTY	962	-0.001341	0.016305	0.000526		

It should be noted however that the mean return after a highly volatile and down session is higher than the mean return after low volatility. In fact, as shown below, the null hypothesis that the mean return after a negative return accompanied by high volatility is equal to zero can not be rejected, while the same null hypothesis is rejected for the mean return following the session with negative return and low volatility.

Table 23 - Comparison of volatile falls with less volatile falls

Variable	Mean	t value	probability
Nretlowvltlty	-0,00134	-2,55025	0,0109
nrethighvlt	-0,00083	-0,96126	0,3367

The perception of an upward move or a downward move in the index might be different from the point of view of brokers. Perhaps just a few points up or down market, or a very small increase or decrease are not regarded as a rise or a fall but rather it might well be regarded as a horizontal move. Therefore, in addition to analyzing mere positive returns, the mean return after sessions with high positive returns with different volatility levels are also evaluated. High positive or large positive returns is defined as return which is at least one standard deviation higher than zero level. It has been found that mean return after session with a large increase and high volatility as measured by the variable range is significantly greater than that of the sessions with large return and low volatility. The relevant Eviews output is shown below:

The same line of reasoning is applied to the case of negative returns, to see whether it makes any difference for the mean return of the next session, if there is a large fall in the index accompanied by a high volatility or low volatility. Similar to the conclusion drawn for the large negative returns, the mean return after significantly down sessions with high volatility is found to be greater than that of the sessions with low volatility and the result is statistically significant as shown below:

Table 24 - Large positive and Large negative returns with high and low ranges

variable	Count	mean	Std. Dev.	Std Err. Of Mean	F-value	F-prob.
HRETHRANGE	305	0.007198	0.030572	0.001751	3,8674	0,0497
HRETLOWRANGE	306	0.002899	0.022928	0.001311		
HNEGRETHRANGE	253	0.001978	0.032329	0.002033	4,7585	0,0296
HNEGRATELOWRANGE	253	-0.003413	0.022368	0.001406		

Lead Lag Relation Between Return And Return Dispersion

Since the brokers are asked about their views for the next period in case of high return dispersion, the sample is also analysed to see the effect of this variable on the mean return for the next period. First the positive returns are taken into account and the return dispersion is found to have no significant impact on the expected returns in case of positive returns. As seen from the output below, the sessions with a positive return and a low return dispersion is not significantly different than the sessions with positive returns and high return dispersion. This result is not in accordance with the result from the survey.

The same analysis is repeated for negative returns and a different conclusion is reached. As seen from the output below, the mean return for the next session after down sessions with low return dispersion is lower than the mean return after down sessions with high return dispersion. Put another way, if all the stocks in the index fall together, it is more probable that the next session will also close lower than the case where some stocks fall in larger percentage than some other stocks in the index.

In order to find the reasons behind the positive expectation for the next session if the return dispersion is low (the case where all the stocks rise together), large positive returns are sampled out from the positive return sample. When this was done , it was found that high returns with low return dispersion have a higher mean expected return for the next session than that of the high return and high return dispersion sessions. This means that if the session is significantly higher than the previous session and if all the stocks increase accordingly then a positive return for the next session should be expected. The expectation is less strong if all the stocks do not rise.

The variable return dispersion also seems to be a matter of concern when the large negative returns are analysed. In other words, if the all the stocks fall in a period with down movement, the mean return for the next period probably will be lower than the case where the returns of index stocks differ much from each other.

Table 25 - Comparison of Returns with different return dispersions

variable	Count	mean	Std. Dev.	Std Err. Of Mean	F-value	F-prob.
PRETHIGHRETDISP	1045	0.002795	0.027422	0.000848	0,5800	0,4464
PRETLOWRETDISP	1045	0.002048	0.015925	0.000493		
NRETHIGHRETDISP	962	0.000658	0.027535	0.000888	11,9640	0,0006
NRETLOWRETDISP	961	-0.002833	0.014857	0.000479		
HRETLOWRETDISP	305	0.006999	0.022738	0.001302	3,1834	0,0749
HIGHRETHIGHRETDIP	306	0.003097	0.030717	0.001756		
HNEGRETHIGHRETDISP	253	0.003200	0.033317	0.002095	10,1550	0,0015
HNEGRETLOWRETDISP	253	-0.004635	0.020476	0.001287		

Daily Returns

Lead Lag Relation Between The Daily Returns

The analysis of returns made above is also repeated for the daily return series, it is found that although the mean return after positive daily return is larger than the mean return after negative returns they are not significantly different from zero and from each other. It should be noted that the mean daily return is positive for both after days with positive returns and after days with negative returns. Returns approximately one standard deviation greater and less than zero are classified as large positive and large negative returns. Although the mean return after positive returns and after large positive returns are not significantly different from each other, the mean return after large positive returns are found to be significantly greater than zero as shown below:

Table 26 - The mean returns after large falls and rises

Variable	Mean	t value	Probability
Largepretson	0,005541	2,553505	0,0111
Largenegretson	0,003654	1,362049	0,1744

The ratio of positive returns after large positive returns is found to be %53 which is not so promising. It is interesting to note that this ratio is even larger, more specifically, it is %53.8 after large negative returns. The mean return after large negative returns is however found to be not significantly different from zero as shown below:

The mean return after large positive and large negative returns are found to be almost equal to each other. Additionally, the mean return after large negative returns are greater than the mean return after negative returns they are found to be not significantly different from each other. In summary, it can be said that, if there is a large rise or large fall in the index during a day, for the next day, the probability of observing a positive return is greater than the probability of observing a negative return, and the magnitude of positive returns is higher after large positive returns.

The mean return following two consecutive positive returns is found to be positive and even a bit greater than the mean return after just a positive return, but the difference is not statistically significant as shown below:

On the other hand, the mean return after there consecutive up days is found to be negative and thus less than the mean return after two consecutive days of positive movement. Therefore the probability of a reversal is higher when there are three consecutive up movements. This conclusion is approved when the ratio of up moves after two consecutive ups and three consecutive ups are compared. The ratio of ups is around 52 % after two consecutive ups, and the ratio of ups is only around 46 % after three consecutive ups. For the down days, the same analysis is repeated and it has been found that the mean daily return after a negative session, two consecutive negative sessions and three consecutive negative sessions are not significantly different from each other. There is also no sign of reversal after two or three down movements in daily index return series.

The mean returns are also analysed to see whether it makes any difference if the index closes above the maximum of the closing values of the last one month period or if the index falls further down to the minimum values of the last one month period. As shown from the output below the mean return after index closes above the maximum of the last one month period is greater than the mean return after the index falls below the minimum of the last one month period. The results can be regarded as almost statistically significant. In other words one can strongly expect a positive market after the index closes above the maximum of the last one month period, but the expected movement is horizontal after the index falls below the minimum of the last one month period.

The expected daily return after the index passes above the maximum of the last one year period and the expected return after the index closes lower than the minimum of the last one year period are found to be not significantly greater than zero.

Table 27 - Comparison of returns with different characteristics

Variable	Count	mean	Std. Dev.	Std Err. Of Mean	F-value	F-prob.
IKIUPSONRASI	532	0.002368	0.032012	0.001388	0,0526	0,8186
PRETSON	1037	0.001966	0.033332	0.001035		
BAYMAX	300	0.006342	0.030004	0.001732	3,5387	0,0606
BAYMIN	170	4.80E-06	0.042634	0.003270		

Effect Of Lagged Volume On Daily Returns

The return volume relation is also analysed for daily positive and negative returns separately. The positive returns are classified into four categories on the basis of volume change as described below:

- The mean returns after the days with positive change in the index and a positive change in volume
- The mean returns after days with positive change in the index and at least a 25% rise in volume
- The mean returns after the days with positive change in the index and a negative change in volume
- The mean returns after days with positive change in the index and at least a 25% fall in volume

The above mentioned four samples are compared to see the effect of a change or a large change in volume accompanied by a rise or a large rise in the index. As shown in the output below, the mean returns are all not significantly different from each other. Thus volume increase or decrease with a rising index does not make any difference in terms of the expectation of the next day's return

Large positive returns are also classified to see whether volume change does make any difference in cases of large increases in the index

- The mean returns after days with large positive change and an increase in volume

- The mean returns after days with large positive change and at least a 25 % increase in volume
- The mean returns after days with large positive change and a fall in volume
- The mean returns after days with large positive change and at least a 25 % fall in volume

The mean returns for the next day in all of the above four cases are found to be almost equal to each other as was the case for positive returns. Therefore the result doesn't change depending on the magnitude of up movements.

Table 28 - Comparison of rises in the index with different volume combinations

Variable	Count	mean	Std. Dev.	Std Err. Of Mean	F-value	F-prob.
POSRET_25VOLRISESON01	322	0.002641	0.035666	0.001988	1,0680	0,3615
POSRETPOSVOLSON	526	0.003631	0.033565	0.001463		
PRET_25FALLINVOL01	279	0.000468	0.032403	0.001940		
PRETNEGVOL	510	0.000302	0.033049	0.001463		
LARGEPOSRET_25FALLINVOL0	279	0.000468	0.032403	0.001940	1,0116	0,3869
LARGEPOSRET_25RISEINVOL0	162	0.004553	0.039008	0.003065		
LARGEPOSRETNEGVOL	222	0.005606	0.036722	0.002465		
LARGEPOSRETPOSVOL	98	0.005393	0.043389	0.004383		

The same volume classification is done for negative returns and large negative returns. The mean returns after large negative returns and negative returns are found to be almost equal to each other regardless of the volume change.

Therefore, in general the amount and sign of change in volume does not give any significant signal about the direction of the market for the next day.

The Effect Of Lagged Volatility And Return Dispersion On Daily Returns

Daily positive and negative returns are each sorted according to the magnitude of the volatility as measured by range which is defined as the difference between the minimum value and maximum value divided by closing value. Each sample is divided into two samples, one being the highest half of the range, and the other belonging to the lowest half of the range. Although the mean return of positive session with high volatility is found to be higher than the mean return after the days with positive return and low volatility, the two samples are not significantly different from each other. For the negative returns, on the other hand, the mean returns after negative return and low range and the mean return after negative return and high range are found to be almost equal to each other. Moreover, the range is found to have no significant effect on the next period's return in case of large positive and large negative returns.

As has been done for finding the effect of lagged volume change on the returns, daily positive and negative returns also are divided into two groups, one being the days with high return dispersion and the others with low return dispersion. As shown in the output below, the mean return after the days with positive returns and high return dispersion is found to be significantly greater than the mean return for the days following the days with positive returns and low return dispersion. This result can be interpreted as that positive returns accompanied by high return dispersion have a higher expected mean return for the next day than the positive returns with low return dispersion.

On the other hand, as shown below, although the mean return after days where all the stocks in the index fall in close magnitude to each other is larger than the days where the return dispersion is high, they are not significantly different from each other.

Table 29 - Positive and Negative Returns with different return dispersions

variable	Count	mean	Std. Dev.	Std Err. Of Mean	F-value	F-prob.
PRHIGHDISP	519	0.004384	0.039745	0.001745	5,4915	0,0193
PRLOWDISP	518	-0.000457	0.025137	0.001104		
NRETHRETDISP	488	0.002081	0.039864	0.001805	1,1179	0,2906
NEGRETLOWRD	486	-0.000155	0.024228	0.001099		

Similarly, for large negative movements in the index, it has been found that the magnitude of the return dispersion does not significantly affect the expected return for the next day.

Summary And Comparison Of The Empirical Analysis With The Expectations Survey

The expectation for the return of the next period for each of the cases with different session to session and daily return, volume, volatility and return dispersion composition are summarized as follows:

i) An increase in the index during a session implies an up market for the next session; however, an increase in the index for the day doesn’t imply an up market for the next day.

ii) A decrease in the index implies a down market for session to session returns, however a decrease in the index during the day doesn’t imply a down market for the next day.

iii) A large increase in the index implies an up market for session to session and for daily returns, and the mean expected return for the next session is significantly higher than the mean expected return for session to session to session mentioned in case 1, while the difference is not statistically significant for daily returns.

iv) A large decrease in the index does not imply a down market, the mean expected return is after large negative returns is negative but not significantly lower than zero. On the other hand the mean daily return after large falls in the index is positive implying some kind of reversal, but the null hypothesis of a zero mean return can not be rejected.

v) The mean return after an increase in the index and increase in volume is higher but not significantly different than the mean return after an increase in the index and a decrease in volume. Therefore, volume change does not produce any up or down signal for the next period. This conclusion is valid for both daily and session to session returns.

vi) The mean daily return and session to session return after a fall in the index and a fall in volume is not significantly larger than the mean return after a fall in the index and a rise in volume.

vii) The mean return after an increase in the index and at least 25 % rise in volume is significantly higher than the mean return after an increase in the index level. This is not the case however for daily returns.

viii) The mean return after a large increase in the index accompanied by an at least 25 % rise in volume is not statistically different than the mean return after a large increase in the index. The same is true for daily returns.

ix) The mean return after a fall in the index and at least 25 % rise in volume is not statistically different than the mean return after a fall in the index. This is true for both session to session and daily returns.

x) The mean return after a fall in the index accompanied by an least 25 % fall in volume is not statistically different than the mean return after a mere fall in the index. This is true for both daily and session to session returns

xi) The mean return after sessions with positive return and high volatility is significantly higher than the mean return after sessions with positive return and low volatility.

xii) The mean return after down sessions with high volatility is not significantly different than the mean return after down sessions with low volatility.

xiii) The mean return after up sessions with low return dispersion is not significantly different than the mean return after up sessions with high return dispersion

xiv) The mean return after a fall in the index accompanied by low return dispersion is significantly lower than the mean return after a fall in the index with high return dispersion

xv) The mean return after a large rise in the index with high volatility is significantly greater than the mean return after a large rise in the index with low volatility.

xvi) The mean return after a large rise in the index accompanied by low return dispersion is significantly higher than the mean return after a large rise in the index with high return dispersion

xvii) The mean return after a large fall in the index with high volatility is significantly greater than the mean return after a large fall with low volatility.

xviii) The mean return after a large fall in the index with high return dispersion is significantly greater than the mean return after a fall in the index accompanied by a low return dispersion.

xix) The mean return after two consecutive up sessions is lower but not statistically different than the mean return after an up session. The mean return after three up sessions are lower and not statistically higher than zero.

xx) The mean return after three consecutive down sessions is higher than the mean return after two consecutive down sessions which is higher than the mean return after a down session. But all the three cases do not have statistically different expected returns.

xxi) The mean return after a session which closes higher than the highest of the last month period is higher but not statistically different from the mean return after an increase in a session.

xxii) The mean return after a session which closes lower than the minimum of the last month period is positive but not significantly different from zero and it is not also significantly different than the mean return after negative returns.

xxiii) There are not many cases found for the condition that the index closes over the maximum of the last one year period and under the minimum of the last one year period. A general finding is that the mean return after the index passes the one year maximum or one year minimum is not significantly different from zero.

CHAPTER IV

DISTRIBUTIONAL PROPERTIES OF RETURNS

Introduction

Returns of the İMKB30 are analyzed for three different time periods, namely, from session to session, from one day to the other and from the first session to the second session of the same day.

Session To Session Returns

First, the distributional properties of the index return series calculated from the closing values and the average values of the index from one session to the other is analyzed. For the whole period between March 1997 and April 2005, a total of 4013 session to session return series are computed.

The following figure and table shows the distribution of the close to close return series by session. The series is called “Ret30seans”. Although the shape of the distribution resembles the normal distribution, the Jarque-Bera statistics is highly significant implying that the distribution is not normally distributed. It should be noted however, that, this statistics can be misleading in some cases. Although not shown here the series is also found be non-normal by using the well-known Kolmogorov Smirnov test. The high kurtosis value is also an indication of thick tails and non-normality. The series is found to have negative skewness, a property which is generally observed in most of the equity markets.

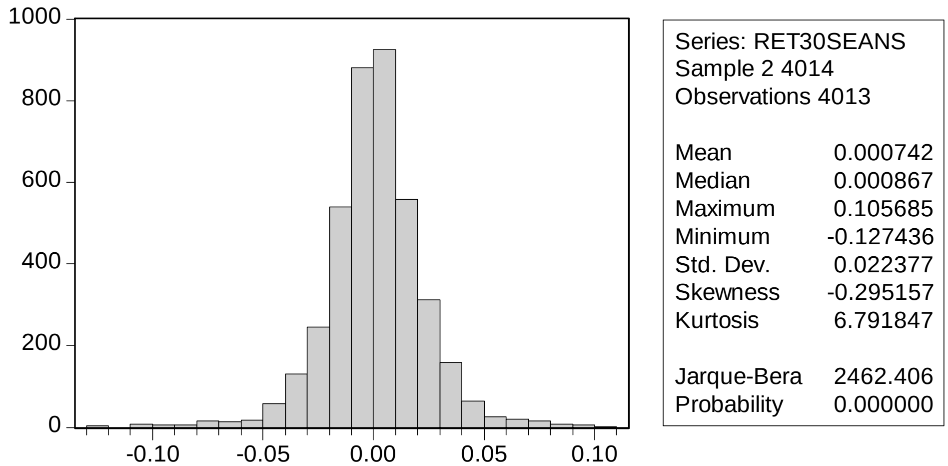

Figure 1 - Distribution of session to session returns

Table 30 - Empirical Distribution Test for session to session returns

Empirical Distribution Test for RET30SEANS
Hypothesis: Normal
Sample(adjusted): 2 4014
Included observations: 4013 after adjusting endpoints

Method	Value	Adj. Value	Probability
Lilliefors (D)	0.064862	NA	0.0000
Cramer-von Mises (W2)	6.210677	6.211451	0.0000
Watson (U2)	6.197138	6.197910	0.0000
Anderson-Darling	37.73300	37.74006	0.0000

Method: Maximum Likelihood - d.f. corrected (Exact Solution)

Parameter	Value	Std. Error	z-Statistic	Prob.
MU	0.000742	0.000353	2.101009	0.0356
SIGMA	0.022377	0.000250	89.57678	0.0000

Log likelihood	9554.653	Mean dependent var.	0.000742
No. of Coefficients	2	S.D. dependent var.	0.022377

The mean of the whole series is positive but it needs to be tested to see whether the average return from one session to the other is zero or not. The result of this hypothesis test is shown in the table below:

Table 31 - T-test for session to session returns

Hypothesis Testing for RET30SEANS
Date: 07/29/05 Time: 10:54
Sample(adjusted): 2 4014
Included observations: 4013 after adjusting endpoints
Test of Hypothesis: Mean = 0.000000

Sample Mean = 0.000742
Sample Std. Dev. = 0.022377

Method	Value	Probability
t-statistic	2.101009	0.0357

As seen from the table above, the null hypothesis that the session to session return is equal to zero is rejected at 5 % significance level, however, it is very close to 5 percent significance level.

An alternative view of the distribution is provided in the table below. The empirical distribution of returns in tabular form is quite helpful in assessing the shape of the distribution. As seen from the table session to session returns are concentrated within the +- 2 % interval. However the extreme values are quite striking. For example, the number of returns below –10% is 12, which is quite high compared to a normal distribution with the same mean and the same standard deviation. If the distribution were normal the probability of observing such a return (i.e. less than or equal to –10 %) is around 0.00001 (1/100,000). In other words, the number of observations within this interval would virtually be zero, if the distribution were normal. From the empirical distribution however, there are 12 observations out of 4013 falling out of the interval, which means that the probability of such an event is considerably higher when compared to normal distribution.

Table 32 - Frequency Distribution of Session to Session Returns
Descriptive Statistics for RET30SEANS
Categorized by values of RET30SEANS
Date: 07/29/05 Time: 11:03
Sample(adjusted): 2 4014
Included observations: 4013 after adjusting
Endpoints

RET30SEANS	Mean	Std. Dev.	Obs.
[-0.14, -0.12)	-0.123173	0.003147	4
[-0.12, -0.1)	-0.105076	0.003392	8
[-0.1, -0.08)	-0.089003	0.006207	12
[-0.08, -0.06)	-0.070701	0.004946	28
[-0.06, -0.04)	-0.046664	0.005224	75
[-0.04, -0.02)	-0.027927	0.005800	376
[-0.02, 0)	-0.008462	0.005357	1420
[0, 0.02)	0.008370	0.005419	1481
[0.02, 0.04)	0.027751	0.005715	470
[0.04, 0.06)	0.046936	0.005295	89
[0.06, 0.08)	0.069198	0.006236	36
[0.08, 0.1)	0.088403	0.005559	12
[0.1, 0.12)	0.103347	0.003305	2
All	0.000742	0.022377	4013

In addition to the calculation of session close to session close returns, session to session average return series is also calculated and graphed to see any differences in empirical distributions .

The following figure shows the empirical distribution average return series. As seen from the figure, the average return series is found to be closer to normal than the distribution of close to close returns. A brief comparison of Figure 1 with (close to close session returns) with Figure 2 shows that the Jarque Berra Statistics is much lower for average returns and the kurtosis value is also lower. Additionally, the skewness is less and values of extreme values are also lower for the average return series. Therefore although the distribution of average returns is still non-normal it is closer to normal than the close to close returns.

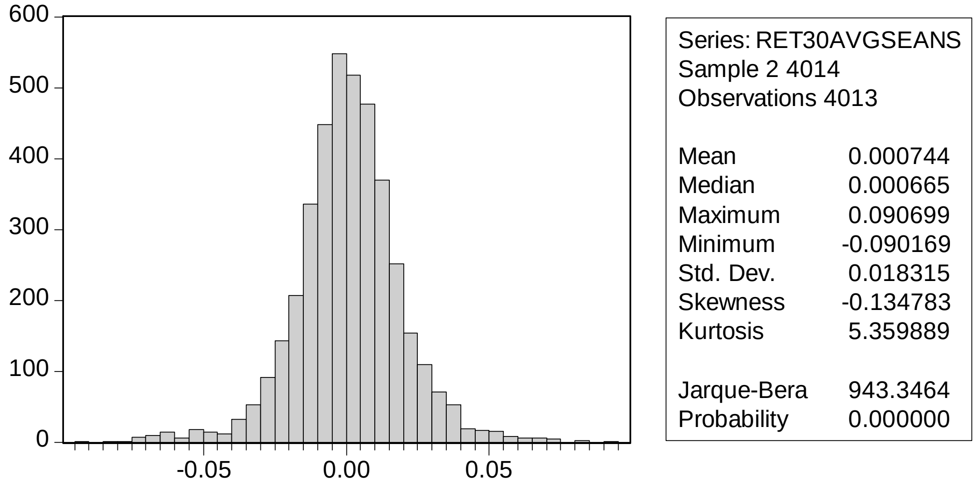

Figure 2 - Distribution of average session to session returns

Intraday Session to Session Returns

Session to session returns should be analyzed carefully in the sense that the time period between the first session and the second session held during a trading day is only two hours while the time period between the two consecutive sessions from one day to the other (i.e. the second session of day T and the first session of day T+1 is 18 hours during weekdays and 66 hours for the weekend and even higher for the religious holiday periods. Therefore dynamics of return distribution might possibly change depending on the time interval between two return values. For this reason the returns between two sessions on the same day are analyzed separately from the returns between two sessions on different days. The following graph shows the distribution of close to close returns for two consecutive sessions on the same day. As seen from the graph the distribution of close to close noon returns is non-normal but closer to normal than the distribution of session to session returns. This finding is in accordance with the results reported by Masulis et al. (1995).

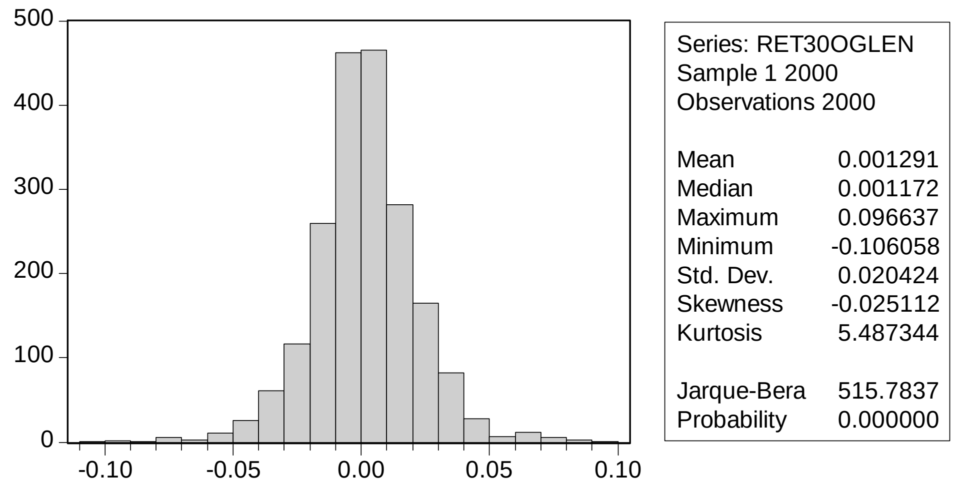

Figure 3 - Distribution of Intraday Session to session returns

It is also worth noting that the mean return of close to close noon returns is greater than the mean return of session close to session close returns. However a simple t test shows that the null hypothesis that session to session and noon returns being equal to each other can not be rejected.

Although the average of close to close noon returns is greater than the average of session close to session close returns (the average of session close to session close returns is 0.000742, the average of noon close to close returns is 0.001291), the risk measured both in terms of standard deviation (the standard deviation of close to session close returns is 0.022377, the standard deviation of noon close to close returns is 0.0204024) and in terms of empirical probabilities of extreme events (the sample size for the noon returns is almost half of the number of the session to session returns and there is only 1 observation less than –10 %, and there is no observation greater than 10 %) is less for the noon returns than that of the session to session returns. This is quite an interesting result and seemingly contrary to the classical risk return trade-off logic.

On the other hand, noon returns calculated from average values are found to have higher Jarque-Bera value and higher kurtosis compared to close to close noon return implying larger deviation from normality compared to close to close returns. In other words noon returns calculated from closing values are closer to normal than that of the average values, while session to session returns calculated from the average values are closer to normal than that of the close to close returns.

The maximum and minimum values of the averages are also quite large implying the higher probability of extreme returns. The standard deviation as measure of risk is also higher for average noon return series, while the average is higher for noon returns calculated from average values of the index. However the mean of the noon returns calculated from averages can not claimed to be significantly higher than the mean of the session to session average returns.

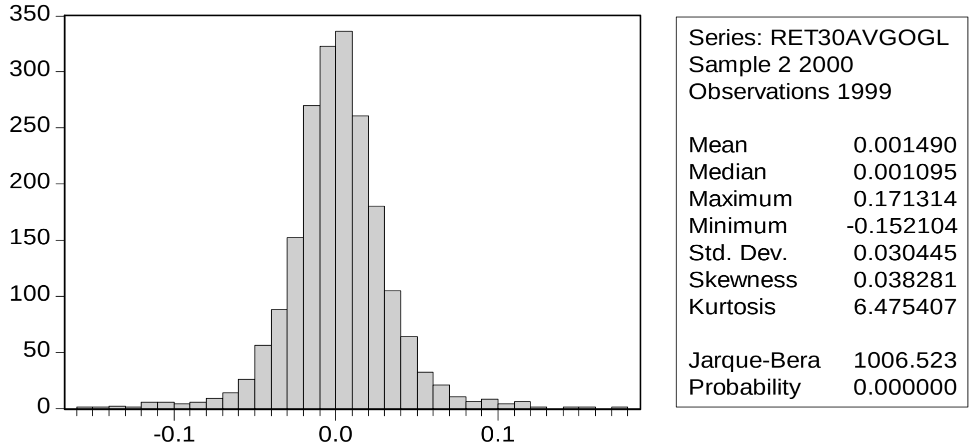

Figure 4 - Distribution of Average Intraday Session to Session Returns

Daily Returns

To complete the analysis of returns distributions, close to close and average daily returns are also calculated and graphed. The following figure shows the distribution of close to close daily returns. As can be seen from the graph, the distribution is non-normal with high kurtosis value, and high and significant Jarque Berra statistics.

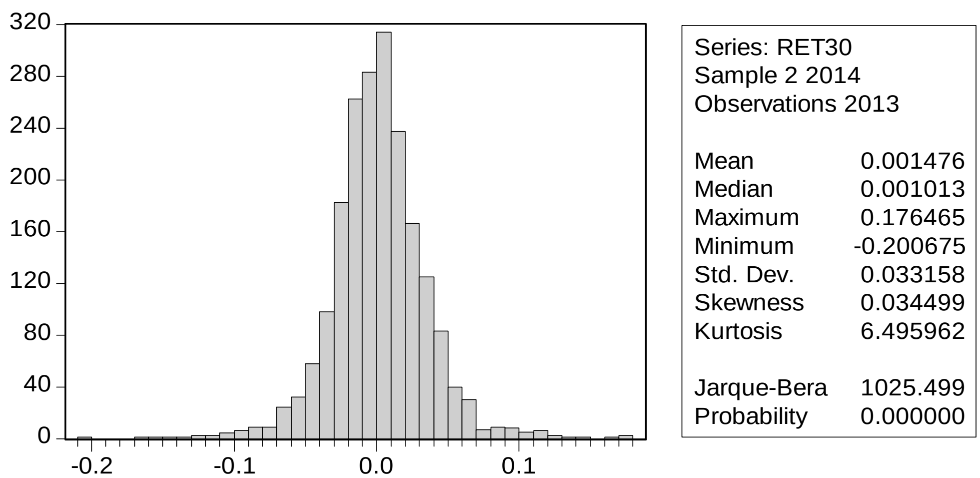

Figure 5 - Empirical Distribution of Daily Returns

The average of close to close daily return is positive but again the t value is not so high as to assure that the daily returns are higher than zero (see below table). The probability value is very close to 5% level and although the returns are positive at 5 % level of significance, care should be taken given the nonnormality of the distribution.

Table 33 - T-test for daily return

Sample(adjusted): 2 2014		
Included observations: 2013 after adjusting endpoints		
Test of Hypothesis: Mean = 0.000000		
Sample Mean = 0.001476		
Sample Std. Dev. = 0.033158		
Method	Value	Probability
t-statistic	1.997035	0.0460

The table below is also very useful in analyzing the distribution of returns. As seen from the table session to session returns are concentrated within the +- 4 % interval. The average of daily returns is higher than the session to session returns but the standard deviation is also higher, almost 50 % higher than that of the session to session return series.

From the table, it can be seen that the number of daily returns below –10 % is 13 out of 2013 observations, while the expected value of such returns should be three if the distribution was normal. On the other hand, the number of daily returns over 10% reaches 18 while this number should be two if the distribution were normal. The probability of extreme positive values for daily returns is considerably higher than session to session returns while the probability of extreme negative returns is higher for session to session returns.

Table 34 - Frequency Distribution of Daily Returns

Descriptive Statistics for RET30			
Categorized by values of RET30			
Sample(adjusted): 2 2014			
Included observations: 2013 after adjusting endpoints			
RET30	Mean	Std. Dev.	Obs.
[-0.22, -0.2)	-0.200675	NA	1
[-0.18, -0.16)	-0.163193	NA	1
[-0.16, -0.14)	-0.151398	0.009944	2
[-0.14, -0.12)	-0.128101	0.006862	3
[-0.12, -0.1)	-0.107681	0.006876	6
[-0.1, -0.08)	-0.089485	0.005211	15
[-0.08, -0.06)	-0.068280	0.005854	33
[-0.06, -0.04)	-0.048573	0.005274	90
[-0.04, -0.02)	-0.027892	0.005521	280
[-0.02, 0)	-0.009540	0.005653	545
[0, 0.02)	0.009146	0.005753	551
[0.02, 0.04)	0.029102	0.005918	291
[0.04, 0.06)	0.047824	0.005477	123
[0.06, 0.08)	0.067177	0.005533	37
[0.08, 0.1)	0.090257	0.007426	17
[0.1, 0.12)	0.108659	0.005795	11
[0.12, 0.14)	0.127119	0.008865	3
[0.14, 0.16)	0.145709	NA	1
[0.16, 0.18)	0.170659	0.008316	3
All	0.001476	0.033158	2013

The distribution of daily returns calculated from the average values of the index is closer to normal when compared to that of the close to close daily returns. As seen from the figure below, the Jarque-Bera statistics drops to almost half of the value of the value that is calculated for close to close return distribution. The extreme values are also lower compared to close to close daily returns. The daily return series calculated from averages being closer to normal than the return series calculated from daily closing values is in accordance with the findings for the session to session returns, but returns are quite different in this sense. For the noon returns, the distribution of the returns calculated from averages deviates larger from normality than that of the close to close returns.

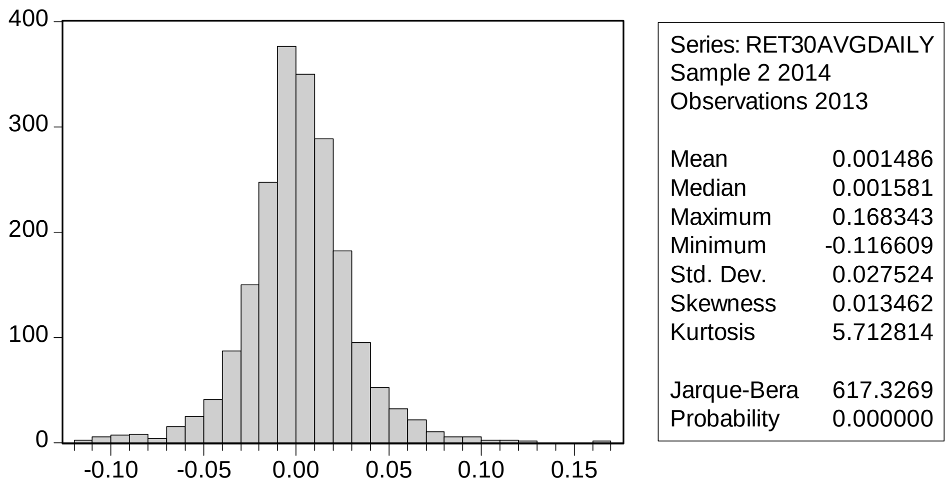

Figure 6 - Empirical Distribution of Average daily returns

Weekday Returns

When the weekday returns are examined separately, i.e. the weekend returns are excluded from the sample, the distribution of the returns calculated from the closing values is again non-normal as shown below. But this time the mean of the daily returns is almost two times higher than the the mean of daily return series including the weekend returns. The standard deviation on the other hand is found to be almost the same. As seen from the higher value of kurtosis and Jarque-Bera statistics the weekday returns deviates larger from normality than the daily returns including both the weekend and weekday returns.

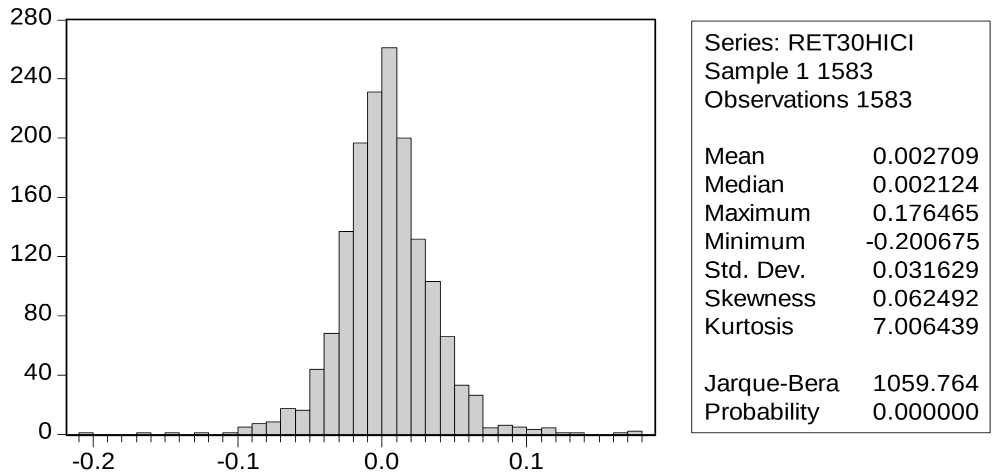

Figure 7 - Empirical Distribution of weekday returns

On the other hand, when the distribution of weekday returns calculated from the daily average values of the index was analysed, it was observed that the distribution gets closer to normal compared to close to close returns. This conclusion can easily be drawn by looking at the value of Jarque-Berra statistics for each of the return distributions.

The mean of daily weekday returns calculated by using the average values however is smaller than the mean of close to close daily weekday returns . Thus, the so called Monday effect encountered in related literature is more visible in close to close daily returns, while it is not that visible in the weekday daily return series calculated from the averages.

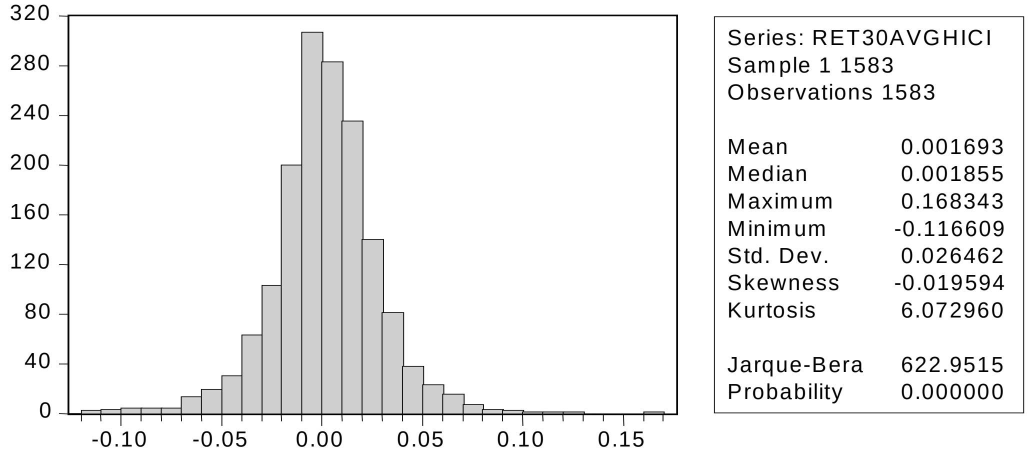

Figure 8 - Empirical Distribution Of Average Weekday Returns

Weekend Returns

Apart from the weekday returns, the distributional properties of weekend returns are also analyzed to see whether they posses something different in terms of the distributions. For this purpose, two types of returns series are calculated again, one being the close to close returns, and the other being the return series calculated from the daily average values of the index. The weekend returns calculated from the closing values and average values are depicted below. Since the whole series of close to close daily returns have positive mean and the weekday close to close returns further have a larger positive mean, the mean of the weekend close to close returns are negative as expected. The standard deviation of the close to close weekend returns is higher while the kurtosis is lower compared to the weekday returns. Although a simple t-test indicates that the means of weekday and weekend returns calculated from closing values are not equal, this result should be evaluated with some care, since the distributions are not normal.

Table 35 - T-test for the weekday and weekend returns

Test for Equality of Means Between Series			
Sample: 1 1585			
Included observations: 1585			
Method	df	Value	Probability
t-test	2011	3.208688	0.0014
Anova F-statistic	(1, 2011)	10.29568	0.0014
Analysis of Variance			
Source of Variation	df	Sum of Sq.	Mean Sq.
Between	1	0.011268	0.011268
Within	2011	2.200830	0.001094
Total	2012	2.212097	0.001099

The weekend returns calculated from the averages are still non-normal but gets even closer to normal distribution (the kurtosis is lower, and the Jarque-Bera statistics is also lower) compared to close to close weekend returns. The mean of the returns calculated from the averages is not negative but very close to zero. However, compared to weekday returns being negatively skewed, weekend returns have positive skewness.

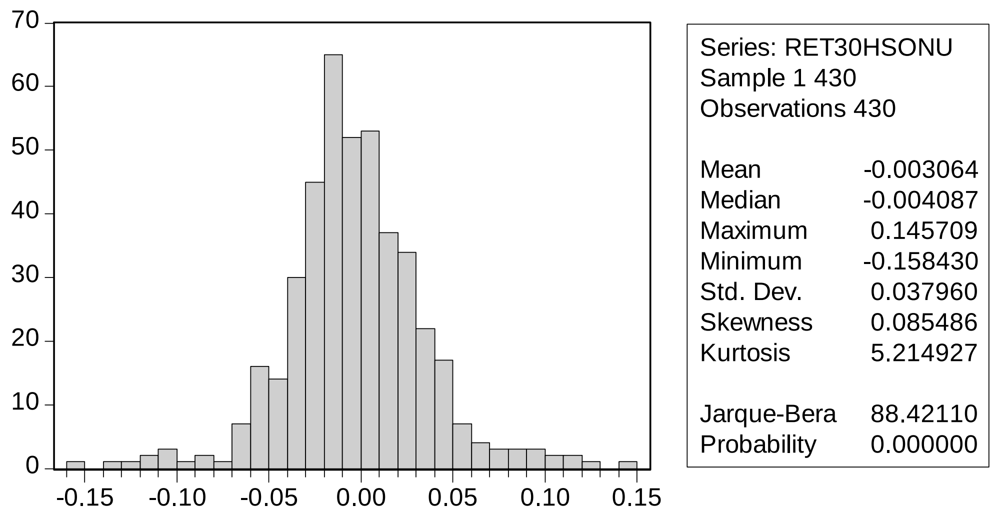

Figure 9 - Distribution of Weekend Returns (Close to Close)

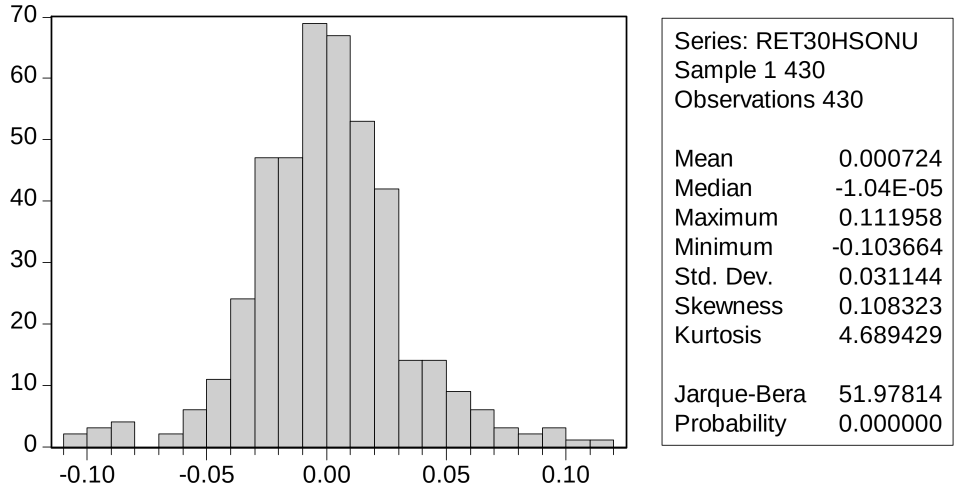

Figure 10 - Distribution of Weekend Returns (Average to Average)

Three Period (Pre-Crisis, Crisis And Post Crisis) Analysis Of The Return Series

The data spans the period from the beginning of 1997 to the first quarter 2005. As known, during the period investigated, Turkey experienced a severe financial crisis in which the Turkish Lira (TL) was devalued, interest rates, especially overnight rates rose substantially and the liquidity of the bond market almost diminished. During crisis periods as such the relationships among market indicators may differ due to changes in portfolio compositions and the changes in the risk appetites of the investors. Therefore, in order to see whether there is any change in the return distributions, the whole period is divided into three sub-periods, namely the pre-crisis period, the crisis period and the-post crisis period.

The last financial market crisis in Turkey may be assumed to have begun in November 2000 with a sudden increase in demand for cash TL. Investors especially foreign investors rushed to buy foreign currency due to the fear of devaluation. The Turkish Central Bank tried to meet the demand for foreign currency both by selling from reserves and by increasing the interest rates. This did not help however and in February 2001 the TL was devalued substantially and the currency anchoring regime was abandoned and the TL was allowed to float freely. The propagations of the crisis was felt after February of 2001, since financial markets calmed down gradually. The overnight interest rates and the exchange rates are very useful to pick the different phases of the crisis. The first upward spike after a relatively long period of decline in overnight rates was observed in November 2000. While there was no big upward move in exchange rates at that time, the crisis is assumed to have started on that date since there was a very large amount of foreign currency demand during that period. The crisis deepened in February 2001 when the TL was devalued substantially. The Ruling Turkish Coalition Government invited Mr. Kemal Derviş, a former World Bank vice-president to take control of the economy. After his appointment as a minister in charge of almost all the major economic and financial units of Turkish Economy, the markets gradually calmed down. After May 2001, the volatility of overnight rates and the exchange rates fell substantially. In summary , the whole period in question is divided into three subperiods described below:

The Pre-crisis Period : March 1997-October 2000

The Crisis Period : November 2000-April 2001

The Post-crisis Period: May 2001-April 2005

Pre-Crisis Period

The distribution of daily returns session returns and noon returns calculated from the closing values (close to close returns) and average values of the index (average to average returns) are all analyzed during the pre-crisis period. All the series are non-normal as seen from the table below. The lower value of Jarque Bera statistics and kurtosis for the average daily returns and session to session returns imply that the distribution of daily returns and session to session returns calculated from average values is closer to normal than the that of the close to close returns. The noon returns however are an exception. The noon returns calculated from averages are more non-normal than the distribution of noon returns calculated from the closing values.

Table 36 - Close To Close Returns

	RET30	RET30OGLEN	RET30SEANS
Mean	0.002657	0.001173	0.001337
Median	0.001123	0.000965	0.001645
Maximum	0.161132	0.087287	0.105685
Minimum	-0.163193	-0.106058	-0.120521
Std. Dev.	0.036149	0.022136	0.024676
Skewness	-0.016541	-0.088443	-0.357692
Kurtosis	4.940202	4.696859	5.599751
Jarque-Bera	140.8917	108.1781	542.2567
Probability	0.000000	0.000000	0.000000
Sum	2.385594	1.046276	2.392862
Sum Sq. Dev.	1.172152	0.436593	1.089336
Observations	898	892	1790

The mean of close to close returns, namely session to session (ret30seans), noon(ret30oglen) and daily returns (ret30) are all positive in the pre-crisis period. The mean daily close to close return during the pre-crisis period is higher than the mean daily return for the whole period, while the noon and session returns in the pre-crisis period are very close to the values for the whole period. The standard deviations of all the close to close return series are also very close to the standard deviation for the same return series for the whole period. Moreover the means of the return series seem to be in logical order; in other words, the mean daily return is the highest, the mean session to session return comes second and the mean of the noon returns comes third. The average time between daily returns is greater than the average time between session to session returns which is again greater than the time between noon returns. The standard deviations follow the same order.

The following table displays the relevant statistics for the average return series in the pre-crisis period.

Table 37 - Average Returns

	RET30AVRG	RET30AVRGOGL	RET30AVRGSNS
Mean	0.002659	0.002679	0.001336
Median	0.003290	0.000988	0.001368
Maximum	0.111958	0.127266	0.073016
Minimum	-0.105206	-0.152104	-0.083807
Std. Dev.	0.030603	0.033792	0.020249
Skewness	-0.051631	-0.073832	-0.206475
Kurtosis	4.400094	5.086841	4.330624
Jarque-Bera	73.74552	162.4853	144.7728
Probability	0.000000	0.000000	0.000000
Sum	2.387365	2.386907	2.392313
Sum Sq. Dev.	0.840054	1.016269	0.733544
Observations	898	891	1790

As seen, the mean returns are all positive, but more specifically the mean of the noon returns calculated from the average values are considerably higher than the mean of close to close noon returns during the pre-crisis period. The volatility of the noon average returns are also higher than the close to close returns in the pre-crisis period.

The Crisis Period

The crisis period starts on the first day of October 2000 and is assumed to end as of the end of April 2001. The selection of this date is of course questionable. However, the end of April 2001 is considered to be an important turning point because the chairman of the IMF at that time announced a 10 billion USD amount of support as a loan to Turkey. The extreme figures during the crisis are most visible in the daily close to close returns, they are as high as almost an 18 % up and as low as more than a 20 % down.

Table 38 - Close To Close Returns

	RET30	RET30OGLEN	RET30SEANS
Mean	-0.000514	0.003736	-0.000338
Median	-0.004024	0.004873	3.53E-05
Maximum	0.176465	0.096637	0.101010
Minimum	-0.200675	-0.096875	-0.127436
Std. Dev.	0.055319	0.031946	0.034657
Skewness	0.097812	-0.179220	-0.286329
Kurtosis	5.282969	4.169614	4.927225
Jarque-Bera	25.59480	7.295314	39.24239
Probability	0.000003	0.026052	0.000000
Sum	-0.060097	0.437115	-0.078795
Sum Sq. Dev.	0.354978	0.118380	0.278660
Observations	117	117	233

The most prominent feature of the statistics is the negative mean returns for daily and session to session returns. In addition to the fall in mean daily and session to session returns, the risk measured by standard deviation and the magnitude of extreme figures increased substantially for these two return series. It is interesting however to observe a positive and even larger mean for noon returns calculated from closing values compared to the pre-crisis period. The crisis period characterized by a positive return in noon returns and negative daily and session to session returns. The risk of the noon returns measured by the standard deviation and the magnitude of extreme values is also interestingly lower than the daily and session to session returns. The statistics of returns calculated from the average values of the index is displayed below. This time all the return series including the noon returns have all negative mean values. The standard deviation of all the three return series show that the largest volatility belongs to the noon return series.

Table 39 - Average Returns

	RET30AVRG	RET30AVRGOGLEN	RET30AVRGSNS
Mean	-0.000627	-0.000764	-0.000301
Median	-0.000825	-0.002005	-0.000446
Maximum	0.168343	0.171314	0.090699
Minimum	-0.116609	-0.149465	-0.090169
Std. Dev.	0.044934	0.050168	0.029077
Skewness	0.240273	0.336051	0.045757
Kurtosis	4.728987	4.935083	3.942261
Jarque-Bera	15.56488	20.28196	8.700899
Probability	0.000417	0.000039	0.012901
Sum	-0.072674	-0.088617	-0.070075
Sum Sq. Dev.	0.232191	0.289434	0.196152
Observations	116	116	233

The shape of the distribution of average returns is again non-normal but the it approaches to normal distribution .

The Post Crisis Period

The return statistics of the post-crisis period is provided below. As in the pre-crisis and crisis period, the distributions do exhibit high kurtosis and high Jarque-Berra statistics implying non-normality. One important point to note is that, the distribution is more non-normal in the post crisis period compared to the pre-crisis period. This is probably due to the fact the effects of the crisis did not fade immediately.

The mean returns are positive and the standard deviations are almost half the crisis period for both the close to close series and return series calculated from the averages. The post crisis period is also characterized by the low values of extreme values across all return series. The distribution of returns after the exclusion of the full year 2001 did again give rise to similar conclusion.

Table 40 - Close To Close Returns

	RET30	RET30OGLEN	RET30SEANS
Mean	0.000647	0.001109	0.000332
Median	0.001394	0.001176	0.000506
Maximum	0.121478	0.089157	0.096065
Minimum	-0.135893	-0.078544	-0.121099
Std. Dev.	0.026165	0.016762	0.017934
Skewness	0.055043	0.115539	-0.100387
Kurtosis	5.448219	5.212299	7.201719
Jarque-Bera	249.7452	204.2974	1465.715
Probability	0.000000	0.000000	0.000000
Sum	0.645456	1.098808	0.659295
Sum Sq. Dev.	0.682567	0.278156	0.639088
Observations	998	991	1988

Table 41 - Average Returns

	RET30AVRG	RET30AVRGOGL	RET30AVRGSEANS
Mean	0.000681	0.000676	0.000336
Median	0.000770	0.001232	0.000235
Maximum	0.077760	0.101278	0.064736
Minimum	-0.092663	-0.122930	-0.073438
Std. Dev.	0.021212	0.023415	0.014440
Skewness	-0.104727	-0.041524	-0.115867
Kurtosis	4.694428	5.665933	5.444463
Jarque-Bera	121.0922	293.4564	499.4105
Probability	0.000000	0.000000	0.000000
Sum	0.678499	0.669570	0.668098
Sum Sq. Dev.	0.448140	0.542242	0.414297
Observations	997	990	1988

The daily returns and session to session returns calculated from averages again approach normal distribution compared to the ones calculated from closing values. Noon returns again exhibit the same trend as in the pre-crisis and crisis period in which the distribution of noon returns calculated from the closing values are closer to normal distribution than those calculated from the average values. The noon return series calculated by using the average values have the highest volatility both in terms of the standard deviation statistics and the magnitude of extreme values during the post crisis period.

A Comparison of the Pre-crisis, Crisis and Post-crisis Periods

The null hypothesis that the mean returns during the pre-crisis period, crisis period and post-crisis period are all equal is tested. As shown in the following output the null hypothesis can not be rejected.

Table 42 - Comparison of Pre-crisis, Crisis and Post Crisis Returns

Test for Equality of Means Between Series				
Included observations: 2000				
Method		df	Value	Probability
Anova F-statistic		(2, 2010)	1.092016	0.3357
Analysis of Variance				
Source of Variation		df	Sum of Sq.	Mean Sq.
Between		2	0.002401	0.001201
Within		2010	2.209696	0.001099
Total		2012	2.212097	0.001099
Category Statistics				
Variable	Count	Mean	Std. Dev.	Std. Err. of Mean
RET30K	117	-0.000514	0.055319	0.005114
RET30KO	898	0.002657	0.036149	0.001206
R30GKS1	998	0.000647	0.026165	0.000828
All	2013	0.001476	0.033158	0.000739

On the other hand, the variances of pre-crisis period, crisis period and post crisis period exhibit a different pattern. As shown below, the null hypothesis of equal variances is rejected. This is regarded as an indication of regime shifts in the volatility of the return series. Alternatively, changing variances also mean that the return series is heteroskedastic.

Table 43 - Variance Equality Test Results

Test for Equality of Variances Between Series				
Sample: 1 2000				
Included observations: 2000				
Method		df	Value	Probability
Bartlett		2	198.0780	0.0000
Levene		(2, 2010)	57.70799	0.0000
Brown-Forsythe		(2, 2010)	56.49496	0.0000
Category Statistics				
Variable	Count	Std. Dev.	Mean Abs. Mean Diff.	Mean Abs. Median Diff.
R30GKS1	998	0.026165	0.019592	0.019583
RET30K	117	0.055319	0.040180	0.040020
RET30KO	898	0.036149	0.026793	0.026748
All	2013	0.033158	0.024001	0.023967
Bartlett weighted standard deviation: 0.033156				

Since the number of observations during the crisis period which is assumed to last only six months is small. The test is repeated by including the sample before the crisis and after the crisis and a similar result is obtained, in the sense that, the variances of the two periods are significantly different from each other as shown below.

Table 44 - Variance Equality Test for the Pre-crisis and Post crisis period

Test for Equality of Variances Between Series					
Date: 05/09/06 Time: 16:14					
Sample: 1 2000					
Included observations: 2000					
Method		df	Value	Probability	
F-test		(997, 897)	1.908717	0.0000	
Siegel-Tukey			6.059410	0.0000	
Bartlett		1	98.02337	0.0000	
Levene		(1, 1894)	56.13344	0.0000	
Brown-Forsythe		(1, 1894)	55.23939	0.0000	
Category Statistics					
			Mean Abs.	Mean Abs.	Mean Tukey-
Variable	Count	Std. Dev.	Mean Diff.	Median Diff.	Siegel Rank
R30GKS1	998	0.026165	0.019592	0.019583	1020.769
RET30KO	898	0.036149	0.026793	0.026748	868.1837
All	1896	0.031301	0.023003	0.022977	948.5000
Bartlett weighted standard deviation: 0.031293					

CHAPTER V

TIME SERIES ANALYSIS

Introduction

A time series can be defined a sequence of observations of a variable that are taken at different periodic time points.Time series can be taken at intervals very close to each other such as seconds or minutes, the data can also be taken at longer time intervals such as days, weeks, months, quarters, years etc. In time series models the basic motivation is that information contained in the past values of a variable or a number of variables might be useful for forecasting future values of the same variable or some other variable. In this thesis the main focus in on the linear time series analysis. In other words, the time series models within the scope of this thesis are aimed at expressing a time series as a linear function of its past values or a linear function of its past and the past values of other explanatory variables. The econometric models within the context of linear time series analysis are mainly Autoregressive (AR) Models, Moving Average (MA) Models, Autoregressive Moving Average (ARMA) Models and Fractionally Integrated Autoregressive and Moving Average models (ARFIMA) and Vector Auto Regressive Models (VAR). VAR model differs from the others due to the inclusion of more than one variable in the analysis.

The main assumption of time series analysis is that the variables should be stationary. If the series is not stationary the statistical inference tests using the classical regression model, even with large-samples become almost meaningless. Therefore stationarity of any series should be tested before progressing in time series modeling.

A General Overview of Autoregressive (AR) and Moving Average Models

As the name implies in the autoregressive (AR) time series model, an observation at time “t” can be written as a function of previous observations. For example the following formula expresses the realization of the series at time “t” in terms of a constant and the observation at time “t-1” and an error term.

$$y_t = \mu + \phi y_{t-1} + e_t \tag{2}$$

where μ is an intercept parameter (constant term), ϕ is an unknown parameter to be estimated and the e_t is the error term which is assumed to be independent and identically distributed (iid) with mean zero and constant variance. In the above formulation the variable denoted by y_t is modeled solely as a function of its lagged value. Since the error term is assumed to be iid with a finite mean and variance, it is called white noise. It should be noted that in time series analysis, the error term need not be distributed normally, it is sufficient for error term to have a well defined distribution. If a random variable having white noise property is normally distributed with mean zero and a constant variance it is termed as Gaussian white noise.

The equation above is called as autoregressive time-series model of first order (AR(1)), since Y depends on its past value in the previous period plus a random disturbance. Since the time series is not deterministic there should a disturbance term in the equation that is called noise term, error, random shock or residual. According to the AR(1) model, the current value of the random variable y_t is centered around $\mu + \phi y_{t-1}$. The AR(1) model also assures that the current value of the variable "y" is not correlated with its previous values other than y_{t-1}. The amount of deviation of y_t around this value can be expressed by the variability of the error term. The variability of the error term is measured by the variance which is assumed to have some constant value say, σ^2.

Autoregressive processes need not be of first order, in other words, the variable Y may depend on its earlier values. The statistical model of an autoregressive process of order p, denoted as AR(p) can be written as:

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-1} + \dots\dots\dots + \phi_p Y_{t-p} + e_t$$

Here again current value of the variable Y is dependent on its p lagged values and conditional on this assumption, Y_t is not correlated with Y_{t-i} where i > p. Therefore, to determine the order of AR process starting from the first lag significance of each of the newly added parameters ϕ_i should be checked. The order p of the AR process is chosen so that the parameter ϕ_i is not equal to zero for i <= p, and zero for i > p.

Another class of time series models is the Moving Average Processes where the Y is expressed as a function of past errors. The moving average representation can best be understood by the following reasoning. Assume that in a stock market information arrival is random which is the case actually. However further assume that the information is not fully absorbed in one trading period (be it a session or a trading day). This implies that the price change next day can be written as;

$$Y_{t+1} = \alpha_1 e_t + e_{t+1}$$

where e_{t+1} is the random disturbance due to the information arrival at time t+1 and the $\alpha_1 e_t$ is the effect of the yesterday information arrival on today's return. As seen the above representation is a moving average process. Generally a moving average process of order q is written as follows:

$$Y = \mu + e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2} + \alpha_3 e_{t-3} + \alpha_q e_{t-q}$$

where u_i's are random disturbances with mean zero and constant variance σ_e^2 and α_i's are unknown parameters. The above functional form says that the variable Y can be written as function of its first second and up to q th lag of past errors

It is also interesting to note that the above moving average representation can also be obtained from AR processes. To illustrate, let's have the following AR(1) process

$$Y_t = \mu + \phi Y_{t-1} + e_t$$

For convenience the intercept parameter μ can be assumed to be equal to zero, which means that the mean of the time series variable is zero. This adjustment does not affect the variances and the covariances of the time series Yt. So we have;

$$Y_t = \phi_1 Y_{t-1} + e_t \; (*)$$

Writing the same formula for Y_{t-1}, the formula becomes;

$$Y_{t-1} = \phi_1 Y_{t-2} + e_{t-1}$$

When above equation is substituted to the equation (*) ; the following representation can be obtained.

$$Y_t = \phi_1 (\phi_1 Y_{t-2} + e_{t-1}) + e_t$$

$$Y_t = \phi_1^2 \; Y_{t-2} + \phi_1 e_{t-1} + e_t$$

Similarly, the above formula can be written in terms of Y_{t-3} and e_i's as follows

$$Y_t = \phi_1^3 \; Y_{t-3} + \phi_1^2 \; e_{t-2} + \phi_1 e_{t-1} + e_t$$

Using the same logic, by repeated substitution it can easily be found that,

$$Y_t = \phi^i Y_{t-1} + \sum_{i=0}^{\infty} \phi^i e_t$$

As one goes further back to previous lags, the first term drops out due to the fact that -1<α1 (stationarity condition). Then Y_t can be denoted as an infinite weighted sum of uncorrelated random disturbance e_t and its lagged values e_{t-i} as follows:

$$Y_t = \sum_{i=0}^{\infty} \phi^i e_t$$

This formulation is called the Moving Average representation (MA) of the AR(1) process and actually any AR process can be represented as an infinite weighted sum of the uncorrelated random disturbances.

In most cases neither of the AR or MA representation may be sufficient, therefore it is very common to use time series models that contain both AR and MA components together, which is called an ARMA(p,q) or ARIMA(p,q) model where p is the order of AR and q is the order of MA model and I stands for integration (differencing) to assure stationarity of the series. The algebraic formula for the ARMA(p,q) model is:

$$Y_t = \mu + +\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + ... + \phi_q Y_{t-p} + \alpha_1 e_t + \alpha_1 e_{t-1} + \alpha_1 e_{t-2} + ... + \alpha_1 e_{t-q}$$

the intercept parameter μ is related to the mean of Y_t and errors denoted by e_i's are uncorrelated random variables with mean zero and a constant variance

In time series analysis model, the most important step is to identify the model. In other words, to identify the appropriate structure (AR, MA or ARMA) and order of model. After identifying the model, the coefficients should be estimated. Box and Jenkins (1970) suggest that appropriate model structure and the order of model should be decided by looking at autocorrelation function (acf) and partial autocorrelation function (pcf) plots. The coefficients of AR models can be estimated by least squares regression. The estimation of MA or ARMA parameters is a more complicated and actually requires an iterative procedure. Finally the model should be checked by ensuring that the residuals are random and the estimated parameters are statistically significant or not. While fitting the model, it is generally suggested that fewest possible number of parameters should be included, because simple is usually better. In time series analysis the best fit can be determined by evaluating the significance of Box-Pierce Q statistics for residuals and also by checking the Akaike Information Criterion(AIC) value, which is written in functional form as follows:

$$\text{AIC} = Log\left[V\left(1+2n/N\right)\right]$$

Where V is the variance of model residuals, N is the length of the time series, and n = p +q, p being the order of AR and q being the order of MA. The best model has the one with minimum value of AIC. The model with the minimum AIV value does not guarantee the fact that residuals are white noise. The randomness of residuals is tested by using the Box-Pierce statistics. Therefore, in identifying the model one should give equal importance to having an approximately minimum value for AIC and producing random residuals. The order of the autoregressive and moving average processes should be chosen such that the number or AR and MA terms should not be large due to the risk of high correlation (collinearity) among regressors (AR and MA terms)

The classical AR, MA and ARMA models are usually very helpful when dealing with series with short memory; in other words the order of AR, MA and ARMA models should not go back to the remote past of a series. Short memory means that the effect of a shock eventually dies out and the length of time that this effect ceases to exist should not be long. Otherwise the series is said to have the long memory or persistence problem, which is documented by financial researchers in the near past. In efforts to tackle the problem of persistence a class of fractional models called Autoregressive fractional integration models (ARFIMA) were introduced by Granger and Joyeux (1980) and Hosking (1981). Long memory or the persistence problem means that an event occurred in the past may affect quite a number future outcomes.

Stationarity

Stationarity is the fundamental assumption of time series analysis. A time series is said to be strictly stationary if any consecutive sample of observations taken from any part of the series random variables follow the same probability density function. This condition is usually thought to be very difficult to verify empirically. In time series analyses, usually the series are assumed to be weakly stationary. A time series is said to be weakly stationary if it has a constant mean (expected value), constant variance and constant lag k auto covariance. More specifically, a stochastic process (Y_t) is weakly stationary or *covariance stationary* if its mean, variance and covariance remain constant over time. That is:

$$E(Y_t) = E(Y_{t-s}) = \mu_Y$$

$$Var(Y_t) = Var(Y_{t-k}) = E\left[(Y_t - \mu_Y)^2\right] = E\left[(Y_{t-k} - \mu_Y)^2\right] = \sigma^2{}_Y$$

$$Cov(Y_t, Y_{t-k}) = \gamma_k = E\left[(Y_t - \mu_Y)(Y_{t-k} - \mu_Y)\right]$$

for all values of t, k

In the finance literature it is common to assume that the series are weakly stationary. In accordance with this assumption, stationarity refers to weak stationarity in this thesis. A related function which is used in the modeling of time series is called the *autocorrelation function* and is given by:

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

where the symbol in nominator stands for the lag k covariance and the symbol in the denominator is the variance of the series. For example, white noise is the simplest stationary process. Defining the error terms as u_t, will result in the following representation.

$$E(u_t) = 0 \qquad E(u_t u_{t+h}) \begin{array}{l} = \sigma^2 \text{ if } h = 0 \\ = 0 \text{ otherwise} \end{array}$$

In other words, the white noise process has a zero mean and a constant variance and no autocovariance.

The first test for stationarity can be done by visual inspection of the autocorrelations plot (correlogram). In the Box-Jenkins approach, for example, if the correlations are high and decline slowly, then the series is said to be nonstationary. Besides visual inspection, stationarity of a series can be tested by the Augmented Dickey Fuller(ADF) test and Philips Perron(PP) Tests.

ADF tests and PP tests for all the variables included in our time series analysis showed that relevant statistics are negative and large in magnitude (less than the critical values). Therefore the series are all said to be stationary. In other words the null hypothesis of unit root is rejected. In all of the stationary tests the series was assumed to have a constant mean with no time trend. The same conclusion is reached even if an extra term for the time trend is added. In the next section the e-views outputs for stationary tests are provided.

Cochrane(2005) provides an excellent explanation with regard to the issue of stationarity in time series analysis as follows:

"Stationarity is often misunderstood. For example, if the conditional covariances of a series vary over time, as in ARCH models, the series can still be stationary. The definition merely requires that the unconditional covariances are not a function of time. Many people use "nonstationary" interchangeably with "has a unit root". That is one form of nonstationarity, but there are lots of others. Many people say a series is "nonstationary" if it has breaks in trends, or if one thinks that the time-series process changed over time. If we think that the trend-break or structural shift occurs at one point in time, no matter how history comes out, they are right. However, if a series is subject to occasional stochastic trend breaks or shifts in structure, then the unconditional covariances will no longer have a time index, and the series can be stationary."

This explanation is quite important for the stationarity of the IMKB30 stock index return series, due to the fact that, the variance before, during and after the crisis year of 2001 have been found to be significantly different from each other. Although this finding seems to be contradictory to the conclusion of the unit root tests, as Cochrane puts it the series can not be claimed to be nonstationary by just taking into account the significant differences in the second moments.

Statistical Properties Of Autoregressive Models

In this section the basic properties of AR(1) and AR(2) models will be provided and a generalization of these models to the AR(p) process will be given. The formulas and derivations are a summary of what can generally be found in classical time series textbooks.

Ar(1) Model

The mean of an AR(1) process can be found by taking the expectation of both sides of the classical AR(1) equation below:

$$y_t = \phi_0 + \phi_1 y_{t-1} + e_t$$

since the error term is assumed to be white noise, its expected value is zero. The expected values of y_t and y_{t-1} are equal to the same constant due to the stationarity assumption. The expected value of a constant is equal to itself. Therefore denoting the expected value of y_t by μ we obtain

$$\mu = \frac{\phi_0}{1-\phi_1}$$

The variance of an AR(1) process can be found by taking the variance of both sides of the equation below

$$y_t = \phi_0 + \phi_1 y_{t-1} + e_t$$

$$Var(y_t) = Var(\phi_0 + \phi_1 y_{t-1} + e_t)$$

The variance of the constant term ϕ_0 above is zero. The error term e_t and y_{t-1} are independent, therefore the equation becomes

$$Var(y_t) = \phi_1^2 Var(y_{t-1}) + Var(e_t)$$

Since the series is stationary, the variance is time invariant. Therefore

$$Var(y_t) = Var(y_{t-1}) = \sigma_y^2$$

Replacing σ_y^2 for $Var(y_t)$ and $Var(y_{t-1})$ and σ_e^2 for variance of the error term we obtain;

$$\sigma_y^2 = \phi_1^2 \sigma_y^2 + \sigma_e^2$$

solving for σ_y^2 we get;

$$\sigma_y^2 = \frac{\sigma_e^2}{1-\phi_1^2}$$

Similarly, the covariance of one period lagged values of y (i.e. y_t and y_{t-1}) can easily be derived by using the following well known formula for covariance;

$$Cov(y_{t,}, y_{t-1}) = E[y_t - E(y_t)][y_{t-1} - E(y_{t-1})]$$

Assuming that, $E(y_t) = E(y_{t-1}) = 0$

$$Cov(y_{t,}, y_{t-1}) = E[y_t y_{t-1}]$$

writing y_t in terms of y_{t-1}, we have,

$$Cov(y_{t,}, y_{t-1}) = E[(\phi_1 y_{t-1} + e_t) y_{t-1}]$$

$$Cov(y_{t,}, y_{t-1}) = \phi_1 E(y_{t-1}^2) + E(e_t y_{t-1})$$

Since the error term e_t and y_{t-1} are uncorrelated, we obtain,

$$Cov(y_{t,}, y_{t-1}) = \phi_1 \sigma_y^2$$

Due to the stationarity assumption which assures that the mean, variance and the covariance of a stationary series is constant across time, this covariance value must be the same for all random variables that are one period apart.

The autocovariance of variables that are two periods apart can easily be calculated by applying the same logic.

$$Cov(y_{t,}, y_{t-2}) = E[y_t - E(y_t)][y_{t-2} - E(y_{t-2})]$$

Here again, $E(y_t) = E(y_{t-2}) = 0$

$$Cov(y_{t,}, y_{t-2}) = E[y_t y_{t-2}]$$

writing yt in terms of y_{t-2}, and making necessary calculations we finally obtain,

$$Cov(y_{t,}, y_{t-2}) = \phi_1^2 \sigma_y^2$$

The autocorrelation function of an AR(1) model is found by the formula for correlation. Correlation is calculated by the formula below

$$Corr(y_{t,} y_{t-k}) = \rho_k = \frac{\text{cov}(y_t, y_{t-k})}{\sqrt{(Var(y_t)\sqrt{Var(y_{t-k})}}}$$

By using the terms calculated above for covariance and variance we get,

$$Corr(y_{t,} y_{t-k}) = \rho_k = \frac{\phi_1^k \sigma_y^2}{\sqrt{\sigma_y^2 \sqrt{\sigma_y^2}}} = \frac{\phi_1^k \sigma_y^2}{\sigma_y^2} = \phi_1^k$$

This result says that the ACF (Corr(yt,yt-k)) of a stationary series decays exponentially with the rate ϕ_1^k. If $\phi_1 > 0$, then the plot of ACF of an AR(1) model shows a smooth exponential decay.. On the other hand, if $\phi_1 < 0$, then the plot alternates between negative and positive values and still decays.

Higher Order AR Models

AR(2) model is represented with the following formulation,

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t$$

To find the mean of y_t we take the expectation of both sides of the above equation as it was done for the AR(1) process. The same reasoning applies here, the expected value of any stationary series across all time points are all equal and the expected value of the error term is zero since it is assumed to be white noise. The following formula is finally obtained for the mean of the series

$$\mu = \frac{\phi_0}{1 - \phi_1 - \phi_2}$$

The variance of an AR(2) process can be found using the expression found for the mean of the series. Writing ϕ_0 in terms of μ, ϕ_1, ϕ_2 and inserting this into the AR(2) equation below

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t$$

we get,

$$y_t - \mu = \phi_1 (y_{t-1} - \mu) + \phi_2 (y_{t-2} - \mu) + e_t$$

At this point let's multiply both sides of the above equation with $(y_{t-k} - \mu)$;

$$(y_t - \mu)(y_{t-k} - \mu) = \phi_1 (y_{t-1} - \mu)(y_{t-k} - \mu) + \phi_2 (y_{t-2} - \mu)(y_{t-k} - \mu) + e_t (y_{t-k} - \mu)$$

When we take the expected value of both sides;

$$Cov(y_t, y_{t-k}) = \phi_1 Cov(y_{t-1}, y_{t-k}) + \phi_2 Cov(y_{t-2}, y_{t-k}) + Cov(e_t, y_{t-k})$$

The last term is zero, since the error term and the series y is independent.

Dividing both sides of the equation by the variance of y_t the following formula is obtained.

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}$$

At this point, let $\rho_0 = 1$, meaning that the correlation of y_t with y_t is by definition equals to one. Then, let k=1 then the formula becomes;

$$\rho_1 = \phi_1 \rho_0 + \phi_2 \rho_{-1}$$

Since the series is assumed to be stationary ρ_{-1} is equal to ρ_1, thus the above formula can be written as;

$$\rho_1 = \frac{\phi_1}{1 - \phi_2}$$

After finding the correlation for k=0 and k=1 it is then easy to find correlation for any value greater than or equal to 2. In mathematical terms;

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \text{ for } k \geq 2$$

Taking the terms on the right of the above equation to the left the formula becomes,

$$\rho_k - \phi_1 \rho_{k-1} - \phi_2 \rho_{k-2} = 0$$

Define L as the lag operator(back shift operator), then the above equation can be written as follows:

$$(1-\phi_1 L-\phi_2 L^2)\rho_k = 0$$

This means that ACF of a stationary AR(2) series can be written as a polynomial of second order. From ordinary algebra, a polynomial of second order defined as $ax^2 + bx + c$ has two characteristic roots and these roots can be found by the following formula

$$x_{1,2} = \frac{-b \pm \sqrt{\Delta}}{2a}$$

where

$$\Delta = b^2 - 4ac$$

If the term Δ is positive then the polynomial has real roots which means that the polynomial can be factored as $(1-x_1L)(1-x_2L)$. This implies that the autocorrelation function is a mixture of two exponential decays. If, on the other hand, the term Δ is negative then the polynomial has complex valued characteristic roots. In this case the plot of ACF shows kind of a damping sinusoidal wave. The types of cases are usually encountered if the series has cyclical components. The characteristic roots of the processes are very important in the sense that they indicate whether the series is stationary or not. More specifically, if the absolute value of the characteristic root(s) of an AR(p) is less than one meaning that the series is stationary, otherwise the series is said to be nonstationary.

The results for ACF of the AR(2) process can be generalized to the general AR(p) model. The ACF can be written as

$$(1-\phi_1 L-\phi_2 L^2-\phi_3 L^3-\ldots\ldots\ldots\ldots\phi_p L^p)\rho_k=0$$

The above equation is a polynomial of order p, therefore it has p roots. The plot of ACF depends on the nature of the roots of the polynomial.

Statistical Properties Of Moving Average (MA) Models

As explained in the previous section moving average(MA) models can be regarded as an infinite order AR model. In general, a moving average process represents time series observations as weighted average of random disturbances. The functional form of a general MA(q) process is provided below

$$Y=\mu+e_t+\alpha_1 e_{t-1}+\alpha_2 e_{t-2}+\alpha_3 e_{t-3}+\alpha_q e_{t-q}$$

The stationarity of the series is again the most fundamental assumption of MA processes as in AR processes. In order to show this let's consider the MA(1) process below:

$$Y_t = \mu + e_t + \alpha_1 e_{t-1}$$

The mean of the variable Y_t can be found by taking the expectation of both sides of the above equation. Since the error term represented by e_t is assumed to be white noise its expected value is zero and the expected value of a constant (μ) is equal to itself, the expected value of the series can be written as; $E(Y_t) = \mu$.

This value does not depend on time, so the expected value of the variable is equal to a constant denoted by the symbol μ. Remembering that stationary series have constant mean, variance and covariance, the constant mean in our case indicates that the series satisfies one of the three conditions of weak stationarity.

Following the same steps as it was done for the AR processes, now let's find the variance of the MA(1) process. The variance of a random variable is given by;

$$Var(Y_t) = E[Y_t - E(Y_t)]^2$$

Since $E(Y_t) = \mu$. the above equation can be written as;

$$Var(Y_t) = E[\mu + e_t + \alpha_1 e_{t-1} - \mu]^2$$

the term μ cancels out and the equation becomes.

$$Var(Y_t) = E[e_t^2 + 2e_t e_{t-1} + \alpha_1^2 e_{t-1}^2]$$

At this point, the property that the error terms represented by the letter "e" are independent and identically distributed (white noise) give rise to the conclusion that the expected value of the product of the error terms e_t and e_{t-1} is zero. Let the variance of the error terms be represented by σ_e. Then the formula for the variance of the series can be written as follows;

$$Var(Y_t) = \sigma_y^2 = \sigma_e^2 + \alpha_1^2 \sigma_e^2$$

The formula for the variance of the series in question does not have any index that changes with time. So the variance of y_t is always equal to a value that is a function of α_1 and σ_e. The first of these two terms is constant, the other term which is the variance of the random error is also constant by definition, since the error term is assumed to be white noise at the first start. White noise means that the error terms belong to a well defined distribution with mean zero and a constant variance. Therefore the second condition that is the variance being constant in weak stationarity assumption is also satisfied Now, to test the third and the last condition of stationarity let us calculate the covariance of the series y_t. The covariance between Y_t and Y_{t-1} can be found as follows:

$$Cov(y_{t,}, y_{t-1}) = E[y_t - E(y_t)][y_{t-1} - E(y_{t-1})]$$

We know that, $Y_t = \mu + e_t + \alpha_1 e_{t-1}$ and $Y_{t-1} = \mu + e_{t-1} + \alpha_1 e_{t-2}$,

Inserting this formulas into the covariance formula, the equation becomes

$$Cov(y_{t,}, y_{t-1}) = \gamma_1 = E\left[(e_t + \alpha_1 e_{t-1})(e_{t-1} + \alpha_1 e_{t-2})\right]$$

$$Cov(y_{t,}, y_{t-1}) = \gamma_1 = \alpha_1 \sigma_e^2$$

Note that one lag covariance is again a function of the variance of the error term and thus it is time invariant.

The covariance between Y_t and Y_{t-2} can be found by using the same logic as follows:

$$Cov(y_{t,}, y_{t-2}) = \gamma_2 = E\left[(e_t + \alpha_1 e_{t-1})(e_{t-2} + \alpha_1 e_{t-3})\right]$$

$$Cov(y_{t,}, y_{t-2}) = \gamma_2 = E\left[(e_t e_{t-2} + \alpha_1 e_{t-1} e_{t-2} + \alpha_1^2 e_{t-1} e_{t-3} + \alpha_1 e_t e_{t-3})\right]$$

Since the error terms are iid, the expected value of all the terms are zero. Thus the following formula is obtained,

$$Cov(y_{t,}, y_{t-2}) = \gamma_2 = 0$$

Applying the same reasoning it can be shown that covariance of the MA(1) process for all lags greater 1 is equal to zero. Thus the autocorrelation function for the MA(1) process is

$$Corr(y_{t,} y_{t-k}) = \frac{\gamma_k}{\gamma_0} = \frac{\alpha_1}{1+\alpha_1^2} \quad \text{for k=1}$$

$$Corr(y_{t,} y_{t-k}) = \frac{\gamma_k}{\gamma_0} = 0 \quad \text{for k>1}$$

Thus the third condition for stationarity which is the covariance being constant, is also satisfied. So the series represented by MA(1) modeling is stationary.

Another important property of MA process is clearly visible after the above formulations. This is the fact that the autocorrelation function becomes equal to zero after the order of the process. It can be shown that the autocorrelation function of a MA(q) process is zero after q lags. At this point let's find the mean, variance and covariance of MA(2) process. The functional form of a MA(2) process is as follows:

$$Y_t = \mu + e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2}$$

The mean of the time series can again be found by taking the expectation of both sides and it can be found that $E(Y_t) = \mu$.

As the formula implies the mean of MA(2) is constant and time invariant.The variance of MA(2) process can also be found by using the following formula,

$$Var(Y_t) = E[Y_t - E(Y_t)]^2$$

$$Var(Y_t) = E[\mu + e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2} - \mu]^2$$

Since the error terms are iid, all the cross product terms drops out since the expected values of all the cross products of error terms belonging to different point in time is zero. Thus the formula becomes;

$$Var(Y_t) = \sigma_y^2 = \sigma_e^2 + \alpha_1^2\sigma_e^2 + \alpha_2^2\sigma_e^2$$

The variance of the MA(2) process does not depend on t, therefore it is constant across all points in time

One period lag covariance of the MA(2) process can be found by using the well known formula for covariance below:

$$Cov(y_{t,}, y_{t-1}) = E[y_t - E(y_t)][y_{t-1} - E(y_{t-1})]$$

Using the fact that,

$E(Y_t) = \mu$, $Y_t = \mu + e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2}$ and $Y_{t-1} = \mu + e_{t-1} + \alpha_1 e_{t-2} + \alpha_2 e_{t-2}$,

And inserting these formulas into the covariance formula we get,

$$Cov(y_{t,}, y_{t-1}) = \gamma_1 = E[(e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2})(e_{t-1} + \alpha_1 e_{t-2} + \alpha_2 e_{t-3})]$$

$$Cov(y_{t,}, y_{t-1}) = \gamma_1 = \alpha_1 \sigma_e^2 + \alpha_1 \alpha_2 \sigma_e^2$$

$$Cov(y_{t,}, y_{t-1}) = \gamma_1 = \sigma_e^2 (\alpha_1 + \alpha_1 \alpha_2)$$

Thus the one period apart covariance is a function of some constants (coefficients of MA terms) and the variance of the error term which is also assumed to be constant since the error terms are assumed to be white noise.

Similarly two period covariance is found as follows:

$$Cov(y_{t,}, y_{t-2}) = E[y_t - E(y_t)][y_{t-2} - E(y_{t-1})]$$

inserting the formula for y_t and y_{t-2} will result in;

$$Cov(y_{t,}, y_{t-2}) = \gamma_{21} = E[(e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2})(e_{t-2} + \alpha_1 e_{t-3} + \alpha_2 e_{t-4})]$$

Using the fact the error term is iid and thus the expected value of cross product of the error terms belonging to different points in time is zero we get the following result

$$Cov(y_{t,}, y_{t-2}) = \gamma_2 = \alpha_2 \sigma_e^2$$

Two period apart covariance is also time invariant which assures that the series is stationary.

The three period apart covariance of the MA(2) process can be written as;

$$Cov(y_{t,}, y_{t-3}) = \gamma_3 = E\left[(e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2})(e_{t-3} + \alpha_1 e_{t-4} + \alpha_2 e_{t-5})\right]$$

Note that the terms in two parentheses that is multiplied by each other contain error terms belonging to different time periods, since the error terms are iid the expected value of their products are all equal to zero, thus the equation becomes;

$$Cov(y_{t,}, y_{t-3}) = \gamma_3 = 0$$

It can be shown that the covariances at all legs greater than two are all equal to zero for MA(2) processes. The corresponding autocorrelation function of the MA(2) is given below.

$$\rho_1 = \frac{\alpha_1(1+\alpha_2)}{1+\alpha_1^2+\alpha_2^2}$$

$$\rho_2 = \frac{\alpha_1}{1+\alpha_1^2+\alpha_2^2}$$

$\rho_k = 0$ for k>2

So MA(2) process has a memory of two periods long. The mean variance and covariances of one period, two period and more than two period values of the series are all time invariant, meaning that the MA(2) process satisfies all the three conditions of stationary series.The autocorrelation function for the general MA(q) process is given by;

$$\rho_k = \sum_{i=0}^{q-k} \alpha_i \alpha_{i+k} \quad \text{for k=0,1,2,3....q}$$

$$\rho_k = 0 \qquad\qquad \text{for k>q}$$

Note that an AR process can be represented as an infinite sum of random errors which is called as "moving average representation". The reverse is true, in the sense some MA processes satisfying a certain condition which is called "invertibility" can be converted to an infinite order AR process.

To show that, let us consider a MA(1) process with no constant term that is its mean is zero. This assumption is made to simplify the proof, it does not affect the validity. So let's start with the classical MA(1) definition written below;

$$Y_t = e_t + \alpha_1 e_{t-1}$$

Now, let's rewrite the equation in terms of Y_t and e_{t-1} ;

$$e_t = Y_t - \alpha_1 e_{t-1}$$

By similar logic e_{t-1} can also be written in the same manner as follows:

$$e_{t-1} = Y_{t-1} - \alpha_1 e_{t-2}$$

If we substitute the above formula into the (*) formula, the equation becomes;

$$Y_t = e_t + \alpha_1 (Y_{t-1} - \alpha_1 e_{t-2})$$

$$Y_t = \alpha_1 Y_{t-1} - \alpha_1^2 e_{t-2} + e_t$$

writing e_{t-2} in terms of Y_{t-2} and e_{t-3} will result in the following formula;

$$Y_t = \alpha_1 Y_{t-1} - \alpha_1^2 (Y_{t-2} + \alpha_1 e_{t-3}) + e_t$$

$$Y_t = \alpha_1 Y_{t-1} - \alpha_1^2 Y_{t-2} + \alpha_1^3 e_{t-3} + e_t$$

Proceeding the calculations in this manner will give rise to an infinite order AR process. The only condition is that the term $\alpha_1^i e_{t-i}$ must converge to zero so that this term can be dropped out of the equation above. Therefore the absolute value of α should be less than 1. If $|\alpha| < 1$ then the process is called as "invertible".

Statistical Properties Of Autoregressive-Moving Average (Arma) Models

An ARMA(p,q) model can be written as

$$y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=0}^{q} \alpha_i e_{t-i}$$

An ARMA(p,q) process is a stationary process. It has all of its characteristic roots in the unit circle. If one or more of the characteristics roots are equal to unity, the process is integrated and it is called an Autoregressive Integrated Moving Average process (ARIMA)

As is done in AR and MA models, enough AR & MA terms should be allowed so that the error term looks like a white noise process. But special care should be given to the case of common factor or the common root problem. In other words, if the AR and MA polynomials have the same roots at some point as the lag parameter p and q are increased, then they are said to have a common root. In such cases the model is said to be over-parameterised, that means that a model with identical properties can be constructed by reducing both p and q by one (Harvey, 1981). It is not useful to further parametrise the ARMA model after facing the common root problem, because this can cause some computational problems and can also cause the coefficients of AR and/or MA terms to become meaningless. A time series Y_t can be represented by an ARMA(p,q) model a follows:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_1 y_{t-2} + ...\phi_1 y_{t-p} + e_t + \alpha_1 e_{t-1} + \alpha_1 e_{t-2} + ... + \alpha_1 e_{t-q}$$

Using lag operator and denoting it by L, the above model can be written as

$$(1 - \phi_1 L - \phi_2 L^2 - .... - \phi_2 L^p) y_t = \phi_0 + (1 - \alpha_1 L + \alpha_2 L^2 + ... + \alpha_q L^q) e_t$$

where p stands for the order of autoregressive part and q stands for the order of moving average part. There are two polynomials in the above equation namely, $(1 - \phi_1 L - \phi_2 L^2 - .... - \phi_2 L^p)$, the AR polynomial and $(1 - \alpha_1 L + \alpha_2 L^2 + ... + \alpha_q L^q)$, the MA polynomial. One important thing to note is that, the common roots of the polynomial should be different, otherwise the order (p,q) of the model can be reduced. Moreover, the characteristic equation of the ARMA(p,q) model is the AR polynomial, meaning that the solutions to the characteristic equation should be less than 1 in absolute value so as to assure the stationarity condition.

At this point let's find the general formulas for the mean, variance and the covariance of the ARMA process. In order to make the discussion simpler let us have the above formula as ARMA(1,1). The process is shown below:

$$y_t = \phi_0 + \phi_1 y_{t-1} + e_t + \alpha_1 e_{t-1}$$

The mean of the series can be found by taking the expectation of both sides:

$$E(y_t) = E[\ \phi_0 + \phi_1 y_{t-1} + e_t + \alpha_1 e_{t-1}\ ]$$

By definition the expected value of the error term is equal to zero and the expected value of a constant is itself, therefore the formula becomes;

$$E(y_t) = \phi_0 + \phi_1 E(y_{t-1})$$

Since the series is stationary its expected value should be constant across all the points in time by definition meaning that

$$E(y_t) = E(y_{t-1})$$

Thus $E(y_t)$ can be represented as;

$$E(y_t) = \mu = \frac{\phi_0}{1-\phi_1}$$

The above formula is quite simple, the mean of the series in ARMA model is exactly the same as its mean in the AR model.The variance of an ARMA(1,1) process can be found by taking the variance of both sides.

$$V(y_t) = V[\ \phi_0 + \phi_1 y_{t-1} + e_t + \alpha_1 e_{t-1}\ ]$$

Now, for simplicity, assume that the constant term denoted by ϕ_0 is equal to zero, this is somewhat equivalent to defining the series in terms of deviation form the mean, i.e. $(y_t - \mu)$, then variance formula can be written as;

$$V(y_t) = V[\ \phi_1 y_{t-1} + e_t + \alpha_1 e_{t-1}\ ]$$

Since the variance is defined as $V(y_t) = E(y_t - \mu)^2$ the variance of the series having an ARMA(1,1) process can be re-written in terms of expected values as follows:

$$Var(y_t) = E(\phi_1 y_{t-1} + e_t + \alpha_1 e_{t-1})^2$$

The above formula can be re expressed as follows:

$$Var(y_t) = \phi_1^2 Var(y_{t-1}) + 2\phi_1\alpha_1 E(y_{t-1}e_{t-1}) + \sigma_e^2 + \alpha_1^2\sigma_e^2$$

The functional form for the term $E(y_{t-1}e_{t-1})$ can be written as;

$$E(y_{t-1}e_{t-1}) = E\left[\ (\phi_1 y_{t-2} + e_{t-1} + \alpha_1 e_{t-2})e_{t-1}\ \right]$$

where y_{t-1} is expressed as a function of y_{t-2} and e_{t-1} and e_{t-2}.

Taking the expectation of each term in the above parenthesis the equation becomes;

$$E(y_{t-1}e_{t-1}) = E(\phi_1 y_{t-2}e_{t-1}) + E(e_{t-1}e_{t-1}) + E(\alpha_1 e_{t-2}e_{t-1})$$

Note that the error term at time t "e_t "and y_{t-1} are not correlated, and e_t and e_{t-1} are not correlated by definition. Therefore the equation simplifies to;

$$E(y_{t-1}e_{t-1}) = \sigma_e^2$$

thus the whole variance equation becomes;

$$Var(y_t) = \phi_1^2 Var(y_{t-1}) + 2\phi_1\alpha_1\sigma_e^2 + \sigma_e^2 + \alpha_1^2\sigma_e^2$$

Since the series is stationary by definition, the variances are equal across time, in other words;

$$Var(y_t) = Var(y_{t-1}) = \gamma_0$$

Thus using the above feature; the formula for the variance of the series can be written as;

$$Var(y_t) = \left(\frac{1 + \alpha_1^2 + 2\phi_1\alpha_1}{1 - \phi_1^2} \right) \sigma_e^2$$

The variance, by definition should be positive, thus the term in the denominator in the formula above, namely $(1 - \phi_1^2)$ should be greater than zero, which means that $|\phi_1|$ < 1. This is exactly the same as the stationary condition of the general AR(1) process. The one lag covariance of the series can be found by usual formula for the covariance as shown below:

$$E(y_{t-1}y_t) = E[y_{t-1}(\phi_1 y_{t-1} + e_t + \alpha_1 e_{t-1})\ ]$$

$$E(y_{t-1}y_t) = E[y_{t-1}\phi_1 y_{t-1} + e_t y_{t-1} + \alpha_1 e_{t-1} y_{t-1}\ ]$$

$$E(y_{t-1}y_t) = E(y_{t-1}^2) + E(e_t y_{t-1}) + \alpha_1 E(e_{t-1} y_{t-1})$$

Note that $E(y_{t-1}^2)$ is the variance which is denoted by γ_0, $E(e_t y_{t-1})$ is equal to zero since e_t and y_{t-1} are uncorrelated. The final term $\alpha_1 E(e_{t-1} y_{t-1})$ is equal to $\alpha_1 \sigma_e^2$, thus the formula becomes;

$$E(y_{t-1} y_t) = \gamma_1 = \phi_1 \gamma_0 + \alpha_1 \sigma_e^2$$

The above formula is valid for 1 lag covariance namely $Cov(y_t y_{t-1})$.To find a general formula for further lags we try lag 2 covariance. Let us now find $Cov(y_t y_{t-2})$

$$E(y_{t-2} y_t) = E\left[y_{t-2}(\phi_1 y_{t-1} + e_t + \alpha_1 e_{t-1})\ \right]$$

$$E(y_{t-2} y_t) = E\left[y_{t-2}\phi_1 y_{t-1} + e_t y_{t-2} + \alpha_1 e_{t-1} y_{t-2}\ \right]$$

$$E(y_{t-2} y_t) = \phi_1 E(y_{t-2} y_{t-1}) + E(e_t y_{t-2}) + \alpha_1 E(e_{t-1} y_{t-2})$$

On the right hand side of the equation above the first term is the 1 lag covariance, the second term is zero since the e_t and y_{t-2} are not correlated, the third term is also zero since e_{t-1} and y_{t-1} are uncorrelated. Thus;

$$E(y_{t-2} y_t) = \gamma_2 = \phi_1 \gamma_1$$

In general it can be shown that;

$$E(y_{t-k}y_t) = \gamma_k = \phi_1\gamma_{k-1}$$

Using the simple formula for autocorrelation, Autocorrelation function for lag 1 can be found as

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \phi_1 + \frac{\alpha_1\sigma_e^2}{\gamma_0}$$

Autocorrelation function for lag 2 and further lags can be found as

$$\rho_k = \phi_1\rho_{k-1}$$

Therefore the ACF of ARMA(1,1) process is exactly the same as that of the AR(1) process after the lag 2, but it behaves different for lag 1. In other words exponential decay starts at lag 2. On the other hand, Partial Autocorrelation Function (PACF) of the ARMA(1,1) process is very similar to that of the MA(1) model.In general, for an ARMA(p,q) process, the autocorrelation function can be written as

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + ... + \phi_p\rho_{k-p} \quad \text{for k>p}$$

This means that for lags k greater than q, the autocorrelation behave like an autoregressive process. In other words, moving average component contributes nothing to the autocorrelation function after q lags meaning that MA component of the ARMA process has a memory of only q periods. The presence of first order autocorrelation can be tested by Durbin Watson test while higher order autocorrelation can be tested by Breusch-Godfrey serial correlation Lagrange Multiplier test. This test is generally suggested to use in large samples and the following null and alternative hypotheses are constructed.

$H_0 : \alpha_1 = \ldots\ldots = \alpha_p = 0$

H_1: At least one of α_I is non zero

Where the coefficients α_i 's cab assumed to be the coefficients of an AR model of order p. This test uses an auxiliary regression for the residuals. In other words residuals are regressed on the original regressors and lagged residuals of up to order p. The test statistic which is asymptotically distributed as χ^2 (p) is nR^2 where n is the number of observations.

The Implications Of Autocorrelation And Partial Autocorrelation Functions

The detection of autocorrelation in a series can also be done by looking at the plots of autocorrelations and partial autocorrelations. In order to see whether there is any autocorrelation or partial autocorrelation of return series correlogram of each series are plotted. Correlogram is a very useful diagram, because by visual inspection, the order of AR and the order of MA components can more easily be understood. More specifically, partial autocorrelations chart is often used to determine the order p of the AR process. To explain in more detail, let's look at the following AR representation:

$$Y_t = \mu + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-1} + ... + \alpha_p Y_{t-p} + e_t$$

Since the p^{th} partial autocorrelation coefficients measure the correlation between Y_t and Y_{t-p} after the effects of Y_{t-1} $Y_{t-2,....}$ Y_{t-p+1} are taken into account, , the parameters α_1 , α_2 , α_3 … α_p are actually the partial autocorrelations in the AR equation above.

On the other hand the autocorrelation function of the MA(1) process becomes zero after lag 1 while the partial autocorrelations declines geometrically. Thus MA(1) process has a memory of one period. This property is very useful when evaluating the correlogram. In other words, the order of MA process can be determined by looking at the lags where the autocorrelations taper off to zero. In summary, in an AR(p) model, Acf declines geometrically, Pacf cuts off abruptly after lag p, while in a MA(q) model Acf cuts off abruptly after lag q, Pacf declines geometrically.

Long Term Dependence (Arfima Models)

If a series exhibits long memory, this means that there is persistent temporal dependence between observations widely separated in time. The autocorrelation of series with long memory decay hyperbolically, meaning a relatively sooth decay compared to quickly declining autocorrelations of short memory series. The An ARFIMA model has three parameters, namely, the order of autoregressive part generally shown by p, the order of moving average component generally called q, and the fractional integration parameter d. An ARFIMA (p,d,q) can be written in functional form as follows:

$$\phi(L)(1\text{-}L)^d\, y_t = \theta(L)\, e_t$$

L is the lag operator,

d is the fractional integration parameter

e_t is the error term being iid with mean zero and a constant variance.

$\phi(L)$ and θ (L) are polynomial in L, up to order p and q respectively. Both of these polynomials should be outside the unit circle to guarantee stationarity and irreversibility.

Granger and Joyeux (1980) and Hosking (1981) show that when the term $(1\text{-}L)^d$ is allowed to assume non-integer values of the variable d, the result is a fractionally differenced time series. The variable "d" stands for the magnitude of fractional differencing. Granger and Joyeux (1980) and Hosking (1981) also show that the series is stationary when d is less than one-half, and invertible when d is greater than minus one-half.

To find the value of the fractional integration parameter different methods are suggested by researchers. In this thesis the Geweke and Porter-Hudak (GPH) algorithm, which is based on frequency domain regression technique, is applied to find "d". To explain the method let $I(\xi)$ be the periodogram (spectrum, the spectral density) of the demeaned series $y_t - \bar{y}$ at frequency ξ, that is ,

$$I(\xi) = \frac{1}{2\pi T}(\sum_{t=1}^{T}(y_t - \bar{y})e^{-it\xi})^2$$

The spectrum, or spectral density of a time series is the Fourier transform of the autocovariance function of a stationary process. The basic idea behind fourier transform is that the data-generating process can be approximated by the sum of stochastic sine waves of variable frequency. The spectrum or the periodogram is the plot of spectral density function against frequency (angular frequency in the range [0 ,π]). In other words, it specifies the contribution each frequency makes to the total variance. The term white noise, actually takes the name from the shape of the its spectral density of its autocovariance function, namely, it has flat spectrum with all frequencies being of the same importance, which is the electromagnetic spectrum of white light.

After defining the periodogram, the spectral regression of the GPH estimator is then computed by regressing logarithmic periodograms on a constant and a nonlinear function of the frequencies as follows.

$$\ln\{I(\xi_i)\} = \beta_0 + \beta_1 \ln\left[\sin^2\left(\frac{\xi_i}{2}\right)\right] + \eta_j$$

where j =1…..v , and (v<<T) is the number of periodogram ordinates used in the regression, $\xi_i = (2\pi j)/T$ and η_j is the error term, and T is the number of observations.

The GPH estimate of d is the negative of the OLS estimate of β_1 in this regression. GPH method has some deficiencies however. One is the fact that a choice the parameter v must be made. The parameter v can be defined as a function of the sample size, the most common choice being T0.5. This choice may lead to biased results, as shown by Tolvi (2003), he says that values around T0.5 lead to very random results. He suggests T0.8 as a better choice. In spite of this deficiency the GPH estimator is found to be robust in case of minor deviations from normality and the existence of ARCH effects which is the case in our logarithmic session to session series. Using the binomial theorem for non-integer powers, the term $(1\text{-L})^d$ can be written by the following polynomial expansion

$$(1-L)^d = \sum_{k=0}^{\infty} (-1)^k \binom{d}{k} L^k$$

$$\binom{d}{k} \equiv \frac{d(d-1)(d-2).....(d-k+1)}{k!}$$

Using the above formulas, $(1\text{-L})^d$ can be written in terms of the gamma function which is denoted by the symbol Γ as follows:

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)} L^k$$

Basic properties of the gamma function is available in Appendix B. When the above expansion is applied to the variable y_t , the following formula can be obtained

$$(1-L)^d y_t = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)} y_{t-k} = \sum \phi_k y_{t-k} = \varepsilon_t$$

The above formulation is actually an infinite order AR process. The autoregressive coefficients in terms of the gamma function are as follows:

$$\phi_k = \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}$$

The coefficients can also be written as an infinite Moving Average Process. To do that, consider the fact that

$$y_t = (1-L)^{-d} \varepsilon_t$$

Assume that $\theta(L) = (1-L)^{-d}$. The above term can be expanded as;

$$(1-L)^{-d} = \sum_{k=0}^{\infty} (-1)^k \binom{-d}{k} L^k$$

$$\binom{-d}{k} \equiv \frac{-d(-d-1)(-d-2).....(-d-k+1)}{k!}$$

or equivalently;

$$\binom{-d}{k} \equiv \frac{d(d+1)(d+2).....(d+k-1)}{k!}$$

Therefore, the MA coefficients denoted by θ_k can be written as;

$$\theta_k = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)}$$

Granger and Joyeux (1980) and Hosking (1981) show that d should be strictly less than 0.5 for stationarity of the series and d should be strictly greater than –0.5 for invertibility. They also show that the autocorrelation coefficients of any fractionally differenced series are of the same sign as d. Campbell, Lo MacKinlay (1997). It has been shown that when d is positive the sum autocorrelations goes to infinity, when d is negative the sum collapses to zero.

For ARFIMA processes where the fractional integration parameter is within the interval (-0.5,0.5), the asymptotic correlation function is approximated by the following formula.

$\gamma(h) \sim Ch^{2d-1}$, as $h \to \propto$

where C is the first order autocorrelation and h denotes the displacement in time. It can easily be inferred from the formula that the correlation function decays at a polynomial rate. If it was decreasing at an exponential rate, the rate of decrease would be higher. Since it decays at polynomial rate, the series is said to have long memory.

In order to see the autocorrelation structure of a series with long term memory a very good example is provided in Campbell et al. (1997). As shown in the table below, the autocorrelation of a fractionally differenced series decays at a much slower rate compared to an AR(1) series. Additionally the rate of decrease in autocorrelations of a series with d>0 is faster than that of a series with d<0. This can also easily be inferred from the formula that approximates the autocorrelation function. Note also the fact the autocorrelations are of the same sign as the fractional integration parameter d.

Table 45 - Behaviour of Autocorrelation Function with different fractional integration values

Lag	d = 1/3 Autocorrelation $\gamma(h)$	d = -1/3 Autocorrelation $\gamma(h)$	AR(1) Autocorrelation $\gamma(h)$
1	0.500	-0.250	0.500
2	0.400	-0.071	0.250
3	0.350	-0.036	0.125
4	0.318	-0.022	0.063
5	0.295	-0.015	0.031
10	0.235	-0.005	0.001
25	0.173	-0.001	2.98×10^{-8}
50	0,137	-3.24×10^{-4}	2.98×10^{-16}
100	0,109	-1.02×10^{-4}	7.89×10^{-31}

Notice that at lag 25 the autocorrelation of AR(1) series drops almost to zero while the autocorrelation of the series with fractional differencing parameter d=1/3, is

0.173, quite high indicating long memory. The series with a minus value of fractional differencing parameter (d= -1/3) decays at a faster rate compared to the one with d=1/3, but decay rate is still quite low compared to the AR(1) process. The autocorrelation of the series with d = -1/3, is different from zero (-0.001) at lag 25, while the autocorrelation of the series AR(1) is virtually zero at the same lag value. Note also the fact that the first order autocorrelation of the series with minus differencing parameter is equal to only the half of the AR(1) process in magnitude. As seen the rate of decay correlations is quite low. The correlation between t and t-1 is 0.80 while it is 0.15 for t and t-20.

The determination process of the fractional parameter d is first done by Geweke and Porter-Hudak (1993) method which is a semiparametric procedure to obtain the least squares estimate of the parameter "d" in a frequency domain regression. The frequency domain regression is generally used to examine the contribution of different frequencies in explaining the variance of a series. The relevant algorithm is downloaded from the RATs internet site and calculation is done using the RATS package.

Findings

Stationarity Tests

Since the return series and other variables are studied by using time series analysis, stationarity tests are conducted for all of them. The results of the tests are provided in Appendix C. In addition to these formal statistical outputs, all the series are also visually inspected to assure the stationarity condition.

As seen the ADF test statistics is more negative than the critical value of even 1 % level. Therefore the null hypothesis of unit root is rejected for the session to session returns calculated from the closing values and average values.

Like the results found for the return series, all the other series such as return dispersion, volume dispersion and returnvolume, ret30vol, volchg, range etc. used in this study were also found to be stationary.

The stationarity is re-checked by PP test and similar results are obtained. Two sample outputs for this test is alsoprovided in Appendix C.

In addition to the stationarity test done for the session to session returns and other variables above, the stationarity of daily returns, daily volume change, daily return dispersion etc are all tested and they are all found to be stationary.

Autocorrelation Tests

The correlogram of close to close session returns, average session returns, daily close to close returns and daily average returns are provided in Appendix D. The Q-statistics is actually known as Box-Pierce Q statistics and larger values of this statistics with very low probability figures at the rightmost column shows the existence of autocorrelation.

For close to close return series, the Q statistics are all large and significant up to 36 lags. However it should be noted that the magnitude of both the Autocorrelation and partial autocorrelation are very close to zero. The largest autocorrelation and partial autocorrelation is observed for the first lag, but the magnitude of Q-statistics imply that the series has the long memory property.

On the other hand the correlogram of average session to session returns reveals the fact that the autocorrelation of average returns is larger and more visible. Especially the first lag autocorrelation is quite larger than the first order autocorrelation of close to close return series. Although the Q statistics are also larger for all the lags up lag 36, the magnitude of autocorrelations and partial autocorrelations are very close to zero.

The correlogram of the daily close to close return series exhibit an interesting property. It can be seen that the autocorrelation and partial autocorrelations are virtually zero for the first four lags and the Q-statistics are very low for these four lags implying the non-existence of serial correlation. Autocorrelations for the lags further past are also very close to zero, but their Q values are not low enough to assure the non-existence of serial correlation. Additionally, the correlogram of daily close to close return series for the period between 2002-2005 exhibit stronger sign of white noise property. The correlogram of average daily returns exhibits similar pattern to that of the session to session average returns. The first two lags have relatively large and significant correlations. Lag three and further lags have very small but still statistically significant correlations.

Comparison of daily returns and session to session returns shows that session to session returns have higher autocorrelations, autocorrelations are quite larger when the returns are calculated from the averages. This property was found to hold even for the average returns belonging to the period 2002-2005.

Autoregressive Model Of Session to Session Return Series

The analysis is first done by using the session close to session close returns denoted by the variable ret30seans. Although the magnitude of autocorrelations and partial autocorrelations are close to zero, the values of Q statistics are large and for this reason the existence of any significant AR and/or MA terms for session to session return series was analysed. As a first step the significant AR term(s) were looked for. The following output shows that the series have a statistically significant AR(1) term. Note that the stationarity condition for general AR(p) processes is that the inverted roots of the lag polynomial lie inside the unit circle. EViews reports these roots as Inverted AR Roots at the bottom of the regression output. There is no particular problem if the roots are imaginary, but a stationary AR model should have all roots with modulus (absolute values) less than one.

Table 46 - AR(1) model of session to session returns

Dependent Variable: RET30SEANS
Method: Least Squares
Date: 10/03/05 Time: 16:50
Sample(adjusted): 3 4014
Included observations: 4012 after adjusting endpoints
Convergence achieved after 2 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.090049	0.015727	5.725651	0.0000

R-squared	0.007021	Mean dependent var	0.000741
Adjusted R-squared	0.007021	S.D. dependent var	0.022379
S.E. of regression	0.022300	Akaike info criterion	-4.768169
Sum squared resid	1.994713	Schwarz criterion	-4.766600
Log likelihood	9565.948	Durbin-Watson stat	1.999858
Inverted AR Roots	.09		

As seen from the above output, the AR(1) coefficient is significant; however, the R-squared statistics is quite low. Additional AR terms were put into the equation to see whether there are any other significant terms. While adding the previous lags the value of adjusted R-squared and the value of Akaike information criterion was checked. The R-squared should get larger and the value of Akaike information criterion should get lower. The point where the Akaike information criterion reaches the minimum value is a good candidate to stop. For example if ar(2) is added, the coefficient of the ar(2) term is found to be statistically insignificant, and the overall fit of the equation as measured by the adjusted R-squared gets worse and the Akaike information criterion gets a higher value meaning that ar(2) term should not be included in the equation. The researcher should also carefully follow the values of the AR roots, since they should all be strictly less than one.The output from the Eviews statistical software package is provided below for ar(1) and ar(2) together.

Table 47 - AR(2) model of Session to session returns

Dependent Variable: RET30SEANS
Method: Least Squares
Date: 10/03/05 Time: 16:53
Sample(adjusted): 4 4014
Included observations: 4011 after adjusting endpoints
Convergence achieved after 2 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.089810	0.015793	5.686825	0.0000
AR(2)	0.001557	0.015792	0.098622	0.9214
R-squared	0.007028	Mean dependent var		0.000734
Adjusted R-squared	0.006780	S.D. dependent var		0.022378
S.E. of regression	0.022302	Akaike info criterion		-4.767800
Sum squared resid	1.993959	Schwarz criterion		-4.764660
Log likelihood	9563.822	Durbin-Watson stat		1.999118
Inverted AR Roots	.10	-.01		

However if the ar(3) term is added the ar(3) term is found to have a statistically significant coefficient. The adjusted R-squared of the equation increases and the Akaike information criterion gets lower compared to that of the equation including only ar(1) and ar(2). The output of the equation including the ar(1) ar(2) and (3) terms is provided below.

Table 48 - AR(3) Model for session to session returns

Dependent Variable: RET30SEANS
Method: Least Squares
Date: 10/04/05 Time: 14:44
Sample(adjusted): 5 4014
Included observations: 4010 after adjusting endpoints
Convergence achieved after 2 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.090066	0.015792	5.703260	0.0000
AR(2)	0.004188	0.015853	0.264165	0.7917
AR(3)	-0.028842	0.015789	-1.826739	0.0678
R-squared	0.007881	Mean dependent var		0.000737
Adjusted R-squared	0.007386	S.D. dependent var		0.022380
S.E. of regression	0.022297	Akaike info criterion		-4.768005
Sum squared resid	1.992058	Schwarz criterion		-4.763294
Log likelihood	9562.849	Durbin-Watson stat		1.997117
Inverted AR Roots	.19+.26i	.19 -.26i	-.28	

By adding and/or deleting the further past lags of the return series and the following model which has the minimum value for Akaike criterion was obtained. The adjusted R-squared value was found as 0.021488. As seen from the output, all the coefficients are significant at 5 % level or more. It is quite interesting to have lag number 32 as a significant parameter in our model, this is regarded as an indication of long memory in return series.

Table 49 - Final AR Model for session to session returns

Dependent Variable: RET30SEANS
Method: Least Squares
Date: 03/14/06 Time: 11:27
Sample(adjusted): 34 4014
Included observations: 3981 after adjusting endpoints
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.091914	0.015715	5.848772	0.0000
AR(4)	0.043258	0.015761	2.744670	0.0061
AR(5)	-0.040281	0.015758	-2.556238	0.0106
AR(8)	0.034088	0.015699	2.171387	0.0300
AR(11)	-0.048708	0.015668	-3.108744	0.0019
AR(17)	0.033644	0.015681	2.145507	0.0320
AR(24)	0.034173	0.015688	2.178305	0.0294
AR(30)	0.075790	0.015659	4.840089	0.0000
AR(32)	0.033817	0.015666	2.158679	0.0309
R-squared	0.023455	Mean dependent var		0.000713
Adjusted R-squared	0.021488	S.D. dependent var		0.022387
S.E. of regression	0.022145	Akaike info criterion		-4.780124
Sum squared resid	1.947927	Schwarz criterion		-4.765906
Log likelihood	9523.837	Durbin-Watson stat		1.996742
Inverted AR Roots	.94	.91+.20i	.91 -.20i	.85 -.37i
	.85+.37i	.75+.53i	.75 -.53i	.64 -.67i
	.64+.67i	.48+.79i	.48 -.79i	.28+.85i
	.28 -.85i	.09+.89i	.09 -.89i	.00+.68i
	.00 -.68i	-.09 -.86i	-.09+.86i	-.28 -.87i
	-.28+.87i	-.47+.79i	-.47 -.79i	-.63 -.67i
	-.63+.67i	-.76+.53i	-.76 -.53i	-.84 -.37i
	-.84+.37i	-.91+.19i	-.91 -.19i	-.94

Adding a constant to the equation does not improve the overall fit and besides the constant term is found to have low significance. Therefore the following equation is obtained.

$$R_t = 0{,}091914 R_{t-1} + 0{,}043258 R_{t-4} - 0{,}040281 R_{t-5} + 0{,}034088 R_{t-8} - 0{,}048708 R_{t-11} + 0{,}033644 R_{t-17} + 0{,}034173 R_{t-24} + 0{,}075790 R_{t-30} + 0{,}033817 R_{t-32} + U_t$$

Although the coefficients in the above equation are statistically significant the adjusted R^2 value (0.0235). Since R^2 is defined as:

$$R^2 = 1 - \frac{Var(e_t)}{Var(y_t)}$$

where Var(y_t) is the variance of original series and the Var(e_t) is the variance of the residuals of autoregressive and/or moving average model. The value adjusted R^2 being equal to only 0.023 means that 2,3 % of the total variance can be explained by the model which is actually very low.

As a final step in our analysis, the residuals of the above model were plotted to see the effectiveness of the overall fit of the model. The correlogram of residuals shown in Appendix D is a very useful tool to assess the effectiveness of the final AR model. As seen, the residuals are uncorrelated up to lag 36 meaning that the error terms of the AR model specified above can be defined as white noise. Another test which is called serial correlation LM test is also done for lags even greater than 36. As the following Eviews output shows, there is no significant autocorrelation in the series. The statistic labeled " Obs*R-squared" is the LM test statistic for the null hypothesis of no serial correlation.

Table 50 - Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.962018	Probability	0.559397
Obs*R-squared	55.70217	Probability	0.633388

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Presample missing value lagged residuals set to zero.

On the other hand the correlogram of squared residuals provided in Appendix D shows that the squared residuals are correlated. This means that there is ARCH and/or GARCH effects in the data .This fact can also be seen by doing a simple ARCH-LM test which is readily available in Eviews. As the following output shows, even at lag 1 there is a significant ARCH effect, in other words the ARCH test results strongly suggest the presence of heteroskedasticity and nonnormality in the residuals.

Table 51 - ARCH LM test for the residuals

ARCH Test:

F-statistic	114.4604	Probability	0.000000
Obs*R-squared	111.3151	Probability	0.000000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Sample(adjusted): 35 4014

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000408	1.98E-05	20.60768	0.0000
RESID^2(-1)	0.167236	0.015632	10.69862	0.0000

R-squared	0.027969	Mean dependent var	0.000489
Adjusted R-squared	0.027724	S.D. dependent var	0.001167
S.E. of regression	0.001151	Akaike info criterion	-10.69657
Sum squared resid	0.005266	Schwarz criterion	-10.69341
Log likelihood	21288.18	F-statistic	114.4604
Durbin-Watson stat	2.072309	Prob(F-statistic)	0.000000

On the other hand, for the average session to session returns the following autoregressive model is obtained for the average return series.

Table 52 - Final AR Model for the average returns

Dependent Variable: RET30AVGSEANS
Method: Least Squares
Sample(adjusted): 32 4014
Included observations: 3983 after adjusting endpoints
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.438184	0.015730	27.85744	0.0000
AR(2)	-0.170307	0.016991	-10.02333	0.0000
AR(3)	0.054283	0.015731	3.450609	0.0006
AR(5)	-0.042461	0.014421	-2.944358	0.0033
AR(8)	0.049263	0.015558	3.166414	0.0016
AR(9)	-0.044130	0.015543	-2.839242	0.0045
AR(11)	-0.061672	0.014416	-4.277970	0.0000
AR(13)	0.035599	0.014486	2.457392	0.0140
AR(18)	0.032518	0.014430	2.253501	0.0243
AR(24)	0.040414	0.014437	2.799290	0.0051
AR(30)	0.058408	0.014404	4.054903	0.0001
R-squared	0.175664	Mean dependent var		0.000716
Adjusted R-squared	0.173588	S.D. dependent var		0.018325
S.E. of regression	0.016659	Akaike info criterion		-5.348987
Sum squared resid	1.102305	Schwarz criterion		-5.331616
Log likelihood	10663.51	Durbin-Watson stat		1.994531
Inverted AR Roots	.93	.90 -.21i	.90+.21i	.84+.37i
	.84 -.37i	.74+.53i	.74 -.53i	.64+.68i
	.64 -.68i	.49 -.79i	.49+.79i	.27+.87i
	.27 -.87i	.10+.91i	.10 -.91i	-.08 -.90i
	-.08+.90i	-.25 -.88i	-.25+.88i	-.44 -.80i
	-.44+.80i	-.61 -.67i	-.61+.67i	-.71+.54i
	-.71 -.54i	-.80 -.37i	-.80+.37i	-.87 -.20i
	-.87+.20i	-.91		

The above output says that the average returns does posses the long memory property that is encountered in the close to close return autoregressive model. One important difference is that the adjusted r-squared is quite better than that of the close to close returns. The correlogram of the autoregressive model including lags 1, 2, 3,5,8,9,11,13,18,24,30 reveals that the residuals are uncorrelated up lag 36, meaning that the model is sufficient in explaining the series. The error terms of this model can be termed as white noise. The residuals are found to be non-normal though. The squared residuals of the autoregressive model found for the average returns are found to be correlated however, implying the existence of an ARCH and/or GARCH effect.

Especially the first two lagged terms of the average returns do have quite large explanatory power compared to the other lag terms, the coefficients of the AR terms other than lag 1 and lag 2 are quite close to zero although their significance level is high. So the model is also formed with only lag 1 and lag2 and the relevant statistics are shown below.

Table 53 - AR modeling of average return series with only two lags

Dependent Variable: RET30AVGSEANS
Method: Least Squares
Date: 04/25/06 Time: 13:35
Sample(adjusted): 4 4014
Included observations: 4011 after adjusting endpoints
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.432021	0.015619	27.66073	0.0000
AR(2)	-0.148546	0.015618	-9.510959	0.0000

R-squared	0.159067	Mean dependent var	0.000737
Adjusted R-squared	0.158858	S.D. dependent var	0.018317
S.E. of regression	0.016799	Akaike info criterion	-5.334507
Sum squared resid	1.131354	Schwarz criterion	-5.331367
Log likelihood	10700.35	Durbin-Watson stat	1.985187
Inverted AR Roots	.22+.32i	.22 -.32i	

The above output leads to us to the conclusion that if the average return of a session is positive, the average return of the next session will also be positive. This is derived due to the positive and significant coefficient of the AR(1) term. On the other hand, one can also deduce that the average return of a session is negatively correlated to the return of the session preceding the previous session.

Autoregressive Model of Daily Return Series

As it has been explained in previous sections, the correlogram of the close to close daily return series imply that the series have almost no autocorrelations. However an autoregressive model to the daily return series calculated from the closing values is still tried The first four lagged terms are found have no significance a result that is also implied by the correlogram. By trial and error as it has been done for the session close to session close series the following AR model has been reached.

Table 54 - AR Model for daily returns

Method: Least Squares
Sample(adjusted): 54 2014
Included observations: 1961 after adjusting endpoints
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(5)	-0.053292	0.022450	-2.373824	0.0177
AR(9)	0.063331	0.022402	2.827029	0.0047
AR(11)	-0.045326	0.022415	-2.022078	0.0433
AR(15)	0.051712	0.022405	2.308023	0.0211
AR(32)	-0.063890	0.022406	-2.851458	0.0044
AR(44)	0.046178	0.022376	2.063747	0.0392
AR(52)	0.052246	0.022362	2.336391	0.0196
R-squared	0.018708	Mean dependent var		0.001460
Adjusted R-squared	0.015695	S.D. dependent var		0.033296
S.E. of regression	0.033034	Akaike info criterion		-3.979020
Sum squared resid	2.132241	Schwarz criterion		-3.959097
Log likelihood	3908.429	Durbin-Watson stat		1.985849
Inverted AR Roots	.95	.95+.12i	.95 -.12i	.91+.24i
	.91 -.24i	.87 -.33i	.87+.33i	.84+.43i
	.84 -.43i	.78+.53i	.78 -.53i	.72+.63i
	.72 -.63i	.63 -.72i	.63+.72i	.53+.77i
	.53 -.77i	.43 -.83i	.43+.83i	.32+.87i
	.32 -.87i	.24+.92i	.24 -.92i	.12 -.95i
	.12+.95i	-.00 -.95i	-.00+.95i	-.12 -.95i
	-.12+.95i	-.24+.91i	-.24 -.91i	-.33 -.87i
	-.33+.87i	-.44 -.85i	-.44+.85i	-.52+.78i
	-.52 -.78i	-.62+.72i	-.62 -.72i	-.72 -.62i
	-.72+.62i	-.78+.53i	-.78 -.53i	-.84+.43i
	-.84 -.43i	-.87+.32i	-.87 -.32i	-.92+.24i
	-.92 -.24i	-.95	-.95 -.12i	-.95+.12i

The final autoregressive model with 7 terms has an R-squared value of 0.015695. As seen from the output, all the coefficients are significant at 5 % level or more. But the adjusted r-squared value is very low, a fact that leads us to say that there seems to be almost no significant linear relationship among the returns through time. On the other hand, the 52^{nd} AR term was found to have a significant coefficient which is regarded as an indication of long memory in close to close return series. The residuals of the above model are found to be similar to white noise, but the squared residuals are found to strongly correlated over time, this finding is in accordance with the conclusion which has been reached for the session to session return series. Therefore the residuals are not linearly dependent but the squared residuals show dependence over time for session to session and daily returns.

For the average daily returns, the results of the autoregressive analysis show similar features to that of the session to session average returns as shown below. Average returns in general have higher adjusted r-squared values. The average returns series is found to have the long memory property which is also the case for close to close returns. From the above output it can be seen that the 52^{nd} AR term have statistically significant coefficient, meaning that there is strong persistence in the series. Note also the fact that the adjusted r-squared value of daily average return series is lower than that of the session to session average return series. This is an expected result since the longer the time interval between the observation the lower the significance of autoregressive effects.

Table 55 - AR Model for the average daily returns

Dependent Variable: RET30AVGD
Method: Least Squares
Date: 05/05/06 Time: 14:38
Sample(adjusted): 54 2014
Included observations: 1961 after adjusting endpoints
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.319494	0.022340	14.30153	0.0000
AR(2)	-0.104786	0.022350	-4.688398	0.0000
AR(5)	-0.071588	0.021339	-3.354797	0.0008
AR(8)	0.053497	0.021410	2.498748	0.0125
AR(33)	-0.047408	0.021365	-2.219024	0.0266
AR(43)	0.045032	0.021348	2.109389	0.0350
AR(52)	0.048017	0.021354	2.248652	0.0246
R-squared	0.107430	Mean dependent var		0.001477
Adjusted R-squared	0.104689	S.D. dependent var		0.027658
S.E. of regression	0.026170	Akaike info criterion		-4.444839
Sum squared resid	1.338239	Schwarz criterion		-4.424916
Log likelihood	4365.164	Durbin-Watson stat		1.997572
Inverted AR Roots	.95	.95+.12i	.95 -.12i	.91+.24i
	.91 -.24i	.87+.32i	.87 -.32i	.85 -.43i
	.85+.43i	.79+.54i	.79 -.54i	.72 -.63i
	.72+.63i	.63+.72i	.63 -.72i	.52 -.77i
	.52+.77i	.45 -.82i	.45+.82i	.34+.88i
	.34 -.88i	.24+.92i	.24 -.92i	.12 -.95i
	.12+.95i	-.01 -.95i	-.01+.95i	-.11 -.92i
	-.11+.92i	-.21 -.92i	-.21+.92i	-.32 -.88i
	-.32+.88i	-.43+.85i	-.43 -.85i	-.54+.78i
	-.54 -.78i	-.63+.69i	-.63 -.69i	-.69 -.62i
	-.69+.62i	-.76+.54i	-.76 -.54i	-.82+.45i
	-.82 -.45i	-.88 -.34i	-.88+.34i	-.92+.21i
	-.92 -.21i	-.92+.11i	-.92 -.11i	-.93

The residual analysis of the above model shows that the model is not fully successful to eliminate the serial correlation of residuals. The squared residuals show strong sign of dependence implying the ARCH/GARCH effects. Moreover, similar results found for the session to session average returns the largest explanation of the model stems from the first two ar terms and although not shown here, it can be said that if the average return of day is positive, the next day’s return will also be expected to be positive.

Moving Average Representation Of Session To Session Return Series

The session close to session close returns are also analysed by using the Moving Average method. As it has been done for the Autoregressive modeling of the return series, by adding and/or deleting the MA terms at different the final model has been reached as shown below.

Table 56 - MA model session to session returns

Dependent Variable: RET30SEANS
Method: Least Squares
Date: 03/14/06 Time: 12:20
Sample(adjusted): 2 4014
Included observations: 4013 after adjusting endpoints
Convergence achieved after 6 iterations
Backcast: -30 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	0.093258	0.015628	5.967446	0.0000
MA(3)	-0.032146	0.015693	-2.048373	0.0406
MA(4)	0.048602	0.015720	3.091736	0.0020
MA(8)	0.039006	0.015646	2.493109	0.0127
MA(11)	-0.044715	0.015628	-2.861156	0.0042
MA(17)	0.039243	0.015648	2.507827	0.0122
MA(30)	0.076970	0.015639	4.921570	0.0000
MA(32)	0.036467	0.015641	2.331567	0.0198

R-squared	0.022736	Mean dependent var	0.000742
Adjusted R-squared	0.021028	S.D. dependent var	0.022377
S.E. of regression	0.022140	Akaike info criterion	-4.780862
Sum squared resid	1.963185	Schwarz criterion	-4.768308
Log likelihood	9600.800	Durbin-Watson stat	1.999321

Inverted MA Roots	.92+.10i	.92 -.10i	.88 -.28i	.88+.28i
	.81 -.46i	.81+.46i	.69 -.60i	.69+.60i
	.55 -.74i	.55+.74i	.38 -.82i	.38+.82i
	.21+.87i	.21 -.87i	.00+.69i	.00 -.69i
	-.02 -.88i	-.02+.88i	-.20+.88i	-.20 -.88i
	-.40 -.83i	-.40+.83i	-.56+.73i	-.56 -.73i
	-.70+.62i	-.70 -.62i	-.81+.45i	-.81 -.45i
	-.89 -.29i	-.89+.29i	-.92+.10i	-.92 -.10i

As it can be seen from the output, quite similar to persistency problem encountered in the AR analysis we have the same issue arising in MA analysis. More specifically, in our final model the MA terms belonging to lag 30 and 32 respectively are statistically significant. The overall fit of this model is slightly poorer than that of the AR model. A constant is also added to the MA equation but it is found to be insignificant at 5 % level. The functional form for the moving average representation of the session close to session close return series are provided below.

$$R_t = 0{,}093258U_{t-1} - 0{,}032146U_{t-3} + .0{,}048602U_{t-4} - 0{,}039006U_{t-8} - 0{,}044715U_{t-11} - 0{,}039243U_{t-17} + 0{,}076970U_{t-30} + 0{,}036467U_{t-32} + \varepsilon_t$$

Although the coefficients are statistically significant, their magnitudes are quite low. The adjusted R2 is very similar to the AR model previously analysed. The explained variance as measured by the R-square statistic is again quite low, the model can explain approximately 2,3 % of total variance.

When the overall fit or the explanatory power of the AR and MA representations is compared the adjusted R-squared values of final AR specification and MA specification are found be very close to each other.

Similar results to those that are found for the AR models are obtained for the correlation of residuals and squared residuals and the distribution of the residuals of MA model . In other words, the residuals are found to be uncorrelated up to lag 30, squared residuals are found to be correlated and the distribution of residuals are non-normal. Thus, our MA model specified above does have error terms which can be called “white noise”.

The same process is repeated for the average session to session return series and the following model is obtained.

Table 57 - MA model for the average session to session returns

Dependent Variable: RET30AVGSEANS
Method: Least Squares
Date: 04/25/06 Time: 11:46
Sample(adjusted): 2 4014
Included observations: 4013 after adjusting endpoints
Convergence achieved after 7 iterations
Backcast: -45 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	0.434650	0.014210	30.58851	0.0000
MA(4)	0.050060	0.014215	3.521561	0.0004
MA(8)	0.050443	0.014157	3.563194	0.0004
MA(12)	-0.033491	0.014206	-2.357579	0.0184
MA(20)	0.040838	0.014198	2.876350	0.0040
MA(30)	0.058482	0.014109	4.145080	0.0000
MA(47)	-0.031351	0.014140	-2.217234	0.0267

R-squared	0.172915	Mean dependent var	0.000744
Adjusted R-squared	0.171676	S.D. dependent var	0.018315
S.E. of regression	0.016669	Akaike info criterion	-5.348779
Sum squared resid	1.113101	Schwarz criterion	-5.337794
Log likelihood	10739.32	Durbin-Watson stat	1.976803

Inverted MA Roots	.92+.12i	.92 -.12i	.91	.89 -.25i
	.89+.25i	.84+.36i	.84 -.36i	.80 -.47i
	.80+.47i	.71 -.58i	.71+.58i	.64+.66i
	.64 -.66i	.54 -.75i	.54+.75i	.43+.81i
	.43 -.81i	.34 -.86i	.34+.86i	.20+.91i
	.20 -.91i	.09+.91i	.09 -.91i	-.03 -.93i
	-.03+.93i	-.17+.92i	-.17 -.92i	-.27 -.87i
	-.27+.87i	-.41+.85i	-.41 -.85i	-.51+.78i
	-.51 -.78i	-.60+.71i	-.60 -.71i	-.71+.63i
	-.71 -.63i	-.77 -.51i	-.77+.51i	-.85 -.43i
	-.85+.43i	-.90 -.31i	-.90+.31i	-.91+.18i
	-.91 -.18i	-.94+.08i	-.94 -.08i	

As seen from the output the moving average model for the average return series also exhibits the long memory property. Similar to the conclusion drawn fro the session to session series, the adjusted r-squared value of moving average model is quite close to that of the AR model. The adjusted r-squared value of the MA model for the average return series is again found to be quite higher than that of the MA model for the close to close return series. A careful look at the MA model shows that the first MA term is quite significant in explaining the average returns as shown below. The model says that if the average return is higher than expected (positive error term) then the next session will also be expected to have positive average return and vice versa.

Table 58 - MA(1) model for the average sessions

Dependent Variable: RET30AVGSEANS
Method: Least Squares
Sample(adjusted): 2 4014
Included observations: 4013 after adjusting endpoints
Convergence achieved after 7 iterations
Backcast: 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	0.417430	0.014350	29.08943	0.0000

R-squared	0.158251	Mean dependent var	0.000744
Adjusted R-squared	0.158251	S.D. dependent var	0.018315
S.E. of regression	0.016804	Akaike info criterion	-5.334195
Sum squared resid	1.132836	Schwarz criterion	-5.332626
Log likelihood	10704.06	Durbin-Watson stat	1.961755
Inverted MA Roots	-.42		

The residual analysis of the average returns show similar features, the squared residuals do have ARCH and/or GARCH property.

Moving Average Representation Of Daily Return Series

Moving average modeling of daily returns are also done separately for close to close and average return series as it has been done for session to session returns. The first result is the fact that, the series do exhibit long memory for both close to close returns and for average returns. The second result is the fact that and the adjusted r-squared value of the MA models for both close to close and average returns are lower than the adjusted r-squared value of autoregressive model of the respective return series. The residuals of the MA model exhibit similar properties to that of the autoregressive model for the same return series The output below exhibits the statistics belonging to the

MA model for daily close to close return series. It can easily be noticed that the adjusted r-squared value for the daily close to close return series is lower than the AR model explained previously. Note also the fact that the nearest MA coefficient belongs to lag number 5, in parallel to what has been found in the AR modeling. It seems that it is almost useless to use a linear autoregressive or moving average model for the daily return series calculated from the closing values of the index. This is mainly because, the both the AR and the MA models have very low adjusted r-squared values and the series also exhibit very long memory with nonlinear residuals.

Table 59 - MA model for the daily returns

Method: Least Squares
Sample(adjusted): 2 2014
Included observations: 2013 after adjusting endpoints
Convergence achieved after 6 iterations
Backcast: -50 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(5)	-0.043610	0.022150	-1.968821	0.0491
MA(9)	0.063461	0.022148	2.865365	0.0042
MA(28)	-0.048447	0.022202	-2.182115	0.0292
MA(32)	-0.063935	0.022236	-2.875276	0.0041
MA(44)	0.051120	0.022257	2.296813	0.0217
MA(52)	0.056839	0.022260	2.553417	0.0107
R-squared	0.015905	Mean dependent var		0.001476
Adjusted R-squared	0.013453	S.D. dependent var		0.033158
S.E. of regression	0.032934	Akaike info criterion		-3.985635
Sum squared resid	2.176914	Schwarz criterion		-3.968921
Log likelihood	4017.541	Durbin-Watson stat		1.987871
Inverted MA Roots	.95 -.05i	.95+.05i	.94+.19i	.94 -.19i
	.88+.29i	.88 -.29i	.86+.38i	.86 -.38i
	.81+.48i	.81 -.48i	.75+.58i	.75 -.58i
	.67 -.68i	.67+.68i	.58 -.76i	.58+.76i
	.48 -.81i	.48+.81i	.38 -.86i	.38+.86i
	.29 -.88i	.29+.88i	.19+.94i	.19 -.94i
	.05 -.95i	.05+.95i	-.05+.95i	-.05 -.95i
	-.19+.94i	-.19 -.94i	-.28 -.88i	-.28+.88i
	-.38 -.86i	-.38+.86i	-.48+.81i	-.48 -.81i
	-.58 -.75i	-.58+.75i	-.68 -.68i	-.68+.68i
	-.76 -.58i	-.76+.58i	-.81+.48i	-.81 -.48i
	-.86+.38i	-.86 -.38i	-.88+.28i	-.88 -.28i
	-.94+.18i	-.94 -.18i	-.95+.05i	-.95 -.05i

The daily returns calculated from averages do however exhibit somewhat a different picture as it was the case in AR modeling of the same return series. The Moving average model for the daily average return series with all significant MA terms included is shown below.

Table 60 - MA model for the average daily return series

Dependent Variable: RET30AVGD
Method: Least Squares
Sample(adjusted): 2 2014
Included observations: 2013 after adjusting endpoints
Convergence achieved after 8 iterations
Backcast: -51 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	0.336608	0.020820	16.16786	0.0000
MA(4)	0.059017	0.020958	2.815989	0.0049
MA(6)	-0.047669	0.020945	-2.275938	0.0230
MA(10)	0.061406	0.020895	2.938800	0.0033
MA(33)	-0.043361	0.020894	-2.075346	0.0381
MA(53)	0.056957	0.020946	2.719205	0.0066
R-squared	0.106828	Mean dependent var		0.001486
Adjusted R-squared	0.104603	S.D. dependent var		0.027524
S.E. of regression	0.026044	Akaike info criterion		-4.455050
Sum squared resid	1.361373	Schwarz criterion		-4.438336
Log likelihood	4490.007	Durbin-Watson stat		2.015712
Inverted MA Roots	.94 -.05i	.94+.05i	.93 -.17i	.93+.17i
	.90+.28i	.90 -.28i	.87 -.38i	.87+.38i
	.81 -.48i	.81+.48i	.74 -.57i	.74+.57i
	.68 -.66i	.68+.66i	.59 -.74i	.59+.74i
	.51+.80i	.51 -.80i	.40+.86i	.40 -.86i
	.30+.89i	.30 -.89i	.19+.93i	.19 -.93i
	.07 -.94i	.07+.94i	-.03 -.94i	-.03+.94i
	-.15+.94i	-.15 -.94i	-.26 -.91i	-.26+.91i
	-.36+.88i	-.36 -.88i	-.47 -.83i	-.47+.83i
	-.56 -.77i	-.56+.77i	-.65 -.71i	-.65+.71i
	-.73+.62i	-.73 -.62i	-.79+.53i	-.79 -.53i
	-.85 -.43i	-.85+.43i	-.89+.32i	-.89 -.32i
	-.93 -.23i	-.93+.23i	-.95	-.95+.11i
	-.95 -.11i			

As seen from the model, the existence of long memory is very evident. On the other hand the adjusted r-squared value is reasonably high and can not be ignored when it is compared the adjusted r-squared value found for close to close daily returns.

However, it should also be noted that, the adjusted r-squared value is again lower than the adjusted r-squared value attained in the autoregressive model for the same return series (average daily return series). The residuals and the squared residuals of the MA model below show similar properties (i.e. ARCH/GARCH effects) to that of the autoregressive model of the same return series (average daily return series).

It does also worth mentioning the fact that, moving average analysis of daily returns show that the first MA term is the most significant term and accounts for the largest part of adjusted R-squared value. Similar to the conclusion drawn for the MA model fitted to the session to session average returns it can be concluded that if there is an unexpected large return (i.e. the error term is positive), then the return of the next day will probably be positive and similarly, if there is an unexpected large negative return in a day, (i.e., the error term is negative), then the return of the next day will probably be negative. This is mainly because the coefficient of the MA(1) term is positive and statistically significant.

Table 61 - MA(1) Model for average daily return series

Dependent Variable: RET30AVGD
Method: Least Squares
Date: 05/05/06 Time: 16:05
Sample(adjusted): 2 2014
Included observations: 2013 after adjusting endpoints
Convergence achieved after 5 iterations
Backcast: 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	0.326265	0.021069	15.48560	0.0000
R-squared	0.093076	Mean dependent var		0.001486
Adjusted R-squared	0.093076	S.D. dependent var		0.027524
S.E. of regression	0.026212	Akaike info criterion		-4.444738
Sum squared resid	1.382334	Schwarz criterion		-4.441952
Log likelihood	4474.628	Durbin-Watson stat		2.001854
Inverted MA Roots	-.33			

Autoregressive And Moving Average Representation Of Session To Session Returns

The time series analysis of the return series is repeated by allowing both AR and MA terms into the equation, a method called ARMA modeling. The first step is to try both AR(1) and MA(1) in the model watch the significance of the coefficients. As seen from the output below, when evaluated together, the coefficients of the AR(1) and MA(1) terms are found to be insignificant. The adjusted R-squared of the equation is even worse than the equation having only AR(1) or MA(1) depicted in the previous sections.

Table 62 - MA(1) Model for session to session to session returns

Dependent Variable: RET30SEANS
Method: Least Squares
Date: 10/03/05 Time: 16:58
Sample(adjusted): 3 4014
Included observations: 4012 after adjusting endpoints
Convergence achieved after 13 iterations
Backcast: 2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.109031	0.173937	0.626840	0.5308
MA(1)	-0.019152	0.174951	-0.109468	0.9128
R-squared	0.007025	Mean dependent var		0.000741
Adjusted R-squared	0.006778	S.D. dependent var		0.022379
S.E. of regression	0.022303	Akaike info criterion		-4.767675
Sum squared resid	1.994704	Schwarz criterion		-4.764536
Log likelihood	9565.957	Durbin-Watson stat		1.999425
Inverted AR Roots	.11			
Inverted MA Roots	.02			

ARMA model is found to be quite sensitive to the relationship between the AR and MA terms. For example, when AR(2) term is added to the equation, the coefficients of all the AR and MA terms become significant as shown below.

Table 63 - An ARMA models for the session to session returns

Dependent Variable: RET30SEANS
Method: Least Squares
Sample(adjusted): 4 4014
Included observations: 4011 after adjusting endpoints
Convergence achieved after 15 iterations
Backcast: 3

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.886251	0.017920	-49.45731	0.0000
AR(2)	0.102320	0.015845	6.457732	0.0000
MA(1)	0.980059	0.008886	110.2981	0.0000
R-squared	0.012171	Mean dependent var		0.000734
Adjusted R-squared	0.011678	S.D. dependent var		0.022378
S.E. of regression	0.022247	Akaike info criterion		-4.772494
Sum squared resid	1.983630	Schwarz criterion		-4.767785
Log likelihood	9574.237	Durbin-Watson stat		1.996809
Inverted AR Roots	.10	-.99		
Inverted MA Roots	-.98			

Note that the adjusted R squared value improves and the Akaike Criterion becomes lower implying a better fit, but the roots of the AR and MA polynomials becomes very close to each other implying the common roots problem.

This problem is encountered more visible when the AR(3) and MA(3) terms are added to the model as shown below:This output means that the order of the ARMA(p,q) model can not go beyond 3 which in turn means that. the overall fit of this model is worse than both the AR and MA models. In ARMA modeling it is usually recommended to make the model as simple as possible, and to use either AR or MA terms not both especially when the data shows long term memory. This is quiet evident in our case. The overall fit could not be improved by using both AR and MA terms.

Table 64 - ARMA(3,3) model for session to session returns

Dependent Variable: RET30SEANS
Method: Least Squares
Date: 04/25/06 Time: 18:43
Sample(adjusted): 5 4014
Included observations: 4010 after adjusting endpoints
Convergence achieved after 17 iterations
Backcast: 2 4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.205880	0.163026	1.262868	0.2067
AR(3)	0.592394	0.102602	5.773690	0.0000
MA(1)	-0.120412	0.162052	-0.743046	0.4575
MA(2)	-0.020969	0.024239	-0.865124	0.3870
MA(3)	-0.629652	0.102115	-6.166137	0.0000

R-squared	0.010814	Mean dependent var	0.000737
Adjusted R-squared	0.009826	S.D. dependent var	0.022380
S.E. of regression	0.022269	Akaike info criterion	-4.769968
Sum squared resid	1.986169	Schwarz criterion	-4.762117
Log likelihood	9568.786	Durbin-Watson stat	1.985815

Inverted AR Roots	.91	-.35+.72i	-.35 -.72i
Inverted MA Roots	.91	-.39+.73i	-.39 -.73i

Analysis of the correlogram of residuals also shows that AR or MA models by themselves produce better results than ARMA model with respect to correlation structure of residuals. As seen below, the residuals of the ARMA model above can not easily be said to have the white noise property. In other words ARMA representation could not achieve to remove the autocorrelation inherent in the series, i.e. the error terms of ARMA equation seem to be correlated.

ARMA modeling is also repeated for the average return series by addition and deletion of the AR and MA terms, the common roots problem was again encountered, when the MA(7) term was added to the model as shown below. Note however the fact that the adjusted r-squared value is larger than that of the ARMA model constructed for close to close return series.

Table 65 - ARMA(6,7) model for average session to session returns

Method: Least Squares
Sample(adjusted): 8 4014
Included observations: 4007 after adjusting endpoints
Convergence achieved after 38 iterations
Backcast: 1 7

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.669171	0.004544	147.2700	0.0000
AR(6)	-0.470575	0.004882	-96.38737	0.0000
MA(1)	-0.249625	0.015255	-16.36382	0.0000
MA(2)	-0.279828	0.011001	-25.43692	0.0000
MA(6)	0.484065	0.011986	40.38618	0.0000
MA(7)	0.180897	0.014784	12.23606	0.0000
R-squared	0.161146	Mean dependent var		0.000742
Adjusted R-squared	0.160098	S.D. dependent var		0.018322
S.E. of regression	0.016791	Akaike info criterion		-5.334422
Sum squared resid	1.128065	Schwarz criterion		-5.324995
Log likelihood	10693.51	Durbin-Watson stat		1.968078
Inverted AR Roots	.91 -.41i	.91+.41i	.10 -.85i	.10+.85i
	-.68+.43i	-.68 -.43i		
Inverted MA Roots	.91 -.41i	.91+.41i	.09+.85i	.09 -.85i
	-.37	-.69 -.43i	-.69+.43i	

The above output shows that the order or AR and MA (p,q) should be less than 7. Thus, the following final ARMA model has been found for the average return series.

Table 66 - The final ARMA model for the average session to session returns

Convergence achieved after 28 iterations
Backcast: 2 7

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-1.094642	0.037225	-29.40617	0.0000
AR(6)	-0.102469	0.034956	-2.931384	0.0034
MA(1)	1.541040	0.040022	38.50506	0.0000
MA(2)	0.509175	0.024897	20.45140	0.0000
MA(6)	0.041359	0.017317	2.388384	0.0170
R-squared	0.169742	Mean dependent var		0.000742
Adjusted R-squared	0.168912	S.D. dependent var		0.018322
S.E. of regression	0.016703	Akaike info criterion		-5.345222
Sum squared resid	1.116505	Schwarz criterion		-5.337366
Log likelihood	10714.15	Durbin-Watson stat		1.996106
Inverted AR Roots	.48 -.32i	.48+.32i	-.12+.60i	-.12 -.60i
	-.82	-.98		
Inverted MA Roots	.34+.26i	.34 -.26i	-.20 -.48i	-.20+.48i
	-.86	-.96		

This model seems quite satisfactory, but the residuals and especially the squared residuals are again found to have significant autocorrelation. The fit as measured by the adjusted r-squared is also poorer compared to the AR and MA models for the average return series.

Autoregressive And Moving Average Representation Of Daily Returns

The same line of reasoning is also applied for the Arma modeling of daily close to close and average returns. ARMA model for the close to close daily returns not shown here, produced a very poor fit due to the common roots problem. For average returns the following final model has been reached.

Table 67 - ARMA model for the average daily returns

Method: Least Squares
Sample(adjusted): 17 2014
Included observations: 1998 after adjusting endpoints
Convergence achieved after 7 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(5)	-0.069424	0.022367	-3.103885	0.0019
AR(9)	0.050358	0.023503	2.142648	0.0323
AR(15)	0.050974	0.022277	2.288208	0.0222
MA(1)	0.329233	0.021172	15.55064	0.0000
MA(8)	0.054790	0.022301	2.456811	0.0141
MA(10)	0.049592	0.021263	2.332313	0.0198
R-squared	0.106634	Mean dependent var		0.001430
Adjusted R-squared	0.104392	S.D. dependent var		0.027540
S.E. of regression	0.026063	Akaike info criterion		-4.453575
Sum squared resid	1.353162	Schwarz criterion		-4.436758
Log likelihood	4455.121	Durbin-Watson stat		2.008821
Inverted AR Roots	.83	.73 -.35i	.73+.35i	.57+.60i
	.57 -.60i	.24+.78i	.24 -.78i	-.08 -.81i
	-.08+.81i	-.41+.73i	-.41 -.73i	-.65 -.46i
	-.65+.46i	-.81+.18i	-.81 -.18i	
Inverted MA Roots	.70+.26i	.70 -.26i	.37 -.61i	.37+.61i
	-.03 -.68i	-.03+.68i	-.43 -.62i	-.43+.62i
	-.77+.26i	-.77 -.26i		

The ARMA model could not be extended to further past lags due to the common root problem. The total fit of the model is also not better than the pure AR and MA models for average daily returns.

Fractional Integration Return Series

As seen from all the above AR and MA representations, both session to session return series and returns calculated from the averages exhibit persistent long memory that is found to be inherent in most financial time series. In recent years, to remedy this problem, increasing number of researchers try to integrate the series fractionally and then apply the moving average and autoregressive methods. This approach or method is called Auto regressive fractionally integrated moving average (ARFIMA) method. In order to overcome the persistence or long memory problem encountered in this analysis ARFIMA method is employed. To do that RATS software was used instead of Eviews, because Eviews does not support this method. The web site of RATS software is www.estima.com and this web site contains many procedures written by a researchers and programmers for employing newly developed and/or complicated algorithms. The procedure used is written by Baum and Barkoulas (1998).

Application Of Arfima Model To Session Close To Session Close Returns

When the procedure was run by using the log session to session return series the following output was obtained for fractional integration parameter “d”.

Geweke-Porter-Hudak Regression, Series LNRET30

Power = 0.50000 Regression Ordinates = 63

Estimated d = 0.07324

Asymp Standard Error = 0.09013 (0.813)

OLS Standard Error = 0.08021 (0.913)

The standard error of the fractional integration parameter is quite high and by using a simple t test the null hypothesis that d is equal to zero (no long memory) can not be rejected. Putting it in another way, let's find the 95 % confidence interval for d. The 95 % confidence interval is found by adding and subtracting 1.96 times the standard error. Using asymptotic standard error the confidence interval is found as (-0.10341,0.2499), while using the OLS standard error the confidence interval is found as (-0.08397,0230452).

Therefore the null hypothesis of no long memory can not be rejected at 95 % significance level. However this conclusion is tentative, because there is another testing procedure for the significance of parameter d called Lagrange Multiplier test proposed by Robinson (1994) which in some cases produces conflicting results to the results of t test.. Another shortcoming of the GPH method is that large outliers in data can bias the estimate of the long memory parameter toward zero (Tolvi 2003).Another important problem of GPH estimator is pointed out by Jensen who shows that GPH estimation of fractional integration parameter is not robust to the value of the first order autoregressive parameter. More specifically the GPH estimator underestimates the true value of d in case of small autocorrelations which is the case we face in our session to session logarithmic returns. Therefore it is assumed that the series have long memory with fractional integration parameter d = 0.07324.

By using this fractional parameter the log session to session series was fractionally integrated. In order to do that the procedures called xgamma.src and arfsim.src written for the RATS statistical packet is used. By using these two procedures a new series which is fractionally integrated (of order d= 0.07324) is generated. One important finding is that , the distribution of the new series becomes very close to normal.

The correlogram of the new series y is plotted and it is observed that the series exhibit nonzero significant autocorrelations up to lag 36a shown below.The significant AR terms in this new series were then tried to be found. Starting with the fist lagged AR terms and gradually adding the further lags the following final output was obtained.

Table 68 - Autoregressive model for the fractionally integrated session to session returns

Dependent Variable: Y
Method: Least Squares
Date: 03/27/06 Time: 19:30
Sample(adjusted): 5 4013
Included observations: 4009 after adjusting endpoints
Convergence achieved after 2 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.062208	0.015790	3.939728	0.0001
AR(2)	0.049392	0.015803	3.125484	0.0018
AR(3)	0.047818	0.015799	3.026601	0.0025
AR(4)	-0.038027	0.015792	-2.407930	0.0161
R-squared	0.010345	Mean dependent var		-0.007530
Adjusted R-squared	0.009603	S.D. dependent var		1.028868
S.E. of regression	1.023916	Akaike info criterion		2.886142
Sum squared resid	4198.854	Schwarz criterion		2.892425
Log likelihood	-5781.273	Durbin-Watson stat		1.999038
Inverted AR Roots	.35 -.22i	.35+.22i	-.32+.35i	-.32 -.35i

The lags further back into the past are found to have insignificant t values. This is quite good, since the long memory property was eliminated Only the first four lags are sufficient to model the process. Although the R-squared value is lower than that of the AR model with the original return series, we have only four lagged term here while we have 9 AR terms up to lag 32 in the previous model.

Moreover our model here does have a higher R-squared value compared to the model with first four AR terms. The correlogram of the residuals also produces a very nice result, namely, the autocorrelations after lag four are all insignificant up lag 36. This means that the model applied to the fractionally integrated series produces white noise and there is no long memory. Correlogram of squared residuals do also exhibit very interesting features. As it can be seen the ARCH effects virtually disappear especially after the lag 5. Lag 5 seems to be an exception. A similar result is obtained by using the MA terms as shown below..

Table 69 - MA modeling of the fractionally integrated series

Dependent Variable: Y
Method: Least Squares
Date: 03/27/06 Time: 19:48
Sample(adjusted): 1 4013
Included observations: 4013 after adjusting endpoints
Convergence achieved after 6 iterations
Backcast: -3 0

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	0.063071	0.015785	3.995567	0.0001
MA(2)	0.054151	0.015796	3.428257	0.0006
MA(3)	0.052201	0.015797	3.304603	0.0010
MA(4)	-0.033856	0.015792	-2.143826	0.0321

R-squared	0.010523	Mean dependent var	-0.006866
Adjusted R-squared	0.009783	S.D. dependent var	1.028763
S.E. of regression	1.023719	Akaike info criterion	2.885757
Sum squared resid	4201.431	Schwarz criterion	2.892034
Log likelihood	-5786.271	Durbin-Watson stat	2.000549

Inverted MA Roots	.32	.05 -.47i	.05+.47i	-.48

Note that, the first four MA terms have significant t values and the other lags are all found to be insignificant Again, similar results have been found for residuals and squared residuals. Moreover the distribution of the residuals of both the Autoregressive and the Moving Average models are found to be normally distributed.

Application Of Arfima Model To Session To Session Average Returns

The autocorrelation and partial autocorrelation of return series calculated from averages shows a different picture than that of the close to close return series. Namely, there is a positive and significant autocorrelation in average return series and a sudden drop to zero in autocorrelations after the first lag. The partial autocorrelation on the other hand is positive and significant for the first lag and negative and significant for the second order lag. All the other lags do have almost zero autocorrelation and partial autocorrelation values. In short the return series calculated from averages exhibit much more significant autocorrelation than session close to session close return series.

On the other hand the average returns their squared values exhibit similar persistency problem in autocorrelation and partial autocorrelation values as found in the close to close return series.

To tackle the persistency problem of the average return series Arfima is again employed. Contrary to the good results obtained for close to close return series, ARFIMA model is seemingly not sufficient to make the autoregressive model better. This is because the average return series is first analyzed to find the fractional differencing parameter. The following output is taken after running the GPH procedure in RATS.

Geweke-Porter-Hudak Regression, Series AVGRET

Power = 0.50000 Regression Ordinates = 63

Estimated d = 0.07222

Asymp Standard Error = 0.09013 (0.801)

OLS Standard Error = 0.07994 (0.903)

The fractional differencing parameter is found to be very close to the one that is found for close to close return series. The main difference is that the magnitude of the autocorrelations in average return series is considerably higher than that of the close to close return series. This may cast some doubt on the significance of d. Because as Jensen pointed out, the fractional integration parameter is biased downward in case low autocorrelation value. Since the average return series does not have this property, the results should be interpreted with some care.

When the fractional differencing parameter was applied to the average return series a new series was obtained and the distributional checks resulted in similar conclusions, i.e. the distribution was close to normal. The new series exhibits significant autocorrelations again a similar result that was found for close to close return series. However when an autoregressive model was tried to be fitted to the new series, it was found that the series does still exhibit some sort of long memory. The original average return series have significant AR terms up to lag 30, the new series have significant AR terms up to lag 14 shown below. The correlogram of residuals and even the squared residuals shows that the serial autocorrelation and dependence is eliminated, but the long term memory problem still remains, the series does still exhibit some form of long memory and additionally, the value of R-squared becomes equal to a very low value, namely it was 0.174581 in the original series with twelve AR terms and 0.158858 with only the first two AR terms (AR(1) and AR(2)), it becomes equal to 0.012029 with 5 AR terms up lag 14 and this value is clearly very low. Thus the fractional integration method applied to average return series proved to be of almost no use.

Table 70 - AR modeling fractionally integrated average return series

Dependent Variable: Y
Method: Least Squares
Sample(adjusted): 15 4013
Included observations: 3999 after adjusting endpoints
Convergence achieved after 2 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.082609	0.015782	5.234414	0.0000
AR(2)	0.047984	0.015792	3.038518	0.0024
AR(5)	0.035538	0.015728	2.259553	0.0239
AR(12)	-0.032258	0.015762	-2.046616	0.0408
AR(14)	0.035665	0.015763	2.262631	0.0237
R-squared	0.013017	Mean dependent var		0.016597
Adjusted R-squared	0.012029	S.D. dependent var		1.017671
S.E. of regression	1.011532	Akaike info criterion		2.862059
Sum squared resid	4086.649	Schwarz criterion		2.869928
Log likelihood	-5717.686	Durbin-Watson stat		2.000679
Inverted AR Roots	.77	.72 -.30i	.72+.30i	.52 -.61i
	.52+.61i	.20 -.79i	.20+.79i	-.18 -.78i
	-.18+.78i	-.52 -.60i	-.52+.60i	-.71 -.31i
	-.71+.31i	-.74		

Similar results have been obtained when the analysis is repeated with MA modeling. Thus the fractional integration did not help very much for the average return series as it did for the close to close returns.

Application Of Arfima Model To Daily Return Series

A similar method is used to fractionally integrate daily close to close returns as it has been used for session to session returns. By using the GPH method, fractional integration parameter of daily returns (close to close) is found as follows:

Geweke-Porter-Hudak Regression, Series DAILYRET

Power = 0.50000 Regression Ordinates = 44

Estimated d = 0.06458

Asymp Standard Error = 0.11111 (0.581)

OLS Standard Error = 0.09916 (0.651)

Using the fractional integration parameter found above, the daily close to close returns are also integrated and a new series is obtained. The first finding is the fact that, the distribution of fractionally integrated daily close to close returns becomes almost normal. Moreover, the new series have been found have no serial correlation, the squared residuals are also found to be uncorrelated. Since the fractionally integrated series of daily close to close returns exhibit almost strict white noise property, no AR, MA or ARMA model is needed.

Since daily average returns are found to have the long term memory problem the fractional integration method is also applied to daily average returns, as it has been done for session to session returns. The following output shows that the fractional integration parameter is 0.05624.

Geweke-Porter-Hudak Regression, Series DAILYAVERDAILY

Power = 0.50000 Regression Ordinates = 44

Estimated d = 0.05624

Asymp Standard Error = 0.11111 (0.506)

OLS Standard Error = 0.09884 (0.569)

After finding the value of the parameter "d" average return series is fractionally integrated by an order of d (0.05624) using the RATS package. The new series is found to be normally distributed as expected.

Additionally, the correlogram of the new series obtained from average returns show that the new series have somewhat a similar serial autocorrelation structure to that of the original average return series (shown below). More specifically the lag 1 and lag 2 autocorrelations are significant, but the magnitudes are smaller than the original series.

After some trial and error the following autoregressive model for the fractionally integrated average returns series was obtained.

Table 71 - AR model fitted to the fractionally integrated average return series

Method: Least Squares
Sample(adjusted): 24 2013
Included observations: 1990 after adjusting endpoints
Convergence achieved after 2 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.073078	0.022343	3.270714	0.0011
AR(23)	-0.049862	0.022301	-2.235822	0.0255

R-squared	0.007346	Mean dependent var	0.024590
Adjusted R-squared	0.006847	S.D. dependent var	1.008111
S.E. of regression	1.004653	Akaike info criterion	2.848167
Sum squared resid	2006.545	Schwarz criterion	2.853791
Log likelihood	-2831.926	Durbin-Watson stat	2.003022

As seen, the long memory could not be eliminated as efficiently as it was done for the close to close return series. The series does have an AR(23) term and moreover the adjusted r-squared value is quite low compare to that of the original average return series.

The correlogram of the residuals of the above AR model is found to have no serial correlation, correlogram of the squared residuals does seem to have no serial dependence. Thus although the long memory is not fully eliminated, and although the fit as measured by the low magnitude of the adjusted r-squared statistics, the residuals of the final model is almost white noise.

CHAPTER VI

VECTOR AUTOREGRESSIVE MODEL OF INDEX RETURNS

General Representation of the VAR Model

A vector autoregressive (VAR) model is distinguished from a classical autoregressive representation by the fact that vectors of variables and coefficient matrices are used instead of scalar variables and their corresponding coefficients.

Up to this section, return series are all analysed by using the autoregressive and moving average models and fractional integration methods. In all of the above methods just a single variable namely the return was analysed and thus the behaviour of returns as a function of its past values was tried to be modeled.. However , other variables such as volume, volume dispersion, return dispersion etc. may have some effect on the return generating process. Therefore this part of the thesis is devoted to analysis of the other variables in the process. In other words, using the past values of all the price and volume variables which are all endogenous, a better fit for the session to session and daily returns of the İMKB30 index was investigated This model is called vector autoregressive model which is distinguished from a univariate autoregression by the fact that single ('scalar') variables are replaced by vectors of variables and all coefficients are replaced by coefficient matrices. A formal representation of the model can be written as follows:

$$y_t = \phi_0 + \Theta_i y_{t-i} + v_t$$

The above notation is used for VAR(i) models, where the subscript i stands for the number of lags. The first term on the right hand side of the above equation is a k dimensional vector of constants , coefficient of the term y_{t-i} is a k by k matrix and v_t is a sequence of serially uncorrelated random error vectors with a contemporaneous variance covariance matrix of shocks Σ. An important property is the fact that, the covariance matrix denoted by Σ must be positive definite, meaning that none of the shocks is perfectly linearly dependent on the others. Note that a real positive definite symmetric matrix can always be transformed to a diagonal matrix by a unique lower triangular matrix with 1's on the diagonal. This is a useful property, because it means the error terms can be redefined as orthogonal to each other. In other words the matrix of dependent error terms can be converted to another matrix containing orthogonal (independent) error terms

More specifically let's take the case of a first order vector autoregressive process [VAR(1)] with two variables, namely y1 and y2. For example, y1 can be the return while y2 can be the volume. The model can be represented by the following set of equations.

$$y_{1,t} = \phi_{10} + \theta_{11} y_{1,t-1} + \theta_{12} y_{2,t-1} + v_{1t}$$

$$y_{2,t} = \phi_{20} + \theta_{21} y_{1,t-1} + \theta_{22} y_{2,t-1} + v_{2t}$$

The coefficients denoted by θij is the (i,j)the element of the 2x2 matrix Θ, and ϕ_{i0} is the i^{th} element of the vector ϕ_0. The two equations above is called a system of simultaneous equations describing the dynamic relationships between the two variables. Since the lagged terms were written as independent variables in each of the above equations this representation was called reduced form. In other words the above equation does not say anything about the concurrent relationships among variables, it does provide information about the lead lag relationships among variables. The first equation says that the variable y_1 can be written as a linear function of its first lag and the first lag of another variable denoted by y_2. The coefficients denoted by θ_{ij} stands for the magnitude of the linear dependence of the dependent variable to the explanatory variables on the right hand side. More specifically, θ_{12} denotes the linear dependence of the variable y_1 at time t on variable y_2 at time t-1, after the effect of the variable y_1 at time t-1 is accounted for. In other words θ_{12} is the conditional effect of $y_{2,t-1}$ on $y_{1,t}$ given $y_{1,t-1}$. The variables in vector autoregressive models can be exogenous or endogenous depending on their nature. In our case since all the variables are calculated from the trade data, they are all endogenous.

In general a VAR(p) model (with two variables) can be written as

$$y_{1,t} = \phi_{10} + \theta_{11} y_{1,t-1} + \theta_{12} y_{1,t-2} + ... + \theta_{1p} y_{1,t-p} + \theta_{1p+1} y_{2,t-1} + \theta_{1p+2} y_{2,t-2} + ... + \theta_{1p+p} y_{2,t-p} + v_{1t}$$

$$y_{2,t} = \phi_{20} + \theta_{21} y_{1,t-1} + \theta_{22} y_{1,t-2} . + ... + \theta_{2p} y_{1,t-p} + \theta_{2p+1} y_{2,t-1} + \theta_{2p+2} y_{2,t-2} + ... + \theta_{2p+p} y_{2,t-p} + v_{2t}$$

Where;

E(v_t)=0

E($v_{1t}v_{2s}$) = σ_{12} for t=s (the error terms at the same time period)

E($v_{1t}v_{2s}$) =0 otherwise

Another assumption of the VAR modeling is the fact all the variables included in the model must be stationary. In fact the variables included in the VAR model should be jointly stationary which means that, in addition to having constant variance and constant auto covariance across time, their cross correlation should also be constant across time.

More specifically,two random variables (x and y) are jointly stationary if both of them are individually stationary and the cross correlation function $\rho_{xy} = (t_1, t_2)$ depends only on the difference between t_1 and t_2.

In VAR modeling AR roots of the polynomial should also be less than one in modulus so that the VAR model is assumed to be stable, otherwise the process can not be modeled as a finite sum. In VAR modeling, examination of the impulse response functions which are responses of all variables in the model to a one unit structural shock to one variable in the model can be very informative The impulse responses are usually plotted on the Y-axis with the periods from the initial shock on the X-axis.

A statistical model of the following form which says that the index return can be modeled as function of its previous values, the previous values of the volume data, the previous values of the other variables such as net volume, range etc. is the main objective of the VAR modeling.

$$R_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i R_{t-i} + \sum_{j=1}^{p} \beta_i Vol_{t-i} + \sum_{j=1}^{p} \lambda_i NetVol_{t-i} \sum_{j=1}^{p} \delta_i Range_{t-i} + \sum_{j=1}^{p} \phi_i DIR_{t-i} + \sum_{j=1}^{p} \varphi_i RD_{t-i} + \sum_{j=1}^{p} \gamma_i RV_{t-i} + \varepsilon_t$$

The squared values of previous returns (a proxy for the variance), and the cubed values of the returns will also be included in the analysis to see their effect on prediction accuracy of the return.

Analysis Of The Session To Session Return Series With The Var Model

In the VAR model, 14 variables, namely, Ret30seans, Ret30sqr, Ret30vol, Ret30cube, Ret_disp, Retvolsns, Vol30chg, Voldispadj, Dir, Range, Artaz, Minfark, Maxfark, ret30avgseans are included. The variables are all calculated form the trade data, therefore they are all assumed to be endogenous.

As a first step, the VAR order selection process must be run. The results for the various selection criteria is shown in the output below. As seen from the table, different lag selection criteria point to different lag lengths. For example, according to Schwarz Information Criterion with lag 3 should be chosen, while HQ criterion says the proper lag length is 5. FPE and AIC selects lag 12, while LR selects 20. Lag 12 was adopted since both of the FPE and AIC suggest that the same lag length.

Table 72 - VAR Order Selection Criteria for session to session returns

VAR Lag Order Selection Criteria
Endogenous variables: RET30SEANS RET30SQR RET30VOL RET30CUBE RET_DISP RETVOLSNS VOL30CHG VOLDISPADJ DIR RANGE ARTAZ MINFARK MAXFARK RET30AVGSEANS
Exogenous variables: C
Date: 06/11/06 Time: 11:43
Sample: 1 4015
Included observations: 3993

Lag	LogL	LR	FPE	AIC	SC	HQ
0	75836.16	NA	1.79E-34	-37.97754	-37.95548	-37.96972
1	102296.0	52720.91	3.46E-40	-51.13249	-50.80157	-51.01518
2	103998.4	3380.093	1.63E-40	-51.88702	-51.24723	-51.66021
3	104950.9	1884.485	1.11E-40	-52.26593	-51.31728*	-51.92963
4	105550.9	1182.837	9.09E-41	-52.46828	-51.21076	-52.02248
5	105997.0	876.3980	8.02E-41	-52.59356	-51.02718	-52.03827*
6	106375.5	740.8508	7.32E-41	-52.68496	-50.80972	-52.02018
7	106586.1	410.6481	7.27E-41	-52.69225	-50.50814	-51.91797
8	106821.5	457.5949	7.13E-41	-52.71201	-50.21904	-51.82824
9	107026.8	397.5115	7.09E-41	-52.71666	-49.91483	-51.72340
10	107274.2	477.3736	6.92E-41	-52.74242	-49.63173	-51.63966
11	107466.6	369.7236	6.93E-41	-52.74058	-49.32102	-51.52833
12	107672.2	393.8136	6.90E-41*	-52.74539*	-49.01697	-51.42365
13	107814.8	272.1164	7.08E-41	-52.71864	-48.68136	-51.28740
14	107964.8	285.3134	7.25E-41	-52.69563	-48.34949	-51.15490
15	108121.8	297.2804	7.40E-41	-52.67606	-48.02106	-51.02584
16	108262.8	266.2456	7.61E-41	-52.64855	-47.68468	-50.88883
17	108392.9	244.5398	7.86E-41	-52.61552	-47.34279	-50.74631
18	108569.4	330.6745	7.94E-41	-52.60577	-47.02417	-50.62706
19	108726.4	293.0460	8.10E-41	-52.58624	-46.69578	-50.49804
20	108897.0	317.2121*	8.21E-41	-52.57353	-46.37420	-50.37583

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

By using the Eviews software, the VAR model with all the fourteen variables cited above was run. The VAR model with 14 variables up to 12 lags has an adjusted R-squared value of 0.083242 for the session returns. The model is also checked by looking at the AR roots in order for the model to satisfy the stationarity condition. All the AR roots are found to fall within the unit circle, which in turn means that the VAR process is stable.

Although the adjusted r-squared value increased about four times with the VAR analysis compared to the adjusted R-squared value (0.0215) of the AR model for the return series including up the lag 32, the VAR model with 14 variables, 12 lags is too complicated. Therefore a more parsimonious model is to be explored. It is clear that when the number of variables and the lag length is quite high and the results become difficult to interpret. In order to find the effect of the lagged values of all the variables on the returns, impulse response functions were all analyzed. From the plot of the impulse response functions for the session to session returns, the variables to be excluded from the VAR model were investigated.

First the variable "ret_disp" which stands for return dispersion was found to have almost no effect on the return, thus it was excluded, then the resulting VAR model was found to have an adjusted r-squared value of 0.0838. As a second step, the variable called "vol30chg" was dropped out of the var model and the VAR model was re-run, it was found that the value of adjusted r-squared statistics remained around 0.083.

And finally the variables called "voldispadj" and "vol30chg" were excluded from the model and the VAR model reduced to a model with 10 variables and 12 lags. The VAR output with 10 variables, namely, Ret30seans, Ret30sqr, Ret30cube, ret30vol, Retvolsns, Dir, Range, Artaz, Maxfark, ret30avgseans with lags up 12 have an adjusted r-squared value of 0.82. It is quite interesting however to see the fact that, Eviews software points to different lag length when the lag selection criteria was run with different number of variables. On the other hand, HQ and SIC seems more stable compared to AIC and FPE, s the number of variables change. For this reason, as suggested by the HQ criterion the VAR model with the first five lagged values of all the variables was run. In this case the adjusted R-squared value dropped to 0.076 which can be regarded as a quite tolerable fall, since the complexity of the VAR equation improved significantly. (From 10 variables up to lag 12 to 10 variables up to lag 5).

The VAR model was also run with three lagged values of all the variables as suggested by the SC criterion. This time an adjusted r-squared value of 0.069 was obtained and this was regarded as a big drop, thus the VAR model with five lagged values was preferred.

The final VAR model with 10 variables and up to five lagged values are shown in the Appendix E . The impulse response graphs are depicted in Appendix F. From the impulse response functions the session returns were found to be affected positively by their own lag 1 values, by the lag 1 value of return squared, by lag1 value of retvolns and artaz. The effects of the lagged values of the variables retvolsns and artaz are more visible than that of the ret30seans and ret30sqr. In short, the results mean that an up market should be expected after

- a volatile session (as measured by the return squared)
- a session in which number of stocks with positive returns exceeds that of the falling stocks
- a session in which the stocks with positive return dominate the market with their high volume.

The effect of volume is quite visible in impulse response functions when one looks at the impulse response function of retvolsns and ret30seans. If there is a positive shock to retvolsns, then the index for the next session is expected to be higher. Thus if the stocks with positive returns do also have large changes in volume, that means there is a positive shock to the retvolsns which in turn implies a rise in the index for the next period. Conversely if either the individual stock returns are lower or the individual volume changes of rising stocks are lower then this means that there is a negative shock to this variable which in turn means that a down market is more probable than an up market. Finally the residuals of the VAR model were examined. As it can be seen from the output below the residuals are correlated after the VAR length.

Table 73 - VAR Residual Portmanteau Tests for Autocorrelations

H0: no residual autocorrelations up to lag h
Sample: 1 4015
Included observations: 4008

Lags	Q-Stat	Prob.	Adj Q-Stat	Prob.	df
1	6.642019	NA*	6.643677	NA*	NA*
2	26.14802	NA*	26.15941	NA*	NA*
3	102.1143	NA*	102.1826	NA*	NA*
4	205.4764	NA*	205.6479	NA*	NA*
5	361.1583	NA*	361.5243	NA*	NA*
6	566.3425	0.0000	567.0162	0.0000	100
7	766.5263	0.0000	767.5502	0.0000	200
8	939.8132	0.0000	941.1836	0.0000	300
9	1166.023	0.0000	1167.903	0.0000	400
10	1360.414	0.0000	1362.780	0.0000	500
11	1489.886	0.0000	1492.608	0.0000	600
12	1622.321	0.0000	1625.441	0.0000	700
13	1774.708	0.0000	1778.324	0.0000	800
14	1933.488	0.0000	1937.660	0.0000	900
15	2092.643	0.0000	2097.413	0.0000	1000
16	2279.780	0.0000	2285.300	0.0000	1100
17	2415.595	0.0000	2421.694	0.0000	1200
18	2553.712	0.0000	2560.434	0.0000	1300
19	2654.199	0.0000	2661.399	0.0000	1400
20	2927.396	0.0000	2935.966	0.0000	1500

*The test is valid only for lags larger than the VAR lag order.
df is degrees of freedom for (approximate) chi-square distribution

For the average returns the picture is different. As it can be seen from Appendix E, the adjusted r-squared for the average returns in the VAR model up to five lags is around 0.3, which is more than three times higher than that of the close to close returns. The impulse response function are also very helpful in explaining the behavior of average returns. The impulse response functions of the average returns are provided in Appendix G. As seen from the graphs, the average returns are affected by the lag 1 value of ret30seans, artaz dir and retvolsns on the same direction. In other words average returns are expected to be higher after a session with positive return, with more increasing stocks than decreasing and with higher retvolsns value.

The VAR model is simplified by having the five variables namely, the session to session returns, average returns, retvolsns, artaz and dir up to two lags, the following output is obtained.

Table 74 - VAR Model with five variables and two lags

Vector Autoregression Estimates
Date: 06/11/06 Time: 13:01
Sample(adjusted): 4 4014
Included observations: 4011 after adjusting endpoints
Standard errors in ( ) & t-statistics in [ ]

	RET30SEANS	RETVOLSNS	DIR	ARTAZ	RET30AVGSEANS
RET30SEANS(-1)	-14.66446	-407.5428	-6.829700	-48.00982	-7.337813
	(4.22142)	(137.036)	(1.42855)	(103.591)	(3.00714)
	[-3.47382]	[-2.97399]	[-4.78088]	[-0.46345]	[-2.44013]
RET30SEANS(-2)	0.207045	3.276458	0.060177	6.888351	0.147337
	(0.08033)	(2.60772)	(0.02718)	(1.97129)	(0.05722)
	[2.57738]	[1.25644]	[2.21366]	[3.49433]	[2.57472]
RETVOLSNS(-1)	0.012949	0.434080	0.003027	0.250478	0.009931
	(0.00162)	(0.05274)	(0.00055)	(0.03986)	(0.00116)
	[7.97115]	[8.23134]	[5.50641]	[6.28319]	[8.58186]
RETVOLSNS(-2)	-0.004397	-0.067792	-0.001709	-0.124433	-0.002702
	(0.00164)	(0.05308)	(0.00055)	(0.04013)	(0.00116)
	[-2.68913]	[-1.27714]	[-3.08919]	[-3.10102]	[-2.31934]

DIR(-1)	14.25397	394.0714	6.668949	38.76591	8.086940
	(4.22738)	(137.229)	(1.43056)	(103.738)	(3.01139)
	[3.37182]	[2.87163]	[4.66176]	[0.37369]	[2.68545]
DIR(-2)	-14.24301	-392.8136	-6.730427	-40.06733	-7.016484
	(4.22008)	(136.992)	(1.42809)	(103.559)	(3.00619)
	[-3.37505]	[-2.86741]	[-4.71287]	[-0.38691]	[-2.33401]
ARTAZ(-1)	0.005970	0.216592	0.001974	0.285588	0.003993
	(0.00135)	(0.04387)	(0.00046)	(0.03316)	(0.00096)
	[4.41798]	[4.93740]	[4.31558]	[8.61206]	[4.14760]
ARTAZ(-2)	0.000911	0.005446	0.000352	-0.003238	0.000553
	(0.00134)	(0.04361)	(0.00045)	(0.03297)	(0.00096)
	[0.67786]	[0.12488]	[0.77493]	[-0.09822]	[0.57758]
RET30AVGSEANS(-1)	14.22919	392.9183	6.731278	35.49162	7.001020
	(4.22621)	(137.191)	(1.43017)	(103.709)	(3.01055)
	[3.36689]	[2.86402]	[4.70664]	[0.34222]	[2.32549]
RET30AVGSEANS(-2)	-0.134972	-1.813701	-0.043344	-3.351489	-0.091417
	(0.04678)	(1.51844)	(0.01583)	(1.14786)	(0.03332)
	[-2.88549]	[-1.19445]	[-2.73820]	[-2.91977]	[-2.74350]
C	-1.44E-05	0.081755	0.000824	-0.005573	-0.000836
	(0.00043)	(0.01380)	(0.00014)	(0.01043)	(0.00030)
	[-0.03385]	[5.92481]	[5.72694]	[-0.53428]	[-2.76228]
R-squared	0.043603	0.046259	0.033031	0.050460	0.275614
Adj. R-squared	0.041212	0.043875	0.030613	0.048086	0.273803
Sum sq. resids	1.920513	2023.798	0.219932	1156.503	0.974557
S.E. equation	0.021912	0.711301	0.007415	0.537704	0.015609
F-statistic	18.23649	19.40115	13.66357	21.25645	152.1916
Log likelihood	9639.088	-4319.471	13985.06	-3197.243	10999.55
Akaike AIC	-4.800842	2.159297	-6.967871	1.599722	-5.479206
Schwarz SC	-4.783573	2.176566	-6.950602	1.616991	-5.461937
Mean dependent	0.000734	0.116152	0.000867	0.001837	0.000737
S.D. dependent	0.022378	0.727438	0.007531	0.551118	0.018317
Determinant Residual Covariance		8.57E-20			
Log Likelihood (d.f. adjusted)		59591.07			
Akaike Information Criteria		-29.68640			
Schwarz Criteria		-29.60006			

As seen, the adjusted r-squared value for the average return is still quite high (0.275). In short the average return is expected to be higher in the next session, if the return calculated from the closing values of the index is positive in the current session. The average return for the next period is also expected to be positive if the variable retvolsns has a positive shock and if the variable dir has a positive shock in the current session. Lag 1 and lag2 values of the variable artaz is also positively correlated with the average returns as seen from the impulse response functions provided in Appendix H.

An interesting point to note is that the statistics called maxfark and minfark does have a relatively high adjusted r-squared value (0.41) in the VAR equation. A quick check with addition of the variable minfark shows that minfark does also have a high adjusted r-squared value in the VAR equation. This finding deserves to be explored in more detail. The variables maxfark and minfark are found to be not affected by their previous values. For example, the autoregressive model for the minfark is provided below. As seen, the model has a very low r-squared value.

Table 75 - AR modeling of the variable Minfark

Method: Least Squares
Sample(adjusted): 22 4014
Included observations: 3993 after adjusting endpoints
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.219238	0.015793	13.88222	0.0000
AR(2)	-0.048952	0.015842	-3.090082	0.0020
AR(5)	-0.035627	0.015441	-2.307324	0.0211
AR(12)	-0.035757	0.015458	-2.313217	0.0208
AR(20)	0.063753	0.015580	4.092057	0.0000

R-squared	0.051726	Mean dependent var	7.092792
Adjusted R-squared	0.050774	S.D. dependent var	262.7453
S.E. of regression	255.9880	Akaike info criterion	13.92939
Sum squared resid	2.61E+08	Schwarz criterion	13.93727
Log likelihood	-27805.03	Durbin-Watson stat	1.992082

Inverted AR Roots	.87	.85+.27i	.85 -.27i	.71 -.52i
	.71+.52i	.53 -.70i	.53+.70i	.28 -.84i
	.28+.84i	.01 -.87i	.01+.87i	-.25 -.84i
	-.25+.84i	-.50 -.70i	-.50+.70i	-.68 -.51i
	-.68+.51i	-.82+.27i	-.82 -.27i	-.85

Impulse response analysis (shown in Appendix I) for the variables minfark and maxfark shows that these variables are affected by the values of the other variables especially by the previous values of the returns to a great extent.

As seen from the impulse response graphs, the variable minfark is not much affected by its previous values, rather it is affected by the lag 1 value of the return in the same direction. In other words a positive shock to the return in t-1, causes the minimum of the next session to be larger than the minimum of the current session.

A positive shock to the lag 1 value of the variable artaz affects the return of the next session on the same direction. In other words, if the ratio of number of rising stocks in ISE30 index gets higher, than the minimum of the next session will probably be higher than the minimum of the current session. Minfark is also influenced, positively from lag 1 value of maxfark, ret30sqr, range and retvolsns. Therefore the minimum of the next session should be expected to be higher than the minimum of the current session, if the current session has a high volatility and high retvolsns value. It is also interesting to see that, a positive shock to the average return has an opposite effect on the minimum of the next session. In other words, lf we have a larger than expected average return, the minimum of the next session is expected to be lower than the current session.

Similarly the variable maxfark is affected by its lag 1 value, the current and lag 1 values of the return, lag1 and lag 2 values of the variable artaz, lag 1 value of retvolsns, and lag 1 value of minfark on the same direction. More specifically it can be said that positive return implies that the maximum of the coming period will be higher than the maximum of the current period. Additionally, if the number of stocks with positive returns (as measured by artaz) are higher, then the maximum of the next period will probably be higher.

On the other hand, maxfark is influenced on the opposite direction by the lag 1 value of ret30sqr and range.. In other words, if the volatility either measured by square of the returns or measured by the range is higher, then the maximum value to be attained for the next period is expected to be lower. if the maximum of the current period is higher than the maximum of the previous period then the maximum of the next period will probably be higher. The effect of lagged values of average returns is similar to the one observed for the variable minfark. If there is positive shock to average return then the maximum of the next period will probably be lower than the current maximum. In order to make the case more explanatory, the variable minfark and maxfark is regressed with the variables previous return and previous return squared. As the following outputs show, the most significant impact to the variables minfark and maxfark comes from the previous return and squared returns.

Table 76 - Relationship between the minfark and the previous returns

Dependent Variable: MINFARK
Method: Least Squares
Date: 06/11/06 Time: 13:56
Sample(adjusted): 3 4014
Included observations: 4012 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RET30SEANS(-1)	5.552386	0.160802	34.52942	0.0000
RET30SQR(-1)	38.96624	2.765626	14.08948	0.0000

R-squared	0.247831	Mean dependent var	0.007119
Adjusted R-squared	0.247644	S.D. dependent var	0.262126
S.E. of regression	0.227364	Akaike info criterion	-0.124027
Sum squared resid	207.2952	Schwarz criterion	-0.120888
Log likelihood	250.7989	Durbin-Watson stat	2.063943

Table 77 - Relationship between the variable maxfark and the previous returns

Dependent Variable: MAXFARK
Method: Least Squares
Date: 06/11/06 Time: 13:57
Sample(adjusted): 3 4014
Included observations: 4012 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RET30SEANS(-1)	5.892340	0.157372	37.44221	0.0000
RET30SQR(-1)	-12.45855	2.706632	-4.602972	0.0000

R-squared	0.265956	Mean dependent var	0.007313
Adjusted R-squared	0.265773	S.D. dependent var	0.259683
S.E. of regression	0.222514	Akaike info criterion	-0.167151
Sum squared resid	198.5459	Schwarz criterion	-0.164012
Log likelihood	337.3043	Durbin-Watson stat	2.107072

As seen, reasonably high adjusted r-squared values have been obtained for each of the variables. The interpretation of the above results are as follows: If the return of a session is positive then the minimum value and the maximum value of the next session will probably be higher than the current session. Conversely if the session return is negative, then the minimum and the maximum of the next session will be lower. On the other hand, if the session is volatile then the minimum of the next session is expected to be higher than the minimum attained during the current session, while the reverse is true for the maximum value.

These results may provide some interesting implications for profitable trading. For example, if any time during a session, trades occur at levels under the minimum of previous session which was an up session, this might be regarded as a buy signal, since the minimum is expected to be higher than the minimum of previous session. On the other hand, if prices exceeds the maximum of the previous session which was a down session, this might be regarded as a sell signal.

Analysis Of The Daily Return Series With The Var Model

Vector autoregressive model for daily returns and all the other variables was also done to see whether there is a similar structure for the daily returns. The same line of reasoning is applied, namely, first step is to find the lag of the VAR model. As it was the case for the session to session returns, different criteria point to different lag lengths as shown below.

Table 78 - VAR Lag Order Selection Criteria

Endogenous variables: RET30 RET30SQR RET30CUBE RET30VOL RET_DISP RETVOLDAY VOL30CHG VOLDISP NETVOLCHG MINFARK MAXFARK DIR RANGE ARTAZ RET30AVGD
Exogenous variables: C
Sample: 1 2015
Included observations: 1795

Lag	LogL	LR	FPE	AIC	SC	HQ
0	19613.94	NA	1.07E-28	-21.83726	-21.79136	-21.82031
1	30984.36	22538.14	4.32E-34	-34.25556	-33.52115	-33.98442
2	32368.25	2719.983	1.19E-34	-35.54680	-34.12388*	-35.02147
3	32949.34	1132.383	8.00E-35	-35.94355	-33.83213	-35.16403*
4	33340.43	755.6006	6.65E-35	-36.12861	-33.32868	-35.09490
5	33637.84	569.6371	6.13E-35*	-36.20929*	-32.72085	-34.92139
6	33845.02	393.3570	6.26E-35	-36.18944	-32.01249	-34.64735
7	34014.95	319.7901	6.66E-35	-36.12808	-31.26262	-34.33180
8	34237.90	415.8466	6.68E-35	-36.12580	-30.57184	-34.07533
9	34423.92	343.8470	6.99E-35	-36.08236	-29.83989	-33.77770
10	34610.98	342.6509	7.30E-35	-36.04009	-29.10911	-33.48124
11	34727.66	211.7779	8.26E-35	-35.91940	-28.29992	-33.10636
12	34859.81	237.6511	9.18E-35	-35.81595	-27.50795	-32.74871
13	34996.71	243.8974	1.02E-34	-35.71778	-26.72128	-32.39636
14	35129.93	235.1295	1.13E-34	-35.61552	-25.93052	-32.03991
15	35280.56	263.3172	1.23E-34	-35.53265	-25.15914	-31.70285
16	35439.44	275.1031*	1.33E-34	-35.45899	-24.39696	-31.37499
17	35580.86	242.5079	1.47E-34	-35.36586	-23.61534	-31.02768
18	35698.77	200.2063	1.66E-34	-35.24654	-22.80750	-30.65416
19	35824.82	211.9445	1.87E-34	-35.13629	-22.00875	-30.28973
20	35942.74	196.2859	2.12E-34	-35.01698	-21.20093	-29.91623

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

The maximum lag length is again implied by the LR test statistic, and the FPE and AIC statistics pointed to the same lag length which is was 5, thus the lag length was chosen to be five for the VAR model of daily returns.

The only additional variable added to the VAR model for daily returns is the variable called netvochg which is calculated by using the TL value of shares actually change hand during a day. The VAR model started with 15 variables and up to the fifth lag. The number of variables included in the VAR model is quite large, therefore the variables were eliminated stepwise by excluding one by one, and monitoring the change in the adjusted r-squared value.

The final VAR model (Appendix J) is constructed by having 10 variables, namely, ret30, ret30sqr, ret30cube, ret30vol, artaz, dir, voldisp, maxfark, netvolchg, ret30avgd up to lag 5. All the AR roots of the VAR model are found to fall within the unit circle, which in turn means that the VAR process is stable. The adjusted R-squared value for the daily close to close return series of this model is 0.053 which is very low but still almost three times higher than the pure autoregressive model (0.016) found in the previous sections. Another important point in the VAR analysis of returns is that the newly added variable netvolchg which is the actual TL value of stocks changed hand during a trading day seems to contribute to the explanatory power of the VAR model.

The impulse response functions (Appendix K) are also analysed to see the lead lag effects of the variables on daily close to close returns. There has been found almost no significant effect of the variables on the daily returns in impulse response functions.

Although the VAR model is not very successful in explaining the dynamics of the close to close returns, it can be seen that the daily average returns in the VAR model have a considerably high adjusted r-squared value, namely the adjusted r-squared for the average returns in the VAR model up to five lags is 0.375 which is even higher than what has been found for close to close session returns. This means that the average returns can be modeled by having the lagged values of itself and the lagged values of the variables calculated such as volume, range etc. which are all calculated by using the trade data. The impulse response function graphs shown below also give very interesting hints for the lead lag effects of the variables on daily average returns.

By looking at the impulse response function for average returns (Appendix L) it can be said that a positive shock to the daily return increase the average return of the next day. If the index closes higher than the average during a day (measured by the variable dir), the average return of the next day will also be expected to be higher than today. A similar conclusion can be drawn for the effect of the variable artaz on daily average return. And the last comment is that, the index increases accompanied by volume increases has some positive effect on the average returns of days further beyond t+1, more specifically t+2 and t+3.

In addition to the variables obtained from the trade data, such as volume, volume dispersion return dispersion etc. the holdings of foreign investors on daily basis was also added to the VAR model to see if the information of the share of foreign investors in the equity market has some effect on the adjusted r-squared value for the return series. The percentage share of foreign investors are published on daily basis in Takasbank web site, data starts from the date May 4th 2004. It was found that the past values of the changes in the total share of foreign investors in equity market does not have any positive effect on the explanatory power of the VAR model.

The variable maxfark in the VAR model (Appendix J) of daily returns does have a large adjusted r-squared value (0.38). The impulse response graphs(Appendix M) for this variable reveal that the maximum of the next day will probably be higher than maximum of today if the index return calculated from closing values is positive, and/or if the index closes higher than the average value (the variable DIR). Additionally, If there is a positive shock to the variable artaz, then the maximum of next day will probably be higher than today's maximum.

Similarly just to see the effects of the other variables, the variable minfark is added to the VAR model and the impulse response function (Appendix M) is analysed. The adjusted r-squared value for maxfark is found to be even higher than the variable minfark. Analysis of the impulse response functions reveal that the minimum of next day is expected to be higher than today's maximum if today is a volatile session (measured by the squared value of the returns). On the other hand, if the index closes above the daily average (measured by the variable DIR), then the minimum of the next day will be probably be higher than today. In case of a positive shock to the lag1 value of the close to close return, then the minimum of the next day will probably be higher than today. The same conclusion can not be drawn for daily average returns. A positive shock to lag 1 value of maxfark and artaz and lag 2 value of ret30vol has also an upward effect on the minfark.

From the portfolio managers point of view some interesting hints can be detected by analyzing the impulse response functions. For example, impulse response functions imply that, if there is positive shock to the close to close return today, then the minimum of the next day will probably be higher than that of today, and maximum of the next day will probably be higher than today's maximum. If the prices in the next day falls below today's minimum, then it is signal to buy. Similar examples can be done, by looking at the combinations of close to close returns, the variables Dir, Artaz and ret30sqr.

CONCLUSION AND DISCUSSION

The main subject of this thesis is the return, volume and volatility dynamics in equity markets, and ISE Equity Market is investigated in particular. An extensive amount of research have been done in this field with an increasing number researchers concentrating on the behavioral finance explanations. The central mission of this study is to uncover equity market return dynamics in Turkish case with a special emphasis on the formation of expectations of market professionals. The empirical evidence verifying the wide spread use of technical analysis methods is looked for, new variables are discovered to contribute to the explanatory power of the analyses and some interesting conclusions are reached.

This dissertation starts by examining the previous research done in explaining return volume dynamics. Previous research seems to focus mainly on trying to explain the return dynamics by using the returns calculated from the closing prices.

An important contribution of this dissertation comes from the starting question which proved to be very fruitful. Why should one use or care about the close to close returns? Should the closing prices be taken as the prices at which trades mostly occur? It should be kept in mind that the closing prices are just the prices of the last trades which can be as small as one lot of the stocks. When someone wants to trade in the equity market it is quite likely that the trade prices will be very different from the closing prices. If a price is taken to measure the returns it will be much more convenient to take the average or the weighted average prices during a trading period, since the probability that the trade price of any trade being close to the average price is generally greater than the probability of being close to the closing price. On the other hand, the fact that mutual funds evaluate their stock portfolios by using the average prices rather than the closes and consequently, the sale or purchase prices of mutual funds by the investors depend on the average prices of stocks further initiated to use some other measure for calculating the returns.

So, the initial question leads us to analyse the average returns as well as the close to close returns. It did prove to be a good decision to add the return series calculated from the averages in the analyses.

On the other hand, since the variable volume is also used in the time series analysis, the return series are calculated on session to session and daily basis. This is due to the fact that the volume data is not available for shorter time intervals in the ISE.

The first finding of this study is that, both close to close and the average returns are found to deviate significantly from the normal distribution, additionally, the return series exhibit fat tail property and heteroskedasticity. These findings are in accordance with the results of almost all the previous studies in the field.

Another finding is that, close to close returns from session to session and from day to day have been found to have very low serial correlation. The daily close to close returns especially for the last three year period of the data exhibited even no serial autocorrelation. Therefore the ISE Equity Market seems to be a very nice example of an efficient market. However, the returns calculated from the averages are found to exhibit significant autocorrelations and partial autocorrelations at lags one and two. This finding can be regarded as evidence against the efficient market hypothesis. This conclusion about the average returns should however be taken with care, because the ISE Equity

Market base prices on which the prices limits are calculated are taken to be the average prices of the previous period. Therefore this may induce some artificial serial dependence in average returns. The researchers are strongly recommended to check the results in different equity markets around the globe to discover the real reasons for this phenomenon.

Due to the low serial correlation of close to close return series, Autoregressive (AR) models and Moving Average Models (MA) produced very small adjusted r-squared values. On the other hand, these models produced larger adjusted r-squared values when the average returns were used as the dependent variable. In general, the pure Autoregressive Models produced better fit than the pure Moving Average Methods and these two methods also generally have larger adjusted r-squared value than the ARMA models. ARMA models generally could not be extended to further lags in the past due to the common root problem.

Both the close to close return series and the average return series are discovered to exhibit the long memory or the persistence problem, a common finding of recent literature in the field. This finding leads us to explore fractional integration methods which are known as Autoregressive Fractionally Integrated Moving Average Methods (ARFIMA). It was found that fractional integration method is more useful for high frequency data especially for session close to session close returns. Fractional integration has been found to transform the close to close return series to a series which is normally distributed and has a short term memory. On the other hand the fractional integration was found to be of not much use for the average session to session returns and for daily returns.

When the AR and MA models are applied to the new series obtained from the fractionally integrated close to close return series, it is observed that the adjusted r-squared value gets worse, but the long term memory problem can be sorted out. The resulting series on the other hand are found to exhibit no hetersoskedasticity.

Another important contribution of this dissertation is the expectation survey which provides interesting clues with regard to the effect of lagged variables such as return, volume, volatility and return dispersion on the expectations of market people.

Technical analysis is found to the most common method used for portfolio management among the brokers. Intuition and/or feelings is discovered to have a considerable weight on the decision making process, providing further support to increasing number of studies focusing on explaining the investor behaviour using psychological theories. It has also been discovered that people use generally more than one method for investing.

The survey revealed the fact that, brokers give special importance to previous return, volume, volatility and other trade variables in forming their expectations, a finding that is seemingly contrary to the efficient markets theory. The survey also revealed that the same information may lead to different expectations among brokers, therefore the well known uniform expectations assumption is also challenged by the results of this survey.

Additionally, the results of the survey are empirically investigated to find any supportive or contradictory evidence from the data. For example, although brokers seem to give special importance to changes in volume in forming their expectations, the empirical evidence generally pointed to the fact that, volume increase or decrease does not contain any evidence for the return of the next period. Volume changes in session to session returns may partially add to the information set for the next period return, but daily returns are not sensitive to previous volume increases or decreases.

Empirical evidence in the ISE Equity Market also suggests that the reversal is more probable than a further fall when the market falls in large amounts. This is partly in accordance with the well-known behavioral finance theory named as “the disposition effect”. According this theory, investors are risk taker in case of down movements, i.e., they generally do not sell the losing stocks. On the contrary, according to our survey, a sharp fall in the index implies a down market for the brokers. The survey generally resulted in the fact that, the brokers in the ISE are trend followers. Empirical analysis of data provided many evidences against such an investor behaviour.

The survey results should be evaluated with care, because the perception of the notions of “up market”, “strongly up market”, “horizontal move”, “down market”, “strongly down market” may differ across brokers. Similarly the notions of high volatility, low volatility, large return, small return etc. may mean different to various investors. Therefore, further research in this field is strongly recommended to use more specific definitions using numerical examples.

Finally, in order to find out the lead lag effects of returns and other trade variables, Vector Autoregressive (VASR) model was also applied. The Adjusted R-squared value for the session close to session close returns improved to some extent, but the improvement was found to be very minor for daily returns calculated from closing values. The adjusted r-squared value for the average returns on the other hand was found to be quite promising. The relatively high adjusted r-squared value found for average returns, may lead to profitable trading strategies. An analysis of the impulse response functions for close to close returns and for the average returns provide some hints towards making more educated guesses for the next period return.

Another important finding of the VAR analysis is that the difference between the minimum of two consecutive periods and the maximum of two consecutive periods has a relatively large adjusted r-squared value in the model implying that some profitable trading strategies can also be constructed by using this relationship. The investors may watch the trade prices and compare the current prices to the previous highs and lows and may come up with a profitable trading strategy. But the profitable trading opportunities should be carefully evaluated since; it is highly probable that if the effect of transaction costs is taken into account, the possible profits might be swept away.

This study is important, because it provides some insights into the expectation formation process of brokers in the ISE. There are interesting signs of investment behaviour that may well shake the foundations efficient markets theory. On the other hand, inclusion of some new variables into the study for uncovering return dynamics is believed to initiate further research in the field. This thesis also shows that that the use of the fractional integration method is not a very useful method to uncover the return dynamics.

Further research in this field is believed to concentrate more on the inclusion of new variables which are mostly overlooked up to now. Since the most extensively explored area is the close to close returns, the market seems to give no opportunity for profitable trading. However, a detailed analysis of the other variables such as average returns, difference between the minimums, difference between the maximums etc. may give rise to new profitable trading opportunities, taking into account, of course, the transactions costs. An expert system, a neural network model taking the lead lag interactions of the close to close returns, the average returns, minimums, maximum, volume and return dispersion etc. is believed to be an interesting challenge to see whether it is possible to beat the market or not.

A more detailed survey with concrete numerical examples is also believed to add to our understanding of the expectation formation process of the investors. Brokers are just a specific example of the total investor pool. The further research should also concentrate on the other investors as well as brokers. The researcher should also be ready however, to face substantial resistance from the respondents in participating the survey, since they are not generally so willing to dedicate their spare time to such a survey.

BIBLIOGRAPHY

Akgiray V., Conditional Heteroskedasticity in Time Series of Stock returns: Evidence and Forecasts, *Journal of Business*, 1989, vol 62 no. 1, pp 55-80

Akgiray, V., Booth, G.G., and Loistl, O. (1989), Stable Laws are Inappropriate for Describing German Stock Returns, *Allgemeines Statistisches Archiv*, 73, 115-21.

Alizadeh,S.,Brandt,M.W.,and Diebold,F.X.(2001),"Range-Based Estimation of Stochastic Volatility Models,"*Journal of Finance* ,June 2002, Vol 57, 3

Aparicio, F. M., Estrada, J., Empirical Distributions of stock returns: European securities markets, 1990-1995,*The European Journal of Finance*, 2001, 7, pp.1-21

Barberis, Nicholas, and Richard Thaler, 2003, A Survey of Behavioral Finance, *Handbook of the Economics of Finance* (Elsevier Science Ltd, North-Holland, Amsterdam

Barkaulas John T., Baum Christopher F., Fractional Dynamics in Japanese Financial Time Series. *Pacific Basin Journal Finance Journal*. Vol:6.pp 115-129

Basçi E., Basçi S., Muradoğlu Gülnur, Do extreme falls help forecasting stock returns? Evidence from world markets. *City University Business School, London Working paper*, 2001, No:6

Basci, E., Özyildirim S., Aydogan K., 1996, A note on price-volume dynamics in an emerging stock market, *Journal of Banking and Finance*, Vol 20 issue 2, pp 389-400

Bloomfield P., 1976, Fourier Analysis of Time Series: An Introduction, *John Wiley& Sons*

Blume, L., Easley, D., O'Hara, M., Market Statistics and Technical Analysis: The Role Volume, *Journal of Finance Vol: 49, No:1*, 1994,153-181.

Bollerslev,T.(1987),"A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Return,"*Review of Economics and Statistics* 69,542 –547.

Bollerslev, T., R. Y. Chou and K. F. Kroner, 1992, ARCH modeling in finance: A review of the theory and empirical evidence, *Journal of Econometrics* 52, 5-59.

Campbell, J. Y., S. J. Grossman, and J. Wang (1993): "Trading Volume and Serial Correlation in Stock Returns," *Quarterly Journal of Economics*, 108, 905–939.

Campbell John Y., Lo Andrew W. , MacKinlay A. Craig, The Econometrics of Financial Markets, 1997, *Princeton University Press*.

Chen, Gong-meng, Firth, Michael and Rui, Oliver M. 2001. “The Dynamic Relation Between Stock Returns, Trading Volume, and Volatility.” *Financial Review, Vol. 36, No. 3*, 153-174.

Christioffersen P.F., Diebold F.X., 2003 Financial Asset Returns, Direction of Change Forecasting and Volatility Dynamics, *Working Paper 04-10, Wharton School Of Business.*

Chordia, T., Swaminathan B., Trading Volume and Cross Autocorrelations in Stock Returns, *The Journal of Finance* Vol. LV, No:2,2000, pp. 913-935

Chowdurry, A. R., Mean Reverting Behavior of Stock returns: Evidence from a panel of Asian and pacific basin countries, *Journal of The Asia Pacific Economy*, *1999*, 4(3), pp. 431-445

Ciner, Ç., The dynamic relationship between returns and trading volume on the Toronto Stock Exchange, *Working paper, 2000, Northeastern University,* 2000

Clark, Peter K, 1973, A Subordinated Stochastic Process Model With Finite Variance for Speculative Prices, *Econometrica,* 41, 135-156.

Cochrane, John H., Time Series for Macroeconomics and Finance, *Manuscript, 2005*, Graduate School of Business, University of Chicago

Connoly R., Stivers C., Momentum and Reversals in Equity Index Returns during periods of abnormal turnover and return dispersion, *Journal of Finance, 2003, Vol: 58, Issue : 4,* pp 1521-1556

Copeland, T.E. "A model of Asset Trading under the assumption sequential information arrival. *Journal of Finance, 31 (Sept. 1976) pp 1149-1168*

Daniel, K., Hirshleifer D. Subrahmanyam A., 1998 Investor Psychology and Security Market under and overreactions, *Journal of Finance* 53, pp 1839-1888

Darrat Ali F., Rahman S., Zhong M., Intraday Trading Volume and Return Volatility of DJIA Stocks: A Note, *Journal of Banking and Finance* 27 (2003) pp 2035-2043

Engle,R.F.(1982),"Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation,"*Econometrica* 50 987 –1007.

Epps, T. W. and M. L. Epps, 1976, The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implication for The Mixture-of-Distributions Hypothesis, *Econometrica,* 44, 305-322.

Fama E.F., 1965. The Behaviour of Stock Market Prices, *Journal of Business* 38, pp 34-105

Fama E.F., Market Efficiency Long Term Returns and Behavioral Finance, *Journal of Financial Economics, 49 (1998) pp. 283-306*

Fan, X., Groenewold N., Wu Y., The stock return-volume relation and policy effects: The case of the Chinese energy sector, *The University of Western Australia, Business School, discussion paper, 2000.*

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition (2nd ed.).* New York: McGraw-Hill.

Garman,M.B.,and Klass,M.J.(1980),“On the Estimation of Price Volatility from Historical Data,"*Journal of Business* 53,67 –78.

Gervals, S., Kaniel R., Mingelgrin D., The high volume return premium, *working paper*, *1999, The Wharton School, University of Pennsylvania*.

Geweke and Porter-Hudak, The Estimation and Application of Long Memory Time Series Models, 1983, *Journal of Time Series Analysis*, pp 221-238.

Granger, C. W. J., and Joyeux, R. (1980), "An Introduction to Long-Memory Time Series Models and Fractional Differencing," *Journal of Time Series Analysis*, 1, 15-29.

Griffiths W.E, Hill R.C., Judge G.G.(1993), Learning and Practicing Econometrics, *John Wiley&Sons,*

Hamilton, J. D., 1994, Time Series Analysis, *Princeton University Press*

Harvey, A. C. (1981), Time Series Models, *Philip Allan Publishers Limited.*

Hosking, J. (1981), "Fractional Differencing," *Biometrika*, 68, pp 165-176.

Huddart, Steven J., Lang, Mark H. and Yetman, Michelle, "Psychological Factors, Stock Price Paths, and Trading Volume" (February 2006). *Working paper, available at SSRN*:

Jennings, R.H., Starks, L.T., Fellingham, J.C., An Equilibrium Model of Asset Trading with Sequential Information Arrival. *Journal of Finance,36 (March 1981), pp 143-161*

Jensen, M., Bayesian inference of long memory stochastic volatility models, *Journal of Time Series Analysis,Vol:25, 2004, pp 895-922*

Kahneman D and A. Tversky (1979) ”Prospect Theory: An analysis of Decision under Risk” *Econometrica*, 47, 263-291.

Kahneman, D., Fredrickson, D.L., Schreiber, C.A., & Redelmeier, D.A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, 4, 401-405.

Karpoff, J.M. 1987. The relation between price changes and trading volume. *Journal of Financial and Quantitative Analysis* 22: pp109-126.

Lamoureux, C.G. & Lastrapes, W.D. 1990. Heteroskedasticity in stock return data : volume versus GARCH effects. *Journal of Finance* 45: 221-229.

LeBaron B., 1992, Persistence of the Dow Jones index on rising volume, *working paper, University of Wisconsin.*

Lee, C.F. , Chen G., Rui, O., 2001, Stock Returns and Volatility on China's Stock Markets, *Journal of Financial Research, Vol: XXIV, No:4,pp 523-543*

Lee C.M.C., Swmainathan B., 2000 Price Momentum and Trading Volume, *The Journal of Finance Vol: LV No:5, October 2000. pp 2017-2069*

Lo, A.,W., Wang, J., Trading Volume : definitions, data analysis, and implications for portfolio theory, *Review of Financial studies, 2000, Vol: 13, No: 2 pp. 257-300*

Lo A. and MacKinlay A.C., 1998, A Non-Random Walk Down Wall Street, *Princeton*

University Press, Princeton, New Jersey

Lo, A., and MacKinlay, A.C., 1988, “ Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test, *Review of Financial Studies”,* 1, pp 41-66.

Mandelbrot, B. 1963 The variation of certain speculative prices. *Journal of Business 36: 394-419*

Masulis R. W., NG Victor K., Overnight and daytime stock return dynamics on the London Stock Exchange: The impacts of Big bang and the 1987 Market Crash, *Journal of Business and Economic Statistics,* October, 1995, Volume 13, No:14

Mougoue M., White A. Marie, Stock Returns and Volatility: An Empirical Investigation of the German and French Equity Markets, *Global Finance Journal, 1996, 7(2)*, pp 253-263

Parkinson,M.(1980),“The Extreme Value Method for Estimating the Variance of the Rate of Return,"*Journal of Business* 53,61 –65.

Richardson T., Peterson D.R., The Cross Autocorrelation Of Size Based Returns Is Not An Artifact Of Portfolio Autocorrelation, 1999, *The Journal of Financial Research, Vol XXII, No: 1, pp 1-13*

Saatcioglu, Kemal, and Laura T. Starks, 1998, The Stock Price-Volume Relationship in Emerging Stock Markets: The Case of Latin America, *International Journal of Forecasting*, 14, 215-225.

Salman F., Risk-return-volume relationship in an emerging stock market, *Research Paper (2000)*

Sentana, E., Wadhwani S., Feedback Traders and Stock Return Autocorrelations: Evidence from a Century of Daily Data *Economic Journal***,** Vol. 102, No. 411 (Mar., 1992) , pp. 415-425

Shefrin, Hersh M. and Meir Statman, 1985, "The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence," *Journal of Finance* 40, 777-792.

Silvapulle, P. and J. S. Choi, 1999, Testing for linear and nonlinear Granger causality in the stock price-volume relation: Korean evidence, *Quarterly Review of Economics and Finance* 39, 59-76.

Taylor S. J., Stock Index and Price Dynamics in the UK and the US: New Evidence from a trading rule and statistical analysis, *The European Journal of Finance, 2000, 6,* pp. 39-69

Tolvi, J., Long memory in a small stock market, *Economics Bulletin,2003, volume 7,pp 1-13*

Tsay R.S., Analysis of Financial Time Series,Wiley Series in Probability and Statistics, 2002, *John Wiley and Sons.*

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. Judgment under Uncertainty: Heuristics and Biases (pp.153–160).*New York: Cambridge University Press.*

Yanxiang Gu, A., Increasing Market Efficiency : Evidence from NASDAQ, *American Business Review, June 2004;22,2,* pp 20-25.

APPENDICES