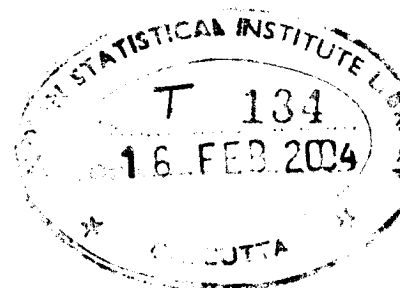


SOME ISSUES ON TIME VARYING RISK PREMIUM IN ARCH-M MODEL

SAMARJIT DAS

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
Calcutta, June 2002



Dedicated to Baba & Ma

ACKNOWLEDGEMENT

I express my deepest sense of gratitude to Professor Nityananda Sarkar, my thesis supervisor, for his constant guidance throughout the entire period of this work. Apart from my indebtedness for having his natural guidance, I am indeed fortunate to work with such a spirited and different personality.

I should not forget to admit my indebtedness to all of my teachers, especially to Mr. Atul Krishna Maity, Mr. Anil Kumar Bera, Professor M. K. Gupta, Professor R.N. Panda, Dr.S.S.Maity and Dr.P.Das, for providing enthusiastic guidance at every important footstep of my career.

I am particularly thankful to my present employer, Reserve Bank of India, for providing me with various resources and also the opportunity to work with personalities like Dr.R.B. Barman, Dr.D.Ray, Dr Balwant Singh, Dr. Ramasastry, who have given me priceless guidance and advices at various levels.

I would also like to express my gratitude to the Indian Statistical Institute, Kolkata, for providing me with all facilities during the entire period of my work.

I have had the privilege and pleasure to work with Dr. Snigdhanu Chatterjee, whose intellectual inputs I would like to acknowledge.

I would also like to take the opportunity to acknowledge the various kinds of help and co-operation I have received from my priceless friends and colleagues: Arijit, Diganta, Rajlaxmi, Suchi di, Brati, Atanuda, Samanta da, Amit da, Sharmistha di, Arnab, Madhuchanda, Indrajit, Sneharthi, Biswabrata, Rajashree-to name just a few. I would also like to put on record my appreciation for my friend Arijit Bishnu who helped me a lot in the computational work.

It is needless to acknowledge the most needed support and encouragement

received from my parents & parents-in-law. Their faith in my abilities and their encouragement have been instrumental throughout this period. I should not forget to acknowledge the courage I received from Kaku and Kakima. From the bottom of my heart, I would like to thank Dada, Mejda and Nantu. Without their sustenance provided to me in the form of allowing me to remain free of any family responsibilities, this stupendous task would have probably taken a much longer time to complete.

Last but not the least, I acknowledge the contribution of Moonmoon, first as my friend and then as my wife, towards the successful completion of this thesis. I remain indebted to her for all her love, encouragement, support and sacrifice. She has been my greatest source of inspiration and without her help this work would not have seen the light of the day.

Calcutta, June, 2002

Samarjit Das

Contents

1	Introduction	7
1.1	Background and a Brief Review on ARCH-M Model	7
1.2	Motivation and Format of the Thesis	13
2	The Anatomy of a Stock Price Index: Testing for Stationarity, Independence, Normality, Long-Memory and Existence of Moments	21
2.1	Introduction	21
2.2	SENSEX Data	23
2.3	Standard Tests	26
2.3.1	Tests for Stationarity	26
2.3.2	Testing for Independence	30
2.3.3	Testing for Normality	36
2.3.4	Testing for Long-Memory	41
2.3.5	On Existence of Moments	44
2.4	Conclusion	51

3 Estimation and Testing of ARCH-M Parameters: Estimating Equations Approach and Bootstrap Technique 53

3.1	Introduction	53
3.2	A Brief Overview of Bootstrap Methodology	57
3.3	Asymptotic Properties Based on Estimating Equations Approach	65
3.4	Derivation of Bootstrap Distribution	69
3.5	Simulation Results	71
3.5.1	Results Based on Asymptotic Approximation	71
3.5.2	Results Based on Application of Bootstrap Methodology	73
3.6	An Empirical Illustration	76
3.7	Conclusions	81

4 An ARCH in the Nonlinear Mean (ARCH-NM) Model 91

4.1	Introduction	91
4.2	The Proposed Model	93
4.3	Estimation	96
4.4	An Illustration	99
4.5	Conclusions	105

5 Testing for Constancy of the Relative Risk Aversion Parameter and Estimating the Time Varying Parameter ARCH-M Model 114

5.1	Introduction	114
-----	------------------------	-----

5.2	Test for Constancy of Relative Risk Aversion Parameter	116
5.2.1	Score Test for Parameter Constancy	117
5.2.2	Test for Constancy of the Relative Risk Aversion Pa- rameter	120
5.3	Monte Carlo Results	122
5.4	Bootstrapping the Test for Size Correction	127
5.5	ARCH-M with Time Varying Risk Aversion Parameter	129
5.6	An Application to SENSEX Data	132
5.6.1	Results of the Test	132
5.6.2	Estimated Time Varying Parameter ARCH-M Model	133
5.7	Conclusions	137
6	Conclusions and Future Outlook	139
6.1	Major Findings	140
6.2	Direction for Further Research	144

Chapter 1

Introduction

1.1 Background and a Brief Review on ARCH-M Model

Since the 1970's it has been observed in many economies that financial and macroeconomic variables like equity prices, treasury bill rates and exchange rates have become more and more volatile in nature. This may be due to more flexible monetary policies pursued in these countries as well as due to their increasing exposure towards various international developments. Accordingly, economic agents are facing increasingly more and more risky environment. Researchers as well as professional economists in the area of capital and business finance have, therefore, been increasingly attracted in recent years towards studying the effect of risk and uncertainty on asset returns, and framing rational decision rules for individuals and institutions for the purpose of selecting security portfolios. The increased importance played by risk and uncertainty considerations in modern economics and finance theory has necessitated the development of new econometric time series modelling techniques that allow the higher-order moments to be time dependent as opposed to the traditional macroeconomic and financial time series modelling approach which mainly

centres on the conditional first moment.

While it has been recognized for quite some time that the uncertainty in speculative prices, as measured by variances and covariances, changes through time (see, for instance, Mandelbrot, 1963; and Fama, 1965), it was not until the introduction of what is now known as *Modern Financial Econometrics* that applied researchers in financial and monetary economics have started explicitly modelling time variation in second or higher-order moments. In a seminal paper in 1982, Engle introduced the autoregressive conditional heteroscedastic (ARCH) model. This model allows the conditional variance to change over time as a function of past errors keeping the unconditional variance constant. It has been observed that such models capture many empirically observed temporal behaviours like thick tail distribution and volatility clustering of many economic and financial variables (see, Bera and Higgins, 1993; Bollerslev, Chou and Kroner, 1992; Bollerslev, Engle and Nelson, 1994; Shephard, 1996; and Gouriéroux, 1997 for excellent surveys on ARCH model and its various generalizations).

The basic ARCH model has been generalized in different directions so as to broaden its applicability. One important generalization of ARCH model is what is known as ARCH in the mean (ARCH-M) model which was first introduced by Engle, Lilien and Robins (1987). This generalization has, in fact, provided a methodology to estimate and to test the conditional version of capital asset pricing model (CAPM). The origin of ARCH-M model may be traced back to the unconditional CAPM of Lintner-Sharpe-Mossin, which is considered to be the most widely used theoretical model for specifying the relation between risk and return. In the early applications of mean-variance capital asset pricing of Sharpe (1964), Lintner (1965) and Mossin (1966), the expected return and its covariance with the market return were assumed to be constant. Over the last two decades, this relationship has been modified

and reoriented in a different manner and used in a wide range of financial applications. The basic risk-return relationship, called the capital asset pricing model of Sharpe (1964), Lintner (1965) and Mossin (1966), provides that for an asset i the relationship is expressed as

$$E(y_i) = b Cov(y_i, r_m), \quad (1.1.1)$$

where y_i is the excess return (in excess of the risk free rate, say treasury bill rate) on asset i , r_m is the excess return on the market portfolio, and b is the factor of proportionality called the market 'beta'. In (1.1.1), expected return of asset i , its covariance with the market and b were assumed, at least implicitly, to be constant. Studies concerning tests of the unconditional version of the CAPM, such as, those by Fama and MacBeth (1973), Gibbons (1982) and Stambaugh (1982) also made such assumptions in their empirical analysis. Initially, most of the researchers had examined the trade off between risk and return among different securities within a given time period. However, in the intertemporal or in the conditional version of the CAPM return/risk or both may vary over time. Gibbons & Ferson (1985) and Ferson, Kendel & Stambaugh (1987) have tested the CAPM at the conditional level, and allowed expected return to vary over time but still keeping the conditional covariances as constant. For a long time it has been observed that risks on returns of different stocks change over time, which, in turn, implies time variation in risk premium. Schwert (1989) has, for instance, found that simple estimates of standard deviation on monthly stock returns vary substantially. Many researchers have studied movements in aggregate stock market volatility. Officer (1973) has related these changes to the volatility of macroeconomic variables. Thereafter, many attempts have been made to relate changes in stock market volatility to changes in expected returns. Notable studies include Fama & Schwert (1977), Pindyck (1984), Poterba & Summers (1988), French, Schwert

& Stambaugh (1988), Bollerslev, Engle & Wooldridge (1988), Abel (1988), Harvey (1989), Nelson (1990), Campbell & Hentschel (1992), Chan, Karolyi & Stulz (1992), Glosten, Jagannathan & Runkle (1993) and Hansson & Hordahl (1997). This phenomenon viz., existence of time varying risk premium has been observed and analysed not only in stock markets but also in foreign exchange markets and the term structure of interest rates (see for instance, Amsler, 1984; Domowitz and Hakkio, 1985; Merton, 1986; and Pesando, 1983)

It may be noted that the finance literature does not provide an appropriate guideline for the determination of the changes in the risk or in the risk premium. It was Engle (1982) who, for the first time, introduced a methodology of modelling time-varying variance which is not assumed to depend on any macroeconomic variables. In essence, this model postulates that the variance in the current period is likely to be large following a large error in the previous period. After a period of five years ARCH-M model was introduced by Engle *et al.* (1987), which provided a new approach by which we could test for and estimate a time varying risk premium. This generalization of ARCH model has indeed revolutionized the time varying risk premium literature.

Considering a world of two assets—one is risky and the other riskless—Engle *et al.* showed that the mean and variance of the returns of risky assets move in the same direction. They also showed that the conditional variance directly affects the expected return on a portfolio. In order to incorporate such aspects in the usual framework of ARCH model, they suggested a new approach in which the ARCH model is extended in a direction so that it allows the conditional variance to influence the mean return. They also found that " variables which apparently were useful in forecasting excess return are correlated with the risk premia and lose their significance when a function of the conditional variance is included as a regressors" (p.392). They specified the ARCH-M model as follows:

$$y_t = x_t' \beta + \lambda \sqrt{h_t} + \epsilon_t, \quad t = 1, 2, \dots, n \quad (1.1.2)$$

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \dots + \alpha_p \epsilon_{t-p}^2, \quad (1.1.3)$$

where $\epsilon_t | \Psi_{t-1} \sim N(0, h_t)$, and $\alpha_0 > 0, \alpha_i \geq 0, i = 1, \dots, p, \sum_{i=1}^p \alpha_i < 1$, y_t is the dependent variable, x_t is the $k \times 1$ vector of independent (exogenous) variables which may also include lagged values of the dependent variable, $\Psi_{t-1} = \{y_{t-1}, x_{t-1}, y_{t-2}, x_{t-2}, \dots\}$ is the information set at time $t-1$, β is the $k \times 1$ vector of regression parameters and ϵ_t is the error term associated with the regression.

To examine the properties of ARCH-M model, let us consider a simple version of (1.1.2) and (1.1.3), *viz.*,

$$y_t = \lambda h_t + \epsilon_t,$$

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2.$$

We have, combining these two,

$$y_t = \lambda \alpha_0 + \lambda \alpha_1 \epsilon_{t-1}^2 + \epsilon_t,$$

Hence, it immediately follows that

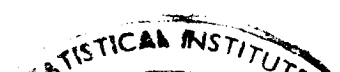
$$E(y_t) = \frac{\lambda \alpha_0}{1 - \alpha_1},$$

and this can be viewed in finance models as the unconditional expected return for holding a risky asset. It is also easy to find that

$$V(y_t) = \frac{\alpha_0}{1 - \alpha_1} + \frac{2(\lambda \alpha_0 \alpha_1)^2}{(1 - \alpha_1)^2 (1 - 3\alpha_1^2)}.$$

Since $V(y_t) = \frac{\alpha_0}{1 - \alpha_1}$ if risk premium is absent, the second component of $V(y_t)$ is due to the presence of risk premium. Further, as shown by Hong (1991),

$$\rho_1 = \text{Corr}(y_t, y_{t-1}) = \frac{2\alpha_1^3 \lambda^2 \alpha_0}{2\alpha_1^2 \lambda^2 \alpha_0 + (1 - \alpha_1)(1 - 3\alpha_1^2)},$$



$$\rho_j = \text{Corr}(y_t, y_{t-j}) = \alpha_1^{j-1} \rho_{j-1}, j = 2, 3, \dots,$$

and thus we find that the ARCH-M effect makes y_t serially correlated. Similar results for GARCH model were obtained by Bollerslev (1988).

If the risk premium term in (1.1.2) is denoted as $\lambda g(h_t)$, then it may be noted that in most applications $g(h_t) = \sqrt{h_t}$ has been used (see, for instance, Bollerslev, Engle and Wooldridge, 1988; Domowitz and Hakkio, 1985) although Engle *et al.* found that $g(h_t) = \ln h_t$ worked better in their estimation of the time varying risk premia in the term structure. However, Pagan and Hong (1991) commented that the use of $\ln h_t$ is somewhat problematic since for $h_t < 1$, $\ln h_t$ will be negative and also when $h_t \rightarrow 0$, the effect on y_t will be infinite.

Insofar as estimation of ARCH-M model is concerned, the method of maximum likelihood under the assumption of conditional normality is used. Engle *et al.* concluded that the likelihood being of the form analyzed by Crowder (1976), the ML estimator $\hat{\theta}$ of the ARCH-M parameter vector $\theta = (\beta', \alpha')'$ has, under sufficient regularity conditions, the property that $(S'S)^{1/2}(\hat{\theta} - \theta_0) \sim$ asymptotically $N(0, I)$, where θ_0 is the true value of θ and S is the matrix of first-order derivatives of the log-likelihood function. Most of the applied work on ARCH-M model have used the Berndt, Hall, Hall and Hausman (1974) algorithm (BHHH) to maximize the log-likelihood function.

Although the asymptotic properties of the estimators of the parameters of ARCH/GARCH model have been studied, there have been hardly any studies concerning (G)ARCH-M model, mainly because of the difficulties posed by the fact that the conditional variance of an GARCH-M model is a nonlinear difference equation. Lee (1991) extended all such asymptotic properties to the GARCH(1,1)-M and IGARCH(1,1)-M models. Further, unlike ARCH model, the information matrix for ARCH-M model is not block diagonal between the

conditional mean and the conditional variance parameters.

There have been numerous applications of ARCH-M model to stock market data, particularly to those of the US and the UK. Some of the important studies are due to Baillie and De Gennaro (1990), Bollerslev, Engle and Wooldridge (1988), Engle, Ng and Rothschild (1990), Hall, Miles and Taylor (1988), Nelson (1990), Black and Fraser (1995), Fraser (1996), Hansson and Hordahl (1997), Elyasiani and Mansur (1999), and Ellul (1999). All of them have used ARCH-M or close relatives of ARCH-M to provide risk premia as functions of conditional variance or covariances of asset returns. All these studies have led to a better understanding of the usefulness of ARCH-M model. Thus, we may conclude by stating that since the introduction of ARCH-M model by Engle *et al.* in 1987, it has become one of the workhorses of econometric research on time varying risk premium.

1.2 Motivation and Format of the Thesis

Given the paramount importance of ARCH-M model in the time varying risk premium literature, there have been not only numerous applications of this model in financial economics but also various theoretical advancements concerning this model. Yet, there appear to be scope for further research into some theoretical aspects of ARCH-M modelling. The broad aspects which this thesis proposes to investigate are the following.

Asymptotic properties of ARCH-M model based on estimating equations approach

The first relates to studying the asymptotic properties of the estimators of the parameters of ARCH-M model. Although the asymptotic properties of the estimators of the parameters of ARCH/ GARCH model have been studied using likelihood/ quasi-likelihood approach, there has been hardly

any such studies concerning ARCH-M/GARCH-M model. This may be due to the fact that unlike the former, the conditional variance of the latter class of models involve nonlinear difference equations. Moreover, the requirement of existence of higher-order moments and the conditions implied thereof are sometimes difficult to justify in return data. It would, therefore, be useful to study the asymptotic properties of the estimators of the parameters of ARCH-M model using a more generalized approach and under more amenable moment conditions. It may be noted that estimation and testing based on (Gaussian) likelihood is sometimes non-robust since the procedure chosen suits only the likelihood of choice. A generalized approach like estimating equations approach, as originally proposed by Godambe (1985), has the advantage that standard likelihood and quasi likelihood based approaches are special cases of estimating equations technique.

Bootstrap technique for ARCH-M model

A related issue of interest and importance is the application of bootstrap methodology for studying ARCH-M model. Although resampling procedure like bootstrap is a widely used technique not only in statistics but also in econometrics, in general, there has been no major work, to the best of our knowledge, on ARCH/ ARCH-M model where bootstrap technique has been used and its statistical properties studied. One possible reason could be the fact that in ARCH/ARCH-M framework, there is conditional heteroscedasticity which, nevertheless, is unconditionally homoscedastic. Thus, it is not at all obvious whether the different types of bootstrap would provide consistent results or not. We propose to use appropriate bootstrap technique for ARCH-M model, and then study some statistical properties of the bootstrap estimator.

Another aspect relates to risk premium which represents the change in the expected rate of return due to a change in risk. As already stated and,

in fact, very well-known also, the most important consideration for ARCH-M model lies in the fact that it explicitly incorporates in its specification the time varying nature of risk premium. It is, therefore, of paramount importance to specify this term as appropriately as possible. This could be done at least in two ways: (i) by considering a very generalized functional form for conditional heteroscedasticity, and (ii) by assuming the relative risk aversion parameter itself to be time dependent.

Generalization of risk premium by considering Box-Cox transformation of h_t

In most applications the functional form of conditional heteroscedasticity (say, $g(h_t)$) has been taken as $g(h_t) = \sqrt{h_t}$. Engle *et al.* (1987) in the original ARCH-M paper found that $g(h_t) = \ln h_t$ worked better in estimating time varying risk premium in term structure. But, Pagan and Hong (1991) have argued that this choice is somewhat restrictive because of the fact that $\ln h_t$ is negative for $h_t < 1$. Studies (see, for instance, Backus and Gregory, 1993; Glosten, Jagannathan and Runkle, 1993; and Harvey, 1989, 1991) have also shown that although theory may assume a monotonic relationship between risk premium and conditional variance, empirical evidences do not always support this. Given these findings, we propose a general form of $g(h_t)$ in the form of Box-Cox (1964) transformation of h_t . We study the properties of the parameters of the proposed model, and then describe an appropriate estimation procedure for the model.

Testing constancy of the relative risk aversion parameter

The economic theory behind ARCH-M model is that economic agents are constant relative risk averse; the assumption seems to be fairly strong and may not be true in many practical situations. If a representative agent maximizes a time additive von Neumann-Morgenstern utility, the mean/variance ratio or the relative risk aversion parameter λ may change due to either change in the

perception towards risk or change in the distribution of wealth or both. This is likely to be more so in an emerging economy like the Indian economy where over the last two decades the stock market has evolved in a spectacular way, and consequently the Indian stock market has undergone some basic changes including possibly changes in the value of λ . Further, during the course of time, the changes in consumption and investment pattern of economic agents might also influence this parameter, and this might lead to its changes over time.

Some empirical studies have found that this parameter need not always be stable. In fact, they have found it to be very unstable across the sample period. For example, based on New York Stock Exchange monthly value-weighted index, French, Schwert and Stambaugh (1987) obtained the estimates of λ to be 1.693, 1.510 and 7.220 for the periods 1928-1984, 1928-1952 and 1953-1984, respectively. Two other notable studies along this line are due to Pindyck (1988) and Chou *et al.* (1992). These findings, therefore, suggest that it should be possible to generalize ARCH-M model with an appropriately defined time varying risk aversion parameter. And, of course, it is important to know in an actual data analysis whether λ is time invariant or not. Towards this end, we propose a test procedure which is essentially based on Chesher's (1983) interpretation of White's (1982) information matrix (IM) test being a test of parameter variation. The size and power performance of this test is then studied with a detailed Monte Carlo study. The test suffers from over-size problem, as would be discussed in the thesis. Hence we suggest bootstrapping the test to correct the size distortion problem.

Estimation of time varying parameter ARCH-M model

In case the proposed test suggests rejection of the null hypothesis of time invariant λ , then it should only be appropriate that a model under time variant λ is proposed. Following Chou, Engle and Kane (1992) we suggest such a

model where the relative risk aversion parameter is assumed to follow a random walk process in the framework of a state-space model. We also obtain rolling sample estimate for the time varying risk aversion parameter, and then compare those with the estimate obtained from the time varying ARCH-M model.

As indicated in the preceding paragraphs, apart from studying the theoretical issues involved and deriving appropriate results, we carry out simulation studies wherever appropriate and so possible, to study the small-sample performance of the estimators of the parameters of the suggested models (which are indeed extensions/ generalizations of the original ARCH-M model) as well as of the proposed tests of hypotheses of relevant parameters.

Application to Indian stock price (SENSEX) data: a closer look

Finally, the usefulness of the proposed models as well as of the tests of hypotheses involving the relative risk aversion parameter λ is now studied with an actual data set. The data set used for this purpose is the return data which is based on the most-widely used Indian stock price index, called the sensitive index (SENSEX). The Bombay Stock Exchange is the premier stock exchange in India, and it alone accounts for over 80 per cent of the total volume of transactions in share in the country. SENSEX is the most widely-used stock price index compiled regularly by the Bombay Stock Exchange comprising 30 most sensitive and well-traded securities. The data is based on the weekly average (closing) price, and cover the period 1st week of January 1984 to 3rd week of October 2000, consisting of a total of 872 observations. Obviously, the resulting empirical findings towards the end of each chapter would throw light on the behaviour of the Indian stock market as represented by SENSEX. Now, it is only meaningful that before carrying out the required empirical exercises, a closer look is taken at the data generating process by studying the standard and not-so-standard properties of these returns based on SENSEX.

Towards this end, we have brought together the relevant literature on tests for stationarity, independence, normality, long-memory, tail behaviour of the underlying distribution, and existence of higher-order moments of univariate time series data, and then applied these tests to our return data set.

The thesis has been organized (chapter-wise) as follows:

Chapter 1: Introduction

The first section of the first chapter attempts to present the background of the thesis along with a brief review of the existing literature on ARCH-M model, especially those in the context of applications to stock markets. Thereafter the motivation of this work is discussed, and this is followed with the chapter-wise format of the thesis.

Chapter 2: The Anatomy of a Stock Price Index: Testing for Stationarity, Independence, Normality, Long-Memory and Existence of Moments

Chapter 2 begins with an introduction. The nature of SENSEX data is described in Section 2.2. Standard tests for stationarity, independence, normality, long-memory, tail behaviour of the underlying distribution, and existence of higher-order moments of univariate time series are briefly described in Section 2.3 along with the findings of their applications to SENSEX data. This chapter closes with some comments in Section 2.4.

Chapter 3: Estimation and Testing of ARCH-M Parameters: Estimating Equations Approach and Bootstrap Technique

Section 3.1 gives the background of the content of this chapter. The basics of bootstrap technique are described in the next section. Section 3.3 proves

the asymptotic properties of the estimators of the parameters of ARCH-M model based on estimating equation approach. Consistency of the bootstrap distribution has been derived in the next section. Monte Carlo and bootstrap based results are discussed in Section 3.5. Findings of the empirical illustration with SENSEX data are given in Section 3.6. The chapter closes with some concluding remarks in Section 3.7.

Chapter 4: An ARCH in the Nonlinear Mean (ARCH-NM) Model

Beginning with an introduction in Section 4.1, the proposed ARCH-MN model is described in the next section along with its properties. Section 4.3 describes the procedure of estimation of the model. The advantages and appropriateness of this generalization of ARCH-M model is illustrated through a application to SENSEX data and the findings are presented in Section 4.4. The chapter closes with some concluding remarks in Section 4.5.

Chapter 5: Testing for Constancy of the Relative Risk Aversion Parameter and Estimating the Time Varying Parameter ARCH-M Model

"Introduction" presents the background of the work contained in this chapter. In Section 5.2, a test for constancy of relative risk aversion parameter λ is proposed. The Monte Carlo results concerning this test are described in the next section. Section 5.4 discusses how bootstrap technique may be used to correct the size problem of the proposed test. Generalization of ARCH-M model where λ is time varying is described in Section 5.5, and the results of application of this model to SENSEX data are presented and discussed in Section 5.6. We conclude this chapter with some remarks in the last section.

Chapter 6: Conclusions and Future Outlook

This is the concluding chapter. The major results and findings of this work are summarized in Section 6.1. Possible limitations of this work are also mentioned in this section. Some avenues for further research in this area are discussed in the next (and also concluding) section.

Chapter 2

The Anatomy of a Stock Price Index: Testing for Stationarity, Independence, Normality, Long-Memory and Existence of Moments

2.1 Introduction

Any stock price index is considered as a representative of the stock market behaviour as a whole, and also thought to be an indicator of the state of the corresponding economy. Movements of such indices are closely observed by market participants, financial analysts, economists and others, and hence appropriate modelling of these series are extremely important. In this context it is well-known that due to volatile behaviour of the stock prices, it is not easy to fit appropriate empirical distributions to such data sets. However, financial models, to start with, are often based on explicit or implicit assumptions of normality along with assumptions like stationarity, existence of higher order moments (including variance) and different kinds of dependencies. For instance, while conventional notion of the efficient market hypothesis presumes that speculative prices can be modelled as random walk processes with no linear dependence in the price changes, the approach since the days of nonlin-

ear dynamics envisages that lack of linear (Granger and Anderson, 1978; Sakai and Tokumaur, 1980; and Scheikman & LeBaron, 1989), dependence does not rule out nonlinear dependence which, if present, would contradict the efficient market hypothesis and may aid in forecasting, especially over short time intervals. Similarly, the theoretical asset pricing models such as the Mean-Variance Theorem of Markowitz (1959), the Capital Asset pricing model (CAPM) of Sharpe (1964), Lintner (1965), the consumption asset pricing model of Lucas (1978) and the Black-Scholes options pricing model of Black and Scholes (1973) are based on assumptions such as normality, uncorrelatedness and the stationarity. All these assumptions, commonly made, are at best approximations to reality. The optimal properties of the estimators and the test statistics under such ideal situations may not be entirely valid for the simple reason that observed data may often violate such assumptions. It is, therefore, useful as well as important to study the extent to which such assumptions are found to hold good for such data. This is more so for this work since all the empirical illustrations herein have been carried out, as stated in Chapter 1, using return data based on Bombay Stock Exchange Sensitive Index (SENSEX).

In this chapter we take a careful look at the properties as well as the data generating process of SENSEX. Towards this end, we present an overview of the different methodologies and test procedures which are available in the literature to study the properties of such time series data, and also report the empirical findings obtained by applying those to returns based on SENSEX data. In this context it is worthwhile to mention that in similar empirical studies, while the usual properties like stationarity, volatility, normality etc. are studied, not much importance is given to model the tail parts of these distributions, i.e., to study the amount of probability mass in the tails of these distributions. But the latter is obviously important from consideration of evaluating extreme risks in the financial markets. The commonly observed

'fat-tailness' property of the return has important implications in risk management, in particular to value-at-risk analysis. Since the normal distribution is often used to fit the empirical distribution of return, any significant discrepancy between the hypothesized normal tails and the actual fat tails may lead to significant errors; the traditional normal based analysis underestimates the frequency of the extreme observations i.e., the observations which lie in the tails of the distribution.

The chapter is organised as follows. The SENSEX data are described in Section 2.2. Standard tests for studying the properties of the time series are reviewed in the next section along with the results of their applications to SENSEX data. Finally, the chapter ends with some concluding remarks in Section 2.4.

2.2 SENSEX Data

The Bombay stock exchange is the premier stock exchange in India, which alone accounts for over 80 per cent of the total volume of transactions in share in the country. SENSEX is the oldest and most widely used stock price index (base, 1978-79=100) compiled regularly by the Bombay Stock Exchange comprising 30 most sensitive and well-traded, liquid and industry representative securities. The compilation of the values is based on the 'weighted aggregates' methods. In this method, the price of a component share in the index is weighted by the number of equity shares outstanding so that each scrip will influence the index in proportion to its respective market importance. The current market value for any particular scrip is obtained by multiplying the price of the scrip by the number of equity shares outstanding. The index on a day is calculated as the percentage of the aggregate market value of the equity shares of all the companies in the sample on that day to the average market

value of the same companies during the base period. The methodology applied in computing SENSEX is the same as the one employed in many of the popular indices such as the Standard & Poor USA, Dow Jones Index, TOPIX, Hang Sang Index, NYSE Composite Index and FT-SE 100 Index. It may be noted that there are other stock indices such as NSE, BSE-100, BSE-200, DOLLEX, to name a few, representing the Indian stock market, but these are relatively new indices.

The study is based on the weekly average closing price of SENSEX. The data cover the period January (1st week), 1984 to October (3rd week), 2000, corresponding in all 872 observations. The analysed series is the first differences of the logarithms of SENSEX values i.e., return at the t -th week $y_t = (\log(p_t) - \log(p_{t-1})) \times 100$, where p_t is the SENSEX value (closing) at the t -th week. Hence, the data represent the continuously compounded rate of return for holding the (aggregate) securities for one week.

We now report the usual statistical description of the return data *viz.*, mean, variance, skewness (squared) and kurtosis coefficient (without adjustment of 3 which is the kurtosis value of normal distribution), which were obtained from 872 SENSEX values. These were found to be 0.31, 27.95, 0.68 and 129.96, respectively. It is evident from the last two figures, that while the kurtosis coefficient is significantly greater than 3, the skewness coefficient being 0.68 conveys some evidence, although not very strong, of asymmetry in the unconditional distribution. The kurtosis coefficient being significantly greater than 3, indicates that the unconditional distribution of the data has heavier tail than the normal distribution. We also observe from the plot of these returns (see Figure 2.2.1) against time that large and small changes (of either sign) in returns tend to be clustered together over time. This phenomenon, which is well known for return series, is known as second order temporal dependence or more commonly, volatility clustering.

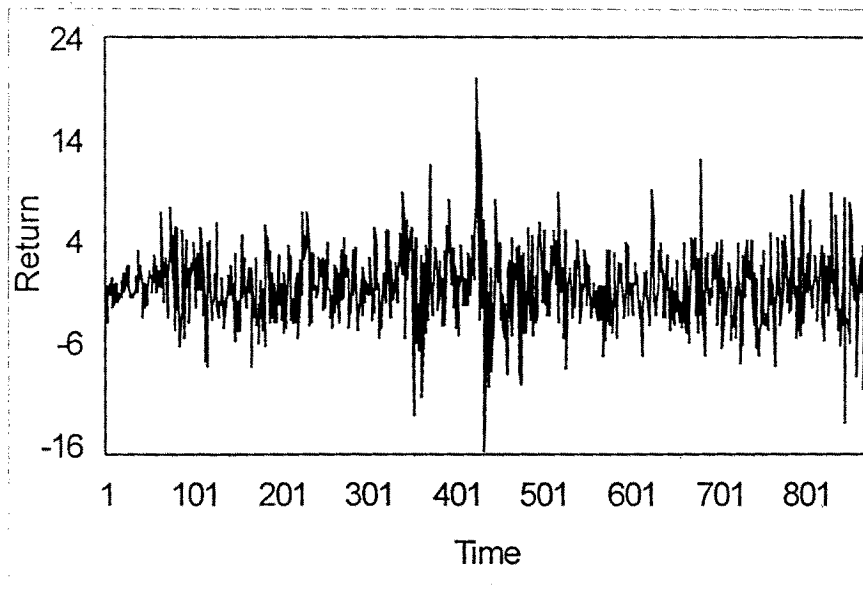


Figure 2.2.1: Plot of the Log-Return Series

2.3 Standard Tests

In the context of modelling financial time series usually tests for stationarity, non-autocorrelation/ independence, normality, long-memory and existence of higher -order moments of the series are carried out. While there are some traditional tests for most of these testing excercises, some recent tests are also available in the literature. In this section, we review some of these tests briefly, and then discuss the results of application of the same to returns on SENSEX.

2.3.1 Tests for Stationarity

Stationarity is a very important property with most of the time series. If a time series process is not stationary then appropriate procedures are used to reduce it to a stationary one. While Box and Jenkins (1976) advocated the use of differencing to achieve stationarity, in the light of unit root revolution and the vast amount of research work done since 1980s, it is extremely important to know beforehand as to whether the non-stationarity is of the nature of trend stationary proces (TSP) or difference stationary process (DSP) so that appropriate de-trending method could be used (*cf.* Nelson and Plosser (1982)). It is also relevent to note that spurious autocorrelation would result, as observed by Nelson and Kang (1981) , whenever a time series generated by DSP is detrended by regression on time or a TSP is detrended by differencing.

As for tests for stationarity, more commonly known as unit root test, the first rigorous test was suggested by Dickey and Fuller (1979, 1981). However, primary because of the problem of serial correlation in the noise of this test, the " augmented" Dickey and Fuller (ADF) of Said and Dickey (1984) is most widely used. This test involves estimating the following equation

$$y_t = \rho y_{t-1} + \sum_{i=1}^p \delta_i \Delta y_{t-i} + \varepsilon_t, \quad (2.3.1)$$

or , equivalently,

$$\Delta y_t = (\rho - 1)y_{t-1} + \sum_{i=1}^p \delta_i \Delta y_{t-i} + \varepsilon_t, \quad (2.3.2)$$

and then testing for the significance of $(\rho - 1)$. While the assumption of a driftless random walk under the null of DF/ADF test may be appropriate for some time series, many often contain a drift parameter. In such cases the estimating equation would be

$$y_t = \alpha_\mu + \rho_\mu y_{t-1} + \sum_{i=1}^p \delta_i \Delta y_{t-i} + \varepsilon_t, \quad (2.3.3)$$

or , equivalently,

$$\Delta y_t = \alpha_\mu + (\rho_\mu - 1)y_{t-1} + \sum_{i=1}^p \delta_i \Delta y_{t-i} + \varepsilon_t, \quad (2.3.4)$$

Now, Perron (1989) pointed out that the unit root test statistic is not capable of distinguishing between the null hypothesis of a random walk with drift and a plausible alternative of y_t being generated by a linear trend buried in stationary noise i.e., y_t being a TSP. An extended testing procedure suggested involves estimating the following regression

$$y_t = \alpha_\tau + \beta t + \rho_\tau y_{t-1} + \sum_{i=1}^p \delta_i \Delta y_{t-i} + \varepsilon_t, \quad (2.3.5)$$

or , equivalently,

$$\Delta y_t = \alpha_\tau + \beta t + (\rho_\tau - 1)y_{t-1} + \sum_{i=1}^p \delta_i \Delta y_{t-i} + \varepsilon_t, \quad (2.3.6)$$

and then testing for the significance of $(\rho_\tau - 1)$. Different F-type tests e.g., $\tilde{\phi}_1$, $\tilde{\phi}_2$ and $\tilde{\phi}_3$, are also considered to test the joint significance of the parameters of the above equations.

While the null hypothesis under $\tilde{\phi}_3$, is $H_0 : \beta = 0, \rho_\tau = 1$, the same for $\tilde{\phi}_2$ is $H_0 : \alpha_\tau = 0, \beta = 0, \rho_\tau = 1$ and $\tilde{\phi}_1$ tests the null hypothesis of $H_0 : \alpha_\mu = 0, \rho_\mu = 1$.

It has been observed that the size and power of the ADF test are severely sensitive to the number of lagged terms p used. Choice of optimal lag structure is an empirical issue. Based on Monte Carlo studies, Hall (1994) and Ng and Perron (1995) have found that both AIC and BIC criteria underestimate the optimum lag length, which in turn results in high size distortions. Considering the problems with AIC and BIC, Hall's (1994) prescription is of general to specific rule, i.e., to start with a large value of p , test the significance of the last coefficient and reduce p iteratively until a significant statistic is found. There are some other suggestions as well by Schwert (1989) and Diebold and Nerlove (1990).

It is well-known that the unit root test statistics do not follow the usual Student's t -distribution or Fisher's F-distribution (in case of joint significance). Depending on the actual specification under the alternative, the appropriate critical values at selected significance levels have been computed by Fuller (1976) , Dickey and Fuller (1981), Guilkey and Schmidt (1989) and Mackinnon (1990). Apart from ADF test, there is another unit root test, called the PP-test, which is due to Phillips (1987) and Phillips and Perron (1988) , which is also often used for testing the presence of unit root in a time series. The advantage with this test is that it allows for a wide range of serial correlation and heterogeneity patterns in the noise term ε_t . The statistic under this test, $Z(\tau)$, can be compared to DF τ test. Different aspects corresponding to this two tests, especially the power performance, have been

widely studied. Some relevant references are Sargan and Bhargava (1983), Park (1990), Kwiatkowski *et al.* (1992) (popularly known as KPSS), Leybourne (1995), Arellano and Pantula (1995), Elliott *et al.* (1996) and Perron and Ng (1996). Some tests based on robust estimation methods, as opposed to ordinary least square (OLS) method, have been suggested by Rudebusch (1992), Andrews (1993), and Andrews and Chen (1994). And, of course, there is also a literature on unit root tests based on Bayesian approach (see the special issue of Journal of Applied Econometrics (1991, no. 4), Econometric Theory (1994, no. 3-4) and Journal of Econometrics (1995, vol.68 for relevant references). However, amongst all such tests, ADF and PP tests are most often used by the practitioners. Whereas PP test is more sophisticated and sometimes more powerful as compared to ADF test for testing unit root, the usefulness has become restrictive due to severe size distortion (Perron and Ng, 1996), especially at the presence of negative moving average component. Several simulation studies (Phillips and Perron (1988), Schwert (1989), Dejong *et al.*, (1992a) have, in fact, shown that PP test has serious size distortion in finite samples when the data are generated by a process having a predominance of negative autocorrelations in first differences. However, Perron and Ng (1996)) have put forward several modifications of the original PP test for unit root, and explained that such size problem would not be appreciable in such modified tests. Dejong *et al.*, (1992b) have argued that on the basis of their Monte Carlo study that PP has very low power (generally less than 0.10) against trend stationary alternatives, but ADF test has power around one-third and thus is likely to be more useful in practice.

The Table 2.3.1 summarizes the unit root test results. The entries of the table show that the SENSEX is a unit root process, which in turn shows that it is a difference stationary process i.e., stationarity is achieved only after taking the first difference. One more observation is that log price does not

Table 2.3.1: Results of Augmented Dickey-Fuller Test

Variable	t-type Test (intercept)		t-type Test (trend)	t-type Test y_{t-1}			F-type Test			Lag (p)
	α_τ	α_μ	β	ρ_τ	ρ_μ	ρ	$\tilde{\phi}_1$	$\tilde{\phi}_2$	$\tilde{\phi}_3$	
log-SENSEX	1.49	1.87	0.90	-1.40	-1.60	1.91	1.75	1.44	0.41	1
Return	2.10	2.09	-1.21	-22.36	-22.32	-22.18	2.19	1.95	0.73	1
5 % critical value	3.11	2.54	2.79	-3.45	-2.88	-1.95	4.71	4.88	6.49	

follow a pure random walk process as a lag of order $p = 1$ has been found in all equations (2.3.1) through (2.3.6) to render the residuals to be white noise. This observation goes against the random walk hypothesis for the Indian stock market.

2.3.2 Testing for Independence

The problem of testing for independence in a sequence of observations is of fundamental importance in statistics. Most of the inferential procedures concerning population characteristics are based on the assumption that the observations form a random sample. However, in many situations the data are collected sequentially so that tests for randomness is desirable. Furthermore, time series observations are generally associated with each other and tests of randomness may therefore be carried out prior to proceeding for further analysis. Formally speaking, dependency is defined in terms of joint and marginal distributions. Two random variables are said to be independent if the joint distribution of these two random variables is same as the product of their in-

dividual marginal distributions. Under normality assumption, simple product moment correlation coefficient is taken as a measure of dependency. However, product moment correlation, in general, measures only linear dependency; it does not tell anything about the nonlinear structure inherent in the data. In other words, under non-normality zero product moment correlation or zero autocorrelations can not be regarded as a measure of independency. Furthermore, studies by Cox (1966) and Bartels (1977) have shown that in small samples the null distribution of the serial correlation coefficient may be badly affected by the departure from normality in the underlying process. The distortion in the null distribution is particularly severe if the underlying process is distributed with long tails. Moreover, Hannan and Kanter (1977) have proved that for the class of stable distributions, the usual asymptotic results also do not hold. Hence, apart from the correlation-based tests, we also describe other tests of dependence all of which may not depend on the assumption of normality of the underlying variable. However, frequency domain based tests like Subba Rao-Gabr (1980) test and parametric model based tests like the one suggested by Keenan (1985) are excluded.

Autocorrelation Based Tests

One of the traditional tests used to examine the independence in a time series is the portmanteau test. If a stationary time series is independently distributed, then the autocorrelation coefficients ρ_k of order k are zero for all lags $k = 1, 2, \dots$. The approximate standard error of the estimator r_k of ρ_k (cf. Kendall and Stuart (1961)) is given by

$$SE(r_k) = \frac{1}{\sqrt{n}}.$$

However, out of a large number of individual autocorrelations some may turn out to be significant due to pure chance factor even when their true values are zeros. To avoid such problems, Ljung and Box (1978) defined a joint test

statistic which is a modification of Box-Pierce (1970) test, commonly known as Ljung-Box statistic, and is defined as

$$Q(k) = n(n+2) \sum_{m=1}^k (n-m)^{-1} r_k^2$$

where sample autocorrelation coefficient up to k lags are present. Under the null hypothesis of $H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0$; $Q(k)$ is asymptotically distributed as $\chi^2(k)$ with k degrees of freedom.

It may be noted that this test statistic assumes (linear) autocorrelation only in the alternative. However, it is also used for testing nonlinear dependency using squared or absolute values of the time series. Based on similar considerations, McLeod and Li (1983) suggested a 'portmanteau' test for testing nonlinear dependence in the series.

The first order autocorrelation for return, absolute return and squared return based on SENSEX has been obtained as -0.22, 0.41 and 0.49, respectively and other higher order autocorrelations for all the three series *viz.*, return, absolute return, and squared return have been found to be insignificant. The $Q(\cdot)$ test statistics for these three series have been obtained as 39.88, 211.32 and 272.38, respectively, at lag 1. It is obvious from these values of the test statistics that the null hypothesis of zero-autocorrelation is rejected for each of returns, absolute return and squared return series at all standard nominal levels of significance.

Turning Point Test

Turning point test is one of the older tests for examining the independence based on number of turning points of a sequence of observations. Let $y(1), y(2), \dots, y(n)$ be an ordered sequence of returns corresponding to the observed returns y_1, y_2, \dots, y_n . Then a turning point is said to occur at time t , ($1 < t < n$) when $y(t-1) < y(t)$ and $y(t) > y(t+1)$ or when $y(t-1) > y(t)$ and $y(t) < y(t+1)$. Let M be the number of turning points from the se-

quence. If the observations of the sequence are independent, then the probability of a turning point at time t is $2/3$. The expected value of M is given as $E(M) = 2(n - 2)/3$, and the variance $V(M) = (16n - 29)/90$. The test statistic to test the randomness of the series of observations is given by

$$Z = \frac{|M - E(M)|}{\sqrt{V(M)}}.$$

It can be shown that under the null of randomness, Z follows a standard normal distribution with mean zero and variance one. The computed Z statistic value for our data has been obtained as 1.20, which is obviously below the 5 per cent critical value of 1.96. We may thus conclude on the basis of this test that the return series based on SENSEX is independent.

Runs Test

Runs test is another traditional test of independence of observations based on number of runs. Given an ordered sequence of two or more types of symbols, a run is defined to be a succession of one or more identical symbols which are followed and preceded by a different symbol or no symbol at all. To define runs test, let n_1 be the number of one type of symbol, say positive (+) signs and n_2 be the number of negative (-) signs. Assuming there is no tie the expected value (*cf.* Gibbons,1971) of \tilde{r} , where \tilde{r} is the total number of runs, is obtained as

$$E(\tilde{r}) = \frac{2n_1n_2}{n_1 + n_2} + 1,$$

and the variance of \tilde{r} is given as

$$V(\tilde{r}) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}.$$

Thus, the test statistic to test for independence is defined as

$$Z = \frac{\tilde{r} - E(\tilde{r})}{\sqrt{V(\tilde{r})}}.$$

Z follows standard normal distribution with mean zero and unit variance (see Gibbons, 1971).

Insofar, as the application of this test to SENSEX data is concerned, the value of the test statistic came out to be 3.57, which is greater (in absolute) than the 5 per cent critical value 1.96 of $N(0, 1)$. Therefore, we reject the null hypothesis of independence at 5 per cent level of significance.

Rank Version of the von Neumann Ratio Test

The rank test as defined by Bartels (1982) is described as follows. Let $R(i)$ be the rank associated with the i -th observation in a sequence of n observations, then the rank test statistic is defined as

$$R = \frac{\sum_{i=1}^{n-1} [R(i) - R(i+1)]^2}{\sum_{i=1}^n [R(i) - \hat{\mu}_r]^2},$$

where $\hat{\mu}_r$ is the mean of the return series. Critical values of R has been given in Bartels (1982). Bartels has also provided explanations for relative superiority of rank test over runs test. The computed value of R for our data has been found to be 565.75; this value is much higher than the Bartels critical value 1.82 at 5 per cent level of significance. Thus, the null hypothesis of independence of returns is decisively rejected by this test.

Brock, Dechert, and Scheinkman (BDS) Test

BDS test is a very powerful and popular test for testing the null hypothesis that a series is independently and identically distributed (*i.i.d.*). It is a diagnostic test since a rejection of the null is consistent with some type of dependence in the data, which could result from a linear stochastic system, a non-linear stochastic system or a non-linear deterministic system. However, additional diagnostic tests are required to determine the source of the rejection. In this test the dependence of y_t is examined through the concept of *correlation integral*, a measure that examines the distance between points, called *embedded* points in say, m -space. The m -space is constructed by considering the m -tuple vector as

$$y^m(t) = [y_t, \dots, y_{t-m+1}], \quad t = 1, 2, \dots, n - m + 1.$$

For each embedding dimension, m , and a choice of epsilon, ϵ , the radius of the hypersphere which determines whether two points are close or not, the correlation integral is defined as

$$c(\epsilon, m, n) = [n_m(n_m - 1)]^{-1} \sum_{t \neq s=1}^{n-m+1} I_\epsilon[y^m(t), y^m(s)], \quad n_m = n - m + 1.$$

The indicator function I is defined as

$$\begin{aligned} I_\epsilon(y^m(t), y^m(s)) &= 1 \text{ if } \|y^m(t) - y^m(s)\| \leq \epsilon \\ &= 0 \text{ otherwise,} \end{aligned}$$

where $\|\cdot\|$ stands for the supremum norm. Thus, the correlation integral will measure the fraction of total pairs of $[y^m(t), y^m(s)]$ for which the distance between the points $y^m(t)$ and $y^m(s)$ is no more than ϵ . If the value of ϵ is so chosen that all the pairs satisfy the condition, then $c(\epsilon, m, n) = 1$. On the other hand if ϵ is so chosen that the condition is never satisfied, then the correlation integral is identically zero. For this reason, the correlation integral sometimes interpreted as a measure of spatial correlation.

Brock and Baek (1991) have demonstrated that under the assumption of *i.i.d.*, the BDS test statistic is asymptotically standard normal ¹. More precisely,

$$W(\epsilon, m, n) = \sqrt{n}[c(\epsilon, m, n) - c(\epsilon, 1, n)^m]/V^{1/2} \Rightarrow N(0, 1),$$

where the expression for V can be found from Hsieh (1989). BDS test is very sensitive to the choice of m and ϵ . Often in empirical studies, the series is so transformed that all the values of the series lie within a unit interval and set the values of ϵ as $\epsilon = 0.9^i$, $i = m - 1$ and the embedding dimension, m , is chosen over the range 2 to 5. Now, we report the results of BDS test on the

¹In this context it is worthwhile to note that BDS test statistic does not follow asymptotic normality when the modelling framework incorporates non-linear dependence.

return data. It may first be noted that we have transformed the return series so as to render the series to lie within the unit interval, and ϵ has been chosen as stated above. As regards the values of m used for the present study, we took $m = 2$ and $m = 3$, which is a very common choice. The BDS test statistic for $m = 2$ and $\epsilon = 0.9$ has been found 6.63, whereas that for $m = 3$ and $\epsilon = 0.81$ has been obtained as 10.14. Thus for both the cases of $m = 2$ and 3, the calculated BDS test values exceed the 5 per cent standard normal critical value of 1.96; and hence we conclude that by BDS test the null hypothesis of independence of returns is found to be rejected in favour of the alternative.

On combing the findings of all these tests of independence on return data, we observe that excepting for turning point test, all other tests suggest that the null hypothesis of independence be rejected. Thus, the evidence overwhelmingly is in favour of dependent returns.

2.3.3 Testing for Normality

Normality is the most familiar and convenient distributional assumption in all branches of statistics including time series analysis. However, though this distributional assumption may be quite appropriate for design of experiment based data, data pertaining to social sciences, especially financial ones, do not often conform to normality assumption. In fact, financial data have usually been found to be heavy tailed as compared to the normal distribution. Although there are a number of tests for normality available in the literature, we present some standard and relevant tests only (see, D'Agostino and Stephens, 1986, for discussions on some other tests).

Normal Probability Plot

The normal probability plot is the most traditional method designed to detect non-normality. It is nothing but the plot of the ordered observations

against the normal order statistics for the appropriate sample size. The order statistics under normality are the expected values of ordered observations from the normal distribution with mean zero and unit variance. The normal order statistics i.e., the i -th order statistic under normality is $x_{(i)} = \Phi^{-1}(p_i)$ where $p_i = (R_i - 3/8)/(n + 1/4)$, R_i is the rank and $\Phi(\cdot)$ is cumulative distribution function of standard normal variable. The plot $x_{(i)}$ against the ordered observations $y_{(i)}$ is called the normal probability plot. Under normality the plot is a straight line, and elongated 'S' shaped curve suggests heavy tailed or light tailed distribution, depending on the direction of 'S'. It may be noted that heavy tailed distributions have a relatively higher frequency of extreme observations than the normal distribution; light tailed distributions have relatively fewer.

The normal probability plot for the return data is given in Figure 2.3.2. The plot evidently suggests a clear departure from normality. The shape also indicates that the distribution may not be skewed though both tails are heavy.

Shapiro-Francia (SF) Test

The Shapiro-Francia (1972) test statistic for normality, a modification of the Shapiro-Wilk (1965) test, provides a direct quantitative measure of the degree of agreement between the normal plot and the expected straight line. The SF statistic is the squared correlation between the observed ordered observations and the normal order statistics. Let z be the vector of centred ordered observations and w be the vector of normal order statistics. Then SF statistic is given by

$$SF = (z'w)^2 / ((z'z)(w'w)).$$

Since it is a measure of association, the small value of SF suggests rejection of the null hypothesis of normality. Critical values are available from Shapiro-Francia (1972).

The value of SF test statistic for the return series has been found as 0.59

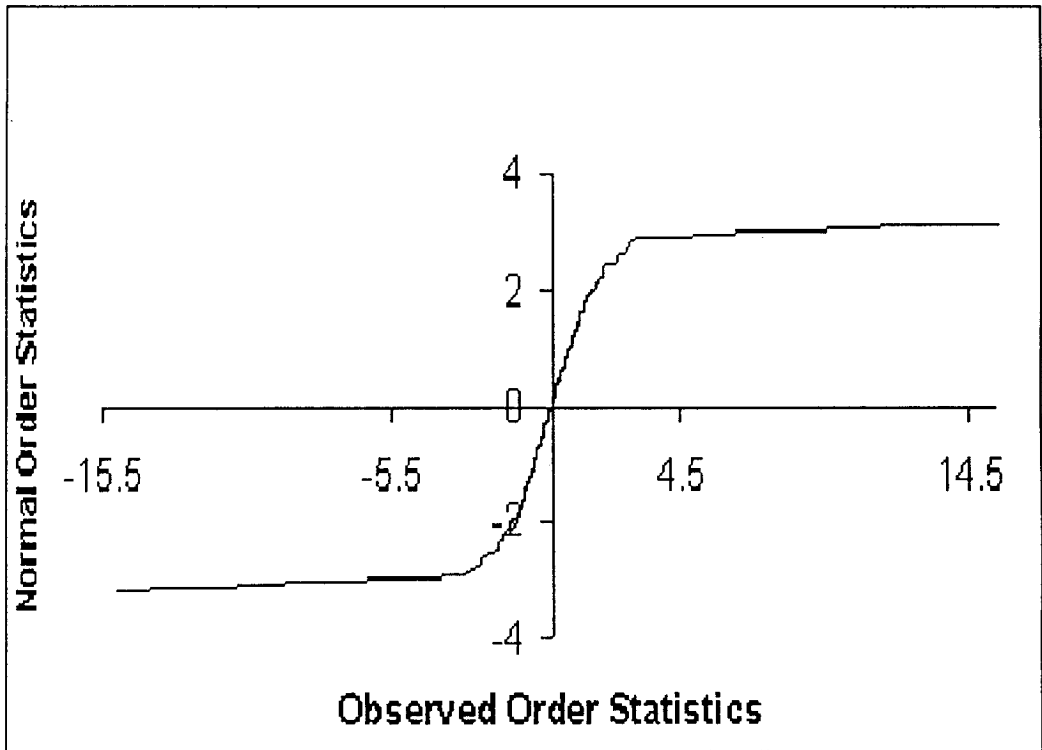


Figure 2.3.2: Normal Probability Plot of Returns

which is far below the 5 per cent critical value of 0.976, suggesting the rejection of null hypothesis of normality.

Jarque-Bera (JB) Normality Test

Normal distribution has some useful properties. Being a symmetric distribution, its all odd-ordered central moments greater than 2 are zero; its skewness (β_1 coefficient) is zero and kurtosis (β_2 coefficient) is 3. Jarque-Bera (1987) normality test is based on these unique properties of normal distribution. JB test statistic is defined as

$$JB = (n/6)\beta_1 + (n/24)(\beta_2 - 3)^2.$$

Under the null hypothesis of normality, JB is distributed as χ^2 with 2 degrees of freedom. JB test statistic value for the return series has been obtained as 585048.44 which is a very large value indeed. The conclusion, therefore, is that the null hypothesis of normality is strongly rejected for SENSEX based return series.

Studentized Range (SR) Test

Sometimes presence of some extreme observations, usually called outliers, in the data may increase the kurtosis coefficient. In such situations SR test is very useful. The test statistic is defined as

$$SR = \frac{(y_{(n)} - y_{(1)})}{\{[1/(n-1)] \sum_{i=1}^n [y_i - \mu]^2\}^{\frac{1}{2}}}.$$

The null hypothesis of normality is rejected when SR value exceeds the selected critical value as obtained from Pearson and Hartley (1966). The computed value of this test statistic for our data has been obtained as 30.74. Since the critical value at 5 per cent level of significance is 5.01, we conclude once again that the null hypothesis of normal distribution for return series is strongly rejected.

Empirical Distribution Function (EDF) Based Test

The empirical distribution function (EDF) is a sample estimate of the population distribution. EDF statistics are measures of the discrepancy between the EDF and a given distribution function under the null hypothesis which in our case is normal. Let $F_n(\cdot)$ be the empirical distribution function and F the population distribution function. One of the most important classes of EDF statistic is the supremum class, of which the most popular one is the Kolmogorov-Smirnov one sample test. This test is defined as

$$D^+ = \sup_x \{F_n(x) - F(x)\}, \quad D^- = \sup_x \{F(x) - F_n(x)\},$$

and

$$D = \max(D^+, D^-).$$

The D^+ test statistic value for our data has been obtained as 0.11 which exceeds the critical value of 0.04 at 5 per cent level of significance, and consequently we conclude that the null hypothesis of normality is rejected.

Another class of measure of association between EDF and the population distribution is what is known as the quadratic class. A most important member of this quadratic class is Anderson-Darling (1954) statistic, which is defined as

$$AD = (-1/n) \left[\sum_{i=1}^n (2i-1) \{ \log x_{(i)} + \log(1-x_{(n+1-i)}) \} \right] - n.$$

Where $x_{(1)} < x_{(2)} \cdots < x_{(n)}$ and $x_{(i)} = \Phi(y_{(i)})$, $0 < x_{(i)} < 1$. The null hypothesis of normality is rejected for large values of AD statistic; the relevant critical values are available from Anderson and Darling (1954). For the given data set, AD test statistic has been calculated as 35.20, and this exceeds the critical value which, for instance, is 2.49 at 5 per cent level of significance. This test, therefore, leads to the conclusion of rejection of normality for return data.

All the tests considered in this sub-section for testing the assumption of normality reject the null hypothesis of normality. We thus conclude that the

assumption of normal distribution for the return data is not tenable on the basis of conclusions drawn by these tests.

2.3.4 Testing for Long-Memory

We have found that the return series is (covariance) stationary, and hence it does not have the usual non-stationary (unit root) type long memory. However, the series may still have long-range dependence in the sense that the series would not satisfy the notion of strong mixing due to Rosenblatt (1956). Heuristically, a time series is strong mixing if the maximal dependence between events at any two dates becomes trivially small as the time span between those two dates increases. Strong mixing can be taken as an operational definition of short-range dependence (see, Lo, 1991). In other words, a (covariance) stationary process may still exhibit long-range dependence, where autocorrelation function decay slowly as compared to the standard stationary ARMA process. One such particular example is the fractionally integrated ARMA (ARFIMA) process with appropriate fractional differences.

The presence of long-memory components in asset returns has important implications in financial economics. For example, optimal consumption/savings and portfolio decisions may become extremely sensitive to the investment horizon if stock returns were long-range dependent. Moreover, conventional time series modelling is also no longer valid.

The importance of long-range dependence in asset markets was first studied by Mandelbrot (1971) and subsequently by Fama and French (1988), Lo and MacKinlay (1988), and Poterba and Summers (1988). One approach to detect evidence of strong dependence in time series is to use the 'range over standard deviation' or 'rescaled range' statistic, or R/S statistic, originally developed by Hurst (1951) and popularized by Mandelbrot (1972). The R/S statistic is

the range of partial sums of deviations of a time series from its mean, rescaled by its standard deviation. This statistic, denoted by \bar{Q}_n , is given by

$$\bar{Q}_n = \frac{1}{s_n} [\max_{1 \leq k \leq n} \sum_{j=1}^k (y_j - \bar{y}_n) - \min_{1 \leq k \leq n} \sum_{j=1}^k (y_j - \bar{y}_n)]$$

where \bar{y}_n is the usual sample mean and s_n is the usual standard deviation of the returns y_1, y_2, \dots, y_n . The first term in the brackets of \bar{Q}_n is the maximum (over k) of the partial sums of the first k deviations of y_i from the sample mean. Obviously, this maximum is always nonnegative. Similarly the second term in the bracket of \bar{Q}_n is always non-positive, and therefore \bar{Q}_n is always non-negative. In several studies Mandelbrot (1972), and Mandelbrot and Wallis (1969) have demonstrated the superiority of R/S statistic to more conventional methods of determining long-range dependence, such as analyzing autocorrelations, variance ratios, and spectral decompositions. Specifically, Mandelbrot and Wallis (1969) have shown by Monte Carlo studies that the R/S statistic can detect long-range dependence in highly non-Gaussian time series with large skewness and kurtosis. Moreover, Mandelbrot (1972) has established the almost sure convergence of this statistic for stochastic processes with infinite variances, which is a distinct advantage over autocorrelations and variance ratios which need not be well-defined for infinite variance processes. Although it has long been established that \bar{Q}_n statistic has the ability to detect long-range dependence, later studies have found that it is sensitive to short-range dependence also, and accordingly Lo (1991) advanced a modified version of \bar{Q}_n which is defined as

$$Q_{q,n} = \frac{1}{s_q} [\max_{1 \leq k \leq n} \sum_{j=1}^k (y_j - \bar{y}_n) - \min_{1 \leq k \leq n} \sum_{j=1}^k (y_j - \bar{y}_n)],$$

where $s_q^2 = s_n^2 + 2 \sum_{j=1}^q w_j \gamma_j$, $w_j = 1 - \frac{j}{q+1}$, $q \leq n$, and γ_j 's are the usual autocovariances. It can be noticed that $Q_{q,n}$ differs from \bar{Q}_n only in its denominator. If y_i is subject to short range dependence, the variance of the

partial sum is not simply the sum of the variances of the individual terms, but it also includes the autocovariances.

Lo (1991) has established that this modified R/S test is robust to many forms of heterogeneity and weak dependence, and hence such modified test covers a broader set of null hypotheses than those of classical ones. Lo has also established that under the null hypothesis of short-range dependence, $V = Q_{q,n}/\sqrt{n}$ converges in distribution to a well-defined random variable; its associated critical values at various significance levels have been also reported in Lo . The V statistic is consistent against a class of long-range dependent stationary Gaussian alternatives as well as all fractionally-differenced Gaussian ARIMA(p, d, q) processes with $d \in (-\frac{1}{2}, \frac{1}{2})$. Of course, if one is interested exclusively in fractionally-differenced alternatives, a more efficient means of detecting long-range dependence might be to estimate the fractional differencing parameter directly as done by Geweke and Porter-Hudak (1983) and Yajima(1988). However, as noted by Lo, the modified R/S test is perhaps most useful for detecting departures from the null hypothesis into a broader class of alternative hypotheses, i.e., it is a kind of ‘portmanteau’ test statistic that may complement a comprehensive analysis of long-range dependence.

To test the long-range dependence of the return series as obtained from SENSEX, we have applied both the classical and the modified R/S tests. For modified R/S test different values of the number of autocovariances, q , ranging from 25 to 250, has been used. The computed value of V statistic for the classical R/S test has been obtained as 1.682, whereas the value of V statistic for the modified R/S has been found slightly smaller than 1.682 for all q , but always greater than 1.40. For both the cases, the observed V statistic value was always within the interval [0.809,1.862], the 95 per cent level of confidence interval, suggesting thereby the acceptance of the null hypothesis of short-range, or equivalently weak dependence. The finding was also true

for different arbitrary sub-groups of approximately 300 observations, e.g., (1-300), (301-600), and (601-871). It is, therefore, appropriate to conclude that the classical R/S as well as the modified R/S statistics do not suggest the presence of long-range dependence in the return data. In other words, any shock to the return series is purely temporary and does not persist for a long horizon.

2.3.5 On Existence of Moments

It is a well-established fact that the distribution of stock returns exhibits heavy tails than the normal distribution. Studies by Mandelbrot (1962, 1963, 1967) and Fama (1965) and subsequently Mittnik and Rachev (1993), to name a few, led them to infer that return distributions are heavy tailed, and also to reject the normality assumption and propose Stable Paretian distribution as a distributional model for asset return.

The observed fatness in the tails of returns has important implications for modelling stock prices. The different financial models are heavily dependent on the existence of second or higher order moments of returns. The distributional form and existence of different moments, especially variance, of returns on financial assets has important implications for theoretical and empirical analyses in economics and finance. For example, asset portfolio, option-pricing theories *etc.* are typically based on distributional moments. This 'fat-tailness' property of the financial variables has vital importance to risk management, and in particular, to Value at Risk (VaR) analysis which directly focusses on the tail (left) of return distribution. Obviously, therefore, the traditional normal based analysis would underestimate the frequency of the extreme events for such financial variables.

In such situation especially in the context of evaluating extreme risks in the

financial markets, modelling the tail may be much more useful in contrast to characterizing the complete distribution, which by itself is often quite difficult. In this sub-section we briefly review the methods available for estimating the tail index or the maximum exponent for such variables.

It may be noted that as the tail becomes more and more thick the higher order moments depending upon the tail thickness are likely to be unduely large. Modelling financial time series sometimes may require very high level of moment conditions. The failure of moment conditions may have adverse effect on the parameter estimates and the underlying test statistics, and consequently the inferences may be misleading. For example, different non-linearity tests including those of autoregressive conditional heteroscedasticity (ARCH) are not robust to the failure of the moment conditions (see, Lima, 1997). In fact, Lima's simulation studies have shown that this problem is more pronounced for extremely heavy tailed distributions.

We begin by defining a heavy tail distribution. Let y be a random variable with distribution function $F(\cdot)$ and density function $f(\cdot)$. Tail behaviour of standard normal distribution which is characterized by Mill's ratio, is given by

$$P[Y > y] \sim f(y)/y = \frac{\frac{1}{2\pi} \exp(-y^2/2)}{y} \text{ as } y \rightarrow \infty$$

i.e., decay rate is exponential, and as we know, bulk of the statistical work deals with such light tails.

In contrast to such normal tail or light tail a typical heavy tail such as Pareto tail with index $\alpha > 0$ is defined as (see, Resnick ,1997 for details)

$$1 - F(y) = P[Y > y] = y^{-\alpha}, \quad y > 1.$$

More generally, we say Y has a heavy tailed distribution function F if

$$1 - F(y) = P[Y > y] = y^{-\alpha} L(y),$$

where $L(\cdot)$ is a slowly varying function, i.e.,

$$L(ty)/L(t) \rightarrow 1 \text{ as } t \rightarrow \infty.$$

Note that when $y > 0$, $E(Y^\beta) < \infty$, for $\beta < \alpha$ and $E(Y^\beta) = \infty$, for $\beta > \alpha$. To assess whether heavy tails are present, and then to estimate the tail index α , exploratory plotting as well as different estimates are available. However, among these Hill estimator as proposed by Hill (1975), is widely used. To define Hill estimator, let us assume that the sequence of returns, y_1, y_2, \dots, y_n drawn from a stationary i.i.d. process whose probability distribution function $F(\cdot)$ is unknown. As earlier defined, $y_{(1)} > y_{(2)} > \dots > y_{(n)}$ are the order statistics. If Y has an exact Pareto distribution, then

$$1 - F(y) = y^{-\alpha}, \quad y > 1.$$

Now, taking logarithms yields a sample from an exponential density with parameter α . Since the mean of the exponential distribution is $1/\alpha$, the maximum likelihood estimator (MLE) of $1/\alpha$ is the sample mean, and thus

$$H_n = \frac{1}{n} \sum_{i=1}^n \ln y_{(i)}$$

is the MLE of $1/\alpha$. If instead of assuming a Pareto distribution, we only assume

$$1 - F(y) = y^{-\alpha}L(y), \quad y \rightarrow \infty,$$

and consider ($k < n$) extreme observations, then Hill estimator of $1/\alpha$ can be defined as

$$H_{k,n} = \frac{1}{k} \sum_{i=1}^k \log \frac{y_{(i)}}{y_{(k+1)}}.$$

The rough idea, as given in Resnick (1997), for using only k extreme observations is that one should only sample from that part of the distribution which looks like Pareto. Under certain conditions, Hill estimator asymptotically follows a normal distribution with mean $1/\alpha$ and variance $1/\alpha^2$.

Although Hill estimator is very simple to compute and has nice distributional property, this estimator is known to be biased. It is also known that Hill estimator is too sensitive to k , the number of extreme values used. Smaller choice of k leads to a high variance and thus an inaccurate estimate. On the other hand, a large k induces an inaccuracy as well, because the true threshold level may be crossed so that the distribution of the observations used in the estimation is not approximately Paretian, and this causes a bias in the estimate. The problem is to find an optimal k . Apart from the usual heuristic approach of studying the plot of Hill estimates for various k and finding the optimal k where the plot is stable, there are less heuristic approaches to find an optimal number of tail observations or the cut-off level. Two such approaches are: Monte Carlo experiment and Bootstrap method, as advanced by Hall (1990).

In Figure 2.3.3, we have plotted Hill estimates, based on 455 positive returns, for different values of k . The plot shows that the maximum estimated value of α is 4.81. To find the optimal value of k , we have performed a series of Monte Carlo experiments. For the purpose of generation of data for the Monte Carlo study, Student's-t distribution has been considered, because this is also a heavy tailed distribution. The tail index of a Student-t distribution is equal to the number of degrees of freedom. Each Monte Carlo experiment consisted of 500 replications of n ($=871$) draws from four different Student-t distribution with degrees of freedom ($\equiv \alpha$) being equal to 1, 2, 3 and 4. Optimum k was chosen by minimizing mean square error (MSE) across the replications. For $\alpha = 1, 2, 3,$ and 4 , optimal choice of k was found to be 129,54,32,19, respectively.

Simulation experiment suggests that the optimal k is inversely related to α , as is expected also; lower α implies fatter tails and hence more observations are clustered in the tail. We now summarize the results with respect to the

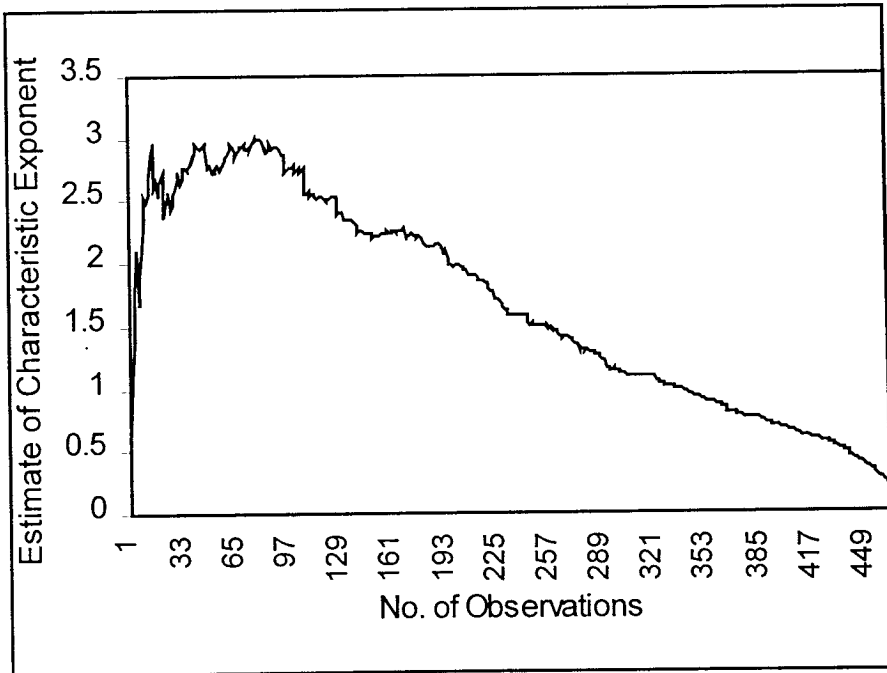


Figure 2.3.3: Sensitivity to the Cut-Off Level

Table 2.3.2: (Right) Tail Index Estimates and Test Statistics

Hypothesis	$H_0 : \alpha < 1$	$H_0 : \alpha < 2$	$H_0 : \alpha < 3$	$H_0 : \alpha < 4$
	$H_1 : \alpha \geq 1$	$H_1 : \alpha \geq 2$	$H_1 : \alpha \geq 3$	$H_1 : \alpha \geq 4$
Estimate of α	2.52	2.95	3.06	3.54
Test Statistic	6.86	2.36	0.11	-0.57
Optimal k	129	54	32	19

right tail index estimates as displayed in Table 2.3.2. We find from this table that both the null hypotheses $H_0 : \alpha < 1$ and $H_0 : \alpha < 2$ against their respective one sided alternative hypotheses are rejected at 5 per cent level of significance as both the observed test statistics exceed 1.64. It is obvious that since second moment exists, so would be the first moment. On the other hand as $H_0 : \alpha < 3$ cannot be rejected at 5 per cent level of significance in favour of the alternative hypothesis of $H_1 : \alpha \geq 3$, the higher order moments do not exist. Since the choice of k may influence empirical findings, we have checked for the sensitivity of the optimal k by taking into consideration other values of k around their respective optimal k values. In the above Monte Carlo study, the MSE values were computed, for all such choices of k , and it was found that the values were almost constant for a range of about 20 values.

The Monte Carlo study has been carried out with data generated from t -distribution with 4 different degrees of freedom. However, the tail behaviour of any true data generating process under consideration may not be very close to the tail of the t -distribution. We have, therefore, used two other procedures, to obtain the optimal cut-off level, of k , and the estimate of α . These are (i) the modified Hill estimator, as prescribed by Huisman *et al.* (2001), and (ii) the bootstrap method (*cf.* Hall, 1990).

Recently, Huisman *et al.* (2001) has put forward a modified version of Hill

estimator which performs better in small samples. To obtain this modified Hill estimate, we need to estimate the following regression equation

$$1/\alpha(k) = \beta_0 + \beta_1 k + \epsilon(k), \quad k = 1, \dots, K.$$

This regression equation is estimated by weighted least squares, weights are given in Huisman *et al.* (2001). The estimated value of β_0 is the modified Hill estimator. K is usually considered as half of the (positive) observations. We have calculated this modified Hill estimator for the return data taking into consideration their rule of thumb for selecting the number of observations for the regression. The modified Hill estimator of α has been obtained as 3.59, which indicates lower fatness of the tail as compared to the conventional Hill estimate. It may be noted that this finding with our return data is consistent with the conclusion of Huisman *et al.*

Now, to describe the results obtained from bootstrap method, we first state that 1000 bootstrap samples each of size of 50 were drawn. Given the sample size of 871 observations, we found that very few positive returns were included in some bootstrap samples when the sample size was taken to be less than 50. We, therefore, choose to fix 50 as the required sample size for our study. The initial estimate of tail index was taken as 2.95 from consideration of our finding on α from the Monte Carlo study. Using these bootstrap samples of size 50 each, we carried out the required computations, and found the cut-off point for k to be 47. Based on this k , we found the tail index to be 3.01, which is very close to the Monte Carlo based estimate.

So far we have considered only right tail using positive returns. However, in the similar manner we can estimate the left tail index as well. For this purpose, the same Monte Carlo based optimal cut-off points have been considered. Table 2.3.3 summarizes the left tail behaviour of the returns.

We find from Tables 2.3.2 and 2.3.3 that the left- tail index estimates

Table 2.3.3: (Left) Tail Index Estimates And Test Statistics

Hypotheses	$H_0 : \alpha < 1$	$H_0 : \alpha < 2$	$H_0 : \alpha < 3$	$H_0 : \alpha < 4$
	$H_1 : \alpha \geq 1$	$H_1 : \alpha \geq 2$	$H_1 : \alpha \geq 3$	$H_1 : \alpha \geq 4$
Estimate of α	2.62	2.98	3.16	3.62
Test Statistic	7.02	2.42	0.29	-0.46
Optimal k	129	54	32	19

are almost the same as that of the right tail index; one of the implications, therefore, is that the distribution of return is almost symmetric.

In this sub-section, we have investigated the tail behaviour of the stock return, instead of looking at the entire distribution. Distribution of stock return has been found to be having fat-tails. Tail index, which is a measure of tail behavior of the underlying variable, has been estimated using the well-known Hill's estimator. As the estimate is likely to be very sensitive towards the number of tail observations used, a Monte Carlo study and a bootstrap method based study have been carried out to find the optimal number of tail observations. Empirical estimate of the tail index shows the existence of the second moment. Further, Hill estimates suggests that third and other higher order moments may not exist for return data. We may thus conclude, on the basis of these empirical findings, that heavy tail distributions like those based on conditional heteroscedasticity, Student's t -distribution etc. are more appropriate for return data.

2.4 Conclusion

In this chapter, we have provided finer insights into the empirical distribution of a widely used stock price index, popularly known as SENSEX. Commonly

used assumptions regarding the data generating process e.g., normality, stationarity, linear and nonlinear dependencies and existence of different orders of moments have been tested applying various methodologies available in the literature. It has been found that the index under study i.e., SENSEX is non-stationary, but return based on SENSEX is stationary. Various normality test suggests that return distribution is far from normal. To test for dependency (linear as well as nonlinear) we have used various classical, nonparametric and autocorrelation based tests along with correlation integral based BDS test. It has been found that there is strong dependency structure in the return data and this goes against much published random walk hypothesis for stock markets. To test the long-range dependence of the return series as obtained from SENSEX; we have applied both the classical and the modified R/S tests. The classical R/S or the modified R/S tests do not suggest the presence of any long-range dependence in the returns. In other words, any shock to this return series is purely temporary and does not persists for a long horizon.

The very fundamental question of existence of moments, as characterized by the tail index, has been studied by considering the widely used Hill estimator. As the Hill estimator is too sensitive to the choice of number of tail observations used, we have performed a Monte Carlo study as well as a bootstrap based exercise to find the optimal number of extreme observations for the Hill estimator. The findings were more or less the same for both the approaches. Finally, Hill estimate suggests that though the existence of second moment can be assumed the same may not be case for higher order moments.

Chapter 3

Estimation and Testing of ARCH-M

Parameters: Estimating Equations Approach and Bootstrap Technique

3.1 Introduction

In their paper in 1987, Engle *et al.* proposed the ARCH-M model where the conditional variance is a determinant of the current risk premium. As regards estimation of this model, they suggested ML method of estimation based on the assumption of conditional normality of the errors, and they concluded that the likelihood being of the form analyzed by Crowder (1976), the ML estimator $\hat{\theta}$ of the ARCH-M parameter vector θ has, under sufficient regularity conditions, the property that $(S'S)^{1/2}(\hat{\theta} - \theta_0) \sim$ asymptotically $N(0, I)$, where θ_0 is the true value of θ and S is the matrix of first order derivatives of the log-likelihood function.

Such likelihood based methods for the original ARCH model have also been studied. The difference, however, is that while researchers (see for details, Tsay, 1987; Bollerslev and Wooldridge, 1992; Lumsdaine, 1996; and Goncalves and Lopes, 1998) have derived sufficient conditions for the asymptotic prop-

erty to hold for the class of models represented by ARCH(q), GARCH(1,1), IGARCH(1,1), there is hardly any such research concerning ARCH-M class of models. This may be due to the fact this latter class of models poses additional difficulties since unlike (G)ARCH model, the conditional variance of (G)ARCH -M model is a nonlinear difference equation. Further, existence of higher order moments are required even for ARCH model, and these may be very difficult to justify in practice. For instance, Pantula (1984) worked out for ARCH(1) model that existence of eight-order moment of the disturbance is required. Weiss (1986) also investigated the same for a similar model, and came up with a weaker sufficient condition in the form of existence of fourth moment of the error. It appears that there is only one major paper by Lee (1991) who extended all these asymptotic distribution results to the GARCH(1,1)-M and IGARCH(1,1)-M models.

Now, it may be noted that most of these studies have used likelihood/quasi likelihood approach for estimating the parameters of the models. While some researchers have also used nonparametric method (see Pagan and Ullah, 1988; and Pagan and Hong, 1991) and generalized method of moments (see Rich *et al.*, 1991) for estimating parameters of ARCH class of models, there is no such study concerning ARCH-M or similar other models.

This chapter is on estimation and inference in ARCH-M model. This would be done using two different approaches: estimating equations approach and resampling procedure like bootstrap technique. In the context of estimation, it can be noted that (Gaussian) likelihood based estimation and inference is sometimes nonrobust, since the procedures adopted typically suit only the likelihood of choice. Also, it is often difficult to verify assumptions concerning the likelihood. In this chapter we first use a very general approach, called the estimating equations approach originally due to Godambe (1985). In this

approach, the estimator is obtained as a solution to certain equations

$$\sum_{t=1}^n \psi_i(X_t, \theta) = 0, \quad i = 1, \dots, l$$

for a vector valued parameter $\theta = (\theta_1, \theta_2, \dots, \theta_l)'$ and X is random variable defined on a sample space with a probability density function, with respect to a σ -finite measure on the sample space.

Godambe and Kale (1991) pointed out that a common feature of estimating methods such as least squares, ML method, method of moments, minimum chi-square is that these methods lead to a set of estimating equations. Thus it can be seen that the standard likelihood and quasi likelihood based approaches are special cases of the estimating equations technique. It may be stated that this approach has the added advantage that the asymptotic properties of most of the usual models involving conditional heteroscedasticity, *viz.*, ARCH, GARCH, ARCH-M, and GARCH-M can be studied in the same framework. Moreover, this approach enables us to consider these models with conditional heteroscedasticity process following ARCH(q), or GARCH(p, q) process where p and q can take integer values greater than 1.

In order to do inference in ARCH-M model with bootstrap methodology, we adopt in this chapter a novel bootstrap technique. In the next section we present an overview of the bootstrap procedure. In the context of ARCH-M model three different ways of bootstrap are possible. First, since the model is based on (Gaussian) likelihood; *parametric bootstrap* using the same likelihood and estimated parameter values is a possibility. This method suffers from the obvious defect of being very vulnerable to likelihood misspecification. Second, one may compute "residuals" based on the estimated ARCH-M model; and then use properly scaled and centered residuals as the population from which a simple random sample is drawn with replacement. The resample model is build on this sample. Third, one may look at some version of "data-pair"

resampling in the present context. Chatterjee and Bose (2000) suggested that for estimators obtained by solving

$$\sum_{t=1}^n \psi(X_t, \theta) = 0,$$

the appropriate generalization of the data pair bootstrap is to use

$$\sum_{t=1}^n w_{n:t} \psi(X_t, \theta) = 0,$$

where $(w_{n:1}, w_{n:2}, \dots, w_{n:n}) \sim \text{multinomial}(n, 1/n, \dots, 1/n)$. It is fairly easy to see that if the likelihood is correctly specified; parametric bootstrap will perform well. The picture is not so clear in the case of residual based bootstrap. In the context of least square estimation in linear regression, Liu and Singh (1992) showed that classical residual based technique is inconsistent if the errors are heteroscedastic. On the other hand, it is more efficient (in asymptotic comparison) than the data pair bootstrap if the errors are homoscedastic.

In the ARCH-M framework, we have conditional heteroscedasticity which, nevertheless, is unconditionally homoscedastic. Thus it is not at all obvious whether residual based bootstrap would provide consistent results or not. It is also possible that residual based bootstrap would provide consistent answer for some parameters, but not for all.

On the other hand, the (appropriate generalization of) data pair bootstrap is expected to be consistent in the presence of heteroscedasticity. Notice that naive use of data pair bootstrap in the linear regression framework will be inconsistent (see, for instance, Shao and Tu, 1995).

We have applied this generalized bootstrap to ARCH-M model, and then proved the consistency property of the resulting bootstrap distribution of the estimators of the parameters involved. A simulation study has also been carried out for the purpose of evaluating the performance of bootstrap as well as the usual normal asymptotic approximation. Finally, computations were

carried out, based on application of generalized bootstrap to return data, in order to draw conclusions on the relative risk aversion parameter. The organisation of this chapter is thus as follows. In the next section we present a very brief overview of the basics of bootstrap technique. The proofs concerning the consistency and asymptotic normality properties of the estimators of ARCH-M parameters obtained from estimating equations are given in Section 3.3. Section 3.4 establishes the consistency property of the bootstrap distribution. The results of the simulation study are presented and then discussed in Section 3.5. The findings on applications of bootstrap technique to SENSEX based return data are discussed in Section 3.6. The chapter ends with some concluding observations in Section 3.7.

3.2 A Brief Overview of Bootstrap Methodology

Most statistical procedures require some knowledge of the sampling distribution of the statistic being used for analysis. For example, constructing confidence sets and testing of hypothesis require the sampling distribution itself or the percentiles of the sampling distribution. On the other hand, in most estimation problems it is important to give an indication of the precision of a given estimate. A simple method is to provide an estimate of the bias, mean square error and variance, which are also often to be estimated from the data. A typical problem in statistics is how to estimate these accuracy measure, or, in general, how to estimate the sampling distribution itself. In the traditional approach, an accuracy measure is estimated by an empirical analogue of an explicit theoretical formula of the accuracy* measure which is based on the theory of weak convergence, and approximate the distribution of the statis-

tic, T_n , by its weak limit as $n \rightarrow \infty$. A major step is to obtain a result like $a_n(T_n - \theta) \Rightarrow N(0, V)$, for some $V > 0$ and some sequence $a_n \rightarrow \infty$. Often V is unknown, and one has to use some estimate of it based on observable data $X = (X_1, \dots, X_n)$.

In situations where the form of the statistic T_n is simple and algebraically amenable, the variance of the limiting distribution can be easily estimated. However, quite often the statistic is so complicated that finding some tractable formula for V is quite tedious. As Efron and Le Page (1992) have said " ... much of statistical history concerns..... finding approximations to ..(*the variance*).. for more general estimators ... and avoiding estimators that do not have such formulas."

Several general approaches have been developed in an attempt to get better approximations to the sampling distribution. These include systematic ways of improving an accuracy based on expansions, including those due to Edgeworth and Gram-Charlier. However, deriving such expansions are not so easy and sometimes may be quite complicated, and may even be impossible. A significant breakthrough was achieved by Tukey in 1958, when he proposed a new method for estimating the variance of T_n , known as jackknife procedure. Another way to estimate the sampling distribution of a statistic is by sampling from the data itself. This is known as bootstrap methodology, as originally proposed by Efron (1979) as a generalization of jackknife.

The important aspect of both the jackknife and the bootstrap is the repeated use of the original sample, by drawing sample-from-the-sample, or *re-sampling*. Thus, these techniques, and their generalizations are broadly called resampling techniques.

Two important papers came up soon after the publication of Efron's ideas about the bootstrap. Bickel and Freedman (1981) showed that consistency was obtainable under very general circumstances for many statistical functionals,

and Singh (1981) established the results that the bootstrap approximation is often better than the classical asymptotic normal approximation in the sense that the bootstrap distribution captures the second order term in the Edgeworth expansion of the original distribution. Thus bootstrap can be considered to be a breakthrough in three aspects:

- It is a computation based technique of obtaining the distribution of a large class of statistics.
- It often produces an estimate that is a better approximation than the asymptotic normal approximation.
- It avoids tedious mathematical derivations.

Over the past two decades a significant part of statistical research concentrated on broadening the class of problems on which bootstrap may be applied, and establishing that by using bootstrap, better approximations are obtainable than what is obtained using more traditional methods. Excellent review of bootstrap and jackknife is available in Efron and Tibshirani (1993), Hall (1992), Shao and Tu (1995), Davidson and Hinkley (1997), and Politis, Romano and Wolf (1999).

In order to illustrate the major ideas in bootstrap, let us discuss it in the context of least squares estimation in linear regression. Two different techniques of bootstrap are popular in this set-up. These are *residual bootstrap* and the *paired bootstrap* (see for details, Efron, 1982). In the case of the residual bootstrap, a sample of size n is drawn with replacement from the centered residuals $\epsilon_t - \bar{\epsilon}$, say $\epsilon_1^*, \dots, \epsilon_n^*$. Here $\epsilon_t = y_t - x_t \hat{\beta}_n$ and $\bar{\epsilon} = \sum \epsilon_t / n$, where x_t is the t -th row of the matrix X and $\hat{\beta}_n$ is the OLS estimator of β . The t^{th} bootstrap data point is defined by $y_t^* = x_t \hat{\beta}_n + \epsilon_t^*$. The bootstrap estimate of the parameter is calculated by applying least squares to this data, pretending that $\hat{\beta}_n$ is unknown.

For paired bootstrap, a sample $(x_1^*, y_1^*), \dots, (x_n^*, y_n^*)$ of size n is drawn with replacement from the observed data $\{(x_t, y_t), t = 1, \dots, n\}$. The bootstrap model is $y_t^* = x_t^* \beta + \epsilon_t^*$, and the bootstrap estimator is obtained by using least squares on this.

The residual bootstrap has been extensively studied by Freedman (1981,1984), Navidi (1989), Lahiri (1992) among others. Among its many nice properties there is this particular property that the bootstrap distribution is consistent even when the dimension p of the parameter tends to infinity with the data size n , as long as $p/n \rightarrow 0$. This shows that bootstrap is consistent under high dimensionality. Other properties are that for fixed p , the bootstrap distribution estimate is second order accurate, and thus it is a better approximation than asymptotic normality.

Liu and Singh (1992) showed two properties of this bootstrap for the problem of estimating the variance of $\hat{\beta}_n$. The first is that it is inconsistent if the errors ϵ_t 's are heteroscedastic, and the second is that it is more efficient than the delete-1 jackknife or the paired bootstrap for estimating the variance if the ϵ_t 's are homoscedastic. In an important and significant contribution, Lahiri (1992) showed that the residual bootstrap with some modifications is second order accurate for general M -estimators in multiple linear regression.

For the paired bootstrap, consistency of the distribution estimate was obtained under very general conditions in Mammen (1993), when the regressors are random and $p^4/n^3 \rightarrow 0$. Liu and Singh (1992) showed that for fixed p and non-random regressors, the paired bootstrap can consistently estimate the variance of β even when ϵ_t 's are heteroscedastic, but as already pointed out, it is less efficient than the residual bootstrap if ϵ_t 's are homoscedastic. It is not very difficult to see that the paired bootstrap distribution estimator is not second order accurate (Mammen, 1993; Hall, 1992; and Hall and Mammen, 1994).

It seems that the residual bootstrap has a very definite edge over the paired bootstrap because of its second order accuracy. Moreover, since it only involves replacement of the unknown parameters by their estimates, and noise by the estimated noise terms, it is potentially applicable to many other statistical problems. The residual bootstrap also has less computational requirements in some problems. Also, in more general problems, there is some ambiguity about the notion of ‘data pairs’, but the paired bootstrap heavily depends on the definition of ‘data pairs’.

All these lead to preferring the residual bootstrap over the paired bootstrap, and it seems that its only shortcoming is that it is not consistent under heteroscedasticity. This issue was addressed early by Wu (1986), Liu (1988) and Mammen (1992a,b,1993,1995), and they came up with a modification of the usual residual bootstrap, called the external or wild bootstrap.

It may be pointed out that the various advantages of using residual based technique has lead to a default understanding that for any problem at hand, a bootstrap technique must mean a residual based technique. Yet it is not necessarily true that a residual based technique is simpler or better than non-residual based techniques. Even for the simple example of least squares in linear regression, using uncentered residuals would generally result in incorrect results. Modifications required for problems like that of M -estimators are often not very simple (see Lahiri, 1992).

Further, we have already commented on the fact that the classical residual bootstrap is not consistent under heteroscedasticity. In our framework, since we deal with conditional heteroscedasticity, we have to be careful about the possible inconsistencies in the use of residual based bootstrap. For this reason, we look at an appropriate generalization of data-pair bootstrap, as suggested by Chatterjee and Bose (2000).

Now let us discuss how the propose bootstrap is a “generalization”. Suppose

X_1, \dots, X_n is an *i.i.d.* sample from some probability distribution function F , and $\theta = E(T(X))$ be the parameter of interest. The estimator for θ is the statistic

$$T_n = \frac{1}{n} \sum_{t=1}^n T(X_t)$$

The bootstrap scheme of Efron (1979), is as follows: from the data X_1, \dots, X_n , draw a simple random sample with replacement, say X_1^*, \dots, X_n^* . Now the bootstrap statistic is defined by

$$T_n^* = \frac{1}{n} \sum_{t=1}^n T(X_t^*).$$

Conditional on the data X_1, \dots, X_n , the distribution of the statistic T_n^* typically closely approximates that of T_n . The variance of T_n may also be estimated by using T_n^* . Conditional on the data, the random component in T_n^* is induced only by the scheme of simple random sampling with replacement from the data, or in other words, by the scheme of drawing ‘sample-from-the-sample’ or resample from the data. Theoretically, this is a randomisation scheme whose distribution is completely known, and thus the distribution of T_n^* can be calculated perfectly and completely. However, since this distribution is generally difficult to obtain exactly, a very routine technique used in bootstrap is to approximate it by Monte Carlo technique. This means that the distribution of T_n^* may be approximated by repeatedly drawing (X_1^*, \dots, X_n^*) from (X_1, \dots, X_n) , say a large B number of times, computing T_n^* for every such resample, and finally computing the histogram formed by these B values. This Monte Carlo step is important from the computational viewpoint since it effectively transfers a problem of mathematical complications to a computational one.

Now, suppose $w_{n,t}$ is the number of times the data point X_t is drawn in the bootstrap sample. Then $(w_{n,1}, \dots, w_{n,n})$ follows a *Multinomial* $(n, 1/n, \dots, 1/n)$

distribution. A different way of writing the bootstrap statistic is

$$T_n^* = \frac{1}{n} \sum_{t=1}^n T(X_t)w_{n:t}$$

Clearly, the distribution of T_n^* conditional on the data is dependent only on $(w_{n:1}, \dots, w_{n:n})$.

It may be noted that we might have started by defining the resample T_n^* -statistics to be for some ‘resample weights’ $w_{n:1}, \dots, w_{n:n}$ appropriately chosen. Then the bootstrap would have been obtained as special cases. By doing that, we would potentially be able to do a number of things: (a) in problems where the bootstrap is known to be inconsistent, weights from some other distribution might possibly achieve consistency; (b) even where the bootstrap is applicable, in order to get ‘the best’ resample estimate one may use this broader class and get resampling approximations that are better according to some notion of efficiency; (c) in practical problems, computations with weights other than those obtained from the bootstrap may be faster and easier; (d) in some problems, the algebra might be easier or more insightful while working with general weights than for particular choices implied by the usual bootstrap; (e) one gets a deeper insight into the technical and philosophical aspects of why does the bootstrap or jackknife work in some problems but not in others; (f) it yields unified way of studying several resampling schemes simultaneously in a given problem, including typically the jackknives and most of the variations of the bootstrap found in literature.

We refer to resampling with weights $w_{n:1}, \dots, w_{n:n}$ as *generalised bootstrap*. Other names by which this general form of resampling is referred to in literature is the *weighted bootstrap* (cf. Barbe and Bertail (1995)) since the emphasis is on the ‘bootstrap weights’ $w_{n:i}$ ’s, and the *exchangeable bootstrap* (Praestgaard and Wellner (1993)) since it is customary to assume that for every fixed n , the weights $w_{n:1}, \dots, w_{n:n}$ are exchangeable. The idea of bootstrapping

with random weights probably appeared first in Rubin (1981). Bootstrapping with exchangeable weights have been treated in Efron (1982), Lo (1987), Weng (1989), Zheng and Tu (1988), Praestgaard and Wellner (1993). Other generalised bootstrap methods may be found in Boos and Monahan (1986), Lo (1991), Hardle and Marron (1991), Mammen ((1992a) and (1993)). Statistical research with generalised bootstrap has largely been dominated by an approach based on empirical processes and weighted empirical processes. Important contributions in this direction include Csorgo and Mason (1989), Gine and Zinn (1989, 1991), Koul (1992), Mason and Newton (1992), Praestgaard and Wellner (1993), Mammen (1996).

Now, in the regression context, direct bootstrap to the observations and bootstrap to the residuals, called the residual based-bootstrap, are widely used. However both the procedures have their respective merits and demerits. For example, naive resampling in a naive fashion from the data may lead to inconsistency in many situations, (see Shao and Tu, 1995 for examples of such cases). Furthermore, the direct method of resampling the data does not embody all the information used in the residual based approach (see in this context Veall, 1987; and Li and Maddala, 1996), and as a results performance of direct bootstrap to the data may be worse than residual-based approach. On the other hand, typical residual based resampling techniques are not robust and may even be inconsistent. For example, the classical residual bootstrap for the least squares estimator in linear regression is inconsistent under heteroscedasticity. Similarly, in the dynamic linear regression set-up, the typical residuals-based estimates produce inconsistent estimates if the structure of the correlation of the residuals is not tractable or misspecified. Furthermore, performance of these two methods are not known in the context of dynamic nonlinear regression set-up with non-normal and conditionally heteroscedastic error. In such situations, generalised bootstrap for estimating equation (*cf.*

Chatterjee and Bose, 2000) may be extremely useful.

Our work in this chapter is mainly concerned with a new class of generalised bootstrap techniques as originally advocated Chatterjee and Bose (2000). They suggested that for estimators obtained by solving

$$\sum_{t=1}^n \psi(X_t, \theta) = 0,$$

the appropriate generalization of the data pair bootstrap is to use

$$\sum_{t=1}^n w_{n:t} \psi(X_t, \theta) = 0,$$

where $(w_{n:1}, w_{n:2}, \dots, w_{n:n}) \sim \text{multinomial}(n, 1/n, \dots, 1/n)$. In the context of regression this new class appropriately captures the idea of a generalised bootstrap based on data pairs. In Section (3.4) we show that the new generalised bootstrap is typically consistent.

3.3 Asymptotic Properties Based on Estimating Equations Approach

In this section we derive the consistency and asymptotic normality of the estimators of the parameters of GARCH(p, q)-M model ¹. As discussed in the "Introduction" of this chapter, we adopt estimating equation approach of Godambe (1985) which is a very general approach in the sense that this has the strengths of both ML and LS methods—the two methods of estimations most frequently used and applied in a variety of statistical applications—and weaknesses of neither.

The basic GARCH(p, q)-M model is specified as

$$y_t = x_t' \beta + \lambda h_t + \epsilon_t, \tag{3.3.1}$$

¹We consider the most general set-up in this class of models, of which the usual ARCH-M is a special one

where

$$E[\epsilon_t | \Psi_{t-1}] = 0,$$

$$E[\epsilon_t^2 | \Psi_{t-1}] = h_t,$$

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \dots + \alpha_q \epsilon_{t-q}^2 + \phi_1 h_{t-1} + \dots + \phi_p h_{t-p}. \quad (3.3.2)$$

and

$$\alpha_0 > 0, \alpha_i \geq 0, i = 1, \dots, q, \phi_i \geq 0, i = 1, 2, \dots, p, \sum_{i=1}^q \alpha_i + \sum_{i=1}^p \phi_i < 1.$$

In (3.3.1) y_t, x_t, Ψ_t have their usual meanings *viz.*, y_t is the value of the dependent variable at time t , x_t is $(k \times 1)$ vector of independent (including the lagged dependent) variables at time t and Ψ_{t-1} is the increasing sequence of σ -field generated by $\{x_{t-1}, x_{t-2}, \dots, \epsilon_{t-1}, \epsilon_{t-2}, \dots\}$. It may be noted that we have not made any assumption regarding the conditional or unconditional distribution of ϵ_t . Further, while the errors of the linear regression equation are serially uncorrelated, they are not independent as they are related through higher moments. Setting $\lambda = 0$ implies that the conditional variance does not have any effect on the dependent variable y_t , and in such a case the equation in (3.3.1) is the standard regression set-up with errors following an ARCH/GARCH process.

In order to derive the asymptotic properties, we use a more general notation:

$$y_t = f(x_t, \underline{\epsilon}_t, \theta) + \epsilon_t, \quad (3.3.3)$$

where $\underline{\epsilon}_t = (\epsilon_{t-1}, \epsilon_{t-2}, \dots)'$, $\theta = (\theta_1, \theta_2, \dots, \theta_l)' \in \Theta \subseteq R^l$ represents the entire parameter vector i.e., $\theta = (\beta', \lambda, \alpha', \phi)'$, $l = k + p + q + 1$, and

$$f : R^k \times R^\infty \times R^l \rightarrow R^1.$$

Similarly, conditional variance h_t , in (3.3.2) may be represented as $h_t = h(\underline{\epsilon}_t, \theta)$

where $h(\cdot)$ is a function defined by

$$h : R^\infty \times R^l \rightarrow (0, \infty).$$

As regards h , we have only assumed that it is a positive function. So long as this holds choice of $h(\cdot)$ is an empirical issue.

We now define

$$u_t = \epsilon_t / \sqrt{h(\epsilon_t)}. \quad (3.3.4)$$

The following assumptions will be made throughout.

A3.1 : u_t is a random variable with mean 0 and variance 1. It may be noted that the second moment for ϵ_t may not exist, the same for u_t still may be finite (see Lumsdaine, 1996).

A3.2: Let ψ be $\ni \psi : R \times \Theta \rightarrow R^l$ and there exists $\theta_0 \in \Theta$ s. t.

$$E[\psi(u_t, \theta_0) | \Psi_{t-1}] = 0 \quad \forall t = 1, 2, \dots, n.$$

Further assume that $\psi = (\psi_1, \psi_2, \dots, \psi_l)'$ is twice differentiable.

A3.3: $k_1 > \limsup_n \frac{1}{n} [\sup_{\|c\|=1} \sum_{t=1}^n E\{c' \psi(u_t)\}^2] (= K_1, \text{ say}) > 0$, k_1 and K_1 are constants. This assumption implies the ψ is non-degenerate.

We now define $G_t = \frac{\partial \psi(u_t, \theta_0)}{\partial \theta}$, a $l \times l$ matrix, and then G as $G = \frac{1}{n} \sum_{t=1}^n G_t$.

A3.4: $\sum_{t=1}^l \sum_{t'=1}^l E[G_{tt'}^2] = O(1)$.

This is essentially a moment condition and it essentially provides restriction on the parameter space.

A3.5: Fix a^{th} co-ordinate from R^l . Let H_{a_t} be the Hessian corresponding to $\psi_a(u_t, \theta)$. Assume there exists $\delta_0 > 0$ s.t. $\sup_{\|c\| < \delta_0} \sum_{t=1}^n \sum_{a=1}^l E \lambda_{\max}^2 [H_{a_t}(\theta_0 + c)] = O(n)$.

A3.6: As $E(G)$ is not necessarily a symmetric matrix, define a symmetric matrix F as $F = [E(G) + E(G)']$. Assume $\lambda_{\min}[F] = K_2 > 0$, K_2 is a constant. Also assume G is a full rank matrix.

Define

$$X_{nt} = v_n^{-1} \left[\left(\sum_{t=1}^n E(G_t) \right)^{-1} c \right]' \psi(u_t, \theta_0),$$

$t = 1, \dots, n$. v_n is defined in Appendix A3.1.

A3.7: Assume that X_{nt} satisfies

$$\max_t |X_{nt}| \xrightarrow{p} 0, \sum_{t=1}^n X_{nt}^2 \xrightarrow{p} 1, \text{Sup}_n E(\max_t X_{nt}^2) < \infty.$$

Now, in the estimating equations approach, the estimate of θ , $\hat{\theta}_n$, is obtained by solving the following estimating equation:

$$\sum_{t=1}^n \psi(u_t, \theta) = 0. \quad (3.3.5)$$

We now define $I = E(G)$, $J = \frac{1}{n} E[\sum_{t=1}^n \psi(u_t) \psi(u_t)']$ and $V = [I^{-1} J I^{-1}]$ at $\theta = \theta_0$. The following theorem establishes the existence of a consistent root of the estimating equation in 3.3.5, which is also asymptotic normal.

Theorem 3.1 *Under assumptions A3.1-A3.7, there exists a sequence of solutions $\{\hat{\theta}_n\}$ of the equation 3.3.5 s. t. $\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow N(0, V)$.*

Proof: See Appendix A3.1.

In the context of maximum likelihood estimation, if the likelihood is assumed correctly, then $J = -I$ i.e., the Hessian matrix (negative) is same as outer product matrix and in that case the asymptotic variance covariance matrix of $\hat{\theta}_n$ reduces to either J^{-1} or equivalently to $-I^{-1}$. On the contrary, if the distributional assumption is not valid, then V is the variance-covariance matrix of the estimate based on inappropriate likelihood function, and $\hat{\theta}_n$ is which is then called the quasi maximum likelihood estimate (QMLE) of θ .

It may be remarked that the theory of estimating equations (and its bootstrap) is not dependent on this particular set of assumptions only. Other alternative assumptions might have been used as well; for example, one could have used Cramer's condition and related assumptions, or convexity related

assumptions. Our set of assumption is one of many such possibilities, which may be more convenient in ARCH-M context. Although convexity related assumptions are often found useful for estimation from log-concave likelihoods, in ARCH-M models the presence of scale parameters makes these assumptions somewhat less convenient.

3.4 Derivation of Bootstrap Distribution

In the last section we have derived consistency and asymptotic normality of the estimators of the parameters of GARCH-M model, and accomplishing this result, it is possible to obtain the asymptotic distribution of different test statistics of relevance. However, in reality, for a given sample size, the distribution of a test statistic can be substantially different from that based on asymptotic theory. While some methods like Edgeworth expansions for analytically obtaining small sample results are available, these involve a lot of tedious algebra, and those are applicable only in some special cases. The bootstrap method initiated by Efron (1979) provides a viable alternative and this has proved to be extremely useful in various practical situations. One of the objectives of deriving the bootstrap distribution is that it may provide a consistent and sometimes more accurate distribution of the underlying test statistics.

Bose and Chatterjee (2001) have developed a theory of bootstrap which is compatible with the framework of Sections 3.3. Let w_1, \dots, w_n be random weights, satisfying certain technical conditions as stated in Bose and Chatterjee (2001).

The bootstrap estimator from the estimating equations is the solution of For the estimating equations problem, the corresponding bootstrap estimator

is the solution of

$$\sum_{t=1}^n w_t \psi(u_t, \theta) = 0, \quad (3.4.1)$$

where $\psi(u_t, \theta)$ is as defined earlier. The concept of resampling with (3.4.1) may be traced back to Freedman and Peters (1984) and Rao and Zhao (1992). A distinguishing feature of resampling with (3.4.1) is that these are bootstrap using equations as opposed to the more traditional approach of resampling using observations.

Obviously, the choice of "bootstrap weights" i.e., w_t 's is crucial. In fact, Bose and Chatterjee (2001) have shown that the paired bootstrap, the Bayesian bootstrap (Rubin (1981)) and the weighted likelihood bootstrap (Newton and Raftery (1994)) can be obtained from their set-up by different choices of bootstrap weights.

Now, suppose w_1, \dots, w_n follow a multinomial distribution with n cells and equal probabilities of $1/n$ each. That is,

$$Pr[w_1 = k_1, \dots, w_n = k_n] = \frac{n!}{k_1! \dots k_n!} n^{-n}$$

whenever $\sum_{t=1}^n k_t = n$, and zero otherwise. Such weights correspond to Efron's classical bootstrap.

Let $F_n(x) = Pr[n^{1/2}(\hat{\theta}_n - \theta_0) \leq x]$, and assume that $F_{nB}(x) = Pr[n^{1/2}(\hat{\theta}_B - \hat{\theta}_n) \leq x]$, where $\hat{\theta}_B$ is the bootstrap estimate of θ conditional on data. Then we have the following theorem.

Theorem 3.2 *In the framework of Theorem 3.1, F_{nB} is a consistent estimator of F_n .*

Formal proof of this theorem is given in Appendix A3.2.

It may be noted that in the context of bootstrap, consistency means that the supremum of differences between F_{nB} and F_n converges to zero in probability.

3.5 Simulation Results

The consistency of bootstrap distribution, loosely speaking, asserts that the bootstrap distribution would be as close as possible to the actual distribution as the sample size goes to infinity. It is also to be noted that the bootstrap distribution could be considered to be an approximation of the actual distribution. Hence, it is of interest to study the accuracy of the bootstrap distribution, and this simulation study intends to do so in terms of performance by confidence interval estimate. In other words, through this Monte Carlo exercise, we study the performance of bootstrap distribution and the asymptotic normal approximation with respect to confidence level or what is also called, coverage probability.

Now, in order to obtain the interval estimate, we need standard error of the estimate concerned. However, since it is not possible to derive the exact distribution of the parameter estimate of ARCH-M model the large sample or the asymptotic distribution is used as an approximation of the distribution. The accuracy of such asymptotic distribution-based confidence interval crucially depends on how good the approximation is. Bootstrap distribution may produce better approximation of the standard error and for that matter of the underlying sampling distribution. Accordingly, more accurate confidence level or coverage probability may be attained by bootstrap distribution.

3.5.1 Results Based on Asymptotic Approximation

We now present the results of a simulation study which has been undertaken to compare the performance of generalized bootstrap and the usual normal asymptotics. First, we discuss the accuracy of the confidence interval for λ

Table 3.5.1: Estimated Coverage Probability

Nominal Confidence Level	Sample Size					
	100	200	300	400	500	1000
90%	0.714	0.756	0.792	0.800	0.804	0.828
95%	0.750	0.796	0.830	0.837	0.848	0.876
99%	0.776	0.850	0.860	0.871	0.886	0.904

² using asymptotic normal approximation. To this end, necessary data have been generated and relevant computations carried out using a program written in GAUSS. Now, we generated y_t 's from the following GARCH-M model

$$y_t = 0.165h_t + \epsilon_t, \quad (3.5.1)$$

$$h_t = 7.54 + 0.56\epsilon_{t-1}^2 + 0.018h_{t-1} \quad (3.5.2)$$

where $\epsilon_t/\sqrt{h_t}$ were obtained as random normal deviate from GAUSS random number generator. For this study, samples of varying sizes (n) have been considered; more specifically, n was taken to be 100, 200, 300, 400, 500, and 1000, and for each of these samples 1000 replications were used. The estimate of the standard error of $\hat{\lambda}$ was obtained from the appropriate element of the Hessian matrix. The performance of the asymptotic distribution was examined through the simulated coverage probability of the interval, and the coverage probability was calculated by counting how many of the 1000 intervals constructed corresponding to 1000 replications contained the true value of λ which is 0.165 for this model. These coverage probabilities are presented in Table 3.5.1.

²Since λ is the most important parameter from consideration of this study, we discuss results concerning this parameters only.

It is evident from Table 3.5.1 that as far as coverage probability is concerned, the performance of the asymptotic approximation based confidence interval is quite poor. However as the sample size increases, the performance gradually improves but it still remains below the nominal level even when the sample size is as large as 1000.

3.5.2 Results Based on Application of Bootstrap Methodology

As we have already discussed, bootstrap technique may be used to improve upon this coverage probabilities. We construct different confidence intervals for λ based on bootstrap distribution. These are : bootstrap-t, bootstrap percentile and bias corrected bootstrap. These methods were proposed by Efron and Tibshirani (1986). Obviously, all these methods have their own advantages and limitations. The bootstrap-t (BT) method, according to another study by Efron and Tibshirani (1993), has good coverage probabilities, but this method is not reliable in practice. While bootstrap-percentile (BP) is more reliable it does not usually yield satisfactory coverage probabilities. Hence, Efron and Tibshirani (1993) proposed another method which essentially corrects the bias in BP method. This again has two versions– the bias corrected (BC) and the bias corrected accelerated (BCA) versions. This last method having two versions have been shown to perform better than both BT and BP methods. Now, the computations involved in finding the acceleration coefficient for BCA is complicated, and hence we have considered the remaining three methods *viz.*, BT, BP, and BC methods for this bootstrap exercise.

We now briefly describe the three methods. Let $\hat{\lambda}$ denote estimate of λ based on full sample, and $\lambda_1^*, \dots, \lambda_B^*$ the bootstrap estimates of λ correspond-

ing to B inner replications.

Bootstrap-t

The bootstrap-t is defined as

$$BT_i = \frac{\lambda_i^* - \hat{\lambda}}{\sigma^*}, i = 1, 2, \dots, B,$$

where σ^* is the standard error of λ^* obtained from the bootstrap estimates.

The bootstrap-t confidence interval is then given by

$$[\hat{\lambda} - \hat{t}_{1-\alpha}\sigma^* \quad \hat{\lambda} + \hat{t}_\alpha\sigma^*],$$

where \hat{t}_α and $\hat{t}_{1-\alpha}$ are the α and $(1 - \alpha)$ percentiles, respectively, of the empirical distribution of BT .

Bootstrap Percentile

The bootstrap percentile is defined as

$$[\lambda_\alpha^* \quad \lambda_{(1-\alpha)}^*],$$

where $\lambda_\alpha^*, \lambda_{(1-\alpha)}^*$ are the 100. α -th and 100. $(1 - \alpha)$ -th, values, respectively, of the empirical distribution of $\lambda_1^*, \dots, \lambda_B^*$. Thus, in practice, after estimating the value of λ for each of the B bootstrap inner replications, we order them, and then take the 100. α -th ordered value as the lower limit of the interval and the 100. $(1 - \alpha)$ -th ordered value as the upper limit of the interval.

Bias Corrected

This method also uses the percentiles of the bootstrap distribution, but not exactly the 100. α -th and 100. $(1 - \alpha)$ -th ordered values as the limits of the interval. Instead, it corrects these values for possible bias in the estimation of λ through a quantity, which, in fact, is the median bias of $\hat{\lambda}$. The bias corrected interval is then defined as

$$[\lambda_{\alpha_1}^* \quad \lambda_{(1-\alpha_2)}^*],$$

Table 3.5.2: Estimated Coverage Probability (Sample Size 100)

	Confidence Intervals		
Nominal Confidence	BT	BP	BC
90%	0.79	0.85	0.85

Table 3.5.3: Estimated Coverage Probability (Sample Size 200)

	Confidence Intervals		
Nominal Confidence	BT	BP	BC
90%	0.81	0.84	0.87

where $\alpha_1 = \Phi(2z_0 + z_\alpha)$ and $\alpha_2 = \Phi(2z_0 + z_{(1-\alpha)})$. The function Φ is the cumulative distribution function of a standard normal variate, z_α is its 100. α th percentile point, and z_0 represents the bias correction term. The value of z_0 , say, \hat{z}_0 , is obtained directly as the proportion of the bootstrap replications where λ^* has value which is less than $\hat{\lambda}$, i.e.,

$$\hat{z}_0 = \Phi^{-1}\left(\frac{b}{B}\right),$$

where B is the number of bootstrap replications having $\lambda^* < \hat{\lambda}$. To study the performance of the proposed bootstrap methods a Monte Carlo experiment was performed with replication 500, and for each replication of the experiment 100 bootstrap replications (called, B) were considered. Sample size has been taken as 100 and 200. As such experiments are prohibitively time consuming, the number of replications as well as inner replications have been kept at moderate values.

Tables 3.5.2 and 3.5.3 provide the insight that bootstrap can substantially improve the performance of the coverage probability of the confidence intervals

³. Further, as expected, BC performed better than BT and BP methods.

3.6 An Empirical Illustration

We now report the results of an application of bootstrap methodology on an actual data set. The exercise has been carried out with weekly returns based on SENSEX covering the period 1st week January 1984 to 3rd week of October 2000. The details of the data set are given in Section 2.2 of Chapter 2. All computations were carried out using program written in GAUSS.

We may recall from Section (2.3) of the previous Chapter that only first-order autocorrelation was found to be significant. We, therefore, fitted an ARMA model to this data. The model turned out to be an AR(1) model as given below.

$$y_t = -0.212y_{t-1} + \epsilon_t, \quad l(\hat{\theta}) = -2663.718. \quad (3.6.1)$$

(-6.397)

(t-ratio is given in parentheses)

³We have tried a comparative study of the different bootstrap methodologies *viz.*, parametric bootstrap, residual bootstrap and the proposed generalized bootstrap, in some simulation experiments, the results of which are not reported here since these are of marginal interest in the present context. Broadly, the results obtained show that the parametric bootstrap confidence intervals are of smallest length if the likelihood is correctly specified. However, the confidence intervals obtained by the residual bootstrap and the proposed generalized bootstrap are only marginally longer than that obtained by the parametric bootstrap, typically differing only in the second decimal place for λ .

The coverage accuracies of all these bootstrap techniques are more or less the same. In view of the above facts, we have concentrated only on the proposed generalized bootstrap for which we have results on consistency and known robustness properties.

Now, in order to test for the presence of conditional heteroscedasticity in the form of ARCH/GARCH process, the squared residuals were regressed on its own different lags upto 25, and it was found by Engle's (1982) χ^2 test for testing the ARCH effect, that all the test statistic values were significant at 1 per cent level of significance suggesting thereby the presence of conditional heteroscedasticity in the returns. We, therefore, analysed the returns in AR(1)-ARCH framework, and the estimated model thus obtained is given by

$$y_t = -0.490y_{t-1} + \epsilon_t, \quad l(\hat{\theta}) = -2508.625, \quad (3.6.2)$$

(-16.6)

$$h_t = 19.2 + 0.57\epsilon_{t-1}^2. \quad (3.6.3)$$

(37.3) (3.1)

(t-ratios are given in parentheses)

All the parameters in the two equations are highly significant at 5 per cent level of significance. In terms of maximum log-likelihood value $l(\hat{\theta})$ also, it is evident that there has been a very significant improvement after ARCH effect has been explicitly incorporated in the model.

Having recognized that ARCH effect is very strong in the data, we now discuss about the suitability of ARCH-M model for this data set. After necessary computations in GAUSS, we found the estimated GARCH-M model as

$$y_t = -0.457y_{t-1} + 0.165h_t + \epsilon_t, \quad l(\hat{\theta}) = -2490.99, \quad (3.6.4)$$

(-13.5) (4.7)

$$h_t = 7.54 + 0.56\epsilon_{t-1}^2 + 0.018h_{t-1}. \quad (3.6.5)$$

$$(20.5) \quad (2.73) \quad (2.59)$$

(t-ratios are given in parentheses)

As before, all the parameters are highly significant at 5 per cent level of significance with maximized log-likelihood value being -2490.99. It is evident from Table 3.6.4 that both the standardized as well as the squared-standardised residuals are serially uncorrelated as evidenced from the values of Ljung-Box test statistics $Q(\cdot)$ and $Q^2(\cdot)$. We also note that for the fitted ARCH-M model, GARCH(1,1) process, as shown in (3.6.5), has been found appropriate and adequate.

The significance of (G)ARCH-M model parameters has so far been studied using the asymptotic (Gaussian) distribution. However, as pointed out earlier, a bootstrap based test is an alternative to the asymptotic test. In general, two basic advantages can result from a bootstrap procedure. First, the bootstrap allows one to devise a valid inference in problems where classical methodology is either very complicated or does not provide any solution at all. A second advantage is that, when applied properly, bootstrap confidence interval and test often have smaller level of error than that based on asymptotic theory (see Beran, 1987; and Hall, 1988). Insofar as testing of hypothesis using bootstrap method is concerned, two important issues are relevant. One, how to generate the bootstrap samples? Two, what test statistic to bootstrap? The first question has been addressed to in the previous section. As for the answer to the second question, some discussions are needed.

Suppose we want to test the null hypothesis $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$. Given an estimator $\hat{\theta}$ of θ , the usual test procedure would be based

on $T = \hat{\theta} - \theta_0$, and the inference is obtained based on the distribution of T . Now, straightforward application of bootstrap procedure suggests the use of $T^* = \theta^* - \theta_0$ in the place of T , where θ^* is the bootstrap based estimate of θ . However, Beran (1987) and Hall (1988) have found that the power of T^* is very poor, specially when samples are generated from a distribution with parameter θ being far away from θ_0 . Hall and Wilson (1991), therefore, considered another distribution based on the empirical distribution of $(\theta^* - \theta_0)$. They also recommended some guidelines for hypothesis testing, viz., to use either $(\theta^* - \hat{\theta})$ or $\frac{(\theta^* - \hat{\theta})}{\hat{\sigma}^*}$ but not $(\theta^* - \theta_0)$ nor $\frac{(\theta^* - \hat{\theta})}{\hat{\sigma}}$, where $\hat{\sigma}^*$ stands for the estimate of standard error of T , based on bootstrap sample where as $\hat{\sigma}$ is the full sample based estimate of the standard error of T . They also found superiority of $\frac{(\theta^* - \hat{\theta})}{\hat{\sigma}^*}$ as compared to $(\theta^* - \hat{\theta})$. Here we follow the Hall and Wilson(1991) recommendation. Repeated sampling allows one to approximate the bootstrap distribution of $(\theta^* - \hat{\theta})$ and $\frac{(\theta^* - \hat{\theta})}{\hat{\sigma}^*}$, and to conduct a test of the relevant hypothesis. For a test of level α , the null hypothesis $H_0 : \lambda = 0$ is rejected in favour of $H_1 : \lambda \neq 0$, if $\hat{\lambda}$ is greater than the $(1 - \alpha)$ percentile of the bootstrap distribution of $(\lambda^* - \hat{\lambda})$. Similarly, based on studentised t-test, the null hypothesis of $H_0 : \lambda = 0$ is rejected in favour of $H_1 : \lambda \neq 0$, if $\hat{\lambda}/\hat{\sigma}$ is greater than the $(1 - \alpha)$ percentile of the bootstrap distribution of $\frac{(\lambda^* - \hat{\lambda})}{\hat{\sigma}^*}$.

As far as the bootstrap distribution is concerned we confined ourselves only to the most important parameter, i.e., the "relative risk aversion" parameter. To find out the bootstrap distribution of λ , 1000 bootstrap estimates of the parameter have been obtained using multinomial based weight as described in the earlier section. From the bootstrap distribution of $(\lambda^* - \hat{\lambda})$ and $\frac{(\lambda^* - \hat{\lambda})}{\hat{\sigma}^*}$, the 95 per centile critical values have been computed as 0.123 and 3.1, respectively. Based on bootstrap distribution of both the above thus mentioned tests, we find that the null hypothesis of $H_0 : \lambda = 0$ is rejected in favour of $H_1 : \lambda \neq 0$, at 5 per cent level of significance.

3.7 Conclusions

In this chapter we have first proved the consistency and asymptotic normality of the estimators of the parameters of GARCH(p, q)-M model obtained by using the estimating equation approach. We have then applied generalized bootstrap to ARCH-M model, and then proved the consistency property of the resulting bootstrap distribution of the estimators of the parameters involved. The results of a detailed simulation exercise for studying the performance of bootstrap as well as the usual normal asymptotic approximation have thereafter been reported and discussed. It has been found that insofar the relative risk aversion parameter is concerned, bootstrap distribution outperforms the asymptotic approximation in terms of confidence level or coverage probability. Finally, by way of an illustration with real data, we have considered weekly returns based on SENSEX. We have found the suitable GARCH-M model for this data set, then obtained the bootstrap distribution of the relative risk aversion parameter of this model, and used it to conclude that this parameter is highly significant.

Appendix A3.1

Proof of Theorem 3.1.

Note that given any $\epsilon > 0, \exists K > 0$ and an integer n_0 such that $\forall n \geq n_0,$
 $Prob \left[\left\| \frac{1}{n} \sum_{t=1}^n \psi(u_t, \theta_0) \right\| > K \right] < \epsilon.$

This is established using Chebyshev's inequality and using assumptions A3.2 and A3.3. Let us now define $\psi(u_t, \theta) = (\psi_1(u_t, \theta), \dots, \psi_l(u_t, \theta))'$.

Let us assume that for each $\psi_a(u_t, \theta)$ the following Taylor series expansion holds:

$$\psi_a(u_t, \theta + \iota) = \psi_a(u_t, \theta) + G^a(u_t, \theta)' \iota + \frac{1}{2} \iota' H_a(u_t, \theta_1) \iota, \quad (3.7.1)$$

where $\theta_1 = \theta_0 + c\iota$ for $0 < c < 1$ and $(G^a)^{l \times 1} = \frac{\partial \psi_a}{\partial \theta}, a = 1, \dots, l.$

Define

$$S_n^{l \times 1}(\iota) = n^{-1/2} \sum_{t=1}^n [\psi(u_t, \theta_0 + n^{-1/2} \iota) - \psi(u_t, \theta_0)] - E[G'] \iota. \quad (3.7.2)$$

Now using above Taylor series expansion

$$S_n(\iota) = n^{-1/2} \sum_{t=1}^n [G'_t(\theta_0) \frac{\iota}{\sqrt{(n)}} + \frac{1}{2n} \sum_{a=1}^l (\iota' \otimes e_a) H_a(u_t, \theta_1) \iota] - E(G') \iota, \quad (3.7.3)$$

where e_a is a $l \times 1$ vector whose a^{th} element is 1 and remaining elements are zeros, \otimes denotes the Kronecker product.

Let $M_1 = \sum_{t=1}^n [G'_t(\theta_0) - EG'_t(\theta_0)][G'_t(\theta_0) - EG'_t(\theta_0)]'$. Note that the assumption A3.4 implies $E\lambda_{max}(M_1) = O(n)$. Now we have

$$\|S_n(\iota)\|^2 \leq 2n^{-2} \|\iota\|^2 \lambda_{max}(M_1) + \frac{1}{2} n^{-3} \|\iota\|^4 \sum_t \sum_{j=1}^n \sum_{a=1}^l \lambda_{max}(H_a(u_t)) \lambda_{max}(H_a(u_j)).$$

Using A3.4 and A3.5

$$E[sup_{\|\iota\| \leq c} \|S_n(\iota)\|]^2 = O\left(\frac{1}{n}\right). \quad (3.7.4)$$

Now we get

$$n^{\frac{1}{2}} t' \psi(\theta_0 + n^{-1/2} t) = t' S_n(t) + n^{-1/2} \psi(\theta_0) + t' E[G'] t, \quad (3.7.5)$$

which again implies

$$\inf_{\|t\|=c} n^{-\frac{1}{2}} t' \sum_{t=1}^n \psi(u_t, \theta_0 + n^{-1/2} t) \geq -c \sup_{\|t\|=c} \|S_n(t)\| + c^2 l_1 - c n^{-1/2} \left\| \sum_{i=1}^n \psi(u_i, \theta_0) \right\| \quad (3.7.6)$$

where $l_1 = \lambda_{\max}(E(G) + E(G'))$.

Choose c large enough and hence from the above inequality

$$\begin{aligned} & \text{Prob}[\inf_{\|t\|=c} \{n^{-1/2} t' \sum_{t=1}^n \psi(u_t, \theta_0 + n^{-1/2} t)\} > 0] \\ & \geq \text{Prob}[n^{-1/2} \left\| \sum_{t=1}^n \psi(u_t, \theta_0) \right\| + \sup_{\|t\|=c} \|S_n(t)\| < c l_1] \\ & = 1 - \text{Prob}[n^{-1/2} \left\| \sum_{t=1}^n \psi(u_t, \theta_0) \right\| + \sup_{\|t\|=c} \|S_n(t)\| \geq c l_1] \\ & \geq 1 - \text{Prob}[n^{-1/2} \left\| \sum_{t=1}^n \psi(u_t, \theta_0) \right\| > c k_3 / 2] - \text{Prob}[\sup_{\|t\|=c} \|S_n(t)\| \geq c k_3 / 2] \\ & \geq 1 - \epsilon, \end{aligned}$$

for sufficiently large n .

Now note that $\psi(u_t, \theta)$ is a continuous function of θ . Hence by using Theorem 6.4.3 of Ortega and Rheinboldt (1970) that, for fixed $\epsilon > 0$, and for all n sufficiently large, $\exists c$ large enough such that

$\psi(\theta_0 + n^{-1/2} t) = 0$ has a root T_n in $\|t\| < c$ with probability $> 1 - \epsilon$.

Set $\hat{\theta}_n = \theta_0 + n^{-1/2} T_n$.

Hence $\hat{\theta}_n$ is a solution of the equation (3.3.5), which satisfies, for fixed $\epsilon > 0$, $\text{Prob}[n^{1/2} \|\hat{\theta}_n - \theta_0\| < c] \geq 1 - \epsilon, \forall n$, large enough.

This proves the consistency of $\hat{\theta}_n$ for θ_0 . To prove asymptotic normality for $\hat{\theta}_n$, let $c \in R^p$ s.t. $\|c\| = 1$, define $v_n^2 = n(I^{-1}c)' J(I^{-1}c)$. Now from (3.7.2)

we get

$$\sqrt{n}E(G')(\hat{\theta}_n - \theta_0) = n^{-1/2}\left[-\sum_{t=1}^n \psi(u_t, \theta_0)\right] + r_{1n}. \quad (3.7.7)$$

where $E[r_{1n}^2] = O(1/n)$, using (3.7.4) The above equation (3.7.6) implies that

$$\left[\sum_{t=1}^n E(G'_t)\right](\hat{\theta}_n - \theta_0) = -\sum_{t=1}^n \psi(u_t, \theta_0) + r_{1n}.$$

$$\begin{aligned} \Rightarrow v_n^{-1}c'(\hat{\theta}_n - \theta_0) &= v_n^{-1}\left[\left(\sum_{t=1}^n E(G_t)\right)^{-1}c'\right]\sum_{t=1}^n \psi(u_t, \theta_0) + r'_{1n} \\ &= \sum_{t=1}^n X_{nt} + r'_{1n}, \end{aligned} \quad (3.7.8)$$

$$(3.7.9)$$

Hence by Theorem 3.2 of Hall and Heyde (1980), and using A3.7 we get

$$\sum_{t=1}^n X_{nt} \xrightarrow{D} N(0, 1). \Rightarrow v_n^{-1}c'(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, 1).$$

$\Rightarrow \sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, nc'v_n^2c)$ Note that $nc'v_n^2c = I^{-1}JI^{-1}$. Hence the theorem.

Appendix A3.2

Proof of theorem 3.2.

It is well-known that when $w_1, \dots, w_n \sim \text{multinomial}(n, \frac{1}{n}, \dots, \frac{1}{n})$, each w_t follows a binomial distribution. Therefore, $E(w_t) = 1$ and $Var(w_i) (= \sigma_w^2, \text{ say}) = \frac{n-1}{n}$. The correlation coefficient ρ , between any two pair of variables is $\rho = -\frac{1}{n-1}$, and this implies asymptotic uncorrelatedness between any two weight variables, i.e., $\rho = O(\frac{1}{n})$. We now define $W_t = \frac{w_t-1}{\sigma_w}$. Let $P_B[\cdot]$ denote the conditional probability of some function of bootstrap sample given the original sample.

Note that $\sum_{t=1}^n \psi(u_t, \hat{\theta}_n) = 0$. Hence we have for some constant $k > 0$

$$\begin{aligned} P_B[\sigma_w^{-1} n^{-\frac{1}{2}} \|\sum_{t=1}^n w_t \psi(u_t, \hat{\theta}_n)\| > K] \\ &= P_B[n^{-\frac{1}{2}} \|\sum_{t=1}^n W_t \psi(u_t, \hat{\theta}_n)\| > K] \\ &\leq kK^{-2} \frac{1}{n} \sum_{t=1}^n \|\psi(u_t, \hat{\theta}_n)\|^2 \end{aligned}$$

Let us assume that for each $\psi_a(u_t, \theta)$ the following Taylor series expansion holds:

$$\psi_a(u_t, \theta + \iota) = \psi_a(u_t, \theta) + G^a(u_t, \theta)' \iota + \frac{1}{2} \iota' H_a(u_t, \theta_1) \iota, \quad (3.7.1)$$

where $\theta_1 = \theta + c\iota$ for $0 < c < 1$ and $(G^a)^{l \times 1} = \frac{\partial \psi_a}{\partial \theta}$, $a = 1, \dots, l$.

Now on the set $\{\|\hat{\theta}_n - \theta_0\| < \frac{\delta_0}{2}\}$

Appendix A3.2

Proof of theorem 3.2.

It is well-known that when $w_1, \dots, w_n \sim \text{multinomial}(n, \frac{1}{n}, \dots, \frac{1}{n})$, each w_t follows a binomial distribution. Therefore, $E(w_t) = 1$ and $\text{Var}(w_i) (= \sigma_w^2, \text{ say}) = \frac{n-1}{n}$. The correlation coefficient ρ , between any two pair of variables is $\rho = -\frac{1}{n-1}$, and this implies asymptotic uncorrelatedness between any two weight variables, i.e., $\rho = O(\frac{1}{n})$. We now define $W_t = \frac{w_t-1}{\sigma_w}$. Let $P_B[\cdot]$ denote the conditional probability of some function of bootstrap sample given the original sample.

Note that $\sum_{t=1}^n \psi(u_t, \hat{\theta}_n) = 0$. Hence we have for some constant $k > 0$

$$\begin{aligned} P_B[\sigma_w^{-1} n^{-\frac{1}{2}} \|\sum_{t=1}^n w_t \psi(u_t, \hat{\theta}_n)\| > K] \\ &= P_B[n^{-\frac{1}{2}} \|\sum_{t=1}^n W_t \psi(u_t, \hat{\theta}_n)\| > K] \\ &\leq kK^{-2} \frac{1}{n} \sum_{t=1}^n \|\psi(u_t, \hat{\theta}_n)\|^2 \end{aligned}$$

Let us assume that for each $\psi_a(u_t, \theta)$ the following Taylor series expansion holds:

$$\psi_a(u_t, \theta + \iota) = \psi_a(u_t, \theta) + G^a(u_t, \theta)' \iota + \frac{1}{2} \iota' H_a(u_t, \theta_1) \iota, \quad (3.7.1)$$

where $\theta_1 = \theta + c\iota$ for $0 < c < 1$ and $(G^a)^{l \times 1} = \frac{\partial \psi_a}{\partial \theta}$, $a = 1, \dots, l$.

Now on the set $\{\|\hat{\theta}_n - \theta_0\| < \frac{\delta_0}{2}\}$

$$\sum_{t=1}^n \|\psi(u_t, \hat{\theta}_n)\|^2$$

$$\leq k \left[\sum_{t=1}^n \|\psi(u_t, \hat{\theta}_0)\|^2 + \|\hat{\theta}_n - \theta_0\|^2 \sum_{t=1}^n \sum_{j=1}^l \|G_j(u_t, \theta_0)\|^2 \right]$$

$$+ \|\hat{\theta}_n - \theta_0\|^4 \sum_{t=1}^n \sum_{j=1}^l \lambda_{max}^2(H_j(u_t, \theta_1))$$

Thus

$$\begin{aligned} P_B[\sigma_w^{-1} n^{-\frac{1}{2}} \|\sum_{t=1}^n w_t \psi(u_t, \hat{\theta}_n)\| > K] &\leq \frac{kK^{-2}}{n} \left[\sum_{t=1}^n \|\psi(u_t, \hat{\theta}_0)\|^2 \right. \\ &+ \|\hat{\theta}_n - \theta_0\|^2 \sum_{t=1}^n \sum_{j=1}^l \|G_j(u_t, \theta_0)\|^2 \\ &+ \|\hat{\theta}_n - \theta_0\|^4 \sum_{t=1}^n \sum_{j=1}^l \lambda_{max}^2(H_j(u_t, \theta_1)) \\ &= U_k(\text{say}) \end{aligned}$$

Note that $U_k = K^{-2}O_p(1)$, so that for fixed $\delta_1, \delta_2 > 0$, by choosing K large enough, we have

$$Prob[P_B[\sigma_w^{-1} n^{-\frac{1}{2}} \|\sum_{t=1}^n w_t \psi(u_t, \hat{\theta}_n)\| > K] > \delta_1] < \delta_2$$

Following the above Taylor series expansion, we can write

$$\psi_a(u_t, \hat{\theta}_n + \frac{c\iota}{\sqrt{n}}) = \psi_a(u_t, \hat{\theta}_n) + \frac{\sigma_w}{\sqrt{n}} G^a(u_t, \hat{\theta}_n)' \iota + \frac{\sigma_w^2}{2n} \iota' H_a(u_t, \theta_1) \iota,$$

where $\theta_1 = \hat{\theta}_n + \frac{c\iota}{\sqrt{n}}$ for some $0 < c < 1$. Let us define

$$S_{nB}^{l \times 1}(\iota) = \frac{1}{\sigma_w \sqrt{n}} \sum_{t=1}^n w_t [\psi(u_t, \hat{\theta}_n + n^{-1/2} \iota) - \psi(u_t, \hat{\theta}_n)] - [G'(\hat{\theta}_n)] \iota.$$

Note that the a -th element of the vector $S_{nB}(\iota)$ is given by

$$\begin{aligned}
S_{nB_a}(\iota) &= \frac{1}{n} \sum_{t=1}^n w_t G^a(u_t, \hat{\theta}_n)' \iota \\
&+ \frac{\sigma_w}{2n^{\frac{3}{2}}} \iota' \sum_{t=1}^n w_t H_a(u_t, \theta_1) \iota - [G'(\hat{\theta}_n)] \iota \\
&= \frac{\sigma_w}{n} \sum_{t=1}^n W_t G^a(u_t, \hat{\theta}_n)' \iota \\
&+ \frac{\sigma_w}{2n^{\frac{3}{2}}} \iota' \sum_{t=1}^n w_t H_a(u_t, \theta_1) \iota
\end{aligned}$$

On the set $\{\|\hat{\theta}_n - \theta_0\| < \frac{\delta_0}{2}\} \cap \{\|\iota\| \leq C\}$, we thus have for large n

$$\begin{aligned}
\|S_{nB}(\iota)\|^2 &\leq \frac{2\sigma_w^2 C^2}{n^2} \lambda_{\max} \left(\sum_{a=1}^l \sum_{t=1}^n \sum_{j=1}^n W_t W_j G^a(u_t, \hat{\theta}_n) G^a(u_j, \hat{\theta}_n)' \right) \\
&+ \frac{\sigma_w^2 C^4}{n^3} \sum_{a=1}^l \sum_{t=1}^n w_t \lambda_{\max} H_a(u_t, \theta_1) \\
&= T_1 + T_2 (\text{say})
\end{aligned}$$

Let $b_n = \frac{\sqrt{n}}{\sigma_w}$. Note that

$$P_B[b_n \sup_{\|\iota\| \leq c} \|S_{nB}(\iota)\| > 2K] \leq \sum_{j=1}^2 P_B[b_n^2 T_j > K] + O_p\left(\frac{1}{\sqrt{n}}\right)$$

Now on the set $\{\|\hat{\theta}_n - \theta_0\| < \frac{\delta_0}{2}\}$

$$\begin{aligned}
&P_B\left[\frac{2b_n^2 C^2 \sigma_w^2}{n^2} \lambda_{\max} \left(\sum_{a=1}^l \sum_{t=1}^n \sum_{j=1}^n W_t W_j G^a(u_t, \hat{\theta}_n) G^a(u_j, \hat{\theta}_n)' \right) > K\right] \\
&\leq \frac{2b_n^2 C^2 \sigma_w^2}{Kn^2} E_B \lambda_{\max} \left(\sum_{a=1}^l \sum_{t=1}^n \sum_{j=1}^n W_t W_j G^a(u_t, \hat{\theta}_n) G^a(u_j, \hat{\theta}_n)' \right) \\
&\leq \frac{2b_n^2 C^2 \sigma_w^2}{Kn^2} \sum_{a=1}^l \sum_{t=1}^n \text{tr} \left(G^a(u_t, \hat{\theta}_n) G^a(u_j, \hat{\theta}_n)' \right) \\
&\quad + \frac{2b_n^2 C^2 \sigma_w^2}{Kn^2} \rho \text{tr} \left(\sum_{a=1}^l \sum_{t=1}^n \sum_{j=1, t \neq j}^n G^a(u_t, \hat{\theta}_n) G^a(u_j, \hat{\theta}_n)' \right) \\
&\leq \frac{kb_n^2 C^2 \sigma_w^2}{Kn^2} \sum_{a=1}^l \sum_{t=1}^n \|(G^a(u_t, \hat{\theta}_n))\|^2 \\
&\leq \frac{kb_n^2 C^2 \sigma_w^2}{Kn^2} \left[\sum_{a=1}^l \sum_{t=1}^n \|(G^a(u_t, \hat{\theta}_n))\|^2 + \|\hat{\theta}_n - \theta_0\|^2 \lambda_{\max}(H_a(u_t, \theta_1)) \right] \\
&= O_p(1)
\end{aligned}$$

and

$$\begin{aligned}
P_B\left[\frac{b_n^2 C^4 \sigma_w^2}{n^3} \sum_{a=1}^l \left(\sum_{t=1}^n w_t \lambda_{\max}(H_a(u_t, \theta_1))\right)^2 > K\right] \\
&= \frac{b_n^2 C^4 \sigma_w^2}{K n^3} \sum_{a=1}^l E_B\left(\left(\sum_{t=1}^n w_t \lambda_{\max}(H_a(u_t, \theta_1))\right)^2\right) \\
&= \frac{b_n^2 C^4 \sigma_w^2}{K n^3} (n + k \sigma_w^2) \sum_{a=1}^l \sum_{t=1}^n \lambda_{\max}^2(H_a(u_t, \theta_1)) \\
&= O_p(1)
\end{aligned}$$

Thus we have $P_B[b_n \sup_{\|u\| \leq c} \|S_{nB}(\iota)\| > 2K] = O_p(1)$. Now it can be seen that

$$\begin{aligned}
inf_{\|u\|=c} \left\{ \frac{1}{\sigma_w \sqrt{n}} \sum_{t=1}^n w_t \psi(u_t, \hat{\theta}_n + \frac{\sigma_w \iota}{\sqrt{n}}) \right\} \\
\geq -C \sup_{\|u\|=c} \|S_{nB}(\iota)\| \\
+ C^2 \hat{l}_{1n} - \frac{C}{\sigma_w \sqrt{n}} \left\| \sum_{t=1}^n w_t \psi(u_t, \hat{\theta}_n) \right\|
\end{aligned}$$

Where $\hat{l}_{1n} = \lambda_{\min}[\frac{1}{2}(G(\hat{\theta}_n)G(\hat{\theta}_n)')]$ By choosing C large enough, we have that on the set $\{\|\hat{\theta}_n - \theta_0\| < \frac{\delta_0}{2}\}$,

$$\begin{aligned}
P_B[\inf_{\|\iota\|=c} \{ \frac{1}{\sigma_w \sqrt{n}} \iota' \sum_{t=1}^n w_t \psi(u_t, \hat{\theta}_n + \frac{\sigma_w \iota}{\sqrt{n}}) \} > 0] \\
&\geq P_B[\frac{1}{\sigma_w \sqrt{n}} \|\sum_{t=1}^n \psi(u_t, \hat{\theta}_n)\| \\
&\quad + \sup_{\|\iota\|=c} \|S_n(\iota)\| < C l_{1n}] \\
&= 1 - P_B[\frac{1}{\sqrt{n}} \|\sum_{t=1}^n \psi(u_t, \theta_0)\| \\
&\quad + \sup_{\|\iota\|=c} \|S_n(\iota)\| > C l_{1n}] \\
&\geq 1 - P_B[\frac{1}{\sigma_w \sqrt{n}} \|\sum_{t=1}^n w_t \psi(u_t, \theta_0)\| > \frac{C k_1}{2} \\
&\quad - P_B[\sup_{\|\iota\|=c} \|S_n(\iota)\| > \frac{C k_1}{2}]]
\end{aligned}$$

Hence for fixed $\delta_1, \delta_2 > 0$, we have that for C large enough, for all n ,

$$\begin{aligned}
\text{Prob}[P_B[\inf_{\|\iota\|=c} \{ \frac{1}{\sigma_w \sqrt{n}} \iota' \sum_{t=1}^n w_t \psi(u_t, \hat{\theta}_n + \frac{\sigma_w \iota}{\sqrt{n}}) \} > 0] < 1 - \delta_1] \\
\leq \text{Prob}[P_B[\frac{1}{\sigma_w \sqrt{n}} \|\sum_{t=1}^n w_t \psi(u_t, \theta_0)\| > \frac{C k_1}{2}] > \frac{\delta_1}{2}] \\
+ \text{prob}[P_B[\sup_{\|\iota\|=c} \|S_n(\iota)\| > \frac{C k_1}{2}] > \frac{\delta_1}{2}] + O(\frac{1}{n}) \leq \delta_2.
\end{aligned}$$

Now note that $\sum_{t=1}^n w_t \psi(u_t, \theta)$ is a continuous function of θ . Hence by using Theorem 6.4.3 of Ortega and Rheinboldt (1970) that, for fixed $\epsilon > 0$, and for all n sufficiently large, $\exists C$ large enough \ni the bootstrap probability that

$\sum_{t=1}^n w_t \psi(u_t, \theta_n + \sigma_w n^{-1/2} \iota) = 0$ has a root T_n in $\|\iota\| < C$ is $< 1 - \epsilon$ with a probability $< \delta$.

Set $\hat{\theta}_B = \hat{\theta}_n + \sigma_w n^{-1/2} T_n$.

Hence $\hat{\theta}_B$ is a solution of the equation 3.4.1, which satisfies, for fixed $\epsilon > 0, \delta > 0$ $\text{Prob}[P_B[\sigma_w n^{1/2} \|\hat{\theta}_B - \hat{\theta}_n\| < C] < 1 - \epsilon] < \delta, \forall n$, large enough.

The rest of the proof follows along similar lines as the proof of Theorem 3.1 by use of standard bootstrap central limit theorem arguments, as in Hall (1992).

Chapter 4

An ARCH in the Nonlinear Mean (ARCH-NM) Model *

4.1 Introduction

The original ARCH-M model, as proposed by Engle *et al.* (1987) and introduced in Chapter 1 (*cf.* equations (1.1.2) and (1.1.3)), contains in its mean specification a term denoted as $\lambda\sqrt{h_t}$. In financial models, this term represents the risk premium, which is obviously time varying in nature in ARCH-M model. In most of the works on ARCH-M or close relatives of ARCH-M models, the maintained hypothesis is that the risk premium can be expressed as an increasing function of the conditional variance of the asset return, say, $g(h_t)$. While in most applications $g(h_t) = \sqrt{h_t}$ has been used (see, for example, Bollerslev, Engle and Wooldridge, 1988; Domowitz and Hakkio, 1985), Engle *et al.* (1987) observed that $g(h_t) = \ln h_t$ worked better in estimating time varying risk premia in the term structure.

* A paper (jointly with Nityananda Sarkar) containing the materials of this chapter and carrying the same title has been published in Sankhya, (2000), Series B, Volume 62, pp. 327-344.

In fact, they have discussed representation of risk premium as some function of the conditional variance. However, as pointed out by Pagan and Hong(1991), the use of $\ln h_t$ is somewhat restrictive in the sense that for $h_t < 1$, $\ln h_t$ will be negative, and for $h_t \rightarrow 0$, effect on y_t will be infinite. It may be pointed out that it is not enough that risk premium is time varying. Since the expected rate of return will depend on the actual risk associated with decisions about y_t , it is imperative that proper functional forms of h_t are used to represent risk premium.

In fact, Backus and Gregory (1993), in a series of numerical examples, have shown that the relation between risk premium and conditional variance of the excess return can have virtually any shape - it can be increasing, decreasing, flat or even nonmonotonic- depending on the parameters of the economy. Thus, although theory may lead to a monotonic relation between risk premium and conditional variance, it does not guarantee it. There are evidences also (e.g., Glosten, Jagannathan and Runkle,1989; and Harvey, 1989, 1991) that the monotonic relation between risk premium and conditional variance is not uniformly supported by the behaviour of actual prices. Given this somewhat unsatisfactory nature of parametric representation of risk premium which is basically an unobservable variable, some researchers like Pagan and Ullah (1988) and Pagan and Hong (1991) have suggested nonparametric methods. However, there are certain limitations in these methods, and hence as yet these cannot be recommended as standard tools of investigation where risk premium is involved (for details, see Pagan and Hong, 1991).

In the light of all these observations, it is clear that a more general and flexible specification of risk premium is called for. In this paper we propose a generalization of ARCH-M model in which $g(h_t)$ is assumed to have a general functional form as given by the Box-Cox (1964) family of power transformations. It is well-known that this family of transformations encompasses all

other standard functional forms as special cases. In this chapter we study the proposed model in which h_t has the usual ARCH specification as given in (1.1.3). It may be noted that the usual ARCH-M specification is a special case of this generalization. Hence, for any given data the adequacy of the usual ARCH-M model against a class of generalized ARCH-M models as proposed here [to be henceforth referred to as ARCH in the nonlinear mean (ARCH-NM) models] can now be studied.

The plan of this chapter is as follows. In Section 4.2 we describe the proposed ARCH-NM model along with its properties. The estimation of the model is described in Section 4.3. In Section 4.4 the advantages and appropriateness of this generalized approach in ARCH-M modelling is illustrated through the application of the proposed model to SENSEX data. The chapter concludes with some comments in Section 4.5.

4.2 The Proposed Model

We propose a generalization of the usual ARCH-M specification by considering the risk premium in equation(1.1.2) as being $\lambda g(h_t)$ where $g(h_t)$ is defined as the Box-Cox (1964) transformation of h_t , i.e.,

$$\begin{aligned} g(h_t) &= \frac{h_t^\xi - 1}{\xi}, \text{ for } \xi \neq 0 \\ &= \ln h_t, \text{ for } \xi = 0. \end{aligned} \tag{4.2.1}$$

Thus, this generalization of the ARCH-M model allows for various nonlinear representations in the mean of y_t through $g(h_t)$. The proposed ARCH-NM model in the regression set-up is, therefore, represented as

$$y_t = x_t' \beta + \lambda g(h_t) + \epsilon_t, \quad t = 1, 2, \dots, n, \tag{4.2.2}$$

where $g(h_t)$ is as defined in (4.2.1), $\epsilon_t | \Psi_{t-1} \sim N(0, h_t)$ and the conditional

variance h_t has the usual ARCH specification ¹ as stated in (1.1.3), i.e.,

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \dots + \alpha_p \epsilon_{t-p}^2,$$

where $\alpha_0 > 0$, $\alpha_i \geq 0 \forall i = 1, 2, \dots, p$.

It may be noted at this stage that risk premium may also be negative, as discussed by Domowitz and Hakkio (1985) and Lintner (1965). However, unlike these models where this may be so due only to the negative sign of λ , in the proposed ARCH-NM model this may also be due to $g(h_t)$ being negative. It is obvious that the model given by the equation in (4.2.1) and (4.2.2) reduces to the standard ARCH-M model when $\xi = 1$ or $\xi = 1/2$, the resulting constants in the transformation being adjusted through the appropriate regression parameters.

It is evident from the equation (4.2.2) the $\{y_t\}$ is autocorrelated, and hence this property can be used to improve the accuracy of the forecasts. Empirical evidences towards this improvement in prediction in time varying risk premia have been provided by Shiller (1979), and Shiller, Campbell and Schoenholtz (1983) for term structure of interest rates, and by Domowitz and Hakkio (1985), Hodrick and Srivastava (1984) and Kendall (1989) for foreign exchange market. Unfortunately, the exact expressions for unconditional mean, variance and autocovariances are very difficult to obtain. However, some approximate expressions may be obtained for ARCH-NM (1) model by considering Taylor series expansion of $g(h_t)$ upto squared term. For the sake of algebraic simplicity in deriving these expressions, the model for y_t in (4.2.2) is assumed to have no exogenous variable except the constant term so that β is a scalar parameter. The approximate expressions for the unconditional mean, variance and autocovariances of y_t for such a model are stated in the following two results; the derivations are given in Appendix A4.1.

¹Obviously, h_t may as well be a GARCH process

RESULT 1. The approximate expressions for the unconditional mean and variance of y_t are given by

$$E(y_t) = \beta + \lambda \left[\frac{\left(\frac{\alpha_0}{1-\alpha_1}\right)^\xi - 1}{\xi} + (\xi - 1) \left(\frac{\alpha_0}{1-\alpha_1}\right)^{\xi-2} \frac{(\alpha_0\alpha_1)^2}{(1-\alpha_1)^2(1-3\alpha_1^2)} \right] \quad (4.2.3)$$

and

$$\begin{aligned} V(y_t) = & \frac{\alpha_0}{1-\alpha_1} + \lambda^2[(2-\xi)\left(\frac{\alpha_0}{1-\alpha_1}\right)^{2(\xi-1)}] \frac{2(\alpha_0\alpha_1)^2}{(1-\alpha_1)^2(1-3\alpha_1^2)} \quad (4.2.4) \\ & + \frac{\lambda^2}{4}(\xi-1)\left(\frac{\alpha_0}{1-\alpha_1}\right)^{2(\xi-1)} K_0 \\ & + \lambda^2[(\xi-1)\left(\frac{\alpha_0}{1-\alpha_1}\right)^{2\xi-3} - 2\left(\frac{\alpha_0}{1-\alpha_1}\right)^{2\xi-1}] \left\{ \frac{6\alpha_0^3\alpha_1^2(1+2\alpha_1+2\alpha_1^2)}{(1-15\alpha_1^3)(1-3\alpha_1^2)(1-\alpha_1)^2} \right\} \end{aligned}$$

where K_0 is appropriately defined, the expression being given in Appendix A4.1.

RESULT 2. The first-order unconditional autocovariance of y_t can be approximated as

$$\begin{aligned} Cov(y_t, y_{t-1}) = & \lambda^2\left[\left(\frac{\alpha_0}{1-\alpha_1}\right)^{2(\xi-1)}(4-4\xi+\xi^2)\right] K_1 \quad (4.2.5) \\ & + \frac{\lambda^2}{2}\left[\left(\frac{\alpha_0}{1-\alpha_1}\right)^{2(\xi-1)}(\xi-1) - \frac{(\xi-1)^2}{2}\left(\frac{\alpha_0}{1-\alpha_1}\right)^{2\xi-3}\right] K_2 \\ & + \frac{\lambda^2}{2}\left[\left(\frac{\alpha_0}{1-\alpha_1}\right)^{2\xi-3}(\xi-1) - \frac{(\xi-1)^2}{2}\left(\frac{\alpha_0}{1-\alpha_1}\right)^{2\xi-3}\right] K_3 \\ & + \frac{\lambda^2}{4}\left[\left(\frac{\alpha_0}{1-\alpha_1}\right)^{2(\xi-2)}(\xi-1)^2\right] K_4, \end{aligned}$$

where the expressions of K_1, K_2, K_3 and K_4 are given in Appendix A4.1.

The higher order autocorrelations of y_t are very cumbersome to evaluate when Taylor series expansion of $g(h_t)$ upto squared term is considered. However, if one considers only the first-order term of the Taylor series expansion of $g(h_t)$, then it is easy to find that $Corr(y_t, y_{t-k}) (= \rho_k, \text{ say}) = \sum_{i=1}^p \alpha_i \rho_{k-i}, k > p$ for an ARCH-NM (p) model. As for $0 < k \leq p$, no such recursive relation exists for $Corr(y_t, y_{t-k})$ and hence the expressions of these autocorrelations are to be separately obtained for each case.

The expressions of $E(y_t)$, $V(y_t)$, $Cov(y_t, y_{t-1})$ etc. for the standard ARCH and ARCH-M models may be obtained as special cases where $\lambda = 0$ and $\xi = 1$, respectively, in the expressions in (4.2.3), (4.2.4) and (4.2.5). Thus, by substituting $\xi = 1$ in (4.2.3) through (4.2.5), we find that

$$E(y_t) = \beta + \lambda \left(\frac{\alpha_0}{1 - \alpha_1} - 1 \right) \quad (4.2.6)$$

$$V(y_t) = \frac{\alpha_0}{1 - \alpha_1} + \frac{2(\lambda\alpha_0\alpha_1)^2}{(1 - \alpha_1)^2(1 - 3\alpha_1^2)} \quad (4.2.7)$$

and

$$Cov(y_t, y_{t-1}) = \frac{2\lambda^2\alpha_0^2\alpha_1^3}{(1 - \alpha_1)^2(1 - 3\alpha_1^2)}, \quad (4.2.8)$$

which are the corresponding expressions for ARCH-M model where risk premium is represented as $\lambda(h_t - 1)$.

We can interpret $E(y_t)$ in the context of finance models as being the unconditional expected return for holding a risky asset. Since in the absence of a risk premium $V(y_t) = \alpha_0/(1 - \alpha_1)$, the second component in (4.2.7) may be considered to be due to the presence of a risk premium which makes y_t more dispersed. Finally, as we know, ARCH-M effect makes y_t serially correlated and this serial correlation is given by (4.2.8). As for the other special case *viz.*, ARCH model, which can be obtained by having $\lambda = 0$, it immediately follows from (4.2.3) through (4.2.5) that $E(y_t) = \beta$, $V(y_t) = \alpha_0/(1 - \alpha_1)$ and $Cov(y_t, y_{t-1}) = 0$. Obviously, these correspond to the case where risk premium is absent.

4.3 Estimation

The model is estimated by the method of maximum likelihood. The log-likelihood based on n observations and *conditional* on the initial values of all

variables, is given by (omitting the constant)

$$L(\theta \mid \Psi_{t-1}) = \sum_1^n l_t(\theta \mid \Psi_{t-1}), \quad (4.3.1)$$

where

$$l_t(\theta \mid \Psi_{t-1}) = -(\ln h_t)/2 - \epsilon_t^2/2h_t, \quad (4.3.2)$$

$\theta' = (\beta', \lambda, \alpha', \xi)$ is an $1 \times l (= k + p + 3)$ vector of all parameters, and $\alpha' = (\alpha_0, \alpha_1, \dots, \alpha_p)$ is the $1 \times (p+1)$ component vector of coefficients in the ARCH specification.

The first order condition of maximization of $l_t(\theta)$ (omitting henceforth from notation conditional on Ψ_{t-1} , for the sake of notational simplicity) yields

$$\frac{\partial l_t(\theta)}{\partial \theta} = \frac{1}{2h_t} \frac{\partial h_t}{\partial \theta} (\epsilon_t^2/h_t - 1) + \epsilon_t/h_t \frac{\partial \beta'}{\partial \theta} x_t + \lambda \frac{\partial g(h_t)}{\partial \theta} + g(h_t) \frac{\partial \lambda}{\partial \theta}. \quad (4.3.3)$$

It can be verified that $\frac{\partial \lambda}{\partial \theta}$ is an $l \times 1$ vector whose $(k+1)$ th element is one and other elements are zeros; $\frac{\partial \beta'}{\partial \theta}$ is a $l \times k$ matrix of elements of zeros and ones with its first $k \times k$ submatrix being an identity matrix and the last $(l-k) \times k$ submatrix being a null matrix. All other derivatives *viz.*, $\frac{\partial l_t(\theta)}{\partial \theta}$, $\frac{\partial h_t}{\partial \theta}$, $\frac{\partial g(h_t)}{\partial \theta}$ are $l \times 1$ vector each. The evaluation of $\frac{\partial l_t(\theta)}{\partial \theta}$ requires the expressions of $\frac{\partial h_t}{\partial \theta}$ and $\frac{\partial g(h_t)}{\partial \theta}$ and the latter can again be expressed in terms of $\frac{\partial h_t}{\partial \theta}$. However, h_t as well as $g(h_t)$ are all functions of the previous innovations and all these derivatives will be of recursive nature. The required derivatives are thus obtained by the usual assumption that the initial values of $\frac{\partial h_t}{\partial \theta}$ and ϵ_t 's do not depend upon the parameters. All the recursive relations are given in Appendix A4.2.

Now let $S_{ti} = \frac{\partial l_t(\theta)}{\partial \theta_i}$, θ_i being the i -th element of the parameter vector θ .

Then

$$\frac{\partial L(\theta)}{\partial \theta} = \sum_t \frac{\partial l_t(\theta)}{\partial \theta} = \begin{pmatrix} S_{11} & S_{21} & \cdots & S_{n1} \\ \vdots & & & \\ S_{1l} & S_{2l} & \cdots & S_{nl} \end{pmatrix} \times \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = S'e \text{ (say),}$$

where e is a $n \times 1$ vector of unity and S is a $n \times l$ matrix, the (t, i) th element of which is $\frac{\partial l_t(\theta)}{\partial \theta_i}$. The first order condition of maximization of the likelihood ensures that $\frac{\partial L(\theta)}{\partial \theta} = S'e = 0$. The information matrix corresponding to the t^{th} observation is $I_t = E\left(\frac{\partial l_t(\theta)}{\partial \theta} \frac{\partial l_t(\theta)}{\partial \theta'}\right) = -E\left(\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'}\right)$, and the same for the sample of n observations is given by $I = E(S'S/n)$. Furthermore, $S'S/n$ is also a consistent estimator of I under certain conditions. It may be noted that in this model conditional mean involves the parameters of the conditional variance, and hence the information matrix is not block diagonal. In order to solve the nonlinear equations for obtaining the ML estimate, we use the well known algorithm suggested by Berndt, Hall, Hall and Hausman (1974). The BHHH algorithm can be written as :

$$\theta^{(i+1)} = \theta^{(i)} + \eta(S^{(i)'}S^{(i)})^{-1}S^{(i)'}e$$

where $\theta^{(i)}$ is the estimate of θ at the i^{th} step of iteration, $S^{(i)}$ the matrix of first order derivatives evaluated at $\theta^{(i)}$, and η the step length parameter. This algorithm ensures the existence of a consistent estimate of θ . If $\hat{\theta}$ be the maximum likelihood estimate of θ thus obtained, then applying Crowder's theorem (1976) we may easily conclude that $(S'S)^{-1/2}(\hat{\theta} - \theta_0) \overset{\Delta}{\sim} N(0, I_l)$, where θ_0 is the true value of the parameter vector θ , and I_l , the identity matrix of order l .

As stated in the preceding section, the adequacy of the usual ARCH-M model for any given data can now be studied by considering the proposed ARCH-NM model as the alternative. In other words, null hypotheses like $\xi = 1$ and $\xi = 1/2$ may be tested against appropriate alternatives given by $\xi \neq 1$ and $\xi \neq 1/2$, respectively, in the ARCH-NM framework. Standard asymptotic tests may be used to carry out these hypotheses testing.

4.4 An Illustration

In this section we report the results of an application of the proposed model to daily ² closing prices on SENSEX covering the period 4 th week of October 1989 till 2nd week of April 1996. The analysed series is the first differences of the logarithms of SENSEX. Hence, the data represent the continuously compounded rate of return for holding the (aggregate) securities for one day. It is evident from the plot in Figure 4.4.1 of this return series that the data exhibit episodes of both low and high volatility. Since dependence in "squared" data signifies nonlinearity and presence of conditional heteroscedasticity, we computed the Ljung-Box test statistic for the "squared" data, denoted as $Q^2(p)$, for lags upto 24. All these values were found to be highly significant (*cf.* Table 4.4.1), suggesting thereby the presence of nonlinear dependence in the return data. We also obtained the skewness and kurtosis coefficients of y_t to be 0.0209 and 4.9399, respectively. Thus, the skewness coefficient conveys some evidence of asymmetry in the unconditional distribution. The kurtosis coefficient being significantly greater than 3 indicates that the unconditional distribution of the data has heavier tail than a normal distribution.

Further, we carried out an ARCH test which yielded the value of the test statistic as 119.121. Obviously, this is highly significant at 1 per cent level of significance. Since nonlinear dependence and a heavy-tailed unconditional distribution are typical characteristics of conditionally heteroscedastic data, we may thus conclude that the daily-level returns may be appropriately analysed by an ARCH model.

²We have already reported in Chapter 3 the fitted ARCH-M model for returns based on weekly closing prices on SENSEX (*cf.* equations 3.6.4 and 3.6.5). Since for this fitted model $g(h_t)$ has been found to be $\ln h_t$, we have carried out the empirical illustration of this chapter with daily closing prices.

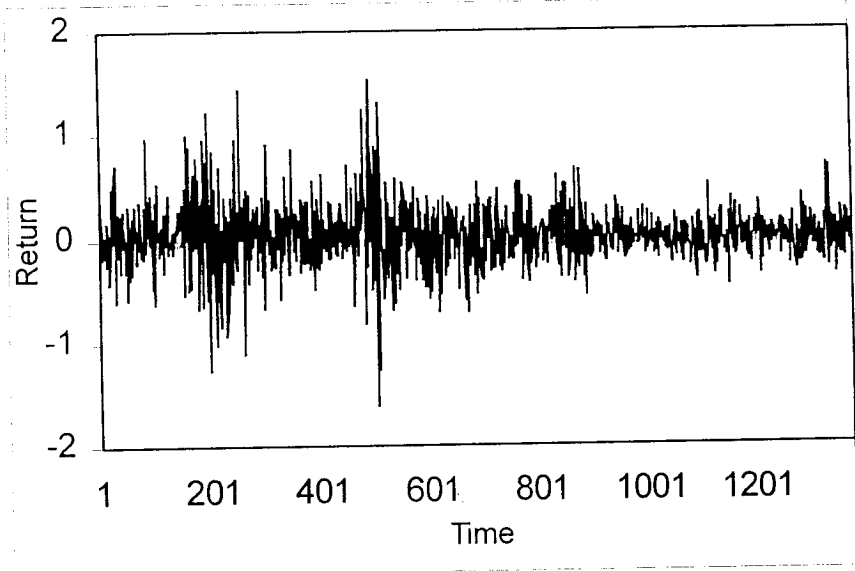


Figure 4.4.1: Time Plot of the Logarithmic Growth Rate of Daily SENSEX

Having recognized that ARCH effect is very strong in the data, we now discuss about the suitability of ARCH-M model or its generalization in the form of the proposed ARCH-NM model, for analysing the return data. To this end it is quite clear from the plot in Figure 4.4.1 that the changes in variances are similar to those hypothesized by ARCH-M/ARCH-NM model. Since both these models yield serial correlation in y_t , we computed Ljung-Box $Q(p)$ test statistic for all lags from 1 to 24 to find if the data support this property. These $Q(p)$ values are summarized in Table 4.4.1. It is evident from the values of $Q(p)$ test statistic that the (linear) correlations are highly significant at 1 per cent level, indicating strongly that ARCH-M or its generalization in ARCH-NM would fit the data well. In what follows we report the empirical findings which, in fact, lend strong support to our approach in which a wider class of models given by ARCH-NM is suggested for proper representation of risk premium. Since it is well-known that GARCH is a better and generalized representation of the conditional variance than ARCH, the empirical exercise was carried out with GARCH representation for h_t . In other words, h_t was assumed to be given by

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 + \eta_1 h_{t-1} + \cdots + \eta_p h_{t-p} \quad (4.4.1)$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$ for $i = 1, 2, \dots, q$ and $\eta_j \geq 0$ for $j = 1, 2, \dots, p$.

As reported below, GARCH (1,1) turned out to be the best model from consideration of maximization of the log-likelihood function. It may also be stated here that for the purpose of this application, it was assumed that excepting for an intercept term there is no other regressor in the model. By following the method of estimation outlined in the preceding section for the proposed ARCH-NM model, we found that the log-likelihood function given

Table 4.4.1: Diagnostic checks of models for SENSEX data

Diagnostics	Observed	GARCH-M	GARCH-NM
	Series	Standardized Residuals	Standardized Residuals
ARCH	119.1211*	8.6736*	0.0934
Q (4)	16.7*	25.0*	40.9*
Q (8)	22.9*	26.0*	42.7*
Q (12)	32.5*	33.1*	45.6*
Q (16)	43.3*	47.3*	57.0*
Q (20)	56.6*	54.7*	62.4*
Q (24)	59.4*	55.9*	64.6*
Q ² (4)	359.0*	34.9*	3.25
Q ² (8)	582.0*	66.3*	4.90
Q ² (12)	786.0*	134.0*	13.5
Q ² (16)	1000.0*	189.0*	17.1
Q ² (20)	1120.0*	225.0*	18.8
Q ² (24)	1210.0*	252.0*	20.5

**indicates significance at 1 per cent level.*

in (4.3.1) was maximized at $\xi = 0.05$. However, the value of the log-likelihood function corresponding to $\xi = 0$ being almost the same as the one at $\xi = 0.05$, (differing in the second decimal place only), we report the best fitted model as

$$y_t = 0.0152 + 0.0018 \ln h_t + \epsilon_t, L(\hat{\theta}) = 3628.77,$$

(2.5347) (2.4699)

$$h_t = 0.00001 + 0.12859 \epsilon_{t-1}^2 + 0.85151 h_{t-1},$$

(4.96431) (7.78673) (55.80221) (4.4.2)

(t-ratios are given in parentheses)

Thus, the chosen model is GARCH-NM of order (1,1) with $\xi = 0$. It is evident that for this model all the parameters in h_t are significant at 1 per cent level of significance, and those in y_t are clearly significant at 5 per cent level and "almost" significant at 1 per cent level. It may be mentioned in this context that the significant positive value of λ is in conformity with the basic understanding in the risk premium literature that positive value of the risk aversion parameter is quite desirable. In order to compare this model with the standard GARCH-M model, we computed the maximum log-likelihood value corresponding $\xi = 1/2$ and found it to be only 3501.21, which is much smaller than the maximized log-likelihood value of the chosen GARCH-NM model. As regards the significance of the estimated GARCH-M model, we find from equation (4.4.3) below that while all the parameters in h_t are significant, the risk aversion parameter λ in y_t is highly insignificant, the t-statistic value being only 0.3031. Thus, not only that the standard GARCH-M model with $\xi = 1/2$ produces a much smaller log-likelihood value as compared to the chosen GARCH-NM model, but also that a straightforward GARCH-M fitting would choose an inappropriate model for the data. In order to formally test

for the adequacy of standard GARCH-M model given the framework of a more generalized model as proposed in this paper, we carried out the likelihood ratio test with the null hypothesis specifying the standard GARCH-M model. The test obviously soundly rejected the null hypothesis in favour of the proposed GARCH-NM model.

$$y_t = -0.0005 + 0.0613 \sqrt{\hat{h}_t} + \epsilon_t, \quad L(\hat{\theta}) = 3501.21,$$

$$(-0.1272) \quad (0.3031).$$

$$h_t = 0.00002 + 0.0991 \epsilon_{t-1}^2 + 0.3941 h_{t-1}, \quad (4.4.3)$$

$$(8.3136) \quad (8.1062) \quad (6.3815)$$

(t-ratios are given in parentheses)

In order to judge the goodness of the fitted GARCH-NM model and compare it with the standard GARCH-M model, we now report in Table 4.4.1 the results of diagnostic checks for the standardized residuals given by $\tilde{\epsilon}_t = \hat{\epsilon}_t / \sqrt{\hat{h}_t}$, where $\hat{\epsilon}_t$ and \hat{h}_t are the ML residual and estimated conditional variance at t , respectively. We observe from this table that the autocorrelations of $\hat{\epsilon}_t$'s are significant; this suggests that there exist some linear correlations in the GARCH-NM standardized residuals. In other words, the autocorrelations of y_t have not been fully incorporated through the GARCH-NM framework. It may, however, be noted from a glance at Table 4.4.1 that the inference with regard to performance by $Q(p)$ test is the same for the usual GARCH-M (with $\xi = 1/2$) standardized residuals as well. Thus, irrespective of the framework being GARCH-M or GARCH-NM, we observe that the residuals exhibit serial correlation. In order to find if this correlation could be rectified by including an ARMA type component in the conditional mean part of the model, we considered an extension of the model in equation (4.2.2) by including y_{t-1} as an

explanatory variable. While this led to some improvement in the serial correlation of these (standardized) residuals, the maximum value of the log-likelihood function reduced quite significantly to 3585.97, as compared to 3628.77 for the original GARCH-NM model of (4.2.2). Thus, we find, that, by ML criterion, the original GARCH-NM model turns out to be a more appropriate model for the given data set.

As far as the autocorrelation structure of $\tilde{\epsilon}_t^2$ is concerned, $Q^2(p)$ based on standardized GARCH-NM residuals show that none of the 24 test statistic values corresponding to 24 lags is significant even at 5 per cent level. We may, therefore, conclude that the residuals contain no more nonlinearity. It is also evident from Table 4.4.1 that $Q^2(p)$ values are highly significant with standardized GARCH-M residuals. Thus, we observe that in terms of diagnostic checking with standardized residuals, the chosen GARCH-NM is a better model than GARCH-M. We have, therefore, established through this illustration that the suggested extension of the standard (G) ARCH-M model (in which the risk premium component is assumed to be a flexible functional form in the sense of Box-Cox transformation) can provide improvement over the standard (G) ARCH-M model.

4.5 Conclusions

In the literature on time-varying risk premium researchers have used conditional variance h_t or $\sqrt{h_t}$ or $\ln h_t$ to represent the risk premium in the model. It is, therefore, natural to argue that a flexible functional form of h_t for representing risk premium should be more useful and appropriate. Some of the existing empirical evidences lend support towards this direction. Keeping this in mind, we have proposed in this chapter a generalization of the ARCH-M model by allowing for nonlinear representation in the mean of the dependent

variable. This has been done by considering the Box-Cox power transformation of the conditional variance representing the risk premium in the model. Obviously, this generalized model encompasses various other functional forms used in the risk premium literature as special cases. The estimation of this model by the method of maximum likelihood has been discussed, and an illustrative example in support of this generalized approach for representation of risk premium has also been given. This illustration with daily returns based on SENSEX demonstrates that the suggested generalization provides improvement over the standard ARCH-M model. Before we conclude, we may state that keeping in mind the importance of functional form of h_t in the performance of ARCH-NM model, possible extensions of our work may be done by considering some alternative functional forms of h_t .

Appendix A4.1

To prove Results 1 and 2, we need the unconditional expectation of $\omega_t' = (\epsilon_t^8, \epsilon_t^4, \epsilon_t^2)$. Engle (1982) proved that $E(\omega_t) = (I - A)^{-1}b$, where

$$A = \begin{bmatrix} 105\alpha_1^4 & 420\alpha_0\alpha_1^3 & 630\alpha_0^2\alpha_1^2 & 420\alpha_0^3\alpha_1 \\ 0 & 15\alpha_1^3 & 45\alpha_1^2 & 43\alpha_1^2\alpha_1 \\ 0 & 0 & 3\alpha_1^2 & 6\alpha_0\alpha_1 \\ 0 & 0 & 0 & \alpha_1 \end{bmatrix}$$

and $b' = (105\alpha_0^4 \quad 15\alpha_0^3 \quad 3\alpha_0^2 \quad \alpha_0)$.

It is quite clear that to obtain $E(\epsilon_t^8)$ we need only the first row of $(I - A)^{-1}$, i.e.,

$$\begin{bmatrix} \frac{1}{1-105\alpha_1^4} \\ -\frac{420\alpha_0\alpha_1^3}{(1-105\alpha_1^4)} \\ \frac{630\alpha_0^2\alpha_1^2(1+30\alpha_1^3)}{(1-105\alpha_1^4)(1-15\alpha_1^3)(1-3\alpha_1^2)} \\ -\frac{420\alpha_0^3\alpha_1(45\alpha_1^5+30\alpha_1^3+6\alpha_1^2+1)}{(1-105\alpha_1^4)(1-15\alpha_1^3)(1-3\alpha_1^2)(1-\alpha_1)} \end{bmatrix}'$$

and therefore

$$E(\epsilon_t^8) = \frac{105\alpha_0^4(1 - 5\alpha_1 + 15\alpha_1^2 - 114\alpha_1^3 + 495\alpha_1^4 - 315\alpha_1^5 - 405\alpha_1^6)}{(1 - 105\alpha_1^4)(1 - 15\alpha_1^3)(1 - 3\alpha_1^2)(1 - \alpha_1)}$$

Similarly it may easily be checked that

$$E(\epsilon_t^6) = \frac{105\alpha_0^3(1 + 2\alpha_1 + 6\alpha_1^2 + 3\alpha_1^3)}{(1 - 15\alpha_1^3)(1 - 3\alpha_1^2)(1 - \alpha_1)},$$

$$E(\epsilon_t^4) = \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)},$$

and

$$E(\epsilon_t^2) = \frac{\alpha_0}{1 - \alpha_1}.$$

PROOF OF RESULT 1. As stated in the text, we take the model as

$$E(y_t) = \beta + \lambda E g(h_t).$$

Now, the Taylor series expansion of $g(h_t)$, around $E(h_t) = E(\epsilon_t^2) (= \bar{h}_t, \text{ say})$ yields

$$g(h_t) = g(\bar{h}_t) + (h_t - \bar{h}_t)g'(h_t) |_{h_t=\bar{h}_t} + \frac{(h_t - \bar{h}_t)^2}{2!}g''(h_t) |_{h_t=\bar{h}_t},$$

taking upto squared term only.

Therefore,

$$E(y_t) = g(\bar{h}_t) + \beta + \lambda \left\{ f(\bar{h}_t) + g''(\bar{h}_t) \frac{V(h_t)}{2} \right\}$$

Now, $V(h_t) = V(\alpha_0 + \alpha_1 \epsilon_{t-1}^2) = \frac{2(\alpha_0 \alpha_1)^2}{(1 - \alpha_1)^2(1 - 3\alpha_1^2)}$, and hence, we have

$$E(y_t) = \beta + \lambda \left[\frac{\left(\frac{\alpha_0}{1 - \alpha_1}\right)^\xi - 1}{\xi} + (\xi - 1) \left(\frac{\alpha_0}{1 - \alpha_1}\right)^{\xi-2} \frac{(\alpha_0 \alpha_1)^2}{(1 - \alpha_1)^2(1 - 3\alpha_1^2)} \right].$$

As for the variance of y_t , it can easily be seen to be

$$V(y_t) = \frac{\alpha_0}{1 - \alpha_1} + \lambda^2 [g'(\bar{h}_t)^2 + \bar{h}_t^2 g''(\bar{h}_t)^2 - 2g'(\bar{h}_t)g''(\bar{h}_t)\bar{h}_t]V(h_t) + \frac{\lambda^2}{4} g''(\bar{h}_t)^2 V(h_t^2) + \lambda^2 [g'(\bar{h}_t)g''(\bar{h}_t) - 2\bar{h}_t g''(\bar{h}_t)^2] Cov(h_t, h_t^2).$$

Since $V(h_t^2) = E(h_t^4) - \{E(h_t^2)\}^2$, we need the expressions for $E(h_t^2)$ and $E(h_t^4)$ for simplifying $V(y_t)$. These latter expressions are as follows.

$$\begin{aligned} E(h_t^2) &= \alpha_0^2(1 + \alpha_1)/(1 - \alpha_1)(1 - 3\alpha_1^2) \\ E(h_t^4) &= \frac{\alpha_0^4(1 + 3\alpha_1 + 15\alpha_1^2 - 6\alpha_1^3 - 50\alpha_1^3 - 1065\alpha_1^5 - 405\alpha_1^6 - 11335\alpha_1^7)}{(1 - 105\alpha_1^4)(1 - 15\alpha_1^3)(1 - 3\alpha_1^2)(1 - \alpha_1)} \\ &\quad + \frac{\alpha_0^4(56700\alpha_1^8 - 9450\alpha_1^9 - 28350\alpha_1^{10})}{(1 - 105\alpha_1^4)(1 - 15\alpha_1^3)(1 - 3\alpha_1^2)(1 - \alpha_1)} \end{aligned}$$

Therefore,

$$\begin{aligned} V(h_t) &= \frac{\alpha_0^4(8\alpha_1^2 - 18\alpha_1^3 + 37\alpha_1^4 - 817\alpha_1^5 + 933\alpha_1^6 - 9160\alpha_1^7 + 6929\alpha_1^8)}{(1 - 105\alpha_1^4)(1 - 15\alpha_1^3)(1 - 3\alpha_1^2)^2(1 - \alpha_1)^2} \\ &\quad - \frac{\alpha_0^4(117555\alpha_1^9 + 154995\alpha_1^{10} + 113400\alpha_1^{11} - 113400\alpha_1^{12} - 85050\alpha_1^{13})}{(1 - 105\alpha_1^4)(1 - 15\alpha_1^3)(1 - 3\alpha_1^2)^2(1 - \alpha_1)^2} \\ &= K_0 \text{ (say)}. \end{aligned}$$

We also note that $Cov(h_t, h_t^2)$ involves $E(h_t^3)$ which is given by

$$\begin{aligned} E(h_t^3) &= E(\alpha_0^3 + 3\alpha_0^2\alpha_1\epsilon_{t-1}^2 + 3\alpha_0\alpha_1^2\epsilon_{t-1}^4 + \alpha_1^3\epsilon_{t-1}^6) \\ &= \frac{\alpha_0^3(1 + 2\alpha_1 + 6\alpha_1^2 + 3\alpha_1^3)}{(1 - 15\alpha_1^3)(1 - 3\alpha_1^2)(1 - \alpha_1)}. \end{aligned}$$

Hence,

$$Cov(h_t, h_t^2) = \frac{6\alpha_0^3\alpha_1^2(1 + 2\alpha_1 + 2\alpha_1^2)}{(1 - 15\alpha_1^3)(1 - 3\alpha_1^2)(1 - \alpha_1)^2}.$$

Thus we obtain the expression for $V(y_t)$.

PROOF OF RESULT 2. The first order autocovariance of y_t , denoted by ν_1 , is

$$\begin{aligned} \nu_1 &= Cov[\beta + \lambda g(h_t), \beta + \lambda g(h_{t-1})] \\ &= \lambda^2 Cov[g(h_t), g(h_{t-1})]. \end{aligned}$$

Again by using Taylor series expansion of $g(h_t)$, with respect to \bar{h}_t and then taking upto the second term only, we have

$$\begin{aligned}
\nu_1 &= \lambda^2 \left[(4 - 4\xi + \xi^2) \left(\frac{\alpha_0}{1 - \alpha_1} \right)^{2(\xi-1)} \right] Cov(h_t, h_{t-1}) \\
&+ \frac{\lambda^2}{2} \left[(\xi - 1) \left(\frac{\alpha_0}{1 - \alpha_1} \right)^{2(\xi-3)} - \frac{(\xi - 1)^2}{2} \left(\frac{\alpha_0}{1 - \alpha_1} \right)^{2\xi-3} \right] Cov(h_t, h_{t-1}^2) \\
&+ \frac{\lambda^2}{2} \left[(\xi - 1) \left(\frac{\alpha_0}{1 - \alpha_1} \right)^{2\xi-3} - \frac{(\xi - 1)^2}{2} \left(\frac{\alpha_0}{1 - \alpha_1} \right)^{2\xi-3} \right] Cov(h_t^2, h_{t-1}) \\
&+ \frac{\lambda^2}{4} \left[(\xi - 1)^2 \left(\frac{\alpha_0}{1 - \alpha_1} \right)^{2(\xi-2)} \right] Cov(h_t^2, h_{t-1}^2).
\end{aligned}$$

To evaluate ν_1 , we note that it involves several covariances which in turn involve the expressions for $E(h_t h_{t-1})$, $E(h_t h_{t-1}^2)$, $E(h_t^2 h_{t-1})$ and $E(h_t^2 h_{t-1}^2)$.

After algebraic simplifications, these expectation terms may be found to be as follows.

$$\begin{aligned}
E(h_t h_{t-1}) &= \frac{\alpha_0^2(1 + \alpha_1 - \alpha_1^2)}{(1 - \alpha_1)(1 - 3\alpha_1^2)} \\
E(h_t h_{t-1}^2) &= \frac{\alpha_1^3(1 + 2\alpha_1 + 2\alpha_1^2 - 9\alpha_1^3 - 12\alpha_1^4)}{(1 - \alpha_1)(1 - 3\alpha_1^2)(1 - 15\alpha_1^3)} \\
E(h_t^2 h_{t-1}) &= \frac{\alpha_0^3(1 + 4\alpha_1 + 6\alpha_1^2 - 15\alpha_1^3 - 42\alpha_1^4 - 36\alpha_1^5 + 90\alpha_1^6)}{(1 - \alpha_1)(1 - 3\alpha_1^2)(1 - 15\alpha_1^3)}
\end{aligned}$$

and

$$\begin{aligned}
E(h_t^2, h_{t-1}^2) &= \frac{\alpha_0^4(1 + 3\alpha_1 + 7\alpha_1^2 + 6\alpha_1^3 - 38\alpha_1^4 - 51\alpha_1^5 + 511\alpha_1^6 + 315\alpha_1^7)}{(1 - \alpha_1)(1 - 3\alpha_1^2)(1 - 15\alpha_1^3)(1 - 105\alpha_1^4)} \\
&- \frac{\alpha_0^4(7050\alpha_1^8 + 27209\alpha_1^9 + 60930\alpha_1^{10} + 99540\alpha_1^{11} + 52380\alpha_1^{12})}{(1 - \alpha_1)(1 - 3\alpha_1^2)(1 - 15\alpha_1^3)(1 - 105\alpha_1^4)}.
\end{aligned}$$

Using the expressions for these expectation terms, the covariances in the expression for ν_1 can be reduced to

$$\begin{aligned} Cov(h_t, h_{t-1}) &= \frac{2\alpha_0^2\alpha_1^3}{(1-\alpha_1)^2(1-3\alpha_1^2)} = K_1(\text{say}), \\ Cov(h_t, h_{t-1}^2) &= \frac{3\alpha_0^3\alpha_1^2(1+4\alpha_1+4\alpha_1^2)}{(1-\alpha_1)^2(1-3\alpha_1^2)(1-15\alpha_1^3)} = K_2(\text{say}), \\ Cov(h_t^2, h_{t-1}) &= \frac{\alpha^3(2\alpha_1+2\alpha_1^2-6\alpha_1^3-12\alpha_1^4+6\alpha_1^5+12\alpha_1^6-90\alpha_1^7)}{(1-\alpha_1)^2(1-3\alpha_1^2)(1-15\alpha_1^3)} = K_3(\text{say}), \end{aligned}$$

and

$$\begin{aligned} Cov(h_t^2, h_{t-1}^2) &= \frac{\alpha_0^4(8\alpha_1^3+79\alpha_1^4+215\alpha_1^5+799\alpha_1^6-1504\alpha_1^7-12201\alpha_1^8)}{(1-\alpha_1)^2(1-3\alpha_1^2)(1-15\alpha_1^3)(1-105\alpha_1^4)} \\ &\quad - \frac{\alpha_0^4(21146\alpha_1^9+11626\alpha_1^{10}-21867\alpha_1^{11}-148323\alpha_1^{12})}{(1-\alpha_1)^2(1-3\alpha_1^2)^2(1-15\alpha_1^3)(1-105\alpha_1^4)} \\ &\quad + \frac{\alpha_0^4(168210\alpha_1^{13}-141480\alpha_1^{14}-157140\alpha_1^{15})}{(1-\alpha_1)^2(1-3\alpha_1^2)(1-15\alpha_1^3)(1-105\alpha_1^4)} \\ &= K_4(\text{say}). \end{aligned}$$

Thus, finally the expression for $Cov(y_t, y_{t-1})$ is given by

$$\begin{aligned} \nu_1 &= K_1\lambda^2 \left[(4-4\xi+\xi^2) \left(\frac{\alpha_0}{1-\alpha_1} \right)^{2(\xi-1)} \right] \\ &\quad + K_2 \frac{\lambda^2}{2} \left[(\xi-1) \left(\frac{\alpha_0}{1-\alpha_1} \right)^{2(\xi-1)} - \frac{(\xi-1)^2}{2} \left(\frac{\alpha_0}{1-\alpha_1} \right)^{2\xi-3} \right] \\ &\quad + K_3 \frac{\lambda^2}{2} \left[(\xi-1) \left(\frac{\alpha_0}{1-\alpha_1} \right)^{2\xi-3} - \frac{(\xi-1)^2}{2} \left(\frac{\alpha_0}{1-\alpha_1} \right)^{2\xi-3} \right] \\ &\quad + K_4 \frac{\lambda^2}{4} \left[(\xi-1)^2 \left(\frac{\alpha_0}{1-\alpha_1} \right)^{2(\xi-2)} \right]. \end{aligned}$$

Appendix A4.2

First order derivatives of $L(\theta | \Psi_{t-1})$ with respect to the parameter vector $\theta' = (\beta', \lambda, \alpha', \xi)$.

The argument of $l_t(\theta | \Psi_{t-1})$ as also the conditional sign are being dropped for convenience.

Case I : $\xi \neq 0$

$$\frac{\partial L}{\partial \theta} = \sum_{t=1}^T \frac{\partial l_t}{\partial \theta} = \sum_{t=1}^T \left(\frac{\partial l_t}{\partial \beta'} \frac{\partial l_t}{\partial \lambda} \frac{\partial l_t}{\partial \alpha'} \frac{\partial l_t}{\partial \xi} \right)'$$

Now, $\frac{\partial l_t}{\partial \theta} = \frac{1}{2h_t} \frac{\partial h_t}{\partial \theta} \left(\frac{\epsilon_t^2}{h_t} - 1 \right) + \frac{\epsilon_t}{h_t} \frac{\partial \beta'}{\partial \theta} x_t + \lambda \frac{\epsilon_t}{h_t} \frac{\partial g(h_t)}{\partial \theta} + g(h_t) \frac{\epsilon_t}{h_t} \frac{\partial \lambda}{\partial \theta}$ and hence

$$\left. \begin{aligned} \frac{\partial l_t}{\partial \beta} &= \frac{1}{2h_t} \frac{\partial h_t}{\partial \beta} \left(\frac{\epsilon_t^2}{h_t} - 1 \right) + \frac{\partial \beta'}{\partial \theta} x_t \frac{\epsilon_t}{h_t} + \lambda \frac{\partial g(h_t)}{\partial \beta} \frac{\epsilon_t}{h_t} \\ \frac{\partial l_t}{\partial \lambda} &= \frac{1}{2h_t} \frac{\partial h_t}{\partial \lambda} \left(\frac{\epsilon_t^2}{h_t} - 1 \right) + \lambda \frac{\partial g(h_t)}{\partial \lambda} \frac{\epsilon_t}{h_t} + g(h_t) \frac{\epsilon_t}{h_t} \\ \frac{\partial l_t}{\partial \alpha} &= \frac{1}{2h_t} \frac{\partial h_t}{\partial \alpha} \left(\frac{\epsilon_t^2}{h_t} - 1 \right) + \lambda \frac{\partial g(h_t)}{\partial \alpha} \frac{\epsilon_t}{h_t} \\ \frac{\partial l_t}{\partial \xi} &= \frac{1}{2h_t} \frac{\partial h_t}{\partial \xi} \left(\frac{\epsilon_t^2}{h_t} - 1 \right) + \lambda \frac{\partial g(h_t)}{\partial \xi} \frac{\epsilon_t}{h_t} \end{aligned} \right] \quad (A4.2)$$

All the above derivatives in (A4.2) involve $\frac{\partial h_t}{\partial \theta}$ as well as $\frac{\partial g(h_t)}{\partial \theta}$, and these may be obtained as follows :

$$\begin{aligned} \frac{\partial h_t}{\partial \beta} &= -2 \left\{ \alpha_1 \epsilon_{t-1} (x_{t-1} + \lambda h_{t-1}^{\xi-1} \frac{\partial h_{t-1}}{\partial \beta}) + \dots + \alpha_p \epsilon_{t-p} \left(x_{t-p} + \lambda h_{t-p}^{\xi-1} \frac{\partial h_{t-p}}{\partial \beta} \right) \right\} \\ \frac{\partial h_t}{\partial \lambda} &= -2\lambda \left(\alpha_1 \epsilon_{t-1} h_{t-1}^{\xi-1} \frac{\partial h_{t-1}}{\partial \lambda} + \dots + \alpha_p \epsilon_{t-p} h_{t-p}^{\xi-1} \frac{\partial h_{t-p}}{\partial \lambda} \right) \\ &\quad - 2(\alpha_1 \epsilon_{t-1} g(h_{t-1}) + \dots + \alpha_p \epsilon_{t-p} g(h_{t-p})) \\ \frac{\partial h_t}{\partial \alpha} &= \eta_t - 2\lambda \left(\alpha_1 h_{t-1}^{\xi-1} \epsilon_{t-1} \frac{\partial h_{t-1}}{\partial \alpha} + \dots + \alpha_p h_{t-p}^{\xi-1} \epsilon_{t-p} \frac{\partial h_{t-p}}{\partial \alpha} \right) \\ \frac{\partial h_t}{\partial \xi} &= 2\lambda \alpha_1 \epsilon_{t-1} \left[h_{t-1}^{\xi-1} \frac{\partial h_{t-1}}{\partial \xi} + \frac{h_{t-1}^{\xi} \ln h_{t-1}}{\xi} - \frac{h_{t-1}^{\xi-1}}{\xi^2} \right] - \\ &\quad \dots - 2\lambda \alpha_p \epsilon_{t-p} \left[h_{t-p}^{\xi-1} \frac{\partial h_{t-p}}{\partial \xi} + \frac{h_{t-p}^{\xi} \ln h_{t-p}}{\xi} - \frac{h_{t-p}^{\xi-1}}{\xi^2} \right] \end{aligned}$$

where $\eta'_t = (1, \epsilon_{t-1}^2, \dots, \epsilon_{t-p}^2)$.

Since $\frac{\partial g(h_t)}{\partial \theta}$ can be easily expressed in terms of $\frac{\partial h_t}{\partial \theta}$, substituting all these above expressions of $\frac{\partial h_t}{\partial \theta}$ in (A4.2) we have the final expression for $\frac{\partial l_t}{\partial \theta}$ as follows

$$\begin{aligned}
\frac{\partial l_t}{\partial \beta} &= -\frac{1}{h_t} \left(\frac{\epsilon_t^2}{h_t} - 1 + 2\lambda \epsilon_t h_t^{\xi-1} \right) \left\{ (\alpha_1 \epsilon_{t-1} \xi_{t-1} + \dots + \alpha_p \epsilon_{t-p} \xi_{t-p}) \right. \\
&\quad \left. + \lambda \left(\alpha_1 \epsilon_{t-1} h_{t-1}^{\xi-1} \frac{\partial h_{t-1}}{\partial \beta} + \dots + \alpha_p \epsilon_{t-p} h_{t-p}^{\xi-1} \frac{\partial h_{t-p}}{\partial \beta} \right) \right\} + \frac{\epsilon_t}{h_t} x_t \\
\frac{\partial l_t}{\partial \lambda} &= -\frac{1}{h_t} \left(\frac{\epsilon_t^2}{h_t} - 1 + 2\lambda \epsilon_t h_t^{\xi-1} \right) \left\{ \left(\alpha_1 \epsilon_{t-1} h_{t-1}^{\xi-1} \frac{\partial h_{t-1}}{\partial \lambda} + \dots + \alpha_p \epsilon_{t-p} h_{t-p}^{\xi-1} \frac{\partial h_{t-p}}{\partial \lambda} \right) \right. \\
&\quad \left. + (\lambda_1 \epsilon_{t-1} g(h_{t-1}) + \dots + \alpha_p g(h_{t-p})) \right\} + g(h_t) \frac{\epsilon_t}{h_t} \\
\frac{\partial l_t}{\partial \alpha} &= \frac{1}{2h_t} \left(\frac{\epsilon_t^2}{h_t} - 1 + 2\lambda \epsilon_t h_t^{\xi-1} \right) \left\{ \eta_t - 2\lambda \left(\alpha_1 h_{t-1}^{\xi-1} \epsilon_{t-1} \frac{\partial h_{t-1}}{\partial \alpha} \right. \right. \\
&\quad \left. \left. + \dots + \alpha_p h_{t-p}^{\xi-1} \epsilon_{t-p} \frac{\partial h_{t-p}}{\partial \alpha} \right) \right\} \\
\frac{\partial l_t}{\partial \xi} &= -\frac{\lambda}{h_t} \left(\frac{\epsilon_t^2}{h_t} - 1 + 2\lambda \epsilon_t h_t^{\xi-1} \right) \left\{ \alpha_1 \epsilon_{t-1} \left(h_{t-1}^{\xi-1} \frac{\partial h_{t-1}}{\partial \xi} + \frac{h_{t-1}^{\xi-1} \ln h_{t-1}}{\xi} - \frac{h_{t-1}^{\xi-1}}{\xi^2} \right) \right. \\
&\quad \left. + \dots + \alpha_p \epsilon_{t-p} \left(h_{t-p}^{\xi-1} \frac{\partial h_{t-p}}{\partial \xi} + \frac{h_{t-p}^{\xi-1} \ln h_{t-p}}{\xi} - \frac{h_{t-p}^{\xi-1}}{\xi^2} \right) \right\} \\
&\quad + \lambda \frac{h_t^{\xi-1} \ln h_t \epsilon_t}{\xi} - \lambda \frac{h_t^{\xi-1}}{\xi^2} \frac{\epsilon_t}{h_t}
\end{aligned}$$

Case II : $\xi = 0$. In this case $g(h_t) = \ln h_t$, and hence we can easily find that the final expressions for the first order derivatives of l_t are as follows :

$$\begin{aligned}
\frac{\partial l_t}{\partial \beta} &= -\frac{1}{h_t} \left(\frac{\epsilon_t^2}{h_t} - 1 + 2\lambda \epsilon_t \right) \left\{ (\alpha_1 \epsilon_{t-1} x_{t-1} + \dots + \alpha_p \epsilon_{t-p} x_{t-p}) \right. \\
&\quad \left. + \lambda \left(\alpha_1 \epsilon_{t-1} \frac{\partial h_{t-1}}{\partial \beta} + \dots + \alpha_p \epsilon_{t-p} \frac{\partial h_{t-p}}{\partial \beta} \right) \right\} + \frac{\epsilon_t}{h_t} x_t \\
\frac{\partial l_t}{\partial \lambda} &= -\frac{1}{h_t} \left(\frac{\epsilon_t^2}{h_t} - 1 + 2\lambda \epsilon_t \right) (\alpha_1 \epsilon_{t-1} h_{t-1} + \dots + \alpha_p \epsilon_{t-p} h_{t-p}) \\
&\quad - \frac{\lambda}{h_t} \left(\frac{\epsilon_t^2}{h_t} - 1 + 2\lambda \epsilon_t \right) \left(\alpha_1 \epsilon_{t-1} \frac{\partial h_{t-1}}{\partial \lambda} + \dots + \alpha_p \epsilon_{t-p} h_{t-p} \frac{\partial h_{t-p}}{\partial \lambda} \right) + \epsilon_t \\
\frac{\partial l_t}{\partial \alpha} &= \frac{1}{2h_t} \left(\frac{\epsilon_t^2}{h_t} - 1 + 2\lambda \epsilon_t \right) \left\{ \eta_t - 1\lambda \left(\alpha_1 \epsilon_{t-1} \frac{\partial h_{t-1}}{\partial \alpha} + \dots + \alpha_p \epsilon_{t-p} \frac{\partial h_{t-p}}{\partial \alpha} \right) \right\}.
\end{aligned}$$

Chapter 5

Testing for Constancy of the Relative Risk Aversion Parameter and Estimating the Time Varying Parameter ARCH-M Model

5.1 Introduction

One of the economic theoretic implications of ARCH-M model is that economic agents are constant relative risk averse; the assumption seems to be fairly strong and may not be true in many practical situations. If a representative agent maximizes a time additive von Neumann-Morgenstern utility, the mean/variance ratio or the relative risk aversion parameter may change due to either change in perception towards risk involved in the financial market or change in the distribution of wealth or both. This is likely to be more so in emerging economies like the Indian economy because of the possible changes in their capital markets as well as in other sectors of the economies, which might happen due to government policies on reforms and liberalization. Moreover, during the course of time, the changes in consumption and investment pattern of different economic agents might also influence the relative risk aversion parameter.

Now, insofar as empirical evidences are concerned, French, Schwert and Stambaugh (1987) have found a striking result to the effect that relative risk aversion parameter is very unstable across the sample periods. Based on New York Stock Exchange (NYSE) monthly value-weighted index, they obtained the parameter estimates to be 1.693, 1.510 and 7.220 for the periods 1928-1984, 1928-1952 and 1953-1984, respectively. They have also found using the S & P daily composite index, that the two subsample (1928-1952 and 1953-1984) estimates are 0.598 and 7.809, respectively. Pindyck (1988) estimated the risk aversion parameter by using NYSE index for the period 1949-83 and 1962-83, and obtained the estimates as 3.447 and 1.672, respectively. With rolling regression technique, Chou *et al.* (1992) found that the mean/variance ratio of S & P ranges from -0.4 to 15.6 with a mean of 5.4 and standard deviation 4.1. All these findings, therefore, indicate that it should be possible to generalize ARCH-M model with an appropriately defined time varying risk aversion parameter.

The effect of neglecting the time varying nature (if present) of relative risk aversion parameter, λ , are two folds. First, one of the inputs required by investors seeking to hold risky asset is the *ex ante* measure of risk premium which, in turn, depends on relative risk aversion parameter. Now if a time invariant λ is estimated using sample information, the estimated risk premium is liable to be erroneous if indeed λ is not stable. So, from investor's point of view test of stability of λ is crucial. Secondly, validity of inferences and post-sample forecasting crucially depends on the underlying parameter being stable. Neglecting parameter variation across observations results in a misspecified likelihood function, which may lead to inconsistent maximum likelihood estimates of relevant parameters (*cf.* White, 1982). It is, therefore, important to know in an actual data analysis whether the relative risk aversion parameter is time invariant or not. Towards this end, we propose,

in this chapter, a test procedure for testing the temporal stability of relative risk aversion parameter. In this test we essentially adopt Chesher's (1983) interpretation of White's (1982) information matrix (IM) test being a test of parameter variation. We then study the size and power performance of this test through a detailed Monte Carlo study. We also consider an ARCH-M model in which the relative risk aversion parameter is assumed to follow a random walk process in the framework of a state space model. This model is estimated by simultaneous application of Kalman filter and ML method. In order to examine the temporal behaviour of λ , the estimate of λ thus obtained as well as those obtained from rolling regression technique have been used.

The present chapter has been organized as follows. The proposed test of constancy of the relative risk aversion parameter is discussed in Section 5.2. In Section 5.3, we summarize the findings of our Monte Carlo experiment. Section 5.4 presents the results of bootstrap-based size corrected test. In Section 5.5, we introduce and then discuss ARCH-M model with time varying λ . The empirical illustration is given in Section 5.6. The chapter concludes with some final remarks in Section 5.7.

5.2 Test for Constancy of Relative Risk Aversion Parameter

A natural and accepted procedure for detecting parameter constancy is to specify any particular probability density, say, $\tilde{g}(\cdot)$, for the underlying parameter and then test for the variance of its distribution being zero. Towards this end, we first present Chesher's (1983, 1984) interpretation of IM test as a test for local parameter variations, and then derive the test statistic for the constancy of the relative risk aversion parameter of ARCH-M model.

5.2.1 Score Test for Parameter Constancy

Following Chesher (1983) and Cox (1983), we focus on deriving Rao's (1948) score test for testing the hypothesis that the variance of the parameter of interest is zero. Here, the score test involves examination of the local behaviour of the log-likelihood function close to the null hypothesis of no parameter variation, and hence it does not require the explicit specification of the alternative hypothesis in the form of any arbitrary distributional assumption on parameter variations. Moreover, the asymptotic properties of the score test are not affected (Bera and Ullah, 1991; Moran, 1971; Self and Liang, 1987; and Bera, Ra and Sarkar, 1998) by the boundary value problem as opposed to likelihood ratio or Wald tests. Following the notations of Chesher (1983), and Bera and Kim (2001), let the *pdf* of the underlying random variable y be $f(y; \theta)$ and suppose, the $l \times 1$ parameter vector θ has density, called the prior density $\tilde{g}(\theta; \theta_0, \Omega)$, where $\theta_0 = E(\theta)$ and $\Omega = \{\omega_{rs}\}$, the variance covariance matrix of θ . Let $h(y; \theta_0, \Omega)$ be the marginal distribution of y , which can be written as

$$h(y; \theta_0, \Omega) = E_{\theta}[f(y; \theta)] = \int f(y; \theta) \tilde{g}(\theta; \theta_0, \Omega) d\theta. \quad (5.2.1)$$

Since, generally we don't have much information on $\tilde{g}(\cdot)$ and we need only local behaviour of $h(y; \theta_0, \Omega)$, we consider a small Ω approximation to this density.

Defining $f_j(y; \theta) = \partial^j f(y; \theta) / \partial \theta^j$, $F_j(y; \theta) = \partial^j \log f(y; \theta) / \partial \theta^j$ $j = 1, 2$. We note that $F_2 = f_2/f - F_1 F_1'$. Now, a second-order Taylor series expansion of $f(y; \theta)$ around θ_0 yields

$f^a(y; \theta_0)^1 = f(y; \theta_0) + f_1(y; \theta_0)(\theta - \theta_0) + \frac{1}{2} tr[f_2(y; \theta_0)(\theta - \theta_0)(\theta - \theta_0)']$, where $tr(\cdot)$ stands for the trace of a matrix. Now, taking expectation with respect

¹Subscript 'a' denotes that the expression is an approximation one

to θ we have from 5.2.1,

$$\begin{aligned}
h^a(y; \theta_0, \Omega) &= E_{\theta_0}[f^a(y; \theta)] \\
&= \int f^a(y; \theta) \tilde{g}(\theta; \theta_0, \Omega) d\theta \\
&= \int \{f(y; \theta_0) + f_1(y; \theta_0)(\theta - \theta_0) + 1/2 \text{tr}\{f_2(y; \theta_0)(\theta - \theta_0)(\theta - \theta_0)'\}\} \tilde{g}(\theta; \theta_0, \Omega) d\theta \\
&= f(y; \theta_0) + 1/2 \text{tr} f_2(y; \theta_0) \Omega \\
&= f(y; \theta_0) + 1/2 \text{tr}(F_2(y; \theta_0) + F_1(y; \theta_0)F_1(y; \theta_0)') f(y; \theta_0) \Omega \\
&= f(y; \theta_0) [1 + 1/2 \text{tr}\{F_2(y; \theta_0) + F_1(y; \theta_0)F_1(y; \theta_0)'\} \Omega] \\
&= f(y; \theta_0) [1 + 1/2 \sum_{r=1}^l \sum_{s=1}^l (F_2^{rs} + F_1^r F_1^s) \omega_{rs}], \tag{5.2.2}
\end{aligned}$$

where F_2^{rs} , F_1^r are the $(r, s)^{th}$ and r^{th} element of $F_2(y; \theta_0)$ and $F_1(y; \theta_0)$, respectively. Now, Chesher and Cox have shown that for Ω close to zero, $h^a(y, \theta_0, \Omega)$ is non-negative and a proper density. Note that test for parameter variation in θ is equivalent to the test for Ω being a zero matrix. Now, we can perform score test on Ω considering $h^a(y, \theta_0, \Omega)$ as the appropriate density. However, as in our case, one may not be interested in the whole parameter vector, but only in a subvector of the parameter vector θ . Therefore, let us now consider the score test of $H_0: \omega_{rs} = 0, (r, s) \in A$ (an index set) against the alternative $H_1: \omega_{rs} \neq 0$ ² given $\omega_{uv} = 0$ for $(u, v) \notin A$. Let $\hat{\theta}_n$ be the likelihood estimate of θ_0 and $\hat{D}_A(\hat{\theta}_n)$ be the estimated score vector which consists of the elements $D_{rs}, (r, s) \in A$, and which can be written as $D_{rs} = \frac{1}{n} \sum_{i=1}^n \frac{\partial \ln h^a(y; \theta_0, \Omega)}{\partial \omega_{rs}}$ evaluated at $\Omega = 0$ and θ_0 replaced by the maximum likelihood estimate, $\hat{\theta}_n$. Let $\hat{V}(\hat{\theta}_n)$ be some consistent estimator of the asymptotic variance covariance matrix of $\sqrt{(n)}\hat{D}_A(\hat{\theta}_n)$. Then the score (Lagrange multiplier) test is obtained as $d = n\hat{D}'_A(\hat{\theta}_n)(\hat{V}(\hat{\theta}_n))^{-1}\hat{D}_A(\hat{\theta}_n)$, which is asymptotically distributed as $\chi^2(\text{central})$ under H_0 with degrees of freedom as the cardinality of A .

²Obviously, $\omega_{rs} > 0$ when $r = s$.

As noted by Bera and Kim (2001), the score function as obtained from equation (5.2.2), is nothing but the criterion function of the IM test which White (1982) originally suggested as a general model misspecification test. This interpretation of the IM test as the score test for the constancy of parameters of the localized likelihood of a model with varying parameters makes it possible to make use of the optimal local power property of the score test. Now, there are mainly two versions of the IM test—the outer product of gradient (OPG) version (Chesher, 1984; and Lancaster, 1984) and the "efficient score" version as proposed by Orme (1990). The OPG version of Chesher is relatively easy to compute because it does not require the third derivative of the log-density function or the analytic expected values of derivatives of the log-density, but Orme's version uses expected values of third derivatives of the log-density. Therefore, the latter is usually more difficult to compute than the Chesher-Lancaster statistic unless it turns out to be of a very simple functional form. The OPG version is very popular because of its computational simplicity, but it is also well-known that it suffers from size distortion problem (see, Taylor, 1987; and Chesher and Spady, 1991). Now, what we propose is to overcome this size distortion problem by using an appropriate bootstrap technique, and hence in our case, we suggest the OPG form of IM test.

Insofar as computation of the test statistic is concerned, we note that using White's (1982) expression for \hat{V}_n leads to White's information matrix (IM) test statistic, but it may be difficult to calculate this as it involves third derivatives of the log-density, $F(y, \theta)$, which appears in $Q_0 = -E(G_2)$, where $G_2 = \frac{\partial G_1}{\partial \psi'}$, $G_1 = \frac{\partial G}{\partial \psi}$, $G = \ln h^a(y; \theta_0, \Omega)$, ψ is the full parameter vector, i.e., $\psi = (\theta_0'; \Omega^c)'$, where the row order of $\Omega^c = \text{vech}(\Omega)$. However, a computationally simpler estimator of $V(\theta_0)$ is available from Chesher. This is given by \tilde{V}_n which is nothing but the inverse of the lower right $(\nu \times \nu)$ submatrix of \tilde{Q}_0^{-1} , $\tilde{Q}_0 = \frac{1}{n} \tilde{G}_1' \tilde{G}_1$, \tilde{G}_1 is the $n \times (l + \nu)$ matrix with i -th row equal to G_1'

with y replaced by y_t and θ_0 replaced by the ML estimator $\hat{\theta}_n$ under H_0 , $G'_1 = [F_1(y; \theta_0)': d(y; \theta_0)']$, $d(y; \theta_0) = [F_2(y; \theta) + F_1(y; \theta_n)F_1(y; \theta)']^c$, $\nu = l(l + 1)/2$ is the number of distinct elements in Ω and $\tilde{Q}_0^{-1} = \begin{pmatrix} \tilde{Q}_0^{\theta\theta} & \tilde{Q}_0^{\theta\Omega} \\ \tilde{Q}_0^{\Omega\theta} & \tilde{Q}_0^{\Omega\Omega} \end{pmatrix}$. The IM test statistic thus simplified is given as $\tilde{d} = nD'_n(\hat{\theta}_n)\tilde{V}_n^{-1}D_n(\hat{\theta}_n)$, which, in turn, reduces to the well-known expression *viz.*, nR^2 , where R^2 is the squared value of the coefficient of multiple determination from the least squares estimation of dependent variable i_n , a column vector of ones, on the matrix of independent variable \tilde{G}_1 .

5.2.2 Test for Constancy of the Relative Risk Aversion Parameter

Since our primary interest lies in the relative risk aversion parameter λ , we derive the explicit form of the test statistic for testing the null hypothesis of constancy for λ . For the purpose of this derivation, we have considered the ARCH-M model (*cf.* equation 1.1.2) along with a GARCH (1,1) process for the conditional variance h_t . Thus, the ARCH-M model along with $h_t \sim GARCH(1, 1)$ may be restated as follows

$$y_t = x'_t\beta + \lambda h_t + \epsilon_t,$$

$\epsilon_t | \Psi_{t-1} \Rightarrow N(0, h_t)$, and

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \phi_1 h_{t-1}.$$

Now, let σ_λ^2 be the variance of the relative risk aversion parameter λ which is now assumed to be a random variable.

Let $l(\theta)$ be the conditional log-likelihood function of y_1, y_2, \dots, y_n , and

$l_t(\theta)$ be the conditional log-density of the t -th observations,, $t = 1, 2, \dots, n$.

Then obviously,

$$l(\theta) = \frac{1}{n} \sum_{t=1}^n l_t(\theta).$$

Now, for our model $l_t(\theta) = \text{constant} - \frac{1}{2} \ln h_t - \frac{\epsilon_t^2}{2h_t}$, since $\epsilon_t | \Psi_{t-1}$ is assumed to follow $N(0, h_t)$.

Now, let $l(\theta_0, \Omega)$ be the log-likelihood for n independent realisations from $h(y; \theta_0, \Omega)$. Then the average score for $\sigma_\lambda^2 = 0$ and $\theta_0 = \hat{\theta}_n$, the maximum likelihood estimator of θ_0 under H_0 , is

$$D_n(\hat{\theta}_n) = \frac{1}{n} \sum_{t=1}^n d(y_t; \hat{\theta}_n). \quad (5.2.3)$$

Now it is quite evident that from (5.2.3) that

$$d(y_t; \hat{\theta}_n) = \frac{1}{2} \left[\frac{\partial^2 l_t(\theta)}{\partial \lambda^2} \Big|_{\hat{\theta}_n} + \left(\frac{\partial}{\partial \lambda} l_t(\theta) \right)^2 \Big|_{\hat{\theta}_n} \right].$$

Given the form of the score function in (5.2.3), it is easy to find \tilde{V}_n and hence the test statistic under $H_0 : \sigma_\lambda^2 = 0$. Since the expression of the IM test statistic (OPG version) would involves $\frac{\partial l_t}{\partial \lambda}$, $\frac{\partial h_t}{\partial \lambda}$, $\frac{\partial^2 l_t}{\partial \lambda^2}$ and $\frac{\partial^2 h_t}{\partial \lambda^2}$, their exact expressions are given below.

$$\frac{\partial l_t(\theta)}{\partial \lambda} = \frac{1}{2h_t} \left(\frac{\epsilon_t^2}{h_t} - 1 + 2\lambda\epsilon_t \right) \frac{\partial h_t}{\partial \lambda} + \epsilon_t,$$

$$\begin{aligned} \frac{\partial h_t}{\partial \lambda} &= \frac{\partial}{\partial \lambda} (\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \phi_1 h_{t-1}) \\ &= -2\alpha_1 \epsilon_{t-1} h_{t-1} + (\phi_1 - 2\alpha_1 \lambda \epsilon_{t-1}) \frac{\partial h_{t-1}}{\partial \lambda}, \end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 l_t(\theta)}{\partial \lambda^2} &= \frac{\partial}{\partial \lambda} \left(\frac{\partial l_t(\theta)}{\partial \lambda} \right) \\
&= \frac{\partial}{\partial \lambda} \left[\frac{\epsilon_t^2}{2h_t^2} - \frac{1}{2h_t} + \frac{\lambda \epsilon_t}{h_t} \right] \frac{\partial h_t}{\partial \lambda} + \frac{\partial \epsilon_t}{\partial \lambda} \\
&= \frac{\partial^2 h_t}{\partial \lambda^2} \left[\frac{\epsilon_t^2}{2h_t^2} - \frac{1}{2h_t} + \frac{\lambda \epsilon_t}{h_t} \right] + \frac{\partial h_t}{\partial \lambda} \left[\frac{\epsilon_t - \lambda h_t}{h_t} - \lambda - \frac{\lambda \epsilon_t h_t + \epsilon_t^2}{h_t^3} \right] \\
&+ \left(\frac{\partial h_t}{\partial \lambda} \right)^2 \left[\frac{\lambda h_t + \lambda \epsilon_t}{h_t^2} \right] - \left(\frac{\epsilon_t}{h_t} + h_t \right), \text{ and}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 h_t}{\partial \lambda^2} &= \frac{\partial}{\partial \lambda} \left(\frac{\partial h_t}{\partial \lambda} \right) \\
&= \frac{\partial}{\partial \lambda} \left[-2\alpha_1 \epsilon_{t-1} h_{t-1} + (\phi_1 - 2\alpha_1 \lambda \epsilon_{t-1}) \frac{\partial h_{t-1}}{\partial \lambda} \right] \\
&= 2\alpha_1 h_{t-1}^2 + \frac{\partial h_{t-1}}{\partial \lambda} [4\alpha_1 \lambda h_{t-1} + \epsilon_{t-1} - 2\alpha_1 \epsilon_{t-1}] + \frac{\partial^2 h_{t-1}}{\partial \lambda^2} [\phi_1 - 2\alpha_1 \lambda \epsilon_{t-1}] \\
&+ \left(\frac{\partial h_{t-1}}{\partial \lambda} \right)^2 2\alpha_1 \lambda^2.
\end{aligned}$$

Obviously, \tilde{d} will follow a χ_1^2 distribution under $H_0 : \sigma_\lambda^2 = 0$.

5.3 Monte Carlo Results

In the last section we have derived the IM test in OPG form for testing the null hypothesis of constancy of the relative risk aversion parameter. We now examine the finite sample performance of this test statistic through a Monte Carlo experiment. The computations were carried out using a program written in GAUSS. For the purpose of studying the performance of the test in terms of its size, we generated y_t 's from the following GARCH-M model:

$$y_t = -0.475y_{t-1} + 0.165h_t + \epsilon_t, \quad (5.3.1)$$

$$h_t = 7.54 + 0.56\epsilon_{t-1}^2 + 0.018h_{t-1}, \quad (5.3.2)$$

Table 5.3.1: Estimated Type-I Error Probabilities ($\lambda = 0.165$)

Nominal Size	Sample Size						
	100	500	600	700	800	900	1000
1%	0.499	0.137	0.082	0.047	0.035	0.012	0.008
2.5%	0.588	0.193	0.123	0.091	0.064	0.027	0.015
5%	0.666	0.276	0.187	0.122	0.101	0.061	0.045
10%	0.772	0.415	0.330	0.232	0.172	0.123	0.094

where $\epsilon_t/\sqrt{h_t}$'s were obtained as random normal deviates from GAUSS random number generator. In order to study the sensitivity of the results we also considered another model where the relative risk aversion parameter λ in (5.3.1) was taken to be -0.165 while all other parameter values in (5.3.1) and (5.3.2) remained the same. We considered samples of size 100, 500, 600, 700, 800, 900, and 1000; and for each of these sample sizes 1000 replications were taken. Obviously, therefore, the maximum possible standard error of the estimated sizes reported in the tables will be $\sqrt{0.5(1 - 0.5)/1000} = 0.0158$. However, in practice the standard error of the estimated size would often be well below 0.0158. The empirical sizes in the tables give (in percentage) number of times (out of 1000) the computed value of the proposed test statistic exceeds the critical values indicated by the nominal sizes (1%, 2.5%, 5%, and 10%) based on χ^2 distribution with one degree of freedom.

From these two Tables 5.3.1 and 5.3.2, it is observed that the test (OPG version) rejects the null hypothesis far too often at all nominal levels considered, although its performance does improve as the sample size increases. We further note that the nominal sizes are more or less attained only when the sample size reaches 1000, which, as it is, seems to be a very large sample

Table 5.3.2: Estimated Type-I Error Probabilities ($\lambda = -0.165$)

Nominal size	Sample size						
	100	500	600	700	800	900	1000
1%	0.511	0.128	0.111	0.046	0.032	0.014	0.011
2.5%	0.586	0.193	0.161	0.090	0.061	0.029	0.018
5%	0.648	0.274	0.232	0.124	0.100	0.056	0.051
10%	0.751	0.434	0.349	0.243	0.181	0.120	0.101

size, in general, but not with financial time series like stock prices which are available at high frequency level. It is thus seen that the test suffers from the well-known over size problem of the OPG version for both the models.

Since there is hardly any use in studying the power performance of the test for sample sizes for which there are extreme size distortions, our investigation on the power performance of the test focussed on samples of size 800, 900, and 1000 only, for which no major size distortion has occurred. For the purpose of power calculation λ_t was taken to be $\lambda_t = \lambda_0 + u_t$ where u_t was assumed to have been generated by $U(0, .8)$ and $U(0, 1)$, referring to two different models under the alternative hypothesis. It may be noted that σ_λ^2 for these two uniform distributions are $\frac{1}{12}$ and $\frac{4}{75}$, respectively. Excepting this all other aspects of data generation remained the same.

The results of power analysis are summarized in Tables 5.3.3, 5.3.4, 5.3.5 and 5.3.6. Tables 5.3.3 and 5.3.5 show that the test seems to have good power when the sample size is reasonably large. Further, we find from the results that power of the test marginally decreases as sample size increases, which is contrary to normal findings. However, the power of the test as calculated may not be very reliable since the test is somewhat biased in size for sample

Table 5.3.3: Estimated Powers ($\lambda_0 = 0.165$ and $u_t \sim U(0, 1)$)

Nominal Size	Sample Size		
	800	900	1000
1%	0.726	0.712	0.700
2.5%	0.832	0.804	0.802
5%	0.910	0.888	0.862
10%	0.970	0.950	0.944

Table 5.3.4: Estimated Size-Corrected Powers ($\lambda_0 = 0.165$ and $u_t \sim U(0, 1)$)

Nominal Size	Sample Size		
	800	900	1000
1%	0.597	0.641	0.711
2.5%	0.713	0.793	0.814
5%	0.799	0.826	0.870
10%	0.847	0.910	0.952

sizes 800 and 900. In order to obtain valid power performance of the test we calculated *size-adjusted* power. The empirical powers of the test with respect to empirical size-corrected critical values are presented in Tables 5.3.4 and 5.3.6. It may be noted that the empirical size-corrected critical values are nothing but the critical values obtained by simulation from the model under the null hypothesis with pseudo-true values of the parameters.

It is evident from Tables 5.3.4 and 5.3.6 that once the size of the test has thus been corrected, the test performs remarkably well in term of consistency property i.e., the power has increased as the sample has increased from 800 to 900 and then to 1000.

Table 5.3.5: Estimated Powers ($\lambda_0 = 0.165$ and $u_t \sim U(0, 0.8)$)

	Sample size		
Nominal size	800	900	1000
1%	0.502	0.465	0.389
2.5%	0.604	0.593	0.506
5%	0.706	0.707	0.634
10%	0.846	0.855	0.809

Table 5.3.6: Estimated Size-Corrected Powers ($\lambda_0 = 0.165$ and $u_t \sim U(0, 0.8)$)

	Sample Size		
Nominal Size	800	900	1000
1%	0.301	0.346	0.401
2.5%	0.368	0.402	0.518
5%	0.493	0.594	0.654
10%	0.612	0.718	0.820

5.4 Bootstrapping the Test for Size Correction

We have already discussed and noted that the IM test in OPG version suffers from serious over size problem resulting in rejection of the null hypothesis far too often than those understood by the assumed nominal level. Some attempts have been made to obtain a size corrected test. Chesher and Spady (1991) have suggested that the critical values be obtained from the Edgeworth expansion through $O(n^{-1})$ of the finite sample distribution of the test statistic. Horowitz (1994), on the otherhand, have proposed a bootstrap-based method to get the size-corrected test. Now, since derivations involving the Edgeworth expansion is quite tedious, we follow the computer-intensive method of Horowitz (1994). In this section we report the results of a limited Monte Carlo study of the proposed IM test where critical values are obtained by using bootstrap procedure as suggested by Horowitz. However, this procedure cannot be applied straightway since in our case the underlying ARCH-M involves conditional heteroscedasticity. Therefore, the usual bootstrapping of the data in this case would not be sensible as bootstrap samples will not be representatives of the original sample simply because the original sample is time dependent, while bootstrap samples are independent. In order to take care of such problems of dependence in the data, especially with the time series data, residual bootstrapping has been suggested in the literature (see, Freedman, 1981, 1984; Bickel and Freedman, 1981; Lahiri, 1992; and Bose, 1988). Following this procedure we bootstrap the standardized residuals.

The sample sizes and the number of replications considered for the Monte Carlo experiment were somewhat limited by the fact that very long computing times are involved. Monte Carlo experiments with bootstrapping are very

time consuming because at each replication of the estimation sample there are many inner replications in which the parameters of the model under null the hypothesis are re-estimated from bootstrap samples. We thus considered samples of size, 100, 200, and 300 only, and each case 1000 replications were taken. For each replication, the number of inner-replications, called the bootstrap size and denoted by B , was taken as 100. The data generation was carried out for the model stated in equations 5.3.1 and 5.3.2. Following Horowitz, we describe below the steps involved in this experiment. Each experiment consisted of repeating the following steps 1000 times:

Step I: We generated an estimation data of size n by random sampling from the model under consideration. This data is, in fact, the same data set that was used in the Monte Carlo study in the previous section, with $\lambda = 0.165$. We then estimated the parameters of the model and calculated the value of the test statistic. Let this be called \tilde{d}_0 .

Step II: The residual series $\hat{\epsilon}_t$ was obtained. The standardized residuals, $\epsilon_t^* = \frac{\hat{\epsilon}_t - \bar{\hat{\epsilon}}}{\sqrt{\hat{h}_t}}$, where $\bar{\hat{\epsilon}}$ is the simple arithmetic mean of the residual series and \hat{h}_t is the estimated conditional variance at time point t , were then computed. A bootstrap sample of size n was generated by random sampling from the *standardized residual* series, and then the corresponding residuals were obtained by using the relation $\hat{\epsilon}_t = \bar{\hat{\epsilon}} + \epsilon_t^* \sqrt{\hat{h}_t}$. Finally, the bootstrap samples of the data were obtained by using the model under consideration but using the estimated parameter values in Step I instead of the true values.

Step III: Using the bootstrap sample of data as obtained in Step II, the model was again estimated and the test statistic \tilde{d} was computed. We then estimated the critical values of the IM test at 2.5 per cent, 5 per cent and 10 per cent significance levels from the empirical distribution of \tilde{d} that was obtained by repeating this step 100 times. Let us denote by $\hat{c}_{n,\alpha}$ the estimated α -level critical value. We reject the model being tested at the nominal level α

Table 5.4.7: Estimated Sizes Based on Bootstrapping ($\lambda = 0.165$)

Nominal Size	Sample Size		
	100	200	300
2.5%	0.037	0.035	0.023
5%	0.061	0.050	0.053
10%	0.092	0.074	0.082

if $\tilde{d}_0 > c_{\hat{n},\alpha}$.

The estimated size figures corrected by bootstrap technique are presented in Table 5.4.7. We find from the entries of this table that the oversize problem of the proposed IM test in OPG version is greatly reduced. For instance, while for $n = 100$ (with nominal level as 2.5 per cent), the estimated size was 0.588 (*cf.* Table 5.3.1), the size of the test with bootstrap correction is 0.037. In fact, most of the entries in Table 5.4.7 corroborate this finding. Thus, we find that these corrected sizes are reduced more or less to their respective nominal sizes.

5.5 ARCH-M with Time Varying Risk Aversion Parameter

In the last section we have described a test for testing the constancy of the relative risk aversion parameter in ARCH-M framework. Now, an obvious modelling issue arises when the proposed IM test suggests in an actual financial data analysis, that the null hypothesis of constancy of relative risk aversion parameter is not tenable. In other words, the question that arises is how the basic ARCH-M model should be extended so that the time vary-

ing nature of λ is explicitly incorporated. In this section, we consider such a model where time varying λ , say λ_t , is assumed to be generated by a stochastic process. As Harvey (1984) has stated, there are basically three classes of models available in the literature which may be adopted to model the stochastic nature of λ_t . First, *random coefficient model* the parameters of which are assumed to vary randomly about a fixed but unknown mean. Second, what Harvey has called as *return to normality model* where λ_t is assumed to follow a stationary ARMA process so that parameters change gradually rather than in a haphazard fashion. Third, *random walk model* where λ_t is not constrained to have a fixed mean, and the model can gradually evolve over time. Now, as advocated by Chou *et al.* (1992), random walk model is quite appropriate for modelling ARCH-M with time varying parameter, and accordingly, we also consider this framework for our analysis.

Given that the relative risk aversion parameter is no longer constant, and that λ_t follows a random walk process, we now have what is called time varying parameter ARCH-M (TVP-ARCH-M) model, and this is represented by the following three equations.

$$y_t = \lambda_t h_t + \epsilon_t, \quad (5.5.1)$$

$$\lambda_t = \lambda_{t-1} + \eta_t, \quad (5.5.2)$$

$$h_t = \tilde{\alpha}_0 + \tilde{\alpha}_1 \nu_{t-1}^2 + \dots + \tilde{\alpha}_q \nu_{t-q}^2 + \phi_1 h_{t-1} + \dots + \phi_p h_{t-p}, \quad (5.5.3)$$

$$\tilde{\alpha}_0 > 0, \tilde{\alpha}_i \geq 0, i = 1, 2, \dots, q, \phi_j \geq 0, j = 1, 2, \dots, p, \sum_{i=1}^q \tilde{\alpha}_i + \sum_{j=1}^p \phi_j < 1,$$

and the disturbances ϵ_t and η_t are assumed to be uncorrelated Gaussian noises with zero means and with variances h_t and Q , respectively. Obviously, this modelling set-up stated in equations (5.5.1) through (5.5.3) is nothing but the state-space model. The equations in (5.5.1) and (5.5.2) are known as the

measurement and the transition equations, respectively, and λ_t is called the state variable.

It may be pointed out that following Chou *et al.*, conditional heteroscedasticity h_t has been specified in (5.5.3) as a GARCH(p, q) model, but this specification of GARCH(p, q) is somewhat modified from the usual GARCH process in the sense that instead of the original squared prediction error ϵ_t^2 of (5.5.1), a newly defined prediction error, ν_t^2 , has been used in (5.5.3). This modification is necessary because both λ_t and h_t are not observable. Now, ν_t is determined by $\nu_t = y_t - E_{t-1}(\lambda_t)h_t = [\lambda_t - E_{t-1}(\lambda_t)]h_t + \epsilon_t$, where $E_{t-1}(\lambda_t)$ is the optimal forecast of λ_t given information up to time $t - 1$. It may be noted that when Q , the variance of the state variable λ_t , becomes small the TVP-ARCH-M model converges to the fixed parameter ARCH-M model.

Insofar as estimation of the parameters are concerned, simultaneous application of Kalman filter and maximum likelihood are carried out. It may be noted that the model in equations (5.5.1) through (5.5.3) has fixed parameters $\tilde{\alpha}_0, \tilde{\alpha}_1, \tilde{\alpha}_2, \dots, \tilde{\alpha}_q, \phi_1, \phi_2, \dots, \phi_p$ and Q which are to be estimated along with the state λ_t , and the variance of ϵ_t i.e., h_t . Estimates of the states are obtained by the Kalman filter conditional on the parameter values. And the variance of the error ϵ_t in measurement equation, h_t , can be obtained through the GARCH equation in (5.5.3), provided parameter values are given. After each pass of the Kalman filter and the GARCH equation, the maximum value of the likelihood can be obtained. These steps are repeated until convergence is achieved.

As the Kalman algorithm requires the initial values for all the fixed parameters as well as for the state variable and the variance of ϵ_t , the parameter estimates from the usual i.e., fixed parameter ARCH-M model can be used as the initial values for the fixed parameters and h_0 . As regards the initial value of the state, λ_0 , a non-informative prior distribution is assumed; i.e., we

assign a large value (say, 1000) to its variance.

5.6 An Application to SENSEX Data

We applied the test of constancy of the relative risk aversion parameter on SENSEX data covering the period 1st week of January 1984 till 3rd week of October 2000, the details of which have been given in Section 2.2 of Chapter 2. The results of this application along with those concerning estimation of the suggested TVP-ARCH-M model with this data set are presented and discussed in this section.

5.6.1 Results of the Test

We may recall from Section 3 of Chapter 3 that the best fitted ARCH-M model with fixed relative risk aversion parameter was found to be as follows.

$$y_t = -0.457y_{t-1} + 0.165h_t + \epsilon_t, \quad l(\hat{\theta}) = -2490.99, \quad (5.6.1)$$

(-13.5) (4.7)

$$h_t = 7.54 + 0.56\epsilon_{t-1}^2 + 0.018h_{t-1}, \quad (5.6.2)$$

(20.5) (2.73) (2.59)

(t-statistic values are given in parentheses)

Although we have discussed earlier as to why λ might vary over time, it is worth mentioning that for an emerging economy like the Indian economy this is all the more likely to be so. Over the last one and half decades the Indian stock market has evolved in a spectacular fashion. In particular, after 1990, due to liberal economic policies pursued by the Government of India, Indian

stock market has undergone reforms in a significant and big way. Moreover, during the course of time, the changes in consumption and investment pattern of the economic agents might also have influenced the risk aversion parameter. In order to find if there is indeed statistical support in this direction from the data, we applied the IM test (OPG version) for constancy of λ , as developed and discussed in Section 5.2. The test statistic value has been found to be 30.21, which is well above the 5 per cent critical value of 3.84. Hence we conclude that the null hypothesis of time invariance of the relative risk aversion parameter is strongly rejected.

Now, we have discussed earlier that this test has severe over-size problem, and that bootstrap technique may be used to reduce this size distortion. To conduct bootstrap experiment with this data set, we generated 500 bootstrap samples (i.e., inner replications). The 5 per cent critical value from the bootstrap distribution thus obtained was 8.12. Therefore, the bootstrap based critical value also indicates that there is parameter variation in λ . We have thus found the test to be very useful for SENSEX data, and the conclusion based on an analysis of this data set is that the relative risk aversion parameter for the Indian stock market is time varying.

5.6.2 Estimated Time Varying Parameter ARCH-M Model

Since the test concludes that the temporal stability of the relative risk aversion parameter is not tenable for SENSEX data, we begin by examining the temporal instability of λ — first by rolling regression and then by fitting TVP-ARCH-M model.

The Rolling Sample Estimation

For rolling sample estimates, we started with the first 250³ observations,

³In this technique, relatively short sample periods are normally used.

and then subsequently added one new observation and excluded the very last observation, keeping the number of observations for each easimate as fixed. ARCH-M coefficients were estimated for each such rolling sample. The rolling estimation procedure thus yield a weekly time series of $\hat{\lambda}$. These are plotted in Figure 5.6.1. We observe from this plot that the series is quite time varying. The coefficient ranges from -0.1796 to 0.2059, with a mean of 0.0199 and standard deviation 0.0901. Therefore, the instability of λ in ARCH-M model is established. The somewhat erratic behaviour, although showing a decreasing tendency at some points, of this coefficient also signifies the inadequacy of the fixed parameter ARCH-M model. In other words, rolling sample technique indicates that a time varying approach may be appropriate. In this context it is also worthwhile to note that rolling sample based estimated series is only an approximation as it uses relatively small sample period, and it is unlikely that weekly-to-weekly changes of the coefficient would be moderately large. It also suffers from the fact that this methodology implicitly assumes parameter constancy for every such period while at the same time estimating a time varying parameter model. To check if the number of observations has any bearing on the conclusion based on rolling sample estimates, we also experimented with samples of size 275, 300, 325, 350, 375 and 400. Interestingly, more or less similar time varying pattern were found.

Results of TVP-ARCH-M Model Estimation

The TVP-ARCH-M model specified in (5.5.1) through (5.5.3) was estimated for this data set ⁴. The computations which involve simultaneous applications of Kalman filter and maximum likelihood method, were carried out using the procedure 'TD_ml' of TSM written in GAUSS.

The autoregressive parameter in the mean equation has been treated as a fixed parameter . As earlier stated, we first used the variance of the diffuse

⁴For the GARCH(p, q) process, p and q were taken as 1.

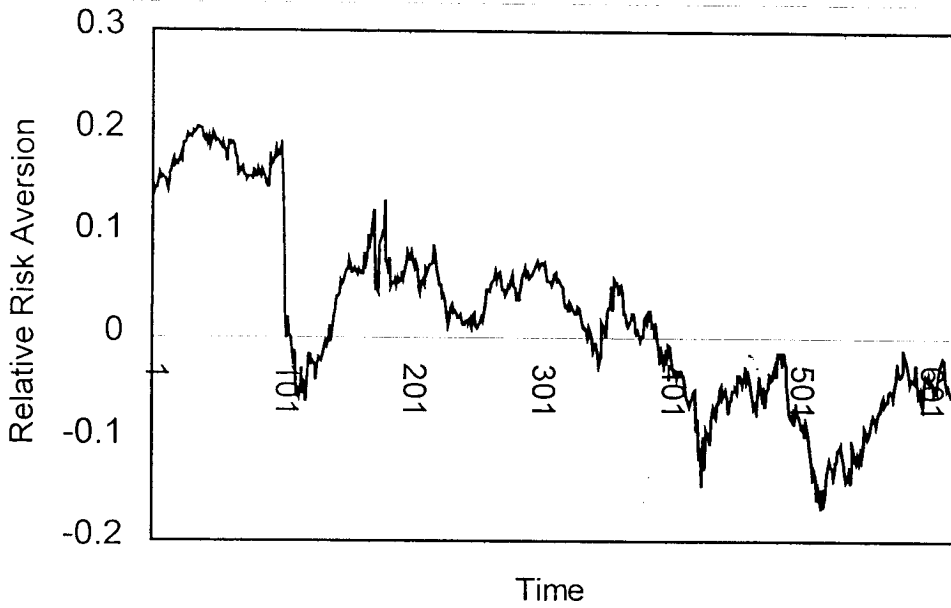


Figure 5.6.1: Rolling Sample Based Estimate of Relative Risk Aversion Parameter

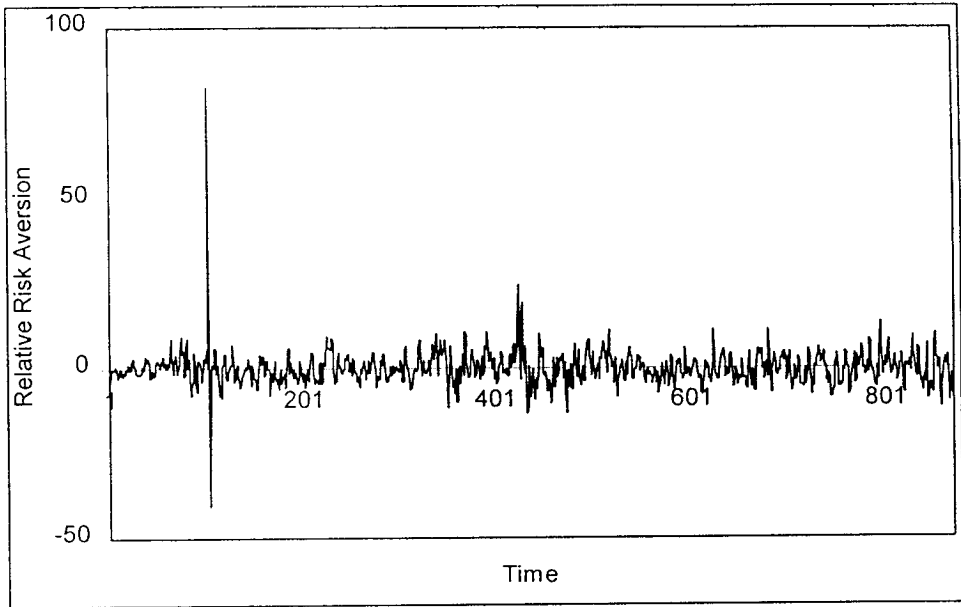


Figure 5.6.2: TVP-ARCH-M Based Estimate of Relative Risk Aversion Parameter

prior distribution to be 1000, which is quite likely. Taking the parameter estimates obtained from the fixed parameter ARCH-M (FP-ARCH-M) model (*cf.* equations (5.6.1) and (5.6.2)) as the initial values for TVP-ARCH-M, the final converged values of the parameters in the conditional variance equation, $\tilde{\alpha}_0$, $\tilde{\alpha}_1$ and ϕ_1 , were, 6.61, 0.11 and 0.001, respectively, which are, incidentally, obtained as not very close to the estimates of the FP-ARCH-M model. The estimated value of Q was also found to very small being 0.17 only. Figure 5.6.2 gives the description of the time path of λ based on this TVP-ARCH-M model. This plot also suggests that λ varies over time. The large variation so obtained in the earlier periods of the sample may be due to the fact that at those points not much information from the data is used. We have also carried out computations with another diffuse with variance of the distribution being 100; however, the features of the plot remained more or less unchanged. It may be pointed out that both the methods-rolling regression and TVP-ARCH-M modelling- suggest and capture the time varying nature of the relative risk aversion parameter of Indian stock market as represented by SENSEX. The slightly decreasing values of the parameter at some time points by rolling regression may be attributed to the fact that while this procedure assumes no specific process for λ_t and it is only an approximation, Kalman filter-ML method of estimation is highly sophisticated, and it assumes some process of generation for λ_t .

5.7 Conclusions

In this chapter, we have proposed an information matrix based (OPG version) test for testing the constancy of the relative risk aversion parameter in ARCH-M model. We have also carried out a detailed simulation study to evaluate the performance of the test in terms of size and power. This test is found to

suffers from size problem. To correct this size distortion we have suggested the application of bootstrap technique and we found that bootstrap method substantially improves the over size problem. The proposed test has been applied on the returns data based on SENSEX , and it has been found that there is sufficient evidence to conclude that the relative risk aversion parameter λ in the Indian stock market is not constant over time. Further, we have studied the temporal behaviour of λ by using rolling regression and also by fitting time varying parameter ARCH-M model. Both these methods confirm the time varying nature of λ in the Indian stock market. While this findings viz., the time variation of the mean/ variance ratio is very interesting, it opens up, at the same time, the issue of providing explanations for this. The explanations may lie in the role of other relevent variables like risky assets other than stocks, or for that matter, the role of rate of inflation in stock price movements. Obviously, this calls for more detailed investigations, and future work in this direction is likely to be very fruitful.

Chapter 6

Conclusions and Future Outlook

This thesis has studied some issues on time varying risk premium in ARCH-M model. Since Engle, Lilien and Robins 's paper in 1987, the importance of ARCH-M model in studying time varying risk premium is very well recognized. While there have been numerous applications of ARCH-M model in financial economics, only few theoretical advancements concerning this model have occurred. Hence, to our understanding there were scopes for further research into some theoretical aspects of ARCH-M modelling, and such work would hopefully further enhance the usefulness of this important model in the time varying risk premium literature.

Towards this end, this thesis has primarily focussed on two broad aspects of ARCH-M modelling. The first one relates to studying some statistical properties of the estimators of the parameters of ARCH(q)-M or, for that matter GARCH(p, q)-M model. The other one is concerned with studying the model after specifying the risk premium term to be as generalized and as appropriate as possible. Now, the last one can be achieved in at least two ways—(i) by considering a very generalized functional form for conditional heteroscedasticity, and (ii) by assuming the relative risk aversion parameter itself to be time dependent. In this work we have examined theoretically all the relevant

issues involved in these aspects of ARCH-M modelling and derived appropriate theoretical results. We have also carried out simulation studies whenever appropriate and possible, so that small-sample properties of the estimators of the parameters of the proposed models or of the tests of hypotheses involving parameters of interest, particularly the relative risk aversion parameter λ , could be available. Further, we have applied the proposed models and the tests to an actual data set on Indian stock price index, called the SENSEX. The major findings and observations of these studies along with limitations, if any, are summarized below.

6.1 Major Findings

We have just stated that the usefulness and appropriateness of the proposed models as well as of the proposed tests of hypotheses of the relevant parameters of ARCH-M model has been studied with an actual data set. The data set used for this purpose is the most widely-used Indian stock price index, called the sensitive index (SENSEX); the span of the data covers 1st week of January 1984 till 3rd week of October 2000. In Chapter 2, we have provided finer insights into the properties of this data set. The properties of this univariate time series studied in this chapter include stationarity, independence, normality, long-memory and tail behaviour along with existence of higher order moments.

We first found, based on unit root tests, that SENSEX is nonstationary, but return based on SENSEX is stationary. This finding on stationarity is in line with those of other major stock indices. Various normality tests like Kolmogorov-Smirnov test, Jarque-Bera test, normal probability plot, etc. strongly suggest that the return distribution is far from being normal. Dependency— both linear and nonlinear —was studied by applying

autocorrelation-based test, Ljung-Box test, various nonparametric tests like turning point test, run test, rank version of the von Neumann ratio test, and finally the BDS test. The results of these tests suggest that there is strong dependency in the returns. To test for the presence of long-memory in the data, we used both the classical and the modified R/S tests. Neither test suggests the presence of long-range dependence in the return data. This means any shock to the series is purely temporary, and does not persist for a long horizon. Finally, we used Hill (1975) estimator to study the existence of moments as characterized by the tail index. Since Hill estimator is very sensitive to the choice of number of tail observations used, we performed a Monte Carlo study as well as a bootstrap based study to find the optimal number of extreme observations for the Hill estimator. We found from this exercise that while the existence of second moment is strongly established, the third and other higher order moments may not, in fact, exist for the return data. We may thus conclude, on the basis of these findings, that heavy tail distributions like those based on conditional heteroscedasticity, student's t-distribution etc. are likely to be more appropriate for return data.

While asymptotic properties of the estimators of the parameters of ARCH/GARCH model have been studied extensively using likelihood/quasi-likelihood approach, there have been no significant studies on ARCH-M/GARCH-M model. Notable exceptions are Engle *et al.* (1987) and Lee (1991) who have used likelihood and quasi-likelihood approaches. It can be noted that (Gaussian) likelihood based estimation and inference is sometimes non-robust, since the procedures adopted typically suit only the likelihood of choice. Also, it is often difficult to verify assumptions concerning the likelihood. In Chapter 3, we have undertaken a different approach. We have used a very generalized approach called the estimating equations approach, originally due to Godambe (1985). The advantage of this approach is that the standard likelihood and

quasi-likelihood based approaches are special cases of this approach.

Further, from consideration of studying the properties of estimators of ARCH-M model we have also applied bootstrap technique in ARCH-M model. In this chapter we have first discussed the bootstrap technique which could be appropriate for ARCH-M model, and then studied some statistical properties of the bootstrap estimator.

The first major result in this chapter is in the form of a theorem where we have proved the consistency and asymptotic normality of the estimators of the parameters of GARCH(p, q)-M model obtained by using the estimating equation approach. Thereafter, we have proved the consistency property of the bootstrap estimator. A detailed simulation exercise was carried out to study the performance of bootstrap technique in ARCH-M model. These results along with those obtained from the usual normal asymptotic approximation have been compared with each other to judge the relative performance of these two approaches in small samples. The finding on this count is that bootstrap distribution outperforms the asymptotic approximation in terms of confidence level or coverage probability for relative risk aversion parameter λ . Insofar as illustration with real data is concerned, we considered weekly returns based on SENSEX. We first obtained the most appropriate GARCH-M model for this data set. The value of λ was found to be highly significant. Bootstrap methodology was then used to obtain the bootstrap distribution of $\hat{\lambda}$, and here again the significance of λ was confirmed.

In this context, we may point out that while we have proved the consistency of bootstrap estimator, and also shown, through a Monte Carlo study, that it is more efficient than the corresponding normal asymptotic approximation in terms of coverage probability, we have not been able to theoretically establish the superiority of the former over the latter.

The choice of proper functional form h_t for representing risk premium is

very important. In the literature on time varying risk premium researchers have often used conditional variance h_t or $\sqrt{h_t}$ or $\ln h_t$ to represent risk premium. Since a generalized and flexible functional form of h_t to represent risk premium should be more useful and appropriate, we have proposed in the Chapter 4 a generalization of GARCH-M model by considering the Box-Cox power transformation of the conditional variance representing the risk premium in the model. Obviously, various other standard functional forms are special cases of this generalization. We have studied the properties of this model, and also described the ML method of estimation of the parameters of this model. Thereafter, we have provided an empirical illustration with SENSEX data to demonstrate that the suggested generalization provides improvement over the standard ARCH-M model. The maximized value of the log-likelihood has indeed improved significantly for the generalized model as compared to the standard ARCH-M model.

It has been recognized for quite some time that the assumption of relative risk aversion parameter λ being time invariant is quite restrictive, and may not indeed be true in many practical situations. There are empirical evidences also towards this end. The relative risk aversion parameter might change due to either change in the perception towards risk or change in the distribution of the wealth or both. This issue is the focus of Chapter 5. Here we have first considered the problem of testing constancy of λ . Towards this end, we have proposed a test which is based on Chesher's interpretation (OPG version) of White's information matrix test. The size and power of the test have then been studied through a detailed simulation study. It is well-known that this test (OPG version) suffers from over size problem. To correct this distortion, we have applied appropriate bootstrap technique i.e., we have used the critical value obtained from bootstrap distribution, and found that bootstrap technique have substantially reduced the over size problem. The application

of the test to SENSEX data has shown that λ is not constant over time in the context of Indian stock market. We have also suggested a model, called the time varying parameter ARCH-M (TVP-ARCH-M) model, in the line of Chou, Engle and Kane (1992), where λ has been assumed to follow a random walk process in the framework of a state-space model. The temporal behaviour of the relative risk aversion parameter λ has been studied in the context of TVP-ARCH-M model. Rolling regression technique has also been applied to study the temporal behaviour of λ . Both the methods have once again confirmed the time varying nature of λ in the context of Indian stock market.

It may be pointed out that having recognized the oversize problem of the proposed test, we tried to overcome this by using an appropriate bootstrap methodology. Our attempt to derive Orme's efficient score version of the test was not really successful.

6.2 Directions for Further Research

Since its inception in 1987, ARCH-M model has become the most important model in studying risk-return relationship in the time varying risk premium framework. As we have discussed in this thesis, researchers would continue to be interested in this model, and hence further researches concerning this model would continue to be done. In the particular context of this thesis where we have worked on some aspects of ARCH-M modelling, we feel that there are scopes for further work in the following directions.

(i) We have used estimating equations approach to establish the consistency and asymptotic normality property of the estimators of the parameters of GARCH(p, q)-M model. Estimating equations approach is a very generalized approach, and it has the advantage that the standard likelihood and quasi-likelihood approaches are special cases of this approach. Since estima-

tion and testing based on (Gaussian) likelihood is sometimes non-robust as the procedure chosen suits only the likelihood of choice, it would be an interesting study if different criteria of estimation are used, and the statistical properties of the resulting estimators are compared amongst the different criteria via the estimating equations approach.

(ii) Resampling procedure like bootstrap technique has not been used so far in the context of ARCH/ARCH-M model. Apparently, there are some difficulties, and hence it is not at all obvious whether the different types of bootstrap would provide consistent results or not. We have considered a particular "generalized bootstrap", and proved the consistency property of the resulting bootstrap estimator. We think there are immense possibilities in this area in future work. For instance, various aspects of application of bootstrap technique in this class of models may be investigated not only with respect to the relative risk aversion parameter but also with respect to other parameters of the model. It is also very important to study the performance of standard bootstrap procedures like residual bootstrap, parametric bootstrap, and find why these are not directly applicable (if that is indeed the case) in ARCH/ARCH-M models. Further, it would be exciting to try to modify some of these bootstrap procedures so that these could be appropriate for ARCH/ARCH-M models. Finally, any attempt to prove, at a theoretical level, the superiority of bootstrap distribution over the normal asymptotic approximation would be a very useful exercise.

(iii) The most important parameter in the ARCH-M model is the relative risk aversion parameter λ , and as we have discussed, this parameter could very well be time varying. With the test proposed in this thesis, it should now be possible to find if, in a given situation, λ is indeed time invariant or not. When we find λ to be time varying, there should be some explanations for this to happen. Future work can be done in this direction. Through em-

empirical exercises, the role of various economic/ financial variables like returns on risky assets other than stock, rate of inflation, portfolio weights etc can be investigated in determining expected stock returns. It would also be useful if further insight into the time varying nature of λ could be established. A further direction of research concerning time invariance test for λ could be in the direction of deriving other more accurate tests like Orme's efficient score version of the IM test.

References

- Abel, A. B. (1988) Stock prices under time-varying dividend risk: An exact solution in an infinite-horizon general equilibrium model, *Journal of Monetary Economics*, 22, 375-393.
- Amsler, Christine (1984) Term structure variance bounds and time varying liquidity premia, *Economics Letters*, 6, 137 - 44.
- Anderson, T. W., and D. A. Darling (1954) A test of goodness of fit, *Journal of the American Statistical Association*, 49, 765-769.
- Andrews, D. W. K. (1993) Exactly median unbiased estimation of first order autoregressive/ unit root models, *Econometrica*, 61, 139-165.
- Andrews, D. W. K., and D. Chen (1994) Approximately median unbiased estimation of autoregressive models, *Journal of Business and Economic Statistics*, 12, 187-204.
- Arellano, C. , and S. G. Pantula (1995) Testing for trend stationarity versus difference stationarity, *Journal of Time Series Analysis*, 16, 147-164.
- Backus, D.V., and A.W. Gregory (1993) Theoretical relations between risk premiums and conditional variances, *Journal Business and Economic Statistics*, 11, 177 - 85.

- Baillie, R. T., and R. P. DeGennaro (1990) Stock returns and volatility, *Journal of Financial and Quantitative Analysis*, 25, 203-21.
- Baillie, R. T., and T. Bollerslev (1990) A multivariate generalized ARCH approach to modelling risk premia in foreign exchange markets, *Journal of International Money and Finance*, 9, 309-324.
- Barbe, P., and P. Bertail (1995) The weighted bootstrap, *Lecture Notes in Statistics*, 98, Springer-Verlag, New York.
- Bartels, R. (1977) Estimation of a first order autoregressive scheme with non-normal stable distributions, *Journal of Statistical Computation and Simulation*, 6, 35-48.
- Bartels, R. (1982) The rank version of Von Neumann's ratio test for randomness, *Journal of the American Statistical Association*, 77, 40-46.
- Bera, A.K., and M.L. Higgins (1993) A survey of ARCH models properties, estimation and testing, *Journal of Economic Surveys*, 7, 305-66.
- Bera, A.K., and S. Kim (2001) Testing constancy of correlation and other specifications of the BGARCH model with an application to international equity returns, *Journal of Empirical Finance*, Forthcoming.
- Bera, A.K., S. Ra, and N. Sarkar (1998) Hypothesis testing for some nonregular cases in econometrics, in *Econometrics: Theory and Practice*, Edited by S. Chakrabarty, D. Coondoo and R. Mukharjee, Allied Publishers, pp. 319-351.
- Bera, A.K., and A. Ullah (1991) Raos score test in econometrics, *Journal of Quantitative Economics*, 7, 189-220.

- Beran, R. (1987) Prepivoting to reduce level error of confidence sets, *Biometrika*, 74, 457-468.
- Beran, R. (1988) Prepivoting test statistics: a bootstrap view of asymptotic refinements, *Journal of the American Statistical Association*, 83, 687-697.
- Berndt, E.K., B.H. Hall, R.E. Hall, and J.A. Hausman (1974) Estimation and inference in nonlinear structural models, *Annals of Economic and Social Measurement*, 4, 653 - 665.
- Bickel, P. J., and D.A. Freedman (1981) Some asymptotic theory for the bootstrap, *Annals of Statistics*, 9, 1196-1217.
- Black, A., and P. Fraser (1995) U.K. Stock returns predictability and business conditions, *Manchester School of Economics and Social Studies* , 63(0), Supplement , 85-102.
- Black, F., and M. Scholes (1973) The pricing of options and corporate liabilities, *Journal of Political Economy*, 81, 637-659.
- Bollerslev, T. P. (1986) Generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics*, 31, 307-327.
- Bollerslev, T. P. (1988) On the correlation structure of the generalized autoregressive conditional heteroscedastic process, *Journal of Time Series Analysis*, 9, 121-131.
- Bollerslev, T., and J.M. Wooldridge (1992) Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances, *Econometric Reviews*, 11, 143-179

- Bollerslev, T., R.F. Engle, and D.B. Nelson (1994) ARCH Models, in R.F. Engle and D.L. McFadden (eds.) *Handbook of Econometrics*, Vol. IV, Elsevier Science.
- Bollerslev, T., R.F. Engle, and J.M. Wooldridge (1988) A capital asset pricing model with time varying covariances, *Journal of Political Economy*, 96, 116 - 31.
- Bollerslev, T., R.Y. Chou, and K.F. Kroner (1992) ARCH modelling in finance : A review of the theory and empirical evidences, *Journal of Econometrics*, 52, 5 - 59.
- Boos, D., and J.F. Monahan (1986) Bootstrap methods using prior information, *Biometrika*, 73, 77-83.
- Borovskikh, Yu. V., and V.S. Korolyuk (1997) Martingale approximation, VSP, Utrecht.
- Bose, A (1988) Edgeworth correction by bootstrap in autoregressions, *Annals of Statistics*, 16, 1709-1722.
- Bose, A., and S. Chatterjee (2001) Generalized bootstrap in non-regular M -estimation problems, to appear in *Statistics and Probability Letters*.
- Box, G.E.P., and D.R. Cox (1964) An analysis of transformations, *Journal of Royal Statistical Society, Series B*, 26, 211 - 43.
- Box, G.E.P., and G.M. Jenkins (1976) Time series analysis : Forecasting and control, Revised edition, San Francisco: Holden-Day.
- Box, G. E. P., and D.A. Pierce (1970) Distribution of residual autocorrelations in autoregressive moving average time series models, *Journal of the American Statistical Association*, 65, 1509-1526.

- Breitung, J., and W. Meyer (1994) Testing for unit roots in panel data: are wages on different bargaining levels cointegrated? *Applied Economics*, 26, 353-361.
- Brock, W. A., and E. G. Baek (1991) Some theory of statistical inference for nonlinear science, *Review of Economic Studies*, 58, 697-716.
- Brock, W. A., W. Dechert, and J. Scheinkman (1986) A test of independence based on the correlation dimension, *SSRI Report No. 8702*, Department of Economics,, University of Wisconsin, University of Houston, and University of Chicago.
- Campbell, John Y., and L. Hentschel, (1992) No news is good news: an asymmetric model of changing volatility in stock returns, *Journal of Financial Economics*, 31, 281-318.
- Chan, K.C., G.A. Karolyi, and R.M. Slutz (1992) Global financial markets and the risk premium on U.S. equity, *Journal of Financial Economics*, 32, 137-167.
- Chatterjee, S. (1999) Generalised bootstrap techniques, Ph. D. thesis, Indian Statistical Institute, Calcutta.
- Chatterjee, S., and A. Bose (2000) Generalised bootstrap for estimating equations, preprint.
- Chesher, A. (1983) The information Matrix Test : Simplified calculation via a score test Interpretation, *Economics Letters*, 13, 45-48.
- Chesher, A. (1984) Testing for neglected heterogeneity, *Econometrica*, 52, 865-872.

- Chesher, A., and R. Spady (1991) Asymptotic expansions of the information matrix test statistics, *Econometrica*, 59, 787-816.
- Chou, R., R.F. Engle, and A. Kane (1992) Measuring risk aversion from excess returns on a stock index, *Journal of Econometrics*, 52, 201-224.
- Cox, D. R. (1966) The null distribution of the first order serial correlation coefficient, *Biometrika*, 53, 623-626.
- Cox, D.R. (1983) Some remarks on overdispersion, *Biometrika*, 70, 269-274.
- Crowder, M.J. (1976) Maximum likelihood estimation for dependent observations, *Journal of the Royal Statistical Society, Series B*, 38, 45 - 53.
- Csorgo, S., and D.M. Nason (1989) Bootstrap empirical functions, *Annals of Statistics*, 17, 1447-1471.
- D'Agostino. R. B., and M. A. Stephens. (1986) Goodness-of-fit techniques, New-York, Marcel Dekker.
- Davidson, A.C., and D.V. Hinkley (1997) Bootstrap methods and their applications, Cambridge University Press, Cambridge.
- Dejong, D. N., J. C. Nankervis, N. E. Savin, and C. H. Whiteman (1992a) The power problems of unit root tests in time series with autoregressive errors, *Journal of Econometrics*, 53, 323-343.
- Dejong, D. N., J. C. Nankervis, N. E. Savin, and C. H. Whiteman (1992b) Integration versus trend trend stationarity in time series, *Econometrica*, 60, 423-433.

- Dickey, D. A., and W. A. Fuller (1979) Distribution of the estimators for autoregressive time series with unit root, *Journal of the American Statistical Association*, 74, 427-431 .
- Dickey, D. A., and W. A. Fuller (1981) Likelihood ratio statistics for autoregressive time series with unit root, *Econometrica*, 49, 1057-1072.
- Diebold, F.X., and M. Nerlove (1990) Unit roots in economic time series; a selective survey, in G.F. Rhodes and T.B. Fomby (eds.) *Advances in Econometrics*, 8, Greenwich, CT, JAI Press, 3-69.
- Domowitz, I., and C.S. Hakkio (1985) Conditional variance and the risk premium in the foreign exchange market, *Journal of International Economics*, 19, 47 - 66.
- Efron, B. (1979) Bootstrap methods : Another look at the jackknife, *Annals of Statistics* , 7, 101-118.
- Efron, B. (1982) The jackknife, the bootstrap and other resampling plans, SIAM, Philadelphia.
- Efron, B. (1992) Six questions raised by the bootstrap, in exploring the limits of bootstrap, ed. Le Page and Billard, Wiley, New York.
- Efron, B., and R. Le Page (1992) Introduction to bootstrap, in exploring the limits of bootstrap, ed. Le Page and Billard, Wiley, New York.
- Efron, B., and R. Tibshirani (1986) Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy, *Statistical Science* 1(1), 54-77.
- Efron, B., and R. Tibshirani (1993) An introduction to the bootstrap, Chapman and Hall, New York.

- Efron, B., and R. Tibshirani (1998) The problem of regions, *Annals Statistics*, 26, 1687-1718.
- Elliott, G., T. J. Rothenberg, and J. H. Stock (1996) Efficient tests for an autoregressive unit root, *Econometrica*, 64, 813-836.
- Ellul, A. (1999) As you like it: an investigation of trading behaviour and price volatility on auction and dealership market architectures, Unpublished Mimeo, London School of Economics.
- Elyasiani, E., and I. Mansur (1999) Sensitivity of the bank stock returns distribution to change in the level and volatility of interest rate: a GARCH-M model, *Journal of Banking and Finance*, 22, 535-563.
- Engle R.F., D.M. Lilien, and R.P. Robins (1987) Estimating time varying risk premium in the term structure : The ARCH-M model, *Econometrica*, 55, 391-407.
- Engle, R. F., V. Ng, and M. Rothschild (1990) Asset pricing with FACTOR- ARCH covariance structure: empirical estimates for treasury bills, *Journal of Econometrics*, 45, 213-237
- Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of variance of U.K. inflation, *Econometrica*, 50, 987-1008.
- Engle, R.F., and T. Bollerslev (1986) Modelling the persistence of conditional variances, *Econometric Reviews*, 5, 1-87.
- Engle, R.F., D.M. Lilien, and R.P. Robins (1987) Estimating time varying risk premium in the term structure: The ARCH-M model, *Econometrica*, 55, 391-407.

- Fama, E., and G. MacBeth (1973): Risk, Return and equilibrium: empirical test, *Journal of Political Economy*, 81, 607-636.
- Fama, E. (1965) The behaviour of stock market prices, *Journal of Business*, 38, 34-105.
- Fama, E., and K. French (1988) Permanent and temporary components of stock prices, *Journal of Political Economy*, 96, 246-273.
- Fama, E., and G. W. Schwert (1977) Asset returns and inflation, *Journal of Financial Economics*, 5, 115-146.
- Ferson, W., S. Kendel, and R.F. Stambaugh (1987) Tests of asset pricing with time-varying expected risk premiums and market betas, *Journal of Finance*, 42, 201-220.
- Fraser, P. (1996) UK excess share returns: firm size, and volatility, *Scottish Journal of Political Economy*, 43, 71-74.
- Freedman, D.A. (1981) Bootstrapping regression models, *Annals of Statistics*, 9, 1218-1228.
- Freedman, D.A. (1984) On bootstrapping two stage least squares estimators in stationary linear models, *Annals of Statistics*, 12, 827-852.
- Freedman, D.A., and S.C. Peters (1984) Bootstrapping a regression equation: Some empirical results, *Journal of the American Statistical Association*, 79, 97-106.
- French, K.R., G. Schwert, and R. Stambaugh, (1987) Expected stock returns and volatility, *Journal of Financial Economics*, 19,3-30.
- Fuller, W. A. (1976) Introduction to Statistical Time series, NY, Wiley.

- Gewke, J. (1986) Modeling persistence of conditional variances: comment, *Econometric Reviews*, 5, 57-61.
- Gibbons. J. D. (1971) Nonparametric statistical inference, McGraw-Hill Kogakusha, LTD.
- Gibbons, M.R.(1982)Multivariate tests of financial models: A new approach,*Journal of Financial Economics*, 10,3-28.
- Gibbons, M.R., and W. Ferson(1985) Testing asset pricing models with changing expectations and an unobservable market portfolio, *Journal of Financial Economics*, 14, 217-236.
- Gine, E., and J. Zinn (1989) Necessary conditions for the bootstrap of the mean, *Annals of Statistics*, 17, 684-691.
- Gine, E., and J. Zinn (1991) Bootstrap general empirical measures, *Annals of Probability*, 18, 851-869.
- Glosten,L.R.,R. Jagannathan, and D.E. Runkle (1993) On the relation between the expected and the volatility of the nominal excess return on stocks, *The Journal of Finance*, XLVIII,1779-1801.
- Godambe, V. P. (1985) The foundations of finite sample estimation in stochastic processes, *Biometrika*, 72, 419-428.
- Godambe, V. P. (1991) (Edited) Estimating Functions, Clarendon Press, Oxford.
- Godambe, V. P., and B.K. Kale (1991) Estimating functions: an overview. in (Eds.) Estimating Functions by Godambe,V. P., Clarendon Press, Oxford.

- Goncalves, E., and N.M. Lopes (1998) Some statistical results on autoregressive conditionally heteroscedastic models, *Journal of Statistical Planning and Inference*, 68, 193-202.
- Gouriéroux, C. (1997) ARCH model and financial applications, Springer, New York.
- Granger, C.W.J., and A.P. Anderson (1978) An introduction to bilinear time series models, Gottingen:Vaden hoeck and Reprecht.
- Granger, C.W.J., and P. Newbold (1976) Forecasting transformed series, *Journal of the Royal Statistical Society*, B, 38, 189 - 203.
- Guilkey, D.K., and P. Schmidt (1989) Extended tabulations for Dickey-Fuller Tests, *Economics Letters*, 31, 355-357.
- Gweke, J. (1986) Modelling the persistence of conditional variances : Comment. *Econometric Reviews*, 5, 57 - 61.
- Gweke, J., and S. Porter-Hudak (1983) The estimation and the application of long memory time series models, *Journal of Time Series Analysis*, 4, 221-238.
- Hall, A. (1987) The information matrix test for the linear model, *Review of Economic Studies*, 54, 257-263.
- 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.
- Hall, P. (1988) Theoretical comparison of bootstrap confidence intervals, *Annals of Statistics*, 16, 927-953.

- Hall, P. (1990) Using the bootstrap to estimate mean square error and select smoothing parameter in nonparametric problem, *Journal of Multivariate Analysis*, 32, 177-203.
- Hall, P. (1992) The bootstrap and Edgeworth expansion, Springer-Verlag, New York.
- Hall, P., and E. Mammen (1994) On general resampling algorithms and their performance in distribution estimation, *Annals of Statistics*, 22, 2011-2030.
- Hall, P., and C.C. Heyde (1980) Martingale limit theory and its application, Academic Press, New York.
- Hall, S.G., D.K. Miles, and M.P. Taylor (1988) A multivariate GARCH in mean estimation of the capital asset pricing model, *Bank of England Discussion paper, Technical Series*, 19.
- Hall, P., and S.R. Wilson (1991) Two guidelines for bootstrap hypothesis testing, *Biometrics*, 47, 757-762.
- Hannan, E. J., and M. Kanter (1977) Autoregressive processes with infinite variance, *Journal of Applied Probability*, 14, 411-415.
- Hansson, B., and P. Hordahl, (1997) Changing risk premia: evidence from a small open economy, *Scandinavian Journal of Economics*, 99(2), 335-350.
- Hardle, W., and S. Marron (1991) Bootstrap simultaneous error bars for non-parametric regression, *Annals of Statistics*, 19, 778-796.
- Harvey, C. (1984) Time series models, Heritage publication, India.

- Harvey, C. (1989) Time varying conditional covariances in tests of asset pricing models, *Journal of Financial Economics*, 24, 289 - 317.
- Harvey, C. (1991) The specification of conditional expectations, Duke University, Department of Economics (mimeo).
- Hentschel, L. (1995) All in the Family : Nesting symmetric and asymmetric GARCH models, *Journal of Financial Economics*, 39, 71-104.
- Higgins, M. L., and A.K. Bera (1992) A class of non-linear ARCH models, *International Economic Reviews*, 33, 137-158.
- Hill, B. M. (1975) A simple general approach to inference about the tail of a distribution, *Annals of Statistics*, 3, 1163-1173.
- Hodrick, R. J., and S. Srivastava (1984) An investigation of risk and return in forward foreign exchange rates, *Journal of International Money and Finance*, 3, 5 - 29.
- Hong, E.P. (1991) The autocorrelation structure for GARCH-M process, *Economics Letters*, 37, 129 - 32.
- Horowitz, J.H. (1994) Bootstrap-based critical values for the information matrix test, *Journal of Econometrics*, 61, 395-411.
- Hsieh, D.A. (1989) Testing non-linear dependence in daily foreign exchanges rates, *Journal of Business*, 62, 339 - 68.
- Hurst, H. (1951) Long term storage capacity of reservoirs, *Transactions of the American Society of Civil Engineers*, 116, 770-799.
- Huisman, R., K.G. Koedijk, J.M. Kool, and F. Palm (2001) Tail-index estimates in small samples, *Journal of Business and Economic Statistics*, 19, 208-217.

- Jarque. C. M., and A. K. Bera (1987) A test for normality of observations and regression residuals, *International Statistical Review*, 55, 163-172.
- Keenan, D. M. (1985) A Tukey nonadditivity -type test for time series non-linearity, *Biometrika*, 72, 39-44.
- Kendall, J. D. (1989) Role of exchange rate volatility in U.S. import price pass-through relationships, unpublished Ph.D. dissertation, Department of Economics, University of California, Davis, CA.
- Kendall, M. G., and M. A. Stuart (1961) The advanced theory of statistics, volume 2, London, Griffin.
- Koul, H.L. (1992) Weighted empiricals and linear models, IMS lecture notes monograph series, 21.
- Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992) Testing the null hypothesis of stationary against the alternative of a unit root, *Journal of Econometrics*, 54, 159-178.
- Lahiri, S.N. (1992) Bootstrapping M-estimators of a multiple linear regression parameter, *Annals of Statistics*, 20, 1548-1570.
- Lancaster, T. (1984) The covariance matrix of the information matrix test, *Econometrica*, 52, 1051-1053.
- Lee, S. W. (1991) Asymptotic Properties of the Maximum-Likelihood Estimator of the GARCH-M and IGARCH-M Models, Mimeo, Department of Economics, University of Rochester.
- Leybourne, S. J. (1995) Testing for unit roots using forward and reverse Dickey-Fuller Regressions, *Oxford Bulletin of Economics and Statistics*,

57, 559-571.

- Li, H. , and G.S. Maddala (1996) Bootstrapping time series models, with discussion, *Econometric Reviews* 15, 115-195.
- Lima, P.J.F. (1997) On the robustness of nonlinearity tests to moment condition failure, *Journal of Econometrics*, 76, 251-280.
- Lintner, J. (1965) The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics*, 47, 13 -37
- Liu, R., and K. Singh (1992) Efficiency and robustness in resampling, *Annals of Statistics*, 20, 370-384.
- Liu, R. (1988) Bootstrap procedures under some non-i.i.d. models, *Annals of Statistics*, 16, 1696-1708.
- Ljung, G. M., and G. E. P. Box (1978) On a measure of the lack of fit in the time series models, *Biometrika*, 65, 297-303.
- Lo, A. Y. (1987) A large sample study of the Bayesian bootstrap, *Annals of Statistics*, 15, 360-375.
- Lo, A. Y. (1991) Bayesian bootstrap clones and a biometry function, *Sankhya, Series A*, 53, 320-333.
- Lo, A. (1991) Long term memory in stock market prices, *Econometrica*, 59, 1279-1313.
- Lo, A., and C. MacKinlay (1988) Stock market prices do not follow random walks: Evidence from a simple specification test, *Review of Financial Studies*, 1, 41-66.

- Lucas, R.E. (1978) Asset prices in an exchange economy, *Econometrica*, 46, 1429-1445.
- Lumsdaine, R. L. (1996) Consistency and asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models, *Econometrica*, 64, 575-596.
- Mackinnon, J.G. (1990) Critical values for cointegration tests, *US San Diego Discussion Paper*, 90-94.
- Mammen, E. (1992 a) When does bootstrap work? Asymptotic results and simulations, *Lecture Notes in Statistics*, 77, Springer-Verlag, New York.
- Mammen, E. (1992b) Bootstrap, wild bootstrap, and asymptotic normality, *Probability Theory and Related Fields*, 93, 439-455.
- Mammen, E. (1993) Bootstrap, wild bootstrap, and high dimensional linear models, *Annals of Statistics*, 21, 255-285.
- Mammen, E. (1995) Bootstrap, wild bootstrap, and generalised bootstrap, preprint available from <ftp://statlab-uni-heidelberg.de/pub/reports/by-series/beitrag11.ps>.
- Mammen, E. (1996) Empirical process of residuals for high dimensional linear models, *Annals of Statistics*, 24, 307-335.
- Mandelbrot, B. B. (1962) Sur certain prix speculatifs: Faits empiriques et modele base sur les processus stables additifs de Paul Levy, *Comptes Rendus*, 254, 3968-3970.
- Mandelbrot, B. B. (1963) The variation of certain speculative prices, *Journal of Business*, 26, 394-419.

- Mandelbrot, B. B. (1967) The variation of some other speculative prices, *Journal of Business*, 40, 393-413.
- Mandelbrot, B. B. (1971) When can price be arbitrated efficiently? A limit to the of the random walk and martingale models, *Review of Economics and Statistics*, 53, 225-236.
- Mandelbrot, B. B. (1972) Statistical methodology for non-periodic cycles: from the covariance to R/S analysis, *Annals of Economic and Social Measurement*, 1, 259-290.
- Mandelbrot, B. B., and J. Wallis (1969) Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence, *Water Resources Research*, 5, 967-988.
- Markowitz, H. (1959) Portfolio selection: efficient diversification of investments, John Wiley, New York.
- Moran, P.A.P. (1971) The uniform consistency of maximum-likelihood estimators, *Proceedings of Cambridge Philosophical Society*, 70, 435-439.
- Mason, D. M., and M.A. Newton (1992) A rank statistics approach to the consistency of a general bootstrap, *Annals of Statistics*, 20, 1611-1624.
- McLeod, A. I., and W. K. Li. (1983) Diagnostic checking ARMA time series models using squared residuals autocorrelations, *Journal of Time Series Analysis*, 4, 269-273.
- Merton, R.C. (1986) On estimating the expected return on the market : An exploratory investigation. *Journal of Financial Economics*, 8, 323 - 61.

- Mittnik., S., and S. T. Rachev (1993) Modelling asset returns with alternative stable distribution, *Econometric Reviews*, 12, 261-330.
- Mossin, J. (1966) Equilibrium in a capital asset market, *Econometrica*, 34, 768-783.
- Navidi, W. (1989) Edgeworth expansions for bootstrapping regression models, *Annals of Statistics*, 17, 1472-1478.
- Nelson, C .R., and C. I. Plosser (1982) Trends and random walk in macroeconomic time series, *Journal of Monetary Economics*, 10, 139-162.
- Nelson, C. R., and H. Kang (1981) Spurious periodicity in inappropriately detrended time series, *Journal of Monetary Economics*, 10, 139-162.
- Nelson, D. B. (1990) Conditional heteroscedasticity in asset returns: a new approach, *Econometrica*, 59, 347-370.
- Newton, M. A., and A.E. Raftery (1994) Approximate Bayesian inference with the weighted likelihood bootstrap (with discussion), *Journal of the Royal Statistical Society*, B, 56, 3-48.
- Ng, S., and P. Perron (1995) Unit root tests in ARMA models with data dependent methods for the selection of the truncated lag, *Journal of the American Statistical Association*, 90, 268-281.
- Officer, Robert.R. (1973) The variability of the market factor of New York Stock Exchange, *Journal of Business*, 46, 434-453.
- Orme, C. (1990) The small sample performance of the information matrix test, *Journal of Econometrics*, 46, 309-331.

- Ortega, J. M. and W.C. Rheinboldt (1970) Iterative solution of nonlinear equations in several variables, Academic Press, New York.
- Pagan, A.R., and A. Ullah (1988) The econometric analysis of models with risk terms, *Journal of Applied Econometrics*, 3, 87 - 105.
- Pagan, A.R., and Y.S. Hong (1991) Nonparametric estimation and the risk premium, in W.A. Barnett, J. Powell and G. Tauchen (eds.), Nonparametric and Semiparametric Methods in Econometrics and Statistics, Cambridge University Press, Cambridge.
- Pantula, S.G. (1986) Modelling the persistence of conditional variances : Comment. *Econometric Reviews*, 5, 79 - 97.
- Park, J. Y. (1990) Testing unit roots and cointegration by variable addition, in T. B. Fomby and G. F. Rodes (eds.), *Advances in Econometrics*, 8, 107-133.
- Pearson, E. S., and H. O. Hartley. (1966) *Biometrika Tables For Statisticians* (vol. 1, 3rd ed.) New york, Cambridge University Press.
- Perron, P. (1989) The great crash, the oil price shock, and the unit root hypotheses, *Econometrica*, 57, 1361-1401.
- Perron, P. and S. Ng (1996) Useful Modifications to some unit root tests with dependent errors and their local asymptotic properties, *Review of Economic Studies*, 63, 435-465.
- Pesando, J.E. (1983) On expectations, term premiums, and the volatility of long-term interest rates *Jour. Monetary Economics*, 12, 467 - 74.
- Phillips, P. C. B. (1987) Time series regression with unit roots, *Econometrica*, 57, 277-302.

- Phillips, P. C. B., and P. Perron (1988) Testing for a unit root in time series regression, *Biometrika*, 75, 335-346.
- Pindyck, R.S. (1984) Risk, inflation, and the stock market, *American Economic Review*, 74, 335-351.
- Pindyck, R.S. (1988) Risk aversion and determinants of stock market behaviour, *Review of Economics and Statistics*, LXX, 183-190.
- Politis, D.N., J.P. Romano, and M. Wolf (1999) Subsampling, Springer-Verlag, New York.
- Poterba, J., and L. Summers (1988) Mean reversion in stock returns: evidence and implications, *Journal of Financial Economics*, 22, 27-60.
- Praestgaard, J., and A.J. Wellner (1993) Exchangeably weighted bootstrap of the general empirical process, *Annals of Probability*, 21, 2053-2086.
- Rao, C.R. (1948) Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation, *Proceedings of the Cambridge Philosophical Society*, 44, 50-57.
- Rao, C.R., and L.C. Zhao (1992) Approximations to the distributions of M-estimates in linear models by randomly weighted bootstrap, *Sankhya, Series A*, 54, 323-331.
- Resnick, S. I. (1997) Heavy tail modelling and teletraffic data, *Annals of Statistics*, 25, 1805-1869.
- Rich, R.W., J. Raymond, and J.S. Butler (1991) Generalized instrumental variables estimation of autoregressive conditional heteroscedastic models, *Economics Letters*, 35, 179-185.

- Rosenblatt, M. (1956) A central limit theorem and a strong mixing condition, *Proceedings of the National Academy of Sciences*, 42,43-47.
- Rubin, D. B. (1981) The Bayesian bootstrap, *Annals of Statistics*, 9, 130-134.
- Rudebusch, G. (1992) Trends and random walks in macroeconomic time series: A re-investigation, *International Economic Review*, 33, 661-680.
- Said, S. E., and D. A. Dickey (1984) Testing for unit roots in autoregressive-moving average models of unknown order, *Biometrika*, 71, 599-607.
- Sakai H., and H. Tokumaru (1980) Autocorrelations of a certain chaos, *IEEE, Transactions on Acoustics, Speech and Signal Processing*, 1, 588-590.
- Sargan, J. D., and A. Bhargava (1983) Testing residuals from linear least squares regression for being generated by the Gaussian random walk, *Econometrica*, 51, 153-174.
- Scheikman, J.A., and B. LeBaron (1989) Non-linear dynamics and stock returns, *Journal of Business*, 62 , 311-37.
- Schwert, G. W. (1989) Tests for Unit roots: A Monte Carlo investigation, *Journal of Business and Economic Statistics*, 7, 147-159.
- Self, S., and K. Liang (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions, *Journal of the American Statistical Association*, 82, 605-610.
- Shao. J., and D. Tu (1995) The jackknife and bootstrap, Springer-Verlag, New York.

- Shapiro, S. S., and R. S. Francia. (1972) An analysis of variance test for normality, *Journal of the American Statistical Association*, 67, 215-216.
- Shapiro, S. S., and M. B. Wilk. (1965) An analysis of variance test for normality, *Biometrika*, 52, 591-611.
- Sharpe, W. (1964) Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance*, 19, 425-442.
- Shephard, N. (1996) Statistical aspects of ARCH an stochastic volatility, in *Time series models in Econometrics, Finance and other fields*, Ed. by Cox, Hinkley and Barndorff-Nielsen, 1-67.
- Shiller, R. J. (1979) The volatility of long-term interest rates and expectations models of the term structure, *Journal of Political Economy*, 87, 1190-1219.
- Stambaugh, R. (1982) On the exclusion of assets from tests of two parameter model, *Journal of Financial Economics*, 10, 237-268.
- Subba Rao, T., and M. M. Gabr (1980) A test for linearity of stationary time series, *Journal of Time Series Analysis*, 1, 145-148.
- Taylor, L. (1987) The size bias of whites information matrix test, *Economics Letters*, 24, 63-67.
- Thavaneswaran, A., and C.C. Heyde (1999) Prediction via estimating functions, *Journal of Statistical Planning and Inference*, 77, 89-101.
- Tsay, R. S. (1987) Conditional Heteroscedastic time series models, *Journal of the American Statistical Association*, 82, 590-604.
- Tukey, J.W. (1958) Bias and confidence in not quite large samples, *Annals of Mathematical Statistics*, 29, 614.

- Veall, M. R. (1987) Bootstrapping the probability distribution of peak electricity demand, *International Economic Review*, 28, 203-212.
- Weiss, A. A. (1986) Asymptotic Theory for ARCH Models: Estimation and Testing, *Econometric Theory*, 2, 107-131.
- Weng, C.S. (1989) On a second order asymptotic property of the Bayesian bootstrap mean, *Annals of Statistics*, 17, 705-710.
- White, H. (1982) Maximum likelihood estimation of Misspecified models, *Econometrica*, 50, 1-25.
- Yajima, Y. (1988) On estimation of a regression model with long memory stationary errors, *Annals of Statistics*, 16, 791-807.
- Zheng, Z., and D. Tu (1988) Random weighting method in regression models, *Sci. Sinica, Series A*, 31, 1442-1459.

