PREDICTABILITY IN THE INDIAN STOCK MARKET: A STUDY FROM AN ECONOMETRIC PERSPECTIVE

DEBABRATA MUKHOPADHYAY

A Dissertation Submitted to the **Indian Statistical Institute** in Partial Fulfilment of the Requirement for the Award of the Degree of **Doctor of Philosophy**

Indian Statistical Institute

Kolkata, India

December 2007

ACKNOWLEDGEMENT

First of all, I express my sincere gratitude to my thesis supervisor Professor Nityananda Sarkar for his inspirational guidance. I am really indebted to him for his kind help throughout the preparation of this thesis.

I would like to gratefully acknowledge an anonymous examiner for very detailed and insightful comments on the original thesis, which have lead to much improvement in presentation in this revised version.

I also express my thankfulness to the authorities of the Indian Statistical Institute, Kolkata, for providing me with a research fellowship and the necessary infrastructure for doing this work. I am also thankful to the University Grants Commission (UGC) for allowing me to join Faculty Improvement Programme (FIP) under Xth plan and to the authorities of Shaympur Siddheswary Mahavidyalaya (College), Howrah, West Bengal, for granting me the necessary leave.

I would also like to thank Professor I. Choi for being kind enough in sending the program for the purpose of computing automatic variance ratio test.

I take this opportunity to acknowledge the help and suggestions I have received from Professor Dipankar Coondo, Professor Pradip Maity, Dr. Suman Bhaumik, Dr. Kaushik Bhattacharya, Dr. Samarjit Das and Mr. Soumyananda Dinda.

I also take this opportunity to express my love and respect to my parents and elder brothers who have always encouraged me for higher studies from the early stages of my education.

Last but not the least, I express my indebtedness to my wife Sanchita for her love, encouragement and sacrifice during the period of my research work.

Contents

CHAPTER 1

Introduction

The first chapter of this thesis presents the motivation of this study along with a brief review of the existing literature on empirical studies on stock market efficiency/predictability including those on India. The format of this chapter is as follows. A brief review of the literature is given in the first section. The motivation of this work is discussed in the next section. Section 1.3 presents a summary of the work done so far on efficiency/predictability in the Indian stock market. The focus and format of the thesis is described in the last section i.e., in Section 1.4.

1.1 A brief review of the empirical studies on stock market efficiency/predictability

The efficient market hypothesis has been one of the central propositions of finance since the publication of Eugene Fama's classical article in 1970. In fact, this hypothesis has remained one of the most controversial yet well-studied propositions in financial economics. The concept of efficient market hypothesis (EMH) which asserts that the asset price changes are unforecastable, can be traced at least as far back as the pioneering theoretical contribution of Bachelier (1900) and the empirical work of Cowels (1933). The modern literature on financial market efficiency began with Samuelson who, in his landmark article in 1965, attempted to show why properly- anticipated prices would fluctuate randomly. He argued that in an informationally efficient market – different from an allocationally efficient or Pareto - efficient market – price changes must be unforecastable if they are properly anticipated i.e., if they fully incorporate the expectations and information of all market participants. In fact, Samuelson (1965, 1972, 1973) demonstrated that an efficient market is one in which the information contained in past prices is instantly, fully and perpetually reflected in the asset's current price. Fama (1970) summarizes this idea by stating that " A market in which prices always 'fully reflect' available information is called 'efficient'" (p.383). More recently, Malkiel (1992) has broadened the definition of informational efficiency as follows: "A capital market is said to be efficient if it fully and correctly reflects all relevant information in determining security prices. Formally, the market is said to be efficient with respect to some information set….. if security prices would be unaffected by revealing that information to all participants. Moreover, efficiency with respect to an information set…. implies that it is impossible to make economic profits by trading on the basis of [that information set]".

Until the early 1980's, it was generally believed that the security markets of developed economies are efficient in the conventional sense. As stated in Malkiel (2003), the accepted view was that when information arises, the news spreads very quickly and it is incorporated into the prices of securities without delay. Thus, neither technical analysis which is the study of past stock prices in an attempt to predict future prices, nor even fundamental analysis which is the analysis of financial information, would enable an investor to achieve returns greater than those that could be obtained by holding a randomly selected portfolio of individual stocks, at least not with comparable risk.

About a quarter century back, much of the finance literature concerning efficient market hypothesis revolved around the random walk model and the martingale model, the two familiar statistical descriptions of unpredictable price changes that were earlier taken to be the implications of efficient markets. A stochastic process $\{P_t\}$ is said to follow a martingale if the following condition is satisfied:

$$E[P_{t+1} | P_t, P_{t-1}, \ldots\ldots\ldots] = P_t \tag{1.1}$$

or, equivalently, $E[P_{t+1} - P_t \mid P_t, P_{t-1}, \ldots\ldots] = 0$ (1.2)

where P_t represents , in this context, a stock price/index at time point t. On the other hand, a random walk model characterizes a stock price/index series where all subsequent price changes represent random departures from previous prices. It is well-known that the simplest version of the random walk model is the independently and identically distributed (i.i.d.) increment in which the dynamics of $\{P_t\}$ are given by the equation

$$P_t = \mu + P_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma^2) \tag{1.3}$$

where μ is the expected price change or drift and ε_t is independently and identically distributed with mean 0 and variance σ^2.

The early tests of random walk hypothesis were the tests of sequences and reversals of the returns on stock prices/indices[1] by Cowels and Jones (1937). While such tests are now of little interest, one of the commonly used tests for random walk hypothesis is due to Lo and MacKinlay (1988). This test uses the property of random walk hypothesis that random walk increments must be a linear function of the time interval. Since the random walk model implies that all autocorrelations of returns are zero, another test statistic which has power against many alternative hypotheses is the Q-statistic due to Ljung and Box (1978). One of the important characteristics of all these tests is that these are carried out in the framework of linear models assuming implicitly linear dependence structure in

---

[1] Return, r_t, is defined as $r_t = \ln P_t - \ln P_{t-1}$. This is indeed the continuously compounded rate of return for holding a financial asset for one unit of time.

the returns on stock prices/indices. But, since the last two- and-a-half decades nonlinear dynamics are being considered while modelling r_t, and appropriate tests to this effect i.e., whether there are indeed nonlinear dependencies, are now available. Further, it is well-established that conditional variance of returns on financial assets like stock indices is not constant over time. This fact has led to the development of autoregressive conditional heteroscedastic (ARCH) and generalised ARCH (GARCH) models (Engle(1982), Bollerslev(1986)). Since returns based on equity prices/indices are most often found to have time dependent conditional variance, ARCH/GARCH model used to take care of the volatility observed in the time series of returns. Obviously, lack of linear dependencies does not rule out nonlinear dependencies, and the presence of the latter contradicts the efficient market hypothesis. There are, in fact, a large number of strong evidence for nonlinear dependencies in stock returns in the context of developed capital markets (Granger and Andersen (1978), Sakai and Tokumaru (1980), Hinich and Patterson (1985) , Scheinkman and LeBaron (1989), Hsieh (1991), Willey (1992), Pandey *et al.* (1997), Opong *et al*. (1999), to cite a few) . Granger and Ansersen (1978) developed bilinear time series model for capturing nonlinearity in asset returns. Sakai and Tokumaru (1980) introduced the idea of chaos to understand the behaviour of asset returns. Hinich and Patterson (1985) were among the firsts in reporting evidence of nonlinearity in daily stock returns of the New York Stock Exchange. In their well-known paper, Scheinkman and LeBaron (1989) found evidence for nonlinearities on the US stock returns, but not for low dimensional chaotic dynamics. Further, Hsieh (1991) cast some doubt as to whether the returns on Standard & Poor's 500 index follow a random walk model. Pandey *et al.* (1997) have reported evidence of nonlinearity in the stock returns of Hong-Kong, Japan and the US. The major study on another developed stock market i.e., the UK stock market, was recently carried out by Opong *et al*. (1999). They found rejection of i.i.d. property of the returns on London Financial Times Stock Exchange (FTSE) indices by using recent advances in chaos. In the context of the other stock markets, particularly those of emerging market economies which obviously include India, only a few such studies are available. Prominent among these – excluding those on India – are due to Sewell *et al*. (1993), Antoniou *et al*. (1997), Ammerman and Patterson (2003), Lim and Hinich (2005). Sewell *et al.* (1993) is the earliest study providing evidence of nonlinearity for emerging stock markets. Antoniou *et al*. (1997) and Ammerman and Patterson (2003) have conducted studies on nonliearity in the stock markets of Turkey and Taiwan, respectively. Lim and Hinich (2005) have found nonlinear behaviour of episodic nature by applying the windowed testing procedure to returns on fourteen Asian stock markets. In fine, all these studies have cast doubt on the conclusion regarding stock market efficiency based only on lack of serial correlation in returns.

1.2 Motivation

Given these empirical evidence in favour of nonlinear dependencies in stock returns, and the recent advances in time series econometric methods, the intellectual dominance of the efficient market hypothesis has become far less universal since the end of last century. Accordingly, many financial economists and econometricians have started emphasizing that stock returns are predictable to some degree. From a different perspective, currently a new school of academic finance, called behavioural finance, is also arguing in favour of predictability of stock returns by suggesting that psychological and behavioural elements have important roles in stock-price determination. Based on past stock price patterns as well as certain "fundamental" valuation metrics, such studies have shown that future stock prices are somewhat predictable (see Malkiel (2003) and Shiller (2003)). In addition, many of these researchers are claiming that presence of predictable patterns will enable investors to earn above average risk adjusted returns. Critics of the efficient market hypothesis also argue that there are several instances in the stock markets of the developed economies like, for example, the stock market crash of October 1987 and the internet bubble of the late 1990's in the U.S.A., which cannot be explained by the behaviour of rational investors, and that psychological consideration must, therefore, be taken into account for explaining such instances. In this context, the concept of rational bubbles may be mentioned so that its role in predictability of returns is clarified. By rational bubbles we mean a situation where asset prices rise far higher than could be explained by fundamental values of the assets. Here the investors appear to anticipate that other investors would drive prices even higher in future. Now, it is perfectly rational for investors to be willing to pay prices that are increasingly divergent from fundamental values as they are being compensated for this by ever-increasing returns. According to this notion, even though investors recognize that the asset is overvalued, they believe that they will be able to sell it at an even higher price at a later date. Thus, the formation and rapid collapse of speculative bubbles does not require investor irrationality. However, there are researchers in the other spectrum, who believe that presence of predictable elements does not necessarily imply rejection of the efficient market hypothesis. They argue that presence of significant transactions cost in the trading process may account for this predictable component. Further, time-varying expected returns due to changing business conditions can also generate predictability. A certain degree of predictability may also be attributable to time-varying stock market risk premia. They have also argued that pricing irregularities and even predictable patterns are temporary in nature, and these will not be able to refute the hypothesis that stock market is efficient (see Campbell *et al*. (1997) for an excellent discussion on this approach). In this context, the conclusion by Malkiel (2003, p. 80) is noteworthy:

" But I suspect that the end result will not be an abandonment of the belief of many in the profession that the stock market is remarkably efficient in its utilization of information. Periods such as 1999 where 'bubbles' seem to have existed, at least in certain sectors of the market, are fortunately the exception rather than the rule. Moreover, whatever patterns or irrationalities in the pricing of individual stocks that have been discovered in a search of historical experience are unlikely to persist and will not provide investors with a method to obtain extraordinary returns."

Despite the existence of such differing views regarding the implications of predictability of stock returns on market efficiency, it can be concluded, based on empirical evidence, that stock returns are most often predictable. The fact that a vast number of researchers have carried out empirical studies to predict financial returns on equity prices/indices shows the fascination and challenges involved in this exercise. It is also relevant, in this context, to note that it is often argued that if stock markets are efficient then it should not be possible to predict stock returns. But this line of argument is not satisfactory. The concept of market efficiency needs to be defined separately from predictability. In fact, it is easily seen that stock market returns are non-predictable only if market efficiency is combined with risk neutrality (*cf*. Pesaran (2003)).

An important point to be noted at this stage is that in most of the studies on predictability/efficiency carried out so far, the underlying models have been assumed to have correctly specified conditional mean. It is now well-recognized that inferences based on models suffering from misspecification could be misleading and incorrect. In fact, as pointed out by Gourieroux and Jasiak (2001, p. ix), "… a statistical model is a simplified image of reality, which is much too complex to be described exactly. Therefore, an econometrician is aware of the fact that a model is necessarily misspecified." Hence they have advocated that empirical findings need to be interpreted with caution. To cite an instance, it may be mentioned that in the context of studies on predictability/efficiency in the framework of ARCH/GARCH, Lumsdaine and Ng (1999) (see also, Weiss (1986), Tong (1990), Giles *et al*.(1993)) have shown that, in general, the popular Lagrange Multiplier (Rao's score) test for testing the null of conditional homoscedasticity leads to over-rejection of the null hypothesis if there is misspecification of conditional mean. It thus becomes important that in order to test for conditional heteroscedasticity in the general context of a possibly misspecified conditional mean, a test for mispecification of conditional mean is required to be carried out first and then appropriate steps need to be taken for guarding against misspecification in the mean function in case the test rejects the null hypothesis of no misspecification of conditional mean. Once the conditional mean is found to be appropriately specified, then the test for conditional heteroscedasticity would yield correct inference. As stated by Lumsdaine and Ng, the misspecification problem referred to here can arise if the functional form and/or conditioning set is misspecified. For linear dynamic models, notable cases of such misspecifications are omitted shifts in the trend function, selecting a lag length in an autoregression that is lower than the true order, failure to account for parameter instability, residual autocorrelation and omitted variables. They have proposed a method based on recursive residuals for adjusting for possible misspecification of the conditional mean function. In this context it is also relevant to note that incorrectly specified conditional mean might as well lead to mispecification of conditional variance. In fact, GARCH model would be correctly specified if only there is no serial correlation. As a way out of this problem in the context of studying serial correlation, Robinson (1991) and Wooldridge (1991a, b) have suggested ways of robustifying tests for serial correlation to allow for possible misspecification of conditional variance.

It may be worthwhile to point out that seasonal patterns, in the returns series, may also affect the specification of the first two conditional moments. This phenomenon may be defined as the tendency of financial asset returns to display systematic patterns at certain times of the day, week, month and year. This observed phenomenon explaining deviation from random walk model is what is popularly known as calendar anomalies/ “seasonal” effects. These anomalies in security returns have been extensively documented. Among the different “seasonal” effects observed in stock returns, the most important one is the different effects across the different days of a week. It was first noted by Fields way back in 1931. Fields observed that the US stock market consistently experienced significant negative and positive returns on Mondays and Fridays, respectively. His observations started receiving increasing attention only after about half a century. Many authors like French (1980), Gibbons and Hess(1981), Lakonishok and Levi (1982), Keim and Stambaugh (1984), Rogalski (1984), Jaffe and Westerfield (1985), Engle *et al.* (1987), Fama and French (1988), Agarwal and Tandon (1994) and

Peiró (1994) have found that returns differ by small yet statiscally significant amounts over different days, and sometimes over different weeks as well as over months. These effects, called the day-of-the week, week-of-the –month and month-of-the year effects, respectively, if present in returns, indicate that stock returns have a predictable pattern in their movements.

1.3 Studies on efficiency/predictability of Indian stock returns

Most of the studies on predictability/efficiency of stock markets are based on developed stock markets, especially those of the US and the UK stock markets. Not much is known about the predictable patterns of the stock markets of emerging market economies which had the fastest growing share markets in the last decade. In particular, if we consider the case of India which is one of the most important emerging market economies, we find that despite having a long history and voluminous turnover, serious, systematic and methodologically-sound studies on predictability of Indian stock returns, are very few. In fact, as commented by Poshakwale (2002), relatively much less is known about the characteristics and dynamics of Indian stock returns.

The first study on efficiency/predictability in the Indian stock market is due to Poshakwale (1996). Based on runs test and tests for serial correlation, he found evidence of violation of weak-form efficiency in Bombay Stock Exchange over the period 1987-1994. In a subsequent study, Gupta and Gupta (1997) re-examined the random walk model in the Indian stock market using data for the period July 1988 to January 1996, and came to the conclusion that the findings were not supportive of the random walk hypothesis. Bhaumik (1997), however, found evidence of market efficiency, although in a very limited framework of analysis. It may be noted that all these studies have mainly used the traditional tests in their efficiency studies and to that extent the scopes of these findings are rather limited. Two other studies on market efficiency in the Indian stock market are due to Basu and Morey (1998) and Kawakatsu and Morey (1999). Although their common objective was to find the effect of economic liberalisation on the efficiency of Indian stock market, their analyses are relevant from the standpoint of market efficiency also. Applying variance ratio test on the monthly all-India share price index data spanning the period July 1987 to October 1996, Basu and Morey found that the aggregate equity prices show signs of being efficient since the mid-1980's. Kawakatsu and Morey, on the other hand, found little evidence that liberalisation has changed the behaviour of Indian stock indices. To the best of our knowledge, the most recent work on efficiency/predictability has been done by Poshakwale in 2002, by applying the BDS test (Brock *et al* . (1996)) for nonlinearity. Poshakwale (2002) has examined the random walk hypothesis by testing nonlinear dependence using both individual stock prices and equally weighted portfolio of 100 stocks for the period January 1, 1990 to November 30, 1998. The major finding of this study is that daily returns from the Indian stock markets do not follow random walk model.

There are also a few studies, although limited in their scopes, focussing on the presence of "seasonal" effects on the Indian stock returns. These studies by Chan *et al*. (1996), Wood and Poshakwale (1997), Choudhry (2000), Bhole and Pattanaik (2002) and Bhattacharya *et al.* (2003) have found evidence of the day-of-the week effect in the returns on Indian stock indices.

1.4 Focus and format of the thesis

We now discuss about the focus of this thesis along with the broad aspects of its coverage.

The thesis primarily aims at carrying out a systematic and comprehensive study on predictability in the Indian stock market. The primary reason for choosing India for such a study is that India is now considered – all over the world – as one of the most important emerging market economies (EME). As coined in 1981 by Antoine W. Van Agtamael, an EME is defined as an economy with low-to-middle per capita income. Such countries constitute 80% of the global population and represent about 20% of the world's economy. Although a loose definition, countries varying from very big (say, China) to very small (say, Tunisia) are included in this category because of their developments and reforms. It is worth noting that amongst the emerging market economies, India is one of the most important one because of its size of population (1114 millions as projected for 2006 in the 2001 census) which is next to China, and its continuing strong and responsible economic performance levels which began sometime after 1992 when India embarked on economic development and reform programmes in right earnest after discarding the decades-old practice of what used to be called 'mixed economy' approach, a peculiar mix of socialistic and capitalistic economic policies and programmes with very strong governmental control at different levels. Given the importance of India in the EMEs, it is natural and meaningful to choose India for the purpose of this study.

As regards our choice of stock market indices for this study, it may be noted that although there are several important characteristics of EMEs, the most important ones are related to reforms in capital market as well as in exchange rate system, which induce increase in both local and foreign investments leading eventually to increase in gross domestic product. Since EMEs are in transition and hence not stable, emerging markets offer an opportunity for investors who are ready to take more risk to their portfolios. Thus, studying the predictability of on stock markets of emerging markets is important in empirical finance, and more so for major EMEs like India.

This thesis on predictability in Indian stock market is essentially an empirical study covering all relevant aspects, as stated below. In order to make the empirical findings not stock index specific since these indices vary both in terms of numbers and composition of equities considered in the aggregation, and in that sense more general and robust, we have studied predictability in the Indian stock market based on all its major stock indices. Further, the study has been done with data at both daily and monthly level frequencies. We now discuss briefly the broad aspects of this study.

(i) *Appropriate specification of both the conditional mean and the conditional variance*

Since there hasn't been any comprehensive study with due consideration to relevant econometric issues, as already mentioned in the preceding sections, on predictability of Indian stock market we have confined, all throughout this work, to the set-up of linear dynamic models with nonlinearity being considered only through the specification of conditional heteroscedasticity. The first issue of importance in this study is regarding appropriate specification of both the conditional mean and the conditional variance. The importance of this issue has already been discussed in Section 1.2. To that end, we first consider the issue of appropriate specification of conditional mean which, apart from other independent variables, requires inclusion of " seasonal effects". Insofar as specification of conditional heteroscedasticity is concerned, it is only appropriate that seasonal effects be considered for the conditional variance as well. It is expected that if returns exhibit seasonal patterns, then volatility may also be affected by similar effects. There is, however, hardly any study in the context of stock market returns where seasonal effects have been explicitly incorporated in the specification of volatility. Since we are emphasizing on appropriate specification of both the conditional mean and heteroscedasticity, we have included seasonal anomalies in the specification of the moments. In the case of analysis with daily level data on Indian stock returns, the limitation is that excepting only a few variables, no other time series on macrovariables or financial variables are available at the daily level frequency. This is obviously not the case with monthly-level data; data on all relevant macro and financial variables of India are available at this frequency.

(ii) Role of macroeconomic and financial variables on predictability of stock returns

There is a growing literature-mostly concerning industrialized countries- showing strong influence of macroeconomic variables on stock markets (see, for instance, Fifield *et al. (*2000), Lovatt and Parikh (2000), Nasseh and Strauss (2000) Hondroyiannis and Papapetrou (2001), Lu *et al.*(2001)). Financial theory asserts that movement in stock prices is related to macroeconomic variables. As macroeconomic variables contain important information for investors it is hypothesized that the stock market participants take these factors into account for estimating appropriate discount rate and the expected flow of dividends from stocks. The empirical evidence, however, are not uniform in providing support to stock returns predictability using macroeconomic variables in the sense that while some studies have found that certain macroeconomic variables have significant effects in explaining returns of some stock indices, others have found no such evidence i.e., significant effects of the same variables for some other stock returns (see, for instance, Balvers *et al.* (1990) and Flannery and Protopapadakis (2002) for details of such findings). Further, it has been observed that predictive ability of some macrovariables with respect to equity returns is quite uneven over time. Durham (2001), for example, has found this for some variables concerning monetary policy.

Researchers have also identified a number of financial variables that appear to predict future stock returns. These include price-earnings ratio, price-book value ratio, dividend yield, short term interest rate etc. (see Campbell and Shiller (1988a, b, 1998), Fama and French (1988) and Ang and Bekaert (2001) to cite a few). Keeping this in mind, we study the predictability of Indian stock returns using standard macroeconomic and financial variables relevant for India by applying both in-sample and out-of-sample tests of return predictability. While the in-sample analysis carried out here employs a predictive regression framework, the out-of-sample forecasts have been analysed using recently developed statistical tests by Clark and McCracken (2001) and McCracken (2004) which are some variants of those proposed by Diebold and Mariano (1995) and Harvey *et al.* (1998), respectively (see Rapach *et al*. (2005), for an application of these tests). Since in-sample forecasts are those generated for the same set of data that is used to estimate the model's parameters, it is expected that such forecasts of a model would be relatively good. Hence, a sensible approach is to use out-of-sample forecasts for model evaluation, and this is what has been done in this thesis.

(iii) Long-run (cointegrated) relationship involving stock index and macro as well as financial variables

The issues so far discussed concern only short-run consideration towards studying the particular issue of efficiency *viz.,* predictability in the Indian stock market since such studies are of paramount importance from the investors' point of view. However, in developed stock markets there have been considerable research efforts towards investigating the issue of market efficiency for several asset prices in the context of long run (see, for instance, Baillie and Bollerslev (1989) and Coleman (1990)). But the notion that cointegration implies market inefficiency has been challenged by some researchers in the recent years. They have argued that the traditional concept of market efficiency in which changes in asset prices are unpredictable lacks necessary economic content. Their view is that market efficiency can be defined more usefully as one of lack of arbitrage opportunity (see Levich (1985) and Ross (1987) in this context). However, there is a consensus in the literature that the existence of a long-run (cointegrating) relationship immediately invalidates the martingale property, one of the two standard models for market efficiency for asset prices. As Caporale and Pittis (1998) have argued that whatever concerns one might have about the identification of a cointegrating relationship with market inefficiency, cointegration tests can still be usefully employed to investigate the predictability of asset returns. Since no such study has been done on India, we have applied the methodology of cointegration involving the Indian stock indices and relevant macro and financial variables in studying the aspect of predictability of Indian stock market in the long run sense.

(iv) A*lternative volatility models and distributional assumptions*

Accurate measures and forecasts of volatility are crucial for the implementation and evaluation of asset and derivative pricing models (*cf.* McMillion and Speight (2004)). Also, proper estimation of volatility is necessary for portfolio analysis and risk management such as value at risk (*cf.* Taylor (2004)). Now, insofar as specification of volatility is concerned, GARCH is the most frequently used model since, as noted by Bollerslev *et al*. (1994), it has turned out to be very useful for describing a wide variety of financial data. Keeping in mind the general usefulness of the basic GARCH model we have carried out our study by taking GARCH as the model for volatility. However, a major limitation of the basic GARCH process is that it is symmetric in that negative and positive shocks have the same effects on volatility. But the empirical literature on returns of risky assets, following Black (1976), has clearly pointed out that future volatility is more affected by negative shocks compared to positive shocks. This phenomenon of asymmetry is known as the so-called 'leverage effect' in the finance literature. There are two standard extensions of the original GARCH model, which capture the above inadequacy of the basic GARCH model. These are known as the threshold GARCH (TGARCH), as proposed by Glosten, Jagannathan and Runkle (1993) and which is also known as the GJR model, and the exponential GARCH or EGARCH model, as introduced by Nelson (1991). We have considered these two models for volatility in our study with the objective of finding if these specifications are more suited (as compared to GARCH) to describe volatility in the Indian stock market. Based on standard forecast evaluation criteria, namely, mean absolute error (MAE) and root mean square error (RMSE), we have compared these two volatility models on the one hand and the standard

GARCH model on the other, to conclude which kind of volatility specification performs better from consideration of predictability of Indian stock returns.

The last issue considered in this thesis relates to the distributional assumption. Linear dynamic models with or without volatility generally assume conditional normality as the distribution of the innovations but this is rarely supported by real data as evidenced by excess kurtosis of return data for almost all stock markets studied in the literature. Because of this nature of fat-tailed conditional distribution of returns, what is most often done is to work with the conditional normal likelihood function.[2] and the standard error is then calculated by using a robust covariance matrix estimator, as suggested by Bollerslev and Wooldridge (1992). Otherwise, alternative conditional distributional assumptions for capturing the excess kurtosis are made. For instance, Bollerslev (1987) advocated evaluating sample log-likelihood under the assumption that innovations follow the standardized t-distribution. In a similar spirit, Nelson (1991) has used the standardized general error distribution (GED) and studied the performance of the model under this alternative assumption of the innovation distribution. We have taken up this issue of alternative distributional assumptions, although in a very limited way, for studying the returns on Indian stock indices. In other words, we have considered these two alternative distributions which allow for additional kurtosis, and then studied if these alternative distributions explain the predictability of Indian stock returns better as compared to the assumption of conditional normality.

The other chapters of this thesis are organized as follows:

CHAPTER 2: Indian Stock Market: Some Relevant Details

This chapter presents some recent facts on the performance of the Indian economy so that India's current status as one of the most important emerging market economies is established. A few major structural and regulatory reforms which were carried out in the Indian stock market during the last one and-a-half decades since its reform process started in the early nineties of the last century, are also stated in this chapter along with some descriptions of the data sets on Indian stock indices and other variables which have been used in this study. This chapter has been organized as follows. The chapter begins with an introduction. The next section discusses the case for India as an emerging market economy. Section 2.3 presents some details on the major reforms in the financial sector of India. This chapter closes with discussions on the data sets.

CHAPTER 3: Testing Predictability of Daily Stock Returns in the Framework of Appropriate Specification

This chapter of the thesis deals with the issue of studying predictability and nonlinear dependence (the latter through GARCH specification for conditional heteroscedasticity) in Indian stock returns at daily level frequency. The framework of this analysis also involves appropriate specification of the conditional first and second-order moments so that the final inferences are free from any possible consequences of misspecification. The format of this chapter is as follows. Section 3.1 gives the introduction of this chapter. The modelling approach is presented in the next section. Empirical findings on testing for parameter stability as well as for misspecification, detecting nonlinear dependencies and testing for the presence of dynamics in higher-order moments are discussed in Section 3.3. Finally some concluding remarks are made in Section 3.4.

---

[2] The estimator is then interpreted as a quasi-maximum likelihood estimator (White (1982)).

CHAPTER 4: Studying Monthly Stock Returns with Macro and Financial Variables: A Predictive Regression Approach

This chapter studies the predictive ability of macroeconomic variables and financial ratios for Indian stock returns at the level of monthly data using both in-sample and out-of-sample forecasting techniques and having due consideration for appropriate specification of the underlying model for return. The organization of this chapter is as follows. This chapter begins with an introduction in Section 4.1. The methodological approach which includes in-sample and out-of-sample tests of predictability and the bootstrap procedure are presented in Section 4.2. Empirical findings are described in the next section. The chapter closes with some remarks in Section 4.4.

CHAPTER 5: Predictability in the Indian Stock Market: A Cointegration Approach

This chapter presents an empirical study on predictability – in the long-run sense in the Indian stock market. Results, based on cointegration analysis of such a study involving a stock index and the relevant macro and financial variables – all at their $I(1)$ level values- are reported and discussed in Chapter 5. The format of this chapter is described below. The first section presents the background of the work contained in this chapter. The cointegration methodology is then discussed in Section 5.2. Empirical findings are presented in the next section. This chapter is concluded with some remarks in the last section.

CHAPTER 6: Predictability of Daily Stock Returns under Alternative Volatility and Distributional Assumptions

This chapter studies the predictability of Indian stock returns under alternative volatility specifications as well as conditional distributional assumptions. Two alternative models of volatility *viz.,* EGARCH and TGARCH, and two alternative conditional distributions for the innovations – standardized Student's $t-$distribution and standardized GED are considered for this study, and comparisons across these models are done using suitable criteria. This chapter is organized as follows. Beginning with a section on introduction in Section 6.1, this chapter presents the modelling approach under alternative volatility and distributional assumptions in the next section. Section 6.3 discusses empirical findings followed by some concluding remarks in Section 6.4.

Chapter 7: Summary and Future Ideas

The last chapter of this thesis begins with an introduction on the research problem. In the next section i.e., in Section 7.2, a summary of the major findings of the entire work are presented along with the limitations of this study. The concluding section contains a few ideas for further work on this problem.

CHAPTER 2

Indian Stock Market: Some Relevant Details

2.1 Introduction

Since this thesis is concerned with studying predictability in the Indian stock market, it is necessary as well as and meaningful to present some relevant details about the Indian stock market and its important stock indices. To this end, we first state some recent facts on the performance of the Indian economy so that India's current status as one of the most important emerging market economies with huge growth potential becomes quite obvious. Since we are concerned with the behaviour of stock market, we then cite a few major structural and regulatory reforms which were carried out in the Indian stock market during the last one and-a-half decades since its reform process started in the early nineties of the last century. Thereafter we present some relevant details including the sources of the data sets on Indian stock indices and other variables which have been used in this study. This chapter has thus been organized as follows. The next section discusses the case for India as an emerging market economy. In Section 2.3, some of the major changes in the financial sector of the Indian economy are briefly described. Finally, some details about the data used in this study are presented and discussed in Section 2.4.

2.2 India as an emerging market economy

The performance of India's economy over the last decade has been quite impressive. After a major economic crisis in 1991, the Government of India initiated bold reform measures and consequently the economy started experiencing a rapid economic growth rate and inflow of increasing foreign investment. Added to these has been the boom in the information technology sector where India is now rated to be virtually the leader. In 2004, India was the fourth largest economy in the world after the USA, China and Japan and also the second largest among emerging market economies in terms of gross domestic product (GDP) based on purchasing power parity. India is now the second fastest-growing major economy in the world with a GDP growth rate of 8.1 per cent at the end of the first quarter of 2005-06. As the figures in Economic Survey 2004-05, show, the Indian economy registered a growth rate of 8.5 percent during 2003-04, the highest ever except for the years 1975-76 (9.1 per cent) and 1988-89 (10.1 per cent). Moreover, during 2001- 04, Indian economy has grown at a rate over 6 per cent. During the first post-reform decade which is considered to be the period covering the years 1992 to 2001, growth rate has been spectacular – being high as 6.5 per cent (Basu (2004)). It can be further noted that during the pre-reform period since late 1970s India's annual growth rate was around 5 per cent. The study by Ramaswamy (2001) has shown that growth in the post-reform period i.e., after 1992, has been significantly higher than in the pre-reform period.

Apart from India's success in terms of high growth rates, an important factor necessary for the sustenance of economic growth is high gross domestic savings and consequently high gross domestic capital formation. In 2001-02, these figures (at current prices) as percentages of GDP were 23.4 and 22.6, respectively. India is also in a strong position as far as foreign exchange reserve is concerned. India's foreign exchange reserve (including gold, SDRs and reserve position in the IMF) is estimated at a level of US$ 128.91 billion on February 4, 2005, which is in excess of India's total external debt of US$ 114 billion at the end of September, 2004. As far as the performance of the Indian economy in terms of important social indicators related to standard of living is concerned, it has a literacy rate of 65 per cent, as per official figures in 2001, and life expectancy at birth, as in 2003, is 63. Thus, although India continues to maintain strong levels of economic performances, it has a long way to go to achieve significant improvements in terms of such social indicators.

India officially liberalized its stock market on November 11, 1992, when it first allowed foreign investors to invest in its stock market (see Bekaert and Harvey (2000) and Kim and Singal (2000)). Since the opening up of the Indian equity markets to foreigners, foreign institutional investment (FII) flows have grown substantially. As the recent issues of Reserve Bank of India (RBI)[1] Monthly Bulletin show, FII has increased from about US$ 1665 million in 1993-94 to US$ 8280 million in 2004-05. Its share in total portfolio flows to India has grown from 47 per cent in 1993-94 to over 93 per cent in 2004-05. During the same period, foreign direct investment has increased from US$ 586 million to US$ 5,535 million. In 2004, 5.8 per cent of gross turnover of National Stock Exchange(NSE) and Bombay Stock Exchange(BSE) was made up by FIIs. Although there are some debate over inherent vulnerabilities with FII flows and their destabilizing effects on equity and foreign exchange markets, it cannot be ignored that India is increasingly becoming an attractive destination to the global investors. This brief presentation of statistical data on India's economic performance thus establishes that India is now one of the fastest growing economies in the world with huge growth potential.

[1] The Reserve Bank of India is the India's Federal or Central Bank.

2.3 Reforms in the financial sector

Since India embarked on a series of structural and regulatory reforms in its economy since the early 1990s to free itself from an extremely fragile financial condition arising out of political instability, sluggish growth and foreign exchange crisis, major policy changes and reforms programmes were initiated in most of the sectors of the economy including, of course, the financial sector. Consequently, some fundamental changes have taken place in the Indian economy as a whole, and in particular, in the financial sector. In the context of Indian capital market which is the focus in this study, a major decision was to form the Security and Exchange Board of India (SEBI) as the regulatory authority of the Indian capital market in the year 1988. Other reform measures initiated in the stock market included the "birth" of a new stock exchange, called the National Stock Exchange (NSE), in 1993 as a competitor to the oldest stock exchange of India *viz.,* the Bombay Stock Exchange(BSE), introduction of computerized screen based trading at both these exchanges and dematerialization of shares. Another major development concerning the secondary segment of the Indian capital market was the introduction of derivative trading in June 2000. SEBI approved derivatives trading based on future contracts at both the BSE and the NSE in accordance with the rules/byelaws and regulations of the stock exchanges.

India has 23 stock exchanges across the country, of which the major ones are the Bombay Stock Exchange at Mumbai (earlier known as Bombay) and the National Stock Exchange at Mumbai. The Securities and Exchange Board of India regulates all the stock exchanges of the country. The Bombay Stock Exchange (BSE) which was established in 1875, is the premier stock exchange of India. As on December 2003, there are 5644 companies listed on the BSE. Insofar as the total market capitalisation is concerned, the figure for BSE stands at Rs.[2] 16984.28 billion on March 2005 (US$ 388.74 billion). Further, during the period April 2004 – March 2005, the total turnover in BSE was Rs 5187.16 billion (US$ 118.73 billion) with an average daily turnover of Rs. 27.06 billion (US$ 0.62 billion). Some more recent statistics to show the spectacular effects of liberalization of Indian economy on the capital market can also be cited. For example, by March 2005, there were 685 registered Foreign Institutional Investors (FIIs). The total purchase by the FIIs in the BSE secondary market during April 2004 to March 2005 was Rs. 599 billion (US$ 13.71 billion). The growth in both investment and number of listed companies in various stock exchanges in India, and also changes in trading rules and in the volume of capital raised from private investors have helped in achieving globalisation and international competitiveness of Indian capital market and sparking a boom in its stock prices.

Although the newest stock exchange of India, the National Stock Exchange has, by now, become the largest stock exchange of the country. It now represents about 45 per cent of the total market capitalisation. During the year 2003-04, NSE reported a turnover of Rs. 1,099,535 crores in the equities segment. After its recognition in April 1993 as a stock exchange under the Securities Contracts (Regulation) Act, 1956, NSE commenced its operations in the wholesale debt market (WDM) segment in June 1994. The Capital Market (Equities) segment started its operations from November 1994 and those in derivatives segment from June 2000. Both the BSE and the NSE provide fully automated, screen-based trading systems. The trading system is order driven, and based

[2] Rupees (in abbreviated form, Rs.) is the Indian currency.

on price-time priority. Both the BSE and the NSE provide continuous trading. Trading cycle is on a daily rolling with settlement $T+2$ basis in the demat form. Both the exchanges provide for settlement through the Clearing House/ Clearing Corporation and have trade/settlement guarantee and investor protection funds.

It may be worth noting that in terms of international ranking of security exchanges, as measured by the number of transactions, the NSE and the BSE have achieved ranks 3 and 5 respectively, in both 2003 and 2004. In fact, it was in 2002 that the NSE replaced the Sanghai stock market to take the third place after NASDAQ and NYSE.

2.4 Some details on the data sets

Any stock market index should capture the behaviour of overall equity market, and also reflect the changing expectations of the stock market about future dividends of the country's corporate sector. Measurements of such an index should also represent the returns obtained by "typical" portfolios in the country concerned. Keeping in mind that any single index may not be very representative from all these considerations, and also that stock indices vary in terms of numbers as well as composition of equities considered in the aggregation, we have, in our thesis, taken a number of standard indices representing Indian stock market. Such a choice of several indices would also be useful in assessing the robustness of our findings on predictability of the Indian stock market.

There are several stock indices, at daily level frequency, of the Bombay Stock Exchange (BSE*) viz.,* BSE sensitive index (BSESENSEX), BSE-100, BSE-200, BSE-500 indices and also some sectoral indices like BSE BANKEX, BSE IT and BSE OIL GAS

INDICES. But the most widely reported index in the national and international media is the BSESENSEX. This index is not only scientifically designed but also based on globally accepted construction and review methodology. First compiled in 1986, BSESENSEX is a basket of 30 constituent stocks representing a sample of large, liquid and representative companies including some major companies in the information technology sector like Infosys Technologies Ltd, Satyam Computer Services Ltd. and Tata Consultancy Services Ltd. Many newly important sectors like finance, pharmaceutical, health-care are also well represented in this index. The BSESENSEX was initially a full-market-capitalisation weighted index. But since September 1, 2003, it follows the free-float market capitalisation methodology of index construction which is now considered to be the widely followed index construction methodology on which majority of global equity benchmarks such as MSCI, FTSE, STOXX, S & P 500, and Dow Jones are based.

BSESENSEX is regarded as the pulse of the Indian stock market. As the oldest index in the country it provides the time series data over a fairly long period of time. With base at 1978-79 = 100, this index has been serving the purpose of quantifying the equity price movements as also reflecting the sensitivity of the Indian capital market in an effective manner. The growth of equity market in India has been phenomenal in the decade gone by. Right from the early nineties the stock market witnessed heightened activities in terms of various bull and bear runs. The BSESENSEX has captured all these events in the most judicial manner.

The launch of BSESENSEX in 1986 was later followed up by introduction of BSE National Index comprising 100 stocks (with base at 1983-84 =100) in January 1989. This index which consists of 100 stocks is thus a broad-based index reflecting the stock price movements on national scale at daily frequency. This index was renamed as BSE-100 index from October 1996. Since 2004, this index also follows the free-float methodology of index construction like the BSESENSEX. The third daily-level index considered in this study is known as DOLLEX which is the dollar conversion of BSE-200. With a view to provide a better representation of the increased number of companies listed, increased market capitalisation and the new industry groups, the Bombay Stock Exchange launched BSE-200 index which is based on equity prices of 200 companies and its dollar equivalence, called the DOLLEX, in May 1994. In the present context of increased willingness shown by foreign investors to participate in trading activities of Indian stock market, and also because of growth in the number of foreign financial institutions in the country, the analysis of such an index i.e., DOLLEX should be very useful. The last index considered in study is the prime index of the National Stock Exchange. This prime index of NSE is known as S & P CNX NIFTY or NIFTY, in short.

We have carried out our analysis using both the daily-level (Chapters 3 and 6) and monthly-level (Chapters 4 and 5) data. Since the construction of the different index series started at different points of time, we have taken different spans for these 4 data sets for the purpose of predicting returns on each of these stock indices at the level of daily frequency. The periods of observations considered for these daily-level data sets on Indian stock indices are as follows:

(i) BSESENSEX series spanning the period January 1986 to December 2000;

(ii) BSE-100 series covering the period January 1991 to December 2000;

(iii) NIFTY series from November 11, 1994 until December 31, 2000;

(iv) DOLLEX series over the period January 1991 to December 2000.

It may be mentioned, in this context, that for dealing with the issue of weekends and holidays in the daily sample, we have used the standard practice of dropping those days so that the resulting series could be treated as equally spaced. Further, the daily closing price has been quoted as the value of the index at daily level. All these daily-level data sets have been downloaded from the official websites of the Bombay Stock Exchange (www.bseindia.com) and the National Stock Exchange (www.nseindia.com).

At the monthly level analysis, we have used the following three major indices *viz.,* BSESENSEX, BSE-100 and NIFTY covering the period April 1996 to December 2002. The choice of this period has been made considering the availability of monthly level data and the findings on structural change(s) at the daily level analysis.

Other data sets : We now briefly state the other variables used in our study along with their data sources. For studies with daily-level data, we have used the call money rate at daily level as the representative of short-term interest rate since this series is market determined and the only one available at daily level frequency in India. The data on call money rate have been downloaded from the website of the Reserve Bank of India (www.rbi.gov.in) since its availability i.e., from January 1991. Insofar as our study with data at the level of monthly frequency is concerned, the set of following macroeconomic and financial variables have been used.

- Domestic industrial production (industrial production index with 1980-81 as the base period ; IP)
- Broad money supply (based on M3; MS)

- Consumer price index (based on consumer price index for industrial workers with 1982-83 as the base period; CPI)
- Nominal exchange rate (rupees per U.S. dollar ($ US) ; FRX)
- Domestic oil price (wholesale price index for fuels with 1981-82 as the base period ; OILD)
- NASDAQ composite stock index (NASDAQ composite of the U.S. ; NSD)
- Foreign direct investment (total monthly inflow in U.S. dollar during a month ; FDI)
- Foreign institutional investment (total monthly inflow in U.S. dollar during a month; FII)
- Short term interest rate I (91-day treasury bill rate ; TBS)
- Short term interest rate II (weighted call money rate; WCALL)
- Long term interest rate (yield on government treasury bill for 10 years' maturity; TBL)
- Foreign exchange reserve (in $US million ; FOREXR)
- Fiscal deficit of the central government (fiscal deficit is measured in Rs. crore; FD)
- Term Spread (difference between TBL and TBS ; SR)
- Price-earnings ratio (P/E)
- Price-book value ratio (P/BV)
- Dividend yield (DY)

The data sources for the three interest rate variables *viz.*, two short-term interest rates and one long term interest rate as well as those for, foreign exchange reserve, fiscal deficit of the central government have been collected from the relevant issues of Handbook of Statistics on Indian Economy, Annual publication of the Reserve Bank of India.

NASDAQ composite has been taken from the website of NASDAQ. The time series on the three financial variables (ratios) *viz.,* P/E, P/BV and DY for a stock index have been obtained from the website of the corresponding stock exchange (BSE/NSE). All the remaining data sets have been obtained from various issues of Monthly Bulletin of Reserve Bank of India. The lengths of the time series for all the macrovariables and the three financial ratios corresponding to BSESENSEX and BSE-100 are the same as that of the span of the monthly-level stock indices *viz.,* from April 1996 to December 2002. It is only for the index NIFTY that the data on the financial ratios are available from January 1999 onwards and hence that part of our study involving these financial variables and other macrovariables as well as NIFTY has been carried out with data covering the period January 1999 to December 2002.

Software used : The relevant software packages used in this study are EVIEWS, GAUSS, JMulti, TSP and the BDS program of W.D. Dechert.

CHAPTER 3

Testing Predictability of Daily Stock Returns in the Framework of Appropriate Specification*

3.1 Introduction

One of the most important and enduring questions in stock markets is whether stock returns are predictable. It is often argued that if stock markets are efficient then it should not be possible to predict stock returns. As discussed in Chapter 1 this line of argument that equates stock market efficiency with non-predictability property is not satisfactory from the point of view of understanding how stock markets operate.

Before the days of nonlinear dynamics, testing for market efficiency usually meant testing the null hypothesis that autocorrelation coefficient of different lags are statistically insignificant. But since 1980s it is well appreciated that lack of linear dependence does not rule out nonlinear dependence which, if present, would contradict Fama's efficient market hypothesis which states, in the context of stock market, that stock returns are unforecastable, and hence it i.e., linear dependence may aid in forecasting, especially over short time intervals. Specifically, Granger and Andersen (1978) and Sakai and Tokumaru (1980) have shown that simple nonlinear models exhibit no serial correlation while containing strong nonlinear dependence. This has, in fact, led several researchers like Granger and Andersen (1978), Hinich and Patterson (1985) and Scheinkman and LeBaron (1989) to look for nonlinear structures in stock returns. With

* A paper (written jointly with Nityananda Sarkar) containing the materials of this chapter, entitled " Testing Predictability and Nonlinear Dependence in the Indian Stock Market" has been published in *Emerging Markets Finance and Trade*, 2005, volume 41, no. 6, pp. 7-44.

increasing power of computers coupled with advances in both nonlinear dynamics and chaos, the volume of research into the re-examination of security returns from the standpoint of market efficiency has increased considerably, and most of these (see for instance, Hsieh (1991), Willey (1992), Sewell *et al*.(1993) and Opong *et al.* (1999)) have cast doubt on the conclusion of market efficiency based only on the lack of serial correlation in returns.

Apart from complicated nonlinear dependence/dynamics, there are two well-known reasons as to why stock prices/indices may deviate from the random walk model. First, of course, is the fact that stock returns are, in general, volatile in nature, and this has given rise to the literature on autoregressive conditional heteroscedasticity (ARCH) and generalized ARCH (GARCH) models (*cf.* Engle (1982) and Bollerslev (1986)) as well as their various extensions. The usual tests for (linear) autocorrelation perform poorly in presence of conditional heteroscedasticity in the returns. In fact, Diebold (1986), Lo and MacKinlay (1988), Silvapulle and Evans (1993), and others have noted that in the presence of ARCH, the serial correlation tests, if not corrected, can result in misleading inferences. The other reason for stock returns to deviate from the random walk model is due to what is known as calendar anomalies. As discussed in Chapter 1, these anomalies, known as the day-of-the-week effect in case of daily-level returns, if present, indicate that stock returns have a predictable pattern in their movements. Additionally, there may be predictable component in stock returns in the form of significant time-varying risk factor.

This chapter of the thesis is concerned with studying predictability and nonlinear dependence (through GARCH specification for conditional heteroscedasticity) in Indian stock returns at daily level frequency with due emphasis on appropriate specification of the conditional first and second-order moments so that the final inferences are free from any possible consequences of misspecification. The importance of appropriate specification in statistical inference has already been mentioned in the first chapter, and accordingly we advocate this modelling approach where tests are first carried out to find if there has been any mispecification in the conditional mean so that necessary adjustments in the mean are done accordingly and then the conditional variance is appropriately specified. It may be noted in this context that one of the most important and useful tests available in the literature for detecting nonlinear patterns i.e., the existence of potentially forecastable structures, is due to Brock *et al.* (1987, revised 1996), to be henceforth denoted as the BDS test. Now, it may be the case that the BDS test for testing nonlinear dependence in ARCH/GARCH adjusted stock returns might reject the null hypothesis of i.i.d. returns owing to existence of dynamics in the higher-order moments. But, when the i.i.d. null is found to be rejected in such residuals, researchers usually do not carry out any further analysis to establish the existence of other such higher-order dependencies in the adjusted returns. In modelling aspects like proper specification of both the conditional mean and the conditional variance have been considered, rejection of null i.i.d. may be attributed to the existence of some dynamics in higher-order moments only.

This chapter has been organized as follows. Section 3.2 presents the modelling approach along with description of some test procedures. Empirical results are discussed in Section 3.3. The chapter concludes with some observations in Section 3.4.

3.2 Modelling approach

In this section we describe the approach to be followed along with the tests to be used in this study with stock returns at daily frequency. Assuming p_t to be the logarithm of stock price index, P_t, return r_t is defined as $r_t = p_t - p_{t-1}$. Fama's efficient market hypothesis (EMH) is very often tested by assuming p_t's to follow a random walk model under the null hypothesis. It is well-known that the simplest version of the random walk model is the independently and identically distributed (i.i.d.) increments in which the dynamics of $\{p_t\}$ are given by the following equation:

$$p_t = \mu + p_{t-1} + \varepsilon_t, \quad t = 1,2,...,n, \text{ and } \quad \varepsilon_t \sim \text{i.i.d.}(0, \sigma^2)$$

where μ is the expected price change or drift. Independence of the increments $\{\varepsilon_t\}$ implies not only that increments are linearly uncorrelated, but that any nonlinear functions of the increments are also uncorrelated. The relevant null and alternative hypotheses are thus stated as H_0 : r_t is i.i.d. and H_1 : r_t is not i.i.d. It may be noted that both linear and nonlinear dependencies are allowed under the alternative.

Now, we first test the stationarity of r_t by applying the augmented Dickey-Fuller (ADF) test due to Said and Dickey (1984) and Phillips-Perron (PP) test (1988). Once the stationarity of r_t is established, we carry out tests for serial correlation in r_t. To this end, we first use the automatic variance ratio test by Choi (1999), as described below.

Automatic Variance Ratio Test : Lo and MacKinlay (1988) suggested a procedure for testing EMH which was quite different from the usual test of serial correlation. The novelty of their test, known as variance ratio test, is that it is robust to many forms of conditional heteroscedasticity. Though variance ratio test is intuitively appealing and known to have optimal properties under certain conditions (*cf.* Faust (1992)), the main limitation of this test is regarding the choice of lag truncation point which in the absence of any objective criterion, is often chosen arbitrarily by researchers. In view of this shortcoming, Choi (1999) suggested that the variance ratio test be calculated by using Andrews(1991) optimal data dependent method. This is what is known as automatic variance ratio test.

For testing the random walk hypothesis, the usual variance ratio estimator of Lo-MacKinlay defined by $VR(l) = Var(p_t - p_{t-l}) / lVar(p_t - p_{t-1})$ equals one at all possible lag truncation points under H_0 of no serial correlation. Hence, Lo and MacKinlay suggested comparing a consistent estimate of $VR\ (l)$ with one to test for EMH. It is, however, clear that the test crucially depends on the arbitrary choice of the lag truncation point. To take care of this limitation, Choi suggested using the quadratic spectral Kernel originally due to Andrews (1991). This is optimal in estimating the spectral density at the zero frequency and hence the lag truncation point is also chosen optimally. Thus, the automatic variance ratio estimator is defined as

$$V\hat{R}(l) = 1 + 2\sum_{i=1}^{n-1} K(i/l)\ \hat{\rho}(i) \tag{3.1}$$

where

$$\hat{\rho}(i) = \sum_{t=1}^{n-i} r_t r_{t+i} / \sum_{t=1}^{n} r_t^2 \quad \text{and} \quad K(x) = \frac{25}{12\pi^2 x^2}\left[\ \frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\ \right].$$

After proper standardization of the variance ratio estimator $V\hat{R}\ (l)$, the standardized statistic becomes $VR = \sqrt{n/l}\ [V\hat{R}\ (l) - 1] / \sqrt{2}$, which is called the automatic variance ratio test statistic. Under the null of no serial correlation, the asymptotic distribution of the statistic has been found to be standard normal (*cf.* Priestley (1980) and Choi (1999)). As this test is a two-sided test, the critical values are taken from both tails of the standard normal distribution.

After we have carried out the automatic variance ratio test, we consider specification of the conditional mean of returns as well as test(s) for detecting possible sources of misspecification of conditional mean; the outcome of the tests would enable us to take steps to ensure that the conditional mean is properly specified. As already discussed in Chapter 1 as well as in the previous section, returns might deviate from i.i.d. assumption for reasons of existence of serial correlations, seasonal effects, time-varying risk factor, conditional heteroscedasticity and other nonlinear dependencies. It is also understandable that any possible misspecification of the conditional mean might include, *inter alia*, exclusion of contemporaneous variables. Thus, consideration of proper specification entails that we include such contemporaneous independent variables in the specification of conditional mean of $r_{t.}$ Researchers like Ang and Bekaert (2001) have found nominal interest rate as the most "popular" predictor of stock returns. Fama and Schwert (1977), Campbell (1987), Lee (1992) and Shiller and Beltratti (1992) have also observed that predictability of excess stock returns could be explained by nominal interest rate.[1]

Taking all these into consideration, we propose the specification of the conditional mean of r_t to be as follows :

[1] As considered in the next chapter, there are other macroeconomic variables which can be possible predictors of stock return; but data for these variables are not available at the level of daily frequency and hence these are not being included in the specification in (3.2).

$$r_t = \sum_{k=1}^{\tilde{m}} \varsigma_k r_{t-k} + \sum_{j=1}^{d} \beta_j D_{jt} + \omega\, i_t + \varepsilon_t, \quad t = 1,2, \ldots, n, \tag{3.2}$$

$$\varepsilon_t | \psi_{t-1} \sim N(0, h_t),$$

where h_t represents conditional variance at time t, D_j's $(j = 1,2, \ldots, d)$ denote the seasonal 0-1 dummies, i_t is the nominal interest rate, $\psi_{t-1} = \{r_{t-1}, r_{t-2}, \ldots, \}$ stands for the information set at time t-1, and $\tilde{m}$ is the appropriate lag value of r_t capturing its autocorrelations and it is determined by Hall's (1994) procedure. The specification in (3.2) may be conveniently written, in vector notation, as

$$r_t = Z_t'\gamma + \varepsilon_t \tag{3.3}$$

where $Z_t' = (r_{t-1}, \ldots, r_{t-m}, D_{1t}, \ldots, D_{dt}, i_t)$ and $\gamma' = (\varsigma_1, \ldots, \varsigma_m, \beta_1, \ldots, \beta_d, \omega)$.

Now, we first test if this conditional mean is correctly specified without any consideration to conditional heteroscedasticity. In case the test rejects the null hypothesis of no misspecification of conditional mean, then suitable procedures are adopted to specify the conditional mean appropriately. Given the specification in (3.2), omitted variables and parameter instability are the two most important sources of misspecification. As mentioned in footnote 1 of this chapter, other relevant macroeconomic variables could not be included in (3.2) because of non-availability of data at this high (i.e., daily) frequency. This non-inclusion may also be due to the fact that often it is not known what these variables are. Primarily because of these two reasons, omitted variables are potentially more important than other sources of misspecification including parameter instability. Obviously, the latter is also relevant and important, especially when the span of the data set is long enough so that structural adjustments are likely to have occurred, as is the case for India. We first consider the issue of testing parameter stability and, to that end, we suggest applying Andrews's (1993) test.

Andrews's Test: The classical test for structural change is attributed to Chow (1960) who first proposed a test for structural break(s) in econometric literature for stationary variables. Chow's test was developed to test the null hypothesis of parameter constancy against the alternative of a known break point *a priori* under the assumption of constant variances. Since both the assumptions *viz.,* the break point being known *a priori* and the variance being constant are highly restrictive, Quandt (1960) suggested testing the null hypothesis of constant coefficients against a more general alternative, where a structural change has occurred at some unknown time and the error variance is also allowed to change. He proposed using appropriate likelihood ratio (*LR*) test for all possible breakdates; but he himself noted, on the basis of a Monte Carlo experiment, that χ^2 distribution is a poor approximation to this test under the null of ' no structural change'. Consequently, the Quandt statistic had no practical application for about three decades. In the early 1990's the problem was primarily solved by Andrews (1993) and Andrews and Ploberger (1994). Andrews (1993) derived the asymptotic distribution of the *LR* test, denoted as *supLR*, for one-time structural change with an unknown change point ((i.e., Quandt test) as well as analogus Wald (*supW*) and Lagrange Multiplier/ Rao's score (*supLM*) tests. These distributions are valid for models with no deterministic or stochastic trends as well as for nonlinear models. Andrews (1993 and 2003) also provided asymptotic critical values which are, incidentally, considerably larger than the comparable χ^2 critical values (for details see, Maddala and Kim (1998) and Hansen (2001)).

One way to see the construction of this statistic is to obtain the *supW, supLR, supLM* statistics as a function of all possible breakdates. However, as pointed out by Hansen (2001), we cannot consider breakdates too close to the beginning or end of the samples, as there are not enough observations to identify the subsample parameters. Conventionally, the search is confined to (0.15,0.85) percent of the observations. This sequence of statistic values are plotted against the candidate breakpoints, and then it is checked if the sequence breaks above the Andrews's appropriate critical value. In case it does, then we conclude that the time series has a structural break.

Treating the date of structural change i.e., the breakdate as an unknown parameter, the issue in case the null hypothesis is rejected is how to estimate the breakdate. To this end, we refer to the theory of least squares estimation developed by Bai and others. Bai (1997a, 1994) has also derived asymptotic distribution of the breakdate estimator. The method of estimation requires that the sample is split at each breakdate and the regression parameters estimated separately on each subsample and also the sum of squared errors for each subsample. Then the sum of squared errors is calculated for the entire sample. The breakdate corresponding to the minimum full-sample sum of squared errors (equivalently the minimum residual variance) is indeed the estimate of the breakdate. This may as well be found by plotting the residual variances against the breakdates.

Insofar as determining multiple breakdates is concerned, Chong (1995) and Bai (1997b) have suggested a sequential approach. The point to be noted is that when there are multiple structural breaks, the sum of squared errors can have a local minimum near each breakdate. Thus, the global minimum can be used as a breakdate estimator, and the other local minima may be viewed, after careful considerations, as candidate breakdate estimators. The sample is then split at the breakdate estimate, and the analysis continues on the subsamples.

In order to ensure that conditional mean is appropriately specified, we next test for any remaining misspecification in the conditional mean for each of these sub-periods. This is done by carrying out a test based on recursive residuals. Such a procedure has been used by many researchers including Lumsdaine and Ng (1999). It may be noted at this stage that any remaining misspecification of the conditional mean may be nonlinear in nature. Our motivation is that any unobserved nonlinearity will be manifested in the recursive residuals, and this nonlinearity may be approximated by functions of the recursive residuals as defined in Brown *et al.* (1975). Kianifard and Swallow (1996) and others have also demonstrated that among many standard tests for model misspecification, use of recursive residuals (rather than standard OLS residuals) increases the power of such tests.

Test of Misspecification : The application of this test proposed by Lumsdaine and Ng (1999) calls for a two-step estimation procedure. Starting from $m^* + 1$ observations where $m^* = \tilde{m} + l + 1$ $(< n)$, recursive estimation of r_t on the regressors specified in the right-hand side of (3.2) over the remaining $n - m^*$ observations are carried out in the first step. This leads to a set of recursive estimate of parameters, $\hat{\gamma}_t$, based on t observations and a set of recursive residuals $\hat{w}_t$ defined as $\hat{w}_t = r_t - Z'_t \hat{\gamma}_{t-1}$. These recursive residuals contain information for updating $\hat{\gamma}_t$ from $\hat{\gamma}_{t-1}$, and cannot be predicted by the regression model given information at time *t*-1. As noted by Lumsdaine and Ng, the recursive residuals are appealing not just because they are easy to compute, but because $\hat{w}_{t-1}$ is in the econometrician's information set at time *t*. This is the reason behind using $\hat{w}_{t-1}$ in (3.4) below at time *t,* rather than $\hat{w}_t$. Obviously, the use of OLS residuals is invalid for the same reason. By construction, these are serially uncorrelated if the model is correctly specified. If, however, the model is misspecified, $\hat{w}_t$ would then contain information about the true conditional mean not captured by the regression function. In the second step we estimate

$$r_t = Z'_t\gamma + f(\hat{w}_{t-1}) + \varepsilon_t{}^{2} \tag{3.4}$$

where $f(\hat{w}_{t-1})$ is a function (likely to be nonlinear) of the recursive residuals $\hat{w}_{t-1}$. In practice, we often try out $f(\hat{w}_{t-1}) = \tau_1\hat{w}_{t-1},\ \tau_2\hat{w}^{\,2}_{t-1}, \tau_3\sum_{i=1}^{t-1}\hat{w}_i$, and $\tau_1\hat{w}_{t-1} + \tau_2\hat{w}^2_{t-1} + \tau_3\sum_{i=1}^{t-1}\hat{w}_i$. If one or more of the τ-coefficients turn out to be statistically significant, we retain the corresponding terms in conditional mean specification of r_t so that the specification does not suffer from any inappropriateness.

Following the approach discussed so far, and also keeping in mind that time-varying risk premia may also be a predictor of stock returns, we now specify the conditional mean of r_t in (3.2), more appropriately as

[2] We use the same notation ε_t for the disturbance term in (3.4) with the understanding that ε_t represents the error associated with the correctly specified mean function.

$$r_t = \sum_{k=1}^{\widetilde{m}} \varsigma_k r_{t-k} + \sum_{j=1}^{d} \beta_j D_{jt} + \omega\, i_t + f(\hat{w}_{t-1}) + \phi\, h_t^{\lambda} + \varepsilon_t\,, \tag{3.5}$$

where h_t is the conditional heteroscedasticity at time point *t* representing risk, and λ is a transformation parameter. Although risk may have a more general representation like Box-Cox transformation as suggested by Das and Sarkar (2000) for ARCH-M model, we consider, keeping in mind its limited role in this study, only three functional forms of risk *viz.*, h_t, $\sqrt{h_t}$ and ln h_t. Now, insofar as proper specification of h_t is concerned,

$$h_t = \alpha_0 + \sum_{j=1}^{d} \theta_j D_{jt} + \alpha_1 \varepsilon_{t-1}^2 + \ldots + \alpha_q \varepsilon_{t-q}^2 + \delta_1 h_{t-1} + \ldots + \delta_p h_{t-p} \tag{3.6}$$

where $D_j's, j = 1,2,\ldots,d$, are seasonal dummies on volatility, $\alpha_0 > 0, \alpha_i \geq 0$ for all $i = 1,\ldots,q$ and $\delta_j \geq 0 \;\forall\; j = 1,\ldots,p$. As noted by Nelson and Cao (1992), this is a sufficient condition for h_t to be positive; weaker sufficient conditions also exist. The reason behind including seasonal dummies for the specification of h_t is the fact that we are here concerned with proper specification of both the first and second order conditional moments. Seasonal dummies have been included in variance in Hsieh (1989). It is conceivable that if returns exhibit seasonal patterns, then so are likely to be the case with observed volatility although the pattern in the latter may not be the same.

The issue of appropriate specification of the conditional variance therefore reduces to proper choice of the values of p and q of the underlying GARCH (p,q) process as specified in (3.6). This is done by carrying out usual diagnostic checking with the standardized residuals as well as with their squared values. It may be noted in this context that usually tests about higher order moments of residuals implicitly assume correct specifications of the lower moments. Since our proposed method of analysis tries to ensure that the conditional mean is properly specified, the routine diagnostic tests should yield appropriate results. In this context it may be noted that the robustification of tests for serial correlation so as to allow for possible misspecification of conditional variance, as suggested by Robinson (1991) and Wooldridge (1991a,b), cannot be carried out because of their non-availability in any standard software package.

BDS Test: We now suggest detecting nonlinear dependencies in the data by using what is known in the literature as the BDS test (Brock *et al*. (1996)). Since it is known that financial data often possess time varying volatilities characterized by GARCH and its variants, the BDS test would be an appropriate test for testing the null of no serial correlation in r_t against the alternative of serial correlation. In fact, under the set-up of BDS test, the null hypothesis is specified as $\{r_t\}$ being i.i.d. and the alternative includes, in addition to serial correlation, higher-order dependencies specified by GARCH as well as other unspecified nonlinear forms. The BDS test statistic measures the statistical significance of the correlation dimension calculations. The correlation integral is a measure of the frequency with which temporal patterns are repeated in the data.

The BDS test statistic is defined as :

$$BDS(m,\xi) = \sqrt{n-m+1}\,\frac{T_m(\xi)}{V_m(\xi)} \tag{3.7}$$

where n is the total number of observations, $T_m(\xi) = C_m(\xi) - C_1(\xi)$, $C_m(\xi)$ and $C_1(\xi)$ are the correlation integrals as defined in Brock *et al.*, $V_m(\xi)$ is the standard error of

$T_m(\xi)$ (ignoring the constant $\sqrt{n-m+1}$) and ξ and m are the distance and dimension respectively, as defined below. This test statistic converges in distribution to *N (*0, 1*)* under H_0. BDS test has the advantage that no distributional assumption needs to be made in using it as a test statistic for i.i.d. random variables. Two parameters are, however, to be chosen by the user. These are the values of ξ (the radius of the hypersphere which determines whether two points are 'close' or not) and m (the value of the embedding dimension). As suggested by Brock *et al*. (1991), Hsieh (1991) and Sewell *et al*. (1993), in most of the cases the values of ξ used are 0.5σ and σ, where σ represents the standard deviation of the linearly filtered data, and the value of m is set in line with the number of observations (e.g., use only $m \leq 5$ if $n \leq 500$). Returns are filtered for linear dependence using (3.5). To examine whether the higher order dependence structure can be adequately captured by GARCH, GARCH standardized residuals are then tested for i.i.d. using BDS test statistic. Since Brock *et al*. and Hsieh have pointed out that the asymptotic standard normal distribution of BDS statistic does not apply to GARCH standardized residuals, appropriate critical values (derived from simulation) for the BDS test applied to the standardized residuals of a GARCH (1,1) are taken from Brock *et al*. (1996) and Brooks and Heravi (1999).

Finally, if the null hypothesis of i.i.d. is not found to be acceptable by the BDS test, we then suspect that it may be due to some dynamics in higher order (greater than second) moments of the residuals. Towards this end, we advocate studying the regressions of higher order standardized residuals on their respective lagged values and test if one or more of the coefficients turn out to be significant. Obviously, in such a case we cannot model the returns further by incorporating such dependencies in an appropriate manner so that the residuals thus obtained would turn out to be i.i.d.

3.3 Empirical findings

As already stated in Chapter 2, this study with daily-level data have been carried out with four sets of data on stock indices representing Indian stock market. The reasons for taking more than one index series has been discussed in Chapter 2. These data sets are the followings: (I) Bombay stock exchange sensitive index (BSESENSEX) at daily level spanning the period January 1986 to December 2000, (ii) Bombay stock exchange national index, currently known as BSE 100, at daily level covering the period January 1991 to December 2000, (iii) NIFTY index of National Stock Exchange from November 1994 until December 2000, and (iv) DOLLEX at daily level over the period January 1991 to December 2000. In dealing with the issue of weekends and holidays in the daily sample, we have used the standard practice of dropping those days so that the resulting series could be treated as equally spaced. Insofar as the data on short-term interest rate is concerned, we have used the call money rate at daily level. Since call money rate is market determined and available at the daily level we have chosen this as short term interest rate.

In this section we report and discuss the results of our analysis with daily level data. The stock return at period *t* is defined by $r_t = \ln P_t - \ln P_{t-1}$, where P_t is the stock price index at period *t*. Hence the analyzed data represent the continuously compounded rates of return for holding the (aggregate) securities for one day.

The visual description of the four return series are given in Figures (3.1) through (3.4) below. All the four series appear to be stationary around zero with no deterministic trend, and exhibit volatilities of varying degrees. The usual statistical details of the data sets *viz.*, the values of mean, standard deviation, the coefficients of skewness and kurtosis as well as the values of Ljung-Box test statistic for testing ARCH in the daily returns of the four series - BSESENSEX, BSE 100, NIFTY and the DOLLEX - are given in Table 3.1. The means and standard deviations of the returns based on the full sample period values of the four indices show that means are not significantly different from zero. The values of coefficient of skewness show that all the distributions except possibly that of NIFTY are skewed, indicating thus their departure from normality. Also, all the excess kurtosis values are much larger than 0. This shows that these three return series *viz*., BSESENSEX, BSE 100 and DOLLEX have fat tails as compared to normal distribution. We also find from Table 3.1 that in all the four series, the Ljung-Box test statistic values for the squared residuals are significant, indicating the presence of second order dependence in returns.

Table 3.1 : Summary Statistics of Returns Based on

BSESENSEX, BSE 100, NIFTY and DOLLEX

Index	Mean	Sd	Skew-ness	Excess kurto-sis	Q^2 (12)	Q^2 (22)	Q^2 (32)	Q^2 (42)
BSESENSEX	0.000595	0.0196	0.275	6.188	697.950	1040.330	1435.380	1490.28
BSE 100	0.000608	0.0186	0.411	7.704	366.030	600.318	699.443	735.753
NIFTY	-0.000011	0.0171	0.0642	2.508	87.270	127.645	131.902	136.965
DOLLEX	0.000113	0.0192	-0.309	14.107	273.829	369.571	508.173	516.445

Note: $Q^2(k)$ represents the value of Ljung-Box statistics of squared returns with k degrees of freedom. Further, all the $Q^2(k)$ values are significant at 1% level of significance.

Figure 3.1: BSESENSEX Daily Return

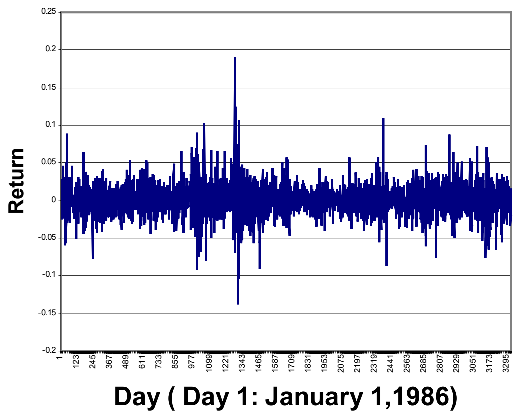


Figure 3.2: BSE100 Daily Return

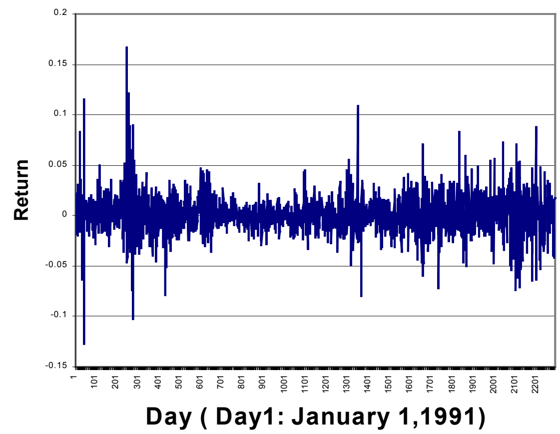


Figure 3.3: NIFTY Daily Return

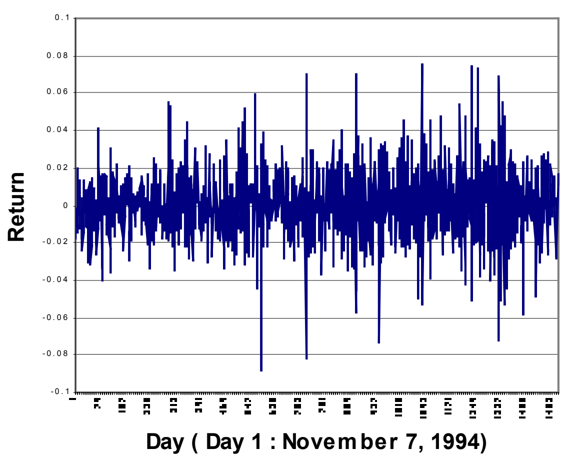
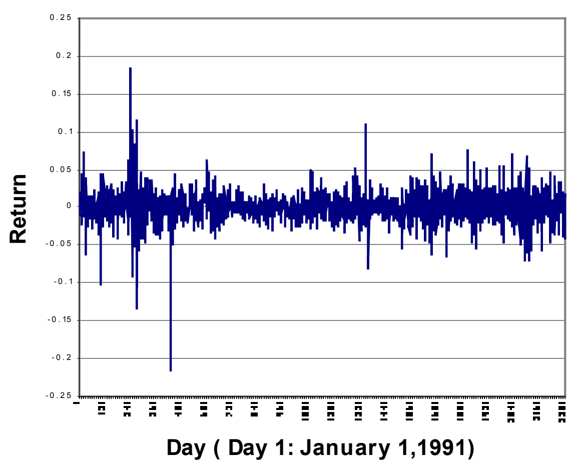


Figure 3.4: DOLLEX Daily Return



We then carried out two standard tests of stationarity of the return series *viz.*, ADF and PP tests. The values of the test statistic for all the four series are presented in Table 3.2.[3] It is evident that all the four return series are stationary at 1 percent level of significance. Once stationarity of returns has been confirmed, we analysed the returns for the four data sets by following the procedure described in the previous section.

Table 3.2 : Unit Root Tests of Stock Returns

Index	ADF	PP	Number of observations
BSESENSEX	-10.721	-3248.166	3323
BSE 100	-9.739	-2260.894	2285
NIFTY	-10.541	-1391.202	1506
DOLLEX	-9.221	-2305.232	2283

Note : All the values of the two test statistics are significant at 1% level of significance. Neither a constant and nor a linear trend term was included as exogenous regressor in the regressions since Figures(I)-(IV) do not show trend or nonzero mean in any of the four series. Maximum lag length was to be 26, 20, 13 and 20 for the four regressions (in that order), respectively.

Excepting for automatic variance ratio test and Andrews's test all the computations have been carried out with TSP 4.3 software package. The computations for automatic variance ratio test have been done with GAUSS and that of Andrews's with EVIEWS 3.1.

[3] In carrying out ADF and PP tests, maximum lag length was chosen to be 26, 20, 13 and 20 for BSESENSEX, BSE 100, NIFTY and DOLLEX, respectively. Since all the four return series do not show any deterministic trend or non-zero mean, stationarity tests were performed using the pure random walk model.

Table 3.3 : Results of Automatic Variance Ratio (AVR) Test

Index	AVR Test Statistic
BSESENSEX	2.950***
BSE100	5.158***
NIFTY	1.677*
DOLLEX	6.554***

*Note : *indicates significant value of the test statistic at 10% level of significance and ***indicates the same at 1% level of significance.*

The seven stages of computations (in order) involved are as follows.

I. Testing for no serial correlation by Choi's automatic variance ratio test;

II. Testing for parameter stability by Andrews's test;

III. Consequent partitioning of the entire time period into sub-periods of stable parameters each by applying the method of least squares estimation for estimating the breakpoint(s), as proposed by Bai (1994, 1997a,1997b), Bai and Perron (1998);

IV. Testing for misspecification of conditional mean based on recursive residuals;

V. Testing for adequacy of GARCH model by standard diagnostic tests based on standardized residuals as well as their squared values;

VI. Testing for nonlinearity by BDS test;

VII. Testing for the presence of dynamics of higher order moments in residuals.

The results of automatic variance ratio test are presented in Table 3.3. We find from this table that the values of the test statistic are 2.950, 5.158, 1.677 and 6.554 for BSESENSEX, BSE 100, NIFTY and DOLLEX, respectively. On comparing these values with the critical values (two-sided) of N (0, 1) distribution we obviously find that the null of no serial correlation is rejected for all the four returns series; although the level of significance for NIFTY series is only 10%. We may thus conclude on the basis of the findings of this test that all the four stock indices are inefficient and hence predictable.

We present below the estimated models for returns based on BSESENSEX, BSE 100, NIFTY and DOLLEX. The model for each of the four series was estimated after considering lags of r_t upto a maximum of 20, five 0-1 dummies representing day-of-the-week effects, and the daily series of call money rate to represent the short-term interest rate, as explanatory variables[4]. In all the four models, the usual Ljung-Box $Q(.)$ test concludes that the residuals are white noise.

<u>BSESENSEX</u>

$$\hat{r}_t = 0.081\ r_{t-1} + 0.050\ r_{t-17} + 0.04\ r_{t-20} + 0.001\ \text{D5} \qquad (3.8)$$

(4.667)*** (2.901)*** (2.328)** (1.748)*

<u>BSE 100</u>

$$\hat{r}_t = 0.119\ r_{t-1} + 0.062\ r_{t-3} + 0.088\ r_{t-9} + 0.051\ r_{t-10} - 0.073\ r_{t-19} + 0.002\ \text{D1} \qquad (3.9)$$

(5.789)*** (3.026)*** (4.236)*** (2.441)** (3.501)*** (2.811)***

<u>NIFTY</u>

$$\hat{r}_t = 0.073\ r_{t-1} - 0.066\ r_{t-6} - 0.051\ r_{t-19} - 0.003\ \text{D1} - 0.002\ \text{D2} + 0.006\ \text{D3} \qquad (3.10)$$

(2.873)*** (2.599)*** (1.996)** (3.229)*** (2.543)** (6.577)***

<u>DOLLEX</u>

[4] In the computations with daily returns on BSESENSEX, call money rate variable could not be included in the model since this data set is available from January 1991 only.

$$\hat{r}_t = 0.153\ r_{t-1} + 0.056\ r_{t-3} + 0.059\ r_{t-5} - 0.049\ r_{t-6} + 0.093\ r_{t-9} + 0.048\ r_{t-17}$$

$$(7.397)^{***}\ (2.699)^{***}\quad (2.824)^{***}\quad (2.319)^{**}\quad (4.514)^{***}\quad (2.267)^{**}$$

$$-\ 0.056\ r_{t-19} + 0.002\ \mathrm{D1} - 0.002\ \mathrm{D2}$$

$$(2.730)^{***}\quad (2.297)^{**}\quad (2.012)^{**} \tag{3.11}$$

[The values in parentheses indicate corresponding absolute values of *t*-statistic; * indicates significance at 10% level , **indicates significance at 5% level, and *** at 1% level of significance.]

Now, if we consider the estimated returns equation for BSESENSEX given in (3.8), we find that, r_{t-1}, r_{t-17} and r_{t-20} are significant; so is D5 dummy. The later finding suggests that Friday effect i.e., the weekend effect is significant. We note from the four equations (3.8) through (3.11) that Monday effect is significant in all the indices except BSESENSEX. In these regressions we have considered a moderately high number of lags so as to remove presence of any autocorrelation in the error term. Specifically, to check the lag length properly we have followed Hall's (1994) procedure. We began with a moderate value for lag length *viz.,* 20, and then reduced it one by one until the last lag was significant. Thereafter all possible subsets were considered in a continuous order, and then using the information-based model selection criterion, called the BIC, as proposed by Schwarz (1978), the model having the minimum BIC value was chosen. Finally, following the standard practice (see, for instance, Brockwell and Davis (1987) and Mills (1993)), the final reported model was obtained after dropping the insignificant intermediate lags and then re-estimating the model with only these lags which were found to be significant.

3.3.1 Testing for parameter stability

We have already mentioned that we have followed Andrews's (1993) procedure for testing structural change. We now report our findings on Andrews's *supW* test for testing parameter stability in the model specified in (3.2). It may be relevant to point out that in this testing procedure, the alternative is taken to be one where a structural change has occurred at some unknown time and the error variance is allowed to change. The issue of breakpoint being assumed to be known *a priori* or taken to be unknown and hence to be determined, still remains a debatable one. Andrews's *supW* test merely tests whether a break exists or not, but does not identify the location of the breaks. As opposed to known breakpoint, we suggest estimation of the breakpoints by applying the least squares method, as proposed by Bai (1994, 1997a, b) and others, which essentially requires searching for a break over the entire time period of observations and then taking the minimum sum of squared errors for the full sample(as a function of the breakpoint) as the estimate of the breakdate. Now, it is conceivable that some prior information often exists about the dates of major shocks (real or financial) and this suggests the approximate location of the breaks like in our case the breakpoints being around the middle of 1992 and also towards the end of 1996, based on the history of the Indian stock market in the recent past. But the fact remains, as pointed out by Hansen (2001), that the results can be highly sensitive if the *a priori* choices are somewhat arbitrary and hence can hardly be considered to be a sound scientific practice, especially because in least squares method the set of candidate breakpoints obviously includes these likely locations.

The Figures (3.5)-(3.8) graphically represent the results of Andrews's test for the four series BSESENSEX, BSE100, NIFTY and DOLLEX, respectively. Figure (3.5) shows that the maximum value of the test sequence for the series BSESENSEX is 22.74, which exceeds the relevant 1% critical value *viz.,* 20.47 corresponding to *p*=4 in the Andrews Table (2003). It is necessary to mention that in Andrews's test we have chosen the trimming parameter as 15 percent. Since the maximum value of 22.74 easily exceeds the 1% Andrews critical value, we conclude that the null hypothesis of no structural break is

Figure 3.5: Testing for Structural Change in BSESENSEX Series

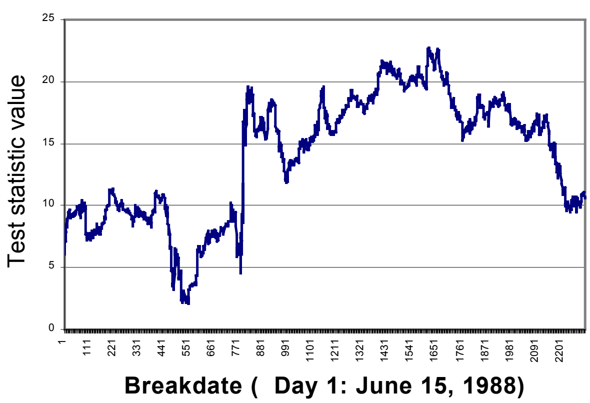


Figure 3.6: Testing for Structural Change in BSE100 Series

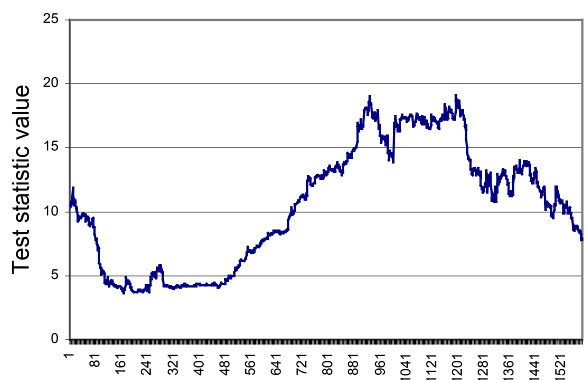


Figure 3.7: Testing for Structural Change in NIFTY Series

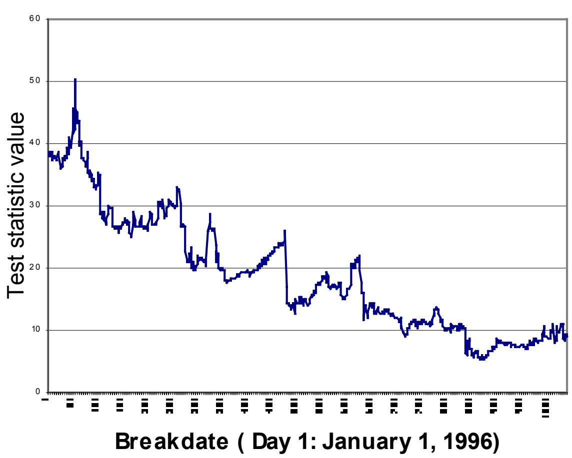
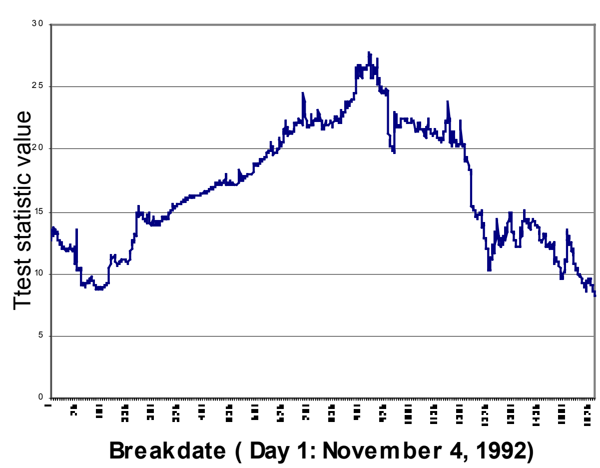


Figure 3.8: Testing for Structural Change in DOLLEX Series



decisively rejected. Thus, our conclusion on BSESENSEX is that the series has a structural break. Through similar exercises we have also found the presence of structural break for each of the remaining three series. The maximum values of the corresponding test statistic sequences for BSE 100, NIFTY and DOLLEX are 19.06, 50.37 and 27.85. Comparing these with respective Andrews's critical values, we find that while NIFTY and DOLLEX clearly indicate presence of a structural break at 5% level of significance, the same for BSE100 holds only at a slightly higher level of significance, the critical value at 5% level corresponding to p= 6 being 20.24.

Figure 3.9: Least Squares Breakdate Estimation of BSESENSEX Series

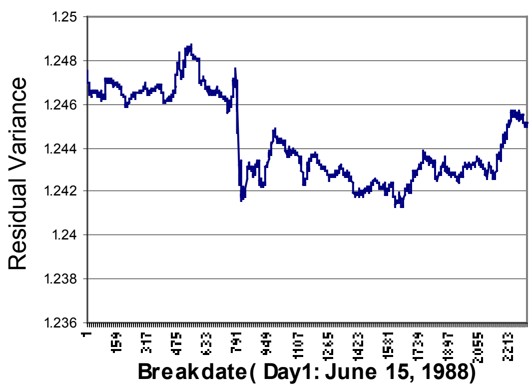


Figure 3.10: Least Squares Breakdate Estimation of BSE100 Series

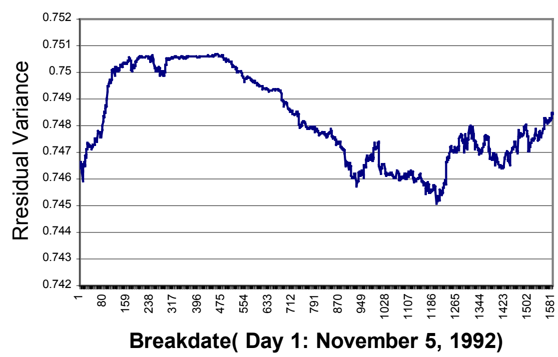


Figure 3.11: Least Squares Breakdate Estimation of NSE Series

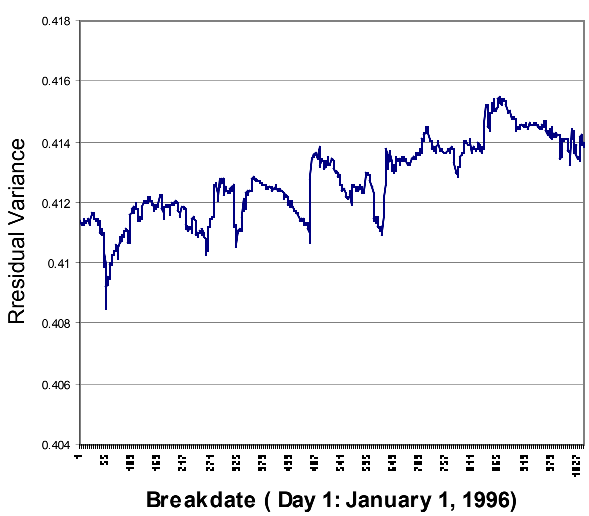
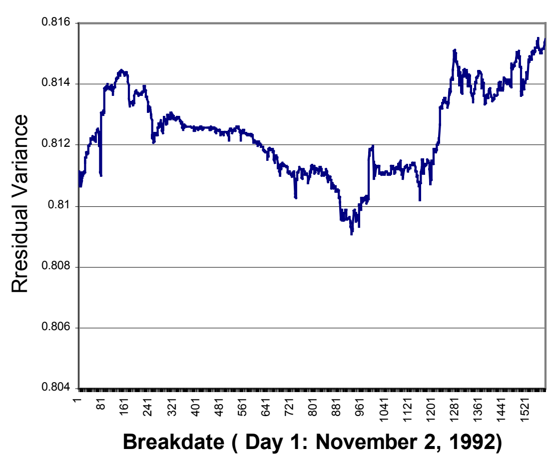


Figure 3.12: Least Squares Breakdate Estimation of DOLLEX Series



The results of breakdate estimation have been shown in Figures (3.9)-(3.12) for the series BSESENSEX, BSE 100, NIFTY and DOLLEX, respectively. For BSESENSEX series i.e., Figure (3.9), the minimum value for the residual variance is found to correspond to the breakdate of February 15,1996. Likewise, the estimated one - point breakdates for BSE 100, NIFTY and DOLLEX series are January 15, 1998, January 12,1996 and 05 December, 1996, respectively. It may be noted that confidence intervals of the breakpoints could as well be reported. But since we need to divide the whole sample period into sub-periods based on this estimation exercise, we are reporting only the point estimates of the breakdates.

We now move on to testing for presence of multiple breakdates, in the series and then estimating those breakdates sequentially, as suggested by Chong (1995) and Bai(1997b). BSESENSEX series is found to have two other breaks on June 18, 1990 and June 17 1992. Thus, BSESENSEX series is found to have three structural breakes. Accordingly, the series is divided into three sub-periods from consideration of structural break. These sub-periods along with similar ones obtained for the other three series are given below.

BSESENSEX : January 1986 - June 1990; July 1992 - February 1996; and February 1996 - December 2000.

BSE100 : May 1992 - December 1996[5]; and December 1996 - December 2000.

NIFTY : November 1994 - January 1996; and January 1996 - December 2000.

DOLLEX : May 1992 - December 1996; and December1996 - December 2000.

---

[5] Bai's estimation procedure produced the residual variance for the full sample to be 0.7451, and this was on 15th January, 1998. However, the value closest to this one was 0.7457, which corresponds to 10th December, 1996. Since the difference in values is practically negligible and the later date conforms to the

We observe from these findings that each of BSESENSEX, BSE100, NIFTY and DOLLEX series, there is a structural break in 1996. This is only expected after four years of reforms programme being pursued with vigour. In this context, it may be pointed out that while we have reported only point estimates of the breakpoints keeping in mind the requirement of dividing the entire sample period into stable sub-periods, interval estimates of breakpoints are also available, and given the continuing nature of reform programmes, the latter would be more meaningful. Thus, for instance, point estimate of break at December 1996, which, in fact, is the case with BSE 100 and DOLLEX series, means, more appropriately, that the break has occurred in an appropriately defined interval around December 1996. While carrying out this exercise on stability, we noted that the sub-period covering July 1990- April 1992, was found to be unstable in the sense that the test suggested the presence of a number of breaks during this period, but none of the resulting sub-periods was found to be stable, and hence this sub-period was excluded from further analysis of BSESENSEX series. This lack of stability during this period may be attributed to the fact that the Indian economy witnessed the biggest ever scam in its capital market from September 1991 to May 1992. Obviously, the confidence of the investors was greatly shaken, and the stock market went through a period of turbulence even after the scam was detected. Since India was facing a serious balance of payments problem towards the end of 1980's and early '90s leading to the decision of opening-up and liberalisation of the Indian economy, this period of July1990- June 1992 itself was somewhat unstable, and this got further compounded by the aforesaid scam.

---

well -established fact that Indian stock market experienced a structural break in 1996, we have taken December 1996 as the breakpoint for BSE100 series.

These findings on partitioning of time periods are thus consistent with the major developments in the Indian capital market and other sectors of the economy during 1990's. We may mention the observations of other researchers in this context. For instance, Kulkarni (1997) has pointed out that the Indian market was in the midst of the bear phase since the early 1990's due to political instability and foreign exchange crisis. We further noted that the biggest scam in the history of the Indian capital market was unearthed in April 1992, and as a consequence the government bestowed more power on the autonomous regulator, Securities and Exchange Board of India (SEBI), which, in turn, immediately swung into action. Further, to curb speculation in BSE, SEBI scrapped the forward trading system (*badla*) mechanism in December 1993, which was, however, reintroduced in January 1996.

It may be noted that the specifications used for carrying out the structural break tests for the four series i.e., equations (3.8)-(3.11), are based on a full sample model selection procedure without assuming any breaks. An alternative to this approach could be to choose a 'fixed' specification e.g., an AR(1) model with five dummies representing day-of-the-week effects. We carried out computations required of this approach, and obtained the same conclusions as earlier on the existence of breaks. As regards the breakpoint estimates, the results were found to be almost the same as those for the full sample case. The sub-periods thus obtained by this approach for the four series are as follows.

BSESENSEX : January 1986 - April 1990; May 1992 - February 1996; and February 1996 - December 2000.

BSE100 : May 1992 - December 1996; and December 1996 - December 2000.

NIFTY : November 1994 - January 1996; and January 1996 - December 2000.

DOLLEX : May 1992 - December 1996; and December1996 - December 2000.

As before, the sub-period covering almost the same span *viz.,* April 1990-April 1992 for BSESENSEX was found to be unstable. We thus find that that the two approaches produce the same empirical conclusions on the existence as well as on the location of structural breaks in the four returns series.

3.3.2 Testing for misspecification:

We now report the results of the recursive residual based test of misspecification of conditional mean. This test has been carried out to find if the conditional mean is still misspecified. After obtaining recursive residual $\hat{w}_t's$, as discussed in Section 3.2, and then including terms like $\hat{w}_{t-1}$, $\hat{w}_{t-1}^2$, $\hat{w}_{t-1}^3$, $\sum_{i=1}^{t-1}\hat{w}_i$ etc., we obtain the estimated models for each of the sub-periods of all the four data sets. These are reported below.

<u>BSESENSEX</u>

Sub-period I (January 1986 - June 1990):

$\hat{r}_t$ = - 0.059 r_{t-1} - 0.066 r_{t-2} + 0.087 r_{t-7} + 0.090 r_{t-8} +0.002D5+ 0.183 $\hat{w}_{t-1}$+ 1.206

(0.383) (1.878)* (2.679)*** (2.339)** (1.119) (1.649)* (1.227)

$\hat{w}_{t-1}^2$ -31.125 $\hat{w}_{t-1}^3$ (3.12)

(1.157)

Sub-period II (July 1992 - February 1996):

$\hat{r}_t$ = 0.117 r_{t-1} - 0.081 r_{t-2} -0.004 D5 - 0.00006 i_t -2.280 r_{t-1}^3

(0.763) (1.792)* (3.091)*** (1.109) (0.037)

\+ 0.212 $\hat{w}_{t-1}$ - 0.019 $\hat{w}_{t-1}^2$ - 46.640 $\hat{w}_{t-1}^3$ (3.13)

(1.377) (0.017) (0.814)

Sub-period III (February 1996 - December 2000):

$\hat{r}_t$ = - 0.112 r_{t-1} - 0.074 r_{t-6} + 0.062 r_{t-9} - 0.056 r_{t-11} - 0.072 r_{t-18}

(0.651) (2.585)*** (2.148)** (1.930)* (2.511)**

- 0.086 r_{t-19} + 0.002 D1 + 0.212 $\hat{w}_{t-1}$ - 1.139 $\hat{w}_{t-1}^2$ - 25.012 $\hat{w}_{t-1}^3$ (3.14)

(0.762) (1.871)* (1.207) (1.536) (1.888)*

<u>BSE 100</u>

Sub-period I (May 1992 – December 1996):

$\hat{r}_t$ = 0.303 r_{t-1} - 0.121 r_{t-2} + 0.043 r_{t-3} + 0.041 r_{t-4} +0.016 r_{t-5} + 0.032 r_{t-12}

(3.788)*** (3.175)*** (1.231) (1.154) (0.476) (1.153)

+ 0.067 r_{t-15} + 0.002D5+ 0.073 $\hat{w}_{t-1}$ -0.619 $\hat{w}_{t-1}^2$ -41.767 $\hat{w}_{t-1}^3$ (3.15)

(2.458)** (2.611)*** (0.880) (0.586) (1.788)*

Sub-period II (December 1996- December 2000):

$\hat{r}_t$ = -0.076 r_{t-1} - 0.061 r_{t-6} + 0.083 r_{t-9} - 0.118 r_{t-19} +0.004D1- 0.003D3

(0.598) (1.939)* (2.625)*** (3.762)*** (3.360)*** (2.333)**

+ 0.265 $\hat{w}_{t-1}$ +1.076 $\hat{w}_{t-1}^2$ -49.781 $\hat{w}_{t-1}^3$ (3.16)

(1.997)** (1.443) (3.404)***

<u>NIFTY</u>

Sub-period I (November 1994 – January 1996):

$\hat{r}_t$ = 0.875 r_{t-1} - 0.448 r_{t-2} + 0.242 r_{t-3} - 0.006 D1

(4.182)*** (3.875)*** (3.523)*** (4.032)***

- 0.386 $\hat{w}_{t-1}$ + 8.259 $\hat{w}_{t-1}^2$ -129.626 $\hat{w}_{t-1}^3$ (3.17)

(1.798)* (2.936)*** (1.088)

Sub-period II (January 1996- December 2000)

$\hat{r}_t$ = -0.123 r_{t-1} - 0.064 r_{t-6} - 0.074 r_{t-19} - 0.003D1- 0.003D2

(1.251) (2.284)* (2.666)*** (2.309)** (2.465)***

+ 0.007 D3 + 0.290 $\hat{w}_{t-1}$ -0.1156 $\hat{w}_{t-1}^2$ -49.781 $\hat{w}_{t-1}^3$ (3.18)

(5.941)*** (2.816)*** (0.152) (4.832)***

<u>**DOLLEX**</u>

Sub-period I (May 1992 – December 1996) :

$\hat{r}_t$ = 0.250 r_{t-1} + 0.081 r_{t-4} + 0.068 r_{t-10} + 0.090 r_{t-15} +0.004D5

(3.593)*** (2.788)*** (2.513)** (3.392)*** (3.816)

+4.324 r_{t-1}^2 - 46.657 r_{t-1}^3 + 0.2093 $\hat{w}_{t-1}$ - 10.09 $\hat{w}_{t-1}^2$ + 85.04 $\hat{w}_{t-1}^3$ (3.19)

(2.434)** (4.712)* (1.197) (5.258)*** (3.517)***

Sub-period II (December 1996- December 2000) :

$\hat{r}_t$ = - 0.057 r_{t-1} + 0.65 r_{t-8} + 0.093 r_{t-9} + 0.070 r_{t-19} +0.005D1

(0.448) (2.079)** (2.774)*** (2.216)** (3.493)***

+0.234 $\hat{w}_{t-1}$ + 0.844 $\hat{w}_{t-1}^2$ - 43.497 $\hat{w}_{t-1}^3$ (3.20)

(1.778)* (1.098) (2.746)***

[The values in parentheses indicate corresponding absolute values of *t*-statistic; * indicates significance at 10% level , **indicates significance at 5% level, and *** at 1% level of significance.]

It is evident from the above estimated equations for BSESENSEX that none of the coefficients associated with $\hat{w}_{t-1}$, $\hat{w}_{t-1}^2$, $\hat{w}_{t-1}^3$, $\sum_{i=1}^{t-1}\hat{w}_i$ is significant even at 5 per cent level of significance in any of the three sub-periods, and hence it can be concluded that there is no further misspecification in the conditional mean in the three sub-periods for BSESENSEX. We find from the estimated equations for the three other stock indices that in some cases the misspecification components are significant. As for instance, $\hat{w}_{t-1}^3$ is found to be significant at 1 per cent level of significance in the equation for Sub-period II of BSE 100. Similarly, we find that the coefficient of $\hat{w}_{t-1}{}^2$ in (3.17) for Sub-period I of NIFTY has the statistic value as 2.936, and this is obviously highly significant. As for DOLLEX, we have, for instance, two significant coefficients in $\hat{w}_{t-1}{}^2$ and $\hat{w}_{t-1}^3$ for Sub-period I. There are, of course, some other significant misspecification coefficients, other than those mentioned, for all these three indices.

3.3.3 Testing for adequacy of GARCH specification

Now that the conditional mean of returns has been properly specified, we estimate this model along with the GARCH assumption for the conditional heteroscedasticity h_t as specified in (3.6). It may be recalled that for reasons stated in the previous section, the specification of h_t involves dummies representing the day-of-the-week effects. The estimated models for r_t and h_t for the four sets of return data are given in Table 3.4 for respective sub-periods.

For each of these estimated models, the adequacy of the estimated GARCH model was examined so that there remained no misspecification in the specification of the second order conditional moment. This was done by studying the behaviour of standardized residuals and squared standardized residuals. The Ljung-Box Q(.) statistic values pertaining to these residuals are given in Table 3.5. It is evident from these computations that barring a few cases, especially those with Sub-period I of DOLLEX

Table 3.4: Estimates of Parameters in Conditional Mean and Variance

Variable	BSESENSEX			BSE100		NIFTY		DOLLEX	
	Sub-period			Sub-Period		Sub-period		Sub-period	
	I	II	III	I	II	I	II	I	II
r_{t-1}	0.099 (2.810)	0.285 (7.205)	0.092 (2.826)	0.346 (10.189)	0.118 (3.131)	0.418 (6.155)	0.190 (5.057)	0.334 (10.429)	0.119 (3.163)
r_{t-2}	-0.078 (2.278)	-0.118 (3.095)	–	-0.139 (3.596)	–	-0.148 (2.356)	–	–	–
r_{t-3}	–	–	–	0.087 (2.547)	–	–	–	–	–
r_{t-6}	–	–	-0.10 (3.591)	–	-0.094 (3.253)	–	-0.083 (2.799)	–	–
r_{t-7}	0.074 (2.153)	–	–	–	–	–	–	–	–
r_{t-8}	0.061 (1.752)#	–	–	–	–	–	–	–	–
r_{t-9}	–	–	–	–	–	–	–	–	–
r_{t-11}	–	–	-0.072 (2.811)	–	–	–	–	–	–
r_{t-15}	–	–	–	–	–	–	–	-0.084 (3.999)	–
r_{t-18}	–	–	-0.072 (2.753)	–	–	–	–	–	–
r_{t-19}	–	–	-0.055 (2.035)	–	-0.090 (3.553)	–	-0.082 (3.121)	–	–
D1	–	–	0.004 (2.734)	–	0.007 (7.776)	-0.005 (3.566)	–	–	-0.008 (8.823)
D3	–	–	–	–	0.002 (1.909)#	–	0.008 (7.399)	–	–
D5	–	0.003 (3.727)	–	0.00195 (2.629)	–	–	–	0.004 (4.508)	–
r^3_{t-1}	–	–	–	–	–	–	-93.01 (5.470)	–	–
i_t	–	-0.00008 (2.170)	–	–	–	–	–	–	–
Constant	0.000002 (0.408)^	–	0.00002 (1.305)#	0.000003 (3.342)	0.00008 (5.918)	0.00001 (1.672)	0.00001 (1.319)#	0.00001 (4.713)	0.00005 (4.197)
ε^2_{t-1}	0.048 (3.342)	0.106 (4.407)	0.148 (4.831)	0.077 (5.564)	0.241 (5.323)	0.199 (3.263)	0.065 (5.051)	0.218 (14.68)	0.166 (4.488)
h_{t-1}	0.888 (23.847)	0.869 (38.212)	0.60 (8.978)	0.904 (68.232)	0.539 (8.556)	0.740 (9.303)	0.875 (37.32)	0.731 (29.912)	0.698 (11.531)
D1	0.00008 (3.024)	–	0.0002 (9.139)	–	–	–	0.00004 (2.338)	–	–
D3	–	–	0.00006 (2.230)	–	–	–	–	–	–
D4	–	0.230 (3.346)	–	–	–	–	–	–	–

Note: The values in parentheses indicate corresponding absolute values of t-statistic. All values in the table are significant at either 1 percent or 5 percent level of significance excepting those (a) indicated by # which are significant at 10 percent level only and (b) marked by ^ which are not significant at any of these standard levels.

Table 3.5 : Diagnostic Checking of Residuals from the Estimated GARCH Models

	Standardized residual $(\hat{\varepsilon}_t / \sqrt{\hat{h}_t})$				
Index	*Q (12)*	*Q (22)*	*Q (32)*	*Q (42)*	LM-statistic$
BSESENSEX					
Sub-period I	3.89	9.16	19.12	26.02	0.153
Sub-period II	7.62	18.13	30.09	41.28	0.244
Sub-period III	14.02	22.72	36.04	47.71	0.008
BSE100					
Sub-period I	10.00	17.30	22.16	32.68	0.009
Sub-period II	17.73	24.61	33.34	42.45	0.251
NIFTY					
Sub-period I	8.57	15.76	18.12	30.56	0.475
Sub-period II	7.73	14.60	22.88	26.87	1.252
DOLLEX					
Sub-period I	19.99	30.15	41.82	61.00*	35.505
Sub-period II	19.78	30.91	40.83	48.68	0.426

Table 3.5 (contd.)

Index	Squared standardized residual $(\hat{\varepsilon}_t^2 / \hat{h}_t)$				
	Q (12)	*Q (22)*	*Q (32)*	*Q (42)*	LM-statistic$
BSESENSEX					
Sub-period I	3.10	9.70	20.32	20.32	0.00003
Sub-period II	9.473	14.36	21.17	24.01	0.008
Sub-period III	7.573	32.21	43.79	45.47	0.083
BSE100					
Sub-period I	10.88	13.16	16.65	26.45	0.005
Sub-period II	3.87	37.96*	42.71	45.79	0.047
NIFTY					
Sub-period I	21.71*	27.91	36.61	43.72	0.008
Sub-period II	7.71	27.17	31.55	37.88	0.018
DOLLEX					

Sub-period I	41.76**	49.28**	52.32**	56.74	194.89**
Sub-period II	3.459	46.33**	49.35*	52.52	0.044

*Note : *indicates significance at 5% level and*
***indicates significance at 1% level.*
Q(k) represents the Ljung-Box statistic value with k degrees of freedom.
$ LM-statistic stands for the usual LM test for GARCH model.

none of these values is significant for all the four data sets and for all the sub-periods indicating thus the adequacy of the estimated GARCH specifications. It may be pointed out that in most of the cases we have found GARCH (1, 1) to be the most appropriate specification along with some day-of-the-week effects. If we look at the estimated values from Table 3.4,we can make the following observations.

(i) Call money rate is found to be not significant in most of the cases except BSESENSEX. Again, for BSESENSEX it is significant for Sub-period II only. Since in Indian capital market BSESENSEX is the most representative index of the economy, these findings on the significance of call money rate pertaining, in particular, to the second half of the last decade suggest that the present policy of the government of India of reducing interest rate to encourage investment in the stock market is likely to be effective in due course.

(ii) Day-of-the week effects are present in most of the cases both in mean as well as in volatility specifications. These findings indicate r_t's are often predictable from consideration of daily pattern. However, there is no uniform day-of-the week effect across all sub-periods for any index. For instance, in BSE SENSEX Friday is significant for Sub-period II, and Monday is significant for Sub-period III. Again for BSE 100, while no day-of-the- week effect is present for Sub-period I, Monday is significant for Sub-period II. But for NIFTY, Monday is significant for Sub-period I, and Wednesday is significant for Sub-period II. For DOLLEX Friday is significant for Sub-period I and Monday is significant for period II. While the presence of Monday and Friday effects are quite understandable, the same for D3 (Wednesday) in sub-period II of NIFTY is explained by the fact that the settlement period for NIFTY is from Wednesday to next Tuesday.

(iii) Some lagged values of r_t are significant in each of these estimated models. While computing we allowed a moderately high lag length so that the error in the model is free from autocorrelation. In most of the cases lag 1 was found to be significant so that the role of previous day's return in prediction is empirically established. However, in some cases, lags 18 and 19 were found to be significant indicating monthly effect in the returns since there are, on the average, 20 observations in a month.

(iv) The daily level seasonality was found to be present in volatility equation in some of the sub-periods of some of the four data sets indicating thus the presence of day-of-week effects in volatility as well. It may further be noted that in most of the cases the day effects in h_t were not the same as in r_t, the exception being Sub-period III of BSESENSEX. It is thus observed that predictability in expected returns is accompanied by a change in the conditional variance in the Indian stock market, but the nature of dependence in mean and in volatility being different, investors ought to consider jointly predicting stock market returns and volatility. Exploitation of both sources of predictability is expected to result in investment strategies that outperform strategies which exploit only one source.

(v) It may also be observed that in one case *viz*., Sub-period I of BSE 100, some polynomial of recursive residuals i.e., $\hat{r}_{t-1}^3$ was found to be significant in the mean part of the model. This means that there are some nonlinearities yet to be captured in the actual model.

(vi) In none of the sub-periods, risk factor has been found to be significant indicating that time-varying risk premia has no role in explaining inefficiency in the Indian capital market.

3.3.4 Detecting nonlinear dependencies

By following the procedure proposed in Section 3.2, we have now reached the step where we are required to detect nonlinear dependencies in the residuals. This detection is indeed important since there are growing evidence that stock prices often show complex regularities. We have, therefore, applied BDS test for detecting i.i.d. system from nonlinear dependencies. The results of this exercise have been presented in Table 3.6.

The first column of the Table 3.6 gives the values of the distance, ξ, measured in terms of half (0.5) and one (1) times the standard deviation of linearly filtered data used in the study; the values of the number of embedding dimensions *m* are given in column 2. The values of BDS test statistics for all the sub-periods are shown in rest of the columns. The series are examined up to ten dimensions when the number of observations exceed 500. For the observations less than or equal to 500, we use embedding dimensions upto 5. It is known that acceptance of the null hypothesis of i.i.d.

observations indicates that the observations are purely random meaning thereby that there are no nonlinear dependencies of any kind. In that event we could conclude that the behaviour of the series has no exploitative value. As for computations, we have used the standardized residuals $\hat{\varepsilon}_t / \sqrt{\hat{h}_t}$, obtained in equation (3.6). Obviously, these are filtered off all linear dependencies of r_t. Also, the appropriate critical values for BDS test have been obtained from Brock *et al.* (1991). As regards the findings, we have found that the BDS test statistic values for all the four stock indices are significant as compared to the nonstandard critical values at standard levels of significance for many of the values of ξ / σ and m considered for this study. These computations indicate, as displayed in Table 3.6, that in most cases, particularly for higher values of m (m > 5), the null of i.i.d. residuals is rejected. Thus, we can conclude on the basis of our analysis that GARCH has been found to be somewhat inadequate in capturing all the nonlinear dependencies in the series.

Table 3.6 : GARCH Adjusted BDS Test Statistic Values

		BSESENSEX		
ξ / σ	m	Sub-period I	Sub-period II	Sub-period III
0.5	2	-0.260[a]	0.145 [a]	0.370[a]
0.5	3	5.554	2.962	2.634
0.5	4	27.700	19.169	33.184
0.5	5	90.757	-12.384	39.113
0.5	6	-7.969	-8.719	-11.842
0.5	7	-5.900	-6.457	-8.872
0.5	8	-4.526	-4.956	-6.889
0.5	9	-3.564[a]	-3.906	-5.494[a]
0.5	10	-2.864[a]	-3.142	-4.471[a]
1	2	1.642*	-2.080	0.241 [a]
1	3	0.872[a]	1.634*	0.085 [a]
1	4	1.933	1.998	1.303*
1	5	5.070	12.872	0.649 [a]

1	6	-9.162	-10.200	12.055
1	7	-6.871	-7.666	-78.310
1	8	-5.339	-5.972	-7.535
1	9	-4.260	-4.777	-6.059
1	10	-3.469	-3.900	-4.539

		BSE 100		NIFTY	
ξ/σ	m	Sub-period I	Sub-period II	Sub-period I	Sub-period II
0.5	2	1.220 [a]	-2.474	-0.538 [a]	2.760
0.5	3	4.067	-5.234	-8.473	3.155
0.5	4	2.540	-5.162	-4.235	-8.277
0.5	5	-1.488	-1.436 [a]	-2.414[a]	-14.380
0.5	6	-1.026 [a]	-1.023 [a]		-10.213
0.5	7	-7.647	-7.660		-7.629
0.5	8	-5.907	-5.946		-5.905
0.5	9	-4.685 [a]	-4.739 [a]		-4.692[a]
0.5	10	-4.547 [a]	-8.363		-3.805[a]
1	2	1.786	-0.898 [a]	-4.082[a]	2.087
1	3	1.618	-0.027 [a]	-8.181	2.935
1	4	4.121	-1.994	-6.284	3.006
1	5	1.032	-4.626	-3.859	-4.253
1	6	-1.174	-7.221		-1.794
1	7	-8.850	-8.363		-8.479
1	8	-6.916	-6.544		-6.621
1	9	-5.550	-5.258		-5.308
1	10	-4.547	-4.313		-4.343

		DOLLEX	
ξ/σ	m	Sub-period I	Sub-period II
0.5	2	-3.420	-3.584
0.5	3	-3.054	-2.360
0.5	4	-17.177	-9.036
0.5	5	-11.191	-13.837
0.5	6	-7.873	-9.850
0.5	7	-5.823	-7.345
0.5	8	-4.463	-5.722
0.5	9	-3.511 [a]	-4.558 [a]
0.5	10	-2.818 [a]	-3.706 [a]
1	2	-1.248[a]	-0.804 [a]

1	3	-0.843 [a]	-0.059 [a]
1	4	-1.045 [a]	-1.955
1	5	9.192	-5.867
1	6	18.809	-2.591[a]
1	7	-6.516	-8.083
1	8	-5.053	-6.322
1	9	-4.022	-5.078
1	10	-3.2667	-4.162

Note : The values of BDS test statistic from standardized residuals are compared with the simulated values given in Brock et al. (1991).
*Values with superscript 'a' indicate non-significance at 5% level. While the values with *indicate significant at 5% level, all others are significant at 1% level of significance.*
ξ*, m and* σ *stand for distance, embedding dimension and the standard deviation of the linearly filtered data, respectively.*

3.3.5 Testing for the presence of dynamics of higher-order moments

Finally, we have carried out the last step of our proposed method of analysis. What we attempt at here is to find out whether there is dynamics in the higher order moments (say, third or fourth) so that the remaining dependencies could be explained. It may be worthwhile to mention that while there are now models that can capture dependencies in higher moments such as skewness and kurtosis (see, for example, Hansen (1994) and Harvey and Siddique (1999) for details of such models), our limited goal, at the end of this chapter, is merely to examine if there are any significant dynamics in the higher-order (to be specific, third and fourth- order residuals) and hence, no further complete modelling with these residuals is being undertaken.

To this end, we note that testing on the higher-order moments is appropriate only after ensuring that all the lower-order moments are completely free from dependence. Accordingly, we have considered dynamic relations based on $\hat{\hat{\varepsilon}}_t^{\,3}$ and $\hat{\hat{\varepsilon}}_t^{\,4}$ where $\hat{\hat{\varepsilon}}_t = \hat{\varepsilon}_t \big/ \sqrt{\hat{h}_t}$ is the standardized residual. From these estimated equations, we have found that no significant dependencies in these moments exist for any of the sub-periods of BSESENSEX and BSE 100 series. But for the other two series *viz*., NIFTY and DOLLEX we have found, as presented in equations (3.21) through (3.24), that standardized residuals have third- and/or fourth-order dependencies on their own lagged values for some of the sub-periods, implying that dynamics in higher-order moments are significant in

NIFTY: Sub-period I $\hat{\hat{\varepsilon}}_t^3 = 0.117\,\hat{\hat{\varepsilon}}_{t-4}^3$ (3.21)

(1.990)*

$\hat{\hat{\varepsilon}}_t^4 = 0.140\,\hat{\hat{\varepsilon}}_{t-2}^4 + 0.140\,\hat{\hat{\varepsilon}}_{t-3}^4$ (3.22)

(2.413)* (2.409)*

DOLLEX: Sub-period I $\hat{\hat{\varepsilon}}_t^3 = 0.440\,\hat{\hat{\varepsilon}}_{t-1}^3 - 0.207\,\hat{\hat{\varepsilon}}_{t-2}^3 + 0.081\,\hat{\hat{\varepsilon}}_{t-3}^3$ (3.23)

(13.976)** (6.127)** (2.571)**

$\hat{\hat{\varepsilon}}_t^4 = 0.598\,\hat{\hat{\varepsilon}}_{t-1}^4 - 0.344\,\hat{\hat{\varepsilon}}_{t-2}^4 + 0.187\,\hat{\hat{\varepsilon}}_{t-3}^4 - 0.07\,\hat{\hat{\varepsilon}}_{t-4}^4$ (3.24)

(18.983)** (9.494)** (5.168)** (2.334)*

[The absolute values of *t*-statistic are indicated within parentheses. * indicates significance at 5% level, ** indicates significance at 1% level.]

explaining inefficiencies in the Indian stock market, as represented by returns on these two series. While the findings on these two series of NIFTY and DOLLEX explain why the null hypothesis of i.i.d. residuals got rejected for these two series, the explanations for the same conclusion in case of BSESENSEX and BSE 100 series may be found in terms of other kinds of dependencies involving the standardized residuals.

3.4 Conclusions

In this chapter we have proposed a systematic approach towards studying stock market efficiency with an aim to identify the factors that make stock returns predictable, after ensuring appropriate specification of first and second order conditional moments. In this approach inefficiency has been defined to include nonlinear dependence in the returns as well. The consideration for appropriate specification is due to the fact that standard tests for conditional heteroscedasticity presume proper specification of the conditional mean, and hence any misspecification of the conditional mean as well as of the conditional variance could lead to misleading inferences about the model on returns and consequently on the predictability or otherwise of the stock market.

This modelling approach has been applied for the Indian stock market. Incorporating short-term interest rate through call money rate, risk by conditional heteroscedasticity, 0-1 dummies for the day-of-the week effects in addition to lagged values of return in the conditional mean, and assuming GARCH specification and 0-1 dummies to represent daily level seasonality in conditional heteroscedasticity, we have applied tests like automatic variance ratio test, Andrews's test, test based on recursive residuals and BDS test, and concluded that there is statistical evidence in favour of two, structural breaks- one in middle of 1992 and the other in late 1996 – which is consistent with the recent history of the Indian stock market, and that the predictability in the Indian stock market represented by four standard daily indices *viz*., BSESENSEX, BSE 100, NIFTY and DOLLEX, can be attributed to serial correlation, nonlinear dependence, day-of-the week effects, parameter instability, conditional heteroscedasticity (GARCH), daily level seasonality in volatility and call money rate (in some sub-periods of some indices only). Further, we have found that the null of i.i.d. residuals was rejected by the BDS test even in the case of standardized residuals of the properly specified models, leading us to conclude that incorporating second order dependence through GARCH is not adequate to capture all potential nonlinearities in the returns for all indices. We have also observed that the remaining nonlinearities could be attributed to the existence of some dynamics in the higher order moments.

CHAPTER 4

Studying Monthly Stock Returns with Macro and Financial Variables: A Predictive Regression Approach

4.1 Introduction

In financial economics one important empirical regularity is that asset returns can be predicted by a set of macroeconomic and financial variables. As changes in these variables contain important information for investors it is hypothesized that the stock market participants take these factors into account for estimating appropriate discount rate and the expected flow of dividends from stocks. The speed and accuracy with which this information is reflected into stock returns is crucial for the understanding of efficient functioning of the stock market.

However, one should be careful in drawing conclusions regarding efficiency or otherwise of the stock market from such evidencel on predictability of stock returns as it has been argued in the recent literature that pricing irregularities and even predictable pattern that appear over time are, in effect, temporary in nature, and hence these may not be able to refute the hypothesis that the stock market is efficient (Malkiel (2003)).

Several recent studies have shown strong influence of several macroeconomic variables as well as some financial variables (ratios)[1] on the established stock markets,

---

[1] By financial variables here we mean financial ratios only; all other variables, in this study, are being taken as macrovariables. For the sake of convenience in expression, both the terms i.e., financial variables and financial ratios are being used concurrently.

mostly in the U.S. and, to a lesser extent, in Europe and Japan. Some of the important studies on this topic are due to Fifield *et al.* (2000), Lovatt and Parikh (2000), Nasseh and Strauss (2000), Hondroyiannis and Papapetrou (2001), and Lu *et al.* (2001). Earlier, the arbitrage pricing theory developed by Ross (1976), and Chen *et al.* (1986) also showed, in the context of industrialized countries, that economic variables have a systematic effect on stock market returns in the sense that economic forces affect discount rates, firm's ability to cash flows and future dividend payouts. Further, there are some studies which have examined the relationship between stock returns and various macroeconomic variables across countries (see, for instance, Solnick (1984), Asprem (1989), Wasserfallen (1989), Ferson and Harvey (1993) Conover *et al.* (1999), and Durham (2001)). In contrast, very few such studies have been carried out on stock markets of the emerging market economies, and these references are: Mookerjee and Yu (1997), Ibrahim (1999), Chong *et al.* (2001) and Wongbangpo and Sharma (2002).

Although most of the studies cited above have found the evidence that stock returns are predictable using macroeconomic and financial variables, the empirical evidencel are not uniform in providing support to stock return predictability using these variables. For instance, while some studies have found that certain macroeconomic variables have significant effects in explaining returns of some stock indices, others have found no such evidencel i.e., no significant effects of the same variables have been observed for some other stock returns (see, for instance, Balvers *et al*. (1990) and Flannery and Protopapadakis (2002) for details of such findings). Further, it has been observed that predictive ability of some macrovariables with respect to equity returns is quite uneven overtime; Durham (2001), for example, has found this for some variables concerning monetary policy. The issue of data mining is also relevant in this context since numerous studies have examined the predictive ability of several macroeconomic variables in the literature, and hence, this makes it difficult to determine the particular macrovariables which are appropriate for explaining stock returns. It is also noteworthy that apart from macrovariables, financial variables (ratios) like price-earning ratio, price-book value ratio and dividend yield have also been found to be the predictors of stock returns (Campbell and Shiller (1988a, b, 1998), Fama and French (1988), Ang and Bekaert (2001)).

It is because of these facts that we first make an attempt to find the predictive ability of the set of relevant macroeconomic and financial variables for returns based on BSESENSEX, BSE 100 and NIFTY.[2] This is done by using in-sample and out-of-sample tests of return predictability. Since in-sample forecasts are those generated for the same set of data that was used to estimate the parameters of the model, it is expected that such forecasts of any model would be relatively good. Hence, a sensible approach is to use out-of-sample forecasts for model evaluation. Based on the findings of this predictive exercise, we decide on the set of macro and financial variables for this study. And then, as in Chapter 3, we study the proposed relationship from consideration of stability and appropriate specification. Thus, in this chapter, we examine the predictability aspect of Indian stock returns using macroeconomic variables and financial variables with the aim to provide an adequate econometric model so that the same can be used for forecasting and other related purposes. Obviously, the other important characteristic of stock returns *viz.,* volatility, is also incorporated in the modelling framework through GARCH

---

[2] Since the data for DOLLEX were not available beyond December 2000, this study could not be done for this particular stock index.

formulation for the errors. In fact, other observed characteristics like seasonal effects etc. are also duly included in the mean specification.

For the purpose of this study, we have taken a set of fourteen relevant macroeconomic variables comprising domestic industrial production (IP), broad money supply (MS), consumer price index (CPI), nominal exchange rate (FRX), two short term interest rates- short term treasury bill rate (TBS) and weighted call money rate (WCALL), long term interest rate (TBL), foreign exchange reserve (FOREXR), domestic oil price (OILD), foreign direct investment (FDI), foreign institutional investment (FII), NASDAQ composite index of the U.S. stock market (NSD), fiscal deficit of the central government (FD) and term spread (SR) and three financial variables *viz.,* price-earnings ratio (P/E), price-book value ratio (P/BV) and dividend yield (DY)[3]. These sixteen variables have been chosen on the basis of evidence of their significant effects in stock returns of other stock markets, as mentioned earlier. Some of the variables such as industrial production, money supply and consumer price index etc. have been included as basic variables. The variables like foreign direct investment, NASDAQ composite are included for the purpose of examining the roles of global factors on Indian stock market. It may be mentioned that general indices like the MSCI World index and FTSE 100 which cover many countries, could have been used as a proxy for the world market, we state that our choice of NASDAQ has been based on the fact that the US dominates the world equity market, and further that the domestic equity markets of emerging economies like India are greatly influenced by the developments in the US equity market. Moreover, insofar as available empirical evidence of such influences on

[3] We have considered these three financial variables since data are not available for other financial variables such as payout ratio, Fed q etc.

Indian stock returns are concerned, we came across only two papers, namely Bhattacharya and Samanta (2003) and Hansda and Ray (2002), where it has been found that NASDAQ influences BSESENSEX and hence we thought that given this evidence, it would be quite appropriate to use NASDAQ for our study. Since data for all the time series on macro variables of India are available at monthly level, not at any other higher frequency, these macro and financial variables along with the three return series have been taken at monthly level value covering the period April 1996 to December 2002. The first month i.e., April 1996 has been chosen considering the availability of the data sets and the observed first break in the series, as obtained in Chapter 3 after the liberalization of the Indian economy in 1992.

While choosing these variables, it is also important to understand the economic meanings of the underlying relationships since financial theory asserts that movement in stock prices is related to macroeconomic and financial variables. To start with, an increase in current real activity increases demand on existing capital stock, which ultimately induces increased capital investment in the future, and the stock market is very likely to anticipate this (see Gallinger (1994), for details). Money supply has a direct effect on stock prices by changing liquidity. Further, as noted by Musilek (1997), money supply also has an indirect effect on stock prices through corporate dividends by increasing or decreasing interest rates. Stock prices are also influenced by changes in interest rates. Since interest rate is an opportunity cost of holding stock, an increase in interest rate is likely to lead to a substitution effect between stocks and other interest bearing assets. It is, therefore, expected that as interest rate declines stock price would rise (*cf.* Musilek (1997)). Inflation rate which is defined as the first difference in consumer price index as percentage of previous periods value also affects the stock market through the output link, as advocated by Fama (1981). There is also a strong evidence on the causal influence of exchange rate on stock prices (see, for instance, Abdalla and Murinde (1997) and Granger *et al.* (2000)). The main implication is that changes in exchange rate affect firm's exports and also the cost of imported goods and production inputs and thus ultimately affect stock prices. In recent times, the linkage between stock return and change in oil price has also been examined (see, for example, Hondroyiannis and Papapetrou (2001)).

In view of financial liberalization in several emerging economies the interdependence among the stock markets in the world has increased (*cf.* John (1993)). The presence of strong economic ties and policy coordination can indirectly link the stock prices of different countries over time (see Choudhry (1997), for example). Jeon and Chiang (1991) have stated that international linkage between different equity markets can also be due to recent deregulation and liberalisation of different markets, improvement and development of communications technology, innovations in financial products and services, increase in the international activities of multinational corporations etc. This international integration, in particular of emerging stock markets, with the rest of the world has also led to link-up of the domestic stock markets with the world financial and economic variables. It is also pertinent to note that foreign investment in the developing countries is now playing a crucial role in restructuring of these economies. There is a growing body of research that studies the hypothesis focussing on the effect of capital flow on stock return in emerging economies (see, for example, Froot *et al.* (2001) and Clark and Berko (1997)). Both Froot *et al.* and Clark and Berko have found that increase in capital flow raises stock market prices, but the studies disagree on whether the effect is temporary or permanent. In view of growing importance of information technology in Indian market, and/or globalization of the Indian capital market we have considered the technology-led NASDAQ composite index, a broad based stock index of the US including over 3000 securities, as an explanatory variable for explaining variations in Indian stock returns. Another important variable i.e., fiscal deficit of the Central government which has drawn keen attention of the analysts since liberalisation, has also been considered as a predictor variable for Indian stock returns.

As regards the effects of financial variables such as dividend yield, its variability can be attributed to the variation of expected cash flow growth (Ang and Bekaert (2001)). In general, financial ratios can predict firm's ability to future cash flows and thus overall stock return of the economy.

As already mentioned, the focus in this chapter is to examine the influence of macroeconomic variables and financial ratios in explaining variations in stock returns of the Indian stock market using in-sample and out-of-sample forecasting techniques with due emphasis on specification and nonlinear dependence in errors as captured through GARCH model. While the in-sample analysis carried out here employs a predictive regression framework, the out-of-sample forecasts have been analysed using recently developed statistical tests by McCracken (2004) and Clark and McCracken (2001) which are some variants of those proposed by Diebold and Mariano (1995) Harvey *et al.* (1998), respectively (see Rapach *et al.*(2005) for an application of these tests). Once the macroeconomic variables with potential predictive ability of returns are selected by these methods, an econometric model is specified by including those variables along with appropriate seasonal dummies. The specification test for the conditional mean is then done, as in the preceding chapter, by the recursive residual based test, due to Lumsdaine and Ng (1999). Modelling of conditional variance is done by assuming GARCH process for the error. Finally, BDS test due to Brock *et al.* (1996) is applied to find if there is any nonlinear dependence still remaining in the series. This chapter proceeds as follows: The next section describes the methodological approach. Section 4.3 discusses the empirical results. This chapter ends with some concluding remarks in Section 4.4.

4.2 Methodological approach

This section describes the methodological approach that has been followed in this chapter. To begin with, the stationarity of all the three monthly stock indices and the sixteen variables are checked by applying the standard unit root tests such as the ADF and PP tests. If any series is found to be nonstationary, then the series is made stationary by taking differences. Unit root test is again carried out to confirm stationarity of the series. Since we have monthly time series for all the macro and financial variables, the data pertaining to these variables are likely to have seasonal effects, and hence for meaningful data analysis, deseasonalization of all the series is necessary. As such, some idea about the nature of seasonality latent in a given time series can be made by plotting the series against the months. However, in the general case of dealing with seasonality there are, apart from some specific filters, two broadly-accepted procedures by which deseasonalization of any series can be carried out. These are - seasonal differencing and use of dummies. In view of the sample size for each series being rather moderate, we have deseasonalized the time series of macroeconomic and financial variables by using appropriate seasonal dummies since seasonal differencing with monthly data would result in exclusion of first 12 observations.

Once all these stationary (in the trend sense) series are thus deseasonalized then the stability of the three return series are checked by applying Andrew's (1993) procedure for testing the presence of structural break(s); and then the breakdates are estimated, based on least-squares procedure (Bai (1994, 1997a, 1997b)) under a fixed specification *viz.,* an AR(1) model for the stationary return.

4.2.1 In-sample predictability

As already mentioned, we have followed the predictive regression framework used in the extant literature for in-sample analysis of return predictability. The predictive regression model takes the following form in our analysis:

$$r_{t+1}^{k} = \alpha + \beta z_t + \delta\ r_t + \varepsilon_{t+1}^{k} \quad , \tag{4.1}$$

where r_t, as before, represents the continuously compounded rate of return for holding the stock for one month, $r_{t+1}^{k} = r_{t+1} + ... + r_{t+k}$ is the rate of return for holding the stock from period $t+1$ to $t+k$, z_t is any particular macroeconomic/financial variable expected to have potential predictive ability for future stock returns and ε_{t+1}^{k} is a disturbance term. The relevant null hypothesis in this case states that the macroeconomic/financial variable z_t has no predictive power for future stock returns (i.e., $\beta = 0$) whereas the alternative says that z_t does have predictive ability for future returns (i.e., $\beta \neq 0$). Our model also includes a control variable r_t, a lagged return term in equation (4.1). The in-sample predictive regression model assesses the predictive ability of z_t by examining the $t-$statistic corresponding to $\hat{\beta}$, the OLS estimate of β in equation (4.1), as well as by the goodness-of-fit measure, R^2. There are two potential econometric problems associated with estimating a predictive regression model like the one in equation (4.1). These are small sample bias (Mankiw and Shapiro (1986) and Stambaugh (1986,1999)) and overlapping observations when $k>1$(Richardson and Stock (1989)). To deal with the problem of overlapping observations, Newey and West (1987) suggested standard errors which are robust to heteroscedasticiry and serial correlation in the disturbance term. In our analysis we have followed the Bartlett Kernel and a lag truncation parameter of $[1.5k]$, where [.] denotes the nearest integer function when calculating Newey and West standard errors. However, even when robust standard errors are used to compute t - statistic there is the potential problem for size distortions when basing inferences on standard asymptotic distribution theory (Goetzmann and Jorion (1993), Nelson and Kim (1993) and Kirby (1997)). In view of this potential size distortions, the recent literature of return predictability has suggested bootstrap procedure which has been followed in our analysis as well.

4.2.2 Out-of-sample predictability

In addition to in-sample tests of return predictability, we also perform out-of-sample tests based on the recursive scheme. The total sample of n observations are divided into in-sample and out-of-sample portions, where the in-sample portion spans the first n_1 observations for r_t and z_t, and the out-of-sample portion spans the last $n-n_1$ observations. The first out-of-sample forecast for the 'unrestricted' model (4.1) is generated using the following recursive scheme. In this procedure, we first apply the ordinary least square (OLS) method on the available data upto period n_1. Suppose the relevant OLS estimates for α, β and δ in equation (4.1) are $\hat{\alpha}_{1,n_1}$ $\hat{\beta}_{1,n_1}$ and $\hat{\delta}_{1,n_1}$. We now construct a forecast error as $\hat{\varepsilon}^k_{1,n_1+1} = r^k_{n_1+1} - \hat{r}^k_{1,n_1+1}$, where $\hat{r}^k_{1,n_1+1} = \hat{\alpha}_{1,n_1} + \hat{\beta}_{1,n_1} z_t + \hat{\delta}_{1,n_1} r_t$. The initial forecast for the "restricted" predictive model is generated in a similar manner, except that we set $\beta = 0$ in equation (4.1), accordingly denote the forecast error corresponding to the restricted model as $\hat{\varepsilon}^k_{0,n_1+1} = r^k_{n_1+1} - \hat{r}^k_{0,n_1+1}$.

We then generate a second set of forecasts, updating the above procedure by one period i.e., by using data available through the period $n_1 + 1$. The corresponding unrestricted and restricted model forecasts errors are $\hat{\varepsilon}^k_{1,n_1+2}$ and $\hat{\varepsilon}^k_{0,n_1+2}$, respectively. This process is repeated through the end of the available sample, leaving us with two sets of $n - n_1 - k + 1$ recursive forecast errors-one each for the unrestricted and restricted models i.e., $\{\hat{\varepsilon}^k_{1,t+1}\}_{t=n_1}^{n-k}$ and $\{\hat{\varepsilon}^k_{0,t+1}\}_{t=n_1}^{n-k}$).

In order to test whether inclusion of z_t improves the out-of-sample forecasts of r^k_{t+1} relative to the benchmark model AR(1), Theil's U, the ratio of the unrestricted model forecast root-mean-squared error (RMSE) to the restricted model forecast RMSE, is used. If U <1, then it indicates predictive ability of z_t. For a formal statistical test concerning superiority of the unrestricted regression model forecasts over the restricted model forecasts, the following two test statistics – McCracken's (2004) $MSE - F$ and the $ENC - NEW$ statistic due to Clark and McCracken (2001) – are used. The $MSE - F$ statistic is used to test the null hypothesis that the unrestricted model forecast mean squared error (MSE) is equal to the restricted model forecast MSE against the one-sided (upper-tail) alternative hypothesis that the unrestricted model forecast MSE is less than the restricted model MSE. The McCracken (2004) $MSE-F$ statistic is given by

$$MSE - F = (n - n_1 - k + 1)\overline{d} / M\hat{S}E_1 \tag{4.2}$$

where $\overline{d} = M\hat{S}E_0 - M\hat{S}E_1$, and

$$M\hat{S}E_i = (n - n_1 - k + 1)^{-1} \sum_{t=n_1}^{n-k} (\hat{\varepsilon}_{i,t+1}^{k})^2, \; i = 0,1 .$$

A significant $MSE-F$ indicates that the unrestricted model forecasts are statistically superior to those of the restricted model. For $k=1$,the $MSE-F$ statistic has a non-standard limiting distribution that is pivotal and a function of stochastic integers of Brownian motion (McCracken (2004)). But, for $k>1$,the $MSE-F$ statistic has a non-standard and non-pivotal limiting distribution and for this a bootstrap procedure is suggested in the lines of Kilian (1999) and Clark and McCracken (2004)).

Another out-of-sample statistic used in our analysis is the $ENC-NEW$ statistic which is related to the concept of forecast encompassing based on optimally constructed composite forecasts. The $ENC-NEW$ statistic due to Clark and McCracken (2001), takes the following form:

$$ENC - NEW = (n - n_1 - k + 1)\overline{c} / M\hat{S}E_1 \tag{4.3}$$

where $\overline{c} = (n - n_1 - k + 1)^{-1} \sum_{t=n_1}^{n-k} c_{t+1}^{k}$ and $c_{t+1}^{k} = \hat{\varepsilon}_{0,t+1}^{k}(\hat{\varepsilon}_{0,t+1}^{k} - \hat{\varepsilon}_{1,t+1}^{k})$.

Under the null hypothesis, the weight attached to the unrestricted model forecast in the optimal composite forecast is zero implying that the restricted model forecasts encompass the unrestricted model forecasts. But under the one-sided (upper-tail) alternative, the weight attached to the unrestricted model forecast in the optimal composite forecast is greater than zero implying that the restricted model forecasts do not encompass the unrestricted model forecasts. The limiting distributional issues of $ENC-NEW$ statistic are similar to those of $MSE-F$ statistic.

4.2.3 The bootstrap procedure

In view of the reasons stated above, we have used bootstrap procedure following Nelson and Kim (1993), Mark (1995), Kothari and Shanken (1997), Kilian (1999), and Rapach *et al.* (2005) for making proper inferences based on both in-sample and out-of-sample predictions. Under the null hypothesis of no predictability, the data are generated by the following system:

$$r_t = a_0 + a_1 r_{t-1} + \varepsilon_{1t}, \tag{4.4}$$

$$z_t = b_0 + b_1 z_{t-1} + ...+ b_p z_{t-p} + \varepsilon_{2t}, \tag{4.5}$$

where the disturbance vector $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t})'$ is assumed to be i.i.d. with covariance matrix Σ. The equations (4.4) and (4.5) are estimated by OLS after using Akaike's information criterion for selecting the value of lag, p, the OLS residuals $\{\hat{\varepsilon}_t = (\hat{\varepsilon}_{1,t}, \hat{\varepsilon}_{2,t})'\}_{t=1}^{n-p}$. For generating a series of disturbances for the pseudo-sample, we randomly draw (with replacement) $n+100$ times from the OLS residuals $\{\hat{\varepsilon}_t\}_{t=1}^{n-p}$, leading to a pseudo-series of disturbance terms $\{\hat{\varepsilon}_t *\}_{t=1}^{n+100}$. Using this and the OLS estimates of the parameters in equations (4.4) and (4.5), and setting the initial observations for r_{t-1} and $z_{t-1},......,z_{t-p}$ equal to zero in the above two equations, we can build up a pseudo-sample of $n+100$ observations for r_t and z_t, $\{r_t *, z_t *\}_{t=1}^{n+100}$. In order to randomize the initial observations we drop the first 100 transient start-up observations, thus leaving us with pseudo-sample of n observations, matching the original sample. We calculate the t-statistic corresponding to β in the in-sample predictive regression model (4.1), and the two out-of-sample statistics given in (4.2) and (4.3) for each pseudo-sample and this is repeated for 1000 times[4]. This process gives an empirical distribution for each of in-sample and out-of-sample statistics. For each statistic, the p-value is simply the proportion of the bootstrapped statistics that are greater than the statistic computed from the original sample.

As our analysis on stock returns predictability deals with macroeconomic and financial variables it is likely that the analysis may suffer from the data-mining problem. While data mining is generally considered to be a serious problem for in-sample tests of predictability, it is not so much for out-of-sample tests. Inoue and Kilian (2005) have recently argued that data mining is serious for both when standard critical values are used. They have, in fact, suggested that the key to controlling data mining is the use of appropriate critical values for both in-sample and out-of-sample predictability tests. Let us first assume that J different macroeconomic and financial variables are considered in the predictive regression framework where each of the variables is a candidate predictor for stock return. The bootstrap process analysed above implicitly assumes that each macro and financial variable is considered in isolation. However, in reality we focus on those variables that give the "best" results. In order to take care of data mining problem for testing predictability, the null hypothesis (H_0) and the alternative hypothesis (H_1) considered by Inoue and Kilian (2005) are as follows:

[4] Since the interest, in this study, is only in the tails of the simulated distribution rather than the mean, it would have been preferable to have far more replications. However, it is primarily because of constraint on computational time that this could not be done.

$H_0: \beta_j = 0 \;\; \forall\, j$ and $H_1: \beta_j \neq 0$ for some j where β_j is the coefficient corresponding to the jth variable z_j as in equation (4.1). The in-sample test statistic used by this

method is $\max_{j \in \{1,.....,J\}} | t_{\hat{\beta} j} |$, where $t_{\hat{\beta} j}$ is the t-statistic corresponding to β_j. The out-of-sample test statistics are the maximal $MSE - F$ and maximal $ENC - NEW$ values. Inoue and Kilian have derived the asymptotic distribution for these statistics and found that the limiting distributions are data-dependent, and hence, they have recommended that bootstrap procedures be used in practice. The bootstrap procedure followed here is similar to that in Rapach *et al.* (2005). Assuming, as before, that there are J macroeconomic and financial variables in the predictive regression framework where each of the variables is a candidate predictor for stock return. The equation (4.5) of the bootstrap procedure discussed above is now modified to take care of all the J variables of the candidate predictors as follows:

$$z_{1,t} = b_{1,0} + b_{1,1} z_{1,t-1} + ... + b_{1,p1} z_{1,t-p1} + \varepsilon_{1,2,t}$$

$$z_{2,t} = b_{2,0} + b_{2,1} z_{2,t-1} + ... + b_{2,p1} z_{2,t-p1} + \varepsilon_{2,2,t}\,.$$

.

.

$$z_{J,t} = b_{J,0} + b_{J,1} z_{J,t-1} + ... + b_{J,pJ} z_{J,t-pJ} + \varepsilon_{J,2,t}$$

where the disturbance vector $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{1,2,t}, ..., \varepsilon_{J,2,t})'$ is assumed to be i.i.d. with covariance matrix Σ. Following the same procedure as discussed earlier, for each pseudo sample, in-sample $t-$statistic corresponding to β_j in the predictive regression model is calculated and the two out-of-sample statistics for each of the $z_{j,t}*$ $(j = 1,..., J)$ in turn are also calculated. Then the largest and the smallest t-statistics, as well as the maximal $MSE-F$ and maximal $ENC-NEW$ statistics are calculated and stored. For each maximal out-of-sample statistic , the 900th, 950th, and 990th values serve as 10%, 5% and 1% critical values when empirical distributions are ordered. On the other hand, for the in-sample t-statistic, the 950th, 975th, and 995th values of the empirical distribution for the largest t-statistic stand for the 10%, 5% and 1% upper-tail critical values of the $\max_{j\in\{1,\ldots,J\}} |t_{\hat{\beta}_j}|$ statistic whereas the 50th, 25th, and 5th values of the empirical distribution for the smallest t-statistic stand for the 10%, 5% and 1% lower-tail critical values of the same $t-$statistic.

Apart from analysing the predictive ability of each macro variable and financial ratios following the procedure as discussed above, and which may be termed as specific to general approach, we also employ a variable selection procedure which is known as general to specific approach. In the context of building up forecast model for stock returns, this has recently been used by Clark (2004). In our efforts towards finding the best forecasting model for stock returns, we also apply this approach for our study. This approach is based on the following predictive regression model in (4.6) in which all the J candidate macro and financial variables are considered in the specification.

$$r_{t+1}^{k} = \alpha + \beta_1 z_{1,t} + \ldots\ldots + \beta_J z_{J,t} + \delta\ r_t + \varepsilon_{t+1}^{k} \tag{4.6}$$

This model is estimated using data only from the in-sample portion of the total sample. We examine each of the t-statistics corresponding to the $z_{j,t}\ j=(1,\ldots\ldots,J)$ variable in (4.6). If the smallest $t-$statistic (in absolute value) is greater than or equal to 1.645, we select the model that includes all the J variables. If the smallest t-statistic is less than 1.645, then that variable is excluded and the model comprises $J-1$ variables. This process is continued until all the $z_{j,t}$ variables included in the model have significant t-statistic values. When the best forecasting model includes at least one $z_{j,t}$ variable, we compare the out-of-sample return forecasts generated by the selected model to the out-of-sample forecasts generated by the benchmark model. Out-of-sample forecasts are then formed recursively and these forecasts from the competing models are compared using the $MSE-F$ and $ENC-NEW$ statistics.

4.2.4 The final model

Following these model selection procedures, we identify the macroeconomic and financial variables having predictive ability for the returns on three Indian stock indices *viz.,* BSESENSEX, BSE 100 and NIFTY. Once these variables (say, $\widetilde{J}(\leq J)$ in number) are identified we specify the appropriate econometric model for Indian stock returns by incorporating those variables along with the past values of returns for capturing serial correlation as well as the seasonal (monthly) dummies for including any deterministic seasonal pattern in returns. The model thus take the following form in which volatility is also duly considered through a GARCH specification for ε_t.

$$r_t = \sum_{i=1}^{\widetilde{m}} \gamma_i r_{t-i} + \sum_{l=1}^{12} w_l D_l + \sum_{j=1}^{\widetilde{J}} \beta_j z_{j,t-1} + \varepsilon_t \tag{4.7}$$

where $\varepsilon_t | \psi_{t-1}$ is assumed to follow $N(0,h_t)$, h_t represents conditional variance at time t, as given by the GARCH specification (3.6) in Chapter 3. D_l's l=1,2,….,12 denote the monthly dummies to capture month-of-the-year effect in returns, $z_{j,t}$ is the seasonally adjusted stationary macroeconomic/financial variable as obtained by applying the in-sample and out-of-sample predictive procedures suggested above ($j = 1,2,....,\widetilde{J}$) , ψ_{t-1} is the information set at time $t-1$, and $\widetilde{m}$ is the appropriate lag value of r_t capturing its autocorrelation. As stated in Chapter 3, $\widetilde{m}$ is determined by Hall's (1994) procedure and other diagnostics tests such as Ljung-Box test statistic.

Combining all these variables appropriately into one vector, say X_t and the coefficients into another vector, ρ, equation (4.6) can be written, in vector notation, as

$$r_t = X_t^{'} \rho + \varepsilon_t . \tag{4.8}$$

In order to take care of mispecification in the conditional mean, we have followed the recursive residual-based test suggested by Lumsdaine and Ng (1999), which has already been described in Chapter 3. Following the procedure similar to that in the preceding chapter, we also test for the adequacy of the GARCH specification for the conditional second-order moment. Once the conditional mean and the conditional variance are properly specified we finally carry out BDS test due to Brock *et al.* (1996) to find if the null of i.i.d. property for residuals in (4.8) is rejected or not.

4.3 Empirical findings

In this section we report and discuss the results on predictability of monthly returns on Indian stock indices by applying the methodology stated in Section 4.2. The data for each series cover the period April 1996 to December 2002. The choice of the starting month April 1996 has been dictated by the fact that data for some of the macrovariables are available from this time point only. Further, since for NIFTY, the data on financial ratios are available from January, 1999 only so this exercise for NIFTY has been done out with macrovariables only. To begin with, we test the stationarity of all the variables used in our analysis. As stated, in addition to Indian stock returns based on the monthly series of BSESENSEX, BSE 100 and NIFTY, our analysis includes fourteen standard macroeconomic variables and three financial variables. These variables along with their definitions are given below.

- Domestic industrial production (industrial production index with 1980-81 as the base period ; IP)
- Broad money supply (based on M3; MS)
- Consumer price index (based on consumer price index for industrial workers with 1982-83 as the base period; CPI)
- Nominal exchange rate (rupees per U.S. dollar ($ US) ; FRX)
- Domestic oil price (wholesale price index for fuels with 1981-82 as the base period ; OILD)
- NASDAQ composite stock index (NASDAQ composite of the U.S. ; NSD)
- Foreign direct investment (total monthly inflow in U.S. dollar during a month ; FDI)
- Foreign institutional investment (total monthly inflow in U.S. dollar during a month; FII)
- Short term interest rate I (91-day treasury bill rate ; TBS)
- Short term interest rate II (weighted call money rate; WCALL)
- Long term interest rate (yield on government treasury bill for 10 years' maturity; TBL)
- Foreign exchange reserve (in $US million ; FOREXR)
- Fiscal deficit of the central government (fiscal deficit is measured in Rs. crore; FD)
- Term Spread (difference between TBL and TBS ; SR)

- Price-earnings ratio (P/E)
- Price-book value ratio (P/BV)
- Dividend yield (DY)

All the variables except consumer price index, foreign institutional investment, fiscal deficit of the Central government and term spread are expressed in logarithmic scale. The stationary/nonstationary status of all the time series have been checked by applying the augmented Dickey-Fuller (ADF) and the Phillips- Perron (PP) tests. First of all, all the series were found to be nonstationary at their level values except the term spread. Thereafter, the first differenced values of all the series except CPI, FII and FD were taken and the unit root tests were run once again. As regards CPI, FII and FD, these were changed to rate variables in percentages, called the inflation rate, growths in FII and FD, respectively, and the unit root tests were done on these series. These latter results are reported in Table 4.1 below. It is important to note that while carrying out unit root tests we have included both drift and linear trend terms along with monthly dummy variables in the ADF estimating equation. We find from Table 4.1 that all the variables in their first (log) differences/ rates in percentages are stationary at standard levels of significance. For most of the series the null of unit root is rejected at 1 percent level of significance by both

Table 4.1 Tests of Unit Root Hypothesis at First Differenced Values

Sample Period: April 1996 to December 2002

Return/variable in First difference	Test statistic value ADF	PP	Maximum lag length
BSESENSEX	-4.276***	-88.118***	2
BSE 100	-4.299***	-77.451***	2

NIFTY	-2.537**	-82.106***	9
IP	-6.227***	-83.978***	2
MS	-6.008***	-97.880***	2
CPI	-3.693**	-33.666***	3
OILD	-1.989**	-68.898***	7
FRX	-3.627**	-78.251***	2
NSD	-4.885***	-68.962***	2
FDI	-3.427**	-85.468***	7
FII	-3.832**	-66.345***	3
TBS	-5.982***	-103.563***	2
WCALL	-5.668***	-92.002***	3
TBL	-5.186***	-82.541***	2
FOREXR	-5.811***	-88.232***	10
FD	-5.169***	-72.736***	2
SR	-3.106*	-38.288**	3
P/E (SENSEX)	-3.567**	-70.438***	3
P/BV(SENSEX)	-3.731**	-69.843***	3
DY(SENSEX)	-3.815**	-65.293***	2
P/E (BSE 100)	-3.633**	-51.778***	2
P/BV(BSE 100)	-4.054***	-58.908***	2
DY(BSE 100)	-3.686**	-54.587***	2

*NOTE: * , ** and *** indicate significance at 10%, 5% and 1% levels of significance, respectively. The critical values for the ADF and PP statistics have been obtained from Dickey-Fuller (1981).*

the ADF and PP tests. But for NIFTY returns, the differenced/rate series of CPI, OILD, FRX, FDI and FII, the null of unit root can be rejected only at 5 percent level of significance by the ADF test although the PP test rejects the null hypothesis for these series at 1 percent level. Thus, all these macro and financial variables would, hereafter, refer to their growths/changes in log values except the three aforesaid variables, *viz.,* CPI, FII and FD, which are percentage changes only.[5] All the 16 series of macro and financial variables have also been deseasonalized, if required, by appropriate regression with monthly dummy variables.

We have already mentioned in Chapter 3 that we test for parameter stability using Andrews's test procedure. In this testing exercise, we have assumed the dynamic specification, $AR(1)$ without an intercept[6] for each of three returns series being considered in this analysis. It may be relevant to point out that the alternative here is taken to be one where a structural change has occurred at some unknown time point and the error variance is allowed to change. Leaving out 15 percent observations at the beginnings as well as at the end for each series, relevant test statistics were computed taking all remaining months as probable breakdates. The maximum values of the three test sequences for the monthly BSESENSEX, BSE 100 and NIFTY series have been obtained as 1.363, 3.301 and 1.227, respectively. All these values are less than the 5% critical value, 8.68, corresponding to $p = 1$ in the Andrews Table (2003). We, therefore,

---

[5] For the sake of convenience, we may not always mention 'growth/change' in respect of these variables; we may merely state the names of the variables although these would refer to their growths or changes, as the case may be.

[6] For all the return series the null hypothesis of zero mean could not be rejected and hence the *AR(1)* specification is taken without any intercept term.

conclude that there is no structural break in all the three monthly return series i.e., all the three return series have parametric stability over the sample period.. Hence the rest of the computations in this chapter are carried out for the entire sample as a whole.

4.3.1 Results of specific-to-general modelling approach

We now examine the predictive ability of each macro and financial variable based on in-sample and out-of-sample forecasting performances. The in-sample regression results for equation (4.1) for each of the return series on BSESENSEX and BSE 100 are reported in Tables 4.2 and 4.3, respectively, and those for NIFTY are presented in Table A4.1 as an appendix to this chapter. These tables also report the values obtained for Theil's U and *MSE-F* and *ENC-NEW* statistics for the out-of-sample forecasts. For the out-of-sample tests, the in-sample portion of the sample ends in December 2000 for each return series, while the out-of-sample portion begins in January 2001. The computations have been carried out by considering four different horizons *viz.,* $k = 1, 3, 6$ and 12 months. We first discuss the in-sample and out-of-sample forecast results based on monthly returns on BSESENSEX for each of the 16 variables. Let us first consider the case of domestic industrial production (IP). Table 4.2 indicates that from consideration of both in-sample and out-of-sample forecasting performances domestic industrial production growth has been found to have no predictive ability for BSESENSEX returns. In a similar way, growth of broad money supply is found to have no predictive ability for BSESENSEX return. Insofar as inflation rate is concerned, we find from Table 4.2 that its in-sample t-statistic value of –2.145 is significant with $p-$value being 0.017, although for 1-month horizon forecast only.

Table 4.2: Results of In-Sample and Out-of-Sample Predictive Ability of the 16 Variables for returns on BSESENSEX (Out-of-sample period : January 2001 – December 2002)

Horizon	1	3	6	12	1	3	6	12
I. Domestic Industrial production (IP)					II. Money supply (MS)			
$\hat{\beta}$	-0.003	-0.001	0.004	0.009	0.003	-0.005	0.009	0.013
$t-$statistic	–0.377	-0.066	0.227	0.892	0.341	-0.498	0.997	0.689
	(0.336)	(0.485)	(0.449)	(0.250)	(0.341)	(0.313)	(0.173)	(0.320)
R^2	0.011	0.001	0.003	0.003	0.01	0.003	0.005	0.004
Theil's U	1.005	1.020	0.999	1.006	1.01	1.003	1.001	0.989
$MSE-F$	-0.234	-0.836	0.034	-0.145	-0.505	-0.149	-0.049	0.294
	(0.435)	(0.176)	(0.336)	(0.442)	(0.361)	(0.506)	(0.572)	(0.128)
$ENC-$	0.097	-0.394	0.019	-0.071	-0.230	-0.043	0.011	0.170
NEW	(0.496)	(0.100)	(0.433)	(0.372)	(0.280)	(0.459)	(0.420)	(0.164)
III. Consumer price index (CPI)					IV. Domestic oil price (OILD)			
$\hat{\beta}$	-0.0193	0.004	-0.023	0.018	0.001	0.005	-0.025	-0.045
$t-$statistic	-2.145	0.247	-1.174	0.554	0.073	0.310	-1.80	-1.490
	(0.017)	(0.432)	(0.149)	(0.377)	(0.499)	(0.420)	(0.007)	(0.179)
R^2	0.066	0.002	0.019	0.007	0.009	0.003	0.023	0.04
Theil's U	0.972	0.999	1.037	1.028	1.012	1.012	1.024	1.098
$MSE-F$	1.362	0.003	-1.346	-0.687	-0.576	-0.518	-0.888	-2.224
	(0.049)	(0.327)	(0.237)	(0.317)	(0.357)	(0.381)	(0.246)	(0.102)
$ENC-$	1.131	0.016	-0.535	-0.313	-0.257	-0.218	-0.325	-0.886
NEW	(0.049)	(0.404)	(0.197)	(0.252)	(0.285)	(0.324)	(0.221)	(0.078)

Table 4.2 (contd.)

Horizon	1	3	6	12	1	3	6	12
	V. Nominal exchange rate (FRX)				VI. NASDAQ composite index (NSD)			
$\hat{\beta}$	-0.015	-0.018	-0.017	-0.026	0.014	0.011	0.018	0.022
$t-$statistic	-1.730	-1.016	-0.889	-1.463	1.515	0.797	1.193	1.352
	(0.042)	(0.171)	(0.208)	(0.163)	(0.058)	(0.22)	(0.133)	(0.161)
R^2	0.047	0.02	0.011	0.014	0.038	0.007	0.011	0.009
Theil's U	1.008	1.170	1.140	1.267	0.963	0.992	1.014	1.120
$MSE-F$	-0.372	-5.942	-4.392	-4.921	1.887	0.356	-0.531	-2.643
	(0.491)	(0.002)	(0.018)	(0.014)	(0.036)	(0.191)	(0.332)	(0.051)
$ENC-NEW$	0.662	-1.698	-1.624	-1.867	1.116	0.240	-0.126	-1.09
	(0.105)	(0.004)	(0.007)	(0.003)	(0.06)	(0.237)	(0.361)	(0.028)
	VII. Foreign direct investment (FDI)				VIII. Foreign institutional investment (FII)			
$\hat{\beta}$	0.005	0.006	0.005	0.008	-0.002	0.005	0.009	0.03
$t-$statistic	0.584	0.814	0.586	0.613	-0.239	0.484	0.647	1.556
	(0.276)	(0.196)	(0.291)	(0.307)	(0.404)	(0.345)	(0.330)	(0.149)
R^2	0.013	0.004	0.003	0.002	0.010	0.002	0.004	0.013
Theil's U	1.058	1.009	1.016	1.058	1.007	1.002	1.035	1.027
$MSE-F$	-2.542	-0.394	-0.607	-1.397	-0.339	-0.086	-1.249	-0.669
	(0.031)	(0.233)	(0.084)	(0.011)	(0.462)	(0.602)	(0.172)	(0.240)
$ENC-NEW$	-0.804	-0.140	-0.230	-0.616	-0.154	-0.04	-0.565	-0.227
	(0.037)	(0.228)	(0.077)	(0.006)	(0.372)	(0.501)	(0.102)	(0.238)

Table 4.2 (contd.)

Horizon	1	3	6	12	1	3	6	12
IX. Short-term interest rate 1 (TBS)					X. Short-term interest rate 2 (WCALL)			
$\hat{\beta}$	0.005	0.008	0.001	-0.001	0.005	-0.003	0.008	-0.001
$t-$statistic	0.598	0.599	0.077	-0.098	0.556	-0.205	0.624	-0.099
	(0.244)	(0.266)	(0.434)	(0.488)	(0.286)	(0.440)	(0.261)	(0.484)
R^2	0.014	0.005	0.003	0.002	0.013	0.002	0.005	0.001
Theil's U	1.002	0.999	1.002	1.008	0.994	0.999	1.005	1.004
$MSE-F$	-0.083	0.006	-0.091	-0.195	0.268	0.053	-0.208	-0.104
	(0.693)	(0.319)	(0.481)	(0.267)	(0.198)	(0.280)	(0.281)	(0.366)
$ENC-NEW$	–0.018	0.019	-0.045	-0.095	0.141	0.027	-0.089	-0.049
	(0.647)	(0.382)	(0.377)	(0.208)	(0.278)	(0.379)	(0.227)	(0.308)
XI. Long-term interest rate (TBL)					XII. Foreign exchange reserve (FOREXR)			
$\hat{\beta}$	0.015	0.04	0.04	0.039	0.005	-0.004	0.006	-0.001
$t-$statistic	1.552	2.476	2.210	2.302	0.561	-0.440	0.503	-0.231
	(0.053)	(0.013)	(0.030)	(0.045)	(0.296)	(0.304)	(0.361)	(0.345)
R^2	0.039	0.074	0.051	0.024	0.013	0.002	0.004	0.001
Theil's U	1.040	0.962	0.931	0.840	1.004	1.004	0.999	0.998
$MSE-F$	-1.795	1.786	2.938	5.434	-0.198	-0.169	0.052	0.047
	(0.087)	(0.029)	(0.012)	(0.000)	(0.524)	(0.361)	(0.246)	(0.274)
$ENC-NEW$	–0.322	2.099	2.085	3.126	-0.090	-0.081	0.028	0.024
	(0.208)	(0.009)	(0.015)	(0.006)	(0.436)	(0.283)	(0.324)	(0.332)

Table 4.2 (contd.)

Horizon	1	3	6	12	1	3	6	12
XIII. Fiscal deficit of the Central government (FD)					XIV. Price-earnings ratio (P/E)			
$\hat{\beta}$	-0.007	-0.032	-0.035	-0.037	0.00003	0.002	-0.005	-0.028
$t-$statistic	-0.804	-4.760	-4.923	-3.288	0.003	0.148	-0.224	-1.553
	(0.232)	(0.011)	(0.013)	(0.040)	(0.519)	(0.491)	(0.459)	(0.196)
R^2	0.017	0.065	0.046	0.029	0.006	0.002	0.003	0.012
Theil's U	1.139	0.968	0.909	0.965	1.008	1.007	1.032	1.112
$MSE-F$	-5.507	1.474	3.974	0.946	-0.392	-0.299	-1.177	-2.491
	(0.032)	(0.04)	(0.005)	(0.060)	(0.469)	(0.511)	(0.233)	(0.078)
$ENC-NEW$	0.244	6.653	3.643	0.528	-0.186	-0.146	-0.535	-0.830
	(0.225)	(0.000)	(0.008)	(0.095)	(0.374)	(0.414)	(0.171)	(0.075)
XV. Price-book value ratio (P/BV)					XVI. Dividend yield (DY)			
$\hat{\beta}$	-0.0001	0.003	0.007	-0.002	-0.002	-0.007	-0.0005	0.007
$t-$statistic	-0.012	0.153	0.316	-0.102	-0.213	-0.438	-0.023	0.308
	(0.519)	(0.436)	(0.387)	(0.519)	(0.415)	(0.346)	(0.457)	(0.448)
R^2	0.006	0.002	0.004	0.001	0.006	0.004	0.003	0.002
Theil's U	1.006	1.005	1.005	1.034	1.008	1.000	1.012	1.078
$MSE-F$	-0.265	-0.209	-0.206	-0.835	-0.404	-0.010	-0.466	-1.815
	(0.589)	(0.543)	(0.479)	(0.206)	(0.436)	(0.674)	(0.382)	(0.115)
$ENC-NEW$	-0.127	-0.102	-.0.098	-0.338	-0.183	0.0001	-0.222	-0.645
	(0.480)	(0.440)	(0.389)	(0.821)	(0.351)	(0.418)	(0.293)	(0.089)

Table 4.2 (contd.)

XVII. Term Spread (SR)

$\hat{\beta}$	0.004	0.016	0.043	0.042
$t-$statistic	0.496	0.759	1.671	0.802
	(0.300)	(0.275)	(0.151)	(0.323)
R^2	0.009	0.014	0.065	0.031
Theil's U	1.003	0.987	0.916	1.136
$MSE-F$	-0.171	0.591	3.638	-2.922
	(0.355)	(0.226)	(0.093)	(0.718)
$ENC-NEW$	-0.324	0.464	2.560	-1.267
	(0.398)	(0.279)	(0.124)	(0.804)

Notes: In this table, all the macro and financial variables along with their abbreviations/symbols have been mentioned in terms of their descriptions in level values although these values refer to their growths/changes; $\hat{\beta}$ and $t-$statistic are the OLS estimate of β in equation (4.1) and its corresponding $t-$statistic value, respectively; R^2 is the goodness of fit measure in equation (4.1) ; Theil's U is the ratio of the RMSE for the out-of-sample forecasts for the unrestricted model to the RMSE for the out-of-sample forecast for the restricted model; *MSE-F* and *ENC-NEW* are the out-of-sample statistics given in equations (4.2) and (4.3), respectively; $p-$ values are given in parentheses.

However, all the out-of-sample test statistics for inflation rate are significant at this horizon ($p-$values of $MSE-F$ and $ENC-NEW$ are both 0.049). Therefore, this exhibits that change in inflation rate has both in-sample and out-of-sample predictive ability for BSESENSEX returns. The change in domestic oil price (OILD) again has no significant predictive ability at any of the horizons considered. The change in nominal exchange rate has some significant in-sample predictive ability at horizon of 1 since t-statistic is significant at 4.02 per cent level of significance, but the out-of-sample test statistic is not significant at this horizon, the $p-$values for $MSE-F$ and $ENC-NEW$ being as high as 0.491 0.105 respectively. However, for other horizons, *viz.,* $k=3$, 6 and

12, all the out-of-sample test statistics are found to be significant. It can be further stated from Table 4.2 that in so far as predictive ability of NASDAQ composite return(NSD) is concerned, it has significant out-of-sample predictive ability at horizon of 1 month ($p = 0.036$) and in-sample t-statistic value being 1.515 with $p-$value 0.058 is significant. at 5.8 per cent level of significance for the same horizon. Further, for $k = 12$, the out-of-sample test statistic is significant since $p-$value is 0.051. Insofar as the two foreign investment variables, FDI and FII, are concerned, we find that in terms of out-of-sample predictive ability for BSESENSEX returns, growth of FDI has a significant role since both the $MSE-F$ and $ENC-NEW$ values of -2.542 and –0.804 are significant at 5 per cent level of significance at k=1 horizon. These two out-of-sample test statistics are also significant at k=12 for FDI but the change in FII has no significant role. Neither any of Theil's U value is less than 1 nor any three of the statistics *viz.,* t-statistic, $MSE-F$ and $ENC-NEW$ value is significant. This implies that growth of foreign capital inflow, as such, has very limited impact on Indian stock returns.

In most of such studies with returns of other stock markets, interest rate (short-term) has been found to be the single most important predictor for domestic stock returns. As already mentioned, we have three interest rate variables – two being short-term rates and one long term rate. The two short-term interest rates are 91-day treasury-bill rate (TBS) and weighted call money rate (WCALL). Both these variables have been found to have no in-sample as well as out-of-sample predictive ability at any of the horizons considered by all the criteria with the sole exception of Theil's U which is less than 1 at k=3 for 91-day treasury bill rate and at k=1 and 3 for weighted call money rate. However, the long-term interest rate (TBL) which represents the yield on government treasury bill for 10 years' maturity has some strong predictive ability at medium and longer horizons. Although TBL shows some in-sample predictive ability for BSESENSEX at $k = 1$ since t-statistic value is significant at 5.3 per cent level of significance, the out-of-sample test statistic is significant by $MSE - F$ criterion. As regards the results for $k = 3$, 6, 12, in-sample t-statistic values are 2.476, 2.210 and 2.302 for these three horizons, which are highly significant since the $p-$values are 0.013, 0.030 and 0.045, respectively and both the $MSE - F$ and $ENC - NEW$ statistics are also significant at all the values of k =3, 6 and 12 and all the Theil's U values are less than 1 for all these horizons. This shows that change in long-term interest rate has strong predictive ability for Indian stock returns as measured by BSESENSEX. This may be related to the nature of expectations of the investors. Foreign exchange reserve (FOREXR) is found to have no in-sample or out-of sample predictive ability for BSESENSEX returns at all horizons. The macro variable fiscal deficit is seen to have quite strong effect on stock return since all the criteria for predictive ability *viz.,* $t-$statistic value ($p-$values being 0.011,0.013 and 0.040 at k=3, 6 and 12, respectively), Theils' U (the values being 0.968, 0.909 and 0.965), $MSE - F$ ($p-$values are 0.032, 0.04, 0.005 and 0.06 at k=1, 3, 6 and 12, respectively) and $ENC - NEW$ ($p-$values are 0.000, 0.008 and 0.095 at k=3, 6 and 12, respectively) show significant values at medium and longer horizons, although at k=1, it has no in-sample significant value but significant out-of-sample predictive ability by $MSE - F$ ($p-$value 0.032) only. Among the three financial variables, only price-earning ratio (P/E) is found to have some out-of-sample predictive ability for BSESENSEX returns since the out-of-sample test statistic values for $MSE - F$ and $ENC - NEW$ are significant at $k = 12$ ($p -$ values are 0.078 and 0.075, respectively). Neither dividend yield (DY) nor price-book-value ratio (P/BV) is significant at any horizon by any of the criteria. This indicates that financial variables have very marginal influence in explaining variations in BSESENSEX returns.

Taking all the above findings together, we can infer that when selection of the variables are made on the basis of specific to general approach, the following macroeconomic and financial variables have predictive ability on BSESENSEX returns: inflation rate (CPI), nominal exchange rate growth (FRX), returns on NASDAQ composite (NSD), growth of foreign direct investment (FDI), change in long term interest rate (TBL), change in fiscal deficit of the Central government (FD), change in price-earnings ratio (P/E).

Since this exercise of variable selection was carried out also for returns based on BSE 100 and NIFTY indices as well, we now briefly report these findings. From Table 4.3 where the computational figures for BSE 100 are presented, we observe that, as in the case of BSESENSEX, domestic industrial production and broad money supply have no in-sample or out-of-sample predictive ability for BSE 100 returns at any of the horizons considered. Again, for inflation rate, the in-sample predictive ability at $k = 1$ is significant, but the same does not hold for the out-of-sample criterion unless the level of significance is taken to be 9.4 per cent. Here, Theil's U is, however, less than 1 for $k = 1$. It is evident from this table that, domestic oil price has no in-sample or out-of-sample predictive ability.

Table 4.3: Results of In-Sample and Out-of-Sample Predictive Ability of the 16 Variables for returns on BSE 100 (Out-of-sample period : January 2001 – December 2002)

Horizon	1	3	6	12	1	3	6	12
	I. Domestic Industrial production (IP)				II. Money supply (MS)			
$\hat{\beta}$	0.001	0.006	0.002	0.010	0.001	-0.007	0.005	0.014
$t-$statistic	0.132	0.284	0.130	0.736	0.117	-0.688	0.507	0.677
	(0.446)	(0.395)	(0.476)	(0.298)	(0.440)	(0.267)	(0.339)	(0.320)
R^2	0.0002	0.001	0.008	0.001	0.0002	0.002	0.009	0.002
Theil's U	1.002	1.005	1.003	1.001	1.009	1.000	1.003	0.989
$MSE-F$	-0.083	-0.224	-0.098	-0.291	-0.409	-0.007	-0.127	0.283
	(0.663)	(0.482)	(0.554)	(0.303)	(0.422)	(0.653)	(0.478)	(0.144)
$ENC-NEW$	–0.038	-0.106	-0.044	-0.142	-0.198	0.028	-0.043	0.166
	(0.567)	(0.378)	(0.445)	(0.248)	(0.326)	(0.372)	(0.426)	(0.180)
	III. Consumer price index (CPI)				IV. Domestic oil price (OILD)			
$\hat{\beta}$	-0.02	0.004	-0.020	0.002	-0.004	0.004	-0.032	-0.063
$t-$statistic	-1.952	0.230	-0.898	0.053	-0.367	0.222	-2.053	-1.803
	(0.021)	(0.439)	(0.215)	(0.491)	(0.347)	(0.468)	(0.060)	(0.122)
R^2	0.048	0.0007	0.016	0.0002	0.002	0.0009	0.023	0.048
Theil's U	0.982	1.003	1.037	1.107	1.002	1.011	1.014	1.073
$MSE-F$	0.863	-0.149	-1.346	-2.393	-0.119	-0.483	-0.656	-1.711

Table 4.3 (contd.)

Horizon	1	3	6	12	1	3	6	12
	(0.090)	(0.642)	(0.239)	(0.123)	(0.650)	(0.423)	(0.335)	(0.148)
ENC – NEW	0.815	-0.062	-0.588	-1.068	-0.055	-0.208	-0.208	-0.568
	(0.094)	(0.542)	(0.177)	(0.088)	(0.539)	(0.361)	(0.324)	(0.150)
	V. Nominal exchange rate (FRX)				VI. NASDAQ composite index (NSD)			
$\hat{\beta}$	-0.013	-0.014	-0.009	-0.033	0.018	0.017	0.019	0.037
$t-$statistic	-1.348	-0.642	-0.467	-1.707	1.618	0.996	1.003	1.352
	(0.101)	(0.268)	(0.342)	(0.128)	(0.054)	(0.196)	(0.184)	(0.105)
R^2	0.023	0.008	0.010	0.013	0.033	0.009	0.014	0.012
Theil's U	1.017	1.189	1.102	1.274	0.965	0.986	1.028	1.069
$MSE-F$	-0.797	-6.429	-3.357	-4.990	1 .746	0.623	-1.035	-1.623
	(0.274)	(0.005)	(0.023)	(0.005)	(0.040)	(0.137)	(0.182)	(0.110)
ENC – NEW	0.200	-2.240	-1.398	-1.934	1.148	0.451	-0.126	-1.09
	(0.756)	(0.001)	(0.014)	(0.001)	(0.056)	(0.164)	(0.144)	(0.068)
VII. Foreign direct investment (FDI)					VIII. Foreign institutional investment (FII)			
$\hat{\beta}$	0.005	0.001	-0.002	0.011	-0.005	-0.005	0.011	0.04
$t-$statistic	0.522	0.094	-0.143	0.806	-0.501	-0.370	0.642	1.594
	(0.320)	(0.472)	(0.597)	(0.240)	(0.316)	(0.379)	(0.299)	(0.117)
R^2	0.004	0.000	0.008	0.001	0.003	0.001	0.010	0.013
Theil's U	1.044	1.007	1.01	1.061	1.00	1.005	1.025	1.046

Table 4.3 (contd.)

Horizon	1	3	6	12	1	3	6	12
$MSE-F$	-1.976	-0.289	-0.378	-1.458	0.019	-0.237	-0.913	-1.129
	(0.064)	(0.267)	(0.157)	(0.009)	(0.281)	(0.508)	(0.121)	(0.150)
$ENC-NEW$	–0.630	-0.135	-0.175	-0.652	0.051	-0.085	-0.407	-0.398
	(0.072)	(0.205)	(0.118)	(0.008)	(0.345)	(0.463)	(0.148)	(0.138)

IX. Short-term interest rate 1 (TBS)					X. Short-term interest rate 2 (WCALL)			
$\hat{\beta}$	0.005	0.006	0.005	0.002	0.006	-0.005	0.011	0.004
$t-$statistic	0.480	0.423	0.367	0.169	0.593	-0.298	0.737	0.312
	(0.300)	(0.329)	(0.349)	(0.408)	(0.270)	(0.379)	(0.204)	(0.341)
R^2	0.003	0.002	0.009	0.000	0.005	0.001	0.011	0.0003
Theil's U	0.999	0.999	1.002	1.003	0.993	1.000	1.004	1.003
$MSE-F$	0.024	0.033	-0.092	-0.065	0.321	0.050	-0.139	-0.082
	(0.278)	(0.295)	(0.482)	(0.434)	(0.165)	(0.260)	(0.334)	(0.338)
$ENC-NEW$	0.024	0.023	-0.044	-0.032	0.168	0.028	-0.057	-0.040
	(0.366)	(0.376)	(0.399)	(0.371)	(0.256)	(0.352)	(0.286)	(0.704)

XI. Long-term interest rate (TBL)					XII. Foreign exchange reserve (FOREXR)			
$\hat{\beta}$	0.020	0.046	0.045	0.040	0.003	-0.002	0.006	-0.003
$t-$statistic	1.892	2.727	2.117	1.963	0.344	-0.234	0.555	-0.478
	(0.030)	(0.012)	(0.059)	(0.072)	(0.377)	(0.379)	(0.369)	(0.299)
R^2	0.045	0.077	0.047	0.015	0.002	0.0002	0.009	0.0002
Theil's U	1.025	0.954	0.954	0.889	1.002	1.003	0.998	0.999

Table 4.3 (contd.)

Horizon	1	3	6	12	1	3	6	12
$MSE-F$	-1.161	2.161	1.892	3.464	-0.108	-0.128	0.062	0.027
	(0.179)	(0.014)	(0.025)	(0.006)	(0.568)	(0.379)	(0.222)	(0.300)
$ENC-NEW$	0.164	2.399	1.440	1.998	-0.050	-0.063	0.032	0.014
	(0.272)	(0.006)	(0.021)	(0.016)	(0.478)	(0.289)	(0.290)	(0.364)
XIII. Fiscal deficit of the Central government (FD)					XIV. Price-earnings ratio (P/E)			
$\hat{\beta}$	-0.009	-0.038	-0.045	-0.048	0.002	0.008	0.014	-0.008
$t-$statistic	-0.895	-6.424	-6.401	-2.917	0.128	0.333	0.332	-0.317
	(0.206)	(0.003)	(0.004)	(0.067)	(0.422)	(0.348)	(0.346)	(0.464)
R^2	0.010	0.063	0.057	0.030	0.0005	0.002	0.010	0.0007
Theil's U	1.075	0.896	0.906	1.014	1.007	1.000	1.000	1.014
$MSE-F$	-3.235	5.413	4.127	-0.362	-0.344	0.017	-0.023	-0.349
	(0.073)	(0.000)	(0.004)	(0.316)	(0.527)	(0.338)	(0.640)	(0.404)
$ENC-NEW$	0.782	6.686	2.981	-0.528	-0.159	0.039	0.010	-0.136
	(0.094)	(0.000)	(0.014)	(0.270)	(0.395)	(0.395)	(0.430)	(0.364)
XV. Price-book value ratio (P/BV)					XVI. Dividend yield (DY)			
$\hat{\beta}$	-0.012	-0.002	0.031	0.032	0.013	-0.002	-0.035	-0.021
$t-$statistic	-0.948	-0.124	0.842	1.178	1.008	-0.089	-0.911	-0.689
	(0.189)	(0.446)	(0.213)	(0.162)	(0.162)	(0.464)	(0.201)	(0.256)
R^2	0.012	0.0005	0.021	0.008	0.014	0.0004	0.024	0.003
Theil's U	0.989	1.012	0.987	0.992	0.985	1.006	0.986	0.988
$MSE-F$	0.526	-0.521	0.522	0.197	0.748	-0.291	0.535	0.313
	(0.154)	(0.413)	(0.170)	(0.269)	(0.112)	(0.539)	(0.191)	(0.263)
$ENC-NEW$	0.362	-0.233	0.340	0.231	0.471	-0.134	0.490	0.206
	(0.195)	(0.330)	(0.217)	(0.241)	(0.161)	(0.577)	(0.196)	(0.308)

Table 4.3 (contd.)

XVII. Term Spread (SR)

$\hat{\beta}$	0.005	0.015	0.044	0.050
$t-$statistic	0.506	0.734	1.481	0.737
	(0.309)	(0.282)	(0.147)	(0.352)
R^2	0.004	0.010	0.052	0.028
Theil's U	1.002	0.995	0.959	1.191
$MSE-F$	-0.113	0.204	1.655	-3.833
	(0.326)	(0.254)	(0.175)	(0.795)
$ENC-NEW$	-0.004	0.231	1.291	-1.685
	(0.398)	(0.320)	(0.201)	(0.881)

Notes: In this table, all the macro and financial variables along with their abbreviations/symbols have been mentioned in terms of their descriptions in level values although these values refer to their growths/changes; $\hat{\beta}$ and $t-$statistic are the OLS estimate of β in equation (4.1) and its corresponding $t-$statistic value, respectively; R^2 is the goodness of fit measure in equation (4.1); Theil's U is the ratio of the RMSE for the out of sample forecasts for the unrestricted model to the RMSE for the out-of-sample for the restricted model; $MSE-F$ and $ENC-NEW$ are the out-of-sample statistics given in equations (4.2) and (4.3), respectively; $p-$values are given in parentheses.

Although there is no significant effect by in-sample criterion of $t-$statistics, nominal exchange rate growth is found to have significant effect by out-of-sample criteria for $k=$ 3, 6, 12. Insofar as returns on NASDAQ composite of the US is concerned, there are significant values by both in-sample and out-of-sample criteria at horizon $k=1$. For other horizons, NSD has no predictive ability for BSE 100 stock returns. As in case of BSESENSEX, growth of FDI has some significant out-of-sample predictive ability at $k=1$ and $k=12$, but the FII has no in-sample or out-of-sample predictive ability at any of the horizons considered. As regards the short-term interest rate variables, TBS and

WCALL, there is no significant in-sample or out-of-sample predictive ability at all. For the long term interest rate variable *viz.,* TBL, the in-sample t- statistic values are significant at all the horizons considered. Out-of-sample test statistic values are also significant at all horizons except for $k=1$. The foreign exchange reserve (FOREXR) variable is found to have no significant role as far as prediction in the stock returns is concerned. Fiscal deficit of the central government (FD) again has both significant in-sample and out-of-sample predictive ability. No financial variable has been found to have any impact on BSE 100 returns as far as predictive ability is concerned. On the basis of these observations we can conclude that changes/growths in the following 6 macrovariables *viz.,* CPI, FRX, NSD, FDI, TBL and FD, have significant predictive ability for BSE 100. No financial variable has been found to have any predictive ability for BSE 100 returns. We thus note that the set of macroeconomic variables found to have significant predictive ability for BSE 100 is exactly the same as that found for BSESENSEX earlier. Insofar as the results concerning NIFTY[6] are concerned, we find that the set of selected macroeconomic variables are the same as those of BSESENSEX and BSE 100, and hence we do not make any further discussions on this index and present the computational findings in Table A4.1 as an appendix to this chapter.

4.3.2 Results of general-to-specific modelling approach We now discuss the findings on selection of macroeconomic variables and financial ratios following the other approach of variable selection *viz.,* general to specific approach. As already pointed out in the preceding section that although two approach of variable selection- specific to general and general to specific- exist in literature, we have put more emphasis on the first

[6] It may be noted that, as stated earlier, the computations for NIFTY was carried out with macro variables only since data for the financial variables corresponding to this index are available only from January 1999.

approach in describing the predictive ability of the macro and financial variables. Accordingly, we briefly report on the findings of the second approach. Following Clark (2004), we combine in-sample general-to-specific model selection with tests of out-of-sample forecasting ability. As in the other approach, our in-sample period covers April 1996 through December 2000 and the out-of-sample is taken to be the period January 2001– December 2002. We first discuss the empirical results for BSESENSEX returns.

From Table 4.4 where these computational figures for returns on BSESENSEX as well as for BSE 100 and NIFTY are given, we observe that the forecasting model selected over the in-sample period at the 1-month horizon includes three macroeconomic variables,

Table 4.4: In-Sample and Out-of-Sample Results of Predictive Ability of the Variables Based on General to Specific Model Selection Approach (Out-of-sample period : January 2001 – December 2002)

I. Returns on BSESENSEX

Horizon	1	3	6	12
Included variables	CPI, FRX, TBL	IP, MS, FRX, WCALL,TBL, FD	MS, CPI, OILD, FRX, FII, TBS, TBL, FOREXR, P/BV	CPI,OILD,NSD, FDI,TBS,WCALL, TBL, FOREXR, P/E, P/BV, DY
Theil's *U*	1.007	1.102	1.391	1.605
$MSE-F$	-0.338	-3.869	-9.190	-7.955
	(0.819)	(0.428)	(0.103)	(0.121)
$ENC-NEW$	1.812	4.234	-2.445	-1.771
	(0.160)	(0.061)	(0.032)	(0.056)

Table 4.4 (contd.)

II. Returns on BSE100

Horizon	1	3	6	12
Included variables	CPI, FRX, TBL, FD, P/E, P/BV	FRX,WCALL, TBL	MS, CPI, OILD, FRX, NSD, TBL, FOREXR	CPI, OILD, FDI, TBS, WCALL, TBL, FOREXR, P/E, P/BV, DY
Theil's *U*	1.108	1.148	1.298	1.791
MSE – F	-4.438	-5.318	-7.731	-8.949
	(0.657)	(0.282)	(0.170)	(0.081)
ENC – NEW	5.295	-0.853	-1.878	-2.354
	(0.018)	(0.283)	(0.907)	(0.091)

III. Returns on NIFTY

Horizon	1	3	6	12
Included variables	CPI, TBL, FD	IP, MS, FRX, NSD, WCALL, TBL, FOREXR,FD	CPI, OILD, FRX, TBL, FOREXR	FRX
Theil's *U*	1.266	1.055	1.148	1.279
MSE – F	-9.037	-2.220	-4.586	-5.050
	(0.036)	(0.512)	(0.262)	(0.208)
ENC – NEW	1.262	2.203	-0.983	-1.896
	(0.222)	(0.083)	(0.797)	(0.038)

Notes: Theil's U is the ratio of the RMSE for the out-of- sample forecasts for the selected model to the RMSE for the out-of-sample for the restricted model; $MSE-F$ and $ENC-NEW$ are the out-of-sample statistics given in equations (4.2) and (4.3), respectively; $p-$ values are given in parentheses.

namely, inflation rate, changes in nominal exchange rate as also long-term interest rate.[7] However, in terms of out-of-sample predictive ability, the forecasting relation obtained thereof is not significant since the $p-$values for the $MSE-F$ and $ENC-NEW$ test statistics are as large as 0.819 and 0.16, respectively. As for the other 3 horizons, we note from this table that the in-sample based variables included are IP, MS, FRX, WCALL, TBL and FD for horizon of 3 ; MS, CPI, OILD, FRX, FII, TBS, TBL, FOREXR and P/BV for horizon of 6 ; CPI, OILD, NSD, FDI, TBS, WCALL, TBL, FOREXR, P/E, P/BV and DY for horizon of 12. But, for out-of-sample performance, predictive ability is significant by $ENC-NEW$ test statistic only since the p-values are 0.061, 0.032 and 0.056 for $k=3$, 6 and 12, respectively. As regards BSE 100, we have found the included variables to be CPI, FRX, TBL, FD, P/E and P/BV (for $k=1$), FRX, WCALL and TBL (for $k=3$), MS, CPI, OILD, FRX, NSD, TBL, FOREXR (for $k=6$), CPI, OILD, FDI, TBS, WCALL, TBL, FOREXR, P/E, P/BV and DY (for $k=12$). But out-of-sample predictability performance was found to be marginally significant at horizon $k=12$ only. In a similar way we have found the following macroeconomic variables for NIFTY by this approach: CPI, TBL and FD (for $k=1$), IP, MS, FRX, NSD, WCALL, TBL, FOREXR and FD (for $k=3$), CPI, OILD, FRX, TBL and FOREXR (for $k=6$), FRX(

[7] The software used for these computations (*cf.* Rapach *et al*.(2005) automatically choose these variables to be the included variables after carrying out the necessary tests, as stated in Section 4.2.

for $k = 12)$. In case of NIFTY, we found two significant out-of-sample values- one each for $MSE - F$ statistic at ($k = 1$ with $p -$value 0.036) and $ENC - NEW$ statistic (at $k = 12$ $p -$value as 0.038). We find that by this criterion, the out-of-sample predictive evidence is, in general, rather weak for all the three return series.

We note that all variables which have been selected by the specific to general approach have also been selected by the general to specific approach for all the three series. Thus, emphasising on the findings of the first approach yet and combining the findings of second approach as well, we finally conclude that as regards the macro variables the same set of the following macro variables are found to have significant predictive ability for BSESENSEX, BSE 100 and NIFTY returns; as for the 3 financial ratios considered, it is found that only for P/E ratio has marginal significance for BSESENSEX returns only. Thus the set of selected macro and financial variables (in terms of growth/change) having predictive ability for Indian stock returns are the following.

BSESENSEX: consumer price index, nominal exchange rate, NASDAQ composite return, foreign direct investment, long-term interest rate, fiscal deficit of the central government and price-earning ratio.

BSE100: consumer price index, nominal exchange rate, NASDAQ composite return, foreign direct investment, long-term interest rate and fiscal deficit of the central government

NIFTY: consumer price index, nominal exchange rate, NASDAQ composite return, foreign direct investment, long-term interest rate and fiscal deficit of the central government.

4.3.3 The estimated models for the residuals Once the significant macroeconomic and financial variables have been chosen, we consider the dynamic linear regression model specified in (4.8) incorporating the relevant macroeconomic variables and financial ratios for each of the three monthly return series. These equations have been first estimated by the OLS and the estimated equations are presented in equations (4.9) through (4.11). While specifying the regression equation, lagged values of r_t have also been considered so that no serial correlation remained in the errors. In addition to lagged values of return, dummy variables were included to take care of any month-of-the year effects in the returns. The estimated models for the three return series are as follows:

<u>BSESENSEX</u>

$$\hat{r}_t = 0.206 r_{t-6} - 0.268 r_{t-7} - 0.027 CPI_{t-1} - 1.909 FRX_{t-1} + 0.047 NSD_{t-1} - 0.001 FDI_{t-1}$$

(1.901)* (2.509)** (2.719)*** (2.616)** (0.556) (0.043)

$$+ 0.465 TBL_{t-1} - 0.00001 FD_{t-1} - 0.152 (P/E)_{t-1} + 0.049 D12 \qquad (4.9)$$

(2.048)** (1.149) (1.207) (1.849)*

<u>BSE100</u>

$$\hat{r}_t = 0.263 r_{t-6} - 0.213 r_{t-7} - 0.031 CPI_{t-1} - 1.907 FRX_{t-1} + 0.076 NSD_{t-1} + 0.001 FDI_{t-1}$$

(2.542)** (2.026)** (2.909)*** (2.409)** (0.816) (0.072)

$$+ 0.721 TBL_{t-1} - 0.00002 FD_{t-1} + 0.075 D11 + 0.061 D12 \qquad (4.10)$$

(3.004)*** (1.373) (2.342)** (2.081)**

<u>NIFTY</u>

$$\hat{r}_t = 0.208 r_{t-6} - 0.208 r_{t-7} - 0.029 CPI_{t-1} - 1.584 FRX_{t-1} + 0.010 NSD_{t-1} - 0.001 FDI_{t-1}$$

(2.011)** (2.056)** (3.292)*** (2.468)** (0.131) (0.078)

$$+ 0.590 TBL_{t-1} - 0.00002 FD_{t-1} + 0.50 D2 + 0.065 D11 + 0.058 D12 \tag{4.11}$$

(3.009)*** (1.586) (1.987)** (2.526)** (2.427)**

((i) All the macro and financial variables in these equations refer to their respective stationary values although their abbreviations/symbols do not indicate the same.

(ii) The values in parentheses indicate corresponding absolute values of the $t-$statistic; *, ** and *** indicate significance at 10%, 5% and 1% levels of significance, respectively).

We note from these estimated models for r_t that not all the respective selected macroeconomic variables obtained in the preceding section for the three returns have been found to be significant when these are included together in obtaining the final model. For instance, returns on the NASDAQ composite index, growth of foreign direct investment and change in fiscal deficit of the central government are not found to be significant in any of the models for return. We have found that in each of the three models, three variables *viz*., inflation rate, nominal exchange rate as well as long run interest rate are found to be significantly present. Further, some monthly dummies are also found to be significant – in particular, we have found $D12$ i.e., the month of December having significant effect in all the three returns series.

Following a procedure similar to that in Chapter 3, we now test for the adequacy of the mean specifications as obtained in equations (4.9), (4.10), (4.11). Following Lumsdaine and Ng (1999), we check first if there are any omission of contemporaneous exogenous variables. Thus, considering the contemporaneous role of the selected variables on each stock return so that the final model is properly specified, we have the following estimated models[8] for the returns on BSESENSEX, BSE 100 and NIFTY.

<u>BSESENSEX</u>

$$\hat{r}_t = -0.277r_{t-1} - 0.273r_{t-7} - 0.021CPI_{t-1} - 0.926FRX_{t-1} + 0.290NSD_t + 0.504P/E_t + 0.037D12$$

(3.154)*** (3.251)*** (2.871)*** (1.768)* (4.674)*** (5.606)*** (1.810)*

(4.12)

<u>BSE100</u>

$$\hat{r}_t = 0.168r_{t-6} - 0.202r_{t-7} - 0.016CPI_{t-1} + 0.413NSD_t - 0.554TBL_t \quad (4.13)$$

(1.868)* (2.103)** (1.838)* (5.156)*** (- 2.680)***

<u>NIFTY</u>

$$\hat{r}_t = -0.213r_{t-7} - 0.020CPI_{t-1} + 0.311NSD_t - 0.370TBL_t + 0.051D2 + 0.046D12 \quad (4.14)$$

(2.169)** (2.582)*** (4.583)*** (2.044)** (2.114)** (2.091)**

((i) All the macro and financial variables in these equations refer to their respective stationary values although their abbreviations/symbols do not indicate the same.

(ii) The values in parentheses indicate corresponding absolute values of the $t-$statistic; *, ** and *** indicate significance at 10%, 5% and 1% levels of significance, respectively).

We note from these equations that although all the return series do not have the same set of significant variables, the difference is only in terms of one or two variables. The inflation rate and returns on NASDAQ composite index are significant for all the three returns, but changes in nominal exchange rate and price-earning ratio are significant only

[8] Final estimated equations with significant variables only are being reported.

for BSESENSEX returns. But the long-term interest rate variable *viz., TBL* is not significant for BSESENSEX, but it has significant effect for BSE 100 and NIFTY returns. The signs of the coefficients of the variables are noteworthy. First of all, inflation rate has negative sign, as expected. The negative sign with the nominal exchange rate explains that as Indian rupee becomes weaker relative to the US dollar, Indian stock returns are expected to fall. As regards the negative sign of the change in coefficient of the long term interest rate is concerned, it may be concluded that it is only expected. We also find that variations in global stock returns such as NASDAQ composite return does have a strong contemporaneous effect in explaining Indian stock returns variations for all the three series. This may indicate integration of the Indian stock market with the world bourses. The significant presence of change in price-earnings ratio in BSESENSEX return obviously demonstrates the important role of this financial variable. However, foreign direct investment, the inflow of which is expected to increase the productive capacity of emerging economies such as India does not have significant role. This may be due to the fact that the issues of foreign direct investment or foreign institutional investment are comparatively new to the Indian capital market, and as a result it does not react to the changes of those variables. The absence of fiscal deficit of the central government in the final model may be due to the fact that its role is being partially captured by the significant presence of long-term interest rate. Our results also show that growth of domestic industrial production i.e., real economic activity of the economy has no significant role as yet in explaining variations in Indian stock returns.

Now in order to check the presence of any remaining autocorrelation in the residuals of the estimated equations, Ljung-Box test was carried out, and for each of the three return series, no autocorrelation for lags up to 12 was found. Diagnostic testing of the residuals thus suggesting presence of no further linear dependence in the residuals, what we need to finally check from consideration of appropriate specification of the mean is if there still remains omission of any lagged higher order values of the variables. Accordingly, we have applied Lumsdaine-Ng test (1999). This test requires including terms like $\hat{w}_{t-1}$, $\hat{w}_{t-1}{}^{2}$ and $\hat{w}_{t-1}{}^{3}$ in the extended mean specification, where $\hat{w}_{t-1}$ is the (t-1) th recursive residual obtained from estimated models – (4.12)- (4.14) . These estimated models for the three returns – based on OLS estimation, have been obtained as follows:

BSESENSEX

$$\hat{r}_t = -0.312r_{t-1} - 0.252r_{t-7} - 0.024CPI_{t-1} - 0.881FRX_{t-1} + 283NSD_t + 0.515DPE_t + 0.039D12$$

(3.134)*** (2.798)*** (2.994)*** (1.514) (4.412)*** (5.574)*** (1.565)

$$+0.226\hat{w}_{t-1} - 0.673\hat{w}_{t-1}{}^{2} - 18.357\hat{w}_{t-1}{}^{3} \qquad (4.15)$$

(0.936) (0.396) (0.728)

BSE100

$$\hat{r}_t = 0.191r_{t-6} - 0.185r_{t-7} - 0.030CPI_{t-1} + 0.405NSD_t - 0.529TBL_t - 0.072\hat{w}_{t-1}$$

(1.927)* (1.827)* (2.081)** (4.804)*** (2.429)** (0.298)

$$+0.423\hat{w}_{t-1}{}^{2} - 4.245\hat{w}_{t-1}{}^{3} \qquad (4.16)$$

(0.390) (0.334)

NIFTY

$$\hat{r}_t = -0.198r_{t-7} - 0.020CPI_{t-1} + 0.305NSD_t - 0.389TBL_{t-1} + 0.050D2 + 0.038D12 - 0.106\hat{w}_{t-1}$$

(1.910)* (2.379)** (4.323)*** (2.017)** (2.023)** (1.537) (0.424)

$$-0.342\hat{w}_{t-1}{}^{2} + 2.656\hat{w}_{t-1}{}^{3} \qquad (4.17)$$

(0.242) (0.122)

((i) All the macro and financial variables in these equations refer to their respective stationary values although their abbreviations/symbols do not indicate the same.

(ii) The values in parentheses indicate corresponding absolute values of the $t-$statistic; *, ** and *** indicate significance at 10%, 5% and 1% levels of significance, respectively).

It is evident from these estimated equations that none of the coefficients associated with $\hat{w}_{t-1}$, $\hat{w}_{t-1}{}^{2}$ and $\hat{w}_{t-1}{}^{3}$ is significant even at 5 percent level of significance in any of the three monthly series, and hence it can be concluded that there is no misspecification in the conditional mean specification of the three series.

Now that it is empirically established that the conditional means of the monthly returns on BSESENSEX, BSE 100 and NIFTY have been properly specified, we now test if there is any conditional variance in the residuals of equations (4.12) through (4.14). The usual LM test statistic values for the three residual series for the estimated equations have been obtained as 2.229, 0.006, 1.506. By comparing the test statistic values with their corresponding χ^2 critical values, it is very obvious that these values are not significant at all and hence we can conclude that there is no conditional. heteroscedasticity in the residuals of these models.

Finally, we have applied the BDS test for testing if the residuals are indeed i.i.d. or that there still remains dynamics in the higher-order moments (say, third or fourth) in the residuals of models in (4.12), (4.13) and (4.14). The results of this test are presented in Table 4.5. We observe from the BDS test statistics values for the three stock returns that while in some cases the null of i.i.d. property for the residuals is not rejected , but in some others it is rejected.

Table 4.5 : BDS Test Statistic Values for Monthly Residuals

		Monthly Return Series		
ξ/σ	*m*	BSESENSEX	BSE 100	NIFTY
0.5	2	-3.253	-2.596	-5.076
0.5	3	-5.982	-4.770	-5.264
0.5	4	-2.961	-2.449	-2.577
0.5	5	-1.673[a]	-1.433[a]	-1.440[a]
1	2	-0.132[a]	-2.421	5.105
1	3	1.746[a]	-0.280[a]	1.388[a]
1	4	-4.318	3.118	-0.361[a]
1	5	-2.662	2.485	-2.699

Note : The values of BDS test statistic from standardized residuals are compared with the simulated values given in Brock et al. (1991).
Values with superscript 'a' indicate non-significance at 5% level All others are significant at 5% level of significance.
ξ*, m and* σ *stand for distance, embedding dimension and the standard deviation of the linearly filtered data, respectively.*

Thus, we can conclude that although no second-order dependence is present in the residuals of the final models, some higher-order dependencies may have remained in the residuals.

4.4 Conclusions

This chapter is basically concerned with studying the predictability aspect of Indian stock returns (at monthly frequency) from consideration of the roles of major macrovariables and financial ratios in such predictions. To put it somewhat differently, the purpose, in this chapter, is to provide, based on monthly data, an adequate econometric model for Indian stock returns for the purpose of forecasting. To this end, a predictive regression approach has been applied to identify the set of relevant macro and financial variables. This has been done from considerations of in-sample forecasts and out-of-sample tests of return predictability. These have involved, *inter alia*, use of Theil's U and McCracken's $MSE-F$ and Clark and McCracken's $ENC-NEW$ test statistics. As in Chapter 3, due consideration, has been given to appropriate specification of both the conditional mean and conditional variance of the final model for returns. This study has been carried out for the returns on three Indian stock indices *viz.,* BSESENSEX, BSE 100 and NIFTY.

The analysis based on predictive regression approach has found a common set of macrovariables as having predictive ability for all the three returns series. These macrovariables are: inflation rate, change in nominal exchange rate, NASDAQ composite return, growth of foreign direct investment and changes in long-term interest rate as well as in fiscal deficit of the central government. In addition, change in price –earnings ratio has also been found to have significant predictive ability, but for returns on BSESENSEX only. However, when final models were estimated incorporating seasonal dummies and other relevant terms incorporating specifications, we have obtained that lagged values of both inflation rate and change in nominal exchange rate and the contemporaneous values of NASDAQ composite return and change in price-earning ratio are significant for returns on BSESENSEX. As regards the other two return series *viz.,* BSE 100 and NIFTY, inflation rate, returns on NASDAQ composite index and change in long-term interest rate have been found to be significantly present, the last two being in contemporaneous values.

It may be pointed out that our finding on inflation rate having significant negative effect is consistent with similar findings for the developed countries, and this is quite obvious from economic reasonings as well. Another interesting finding relates to the role of long-term interest rate. Although long-term interest rate change has been found to have no significant effect on BSESENSEX, but for the other two returns series, we have found negative relations between stock return and interest rate cahnge. This shows that India’s present policy of lowering interest rate has favourable effect on stock market. The significant positive impact of NASDAQ composite return in explaining variations in Indian stock returns implies integration of the Indian capital market with the world bourses. Finally, we have found that volatility at the monthly level data is not significant at all for all the three series, and hence no further analysis with volatility specification was required to be done.

APPENDIX

Table A4.1: Results of In-Sample and Out-of-Sample Predictive Ability of the 13 Macrovariables for returns on NIFTY (Out-of-sample period :January 2001 – December 2002)

Horizon	1	3	6	12	1	3	6	12
	I. Domestic Industrial production (IP)				II. Money supply (MS)			
$\hat{\beta}$	-0.006	-0.003	0.006	0.007	0.007	-0.005	0.010	0.016
$t-$statistic	-0.747	-0.175	0.371	0.646	0.833	-0.491	1.077	0.953
	(0.220)	(0.435)	(0.380)	(0.328)	(0.210)	(0.315)	(0.179)	(0.242)
R^2	0.012	0.001	0.006	0.001	0.014	0.002	0.009	0.004
Theil's U	1.010	1.024	0.997	1.011	1.018	1.000	1.002	0.988
$MSE-F$	-0.457	-1.014	0.111	-0.294	-0.855	-0.008	-0.061	0.304
	(0.403)	(0.142)	(0.264)	(0.333)	(0.222)	(0.683)	(0.568)	(0.137)
$ENC-NEW$	–0.159	-0.472	0.057	-0.143	-0.296	0.012	0.024	0.180
	(0.391)	(0.081)	(0.343)	(0.258)	(0.233)	(0.590)	(0.395)	(0.168)
	III. Consumer price index (CPI)				IV. Domestic oil price (OILD)			
$\hat{\beta}$	-0.020	0.001	-0.030	0.011	-0.002	0.003	-0.024	-0.043
$t-$statistic	-2.335	0.08	-1.527	0.385	-0.02	0.189	-2.046	-0.043
	(0.006)	(0.474)	(0.104)	(0.419)	(0.514)	(0.444)	(0.054)	(0.157)
R^2	0.071	0.0008	0.026	0.002	0.005	0.001	0.026	0.037
Theil's U	0.973	1.003	1.041	1.040	1.011	1.011	1.022	1.115
$MSE-F$	1.347	-0.120	-1.459	-0.971	-0.528	-0.474	-0.815	-2.535
	(0.059)	(0.645)	(0.203)	(0.266)	(0.356)	(0.418)	(0.262)	(0.087)

Table A4.1 (contd.)

Horizon	1	3	6	12	1	3	6	12
ENC – NEW	1.209	-0.057	-0.552	-0.455	-0.238	-0.208	-0.291	-1.047
	(0.051)	(0.543)	(0.182)	(0.200)	(0.730)	(0.345)	(0.253)	(0.051)
	V. Nominal exchange rate (FRX)				VI. NASDAQ composite index (NSD)			
$\hat{\beta}$	-0.012	-0.017	-0.016	-0.028	0.014	0.013	0.018	0.027
$t-$statistic	-1.493	-0.967	-0.882	-1.522	1.624	1.008	1.372	1.763
	(0.052)	(0.193)	(0.227)	(0.121)	(0.049)	(0.175)	(0.129)	(0.105)
R^2	0.033	0.020	0.014	0.016	0.038	0.010	0.015	0.012
Theil's U	1.009	1.186	1.143	1.279	0.959	0.988	1.010	1.072
$MSE-F$	-0.429	-6.367	-4.449	-5.050	2 .095	0.528	-0.360	-1.695
	(0.431)	(0.000)	(0.012)	(0.007)	(0.034)	(0.160)	(0.378)	(0.110)
ENC – NEW	0.407	-1.867	-1.647	-1.896	1.267	0.396	-0.049	-0.060
	(0.140)	(0.000)	(0.005)	(0.001)	(0.056)	(0.173)	(0.456)	(0.060)
	VII. Foreign direct investment (FDI)				VIII. Foreign institutional investment (FII)			
$\hat{\beta}$	0.003	0.006	0.001	0.007	-0.002	0.009	0.008	0.029
$t-$statistic	0.308	0.844	0.122	0.470	-0.196	0.719	0.537	1.594
	(0.411)	(0.185)	(0.453)	(0.342)	(0.433)	(0.279)	(0.325)	(0.117)
R^2	0.006	0.003	0.005	0.0008	0.005	0.004	0.007	0.011
Theil's U	1.040	1.014	1.009	1.051	1.006	1.002	1.034	1.026
$MSE-F$	-1.829	-0.585	-0.330	-1.232	-0.317	-0.095	-1.243	-0.639
	(0.082)	(0.144)	(0.187)	(0.017)	(0.456)	(0.569)	(0.147)	(0.257)

Table A4.1 (contd.)

Horizon	1	3	6	12	1	3	6	12
ENC – NEW	–0.665	-0.215	-0.145	-0.550	-0.144	-0.033	-0.560	-0.209
	(0.080)	(0.124)	(0.146)	(0.009)	(0.377)	(0.506)	(0.096)	(0.236)
IX. Short-term interest rate 1 (TBS)					X. Short-term interest rate 2 (WCALL)			
$\hat{\beta}$	0.004	0.005	0.003	-0.001	0.004	-0.003	0.009	-0.002
$t-$statistic	0.511	0.425	0.223	-0.069	0.528	-0.204	0.737	0.312
	(0.295)	(0.305)	(0.386)	(0.516)	(0.295)	(0.413)	(0.228)	(0.579)
R^2	0.008	0.003	0.005	0.000	0.008	0.001	0.008	0.000
Theil's U	1.00	1.00	1.00	1.005	0.995	0.999	1.005	1.004
$MSE-F$	-0.031	-0.02	-0.06	-0.137	0.236	0.052	-0.193	-0.097
	(0.715)	(0.674)	(0.534)	(0.325)	(0.198)	(0.265)	(0.302)	(0.350)
ENC – NEW	-0.001	-0.002	-0.029	-0.067	0.123	0.027	- 0.081	-0.047
	(0.619)	(0.605)	(0.444)	(0.252)	(0.276)	(0.356)	(0.252)	(0.302)
XI. Long-term interest rate (TBL)					XII. Foreign exchange reserve (FOREXR)			
$\hat{\beta}$	0.014	0.035	0.035	0.034	0.002	-0.008	0.004	-0.001
$t-$statistic	1.562	2.560	1.803	2.010	0.277	-0.834	0.418	-0.152
	(0.0610)	(0.019)	(0.066)	(0.068)	(0.398)	(0.177)	(0.363)	(0.392)
R^2	0.036	0.067	0.042	0.018	0.006	0.005	0.006	0.000
Theil's U	1.033	0.962	0.941	0.870	1.002	1.006	0.999	0.998
$MSE-F$	-1.520	1.796	2.435	4.172	-0.113	-0.268	0.045	0.027
	(0.110)	(0.021)	(0.014)	(0.000)	(0.588)	(0.302)	(0.262)	(0.246)
ENC – NEW	-0.226	1.956	1.595	2.312	-0.054	-0.122	0.024	0.024
	(0.282)	(0.008)	(0.021)	(0.004)	(0.482)	(0.214)	(0.316)	(0.315)

Table A4.1 (contd.)

Horizon	1	3	6	12	1	3	6	12
XIII. Fiscal deficit of the Central government					XIV. Term Spread			
$\hat{\beta}$	-0.009	-0.028	-0.032	-0.036	0.004	0.014	0.043	0.040
$t-$statistic	-1.075	-5.904	-4.395	-3.276	0.512	0.801	1.940	0.815
	(0.137)	(0.006)	(0.019)	(0.056)	(0.294)	(0.250)	(0.108)	0.340
R^2	0.020	0.054	0.045	0.028	0.006	0.012	0.071	0.029
Theil's U	1.230	0.928	0.921	0.969	1.003	0.987	0.905	1.129
$MSE-F$	-8.139	3.552	3.385	0.837	-0.123	0.579	4.189	-2.804
	(0.013)	(0.006)	(0.005)	(0.078)	(0.358)	(0.222)	(0.094)	0.719
$ENC-NEW$	0.299	5.536	2.653	0.459	-0.013	0.428	2.767	-1.225
	(0.194)	(0.003)	(0.010)	(0.122)	(0.405)	(0.275)	(0.134)	(0.806)

Notes: In this table, all the macro and financial variables along with their abbreviations/symbols have been mentioned in terms of their descriptions in level values although these values refer to their growths/changes; $\hat{\beta}$ and $t-$statistic are the OLS estimate of β in equation (4.1) and its corresponding $t-$statistic value, respectively; R^2 is the goodness of fit measure in equation (4.1); Theil's U the ratio of the RMSE for the out-of-sample forecasts for the unrestricted model to the RMSE for the out-of-sample forecast for the restricted model; $MSE-F$ and $ENC-NEW$ are the out-of-sample statistics given in equations (4.2) and (4.3), respectively; $p-$ values are given in parentheses.

CHAPTER 5

Predictability in the Indian Stock Market : A Cointegration Approach

5.1 Introduction

In the preceding chapter, we have investigated the relationship between Indian stock returns and growths/changes in macroeconomic and financial variables. However, in that chapter we have analysed the role of macroeconomic variables and financial ratios[1] in explaining the behaviour of the Indian stock prices by taking all the variables including the stock prices in their stationary values so that the predictability aspect of the Indian stock market was studied there in terms of stationary stock returns. Obviously, forecasts based on such a model would be meaningful for short-run periods only. Now, our objective in this chapter is to study the relationship, if any, at the level values of the variables i.e., between non-stationary indices and the non-stationary macro and financial variables, by using the cointegration methodology so that existence of such a relation(s) would suggest predictability in the Indian stock market in the long-run sense.

Over the years there have been considerable research efforts to understand the issue of market efficiency in the context of long-run for several asset prices (Baillie and Bollerslev (1989) and Coleman (1990)). There is a notion following Granger (1986) that cointegration implies market inefficiency. But, this has been challenged by some researchers who have pointed out that the traditional approach to efficient markets in

[1] As in Chapter 4, financial variables would mean the three financial ratios considered in our study. For the sake of convenience of expression, we would be using these two terms *viz.*, financial variables and financial ratios, concurrently.

which changes in asset prices are unpredictable lacks necessary economic content. Further, they have argued that market efficiency can be defined more usefully as a situation where there is a lack of arbitrage opportunity in the markets (see, in this context, Levich (1985) and Ross (1987)). However, there is a consensus in the literature that the existence of a long-run (cointegrating) relationship immediately invalidates the martingale property which is the standard model for market efficiency for asset prices. As our main focus in this thesis is on predictability in the Indian stock market, the arguments made by Caporale and Pittis (1998) is noteworthy. They have pointed out that whatever concerns one might have about the identification of a cointegration relationship with market inefficiency, cointegration tests can still be usefully employed to investigate the predictability of asset returns.

It may be relevant to state that barring a few like Ibrahim(1999), most of such studies on cointegration involving stock prices and macro and financial variables, are confined to bivariate set-up where relationship between a stock price index and one relevant variable has been studied. While such bivariate studies may be useful, although to a limited extent, from the point of view of economic understanding , these would be constrained by the omission of other relevant variables from such relations and, to that extent, the cointegration result may be misleading since there may very well be other variables having similar comovements. Our approach therefore, is to study long-run relations involving a specific stock index and all other relevant macro and financial variables.

To examine the long-run relationship between stock returns and macroeconomic and financial variables, we have applied the standard VAR-based cointegration methodology developed by Johansen (1988, 1991, 1995) and Johansen and Juselius (1990). But, since cointegration relations do not appear explicitly in the VAR framework, a more convenient modelling set-up obtained by rewriting the VAR model, known as the vector error correction model (VECM), is used for cointegration analysis. We carry out cointegration analysis separately for each of the three stock indices *viz.*, BSESENSEX, BSE 100, NIFTY[2] and the respective appropriate set of macrovariables and financial ratios. The purpose behind these separate cointegration exercises is to check if the predictability status prevailing in the Indian stock market from the point of view of cointegrated relationship involving a stock index and the variables, is the same for all the three stock indices considered. The conclusion, in that case, would be robust in the sense that the particular choice of the index would not be really important to conclude about the long-run behaviour of stock price index and macro and financial variables. In case the findings differ, the inference would then be stock index specific. The data for this analysis are at monthly frequency and have been taken for the period covering April 1996 to December 2002. This chapter has been organized as follows. The next section presents the cointegration methodology very briefly. The third section i.e., Section 5.3, deals with the empirical findings and the last section presents some concluding remarks.

5.2 Cointegration methodology

Since cointegration necessitates that all the variables be integrated of order one, the first step in testing for cointegration requires testing for the order of integration for each of the variables involved in this exercise. The standard unit root tests like the augmented Dickey

[2] It is for the same reason as stated in Chapter 4 *viz*., lack of availability of data on the financial ratios, that this exercise could not be done for the DOLLEX series.

– Fuller (ADF) test by Said and Dickey (1984) and/or Phillips – Perron (PP) test (1988) are applied to each of the macro and financial variables to infer about their orders of integration. If the results of unit root tests indicate that each of the variables is $I(1)$, then we test for cointegration. We now briefly describe the Johansen's procedure for testing for cointegration and estimation of a system of equations. Further, we mention the three different cases arising out of consideration towards inclusion of deterministic terms in the cointegrating equation, and also the differences between what are called the 'intercept version' and the 'mean and/or trend adjusted version'.

Johansen's procedure: This procedure begins with a vector autoregressive (VAR) model. For a set of K time series variables, $y_t = (y_{1t} y_{2t} \dots y_{Kt})'$, each being $I(1)$, a VAR model captures their dynamic interactions. The basic model of order p, called $VAR(p)$, has the form

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + u_t, t = 1, \dots, n \tag{5.1}$$

where $A_i (i = 1,2,....., p)$ are $(K \times K)$ coefficient matrices and $u_t = (u_{1t} u_{2t} \dots u_{Kt})'$ is the unobservable error term. u_t is usually assumed to be an independent white noise process with zero mean and time invariant, positive definite covariance matrix $E(u_t u_t') = \Sigma_u$ i.e., u_t's are independent stochastic vectors with $u_t \sim (0, \Sigma_u)$. Several extensions of this basic model (5.1) are usually necessary to represent the main characteristics of a data set of interest. It is quite obvious that including deterministic terms, such as an intercept, a linear time trend or seasonal dummy variables may be required for a proper representation of the data generating process (DGP). We shall consider these extensions later.

Johansen's procedure applies the method of maximum likelihood (ML) to the $VAR(p)$ model assuming that the errors i.e., $u_t's$, are Gaussian. Although, the VAR model specified in (5.1) is general enough to accommodate variables with stochastic trends, it is not the most suitable representation if one is primarily interested in cointegrating relations because they do not appear explicitly in the representation. By subtracting y_{t-1} from both sides of (5.1) and then rearranging terms, (5.1) can easily be shown to reduce to

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \ldots\ldots.. + \Gamma_{p-1} \Delta y_{t-p+1} + u_t, \tag{5.2}$$

where $\Pi = -\left(I_K - \sum_{i=1}^{p} A_i \right)$ and $\Gamma_i = -\sum_{j=i+1}^{p} A_j, i = 1,2,\ldots, p-1$. It is worthwhile to note that (5.2) is nothing but the vector error correction model (VECM) or, more appropriately, VECM of order $p-1$ (notationally, VECM $(p-1)$) of (5.1), and therefore, a more convenient modelling set-up for cointegration analysis.

By our assumption of y_t being a $K \times 1$ vector of $I(1)$ variables, $\Delta y_{t-1},\ldots,\Delta y_{t-p+1}$ are all $K \times 1$ vectors of $I(0)$ variables, but y_{t-1} in the right hand side of (5.2) is $I(1)$. Hence, in order that the system of equations is consistent, Πy_{t-1} must also be $I(0)$. Thus, it contains the cointegrating relations. Γ_i, i = 1, 2,..., p-1, are often referred to as the short-run parameters, and Πy_{t-1} is sometimes called the long-run part of VECM. It may be noted that a VECM is a restricted VAR that has cointegration restrictions built into the specification so that it is used for nonstaionary series that are known to be cointegreted. The VECM specification restricts the long-run behaviour of the endogenous variables to converge to their cointegrating relationships while allowing a wide range of short-run dynamics. The cointegration term is the error correction term as it captures the deviation from the long-run equilibrium through a series of partial short-run adjustments.

Now, for Πy_{t-1} to be $I(0)$, Π should not be of full rank. Let its rank be *r*. So, Π can be written as product of two matrices α and β i.e., $\Pi = \alpha\beta'$ where each of α and β is a $K \times r$ matrix of rank r. Then $\beta' y_{t-1}$ are the r cointegrating relations. The rank of Π is, therefore, referred to as the cointegrating rank of the system. β' is the matrix of coefficients of the cointegrating vectors (or, cointegrating matrix, in short) and α is the matrix of weights attached to the cointegrating relations, or, sometimes called the loading matrix[3]. The matrices α and β are not unique, and thus there are many possible α and β matrices that contain the cointegrating relations or linear transformations of them. In fact, using any nonsingular $r \times r$ matrix B, we have a new loading matrix αB and new cointegrating matrix $\beta B^{'-1}$, which satisfy $\Pi = \alpha B(\beta B^{'-1})'$. Consequently, cointegrating relations with economic content cannot be extracted purely from the observed time series. Some non-sample information is required to identify them uniquely.

It is obvious that in the VECM framework, the cointegration rank r has to be determined and the lag order is also required to be known. Insofar as the latter issue is concerned, similar procedures as followed for determining lag order of an univariate stationary time series, are available. In other words, sequential testing procedures of Hall

---

[3] In the particular case where K=2, the scalar α is often called the parameter representing the speed of adjustment to long-run.

(1994) of general to specific kind and the generalized version of information-based model selection criteria are used.

Johansen (1988, 1991) has proposed two test statistics for testing the presence of cointegration , and these are known in the literature as the trace test and the maximum eigenvalue test. Sequential testing procedures based on likelihood ratio (LR)–type tests are possible statistical tools for this testing problem. It is important to note at this point that the deterministic trend terms have an important impact on the null distribution of LR-type test of cointegration. Including deterministic terms such as an intercept, a linear time trend or seasonal dummy variables are required for an appropriate representation of the DGP. One way to include deterministic terms is to assume that the nonstationary time series y_t comprises two additive terms i.e.,

$$y_t = \mu_t + \tilde{y}_t \quad , \tag{5.3}$$

where μ_t represents the deterministic part and $\tilde{y}_t$ the stochastic part. The latter i.e., $\tilde{y}_t$ is assumed to have a VAR or a VECM representation as in (5.1) and (5.2). Now, for the purpose of presenting cointegration tests, let us consider $\mu_t = \mu_0 + \mu_1 t$ so that $y_t = \mu_0 + \mu_1 t + \tilde{y}_t$. There are three possible cases depending on particular assumptions about the deterministic part of y_t.

Case(i): μ_0 arbitrary and $\mu_1 = 0$ i.e., there is a constant mean but no deterministic trend term;

Case(ii): A linear deterministic term in the DGP so that $\mu_1 \neq 0$ but $\beta' \mu_1 = 0$ and μ_0 is arbitrary;

Case(iii): Both μ_0 and μ_1 are arbitrary.

We first consider the first case. In this case $y_t - \mu_0 = \tilde{y}_t$, and hence $\Delta y_t = \Delta\tilde{y}_t$. Thus, from consideration of VECM of $\tilde{y}_t$, the VECM of y_t can be either of the following two forms-- the mean adjusted form due to Saikkonen and Luukkonen (1997) and Saikkonen and Lütkepohl (2000b) (equation (5.4) below), and the intercept form due to Johansen (1995) (equation (5.5) below):

$$\Delta y_t = \mu_0^* + \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + u_t$$

$$= \Pi^* \begin{bmatrix} y_{t-1} \\ 1 \end{bmatrix} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + u_t \ , \qquad (5.4)$$

and $$\Delta y_t = \Pi(y_{t-1} - \mu_0) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + u_t \qquad (5.5)$$

where $\Pi^* = \lfloor \Pi : \mu_0^* \rfloor$ is of order ($K \times (K+1)$) and $\mu_0^* = -\Pi\mu_0$. It may be noted that in the latter case, the intercept term can be absorbed into the cointegrating relations and thus $\Pi^* = \alpha\beta^{*\prime}$ has rank *r*. Both the VECM versions can obviously be used for the purpose of testing the cointegration rank. The intercept version (5.4) has been considered by Johansen (1995) and the critical values for the LR test, called the trace test, have also been provided by him as well as by Osterwald-Lenum (1992).

The trace test statistic, denoted by λ_{trace}, under the null hypothesis of (at most) r cointegrating vectors is given by

$$\lambda_{trace} = -n \sum_{i=r+1}^{K} \ln(1 - \hat{\lambda}_i) \ , \qquad (5.6)$$

where $\hat{\lambda}_{r+1}, \hat{\lambda}_{r+2}, \ldots, \hat{\lambda}_K$ are the eigenvalues obtained by applying reduced rank (RR) regression technique to (5.4). This test requires a sequence of null hypotheses beginning with $r = 0$ to be tested until the null hypothesis cannot be rejected for the first time. The other cointegration test proposed by Johansen, *viz*., the maximum eigenvalue test, denoted by $\lambda_{\max}$, tests the null hypothesis of r cointegrating vectos versus the alternative of $(r+1)$ cointegrating vectors, and this is based on the following statistic:

$$\lambda_{\max} = -n \ln\left(1 - \hat{\lambda}_{r+1}\right). \tag{5.7}$$

Like the trace test, this test statistic also has a non-standard limiting distribution and this distribution depends on the specification of the deterministic term. Further, the critical values are available in Johansen (1995). While Johansen and Juselius (1990) have suggested that the maximum eigenvalue test may perform better, Lütkepohl *et al*. (2001) have found, based on their study comparing between trace and maximum eigenvalue tests, that the former sometimes has slightly more distorted sizes than the latter in small samples but at the same time it i.e., the former, may also have some power advantages. As regards the mean adjusted form of VECM given in (5.5), two-step procedures are considered in which the mean term, μ_0, is first estimated by a feasible GLS method, and then substituting the estimate of μ_0 thus obtained in (5.5) a LR-type test based on a RR regression of (5.5) is applied. The resulting test statistic has an asymptotic distribution which is different from that obtained from the intercept form. The critical values of this asymptotic null distribution are given in Saikkonen and Luukkonen (1997) and Saikkonen and Lütkepohl (2000b). The VECM forms used in the LR-type test for cointegration for the other two cases, namely, *Case (ii)* and *Case (iii)*, can similarly be obtained (see Lütkepohl (2004) for these details).

It may be worthwhile to point out that a proper choice of the deterministic term is important for tests of the cointegrating rank. Of course, if a linear time trend is regarded as a possibility, this could be included in the fully general form to be on the safe side. However, as pointed out by Doornik *et al*. (1998) and also Hubrich *et al*. (2001), including an unnecessary trend term may result in loss of power. Therefore, it is necessary to be careful about the proper trend specification in applied work. This is usually done based on subject matter considerations or a visual inspection of the plots of the time series under study or findings of some statistical exercises.

It may also be important to note that seasonal dummy variables may as well be added to the deterministic terms considered in the above three cases; but this will not change the asymptotic properties of the tests (*cf.* Lütkepohl and Krätzig (2004)). However, other dummy variables like those representing structural shifts lead to changes in the asymptotic distributions of the Johansen – type tests for the cointegrating rank. In fact, in case of structural shift in the level of the DGP, the null distribution, as shown by Johansen *et al.* (2000), depends on the unknown points where the shifts have indeed occurred in the period over which observations are available. In contrast, Saikkonen and Lütkepohl (2000a) have extended their approach to DGPs with level shifts and have shown that, for their tests, the limiting null distributions do not change [see Lütkepohl (2004) for details].

5.3 Empirical findings

This section presents the empirical findings and relevant discussions. At the very beginning of a cointegration exercise, it is required to be checked that all the variables are $I(1)$. As, already given in Table 4.1 of the preceding chapter, both the ADF and PP tests for unit roots, clearly show that along with the three stock indices *viz.*, monthly closing prices of BSESENSEX (X1), BSE 100 (X2), NIFTY (X3)[4], all the 13 macro variable and 3 financial ratio series are $I(1)$ in their level values.

Now, the first step towards carrying out the cointegration exercise, is to decide on the variables to be included in the study. Since the set of variables found to influence any stock market is rather large while the number of observations at monthly frequency being somewhat moderate i.e., 81 for our study, it is not computationally feasible to run cointegration even with a choice of a small value for the lag size. What we have done, therefore, is that we have considered those variables at their $I(1)$ level values which have been found to have significant effects in predicting stock returns (*cf.* Chapter 4) along with all possible combinations of the remaining variables of the set subject to, of course, computations being possible without too many identifying restrictions. Based on the findings of such cointegration exercises, we finally choose and report the most economically meaningful and plausible cointegrating relation, provided, of course, at least one cointegration is found to exist.

Insofar as actual computations are concerned, we first find the cointegrating rank

---

[4] Notations X1, X2 and X3 for the three index series are being used for the purpose of convenience in tabular representations.

r and then estimate the cointegrating vector β for the system of equations involving stock price and macro and financial variables for each of the three stock indices. Using Johansen's ML – based reduced rank (RR) regression procedure as discussed in the previous section, λ_{trace} and $\lambda_{\max}$ test statistics were computed to determine the number of cointegrating vectors. As regards the choice between the two versions for treating the deterministic terms – the mean / trend adjusted version and the intercept version- we have applied Johansen's intercept version. The computations have been done by using the JMulTi software available in Lütkepohl and Krätzig (2004).

The values of the two test statistics for cointegration model with the five macrovariables and one of the three stock indices along with a constant and seasonal dummy variables are given separately for each of three stock indices. This set of macrovariables i.e., domestic industrial production (IP), consumer price index (CPI), nominal exchange rate (FRX), foreign direct investment (FDI) and long-term interest rate (TBL) at their level values and any one of the stock index (X1/ X2/ X3) constitute the set of variables for the cointegrated system of equations since cointegration has not been found with any other macro and financial variables. It may further be noted that the set of variables is the same for all the three indices and that no financial ratio has been found to be relevant for cointegration analysis. Since in recent years emphasis has been given on appropriate consideration of deterministic terms such as constant mean, linear trend and seasonal dummy variables in cointegration analysis for a proper representation of the DGP, we have carried out our analysis considering various combinations of these components of the deterministic term and reported the one which is found to be quite satisfactory.

We observe from the following table that for the system involving X1/X2/X3, IP, CPI, FRX, FDI and TBL, both the λ_{trace} and $\lambda_{\max}$ statistics indicate that the null of no cointegration (i.e., $H_0: r = 0$) is rejected at 5% level of significance.

Table 5.1 Results of Cointegration Tests

	Test statistic value					
Null hypothesis	BSESENSEX (X1)		BSE 100 (X2)		NIFTY(X3)	
	λ_{trace}	$\lambda_{\max}$	λ_{trace}	$\lambda_{\max}$	λ_{trace}	$\lambda_{\max}$
$r = 0$	176.09*	76.03*	180.68*	79.05*	176.77*	76.25*
$r = 1$	100.06*	57.71*	101.63*	58.76*	100.52*	57.87*
$r = 2$	42.35	22.37	42.87	25.87	42.65	23.79
$r = 3$	19.98	11.06	17.00	9.72	18.86	11.09
$r = 4$	8.92	6.30	7.28	4.88	7.77	5.36
$r = 5$	2.62	2.62	2.40	2.40	2.41	2.41

Notes: * indicates significant values at 5% level of significance. Critical values have been taken from Osterwald -Lenum (1992, Table 1*, p. 467) and Johanesn (1995). The lag order for each system of equations has been taken to be 1 since this is the lag value which has been optimally determined by the generalized AIC criterion.

For instance, the value of λ_{trace} statistic corresponding to $r = 0$ is 176.09 which is higher than the corresponding critical value of 102.14 at 5% level of significance, and hence the conclusion is that the null of 'no cointegration' is rejected in favour of the alternative of 'cointegration' involving BSESENSEX, domestic industrial production, consumer price index, nominal exchange rate, foreign direct investment and long-term interest rate. Next we find that the null of $r = 1$ is also rejected by the λ_{trace} statistic at the 5% level of significance as 100.06 is higher than the corresponding critical value, 76.07. However, the null of $r = 2$ cannot be rejected since the computed value of 42.35 is smaller than the corresponding critical value i.e., 53.12. The cointegration test thus suggests that that there exists two cointegrating relations involving BSESENSEX, domestic industrial production, consumer price index, nominal exchange rate, foreign direct investment and long-term interest rate. The same conclusion is also obtained on the basis of $\lambda_{\max}$ test statistic since its computed values for the null of $r = 0$ and $r = 1$ i.e., 76.03 and 57.71, respectively, are greater than the corresponding critical values of 40.30 and 30.40 at 5% level of significance. However, as in the case of λ_{trace}, the null of $r = 2$ cannot be rejected by the $\lambda_{\max}$ test statistic as the test statistic value 22.37 is now smaller than the corresponding critical value, 28.14, at 5% level of significance. Thus, it is clearly found that there exists two cointegrating relations for the system method comprising BSESENSEX and the other macroeconomic variables. In this context it is worthwhile to mention that as far as the choice of lag order is concerned, it has been found that all the criteria for selecting the proper lag length, including the generalized AIC, suggest inclusion of only the first lag although in the computations the maximum number of lags allowed was 6. Accordingly, we have reported the empirical results on cointegration with lag order 1.

We have obtained similar results for the other two system of dynamic equations where BSE 100 and NIFTY are the relevant stock indices. In both the cases, λ_{trace}

statistic shows that the null of 'no cointegration' is rejected at 5% level of significance. Again, like the earlier results, the null of rank 1 is rejected for both the systems. However, the null of rank being 2 cannot be rejected for both the system of variables. The same conclusion on the cointegrating rank is obtained by the $\lambda_{\max}$ as well. As before, the results have been reported with lag order 1. We thus conclude that there are two cointegrating relations i.e. $r = 2$ for all the three system methods for the three indices. As the cointegration rank r and lag order are now determined, we present the results on estimated VECM to understand the nature of interdependence and short term dynamics involving these variables. The equations (5.8) through (5.10) present the VECMs for the systems with the stock index being BSESENSEX, BSE 100 and NIFTY, respectively. As already mentioned, we have included the seasonal dummies in the VECM to capture the month-of-the-year effect.

$$
\begin{bmatrix} \Delta X1_t \\ \Delta IP_t \\ \Delta CPI_t \\ \Delta FRX_t \\ \Delta FDI_t \\ \Delta TBL_t \end{bmatrix}
=
\begin{bmatrix}
-0.106 & 6.502 \\
(-2.297) & (2.253) \\
0.001 & -0.055 \\
(0.849) & (-0.895) \\
0.002 & -0.134 \\
(4.544) & (-5.009) \\
0.000 & -0.010 \\
(1.852) & (-1.682) \\
-0.045 & 1.267 \\
(-2.286) & (1.024) \\
0.000 & -0.009 \\
(1.859) & (-1.953)
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 193 & -1162 & 43 & 1875 \\
(.....) & (....) & (3.14) & (-2.06) & (7.26). & (3.72) \\
0 & 1 & 3.81 & -27.32 & 0.67 & 40.45 \\
(.....) & (....) & (3.99) & (-3.13) & (7.23) & (5.18)
\end{bmatrix}
\begin{bmatrix} X1_{t-1} \\ IP_{t-1} \\ CPI_{t-1} \\ FRX_{t-1} \\ FDI_{t-1} \\ TBL_{t-1} \end{bmatrix}
+
$$

$$
\begin{bmatrix}
-0.054 & -2.706 & -23.76 & -40.846 & 0.117 & 161.346 \\
(-0.453) & (-0.495) & (-2.496) & (-0.657) & (0.447) & (1.870) \\
-0.000 & -0.219 & 0.035 & -1.236 & -0.005 & 2.046 \\
(-0.042) & (-1.880) & (0.173) & (-0932) & (-0.953) & (1.112) \\
-0.002 & -0.014 & 0.364 & 0.116 & 0.001 & -0.206 \\
(-1.962) & (-0.271) & (4.120) & (0.201) & (0.382) & (-0.256) \\
-0.000 & 0.009 & -0.003 & -0.155 & 0.000 & 0.023 \\
(-2.038) & (0.844) & (-0.168) & (-1.250) & (0.286) & (0.133) \\
0.063 & -4.520 & -0.917 & -29.251 & 0.051 & 38.144 \\
(1.225) & (-1.932) & (-0.225) & (-1.097) & (0.451) & (1.032) \\
-0.000 & -0.003 & 0.027 & 0.053 & 0.000 & -0.185 \\
(-0.641) & (-0.401) & (1.791) & (0.549) & (0.110) & (-1.374)
\end{bmatrix}
\begin{bmatrix}
\Delta X1_{t-1} \\
\Delta IP_{t-1} \\
\Delta CPI_{t-1} \\
\Delta FRX_{t-1} \\
\Delta FDI_{t-1} \\
\Delta TBL_{t-1}
\end{bmatrix}
+
$$

$$
\begin{bmatrix}
-1668 & -294 & -273 & -374 & -425 & -217 & -59 & -145 & -81 & -349 & -408. & -29 \\
(-1.46) & (-1.88) & (-1.75) & (-2.26) & (-2.63) & (-0.77) & (-0.43) & (-0.88) & (-0.60) & (-2.54) & (-2.92) & (-0.20) \\
46 & -21 & -33 & -1 & -56 & -30 & -32 & -23 & -23 & -23 & -29 & -16 \\
(1.92) & (-6.39) & (-10.07) & (-0.304) & (-16.38) & (-5.11) & (-10.76) & (-6.49) & (-7.95) & (-7.95) & (-9.82) & (-5.21) \\
52 & 7.85 & 4.48 & 5.46 & 10.28 & 3.20 & 6.22 & 5.38 & 2.19 & 4.23 & 7.81 & 3.56 \\
(4.93) & (5.38) & (3.08) & (3.55) & (6.84) & (1.22) & (4.81) & (-3.53) & (1.74) & 3.31) & (6.01) & (2.60) \\
2.00 & 0.10 & 0.48 & 0.20 & 0.10 & 0.85 & 0.04 & 0.30 & 0.17 & 0.09 & 0.11 & 0.71 \\
(0.88) & (0.31) & (1.53) & (0.61) & (0.31) & (1.52) & (0.13) & (0.91) & (0.65) & (0.33) & (0.38) & (2.41) \\
1277 & 3.60 & -70 & 27 & 27 & -121 & 18 & -78 & 51 & -35 & -50 & -60 \\
(2.600) & (0.054) & (-1.06) & (0.38) & (0.38) & (-1.00) & (0.32) & (-1.11) & (0.88) & (-0.59) & (-0.83) & (-0.95) \\
3.13 & 0..61 & -0.003 & 0.07 & 0.183 & -0.145 & -0.204 & -0.206 & -0.162 & 0.111 & -0.180 & -0.337 \\
(1.75) & (2.45) & (-0.01) & (0.258) & (0.724) & (-0.329) & (-0.937) & (-0.802) & (-0.765) & (0.518) & (-0.821) & (-1.461)
\end{bmatrix}
$$

$$
\begin{bmatrix}
consa\tan t \\
D_{1t} \\
D_{2t} \\
D_{3t} \\
D_{4t} \\
D_{5t} \\
D_{6t} \\
D_{7t} \\
D_{8t} \\
D_{9t} \\
D_{10t} \\
D_{11t}
\end{bmatrix}
+
\begin{bmatrix}
\hat{u}_{1t} \\
\hat{u}_{2t} \\
\hat{u}_{3t} \\
\hat{u}_{4t}
\end{bmatrix}
\qquad (5.8)
$$

(The figures in parentheses are the corresponding t-statistic values.).

$$
\begin{bmatrix} \Delta X2_t \\ \Delta IP_t \\ \Delta CPI_t \\ \Delta FRX \\ \Delta FDI_t \\ \Delta TBL \end{bmatrix}
=
\begin{bmatrix}
-0.105 & 4.062 \\
(-2.100) & (1.980) \\
0.001 & -0.055 \\
(0.704) & (-0.759) \\
0.004 & -0.170 \\
(5.010) & (-5.440) \\
0.000 & -0.011 \\
(1.796) & (-1.652) \\
-0.092 & 2.187 \\
(-2.578) & (1.500) \\
0.000 & -0.012 \\
(2.104) & (-2.195)
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 128 & -784 & 28 & 1247 \\
(.....) & (.....) & (3.30) & (-2.21) & (7.36) & (3.90) \\
0 & 1 & 3.76 & -27 & 0.67 & 40 \\
(.....) & (......) & (4.06) & (-3.18) & (7.30) & (5.27)
\end{bmatrix}
\begin{bmatrix} X2_{t-1} \\ IP_{t-1} \\ CPI_{t-1} \\ FRX_{t-1} \\ FDI_{t-1} \\ TBL_{t-1} \end{bmatrix}
+
$$

$$
\begin{bmatrix}
0.14 & -0.015 & -12.175 & -8.304 & 0.148 & 99.593 \\
(1.136) & (-0.004) & (-2.088) & (-0.222) & (0.940) & (1.963) \\
0.003 & -0.236 & 0.047 & -1.196 & -0.005 & 2.513 \\
(0.682) & (-1.967) & (0.227) & (-0.902) & (-0.940) & (1.399) \\
-0.002 & -0.000 & 0.337 & 0.002 & 0.001 & 0.356 \\
(-0.886) & (-0.004) & (3.788) & (0.003) & (0.349) & (0.460) \\
-0.001 & 0.011 & -0.005 & -0.144 & 0.000 & 0.027 \\
(-2.112) & (1.005) & (-0.267) & (-1.152) & (0.334) & (0.162) \\
0.092 & -5.086 & -0.076 & -26.214 & 0.053 & 29.619 \\
(1.046) & (-2.107) & (-0.018) & (-0.985) & (0.479) & (0.822) \\
-0.000 & -0.002 & 0.024 & 0.042 & 0.000 & -0.147 \\
(-0.246) & (-0.258) & (1.615) & (0.431) & (0.084) & (-1.115)
\end{bmatrix}
\begin{bmatrix} \Delta X2_{t-1} \\ \Delta IP_{t-1} \\ \Delta CPI_{t-1} \\ \Delta FRX_{t-1} \\ \Delta FDI_{t-1} \\ \Delta TBL_{t-1} \end{bmatrix}
+
$$

$$
\begin{bmatrix}
-969 & -174 & -114 & -226 & -293 & -30 & -31 & -43 & -7.5 & -161 & -159 & 30 \\
(-1.22) & (-1.85) & (-1.19) & (-2.26) & (-2.96) & (-0.02) & (-0.38) & (-0.44) & (-0.10) & (-1.96) & (-1.91) & (0.35) \\
48 & -21 & -34 & -1.39 & -55 & -31 & -31 & -22 & -22 & -23 & -29 & -16 \\
(1.73) & (-6.39) & (-9.97) & (-0.39) & (-15.88) & (-5.12) & (-10.80) & (-6.59) & (-7.94) & (-9.84) & (-9.84) & (-5.23) \\
66 & 7.96 & 4.86 & 5.58 & 11 & 3.79 & 6.30 & 5.26 & 2.35 & 4.26 & 8.04 & 3.94 \\
(5.51) & (5.54) & (3.31) & (3.65) & (7.36) & (1.45) & (4.97) & (3.52) & (1.92) & (3.40) & (6.31) & (2.94) \\
2.64 & 0.11 & 0.54 & 0.26 & 0.08 & 0.93 & 0.05 & 0.34 & 0.21 & 0.13 & 0.14 & 0.76 \\
(1.00) & (0.35) & (1.68) & (0.78) & (0.26) & (0.103) & (0.19) & (1.08) & (0.77) & (0.47) & (0.52) & (2.60) \\
934 & 1.98 & -85 & 19 & 15 & -140 & 19 & -78 & 49 & -35 & -54 & -69 \\
(1.65) & (0.03) & (-1.25) & (0.27) & (0.22) & (-1.15) & (0.32) & (-1.12) & (0.856) & (-0.61) & (-0.91) & (-1.11) \\
4.24 & 0.61 & 0.03 & 0.08 & 0.24 & -0.10 & -0.20 & -0.22 & -0.16 & 0.11 & -0.17 & -0.31 \\
(2.06) & (2.49) & (0.11) & (0.29) & (0.95) & (-0.23) & (-0.94) & (-0.86) & (-0.75) & (0.510) & (-0.78) & (-1.38)
\end{bmatrix}
$$

$$\begin{bmatrix} consa\tan t \\ D_{1t} \\ D_{2t} \\ D_{3t} \\ D_{4t} \\ D_{5t} \\ D_{6t} \\ D_{7t} \\ D_{8t} \\ D_{9t} \\ D_{10t} \\ D_{11t} \end{bmatrix} + \begin{bmatrix} \hat{u}_{1t} \\ \hat{u}_{2t} \\ \hat{u}_{3t} \\ \hat{u}_{4t} \end{bmatrix} \tag{5.9}$$

(The figures in parentheses are the corresponding t-statistic values.).

$$\begin{bmatrix} \Delta X3_t \\ \Delta IP_t \\ \Delta CPI_t \\ \Delta FRX \\ \Delta FDI_t \\ \Delta TBL \end{bmatrix} = \begin{bmatrix} -0.109 & 2.047 \\ (-2.346) & (2.257) \\ 0.004 & -0.089 \\ (1.252) & (-1.287) \\ 0.007 & -0.141 \\ (4.137) & (-4.568) \\ 0.001 & -0.012 \\ (2.043) & (-1.883) \\ -0.163 & 1.627 \\ (-2.269) & (1.161) \\ 0.001 & -0.011 \\ (2.041) & (-2.129) \end{bmatrix} \begin{bmatrix} 1.00 & 0 & 64.496 & -408.1 & 13.35 & 593 \\ (....) & (....) & (3.382) & (-2.337) & (7.20) & (3.80) \\ 0 & 1 & 3.823 & -27.561 & 0.663 & 40.085 \\ (....) & (.....) & (4.035) & (-3.177) & (7.191) & (5.172) \end{bmatrix} \begin{bmatrix} X3_{t-1} \\ IP_{t-1} \\ CPI_{t-1} \\ FRX_{t-1} \\ FDI_{t-1} \\ TBL_{t-1} \end{bmatrix} +$$

$$\begin{bmatrix} -0.023 & -1.553 & -7.388 & -9.267 & 0.043 & 42 \\ (-0.191) & (-1.012) & (-2.763) & (-0.543) & (-0.594) & (1.767) \\ -0.001 & -0.199 & 0.007 & -1.402 & -0.006 & 2.214 \\ (-0.136) & (-1.708) & (0.034) & (-1.081) & (-1.022) & (1.219) \\ -0.006 & -0.018 & 0.365 & 0.304 & 0.001 & -0.124 \\ (-1.416) & (-0.340) & (0.002) & (0.523) & (0.560) & (-0.153) \\ -0.002 & 0.010 & -0.005 & -0.146 & 0.000 & -0.003 \\ (-2.540) & (0.971) & (-0.260) & (1.212) & (0.427) & (-0.021) \\ 0.112 & -4.408 & -0.950 & -32.923 & 0.02 & 28.643 \\ (0.602) & (-1.858) & (-0.230) & (-1.249) & (0.375) & (0.776) \\ -0.000 & -0.003 & 0.025 & 0.055 & 0.000 & -0.173 \\ (-0.560) & (-0.313) & (1.698) & (0.575) & (0.127) & (-1.301 \end{bmatrix} \begin{bmatrix} \Delta X3_{t-1} \\ \Delta IP_{t-1} \\ \Delta CPI_{t-1} \\ \Delta FRX_{t-1} \\ \Delta FDI_{t-1} \\ \Delta TBL_{t-1} \end{bmatrix} +$$

$$
\begin{bmatrix}
-483 & -80 & -74 & -149 & -121 & -98 & -25 & -56 & -24 & -110 & -116 & -16 \\
(0.00) & (-1.84) & (-1.67) & (-3.21) & (-2.64) & (-1.25) & (-0.654) & (-1.22) & (-0.63) & (-2.62) & (-2.98) & (-0.39) \\
57 & -21 & -33 & -0.86 & -56 & -30 & -32 & -23 & -23 & -24 & -29 & -16.37 \\
(2.17) & (-6.42) & (-9.79) & (-0.24) & (-16.09) & (-5.04) & (-10.96) & (-6.64) & (-8.10) & (-8.10) & (-9.97) & (-5.243) \\
55 & 8.25 & 4.74 & 5.67 & 11 & 3.34 & 6.49 & 5.51 & 2.38 & 4.40 & 8.06 & 3.63 \\
(4.680) & (5.54) & (3.12) & (3.58) & (6.94) & (1.24) & (4.93) & (3.55) & (1.85) & (3.38) & (6.08) & (2.59) \\
2.72 & 0.15 & 0.56 & 0.27 & 0.10 & 0.89 & 0.04 & 0.31 & 0.16 & 0.08 & 0.11 & 0.70 \\
(1.11) & (0.50) & (1.78) & (0.83) & (0.29) & (1.60) & (0.15) & (0.95) & (0.59) & (0.31) & (0.40) & (2.40) \\
1141 & -6.79 & -77 & 22 & 3.98 & -124 & 14 & -77 & 47 & -38 & -57 & -65 \\
(2.13) & (-0.10) & (-1.11) & (0.31) & (0.06) & (-1.02) & (0.23) & (-1.09) & (0.80) & (-0.64) & (-0.95) & (-1.03) \\
3.78 & 0.63 & 0.04 & 0.08 & 0.23 & -0.12 & -0.20 & -0.21 & -0.17 & 0.11 & -0.18 & -0.33 \\
(1.95) & (2.57) & (0.143) & (0.34) & (0.90) & (-0.26) & (-0.93) & (-0.84) & (-0.79) & (0.51) & (-0.81) & (-1.45)
\end{bmatrix}
$$

$$
\begin{bmatrix} consa\tan t \\ D_{1t} \\ D_{2t} \\ D_{3t} \\ D_{4t} \\ D_{5t} \\ D_{6t} \\ D_{7t} \\ D_{8t} \\ D_{9t} \\ D_{10t} \\ D_{11t} \end{bmatrix} + \begin{bmatrix} \hat{u}_{1t} \\ \hat{u}_{2t} \\ \hat{u}_{3t} \\ \hat{u}_{4t} \end{bmatrix} \qquad (5.10)
$$

(The figures in parentheses are the corresponding t-statistic values).

Let us first consider the estimated VECM involving the variables BSESENSEX, IP, CPI, FRX, FDI and TBL as presented in (5.8). In this case neither a constant nor a linear trend term has been assumed in the long-run relationship. We can notice from (5.8) that there are two cointegrating relations involving BSESENSEX and the other five macroeconomic variables. Since BSESENSEX (X1) has been chosen as the first variable in this model, the coefficient associated with this variable has, therefore, been normalized to 1 and the coefficient associated with the next variables in our model *viz*., domestic industrial production (IP), normalized to 0 in the estimation procedure. In the second cointegrating relation, the normalization rule has been interchanged between these two variables i.e., BSESENSEX and IP, and thus we have the coefficient of BSESENSEX (X1) normalized to 0 and the coefficient associated with industrial production (IP) is normalized to 1. Looking at the estimates of loading coefficients (or speed of adjustment parameter) denoted by α in the preceding section, we find that some of these coefficients in the equations with $\Delta X1_{t-1}$, ΔCPI_{t-1}, ΔFDI_{t-1} as the dependent variables for the first cointegrating relation are significant at 5% level of significance, and similarly for the second cointegrating relation some coefficients attached to $\Delta X1_{t-1}$, ΔCPI_{t-1} and ΔTBL_{t-1} have been found to be significant. This implies that the cointegrating relations resulting from our normalization rules enter significantly in some of these equations involving these variables so that cointegration holds good. We can also observe from (5.8) that some of the lagged differences of the variables are significantly present. The constant and some seasonal dummies have also been found to be significant. Thus both the cointegration test and the VECM results show that there exists a long-run relationship between stock price and major macroeconomic variables – thus describing their interdependence in the Indian stock market. This long-run relation obtained from the ML estimation with the normalization rule–coefficient of BSESENSEX being one and that of IP being zero – is given by

$$X1_t = -193CPI_t + 1162FRX_t - 43FDI_t - 1875TBL_t + EC_{1t} \qquad (5.11)$$

(-3.11) (2.06) (-7.26) (-3.72)

($t-$ratio values are given in parentheses)

where EC_{1t} stands for the error-correction term. We find from these $t-$ratios that CPI, FRX, FDI and TBL are all significantly present in the cointegrating relation. Similarly, we have the results for the second cointegrating equation as follows.

$$IP_t = -3.81CPI_t + 27.32FRX_t - 0.67FDI_t - 40.45TBL_t + EC_{1t} \tag{5.12}$$

(-3.99) (3.13) (-7.23) (-5.18)

In this cointegrating relation also we find that the variables CPI, FRX, FDI and TBL are significant. However, from the point of view of our study, the first relation is obviously more relevant.

As already mentioned, we have also carried out our analysis considering other combinations of deterministic terms. For example, we have obtained the following cointegrating relation when constant mean, linear trend and seasonal dummy variables are included in the VECM and both constant mean and linear trend are assumed in the long-run relationship. However, both the components of the deterministic term in the cointegrating relation were found to be not significantly different from zero at conventional levels of significance. Therefore, the model was reestimated

$$X1_t = 390IP_t + 0.728CPI_t + 2463FRX_t + 7.324FDI_t - 1511TBL_t + EC_{1t} \tag{5.13}$$

(6.085) (0.017) (4.302) (1.734) (2.494)

(*$t-$ratio values are given in parentheses*)

We notice that although IP, FRX, FDI and TBL are all found to be significant at varying levels of significance but CPI is found to be insignificant. We also carried out cointegration analysis for other combinations and the results, like this case, were not quite satisfactory. In view of this we conclude that the reported cointegrating relation in (5.11) is the most meaningful one.

The findings on the other two systems of equations involving BSE 100 and NIFTY have been similar as in case of BSESENSEX. We observe from equations (5.9) and (5.10) that there are two coniegrating relations for both the system of equations involving BSE 100 and NIFTY, respectively. The relevant ones for BSE 100 and NIFTY are given below:

$$X2_t = -128CPI_t + 784FRX_t - 28FDI_t - 1247TBL_t + EC_{2t} \tag{5.14}$$

(-3.30) (2.21) (-7.36) (-3.90)

$$X3_t = -64.5CPI_t + 408.1FRX_t - 13.35FDI_t - 593TBL_t + EC_{3t} \tag{5.15}$$

(-3.38) (2.34) (-7.20) (-3.80)

($t -$ ratio values are given in parentheses)

From these two equations, we find that for BSE 100 and NIFTY also, the variables CPI, FRX, FDI and TBL are significant in the cointegrating relations.

This cointegration analysis, therefore, clearly demonstrates that stock index and the macroeconomic variables involving domestic industrial production, consumer price index, nominal exchange rate, foreign direct investment and long-term interest rate have long-run equilibrium relationships. A relevant point to mention, in this context, is that that other important macroeconomic variables such as money supply and short-term interest rate have been found to have no significant influence on stock prices in the long-run sense. The strong cointegrating relationship between stock price and the stated macroeconomic variables can be explained by existing economic theories. For instance, the significant negative coefficient of interest rate is consistent with the economic argument that interest rate representing cost of capital and opportunity cost to the investors is negatively related with stock prices. The significant negative relationship between stock price and consumer price index is clearly understood as the inflation rate which can be defined on the latter affects stock market through output link. The presence of nominal exchange rate in the cointegrating relations shows that policies having effects on exchange rate have ramifications for the stock market as well. Lastly, significant positive effect of foreign direct investment is crucial from the point of view of India's policy of inviting foreign investors, and it is only expected that after liberalisation India has moved along in that direction. But, we find from the observed long-run relationships that foreign direct investment has a negative effect on stock prices. Although this is a bit unusual, we may note that in a study Kumar and Pradhan (2002) have also found that FDI had a significant negative effect on domestic investment for a number of developing countries including India. They have also found that FDI is growth-neutral for India. These empirical findings, therefore, suggest that FDI has crowded out India's domestic investment- at least in the initial phase of liberalisation. Thus, our analysis in this chapter shows that insofar as India is concerned, the macroeconomic fundamentals such as real economic activity, inflation, investment and exchange rate, have strong interdependence with Indian stock index and hence policies which affect these macrovariables will also have effects on the stock market in the long-run.

5.4 Conclusions

In this chapter, we have carried out a cointegration analysis involving a particular Indian stock index and a set of macro and financial variables in order to find long-run relations, if any, involving these variables. Existence of such relation(s) would then suggest predictability of Indian stock market in the long-run sense. This exercise has been done separately for each of the three stock indices considered *viz*., BSESENSEX, BSE 100 and NIFTY. Following essentially Johansen's approach, our findings based on the trace and maximum eigenvalue tests indicate that there exist cointegrating relations involving a stock index and the following macro variables *viz.,* domestic industrial production, consumer price index, nominal exchange rate, foreign direct investment and long-term interest rate. Further, the cointegrating relations corresponding to each of the three stock indices involve the same set of macroeconomic variables as stated above. Neither any of the three financial ratios nor any of the remaining eight macro variables has been found to be significantly involved in the cointegrating regressions. We also note from the cointegrating regressions that the observed long-run relations show a negative relationship between stock index and consumer price index, but a significant positive relationship between nominal exchange rate and stock index. These findings are in the expected lines as supported by existing economic theories. The evidence of a negative effect of foreign direct investment on stock index, in the long run, is, although a little intriguing, but yet consistent with similar findings in other emerging economies.

CHAPTER 6

Predictability of Daily Stock Returns under Alternative Volatility and Distributional Assumptions

6.1 Introduction

It has been observed that asset returns tend to cluster i.e., large changes tend to follow large changes and small tend to follow small. This phenomenon is called volatility. Proper modelling of volatility is an essential part of any predictability study on asset returns. Moreover, proper estimation of volatility is also necessary for portfolio analysis and risk management such as value-at-risk methodology (Taylor (2004)). The phenomenon of time-varying volatility which contradicts the hypothesis of normally distributed price changes put forth by Bachelier (1900), led to the generation of a vast body of literature on volatility, which began with the seminal contribution by Engle (1982). To capture the volatility of asset returns, Engle (1982) proposed the class of autoregressive conditional heteroscedastic , or ARCH, models. In order to model a persistent movement in volatility without allowing large number of coefficients in an ARCH model, Bollerslev (1986) suggested the generalized autoregressive conditional heteroscedastic (GARCH) model. GARCH(1,1) is the most commonly used model in the family of GARCH models. According to Bollerslev *et al.* (1994), GARCH(1,1) model has turned out to be very useful for describing a wide variety of financial data . Keeping in mind the general usefulness of GARCH models, we have described the volatility of Indian stock returns in Chapter 3 in the GARCH(1,1) framework, and have indeed found GARCH(1,1) to be the most appropriate model for volatility. However, it is obvious from the specification of a GARCH model that this class of models are symmetric in that negative and positive shocks have the same effect on volatility. But the empirical literature on returns of risky assets following Black (1976), has pointed out that future volatility is affected more by negative shocks than by positive shocks. This asymmetry in equity returns is typically attributed to ‘leverage effect’, which basically means that a fall in the value of a firm’s stock causes the firm’s debt to equity ratio to rise. This leads shareholders who bear the residual risk of the firm, to perceive their future cash-flow stream as being relatively more risky.

Nelson (1991) extended the usual GARCH model in a way so that in addition to volatility, the leverage effect is also captured. This model, known as the exponential GARCH (EGARCH) model, has found immense applications in studies concerning stock market returns. Building on the success of the EGARCH model to represent asymmetric responses in the conditional variance to positive and negative errors, some other extensions have been suggested in which a general shape of the conditional variance has been proposed. For instance, Glosten, Jagannathan and Runkle (1993) and Zakoian (1994) have independently suggested which is called the threshold GARCH (TGARCH) model.

The first objective in this chapter is to consider these two asymmetric volatility models *viz.,* the EGARCH and TGARCH models to predict the Indian stock returns on BSESENSEX, BSE 100 and NIFTY indices[1] and then find whether these volatility models can predict the returns better as compared to those by the GARCH model. The computations are done by using the standard forecasting criteria like the mean absolute

[1] At the time of actual computations, the data on DOLLEX stock index for the hold-out sample were not available, and hence this index has been kept out of this exercise.

error (MAE) and the root mean squared error (RMSE) of out-of-sample return forecasts. This study has been carried out with daily level data covering the post-liberalisation period spanning January 1996 to December 2000. The choice of this in-sample period has been made from consideration of the fact that in Chapter 3 where analysis of daily- level returns was done with the assumptions of GARCH volatility model and Gaussian conditional distribution, the sub-period from January 1996 to December 2000 was found to be the last stable sub-period for all the return series considered in this chapter. Since the purpose here is to compare across different volatility models and distributional assumptions insofar as their effects on predictability of daily returns are concerned, we have confined ourselves to this choice of in-sample observations. For out-of-sample forecasting exercise, the hold-out sample has been taken to be the next three months i.e., the period from January 2001 to March 2001.

Apart from this exercise on alternative volatility models for the Indian stock returns, this chapter is also concerned with the issue of appropriate (conditional) distributional assumption of the error term. Most often conditional normality is considered to be the distribution, but unfortunately this is rarely supported by economic and financial data exhibiting volatility. In fact, it is a widely accepted fact that most financial data exhibit leptokurtosis and sometimes asymmetry in return distributions (see, for example, Kon (1984), Mills (1995) and Peiró (1999)). In case of our return data also, this is clearly evident, as noted in Chapter 3. In such cases, what is often done is to continue working with the conditional normal likelihood function but in that case the resulting estimators of both the mean and variance parameters are called the quasi-maximum likelihood estimator (QMLE). Weiss (1986) has shown that the asymptotic properties of the QMLE of GARCH model in the usual regression framework is consistent as long as the first two conditional moments are correctly specified. Later, Bollerslev and Wooldridge (1992) have suggested a procedure for obtaining a robust estimator of the covariance matrix, and this can be used for the purpose of computing an appropriate standard error of QMLE.

Alternatively, one can apply the parametric models which have been proposed for capturing excess kurtosis through alternative distributional assumptions about the return. For instance, Bollerslev (1987) and Tucker (1992), among others, have advocated evaluating sample log-likelihood under the assumption that innovations ε_t follow a standardized Student's t-distribution. In a similar spirit, Nelson (1991) has suggested using a standardized general error distribution (GED).

Thus, in this chapter we have considered alternative volatility models as well as alternative conditional distributions for Indian stock returns and estimated the parameters involved thereof by appropriate estimation procedures – QML and/or ML– and then finally evaluated their performances by suitable forecasting criteria. As regards the distributional assumption of the error under volatility specifications given by GARCH(1,1), EGARCH and TGARCH, we have three specific distributions *viz.,* normal, standardized t -distribution and standardized GED for all the three stock returns. Comparing across these various volatility specifications and distributional assumptions, we finally try to conclude which volatility specification and distributional assumption fit the Indian stock returns best. This chapter has been organized in the following format. The next section presents the modelling approach, the third section deals with the empirical results. The chapter closes with some concluding remarks in Section 6.4.

6.2 Modelling approach

In this section we first specify the model along with the GARCH(1,1) specifications and the two other volatility models capturing leverage effect *viz.*, Nelson's EGARCH and Glosten, Jagannathan and Runkle (GJR)'s TGARCH.

6.2.1 Alternative volatility assumptions

Model for returns with GARCH(1,1) volatility specification: We recall the conditional mean and variance specifications as in Chapter 3. It may be noted, in this context, that the risk aversion parameter, ϕ, in (3.5) was found to be insignificant for all the return series and hence the risk premia term i.e., ϕh_t^{λ} is being dropped here.

Accordingly, we have

$$r_t = \sum_{k=1}^{\tilde{m}} \varsigma_k r_{t-k} + \sum_{j=1}^{d} \beta_j D_{jt} + \omega i_t + f(\hat{w}_{t-1}) + \varepsilon_t , \varepsilon_t |_{\Psi_{t-1}} \sim N(0, h_t) \qquad (6.1)$$

$$h_t = \alpha_0 + \sum_{j=1}^{d} \theta_j D_{jt} + \alpha_1 \varepsilon_{t-1}{}^2 + \delta h_{t-1} \qquad (6.2)$$

where D_j' s $j = 1,2,....,d$ are daily dummies , $\alpha_0 > 0, \alpha_1 \geq 0$ and $\delta \geq 0$, and i_t is the call money rate variable at time point t.

Model for returns with TGARCH(1,1) volatility specification: The TGARCH(1,1) model is a simple extension of GARCH(1,1) model with an additional term added to take account of the possible leverage effect in the data. The conditional mean specification is the same as in (6.1). The conditional variance is now specified as

$$h_t = \alpha_0 + \sum_{j=1}^{d} \theta_j D_{jt} + \alpha_1 \varepsilon_{t-1}{}^2 + .\delta h_{t-1} + .\gamma \varepsilon_{t-1}{}^2 I_{t-1} \qquad (6.3)$$

where $I(.)$ is an indicator function so that $I_{t-1} = 1$ if $\varepsilon_{t-1} < 0$ and = 0 otherwise. For a leverage effect in the returns, one would find $\gamma > 0$. The non-negativity restrictions on the other parameters are now $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\delta > 0$ and $\alpha_1 + \gamma \geq 0$. In this model, positive news have an impact of α_1 while negative news have a effect of $\alpha_1 + \gamma$, and thus negative news have a greater effect on volatility if $\gamma > 0$.

Model for returns with EGARCH(1,1) volatility specification : In addition to TGARCH model, many parametric as well as nonparametric volatility models are designed to capture the asymmetries in the conditional variance. One such specification is the well-known EGARCH model proposed by Nelson (1991). In this specification, h_t is expressed in logarithmic transformation, and given by[2]

$$\ln h_t = \alpha_0^* + \delta^* \ln(h_{t-1}) + \gamma^* \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} + \alpha_1^* \left(\frac{|\varepsilon_{t-1}|}{\sqrt{h_{t-1}}} - \frac{|\varepsilon_t|}{\sqrt{h_t}}\right). \tag{6.4}$$

Now, if α_1^*>0, the process (6.4) generates volatility clustering. In addition, if $\gamma^* < 0$, then there will be a negative relationship between volatility and returns which is the leverage effect. The model specified in (6.4) has an obvious advantage over the conditional GARCH model in that conditional variance in this specification will always be positive irrespective of the sign of the parameters since h_t is specified in logarithmic scale. However, as pointed out by Engle and Ng (1993), one particular drawback of the EGARCH model is that owing to the exponential structure of h_t, the model may tend to overestimate the impact of outliers on volatility.

[2] Dummy variables were not included in this volatility model since inclusion of them created some computational problems.

Insofar as comparisons amongst these volatility specifications are concerned, we have used the out-of-sample forecast performance of returns. To this end, the usual criteria such as the mean absolute error (MAE) and the root mean squared error (RMSE) have been applied. Denoting the s-step ahead forecast of return based on t in-sample observations as $f_{t,s}$ the mean absolute error (MAE) is defined as

$$MAE = \frac{1}{n - n_1} \sum_{t=n_1+1}^{n} | r_{t+s} - f_{t,s} | \qquad (6.5)$$

which measures the average absolute forecast error. Here, as usual, n is the total sample size i.e., sum of in-sample and out-of-sample sizes and (n_1+1)th sample is the first out-of-sample forecast observation. The in-sample model estimation initially runs from observations 1 to n_1 and observations $n_1 + 1$ to n are available for out-of-sample estimation forecasting i.e., the hold-out sample size is $n - n_1$.

Another criterion *viz.,* the root mean squared error (RMSE) is defined as

$$RMSE = \sqrt{\frac{1}{n - n_1}} \sum_{t=n_1+1}^{n} (r_{t+s} - f_{t,s})^2 \qquad (6.6)$$

RMSE is a conventional criterion which clearly weights greater forecast errors more heavily than smaller forecast errors in the forecast error penalty. This may, however, also be viewed as an advantage if large errors are not disproportionately more serious, although the same critique could also be applied to the so-called least squares methodology. As these statistics in (6.5) and (6.6) are unbounded from above, taken individually, little can be said from the value of RMSE or MAE. Instead, the MAE or RMSE from one model should be compared with those of other models for the same data and forecast period, and the model with the lowest value of the error measure would be considered as the best model.

6.2.2 Alternative distributional assumptions

For all the volatility specifications considered in this chapter, namely, GARCH(1,1), EGARCH(1,1), and TGARCH(1,1), the parameters in the conditional mean as well as in the conditional variance have been estimated by maximum likelihood method under the assumption of conditional normality which is more in the nature of an ad hoc assumption rather than based on any statistical or economic reasoning. It has been found in the empirical literature on GARCH model that conditional normality of speculative returns is more an exception than the rule. Therefore, the issues on estimation and inference arising out of conditional non-normality of speculative returns have also got attention of the researchers in the GARCH literature. It may be mentioned in this context that Bollerslev and Wooldridge (1992) suggested a variance-covariance matrix estimator (which is) robust to non-normality where Gaussian log-likelihood function is maximized ignoring non-normality of innovations. This procedure, as already stated in the preceding section, is known as quasi maximum likelihood or QML procedure. Asymptotic theory on properties of the QML estimator in univariate GARCH model is well developed due to Weiss (1986). Consistency and asymptotic normality of the QML estimator have been shown for a wide variety of strictly stationary GARCH processes. In view of these fine properties of the QML estimator, we apply, in addition to the ML procedure, this estimation procedure also for our models on Indian stock returns with EGARCH and TGARCH volatility specifications.

However, it is to be noted that when normality assumption is violated, it is no longer possible to provide forecasting intervals. Further, maximum likelihood estimation under misspecification of the (non-Gaussian) conditional distribution, may yield inconsistent parameter estimates (Newey and Steigerwald (1997)). Under these circumstances, parametric models with explicit distributional assumptions, have been suggested. For instance, Bollerslev (1987) has advocated that innovations follow the standardized Student's *t*-distribution. On the other hand, Nelson (1991) has suggested the standardized general error distribution (GED). The probability density functions of these two alternative distributions are given below.

Standardized Student's $t-$distribution: The innovation ε_t is said to follow a standardized Student's $t-$distribution with degrees of freedom v, mean zero, and variance h_t involving the parameters η if it has the following probability density function:

$$f(\varepsilon_t |_{\eta,v}) = \frac{v^{v/2}\Gamma(\frac{v+1}{2})}{\sqrt{\Pi}\Gamma(v/2)\sqrt{\frac{(v-2)h_t}{v}}}(v + \frac{v.\varepsilon_t^{\ 2}}{(v-2)h_t})^{-\frac{v+1}{2}} \tag{6.7}$$

where $\Gamma(.)$ denotes the gamma function, $\Gamma(s) = \int_0^\infty x^{s-1}\exp(-x)dx$, s>0. As $v \to \infty$, the density in (6.7) coincides with the Gaussian density.

Generalized error distribution (GED): The random variable ε_t having mean zero and variance h_t involving the parameter vector η, is said to follow the generalised error distribution if its probability density function is given by

$$f^*(\varepsilon_t \,|_{\eta,v}) = v\exp(-\frac{1}{2}\left|\frac{\varepsilon_t}{\lambda\sqrt{h_t}}\right|^{v})[2^{\frac{v+1}{v}}\Gamma(\frac{1}{v})\lambda\sqrt{h_t}\,]^{-1} \tag{6.8}$$

where v is now called the shape parameter and λ is defined as

$$\lambda = [\frac{\Gamma(\frac{1}{v})}{2^{\frac{2}{v}}\Gamma(\frac{3}{v})}]^{0.5}. \tag{6.9}$$

In case $v = 2$, the density in (6.8) is equal to the density of $N(0, h_t)$ and the distribution becomes leptokurtic if $v < 2$. For $v = 1$, the GED coincides with the double exponential distribution. Further, as $v \to \infty$, the GED approximates the rectangular distribution.

The GARCH model and the other volatility models require joint estimation of the conditional mean and the conditional variance. These computations under the alternative assumptions of standardized Student's t-distribution and GED have been carried out by the software called JMulTi (*cf.* Lütkepohl and Krätzig (2004)) where the OLS residuals from the mean equation are first obtained and then the parameters in the conditional variance under alternative distributions are obtained. This procedure continues until convergence.

6.3 Empirical results

In this section we discuss the empirical results. We first present the estimation results on volatility modelling of Indian stock returns at daily frequency, based on three standard stock indices *viz.,* BSESENSEX, BSE 100 and NIFTY covering the period January 1996 to December 2000. We have already noted in Chapter 3 that daily level returns exhibit significant volatility for all the series considered for this study. Accordingly, the usual GARCH model was considered and parameter for both the conditional mean and the conditional variance were obtained and these have been presented in Table 3.4 of Chapter 3. We now report the results of estimation under TGARCH and EGARCH volatility specifications under the assumption of conditional Gaussian distribution. Parameters estimates based on both ML and QML methods are presented in Table 6.1 and 6.2 for TGARCH(1,1) and EGARCH(1,1), respectively. The orders of both the volatility models have been taken to be (1,1) since this choice has been found, as in GARCH(1,1), to be adequate in most such empirical studies. Insofar as the choice of $\widetilde{m}$, own lag length of r_t is concerned, we have chosen a sufficiently large value of 20 and the estimated model has been reported with the significant ones only. The first major observation from Table 6.1 is that all the parameter estimates except the coefficients of r_{t-1} and $\varepsilon_{t-1}{}^2 I_{t-1}$ in NIFTY series are very close between ML and QML estimates. Thus, there is practically no

Table 6.1 : Estimates of Parameters in Conditional Mean and Conditional Variance under TGARCH(1,1) Volatility Model and Gaussian Conditional Distribution

	BSESENSEX		BSE 100		NIFTY	
Variable	MLE	QMLE	MLE	QMLE	MLE	QMLE
r_{t-1}	0.083 (2.712)***	0.083 (2.693)***	0.112 (3.152)***	0.112 (3.196)***	0.188 (4.774)***	0.078 (2.565)***
r_{t-6}	-0.102 (3.869)***	-0.102 (3.565)***	-0.088 (2.999)***	-0.088 (2.776)***	-0.090 (3.158)***	-0.087 (3.086)***
r_{t-11}	-0.067 (2.573)***	-0.067 (2.380)***	_	_	_	_
r_{t-18}	-0.073 (2.838)***	-0.073 (2.649)***	_	_	_	_

Table 6.1 (contd.)

r_{t-19}	-0.061	-0.061	-0.098	-0.098	-0.085	-0.081
	(2.400)**	(2.295)**	(3.853)***	(2.976)***	(3.449)***	(3.535)***
$D1$	0.003	0.003	0.007	0.007	_	_
	(2.283)**	(1.896)*	(7.045)***	(3.841)***		
$D3$	_	_	_	_	0.008	0.008
					(7.272)***	(7.292)***
constant	0.000008	0.00008	0.00008	0.00008	0.00003	0.00004
	(0.598)	(0.442)	(5.454)***	(3.572)***	(2.625)***	(3.173)***
ε_{t-1}^{2}	0.043	0.043	0.094	0.094	0.048	0.059
	(1.906)**	(1.397)	(3.044)***	(1.636)#	(2.683)***	(2.986)***
$\varepsilon_{t-1}^{2} I_{t-1}$	0.169	0.169	0.225	0.225	0.085	0.102
	(3.981)***	(2.922)***	(4.171)***	(2.201)**	(2.688)***	(2.684)***
h_{t-1}	0.664	0.664	0.575	0.575	0.776	0.713
	(11.416)***	(8.954)***	(9.679)***	(6.693)***	(16.931)***	(12.067)***
$D1$	0.0002	0.0002	_	_	0.00004	0.00005
	(8.913)***	(3.666)***			(2.859)***	(3.013)***
$D3$	0.00007	0.0001	_	_	_	_
	(2.491)***	(1.560)				
Max log$	3131.519	3131.519	2554.448	2554.448	3220.839	3210.953

*NOTE: (i) The values in the parentheses indicate the corresponding t-statistic values(absolute). * indicates significance at 10% level whereas ** indicates significance at 5% level.*** indicates significance at 1% level.*

(ii) $: *This row presents the maximum log-likelihood values.*

statistical reason to choose one estimation method over the other. Computational figures presented in Table 6.1 clearly show that there is presence of asymmetry in volatility of Indian stock returns. For instance, if we consider returns based on BSESENSEX, the ML estimate of the term capturing leverage effect i.e., ${\varepsilon_{t-1}}^2 I_{t-1}$ is 0.169 and its $t-$statistic value is 3.981 which is significant at 1% level. This term remains significant with QML estimation procedure also. In fact, in all the three return series, this term representing asymmetry is highly significant and positive – irrespective of the estimation procedure used – indicating thereby that volatility rises more after a large negative shock than an equal large positive shock in the Indian stock market. The $t-$statistic values of the coefficient corresponding to this term for BSE 100 and NIFTY under ML estimation are 4.171 and 2.688, respectively, and these are, as indicated in Table 6.1, highly significant. It is also noteworthy that a comparison with the GARCH(1,1) volatility specification in Chapter 3 (*cf.* Table 3.4) shows that in the mean specification of the returns series, the set of significant variables are found to be the same with the same sign for both TGARCH and EGARCH models for all the three stock returns except the dummy variable, D3, for BSE 100, which has now been found to be insignificant. For example, in case of BSESENSEX returns, the variables r_{t-1}, r_{t-6}, r_{t-11}, r_{t-18}, r_{t-19} and $D1$ have been found to be significant in the mean part of the estimated model for returns with GARCH(1,1) volatility specification (*cf.* 4th column of Table 3.4 referring to the same period) and we note from Table 6.1 that exactly the same variables in the conditional mean are significant with TGARCH(1,1) specification for volatility, and further that their respective signs are also the same in the two models. Since ML and QML methods estimation produce almost the same estimates, their maximized log-likelihood values are likely to be very close and we find from the entries in the last row that this is indeed the case for all the three series, the highest difference in the log-likelihood value being around 10 in case of NIFTY where, as already mentioned, there are some differences in the estimates of two parameters.

We now check whether estimation results with EGARCH(1,1) specification, which are presented in Table 6.2, also indicate the presence of asymmetry in volatility of Indian stock returns. The EGARCH results presented in Table 6.2 shows similar findings as that for the TGARCH specification. In all the return series, the asymmetric term γ^* is negative and significant. For example, in case of BSESENSEX the estimated value of γ^* is –0.099 under ML estimation, and its $t-$statistic value is 4.935 which clearly indicates significance at 1% level of significance. As 'leverage effect' or asymmetry in volatility is found to be significantly present in both the TGARCH and EGARCH models for all the three series, we can infer that asymmetric response of volatility to positive and negative shocks is an important phenomenon with Indian stock market. One can further note that the maximum log-likelihood values for TGARCH and EGARCH models are very close for all the three return series, although for BSESENSEX, the value is somewhat smaller for EGARCH model suggesting that TGARCH is a better volatility representation than EGARCH for this series.

Table 6.2 : Estimates of Parameters in Conditional Mean and Conditional Variance under EGARCH(1,1) Volatility Model and Gaussian Conditional Distribution

	BSESENSEX		BSE 100		NIFTY	
Variable /Coefficent[@]	MLE	QMLE	MLE	QMLE	MLE	QMLE

Table 6.2 (contd.)

r_{t-1}	0.079	0.079	0.112	0.112	0.184	0.075
	(2.607)***	(2.562)***	(3.232)***	(3.169)***	(4.750)***	(2.402)**
r_{t-6}	-0.106	-0.106	-0.086	-0.086	-0.101	-0.095
	(3.879)***	(3.460)***	(2.997)**	(2.721)***	(3.801)***	(3.283)***
r_{t-11}	-0.053	-0.053	_	_	_	_
	(1.897)*	(1.892)*				
r_{t-18}	-0.083	-0.083				
	(3.424)***	(2.558)**				
r_{t-19}	-0.057	-0.057	-0.083	-0.083	-0.085	-0.009
	(2.226)**	(1.933)*	(3.500)***	(2.380)***	(3.612)***	(2.665)***
$D1$	0.004	0.004	0.007	0.007	_	_
	(3.833)***	(2.479)**	(7.170)***	(4.084)***		
$D3$	_	_	_	_	0.008	0.008
					(7.532)***	(7.266)***
$\hat{\alpha}_0^*$	-1.314	-1.314	-1.909	-1.970	-1.075	-1.307
	(4.726)***	(2.898)***	(5.844)***	(3.855)***	(4.701)***	(3.021)***
$\hat{\alpha}_1^*$	0.196	0.196	0.380	0.380	0.196	0.222
	(4.757)***	(3.996)***	(6.891)***	(5.746)***	(7.067)***	(2.927)***
$\hat{\gamma}^*$	-0.099	-0.099	-0.111	-0.111	-0.065	-0.076
	(4.935)***	(2.231)**	(4.520)***	(1.941)**	(3.468)***	(2.031)***
$\hat{\delta}^*$	0.856	0.856	0.790	0.790	0.887	0.861
	(26.568)***	(15.648)***	(19.890)***	(12.812)***	(32.406)***	(17.847)***

Table 6.2 (contd.)

Max log$	3106.505	3106.505	2555.651	2555.651	3221.997	3211.098

NOTE: (i) # : *The description of some of the variables in volatility specification being quite large, coefficients associated with these have been denoted in this column; for the mean part, however, the variables are being denoted.*

*(ii) The values in the parentheses indicate the corresponding t-statistic values(absolute). * indicates significance at 10% level whereas ** indicates significance at 5% level. *** denotes significance at 1% level.*

(iii) $: *This row presents the maximum log-likelihood values*

Now, we examine the performance of the volatility models based on out-of-sample forecasting performance of r_t by using the criteria of MAE and RMSE. In this comparison apart from TGARCH(1,1) and EGARCH(1,1) models, GARCH(1,1) is also considered. As already mentioned, our forecasting exercise employs in-sample data from January 1996 to December 2000 at daily level. The hold-out sample has been taken to be January 2001 to March 2001, comprising 63 observations to construct forecasts and then find the forecast accuracy by these two criteria. We have applied the dynamic forecasting technique to calculate multi-step forecasts starting from the first period in the hold-out sample.

Table 6.3 presents the values of MAE and RMSE for all the three return series. These clearly demonstrate that each of the two criteria *viz.,* MAE and RMSE has produced almost the same value for all the three volatility models. This holds for returns on BSESENSEX and NIFTY entirely and to a great extent for returns on BSE 100. As for instance, the MAE values for BSESENSEX are 0.0156, 0.0155 and 0.0156 for GARCH(1,1), TGARCH(1,1) and EGARCH(1,1) models, respectively, but the same for

BSE100 are 0.0247, 0.0179 and 0.0179. A closer look at MAE and RMSE values for BSE 100 series shows that both the MAE and RMSE values are the same for TGARCH(1,1) and EGARCH(1,1) models but both these values are less than the corresponding value for GARCH(1,1) model. It may thus be concluded that for returns based on BSE 100, GARCH(1,1) performs worse than TGARCH and EGARCH models by these two forecasting criteria.

Table 6.3: Out-of-Sample Forecasting Performances of Indian Stock Returns under Alternative Volatility Models and Gaussian Conditional Distribution

(In-Sample Period : January 1996 to December, 2000 ; Hold-out sample January 2001 to March 2001)

	GARCH(1,1)		TGARCH(1,1)		EGARCH(1,1)	
Stock Return	MAE	RMSE	MAE	RMSE	MAE	RMSE
BSE SENSEX	0.0156	0.0208	0.0155*	0.0207*	0.0156	0.0208
BSE100	0.0247	0.0286	0.0179*	0.0244*	0.0179*	0.0244*
NIFTY	0.0148*	0.0197*	0.0148*	0.0197*	0.0148*	0.0197*

Notes: (i) Since ML and QML estimates have produced almost the same results, we have reported results based on QML estimation only.

(ii) The forecast evaluation criteria used viz., MAE and RMSE stand for means absolute error and root mean squared error, respectively.

*(iii) * denotes the lowest value among the three models by a particular criterion.*

Thus, considering both the maximized log-likelihood values and the out-of-sample forecasting performances, one can conclude that EGARCH and TGARCH perform equally well for all three Indian stock indices. Thus, it is seen that insofar as Indian stock returns are concerned, it really does not matter whether TGARCH or

EGARCH model is considered for capturing the asymmetry in volatility. However, as regards comparing these two volatility models against the GARCH model which disregards asymmetry, we may conclude that for BSESENSEX and NIFTY series there is hardly any difference but for BSE 100, the findings suggest that EGARCH/TGARCH is somewhat better than the simple GARCH model.

The analysis of volatility models so far carried out are based on the distributional assumption of conditional normality. We now discuss the empirical findings on alternative assumptions about the conditional distribution. This concluding part of our analysis considers only two volatility models viz., GARCH(1,1) and TGARCH(1,1)[3]. Estimation and diagnostic results with GARCH(1,1) volatility model under alternative conditional distributional assumptions are given in Table 6.4. We have applied ML estimation method under the three alternative conditional distributional assumptions *viz.,* the usual normal distribution, the standardized Student's t-distribution and the generalized error distribution (GED). We find that the volatility parameters i.e., α_1 and δ of GARCH(1,1) specification are significant for all the three series and for all three distributions considered. Further, it is also evident that excess conditional kurtosis, denoted by v which is a parameter for both the $t-$distribution and GED, is significantly present in all the three series. The t-ratios are 6.875 and 24.071 for BSESENSEX, 6.284 and 20.794 for BSE 100 and 6.328 and 22.199 for NIFTY, and all these are highly significant values. Further, the estimate of parameter α_1, representing volatility clustering is significantly pronounced under all the three distributions. For instance, if

[3] The computations under EGARCH model is not available in the JMulti software package.

Table 6.4 Estimates of Parameters and Diagnostics for Residuals $\hat{\varepsilon}_t$ under GARCH(1,1) Volatility Model and Different Distributional Assumptions

	BSESENSEX			BSE 100			NIFTY		
	$N(0,h_t)$	$t(v,0,h_t)$	$GED(0,h_t)$	$N(0,h_t)$	$t(v,0,h_t)$	$GED(0,h_t)$	$N(0,h_t)$	$t(v,0,h_t)$	$GED(0,h_t)$
Estimate/test statistic value									
$\hat{\alpha}_0$	0.000047	0.00003	0.00003	0.00003	0.00009	0.00006	0.00002	0.00002	0.00002
	(3.256)**	(2.346)*	(2.393)*	(2.393)*	(5.953)**	(3.061)**	(3.944)**	(2.302)*	(2.421)*
$\hat{\alpha}_1$	0.085	0.093	0.087	0.220	0.194	0.208	0.063	0.075	0.069
	(3.659)**	(3.131)**	(2.974)**	(5.103)**	(3.797)**	(3.576)**	(5.373)**	(3.424)**	(3.492)**
$\hat{\delta}$	0.772	0.821	0.808	0.544	0.660	0.596	0.866	0.848	0.857
	(12.454)**	(14.401)**	(12.744)**	(8.285)**	(8.261)**	(6.363)**	(38.413)**	(18.248)**	(20.886)**
$\hat{v}$	_	6.657	1.412	_	5.939	1.352	_	6.725	1.390
		(6.875)**	(24.071)**		(6.284)**	(20.794)**		(6.328)**	(22.199)**
Max log$^{\$}$	3094.390	3136.74	3122.21	2543.99	2582.36	2572.42	3207.29	3241.17	3234.05

Table 6.4 (contd.)

No remaining ARCH (LM, 1 lag)									
$F-Test$	0.012	0.038	0.034	1.149	0.842	1.102	2.426	2.255	2.354
$p-value$	0.915	0.845	0.853	0.284	0.359	0.294	0.120	0.133	0.125
Lomnicki-Jarque-Bera (LJB) test									
LJB	560.420	701.008	636.761	462.335	567.061	504.078	304.612	324.688	314.462
$p-value$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

*Notes: The values in the parentheses indicate of the corresponding t-statistic values(absolute). ** indicates significance at 1% level whereas * indicate significance at 5% level. $ this row presents the maximum log-likelihood values.*

we find that $\hat{\alpha}_1$ is 0.085 under normality whereas for $t-$ distribution and GED, the corresponding values are 0.093 and 0.087, respectively. Now, we look at the maximum log-likelihood values under the three alternative families of distributions in order to be able to conclude, if at all, which distributional assumptions are suitable, under a specific volatility model, for each of three return series considered. To that end, we note that the density of GED reduces to that of normal when the shape parameter takes the value 2 i.e., $v = 2$ and hence the latter belongs to the family of GED. Likewise, normal belongs to the family of standardized Student's $t-$distribution since the density of the latter coincides with the former as $v \to \infty$, where v now stands for the degrees of freedom. We can, therefore, carry out formal tests like the likelihood ratio (LR) test to find if normality is the appropriate distributional assumption against the alternative of GED. In a similar spirit, such a test can be envisaged for normality against the alternative of standardized Student's t- distribution. However, a formal test is not possible since the underlying null hypothesis $H_0 : v \to \infty$ is not well specified from the point of view of a proper test of hypothesis. Obviously, such a test involving $t-$distribution and GED cannot be performed at all.

Now, the maximized log-likelihood values for the three families of distributions *viz*., Gaussian distribution, standardized Student's $t-$distribution and GED under GARCH(1,1) volatility model have been obtained as 3094.39, 3136.74 and 3122.21, respectively, for the return series on BSESENSEX. Applying the LR test under the null hypothesis : $H_0 : v = 2$ against the alternative hypothesis $H_1 : v \neq 2$ under the GED family, we find the test statistic value is 84.7 which is very high as compared to the χ_1^2 critical value of 6.63 at 1 percent level of significance, and hence the null hypothesis is strongly rejected in favour of GED. A similar conclusion regarding Gaussian distribution against the standardized Student's $t-$distribution can perhaps be drawn heuristically, but not on the basis of a formal test. Thus we can conclude that the assumption of normality is not an appropriate distributional assumption against the alternative of GED as well as, in a similar spirit, against standardized Student's $t-$distribution for BSESENSEX return series under GARCH(1,1) volatility model. It is obvious from Table 6.4 that the same conclusion holds for each of the other two return series on BSE 100 and NIFTY. Therefore, we can finally conclude that, under the assumption of GARCH(1,1) volatility model, the conditional distributional assumption of normality is not statistically quite tenable for all the three return series.

Diagnostic statistics computed for $\hat{\varepsilon}_t$, the implied GARCH(1,1) residuals, are also given in Table 6.4. It turns out, as seen Chapter 3 as well, that GARCH(1,1) is convenient and adequate for capturing the time-varying variances of all the three Indian stock returns. LM test statistic values for testing the hypothesis of homoscedasticity against conditional heteroscedasticity are found to be insignificant for the residuals of all the three distributional assumptions. Finally, we note that Lomnicki-Jarque-Bera test on normality shows strong evidence against the Gaussian distribution.

As we have found in this chapter that although GARCH(1,1) specification adequately captures the volatility of the three series, there is however, some evidence of leverage effect in the Indian stock returns. Hence, similar computations were carried out with TGARCH(1,1) volatility model under the three different distributional assumptions. Detailed estimation and diagnostic results under TGARCH(1,1) volatility model are

Table 6.5 Estimates of Parameters and Diagnostics for Residuals $\hat{\varepsilon}_t$ under TGARCH(1,1) Volatility Model and Different Distributional Assumptions

	BSESENSEX			BSE 100			NIFTY		
	$N(0,h_t)$	$t(v,0,h_t)$	$GED(0,h_t)$	$N(0,h_t)$	$t(v,0,h_t)$	$GED(0,h_t)$	$N(0,h_t)$	$t(v,0,h_t)$	$GED(0,h_t)$
Estimate/ test statistic value									
$\hat{\alpha}_0$	0.00005	0.00004	0.00004	0.00009	0.00007	0.00008	0.00003	0.00003	0.00003
	(3.218)**	(2.694)**	(2.757)**	(3.494)**	(3.303)**	(3.427)**	(3.356)**	(2.520)**	(2.609)**
$\hat{\alpha}_1$	0.013	0.034	0.246	0.084	0.089	0.086	0.040	0.049	0.045
	(0.839)	(1.303)	(1.000)	(2.532)	(2.129)	(2.009)	(2.612)**	(2.067)	(1.918)
$\hat{\gamma}$	0.144	0.126	0.135	0.230	0.209	0.220	0.084	0.068	0.076
	(3.426)	(2.644)	(2.752)	(3.439)	(2.681)	(2.810)	(2.930)	(1.896)	(2.015)
$\hat{\delta}$	0.762	0.793	0.776	0.564	0.632	0.595	0.803	0.825	0.811
	(11.378)	(13.117)	(11.570)	(5.163)	(7.725)	(6.747)	(17.654)	(16.195)	(15.264)
$\hat{v}$	_	6.977	1.442	_	6.287	1.377	_	6.972	1.405
		(6.634)	(24.210)		(6.187)	(21.559)		(6.215)	(22.458)

Table 6.5 (contd.)

Max log$^{\$}$	3104.44	3142.37	3128.56	2552.23	2587.93	2578.10	3211.74	3244.00	32366.66
No remaining ARCH (LM, 1 lag)									
$F-Test$	0.435	0.122	0.233	0.455	0.654	0.535	0.038	0.014	0.007
$p-value$	0.510	0.727	0.630	0.500	0.419	0.465	0.845	0.906	0.935
Lomnicki-Jarque-Bera (LJB) test									
LJB	489.662	594.417	540.573	485.12	543.627	512.188	296.451	305.132	311.430
$p-value$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

*Notes: The values in the parentheses indicate of the corresponding t-statistic values(absolute). ** indicates significance at 1% level whereas * indicate significance at 5% level .$ this row presents the maximum log-likelihood values.*

given in Table 6.5. It is evident from the entries in the third row of this table that γ is significant and positive under all the distributional assumtions for all the three series, and hence we may conclude that irrespective of the conditional distribution assumed, there is significant leverage effect in all the series.

Also, as in the case of GARCH(1,1) volatility specification, estimated values of v indicate that the excess kurtosis of the conditional distribution is significant for all the three series. Further, computational results for the returns on BSESENSEX series, for instance, show that the maximum log-likelihood values for the three assumed distributions *viz*., normal distribution, standardized Student's $t-$distribution and GED for this volatility model of TGARCH(1,1) are 3104.44, 3142.37 and 3128.56, respectively. Applying the LR test, it is evident that as in case of GARCH (1,1) volatility model, the assumption of normal distribution is rejected against GED and, heuristically speaking, against the alternative of standardized Student's $t-$distribution also. This is true for all the three series. As regards residual diagnostics, the findings are the same as in GARCH(1,1) i.e., TGARCH(1,1) is adequate as a volatility model for Indian stock returns under all the three distributional assumptions.

6.4 Conclusions

In this chapter, we have studied predictability of Indian stock returns under alternative volatility specifications as well as conditional distributional assumptions. To be more specific, we have considered (i) two alternative models of volatility *viz.,* EGARCH and TGARCH models for representing the phenomenon of 'leverage effect' in returns, and (ii) two alternative conditional distributions for the innovations – standardized Student's $t-$distribution and standardized GED – so that the leptokurtic property of the return distribution is supported. To find which of the three assumed volatility models (including GARCH) best describes the volatility prevalent in the Indian stock returns, standard forecasting criteria like the MAE and RMSE corresponding to out-of-sample period have been used. As regards which distribution (including Gaussian) produces the best fit to the empirical process, the criterion of maximized log-likelihood value has been chosen.

In our empirical analysis, we have found significant presence of 'leverage effect' in the Indian stock returns by both the EGARCH and TGARCH models. The values of MAE and RMSE for the hold-out sample clearly demonstrate that under Gaussian distribution the three volatility models GARCH(1,1), TGARCH(1,1) and EGARCH(1,1) – perform almost equally well by both these criteria. In fact, this holds for returns on BSESENSEX and NIFTY entirely and for returns on BSE 100 to a great extent. Thus, it can be concluded that between EGARCH and TGARCH models, the two volatility models capturing 'leverage effect' in returns, there is practically no difference in terms of the chosen criteria. However, as regards comparing these two volatility models against the standard GARCH model which disregards asymmetry, we have found that for both the BSESENSEX and NIFTY series there is hardly any difference, but for BSE 100, the MAE and RMSE values suggest that the EGARCH/TGARCH model is slightly better than the simple GARCH model.

Insofar as the performance of the models under alternative distributional assumptions are concerned, we find that the volatility parameters are significant for all the three series under all the three distributions considered. Further, it is also evident that the excess conditional kurtosis, which is a parameter in both standardized $t-$distribution and GED, is significantly present in all the three series indicating thereby that the standard conditional normality assumption is not adequate. We have also found that the assumption of normality for the conditional distribution is not statistically quite tenable against GED and also perhaps against the standardized Student's $t-$distribution under both the GARCH(1,1) and TGARCH(1,1) volatility models, for all the three return series considered in this study.

CHAPTER 7

Summary and Future Ideas

7.1 Introduction

A systematic and comprehensive study on predictability in the Indian stock market has been carried out in this thesis. For this study, time series of the four major stock indices of India *viz*., BSESENSEX, BSE 100, NIFTY and DOLLEX at daily as well as monthly frequencies, have been used. All throughout this thesis, attempts have been made to deal with the relevant econometric issues by applying available modern econometric techniques. To this end, due emphasis has been given, wherever appropriate, on modelling issues such as correct specification of the conditional mean as well as of the conditional variance, and testing for the adequacies of these specifications so that inferences on predictability are not based on misspecified models. The consideration for appropriate specification is due to the well-known fact that any misspecification(s) of the first two conditional moments may lead to misleading inferences about the models on returns and consequently on the predictability of the stock market. Likewise, we have applied predictive regression technique to identify those macro and financial variables which have predictive ability for returns. This has been done both in terms of in-sample forecasts and out-of-sample tests of return predictability. Further, cointegration analysis has also been used to study the long-run relationship involving stock indices and relevant macro and financial variables. Econometric issues like alternative specifications (other than GARCH) of the volatility model and non-Gaussian assumptions for the conditional distribution of the innovations have also been considered in our study. In the next section of this chapter we summarize the major findings of this work, and also mention about the limitations of our study. This chapter as also this thesis ends with some discussions on a few ideas for further work on the predictability aspect of Indian stock market.

7.2 Major findings

Before we briefly discuss the major findings, it is appropriate to point out that the time series used in this study covers 31 December 2000 (for Chapters 3 and 6) and December 2002 (for Chapters 4 and 5). Since the analysis excludes the recent years, it is prudent to discuss briefly the likely impact of excluding the most recent period on the empirical findings of the thesis. To that end, it may be pointed out that the reform process in India has significantly slowed down during the last few years. This has been noted by researchers and analysts (see, for instance, Swamy (2005) and Rakshit (2007)). This is primarily because of sharp differences among the political parties of the ruling coalition, mainly in respect of liberal economic policies to be followed. The present Indian government is a coalition of a number of political parties of divergent political ideologies and economic programmes, and is supported by communists from outside. There are economic factors as well behind this slow-down, as discussed in the cited references. Since the thesis deals with studying the aspect of predictability prevailing in the Indian stock market, our contention is that the major conclusions on predictability are unlikely to change since there has not been any major structural adjustments in the Indian economy during the last few years. Nor have there been any major reform initiatives in the recent past in any of the important sectors of India. Undoubtedly, like any other empirical work, computational figures may change if the length of the time series is increased. But, some improvement in efficiency of the Indian economy notwithstanding, we think that the likely impact of not including the most recent period should be marginal, if at all, in terms of inferences on predictability of India's stock market.

This thesis is concerned with studying the predictability aspect of the Indian stock market, and hence we have first presented some relevant facts about the Indian stock market so that India's present position as a major emerging market economy is briefly established. Additionally, in Chapter 2, we have mentioned about some of the major structural and regulatory reforms which were introduced since India embarked on liberalising its economy earnestly and steadily in 1992. The content of this chapter should be considered to be the bare minimum exposition about the stock market of India in the post-liberalisation period.

In the third chapter, we have carried out a study on predictability of Indian stock returns based on daily level data. The modelling approach followed here involves a linear dynamic model for the conditional mean of return and nonlinear dependence for the conditional variance in the form of GARCH model. Further, the approach also envisages appropriate specification of both the conditional mean and conditional variance so that the data analysis does not produce misleading inferences owing to any probable inappropriate specification of these moments. To start with, a battery of rigorous tests have been used to find if there are structural break(s) as well as serial correlations in the returns on the four stock index series *viz*., BSESENSEX, BSE 100, NIFTY and DOLLEX, considered for this study, and also to test for misspecification in the conditional mean. Thereafter, the model for return has been specified by incorporating in the conditional mean the independent variable, call money rate, to represent short-term interest rate, 0-1 dummies to capture the day-of-the week effects in returns, polynomials of recursive residuals to represent mispecification in the mean function and, of course, lagged values of return to take into account the serial correlations. The day-of-the week dummies have also been included in the GARCH specification so as to represent daily level seasonality in conditional heteroscedasticity.

The first major finding in Chapter 3 suggests that there is statistical evidence in favour of two structural breaks – one in mid-1992 and the other in early or late 1996, depending on the index. This finding is quite consistent with the recent history of the Indian stock market. Insofar as predictability is concerned, it has been found that returns for all the four series are predictable, and this observed predictability in daily returns is due to serial correlation, conditional heteroscedasticity, significant day-of-the week effects both in conditional mean and conditional variance, short-term interest rate as represented by call money rate (in some sub-periods of some indices only) and some dynamics in higher order moments. The standardized residuals of the fitted return models for the four return series were finally subjected to a test for null of i.i.d property. Based on the BDS test, we have found that the null is rejected for some combinations of m and ξ/σ values leading us to conclude that the GARCH model for volatility is not adequate for capturing the entire nonlinear dependencies in the returns on all the four stock indices.

This study with daily-level returns has been constrained by the fact that other exogenous variables – macroeconomic or financial- could not be added in the model specification since, as in other countries, data for most of such variables at daily level frequency are not available for India also. Finally, we may pose the question: Can the investors in the Indian stock market really exploit the sources of observed dependencies in daily returns, which are primarily either their own past or simply calendar dummies, for the purpose of making profit? Based on our findings on predictability, we may conclude that investors can, in fact, benefit from such relationships. However, modern financial economics, particularly the newly developed behavioural finance, suggests that some perfectly rational factors such as transactions costs and liquidity may account for such predictability. For example, in the presence of transactions costs arbitrageurs may not have enough opportunity to earn excess profit although some of these costs, such as trading costs, are probably not large in liquid markets. Transactions costs associated with covering short position in a not-so-liquid market can be substantial and can limit arbitraging opportunities (Shleifer (2000)). The resulting profit making may not also be implementable if there is not enough liquidity in the market to enable the trades to take place when the strategy says these should. Despite some limitations of statistical tests based on large data sets, like, for instance, standard tests can find statistically significant effects even when there are small deviations from the null hypothesis, or the fact that it is likely that most empirical studies evaluating –sample predictability overstate out-of-sample predictability because of overfitting, finite sample biases and data snooping, the fact remains that the majority of the literature focus on the statistical evidence of predictable time variation in expected returns. Further, the econometric evidence for structural change, or stated differently, the non-existence of a stable relationship for prediction of returns is quite a common finding in economic time series modelling, especially if the span of the series is long enough since " structural change is pervasive in economic time series relationships" (Hansen (2001), p 127). In such cases it is prudent to keep in mind that only short-term predictions should better be made, and that the relations should be updated with passage of time when more observations become available.

After the study with daily returns, we have considered a similar study with monthly returns. Now at monthly level frequency, data for all major macrovariables and financial ratios are available for India. However, the existing studies between stock returns and such variables show a lack of uniformity in identifying the set of macro and financial variables as predictors of return. In other words, empirical evidence do not identify, most often, the same set of variables across different such studies; rather, these studies suggest varying sets some variables being common. To deal with this problem of choosing the appropriate set of macrovariables and financial ratios for predicting the monthly stock returns of India based on three stock indices *viz.,* BSESENSEX, BSE 100 and NIFTY, we have applied what is known as predictive regression approach. In this procedure, we find the predictive ability of each of these variables separately for return by using in-sample forecasts and out-of-sample tests of return predictability. In addition to analysing the predictive ability of each of these variables, we have also used a procedure that combines both general-to-specific and specific-to-general model selection procedures with out-of-sample tests of predictability in an effort to identify and test the 'best' forecasting model of stock returns of India. Following this procedure, we have first identified the relevant set of macro and financial variables from a set of thirteen macrovariables and three financial ratios which have been found to affect predictability of return in other such studies. Because of non-availability of data for some variables prior to April 1996, this work has been done with data covering the period April 1996 to

December 2002 which, incidentally, is a stable period with no structural break, as evidenced in Chapter 3.

The results of this study, presented in Chapter 4, have found the following six macrovariables *viz.,* inflation rate, change in nominal exchange rate, NASDAQ composite return, growth of foreign direct investment and changes in long-term interest rate as also in fiscal deficit of the central government which have significant predictive ability for returns. This set of macrovariables having predictive ability for returns has been found to be the same for all the three stock indices. Among the three financial ratios, only change in price-earnings ratio was found to be significant and that too for returns on BSESENSEX only. Once the significant variables have been identified, a dynamic linear regression model incorporating these variables, monthly seasonal dummies and other terms from consideration of appropriate specification has been considered, and we have found the lagged values of both inflation rate and change in nominal exchange rate and the contemporaneous values of NASDAQ composite return and change in price-earnings ratio to be significant predictors for returns on BSESENSEX. For returns on BSE 100 and NIFTY, the predictors which have been found significant are: lagged value of inflation rate and contemporaneous values of returns on NASDAQ composite index and change in long-term interest rate. It may also be noted that volatility was not found to be significant for any of the three return series at monthly level. This latter finding is not uncommon since similar evidence for other data sets have also been noted. When returns are at low frequency, like monthly or quarterly level, the clusterings of large (small) deviations which are pronounced at higher frequency, say at daily level, get clubbed together at the aggregated frequency level, and hence episodes of volatility are much reduced or non-existent in the monthly or quarterly level returns data. Further, a look at the signs of the coefficients in the fitted models shows that these signs are as expected and, in fact, these are supported by economic theories as well. While our finding of significant negative coefficient value for inflation rate being easily explainable, the one concerning negative relation between stock returns and change in long-term interest-rate for both BSE 100 and NIFTY only shows that the current policy of lowering interest rate in India has, as suggested by economic reasonings, favourably affected the Indian stock market. Similarly, the significant positive impact of NASDAQ composite return on Indian stock return implies the growing integration of Indian capital market with other major capital markets in the desired direction.

Finally, we may raise the same question as in Chapter 3. In this present study, we have been primarily concerned with testing for return predictability using macrovariables and financial ratios in populations and not necessarily whether an investor in real time could have used one or more of these variables to earn super-normal profits. While our use of out-of-sample tests somewhat mimics the situation of an investor in real time, unless we take into account data revisions by using real-time data when forming a portfolio, we would not really be able to infer about the advantage of return predictability from the investors' point of view. If, however, a macro or financial variable is typically not subject to revision and are available immediately, say, for example, long-term interest rate in our case, then our finding on the relevance of this variable on predictability of return is likely to be relevant for real-time also.

So far we have studied the aspect of predictability of return in the Indian stock market. Stock return being a stationary variable, its predictions would refer to short-run periods only. While it is true that investors are primarily concerned with returns, rather than stock price/index, and to that extent studying predictability of return is very important for any stock market, yet a study on predictability in any stock market should also envisage studying the relationship, if any, between a stock price/index and the relevant macro and financial variables – all at their nonstationary level values – so that it can be found if there exists any long-run relation involving the stock index and these variables. In the event of such a relationship indeed existing in any stock market, it can be inferred that the stock price/index has a comovement with the variables over time and hence the stock market is predictable in the long-run sense. It is well-known that if all the variables are $I(1)$ variables, such a study of finding long-run relation(s) involving stock return and other relevant variables can be carried out by using the methodology of cointegration. Thus, in this thesis, we have next taken up, in Chapter 5, a cointegration analysis involving a stock index and a set of relevant macrovariables and financial ratios- each being $I(1)$ - for the purpose of studying long-run predictability in the Indian stock market. This exercise on cointegration has been done separately for three indices *viz*., BSESENSEX, BSE 100 and NIFTY. Following essentially Johansen's approach in which allowance was made for including a constant, a linear trend term and monthly dummy variables in the model, we have found, based on the trace and maximum eigen value tests, that cointegrating relations indeed exist and the set of variables found to be significant in the cointegrating relations are domestic industrial production, consumer price index, nominal exchange rate, foreign direct investment and long-term interest rate. Further, this set of macrovariables is the same for all three cointegrating exercises corresponding to the three stock indices. None of the financial ratios has been found to have any significant effect in the cointegrating regressions. From the cointegrating regressions, we note that two significant macrovariables, namely, consumer price index and nominal exchange rate have negative and positive signs, respectively, in explaining the return in the long-run sense. These are quite understandale and expected since these are supported by the standard economic theories involving these variables. What, however, is a little uncommon is the finding that foreign direct investment and stock indices in case of India has, in the long run, a negative relationship. The theory would normally suggest a positive relationship between the two. It may, however, be noted that such findings have also been obtained by other researchers with other stock indices, primarily of developing/emerging economies. The explanation probably lies in the fact that in the initial stage of liberalisation, as in the case with India, domestic investment is been crowded out by FDI.

The last part of the thesis is concerned with studying the predictability of Indian stock return under alternative volatility specifications as well as conditional distributional assumptions. In Chapter 3, we have studied predictability of daily returns based on daily closing stock indices of the Indian stock market. The assumptions of GARCH model for volatility and Gaussian distribution for conditional distribution for innovations in that chapter were made from consideration of the facts that these assumptions have most often been made in empirical studies and the results have been found to be quite satisfactory, and that the relevant econometric theories are well-developed for the underlying model for return under these assumptions. Now, in spite of these facts, it is also true that there is widespread evidence of 'leverage effect' in stock returns and the standard GARCH model is unable to capture this effect. Further, it is a widely accepted fact that most return data exhibit leptokurtosis and hence the assumption of conditional normal distribution, strictly speaking, is not tenable. Keeping these two observed properties of returns in mind, we have, in Chapter 6, considered (i) two alternative volatility models *viz.,* the EGARCH and TGARCH models and (ii) two alternative conditional distributions for the innovations *viz.,* the standardized Student's $t-$distribution and the standardized generalized error distribution. The performance of these volatility models (including GARCH) has been evaluated by standard forecasting criteria like the MAE and RMSE corresponding to the out-of-sample period, and the comparison (in the sense of best fit to the return distribution) among the three distributional assumption (including the Gaussian distribution) has been done in terms of maximized value of log-likelihood function.

This study has been carried out with daily return data for the post-liberalisation period spanning January 1996 to December 2000 which is a stable period, as found by structural break analysis in Chapter 3. The first major finding in this chapter is that there is significant presence of 'leverage effect' in the Indian stock returns by both the EGARCH and TGARCH models. Thus, while GARCH(1,1) model was found, in Chapter 3, to adequately capture the volatility in the Indian stock returns, we find from this exercise that alternative volatility specifications like the EGARCH and TGARCH models also adequately represent the volatility of the series. In fact, the values of MAE and RMSE for the hold-out sample show that under Gaussian distribution all the three volatility models –GARCH(1,1), TGARCH(1,1) and EGARCH(1,1) – perform equally well for returns on both BSESENSEX and NIFTY, and almost so for returns on BSE 100. As regards the superiority between the two volatility models representing asymmetry, i.e., TGARCH and EGARCH models, we have found that there is practically no difference between the two by both the MAE and RMSE criteria. Insofar as the comparison between the GARCH model which does not capture asymmetric property of returns and the two asymmetric volatility models (i.e., TGARCH and EGARCH models) is concerned, we have found no evidence of any significant difference between these two classes of models for both BSESENSEX and NIFTY series, but the latter class of models has been found to perform slightly better than the simple GARCH model for returns on BSE 100 stock index. Now, our major finding on the performance of the return models under alternative distributional assumptions is that the volatility parameters are significant for all the three return series under all the three distributions *viz.*, Gaussian, standardized $t-$distribution and GED. Further, it has also been found that the parameter representing excess kurtosis in $t-$distribution and GED is significant for all the three return series. Thus, it shows that the assumption of conditional normal distribution for innovations is not quite appropriate for returns on Indian stock indices. Finally, we have found that the assumption of normality for the conditional distribution is not found to be acceptable against GED as well as perhaps against the standardized Student's $t-$distribution, under both GARCH(1,1) and TGARCH(1,1) volatility models, for all the three return series on BSESENSEX, BSE 100 and NIFTY.

To sum up, we have, in this thesis, carried out a systematic and comprehensive study on predictability in the Indian stock market where due considerations have been given to various econometric as well as economic issues like structural break, appropriate specification of first- and second-order conditional moments, alternative volatility models as well as conditional distributions for the innovations, short-run and long-run predictability and roles of macrovariables as well as financial ratios. Applying various econometric techniques including structural break analysis, specification analysis, predictive regression approach, cointegration analysis, forecasting criteria like the MAE and RMSE and the recently- developed out-of-sample tests of predictive ability on all the time series data representing India's major stock indices/returns at daily and/or monthly levels of frequency, we have found that, despite reforms in the Indian economy during the last one-and-a half decades, the Indian stock market remains predictable. This finding underscores in case of Indian stock market, the notion held by some researchers that stock returns have only a small predictable component and are inherently difficult to predict, and this obviously has significant implications for the efficacy of the stock index futures as a vehicle for investment in India. Yet, given the nature of debate in the literature on market efficiency/predictability, as mentioned in Chapter 1, the question may still arise: Can the observed predictability in the Indian stock market be really useful for investors in fashioning an investment strategy that will dependably earn excess returns? The key factor to the answer is whether any patterns of serial correlation as well as second order dependencies and seasonal (daily/monthly) effects, deterministic roles of other variables etc. are consistent over time. In fact, as Schwert (2001) and Malkiel (2003) among others, have noted, many predictable patterns seem to disappear after they are published in the finance literature. One plausible explanation could be that practitioners, perhaps, learn quickly about any true predictable pattern and exploit it to the extent that it becomes no longer profitable. It is quite possible that such patterns are not stable enough to generate consistently superior investment results. Finally, the issue of transactions costs and liquidity needs to be discussed in this context. Just finding some statistical evidence of predictability is not sufficient to conclude either that there are arbitrage opportunities or that the market is inefficient since transactions costs may outweigh any profit opportunities. Further, the strategies towards exploiting the predictability may not be implementable if there is not enough liquidity in the market to enable the trade to take place. This may hold for the Indian stock market as well!

7.3 Ideas for future work

India is now considered to be one of the most important emerging market economies with a huge potential for future growth. Naturally, studies concerning various important economic and financial variables of India would continue to grow. Our study, in this thesis, has focussed on the Indian stock market, that too on a particular aspect of the Indian stock market, namely, predictability. Confining ourselves to the aspect only, we present below some ideas towards future work on the Indian stock market.

In our study, we have used the framework of a dynamic model with linear conditional mean and nonlinear conditional heteroscedasticity. Such models are, in fact, commonly and extensively used in empirical finance. But, there is a growing literature on nonlinear time series models that can be potentially useful for modelling and forecasting economic and financial variables. Obviously, therefore, future work on modelling returns on Indian stock indices should involve these recent advances in nonlinear time series models such as regime-switching models like the threshold autoregressive (TAR) model, self-exciting TAR (SETAR) model and smooth transition autoregressive (STAR) model. These models may be used to study the predictability of Indian stock returns first without consideration to volatility and then with different models for volatility. Comparison across various models may be carried out using out-of-sample forecasting criteria including recently-developed tests of predictability.

Another direction of future work could be in the area of volatility models used for representing volatility in returns. During the last decade a good many model for volatility - mostly semiparametric and non-parametric have appeared in the literature. Earlier developments on volatility modelling were parametric, but the recent literature has moved towards semi-parametric and even fully nonparametric directions. The parametric procedures rely on explicit functional form assumption about the volatility under consideration. In the ARCH class of volatility models in discrete-time frame, the expectations are made in terms of directly observable variables. There are also continuous-time stochastic volatility models which involve latent state variables. The nonparametric procedures, on the other hand, are generally free from such functional form assumptions and hence afford estimates of notional volatility that are flexible yet consistent. These procedures include ARCH filters and smoothers which are formulated to measure the volatility over infinitesimally short-horizons, as well as realized volatility measures (see Andersen *et al*. (2005) and Linton and Mammen (2005), for details on such models) for fixed length time intervals. More clearly, ARCH filters and smoothers are based on the assumption of ever more observations over ever finer time intervals (a double limit theory), while the realized volatility measures build on the idea of an increasing number of observations over fixed length time intervals (a single limit theory). It would, therefore, be an interesting exercise to study the predictability of Indian stock returns using these recently-developed volatility models and compare their performances with those of the parametric models for volatility.

The issue of testing for predictability for predictability in population versus real-time predictability should be another topic of useful empirical research in the context of

Indian stock market. While our use of out-of-sample tests reflects, by and large, the situation of an investor in real time, we would also have to take account of data revisions by using real-time data when forming a portfolio, and hence it would be interesting in future research to test for predictability of the Indian stock return using macro and financial variable in real time. It would also be interesting to examine whether some macro and/or financial variables perform significantly better in the Indian stock market if one allows for time-varying effects of these variables on stock returns. This would allow, for example, the effects of macro and financial variables on returns to vary across the stage of the business cycle.

We conclude by mentioning another area of future research which concerns the practical use of these linear and nonlinear time series models. As we understand, the informative content in returns data, or for that matter in any financial data (that can be exploited for generating reliable forecasts) is not equally distributed over the observation. To put it differently, some data points are more important than others in the sense that they can be predicted relatively easily or that they can serve a useful basis for generating forecasts. In practical terms, this would imply that, even though model is specified for all observations, it is used only infrequently. Inclusion of such realistic considerations as well as other known and not-so-known empirical evidence into theoretical and empirical models seems to be a challenge for the future.

References

- Abdalla, I.S.A. and V. Murinde (1997). " Exchange Rate and Stock Price Interactions in Emerging Financial Markets : Evidence on India, Korea, Pakistan and the Philippines." *Applied Financial Economics* 7 : 25-35.
- Agrawal, A. and K. Tandon. (1994). " Anomalies or Illusions ? Evidence from Stock Markets in Eighteen Countries." *Journal of International Money and Finance* 13 : 83-106.
- Ammermann, P.A. and D.M. Patterson (2003). " The Cross-sectional and Cross-temporal Universality of Nonlinear Serial Dependencies : Evidence from World Stock Indices and the Taiwan Stock Exchange." *Pacific-Basin Finance Journal* 11 : 175-195.
- Andersen, T.G.; T. Bollerslev, and F.X. Diebold (2005). " Parametric and Nonparametric Volatility Measurement" in *Handbook of Financial Econometrics*, ed. L.P. Hansen and Y. Ait-Sahalia, North-Holland, Amsterdam.
- Andrews, D.W.K. (1991)."Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica* 59: 817-858.
- Andrews, D.W.K. (1993). "Tests for Parameter Instability and Structural Change with Unknown Change Point." *Econometrica* 61 : 821-856.
- Andrews, D.W.K. (2003). "Tests for Parameter Instability and Structural Change with Unknown Change Point: A Corrigendum." *Econometrica* 71 : 395-397.
- Andrews, D.W.K., and W. Ploberger. (1994). "Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative." *Econometrica* 62 : 1383-1414.

- Ang, A. and G. Bekaert. (2001). "Stock Return Predictibility : Is It There*?", NBER Working Paper 8207*, National Bureau of Economic Research, Inc.
- Antoniou, A. ; N. Ergul and P. Holmies. (1997). " Market Efficiency, Thin Trading and Non-linear Behaviour : Evidence from an Emerging Market." *European Financial Management* 3 : 175-190.
- Asprem, M. (1989). " Stock Prices, Asset Portfolios and Macroeconomic Variables in Ten European Countries." *Journal of Banking and Finance*13: 589-612.
- Bachelier, L. (1900). "Theory of Speculation." *In The Random Character of Stock Market Price*, ed. P. Cootner. 17-78, Reprint, Cambridge: MIT Press, 1964.
- Bai, J. (1994). "Least Squares Estimation of a Shift in Linear Processes." *Journal of Time Series Analysis* 15, no. 5: 453-472.
- Bai, J. (1997a). "Estimation of a Change Point in Multiple Regression Models." *Review of Economics and Statistics* 79, no. 4: 551-563.
- Bai, J. (1997b). "Estimating Multiple Breaks One at a time." *Econometric Theory* 13, no. 3 : 315-352.
- Bai, J., and P. Perron. (1998). "Estimating and Testing Linear Models with Multiple Structural Changes." *Econometrica* 66, no. 1 : 47-78.
- Baillie, R.T. and Bollerslev, T. (1989). " Common Stochastic Trends in a System of Exchange Rates." *Journal of Finance*, 44, no. 1 : 167-181.
- Balvers, R. ; T. Cosimano and T. McDonald. (1990). " Predicting Stock Returns in an Efficient Market." *Journal of Finance* 45 : 1109-1118.

- Basu, K. (2004). The Indian Economy : Upto 1991 and Since In *India's Emerging Economy : Performance and Propectus in the 1990s and Beyond* ed. K. Basu , MIT Press.
- Basu, P., and M. Morey. (1998). "Stock Market Prices in India after Economic Liberalisation." *Economic and Political Weekly* 33, no. 7 : 355-358.
- Bekaert, G. and C. R. Harvey. (2000). " Foreign Speculators and Emerging Equity Markets." *Journal of Finance* 55: 565-613.
- Bhattacharya, K., and G.P. Samanta (2003). ""A Tale of Two Indices: The Story of the NASDAQ and the SENSEX." *Journal of Quantitative Economics* 17, no. 2 89-102.
- Bhattacharya, K. ; N. Sarkar, and D. Mukhopadhyay. (2003). "Stability of the Day of the Week Effect in Return and in Volatility at the Indian Capital Market: A GARCH Approach with Proper Mean Specification." *Applied Financial Economics* 13, no. 8: 553-563.
- Bhaumik, S.K. (1997). "Stock Index Futures in India : Does the Market Justify its Use?" *Economic and Political Weekly* 32, no. 41: 2608-2611.
- Bhole, L.M., and S. Pattanaik. (2002). "The State of Indian Stock Market Under Liberalisation." *Finance India* 16, no. 1: 159-80.
- Black, F. (1976). Studies of Stock Price Volatility Changes In *Proceedings of the 1976 Meeting of the Business and Economics Statistics Section,* American Statistical Association, 177-181.
- Bollerslev, T. (1986)."Generalised Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31: 307-327.

- Bollerslev, T. (1987). “ A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return.” *Review of Economics and Statistics* 69 : 542-547.
- Bollerslev, T. and J. Wooldridge. (1992). “ Quasi Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances.” *Econometric Reviews* 11 : 143-172.
- Bollerslev, T. ; R.F. Engle and D.B. Nelson. (1994). ARCH Models In *Handbook of Econometrics vol. 4 ,* ed. R. Engle and D.McFadden Elsevier Science Publishers, Amsterdam.
- Brock, W.A. ; W. Dechert, and J. Scheinkman.(1987). “A Test for Independence Based on the Correlation Dimension.” Working Paper, University of Wisconsin-Madison, University of Houston and University of Chicago..
- Brock, W.A.; W. Dechert; J. Scheinkman and B. LeBaron. (1996).“A Test for Independence Based on the Correlation Dimension.” *Econometric Reviews* 15, no.3: 197-235.
- Brock, W.A., Hsieh, D.A. and LeBaron, B. (1991). *Nonlinear Dynamics, Chaos and Instability: Statistical Theory and Economic Evidence*, Cambridge, MA: MIT Press.
- Brockwell, P.J., and R.A. Davis. (1987). *Time Series: Theory and Methods*, Springer Verlag, Newyork.
- Brooks, C., and S.M. Heravi.(1999). “The Effect of (Mis-Specified) GARCH Filters on the Finite Sample Distribution of the BDS Test.” *Computational Economics* 13, no. 2: 147-162.

- Brown, R.L. ; J. Durbin, and. J.M. Evans. (1975). "Techniques for Testing the Constancy of Regression Relationship over Time." *Journal of the Royal Statistical Society* Series B, 37 : 149-192.
- Campbell, J.Y. (1987). "Stock Returns and the Term Structure." *Journal of Financial Economics,* June*,* 18 : 373-399.
- Campbell, J.Y. ; A.W. Lo and A.C. MacKinlay.(1997). *The Econometrics of Financial Markets*. Princeton : Princeton University Press, 1997.
- Campbell, J.Y. and R.J. Shiller. (1988a). " The Dividend-Price Ratio Expectations of Future Dividends and Discount Factors." *Review of Financial Studies* 1 : 195-228.
- Campbell, J.Y. and R.J. Shiller. (1988b). " Stock Prices, Earnings and Expected Dividends." *Journal of Finance* 43 : 661-676.
- Campbell, J.Y. and R.J. Shiller. (1998). " Valuation Ratios and the Long-run Stock Market Outlook." *Journal of Portfolio Management* Winter 11-26.
- Caporale, G.M. and N. Pittis . (1998). " Cointegration and Predictability of Asset Prices." *Journal of International Money and Finance* 17, no. 3 : 441-453.
- Chan, M.W.L.; A. Khanthavit and T. Hugh (1996). " Seasonality and Cultural Influences on Four Asian Stock Markets." *Asia Pacific Journal of Management*, 13 : 1-24.
- Chen N. ; R. Roll and S. Ross (1986). " Economic Forces and the Stock Market." *Journal of Business* 59 : 383-403.
- Choi, I. (1999). "Testing the Random Walk Hypothesis for Real Exchange Rates." *Journal of Applied Econometrics* 14, no. 3 : 293-308.

- Chong, L.; H. Tan and A. Baharumshah (2001) "The Stock Market, Macroeconomic Fundamentals and Economic Growth in Malaysia." *Asia Pacific Journal of Economics and Business* 5: 44-65.
- Chong, T.T. (1995). "Partial Parameter Consistency in a Misspecified Structural Change Model." *Economics Letters* 49, no. 4 : 351-57.
- Choudhry, T. (1997). " Stochastic Trends in Stock Prices : Evidence from Latin American Markets." *Journal of Macroeconomics* 19: 285-304.
- Choudhry, T. (2000). "Day of the Week Effect in Emerging Asian Stock Markets : Evidence from the GARCH Model." *Applied Financial Economics* 10 : 235-42.
- Chow, G.C. (1960). "Tests of Equality between Sets of Coefficients in two Linear Regressions." *Econometrica* 28, no. 3 : 591-605.
- Clark, T.E. (2004). " Can Out-of-Sample Forecast Comparisons help Prevent Overfitting?" *Journal of Forecasting* 23 : 115-139.
- Clark, J. and E. Berko. (1997). " Foreign Investment Fluctuations and Emerging Market Stock Returns : The Case of Mexico, *Staff Report* 24, Federal Reserve Bank of New York, NY.
- Clark, T.E. and M.W. McCracken. (2001). " Tests of Equal Forecast Accuracy and Encompassing for Nested Models." *Journal of Econometrics* 105 : 85-110.
- Clark, T.E. and M.W. McCracken. (2004). " Evaluating Long-Horizon Forecasts." *University of Missouri Columbia, Manuscript*.
- Coleman, M. (1990). " Cointegration-based Tests of Daily Exchange Rate Market Efficiency, *Economics Letters*, 32 : 53-59.

- Conover, C.M.; G.R. Jensen and R. R. Johnson (1999). " Monetary Environments and International Stock Returns." *Journal of Banking and Finance* 23 : 1357-1381.
- Cowles, A. (1933)."Can Stock Market Forecasters Forecast?" *Econometrica* 1: 309-324.
- Cowles, A., and H. Jones (1937). " Some A Posteriori Probabilities in Stock Market Action." *Econometrica* 5 : 280-294.
- Das, S., and N. Sarkar. (2000). "An ARCH in the Nonlinear Mean (ARCH-NM) Model." *Sankhy$\overline{a}$. Series B* 62, Part 2 : 327-344.
- Dickey, D., and W.A. Fuller (1981). " Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root." *Econometrica* 49: 1057-72.
- Diebold, F. (1986). "Testing for Serial Correlation in the Presence of Heteroscedasticity." In *Proceedings of the American Statistical Association, Business and Economic Statistics Section* : pp 323-328. Alexandria, VA : American Statistical Association.
- Diebold, F.X. and R.S. Mariano (1995). " Comparing Predictive Accuracy." *Journal of Economics and Business Statistics* 13 : 253-263.
- Doornik, J.A.; D.F. Hendry and B. Nielsen. (1998). " Inference in Cointegrating Models : UK M1 Revisited." *Journal of Economic Su*rveys 12 : 533-572.
- Durham, J.B. (2001). " The Effects of Monetary Policy on Monthly and Quarterly Stock Returns : Cross Country Evidence and Sensitivity Analyses." Federal Reserve Board of Governors, Finance and Economics, Discussion Series Paper : 2001-42.
- Engle, R.F. (1982). "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation." *Econometrica* 50 : 987-1008.

- Engle, R.F. and C.W.J. Granger (1987). "Cointegration and Error Correction: Representation, Estimation and Testing." *Econometrica* 55 : 251-276.
- Engle, R.F. ; D.M Lilien.; and R.P. Robbins. (1987). "Estimating the Time Varying Risk Premia in the Term Structure : the ARCH-M Model." *Econometrica* 55 : 391-408.
- Engle, R.F. and V.K. Ng (1993). " Measuring and Testing the Impact of News on Volatility." *Journal of Finance* 48 : 1749-1778.
- Fama, E.F. (1970). "Efficient Capital Markets : A Review of Theory and Empirical Work." *Journal of Finance* 25, no. 2 : 383-417.
- Fama, E.F. (1981). " Stock Returns, Real Activity, Inflation and Money.*" American Economic Review* 71: 545-565.
- Fama, E.F., and K.R. French. (1988). "Dividend Yields and Expected Stock Returns." *Journal of Financial Economics* 22 : 3-27.
- Fama, E.F., and G.W. Schwert. (1977). "Asset Returns and Inflation." *Journal of Financial Economics* 5, issue 2 : 55-69.
- Faust, J. (1992). "When are Variance Ratio Tests for Serial Dependence Optimal?" *Econometrica* 60 : 1215-1226.
- Ferson, W., and C. Harvey. (1993). " The Risk and Predictability of International Equity Returns." *Review of Financial Studies* 6 : 527-566.
- Fifield S.; D. Power and C. Sinclair. (2000). " A Study of Whether Macroeconomic Factors Influence Emerging Market Share Returns." *Global Economy Quarterly* 1 : 315-335.

- Fields M.J. (1931). “ Stock Prices : A Problem in Verification.” *Journal of Business* 7 : 415-418.
- Flannery, M.J. and A.A. Protopapadakis. (2002). “ Macroeconomic Factors Do Influence Aggregate Stock Returns.” *Review of Financial Studies* 15: 751-782.
- French, K.R. (1980). “Stock Returns and the Weekend Effect.” *Journal of Financial Economics* 8 : 55-69.
- Froot, K.A.; P.G. O’Connel and M.S. Seasholes (2001). “ The Portfolio Flows of International Investors.” *Journal of Financial Economics* 59: 151-194.
- Gallingar, G.W. (1994). “ Causalty Tests of the Real Stock Return- Real Activity Hypothesis.” *Journal of Financial Research* 17 : 271-288.
- Gibbons, M. and P. Hess. (1981). “ Day of the Week Effects and Asset Returns.” *Journal of Business* 54 : 579-596.
- Giles, D.E.A. ; J.A. Giles., and J.K. Wong. (1993). “Testing for ARCH-GARCH Errors in a Misspecified Regression.” *Computational Statistics* 8 : 109-126.
- Glosten, L. ; R. Jagannathan and D. Runkle.(1993). “ Relationship between the Expected Value and the Volatility of the Nominal Excess Return on Stocks.” *Journal of Finance* 48 : 1779-1801.
- Goetzmann, W.N. and P. Jorion. (1993). “ Testing the Predictive Power of Dividend Yields.” *Journal of Finance*, 48 : 663-679.
- Gourieroux, C. and J. Jasiak (2001). *Financial Econometrics : Problems, Models, and Methods*, Princeton University Press.
- Granger, C.W.J. (1986). “ Developments in the Study of Co-integrated Economic Variables.” *Oxford Bulletin of Economics and Statistics* 48 : 213-228.

- Granger, C.W.J., and A.P. Andersen. (1978). *An Introduction to Bilinear Time Series Models.* Gottingen : Vadenhoeck and Reprecht.
- Granger C.W.J. ; B.N. Huang and C.W. Yang (2000). " A Bivariate Causality between Stock Prices and Exchange Rate : Evidence from Recent Asian Flu." *Quarterly Review of Economics and Finance* 40 : 337-354.
- Gupta, O.P., and V.Gupta. (1997). "A Re-examination of Weak-form Efficiency of Indian Stock Market." *Finance India* 11, no. 3 : 619-632.
- Hall, A. (1994). "Testing for a Unit Root in Time Series with Pretest Data-based Model Selection." *Journal of Business and Economic Statistics* 12 : 461-470.
- Hansda, S.K., and P. Ray. (2002). " BSE and NASDAQ : Globalisation, Information Technology and Stock Prices." *Economic and Political Weekly*, February, 459-468.
- Hansen, B.E. (1994). " Autoregrssive Conditional Density Estimation." *International Economic Review* 35: 705-730.
- Hansen, B.E. (2001). "The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity." *Journal of Economic Perspectives* 15 : 117-128.
- Harvey, D.L.; S. Leybourne and P. Newbold. (1998). " Tests for Forecast Encompassing ." *Journal of Business and Economic Statistics* 16 : 254-259.
- Harvey, C.R., and A. Siddique. (1999). " Autoregressive Conditional Skewness." *Journal of Financial and Quantitative Analysis* 34 : 465-487.
- Hinich, M.J., and D.M. Patterson. (1985). "Evidence of Nonlinearity in Daily Stock Returns." *Journal of Business and Economic Statistics* 3 : 69-77.
- Hondroyiannis G. and E. Papapetrou. (2001). " Macroeconomic Influences on the Stock Market." *Journal of Economics and Finance* 25 : 33-49.

- Hsieh, D.A. (1989). "Testing for Nonlinear Dependence in Daily Foreign Exchange Rates." *Journal of Business* 62, no. 3: 339-368.
- Hsieh, D.A. (1991). "Chaos and Nonlinear Dynamics : Application to Financial Markets." *Journal of Finance* 46, no. 5: 1839-1877.
- Hubrich, K. ; H. Lütkepohl and P. Saikkonen (2001). " A Review of Systems Cointegration Tests." *Econometric Reviews* 20 : 247-318.
- Ibrahim M.H. (1999). " Macroeconomic Variables and Stock Prices in Malaysia : An Empirical Analysis." *Asian Economic Journal* 13 : 219-31.
- Inoue, A. and L. Kilian (2005). " In-Sample or Out-of-Sample Tests of Predictability : Which One Should We Use?" *Econometric Reviews* 23, no. 4: 371-402.
- Jaffe J. and R. Westerfield. (1985). " The Week-End Effect in Common Stock Returns : The International Evidence." *Journal of Finance* 40 : 433-454.
- Jeon B. and T. Chiang (1991). " A System of Stock Prices in World Stock Exchanges : Common Stochastic Trends for 1975-1990." *Journal of Economics and Business* 43 : 329-338.
- Johansen, S. (1988). "Statistical Analysis of Cointegration Vectors." *Journal of Economic Dynamics and Control* 12 : 231-254.
- Johansen, S. (1991). "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica* 59: 1551-1581.
- Johansen, S. (1995). Likelihood-Based Inference in Cointegrated Vector Autoregressive Models, Oxford University Press, Oxford.

- Johansen, S. and K. Juselius (1990). "Maximum Likelihood Estimation and Inference on Cointegration- with Applications to the Demand for Money." *Oxford Bulletin of Economics and Statistics* 52 : 169-210.
- Johansen, S. ; R. Mosconi and B.Nielsen (2000). "Cointegration Analysis in the Presence of Structural Breaks in the Deterministic Trend." *Econometrics Journal* 3 : 216-249.
- John M. (1993). " Emerging Equity Markets in the Global Economy." *Federal reserve Bank of New York* 18: 54-83.
- Kawakatsu, H., and M.R. Morey. (1999). "Financial Liberalisation and Stock Market Efficiency : An Empirical Examination of Nine Emerging Market Countries." *Journal of Multinational Financial Management* 9 : 363-371.
- Keim, D.B., and R. Stambaugh. (1984). "A Further Investigation of the Weekend Effect in Stock Returns." *Journal of Finance* 39, no. 3: 819-835.
- Kianifard, F., and W.H. Swallow. (1996). "A Review of the Development and Application of Recursive Residuals in Linear Models." *Journal of the American Statistical Association* 91 : 391-400.
- Kilian, L.(1999). " Exchange Rates and Monetary Fundamentals: What do we learn from Long-horizon Regressions?" *Journal of Applied Econometrics* 14: 491-510.
- Kim, E.H. and V. Singal. (2000). " Stock Market Openings : Experience of Emerging Economies." *Journal of Business* 73: 25-66.
- Kirby, C. (1997). " Measuring the Predictable Variation in Stock and Bond Returns." *Review of Financial Studies*, 10 : 579-630.

- Kon, S.J. (1984). " Models of Stock Returns – A Comparison." *Journal of Finance* 39: 147-165.
- Kothari, S.P. and J. Shanken. (1997). " Book-to-Market, Dividend Yield, and Expected Market Returns : A Time Series Analysis." *Journal of Financial Economics* 44 : 169-203.
- Kulkarni, V. (1997). "'Badla' : the Mumbai Derivative." *Economic and Political Weekly* 32, no. 42: 2747-2752.
- Kumar, N. and J.P. Pradhan. (2002). " Foreign Direct Investment, Externalities and Economic Growth in Developing Countries : Some Empirical Explorations and Implications for WTO Negotiations on Investment." RIS Discussion Paper 27, Research and Information System for the Non-aligned and Other Developing Countries, New Delhi.
- Lakonishok J. and M. Levi. (1982). " Weekend Effects on Stock Returns : A Note." *Journal of Finance* 37 : 883-889.
- Lee, B.S. (1992). "Causal Relations Among Stock Returns, Interest Rates, Real Activity and Inflation." *Journal of Finance* 48: 1591-1603.
- Levich, R.M. (1985). " Empirical Studies of Exchange Rates: Price Behaviour, Exchange Rate Determination and Market Efficiency" In *Handbook of International Economics 2*, ed. R.W. Jones and P.B. Kenen, Elsevier Science, New York.
- Lim, K.P. and M. Hinich. (2005). " Cross-temporal Universality of Non-Linear Serial Dependencies in Asian Stock Markets." Economics Bulletin 7 : 1-6.
- Linton, O. and E. Mammen (2005). " Estimating Semiparametric ARCH (∞) Models by Kernel Smoothing Methods." *Econometrica* 73 , no. 3, 771-836 .

- Ljung, G.M., and G.E.P. Box. (1978). "On a Measure of lack of fit in Time Series Models." *Biometrika* 66 : 67-72.
- Lo, A.W., and A.C. MacKinlay. (1988). "Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test." *Review of Financial Studies* 1, no. 1: 41-66.
- Lovatt, D. and A. Parikh. (2000). " Stock Returns and Economic Activity : The UK Case." *European Journal of Finance* 6 : 280-297.
- Lu G.M. ; K. Metin and R. Argac. (2001). " Is there a longrun relationship between Stock Returns and Monetary Variables : Evidence from an Emerging Market." *Applied Financial Economics* 11, no. 6, 641-649.
- Lumsdaine, R.L., and S. Ng. (1999). "Testing for ARCH in the Presence of a Possibly Misspecified Conditional Mean." *Journal of Econometrics* 93, no. 2 : 253-279.
- Lütkepohl, H. (2004). "Vector Autoregressive and Vector Error Correction Models in H. Lütkepohl and M. Krätzig (eds.), *Applied Time Series Econometrics*, Cambridge University Press, Cambridge, U.K.
- Lütkepohl, H., and M. Krätzig (2004). (eds.), *Applied Time Series Econometrics*, Cambridge University Press, Cambridge, U.K.
- Lütkepohl, H. ; P. Saikkonen and C. Trenkler (2001). " Maximum Eigenvalue versus Trace Tests for the Cointegrating Rank of a VAR Process." *Econometrics Journal* 4: 287-310.
- Maddala, G.S., and I. Kim. (1998). *Unit Roots, Cointegration and Structural Change.* Cambridge: Cambridge University Press.

- Malkiel, B. (1992). "Efficient Market Hypothesis." In *New Palgrave Dictionary of Money and Finance*, ed. P. Newman, M. Milgate and J. Eatwell, London: Macmillan.
- Malkiel, B. (2003). "The Efficient Market Hypothesis and It's Critics." *Journal of Economic Perspectives* 17, no.1 : 59-82.
- Mankiw, N.G. and M.D. Shapiro. (1986). " Do we reject too often ?" *Economics Letters* 20 : 139-145.
- Mark , N.C. (1995). " Exchange Rates and Fundamentals: Evidence on Long-horizon Predictability." *American Economic Review* 85 : 201-218.
- McCracken, M.W. (2004). " Asymptotics for Out-of-Sample Tests of Granger Causality ." University of Missouri-Columbia, Manuscript.
- Mills, T.C. (1993). *The Econometric Modelling of Financial Time Series*, Cambridge University Press, Newyork.
- Mills, T.C. (1995). " Modelling Skewness and Kurtosis in London Stock Exchange FT-SE Index Return Distributions." *The Statistician* 44 : 323-332.
- Mookerjee, R. and Q. Yu. (1997). " Macroeconomic Variables and Stock Prices in a Small Open Economy : The Case of Singapore." *Pacific-Basin Finance Journal* 5 : 377-388.
- Musilek, P. (1997). " Changes in Macroeconomic Variables and the Stock Market Finance A Uver 47 : 150-162.
- Nasseh, A. and J. Strauss. (2000). " Stock Prices and Domestic and International Macroeconomic Activity : A Cointegration Approach, *Quarterly Review of Economics and Finance* 40 : 229-245.

- Nelson, D.B. (1991). “ Conditional Heteroskedasticity in Asset Returns : A New Approach.” *Econometrica* 59 : 347-370.
- Nelson, D.B., and C.Q. Cao. (1992). “Inequality Constraints in the Univariate GARCH Model.” *Journal of Business and Economic Statistics* 10 : 229-235.
- Nelson, C.R. and M.J. Kim. (1993). “ Predictable Stock Returns : The Role of Small Sample Bias, *Journal of Finance* 48 : 641-661.
- Newey, W.K. and D.G. Steigerwald. (1997). “ Asymtotic Bias for Quasi-maximum Likelihood Estimators in Conditional Heteroskedasticity Models*.” Econometrica* 65 : 587-599.
- Newey, W. and K.J West. (1987). “ A Simple Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* 55 : 703-708.
- Opong, K.K. ; G. Mulloholland. ; A.F. Fox., and K. Farahmand. (1999). “The Behaviour of Some UK Equity Indices : An Application of Hurst and BDS Tests.” *Journal of Empirical Finance* 6, no. 3 : 267-282.
- Osterwald-Lenum, M. (1992). “A Note with Fractiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics: Four cases.” *Oxford Bulletin of Economics and Statistics* 54 : 461-72.
- Pandey, V.; T. Kohers and G. Kohers. (1997). “ Nonlinear Determinism in the Equity Markets of Major Pacific Rim Countries and the United States.” Southwestern Finance Association Conference, New Orleans, LA.
- Paulsen, J. (1984). “Order Determination of Multivariate Autoregressive Time Series with Unit Roots.” *Journal of Time series analysis* 5 : 115-127.

- Peiró, A. (1994). " Daily Seasonality in Stock Returns : Further International Evidence." *Economics Letters* 45 : 227-232.
- Peiró, A. (1999). " Skewness in Financial Returns." *Journal of Banking and Finance*, 6 : 847-862.
- Pesaran, M.H. (2003). " Market Efficiency and Stock Market Predictability." Mimeo.
- Phillips, P.C.B., and P. Perron. (1988). "Testing for Unit Roots in Time Series Regression." *Biometrika* 75 : 335-346.
- Poshakwale, S. (1996). "Evidence on Weak form Efficiency and Day of the Week Effect in the Indian Stock Market." *Finance India* 10, no. 3: 605-616.
- Poshakwale, S. (2002) "The Random Walk Hypothesis in the Emerging Indian Stock Market." *Journal of Business Finance and Accounting* 29, nos. 9-10: 1275-1299.
- Priestley, M.B. (1980). "State-Dependent Models : A General Approach to Nonlinear Time Series Aanalysis." *Journal of Time Series Analysis* 1: 47-71.
- Quandt, R. (1960). "Tests of the Hypothesis that a Linear Regression Obeys Two Separate Regimes." *Journal of American Statistical Association* 55: 324-330.
- Rakshit, M. (2007). " Inflation in a Developing Economy : Theory and Policy." *ICRA Bulletin : Money and Finance*, September issue.
- Ramaswamy, K.V. (2001). " Economic Reforms, Industrial Structure and Performance : The Case of Consumer Durable Goods Industry in India." In *Economic Reform and Industrial Structure in India,* ed. S.Uchikawa, Tokyo : Institute of Developing Economies.

- Rapach, D.E.; M.E. Wohar and J. Rangvid. (2005). " Macro Variables and International Stock Return Predictability." *International Journal of Forecasting* 21 : 137-166.
- Richardson, M. and J.H. Stock. (1989). " Drawing Inferences from Statistics Based on Multiyear Asset Returns." *Journal of Financial Economics* 25 : 323-348.
- Robinson, P.M. (1991). "Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression." *Journal of Econometircs* 47, no. 1 : 67-84.
- Rogalski, R.J. (1984). "New Findings Regarding Day-of-the Week Returns Over Trading and Non-trading Periods : A Note." *Journal of Finance* 39, no. 5: 1603-1614.
- Ross, S. (1976). " The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory* 13 : 341-360.
- Ross, S. (1987). " The Interrelations of Finance and Economics : Theoretical Perspectives." *American Economic Review* 77: 29-34.
- Said, S.E., and D.A. Dickey.(1984). "Testing for Unit Roots in Autoregressive Moving-Average Models with Unknown Order." *Biometrika* 71 : 599-607.
- Saikkonen, P. and R. Luukkonen (1997). "Testing Cointegration in Infinite Order Vector Autoregressive Processes." *Journal of Econometrics* 81 : 93-129.
- Saikkonen, P. and H. Lütkepohl (2000a). "Testing for the Cointegrating Rank of a VAR Process with Structural Shifts." *Journal of Business & Economic Statistics* 18 : 451-464.

- Saikkonen, P. and H. Lütkepohl (2000b). "Trend Adjustment Prior to Testing for the Cointegrating Rank of a Vector Autoregressive Process." *Journal of Time Series Analysis* 21 : 435-456.
- Sakai, H., Tokumaru H. (1980). Autocorrelations of a Certain Chaos, *IEEE Transactions on Acoustics*, *Speech and Signal Processing* 1 : 588-590.
- Samuleson, P. (1965). "Proof that Properly Anticipated Prices Fluctuate Randomly." *Industrial Management Review* 6 : 41-49.
- Samuleson, P. (1972). " Mathematics of Speculative Price." In *Mathematical Topics in Economic Theory and Computataion*, ed. R. Day and S. Robinson, Society for Industrial and Applied Mathematics, Philadelphia, PA.
- Samuelson, P. (1973). " Proof that Properly Discounted Present Values of Assets Vibrate Randomly." *Bell Journal of Economics and Management Science* 4 : 369-374.
- Scheinkman, J., and B. LeBaron. (1989). "Nonlinear Dynamics and Stock Returns." *Journal of Business* 62 : 311-337.
- Schwarz, G. (1978). " Estimating the Dimesion of a Model." *Annals of Statistics* 6, 461-4.
- Schwert, G.W. (2001). " Anomalies and Market Efficiency." In *Handbook of the Economics of Finance* ed. G. Constantinides *et al*. Amsterdam: North Holland, Chapter 17.
- Sewell, S.P. ; S.R. Stansell. ; I. Lee., and M.S. Pan. (1993). "Nonlinearities in Emerging Foreign Capital Markets." *Journal of Business Finance and Accounting* 20, no.2: 237-248.

- Shiller, R.J., and A.E. Beltratti. (1992). "Stock Prices and Bond Yields: Can Their Comovements be Explained in Terms of Present Value Models?" *Journal of Monetary Economics* 30: 35-46.
- Shiller, R.J. (2003). " From Efficient Markets Theory to Behavioral Finance." *Journal of Economic Perspectives* 17: 83-104.
- Shleifer, A.(2000). *Inefficient Markets : An Introduction to Behavioral Finance*, Oxford University Press, New York.
- Silvapulle, P., and M. Evans. (1993). "Testing for Serial Correlation in the Presence of Conditional Heteroskedasticity." *Discussion Paper, La Trobe University, Australia*.
- Solnick, B. (1984). " The Relationship between Stock Prices and Inflationary Expectations : The International Evidence." *Journal of Finance* 38 : 35-48.
- Stambaugh, R.F. (1986). " Biases in Regressions with Lagged Stochastic Regressors." *Graduate School of Business, University of Chicago, Working Paper No. 156.*
- Stambaugh, R.F. (1999). " Predictive Regressions." *Journal of Financial Economics* 54 : 375-421.
- Swamy, S. (2005). " Political Structure and Economic Reforms." *Economic and Political weekly* 40.
- Taylor, J.W. (2004). " Volatility Forecasting with Smooth Transition Exponential Smoothing." *International Journal of Forecasting* 20 : 273-286.
- Tong, H. (1990). *Nonlinear Time Series*. Oxford: Oxford Science Publications.

- Tucker, A.L. (1992). " A Reexamination of Finite and Infinite-Variance Distributions as Models of Daily Stock Returns." *Journal of Business and Economic Statistics*, 10, 73-81.
- Wasserfallen, W. (1989). " Macroeconomic News and the Stock Market : Evidence for Europe" *Journal of Banking and Finance* 13 : 613-626.
- Weiss, A.A. (1986). "Asymptotic Theory for ARCH Models : Estimation and Testing." *Econometric Theory* 2 : 107-131.
- White, H. (1982). " Maximum Likelihood Estimation of Misspecified Models." *Econometrica* 50 : 1-25.
- Willey, T. (1992). "Testing for Nonlinear Dependence in Daily Stock Indices." *Journal of Economics and Business* 44, no. 1.: 63-76.
- Wongbangpo P. and S. Sharma. (2002). " Stock Market and Macroeconomic Fundamental Dynamic Interactions : ASEAN-5 Countries." *Journal of Asian Economics* 13 : 27-51.
- Wood, D. and S. Poshakwale (1997). " Market Anomalies in Selected Emerging and Developed Stock Markets." *Asia-Pacific Journal of Accounting* 4 : 73-92.
- Wooldridge, J.M. (1991a). "On the Application of Robust, Regression-based Diagnostics to Models of Conditional Means and Conditional Variances." *Journal of Econometrics* 47 : 5-46.
- Wooldridge, J.M. (1991b). "Specification Testing and Quasi-maximum Likelihood Estimation." *Journal of Econometrics* 48: 29-55.
- Zakoian, J.M. (1994). " Threshold Heteroskedastic Functions." *Journal of Economic Dynamics and Control* 18 : 931-955.