Durham E-Theses

An investigation into the role and impact of the volume of trade in UK futures markets

Tomsett, Mark Philip

How to cite:

Use policy

Academic Support Office, Durham University, University Office, Old Elvet, Durham DH1 3HP
e-mail: e-theses.admin@dur.ac.uk Tel: +44 0191 334 6107
http://etheses.dur.ac.uk

AN INVESTIGATION INTO THE ROLE AND IMPACT OF THE VOLUME OF TRADE IN UK FUTURES MARKETS

MARK PHILIP TOMSETT

SUBMITTED IN PARTIAL REQUIREMENT FOR THE DEGREE OF PhD

DEPARTMENT OF ECONOMICS AND FINANCE

UNIVERSITY OF DURHAM

1999

Mark Philip Tomsett

An Investigation into the Role and Impact of the Volume of Trade in UK Futures Markets

Abstract

In this thesis a detailed examination is carried out into the role and impact of the volume of trade in UK futures markets. While the success of a market may be judged by the number of investors that it attracts, how does the behaviour of individuals influence such key variables as price volatility and the cost of trading? The empirical work carried out here allows a unique appreciation of issues that have important implications for policy makers, investors and the practitioner.

Motivated by a desire to understand whether volatility is destabilising or a reflection of fundamental factors, as well as the nature of the distribution of price returns, the relationship between volume and price movements is investigated in detail. The preliminary analysis suggests an important role for the flow of information which is confirmed by the rigorous testing of Anderson's (1996) specification of the Mixture of Distributions Hypothesis. The exploitation of this model allows an in-depth analysis of the information process including the identification of the informed and uninformed components of volume. There is also an investigation into the possibility that the volume statistic itself has an informative value. Using the Blume et al. (1994) approach the results suggest that, for a variety of futures contracts, the markets show a high degree of information dispersion.

The need to attract investors has never been more acute than in today's competitive financial environment. It is therefore important to obtain a good appreciation of the relationship between volume and the cost of trading. This thesis includes a comprehensive intra-day study of the relation within a simultaneous econometric framework that exploits state-space models to investigate how markets react to unexpected levels of trading. The results question the dominance of inventory cost models and suggest that patterns of trade have become more predictable since contract inception, despite increases in volume.

To Sam, Mum, Dad and Ben for all your love and support.

DECLARATION OF AUTHORSHIP AND COPYRIGHT

ACKNOWLEDGEMENTS

CONTENTS

LIST OF TABLES

CHAPTER FOUR

LIST OF FIGURES

CHAPTER FIVE

CHAPTER ONE: INTRODUCTION

1.1 BACKGROUND

The futures industry[1] has been associated with some of financial markets' darkest days, notably the 1987 crash. However, the popularity of its instruments continues to grow. This can be partially attributed to the two important social functions that they serve; the transference of risk and price discovery.

The concept of risk plays a key role in capital markets. The bearing of risk has its rewards, but investors have idiosyncratic risk preferences. Futures markets allow investors to meet their demand objectives by transferring risk from one individual unwilling to bear it to another with a higher level of risk tolerance; the matching of complementary capital requirements. The advantages of using futures markets for this purpose can only be properly judged by considering the alternatives.

One option is to use forward markets. They serve a similar purpose of allowing investors to hedge against movements in the value of the underlying asset. They suffer, however, from the difficulties inherent in having to find an individual to make the other side of a trade, and the danger of default by the counterparty to any contract agreement. The second option is to try to diversify away risk in the cash market. Here too, there is the problem of each investor having to search for a reciprocal trade. In addition, individuals are often prevented from quickly offloading unwanted inventory by short sales restrictions.

Futures markets have tried to address these problems, and thereby increase their attractiveness, by using standardised contracts, organising trading on centralised exchanges, and using clearing houses to monitor contract performance. They are also not subject to short sales restrictions.

---

[1] For a good guide to the principles of futures markets see Edwards and Ma (1992).

The second important role of futures markets is to provide information about prices. The price of a futures contract should reflect the expected future price of the underlying asset and represent a forecast based on the aggregate opinions of the investors in the market place. Such information may be vital to ensure the efficient allocation of resources. Firms, for example, would be able to plan future production schedules more effectively. Although there is some evidence that futures prices are not perfect predictors of future spot prices, they will continue to be used where other forecasting services are less reliable and/or more expensive.

Whether a futures market fulfils these two requirements is very much dependent on its ability to provide liquidity at relatively low cost. A futures market can help itself through the design of the contracts that it offers and the careful organisation of the trading process. Therefore, in order to survive and grow a market needs to instil confidence in its capacity to play the roles of effective risk management and price discovery. As Carlton (1984: 237), notes,

'[T]heir objective is to succeed by generating volume.'

Clearly, the more traders that are in a market the greater the liquidity as it becomes easier for investors to find reciprocal trades. However, although a futures market may be judged by the amount of investment that it generates, do we know anything about the impact of the volume of trade?

1.2 MOTIVATION AND THESIS PLAN

The motivation for this thesis is a desire to obtain a better understanding of the role and impact of the volume of trade in futures markets. A thorough understanding of the issues surrounding the trading decisions of individuals is important for both policy makers and investors. Unfortunately, as will become apparent, the existing literature is limited in its scope; concentrating on US markets, with a bias towards equities, and often informing policy recommendations based on anecdotal evidence. This study will address these weaknesses by carrying out a rigorous examination of the functioning of UK futures markets. It is unique in specifically considering the volume

of trade in such a context. This thesis looks, in particular, at the relationship between volume and price volatility, and volume and the cost of trading.

The four pieces of empirical work that are carried out in this thesis adopt a confirmationist view to econometric study. This approach represents a traditional econometric methodological approach that appears at odds with the more fashionable 'general to specific' modelling techniques originally advocated by Hendry (1979) and Hendry and Mizon (1978).

The process of econometric analysis within the traditional framework is to begin by forming a prediction generated from a main hypothesis. This prediction is then used to construct a regression specification that can be estimated using an appropriate method. Examination of the regression residuals for certain desirable characteristics indicates whether the main hypothesis can be tested. If these characteristics are not evident then this particular form of the testable specification of the main hypothesis is rejected.

One of the criticisms of this approach is that it is too easy to adopt a strategy of running regressions until a 'verifying equation' is discovered, either by changing the specification of the regression or by choosing an estimation technique that provides the 'right' result. It is important, therefore, to carry out any necessary modifications to a model within a structure that preserves the integrity of the study. Economic theory must play a role in this process. If a model fails because either the error terms exhibit systematic bias or the main hypothesis, as represented by the regression, is rejected, it is the underlying theory that must provide the driving force of any re-specification. As Darnell and Evans (1990) argue, statistical considerations only identify the need for re-specification. It is economic theory that identifies the direction of that change. Models are then tested within a culture of falsification. Repeated rejection of the main hypothesis allows the modeller to question the validity of the underlying theory.

The approach of Hendy-Mizon is the result of criticism of studies that have failed to adopt a strict strategy of falsification under the traditional model. As Darnell and Evans (1990) explain, their methodology centres on the view that although economic

theory may lead us to 'long-run equilibrium' behaviour, economic data is generated using dynamic representations of variables within regression models. The standard method of achieving this is to use lagged values of the variables in question. This is usually in the form of an overparameterised 'general' model specification, which is reduced to a more parsimonious model through a process of sequential testing.

In contrast to the strict traditional approach, this methodology approaches empirical study from a verificationist point of view. Indeed Darnell and Evans (1990) argue that the use of lags typically fails to reflect theory explicitly. They state (1990: 84)

> "The use of empirical analysis in the attempt to refute economic hypotheses requires far more careful selection of the original model that is demonstrated by those who advocate 'general to specific modelling."

These criticisms support the adoption of the traditional methodological process in this thesis.

Futures markets have been criticised for attracting investors whose herd mentality results in volatile contract prices. Such volatility is believed to have a destabilising influence on, not only derivative markets, but asset markets in general. The alternative view is that price movements merely reflect a change in the fundamental information set. In this sense one might expect links to exist between price volatility and economic activity.

A detailed understanding of this relationship is important for a number of reasons. As Karpoff (1987) points out, one of the most interesting issues relates to the structure of financial markets. The theoretical models that are used to analyse the volume-volatility relation use the rate of information flow to the market as a key determinant of the strength of any association between these two variables. Apart from the fact that the testing of these models has generally been unconvincing, the paucity of

studies of futures markets misses an opportunity to understand an asset that is quite distinct from other securities. There is implicit evidence to suggest[2] that information is impounded into futures prices at a faster rate than it is into equity prices. In addition, the particular characteristics of futures contracts noted above, and the relatively low costs of trading involved, are likely to attract a group of investors that differs from those who trade in the underlying asset. It has also been suggested[3] that it is the trading in futures markets that leads to the improved quality and speed of information flow to spot markets. Therefore, an understanding of the volume-volatility relation will not only provide an insight into the structure of futures markets from the point of view of the role and impact of volume, but it will also increase our appreciation of an asset with unique properties. The implications for investors and regulators are twofold. It will aid decision making with regard to whether derivatives are a 'safe' investment. It will also provide guidance to policy makers keen to avoid the inefficient allocation of resources that is the result of excessive speculation and the manipulation of prices.

The volume-volatility relation is also a crucial element in the debate over the distribution of prices. As Karpoff (1987) notes, empirical studies suggest that price returns follow a leptokurtic distribution. The theoretical models in this field argue that it is important to distinguish between real time and event time. They hypothesise that the real time phenomenon of non-normality is the result of the arrival of pieces of information. However, the frequency of these arrivals is measured in so-called event time. If this information is brought to the market by investors it is not unreasonable to assume that the volume of trade has an important role to play in describing price distributions.

It is these key issues that motivate the second chapter of this thesis. It looks in detail at the relationship between volume and volatility for a range of UK futures contracts, each of which have their own characteristics in terms of the type of trader they attract, the number of expiration dates per year, seasonal factors etc. By utilising standard

---

[2] See Antoniou and Holmes (1995).
[3] See Antoniou and Holmes (1995).

econometric techniques it offers a re-evaluation of the current literature within the context of derivative markets. Indeed, in the process of examining the themes discussed above it becomes evident that further investigation of the volume-volatility relationship is necessary before a clear picture of the importance of the volume of trade can be obtained.

The study carried out in chapter 2 is, to some extent, traditional in its approach because it concentrates on explaining the role of volume by looking at its interaction with other economic variables. Chapter 3 looks at volume from a slightly different angle and considers whether there is information inherent in the volume statistic? The study looks, in particular, at issues of information precision and dispersion among investors in UK futures markets. The discovery that volume has an informative role would have important implications for the behaviour of investors who traditionally concentrate on price movements to construct their demand schedules. An investigation of this type has not been carried out previously on a range of different contracts. It is, therefore, a unique opportunity to further examine the idiosyncrasies of derivative assets. In particular, the method of analysis allows us to determine how the mix of investors, whether informed or uninformed, varies between, say, financial futures and commodity futures.

Chapters 2 and 3 together represent the preliminary stages of the investigation into the role and impact of the volume of trade. They raise issues and provide results that are exploited by chapters 4 and 5 where the level of examination increases to allow a greater depth of analysis and interpretation.

Chapter 4 returns to the question of the nature of the volume-volatility relationship. As indicated above it is clear that it is only possible to make tentative conclusions as to the exact nature of the link between these two variables. A common feature of the majority of the empirical work in this field is the mistaken belief that it has the ability to distinguish between the different theories of the volume-volatility relationship based on, what is largely, anecdotal evidence. Chapter 4 addresses this problem by exploiting a modern econometric technique to carry out a direct test of the Mixture of Distributions Hypothesis. The unique specification of this model that is examined,

allows an analysis of the UK futures market trading process that has not been possible previously. In particular, the issues that motivate the work of chapter 2 are examined in more depth, providing greater detail on the nature of price return distributions and the role of information in determining the relationship between volume and volatility. The ability to distinguish between the informed and uninformed components of volume, within the specification of the theoretical model exploited in this chapter, is vital in this regard. Another distinctive feature of this chapter is the use of two procedures that allow the construction of futures price and volume samples, taking into account both expiration and roll-over effects. Where the work of chapter 2 has important implications for investors and policy makers, this chapter is able to offer further guidance on issues related to whether futures markets need to be regulated and the possible dangers of restricting an individual's ability to trade. Such subjects need to be carefully considered where the efficient functioning of the market is an important objective of policy makers.

The modern asset market operates in a very competitive environment. The London International Financial Futures and Options Exchange (LIFFE) is the world's biggest non-US derivatives market. More recently, the development of automatic trading systems, the growth of Euro-zone exchanges, and merger activity between former rivals, has begun to put pressure on LIFFE's position of superiority. When the success of a market is judged primarily by the volume that it generates and investors demand liquidity at low cost the relationship between the cost of trading and the volume of trade becomes a fundamental issue.

Chapter 5 carries out a detailed investigation of this important relationship for two financial futures contracts traded on LIFFE where costs are proxied by the bid-ask spread. One of the key achievements is the resolution of some of the issues related to the conflict between inventory cost and information cost models of the spread; the benefits of a liquid market and the costs associated with the increased probability of trading with informed investors at high levels of volume. Another distinctive feature is that this examination of volume and the spread is carried out at an intra-day level that allows us to analyse the patterns in the trading process. Investors may be particularly interested to discover when the market is busiest and how the spread

varies between the open and close of trade. These factors are likely to be an important factor in constructing a strategic investment policy that considers which market conditions are most suitable for the individual. The previous unavailability of high frequency data for UK futures markets has meant that few studies of this nature have been undertaken to date.

Studies relating to various aspects of the spread have again tended to concentrate on non-UK equity markets. The idiosyncrasies of futures markets, in addition to those already mentioned, particularly the fact that bid and ask prices are non-binding, appear to have deterred empiricists with the result that the issues relating to derivatives markets have been under-investigated. This important weakness in the literature, that is also one of the motivating factors for the other empirical chapters of this thesis, is further addressed in chapter 5.

A common criticism of derivatives markets, and futures markets in particular, has been their apparent inability to function during periods of intense pressure. Among the measures adopted to try to maintain the stability of futures markets, notably after the 1987 crash, has been the use of trading halts. However, it is not altogether clear that such mechanisms are likely to be successful. They may simply delay the inevitable in situations where investors continue to hold information that has not been revealed to the market. A similar situation can occur where market-makers feel that market conditions are such that they are at a distinct disadvantage in any trade and therefore set spreads that are prohibitively wide. The ability of the market to cope during periods of high activity or unexpected levels of trading is an important indicator of its capacity to adapt to potential crises. Chapter 5 investigates this key issue by considering the impact of the expected and unexpected components of volume on the spread. It is unique in exploiting an econometric technique that avoids many of the weaknesses inherent in the standard approaches of generating such variables. Chapter 5 also provides guidance to policy-makers with regard to how volume reacts to changes in trading fees. This is possible within a methodological framework that is rarely exploited by empiricists who tend to disregard the possibility of a simultaneous relationship between volume and the spread. There needs to be some care that investors are not deterred from investing on LIFFE by high costs when

alternative investment opportunities are becoming so readily available. It is, therefore, crucial that there is some awareness of the sensitivity of investors to changes in costs.

These four empirical chapters together allow an appreciation of the role and impact of volume that has, until now, not been possible for UK futures markets. Significantly, they address fundamental issues that should be of interest to those other than simply the academic. A summary of the achievements and suggestions for further research are provided in Chapter 6.

CHAPTER TWO: VOLUME-VOLATILITY RELATIONS FOR UK FUTURES MARKETS

2.1 INTRODUCTION

The volatile nature of prices in financial markets is a much investigated but still misunderstood phenomenon. Supporters of the 'casino' view[1] argue that the excessive movement of prices provides opportunities for profit for a few at the expense of others. The alternative so-called 'information' view is that price volatility is simply a reflection of changes in fundamental economic factors, or information and expectations about them.

The aim of this study is to obtain a better understanding of price volatility by considering its links with economic activity, in particular, the volume of trade. A detailed understanding of this relationship is important in aiding our comprehension of the structure of financial markets. As will become apparent, the key element in the theoretical models that are used to analyse the volume-volatility relation is the rate of information flow to the market. However, the empirical work in this field is generally unconvincing, and tends to neglect the issues particular to futures markets. The idiosyncratic nature of futures contracts (standardised contracts, organised trading on centralised exchanges, the use of clearing houses to monitor contract performance, and the relatively low costs of trading involved), suggests that they are quite distinct from other securities. These peculiarities may manifest themselves in different trading patterns or by attracting a group of investors who differ from those who trade in the underlying asset. As has already been noted in chapter 1, there is evidence that indicates[2] that the speed with which information is impounded into futures prices exceeds that in equity markets.

---

[1] See Miller (1991: 130)
[2] See Antoniou and Holmes (1995).

There is also implicit evidence to suggest that futures trading is responsible for the improved quality and speed of information flow to spot markets. Therefore, by increasing our understanding of the volume-volatility relation we will gain an insight into the structure of futures markets from the point of view of the role and impact of volume, and we will also increase our appreciation of an asset with unique properties. This should be of interest to practitioners wary of derivative assets and to policy makers responsible for market regulation. If information does play an important role in determining the volatility of prices, care must be taken to ensure that trading restrictions do not prohibit the efficient operation of the market.

The debate over the distribution of prices is also very much dependent on the nature of the volume-volatility relation. The fat-tailed distributions that are commonly observed in price return data are believed to be caused by an underlying process whose realisation is distinct from those measured in conventional time. This process is hypothesised to be information flow. Although price series are conventionally measured in real time the frequency of these information arrivals is measured in so-called event time. If it is the actions of investors that reveal this information to the market it is not unreasonable to assume that the volume of trade has an important role to play in describing the distribution of prices.

It is these key issues that motivate the work in this chapter. By considering the relationship between volume and volatility for a variety of UK futures contracts, each of which have their own characteristics in terms of the type of trader they attract, the number of expiration dates per year, the impact of seasonal factors etc., it is hoped that a new insight into this field will be gained. This will also allow a re-evaluation of the current literature. Indeed, the empirical work carried out here suggests that further investigation of the volume-volatility relation is necessary before a clear picture of the importance of the volume of trade can be obtained. This is addressed in chapter 4.

The rest of this chapter is structured in the following way. Section 2.2 explains in more detail the importance of the volume of trade; the determinants of volume and the theoretical basis of the relationship between volume and volatility. Section 2.3 provides an overview of the alternative approaches that have been used to investigate

the possible links between volume and volatility as part of a comprehensive literature review. The purpose of this chapter is to provide an exploratory analysis, raising issues that will be investigated more extensively in later chapters. Therefore, this study exploits established techniques to investigate the volume-volatility relation. This will allow us to examine the suitability of such methods in providing a test of the various theories of the link between volume and volatility. A description of these methods is provided in section 2.4.

The use of futures markets data presents a number of problems with regard to collecting a sample of observations. Section 2.5 describes how the sample was constructed and also presents the results of this study that tries to answer the question, "does a relationship exist between volume and volatility for UK futures markets and if so why?" Section 2.6 concludes and suggests ideas for future research.

2.2 THEORETICAL BACKGROUND

This section looks at some of the theoretical models that have been used to explain the volume of trade within financial markets and presents those that are our primary concern; models that address the relationship between volume and price changes.

2.2.1 THE DETERMINANTS OF THE VOLUME OF TRADE

Admati and Pfleiderer (1992) argue that one of the shortcomings of traditional asset pricing theories is that they are unable to model the dramatic changes in trading volume that occur over relatively short time periods. An important element in developing a theoretical model is an appreciation of the concept of investor heterogeneity. The decision to trade is based on individual investor preferences and the belief that by entering the market it is possible to obtain asset pay-offs. The level of actual volume will depend on the degree of heterogeneity of investors and the efficiency of the trading process.

The assumption that trade only occurs where there are gains to the investor is the basis of the model developed by Karpoff (1986). He defines the gains to trade as the situation where one investors' reservation price for selling a share is lower than another investors' price for buying a share. He argues that the volume of trade is

12

positively related to the same random variables that generate reservation prices, and that volume levels will be higher on average in a Walrasian auction where all gains to trade are realised, than they are in an inefficient system of random pairings of buyer and seller. If markets do not exhaust all gains in the first round of trading there may be some volume persistence.

It should be noted, however, that futures markets do not operate as a Walrasian auction. As Sutcliffe (1993) points out, in the Walrasian ideal trading occurs much like a conventional auction where the desires of buyers and sellers are matched by an auctioneer. The auctioneer sets a price at which the parties involved will declare the amounts that they are willing to buy or sell. Futures markets, on the other hand, operate without this trading facilitator, and buyers and sellers present bids and offers simultaneously. A trade occurs where these two prices match, usually after a period of adjustment. In this sense, the Walrasian auction represents a periodic trading process while trading in futures markets, under this so-called double auction, is continuous.

This process of relating volume to differences in investors' preferences and endowments is fairly straight forward. The complexity of this issue is increased, however, when one considers the scenario where heterogeneity among investors is due to investors observing different pieces of private information. 'No trade' theorems show that there are situations where differences in information will not generate any trade. Trading based on private information will only occur where there are other motives to trade. It is often assumed that to generate trade there must be some investors who have liquidity or hedging motives for entering the market. The demands of these investors are unrelated to the future payoffs and their trading is only responsible for some of the activity in the market. The rest is generated by speculators who might be privately informed about asset payoffs.

Another development in the models designed to explain the determinants of volume is identified by Admati and Pfleiderer (1992) to be the use of rational expectations equilibria (REE). They assume that investors have the ability to infer from asset prices the information held by other investors in the market. The combination of this information with that they may already hold, defines an investors' demand schedule.

13

The REE model developed by Grossman and Stiglitz (1980) assumes that investors can pay a cost to observe a common piece of private information. They show that the volume of trade between the informed and uninformed traders is a decreasing function of the precision of the informed traders' information. When the informed traders' private information is very precise, prices reveal most of the information to the uninformed traders so that in equilibrium all traders have similar beliefs. This lack of heterogeneity inhibits trade.

Pfleiderer (1984) extends work on information aggregation in markets and considers a noisy REE model. This contrasts with the Grossman and Stiglitz (1980) model because their model assumes informed traders observe distinct private signals. Each trader observes a signal equal to the final pay-off of an asset plus a random error term, where the error terms of different traders are statistically independent. Pfleiderer argues that if the precision of private signals is increased, the dispersion of traders' forecasts, based on private signals and the equilibrium price, decrease. This effect tends to decrease volume. However, risk averse investors will trade more aggressively on their information if they believe it to be more precise. This tends to increase volume. This second effect dominates the first assuming that the error terms are independent. If there is a common error, volume is a poor predictor of investor forecast diversity.

In many situations the level of trading volume partially determines the costs and benefits of trading. In some cases a 'feedback loop' arises where the level of volume affects the gains to trade and this in turn affects the level of volume. A trader's decision to enter a market may depend on how many other traders enter, since the number of traders is a determinant of the volatility of prices. This view is consistent with the notion that trade generates trade. It also arises in situations where traders are asymmetrically informed. The self-generating trade scenario is illustrated by Kyle (1985) who develops a model where a single informed trader and liquidity traders submit orders to a risk-neutral supplier of immediacy, a so-called market-maker. He argues that informed traders will try to exploit those who are less informed by trading more heavily on the information that they hold. The greater the number of liquidity

traders in a market, the bigger the positions taken by the informed trader. Therefore, the increase in the volume of liquidity trading leads to an increase in the volume of informed trading.

An alternative explanation for the demand function of investors is provided by Lakonishok and Smidt (1989). They identify a number of tax and non-tax motives for trading. Tax related motives are associated with a desire by investors to limit their losses on capital gains during the year. Non-tax related motives include window dressing, portfolio rebalancing, and contrarian strategies. Lakonishok and Smidt show that the dynamic relation is negative for tax-related trading motives and positive for non-tax-related motives.

Noise trader models reconcile the difference between the short and long-run autocorrelation properties of aggregate stock returns. Aggregate stock returns are positively autocorrelated in the short-run, but negatively autocorrelated in the long-run. This phenomenon is discussed in detail by De Bondt and Thaler (1985) and Fama and French (1988). Since noise traders do not trade on the basis of economic fundamentals, they impart a transitory mispricing component to stock prices in the short-run. The temporary component disappears in the long-run, producing a mean reversion in stock returns. A positive causal relation from volume to stock returns is consistent with the assumption made in these models that the trading strategies pursued by noise traders cause stock prices to move. A positive relation from stock returns to volume is consistent with the positive feedback strategies of noise traders, for whom the decision to trade is conditioned on past stock price movements.

2.2.2 THE RELATIONSHIP BETWEEN VOLUME AND VOLATILITY

Although the models above can tell us why investors choose to trade they are less illuminating about what is our primary concern; the relationship between volume and price changes. The models is this area are based on the premise that much of the information coming into markets comes from private information revealed through the trading process itself. The two dominant models in this field are the Sequential Information Model and the Mixture of Distributions Hypothesis. This section provides a brief overview of these models which will be discussed in more detail when we return to the volume-volatility issue in chapter 4.

The Sequential Information Model (SIM) of Copeland (1976) is based on the idea that each investor in turn receives a piece of information and then acts on that information before it becomes public knowledge. In this model uninformed investors are assumed to be unable to extract information from prices or from the actions of others. This trading process results in a series of incomplete equilibria. It is only when all traders have received the information signal that a final equilibrium is established. Since different investors will interpret the information differently the path of prices and volume will depend on the sequence in which individuals have become informed. Hence, the positive relation between the volume of trade and price volatility. In his simulation Copeland finds that a positive correlation does exist between volume and the absolute price change, and also that volume is highest when investors are either all optimists or all pessimists.

The Mixture of Distributions Hypothesis (MDH) is based on the difference between price changes that occur over periods of calendar time, and information arrivals that occur over an equal number of periods of so-called event time. We conventionally discriminate between different observations in a series according to calendar time. Supporters of the MDH argue that in fact we need to consider the underlying process that produces the different observations. Clark's (1973) seminal paper attempts to model the joint distribution of daily stock price changes and volume. Daily price changes of speculative assets appear to be uncorrelated with each other and symmetrically distributed, but the distribution is kurtotic relative to the normal distribution. He believes that this is caused by variations in the flow of an underlying

16

process; the rate of information arrival. The assumption underlying his empirical work is that this process can be proxied by trading volume.

Harris (1987) extends the MDH to allow further investigation of the joint distribution of price change and volume. In his model the daily price change is seen as the sum of a variable number m of independent within-day price changes. It is intuitively attractive to interpret m as the number of within-day information arrivals. Therefore, the conditional variance of the price change is considered to be an increasing function of the rate at which new information enters the market. This correlation between volume and price changes results because volume is also an increasing function of the number of within-day information arrivals.

The discussion above illustrates the variety of models that have been developed to explain both the determinants of volume and its relationship with price volatility. They will provide the basis for the investigation carried out in this chapter. The next section considers the different approaches that have been used to analyse these theoretical links between the movement of prices and the volume of trade.

2.3 LITERATURE REVIEW

The literature in this field is large and diverse. This section takes a broad look at the studies that have investigated the relationship between volume and price variability. The discussion takes a particular interest in the empirical work that has utilised similar methodologies to those that will be exploited in this chapter. The two main econometric techniques used here, causality tests and autoregressive conditional heteroscedasticity (ARCH) modelling have been widely used in contemporary economic analysis. A comprehensive review of the early empirical work in this field is provided by Karpoff (1987).

Grammatikos and Saunders (1986) investigate whether the volume-volatility relationship is simultaneous or sequential for five different foreign currency futures. The daily data covers the period from March 1978 to March 1983. Following Harris' (1986) specification of the MDH, due to the random variations in the directing variable, price variances may be changing through time. Grammatikos and Saunders

argue that such a possibility may exist for futures contracts, in which information arrival may be maturity dependent. With time, the information may have a greater impact due to the resolution of uncertainty. The implication is that the time-to-maturity may be the directing variable. Rather than create a single 'price' for a single 'composite' futures contract they use disaggregated data and also look at two measures of price variability - the 'classic' and the Garman-Klass (1980) measure[3]. Their results suggest that the new Garman-Klass method of calculating variance is superior to traditional methods. They also exclude any observations that fall in the expiration month to avoid so-called 'delivery complications'. The explanatory power of the time-to-maturity is assessed using the Pearsonian correlation coefficient. Their results suggest that the time-to-maturity is not the directing variable. This is because the relationship between the time-to-maturity and volume is different to the relationship between the time-to-maturity and price variability.

Grammatikos and Saunders also exploit the Geweke et al. (1983) causality test methodology; first to test whether the volume of trading causes price variability, and secondly to test whether price variability causes volume of trading. In each case the regressions are run with three lead and lag coefficients. In the majority of cases futures contract price variability and trading volume are contemporaneously correlated. There is, however, a significant number of cases in which a sequential relation between price variability and volume appears to be present. They do not attempt to speculate as to why these links occur other than to report that they exist.

In a related study Jain and Joh (1988) look at common stock trading volume and returns on the New York Stock Exchange (NYSE). They examine both intra- and inter-day differences in the patterns of trading using hourly trading data over the period from 1979 to 1983. They also examine whether a relationship exists between volume and returns using the causality methodology, and whether the relation is

---

[3] The classical variance estimator based on closing prices (C) is given by:
$$\sigma^2 = (C_t - C_{t-1})^2$$
The Garman-Klass estimator is derived from daily high(H), low(L), closing (C), and opening (O) prices:
$$\sigma^2 = 1/2[\ln(H) - \text{Ln}(L)]^2 - [2\ln(2) - 1][\ln(O) - \ln(C)]^2$$

different for positive and non-positive price changes. Having decided that there are differences during the day and between days they create dummy variables to account for both of these effects and to distinguish between positive and negative returns. In order to control for the predictive ability of its own past values, both returns and volume are transformed by an autoregressive integrated moving average (ARIMA) filtering process. The residuals are viewed as that part of the series that cannot be predicted from its own past history. Although Jain and Joh describe the statistical motivation for this process, they do little to provide an economic justification for their use of this filtering procedure.

The results of their Granger-Sims causality tests suggest that there is a strong positive contemporaneous correlation between volume and the absolute value of returns. This, they argue, is consistent with the MDH. They also find that lagged values of the return variable have a significant impact on the volume variable that they believe is consistent with the SIM. In contrast there is only weak evidence of causality from volume to returns. Jain and Joh argue that in an informationally efficient market volume should not be useful in predicting returns. They also that find that the relationship is significantly different when returns are positive from when they are non-positive. This result is perhaps unsurprising in a stock market scenario where restrictions exist on the short selling of assets. These differences would not be expected to materialise in futures markets where such costs are not imposed.

Hiemstra and Jones (1994) test for linear and non-linear relationships between stock returns and percentage changes in trading volume, based on daily data for the Dow Jones Industrial Average for the period 1915 to 1940, and the Dow Jones 65 Composite Index for the period 1941 to 1990. Their examination of the linear relationship between returns and volume is very similar to the approaches described above and exploits the Granger test procedure. However, they also examine the possibility of non-linearities based on the residuals from the linear causality model. They argue that large price swings and abrupt changes in stock market volatility can only be properly modelled with non-linear models. They also claim that there is evidence of non-linearities in the volume series. Therefore, the causal relation between the two may also be non-linear. Hiemstra and Jones argue that their use of

the non-linear test is justified by the results. They find evidence of linear causality from returns to volume but not from volume to returns. The second test, however, reveals that bi-directional non-linear causality exists between the two variables. They, therefore, conclude that the linear test is inappropriate because it is unable to detect the true underlying volume-return relationship.

The next stage of their study examines whether the non-linear predictive power of trading volume for stock returns can be attributed to volume serving as a proxy for the daily flow of information into the market. Anderson (1996) notes that the common-factor model provides an explanation for the volatility persistence associated with ARCH in daily stock returns when Clark's (1973) independent and identically distributed (iid) assumption for information is relaxed. Therefore, evidence of non-linear Granger causality could be due to volatility effects associated with information flow. Hiemstra and Jones filter the stock return series using the exponential ARCH (EGARCH) methodology. However, their results suggest that although the bi-directional non-linear causality between the two variables is now less strong, it is still significant. They conclude that the causal link between returns and volume is not wholly explained by information flows.

An almost identical approach is adopted by Fujihara and Mougoué (1997) to investigate the causal relationship between volume and volatility for crude oil, heating oil and unleaded gasoline futures traded on the New York Mercantile Exchange (NYMEX) between 1984 and 1993. They also question the suitability of the linear causality methodology in a study of this type. They argue that the evidence of strong bi-directional non-linear causality is consistent with both the SIM and the MDH, and the noise trading models of DeLong et al. (1990).

Schwert (1989) analyses the relationship between stock volatility and real and nominal macroeconomic volatility, economic activity, financial leverage, and stock trading activity. This is based on monthly and daily data from a variety of different US sources over the period from 1885 to 1987. In what is a very comprehensive study, Schwert considers the possible links between returns and volume for NYSE stocks.

He runs the following regression which is estimated by generalised least squares (GLS):

$$\hat{\sigma}_{st} = \alpha_0 + \frac{\beta}{(1-\delta L)} \text{Vol}_t + u_t \qquad (2.1)$$

This model relates stock volatility ($\hat{\sigma}_{st}$) to a distributed lag of past share volume (Vol_t) growth, where the coefficient of volume growth decreases geometrically. The results suggest a positive contemporaneous relation between stock volatility and volume using monthly data. Using daily data the relationship still holds, but the lagged values of volume become more significant. Further vector autoregressive (VAR) modelling of volatility regressed on volume lagged up to twelve periods provides additional support for a contemporaneous, rather than sequential, relationship between the two variables. Schwert concludes by admitting that it is not possible to say whether this relation is due to 'trading noise' or to the flow of information to the stock market.

Lang et al. (1992) devise tests that distinguish between competitive (Walrasian), fully revealing rational expectations and noisy rational expectations equilibria based on their predictions concerning trading volume around public information signals. They use simple regression analysis and regress volume on a number of variables each designed to distinguish between the different models. They argue that if price fully reveals market information, traders will be indifferent to holding different quantities of the asset at this price. Price changes would then be uncorrelated with asset holdings and volume. Using data surrounding quarterly earnings announcements for 101 firms in the period from April 1984 to March 1986, Lang et al. find a link between price changes and volume and conclude that this is evidence of a noisy market.

An alternative approach to modelling the difference between the trading response to positive and negative returns is the use of state-space techniques by McCarthy and Najand (1993). They look at the volume-price change *per se* and the volume-absolute price change relationships for foreign currency futures traded on the Chicago Mercantile Exchange (CME) over the period from 1979 to 1990. Like Grammatikos and Saunders (1986) they use daily data excluding the expiration month to avoid 'delivery complications'. McCarthy and Najand argue that the use of state-space

models has a number of advantages; it allows simultaneous determination of the causal link and the relationship between variables, it avoids the subjective nature of VAR modelling, allowing the mathematical determination of the number of necessary dimensions, it allows greater insight into the lead/lag relationship, and it allows us to see whether or not causality is unidirectional.

Their investigation into the volume-price change *per se* relationship originally extends from Epps (1975). The hypothesis is that bulls consider assets to be riskier than do bears and, from this, the bulls' demand function is steeper than that of the bears. The implication is that the ratio of volume to a positive price change would be greater than that of volume to a negative price change. The drawback is that this hypothesis also implies investor irrationality where pertinent information is systematically ignored. Jennings et al. (1981) extend Copeland's (1976) model to include margin requirements and short selling. Since a short sale is more costly than a long position, those investors undertaking short positions face a demand curve which is less responsive to price changes. Thus, the volume generated by optimistic traders exceeds that of pessimistic traders. This is the phenomenon observed by Jain and Joh (1988). Therefore, volume rises with price increases, while price decreases are associated with falls in volume. However, as already mentioned, in futures markets the costs of taking long or short positions is symmetric. Karpoff (1987) argues that this is the reason why there is little evidence of a significant correlation between returns and volume in the futures market literature.

McCarthy and Najand find no relationship between volume and price change *per se*, and also no contemporaneous relationship between absolute price changes and volume. There is, however, evidence that volume lagged up to two periods is causally related to absolute changes in prices. This, they argue is consistent with the SIM but not the MDH. The state-space modelling of the relationship between volume and absolute changes in price also reveals a relationship, where returns are lagged up to two periods. They suggest that the negative signs on the volume variables are indicative of volume's stabilising influence on volatility.

These results are in contrast to those of Smirlock and Starks (1985). They test the asymmetry hypothesis for the volume-volatility relationship using the Wilcoxon non-parametric test. This allows comparisons of the ratio of volume to absolute price change on down-ticks to that of the ratio on up-ticks. Using transaction data for all NYSE stocks for the period June 15 to August 21 1981, Smirlock and Starks compare days when earnings announcements are made to those when there is no known information dissemination. Their results indicate strong support for the hypothesis that volume is higher on up-ticks than on down-ticks on the day when there is information arrival.

Malliaris and Urrutia (1998) postulate several hypotheses with regard to the volume-volatility relationship and test them with data for agricultural commodity futures contracts. By postulating that volume is a function of price and time, Malliaris and Urrutia investigate the hypotheses that prices and volume both follow a random walk, that futures prices and the corresponding volumes of trading are interrelated and can affect each other, and that the volatility of trading volume is a function of the futures price volatility. They use a combination of randomness and stationarity tests, Granger causality, cointegration techniques and regression analysis. Malliaris and Urrutia investigate corn, wheat, oats, soyabean, and soyabean meal futures contracts. They find that price and trading volume are non-stationary in levels, but stationary in the first differences. No causality between price and volume appears to exist, but price and volume are cointegrated with volume following and adjusting to price movements. Their final conclusion is that price and price volatility are determinants of trading volume and price volatility influences the volatility of the volume of trade.

Lamoureux and Lastrapes (1990), like Hiemstra and Jones (1994), work on the premise that the presence of generalised ARCH (GARCH) effects is based on the hypothesis that daily returns are generated by a mixture of distributions in which the rate of information arrival is the stochastic mixing variable. They use daily stock returns for twenty actively traded US stocks and show that volume can be used as a proxy for information. They use the following simple GARCH model for twenty actively traded stocks:

$$r_t = \mu_{t-1} + \varepsilon_t \qquad (2.2)$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1} + \alpha_2 h_{t-1} + (\alpha_3 V_t) \qquad (2.3)$$

where r_t represents the rate of return, μ_{t-1} is the mean return conditional on past information, and V_t represents the volume of trading. After the introduction of contemporaneous volume into the conditional variance equation the ARCH effects disappear for the majority of stocks considered. They argue that volume is, therefore, a good information proxy. This study can be criticised for its use of contemporaneous volume in the conditional variance equation which introduces simultaneity bias. This point, originally noted by Karpoff (1987), is not lost on Lamoureux and Lastrapes. They try to exogenise the volume variable by using lagged and fitted values. They argue that the poor explanatory power of these variables is due to their inability to act as an instrument for contemporaneous volume.

Najand and Yung (1991) use the Lamoureux and Lastrapes model to investigate the volume-volatility relation for Treasury-bond futures traded on the Chicago Board of Trade (CBOT) over the period from 1984 to 1989. However, unlike the study by Lamoureux and Lastrapes, the GARCH effects remain even when volume is included. This suggests that the time series of Treasury-bond futures prices exhibit significant levels of second order dependence, and they cannot be modelled as white noise processes. With respect to the volume-volatility relationship they find a correlation exists for only two of the six years in the study, and also does not exist for the sample as a whole. However, the problem of simultaneity bias is still evident and these results should be treated with caution. They try to exogenise volume by using lagged values of the variable. Its significance in all but one case suggests that lagged volume is a good proxy for contemporaneous volume. They argue that the two sets of results together indicate a positive relation between volume and price variability, even though volume is not able to account for all of the GARCH effects in the data.

Locke and Sayers (1993) also question the results of Lamoureux and Lastrapes. They use an equation of the form:

$$r_t^2 = \alpha + \gamma I_t + \sum_{i=1}^{n} \beta_i r_{t-i}^2 + \varepsilon_t \qquad (2.4)$$

where I_t represents information arrival and r_t^2 represents the variance of the returns from the S&P 500 futures index. They investigate whether variance persistence is

removed after controlling for a number of information proxies, including contract volume, floor transactions, the number of price changes and order imbalance. Their results suggest that all information proxies are capable of explaining a significant amount of return variance, but there remains evidence of variance persistence. Their conclusion is that trading *per se* will not explain persistence in returns volatility.

Foster (1996) carries out a comprehensive study of crude oil futures that combines a number of the elements of the studies by Jain and Joh (1988), McCarthy and Najand (1993), and Najand and Yung (1991). The first area of interest is the general relationship between trading volume and price variability. He tests whether the predicted contemporaneous volume-volatility relation holds over an alternative where lagged volume explains current price variability. He argues that if the latter holds this violates the notion of informational efficiency. The second area of interest considers whether the level of trading volume associated with a price rise is different to that associated with a price fall. Finally, Foster considers whether the size of the futures market affects the volume-volatility relationship. He considers large markets versus small markets and initial market phases versus mature market phases.

The volume-price variability relation is tested using the GARCH methodology, justified along the same lines as Lamoureux and Lastrapes (1990). Foster believes that the two problems that arise in using this methodology are that of simultaneity bias, and the fact that the introduction of volume into the GARCH equation is more a test of whether volume represents a proxy for information than a test of the volume-volatility relation. Foster, therefore, uses the generalised method of moments (GMM) as an additional test. The model takes the following form:

$$V_t = \beta_0 + \beta_1 h_t + \beta_2 V_{t-1} + \beta_3 V_{t-2} + \varepsilon_t \qquad (2.5)$$

$$h_t = \phi_0 + \phi_1 V_t + \phi_2 V_{t-1} + \phi_3 h_{t-1} + \upsilon_t \qquad (2.6)$$

where h_t represents the volatility variable, and V_t is the volume variable.

The primary finding using data on Brent crude and WTI crude from the International Petroleum Exchange (IPE) and NYMEX respectively over the period from 1990 to 1994, is that volume is not an adequate proxy for the rate of information flow, but that

volume and volatility are largely contemporaneously related and are both driven by the same factors, assumed here to be information. He finds from both the GARCH and GMM methodologies that lagged volume can also explain current variability, but argues that this suggests a degree of inefficiency rather than a rejection of the MDH. Foster also finds, with regard to the market size effect, that the magnitude of trading volume does not have implications for volatility or the volume-volatility relationship other than that which would be expected with increased liquidity. This implies a rejection of the SIM which supports increasing volatility with market growth.

Gallant et al. (1992) do not test a specific economic model. Their investigation of NYSE data covering the period from 1928 to 1987 is simply an analytical study of a long run of data. Like many of the other papers considered here, they argue that more can be learned about the market, and in particular volatility, by studying prices in conjunction with volume, instead of prices alone. Among their objectives is a desire to analyse the relationship between volume and volatility in an estimation context, and to investigate the intertemporal relationships among prices, volatility, and volume. They begin by using simple graphical methods before estimating a semi-non-parametric model of the conditional joint density of market prices and volume, as proposed by Gallant and Tauchen (1989).

They begin by eliminating systematic effects, including turn of the year and weekday effects, from raw Standard and Poor (S&P) price change data and NYSE aggregate volume data. This entails fitting a series expansion to the bivariate conditional density. The leading term in the expansion is a VAR model with an ARCH-like error process. Higher order terms accommodate departures from the model, for example, the complicated nature of the bivariate conditional variance function, the thick tailed error density characteristic of financial price change data, the non-linear interactions between volume and prices, and the temporal dependence of the volume series. Their results suggest that trading volume is positively and non-linearly related to the magnitude of the daily price change. This appears to hold for both the unconditional distribution of price changes and volume and the conditional distribution given past price changes and volume. They also find that price changes lead to volume movements in a symmetric relation.

Jones et al. (1994) exploit the Schwert (1990) methodology to investigate the relation between volatility and economic activity. They use the absolute residuals from the following model:

$$R_{it} = \sum_{k=1}^{n} \hat{\alpha}_{ik} D_{kt} + \sum_{j=1}^{m} \hat{\beta}_j R_{it-j} + \hat{\varepsilon}_{it} \tag{2.7}$$

where R_{it} is the return of security i on day t and D_{kt} are day of the week dummies used to capture differences in mean returns. These are then regressed on the market trading and volume variables. They use daily volume and number of transactions data together with returns on 853 US securities calculated from the average of closing bid-ask quotes. The sampled securities are sorted into five portfolios according to market value in an attempt to control for size-related systematic components of the volume-volatility relationship. Their results suggest that it is the frequency of trade and not its volume that generates volatility. Therefore, volume has no information content beyond that contained in the number of transactions.

The discussion above illustrates the variety of approaches that have been used to investigate the relationship between price volatility and the volume of trade in financial markets. Although this review is by no means exhaustive it does allow us to draw out some interesting points. The first of these is that very few studies actually appear to have a strong sense of economic purpose. The discovery of links between volume and volatility is certainly interesting, but this not always apparent from reading the different studies. They make tentative conclusions that the results confirm certain models, but the evidence appears to be largely anecdotal.

The issue of causality certainly needs further investigation. The literature mentioned above seems to be confused as to whether linkage between volume and volatility is evidence of economic causation or simply correlation.

The studies that exploit the more sophisticated techniques of ARCH and GARCH also suffer from a lack of clarity. What is the underlying hypothesis that is being tested? Is it actually the MDH or the SIM, or are we really investigating whether volume is a

proxy for information? Further examination of the role of volume is, therefore, required.

This study aims to address these issues and also to develop our understanding of UK derivative markets which have been largely neglected in the empirical work to date. The few studies that have considered futures markets suggest that there are issues particular to these securities concerning the role of volume that differ from those of equity markets. Therefore, further investigation is necessary. In addition, there has also been a tendency, with the odd exception, to concentrate on financial contracts. This study looks at a mixture of financial and commodity futures to determine whether differences exist in terms of the nature of the volume-volatility relation. The empirical work in this chapter also aims to add to the existing literature by answering the following questions:

- does a relationship exist between volume and volatility?
- is the level of daily trading volume an important factor in this relationship?
- does the total amount of volume in the market have a bigger impact than the volume of trade relating to a particular contract?
- does the trading during the expiration month hide the 'true' relationship?
- does volume act as a proxy for information in futures markets?

2.4 METHODOLOGY

As stated earlier, the aim of this chapter is to provide a better understanding of the role of volume in futures markets, in particular its association with price volatility. The previous section highlighted a number of weaknesses in the approaches of other studies in this field. The main criticism that can be made is that it is not always clear what the economic hypothesis is intended to be. There appears to be little consensus regarding the aim of causality tests which prompts conclusions that are not wholly supported by the results. The use of GARCH analysis also appears to be based on ambiguous objectives.

This section of the chapter considers both of these techniques in some detail, in particular, outlining exactly what they are capable of telling us about the relationship between volume and volatility. The aim is to exploit these methods within certain parameters and therefore avoid making conclusions that cannot be justified.

2.4.1 THE CAUSALITY APPROACH

As Harvey (1981a) points out, the issue of cause and effect is central to any scientific enquiry. Economics, despite its best efforts, is still considered to be outside the group of so-called real sciences that includes mathematics and physics. The main reason for this is that controlled experiments are not possible. This makes it very difficult to say with any certainty that a cause and effect relationship actually exists. The use of econometric models has traditionally adopted an approach where economic theory drives the specification of the model. The direction of causation is, therefore, assumed rather than tested. The concept of 'causality' has arisen out of a need to test these assumptions.

Our notions of causality are commonly based on the work of Granger (1969). Within this framework there are two basic rules; the future cannot predict the past, and the variables under consideration must be stochastic. A variable x is then said to 'cause' a variable y if taking account of past values of x enables better predictions to be made for y, all other things being equal.

However, as pointed out by Harvey (1981a), this notion of causality is a purely statistical one, and it does not correspond to any acceptable definition of cause and effect in the philosophical sense[4]. A more appropriate term would probably be 'predictability'. This distinction is crucial and appears to have been largely missed by the majority of empiricists in this field. Evidence of causality defined in this manner allows us to say something about the correlation structure between variables, but it does not determine causality in an economic sense.

---

[4] For a detailed discussion of the issues see Zellner (1984: 35-74)

29

Therefore, in this study, causality tests will simply be used within these limits to determine whether links exist between volume and volatility.

The following, more detailed description of the concept of Granger causality, follows that given by Harvey (1981a). The central idea is that x causes y if taking account of past values of x leads to improved predictions for y. More accurately, Granger's definition involves a reduction in forecasting variance with respect to a given information set.

Let U be an information set including all past and present information, and let $\overline{U}$ denote the same set, but excluding present information. Similarly, let X denote all past and present information on the variable x, i.e. X=(x_t, $\tau \le t$), and let $\overline{X}$ be the past information alone, i.e. $\overline{X}$=(x_t, $\tau < t$). The variable x is then said to cause y if the one step ahead predictor of y, $\overline{y}$, based on all past information has a smaller mean square error than the predictor of y based on all past information excluding x. Thus, x causes y if:

$$MSE\left(\overline{y}|\overline{U}\right) < MSE\left(\overline{y}|\overline{U} - \overline{X}\right)$$

Similarly, *x* causes *y* instantaneously if:

$$MSE\left(\overline{y}|U\right) < MSE\left(\overline{y}|U - X\right)$$

The problem with Granger causality as it stands is that U represents all available information. Granger suggests the concept of all relevant information as an alternative. The decision regarding what is and what is not relevant information is, however, fraught with problems. Economic theory must play a role at this stage but this assumes that the theory is correct *a priori*.

In order to carry out a test of Granger causality it is assumed that the relevant information set, U, consists only of information on the two variables x and y. If there are no expectations as to which of these variables is exogenous and which is endogenous a suitable framework for testing causality is a general unrestricted VAR model. This consists of regressing each current variable in the model on all of the variables in the model lagged a certain number of times. This can be written as:

$$Z_t = \sum_{i=1}^{k} A_i Z_{t-i} + \varepsilon_t \qquad (2.8)$$

where

$$Z_t = \begin{bmatrix} y_t \\ x_t \end{bmatrix} \qquad (2.9)$$

Z_t is a column of vectors on the current values of all the variables in the model. ε_t is a column vector of random errors. Since the right hand side of this model contains only lagged variables, assuming no autocorrelation, it can be estimated equation by equation by ordinary least squares (OLS). It can also be estimated using a multivariate regression technique. However, Charemza and Deadman (1997) point out that since no restrictions are placed on the coefficients in the equation above, multivariate least squares estimators are no more efficient than those of OLS.

The most commonly adopted technique that can be used to establish whether there is causality between two variables is the Granger test. The following description of the test follows that given by Charemza and Deadman (1997). Consider an equation describing y_t in an unrestricted bivariate VAR model, that is, one describing relations between two variables, x and y that are assumed to be stationary. The equation may be written as:

$$y_t = A_0 D_t + \sum_{j=1}^{k} \alpha_j y_{t-j} + \sum_{j=1}^{k} \beta_j x_{t-j} + \varepsilon_t \qquad (2.10)$$

where D_t captures the non-stochastic variables of the equation and A_0 is a vector of parameters. If $\beta_1 = \beta_2 = \ldots = \beta_k = 0$ then x does not cause y in the Granger sense. This can be tested using the log-likelihood ratio statistic.

An important consideration when using causality tests is the determination of the appropriate lag length for the model being used. If autocorrelation is present in the residuals, tests using Lagrange multiplier statistics are no longer reliable. At the same time it is necessary to ensure that irrelevant variables are excluded. The appropriate order can be determined by calculating log-likelihood ratio statistics. This allows the testing of the hypothesis that the order of the VAR is k against the alternative that it is K, for $k = 0,1,2,\ldots,K-1$.

2.4.1.1 Causality Tests and the Volume-Volatility Relation

The theory has now been discussed in detail. How can this be exploited to tell us something about the relationship between volume and price volatility? Causality testing of the volume-price variability relation has not produced a wealth of empirical work. The two prominent studies that have used this technique, those by Grammatikos and Saunders (1986) and Jain and Joh (1988) were discussed in section 2.3. While the former study merely reports the results of the causality tests, Jain and Joh (1988) argue that evidence of causality between price volatility and volume is supportive of the theoretical models of this relationship. The advantage of using the VAR methodology rather than more sophisticated complex simultaneous models is its simplicity. However, as pointed out earlier, this restricts the scope of any conclusions that can be made when using this method. This study will, therefore, only use causality tests to identify whether a relationship exists between the variables in question and to determine the direction of causality. This will be an important first step in our examination of the role of volume in derivative markets. The task then is to explain why this relationship might exist? This requires the exploitation of another statistical technique.

2.4.2 ARCH MODELLING

The concept of ARCH modelling has its origins in the work of Bachelier (1900) and Mandelbrot (1963). It was Mandelbrot, in particular, who noted that the distributions of many economic and financial variables are characterised by fat tails and the clustering of observations. Engle's (1982) ARCH model was the first to capture these effects within a formal framework.

This description of Engel's model closely follows that given by Bera and Higgins (1993). The central theme is that these so-called ARCH effects can be accounted for by an autoregressive error process. Let the dependent variable Y_t be generated by:

$$Y_t = X'_t \xi + \varepsilon_t \qquad t = 1,...,T \qquad (2.11)$$

where X_t is a $k \times 1$ vector of exogenous variables, which may include lagged values of the dependent variable, and ξ is a $k \times 1$ vector of regression parameters. ε_t is the

stochastic error term which is assumed to be conditional on the information set Ψ available at time t-1. This information set contains lagged values of both the endogenous and exogenous variables in the model. This is written more formally as:

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

where

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \ldots + \alpha_q \varepsilon_{t-q}^2 \qquad (2.12)$$

This is known as the conditional variance equation since h_t is a function of the information set. To ensure that the variance is positive both α_0 and α_i are restricted to be greater than zero. It is this equation that plays a very important role in describing the characteristics of the data, in particular, the periods of volatility common in financial series. The aim is to capture the clustering of shocks that causes the volatility. It can be seen, looking at the equations above, that any shock will result in a diversion of Y_t away from its conditional mean. Depending on the form of the shock ε_t will have a large positive or negative value. The conditional variance will increase with any shocks to the system since the lagged error terms appear as squared values. Therefore, large (small) errors of either sign tend to be followed by a large (small) error of either sign. The number of lags gives some measure of the persistence of the shock. A large value of q would be indicative of a long period of volatility.

Bollerslev (1986) proposed an extension to this approach, which he termed GARCH. He suggested that the conditional variance should be specified as:

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \ldots + \alpha_q \varepsilon_{t-q}^2 + \beta_1 h_{t-1} + \ldots + \beta_p h_{t-p} \qquad (2.13)$$

where the inequality restrictions

$$\alpha_0 > 0$$

$$\alpha_i \geq 0 \text{ for } i = 1, \ldots, q$$

$$\beta_j \geq 0 \text{ for } j = 1, \ldots, p$$

are imposed to ensure that the conditional variance is strictly positive. A GARCH process with orders p and q is denoted as GARCH(p, q). This can be distinguished from the Engle (1982) specification by the fact that the conditional variance is dependent on its own lagged values as well as those of the squared error term. Therefore, h_t effectively depends on all past values of ε_t^2 and can be used to represent

a high order ARCH process. This model is popularly estimated using the maximum likelihood procedure.

2.4.2.1 ARCH Modelling and the Volume-Volatility Relation

As with the causality tests we need to see how the GARCH methodology can be exploited to investigate the relationship between the volume of trade and price variability. The main motivation for its use is provided by the work of Lamoureux and Lastrapes (1990). They argue that the fat-tailed distributions prevalent in financial data are caused by the arrival of information in the market. This argument uses as its basis the MDH; the theory that the observed variation is caused by variation in an underlying process.

Lamoureux and Lastrapes (1990) begin by assuming that the unexpected price change in a day, ε_t, is the summation of a number of intra-day price equilibria;

$$\varepsilon_t = \sum_{i=1}^{n_t} \delta_{it} \qquad (2.14)$$

where δ_{it} is the ith equilibrium price increment in day t, and n_t represents the number of daily information arrivals.

They also assume that if δ_i is iid with mean zero and variance σ^2, then if n_t is sufficiently large, by virtue of the Central Limit Theorem:

$$\varepsilon_t | n_t \sim N(0, \sigma^2 n_t)$$

The link between GARCH models and the economic theory is provided by assuming that the daily number of information arrivals follows an autoregressive process. This can be represented by the following equation:

$$n_t = \alpha_0 + \alpha_1 \sum_{t=i}^{m} n_{t-i} + u_t \qquad (2.15)$$

Lamoureux and Lastrapes then define a variance term:

$$\Omega_t = E(\varepsilon_t^2 | n_t) = \sigma^2 n_t \qquad (2.16)$$

When this term is combined with the autoregressive structure of information arrival:

$$\Omega_t = \sigma^2 \alpha_0 + \alpha_i \sum_{i=1}^{m} \Omega_{t-i} + \sigma^2 u_t \qquad (2.17)$$

Lamoureux and Lastrapes argue that this equation captures the type of persistence in conditional variance that can be modelled by estimating a GARCH model. The difficulty, however, is that the information flow variable, n_t, is not directly observable. Trading volume is, therefore, proposed as a proxy. This is consistent with the approach adopted by Clark (1973). As Foster (1996) explains, if volume is exogenous to the volatility in the system it can be entered into a GARCH model as follows;

$$Y_t = X'_t \xi + \varepsilon_t$$

$$\varepsilon_t \big| (V_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) \sim N(0, h_t)$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} + \gamma V_t \qquad (2.18)$$

where Y_t is the price returns variable and V_t is trading volume. If volume can account for all of the GARCH effects in the data then γ will be positive and statistically significant, and α_i and β_j will be small and statistically insignificant.

This approach appears to be a direct test of the MDH and would, therefore, reveal a great deal about the relationship between volume and volatility. It does, however, have one or two shortcomings that ultimately prevent such a test. Although it is close to the spirit of the Clark (1973) study, more modern versions of the MDH, in particular Harris (1987) and Tauchen and Pitts (1983), argue that the link between volume and volatility occurs because they are both driven by the same directing variable, namely information. A weakness of this method is that it only approaches the relationship between the two variables in terms of the impact of volume on volatility. Therefore, this GARCH based approach is not a direct test of the MDH. As Foster (1996) points out, it is really a test of whether volume is a proxy for information.

It is also necessary to be aware of a possible simultaneity problem. As Lamoureux and Lastrapes (1990), Najand and Yung (1991) and Foster (1996) note, volume may not be exogenous as assumed by the model above. One possible solution, if volume is endogenous, is to re-estimate the model using lagged values of volume. This, of

course, assumes that lagged volume is a good proxy for current volume which may not be the case.

Ultimately, however, despite the fact that it is not a direct test of any of the economic theories relating volume and volatility, it does provide us with an important first step in being able to describe the possible role of volume in financial markets.

The next section presents the empirical results of the investigation of the relationship between volume and volatility using the two approaches that have just been discussed in detail.

2.5 EMPIRICAL RESULTS

In this section the causality and ARCH techniques will be used to investigate the relationship between volatility and the volume of trade in UK futures markets. The aim is to establish whether there is a link between the two key variables and, if so, why this link might occur. The discussion begins by looking at the construction of the data set and some summary statistics before implementing the techniques described in section 2.4.

2.5.1 DATA AND PRELIMINARY ANALYSIS

The data set is constructed from daily settlement prices for nearby futures contracts written on five different commodities and financial instruments together with their corresponding daily trading volumes. Table 2.1 provides the contract details, the period covered and the number of observations in each sample. There is also some indication of whether or not the contract is heavily traded relative to similar contracts on the same market. This is achieved simply by comparing daily trading volumes for each contract over the period under investigation. As already mentioned in section 2.3, the aim of investigating the volume-volatility relation for a variety of futures contracts is to determine the particular characteristics of its relationship for different assets. The nature of these differences is discussed more fully later.

The nearby futures contract is selected since it attracts the greatest amount of trading activity. This minimises problems due to long periods without volume where prices

are 'stale'[5]. The futures returns series are calculated as the first difference of the log of settlement prices.

As Sutcliffe (1993) notes, this has advantages over the use of actual price changes, particularly where an index is being studied. The general trend of market indices is upward. Since price changes are scale dependent the data will reflect this trend. Logarithmic price changes have the advantage of not being scale dependent[6].

The measurement of volatility is a matter of much debate[7]. The consensus appears to be that the best measures of volatility use a large amount of information. The Garman-Klass (1980) measure, for example, uses daily high, low, open and closing prices. Ultimately, however, the choice is restricted by the observations available and the length of period over which volatility needs to be calculated. This study uses the squared value of the first difference of the natural logarithm of settlement prices. This measure is used in a number of studies, in particular Grammatikos and Saunders (1986). Since this study is in some sense a re-examination of their work, this seems appropriate.

Daily volume is calculated as the total number of contracts traded in the nearest contract per day with one exception. The sample data from the London Metals Exchange (LME) is a little different in that volume reflects the total amount of trade on the market for that day, as opposed to a single contract, and the returns series is constructed from three month forward rather than futures prices. The distinction between the two volume measures should provide an interesting comparison. It may show whether a distinction can be made between the total amount of information in a market and that specific to a particular contract.

---

[5] So-called stale prices occur where prices do not reflect economic activity. If there is no trading, prices are set based on the closing price of the last day when trading did occur.
[6] For a detailed discussion of the issues see Sutcliffe (1993: 155-162)
[7] For a detailed discussion of the issues see Sutcliffe (1993: 176-179)

Table 2.1. UK Futures Contract Details

Futures Contract	Sample Period	No. of observations	Activity Indicator	Market
FTSE 100	2/1/92-21/6/96	1129	high	LIFFE
Long Gilt	2/1/92-23/5/96	1133	high	LIFFE
Brent Oil	16/1/92-18/6/96	1125	high	IPE
Cocoa No. 7	2/1/92-31/7/96	1158	high	LCE
Tin	3/1/92-28/6/96	1135	low	LME

Note: LCE is the London Commodity Exchange, LME is the London Metals Exchange, LIFFE is the London International Financial Futures and Options Exchange, and IPE is the International Petroleum Exchange.

The link between volume and volatility may also be affected by the intrinsic nature of the flows of information to both the futures market and the underlying spot market. The Gilt contract is likely to be affected by trends in Government spending. If the announcement of Government spending plans occurs relatively infrequently, does this mean that mean volatility in the Long Gilt contract will be relatively low? The return and volatility patterns of the FTSE 100 futures contract are connected to the sorts of firms that the index represents; large company stocks may be more or less volatile. It is also a characteristic of this market that it is able to absorb large volumes of buy and sell orders without large changes in prices.

One of the problems with data surrounding the contract expiration is that, during the delivery month, trading shifts gradually to the contract with the next nearest delivery date. This is often an attempt by hedgers and speculators to avoid the transaction costs incurred from having to roll over their position in the near future. Trading volume on the nearby contract, therefore, tends to fall as expiration approaches, resulting in higher transaction costs for existing traders, which in turn motivates them to switch to the more liquid next nearest futures contract. Two data sets were, therefore, constructed, one of which eliminates the last twenty days of trading in each contract to determine the influence of these so-called 'delivery complications'. The only exceptions to this are the metals contracts which reflect forward positions.

The trends that are prevalent in volume data, whether as a result of seasonal effects, or the manifestation of rollover movements, reduce the power of our testing procedures. Although this study is not trying to test any of the theoretical models explicitly, the underlying theme is the role of information in determining the relationship between volume and volatility. To distinguish between these different influences it is important, therefore, to eliminate all trends that are not due to the arrival of news into the market. The filtering out of the trends follows the procedure adopted by Anderson (1996). Theory provides very little guidance here as to the most appropriate method. The method used here estimates a trend component that produces a 'normal' or expected volume series, and the detrended series is then obtained by dividing each trading figure with the corresponding 'normal' volume for that trading day. The adoption of a sixty three day moving average centred on the estimated trend component is justified by the four expiration dates for the financial futures and the possible effects of the change in season on the commodity futures. A common detrending method is necessary to ensure that the results across commodities are comparable.

Tables 2.2 and 2.3 show the mean daily return and daily volume for the eight different futures under investigation, with and without detrended volume, as well as mean daily price variability, standard deviation, and measures of skewness and kurtosis.

There are one or two points worth noting from the tables. The first is that, for both the filtered and unfiltered data, the mean returns tend to be very small with little dispersion about the origin. By excluding the delivery month the volume data in table 2.2 has become less predictable with the standard deviation rising in all cases except the Long Gilt contract. This may in some way have taken out the impact of roll-over effects. This result is supported by the filtered series suggesting, as perhaps expected, that there is less news-related activity as the contract nears expiration. The majority of traders are simply closing out positions. All returns series exhibit evidence of thicker tails than normal. This has important implications with regard to possible GARCH testing. The existence of leptokurtic distributions is exactly what the ARCH methodology is designed to capture.

39

The most volatile market is the cocoa market. Cocoa production is greatly affected by weather conditions, disease, insects, crop care and political conditions in the producing countries. Production is limited to countries not more than 20 degrees north or south of the equator. The world's leading producer is the Ivory Coast, followed by Brazil, Ghana, Malaysia, and Indonesia. Variations in political and economic stability in these countries over the last few years have undoubtedly had an adverse effect on the stability of market prices. In contrast, the suggestion that the mean volatility of the Gilt contract would be relatively low is supported by the evidence. This may be due to the idiosyncrasies of information flows from the Government, or it may simply reflect the maturity of the market in line with the arguments of Tauchen and Pitts (1983).

Perhaps the most surprising finding, however, is that there is no direct correlation between those contracts with the highest average daily trading volume and those with the highest price return variability. This may reflect the mix of investors trading in each contract. The high volume of trading in the financial futures markets possibly captures large numbers of uninformed traders whose actions have less impact on prices. This is in contrast to the commodity futures, particularly cocoa, where there is evidence of relatively lower volume and higher volatility. Although cocoa is heavily traded in comparison with other commodities, the figures suggest that the market has little depth. If there are informed traders present their actions will result in volatile prices.

Table 2.2: UK Futures Contract Return, Volume and Volatility Summary Statistics for

the Unfiltered Data (1992-1996)

	Including Expiration Month				Excluding Expiration Month			
	Mean	S.D.	Skewness	Kurtosis	Mean	S.D.	Skewness	Kurtosis
FTSE 100								
Return	0.274E-03	0.885E-02	0.067	1.714	0.238E-03	0.898E-02	0.067	1.534
Volume	11744.8	4661.2	1.100	2.109	9119.4	5412.0	0.400	0.394
σ^2	0.783E-04	0.151E-03	7.691	93.819	0.805E-04	0.152E-03	7.493	90.929
Long Gilt								
Return	0.147E-03	0.533E-02	0.023	3.270	0.131E-03	0.539E-02	-0.031	3.119
Volume	36462.2	31898.5	1.000	1.522	50149.7	27492.3	1.335	3.024
σ^2	0.284E-04	0.653E-04	9.243	149.603	0.291E-04	0.657E-04	9.092	146.701
Brent Oil								
Return	0.454E-03	0..408E-	0.012	3.409	0.158E-03	0.133E-01	0.122	3.180
Volume	19265.1	6480.8	0.454	0.607	12743.7	8071.9	1.055	1.236
σ^2	0.198E-03	0.461E-03	7.344	81.491	0.178E-03	0.405E-03	6.626	65.3133
Cocoa								
Return	0.170E-03	0.162E-01	0.406	2.415	-0.001	0.143E-01	0.214	2.119
Volume	1116.4	1320.4	2.873	20.843	2026.2	1515.9	2.773	14.965
σ^2	0.263E-03	0.553E-03	5.086	35.324	0.205E-03	0.415E-03	5.170	37.554
Tin								
Return	0.964E-04	0.122E-01	0.068	7.061				
Volume	3667.9	1921.8	1.042	1.793				
σ^2	0.149E-03	0.449E-03	13.531	279.914				

Note: σ^2 represents the variance of returns measured by $(\text{Return}_t)^2$. S.D. is the standard deviation for

each series, x_t, measured as: $(\sum_{t=1}^{n} (x_t - \bar{x})^2 /(n-1))^{1/2}$, where n is the number of observations.

$$Kurtosis = \left\{ \frac{n}{(n-1)(n-2)(n-3)} \sum_{j=1}^{n}((x_j - \bar{x})/s)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

$$Skewness = \frac{n}{(n-1)(n-2)} \sum_{j=1}^{n}((x_j - \bar{x})/s)^3, \text{ where s is the sample standard deviation and n is the number}$$

of observations in the sample.

Table 2.3: UK Futures Contract Return, Volume and Volatility Summary Statistics for the Filtered Data (1992-1996)

	Including Expiration Month				Excluding Expiration Month			
	Mean	S.D.	Skewness	Kurtosis	Mean	S.D.	Skewness	Kurtosis
FTSE 100								
Return	0.293E-03	0.896E-02	0.064	1.685	0.257E-03	0.909E-02	0.061	1.520
Volume	0.998	0.329	0.911	2.111	0.995	0.561	0.106	0.079
σ^2	0.804E-04	0.154E-03	7.557	90.106	0.826E-04	0.155E-03	7.354	87.147
Long Gilt								
Return	0.143E-03	0.542E-02	0.033	3.161	0.127E-03	0.548E-02	-0.022	3.019
Volume	0.995	0.772	0.466	0.555	0.996	0.418	0.891	2.265
σ^2	0.294E-04	0.669E-04	9.050	143.002	0.301E-04	0.674E-04	8.900	140.152
Brent Oil								
Return	0.495E-03	0.139E-02	0.044	3.694	0.127E-03	0.126E-02	0.051	3.095
Volume	1.000	0.298	0.274	0.517	0.993	0.585	1.037	1.133
σ^2	0.193E-03	0.461E-03	7.664	86.460	0.159E-03	0.360E-03	7.774	95.088
Cocoa								
Return	0.387E-03	0.161E-01	0.473	2.382	-0.001	0.143E-01	0.244	2.108
Volume	1.072	1.276	3.209	27.573	0.997	0.724	3.220	25.060
σ^2	0.260E-03	0.548E-03	5.234	37.726	0.204E-03	0.413E-03	5.177	38.065
Tin								
Return	0.128E-03	0.124E-01	0.677	6.895				
Volume	0.996	0.348	1.156	2.452				
σ^2	0.154E-03	0.460E-03	13.244	267.166				

Note: σ^2 represents the variance of returns measured by $(\text{Return}_t)^2$. S.D. is the standard deviation for each series, x_t, measured as: $(\sum_{t=1}^{n}(x_t - \bar{x})^2 /(n-1))^{1/2}$, where n is the number of observations.

$$\text{Kurtosis} = \left\{ \frac{n}{(n-1)(n-2)(n-3)} \sum_{j=1}^{n}((x_j - \bar{x})/s)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

$$\text{Skewness} = \frac{n}{(n-1)(n-2)} \sum_{j=1}^{n}((x_j - \bar{x})/s)^3$$, where s is the sample standard deviation and n is the number of observations in the sample.

Tables 2.4 and 2.5 show contemporaneous correlations between volume and price return variability for all five contracts. The results from the filtered and unfiltered data indicate, with one exception, that there is a positive contemporaneous relationship between the volume of trade and volatility. The exception is the cocoa contract. It is only when the expiration month is excluded that a positive relationship becomes apparent. This suggests that the 'delivery complications' referred to earlier obscure the relationship between the two key variables. In fact, for all of the contracts there is a difference between the results including and excluding the expiration month. The relationship also appears to differ between the filtered and unfiltered data sets. It

is difficult to say exactly why these variations occur, but they do illustrate the importance of acknowledging the potential influence of underlying trends in the data.

Table 2.4: UK Futures Market Volume and Volatility Contemporaneous Correlation Coefficients for the Unfiltered Data

	FTSE 100	Long Gilt	Brent Oil	Cocoa	Tin
Inc Exp	0.362*	0.137*	0.332*	-0.044	0.327*
Exl Exp	0.217*	0.329*	0.266*	0.346*	-

Note: Inc Exp and Exl Exp include and exclude the expiration month respectively. The significance of the correlation coefficients can be tested using the following test statistic:

$\dfrac{r}{\sqrt{(1-r^2)/(n-2)}}$, where r is the correlation coefficient and n is the number of observations in the

sample. Under the null hypothesis of no correlation this is distributed as a Student t-test with n-2 degrees of freedom. The asterisk indicates that the null hypothesis is rejected for an alternative that a correlation exists at the 5% significant level.

Table 2.5: UK Futures Market Volume and Volatility Contemporaneous Correlation Coefficients for the Filtered Data

	FTSE 100	Long Gilt	Brent Oil	Cocoa	Tin
Inc Exp	0.421*	0.155*	0.318*	-0.022	0.277*
Exl Exp	0.258*	0.344*	0.149*	0.365*	-

Note: Inc Exp and Exl Exp include and exclude the expiration month respectively. See table 2.4 for details of the test of significant correlation. The asterisk indicates that the null hypothesis is rejected for an alternative that a correlation exists at the 5% significant level.

This issue will be considered further in chapter 4. These results do, however, provide some preliminary justification for this study. The next stage is to consider the volume-volatility relationship in more detail.

2.5.2 CAUSALITY TESTS

The use of the VAR methodology for the purposes of testing causality between two variables is dependent on the stationarity of the individual series. Charemza and Deadman (1997) point out that if the variables are nonstationary these tests are only approximately valid, or may not be valid at all. Tables 2.6 and 2.7 provide the Dickey-Fuller (1979) test statistics for the presence of unit roots[8] in the volume and return variance series.

Table 2.6: Unit Root Tests of the Volume and Return Variance Series for the Unfiltered Data

Contract	Variable	Inc Exp	Exl Exp
FTSE 100	Volatility	-20.697	-28.447
	Volume	-8.109	-8.547
Long Gilt	Volatility	-31.566	-31.256
	Volume	-6.353	-6.286
Brent Oil	Volatility	-13.124	-9.714
	Volume	-12.076	-14.788
Cocoa	Volatility	-20.393	-30.675
	Volume	-10.043	-17.700
Tin	Volatility	-14.551	-
	Volume	-9.721	-

Note: The Dickey-Fuller test is carried out under a null hypothesis of non-stationarity against the alternative of stationarity. If the absolute value of the test statistic exceeds the absolute value of the critical the null is rejected. The critical value for the Dickey-Fuller test is -2.864 at the 5% significance level. Inc Exp and Exl Exp include and exclude the expiration month respectively.

In all cases the absolute value of the test statistic exceeds the absolute value of the critical value. Therefore, the volume and volatility series for every contract is stationary and we can proceed to test for causality.

---

[8] See Charemza and Deadman (1997) section 5.3 for a detailed discussion of the process of testing for the order of integration.

The causality tests were carried out using the Granger test within the VAR methodology as described in section 2.4.1. In each case the appropriate lag length was determined using the adjusted likelihood ratio test statistic, and each individual equation was checked to ensure that there was no serial correlation in the residuals using the Breusch-Godfrey Lagrange Multiplier test statistic.

Table 2.7: Unit Root Tests of the Volume and Return Variance Series for the Filtered Data

Contract	Variable	Inc Exp	Exl Exp
FTSE 100	Volatility	-28.408	-27.723
	Volume	-24.034	-9.010
Long Gilt	Volatility	-30.828	-30.496
	Volume	-20.765	-14.482
Brent Oil	Volatility	-12.680	-12.198
	Volume	-25.657	-14.693
Cocoa	Volatility	-19.759	-29.633
	Volume	-9.784	-24.602
Tin	Volatility	-14.184	-
	Volume	-23.684	-

Note: The Dickey-Fuller test is carried out under a null hypothesis of non-stationarity against the alternative of stationarity. If the absolute value of the test statistic exceeds the absolute value of the critical value the null is rejected. The critical value for the Dickey-Fuller test is -2.864 at the 5% significance level. Inc Exp and Exl Exp include and exclude the expiration month respectively.

Tables 2.8 and 2.9 show the results of the causality tests using the Granger method. At first glance the results appear to show that in a large number of cases there is no causality between volume and volatility. This is particularly true where the data set takes account of trends in the data as well as excluding the expiration month. Once again, the differences in the results between the data sets illustrates the importance of

being aware of the impact that these trends have on the outcome of statistical test procedures.

Assuming that it is preferable to eliminate trends in the data, only the FTSE 100 and tin contracts show that volume causes volatility in the Granger sense. Thus, lagged values of volume have a determining influence on current volatility. This could be interpreted as evidence that within these markets there is some sort of persistence effect. The information contained in volume takes some time to become fully revealed in prices. This might seem an unusual result for the FTSE 100 where one would expect, due to the very competitive nature of the market, that the value of information has a high rate of decay. In a market where the volume of trade is much lower, like tin, it may not be in an investor's best interests to reveal their intentions too quickly. It may be beneficial to hide the information in a series of smaller trades over a longer trading period.

In terms of volatility Granger causing volume, a statistically significant relationship exists only for the FTSE 100 and Long Gilt contracts. Jain and Joh (1988) argue that this can occur from changes in price that trigger stop-loss orders[9], or from investors who are slow to react to price movements. The mix of investors in these two markets, particularly the existence of uninformed or 'noise' traders, who react to actions of others, may help to explain such phenomena. This suggests that the lack of causality between volatility and volume for the other contracts might be explained by the relatively smaller numbers of such trend following investors.

The fact that there does not appear to be strong bi-directional causality, except for the FTSE 100 contract, could be mistakenly interpreted as evidence that links between volume and volatility are relatively poor, or non-existent.

---

[9] A stop order only becomes a market order if a certain price level is reached. A stop order to buy instructs a broker to act on behalf of an investor and buy at whatever price is available once the stop price has been reached. A stop order to sell works in the opposite direction. The broker will sell as soon as the stop price is reached. This instrument is useful if any adverse movement in prices is likely to result in losses for the investor. A stop order will prevent any further losses (a stop loss).

It is important to bear in mind, however, that this test can only determine the impact of lagged values of one variable on the current value of the other. Although it is not possible to explicitly test any of the underlying theories within this framework, the absence of causality should not really be a surprise.

Table 2.8: Granger Causality Tests of the Relationship Between Volume and Volatility based on the Unfiltered Data

Contract	Incl Exp		Exl Exp	
	Vol→Var	Var→Vol	Vol→Var	Var→Vol
FTSE 100	30.939*	40.481*	16.625	27.841*
Long Gilt	20.459*	35.665*	22.884	38.897*
Brent Oil	10.971**	9.111	12.758**	11.295
Cocoa	21.170	1.901	7.255	13.848
Tin	34.540*	11.254	-	-

Note: The Granger test is based on the following VAR model where y_t and x_t represent volume or volatility depending on the equation being considered:

$$y_t = \alpha_0 + \sum_{j=1}^{k} \alpha_j y_{t-j} + \sum_{j=1}^{k} \beta_j x_{t-j} + \varepsilon_t$$

If $\beta_1 = \beta_2 = \ldots = \beta_k = 0$ then x does not cause y in the Granger sense. The statistics above represent the likelihood ratio statistics based on this test. Vol→Var and Var→Vol indicate a test that volume 'causes' volatility and volatility 'causes' volume respectively. * indicates that the null hypothesis of no Granger causality is rejected in favour of the alternative that x does cause y in the Granger sense at the 5% level of significance. ** indicates rejection at the 10% level of significance. The critical value is determined by the number of lags in the VAR model which are not quoted here.

As Sutcliffe (1993) points out, if the link between volume and volatility occurs because they are both driven by the same underlying variable, as in the MDH, there is no reason to assume that one variable should Granger cause another. However, under this model there is likely to be some evidence of a contemporaneous correlation. Such links have already been shown to exist in section 2.5.1.

These are very interesting results. They suggest that there are links between volume and volatility but the relationship is not as obvious as previous research indicates. In

particular, it questions the work of Jain and Joh (1988) who assume that evidence of causality is supportive of the MDH. Further consideration should also be given to those studies, for example Hiemstra and Jones (1994), who argue that the failure to find evidence of linear causality is a weakness in the methodology rather than an economically interpretable result[10]. If information is the missing factor in this analysis we need to investigate volume and volatility in more detail.

Table 2.9: Granger Causality Tests of the Relationship Between Volume and Volatility based on the Filtered Data

Contract	Incl Exp		Exl Exp	
	Vol→Var	Var→Vol	Vol→Var	Var→Vol
FTSE 100	17.138	29.899*	27.664*	35.965*
Long Gilt	23.312**	49.550*	13.213	33.951*
Brent Oil	21.542*	8.479	5.990	9.731
Cocoa	25.432	15.230	20.950	24.253
Tin	11.097*	4.004	-	-

Note: The Granger test is based on the following VAR model where y_t and x_t represent volume or volatility depending on the equation being considered:

$$y_t = \alpha_0 + \sum_{j=1}^{k} \alpha_j y_{t-j} + \sum_{j=1}^{k} \beta_j x_{t-j} + \varepsilon_t$$

If $\beta_1 = \beta_2 = \ldots = \beta_k = 0$ then x does not cause y in the Granger sense. The statistics above represent the likelihood ratio statistics based on this test. Vol→Var and Var→Vol indicate a test that volume 'causes' volatility and volatility 'causes' volume respectively. * indicates that the null hypothesis of no Granger causality is rejected in favour of the alternative that x does cause y in the Granger sense at the 5% level of significance. ** indicates rejection at the 10% level of significance.. The critical value is determined by the number of lags in the VAR model which are not quoted here.

2.5.3 GARCH ANALYSIS RESULTS

The GARCH analysis carried out in this chapter follows the approaches of Lamoureux and Lastrapes (1990), Najand and Jung (1991), and Foster (1996). The return series of

---

[10] The usefulness of non-linear causality techniques is also questioned by Brooks (1998). He argues that they have limited use in forecasting where there is little guidance as to the form of the volume-volatility relation.

each contact is modelled within a GARCH (1,1) framework. Section 2.5.1 has already indicated the presence of leptokurtic return distributions consistent with the phenomena that this approach seeks to explain.

Volume is then used to see if it can account for the GARCH effects in the returns series. This assumes that volume is an exogenous variable in the system. In order to establish whether simultaneity bias is present in the model it is also estimated using lagged volume to proxy for current levels of trading. In acknowledging the impact of trends on the results, this analysis uses only the filtered data that excludes the expiration month. The maximum likelihood estimation procedure[11] is exploited in each case. Various starting values were used to check the robustness of the results to ensure that global maxima were achieved. Tables 2.10, 2.11, and 2.12 present the results.

These results are very revealing. Table 2.10 indicates that the returns series for every contract under consideration can be modelled using a GARCH specification. Closer examination of the coefficient values indicates the presence of integrated GARCH (IGARCH) effects. Engle and Bollerslev (1986) argue that if the sum of a_1 and b_1 is equal to one, this implies that there is persistence in the conditional volatility.

The most striking results, however, are presented in table 2.11. When volume is added to the conditional variance equation, in the majority of cases, changes occur in the coefficients a_1 and b_1. Indeed, in all cases b_1 becomes smaller and, with the exception of the tin contract, insignificant. The value of the a_1 coefficient, with the exception of the Brent Oil contract, where it becomes smaller and insignificant, either remains relatively stable, or increases slightly in value. This suggests that the non-normality in the returns series can be largely accounted for by contemporaneous volume, which for each contract is statistically significant. If the fat-tailed distributions in returns are caused by the variance of information flows then volume is

---

[11] Estimation was carried out using the RATS econometrics package. The programs used are available on request.

acting as a good proxy for that information. It is not a perfect proxy since the lagged

squared residuals, in most cases, continue to be significant[12].

Table 2.10: GARCH Analysis of the Volume-Volatility Relation based on the Filtered

Data Excluding the Expiration Month

GARCH	Contract				
Coefficient	FTSE 100	Long Gilt	Brent Oil	Cocoa	Tin
a_0	0.817E-06	0.832E-06*	0.184E-05*	0.398E-05*	0.102E-04*
	(1.192)	(2.645)	(2.614)	(2.241)	(6.139)
a_1	0.037*	0.039*	0.045*	0.040*	0.192*
	(3.750)	(4.169)	(5.822)	(4.833)	(10.401)
b_1	0.953*	0.933*	0.945*	0.944*	0.754*
	(69.428)	(53.162)	(104.495)	(77.065)	(34.741)

Note: GARCH (1,1) Estimation.

$$R_t = \alpha + \varepsilon_t$$
$$h_t = a_0 + b_1 h_{t-1} + a_1 \varepsilon_{t-1}^2$$

where R_t represents the futures price returns series. The T-statistics for each coefficient (written to 3 decimal places) are in brackets. * denotes significance at the 5% level.

As already noted, the a_1 and b_1 coefficients do not become insignificant when volume is added to the conditional variance equation of the tin contract. In this case although volume accounts for some of the GARCH effects, it is a less good proxy for information than exhibited in the other contracts. This may be because in this case volume reflects the total amount of activity in the tin market. Therefore, the information captured in this variable may not reflect the idiosyncrasies of the three month forward contract.

---

[12] For an explanation of why volume may not be a perfect proxy for information see chapter 4, section 4.2.1, and the discussion of the Tauchen and Pitts (1983) specification of the MDH.

The previous studies that have exploited the GARCH methodogy in this way have all pointed out the potential problem of assuming that volume is exogenous to the system. Table 2.12 presents the results of replacing volume with lagged volume in the conditional variance equation. In every case lagged volume is not significant and the a_1 and b_1 coefficients are very similar to those presented in table 2.10. This suggests that either the results in table 2.11 are incorrect, or more simply, that lagged volume is not a good proxy for contemporaneous volume.

Table 2.11: GARCH Analysis of the Volume-Volatility Relation based on the Filtered
Data Excluding the Expiration Month; Contemporaneous Volume

GARCH	Contract				
Coefficient	FTSE 100	Long Gilt	Brent Oil	Cocoa	Tin
a_0	0.514E-04*	0.161E-10	0.632E-04*	0.605E-04*	0.961E-15
	(3.794)	(0.000)	(8.628)	(4.730)	(0.000)
a_1	0.047	0.054*	0.185E-04	0.040	0.273*
	(1.553)	(2.010)	(0.001)	(1.549)	(7.931)
b_1	0.144E-04	0.201E-05	0.803E-10	0.030	0.302*
	(0.000)	(0.000)	(0.000)	(0.572)	(7.868)
γ	0.264E-05*	0.263E-04*	0.961E-04*	0.120E-03*	0.572E-04*
	(4.725)	(14.178)	(9.269)	(11.923)	(6.284)

Note: GARCH (1,1) Estimation.

$$R_t = \alpha + \varepsilon_t$$
$$h_t = a_0 + b_1 h_{t-1} + a_1 \varepsilon_{t-1}^2 + b_2 VOL_t$$

where R_t represents the futures price return series and Vol_t represents volume. The T-statistics for each coefficient (written to 3 decimal places) are in brackets. * denotes significance at the 5% level.

Table 2.12: GARCH Analysis of the Volume-Volatility Relation based on the Filtered

Data Excluding the Expiration Month; Lagged Volume

GARCH	Contract				
Coefficient	FTSE 100	Long Gilt	Brent Oil	Cocoa	Tin
a_0	0.817E-06	0.833E-06	0.184E-05	0.336E-05	0.205E-05
	(0.749)	(1.424)	(0.580)	(1.058)	(0.357)
a_1	0.037*	0.039*	0.045*	0.039*	0.187*
	(3.613)	(4.939)	(5.850)	(4.579)	(9.802)
b_1	0.953*	0.933*	0.945*	0.944*	0.756*
	(68.362)	(67.288)	(101.268)	(74.743)	(33.788)
γ	0.106E-10	0.858E-11	0.983E-12	0.191E-12	0.858E-05
	(0.000)	(0.000)	(0.000)	(0.000)	(1.578)

Note: GARCH (1,1) Estimation.

$$R_t = \alpha + \varepsilon_t$$
$$h_t = a_0 + b_1 h_{t-1} + a_1 \varepsilon_{t-1}^2 + b_2 \, VOL_{t-1}$$

where R_t represents futures price returns series and Vol_{t-1} represents lagged volume. The T-statistics for each coefficient (written to 3 decimal places) are in brackets. * denotes significance at the 5% level.

2.6 CONCLUSION

These results represent an important first step in determining the role of volume in futures markets, and in particular its relationship with price volatility. This chapter has examined this relationship using established techniques that have already been exploited in this field. However, this work differs from the work of others because it has used these techniques without assuming that they represent a strict test of any of the theories that might explain the volume-volatility relation. The results certainly appear to provide *prima facie* support for the role of information in determining this linkage, but this is a matter that requires further investigation.

The main interesting points to draw from this analysis are:

1. The contemporaneous links between the volatility of futures prices and the volume of trade, and the apparently conflicting results of the causality tests.

2. The use of volume to capture the non-normality in futures price returns.

3. Trends in data, in particular those surrounding expiration, must be properly accounted for to ensure that they do not affect the statistical significance of any tests.

These points help to raise some issues that provide the basis for the next three chapters. As already mentioned, the aim of this study was to carry out a preliminary investigation into the role of the volume of trade. Although we have made an important first step, we have still to determine its precise function. We have also said relatively little about the possible impact of the distinction between informed and uninformed traders. In chapter 4 a detailed investigation is carried out into the relationship between volume and volatility; specifically a direct test of the MDH. Chapter 5 further extends our understanding of the role of volume by considering its impact on the cost of trading in a futures market.

The impact of trends in data also needs further investigation. This chapter illustrated the importance of such considerations while using relatively ad-hoc detrending methods. Chapter 4 will exploit a more sophisticated method that deals explicitly with 'delivery complications'.

The use of a range of contracts has provided an important preliminary insight into the trading processes of derivative assets. It has been shown that the highest volume contracts are not necessarily those with the greatest levels of volatility. The causality tests reveal that the FTSE 100 contract is unique in exhibiting bi-directional causality between the key variables. In contrast, the GARCH analysis suggests a certain degree of similarity between the contracts with regard to the role of volume. The exception is the tin three month forward contract where the measure of volume reflects total trade in tin contracts on the LME. The overall impression, however, is that the explanatory powers of the techniques used in this chapter are not sufficient to allow us to describe the trading processes in more detail. This will be addressed in the following chapters.

While the results of this chapter are informative about whether volume is a proxy for information, it is necessary to investigate this issue further by considering different aspects of information and the volume of trade. In addition to the ability of volume to account for GARCH effects in return series, Blume et al. (1994) argue that there is information inherent in the volume statistic itself. The next chapter looks explicitly at this issue by exploiting their work to examine whether volume can tell us anything about the dispersion and precision of information.

CHAPTER THREE: THE INFORMATION VALUE OF THE VOLUME OF TRADE

3.1 INTRODUCTION

The aim of this chapter is to further our understanding of the role of volume in UK futures markets. Chapter 2 made the first tentative steps in trying to explain the relationship between volume and volatility. The results suggested that a possible reason for the existence of a link between these two variables is that they are both driven by the same underlying process; the flow of information. The emphasis in this chapter changes to consider whether volume, in addition to acting as a proxy for information, may also be able to tell us something about its quality and dispersion.

The theory that volume can, in itself, provide information, has been developed by Blume et al. (1994). It uses as its basis a scenario where prices are not fully revealing and volume can be used as a tool to aid technical analysis. Section 3.2 explains the theoretical background to their work in detail.

Although the co-movement of volume and price changes has been well documented, as mentioned in chapter 2, very few studies are able to explain why this link occurs. Despite the approach of Blume et al. (1994) being largely consistent with other work in this field, very few empiricists have adopted their model. Section 3.3 is, therefore, a relatively short review of the work in this area. The model of Blume et al. (1994) appears, at first glance, to be quite complex. In fact, application of the concepts to the data is a relatively simple process. Section 3.3 describes this methodology. The results of this study are presented and discussed in section 3.4 and section 3.5 concludes.

3.2 THEORETICAL BACKGROUND

The work of Blume et al. (1994) is motivated by a desire to obtain a better understanding of the role of volume in financial markets. In particular, they focus on the information inherent in the volume statistic, and what traders can learn from observing volume. Blume at al. (1994) use as their starting point an argument that conventional models in the microstructure literature are limited by only using price as the information mechanism. The two models that they consider are those of Brown and Jennings (1989) and Grundy and McNichols (1989).

These two models analyse the price adjustment process of an asset within a rational expectations framework. The determination of the equilibrium point is dependent on investors' price and demand functions. Individuals base their assessment of the equilibrium price function on an information set that includes the current price of an asset and its supply. Investors use this price function to form their demand schedules. In the equilibrium an investor's assessment of the price will be correct and demand will equal supply.

Brown and Jennings (1989) show that prices are not fully revealing. Therefore, an investor cannot obtain a 'complete' information set by observing a single price. Brown and Jennings argue that this implies a role for technical analysis; an investor can 'learn' by looking at the sequence of prices. In this scenario Blume et al. (1994) show that if individuals are allowed to observe volume then prices become fully revealing, and the role of technical analysis dissipates; there is no benefit to studying the sequence of prices.

Introducing volume to the Grundy and McNichols (1989) model reveals that volume does not contain any useful information. The restrictive assumptions that Grundy and McNichols impose to construct their model mean that the analysis of volume is very difficult. This is essentially because volume does not have a distinctive distribution from which inferences can be drawn.

Blume et al. (1994) address these shortcomings, that volume provides either too much or too little information, by developing their own model of the equilibrium price

process. This description of their model follows closely those given in both their original paper and in O'Hara (1997).

They begin by assuming that investors maximise utility functions of the form:

$$U(w_i) = -\exp(-w_i) \tag{3.1}$$

where w_i is investor i's terminal wealth. They introduce two assets; one is assumed risky, while the other is riskless. The final value of the risky asset, Ψ, is a random variable which is normally distributed with mean Ψ_0 and variance $1/\rho_0$. There are N traders each of whom receives an information signal relating to the value of the asset. A proportion μ of these traders (group 1) receives a signal at date t given by:

$$y_t^i = \phi + \omega_t + \varepsilon_t^i \tag{3.2}$$

where ω_t is a common error distributed as $N(0, 1/\rho_\omega)$ and ε_t^i is an idiosyncratic error distributed as $N(0, 1/\rho_t^i)$. ρ_t^i represents the precision of the signal which varies randomly over time. A proportion (1-μ) of the traders (group 2) receive a similar signal but the distribution of the idiosyncratic error is $N(0, 1/\rho^2)$ and so the precision is fixed over time. All of the information, apart from the ρ_t^i's is, therefore, known to all traders.

Blume et al. (1994) demonstrate that this represents a complex information problem. In a sense, the differences in signal make the investors in group 1 informed individuals, while those in group 2 are uninformed. These differences plus the common error (ω_t) make determining the 'true' value of the asset quite a challenge. Blume et al. simplify the task by noting that conditional on ω_t, each signal y_t^i is distributed as $N(\theta_t, 1/\rho_t^i)$ for investors in group 1 and $N(\theta_t, 1/\rho^2)$ for traders in group 2, where $\theta_t = \phi + \omega_t$. Under the Law of Large Numbers as $N \to \infty$ the mean signal in each group converges to θ_t; the true value plus a common error.

The task is to determine the equilibrium price, which can be found by analysing investor demand for the risky asset and calculating the price that eliminates excess demand. Blume et al. show that the equilibrium price as $N \to \infty$ can be replaced by:

$$p_1 = \frac{\rho_0 \varphi_0 + [\mu \rho_1^{s1} + (1-\mu)\rho^{s2}]\theta_1}{\rho_0 + [\mu \rho_1^{s1} + (1-\mu)\rho^{s2}]} \qquad (3.3)$$

where ρ_1^{s1} is the signal variance, defined by $\rho_1^{s1} = \rho_\omega \rho_t^1/(\rho_\omega + \rho_t^1)$ and similarly $\rho^{s2} = \rho_\omega \rho^2/(\rho_\omega + \rho^2)$. The difficulty is that this price is not revealing. There is a problem of asymmetric information. Investors in group 1 know ρ_1^{s1}, and so can infer θ_1, which tells them everything about the value of the asset. Those in group 2 cannot determine the value of the asset because they do not have all of the information. Blume at al. argue that this creates the incentive for investors in group 2 to use volume to help determine the value of the signal (θ_1).

Volume is calculated by summing the absolute value of demands at price p_1 and dividing by 2. In terms of per capita volume this is written as:

$$V_1 = \frac{1}{2N}\left(\sum_{i=1}^{N_1}\left|\rho_0(\varphi_0 - p_1) + \rho_1^{s1}(y_i^1 - p_1)\right| + \sum_{i=N_1+1}^{N}\left|\rho_0(\varphi_0 - p_1) + \rho^{s2}(y_i^1 - p_1)\right|\right) \qquad (3.4)$$

where N_1 is the number of investors in group 1. Blume et al. point out that because of the use of absolute values volume is not normally distributed, and it is, therefore, difficult to describe its behaviour. Some characterisation of the distributional properties of volume is necessary to understand how investors interpret the information in volume, and to establish its statistical background. Blume et al., therefore, re-write this volume expression in the following form:

$$V_1 = \frac{\mu}{2}\left[2\frac{\rho_1^{s1}}{(\rho_1^1)^{1/2}}\phi\left(\frac{\delta^1(\rho_1^1)^{1/2}}{\rho_1^{s1}}\right) + \delta^1\left(\Phi\left(\frac{\delta^1(\rho_1^1)^{1/2}}{\rho_1^{s1}}\right) - \phi\left(\frac{-\delta^1(\rho_1^1)^{1/2}}{\rho_1^{s1}}\right)\right)\right]$$
$$+ \frac{1-\mu}{2}\left[2\frac{\rho^{s2}}{(\rho^2)^{1/2}}\phi\left(\frac{\delta^2(\rho^2)^{1/2}}{\rho^{s2}}\right) + \delta^2\left(\Phi\left(\frac{\delta^2(\rho^2)^{1/2}}{\rho^{s2}}\right) - \phi\left(\frac{-\delta^2(\rho^2)^{1/2}}{\rho^{s2}}\right)\right)\right] \qquad (3.5)$$

where ϕ is the standard normal density, Φ is the standard normal cumulative distribution function, and $\delta^i = \rho_0(\varphi_0 - p_1) + \rho^{si}(\theta_1 - p_1)$. Using this equation it is possible to determine the relationship between price, the value of the signal, and signal quality. Blume et al. demonstrate how it can be used by assuming that $\rho^2 = 0$ and $\rho_1^1 > 0$. The relationship between changing volume and changes in ρ_1^1 can be represented by:

$$\partial V_1 / \partial \rho_1^1 = \frac{\mu}{2} \phi \left(\delta^1 \frac{\left(\rho_\omega + \rho_1^1 \right)}{\rho_\omega \left(\rho_1^1 \right)^{1/2}} \right) \left(\frac{\rho_\omega}{\left(\rho_1^1 \right)^{1/2}} \right) \frac{\left(\rho_\omega - \rho_1^1 \right)}{\left(\rho_\omega + \rho_1^1 \right)^{1/2}} \qquad (3.6)$$

It can be seen that where $\rho_1^1 < \rho_\omega$, per capita volume increases as ρ_1^1 increases and falls where $\rho_1^1 > \rho_\omega$. In this way volume can be used to obtain information on the signal quality, ρ_1^1, and therefore the signal value θ_1. The model also demonstrates that new information is as likely to result in low volume as high volume. Investors will not trade where signals are of low quality, and will also avoid investing where signals are of such high quality that everyone agrees on the value of the information.

Blume et al. demonstrate that in equilibrium, for a fixed level of signal precision, volume is strictly convex (or V-shaped) in price. The quality and dispersion of the information determines the steepness and dispersion of the V-shape. Therefore, investors can use volume to discriminate between the quality of information and the direction of information effects impounded in price.

It is this model that will be exploited in this chapter to further examine the role of volume in UK futures markets. The next section looks at the few studies that have analysed volume in this manner.

3.3 LITERATURE REVIEW

The literature in this field, as intimated earlier, is not extensive. The issues concerning the information value of volume in this context, have only been developed relatively recently. A review of the studies that have looked at the relationship between volume and volatility has already been carried out in section 2.3. What is clear is that they are limited in their ability to explain the role of volume. There is anecdotal evidence in support of the various theories but little in the way of strong tests.

The usual approach to examining the role of volume is to look at the correlation of volume with variables such as investor heterogeneity. Wang (1994) uses dividend information and private investment opportunities to discriminate between the volume

associated with informed and uninformed trading respectively. Kim and Verrecchia (1991) investigate the changes in volume that occur around public information announcements. If this information can be interpreted in a number of different ways then volume is likely to increase. Blume et al. (1994) argue that these two studies represent a different approach to understanding the role of volume because they do not infer information from the volume statistic itself. The information is only derived from its association with other variables.

A review of the early work looking at the relation between price changes and volume is carried out by Karpoff (1987). An example of a recent study is that of Gallant et al. (1992). They carry out an extensive statistical analysis of daily price and volume data for the S&P composite index over the period from 1928 to 1985. Among the techniques that they use is a scatter plot of detrended price changes versus standardised volume. The plot shows that large price movements are associated with large volume. The V-shaped scatter that is produced, although not explained in any detail, is consistent with that predicted by Blume et al. (1994).

The only study to have directly exploited the Blume et al (1994) approach is that by Foster (1996). He uses price and volume data for Brent and WTI crude oil futures for the period from 1990 to 1994. The scatter plots of detrended logarithmic volume against price returns reveal that volume is strictly convex in price. If a fixed level of precision is assumed then the results are suggestive of a wide dispersion of information. It is also possible to identify a symmetric response by volume to both positive and negative price changes. This is consistent with Karpoff's (1987) argument that this phenomenon is likely to be particular to futures markets which are not subject to the short selling restrictions imposed on spot markets.

This study aims to add to the existing literature by investigating the information role of volume in UK futures markets. We have already seen that for the five contracts considered in chapter 2 volume can account for a large proportion of the GARCH effects in price returns. If this is because, as we suspect, volume is acting as a proxy for information, can volume also tell us something about the quality and dispersion of that information? In particular this study will seek to investigate the following issues:

- does the relationship between volume and price changes differ across contracts in terms of the quality and dispersion of information?
- is the response of volume to positive and negative price changes symmetric?
- is quality and dispersion affected by the mix of investors in a market?

The next section explains the methodology of the Blume et al. (1994) approach.

3.4 METHODOLOGY

The methodology of the Blume et al. (1994) approach is very simple. They construct an 'ideal' set of data and use a scatter plot to examine the relation between price changes and volume. They then use the pattern of the scatter to determine the precision and dispersion of the information signal in their data. They show that as the precision of information increases, the V-shape relation between price changes and volume becomes more sharply defined, and reduces the dispersion of points in the scatter plot.

Blume et al. examine the dispersion of information by varying the number of traders in their high precision group, as defined in section 3.2. As the number of well informed investors increases and information is more widely dispersed the V-shape flattens out to form a ∇-shape.

The interpretation of scatter plots of price change and volume, therefore, requires some care. As Foster (1996) points out, it is very difficult to model changes in precision. It is more sensible to assume a given level of precision and then to concentrate on levels of dispersion and the volume-price change relation. The next section presents scatter plots associated with the UK futures data and discusses the implications.

3.5 EMPIRICAL RESULTS

In this section the Blume et al. (1994) approach is applied to five UK futures contracts to extend our understanding of the role of volume in derivative markets.

3.5.1 DATA

The data used in this study exploits the volume and returns series constructed in chapter 2. The importance of taking account of trends in the data has already been illustrated. Therefore, the analysis is restricted to the filtered data that eliminates the expiry month. The use of a price returns series differs slightly from the model specified by Blume et al. The implication is that it is the logarithmic change in price rather than price levels that is being analysed. It is necessary, therefore, to also use volume in logarithmic terms. This is consistent with the approach adopted by Foster (1996).

3.5.2 RESULTS

The scatter plots of logarithmic volume against price returns for the filtered series excluding the expiration month are presented in figures 3.1 to 3.5. The first important result is that in all cases the plots illustrate that the relationship between volume and price changes appears to be approximately symmetric. Therefore, the response by investors to either positive or negative price changes is the same. Karpoff (1987) argues that this is to be expected in futures markets, where the absence of restrictions on short selling eliminates the bias in the response to information that results in a fall in investor demand.

It is clear from looking at the diagrams that they are not as sharply defined as those produced by Blume et al. (1994) based on a manufactured data set. It is, therefore, not possible to make inferences regarding the precision of the information. This justifies our decision to assume a given level of information precision and to concentrate on its dispersion. The scatter plots indicate that the information in the markets for all of the futures contracts considered here is relatively well dispersed. One might expect the ratio of informed to uninformed traders to be lowest for the financial futures which are believed to attract large numbers of feedback traders. It is difficult, however, to discriminate between the contracts in this sense because they all exhibit a ∇-shape. One might, however, tentatively infer from the relatively well-defined base of the FTSE 100 contract that, contrary to our expectations, there is a relatively high proportion of informed traders in this market. The high dispersion of information in

62

all of the contracts suggests that the relationship between volume and price changes is not driven by so-called 'noise trading'.

In terms of what an investor can learn from the analysis of volume in this context it suggests that they need to exercise caution, since there is a high probability that the person they are trading with is carrying information.

Figure 3.1: Plot of Detrended Log Volume Against Price Returns for the FTSE 100 Futures Contract (1992-1996)
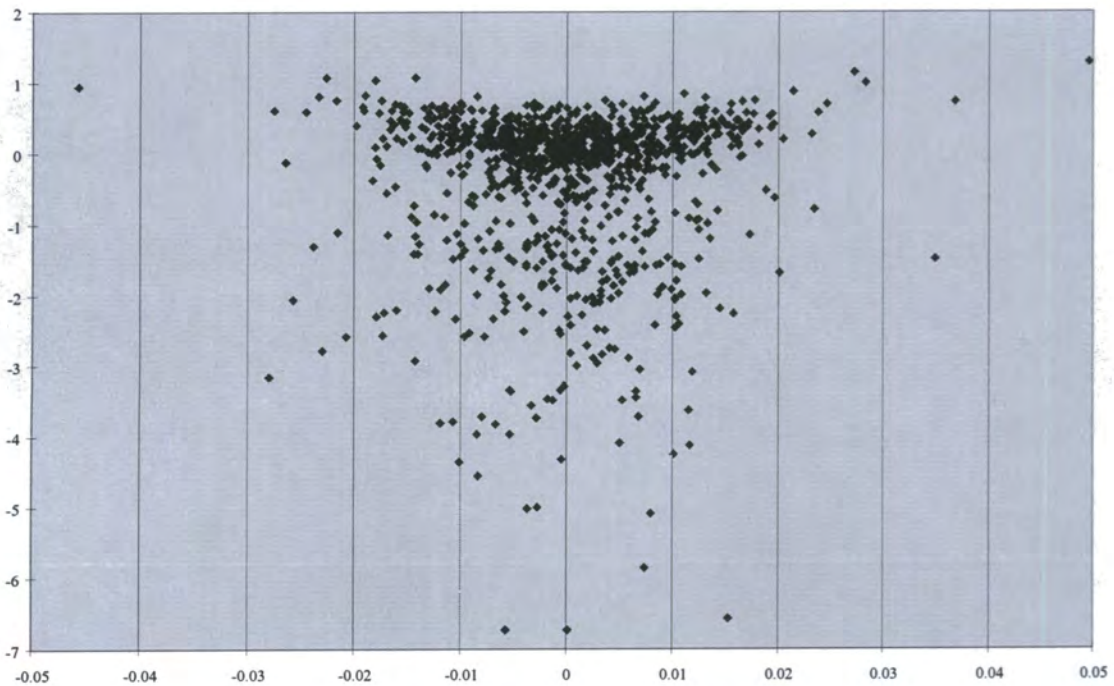
Figure 3.2: Plot of Detrended Log Volume Against Price Returns for the Long Gilt
Futures Contract (1992-1996)
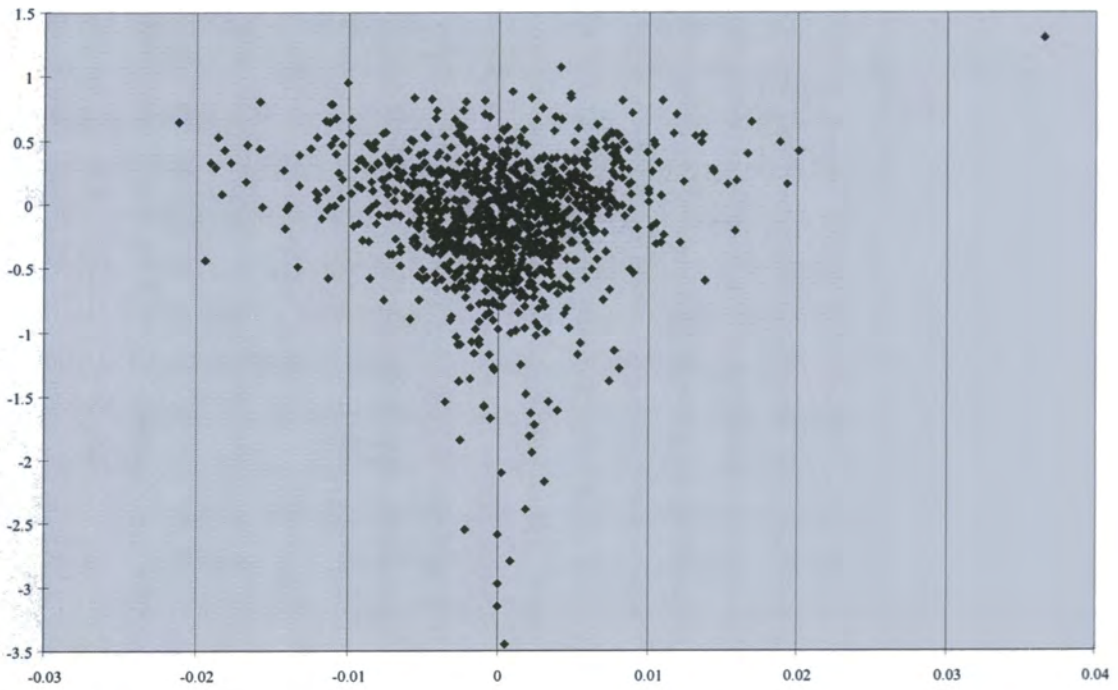
Figure 3.3: Plot of Detrended Log Volume Against Price Returns for the Brent Oil
Futures Contract (1992-1996)
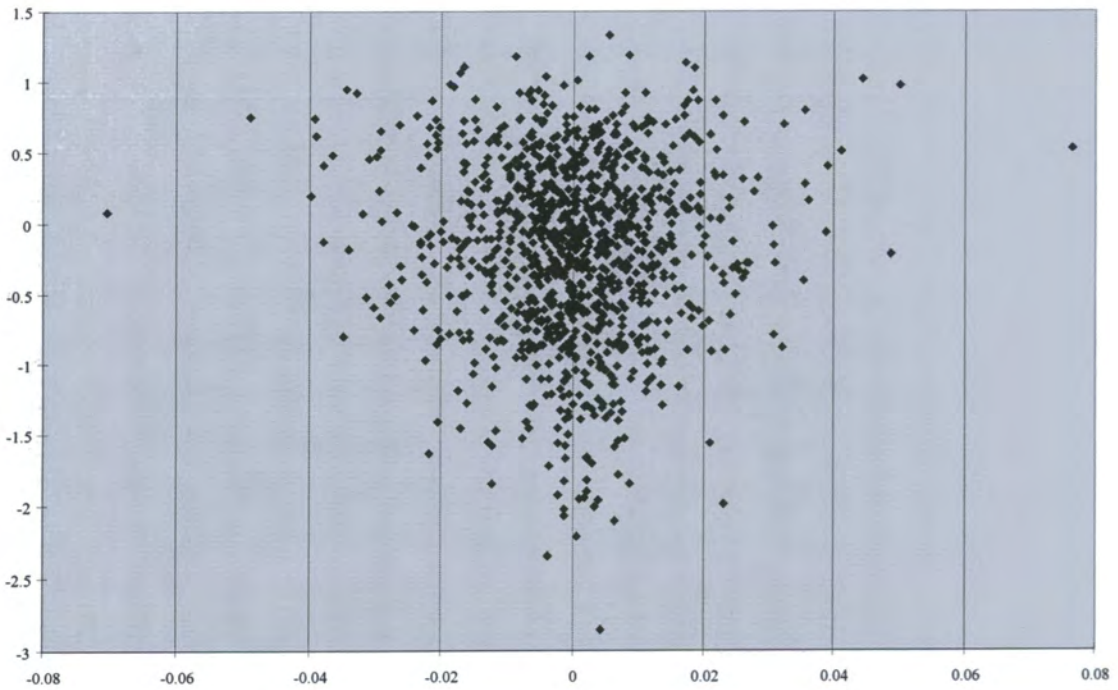
Figure 3.4: Plot of Detrended Log Volume Against Price Returns for the Cocoa
Futures Contract (1992-1996)
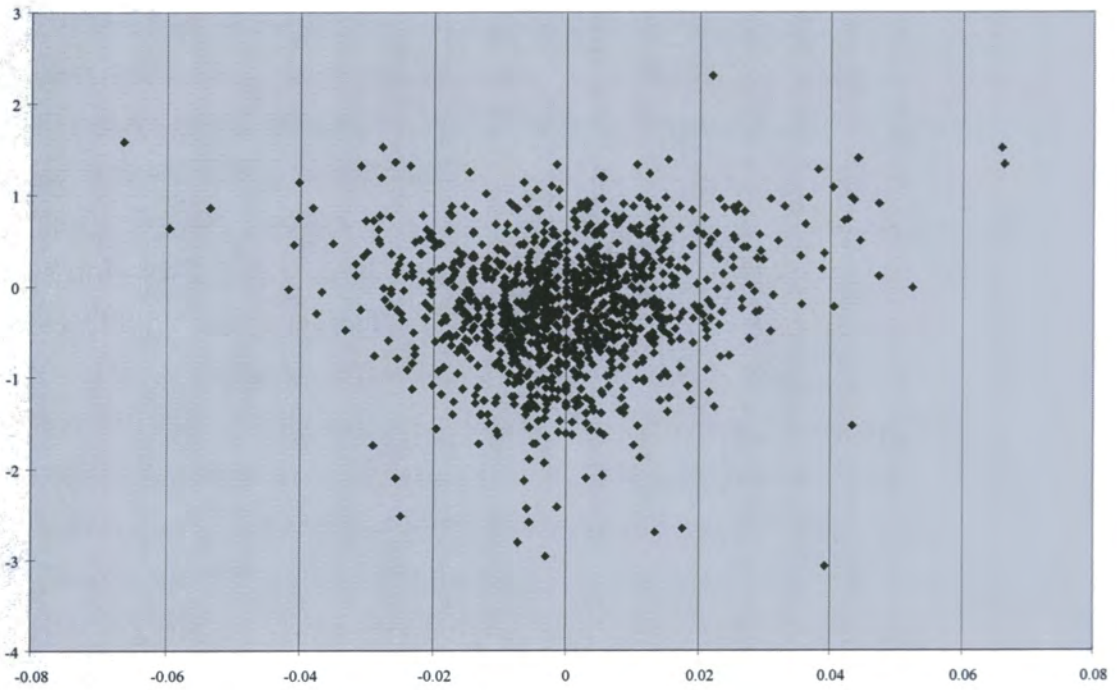
Figure 3.5: Plot of Detrended Log Volume Against Price Returns for the Tin
Futures Contract (1992-1996)
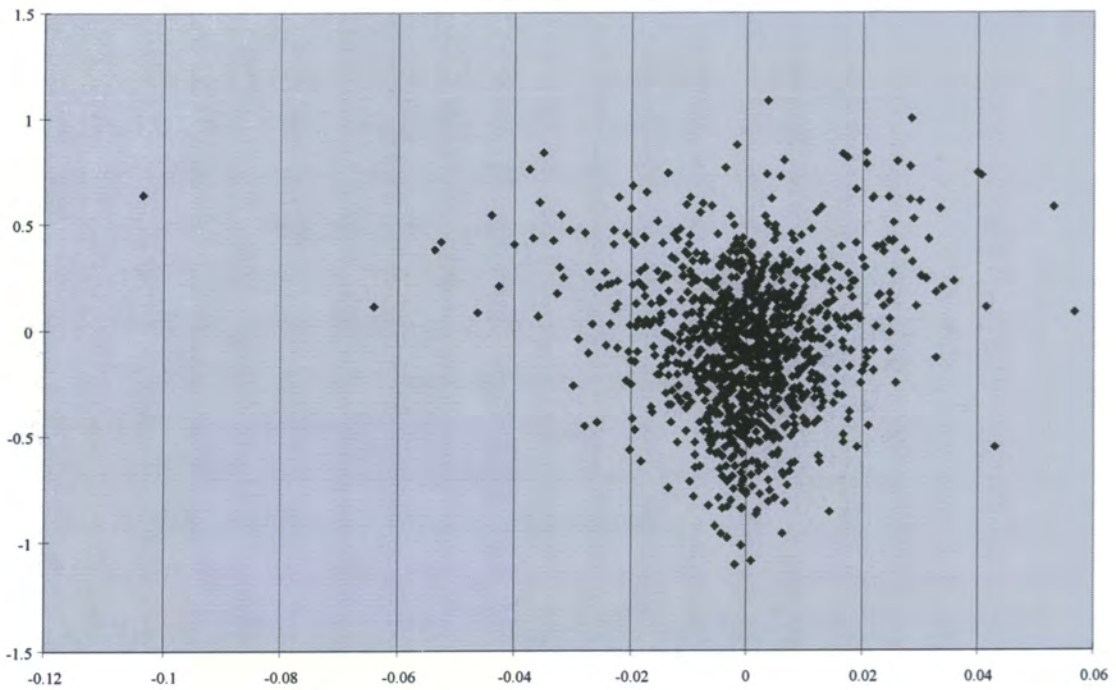
3.6 CONCLUSION

The exploitation of the Blume et al (1994) methodology in this chapter has added to our understanding of the role of volume in derivative markets. The aim was to investigate the role of volume in determining the precision and dispersion of information. It is clear, however, from this analysis that modelling information precision is not easy using 'real' data. In terms of information dispersion this study was able to make an important discovery. The results suggest that for all of the futures contracts considered information is widely dispersed. This is contrary to the popular view of derivative markets that they are dominated by uninformed feedback traders whose destabilising actions can result in the breakdown of asset markets. This study also provides evidence of a symmetric response by investors to positive and negative price movements. A useful development would be to define a suitable measure of statistical inference that could be used with these scatter plots to determine the strength of any relationship.

The continuing theme in these first two empirical chapters has been the role of information in defining the relationship between volume and price changes. Although this chapter marks an improvement from chapter 2 in that a specific theory is being tested, whether or not it holds is really in the eye of the observer. If we are really going to understand the volume-volatility relation we need to test the theoretical models directly in a robust statistical environment. This task is undertaken in the next chapter.

CHAPTER FOUR: MODELLING THE RELATIONSHIP BETWEEN THE VOLUME OF TRADE AND PRICE VOLATILITY IN FUTURES MARKETS: A DIRECT TEST OF THE MIXTURE OF DISTRIBUTIONS HYPOTHESIS

4.1 INTRODUCTION

Chapters 2 and 3 considered a number of different aspects of the impact of volume in futures markets. Among the approaches used were various tests aimed at identifying a relationship between the volume of trade and price variability. It was not possible, however, to explicitly say why this relation occurs. A large body of work has tried to address this issue, and the common theme is the idea that the flow of information into a market can somehow influence the strength of the relationship between these two variables.

The two most commonly quoted models, on which much of the empirical work is based, are Clark's (1973) MDH, and Copeland's (1976) SIM. The first part of this chapter will consider in detail the development of the MDH and the SIM and it will be argued that, of the two, the MDH is the more dominant, and therefore becomes the model of choice. It will also consider more recently developed models of the volume-volatility process.

The second section of this chapter looks in detail at the various empirical studies of these models. As is made clear, however, although the volume of work in this field is extensive, very few papers actually test the theoretical models directly. This was also apparent from the work carried out in chapter 2. Much of the support for the MDH and the SIM is based on, admittedly quite convincing, anecdotal evidence. The exceptions to this are the studies of Richardson and Smith (1994), Lamoureux and Lastrapes (1994) and Anderson (1996). The difficulty that researchers face is that while the theoretical models are based on the impact of information flows, this is a very difficult variable to measure. This study adopts the approach used by Richardson and Smith (1994) and Anderson (1996) to carry out a direct test of the MDH. Their

work is based on the premise that the MDH implies certain characteristics of the data. It is possible to construct a series of moment based conditions that describe these characteristics and test whether or not the conditions hold for real price volatility and volume series.

In this study GMM has been chosen to test the hypothesis that the MDH holds. This methodology is described in detail in section 4.4, including why it is the most appropriate in this case. In addition, two techniques are discussed that help in the construction of the return and volume series from the raw data. Previous studies that use futures data pay varying degrees of attention to the implications of the methods they use to put their series together. Chapter 2 illustrated the differences that can occur between trended and de-trended data. However, although it is very important to account for trends already present in the data, it is equally important to avoid introducing others simply because a series has been constructed in a particular way. The aim of this study is to concentrate on information based activity within a market. The impact of seasonality due to, for example, contract expiration, and the growth of the market, should, therefore, be minimised. The approaches used here achieve this objective.

The empirical section of this chapter, section 4.5, presents the test of the MDH. The characteristics of the data are described at some length before testing whether or not the MDH actually 'fits'. In addition, the results of this test allow an analysis of the information process within a futures market and thus provide important insights into the role of information in futures markets. What is shown is that, contrary to the studies of Richardson and Smith (1994) and Lamoureux and Lastrapes (1994), the MDH does hold for the futures markets examined here. Support is therefore provided for much of the empirical work in this field without having to rely on anecdotal evidence to support the theory. Section 4.6 concludes.

4.2 MODELLING PRICE CHANGES AND VOLUME

Chapter 2 carried out a preliminary investigation into the relationship between volume and the volatility of prices in derivative markets. Section 2.2 looked at the various models in this field including those concerned with the determinants of volume, and

those dealing explicitly with the link between volume and volatility. The discussion of the latter group of models was necessarily brief, and intended only to give a flavour of the theoretical work. In this section, the two most commonly exploited models, the MDH and the SIM, will be discussed in detail.

4.2.1 THE MIXTURE OF DISTRIBUTIONS HYPOTHESIS

Clark's (1973) seminal paper is motivated by a desire to model the characteristics of financial data. Much of our understanding of this data is based on the application of the Central Limit Theorem that allows us to assume that price returns are normally distributed. A closer examination of price changes, supported by the second chapter, reveals that they are not normally distributed. Rather, in comparison to a normal distribution, there are too many small and too many large observations giving the impression of fatter tails than would be normal. This distribution is described as leptokurtic.

Clark argues that the distribution of price changes is subordinate to a normal distribution. The price series evolves at different rates during identical intervals of time. The Central Limit Theorem cannot be applied because the number of individual effects added together to give the price change during a day is variable and random. Clark argues that the movement of prices that occur across different trading days is caused by the variation in the flow of information that is available to investors. More specifically, the more new information that is available, the greater the volume of trade, and the faster the evolution of the price process.

The important concept here is the idea of subordinated processes. Feller's (1971) description of this concept, as adopted by Clark, is based on the premise that we can index a discrete stochastic process by a discrete variable, which when using time series data we assume to be equal to the frequency of the data. Thus $X_0, X_1, ..., X_{t+1}$ is our process indexed by, e.g. minutes, days, etc., and X_t is the realisation of the stochastic process at time t. Feller argues that rather than using time to index the process we could index it using numbers that themselves are the realisation of a stochastic process. Our process is, therefore, represented by $X_{t1}, X_{t2}, X_{t3}, ...,$ where

$t1 \leq t2 \leq t3...$ If $T(t)$ represents the underlying stochastic process which forms $X(T(t))$, it is said to be subordinated to $X(t)$ and is defined as the directing process. The distribution of $\Delta X(T(t))$, where $\Delta X_t = X_t - X_{t-1}$, is then said to be subordinate to the distribution of $\Delta X(t)$.

In the scenario above $\Delta X(t)$ represents the evolution of the price process, while $T(t)$ is a clock measuring the speed of evolution. $X(T(t))$ represents the price process itself. In this way price changes and the rate of flow of information have been reconciled within one model. Clark argues that an obvious measure of this speed of evolution is trading volume and tests this by investigating the relationship between trading volume and price change variance for a sample of daily data on cotton futures for the period 1945 to 1958. Clark shows that grouping the price variance data into volume classes can account for a large amount of the excess kurtosis present in the data. He also shows that rather than a linear relationship between the two variables, there is evidence of a curvilinear relationship. In comparison with a linear model the formulation $\sigma^2 = k V^\beta$ is superior. This curvilinear relationship makes intuitive sense. When new information flows into a market, if that information can be interpreted in a number of different ways, then large price changes will be coincident with large trading volumes. If, on the other hand, traders are in agreement about the impact of a piece of news, the price change may result in low volume. Clark goes further to determine the distribution of the price change series by looking at the distribution of the directing process. He suggests that the distribution is lognormal-normal. That is to say that the directing process, measured by volume, is lognormally distributed, and price changes are normally distributed when adjusted for operational time, i.e. volume. Clark argues that these results provide strong evidence in favour of a subordinated stochastic process model.

Epps and Epps (1976) take a slightly different approach to Clark and derive the relationship between price changes and volume from first principles, rather than testing a model that appears to fit the data. They derive demand equations for two types of trader; those that sell stock following the arrival of new information and those that buy stock. In this way the price variability-volume relationship arises because the

volume of trading is positively related to the extent to which traders disagree when they revise their reservation prices. The model they derive, $\log \sigma^2 = \alpha + \beta \log V^2 + \log u$, is very similar to that of Clark. The tests using OLS and maximum likelihood both support the hypothesis that if the theory holds then β should be significantly positive.

While the complementary approaches of Clark (1973) and Epps and Epps (1976) provide an insight into the volume-volatility relation in financial markets they really only represent the first stages of the development of the MDH. Tauchen and Pitts (1983) extend the theory further to derive the distribution of the price change and trading volume over any interval of time within the trading day. The description of their model, given below, follows their paper closely and exploits the same notation.

Tauchen and Pitts (1983) begin by setting up their trading scenario in a futures market where there is a fixed number of, J, individual investors in a single contract. These investors are assumed to act on each new piece of information as it arrives at the market until equilibrium is reached. In this way information moves the market from one equilibrium to another. If P_{ij}^* represents the jth trader's reservation price and P_i is the current market price, the investor's desired position, Q_{ij}, at the time of the ith within-day equilibrium can be represented by:

$$Q_{ij} = \alpha[P_{ij}^* - P_i] \tag{4.1}$$

where j=1,2,...,J, and $\alpha > 0$ is constant. If the reservation price exceeds the current price, i.e. Q_{ij} is positive, the investor will aim to hold a long position in the contract. If the reservation price is less than the current price, i.e. Q_{ij} is negative, the investor will aim to hold a short position in the contract. Reservation prices will differ across investors because of different expectations about the future and from different risk transfer requirements. Equilibrium in this market will, therefore, occur where the sum of the individual investor equilibria is equal to zero, i.e. $\sum_{j=1}^{J} Q_{ij} = 0$. Thus, the average of the reservation prices, given by:

$$P_i = \frac{1}{J} \sum_{j=1}^{J} P_{ij}^* \tag{4.2}$$

clears the market.

Tauchen and Pitts (1983) then model the process that moves the market from one equilibrium to another following the arrival of a piece of information. They show that the average changes in investors' reservation prices can be measured by ΔP_i where,

$$\Delta P_i = \frac{1}{J} \sum_{j=1}^{J} \Delta P_{ij}^* \tag{4.3}$$

and $\Delta P_{ij} = P_{ij}^* - P_{i-1,j}^*$ is the increment to the jth trader's reservation price. The volume induced by the information arrival is equal to half the sum of the absolute values of the changes in the traders' positions;

$$V_i = \frac{1}{2} \sum_{j=1}^{J} \left| Q_{ij} - Q_{i-1,j} \right| \tag{4.4}$$

Using the equation above this can be written in terms of price changes as:

$$V_i = \frac{\alpha}{2} \sum_{j=1}^{J} \left| \Delta P_{ij}^* - \Delta P_i \right| \tag{4.5}$$

To determine the distributional properties of price and volume Tauchen and Pitts assume a variance-components model;

$$\Delta P_{ij}^* = \phi_i + \psi_{ij} \tag{4.6}$$

$$E[\phi_i] = E[\psi_{ij}] = 0 \tag{4.7}$$

$$var[\phi_i] \equiv \sigma_\phi^2 \tag{4.8}$$

$$var[\psi_{ij}] \equiv \sigma_\psi^2 \tag{4.9}$$

where the ϕ's and the ψ's are assumed to be mutually independent across traders and through time. ϕ_i is the part of the price change that is common to all traders. ψ_{ij} reflects the component that is specific to the jth trader. If the common component is large relative to the specific component this reflects agreement among traders about the interpretation of new information. The converse is true for relatively large realisations of the specific component. The assumption of mutual independence allows Tauchen and Pitts to assume that there are no delays in the receipt of new information.

The changes in price and volume due to the arrival of news can now be written as:

$$\Delta P_i = \phi_i + \overline{\psi}_i \tag{4.10}$$

$$\overline{\psi}_i \equiv \frac{1}{J}\sum_{j=1}^{J}\psi_{ij} \tag{4.11}$$

$$V_i = \frac{\alpha}{2}\sum_{j=1}^{J}\left|\psi_{ij} - \overline{\psi}_i\right| \tag{4.12}$$

Note that the common component ϕ_i is not a determining factor of trading volume. This is in line with the phenomenon of price changes but little or no trading volume. Tauchen and Pitts then assume that the variance components ϕ_i and ψ_{ij} are normally distributed. This allows the following results regarding the joint distribution of the price change and the trading volume:

i) The price change ΔP_i is normally distributed.

ii) For large J the volume V_i is approximately normally distributed.

iii) ΔP_i and V_i are stochastically independent.

iv) Their first two moments are

$$\mu_1 \equiv E[\Delta P_i] = 0 \tag{4.13}$$

$$\sigma_1^2 \equiv Var[\Delta P_i] = \sigma_\phi^2 + \frac{\sigma_\psi^2}{J} \tag{4.14}$$

$$\mu_2 \equiv E[V_i] = \left(\frac{\alpha}{2}\right)\sigma_\psi\sqrt{\frac{2}{\pi}}\left(\sqrt{\frac{J-1}{J}}\right)J \tag{4.15}$$

$$\sigma_2^2 \equiv Var[V_i] = \left(\frac{\alpha}{2}\right)^2\sigma_\psi^2\left(1 - \frac{2}{\pi}\right)J + \frac{\sigma_\psi^2}{J} \tag{4.16}$$

Tauchen and Pitts (1983) argue that the important thing to notice is that the moments of price and volume are linked by their common dependence on the specific term ψ_{ij}. More specifically, the variance of the change in price and the expected volume are both increasing functions of its variance.

The next stage of the model pulls together the elements discussed above but uses the intra-day analysis to say something about the daily joint probability distribution of price change and volume. I, the number of daily equilibria, is assumed to be random because the number of new pieces of information arriving to the market each day varies significantly.

Summing the within-day price changes and trading volumes gives the daily values,

$$\Delta P = \sum_{i=1}^{I} P_i \qquad (4.17)$$

$$\Delta P_i \sim N(0, \sigma_1^2)$$

$$V = \sum_{i=1}^{I} V_i \qquad (4.18)$$

$$V_i \sim N(\mu_2, \sigma_2^2)$$

Thus, both the daily price change and trading volumes are mixtures of independent normals with the same mixing variable, I. Conditional on I the daily price change ΔP is $N(0, \sigma_1^2 I)$ and the daily volume is $N(\mu_2 I, \sigma_2^2 I)$. This allows the model to be written as:

$$\Delta P = \sigma_1 \sqrt{I} Z_1 \qquad (4.19)$$

$$V = \mu_2 I + \sigma_2 \sqrt{I} Z_2 \qquad (4.20)$$

where Z_1 and Z_2 are $N(0,1)$ variables and Z_1, Z_2 and I are mutually independent. Tauchen and Pitts use these two expressions to show the existence of the price variability-volume relationship:

$$\begin{aligned} Cov(\Delta P^2, V) &= E[\Delta P^2 V] - E[\Delta P^2]E[V] \\ &= \sigma_1^2 \mu_2 E[I^2] - \sigma_1^2 \mu_2 (E[I])^2 \qquad (4.21) \\ &= \sigma_1^2 \mu_2 Var[I] > 0 \end{aligned}$$

It is clear that the mixing variable, I, is crucial in the relation between the two variables. If there is no variation in this mixing variable the relationship vanishes. This model, therefore, makes explicit the work of the earlier modellers, in particular Clark (1973). It also makes clear, from the volume specification written in the form above, that trading volume is an imperfect proxy for the mixing variable. This may help to explain the failure of volume to capture all of the GARCH effects in returns in the analysis carried out in chapter 2. For volume to act as a perfect proxy σ_2 would have to be equal to zero. Tauchen and Pitts (1983) argue that there is no reason why this restriction should hold.

The Tauchen and Pitts (1983) specification of the MDH has become recognised as the standard model of the relationship between volume and volatility. However, recent

empirical work[1] has cast doubt on its validity. The specification of the Anderson (1996) model is an attempt to improve upon the work of Tauchen and Pitts (1983). In particular, the theory is much more focused on the daily frequency and reconsiders the distributional assumptions attached to the arrival of investors in a market.

The structure of the Anderson (1996) model is based upon the price discovery model of Glosten and Milgrom (1985). This discussion follows closely the description given by Anderson in his paper and exploits the same notation.

It is assumed that there is a single market for an asset with a random liquidation value of V some time in the future, in which three groups of investors transact; a specialist (or market-maker), and informed and uninformed traders. The specialist offers a bid and an ask price and investors decide whether or not to act on them. Informed traders act on private information that moves the market away from equilibrium until prices fully reveal all information and equilibrium is restored. Anderson models these movements from one temporary equilibrium to another by looking at the cumulative price and volume movements that occur[2].

Investors obtain information that is either publicly available or specific to themselves, or that can be interpreted from transaction prices. C_τ represents the common information set at time τ. Φ_τ represents the investors' information set which includes common plus any private information. The specialist determines the value of the asset at time τ as the expected value of the asset based on his current information set, S_τ, i.e. $P_\tau = E\left[V|S_\tau\right]$. This is not the quoted price from the specialist since information is gained from the actions of traders who come to the market.

---

[1] See Hiemstra and Jones (1994), Lamoureux and Lastrapes (1994), and Richardson and Smith (1994) in section 4.3.

[2] The inclusion of a sequential trading process within Anderson's (1996) mixtures model is an important factor in our decision to use the MDH rather than the SIM in this investigation of the volume-volatility relation.

Let A_τ denote the event that a trader purchases the asset at the ask price, and B_τ the event that an agent sells the asset at the bid price. The bid and ask prices then become $P_\tau^a = E[V|S_\tau \cup A_\tau]$ and $P_\tau^b = [V|S_\tau \cup B_\tau]$. The new prices are based on the specialist's new extended information set. These are considered 'fair' prices for the asset and imply that the trade has an expected value of zero to the specialist. The fact that the expected profit on each trade is zero is designed to reflect the competition that occurs between market-makers. The 'fair' price is, as O'Hara (1997), points out, a feature of rational expectations models where trades reveal information. Glosten and Milgrom (1985) argue that prices therefore follow a martingale with respect to the common and the specialist's information sets[3].

Anderson assumes that uninformed investors arrive at the market according to a constant Poisson arrival process with intensity m_0 per day. They either buy or sell one unit of the asset with probability of one half. They differ from the informed traders who make decisions based on their information sets. Information signals received are correlated but not identical since it is the disagreements in interpretation of these signals that generates trading. These differences are resolved as the market approaches equilibrium. So far this is not dissimilar to the approach adopted by Tauchen and Pitts (1983). It is the next stage where the distinction between the two models becomes apparent.

Let the transaction price recorded during the jth temporary equilibrium of day t be $P_{j,t}, j = 1, \ldots, J_{t-1}$. J_t denotes the total number of information arrivals on day t, which is assumed to be random but large. The return over the whole day is represented by:

$$R_t = \sum_{j=1}^{J_t} \ln\left(\frac{P_{j,t}}{P_{j-1,t}}\right) = \sum_{j=1}^{J_t} \eta_{j,t} \qquad (4.22)$$

$$\eta_{j,t} \sim \text{i.i.d.}(0, \sigma_\eta^2)$$

[3] If the sequence of transaction prices is represented by P_τ, S_τ is the specialist's information set and C_τ represents the common information set, we know $P_\tau = E[V|S_\tau]$. Therefore, $E[P_{\tau+1}|S_\tau] = E[E[V|S_{\tau+1}]|S_\tau] = E[V|S_\tau] = P_\tau$, and since S_τ is a subset of C_τ, prices form a martingale with respect to both the common and the specialist's information set.

Anderson assumes that although J_t is large it varies significantly over the sequence of trading days. To accommodate the distinctive feature of the flow of information within the model Anderson introduces the concept of a benchmark day with a fixed large number of arrivals, J. If K_t denotes the intensity of information arrivals relative to the benchmark, then, $J_t = K_t J$. Incorporating this into the return distribution given above, the benchmark day with J arrivals generates a random return with mean zero and variance $\sigma^2 = J\sigma_\eta^2$. Therefore, intra-day return components are represented by $\eta_{j,t} = \sigma\varepsilon_{j,t}/J^{1/2}$, where $\varepsilon_{j,t}$ is iid., with zero mean and unit variance. The return equation can now be rewritten as:

$$R_t = \sigma K_t^{1/2} \frac{1}{(JK_t)^{1/2}} \sum_{j=1}^{JK_t} \varepsilon_{j,t} \qquad (4.23)$$

Thus for large J the conditional distribution of daily returns can be written in the form,

$$R_t | K_t \sim N(0, \sigma^2 K_t) \qquad (4.24)$$

Like Tauchen and Pitts (1983) this reflects a subordinated stochastic process driven by the intensity of information arrivals.

Anderson breaks down daily volume into informed (I), and uninformed, or noise (N) components, i.e., $V_t = IV_t + NV_t$. As already noted above, the noise trading is assumed to be driven by a stochastic process that has a constant arrival intensity of m_0 per day. Therefore, the uninformed component of volume, NV_t, is directed by a time-invariant Poisson process, $Po(m_0)$. The systematic variation in trading volume is due solely to fluctuations in the informed volume.

In the Anderson model although the intensity of the flow of information may be high, this does not mean that informed volume is also high. Anderson argues that this is because the probability that a given informed trader acts on a piece of information is small. This may be for a variety of reasons; the low probability of picking up relevant information, public news may reveal the information before it reaches the trader, other informed traders may reach the market first and reveal the information through their trading preferences, and the specialist may adjust the bid and ask prices against the trader if there is a suspicion of private information in the market.

Anderson argues that under these conditions each informed trader, on average, only makes a few transactions per day. The probability of acting on a piece of news is also affected by the informational content of the news signal. Anderson argues that a number of factors equalise this probability across different types of information arrivals. He provides the following example. Consider a situation where the arrival of a piece of news results in a large price revision. More insiders may be informed and find it profitable to trade. However, as a result of the information revealed through trading, the market maker will adjust the bid and ask prices making profitable opportunities that much harder to find. The opposite occurs for an arrival with less informational content. If the amount of insider trading is less concentrated, the bid and ask prices are less likely to change and the amount of time over which profitable opportunities can be exploited is extended. It is these arguments that Anderson uses to justify the distributional characteristics of conditional volume.

He shows that the limited variation in the probability of trading induced by a single news arrival can be modelled by a Binomial distribution which is approximated by a Poisson distribution in large samples. Let the expected number of trades by an insider be μ on a day with J arrivals. The mean probability that an insider acts on a piece of news is μ/J. Under the arguments above the conditional distribution of the daily informed volume is therefore given by:

$$IV_t | K_t \sim P_o(IK_t\mu) \tag{4.25}$$

This can then be combined with the noise component to define the distribution for overall daily trading volume:

$$V_t | K_t \sim Po(m_0 + IK_t\mu) \tag{4.26}$$

Anderson defines $m_1 = I\mu$ as the factor of proportionality that measures the fluctuation of information induced volume. This helps in the estimation of the model since both I and μ are unobservable. He further argues that since the scale of K_t is largely arbitrary, by setting $\sigma = 1$ in the return equation the scale of m_0 and m_1 is fixed. Therefore the return equation becomes:

$$R_t | K_t \sim N(0, K_t) \tag{4.27}$$

78

The volume equation also needs some adjustment because we have not considered the trends that are often prevalent in trading statistics. The estimation of the trend that is carried out in the empirical section of this chapter is, despite our best efforts, unlikely to capture all of the characteristics of the data. Anderson uses a constant, c, to reflect the proportion of the trend that has been accounted for. The detrended volume series $\hat{V}_t$ is, therefore, equal to c times the theoretical volume in the model. Thus, the volume equation becomes:

$$\hat{V}_t | K_t \sim c . Po(m_0 + m_1 K_t) \qquad (4.28)$$

These two final equations represent the empirical specification of Anderson's model of the MDH.

4.2.2 THE SEQUENTIAL INFORMATION MODEL

An alternative model of the relationship between price change and the volume of trade is provided by the Sequential Information Model. Copeland (1976) develops a model where individuals receive information one at a time and in a random order. The market is initially in equilibrium and all traders possess identical sets of information. The arrival of news brings about an adjustment in each individual's demand curve. A new equilibrium is established once all individuals have received the news and adjusted their demand curves accordingly.

The model is based on the assumption that all traders have homogenous demand curves with identical slopes and intercepts in the initial equilibrium before the new piece of information is generated. The curves are also assumed to shift up, if the trader is optimistic about the news, or down, if the trader is pessimistic, by an equal amount, δ. Uninformed traders do not infer the content of the information from the actions of others and short sales are prohibited.

The market of N traders is made up of k optimists, r pessimists, and N-k-r uninformed investors. The values of k and r are dependent on the order in which investors become informed. The short sales restriction means that the volume generated by a pessimist is generally less than the volume generated by an optimist. The implication is that

given the price change and trading volume, when the next trader becomes informed depends upon both the previous pattern of who has been informed and whether the next trader is an optimist or a pessimist. Thus, the total volume after all traders become informed depends on the path by which the final equilibrium is reached. Copeland develops expressions for the changes in price with and without the short sales restriction and a probability model for the expected number of trades. The simulation analysis of this model produces some rather interesting results.

One of the most curious results is that the minimum volume occurs where the disagreement among traders is relatively large, while the maximum volume occurs where there is complete unanimity of opinion. Copeland puts this down to the short sales constraint. The simulation also reveals that maximum price changes coincide with maximum volume and that price changes and volume have the same minimum. Copeland argues that his model therefore predicts a positive relation between the absolute value of price changes and volume.

Copeland's model has been extended further by Jennings et al. (1981). Instead of a short sales constraint they impose a margin requirement upon market participants. An investor who sells short is not entitled to the proceeds from the sale and must put up the margin requirement with the broker until this short position is covered. Both long and short investors are liable to this transaction cost but for the latter the penalty is assumed to be greater.

The model, initially excluding the margin requirement, indicates that the volume and price change caused by a single investor depend only on the total number of traders, and are independent of the numbers of optimists and pessimists. It also argues that the largest change in price occurs when all traders agree on the meaning of a piece of information. This latter result is still supported when the margin requirement is imposed. Jennings et al. have problems, however, in supporting Copeland's (1976) assertion that the relation between absolute price changes and volume is strictly positive. Their analysis suggests that the correlation between the two variables depends on the margin requirement, the riskless rate of interest and the mix of

optimists and pessimists in the market. The greater the proportion of optimists the stronger the positive relation becomes.

Jennings and Barry (1983) go one stage further and allow informed traders to take speculative positions within the market. The basic premise is to examine the investment decisions of those who are first to receive new information within a market. If a trader is aware that he or she holds an informational advantage then Jennings and Barry demonstrate that they will adjust their expectations of future trading opportunities. This may affect the volume of trade, and the variability of prices. Their model suggests that price adjustment occurs more rapidly in those markets where speculation is present. They also find evidence of a positive contemporaneous correlation between price change and volume, due to the association of both variables with the amount of portfolio revision desired by a market participant receiving new information. Their model predicts that the first informed investor would cause a relatively large price change and volume reaction, while subsequent traders would have an increasingly reduced impact.

4.2.3 ALTERNATIVE APPROACHES

Admati and Pfleiderer (1988) develop a model of intra-day trading designed to answer three important questions. Why does trading tend to be concentrated in particular time periods during the trading day? Why are returns (or price changes) more variable in some periods and less variable in others? And why do periods of higher trading volume also tend to be the periods of highest return volatility? These questions are prompted by observations in the intra-day trading patterns of Exxon shares in 1981 that show trading volume concentrated at the beginning and ends of the trading day. The variances of returns and price changes appear to follow a similar U-shaped pattern. They argue that the patterns observed in the data can be explained by the optimising decisions of traders.

Their model considers the interaction between informed and uninformed, or liquidity, traders. The liquidity traders are divided into two groups. The nondiscretionary liquidity traders must trade a particular number of shares at a particular time. The discretionary liquidity traders can be strategic in choosing when to execute their trades

within a given period of time. This latter group are assumed to act to minimise the expected cost of their transactions.

The behavioural characteristics of each group of investors produce a trade generating trade scenario. Liquidity traders want to trade where their actions are not going to change prices. Informed investors need to exploit the information that they hold and therefore look to trade when the market is at its 'thickest'. The combination of these two effects brings these two groups together. There is then an incentive for other investors to become informed. Admati and Pfleiderer argue that discretionary liquidity traders accrue welfare benefits by entering the market in such a situation due to the competition between informed investors. They also argue that if the information in the market is diverse then the variability of prices is likely to increase during this concentrated period of market activity.

Foster and Viswanathan (1995) build a model that has elements of both the Anderson (1996) and Admati and Pfleiderer (1988) approaches. In their model there is a market-maker, traders who must incur a cost to obtain a piece of information, and liquidity traders, who are all trading in a single asset whose liquidation value is changing each period. The premise is to combine aspects of speculative trading and stochastic volatility models. They make assumptions regarding the distribution of the asset, the orders of the liquidity traders, and the private information signal underpinned by an unobservable latent process. It is essentially a mixture of normals specification. The difference between this and the approaches discussed above is the introduction of the speculation element. The model makes a number of propositions including; a positive correlation between volume and the variance of price changes, and the conditional heteroscedasticity of price changes and volume. Unfortunately, although an interesting approach, the speculation model appears to be unable to support its claims when tested using real data.

4.2.4 THE BEST MODEL?

The attraction of the Mixture of Distributions Hypothesis is its intuitive appeal. The characteristics of return distributions have been readily observed, and the links between price volatility and the volume of trade appear to be accepted as the norm. There have been, however, very few attempts to explain this link. The MDH offers a plausible economic model of this process that describes the market in some detail.

The Sequential Information Model is less appealing. It provides an insight into the microstructure of financial markets but it does not explain the distributional properties of asset returns. Copeland's (1976) model has a number of characteristics that make it much less attractive than the MDH. As Karpoff (1987) points out, one of these is the idea that traders do not learn from the actions of others. His model also implies that the volume of trade within a market is greatest when all traders agree on the meaning of information. In practice, trading usually only occurs where there is asymmetric information. If all traders agree on the interpretation of a piece of information this is likely to reduce, rather than increase, the number of individuals willing to trade. Karpoff (1987) also criticises the assumption that disagreement among traders can be represented by an identical response, although in opposite directions.

Empirical studies (e.g. Jain and Joh (1988)) have used causality tests to distinguish between the two models; a simultaneous relationship implies the MDH, a sequential relationship implies the SIM. Karpoff (1987) argues that in fact the MDH subsumes the SIM. He points out that while the model of Epps and Epps (1976) requires the simultaneous receipt of information by investors, the model of Tauchen and Pitts (1983) is less restrictive. Their mixtures model assumes a process of successive market equilibria. This may be the result of a single piece of news being slowly disseminated by market agents, or the result of news being simultaneously received by all traders. The Anderson (1996) model described above illustrates quite clearly how the SIM can be reconciled within a mixtures based approach.

The models discussed in section 4.2.3, although interesting, do not have the same appeal from the point of view of an empirical test of the volume-volatility relationship. The Admati and Pfleiderer (1988) approach does not allow us to

examine the trading process in the detail offered by Anderson's (1996) specification of the MDH, particularly with regard to explaining the distribution of price changes. The more complex model of Foster and Viswanathan (1995) is let down by its apparent inability to withstand empirical scrutiny. These weaknesses, the criticisms that can be made of the SIM and the fact that it can be subsumed by the mixtures model make the MDH our preferred model.

4.3 LITERATURE REVIEW

The motivation for much of the work on the Mixture of Distributions Hypothesis has been the phenomenon associated with the distributions of asset prices. Taylor (1985) has carried out a comprehensive investigation into the behaviour of futures prices over time. The study concentrates on daily data for eight agricultural and financial commodities spanning a period from 1961 to 1981. He finds that for all of the futures contracts the distributions display excess kurtosis which rules out a normal distribution. He also finds evidence that the variance of each series changes over time.

Wood et al. (1985) investigate the behaviour of returns for transaction data from a large sample of NYSE stocks. Their results suggest that the return-generating process varies systematically across the trading day and overnight. They look at the distribution characteristics for the opening of the trading day, the end of the day, and the trading period in-between. Returns at both the beginning and end of the day periods have distributional characteristics consistent with those that the MDH is designed to explain. They also find that when the first and last thirty minutes of each trading day are excluded, market returns are normally distributed and any autocorrelation effects are substantially reduced. They argue, therefore, that the phenomena associated with returns series aggregated over longer periods can be attributed to the price effects that give intra-day return series their U-shaped pattern.

The seminal work in this field is that of Clark (1973). However, although Clark was one of the first to suggest that the distributional properties of returns could be linked to the concept of subordinated processes, his empirical work is less convincing. The first stage of his study involves taking two samples of 1000 daily observations on

cotton futures, covering the periods 1947 to 1950 and 1951 to 1955. The sample is split into twenty groups measuring price change variance, by increasing volume. He notes that while there is evidence of leptokurtosis when the sample is taken as a whole, this value is much reduced when price changes with similar volumes are considered. The analysis also suggests a curvilinear relationship between price variance and trading volume. Clark investigates this further by hypothesising two models; $\sigma^2 = A e^{\alpha V}$ and $\sigma^2 = B V^{\beta}$, where σ^2 represents the price variance, and V represents volume. The results of trying to fit these models to the data support the second specification as the better of the two. Represented in this way Clark argues that trading volume acts as an instrument that measures the speed of the evolution of the price process. It allows the distinction to be made between normal time and operational time.

Clark then uses these equations to model the distribution of price changes. Although the results are not perfect, the kurtosis values and the Kolmogorov-Smirnov (KS) tests against normality do suggest that price changes are normally distributed when adjusted for volume. Clark tests the distribution of the underlying process by looking at the distribution of volume. The results indicate that volume is lognormally distributed as opposed to normally distributed. Therefore, much of the analysis supports the idea of a subordinated process model, where volume is used to proxy the directing process.

Although the subordinated process argument appears to fit, it has been suggested that the data could be modelled by the class of Paretian stable distributions. This family of distributions have high unbounded kurtosis values and infinite variance. At a glance they would appear to describe the data quite well. Clark tests these two models against each other and using Bayesian analysis of the posterior distributions and, comparing the KS statistics against the maximum likelihood distributions, rejects the stable distributions hypothesis.

The approach of Epps and Epps (1976), which is very similar to that of Clark (1973) starts with OLS estimation of their model; $\log \sigma^2 = \alpha + \beta \log V^2 + \log u$. They use

daily data on prices and volumes for twenty stocks traded on the NYSE for the month of January 1971. Under the hypothesis that the theory holds, β should be significantly positive. The use of OLS in this situation is, however, unsuitable particularly given the heteroscedasticity in the disturbances. Epps and Epps therefore repeat the estimation using maximum likelihood. Although the distribution of the disturbances must be known, Epps and Epps carry out the estimation assuming the distribution of the disturbances is normal and then assess the assumption by investigating the disturbances directly. The results suggest that the variance of returns is a function of volume, thus supporting the work of Clark (1973). They do, however, offer a number of caveats. They believe that Clark's model is mis-specified, which may account for the fact that while the distribution of returns is far less leptokurtic, once the returns have been adjusted for volume, it is still not strictly normal. The results of the maximum likelihood estimation can also be questioned, since the distribution of the disturbance terms is not normal. In support of Clark, however, they argue that there is no evidence that price changes can be modelled by a stable distribution.

Morgan (1976) also provides evidence that the variance of returns on common stocks is not constant through time, but is related to the volume of shares traded. He tests two hypotheses; that variance depends on volume and that returns are conditionally normal. Morgan assumes a normal distribution for asset returns, y_t, with constant mean θ; $y_t \sim N(\theta, \sigma^2 \phi_t(\lambda))$. This assumes that the variance of returns is proportional to some function $\phi_t(\lambda)$, which Morgan assumes is increasing in volume. The data on prices and volume relate to a sample of stocks traded on the NYSE. A total of seventeen stocks for the period 1962 to 1965 are chosen for analysis on a daily level and forty-four stocks for the period 1926 to 1968 are chosen at monthly frequency. The first hypothesis is tested by determining whether λ has a value of zero by looking at its posterior distribution. The second hypothesis is investigated by transforming the data into the form; $(y_t - \phi) f_t(\lambda) \sim N(0, \sigma^2)$, and measuring the kurtosis in various ways. The results suggest that λ is not zero and that kurtosis falls once the data has been transformed by volume. Morgan argues that volume is therefore important in determining the distribution of returns.

Westerfield (1977) also looks at evidence to support the subordinated process model. He uses dividend adjusted return relatives for 315 common stocks listed on the NYSE from the period January 1968 to September 1969, and the number of shares traded daily for each security. Analysis of the measures of the sample moments reveal that the daily price changes have a leptokurtic distribution. Westerfield follows a similar process to Clark (1973) and ranks estimates of price change variance into volume classes. The two variables appear to have a positive relationship. He then standardises the variance of daily price change in each group by dividing by the securities' total variance of price change. The purpose is to investigate whether the variance of the price change will vary with increments of the directing process as predicted by the MDH. The results indicate that the larger than average price changes (both positive and negative) are associated with relatively high levels of trading over the same calendar time intervals. Westerfield also runs the two linear regressions hypothesised by Clark (1973) to determine the relationship between the conditional variance of price change and volume. He argues that because there is a significant relationship between the two for most securities, trading volume can be used as an instrumental variable in measuring transaction time. To support this further Westerfield shows how the kurtosis values are much reduced when price change is ranked by the volume of trade across the whole sample. Although much of the evidence supports the subordinated process model, Westerfield extends his analysis to compare this theory with that of the Paretian stable model. The comparison of the theoretical standardised probability distribution functions with those observed from the data using the KS and Chi-Square statistics support, somewhat tentatively, the subordinated model.

The comparison of these two different explanations for the distribution of asset returns is also carried out by Upton and Shannon (1979). They also test whether the assumption of lognormality for returns of common stocks holds over a number of frequencies from monthly up to annual periods. The basic data set consists of 235 monthly returns for each of 50 companies randomly selected from the NYSE listing over the period from January 1956 to July 1975. In an attempt to determine whether the characteristics of individual stock returns also hold for groups, two portfolios are created of ten stocks each. The difference between the two is that in the first the

portfolio is balanced to equal proportions every month, while in the second a buy and hold strategy is employed. Analysis using the KS statistic reveals that lognormality only holds over longer time horizons for individual stocks. It is much harder to reject this assumption for the portfolios across any frequency. In addition, evidence of leptokurtosis appears to be less apparent over higher frequencies. Upton and Shannon also use the Studentized Range statistic and analysis of the α-characteristic to compare the competing hypotheses. The Studentized Range statistic which can compare the two characteristic distributions favours the subordinated model. The hypothesis that the α-characteristic is less than two, as predicted by the stable model, also fails to be accepted.

Tauchen and Pitts (1983) test their model, described in section 4.2, using a data set of 876 observations on the daily price change and volume of trading on the 90-day T-bills futures market for the period January 1976 to June 1979. They estimate the parameters in their joint distribution model by maximum likelihood, which allows the conditional expectation of the squared price change to be known given the volume. This avoids many of the problems of the studies above in hypothesising and testing numerous regressions to find the 'correct' functional form. They find that the model predicts the observed data very closely but warn that any relationship may be obscured if trends in the volume data are not filtered out.

Harris (1986, 1987) takes the model of Tauchen and Pitts (1983) and tests the implications for a cross-sectional sample of securities and transactions data respectively. Harris argues that if the observations are in accordance with the theoretical implications then the MDH holds. Harris identifies six testable implications of the Tauchen and Pitts specification of the MDH:

1. The marginal distribution of daily returns is kurtotic relative to the normal.
2. The marginal distribution of daily volume is skewed to the right.
3. The squared return is correlated with the daily volume of trade.
4. Interval measures of price variance change through time if the probability distribution of the directing process changes through time.
5. The marginal distribution of returns is skewed.

88

6. Volume and returns are slightly correlated.

These predictions are dependent upon variation in the directing variable. In addition, Harris argues, if it is assumed that the distribution of the directing variable is not the same across all securities and the coefficient of variation of this distribution varies across securities, then sample measures of return skewness and kurtosis, of volume skewness, of return-with-volume correlation, and of squared-return-with-volume correlation will be positively correlated across securities. He also argues that if some directing distributions are more stationary than others, sample measures of price and volume heteroscedasticity will be positively correlated across securities.

The sample consists of prices and volumes for 479 securities traded on the NYSE between January 1976 and December 1977. This period is chosen as one in which there is relatively little growth in the volume of trade. The results indicate that the predictions of the MDH are supported. Harris argues that since the directing variable is often assumed to be the rate of information arrival, then these rates of arrival differ across securities.

In addition to the six predictions of the mixtures hypothesis outlined above, Harris identifies a number of additional predictions for transactions data under the assumption that transactions occur at a uniform rate in event time. Harris also assumes that the number of transactions in the market are proportional to the number of information events.

1. The number of transactions is correlated with the price change, the square of the price change and volume.

2. The correlation coefficients should be largest for volume, second largest for the square of the price change, and smallest for the price change itself.

3. Autocorrelation in the time series of the number of transactions should be stronger than that found in any other daily series.

Harris also makes a number of predictions concerning the distribution of daily price changes conditional on the daily number of trades;

4. The adjusted series should exhibit more symmetry and be less leptokurtic.

5. The adjusted series should show reduced levels of heteroscedasticity.

6. The adjusted price change and the squared adjusted price change should not be as autocorrelated as their unadjusted counterparts.

If transactions are assumed to occur at a uniform rate in event time, then the number of information arrivals within different transaction intervals of some fixed length should be constant. This has three additional implications;

7. Price changes and volume measured over a fixed transaction interval should be more normally distributed the longer the transaction interval.

8. Transaction interval price changes and squared price changes should not be correlated with transaction interval volumes.

9. Transaction interval price changes and volumes should not be autocorrelated.

The data set consists of fifty securities traded on the NYSE between December 1981 and January 1983. Price changes and volume were computed over fixed intervals of 1, 10, 50, and 100 transactions and over daily intervals. The results on the whole support the predictions above and Harris concludes that the daily number of transactions may be a good estimate of a time-varying evolution rate. He notes, however, that this is based on indirect evidence since the information evolution rate is not directly observable.

The evidence presented above is very much in support of the MDH and the theory of subordinated processes. There have, however, been some dissenting voices among the early empirical studies.

Harris and Gurel (1986) examine price and volume changes surrounding changes in the composition of the S&P 500. They distinguish between the informational price effects of information-bearing transactions, and events which are unlikely to bring new information to the market. They argue that analysis of the former is difficult, since it requires an empirical model of the information price effect. Harris and Gurel, therefore, concentrate instead on price pressures that they believe are information free. Changes in the composition of the S&P 500 cause demand to change with very little informational basis. A study of their effects on prices and volume may identify price pressures in the absence of new information. Harris and Gurel consider all changes in the S&P 500 list for the period 1973-1983 concentrating primarily on additions to the

list. Close examination of mean volume and mean returns surrounding increases to the list reveal increases in both variables. They argue that since the increases in price are consistently reversed it is unlikely that information causes the initial increase. The implication is that some care must be taken before assuming that all price and volume movements will be principally information driven.

French and Roll (1986) investigate the difference between the volatility of asset prices during exchange trading hours to that during non-trading hours. The evidence suggests that the former exceeds the latter and French and Roll consider three explanations for this phenomenon. The first suggests that volatility is caused by public information which is more likely to arrive during trading hours, the second is that volatility is caused by private information which affects prices when informed investors trade, and the third is that volatility is caused by pricing errors that occur during trading. Their sample covers all common stocks traded on the NYSE and American Stock Exchange (AMEX) between 1963 and 1982. This twenty year period is then broken down into ten two-year subperiods. Return variances are calculated for weekdays, weekends, holidays, and holiday weekends during each subperiod. Examination of these results confirms that trading hours are more volatile than non-trading hours. French and Roll then try to distinguish between the three hypotheses by looking at the effects of exchange holidays and trading breaks due to elections, and the autocorrelation of returns. Their results suggest that, in contrast to Harris and Gurel (1986), despite a small percentage of the price variance being attributable to mispricing, the greatest impact on the market is provided by the flow of information. The difference in variance between the two periods can be attributed to differences in the flow of information between trading and non-trading hours.

Two other empirical methods in this field, that have already been discussed in detail in chapter 2, are causality testing and GARCH modelling. The causal relationship between volume and volatility is examined by Grammatikos and Saunders (1986), Jain and Joh (1988), and Hiemstra and Jones (1994). The modelling of price returns using the GARCH methodology is exploited by Lamoureux and Lastrapes (1990) and Foster (1996). The details of these studies are presented in section 2.3.

In a more recent paper, Lamoureux and Lastrapes (1994) use a signal extraction approach within the mixture framework to extract a time series on the unobservable information flow. This series together with stock data on volume and returns is used to test whether the mixtures model is consistent with GARCH effects. They use the following model:

$$r_t = \sigma_1 Z_{1t} \sqrt{F_t} \tag{4.29}$$

$$V_t = \mu_2 F_t + \sigma_2 Z_{2t} \sqrt{F_t} \tag{4.30}$$

$$F_t = \alpha_0 + \alpha F_{t-1} + \phi_t \tag{4.31}$$

where r_t is the stock return on day t, V_t is the daily trading volume, F_t is a latent mixing variable, Z_1 and Z_2 are mutually and serially independent stochastic processes with zero mean and unit variance, and ϕ_t is a serially independent random variable with zero mean that is restricted to ensure that F is always non-negative. The first two equations correspond to the Tauchen and Pitts (1983) model, while the third is based on the assumption that the information arrival process is serially uncorrelated. The signal extraction process finds a value of F_t that sets the observed values of r^2_t and V_t as close as possible to the conditional means predicted by the model. The returns are then adjusted and tested for serial dependence. The absence of GARCH effects would support the mixtures model.

Lamoureux and Lastrapes use a sample of daily returns and volume data for 10 individual companies for the period from January 1967 to December 1987. The results suggest that accounting for serial dependence in the information arrival process does not eliminate GARCH effects. In contrast to their earlier study, Lamoureux and Lastrapes, therefore, question the ability of the mixtures model to account for the characteristics of return data.

The discussion above illustrates the number and the range of studies that have tried to explain the volume-volatility relation using the Mixture of Distributions Hypothesis. The evidence in its favour, however, is largely based on tests of, or observations that comply with, the model's implications rather than providing direct tests of the MDH.

In contrast, Richardson and Smith (1994) carry out a direct test of the MDH using the GMM methodology. One of the difficulties of testing the model is that the directing process, assumed to be the information flow, is unobservable. Although there is very little guidance given by the theory, distributional assumptions are often made regarding the information flow, which have important implications with regard to empirical work. The model does, however, place restrictions on the unconditional moments of the changes in price and volume and on their cross moments. Because I, the information variable, enters their conditional moments in a similar way, all higher moments are a function of only price and volume and the central moments of I. Therefore the unconditional moments and cross-moments of the unobservable variables (the change in price and volume), will place over-identifying restrictions on the data. Richardson and Smith argue that under weak assumptions the MDH can thus be tested directly. Their data set consists of daily prices and volume for the Dow Jones 30 firms for the sample period from 1982 to 1986. Their results suggest that the MDH is not a good model for explaining variations in the data.

Anderson (1996) also exploits the GMM methodology to carry out a direct test of the standard mixtures model against his own specification (see section 4.2.1). He constructs a continuously compounded return series from daily closing prices of common IBM stock for the period from 1973 to 1991. The corresponding volumes are detrended using a non-parametric kernel regression and a centred moving average. The results suggest that the characteristic phenomena associated with asset returns can be explained by a subordinated process and reveal that the new specification of the model vastly outperforms the standard version of the MDH.

While there has clearly been considerable analysis undertaken of the volume-volatility relationship, it is only recently that direct tests of the MDH have been undertaken. However, very few such studies have been attempted to date. In addition, while the work by Richardson and Smith (1994) and Anderson (1996) provides direct tests of the MDH, very little work has been undertaken either for UK traded assets or for futures markets. If a full understanding of the volume-volatility relationship and the MDH is to be gained, it is clearly important that more empirical work is carried out

for assets traded in countries outside the USA and for derivative assets. Such an analysis is undertaken in this chapter.

As pointed out in chapters 1 and 2, the relationship between volume and volatility is a very important issue. In order to make policy recommendations and to inform our investment decisions a detailed understanding of the trading process is vital. The empirical work carried out in chapter 2 suggested that information plays an important role in defining the relationship between the two variables. However, the exact nature of this role is still unclear.

In Chapter 2 it was shown that it is very important to eliminate trends in volume data. Trends can obscure the underlying relationship between volume and volatility. A number of studies using futures data remove the trading that occurs in the month before expiration to avoid 'unusual' results. This means that a large amount of information that comes into the market during this month is lost. This will undoubtedly affect any empirical investigation where the flow of information is expected to form a pivotal role.

This study aims to add to the existing literature in three important ways.
- the exploitation of a specification of the MDH that allows it to be tested directly using a standard econometric technique. As noted in this section, and in chapter 2, the majority of studies have looked at the data searching for evidence that some of the implications of the MDH hold rather than testing the model directly.
- the use of a specification of the MDH that allows us to investigate the characteristics of the information process and to discriminate between the different components of the volume of trade; informed and uninformed trading.
- the use of the Holmes-Rougier (1997) roll-over adjustment. The use of the roll-over adjustment allows the expiration month to be included when constructing the sample, as well as detrending the data.
- the use of futures market data for the UK. Although the original Clark (1973) study used futures data, the majority of papers have concentrated

on stock data from US markets. This study will look exclusively at UK futures data over a range of commodities and therefore provide a previously unavailable insight into the relationship between volume and volatility.

4.4 METHODOLOGY

Anderson's (1996) modified version of the MDH can be neatly expressed as a series of twelve equations relating to different characteristics of the volume-volatility process. It is quite simple to form orthogonality or moment conditions from these equations and to therefore exploit GMM. This section includes a brief description of the GMM methodology, and discusses its advantages over other techniques.

One of the problems of using financial data is deciding how to construct the price series from a number of contracts that are all being offered at the same time. This study utilises a method that is relatively easy to construct and avoids the introduction of trends that can occur if the prices of contracts are simply spliced together.

This study also considers a method of addressing the problem of roll-over effects in futures volume data. It is important in this study to remove the impact of non-information based trends in trading as far as possible. The Holmes-Rougier (1997) adjustment allows us to do this.

4.4.1 THE GENERALISED METHOD OF MOMENTS

The generalised method of moments is a direct extension of the method of moments and is an ideal technique to use to obtain consistent parameter estimates of a model where efficiency is of secondary importance. Its main advantage over other techniques, for example OLS, is that it is less restrictive in terms of the assumptions that must be made regarding the model under investigation. It is worth at this stage looking at some of the ideas that underpin GMM.

The method of moments works on the principle that in random sampling, a sample statistic will converge in probability to some constant. This constant will be a function of the unknown parameters of the distribution. By equating the functions

with the moments the equations can be solved to provide the parameter estimates. The method of moments, as explained by Barr (1997), essentially sets up a series of orthogonality conditions. More explicitly we can say if u_t and z_t are orthogonal to each other then the expected value of their product is zero:

$$E(u_t z_t) = 0 \qquad (4.32)$$

If z_t is in fact a constant, d, then:

$$E(u_t z_t) = E(u_t d) = E(u_t)d \qquad (4.33)$$

Utilising the method of moments, if the expected value of u_t converges to a constant, μ_u, then we can rewrite our orthogonality condition as:

$$\left(E(u_t) - \mu_u\right)d = 0 \qquad (4.34)$$

This approach can be translated to regression functions. Consider the following simple function of the variables y_t and x_t, the error term ε_t, and the parameters α and β:

$$y_t = F(\alpha, \beta, x_t) + \varepsilon_t \qquad (4.35)$$

From this equation we can construct an orthogonality condition using the error term and the constant as above:

$$E\left(\varepsilon(y_t, x_t, \alpha, \beta,) - \mu_\varepsilon\right)d = 0 \qquad (4.36)$$

If we set the constant equal to one it is clear that our orthogonality condition is also a moment condition. The error term may be orthogonal not only to a constant, but also to a number of variables. These represent the instruments that are crucial in obtaining parameter estimates. Let $\varepsilon(y_t, x_t, \alpha, \beta) \equiv \varepsilon_t(\theta)$ and let $h_t = (h_{1t}, \ldots, h_{nt})$, where h_1 to h_n each represent a different instrument. If $f_t(\theta) = \varepsilon(\theta)h_{1t}, \ldots, \varepsilon(\theta)h_{nt}$, then the orthogonality condition can be written as:

$$E\left[f_t(\theta)\right] = 0 \qquad (4.37)$$

The method of moments requires us to find estimates of θ, ($\hat{\theta}$) such that the condition above is satisfied. In other words the following function must be as close to zero as possible:

$$g_n(\hat{\theta}) = n^{-1} \sum_{t=1}^{n} f_t(\hat{\theta}) \qquad (4.38)$$

or

$$g_n(\hat{\theta}) = n^{-1} H' \varepsilon(\hat{\theta}) \qquad (4.39)$$

96

where H' represents the vector of instrumental variables.

As noted above, under the method of moments, there are exactly the same number of orthogonality conditions as there are parameters to be estimated. The generalised method of moments is designed to deal with situations where this is not the case and estimation is more difficult. The approach to this problem is very similar to OLS, and more particularly, GLS. GMM minimises a quadratic form of the type:

$$Q_n = g_n(\theta)' W_n g_n(\theta) \tag{4.40}$$

where W_n is a weighting matrix equal to the covariance matrix of the orthogonality conditions. This result of Hansen (1982) is outlined more clearly below.

The first-order condition for finding the minimising parameter values is given by:

$$D_n(\hat{\theta})' W_n g_n(\hat{\theta}) = 0 \tag{4.41}$$

where D_n is a matrix of partial derivatives;

$$D_n(\hat{\theta}) = \partial g_n(\hat{\theta}) / \partial \hat{\theta} \tag{4.42}$$

Large sample theory provides us with the following result regarding the asymptotic distribution of our estimated parameters:

$$\hat{\theta} \sim N(\theta, \Sigma) \tag{4.43}$$

where Σ, the asymptotic variance, is given by:

$$\Sigma = (D'_0 W D_0)^{-1} D_0 WSW D_0 (D'_0 W D_0)^{-1} \tag{4.44}$$

S is given by:

$$S = n^{-1} D'_n \Omega D'_n \tag{4.45}$$

where Ω represents the covariance matrix of the error terms. This is a very important result for hypothesis testing of the model under investigation as we will see below.

Hansen (1982) shows that the optimal weighting matrix is where:

$$W = S^{-1} \tag{4.46}$$

The asymptotic variance then simplifies to:

$$\Sigma = (D'_0 S^{-1} D_0)^{-1} \tag{4.47}$$

The final part of this process is to check that the moment restrictions implied by the model under investigation are valid. If the hypothesis of the model that led to the

moment equations is incorrect, at least some of the moment restrictions will be systematically violated. Hansen (1982) uses the quadratic form above to construct a chi-square test statistic with degrees of freedom equal to the number of over-identified orthogonality conditions (i.e. the number of orthogonality conditions minus the number of parameters). Under the null hypothesis that the model is correct we have:

$$nQ \sim \chi^2_{df} \qquad (4.48)$$

The discussion above provides a brief outline of the GMM methodology. The reasons for using it in this study are outlined in more detail below.

4.4.1.1 What are the advantages of using GMM?

As noted above, more conventional estimation techniques, for example Ordinary Least Squares, can only be used under quite a restrictive set of assumptions. GMM forms part of a wider class of models which exploits large sample theory to generate results that hold under conditions much weaker than those of so-called classical regression theories. GMM is used where a consistent estimator of a parameter is required but efficiency is secondary. Crucially it avoids the assumptions relating to the error terms of models that can make estimation so difficult.

In terms of what it can show economically, GMM offers a distinct advantage in allowing the direct testing of a model of the volume-volatility relationship. The Mixture of Distributions Hypothesis models the impact information has on prices and volume. Lamoureux and Lastrapes (1990) show that this hypothesis can be written in terms of ARCH models. Evidence in support of the MDH is also provided by Harris (1987). Richardson and Smith (1994) argue, however, that these results are anecdotal. That is, the type of distribution patterns generated from daily data appear consistent with those from a mixed distribution model. Few direct tests of the MDH have been carried out, partly due to the fact that the flow of information is unobservable. A further complication is the model's implied heteroscedasticity and autocorrelation properties of price changes.

A direct test of the MDH is possible because the model imposes restrictions on the joint moments of price changes and volume as a function of only a few parameters.

This allows the formation of overidentifying restrictions on the data which can be tested using GMM. The characteristics of the distribution of the random flow of information can then be estimated and used to provide details on a number of market microstructure issues. The orthogonality conditions implied by Anderson's (1996) modified version of the MDH are given in the empirical section of this chapter.

4.4.2 THE CONSTRUCTION OF A CONTINUOUS PRICE SERIES FOR A FUTURES CONTRACT

This chapter also makes use of two techniques that will help in our construction of the data series. The first of these is the Rougier (1996) contiguous price index. Rather than splicing the series together as in chapter 2, the Rougier index reflects a time weighted mean of the prices of the nearest and next nearest contracts;

$$F^* = \frac{k - t}{v} F_k + \frac{v - (k - t)}{v} F_{k+v} \qquad (4.49)$$

where F_k is the price of the nearest contract, F_{k+v} is the price of the next nearest contract, v is the time between the expiry of two adjacent contracts, and k-t represents the time to expiry. Therefore, as the nearest contract approaches expiration the weighting of the index shifts to place the emphasis on the next one. The problem with splicing contracts together to form a returns series is that it can introduce expiry related seasonality in addition to trends already present in the data.

Rougier (1996) argues that his index addresses a number of the problems inherent in using futures data. Unlike the Clark (1973) approach which requires open interest, the series above requires very little extra information and is relatively easy to construct. It also reduces the impact of time trends due to expiry and is independent of the period over which it is calculated. As Rougier points out, a potential problem with this method is that it only takes account of two contracts trading at any one time. However, for the majority of futures, trading tends to be concentrated in the nearest and next nearest contracts.

4.4.3 THE CONSTRUCTION OF A CONTINUOUS VOLUME SERIES FOR A FUTURES CONTRACT.

As well as problems relating to the returns series, it is also important that non-information based trends are eliminated from the volume data. Trading volume for futures contracts often exhibits trends due to the roll-over effect that occurs as contracts near expiration. The task is to obtain some measure of the roll-over effect that can then be extracted from the data. Holmes and Rougier's (1997) roll-over adjustment uses the volume of trade which occurs during the day and the open interest at the end of the day to generate an upper bound for roll-over.

As expiration approaches there are three types of trade that can occur; opening a position, closing a position, and rolling over a position. By concentrating on the nearest and next nearest contracts it is possible to identify five key variables and to derive the relationship between them. Consider the following:

$$v' = n'_o + n'_c + n_r \qquad (4.50)$$

$$v'' = n''_o + n''_c + n_r \qquad (4.51)$$

$$\Delta o' = n'_o - n'_c - n_r \qquad (4.52)$$

$$\Delta o'' = n''_o - n''_c + n_r \qquad (4.53)$$

where a single prime refers to the nearest contract and the double prime relates to the next nearest contract. v is the daily trading volume, Δo is the change in open interest, n_o is the number of contracts opened, n_c is the number of contracts closed, and n_r is the rollover volume. All of the variables must be positive. Holmes and Rougier show that these equations must solve to give the following upper and lower bounds for rollover:

$$0 \le n_r \le \min\{\frac{1}{2}(v' - \Delta o'), \frac{1}{2}(v'' + \Delta o'')\} \qquad (4.54)$$

If the upper bound is a good proxy for roll-over, then roll-over adjusted volume is created by subtracting twice the value of the upper bound for roll-over on any particular day from the total volume of trade on that day. Holmes and Rougier show that in the case of a sample of S&P 500 volume data for the next and next nearest contracts over the period 2/1/90 to 18/9/96, this technique proves very successful in eliminating the roll-over effects that occur at expiration.

The next section of this chapter uses GMM, the contiguous returns index, and the roll-over adjustment to examine Anderson's (1996) modified version of the MDH.

4.5 EMPIRICAL RESULTS

In this, the empirical section of the chapter, a direct test of the MDH is carried out by investigating the implied moments of Anderson's (1996) version of the mixtures model. The generalised method of moments methodology is used, like Anderson, to carry out this test. The MDH assumes that the flow of information is the driving force behind the association between the volume of trade and price return volatility. As will be explained, it is crucial that the data reflects, as far as possible, movements due to information arrival in the market and that trends due to, for example, the growth of the market are eliminated. Anderson spends a lot of time ensuring that his data is stationary before carrying out a test of his model. This study goes a step further than any previous studies that have used futures data and actually makes allowances for the roll-over effect that occurs as contracts reach expiration. This is achieved without losing the impact of important pieces of information that may arrive during this period.

4.5.1 DATA AND PRELIMINARY ANALYSIS

This section provides detailed analysis of the return and volume data used in this investigation. These preliminary results are very important. The MDH is believed to explain certain characteristics of return and volume data and the relationship between them. The process of carrying out some simple statistical tests of the data allows the possible identification of these characteristics. There is little point in using a model to explain why volume and returns exhibit certain traits if these traits do not exist in the first place.

This test of the MDH is carried out using the daily returns and volume for three futures contracts; the FTSE 100, Long Gilts and Brent Oil. Data covering the period from January 1992 to July 1996 was supplied by LIFFE for the first two contracts, and the Brent Oil data, covering the same period, was supplied by the IPE.

These three contracts have been chosen for a number of economic and practical reasons. Firstly, it has already been shown in chapter 2 that for each contract the relationship between volume and volatility exhibits the intrinsic qualities that have been considered consistent with the MDH. The contracts also have particular characteristics, alluded to in chapter 2, that may allow some interesting comparisons to be made. Each futures contract tends to reflect the characteristics of the underlying spot market. In comparison with the FTSE 100, Long Gilts are often considered a relatively safe investment and it might therefore be expected that this futures market will attract a greater proportion of risk-averse investors. Thus, the distinction between information based and non-information based trading may be more important for the FTSE 100 futures contract than for the Long Gilts futures contract. Hedging in commodity markets, for example Brent Oil, tends to form a larger percentage of overall trading compared to financial markets, so it might be expected that the proportions of noise and news trading will differ relative to the other contracts.

A further expectation is that the information processes of the FTSE 100 and Brent Oil markets may be linked. A large number of firms in the FTSE 100 will be affected by the price of oil, either because of direct links with the oil industry or because oil is an important part of the production process. Therefore, there will be information that is common to both markets.

There are also a number of practical reasons for choosing these three contracts. Since they all represent highly liquid markets there are unlikely to be a large number of days, if any, when there is no volume and where prices reflect the last day of active trading which may have been some time ago.

An important part of this chapter is the use of the Holmes-Rougier (1997) roll-over adjustment and the Rougier (1996) contiguous price index. The tin contract used in chapters 2 and 3 is not suitable for this purpose because it represents a forward contract rather than a series of individual contracts with definite expiration times. The cocoa futures contract is also unsuitable because the relationship between open interest and volume is not as well-defined as it is for the contracts under investigation in this chapter. The expectation is that daily volume is at least as great as the

corresponding daily change in open interest. This relationship does not hold, however, in certain commodity markets. The reason for this is a procedure called "re-allocation", which has been in operation since 1992, but is now under review. For some contracts LIFFE allows single clients to hold a long and a short position with different members. Should the client wish to close out this position, then one member can transfer their position to the other, who will then close it out. The "re-allocation" will be announced to the market, but it will not be counted as traded volume. Thus, the change in open interest may exceed volume. Normally this would only be expected in a data set supplied by a market if values have been incorrectly inputted into the spreadsheet. Therefore, in an attempt to screen the data all observations used in this study where the daily change in volume is less than the corresponding open interest are removed. The corresponding price return observations are also removed. This process reduces the initial data set of 1131 observations down to one of 1121 observations for the FTSE 100 contract, and from 1135 down to 1093 observations for the Long Gilt contract.

The Brent Oil contract requires a little more adjustment. Close examination of the data reveals that large drops in trading occur after the expiry of the October contract, for 1993, 1994, and 1995. In 1992 the drop in trading occurs at the expiration of the November contract. This lack of volume lasts up to twenty-four trading days. This may be explained by the fact that the next contract does not expire until January. This gap is unusual in the Brent Oil market where for the rest of the year a contract expires every month. This seasonal trend is excluded after the data is screened and reduces the original 1184 observations to 1029 observations.

The returns series is constructed from settlement prices and the Rougier (1996) contiguous price index. The returns are calculated as the difference in the logarithm of daily prices, consistent with previous chapters. Preliminary analysis of the contiguous returns index is shown in table 4.1.

The first thing to notice is that the mean for each contract is very close to zero. The returns also display excess kurtosis relative to a normally distributed series and are positively skewed, with the exception of Long Gilts. It should be noted, however, that

although these are conditions that the MDH attempts to explain, the estimation of higher order moments can be affected by sample outliers so these preliminary statistics should only act as an initial guide. Further evidence that returns are not independently drawn from a normal distribution is provided by looking at the autocorrelation of the returns series.

Table 4.1: Summary Statistics for the Price Return Series of the FTSE 100, Long Gilt and Brent Oil Futures Contracts for the Period 1992 to 1996

Contract	Obs	Mean	St Dev	Skewness	Kurtosis
FTSE 100	1121	0.299E-03	0.885E-02	0.061	1.660
Long Gilt	1092	0.770E-04	0.536E-02	-0.036	3.345
Brent Oil	1029	0.114E-03	0.014	0.044	3.132

Note: Obs is the number of observations in the sample. St Dev is the standard deviation for the return series, measured as: $(\sum_{t=1}^{n}(x_t - \overline{x})^2 /(n-1))^{1/2}$, where x_t is the price return series and n is the number of observations.

$$\text{Kurtosis} = \left\{ \frac{n}{(n-1)(n-2)(n-3)} \sum_{j=1}^{n}((x_j - \overline{x})/s)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}.$$

$$\text{Skewness} = \frac{n}{(n-1)(n-2)} \sum_{j=1}^{n}((x_j - \overline{x})/s)^3, \text{ where s is the sample standard deviation.}$$

The two most common tests of autocorrelation are the Ljung-Box (1978) and Box-Pierce (1970) statistics. Although the Box-Pierce (1970) is a very popular test Ljung and Box (1978) argue that it produces lower than anticipated test statistic results, particularly for series that do not exhibit standard normal characteristics. We therefore use the Ljung-Box (1978) test represented by:

$$Q^* = n(n+2)\sum_{j=1}^{p}\left(\frac{R_j^2}{n-j} \right) \tag{4.55}$$

where R is the autocorrelation parameter, p is the order of autocorrelation and n is the number of observations in the sample. The test statistic Q^* is distributed as a χ^2-distribution with degrees of freedom equal to the order of autocorrelation under investigation.

In tables 4.2, 4.3, and 4.4, the test for autocorrelation in returns, absolute returns and squared returns under the assumption of no serial correlation is distributed as a χ^2-distribution with 10 degrees of freedom. The probability values are in parentheses. The critical values for a χ^2-distribution with 10 degrees of freedom are 15.99 and 18.31 at significance levels of 10 percent and 5 percent respectively. The Ljung-Box test statistic for the FTSE 100 returns series is greater than the critical value indicating that there is evidence of correlation between successive returns. The Ljung-Box test statistics for the Long Gilt and Brent Oil returns series are just under the ten percent critical value. The assumption of no serial correlation is therefore not rejected. If the returns series is iid then squared returns and absolute returns should also be iid. This is not borne out by the evidence. In each case the Ljung-Box statistic shows that a relationship exists between successive observations.

Evidence of autocorrelation can be illustrated further by considering the autocorrelation plots of returns, squared returns, and absolute returns in figures 4.1, 4.2 and 4.3. At a purely visual level it is possible to see that successive observations for each series do not appear to be independent. One positive (negative) movement is often followed by another positive (negative) movement. A truly independent series would oscillate around the origin. The dotted lines on each graph indicate the 5 per cent confidence intervals for first order serial correlation calculated from the Ljung-Box statistic. The graph shows clearly that the returns series do not lie entirely within these bands.

It can be clearly seen in these figures that the three series exhibit autocorrelation. The series do not, therefore, appear to have normally distributed iid. returns. It was this dependency in higher order moments that allowed the modelling of each returns series as a GARCH specification in chapter 2.

Table 4.2: Autocorrelation Analysis of the Returns Series for the FTSE 100 Futures
Contract (1992-1996)

Sample	1	2	3	4	5
Returns	-0.014	-0.269E-03	-0.021	-0.011	0.019
|Returns|	0.104	0.114	0.066	0.048	0.074
Returns^2	0.123	0.085	0.048	0.021	0.024
	6	7	8	9	10
Returns	-0.078	-0.082	0.052	-0.049	0.009
|Returns|	0.138	0.134	0.079	0.080	0.074
Returns^2	0.073	0.086	0.069	0.031	0.081

Note: Ljung-Box Test: χ^2_{10} (Returns)=21.54 (0.018), χ^2_{10} (|Returns|)=57.36 (0.000),

χ^2_{10} (Returns^2)=102.49 (0.000).

Table 4.3: Autocorrelation Analysis of the Returns Series for the Long Gilt Futures
Contract (1992-1996)

Sample	1	2	3	4	5
Returns	0.002	-0.018	0.127	-0.063	0.220
|Returns|	0.067	0.099	0.136	0.025	0.137
Returns^2	0.052	0.051	0.101	0.023	0.086
	6	7	8	9	10
Returns	-0.004	-0.073	-0.010	-0.002	0.045
|Returns|	0.100	0.108	0.080	0.084	0.111
Returns^2	0.053	0.070	0.025	0.047	0.052

Note: Ljung-Box Test: χ^2_{10} (Returns)=13.69 (0.188), χ^2_{10} (|Returns|)=109.72 (0.000),

χ^2_{10} (Returns^2)=40.24 (0.000).

Table 4.4: Autocorrelation Analysis of the Returns Series for the Brent Oil Futures Contract (1992-1996)

Sample	1	2	3	4	5
Returns	-0.024	-0.030	-0.063	-0.060	-0.004
|Returns|	0.076	0.112	0.074	0.171	0.098
$Returns^2$	0.050	0.092	0.047	0.248	0.076
	6	7	8	9	10
Returns	0.027	0.023	0.012	0.065	-0.019
|Returns|	0.117	0.078	0.120	0.047	0.075
$Returns^2$	0.041	0.079	0.085	0.026	0.042

Note: Ljung-Box Test: χ^2_{10} (Returns)=15.65 (0.110), χ^2_{10} (|Returns|)=108.59 (0.000),

χ^2_{10} ($Returns^2$)=101.65 (0.000).

A further justification for this study can be provided by considering the cross-correlations between return volatility and trading volume for the whole sample and yearly subsamples. The yearly subsamples are used to reveal trends that, we suspect, are present in the volume data. Table 4.5 shows that quite a strong contemporaneous relationship exists between the variables. This strong correlation supports much of the early work done in this field (see section 4.3) even though at this stage we have yet to fully analyse the volume data.

Figure 4.1: Plots of the Autocorrelation Function of the Returns, Absolute Returns, and Squared Returns Series for the FTSE 100 Futures Contract (1992-1996)
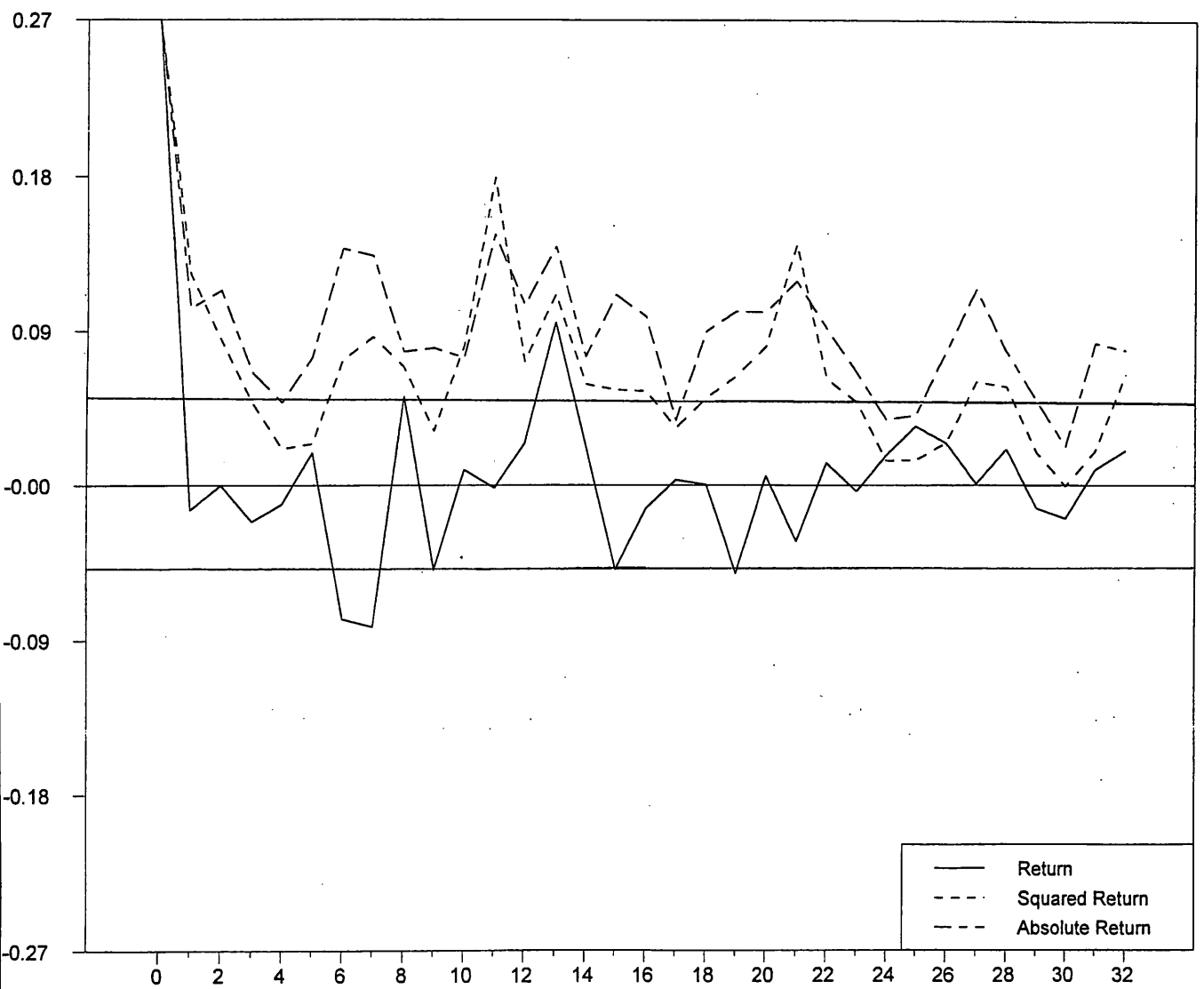
Figure 4.2: Plots of the Autocorrelation Function of the Returns, Absolute Returns, and Squared Returns Series for the Long Gilt Futures Contract (1992-1996)
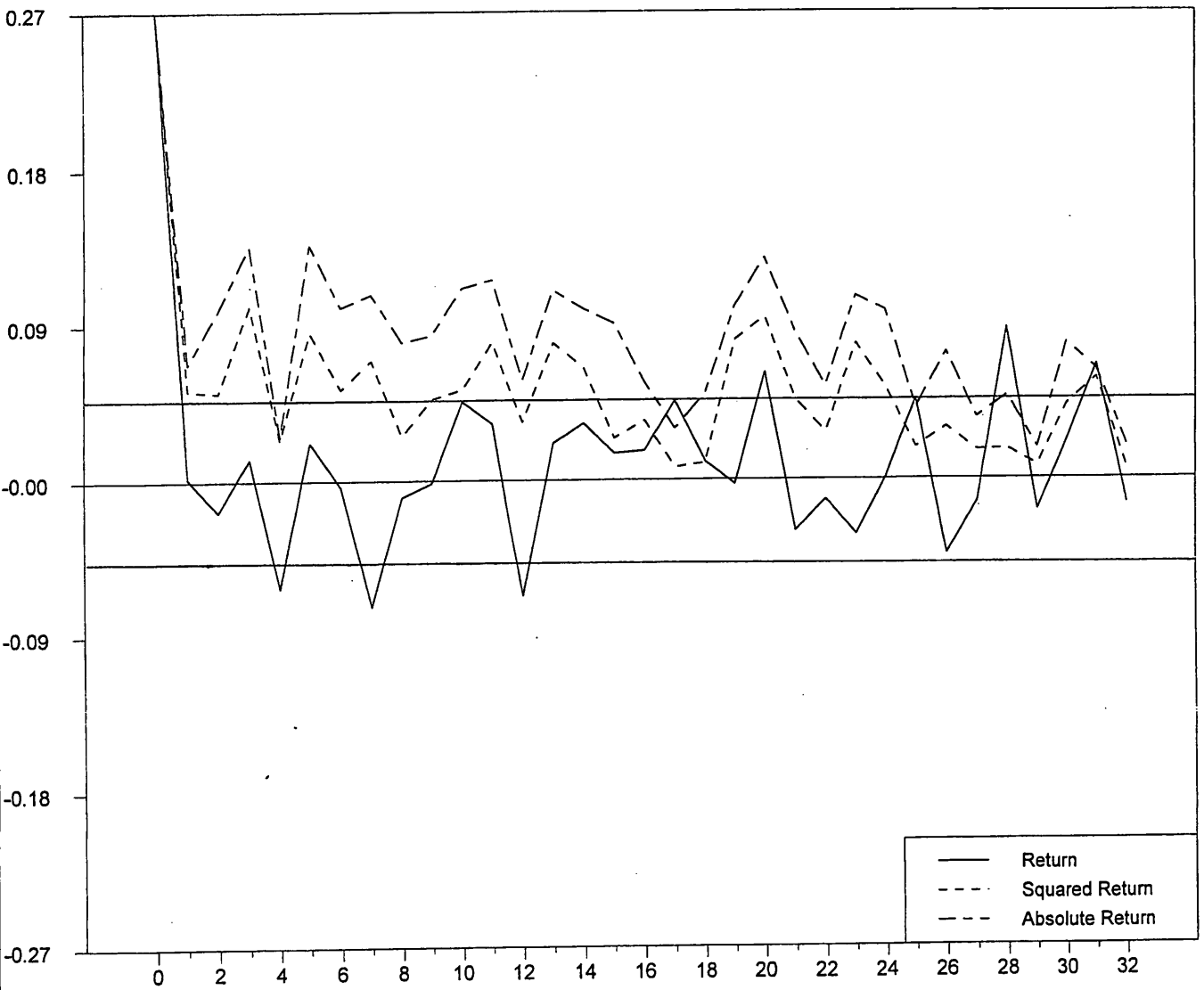
Figure 4.3: Plots of the Autocorrelation Function of the Returns, Absolute Returns, and Squared Returns Series for the Brent Oil Futures Contract (1992-1996)
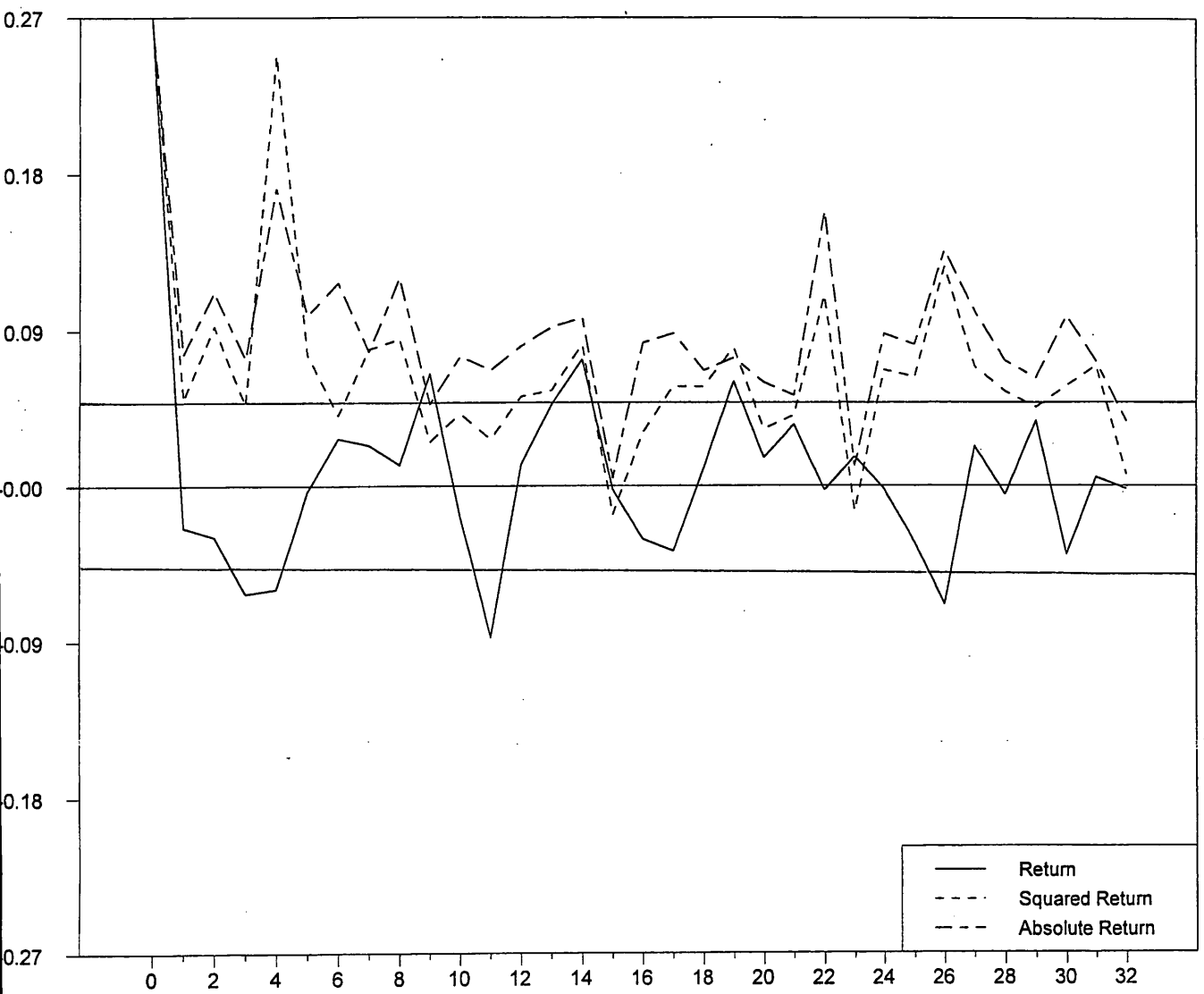
It is interesting to note how, in comparison with the results in chapter 2, the contemporaneous relationship for the full sample is stronger. This further illustrates how important the construction of the data set can be in determining the underlying relationship between variables.

Table 4.5: Cross-Correlations between Squared Returns and Trading Volume for the FTSE 100, Long Gilt and Brent Oil Futures Contracts

	$\text{Corr}(R_t^2, V_t)$		
Sample	FTSE 100	Long Gilt	Brent Oil
1992-1996	0.302	0.344	0.325
1992	0.508	0.449	0.208
1993	0.437	0.362	0.367
1994	0.335	0.308	0.276
1995	0.129	0.448	0.246
1996	0.397	0.345	0.491

Note: R_t^2 is the squared return series and V_t is volume.

The volume series for each contract was initially constructed, after the observations had been screened, using total volume from the nearest and next nearest contracts being traded. One of the problems of using contemporary volume data is that it tends to exhibit significant upward trends. Since this is unlikely to be attributable to information, otherwise we would experience a news explosion, it must be attributed to the steady growth in the popularity of futures market trading. The three contracts considered in this chapter are all relatively new. The Long Gilt contract was the first government bond futures contract launched in Europe and began trading on LIFFE in 1982. The FTSE 100 futures contract traces its inception back to 1984, while the Brent Oil contract, despite being initially launched in 1983, was re-launched in 1988. The last decade has seen a dramatic growth in the volume of futures trading. This has been partly attributed to the development of financial futures, but can also be explained by increased awareness of the function of secondary markets. Another

important factor is the growing use of futures markets by large financial institutions to manage portfolios and other risky assets.

As has already been mentioned in this and other chapters, such trends can obscure what is really happening in the market in terms of the arrival and dissemination of news and make the testing of a model like the MDH very difficult. Table 4.6 gives some indication of the growth of the volume of trade in all three contracts over the period of the sample. Although the table indicates that the general trend in the volume of trade is upwards there is evidence of significant falls during the 1994-95 period. This drop in trading is well documented. Despite record 1994 volumes being reported on the world's major exchanges, the situation had changed by the middle of 1995. The blame for this pessimism on the derivative markets has been placed on a number of high-profile corporate losses. The collapse of Barings, the experiences of Procter and Gamble with interest rate swaps, Metallgesellschaft's dealings in oil markets, and Orange County's losses in gilt markets all conspired to create a very nervous market. The devaluation of the Mexican currency and the collapse of a number of emerging markets is also believed to have had an important impact on trading volume. Lapper (1995b) argues that the losses experienced by many of the banks and security houses dealing in derivatives resulted in dealers having to operate under much tighter controls.

The overall upward trend in trading volume is further supported by figures 4.4, 4.5 and 4.6. The fitted trendline clearly indicates an increase in trading over the period 1992 to 1996. It is possible to eliminate this trend by taking a log transformation of the volume data as shown in figures 4.7, 4.8 and 4.9. At a quick glance the log transformation appears to remove the upward trend and it also appears to stabilise the variance of the volume series. However, like Anderson (1996), we believe that detrending the data in this way and therefore assuming that growth in the volume of trading is constant, is too restrictive. The negative growth in trading volume between 1994 and 1995 for each contract would suggest that a more sympathetic detrending procedure is required. The data is therefore detrended using a two-sided one year weighted rolling mean.

Table 4.6: Futures Contract Annual Growth Rates in Trading Volume (1992-1996)

	Percentage Annual Growth		
Sample	FTSE 100	Long Gilt	Brent Oil
1992-1996	20.151	35.668	38.329
1993-1994	37.100	62.908	12.477
1994-1995	-20.222	-27.518	-3.723
1995-1996	4.679	8.475	16.336

A one-sided weighted mean, as described by Brockwell and Davis (1987)[4], is used at the beginning and end of the sample where the two sided technique can not be used. There is little guidance provided in the literature regarding the choice of the length of the weighted mean. Anderson (1996) uses a two year weighted mean, but, given the evidence above in table 4.6, this would miss events that occur at a higher frequency. Holmes and Rougier (1997) use a moving average of 63 days, based on the length of time between successive contract expirations[5]. It is felt that although this could be applied in this case, it represents a 'belt and braces' approach to coping with the problems caused by roll-over which is unnecessary given the use of the roll-over adjustment.

Another issue here, one that did not concern Anderson (1996), is the problem of roll-over. As a contract expires investors will often close out their positions in the expiring contract and open positions in the next nearest contract. This leads to a large amount of market activity that is not information driven. Figure 4.10 shows the autocorrelation plot for the trading volume of the FTSE 100 futures contract.

---

[4] We used the following formula for the two-sided weighted moving average W_t; $W_t = (2q+1)^{-1} \sum_{j=-q}^{q} X_{t+j}$, where $q+1 \leq t \leq n-q$. X_t is the series to be weighted and q is a non-negative integer. In our example X_t is the volume series and $q = 126$, half the number of trading days in one year. The one sided moving average is given by: $m_t = \sum_{j=0}^{n-t} \alpha(1-\alpha)^j X_{t+j}$, where $t = 1,...,q$. We used $\alpha = 0.3$. Brockwell and Davis (1987) argue that there is little to choose between 0.3, 0.2 or 0.1.

[5] As exploited in chapter 2.

Figure 4.4: The Growth in the Volume of Trade for the FTSE 100 Futures Contract (1992-1996)
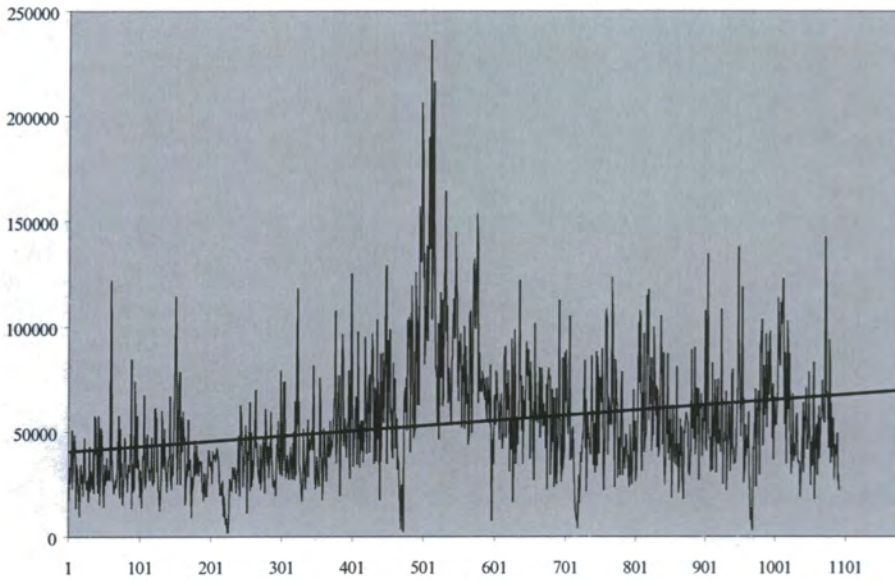
Figure 4.5: The Growth in the Volume of Trade for the Long Gilt Futures Contract (1992-1996)
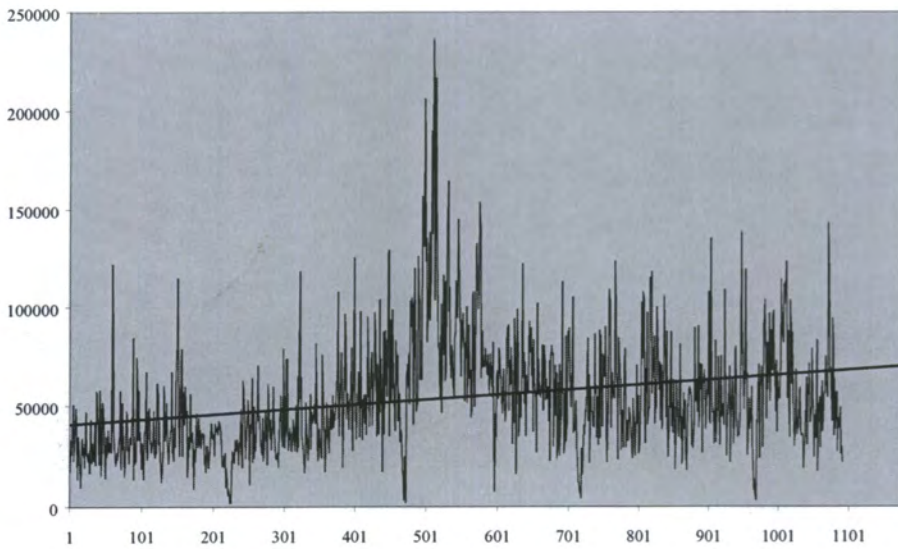
Figure 4.6: The Growth in the Volume of Trade for the Brent Oil Futures Contract
(1992-1996)
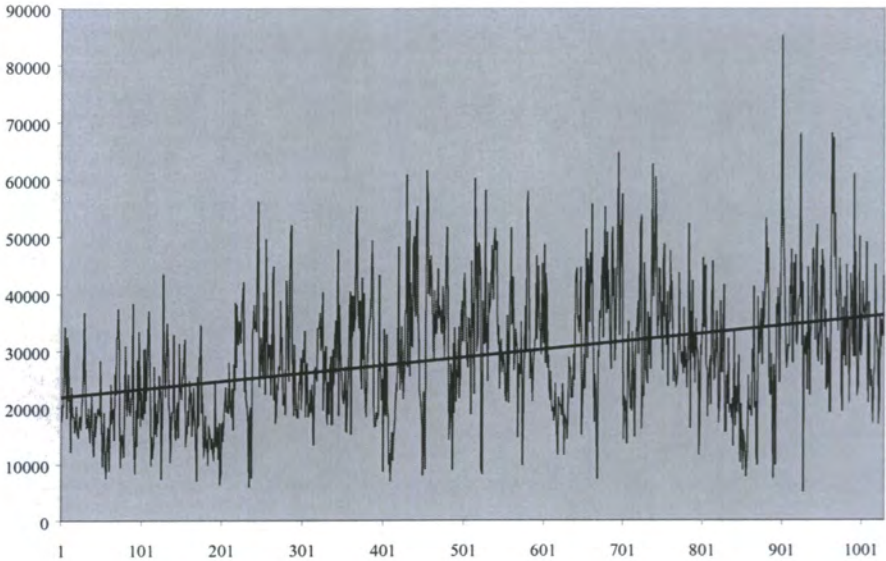
Figure 4.7: The Logarithmic Transformation of the FTSE 100 Futures Contract
Trading Volume (1992-1996)
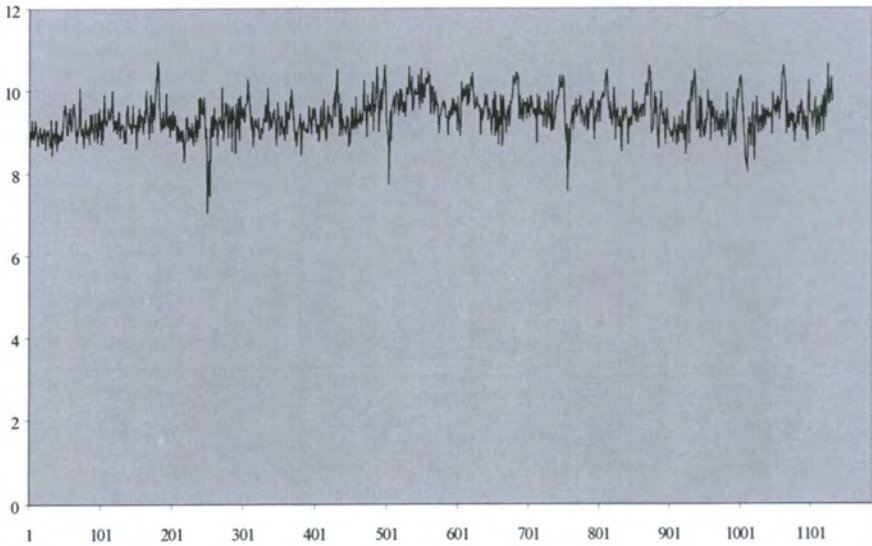
Figure 4.8: The Logarithmic Transformation of the Long Gilt Futures Contract
Trading Volume (1992-1996)
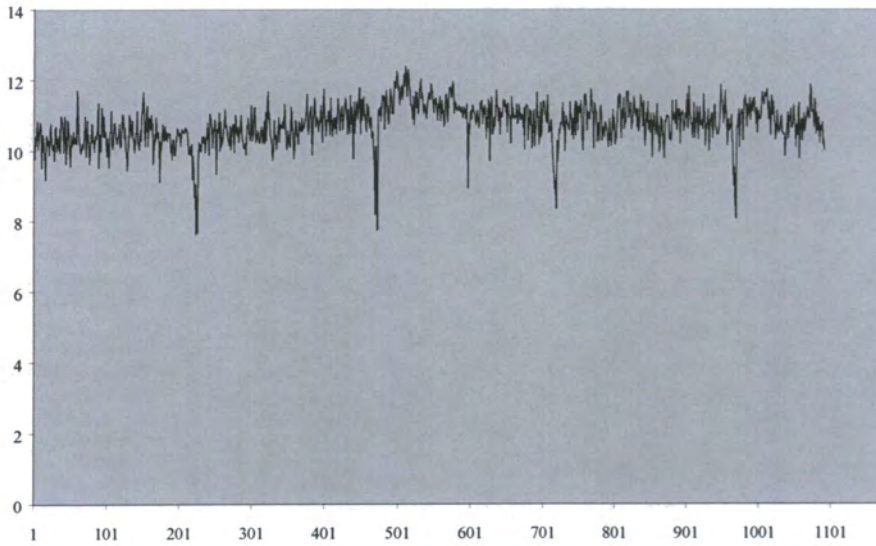
Figure 4.9: The Logarithmic Transformation of the Brent Oil Futures Contract
Trading Volume (1992-1996)
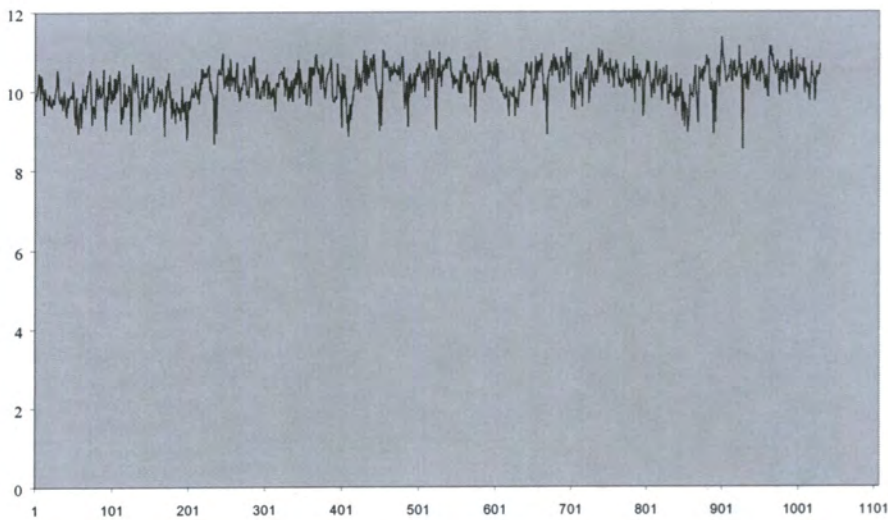
The volume series, represented by the sum of the trading volume for the nearest and next nearest contract on any given day, exhibits peaks in its autocorrelation function that coincide with the quarterly nature of contract expiration. Each of these peaks indicates that there is a significant period where the correlation of daily trading increases before reaching a peak and a subsequent fall. Closer examination of the data reveals quite clearly the increase in trading at expiration due to either traders having to meet their obligations, or more significantly, roll-over. One or two days after expiration the level of trading then falls back to 'normal' levels. This phenomenon can also be seen in figure 4.12, the autocorrelation plot of Brent Oil trading volume. The peaks appear more frequently because contracts expire more often during a given period relative to the FTSE 100. In one year up to eleven contracts expire in the Brent Oil market compared to the FTSE's four. Volume in the Long Gilt market, however, does not display such obvious autocorrelation characteristics. The periodicity of expiration is four per year, similar to the FTSE 100, and there will almost certainly be some roll-over. Figure 4.11 shows two small peaks, (the horizontal axis is shorter than in figures 4.10 and 4.12 because with a longer axis it is very difficult to identify these small peaks), indicating that the correlation of trades around expiration is much smaller in this market. A close examination of the data also reveals that, unlike the other two contracts, there are only small increases in volume at this time.

The challenge is to eliminate the non-information based roll-over trading while preserving the information based trading in the market. Therefore, before reweighting each volume series the Holmes-Rougier (1997) roll-over adjustment is applied. Figures 4.13 and 4.15 show that the impact of the expiration effects has been significantly reduced. The difference between figures 4.14 and 4.11 is barely perceptible which is unsurprising given the much smaller impact of roll-over in the Long Gilt market.

Figure 4.10: The Plot of the Autocorrelation Function of Trading Volume for the
FTSE 100 Futures Contract (1992-1996)
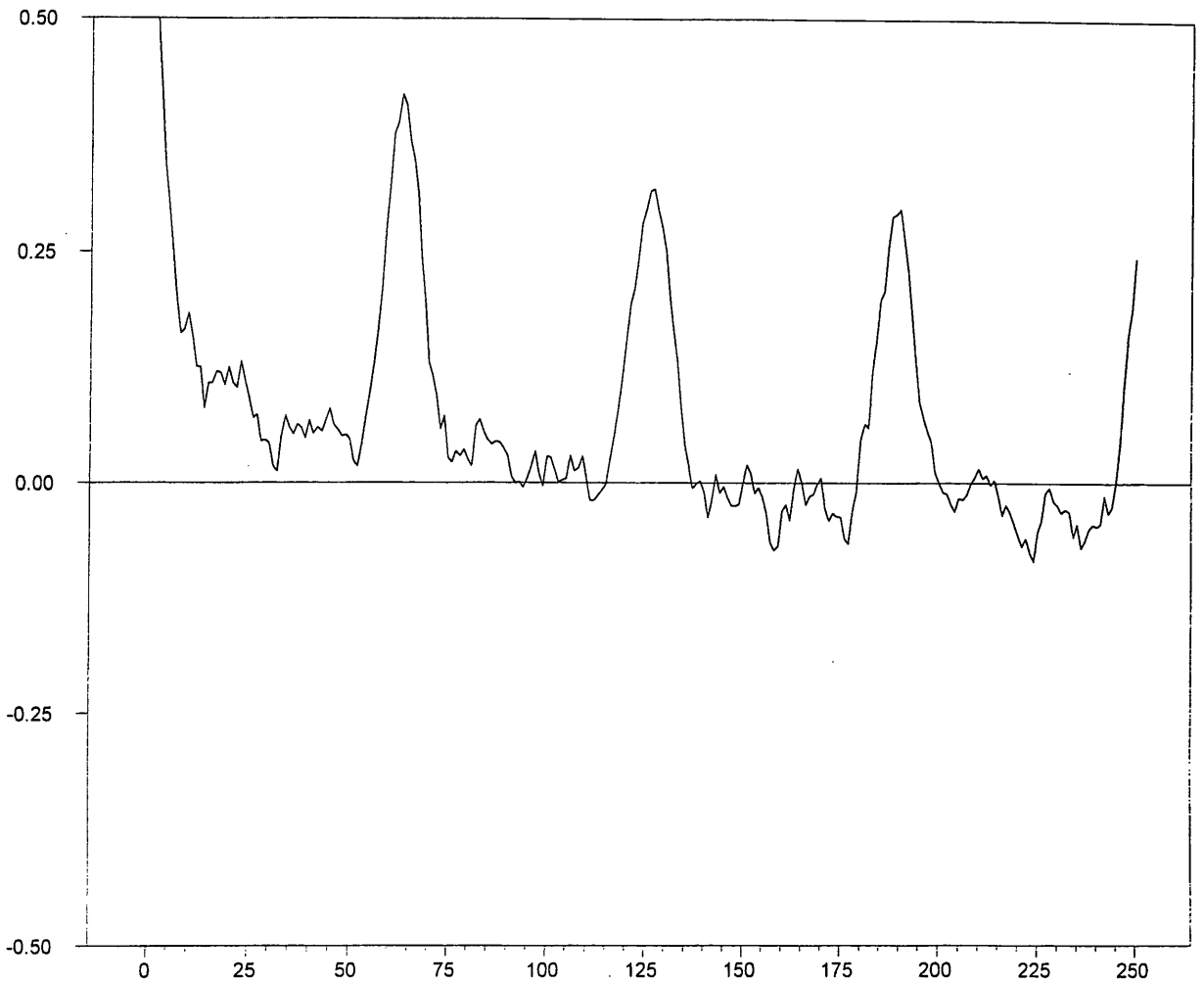
Figure 4.11: The Plot of the Autocorrelation Function of Trading Volume for the
Long Gilt Futures Contract (1992-1996)
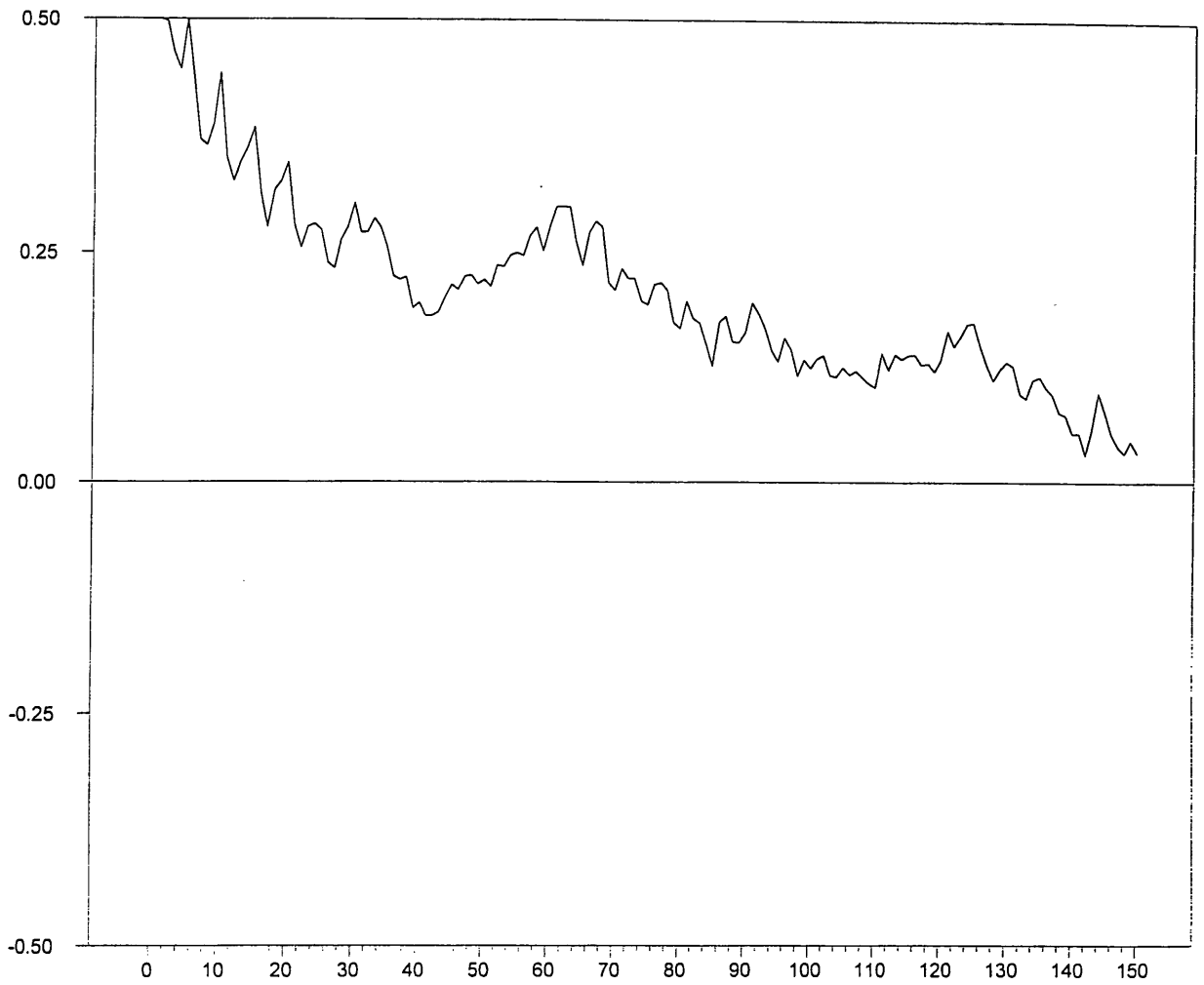
Figure 4.12: The Plot of the Autocorrelation Function of Trading Volume for the Brent Oil Futures Contract (1992-1996)
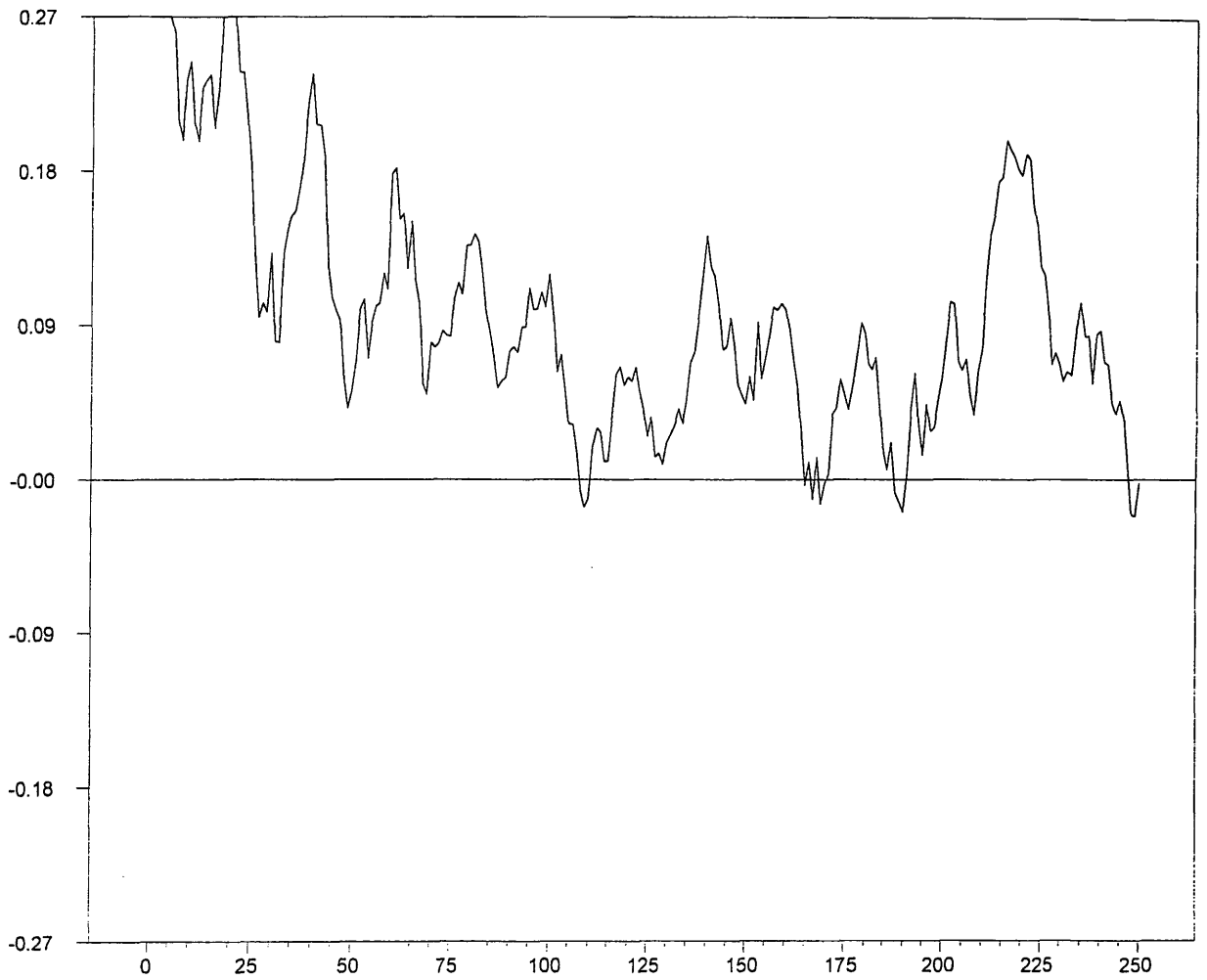
Figure 4.13: The Plot of the Autocorrelation Function of Trading Volume Minus
Rollover for the FTSE 100 Futures Contract (1992-1996)
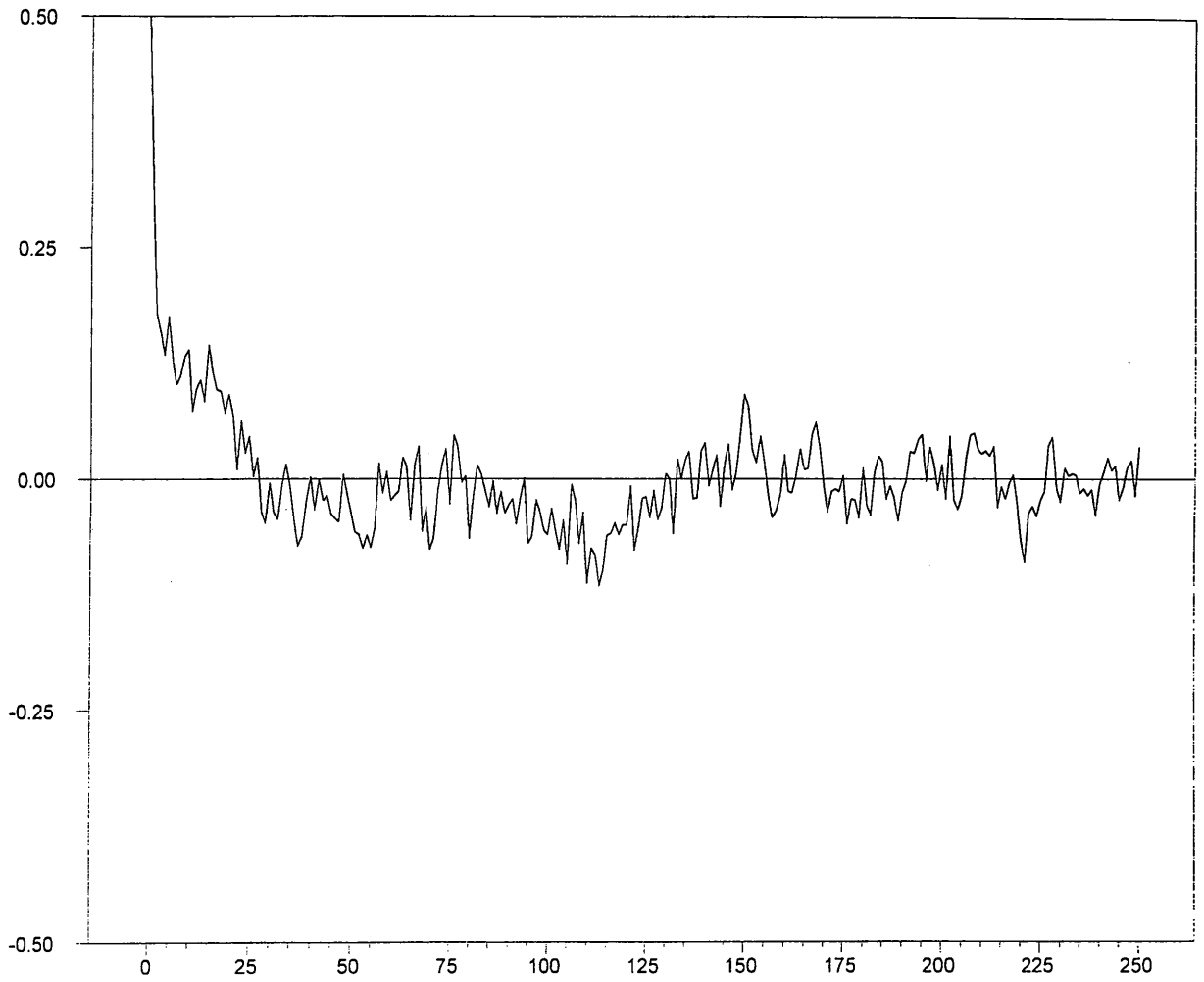
Figure 4.14: The Plot of the Autocorrelation Function of Trading Volume Minus
Rollover for the Long Gilt Futures Contract (1992-1996)
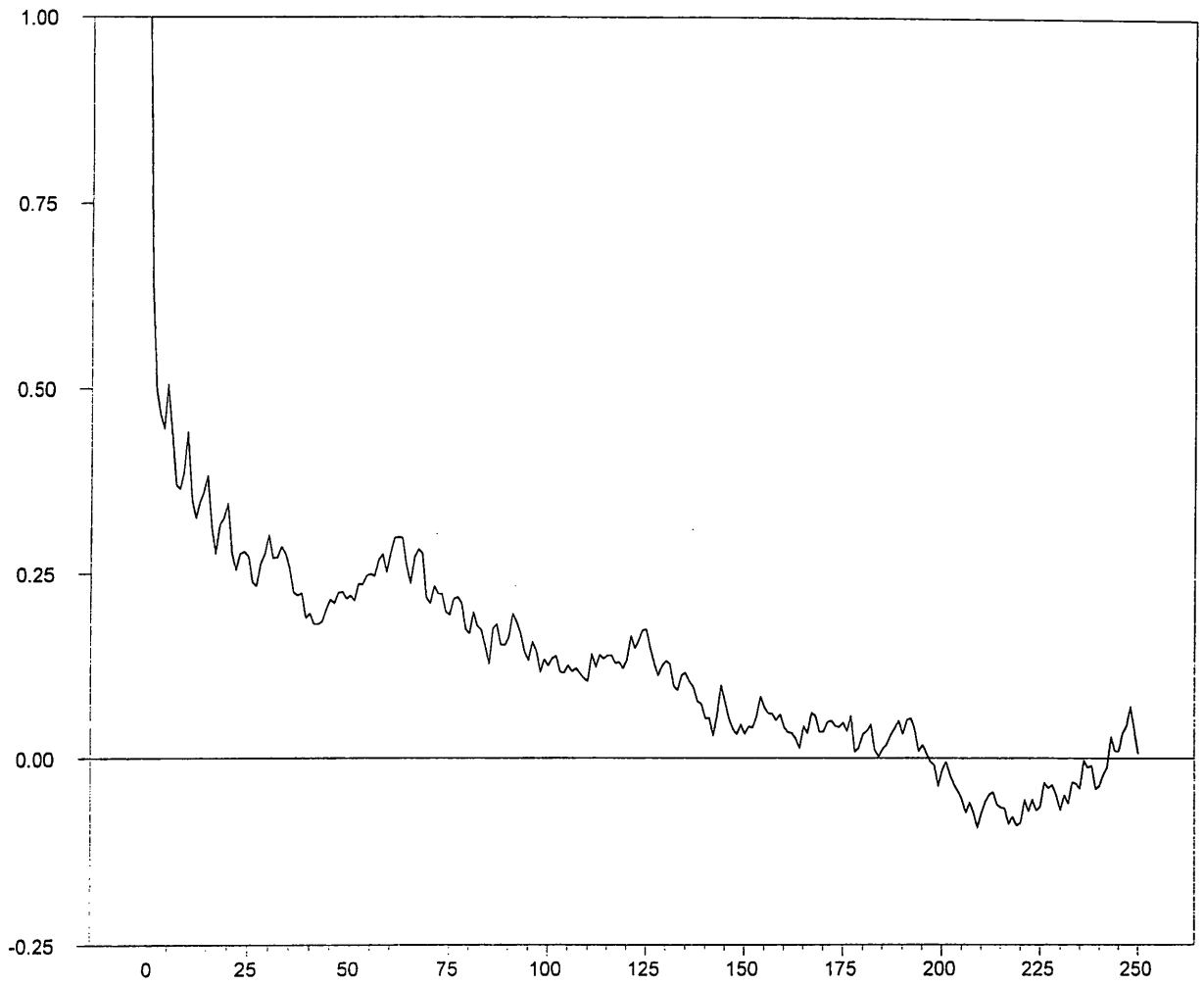
Figure 4.15: The Plot of the Autocorrelation Function of Trading Volume Minus
Rollover for the Brent Oil Futures Contract (1992-1996)
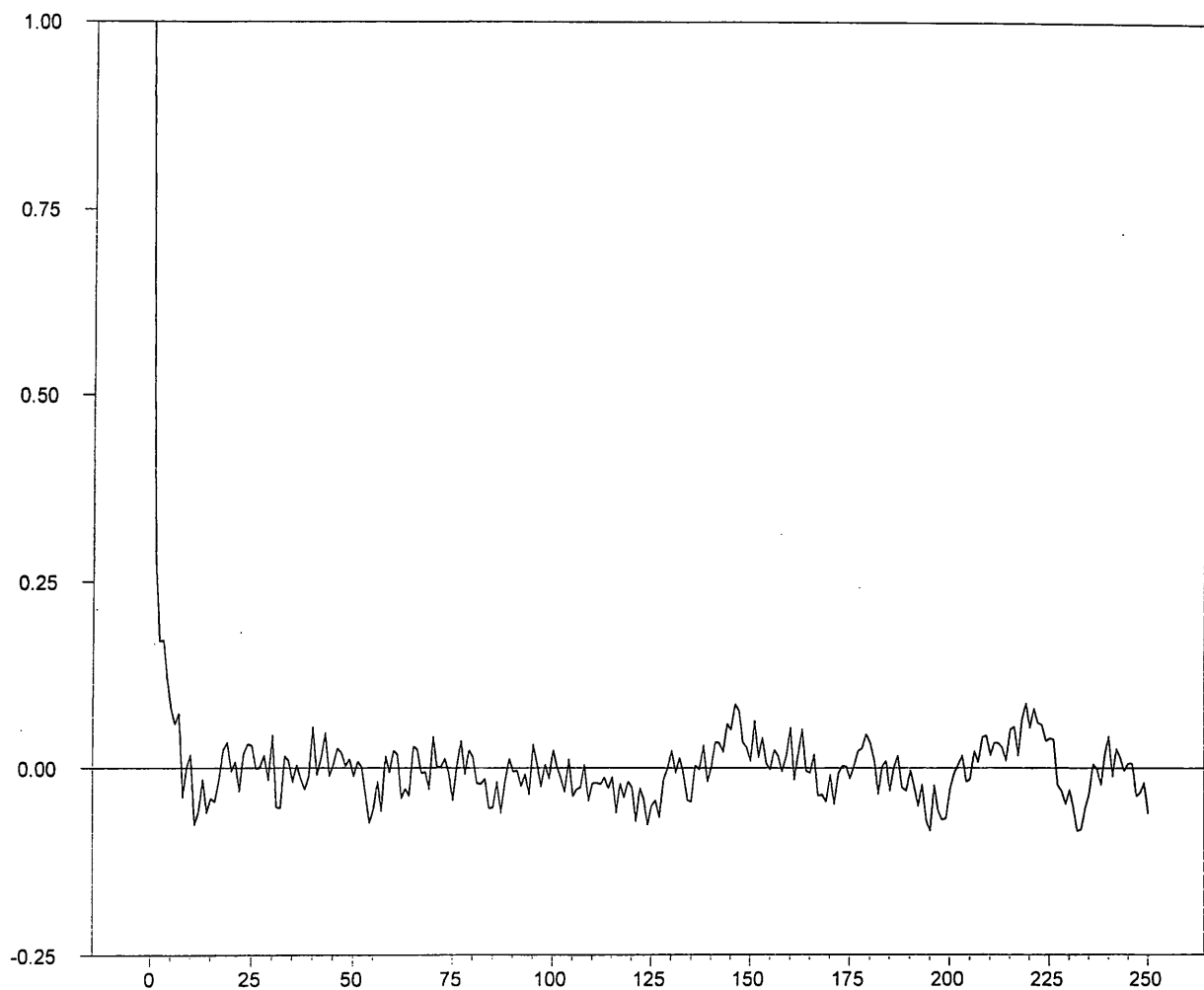
One of the problems that Holmes and Rougier address is whether or not it is important to consider a range of possible values of the roll-over volume and thereby construct an optimal roll-over adjustment. Figure 4.16 considers the effect of using a proportion φ of the upper limit for the FTSE 100 futures contract roll-over, i.e.

$$v^* = v - 2\varphi n_r \qquad (4.56)$$

where v^* and v represent adjusted volume and unadjusted volume respectively. Values of φ equal to 0.5, 0.75 and 1 are tried. Figure 4.16 illustrates clearly that the roll-over adjustment is most effective when φ is equal to one. At values less than one there is still evidence of seasonality due to roll-over in the volume series. A similar process carried out for the other two contracts produced the same result.

The weighting procedure and the roll-over adjustment together represent the first part of the detrending process. They have essentially estimated a trend component that produces an expected volume series. The detrended series for each contract is then generated by dividing the actual volume on a particular day by the expected volume on that same day. Summary statistics for each new volume series are given in tables 4.7, 4.8 and 4.9.

All three tables show that the mean of the whole sample and across five subsamples is near unity which would be expected given our detrending procedure. The standard deviation appears on the whole to be stable indicating that some degree of stability has been achieved across each sample. The only subsample out of line appears to be 1996 for both the Long Gilt and the FTSE 100 contracts. Although, in each case, the mean is close to one, the standard deviation is much lower than the average for the other four subsamples. It is worth noting that 1996 does not represent a full year of data.

Figure 4.16: Plots of the Autocorrelation Function for the Detrended FTSE 100 Futures Contract for Different Magnitudes of the Rollover Adjustment
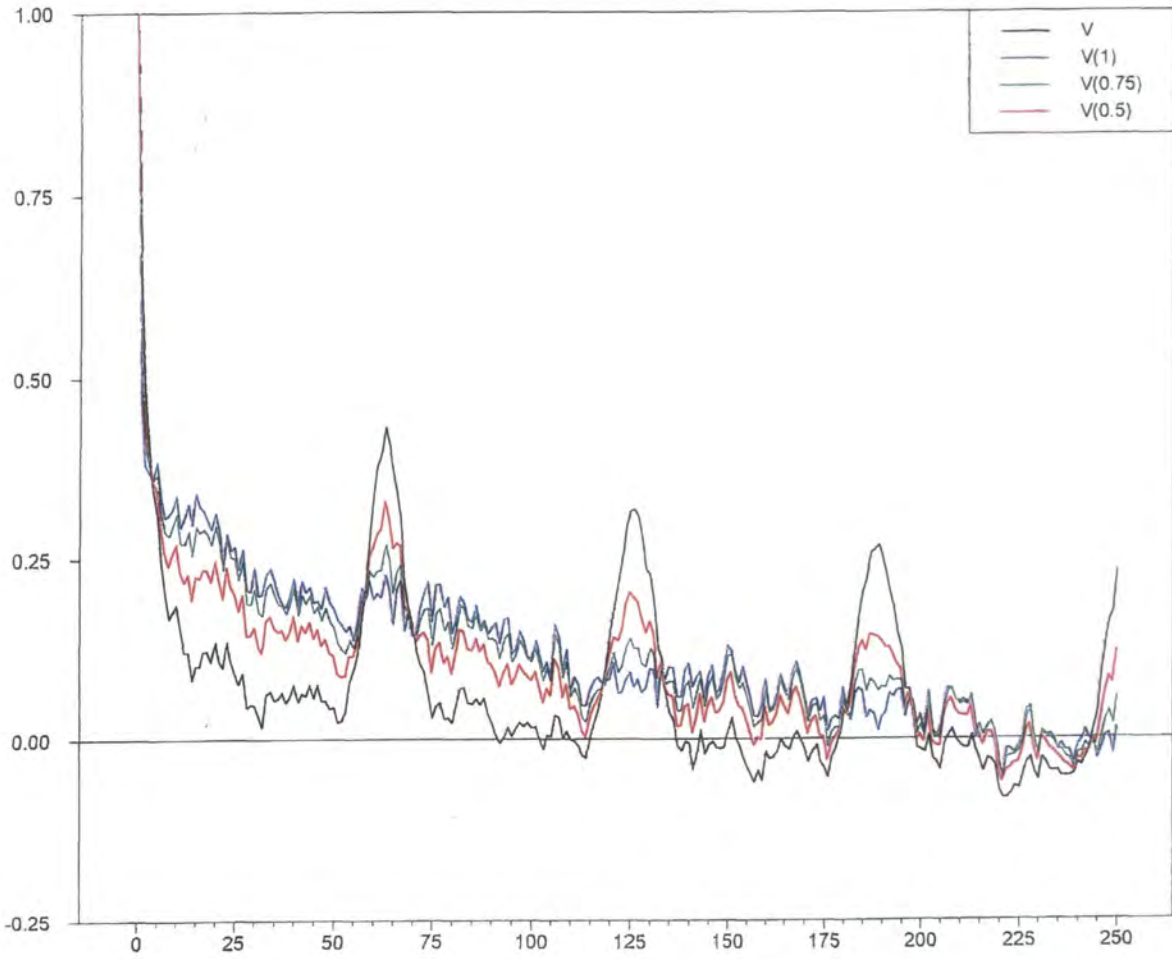
While the standard deviation is relatively stable the higher order moments do not appear to be quite as predictable. Such calculations are adversely affected by outlying observations, but it may also signal estimation problems at the GMM stage of this investigation. Another important observation is that the skewness is positive in the full sample and across all sub-samples. As Anderson (1996) points out, any theoretical model of the volume-volatility relation must be able to explain such data characteristics. It is worth noting that for all three contracts the level of skewness is higher for the volume than it is for the returns samples in tables 4.1, 4.2, and 4.3. A possible reason for this is suggested in the next section.

Tables 4.10-4.15 show the autocorrelation coefficients for squared returns and volume up to the thirty second order. Harris (1987) argues that an implication of the MDH is that autocorrelation coefficients should be largest for volume relative to squared returns. He argues that the impact of information on the autocorrelation of each series is dependent on the fraction of the variation in the series that is explained by the variation in the intensity of information flows. This fraction is large for volume because the conditional mean of the volume distribution is large relative to the conditional variance for all information arrivals. The converse is true for squared returns. It is quite possible for squared returns to be high when the number of information arrivals is low and vice versa.

Table 4.7: FTSE 100 Futures Contract Summary Statistics for Detrended Volume

	Sample					
	1992-96	1992	1993	1994	1995	1996
Mean	1.000	1.002	0.967	1.045	0.985	1.012
Std. Dev	0.408	0.408	0.404	0.413	0.463	0.245
Skewness	1.822	3.478	1.620	0.828	1.162	0.302
Kurtosis	6.484	20.332	3.661	0.548	2.979	0.312

Note: Std Dev is the standard deviation for the return series, measured as: $(\sum_{t=1}^{n}(x_t-\bar{x})^2/(n-1))^{1/2}$,

where x_t is the price return series and n is the number of observations.

$$Kurtosis = \left\{\frac{n}{(n-1)(n-2)(n-3)}\sum_{j=1}^{n}((x_j-\bar{x})/s)^4\right\} - \frac{3(n-1)^2}{(n-2)(n-3)}.$$

$$Skewness = \frac{n}{(n-1)(n-2)}\sum_{j=1}^{n}((x_j-\bar{x})/s)^3 \text{, where s is the sample standard deviation.}$$

Table 4.8: Long Gilt Futures Contract Summary Statistics for Detrended Volume

	Sample					
	1992-96	1992	1993	1994	1995	1996
Mean	0.984	0.967	0.916	1.012	0.985	1.000
Std. Dev	0.396	0.393	0.360	0.450	0.416	0.252
Skewness	0.947	0.986	0.761	0.898	0.834	0.228
Kurtosis	2.031	2.887	1.132	2.024	0.442	-0.258

Note: Std Dev is the standard deviation for the return series, measured as: $(\sum_{t=1}^{n}(x_t-\bar{x})^2/(n-1))^{1/2}$, where

x_t is the price return series and n is the number of observations.

$$Kurtosis = \left\{\frac{n}{(n-1)(n-2)(n-3)}\sum_{j=1}^{n}((x_j-\bar{x})/s)^4\right\} - \frac{3(n-1)^2}{(n-2)(n-3)}.$$

$$Skewness = \frac{n}{(n-1)(n-2)}\sum_{j=1}^{n}((x_j-\bar{x})/s)^3 \text{, where s is the sample standard deviation.}$$

Table 4.9: Brent Oil Futures Contract Summary Statistics for Detrended Volume

	Sample					
	1992-96	1992	1993	1994	1995	1996
Mean	0.992	0.965	1.032	0.988	0.981	0.998
Std. Dev	0.291	0.238	0.334	0.271	0.333	0.251
Skewness	0.682	0.331	0.681	0.197	0.890	0.797
Kurtosis	1.870	1.439	0.902	0.255	2.527	2.267

Note: Std Dev is the standard deviation for the return series, measured as: $(\sum_{t=1}^{n}(x_t - \bar{x})^2 /(n-1))^{1/2}$, where

x_t is the price return series and n is the number of observations.

$$Kurtosis = \left\{ \frac{n}{(n-1)(n-2)(n-3)} \sum_{j=1}^{n} ((x_j - \bar{x})/s)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)} .$$

$$Skewness = \frac{n}{(n-1)(n-2)} \sum_{j=1}^{n} ((x_j - \bar{x})/s)^3 , \text{ where s is the sample standard deviation.}$$

In the tables below it is quite apparent that there is support for Harris' assertion since the autocorrelation coefficients for volume are greater in general than those for the squared returns series, particularly up to lags of fifteen for the FTSE 100 and Long Gilts, and for lags up to eight for Brent Oil.

In this section the return and volume series for each contract have been analysed in some detail. This has allowed the identification of some important characteristics of the data that the MDH must be able to explain if it is to provide a good explanation of the relationship between volume and volatility. It has also allowed the checking of the data for 'bad' observations and to identify and account for trends that may have an important impact on the interpretation of the results at the next stage. Therefore, having discussed and tested the data it is possible to carry out the direct test of the MDH.

Table 4.10: FTSE 100 Futures Contract Autocorrelations for the Detrended Volume series (V_t)

$\text{Corr}(V_t, V_{t-j})$									
j	1	2	3	4	5	6	7	8	9	10
	0.31	0.18	0.16	0.13	0.18	0.13	0.10	0.11	0.13	0.14
j	11	12	13	14	15	16	17	18	19	20
	0.07	0.10	0.11	0.08	0.14	0.12	0.10	0.09	0.07	0.09
j	21	22	23	24	25	26	27	28	29	30
	0.07	0.01	0.06	0.03	0.05	0.02	0.02	-0.04	-0.05	-0.00

Table 4.11: FTSE 100 Futures Contract Autocorrelations for the Squared Return Series (R_t^2)

$\text{Corr}(R_t^2, R_{t-j}^2)$									
j	1	2	3	4	5	6	7	8	9	10
	0.12	0.08	0.05	0.02	0.02	0.07	0.09	0.07	0.03	0.08
j	11	12	13	14	15	16	17	18	19	20
	0.07	0.11	0.06	0.06	0.06	0.05	0.03	0.05	0.06	0.08
j	21	22	23	24	25	26	27	28	29	30
	0.14	0.06	0.05	0.01	0.01	0.02	0.06	0.06	0.02	-0.00

Table 4.12: Long Gilts Futures Contract Autocorrelations for the Detrended Volume Series (V_t)

$Corr(V_t, V_{t-j})$									
j	1	2	3	4	5	6	7	8	9	10
	0.40	0.19	0.13	0.14	0.26	0.17	0.09	0.11	0.14	0.22
j	11	12	13	14	15	16	17	18	19	20
	0.07	0.02	0.07	0.11	0.14	0.04	0.01	0.09	0.10	0.14
j	21	22	23	24	25	26	27	28	29	30
	0.02	-0.01	0.03	0.04	0.02	-0.02	-0.04	0.00	0.01	0.05

Table 4.13: Long Gilts Futures Contract Autocorrelations for the Squared Return Series (R_t^2)

$Corr(R_t^2, R_{t-j}^2)$									
j	1	2	3	4	5	6	7	8	9	10
	0.05	0.05	0.10	0.02	0.09	0.05	0.07	0.02	0.04	0.05
j	11	12	13	14	15	16	17	18	19	20
	0.08	0.03	0.08	0.06	0.02	0.03	0.01	0.01	0.08	0.08
j	21	22	23	24	25	26	27	28	29	30
	0.04	0.03	0.08	0.05	0.02	0.03	0.02	0.02	0.01	0.04

Table 4.14: Brent Oil Futures Contract Autocorrelations for the Detrended Volume Series (V_t)

$Corr(V_t, V_{t-j})$										
j	1	2	3	4	5	6	7	8	9	10
	0.28	0.17	0.17	0.12	0.08	0.06	0.07	-0.04	0.00	0.02
j	11	12	13	14	15	16	17	18	19	20
	-0.01	-0.06	-0.02	-0.06	-0.04	-0.05	-0.02	0.02	0.04	-0.00
j	21	22	23	24	25	26	27	28	29	30
	0.01	-0.03	0.02	0.03	0.03	-0.00	0.00	0.02	-0.02	0.04

Table 4.15: Brent Oil Futures Contract Autocorrelations for the Squared Return Series (R_t^2).

$Corr(R_t^2, R_{t-j}^2)$										
j	1	2	3	4	5	6	7	8	9	10
	0.05	0.09	0.05	0.25	0.07	0.04	0.08	0.08	0.03	0.04
j	11	12	13	14	15	16	17	18	19	20
	0.02	0.05	0.05	0.08	-0.02	0.03	0.06	0.06	0.08	0.03
j	21	22	23	24	25	26	27	28	29	30
	0.04	0.12	-0.01	0.07	0.06	0.13	0.07	0.05	0.05	0.06

4.5.2 TESTING THE MIXTURE OF DISTRIBUTIONS HYPOTHESIS

As has been noted above, if the MDH holds it must be able to explain certain characteristics of the data. Anderson (1996) develops twelve equations each designed to address a different implication of the model. This system of unconditional return, volume, and cross moments is given below:

$$E[R_t] = \bar{r} \tag{4.57}$$

$$E|R_t - \bar{r}| = (2/\pi)^{1/2} E[K_t^{1/2}] \tag{4.58}$$

$$E[(R_t - \bar{r})^2] = E[K_t] = \overline{K} \tag{4.59}$$

$$E|R_t - \bar{r}|^3 = 2(2/\pi)^{1/2} E[K_t^{3/2}] \tag{4.60}$$

$$E[(R_t - \bar{r})^4] = 3E[\overline{K}^2 + var(K_t)] \tag{4.61}$$

$$E[\hat{V}_t] = c(m_0 + m_1 \overline{K}) = \overline{V} \tag{4.62}$$

$$E[(\hat{V}_t - \overline{V})^2] = c\overline{V} + c^2 m_1^2 var(K_t) \tag{4.63}$$

$$E[(\hat{V}_t - \overline{V})^3] = c^2 \overline{V} + 3c^3 m_1^2 var(K_t) + c^3 m_1^3 E[K_t - \overline{K}]^3 \tag{4.64}$$

$$E[R_t \hat{V}_t] = \bar{r}\overline{V} \tag{4.65}$$

$$E[|R_t - \bar{r}|(\hat{V}_t - \overline{V})] = c(2/\pi)^{1/2} m_1 (E[K_t^{3/2}] - E[K_t^{1/2}]) \tag{4.66}$$

$$E[(R_t - \bar{r})^2 \hat{V}_t] = \overline{V}\overline{K} + m_1 var(K_t) \tag{4.67}$$

$$E[(R_t - \bar{r})^2 (\hat{V}_t - \overline{V})^2] = c\overline{K}\overline{V} + c^2 m_1 var(K_t) + c^2 m_1^2 [E[K_t - \overline{K}]^3 - \overline{K} var(K_t)] \tag{4.68}$$

where the two observable series are R_t, the returns series, and $\hat{V}_t$, the detrended volume series. K_t represents the information intensity variable. $\bar{r}$ is a constant designed to allow for the possibility of a mean return that is non-zero. m_0 and m_1 are the noise and informed components of volume respectively. The positive constant, c, as explained in section 4.2, is added to the model because the parameters in the model are not invariant to the detrending that has been carried out.

It is worth at this stage explaining each equation and the part of the mixtures process that it relates to.

$E[R_t] = \bar{r}$; as explained above, this allows for the possibility of a nonzero mean return.

$E|R_t - \bar{r}| = (2/\pi)^{1/2} E[K_t^{1/2}]$; since the conditional return is normally distributed, the expected return has this form dependent on information intensity. The expectation is that this will be close to zero in an efficient market.

$E[(R_t - \bar{r})^2] = E[K_t] = \bar{K}$; the variance of returns is assumed to be dependent on the intensity of information arrivals. This forms the basis of the subordinated process argument.

$E|R_t - \bar{r}|^3 = 2(2/\pi)^{1/2} E[K_t^{3/2}]$; this is the skewness equation. The expectation is that returns will be slightly skewed to the right. This effect is dependent on information intensity. The right skewness occurs because the distribution reflects average return centred on zero and larger returns which are less common. Most information arrivals do not result in great return opportunities particularly when measured at the daily frequency.

$E[(R_t - \bar{r})^4] = 3E[\bar{K}^2 + var(K_t)]$; this is the kurtosis equation. This is one of the most observed characteristics of the MDH. Under the MDH, returns driven by information will be leptokurtic.

$E[\hat{V}_t] = c(m_0 + m_1 \bar{K}) = \bar{V}$; under Anderson's modified specification of the MDH, volume is driven by noise and informed trading.

$E[(\hat{V}_t - \bar{V})^2] = c\bar{V} + c^2 m_1^2 var(K_t)$; the variance of volume has a common component and a component driven by the variance of information intensity. The variance of volume is important to our predictions of the autocorrelation of the observed volume

133

series. The autocorrelation of the volume series should be greater than that of the squared returns series, because the fraction of the variance due to variance in information intensity is greatest for volume.

$$E\left[\left(\hat{V}_t - \overline{V}\right)^3\right] = c^2\overline{V} + 3c^3m_1^2\,\mathrm{var}(K_t) + c^3\,m_1^3\,E\left[K_t - \overline{K}\right]^3\,;$$ the skewness of volume is very much dependent on the information process. The expectation is that volume is positively skewed but to a greater extent than returns. The skewness occurs because both the mean and variance of volume are dependent on the information process. Harris (1987) argues that the difference between the skewness of volume and the skewness of returns occurs because for price changes the mean is small relative to the variance.

$$E\left[R_t\hat{V}_t\right] = \overline{r}\overline{V}\,;$$ this represents the cross moment between return and volume. The covariance between return and volume is expected to be very weak (if not zero). This can be explained using the Tauchen and Pitts (1983) model. In section 4.2 it was shown how the change in price and volume can be represented as a variance components model. The change in price has a common and a mean specific component. Volume is represented by the deviation from the mean of the specific component. The variation in the mean of the specific component will be small relative to variation about the mean and, therefore, the relation between volume and price changes should be very small.

$$E\left[|R_t - \overline{r}|\left(\hat{V}_t - \overline{V}\right)\right] = c(2/\pi)^{1/2}m_1\left(E\left[K_t^{3/2}\right] - E\left[K_t^{1/2}\right]\right)\,;$$ this represents the cross moment of the deviation of return from its mean and volume from its mean, which is a function of information intensity only. The expectation is that this will be positive, although it could potentially be quite small, particularly if the deviation of returns from its mean is, as we would expect, quite small.

$$E\left[\left(R_t - \overline{r}\right)^2\hat{V}_t\right] = \overline{V}\overline{K} + m_1\,\mathrm{var}(K_t)\,;$$ this is key. The relationship between return variance and volume exists because variance and volume are both related to the underlying information process. This specification is slightly different from that of

Tauchen and Pitts (1983), who argue that the relationship vanishes if there is no variation in the information flow. In this equation both the informed and noise components of volume have an impact on the relationship that is expected to be positive.

$$E\left[(R_t - \bar{r})^2 (\hat{V}_t - \overline{V})^2\right] = c\overline{KV} + c^2 m_1 \, var(K_t) + c^2 m_1^2 \left[E\left[K_t - \overline{K}\right]^3 - \overline{K} \, var(K_t)\right]; \quad \text{the}$$

covariance of return variance and volume variance is a function of information intensity and mean volume. It is expected to be positive. The expectation is that the variance of volume will be greater than that of squared returns, but given their dependence on a common mixing variable the correlation should be strong.

The five return moments, 3 volume moments, and four cross moments help in the testing of the MDH. As noted above, between them they represent the important observed characteristics of return and volume and the relationship between them. The MDH implies that each of the observed characteristics can be explained by the information process. From these equations it is possible to form orthogonality conditions as shown in section 4.4. The theory does not specify any lagged volume and return relationships and so the orthogonality conditions are created using a constant as the only instrument. These orthogonality conditions can then be estimated using GMM. Since only volume and returns are directly observable there are nine so-called free parameters. Together they form the parameter vector given by:

$$\left(\bar{r}, E\left[K_t^{1/2}\right], \overline{K}, E\left[K_t^{3/2}\right], var(K_t), E\left[K_t - \overline{K}\right]^3, m_0, m_1, c\right).$$

With nine free parameters and twelve orthogonality conditions there are three over-identifying restrictions. This allows the use of the Hansen (1982) test (see section 4.4) with a distribution of χ_3^2. If the twelve equations represent the MDH and the test statistic is above the critical value then the MDH, as described by the orthogonality conditions, does not hold. Conversely, if the test statistic is less than the critical value then the MDH, as described by the orthogonality conditions above, does hold.

As well as investigating whether or not the MDH holds it is possible to say something about the underlying information process for each contract by looking at the point

estimates of the information intensity parameters. $\overline{K}$ reflects the average daily information intensity across the sample. As with the other components of the information process, the expectation is that this will be positive. Information intensity may be high or low but it cannot be negative. $Var(K)_t$ is an indicator of whether there is a lot of variation in the level of information intensity. $\overline{K}$ and $var(K_t)$ together can indicate whether news comes in on a regular basis or whether news tends to be more unpredictable with some days when information intensity is high and other days when the intensity is low. The other moments of the information process, $E[K_t^{1/2}]$ and $E[K_t^{3/2}]$ are expected to be positive. The size of $E[K_t - \overline{K}]^3$ depends on the variance of information intensity. If the variance is small relative to the mean this will also be a small value and vice versa. $\overline{r}$ is expected to be close to zero reflecting the lack of profit opportunities in the market.

It is also possible to say something about the relative impacts of the noise and informed components of volume. If those who argue that futures trading is little more than sophisticated gambling are to be believed, we might expect the noise component of volume to be the largest of the two. By looking at the point estimates for m_1 and m_0 for each contract, it is possible to identify the types of trader operating in each market. We can therefore determine whether the relative values are in line with our earlier expectations.

The estimation of the model represented by the twelve equations above is not an easy task. The biggest problem is the large number of point estimates that are required. Convergence is very much dependent on choosing the right starting values. The econometric package used here is TSP[6]. It gives an indication of which point estimates are furthest away from their starting values. The approach exploited here is therefore to restrict those parameters whose point estimates are varying the most. The model is then re-run to find the best values for the remaining parameters before the 'trouble' parameters are put back into the estimation.

---

[6] The programs are available on request.

This is combined with estimating the system in the form of smaller subsets of equations. The weighting matrix and the parameter estimates are then iterated until convergence. The values given in tables 4.16, 4.17 and 4.18 reflect final parameter estimates.

Another difficulty in testing a model of this type is ensuring that the covariance matrix and therefore the weighting matrix have been properly estimated. It is crucial that the estimation of the covariance matrix is adjusted for possible heteroscedasticity and serial correlation between the error terms. The common approach is to select a number of lags which are then weighted by a kernel density estimator to guarantee that the covariance matrix is positive semi-definite. Andrews (1991) provides some guidance in this area and compares a number of different estimators as well as determining an optimal lag structure dependent on the sample size. The software used to estimate this matrix has two options. The Bartlett heteroscedasticity and autocorrelation consistent (HAC) estimator favoured by Newey and West (1987) is shown by Andrews (1991) to be the least effective of a group of kernel HAC estimators. We instead chose the Parzen kernel estimator.

Tables 4.16, 4.17 and 4.18 show the results of the estimation of our twelve equation system using GMM. They include the point estimates for the nine free parameters as well as their standard errors.

Hansen's (1982) test of overidentifying restrictions has a test statistic of 10.251 for the FTSE 100 futures contract. The critical value of χ_3^2 at the one percent level is 11.34. Since the test statistic is less than the critical value we can say that at the ninety-nine percent confidence level the twelve moment equations above implied by MDH can explain the characteristics of the data. The statistics for the Long Gilts and Brent Oil futures contracts are 9.998 and 9.430 respectively. Therefore in all three markets the MDH does hold, i.e. information is the driving force behind the moments of volume and return and their cross moments. This is the result that Anderson (1996) finds for his selection of stocks quoted on the NYSE.

Table 4.16: GMM estimation of Anderson's specification of the MDH for the FTSE
100 Futures Contract (1992-1996)

Parameter	Point Estimates	Standard Error
$\bar{r}$	0.392E-03	0.583E-03
$E\left[K_t^{1/2}\right]$	0.812E-02	0.138E-02
$\overline{K}$	0.948E-04	0.301E-04
$E\left[K_t^{3/2}\right]$	0.175E-02	0.223E-03
$\text{var}\left[K_t\right]$	0.788E-05	0.324E-07
$E[K_t-\overline{K}]^3$	0.769E-02	0.224E-02
m_0	5.838	1.503
m_1	69.804	2.275
C	0.050	0.060

Note: The value of the test statistic for Hansen's (1982) test of overidentifying restrictions is 10.251.
Under the null hypothesis that the moment equations are satisfactory the test statistic has a chi-square
distribution with three degrees of freedom.

As well as this very important result we can comment a little further by looking at the
point estimates. With one important exception, in the majority of cases the small
standard errors relative to the size of the point estimates suggests that they have been
accurately measured. Given the problems of estimating higher order moments, this is
a pleasant surprise. Before discussing the point estimates it is important to bear in
mind that their significance is not tested explicitly. The ideal procedure would be to
set up significance tests for each variable, run a restricted model for each test and then
compare the restricted and unrestricted models using a likelihood ratio test. However,
given the estimation problems involved this was considered to be impractical.

Table 4.17: GMM estimation of Anderson's specification of the MDH for the Long Gilt Futures Contract (1992-1996)

Parameter	Point Estimates	Standard Error
$\bar{r}$	0.709E-04	0.552E-04
$E\left[K_t^{1/2}\right]$	0.500E-02	0.194E-03
$\bar{K}$	0.300E-04	0.245E-05
$E\left[K_t^{3/2}\right]$	0.200E-04	0.206E-06
$var[K_t]$	0.250E-06	0.664E-11
$E[K_t-\bar{K}]^3$	0.799E-02	0.697E-03
m_0	0.090	0.835
m_1	350.068	10.289
c	0.015	0.101E-05

Note: The value of the test statistic for Hansen's (1982) test of overidentifying restrictions is 9.998. Under the null hypothesis that the moment equations are satisfactory the test statistic has a chi-square distribution with three degrees of freedom.

Let us consider each of the point estimates in turn:

$\bar{r}$; the mean return is positive but very small in all cases. For the Long Gilt contract the standard error is actually quite large relative to the point estimate value. This may suggest that the assumption of a non-zero return does not hold in this market. At the daily frequency the expectation of large returns is small. In liquid futures markets, like the three considered here, profit opportunities may only last a matter of minutes.

$E\left[K_t^{1/2}\right]$ and $E\left[K_t^{3/2}\right]$; the expected square root and the expected cube root of daily information intensity are both positive as expected. The relative magnitudes of these two moments in all cases are also in line with expectations.

Table 4.18: GMM estimation of Anderson's specification of the MDH for the Brent Oil Futures Contract (1992-1996)

Parameter	Point Estimates	Standard Error
$\bar{r}$	0.398E-03	0.772E-04
$E[K_t^{1/2}]$	0.825E-02	0.998E-03
$\bar{K}$	0.992E-04	0.320E-04
$E[K_t^{3/2}]$	0.175E-02	0.123E-04
$var[K_t]$	0.559E-05	0.915E-10
$E[K_t-\bar{K}]^3$	0.828E-02	0.469E-02
m_0	1.471	0.656
m_1	35.377	6.190
c	0.034	0.015

Note: The value of the test statistic for Hansen's (1982) test of overidentifying restrictions is 9.430. Under the null hypothesis that the moment equations are satisfactory the test statistic has a chi-square distribution with three degrees of freedom.

$\bar{K}$; the mean information intensity is positive in all cases as expected. The small values suggest that on the whole information intensity is low. Mean information intensity is lowest in the Long Gilts market. This is not a surprise given that the information likely to have the greatest impact on the futures price, for example, government spending figures, interest rate changes, etc., arrives regularly but infrequently relative to other contracts. This suggests quite a strict structure to the information process. The general implication of these results, however is that, given that the MDH holds, and that the link between the volume of trade and price volatility is strong, driven by the underlying information process, although information may arrive infrequently, its impact is significant. This is what we would expect in a market driven by a subordinated process. If no new information is coming into the market trading will be relatively stable. Prices reflect information available in the market. If there is no news, prices will not move away from equilibrium. We have yet to discuss the model's implications for the components of volume, but even noise traders are simply reacting to traders who initiate trading by acting on a piece of news that they

believe offers them an advantage over other market agents. While news may result in little volume if there is common interpretation of its implications, it is information that ultimately drives the market away from equilibrium.

$Var(K_t)$; the variance of daily information intensity for each contract is positive as expected. The point estimates of the variance are also small relative to the mean. The implication is that all three markets are used to regular information flows. This may be attributed to periodic macroeconomic announcements, or news from firms that reveal company account details at regular intervals. This is even more relevant to the Long Gilt market and supports the comments made above.

$E[K_t - \overline{K}]^3$; this is positive for all contracts, in line with the initial expectations. In each case the values are quite small indicating that information intensity does not deviate far from its mean.

It is also important to note the similar magnitudes for the information coefficients of the FTSE 100 and Brent Oil contracts. This suggests that there are close links between the two markets. A large number of the companies that make up the FTSE 100 also have close links with Brent Oil. This is because they are either oil companies, energy producers or companies for whom the price of oil will have an impact on production costs. In fact, the Brent Oil futures contract is the benchmark by which two-thirds of the world's internationally traded crude oil supplies are priced. Its impact is therefore widespread.

c; reflects the adjustment made to the volume specification as a result of the detrending process. In each case the observed volume will be smaller than the stationary volume specified in the theoretical model.

m_0 and m_1; the point estimates relating to the informed and noise components of volume allow some interesting observations to be made. The first is that for both the FTSE 100 and Brent Oil contracts the informed component is much greater than the noise component. For the Long Gilt contract the informed component far exceeds the

uninformed component. In fact the size of the standard error relative to the point estimate suggests that noise trading may not be an important factor. In all three markets trading volume appears to be driven primarily by informed agents. Those who feel that the pieces of news that they hold offer an exploitable opportunity outweigh those reacting to these news induced movements. This is in contrast to Anderson (1996) who finds that the noise component of volume tends to outweigh the informed component. These results also have interesting policy implications. Critics of futures markets argue that the impact of noise traders in futures markets is primarily one of destabilisation. The argument is often that the particular characteristics of futures markets; specified delivery dates, a narrowly defined deliverable commodity, etc., create an environment conducive to destabilising activity. The results above would seem to show that, in fact, the impact of noise trading is very small and for Long Gilts virtually non-existent. This result might be expected, particularly given the way that the data has been treated. Initial analysis of the data revealed significant peaks in the autocorrelation function of volume at the same time as contract expiration. These peaks were identified as periods when a large proportion of traders roll-over their positions from the nearest to the next nearest contract. Using the Holmes-Rougier (1997) roll-over adjustment, this non-information based trading volume has been eliminated, leaving just the informed component. It could be argued that if a large proportion of the trading occurs in the expiration month, as appeared to be the case from looking at the data, and the noise component in the model picks up all non-information based trading, then m_0 will be small relative to m_1[7]. This highlights one of the criticisms that can be made of the model. It is assumed that the parameters m_0 and m_1 are constant over the sample period. The point estimates considered here are, therefore, essentially considering average values for the two components of volume. Thus, the point estimates fail to provide any indication of the

---

[7]Estimation of the twelve orthogonality conditions using GMM without detrending the data for the FTSE 100 was tried and it proved very difficult to get the model to converge. It is difficult to assess, therefore, the impact of the Holmes-Rougier (1997) adjustment. The lack of convergence is not surprising given our belief that trends in the data can obscure the underlying relationship between volume and squared returns. In contrast data for the Long Gilt contract did converge (with little difference in coefficient values), but the effect of roll-over is small in this market. The implication is that there is noise in the market separate from roll-over effects.

change in balance between noise and information based trading that may occur across the sample.

4.5.3 WHAT IS THE IMPORTANCE OF THESE RESULTS?

It has been shown that for the FTSE 100, the Long Gilt and the Brent Oil futures contracts the observed characteristics of the data can be explained by the Mixture of Distributions Hypothesis. This supports the findings of much of the empirical work based on anecdotal evidence. It is in contrast, however, to the results of attempts at a direct testing procedure carried out by Lamoreux and Lastrapes (1994) and Richardson and Smith (1994). This may be due to a number of differences between this study and theirs; they used different specifications of the MDH, they considered spot data for which the MDH genuinely may not hold, or they may have failed to adjust for trends in the data.

Therefore, for the contracts considered here, the link between the volume of trade and return volatility can be attributed to the flow of information. The movement of prices and the market activity of traders are both driven by the same underlying process. This is the first time that this link has been confirmed for UK futures markets.

It has also been possible to identify some of the characteristics of the information process. It has been shown that, contrary to popular belief, volume is dominated by informed rather than noise trading. However, the fact that noise trading is so low looks like a slightly odd result. Noise trading is often assumed to facilitate investment. Do we need to redefine our definition of noise? Anderson argues that noise traders arrive at the market at a constant rate. Maybe, noise traders are more discriminating. They may be uninformed only in the sense that the information that they hold has already been revealed in prices. This suggests closer links with information arrival. The trading model underlying Anderson's (1996) specification of the MDH implies that they are uninformed only because their information set is smaller than that of the specialist. It is possible that m_1 is in fact capturing volume associated with information rather than informed trading *per se*.

Anderson's (1996) model as tested here is essentially static, particularly with regard to the information process. In the same paper, Anderson develops an interesting dynamic model of the flow of information that involves adding a GARCH-type specification of the information intensity variable. One problem with this model, however, is the seemingly arbitrary choice of the number of lags on key variables that do not appear to be justified by any theory. There is clearly work to be done in this area.

These results also need to be considered in terms of their impact on current research issues and from the point of policy objectives. These results are particularly supportive of moves within the market microstructure literature away from inventory models to those that consider the impact of information. The traditional view assumes that specialists, or market-makers, set prices based on exogenous parameters, balancing risks over time. If market-makers can actually learn from trades it suggests that prices are not independent of information. One particular field of research is that involving strategic trading issues. The results of this study, that a large proportion of investors are informed, presents an interesting problem. The likelihood is that prices will reveal information much more quickly. However, if prices are revealing, the incentive to collect costly information is reduced. Holden and Subrahyman (1992) argue that where there are a large number of informed investors market depth is low. This is not what is observed in practice. The FTSE 100 contract, in particular, is able to absorb large quantities of trading without significant price changes. How can these apparent contradictions be resolved?

O'Hara (1997) argues that the key is to move away from competitive to strategic trading. If investors know that their 'private' information will be quickly revealed in prices they will trade more carefully. This suggests an imperfect competitive equilibrium. Prices then become less revealing and there is an incentive to obtain information because of the possibility of returns.

These results also have implications regarding the work of Blume et al. (1994). Their model of the relationship between volume and price changes is essentially information driven. This study supports that conjecture. It also supports our own results in

144

chapter 3 that information is very dispersed due to the fact that there are more informed than uninformed investors.

Another area of interest is the impact these results have in terms of the supposed destabilising nature of trading in derivative markets. We have established that the link between volume and volatility is driven by information. In addition, the investors in this model are primarily motivated by acting on information. This is crucially important, suggesting that artificial restrictions on price movements or the volume of trade could have very harmful effects on the successful operation of the UK futures contracts considered in this study. Reductions in liquidity make it harder for investors to meet their risk requirements. Market-makers, in particular, are likely to pass on the costs of holding unwanted inventory through higher transaction costs in terms of bid and ask prices.

The real achievement of this study, therefore, has been to provide an explanation of why the link between volume and price volatility occurs based on a comprehensive test that directly models the subordinated process. Thus, it has been possible to obtain and to discuss in detail a result that distinguishes this work from the vast majority of previous empirical studies.

4.6 CONCLUSION

In this chapter an extensive investigation has been carried out into one of the most important aspects of trading volume - its relationship with price variability. This builds on the ideas developed from the work in chapter 2. The theoretical models that seek to explain this relation, and the empirical studies that have tested them, have been considered in detail. What becomes apparent is the paucity of direct tests of any of the explanatory models. Supportive evidence, though convincing, is largely anecdotal.

The achievement of this chapter has been the selection of the most theoretically and intuitively appealing model, the Mixture of Distributions Hypothesis, and a direct test of its validity. This has been achieved by initially taking great care with the

construction of each price return and trading volume series, and then by exploiting the GMM methodology.

The estimation was difficult but has produced some very exciting results. This is the first time that such a test has been carried out for a UK futures market and it has been possible to show that the MDH holds. The methodology also allowed the description, in some detail, of the information process that drives the volume-volatility relation, and the composition of daily trading volume.

This study is not without its faults. Anderson's modified form of the MDH relies on certain restrictive assumptions and there has been no discussion of the dynamic nature of the information process. Ultimately, however, the achievement is the combination of a theory that empiricists believe to be true with an estimation technique that demonstrates its validity. This result has important policy implications particularly with regard to the discovery that the three markets investigated here are dominated by informed trading. If intervention is prompted by market failure then the assumption must be that market agents do not know how to use their information properly. Regulatory bodies need to be careful that their actions are not due to perceptions of allocative inefficiency rather than based on sound economic fact.

In terms of further research, one particularly interesting issue that deserves further investigation is the role of volume in determining the costs of trading. The Glosten and Milgrom (1985) model that underpins Anderson's (1996) specification of the MDH is based around the concept of a market-maker whose information set changes with each new trade. We have discovered that the volume of trade is dominated by informed investors. How will this affect the bid and ask prices set by the market-maker? The issue of volume and the cost of carrying out transactions in a futures market will be considered in the next chapter.

CHAPTER FIVE: AN INVESTIGATION INTO THE RELATIONSHIP BETWEEN THE BID-ASK SPREAD AND THE VOLUME OF TRADE

5.1 INTRODUCTION

In chapter 4 the relationship between price return volatility and the volume of trade was investigated by carrying out a direct test of the Mixture of Distributions Hypothesis. This was a development of the discovery in chapters 2 and 3 that volume plays an important role in derivative markets consistent with the supposition that it acts as a bearer of information. Indeed, the results in support of the MDH from chapter 4 suggest that it is information that dominates the volume-volatility relation.

The specification of the MDH that was tested in chapter 4 is based on the trading model of Glosten and Milgrom (1985). It suggests that the setting of bid and ask prices by specialists, or market-makers, is partly determined by the actions of investors who arrive at the market. Their decisions to buy or sell quantities of the asset convey information to the market-maker, who adjusts prices accordingly.

This raises some interesting issues. If volume is dominated by informed investors, as suggested by chapter 4, how does this affect the setting of prices by those less well informed? Is there a danger that prices might be set that deter trading and, therefore, reduce the price discovery and liquidity roles of a derivatives market?

The aim of this chapter, therefore, is to carry out a detailed investigation into the relationship between the volume of trade and the determination of the bid-ask spread. This represents the difference between the lowest available quote to sell the asset (the ask price) and the highest available quote to buy the asset (the bid price). Thus, an investor attempting a so-called 'round-trip' exchange by buying and selling the asset,

immediately incurs a cost due to the spread. This cost is the price that a market-maker charges for the service of providing this immediacy[1].

The first part of this chapter looks in detail at the different models of the spread; those related to inventory costs and those related to information costs. Essentially the conflict between them is that while there are advantages to the market-maker of operating in a market where the frequency of transactions is high there are also potential risks. The advantage of high transaction frequency is that the risk of holding unwanted assets for a long period is reduced. At the same time the market-maker is exposed to the risk of being exploited by investors holding superior information. This study aims to resolve this conflict by assessing the relative costs and benefits of each scenario. As mentioned above, this is a particularly interesting issue bearing in mind the results from chapter 4 which indicated that trading in UK futures markets is dominated by informed investors. Section 5.2 also considers the theoretical work related to intra-day trading patterns. The empirical work in this field, particularly with regard to futures trading is very limited. This study therefore aims to extend our understanding of the behaviour of derivatives markets by considering the relationship between volume and the spread using high frequency, transactions data.

An analysis of the theoretical literature reveals that the emphasis is placed on describing behaviour in equity markets. Futures markets have their own idiosyncrasies that distinguish them from other markets. Section 5.2, therefore, also examines how the theory can be related to the operation of futures markets. Section 5.2 concludes by discussing in more detail the important issues that this study will address.

Section 5.3 of this chapter takes a comprehensive look at the various empirical studies that have investigated aspects of the bid-ask spread. What becomes apparent is that very few carry out any detailed analysis of the relationship with the volume of trade or consider any of these issues in the context of futures markets.

---

[1] Note that the investors also incur costs due to brokerage fees, search costs, etc.

A number of the theoretical and empirical studies that look at intra-day patterns in trading suggest that the incidence of volume is to some extent predictable. One of the aims of this study is to investigate this issue further by not only considering the relationship between volume and the spread, but also by looking at how different components of volume, specifically expected and unexpected trading, affect the market. Section 5.4 of this chapter looks in detail at the methodological background to the generation of these two series.

Another weakness of the empirical work is that it tends to treat the relationship between volume and the spread as unidirectional. The impact of volume on the spread is considered without an appreciation of the fact that the spread is also likely to be a determinant of volume. The aim of this study is to use a regression technique to describe this expected inter-dependency. This issue of simultaneity requires the use of an alternative estimation technique to the more conventional method of Ordinary Least Squares. Section 5.4 looks in detail at the estimation of simultaneous models and also considers the different methods of calculating the spread in a market where bid and ask quotes are non-binding.

The empirical section of this chapter, section 5.5, presents the results of the estimation of the regression model used to investigate the relationship between volume and the spread for two financial futures contracts traded on LIFFE; the FTSE 100 and the Long Gilt. The specification of this model is discussed at some length as well as how the data was constructed. There is also some preliminary discussion of the variables and their variation across the trading day. The results allow us to resolve some very important issues that have significant implications for both market-makers and regulators, as well as highlighting possibilities for future research. Section 5.6 concludes.

5.2 THEORETICAL BACKGROUND

This section presents the two main classes of theories of the bid-ask spread. The first of these addresses the important role of the volume of trade in reducing the risk that a market-maker incurs in holding outstanding assets. These are more usually known as the inventory cost models. The second group of theories of the bid-ask spread, the

information cost models, considers the impact of the volume of trade in terms of the probability that some investors will hold better information than those setting the prices.

This section also looks at the theories that attempt to hypothesise why the volume of trade, as well as volatility and the cost of trading, might vary during the period when the market is open.

The majority of the theoretical work in this area centres on the microstructure of equity markets. The translation of these models to futures markets requires some appreciation of the idiosyncratic nature of derivatives trading. This is addressed in the third part of this section.

Finally, this section discusses some of the issues that arise out of the theoretical work that will be investigated further in this study.

5.2.1 INVENTORY COST MODELS OF THE BID-ASK SPREAD

The seminal paper on the modelling of the bid-ask spread is that of Demsetz (1968). He presents a static model of the spread as one part of the cost of transacting in a market. The other major cost is represented by brokerage fees. He argues that the spread can be considered as the cost of immediacy. If an individual approaches a market to either buy or sell shares, it is purely by chance that another individual will arrive at the same time to take the other side of that trade. Therefore, to ensure the demand for immediacy is met, specialists, or market-makers, will complete the trade before reversing their new position at a later stage. This service will only be provided at a cost represented by the spread. Demsetz provides the following neat argument to illustrate the demand and supply of this service.

Consider figure 5.1 where D1 and S1 represent respectively the demand for and supply of immediacy in a market for asset X.

Figure 5.1: The Cost of Immediacy and the Determination of the Bid-Ask Spread
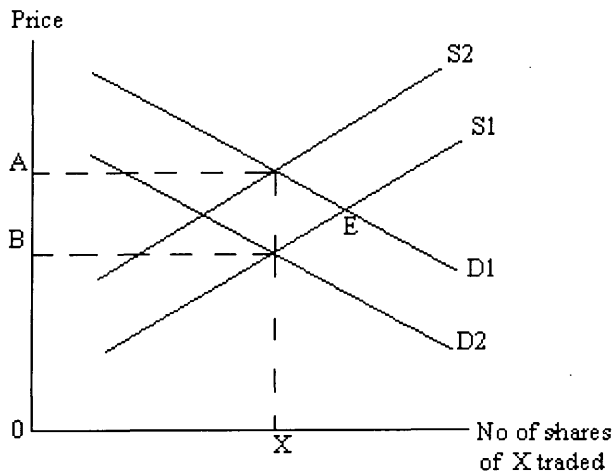
Source: Demsetz (1968)

Demsetz demonstrates that although E can be conventionally considered as an equilibrium, it is more helpful to view it as the average price at which the asset X can be exchanged. This is the price that prevails if exchange can occur immediately.

Assume that a market-maker exists who stands ready to buy or sell at stated prices as soon as an order reaches the market. The cost of standing ready means that the market-maker will only be willing to buy X at a price below E, and sell at a price above E. The difference between the two prices represents the bid-ask spread.

In figure 5.1 if S1 represents the supply curve of those who wish to sell immediately, S2 represents the supply curve of those willing to wait in order to keep their orders active. S2 lies above S1 to cover the cost of waiting. The ask price, A, is therefore represented by the intersection of D1 and S2.

A similar argument is used to establish the bid price B. The demand of those who are willing to wait to buy shares will be slightly lower than those who wish to buy shares immediately. The difference between E and B represents the cost of providing the service of standing ready to buy shares as sell orders reach the market. The difference

151

between A and B is the spread. This is often referred to as the cost of a 'round-trip' exchange. A person who buys an amount of the asset and then wishes to sell it immediately will suffer a loss equal to the size of the spread. It is important to note, however, that the investor also incurs costs due to brokerage fees, etc., every time that a transaction is made.

Demsetz highlights the importance of the market-maker in this process. He makes parallels between the cost of immediacy and the inventory mark-up charged by a retailer or wholesaler. The market-maker's main source of income is trading carried out for a personal account, but there is also the possibility of making a profit from the spread. Demsetz argues, however, that the ability of the market-maker to set a spread above cost depends on the level of competition in the market. This can arise from; rivalry for the specialist's job, other specialists, competing markets, traders who bypass the market-maker and complete trades themselves, and competition provided by those who submit limit orders rather than market orders. Limit orders represent the specific price at which an investor will transact. The last of these factors is quite an interesting one. The bid price and the ask price are effectively the limit orders set by the market-maker. Individuals arriving at the market with limit orders will set them slightly below the current ask and above the current bid price. If no market orders arrive to initiate trading the market maker may be forced to set more competitive prices to ensure that trading takes place.

Demsetz argues, however, that the most important determinant of the spread is waiting costs. If the frequency of market orders is high then any given set of bid and ask prices will be acted on more quickly. Those at the front of the queue of limit orders, i.e. with the most competitive prices, will therefore face low waiting costs. Those wishing to get to the front of this queue must set lower ask and higher bids than those already well placed. The key element here is the time between transactions. If the frequency of transactions is high, the cost of waiting is driven down. Demsetz refers to these as scale economies. He argues that the inverse relationship between the spread and the number of transactions is likely to dominate any increasing marginal costs, due to congestion in the market caused by a large number of market orders arriving in a short space of time.

Stoll (1978) provides a more explicit model based on the inventory cost hypothesis. It focuses on the problem that market makers are forced to carry inventories that differ from their optimal portfolios. Stoll explains the hypothesis using the diagram in figure 5.2. The x and y axes represent the standard deviation of returns and expected returns respectively. The market-maker's efficient frontier, RfE, represents combinations of an efficient portfolio of risky assets, point E, and the risk free asset, with a yield equal to Rf. N is assumed to be the optimum portfolio position for the market-maker. Movement away from N represents non-optimal portfolio positions since the market-maker has to move away from indifference curve U* to a lower indifference curve. Stoll (1978) labels this portfolio as the trading account. The market-maker's portfolio becomes de-diversified by long or short positions in the trading account. The new portfolio consisting of the trading account plus the investment account is described by line A1NB.

Figure 5.2: The Inventory Cost Model

Source: Stoll (1978).

In figure 5.2 the movement along the line from N to A1 represents an undiversified long position financed by borrowing at Rf. (A movement along the line from N to B would represent an undiversified short position.) The total cost to the market-maker of having to hold a non-optimal portfolio is equal to g', which is the amount that

customers have to pay to keep the market-maker on the initial indifference curve U*. This cost accounts for the de-diversification and the risk that the market-maker has to bear while not holding a preferred position. Note that a cost would still be incurred if the market-maker was able to remain on RfE and maintain a diversified portfolio due to being on a lower indifference curve. If the market-maker is at a non-optimal position A1, the cost of another transaction is the difference between the percentage cost at A1 and that at the new position following the transaction. A movement from A1 to A2 would actually lead to a fall in costs, (g'' is less than g'), because the market-maker has been able to increase diversification and reduce risk.

Stoll uses this framework to derive a function to describe explicitly the costs incurred by the market-maker in supplying this service of immediacy. His one period model is extended to a multi-period context by Ho and Stoll (1981). Both models illustrate the importance of return variance and transaction size in determining the spread in terms of increased risk to the market-maker. The multi-period model also demonstrates, in line with Demsetz (1968), how the costs, and therefore the size of the spread, increase the longer the market-maker has to wait between trades.

5.2.2 INFORMATION COST MODELS OF THE BID-ASK SPREAD

One aspect of the cost of providing immediacy touched on by Stoll (1978), but not developed in any great detail, is the cost faced by a market-maker in carrying out transactions with individuals who possess superior information. The first substantial work in this field is attributed to Copeland and Galai (1983) which develops earlier work by Bagehot (1971). They argue that the dealer in a market is faced with two types of trader; those who are informed and those who are uninformed. These uninformed traders are commonly called noise traders. This does not necessarily mean that they do not carry information. If they do hold information it will not have any bearing on price, because the news has already been revealed to the market. Informed traders carry private information that allows them to evaluate the future value of an asset more accurately than the market-maker or the noise traders. The market-maker therefore has to trade-off losses that are incurred from trading with informed traders with gains that can be made by trading with uninformed traders.

In Copeland and Galai's model the bid-ask spread is considered in terms of the dealer's expected costs and revenues. The expected losses to informed traders are a function of the probability that the next trader is informed, P_I, the dealer's knowledge of the underlying process driving price changes, $f(S)$, and on the bid price, B, and the ask price, A, that have been set. It is assumed that the probability that an individual is informed is less than unity. If all traders were informed the market-maker could only lose.

Copeland and Galai develop their model under two different scenarios relating to the time between the quoting of prices by the market-maker and the arrival of a trader. Under the instantaneous quote scenario the market-maker waits until a trader arrives at the market before offering a quote. Under the open quote scenario the market-maker offers the quote immediately and then waits for the arrival of traders. It is open to debate as to which of these is the most realistic, but the predictions in each case are very similar.

This model is illustrated in figure 5.3 by considering only one half of the spread; the ask spread. In this diagram WX represents the market-maker's expected costs from informed trading. YZ represents the losses if all traders are informed. As the spread increases it is clear from the diagram that expected losses to informed traders will fall. The market-maker earns money from those uninformed traders who are willing to accept A-S or S-B (not represented in this diagram), where S is the 'true' price of an asset, as the cost of liquidity. Copeland and Galai derive the market-makers expected revenue curve (OV) by multiplying the unconditional gain per transaction (the 45° line OQ) by the percentage of uninformed traders, P_U, where $P_U = 1-P_I$. This is represented by line OR.

Figure 5.3: The Information Cost Model

£   Q

Y

W

R

V

Z

X

O   (A-S)*   (A-S)**   The Ask Spread A-S
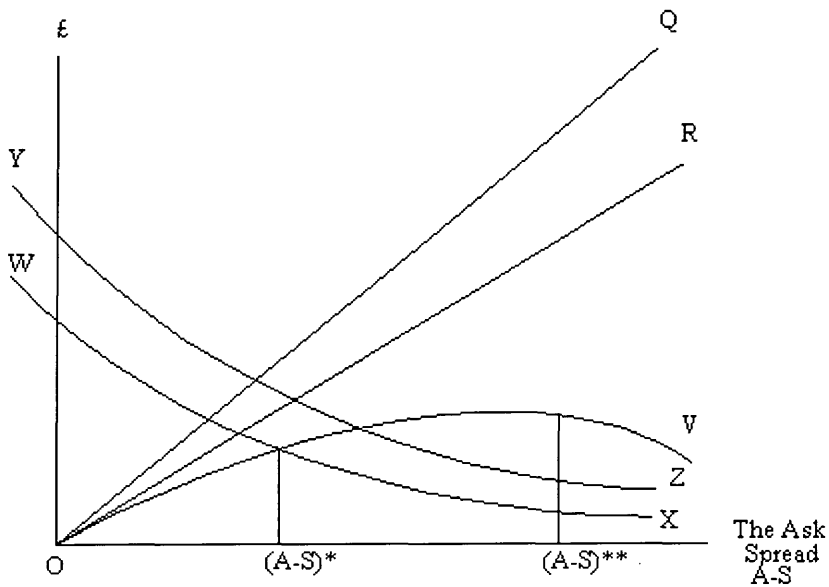
Source: Copeland and Galai (1983).

If the probability that an uniformed trader will buy at the asking price is given by P_{AU}, then this will decline as the spread increases. The revenue line OR multiplied by P_{AU} gives the expected revenue curve which will be concave if P_{AU} decreases monotonically as a function of the asking price.

The aim of the market-maker, assuming risk neutrality, is to set the bid-ask spread to maximise expected profit. If there is only one dealer in the market the ask price will be set at A**, to maximise the difference between the expected revenue and cost functions. In a competitive dealer market the ask price is set at A* where costs and revenue are equal. Therefore, if the percentage of informed traders increases, then the expected dealer costs increase relative to revenues and the ask price increases.

Copeland and Galai (1983) admit that this is a slightly simplified model of the way that a market operates, but it does allow them to show some important results. If the variance of returns increases, pushing the market-maker's expected cost function (WX) to the right, the ask price is raised. This is in line with the inventory cost models. The most significant result, however, is that the bid-ask spread increases in accordance with the number of informed traders in the market. Copeland and Galai

156

argue that if the probability of informed trading is higher for thinly traded stocks then this implies an inverse relation between the spread and trading volume. This assumes that the size of the transaction is constant. They predict that the probability of informed trading rises with the size of the transaction[2]. The concept of the time between trades that is so important to the models of Demsetz (1968) and Ho and Stoll (1981), is incorporated into the open quote scenario. In line with these studies, Copeland and Galai show that costs rise with the expected duration of the quote. These costs are likely to be lowest where there is more frequent trading. Thus this model incorporates elements of both the information and the inventory cost hypotheses.

The original motivation for the work in this chapter is the paper by Glosten and Milgrom (1985) which also looks at the relationship between information and the bid-ask spread. Although they use a slightly different analytical framework, the predictions of their model are very similar to those of Copeland and Galai (1983). The main difference is that Glosten and Milgrom look at the dynamic nature of the spread with particular reference to how market-makers process privately held information. Unlike Copeland and Galai (1983) they do not assume that private information is revealed immediately after each trade. Instead they assume that there will be further trading until information is revealed that resolves the informational differences between informed traders and the rest of the market. Therefore market-maker and trader predictions of the 'true' value of an asset will converge as private information is fully revealed in prices. They also argue that the spread will widen if the quality of information held by traders increases, or if informed traders become more numerous relative to uninformed traders. Another interesting aspect of their paper deals with the situation where informed traders hold such a strong position that the dealer is unable to break even. In this situation the market may shut down. This may, however, exacerbate the problem if a higher ask and a lower bid than expected is set when trading resumes. There may also be a welfare loss if a trader with potentially valuable information is unable to trade. They show that while the inventory costs of

---

[2] This is also predicted by the model of Easley and O'Hara (1987). They argue that informed traders prefer to trade larger amounts at any price. This quantity bias is not shared by uninformed traders.

trading are predicted to lead to negative serial correlation[3] between prices, those due to information effects do not. In fact they show that transaction prices form a martingale. This distinction between transitory and permanent effects has been exploited in empirical work.

5.2.3 MODELLING THE PATTERNS OF TRADE

The inventory and information cost models described above provide a good background to the existence and the determination of the spread particularly with regard to its relationship with the volume of trade. The next stage is to look at the modelling of the spread and volume at the intra-day level. Is it possible to predict patterns in these two key variables during the hours that the market is open?

One of the seminal papers in this area is that of Admati and Pfleiderer (1988). As discussed in detail in chapter 4, the model is designed to answer three important questions. Why does trading tend to be concentrated in particular time periods during the trading day? Why are returns (or price changes) more variable in some periods and less variables in others? And why do periods of higher trading volume also tend to be the periods of highest return volatility? These questions arise from observations based on intra-day trading patterns of Exxon shares in 1981. Both volume and volatility appear to follow a U-shape with concentrations coincident with the opening and closing of trade.

Admati and Pfleiderer explain this phenomenon in terms of the interaction of informed and uninformed traders. Their model is essentially based on the argument that trade generates trade. Their model is aided by dividing the group of uninformed traders into those who can use discretion with regard to when they trade and those who do not have this choice.

Admati and Pfleiderer show that in equilibrium discretionary traders will choose to trade at the same time of day, since their trading is unlikely to affect prices when trading is 'thick'. Although this attracts informed traders, Admati and Pfleiderer show

---

[3] See Roll (1984a).

that this minimises the costs of discretionary traders. Rather than increase adverse selection costs, they are driven down by the competition that occurs between informed traders. If the group of discretionary traders is split into 'large' traders and 'small' traders, large traders will avoid incurring large price discounts or premiums when the market is thin. Smaller discretionary traders can choose to trade at any time. Non-discretionary traders are likely to concentrate their trades at the beginning and end of each day since they represent the first and last opportunities that they can trade. Thus, under the Admati and Pfleiderer model, trading will be concentrated at the opening and closing of the market. They also show that the concentration of informed traders at these times increases the informativeness of prices which therefore exhibit increased variability. The emphasis here is on the relationship between volume and volatility. Although the bid-ask spread is not mentioned explicitly, the implication of Admati and Pfleiderer's trade generating trade argument is that volume will be highest when the cost of trading is at its lowest.

Foster and Viswanathan (1990) develop a similar model to Admati and Pfleiderer (1988) that looks at inter-day trading by informed traders and uninformed traders who have some discretion over when they trade. The advantage that an informed trader has over other market participants is gradually reduced as some part of the private information held is revealed through a daily public announcement. Discretionary traders will therefore delay entering the market until this information is revealed. At the same time, informed traders, knowing that an announcement will be made, trade more aggressively on the news that they hold in the interim. Thus, more information is revealed through trading. They argue that if private information accumulates over the course of a weekend, then the cost of trading on a Monday is likely to be higher than during any other day of the week. The two key results of Foster and Viswanathan's work are that the volume of trade will be lowest when trading costs are highest and, contrary to the Admati and Pfleiderer (1988) model, this coincides with the period when prices are at their most variable.

A slightly different result is provided by Subrahmanyam (1991) who utilises the Admati and Pfleiderer (1988) framework to show that high volume and high costs are not necessarily inversely related. This is in response to empirical work[4] that suggests that the bid-ask spread follows a similar intra-day U-shape to volume. He argues that the results of Admati and Pfleiderer (1988) are dependent on informed traders being risk-neutral. Subrahmanyam argues that if informed traders are risk-averse then increased trading on their part will increase the costs of other traders in the market. Assuming that discretionary traders will choose to avoid such periods the burden must fall on non-discretionary traders who have no choice about when to trade.

Subrahmanyam's (1991) model is questioned by Brock and Kleidon (1992) who believe that information based arguments are not, on their own, sufficient to explain the coincidence of high volume and high costs as measured by the spread. The main emphasis of their work is in considering the impact of exogenous factors on trading. They exploit the work of Merton (1971) to show that transaction demand at the open and close of the day is less elastic than at other times of the day. They attribute this to two separate effects. The first is that information accumulates overnight but there is no opportunity to trade. At the opening of the market, the portfolio holdings of traders will not be at their optimum and a period of adjustment ensues. The second effect is that in anticipation of being unable to trade overnight, and since optimal portfolios at the close will differ from those that are optimal during a period of continuous trading, traders will avoid the risk of holding open positions during non-trading hours by closing out at the end of the day.

It is not altogether clear that the closing of positions at the end of the day in this way should be uniformly regarded as uninformed trading. It is quite conceivable that traders will avoid holding open positions because they know that 'harmful' information is due to be made public while the market is closed. Such an action could be regarded as informed rather than noise trading.

---

[4] See Ekman (1992).

Brock and Kleidon also consider non-portfolio motivated patterns of trading. If brokers are given limit orders to use their discretion over the trading day, the need to fill orders increases as the close of trade approaches. Differential demands across trading times may also occur if investors receive payoffs that depend on the time of day at which trading occurs. If, for example, portfolio managers are judged according to the performance of a benchmark index portfolio, e.g. FTSE 100, which is valued based on closing prices, managers will try to trade as close to the end of trading as possible.

Brock and Kleidon also consider the effects of information on the variance of prices and its impact on bid-ask spreads. They argue that if information accumulates while the market is closed, then the variance of prices at the opening of trade will be higher than at other times of the day. They also argue that because prices serve to aggregate information across traders and since they are unobservable during non-trading hours, there will be a greater divergence of beliefs. Under these conditions of increased uncertainty, the Brock and Kleidon model predicts a widening of bid-ask spreads at the opening and closing of trade.

5.2.4 EXTENDING THE THEORIES TO FUTURES MARKETS

The theoretical work described above is based on the microstructure of equity markets. Daigler (1997) argues that derivative markets, and in particular futures markets, must be considered separately because of their idiosyncratic trading systems and because they may not be affected by the same factors that affect stock prices.

One of the most important differences between stock markets and futures markets is that futures trading is organised as an auction market. Under this scenario buyers and sellers interact directly in a trading pit or ring on an exchange floor[5]. These traders act as brokers for hedgers and speculators who wish to carry out transactions in a futures market. The market also contains individual traders who trade for their own account rather than acting through a broker. A subset of this latter group are the scalpers, who

---

[5] This open-outcry system has recently been abolished by LIFFE in favour of an automated computer based trading structure.

although they are under no obligation to do so, offer bid and ask prices in the hope of making a profit.

The difficulty is in determining whether scalpers are actually market-makers who provide liquidity to incoming market orders, or whether they are simply another form of speculator. In a study of scalper behaviour on the New York Futures Exchange, Silber (1984) argues that scalper returns can be directly related to the quoted spread and the frequency of transactions. Therefore, the scalper in a futures market can be compared to the market-maker in an equity market. Further supporting evidence is provided by Kuserk and Locke (1993) whose study indicates that scalpers play a significant role in determining the level of trading in a market by accounting for nearly half of all trading volume.

The hectic nature of the trading process on a futures market means that the recording of bid and ask quotes by pit observers can be difficult. In addition these prices are not binding and therefore price observers tend to record only the prices at which transactions occur. This necessitates the calculation of an effective spread. The problems inherent in this calculation are addressed in section 5.4.

The argument by Daigler (1997) that futures markets deserve special consideration because they are affected by different factors to equity markets is not to the detriment of this study. The generic nature of many futures contracts and the possible impact of different information that might affect stocks (e.g. macroeconomic news) actually make this analysis more intriguing.

Some consideration must also be given to the adaptation of the inventory cost models to futures markets. It is important not to place too much emphasis on the idea that scalpers will carry large amounts of inventory. Unlike equity market-makers, they try to hold a so-called 'flat book' at all times. However, they will still incur some element of inventory risk where they cannot offload outstanding positions immediately.

In addition there are some interesting issues raised by Locke and Venkatesh (1997). The most significant of these is possibly the concern over the assumption that all trade goes through the market-maker. This of course may not be true, but it is difficult to determine the extent of such activity. There are at present no publicly available records of market-maker transactions for LIFFE contracts.

Ultimately, however, the translation of the models of the spread from equity markets is possible as long as the underlying differences are fully appreciated.

5.2.5 ISSUES TO BE ADDRESSED

The main focus of this thesis is how the volume of trade impacts on various aspects of futures trading. It appears, from the theoretical work presented above, that the main debate in this field is whether the volume of trade causes an increase or a decrease in the cost of trading, as measured by the bid-ask spread. On balance the theory comes down on the side of the argument that costs are driven down by the number of trades in a given period. The benefits of high frequency trading are believed to outweigh the costs of trading with informed investors.

The work in chapter 4 indicates that trading in the FTSE 100 and Long Gilts futures contracts is dominated by informed investors. How will the spread be determined for these contracts, where there are large volumes of trading and the probability that a market-maker is dealing with informed investors is very high?

As the theory above indicates, the situation becomes even more interesting when these issues are considered at an intra-day level. There is some support for volume being highest at the beginning and the end of the trading day, but the relationship with the spread is unclear. There are separate issues of trading at lowest cost and inelastic demand that imply totally different patterns in the cost of trading. An empirical study into this relationship for high frequency data will help to resolve some of these conflicting arguments.

Another interesting question that arises from the theory is that if the volume of trade exhibits certain patterns, does this make it easier for market-makers to set bid and ask

prices? Is it possible to predict certain times of the day when the volume of trade will be highest? Consider the following scenario which brings together some of the ideas mentioned above. Information accumulates while the UK futures market is closed. It can be argued that informed traders, trading on private information or different interpretations of public information, will try to act as quickly as possible on their 'news' as soon as the market reopens. This will attract uninformed traders motivated by a number of possible pretexts; they believe that informed traders will compete away any individual advantages, they are simply following a trend, or they are informed traders whose information has already been revealed in prices. It is also possible[6] that informed traders will initially trade like uninformed traders to lay a false trail and only trade in the 'correct' manner later in the day.

The end of the trading day is a period of particularly high demand by uninformed traders whose primary concerns are portfolio considerations, etc. This will attract informed traders keen to hide their intentions among the trades of others. There will also be a high demand from risk-averse traders wanting to close out positions that they believe will become exposed while information accumulates overnight.

The point of this discussion is to demonstrate that it is quite conceivable that the patterns of trade are predictable. A market-maker may not know who is informed or what that information might be, but is able to form expectations regarding the timing of trades and the likelihood that an investor is carrying 'news'. In a market where the proportion of informed traders is very high the market-maker is likely to err on the side of caution and set a relatively wide spread. As the market becomes more established it is likely to be easier for the market-maker to form these expectations. The question that now needs to be asked is if trading is 'stable' how does the market-maker react to unexpected trade? Although trading during thin periods may be carried out by those traders for whom the intra-day decay of private information is high[7], it does not have to occur during a thin period to be unpredictable. Presumably, however, that is when its impact is likely to be greatest. If it is possible to distinguish

---

[6] See Foster and Viswanathan (1994).
[7] See Barclay et al. (1990).

between the impacts of expected and unexpected trading then this will provide a valuable insight into how the market deals with shocks and whether it is capable of adapting to periods of unexpected trade without damaging the functioning of the market.

The aim of this study is to consider these relationships during two different time periods; close to the inception of the contract and when the contract is well established. This will provide an insight into how the spread is determined when a contract has little trading history and trading patterns may be less predictable. Does a higher level of uncertainty actually make the market more adaptable and better able to deal with shocks?

The next section looks at how various studies have investigated the relationship between the volume of trade and the bid-ask spread.

5.3 LITERATURE REVIEW

The empirical work in this field uses a variety of approaches to investigate various aspects of the bid-ask spread. This section provides an overview of the most important studies and those of particular relevance to the analysis in this chapter.

The most commonly adopted approach is to investigate the determinants of the spread where expectations are based on Demsetz's (1968) inventory cost model. In his own seminal paper Demstez uses the following regressions, each estimated individually by OLS:

$$S = \alpha_0 + \alpha_1 P + \alpha_2 \ln T + \alpha_3 M + \varepsilon_t \qquad (5.1)$$

$$S = \alpha'_0 + \alpha'_1 P + \alpha'_2 \ln N + \alpha'_3 M + \varepsilon'_t \qquad (5.2)$$

$$T = \beta_0 + \beta_1 N + v_t \qquad (5.3)$$

where S is the bid-ask spread, T is the number of transactions per day, P is the price per share, N is the number of shareholders and M is the number of markets on which the security is listed. The expectations under Demsetz's model are that S and T will be negatively related, while S and P will be positively related. This is because Demsetz believes that the spread per share will increase with price to maintain the

cost of transacting per pound exchanged. If this does not occur then those submitting limit orders will find it profitable to narrow spreads on securities where the spread per pound is larger. The number of markets on which a stock is listed is likely to be a reflection of its popularity and of the competitive forces on the spread, therefore S and M will be negatively related. He uses a random selection of 200 securities listed on the NYSE. Observations on these variables are then averaged for trading on two days, January 5 and February 28, 1965. His results are entirely in line with his predictions. The most significant result, however, is that the cost of trading, as proxied by the spread, and market activity are negatively related. The finding that lnT has a bigger impact on the spread than lnN is, he feels, surprising.

A similar result regarding the relationship between volume and the spread is provided by Tinic and West (1972) in an investigation of the impact of competition between market dealers in an automated exchange system for two periods in 1962 and 1971. Using an approach closely resembling that of Demsetz (1968), they provide further evidence that a negative relationship exists between the volume of trade and the bid-ask spread.

One of the few studies to consider derivatives markets is that of Goss and Avsar (1998). They also investigate the hypothesis that volume and the spread are negatively related. They test this relationship using monthly data on six different futures contracts quoted on the Sydney Futures Exchange from 1980 to 1991. Both variables are tested for stationarity using unit root tests. Since volume is integrated of the first order the following difference equation is estimated using the instrumental variable technique:

$$\Delta V_t = \alpha_0 + \alpha_1 \Delta APB_t + \varepsilon_t \qquad (5.4)$$

where V_t is the volume at time t, APB_t is the bid-ask spread at time t and ε_t is the error term. The results suggest that for the majority of contracts the two variables under investigation have a negative association.

Gwilym and Buckle (1996) carry out a test of the hypothesis that volume and the spread are inversely related using data on bid and ask prices for American- and European-style index options. Their expectation is that spreads on American-style

FTSE 100 options, (which can be exercised at any time up to maturity), are likely to be lower than those on European-style options, (which can only be exercised on the day of expiration), for the same contract, because the market for the former is well established and characterised by higher trading volumes. They believe that the time between trades is a more important factor than any adverse selection costs due to information asymmetries. Using daily data on bid-ask quotes for FTSE 100 index options priced on the LIFFE for the period January 1993 to March 1994, they show, by way of simple distribution and standard statistical analysis, that lower spreads are associated with American-style than with European-style options, i.e. at higher levels of trading.

One approach to distinguishing between the components of the spread and hence between the different inventory and information cost models is that proposed by Stoll (1989). His study centres on assigning probabilities to the movements of bid and ask prices based on the underlying assumptions of the different models. For example, under the inventory cost model, prices after a sale will be lowered to increase the probability of the next transaction being a purchase to offset an unwanted inventory holding. Under the information cost hypothesis, however, the likelihood of a purchase equals the likelihood of a sale once prices have been adjusted to reflect new information.

Stoll (1989) then models the covariance of transaction prices, $cov(P_T)$, and quoted end of day prices, $cov(P_Q)$ as:

$$cov(P_T) = a_0 + a_1 S^2 + u_t \qquad (5.5)$$

$$cov(P_Q) = b_0 + b_1 S^2 + v_t \qquad (5.6)$$

where S represents the bid-ask spread, u is a random error, and:

$$a_1 = \delta^2 (1 - 2\pi) - \pi^2 (1 - 2\delta) \qquad (5.7)$$

$$b_1 = \delta^2 (1 - 2\pi) \qquad (5.8)$$

In the equations above δ is the price reversal (as a fraction of the spread) and π is the probability of a price reversal. Using data for National Market Securities (NMS) quoted on the National Association of Securities Dealers Automated Quotations (NASDAQ) system between October and December 1984, and a combination of intra-

day and end of day prices he decomposes the spread into an inventory cost, an information cost, and a processing cost component. The results suggest that the information cost and inventory cost components both account for roughly forty per cent of the spread with processing costs making up the other twenty per cent. He also finds that while the size of the spread changes across the stocks, the components of the spread appear to be relatively unchanged. The other interesting result is that covariances calculated from the transactions data are negatively associated with the square root of spreads. Stoll argues that this further supports the existence of an inventory cost effect in the spread.

Gerber (1996) uses the same technique to analyse the structure of the Italian bond secondary market, using daily bid and ask prices for 15 bonds over a period between May 1988 and January 1989. This data is also used to construct a weekly series. She does not calculate the different components of the spread but instead concentrates on the relationship between price covariances and the squared spread. Unlike Stoll (1989) she finds a positive relationship between the two variables at the daily frequency. A significant negative relationship only occurs when the weekly data is used. She argues that this could be due to the fact that dealers, while risk-averse, tend to adjust their inventory slowly following a transaction. This period of adjustment may cover more than one day.

Krinsky and Lee (1996) use the Stoll (1989) technique to investigate the components of the spread around earnings announcements. Their data set includes intra-day price and volume data on securities quoted on the NYSE and the AMEX as well as earnings announcements made during the period January 1989 to December 1990. Their expectation is that the period immediately before an earnings announcement is characterised by an increased level of information asymmetry. Under the information cost model dealers will therefore increase the spread accordingly. They also expect a similar phenomenon to occur following the announcement as dealers try to protect themselves from those who can interpret the results to gain an advantage. Their results provide support for both of these effects and suggest that while information costs rise around earnings announcements the inventory cost actually falls. They

argue that this is because at these times dealers can take advantage of the 'thickness' of the market.

Affleck-Graves et al. (1994) exploit the Stoll (1989) technique to investigate the differences in the composition in the spread for stocks traded on auction based exchanges and those traded using automated trading systems. They hypothesise that processing costs are lower on auction markets because of the greater direct interaction of public orders. They also hypothesise that multiple dealers on automated exchanges are able to compete away inventory costs far more easily than specialists in auction markets. The markets that they consider are the NYSE and AMEX (auction based) and the NASDAQ system. Using data on transaction prices and bid-ask quotations for the months of March and April 1985, they show that while processing costs are lower for the auction traded stocks, the differences in inventory cost between the two market types are not statistically significant. They also show that the information cost component is much greater for stocks traded on the NYSE and the AMEX.

Glosten and Harris (1988) take a slightly different approach to this problem by decomposing the spread into just two components, one due to information costs and all other costs captured in the second component. The system that they estimate, which is not discussed here, shows that for a data set consisting of a total transaction record for every common stock traded on the NYSE over the period December 1981 to January 1983, the adverse information costs are an important factor in determining the spread. It remains, however, a small component of the overall costs.

Huang and Stoll (1997) provide an approach that tries to reconcile these two different methods of calculating the components of the spread. They argue that the previous specifications of the spread components approach suffer because they do not take account of trade size and are very sensitive to assumptions about the relationship between orders and trades. Their model is used to investigate the components of the spread for intra-day trade and quote data of the 20 most actively traded stocks in the Major Market Index (MMI) for the year 1992. Their results suggest information and inventory costs represent fairly small proportions of the spread. It is only when trade

size is taken into account that these components appear to increase significantly at higher volumes.

While the Stoll (1989) method has proved very popular it does not reflect the dynamic changes that may occur in the components of the spread, since the estimates of the different costs are only point values. There is also, with the exception of Huang and Stoll (1997), very little consideration given to how the volume of trade might affect these costs. Some appreciation of the dynamic nature of the trading process can be obtained by looking at the distribution of, for example, bid and ask prices, volume, and volatility across the trading day. One of the first papers to identify a U-shaped pattern in intra-day returns and variance was that of Wood et al. (1985). Using data on approximately 1000 stocks listed on the NYSE for two separate periods, September 1971 to February 1972 and the whole of 1982, they show that significant differences in the returns and variance occur throughout the day. Their evidence suggests that both of these variables are at their highest point during the open and closing periods of the trading day. This result is confirmed by Jain and Joh (1988) using price and volume data for the S&P 500 index over a five year period from 1979 to 1983.

Ekman (1992) investigates intra-day patterns in the S&P 500 index futures market. He argues that while the evidence of U-shaped patterns in price and volume data is consistent with the information models of Admati and Pfleiderer (1988) and Foster and Viswanathan (1990), other explanations should be considered. Firstly, results may be biased by the widespread use of relatively small samples. The results may be specific to equity markets and may not apply to the different microstructure conditions of futures markets. Finally, the patterns may be caused by the effects of non-synchronous trading.

Ekman argues that his use of a relatively long six-year sample and a single asset rather than a constructed index will help to address the problems of sample specificity and non-synchronous trading. His data set consists of time and sales data for the S&P 500 futures index quoted on the CME for the period from January 1983 to November 1988. His main variables of interest are returns, absolute returns as a proxy for return variance, the number of trades, the autocorrelation of returns and the percentage of

170

price reversals as proxies for the autocorrelation of transaction returns. He tests the equality of the means of each variable across intra-day intervals for each trading day of the week.

His results provide evidence of U-shaped patterns in both absolute returns and in the number of trades that is consistent with other empirical work in equity markets. This is inconsistent with the arguments that these results may be due to non-synchronous trading or small sample sizes. He argues that the rise in the end of day return after the spot market closes is evidence of different informational processes within each market. He also finds that there is an S-shaped intra-day autocorrelation pattern, consistent with the arguments of Glosten and Milgrom (1985) that information traders cause the autocorrelation coefficient to fall towards zero. Autocorrelation appears to be low at the open and close of trading suggesting that the impact of informed traders is highest at those times. Just after the spot market closes, however, the autocorrelation coefficient rises as the balance of trade tips towards more uninformed individuals.

Jordan et al. (1988) carry out a similar study looking at information and trading effects in the intra-day variability of soyabean futures prices using time and sales data for the period from January 1978 to October 1984. The variance of price changes is used as the measure of variability with the relative variability across the five intra-day periods measured by the ratios of variances in periods 1, 2, 4, and 5 to the variance in period 3. The statistical significance of any differences is tested using a non-parametric technique, on the basis that neither the variances nor the variance ratios are likely to be normally distributed.

Jordan et al. test the hypothesis that periods following news releases and the overnight and weekend suspensions of trading will be characterised by high volatility as information flows into the market. Their results suggest that volatility is highest at the opening of trade, particularly on a Monday, and directly following the publication of the relevant soyabean farming reports. They also find high levels of volatility at the market close that they are unable to attribute to information. They argue that this is more likely to be caused by those simply closing out positions to avoid overnight risk.

171

Daigler (1997) provides evidence in support of U-shaped trading and volatility patterns for three futures contracts. He considers a contract that has both an overnight and a daytime trading session (US Treasury bonds), a very active contract (S&P 500), and a contract with extended trading hours (MMI). He uses transaction data covering the period from 1988 to 1989. His results suggest that macroeconomic information does not play a large role in the increased activity at the open of trading and that trading in both the S&P 500 and the Treasury bonds contract is more active when the underlying cash market is open. He also argues that information has a greater relative impact on volatility than it does on volume.

Although the studies discussed above provide an insight into the trading patterns that occur in financial markets, none of them specifically consider the intra-day patterns of the bid-ask spread. Chan et al. (1995) look at the bid-ask spread for both NYSE stocks and options quoted on the Chicago Board Options Exchange (CBOE). The data is collected for the first quarter of 1986 and standardised bid-ask spread, return volatility and trading volume variables are generated. Intra-day differences are tested using the GMM methodology in a similar manner to the approach adopted by Foster and Viswanathan (1993). The results suggest that, while volume and volatility exhibit U-shaped patterns in both markets, the spread is U-shaped in the stock market but not in the options market. In the options market the spread is high at the open of trading but is lowest at the close. The results of the NYSE data appear to be inconsistent with the Admati and Pfleiderer (1988) information cost hypothesis and more in line with the arguments of Brock and Kleidon (1992). They argue that the diminishing intra-day spread that occurs in the options market can be explained by the model of Madhavan (1992) that predicts that information asymmetry is gradually reduced as information is revealed through trade prices.

Brock and Kleidon (1992) provide support for their own hypothesis using intra-day data on 462 stocks in the S&P 500 traded on the NYSE between October 1 and October 15 1987. They document a U-shaped pattern in both volume and the bid-ask spread across the trading day.

Leng (1996) investigates the intra-day patterns of four different variables for Deutsche Mark and Japanese Yen futures covering the period from November 1988 to November 1992. The variables under investigation are the autocorrelation of price changes, the realised bid-ask spread, price volatility and the number of trades. The main aim of his study is to see how these variables react to the release of US macroeconomic news. The results suggest that although there is evidence of a U-shape in intra-day volume, the other three variables exhibit an inverse U-shape. A lower spread at the open and close of trading accompanied by high volume is consistent with the inventory cost hypothesis and the work of Admati and Pfleiderer (1988).

Foster and Viswanathan (1993) look at variations in volume, volatility, and intra- and inter-day trading costs to test their model[8] that there are differences in the patterns of each of these variables due to information arrivals. With regard to volume, Foster and Viswanathan argue that, in a market of informed and uninformed discretionary traders, if public information is precise and the informed trader has more private information, then discretionary traders delay their trades. This makes it easier for a dealer to interpret the actions of an informed trader. Consequently volume is lower and trading costs are higher on Monday than on any other day of the week. Variations in volume are tested using the following equation which is estimated using GMM:

$$V_t = V + \sum_{i=1}^{n-1} 1_{d=i} \eta_i + \varepsilon_t \qquad (5.9)$$

where n is equal to either five, for the inter-day study, or 7, for the hourly intra-day investigation. V_t is the volume on day t which is composed of a fixed effect, V, an adjustment for the different periods, η_t, and an error term with an expected value of zero. They then use a chi-square test on the dummy variables to determine whether there are significant differences in trading volume during the periods under investigation. Foster and Viswanathan use data on stocks listed on the NYSE and the AMEX for the year 1988 divided into deciles according to their relative levels of trading activity. Their results suggest that there are variations in trading volume across the week, but only for the most actively traded stocks. At the intra-day level

---

[8] Foster and Viswanathan (1990).

they are unable to reject the hypothesis that the first half hour of trading exhibits higher volume for all of the groups of stocks under consideration.

Foster and Viswanathan carry out a similar investigation into variations in return volatility. They argue that this enables a better understanding of when prices are more informative. Their results suggest that significant differences occur only at the intra-day level and that the periods of highest volatility are also those when volume is at its highest.

The largest part of their work is devoted to investigating variations in trading costs and the number of transactions. They use a two equation model where quantity traded and the price change are the two dependent variables. Dummy variables are used to recognise on which day of the week, or which hour of the day, each transaction occurs. The first equation gives a conditional expected value for the transaction at time t. Price changes and quantity traded are both lagged by five periods and dummies are added for day of the week or hour of the day effects. The second equation gives the price change as a function of the order that was not expected by the dealer. A variable is included in this equation to represent the amount by which the dealer adjusts the transaction price for each share of unexpected order flow. This acts as some measure of the adverse cost component. The dummies are included to estimate variations in the fixed and adverse selection cost components of the price change. Each equation is estimated by OLS. Their results suggest that while the fixed component of trading costs shows very little variation, the adverse selection component is highest during the first half an hour of trading, falls during the middle of the day and then increases at the close of trade. They are also higher on a Monday relative to other days of the week. The fact that these periods of high intra-day trading costs are coincident with periods of high volume and high return volatility appears to reject the implications of Admati and Pfleiderer's (1988) model.

Hasbrouck (1988) takes a slightly different approach to testing the hypotheses of the bid-ask spread by looking at certain features of trades and the movements of bid and ask quotes in an attempt to identify characteristics consistent with either the asymmetric information cost or the inventory cost models. This centres on the

development of simple models of trade and quote behaviour that predict that, for the information cost hypotheses, quote revisions will be serially uncorrelated and the impact of trades on quotes is persistent. With regard to inventory cost, quote revisions are serially correlated and the impact of trades on quotes is temporary. This analysis is carried out using a moving average specification for the number of trades and a specification that involves the regression of quote revisions against a buy/sell indicator and a variable to take account of the size of an order. The data consists of time-stamped quote and transaction records for stocks listed on the NYSE over the period from March to April 1985. The results suggest that only low volume stocks exhibit significant negative correlation inventory cost effects. In contrast, he finds evidence of the persistent impact of quote revisions in line with the information cost hypothesis. There is also evidence that order-size is important in determining quote revisions reflecting, he argues, that large orders convey more information.

Hasbrouck (1991) extends this approach in a more general study that allows the use of broader information sets, for example, histories of quote revisions and non-linear functions of trade variables. Using transactions data for firms quoted on the NYSE and the AMEX for the first quarter of 1989 his results are very similar to his 1988 study, namely that volume and the spread are negatively related and that information costs rise with the size of the trade. In addition, his results suggest that the total impact of trades is not immediate, with some lag before all the information is revealed.

One of the problems of investigating the determinants of the bid-ask spread is that, in constructing regression equations containing the variables of interest, issues of simultaneity are often ignored. George and Longstaff (1993) examine the relationship between bid-ask spreads and trading activity in the S&P 100 index options market. They use intra-day trade and sales data for the index quoted on the CBOE during 1989. To account for the fact that the spread and any measure of trading activity may be jointly determined, they estimate the following equations using two stage least squares for both call and put options:

$$BA_i = \alpha_0 + \alpha_1 DUM_i + \alpha_2 P_i + \alpha_3 L_i + \alpha_4 T_i + \alpha_5 R_i + \epsilon_i \qquad (5.10)$$

$$L_i = \gamma_0 + \gamma_1 BA_i + \gamma_2 T_i + \gamma_3 T_i^2 + \gamma_4 M_i^2 + v_i \qquad (5.11)$$

where BA_i is the spread for the ith option, P_i is the option price, T_i denotes the time to expiry, R_i is a measure of the relative risk of the option given by the squared delta, L_i is a measure of the liquidity of the option, DUM_i is a dummy variable which takes the value one if the put or call option has a price above \$3, and M_i^2 is the squared difference between the S&P 100 index value and the strike price of the call or put option. This last variable is included since the expectation is that trading tends be higher for at-the-money options.

The results suggest that trading activity is a very important determinant of the spread. The measure of liquidity that they use; the time between trades, indicates that as the frequency of trades decreases the cost of trading rises. They also indicate that the time to maturity is an important factor in the spread set by dealers. As expiration approaches market-making becomes more risky; a fact reflected in a higher spread. George and Longstaff also estimate a four system equation incorporating the spread and liquidity equations for both the put and the call options to examine these relationships across options. The results confirm much of the work from the first set of equations and also suggest that put and call options can be regarded as substitutes. The spreads for puts are related to the spreads of calls reflecting, according to George and Longstaff, that dealers use information common to both to set bid and ask prices.

Wang et al. (1994) use a similar simultaneous estimation approach, but with the effective spread and price volatility of the S&P 500 futures index as the key variables. They model the spread and volatility as functions of average volume per trade, the number of market-makers, the number of transactions lagged by one period, treasury bill futures volatility, and dummy variables for each half-hour interval of trading during the day. Wang et al. believe that the relationship must be modelled in this way to take account of the close association between the spread and price risk proxied by price volatility. They also believe that it helps to separate liquidity and information effects on volatility. Under the information cost and inventory cost hypotheses volume could have either a positive or negative effect on the spread. The number of market makers is a proxy for competition in the trading pit and is expected to reduce spreads. The half hour dummies are designed to account for differences that occur in

176

the key variables across the trading day that cannot be accounted for by the other variables in the system.

The volatility equation contains two proxies for information effects. Wang et al. believe that the lagged number of transactions is a good proxy for information arrival. The close association of treasury bill movements with key economic announcements also makes it a good information proxy. It is also serves a useful purpose as the exogenous variable necessary to allow estimation by two-stage least squares (2SLS). Intra-day time and sales data is taken from the CME for the periods surrounding the 1987 crash and the year 1988. The results show that volatility is a significant positive determinant of the spread. They also show that the volume variable has a positive effect on the spread, in line with the information cost models, but it is only significant before and after the crash. Wang et al. argue that this is evidence of a structural change in the crash period. The coefficients for the number of market makers have the expected sign and the treasury bill information proxy also has a significantly positive impact. The other information proxy is shown to be insignificant. The other interesting result is that the dummy variables are insignificant suggesting that phenomena such as the much documented U-shape is accounted for by the other variables in the model.

The same approach is used by Wang et al. (1997) to model the simultaneous relationship between the volume of trade and the bid-ask spread. They exploit a two equation model similar to that used in Wang et al. (1994) but which is estimated at the daily rather than the intra-day level. They consider the most active contracts from a sample of financial, agricultural and metal futures covering the period from January 1990 to April 1994. The hypothesis that volume and the spread are jointly determined is tested using the Hausman (1978) specification test. The hypothesis is not rejected. In contrast to the intra-day study, the results from this analysis reveal a negative relationship between volume and the spread for all of the contracts considered.

It is perhaps surprising that given the simultaneity that is identified between volume, volatility and the spread, Wang et al. (1997) continue to use a two equation system. In

particular, they acknowledge the close empirical and theoretical links between volume and volatility, but then choose to ignore the simultaneity issue that arises.

Demos and Goodhart (1996) analyse the relationship between volatility, the average spread and the number of quotations using a two step procedure. A VAR approach, using the Box-Cox transformation, is used to find the best functional form between the variables. The resulting simultaneous system is then estimated by 2SLS. Demos and Goodhart use data on intra-day trading activity and returns for the Deutschmark/Dollar and Yen/Dollar exchange rates used on the interbank market. Using a combination of the variables in question and a set of dummies to account for temporal half-hourly effects, Demos and Goodhart show that volatility and the average spread are determined simultaneously, while the number of quotations affects the spread through volatility only. The analysis of the dummy variables allows links to be made between periods of high volatility and the release of public information. They also find that the relationship between the spread and volume is more in line with the arguments of Foster and Viswanathan (1990) than those of Admati and Pfleiderer (1988).

The question of the impact of the number of dealers in a market and the spread, as originally hypothesised by Demsetz (1968), is addressed in a simultaneous model by Laux (1995). A cross-sectional study is carried out on 829 NMS stocks for the period November 1984. These stocks represent the most frequently traded stocks quoted on the US over-the-counter equity market. Laux shows that institutional investors have an important role to play in providing competition to established market-makers and reducing the size of the spread.

Choi and Subrahmanyam (1994) carry out an investigation into the determinants of the bid-ask spread from the context of links between spot and futures markets. They argue that the links occur because futures trading draws uninformed traders away from stock markets and encourages trading on market-wide information, because futures market indexes are not subject to high levels of firm-specific information asymmetries. The hypothesis that they test is that if futures markets attract uninformed traders then dealers in stock markets will increase spreads to protect

themselves from the remaining informed traders. They estimate a single equation model of the determinants of the spread using the generalised least squares methodology for S&P 500 and non-S&P 500 stocks around the period that the MMI futures contract was introduced in 1984. Their results support their central hypothesis and fail to provide evidence that futures markets actually reduce spreads by creating liquidity.

One of the problems with much of the empirical work is that it does not address the fact that there is some overlap between the two main hypotheses discussed in Section 5.2. The bid-ask spread is likely to reflect both information and inventory control effects. One approach to this problem has been attempted by Ma et al. (1992). They aim to improve our understanding of the determinants of the spread by splitting it up into its noise and information components. They look at the effective spread[9] for futures contracts on four commodities; Treasury bonds, silver, corn, and soyabeans. This is based on a data set of transaction prices for contracts quoted on the CBOT over an approximately 1000 day period between 1980 and 1985. The construction of this data set involves a four stage screening process. This is essentially designed to ensure that enough observations are available to avoid biases caused by infrequent trading. Preliminary analysis of the effective spread across the day indicates that there is a statistically significant difference between the spread at the open and close of trading compared with the rest of the day, supporting the U-shape hypothesis.

This evidence does not, on its own, provide support for either the information or inventory cost hypotheses. Ma et al. (1992) argue that while the normal expectation is that greater liquidity actually lowers the spread, it is conceivable that positively correlated trades may actually increase spreads if dealers find themselves trapped, holding unwanted inventory that they are unable to unload because trades are all moving in one direction.

Therefore, based on this argument, Ma et al. separate the so-called noise effects from the information effects by filtering out those short-term price movements which

exhibit evidence of positive correlation. They acknowledge, however, that information can have similar effects and so their results may understate the impact of news. Noise is proxied by the time that sequential price changes during a particular time interval are of the same, rather than the opposite sign. Their results suggest that statistically noise is significantly greater at the open and close of trading. In addition their simple regression of the effective spread against noise provides evidence of a positive relationship. It is the residual from this equation that Ma et al. assume to be the information component of the spread. Analysis of this variable across the trading day suggests significant increases in the amount of information flowing into the market at both the open and close of trading. This is largely supportive of the information cost hypothesis, but suggests either that non-informed traders also tend to be non-discretionary traders, or that the cost of higher spreads due to information based trading at the open and close of trading is offset by the benefits of a 'thick' market.

It is clear from the discussion above that the empirical work in this area is rich and diverse. Nevertheless there are a number of shortcomings, some of which have already been noted, that need to be addressed.

The initial impression is that there have been few studies analysing the cost of trading in futures markets. The emphasis on equity and option markets may reflect the relative ease with which spread data can be obtained for these assets. The difficulties inherent in calculating the spread for futures contracts, where official bid and ask price are non-binding, is discussed in more detail in the next section.

Although there is a large literature investigating different aspects of the spread, an in-depth analysis of the role of volume appears to have been neglected. The Stoll (1989) approach, although quite interesting in terms of addressing the different elements of the cost of trade, is unable to say anything about the impact of the volume of trade[10].

---

[9] See Section 5.4 for a discussion on the difference between the effective and the quoted bid-ask spread.
[10] Huang and Stoll (1997) is an exception.

The studies that look at the patterns in key variables rarely provide any statistical justification for their conclusions. The economic links between the spread and volume are largely based on anecdotal evidence. In addition, a number of the studies do not actually consider the spread, and inferences on the cost of trading are made by implication.

Another weakness of the literature is its general failure to consider the impact of the spread on volume. The potentially most comprehensive approaches of Wang et al. (1994) and Wang et al. (1997), are spoilt by the apparent contradiction between the two studies. In using a simultaneous modelling technique the discovery of a bi-directional relation between three variables; volume, volatility and the spread, is ignored to allow a two equation specification.

This study aims to address these shortcomings and to add to the existing literature in the following key areas:

- an extensive study of not only volume and the spread but also of the impact of the expected and unexpected components of volume. A number of the studies discussed in this section have identified intra-day patterns in the inter-relationship between the volume of trade and the bid-ask spread. If these patterns are to some extent predictable how does the market react to an unexpected shock? It is not sufficient to regard all informed and uninformed trading as, respectively, unexpected and expected events. The correlation between the two groups of traders makes this distinction unsuitable. An understanding of the impact of these shocks, that has not been previously attempted, is vital particularly from the point of view of maintaining the smooth functioning of the market and regulation issues.
- the use of futures market data for the UK. The majority of the empirical work in this field concentrates on equities and rarely looks beyond the US trading system. This study will look exclusively at two UK futures contracts, the FTSE 100 and Long Gilts, using high frequency data that has only recently become available. This will therefore provide a unique

insight into the intra-day relationship between the bid-ask spread and the volume of trade.

5.4 METHODOLOGY

The investigation of the intra-day relationship between the bid-ask spread and the volume of trade requires two major econometric techniques. The first of these, state-space modelling, allows a time series, in this case volume, to be separated into its 'expected' and 'unexpected' components. This section takes a brief look at this approach and considers its appeal in relation to other techniques for identifying the components of a time series.

The second technique is simultaneous equation modelling. If the bid-ask spread and volume are determined simultaneously then more conventional estimation techniques, for example OLS, are unsuitable. This section considers the theory behind this approach, how simultaneity can be determined and its suitability for this study.

This section also looks at the problem of estimating the bid-ask spread from futures price data. As has already been discussed in sections 5.2 and 5.3, one of the peculiarities of futures markets is that the quoted spread is not a binding agreement and is frequently not recorded. Therefore, an effective spread must be calculated. The various different estimators that are available and their individual advantages and disadvantages are discussed in detail.

5.4.1 STATE SPACE MODELLING AND THE KALMAN FILTER

This approach to the modelling of time series has its origins in engineering science. The seminal work by Kalman (1960) and Kalman and Bucy (1961) has, however, found applications in economics following the work of Harvey (1981b, 1994). This description of state-space modelling relies heavily on these two references.

The basic premise is that an N x 1 vector of observable variables y_t can be described by an m x 1 vector of unobservable state variables α_t in an equation of the form:

$$y_t = Z_t \alpha_t + S_t \varepsilon_t, t = 1, \ldots, T \tag{5.12}$$

where Z_t and S_t are fixed matrices of order N x m and N x n respectively. The n x 1 vector of disturbances, ε_t, has zero mean and covariance matrix, H_t. This is known as the measurement equation.

The state vector, although unobservable, is assumed to be governed by the following process:

$$\alpha_t = T_t \alpha_{t-1} + R_t \eta_t, t = 1,\ldots,T \qquad (5.13)$$

where T_t and R_t are fixed matrices of order m x m and m x g respectively, and η_t is a g x 1 vector of disturbances, with mean zero and covariance matrix Q_t. This is known as the transition equation.

It is assumed that the disturbances in both the measurement and transition equations are serially uncorrelated. They are also assumed to be uncorrelated with each other for all time periods and with the initial state vector, α_0. These assumptions can be represented in matrix form as:

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim WN \left[0, \begin{pmatrix} H_t & 0 \\ 0 & Q_t \end{pmatrix} \right], t = 1,\ldots,T$$

and

$$E[\alpha_0 \eta'_t] = 0, \ E[\alpha_0 \varepsilon'_t] = 0, t = 1,\ldots,T$$

where WN stands for white noise. Z_t, S_t, H_t, T_t, R_t, and Q_t, are known as the system matrices and are often based on unknown parameters. The estimation of these so-called hyperparameters plays an important part in state-space modelling.

The transition equation and the measurement equation together represent the state-space form and within this framework it is possible to construct a number of different model specifications. It is possible, for example, to represent autoregressive moving average (ARMA) models in state space form. Consider the following MA(1) model:

$$y_t = \varepsilon_t + \theta \varepsilon_{t-1}, t = 1,\ldots,T \qquad (5.14)$$

If the state vector is defined as $\alpha_t = \begin{bmatrix} y_t & \theta \varepsilon_t \end{bmatrix}'$ then the MA(1) model can be written as:

$$y_t = [1 \quad 0]\alpha_t \qquad (5.15)$$

$$\alpha_t = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\alpha_{t-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}\varepsilon_t \qquad (5.16)$$

The aim in setting up the state-space formulation, particularly in the transition equation, is to convey a large amount of information in as few elements as possible.

Once the model has been written in state-space form the next stage is to implement the Kalman filter algorithm. This is a recursive procedure for computing the optimal estimator of the state vector at time t, based on the information available at time t. This is carried out in two stages. The so-called prediction equations form the optimal predictor of the next observation, while the updating equations incorporate this observation into the estimator of the state vector. Its derivation is based on the assumption that the disturbances and the initial state vector are normally distributed. Under this assumption the current estimator of the state vector is the best available, as are the predictor and the updated estimator. A similar result holds in the absence of normality, but only within the class of estimators and predictors which are linear in the observations.

Smoothing describes the application of these recursive techniques in reverse, once all the observations have been processed. Therefore, because more information is being used relative to the normal filtered estimates, this provides the optimal means of extracting estimates of the state variables from the observations. The general form of the Kalman filter using the state-space model above can be described in the following manner[11].

Let a_{t-1} denote the optimal estimator of α_{t-1} based on all the observations available at t-1. Let P_{t-1} denote the m x m covariance matrix of the estimation error. Therefore:

$$P_{t-1} = E[(\alpha_{t-1} - a_{t-1})(\alpha_{t-1} - a_{t-1})'] \qquad (5.17)$$

The prediction equations are given by[12]:

---

[11] See Harvey (1981b) for a more detailed derivation.
[12] The subscript t/t-1 used here indicates, for example in the case of $a_{t/t-1}$, that it represents the estimator of α_t at time t-1.

$$a_{t/t-1} = T_t a_{t-1} \tag{5.18}$$

and

$$P_{t/t-1} = T_t P_{t-1} T'_t + R_t Q_t R'_t, t = 1,\ldots, T \tag{5.19}$$

The updating equations are given by

$$a_t = a_{t/t-1} + P_{t-1} Z'_t F_t^{-1}(y_t - Z_t a_{t/t-1}) \tag{5.20}$$

and

$$P_t = P_{t/t-1} - P_{t/t-1} Z_t' F_t^{-1} Z_t P_{t/t-1} \tag{5.21}$$

where

$$F_t = Z_t P_{t/t-1} Z'_t + S_t H_t S'_t, t = 1,\ldots, T \tag{5.22}$$

The prediction error

$$v_t = y_t - Z_t a_{t/t-1}, t = 1,\ldots, T \tag{5.23}$$

is an N x 1 vector. It has zero mean and covariance matrix F_t. It plays an important role, as can be seen above, in updating the state vector by 'correcting' $a_{t/t-1}$.

The next stage in this process is evaluating the specification of the state-space model. Assuming that ε_t and η_t are normally distributed, the starting values can be specified in terms of a_0 and P_0, $a_{1/0}$ and $P_{1/0}$. With these initial conditions the Kalman filter will yield the 'best' estimator of y_t. This, together with the corresponding prediction error, v_{t-1}, allows the evaluation of the likelihood function. Each different specification of the state-space model implies its own likelihood function. This can be maximised with respect to any of the unknown parameters using a variety of available optimisation algorithms.

The setting of α_0 and P_0 is not a simple task unless genuine prior information is available[13]. Harvey (1994) argues that one solution is to initialise the Kalman filter at $t = 0$ as $a_0 = 0$, and to set $P_0 = \kappa I$, where I is the identity matrix and κ is a positive scalar. If κ is set to a large finite number it is possible to estimate a_t and P_t exactly for large values of t.

---

[13] For a review of some of the different methods see Harvey and Peters (1990).

It is also necessary to set starting values for the estimation of the other hyperparameters in the state-space model, particularly those relating to the variance structures of the disturbance terms in both the transition and measurement equations. It is easier to discuss this process for particular model specifications and so it is left until the end of this section.

The choice of model is based primarily on the characteristics of the observed series. It is possible to define models that take account of cycles, seasonal components, trends and, as shown above, more complicated ARMA structures. The evaluation of different models is usually carried out by comparing the goodness of fit inside and outside the sample period. The prediction error variance is often used as a basic measure of the goodness of fit within a sample. Post-sample predictions can be made once the parameters of the model have been estimated. The sum-of squares of the one-step prediction errors then give a measure of forecasting accuracy. These values can be compared across models.

As well as testing alternative state-space models, it is also possible to test whether or not a particular model has been mis-specified using various diagnostic procedures. In a well-specified model the residuals should be approximately normally and independently distributed. An investigation can be carried out by simply plotting the residuals, or more accurately by looking for evidence of serial correlation through the Ljung-Box statistic, heteroscedasticity, and the values for skewness and kurtosis.

5.4.1.1 What are the advantages of using the Kalman Filter approach?

The appeal of the state-modelling approach can be explained in a number of ways. The usual approach to identifying the expected and unexpected components of time series data is to exploit the ARIMA methodology. Bessembinder and Seguin (1992, 1993) and Jain and Joh (1988) use such methods to model prices and volume. The expected component is assumed to be the predictable part of the series while the residual is the unexpected component. Bessembinder and Seguin in particular exploit an equation that includes a series of dummy variables to form the explanatory part of the equation that they do little to justify. Harvey and Todd (1983) argue that such methods are often inappropriate. Their main concern is that the methods of

determining the specification of an ARIMA model; the correlogram and the partial autocorrelation function, are not always reliable with the result that incorrect models are sometimes fitted. They also argue that the use of automatic selection procedures, for example the Akaike Information Criterion, are even less dependable. Their paper attempts to demonstrate this by comparing the forecasting performance of the two approaches. The results are not exactly definitive, but they do indicate that the state-space models perform at least as well as the ARIMA models.

Another factor in the appeal of state-space models, as mentioned above, is that the individual components of the model can be associated with the particular characteristics of a series. The decomposition of a series into its component parts can be achieved using the ARIMA approach[14] but it is, according to Harvey and Todd (1983), a very complex procedure.

5.4.1.2 Volume and the Kalman Filter

The focus of this study, as well as looking at the relationship between volume and the spread, is to consider the relative impacts of the expected and unexpected components of volume on the cost of trading. The weaknesses in previous methods of achieving this discrimination have been discussed. The appeal of the state-space approach is that it allows the separation of volume into its two constituent elements based on the bare minimum of information. In fact the only assumption that is being made is that these two elements actually exist.

The specifications used in this study assume that volume has an underlying component and an irregular component. The irregular component is a sequence of random variables. The regular component is equal to the level in the previous period plus a white noise disturbance. Therefore, it is modelling the shocks to the system that occur during the trading day, (e.g. the U-shaped volume identified in the empirical work discussed in section 5.3). It is this simplicity that holds much of the appeal.

Two simple models are used in this study. The first of these is the 'signal plus noise' model. The measurement and transition equations are written as:

$$Y_t = \mu_t + \varepsilon_t \tag{5.24}$$

$$\mu_t = \mu_{t-1} + \eta_t \tag{5.25}$$

where μ_t and ε_t are the signal and noise components. Y_t is the observable variable and ε_t and η_t are distributed independently with zero mean and variances σ_ε^2 and σ_η^2 respectively. This model can be written in state-space form as:

$$Y_t = \alpha_t + \varepsilon_t \tag{5.26}$$

$$\alpha_t = \mu_t = \mu_{t-1} + \eta_t \tag{5.27}$$

where α_t is the state vector. The second specification is the 'local linear trend' model. The measurement and transition equations are written as:

$$Y_t = \mu_{t-1} + \varepsilon_t \tag{5.28}$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \tag{5.29}$$

$$\beta_t = \beta_{t-1} + \zeta_t \tag{5.30}$$

where ε_t, η_t and ζ_t are distributed with zero mean and variances σ_ε^2, σ_η^2 and σ_ζ^2 respectively. The state-space form of this model is given by:

$$Y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \alpha_t + \varepsilon_t \tag{5.31}$$

$$\alpha_t = \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \zeta_t \end{bmatrix} \tag{5.32}$$

The difference between these two specifications is the addition of the trend variable in the second model. This is included under the hypothesis that there may be an upward trend in the volume of trade induced by the approaching expiration of the futures contract[15].

The difficulty that arises when estimating these models is that starting values have to be identified. The use of the diffuse prior has already been discussed. However, starting values still need to be provided for the variances of the disturbance terms.

---

[14] See Hillmer and Tiao (1982).
[15] Harvey (1994) argues that to model the trend as forward looking, β_t, or backward looking as β_{t-1} is a matter of taste.

There are two in the 'signal plus noise' model; σ_ε^2, σ_η^2, while 'the local linear trend' model requires three, σ_ε^2, σ_η^2 and σ_ζ^2. Harvey and Peters (1990) point out, however, that the process of optimisation allows the variances to be specified with respect to one of the group. Therefore, one variance starting value needs to be specified for the first model and two for the second. Harvey (1994) suggests that one way to obtain these values is to look at the autocorrelation of first differences of the series under investigation. For the 'noise plus signal' model the relative variance, $q = \sigma_\eta^2/\sigma_\varepsilon^2$, may be estimated as:

$$\hat{q} = -2 - r^{-1}(1) \qquad\qquad (5.33)$$

where $\hat{q}$ is the estimator of q and r(1) is the first-order sample autocorrelation. The same equation can be used to generate estimates of $\sigma_\eta^2/\sigma_\varepsilon^2$ and $\sigma_\zeta^2/\sigma_\varepsilon^2$ for the 'local linear trend' model using the first and second order autocorrelations respectively.

Once these two models have been estimated the most appropriate specification can be chosen using the methods described above.

5.4.2 MODELLING SIMULTANEOUS EQUATION SYSTEMS

The aim of this study is to examine the relationship between volume and the bid-ask spread. Section 5.3 of this chapter pointed out that very few studies have acknowledged the possibility that volume and the bid-ask spread may be jointly determined. It seems reasonable to expect that while market-makers will adjust prices to the flow of volume, at the same time traders' investment decisions will be determined by how much it costs to carry out a transaction. This could be because either the costs determine any profit opportunity or their signalling properties indicate who may be in the market. This is essentially a supply and demand model and it is important that each part of the model is properly identified. This would not be possible in a single equation model. Therefore, this investigation is carried out in a two equation framework under the assumption that the variables of interest may be jointly determined. Simultaneous equation estimation can be carried out using 2SLS. The use of this method relies on certain preconditions that must be inherent in the model. These conditions, and the reasons why this method is preferable to other estimation techniques, are discussed below.

189

An important concept in the development of a simultaneous equation system is the distinction between endogenous and exogenous variables. As Stewart (1991) points out, however, there is a certain amount of ambiguity regarding the definition of these terms. The usual approach is to regard endogenous variables as those whose behaviour is determined by the model, while exogenous variables are taken as given. Problems arise usually in relation to the concept of exogeneity. This rather loose term can describe predetermined variables and those that are strictly exogenous. A predetermined variable is independent of current and future values of the disturbance to that equation. A strictly exogenous variable is independent of all future, past and present disturbances.

The following description follows closely that given by Stewart (1991) and exploits the same notation. Consider the following structural form of a simultaneous equation model:

$$A y_t = \Gamma z_t + u_t; t = 1, \ldots, n \tag{5.34}$$

where y_t is a G x 1 vector of current observations on endogenous variables, z_t is a K x 1 vector of observations on predetermined variables, u_t is a G x 1 vector of disturbances to each of the equations at time t, and A and Γ are matrices of parameters, with dimensions G x G and G x K respectively. A is assumed to be a non-singular matrix thereby ensuring that y_t is uniquely determined by x_t and u_t. If there is no serial correlation between the disturbances, the vector of predetermined variables can include current and lagged exogenous variables, and lagged endogenous variables.

It is also assumed that the disturbance vector has the following properties:

$$u_t \sim IID(0, \Sigma); t = 1, \ldots, n$$

where the diagonal elements of Σ represent the variance terms for the individual elements of u_t. The off-diagonal elements represent the covariances between the disturbances of the different equations in the model. Σ is assumed to be positive definite. This rules out the possibility that there is an exact linear dependency between the disturbances.

The structural model above can be solved for y_t to give the following reduced form:

$$y_t = A^{-1}\Gamma z_t + A^{-1}u_t, t = 1,\ldots,n \qquad (5.35)$$

or, more simply,

$$y_t = \Pi z_t + v_t; t = 1,\ldots,n \qquad (5.36)$$

A feature of this structural system is that multiplying each side by a non-singular matrix F produces a new system of equations:

$$FAy_t = F\Gamma z_t + Fu_t; t = 1,\ldots,n \qquad (5.37)$$

that when solved for y_t has the same reduced form as equation 5.34, since

$$y_t = (FA)^{-1}F\Gamma z_t + (FA)^{-1}Fu_t$$

$$y_t = A^{-1}F^{-1}F\Gamma z_t + A^{-1}F^{-1}Fu_t$$

$$y_t = A^{-1}\Gamma z_t + A^{-1}u_t; t = 1,\ldots,n$$

This linear transformation above replaces the equations of the original structure with a set of G linear combinations of the form:

$$f'_i A y_t = f'_i \Gamma z_t + f'_i u_t; i = 1,\ldots,G; t = 1,\ldots,n \qquad (5.38)$$

Where f'_i is row i of F. However, as Stewart (1991) points out, this creates a potential problem; identification. Since equations 5.34 and 5.37 are so similar it is not possible to say whether the estimation of 5.34 is actually estimating the parameters of 5.34, or the parameters of a set of linear combinations as in 5.37. Therefore, to ensure identification it is necessary to impose restrictions on the elements of A and Γ (or indeed Σ). The basic premise is that in order to estimate a system of simultaneous equations there must be at least as many structural equations as there are endogenous variables.

The two most common methods of determining whether a system of equations is identified are the order and rank conditions. The order condition is a necessary restriction for the identification of a structural equation and requires that the number of linear restrictions on the equation must be at least equal to the number of structural equations G, minus one. The rank condition is a necessary and sufficient restriction. It requires that at least one non-zero determinant can be constructed from the

coefficients of the variables excluded from that particular equation, but included in other equations in the model.

The next problem to address is the estimation of a system of simultaneous equations. There are essentially three alternative methods available. They are the naive approach, the limited-information approach, and the full-information approach. The naive approach takes the reduced form of each equation and estimates it using OLS. It therefore ignores any information that might be contained in the other equations of the system, particularly regarding the identity of the endogenous and the exogenous variables. It is possible to show that using this method results in estimators that are biased and inconsistent because of the inclusion of endogenous variables among the set of explanatory variables.

The limited information approach also estimates one equation at a time but, unlike the naive approach, it distinguishes between endogenous and exogenous variables. It also takes account of which variables are included in other equations but excluded from the one being estimated. The class of estimators used in this estimation includes; indirect least squares, limited-information maximum likelihood, and the most common, two-stage least squares. They are sometimes referred to as instrumental variable estimators.

The full information approach estimates the entire system of equations simultaneously using all the available information. It estimates all the structural parameters and all identifying restrictions on each equation of the system. This approach utilises two specific estimators, three stage least squares (3SLS), and full information maximum likelihood (FIML).

A comparison, using Monte Carlo techniques, of these different approaches is carried out by Intriligator (1978). He argues that OLS estimators have the largest bias of all the estimators considered, which outweighs any benefits from retaining the Gauss-Markov property of minimum variance. They do have their uses, however, in performing preliminary regressions or in recursive models where alternative techniques are unnecessary.

Among the limited information estimators, Intriligator finds that the 2SLS estimator performs best in terms of both bias and mean squared error. Although problems can arise regarding multicollinearity, it has the additional advantage of being the most stable in terms of specification errors.

If the system of simultaneous equations is correctly specified and the variables are correctly measured, the full information approaches appear to provide the best estimators with regard to bias and mean square error. The prerequisite of correct specification is, however, a vital one. If this does not hold then the estimators actually perform worse than those of the limited information approach. The nature of full-information estimation means that an error in any one equation will be transferred throughout the whole system. 2SLS confines the problem only to the particular equation that is being estimated at the time.

Therefore, 2SLS is chosen for this study as a superior method to OLS for the investigation intended here, and to avoid the potential problems inherent in using 3SLS or FIML.

The following explanation of 2SLS, provided by Barr (1997), follows closely the notation in the previous chapter. This is appropriate since GMM and instrumental variable estimation are very closely linked. Consider the following linear model

$$Y = X\beta + \varepsilon \qquad (5.39)$$

Suppose that the set of instruments is represented by H. Therefore, following the previous notation:

$$g_T(\hat{\beta}) = T^{-1}H'\varepsilon(\hat{\beta}) \qquad (5.40)$$

To generate the parameter estimates it is necessary to minimise:

$$Q_T(\beta) = g_T(\beta)'W_T\, g_T(\beta) \qquad (5.41)$$

The first order condition for the solution is given by:

$$D_T(\hat{\beta}) = \frac{\partial\, g_T(\hat{\beta})}{\partial\hat{\beta}} \qquad (5.42)$$

In this case:

$$g_T = T^{-1} H' \varepsilon \tag{5.43}$$

$$g_T = T^{-1} (H'Y - H'X\beta) \tag{5.44}$$

Therefore:

$$D_T(\hat{\beta}) = -T^{-1} X'H \tag{5.45}$$

If this is substituted into the first-order condition it is possible to show:

$$X'HW_T H'Y = X'HW_T H'X\hat{\beta} \tag{5.46}$$

If the system is just-identified then $X'HW_T$ can be cancelled on both sides to give:

$$\hat{\beta} = (H'X)^{-1} H'Y \tag{5.47}$$

In an over-identified model the estimator is given by;

$$\hat{\beta} = (X'HW_T H'X)^{-1} X'H W_T H'Y \tag{5.48}$$

The weighting matrix W_T that is obtained form the general formula is given by

$$W_T = H'\Omega H / T \tag{5.49}$$

Barr (1997) points out that the estimation of Ω still needs to be carried out. However, where the errors satisfy the Gauss-Markov conditions of no serial correlation and no autocorrelation such that, $\hat{\Omega} = \hat{\sigma}^2 I$, the variance terms cancel to give the 2SLS estimator:

$$\hat{\beta} = (X'H(H'H)^{-1} H'X)^{-1} X'H(H'H)^{-1} H'Y \tag{5.50}$$

The actual procedure is carried out by regressing the explanatory variables on the instruments, and then regressing the endogenous variables on the fitted values from the first regression. Hence, the name two stages least squares.

Although the expectation is that volume and the bid-ask spread are simultaneously determined it is important to check whether such a relationship actually exits. This can be carried out using the Hausman (1978) Specification Test. This essentially tests whether the endogenous variable is related to the error term. The test procedure involves regressing the endogenous variable on all of the predetermined variables (i.e. the reduced form equation) to obtain the fitted values and the residuals. These are then placed into the structural equation. If this equation is estimated to reveal that the residual term is statistically significant then a simultaneous relationship exists.

Pindyck and Rubinfeld (1991) suggest that using actual rather that fitted values improves the efficiency of the estimation and this is the approach adopted here.

5.4.3 CALCULATION OF THE EFFECTIVE BID-ASK SPREAD.

The calculation of the bid-ask spread in futures markets is not an easy task. That is to say, that while the numbers can be easily compiled, it is not altogether clear which of the various estimators is the most suitable. Unlike in equity markets, where bid and ask prices are quoted continually, in futures markets such prices are usually only quoted when trading is slow to initiate transactions. They may therefore bear little relation to 'true' prices. Unless bid and ask prices are quoted at exactly the same time as transaction prices, their use in the calculation of futures markets spreads is likely to result in biased estimates. Therefore, empirical studies using futures market data tend to rely on calculations of the effective spread generated from transaction prices.

One of the most commonly used estimators is that derived by Roll (1984a). He demonstrates that the first-order serial covariance of price changes may be used as an estimator of the effective spread. If the price change on a transaction t, ΔP_t, is given by:

$$\Delta P_t = s\,D_t + \varepsilon_t, \varepsilon_t \sim \text{IID}(0, \sigma^2) \tag{5.51}$$

where s is the spread, D_t is a dummy variable taking the value -1 if a transaction at the bid is followed by one at the ask, 0 if a transaction at the bid (ask) is followed by another at the bid (ask), and 1 if a transaction at the ask is followed by one at the bid. Roll assumes that the market is informationally efficient, that buy and sell orders arrive with equal probability, and that the underlying distribution of price changes is stationary. He then shows that:

$$\text{cov}(\Delta P_t, \Delta P_{t-1}) = \text{cov}(s\,D_t, \varepsilon_{t-1}) + \text{cov}(s\,D_{t-1}, \varepsilon_t) + \text{cov}(\varepsilon_t, \varepsilon_{t-1}) + s^2 \text{cov}(D_t, D_{t-1}) \tag{5.52}$$

In an informationally efficient market there should be no relationship between the dummy variable and the error term. Therefore, the first three terms in the expression above are zero. Roll shows that by counting the number of possible price paths between the bid and ask price over two consecutive trades, $\text{cov}(D_t, D_{t-1}) = 1/4$. By rearranging the equation above it can be shown that the effective spread then becomes:

$$s = 2(-\operatorname{cov}(\Delta P_t, \Delta P_{t-1}))^{1/2} \qquad (5.53)$$

An alternative method of moments estimator is provided by Smith and Whaley (1994). They model price change, ΔP_t, in a similar fashion to Roll (1984a):

$$\Delta P_t = \delta_t s + u_t \qquad (5.54)$$

where δ_t, is defined much like D_t in the equations above, except that there is only a realisation when there is a record on the futures time and sales report, which is generally after a price change. Repeated offers at the bid and ask price are eliminated. The spread is represented by s, and u_t is a normally distributed innovation associated with each price change. Smith and Whaley obtain estimates of s and σ_u^2, the variance of u_t, from the first two moments of the empirical distribution of the absolute value of the price change.

Bhattacharya (1983) estimates the bid-ask spread from price series by considering only those prices which are the result of reversing price movements. The spread is then calculated as the mean value for all cases where the sequential price changes reverse signs.

One of the simplest estimators of the spread is that proposed by Thompson and Waller (1988). They estimate the spread as the average of absolute price changes from tick to tick over a specified period of time interval, n:

$$\text{Spread} = \frac{1}{n} \sum_{i=1}^{n} |P_t - P_{t-1}| \qquad (5.55)$$

Locke and Venkatesh (1997) argue that the only way to measure the transaction costs directly is to use data on the aggregate dollar flow from customers to market-makers. This is in line with the work of Demsetz (1968) who advocates such an estimator. Unfortunately, however, floor trader data is rarely available.

The dilemma here relates which of these methods to use. The Locke and Venkatesh (1997) approach can be rejected immediately simply because this study does not have access to such detailed information. The Roll (1984a) estimator has been widely

criticised in the literature because the estimation of the covariance often produces positive values. Followill and Rodriguez (1991) and Smith and Whaley (1994) find that over 25% of the covariance values result in an imaginary value for the spread. The modified Roll estimator used by Laux and Senchack (1992) produces better results but the fundamental problem of not taking account of the fact that prices may follow positive trends remains.

The difficulty in generating a sensible series for econometric analysis also affects the Bhattacharya (1983) estimator. By eliminating all non-reversing prices a large part of the sample is lost. Ma et al. (1992) argue that this may result in the understating of the spread if occasional transactions take place inside the market-maker's bid-ask spread. This is in contrast to the measure of the spread proposed by Thompson and Waller (1988). Smith and Whaley (1994) and Ma et al. (1992) point out that it implicitly assumes that the expected price change and the variance of future price changes is zero. This latter assumption may be unrealistic if the absolute value of successive price changes is affected by the changes in the underlying prices whenever new information arrives at the market. Thus, this estimator of the spread may have an upward bias. Despite these criticisms it continues to be used by the Commodity Futures Trading Commission.

The moments based estimator of Smith and Whaley (1994) is compared to the Roll (1984a) estimator by Locke and Venkatesh (1997), along with their own 'direct' measure. They consider twelve different futures contracts quoted on the Chicago Mercantile Exchange over a period from January to June 1992. The three estimators of trading costs are compared with a customer-market-maker spread that represents the difference between the average price at which customers buy from market-makers and the average price at which customers sell to market-makers. Although they, unsurprisingly, find that their estimator of the spread produces the most consistent results, they also find that the Roll estimator underestimates, while the moments estimator overestimates the spread.

These results are not entirely helpful. None of the estimators appears to be universally superior. The decision to adopt the Thompson and Waller (1988) estimator in this

study is based on the fact that it continues to be used in empirical work and the futures industry. It is important, however, to be aware of the potential biases at the interpretation stage.

The next section presents the empirical investigation into the relationship between the bid-ask spread and the volume of trade.

5.5 EMPIRICAL RESULTS

In this, the empirical section of the chapter, a regression model is used to examine the relationship between the bid-ask spread and the volume of trade for the FTSE 100 and Long Gilt futures contracts[16]. One of the aims of this study is to investigate this relationship for high frequency data, i.e. at the intra-day level. This section looks at how the data set was constructed from a sample of transaction price and volume details. Preliminary analysis is carried out using summary statistics and graphs of volume and the spread. The main emphasis, however, is the regression analysis. A justification for the specification is provided as well as detailed analysis of the results and their implications.

5.5.1 EMPIRICAL MODEL SPECIFICATION

The specification of the model employed in this study uses as its basis the work of Martell and Wolf (1987) and Wang et al. (1997). It is not possible to simply use the spread and volume on their own as both explanatory and dependent variables. The aim, therefore, is to put together a model incorporating a number of possible different determinants of the key variables. This serves two purposes; allowing the identification of the two equations and providing more information regarding the operation of futures markets. The discussion that follows looks at the variables considered with these objectives in mind.

---

[16] It would have been nice to investigate the volume/bid-ask spread relation for some of the other contracts considered in the earlier chapters. Unfortunately the data was not available at the time of this study.

Consider the following model:

Bid-Ask Spread = f(trading volume, price, average volume per trade) (5.56)

Trading Volume = f(bid-ask spread, volatility, volatility of short gilts, price of short gilts) (5.57)

In equation (5.56) the impact of volume on the spread will provide the answers to some of the key issues that this study seeks to resolve. The bulk of the theoretical work suggests that this relation should be negative since the benefits to market-makers of transactions occurring at high frequency outweigh the costs of trading with informed traders. As well as the relationship between the spread and total volume, the intention is to investigate whether differences exist between the relative impacts of expected and unexpected volume. Equation (5.56) will therefore also be estimated using, in turn, the two different components of volume generated from the Kalman Filter process.

It is also our aim to provide information regarding the impact of costs on the volume of trade. This forms the second main element of this investigation. Although an investor must also consider such costs as margin requirements, brokerage fees, etc., the cost of the so called 'round-trip', of simultaneously buying and selling a contract, must play a role in an investor's demand function. The expectation is that as these costs rise demand will fall and the relationship between the spread and the volume of trade will be negative. It will be interesting to see how this differs between the expected and unexpected components of volume. Therefore, at the same time as the different components of volume are put into equation (5.56), they will also be put into equation (5.57).

Easley and O'Hara (1987) argue that a key indicator of whether informed investors are present in a market is trade size. They show that the amount of information that an individual holds is positively correlated to the quantity of an asset that is traded. This can be measured by the average volume per trade. Another argument in favour of a positive relation between average volume per trade and the spread is that large trades

may force the market-maker to hold an undesirable inventory position. The market-maker will, therefore, increase the spread to offset the increased risk that such a position implies. At the same time it can be argued that it is in the interests of informed traders to deliberately mask their identities (as informed traders) by trading in smaller numbers of contracts. If market-makers do infer how much information a trader holds by the number of contracts they trade, costs will rise accordingly.

The impact of price on the bid-ask spread is, according to Demsetz (1968), and others[17], likely to be positive. He argues that the spread will increase in line with price to equalise the cost of transacting per pound exchanged. If this does not occur then those submitting limit orders will find it profitable to narrow spreads on securities where the spread per pound is larger.

Volatility is included in equation (5.57) to provide a further insight into its relationship with volume. The work in the previous chapters has established that a link exists between volume and volatility due to the fact that they are both driven by information. However, in this chapter it will be possible to say something about this link at an intra-day level. The relationship between these two variables is, based on the earlier work, expected to be positive.

The use of the volatility of short gilts variable in equation (5.57) is intended to show whether there is a common element to the information that moves around futures markets. It has been argued[18] that futures markets are primarily affected by macroeconomic information that is not market specific. If this is true volume will be positively related to volatility in another market. If information is not common to different markets then this variable would not be expected to have a significant impact on volume.

The price of short gilts in equation (5.57) is designed to capture any opportunity cost effects. If the cost of short gilts rises one might expect the demand for alternative

---

[17] See Tinic and West (1972), and Stoll (1978).
[18] See Daigler (1997)

200

investments to rise. This assumes, of course, that there is some degree of homogeneity between different futures contracts. The size of the coefficient on this variable will provide an indication of this substitutability.

These variables provide the basis for the proposed simultaneous investigation into the relationship between the bid-ask spread and the volume of trade.

5.5.2 DESCRIPTION OF THE DATA

The high frequency data for this empirical work was obtained from the LIFFE. The data was sampled at the transaction level, i.e., at the highest possible frequency, to provide information on prices and the volume of trade for the FTSE 100 and Long Gilt futures contracts. An important consideration when dealing with futures market data, as identified in chapter 4, is the problem of roll-over that occurs as contracts near expiration. It is, therefore, important to avoid as far as possible simply eliminating trading data of the last few days of a contract's life because this invariably excludes a large amount of the information that is coming into the market. Unfortunately, it was not possible to obtain open-interest details for transaction data so the Holmes-Rougier (1997) roll-over adjustment could not be exploited in this study. Thus, in order to avoid the problems of trying to form a continuous series from a number of contracts with different expiry dates, data was considered for a single contract during its most actively traded period. The September contract is used here and it is assumed that the results obtained are representative of any contract that could have been chosen. Observations were restricted to the period between the expiration of the June contract and that of the September contract to ensure that the data represents a highly liquid sample.

As mentioned in section 5.2 the intention is to see whether the relationship between the spread and volume, for a particular contract, alters as the market becomes more established. Therefore, data was collected for each contract in 1986, just after inception, and in 1996. Details on the two contracts are given in table 5.1.

Table 5.1: Contract Details for the FTSE 100 and Long Gilt Futures Contracts

FTSE 100 Futures Contract		
1986	Last Trading Day	One business day prior to the last in the delivery month
	Times	8.34-17.30
	Inception	03/05/84
	Months	March, June, Sept, Dec
1996	Last Trading Day	3rd Friday in delivery month
	Times	8.35-16.10 (16.32-17.30 APT)
	Months	March, June, Sept, Dec
Long Gilts Futures Contract		
1986	Last Trading Day	One business day prior to the last in the delivery month
	Times	8.00-18.00
	Inception	18/11/82
	Months	March, June, Sept, Dec
1996	Last Trading Day	Two business days prior to the last in the delivery month
	Times	8.00-16.15 (16.30-18.00 APT)
	Months	March, June, Sept, Dec

Note: APT refers to the period of automated trading recently introduced by LIFFE.

The table shows that all of the contracts can be traded throughout the day. The introduction of automated trading will also allow us to examine whether there are any differences in terms of costs and trading patterns between an open-outcry auction market and a computerised trading system. This is particularly relevant given the decision by most leading financial markets, (CBOT is a notable exception), to end traditional trading methods with the aim of providing cheaper and more efficient trading. The long trading day of the futures markets tends to extend beyond those of the underlying stock. It might, therefore, be possible to make an interesting comparison between trading patterns when the underlying stock is being traded and when it is not.

Regression analysis at transaction frequency is very difficult. Therefore, in line with numerous other studies, for example Ma et al. (1992) and Wang et al. (1994), each day was split into half-hour intervals starting at the top or bottom of the hour closest to the opening of the market. The data was then used to generate a series of variables based on the regression specification outlined above. Unfortunately, it was not possible to generate a variable for every half-hour of the sample period. This is because occasionally there were simply too few transactions.

The problem is to try to maximise the amount of data used to generate each variable while at the same time trying to maximise the number of half-hour observations. The threshold number of observations was chosen by looking at the data and calculating how many intervals would be lost for various limits on the minimum number of transactions. It should be noted that for the 1986 FTSE 100 contract trading during the middle of the day was very low. The 'lunch-time' intervals were therefore combined to ensure that the whole trading day could be represented.

Table 5.2 provides details on the number of transactions used to generate the sample used in this study. The number of trading days differs slightly between the contracts in any one year because of the different expiration dates. The sample size indicates the number of half-hour intervals used in each sample. It also indicates the imposed threshold value on the number of transactions required per interval.

Table 5.2: Sample Details for the FTSE 100 and Long Gilt Futures Contracts

	Year	FTSE 100	Long Gilts
No of Transaction Records	1986	6095	27718
	1996	77783	32620
Threshold Value	1986	5	10
	1996	20	10
Sample Size	1986	447	539
	1996	1024	888
No of Trading Days	1986	64	58
	1996	64	65

It is important to be aware at this stage that for the 1996 data floor trading does not occur across the whole interval. Trading in the FTSE 100 contract in particular is carried out on both the open-outcry and automated systems between 1600 and 1630 hours. In order to allow the discrimination of the two systems the APT data for this period was eliminated. Once the transaction data had been collected into 30-minute intervals the different variables were generated.

The bid-ask spread was calculated using the Thompson-Waller (1988) measure as outlined in section 5.4 of this chapter. The price data was also used to generate the average price. The choice of the price volatility measure requires some care. A number of studies[19] use absolute returns as a measure of volatility. The problem, however, is that this is very similar to the calculation of the spread. Thus, to avoid potential multicollinearity problems at the estimation stage, volatility was calculated using the Garman-Klass (1980) measure.

---

[19] See for example Ekman (1992).

This can be defined as:

$$\text{Variance} = 1/2[\ln(\text{High}) - \ln(\text{Low})]^2 - [2\ln(2) - 1][\ln(\text{Open}) - \ln(\text{Close})]^2 \quad (5.58)$$

where High Low, Open and Close are respectively the maximum, minimum, opening and closing prices in an interval. This is a widely used measure of volatility; see for example Grammatikos and Saunders (1986) and Daigler (1997), who consider it to be superior to alternative methods of calculation.

Total volume was calculated as the sum of the number of contracts traded in each interval. Average volume was calculated as the total volume divided by the total number of transactions.

The specification of the regression model used in this study requires that variables calculated from the Short Gilt futures contract must also be generated at the same time as those of the FTSE 100 and Long Gilt futures contracts. Therefore, all of the considerations that apply to the contracts on which this analysis centres also apply to the Short Gilts. In some intervals it was not possible to match the price volatility and average price of the Short Gilt contract to one in either of the other two contracts. Therefore, approximately thirty values in each sample were replaced by the weekly average for that variable.

One of the problems of trying to generate these variables is that they can be extremely sensitive to outliers. These can be caused due to simple input error on the part of the market. It is likely that, particularly during periods of high activity, some trading will either be missed altogether by those recording the events or incorrect details are put into the records. Therefore, a univariate test was carried out on each of the variables to check for possible outliers. It appeared that in each sample between ten and twenty prices, and some volume details, were of a completely different scale to those around them. Therefore, rather than exclude these observations entirely they were replaced by the maximum possible value allowed within a 95% confidence interval.

Analysis involving transaction data is not a simple task. As the discussion above illustrates, the construction of a data set is a very time consuming exercise. The task can be simplified, however, by exploiting a suitable computer programming language.

This study uses programs written in Visual Basic to carry out the majority of the manipulation required to obtain the sample of observations used in the empirical work of this chapter. They proved particularly useful in screening the data and calculating the variables for each interval. The programs can be supplied on request.

5.5.3 PRELIMINARY ANALYSIS

The summary statistics for the FTSE 100 and Long Gilt variables generated for each 30-minute interval of the trading day are given in tables 5.3a and 5.3b. These summary statistics allow us to make some preliminary comments about the variables and their possible inter-relationships. The first thing to notice is how, for both contracts, the mean spread and its variation have fallen between 1986 and 1996. One might expect the relative spread to be high closer to the inception of a contract for two reasons. Firstly, because the volume of trade is lower, market-makers are at greater risk of holding an undesirable position, because they are unable to obtain the benefits of trading that occurs at high frequency. Secondly, if a market is not yet fully established those who trade in it are at least likely to be partially informed. The risks to less informed investors in a new market are that much greater if they try to follow trading rules, etc., based on a relatively short trading history.

This may deter them from entering the market. Market-makers may know this and set prices accordingly to protect themselves from those who are better informed. The argument that the 1986 contracts are more risky is supported by the statistics on the mean volatility. The figures suggest that volatility has also fallen since the early life of both contracts. However, this needs to be interpreted carefully. Relatively high volatility and low volume indicates that the markets may have been dominated by informed traders, hence the higher spreads. The lower volatility and higher volumes of 1996 indicate that either the proportion of informed individuals has fallen, or as is perhaps more likely, the increased market depth means that it is harder to move the market.

Table 5.3a: Summary Statistics for the Variables Calculated from the Price and Volume Data for the FTSE 100 and Long Gilt Futures Contracts

Variable	FTSE 100		Long Gilts	
	1986	1996	1986	1996
Sample Size	447	1024	539	888
BA				
Mean	0.781	0.471	0.017	0.012
STD Dev	0.476	0.111	0.029	0.023395
Max	2.175	0.739	0.145	0.091
Min	0.077	0.213	0.833E-03	0.213E-03
TOTVOL				
Mean	43.372	358.771	362.887	1420.300
STD Dev	27.523	270.092	286.433	956.128
Max	115.188	1081.000	1061.200	3674.300
Min	6.0	25.0	29.0	46.0
VOLATILITY				
Mean	0.308E-05	0.163E-05	0.102E-03	0.120E-05
STD Dev	0.547E-05	0.216E-05	0.002	0.294E-05
Max	0.344E-04	0.142E-04	0.053	0.100E-04
Min	0.108E-09	0.299E-07	0.761E-09	0.973E-09

Note: BA is the Thompson-Waller (1988) estimated bid-ask spread. TOTVOL is the total number of contracts traded. VOLATILITY is the Garman-Klass measure of volatility. STD Dev is the standard deviation.

It is also interesting to note that if the hypothesis that informed traders trade in larger bundles is true, then the increases in average volume that have occurred while the spread has fallen, indicate that perhaps market-makers do not base their pricing decisions around the incidence of informed trading. They may be confident that, because of the high frequency of trading, the probability of finding an offsetting position is quite high.

All of these comments are purely speculative but they do indicate that there are some interesting issues to be investigated.

Table 5.3b: Summary Statistics for the Variables Calculated from the Price and Volume Data for the FTSE 100 and Long Gilt Futures Contracts

Variable	FTSE 100		Long Gilts	
	1986	1996	1986	1996
AV PRICE				
Mean	1625.600	3989.200	120.698	106.756
STD Dev	44.536	100.515	1.296	0.699
Max	1709.400	3989.200	123.512	108.108
Min	1545.500	3614.000	118.252	105.346
AV VOL				
Mean	4.089	5.238	7.432	41.285
STD Dev	2.649	4.346	3.722	18.165
Max	33.000	90.909	17.170	90.718
Min	1.000	1.000	2.609	3.754

Note: AV PRICE is the average price. AV VOL is the average number of contracts traded. STD Dev is the standard deviation.

The theoretical and empirical work discussed in the earlier sections of this chapter suggests that it should be possible to observe intra-day patterns in trading volume and the bid-ask spread. To investigate this issue graphs were generated based on the average values of total volume and the spread calculated for every interval between the opening and the closing of trade. The time measured on the x-axis represents the time at the end of the interval.

Figures 5.4-5.11 allow us to make a number of interesting observations. The theory suggests that high demand from investors at the beginning and end of the trading day produces a U-shape in the volume of trade. This can be observed for the periods of open-outcry trading for all of the contracts apart from the 1986 FTSE 100 contract.

A special note should be made of the last three intervals of Long Gilt and the last two intervals of FTSE 100 trading. These correspond to the period when open-outcry is replaced by automated trading. On both markets there is a significant fall in the total number of contracts traded. There are two possible explanations for this. First, that this period coincides with the closure of the underlying spot markets. This prevents arbitrageurs matching trades in both spot and futures assets and it closes a potential source of information. Second, that the drop in volume may simply be due to the fact that investors are wary of trading on unfamiliar automated exchange systems.

The pattern of trading on the 1986 FTSE 100 contract is difficult to explain. Trading is high at the beginning of the day but, with one exception, tails off to its lowest point at the close of trading. One possible explanation is that informed investors trade aggressively at the opening of trading and as the information is gradually revealed through prices, the incentive to trade is reduced. The opportunity of hiding behind uninformed traders at the end of the day is, as discussed above, perhaps less likely in a new market. This pattern of decline is not particularly uniform and may therefore reflect the fact that trading is simply unpredictable. Patterns of trading have not been established and, apart from the opening of trade, there is no particular rationale to concentrate trading at any other point in the day.

Figure 5.4: The Intra-Day Bid-Ask Spread for the FTSE 100 Futures Contract (Sept 1986)

FTSE 100 Sept 1986 Contract
Bid-Ask Spread
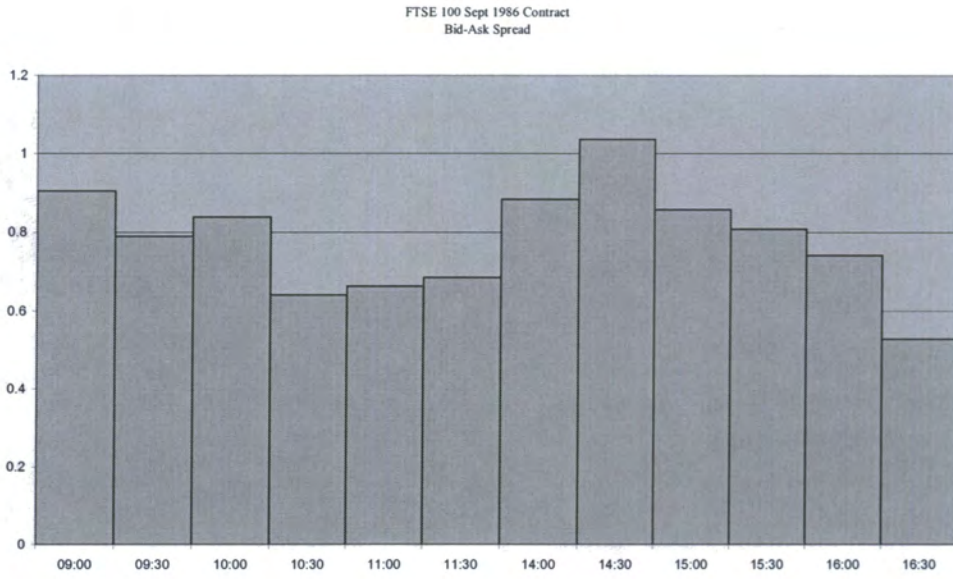
Figure 5.5: The Intra-Day Volume of Trading for the FTSE 100 Futures Contract (Sept 1986)
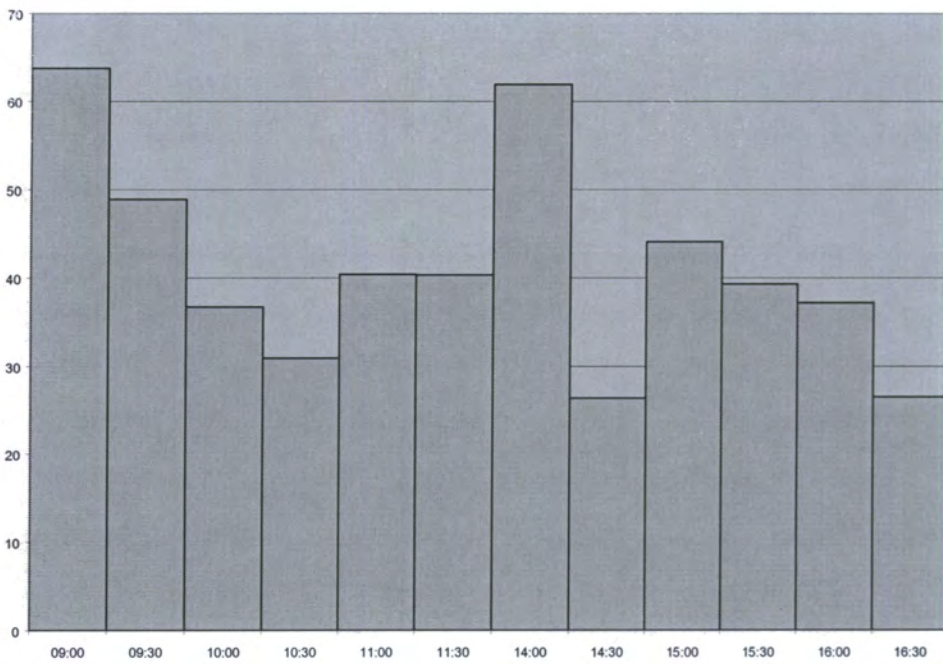
Figure 5.6: The Intra-Day Bid-Ask Spread for the FTSE 100 Futures Contract (Sept 1996)
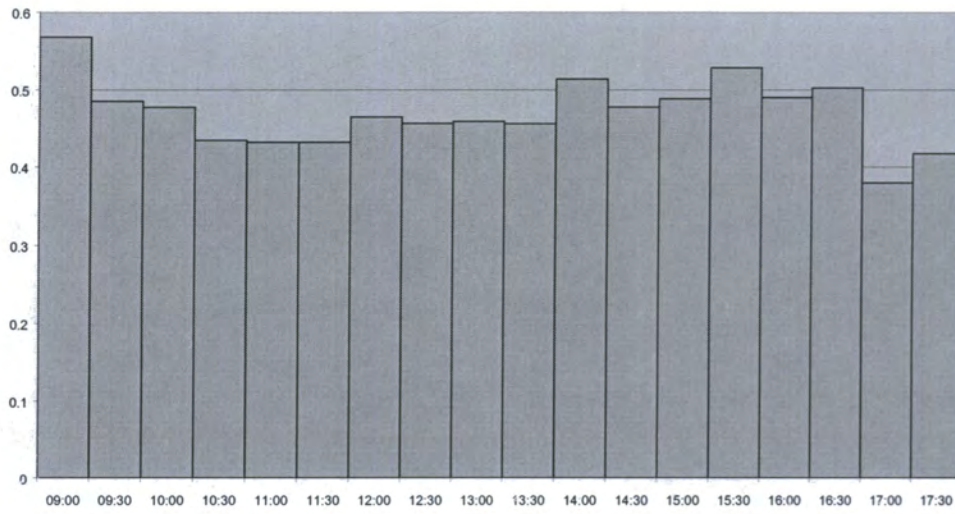
Figure 5.7: The Intra-Day Volume of Trading for the FTSE 100 Futures Contract (Sept 1996)
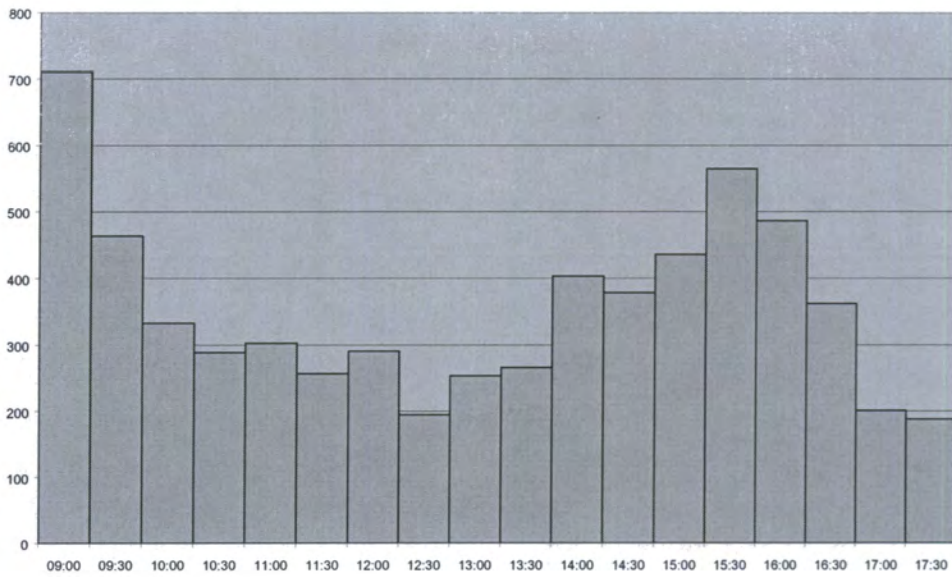
Figure 5.8: The Intra-Day Bid-Ask Spread for the Long Gilt Futures Contract (Sept 1986)
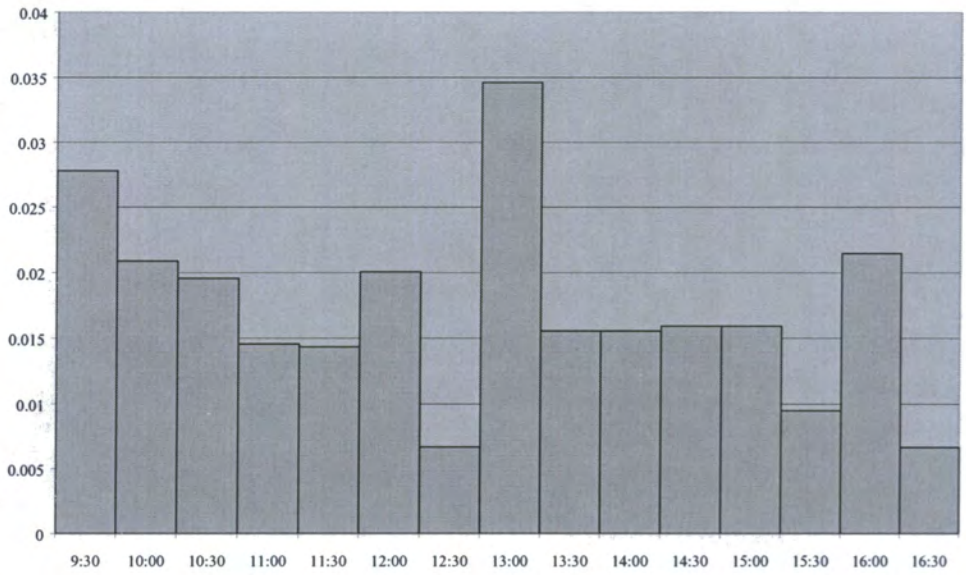


Figure 5.9: The Intra-Day Volume of Trading for the Long Gilt Futures Contract (Sept 1986)
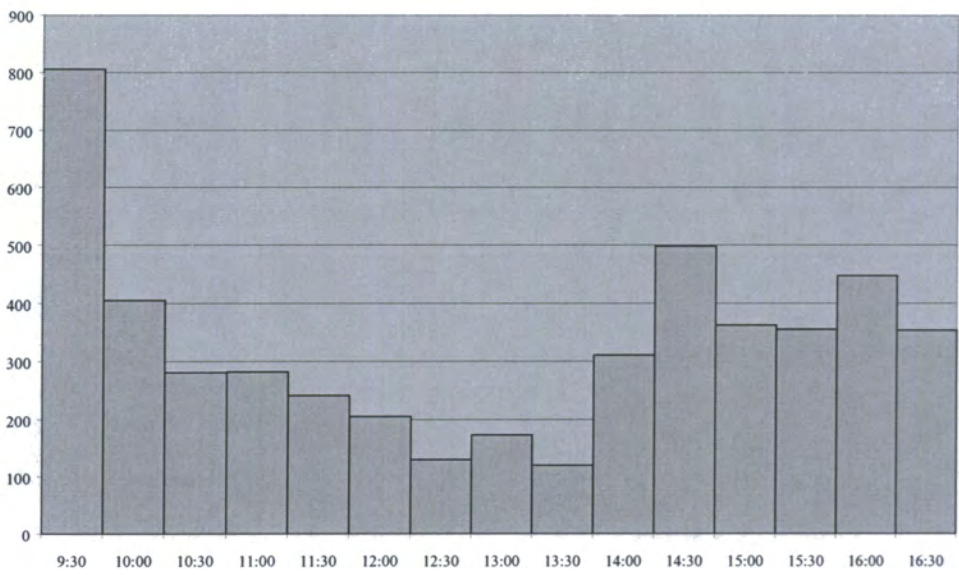
Figure 5.10: The Intra-Day Bid-Ask Spread for the Long Gilt Futures Contract (Sept 1996)
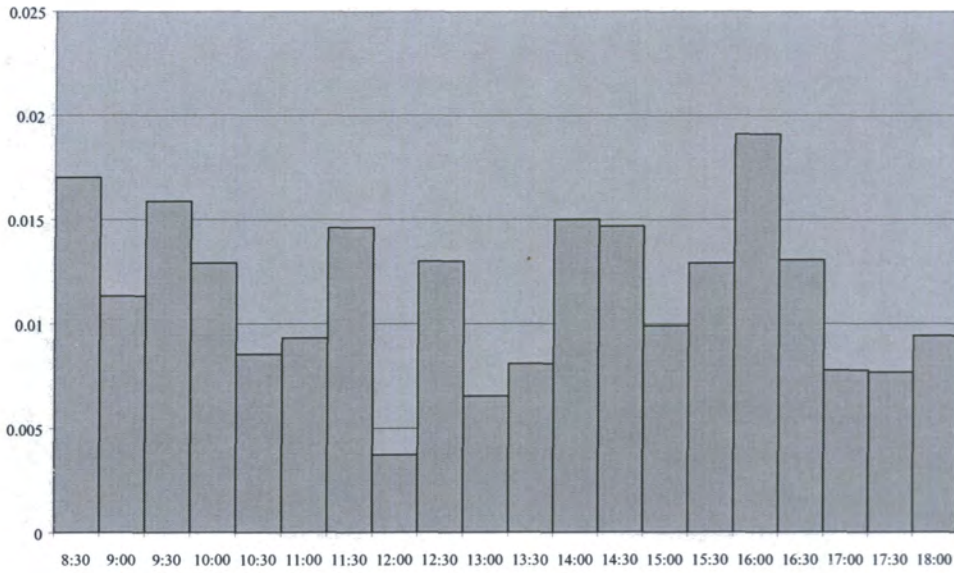
Figure 5.11: The Intra-Day Volume of Trading for the Long Gilt Futures Contract (Sept 1996)
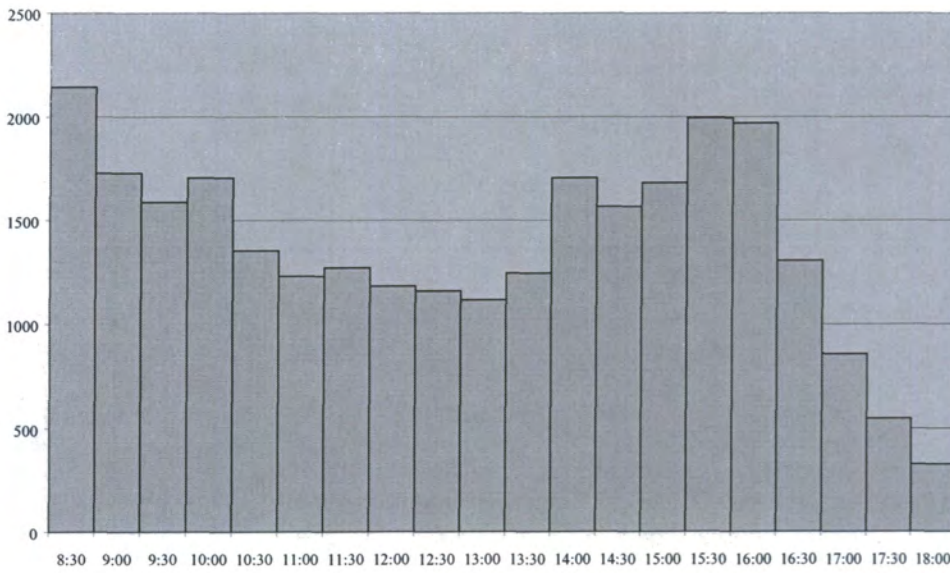
Analysis of the plots of the bid-ask spread reveals the coincidence of high trading activity with the highest average values of the spread, particularly for the 1996 contracts. This appears to reject the hypothesis that volume and the spread are negatively related. Note how the spreads during the period of automated trading are lower than those during the rest of the trading day. This supports the arguments of those in favour of a completely automated system at LIFFE who believe that it will lead to significantly lower costs. However, it must be remembered that volumes are also very low at this time.

The bid-ask spread patterns for the 1986 contracts are harder to explain. Although the opening spread is quite high there appears to be less predictability in 1986 relative to 1996. In 1986 both contracts, particularly the FTSE 100, are relatively new. If a market is still in its infancy, market-makers may still be finding their way in terms of reading investor behaviour and setting the appropriate spread. If investors do not follow particular patterns of trade it is less likely that spreads will exhibit any structure.

Once again, these comments are purely speculative at this stage, but they do suggest that there are indeed differences between the bid-ask spread and volume across the intervals that make up the trading day. In order to capture these differences that may not be explained by the variables in the regression model, a set of dummy variables was constructed; one for each 30 minute interval. For the FTSE 100 contracts there are 12 intervals in 1986 and 18 in 1996. For the Long Gilt contracts there are 15 intervals in 1986 and 20 in 1996.

5.5.4 THE EXPECTED AND UNEXPECTED COMPONENTS OF VOLUME

One of the important issues of this study is to investigate whether market-makers react differently to the expected and unexpected components of volume. These two series can be extracted from total volume using the state-space modelling technique described in section 5.4. Two different models were used for this purpose, to determine the most appropriate specification.

The 'signal plus noise' and the 'local linear trend' models were estimated using the log of total volume as the observed variable[20], as suggested by Harvey (1994). One of the difficulties in carrying out this estimation is the specification of the starting values. The simple noise plus signal model contains two unknown variance parameters, σ_ε^2 and σ_η^2, while the local linear trend contains three, σ_ε^2, σ_η^2 and σ_ζ^2. There is also the problem of setting the starting values for the state-vector. With regard to the latter the diffuse prior approach, as outlined in the methodology, was adopted. The starting values for the variances were generated using the autocorrelation based statistic proposed by Harvey (1994) and also described in section 5.4. The Kalman filter was used to produce the predicted series which was then smoothed using a reversal of the filter process. The estimates of the smoothed series are therefore based on the full information that the whole sample provides. For each model the predictive residuals were examined to determine the suitability of each model and to distinguish between them. The results are presented in tables 5.4 and 5.5[21].

The standard approach of determining whether a model is well specified is to check that the predictive residuals are approximately normally and independently distributed (nid) and to use the variance as an indicator of fit. Tables 5.4 and 5.5 reveal that in terms of the nid condition, the 'local linear trend' models perform poorly.

The 'signal plus noise' models conform to approximate nid[22] only. The 'signal plus noise' models are also superior in terms of minimum variance. Therefore, smoothed state vector and the direct residuals were extracted from the 'signal plus noise' output for each contract during 1986 and 1996. These time series represent the expected and unexpected components of volume respectively.

---

[20] The package used to do this was TSP version 4.4.

[21] Note that these statistics are not based on the full sample of observations. This is because for each model observations are used to generate the starting values in the state vector. Therefore, one observation is lost in the signal plus noise model and two are lost in the local linear trend model.

[22] It is possible to force the predictive residuals to fit the nid condition more closely but this approach is not widely practiced and tends to prohibit the isolation of the two components of a series.

Table 5.4: Predictive Residual Analysis for the FTSE 100 Futures Contract

	FTSE 100			
	S+N	LLT	S+N	LLT
Year	1986	1986	1996	1996
No of Obs	446	445	1023	1022
Mean	-0.004	0.676E-03	-0.525E-03	-0.245E-03
Variance	0.482	1.867	0.596	1.068
Skewness	0.035	0.016	0.390	0.221
Kurtosis-3	-0.052	-0.067	0.066	-0.081
LB for SC	6.203 (0.013)	147.154 (0.00)	4.968 (0.026)	81.558 (0.00)
LB for Hetero	0.008 (0.928)	59.614 (0.00)	3.777 (0.052)	8.350 (0.004)

Note: S+N is the signal plus noise model and LLT is the local linear trend model. LB for SC and LB for Hetero are the tests of the null hypothesis of no serial correlation and homoscedasticity based on autocorrelation tests of the predictive residuals and the squared predictive residuals respectively. LB is the Ljung-Box statistic. The values in brackets represent the p-values.

Table 5.5: Predictive Residual Analysis for the Long Gilt Futures Contract

	Long Gilts			
	S+N	LLT	S+N	LLT
Year	1986	1986	1996	1996
No of Obs	538	537	887	886
Mean	-0.0034	0.0029	-0.0017	0.8870E-03
Variance	0.8019	1.7806	0.6410	1.7400
Skewness	0.1133	0.2700	0.1911	0.4053
Kurtosis-3	-0.1122	-0.4248	0.3840	0.6149
LB for SC	1.8971 (0.170)	105.0868 (0.00)	2.8372 (0.092)	251.7886 (0.00)
LB for hetero	3.4609 (0.063)	20.2683 (0.00)	11.8431 (0.001)	92.6373 (0.00)

Note: S+N is the signal plus noise model and LLT is the local linear trend model. LB for SC and LB for Hetero are the tests of the null hypothesis of no serial correlation and homoscedasticity based on autocorrelation tests of the predictive residuals and the squared predictive residuals respectively. LB is the Ljung-Box statistic. The values in brackets represent the p-values.

The natural inclination is to detrend the total volume series based on the experiences of the previous chapters. Harvey (1989) stresses that it is important not to do this prior to the use of the Kalman filter, particularly if the specification under examination

216

includes a trend variable. An interesting point to come out of the analysis here is that fitting a trend is not suitable for this volume series.

The summary statistics for the expected and unexpected components of volume are given in table 5.6.

Table 5.6: Summary Statistics for Expected and Unexpected Volume

Variable	FTSE 100		Long Gilts	
	1986	1996	1986	1996
Sample Size	447	1024	539	888
EXVOL				
Mean	36.583	319.402	302.596	1220.800
STD Dev	9.087	181.519	163.074	555.630
Max	62.771	923.804	823.423	3157.400
Min	18.036	32.089	55.217	236.560
UNEXVOL				
Mean	1.150	1.039	1.092	1.088
STD Dev	0.638	0.293	0.456	0.441
Max	4.790	2.284	2.808	2.940
Min	0.212	0.431	0.294	0.194

Note: EXVOL and UNEXVOL are the expected and unexpected components of volume respectively. They have been generated by taking the exponential of the smoothed state series and the direct residuals. STD Dev is the standard deviation.

It is interesting to note how the unexpected component of volume is small relative to the expected component. This suggests that 'random' trades are quite rare; the majority of volume can be considered 'predictable'. These figures provide further support for the hypothesis that the patterns of trade have become more established over time. Both contracts indicate that the levels of unexpected trade have fallen between 1986 and 1996. In addition, the ratios of unexpected volume to expected volume have declined dramatically during this period. This is partly a reflection of the increasing popularity of financial futures. Table 5.3a indicated that although the absolute levels of volume are highest for the Long Gilt contract, the FTSE 100 contract shows a higher level of growth in trade. This could be attributed to the

217

relative immaturity of the contract in 1984. This growth appears to have been picked up in the expected component of volume.

The fact that the unexpected levels of trading in both contracts have fallen only slightly over time may indicate that profitable opportunities to 'surprise' the market are uncommon. Whether this is due to assiduous spread setting by market-makers should become clear in the next stage of the analysis. The figures in table 5.6 will be important in determining how the different components of volume affect the spread and whether the impacts are consistent across markets.

5.5.5 REGRESSION ANALYSIS

In this part of the empirical section the results of the estimation of the two equation model of the bid-ask spread and volume are presented and discussed.

The final specification of the model including the dummy variables can be written as:

$$Y_{1t} = \alpha_0 + \alpha_1 Y_{2t} + \alpha_2 X_{1t} + \alpha_3 X_{2t} + \sum_{i=2}^{K} \delta_i D_{it} + u_{1t} \qquad (5.59)$$

$$Y_{2t} = \beta_0 + \beta_1 Y_{1t} + \beta_2 X_{3t} + \beta_3 X_{4t} + \beta_4 X_{5t} + \sum_{i=2}^{K} \phi_i D_{it} + u_{2t} \qquad (5.60)$$

where
Y_{1t} = the bid-ask spread in period t (a half-hour interval);
Y_{2t} = the total volume/expected component of volume/unexpected
      component of volume during period t;
X_{1t} = the average price of the contract during period t;
X_{2t} = the average volume per transaction during period t;
X_{3t} = the price volatility of the contract during period t;
X_{4t} = the price volatility of the Short Gilt futures contract during period t;
X_{5t} = the average price of the Short Gilt futures contract during period t;
D_{it} = a dummy variable taking the value 1 if the observation belongs to the ith
      half-hour period and 0 otherwise. K is the maximum number of half-
      hour intervals during the day[23];

---

[23] Note that with an intercept in the model one less dummy than actual intervals is used. This avoids falling into the dummy-variable trap.

u_{1t}, u_{2t} = the random disturbance terms with zero mean and constant variance.

Following Wang et al. (1997) all of the variables were transformed into logarithmic form. This serves two purposes; it stabilises the variance of the error terms to aid estimation and it allows the variable coefficients to be interpreted in terms of elasticities.

The first stage in attempting to model a simultaneous relationship is to ensure that the system is identified. This means that numerical estimates of the parameters of a structural equation can be estimated from the reduced-form coefficients. Further details are given in the methodology section of this chapter. Consider as an example the identification of the model looking at the 1996 FTSE 100 futures contract. Under the order condition, the number of predetermined variables in the model, less the number in a particular equation, must be at least as big as the number of endogenous variables in an equation minus one. In this model there are 23 predetermined variables, 20 in the bid-ask equation and 21 in the volume equation. Each equation contains a single endogenous variable. Therefore, both equations are over-identified. The rank condition requires that at least one non-zero determinant can be constructed from the coefficients of the variables excluded from that particular equation, but included in other equations in the model. It is clear that in this model there is more that one non-zero determinant in each equation. Thus, the rank condition is satisfied. It is also possible to show that the rank and order conditions are satisfied for the three other models used in this study.

The next important step is to check that a simultaneous estimation technique is suitable for this data. The results of the Hausman (1978) specification test, as described in section 5.4, are presented in table 5.7. Since each model is to be run, in turn, using three different volume variables, three statistics are provided for each contract in 1986 and 1996.

Table 5.7: Hausman Specification Tests

Contract		Volume Variable	Residual Coefficient
FTSE 100	1986	Total Volume	1.155 (2.986)
		Expected Volume	0.524 (3.009)
		Unexpected Volume	0.631 (2.011)
	1996	Total Volume	0.307 (0.442)
		Expected Volume	0.662 (1.203)
		Unexpected Volume	-0.355 (-1.320)
Long Gilts	1986	Total Volume	28.331 (24.538)
		Expected Volume	15.364 (16.995)
		Unexpected Volume	12.967 (17.539)
	1996	Total Volume	1.693 (4.465)
		Expected Volume	0.724 (2.864)
		Unexpected Volume	0.969 (4.427)

Note: The numbers in parentheses denote the Student T statistics. Under the null hypothesis of no simultaneity the critical value at the 5% significance level is 1.96. If the absolute value of the test statistic exceeds the critical value the null hypothesis is rejected.

Under the null hypothesis of no simultaneity the significance of the residual term was tested using the Student T-test. It is clear from the table that for all of the contracts under investigation, with the exception of the FTSE 100 1996 contract, the bid-ask spread and volume are jointly determined over the period of investigation. Therefore, a simultaneous estimation technique is appropriate. The two equations in the model of the 1996 FTSE 100 contract must be estimated separately using OLS. This result suggests that while there may be a relationship between volume and the spread for the 1996 FTSE 100 contract, which may be bi-directional, it is not strong in statistical terms.

Tables 5.8 to 5.15 provide the details of the estimation of the bid-ask spread and volume equations.

Table 5.8: Results from the Simultaneous Equation Estimation Where the Bid-Ask
Spread of the FTSE 100 September 1986 Contract is the Dependent
Variable.

Variable	Equation 1	Equation 2	Equation 3
Constant	-62.530**	-98.786**	-30.879
Y_{2t}	2.842**	6.371**	4.703**
X_{1t}	7.689**	10.837**	4.792
X_{2t}	-2.809**	-1.629**	-3.450**
D_{2t}	0.674**	-0.162	1.231**
D_{3t}	1.390**	-0.109	2.369**
D_{4t}	1.392**	-0.332	2.516**
D_{5t}	1.335**	0.018	2.171**
D_{6t}	1.273**	-0.427	2.396**
D_{7t}	0.514*	0.226	0.684*
D_{8t}	2.124**	0.364	3.244**
D_{9t}	0.953**	-0.216	1.726**
D_{10t}	1.138**	0.068	1.832**
D_{11t}	1.5137**	0.113	2.413**
D_{12t}	1.407**	-0.657**	2.763**
GR^2 (bar)	0.646	0.634	0.605

Note: All three equations are estimated by 2SLS. Y_{2t} = the total volume in equation 1, the expected component of volume in equation 2, and the unexpected component of volume in equation 3 during period t (a half-hour interval); X_{1t} = the average price of the contract during period t; X_{2t} = the average volume per transaction during period t; D_{2t} to D_{12t} are the interval dummies for the trading day. GR^2 (bar) is the generalised R-bar-squared measure of fit proposed by Pesaran and Smith (1994). ** indicates significance at the 5% level. * indicates significance at the 10% level. White's adjusted disturbances have been used where appropriate.

Table 5.9: Results from the Simultaneous Equation Estimation Where the Volume of
Trade of the FTSE 100 September 1986 Contract is the Dependent
Variable.

Variable	Equation 1	Equation 2	Equation 3
Constant	-9.244	48.006	-57.250
Y_{1t}	-1.702**	-0.655**	-1.047**
X_{3t}	0.688**	0.271**	0.417**
X_{4t}	0.152	-0.001	0.017
X_{5t}	4.888	-9.130	14.019
D_{2t}	-0.233*	0.047	-0.281**
D_{3t}	-0.184	0.144*	-0.328**
D_{4t}	-0.289*	0.142*	-0.431**
D_{5t}	-0.028	0.154**	-0.182
D_{6t}	-0.345*	0.150*	-0.495**
D_{7t}	-0.043	-0.001	-0.0417
D_{8t}	-0.318	0.136	-0.453**
D_{9t}	-0.318**	0.062	-0.379**
D_{10t}	-0.172	0.106	-0.278**
D_{11t}	0.042	0.201**	-0.159
D_{12t}	-0.542**	0.127*	-0.669**
GR^2 (bar)	0.277	0.123	0.290

Note: All three equations are estimated by 2SLS. The regressands in equations 1, 2, and 3 are total
volume, the expected component of volume and the unexpected component of volume in period t (a
half-hour interval) respectively; Y_{1t} = the bid-ask spread in period t; X_{3t} = the price volatility of the
contract during period t; X_{4t} = the price volatility of the Short Gilt futures contract during period t; X_{5t} =
the average price price of the Short Gilt futures contract during period t; D_{2t} to D_{12t} are the interval
dummies for the trading day. GR^2 (bar) is the generalised R-bar-squared measure of fit proposed by
Pesaran and Smith (1994). ** indicates significance at the 5% level. * indicates significance at the
10% level. White's adjusted disturbances have been used where appropriate.

Table 5.10: Results from the Simultaneous Equation Estimation Where the Bid-Ask Spread of the FTSE 100 September 1996 Contract is the Dependent Variable.

Variable	Equation 1	Equation 2	Equation 3
Constant	12.812**	14.201**	15.821**
Y_{2t}	0.139**	0.144**	0.148**
X_{1t}	-1.713**	-1.887**	-1.998**
X_{2t}	-0.107**	-0.064**	0.005
D_{2t}	-0.110**	-0.150**	-0.105**
D_{3t}	-0.094**	-0.143**	-0.113**
D_{4t}	-0.168**	-0.220**	-0.212**
D_{5t}	-0.166**	-0.209**	-0.231**
D_{6t}	-0.165**	-0.214**	-0.238**
D_{7t}	-0.076*	-0.119**	-0.170**
D_{8t}	-0.058	-0.118**	-0.161**
D_{9t}	-0.083*	-0.143**	-0.176**
D_{10t}	-0.082**	-0.142**	-0.158**
D_{11t}	-0.018	-0.053	-0.069*
D_{12t}	-0.070**	-0.128**	-0.114**
D_{13t}	-0.095**	-0.139**	-0.097**
D_{14t}	-0.038	-0.073**	-0.035
D_{15t}	-0.092**	-0.133**	-0.101**
D_{16t}	-0.010	-0.068*	-0.084**
D_{17t}	-0.276**	-0.339**	-0.320**
D_{18t}	-0.215**	-0.283**	-0.231**
R^2 (bar)	0.253	0.243	0.212

Note: All three equations are estimated by OLS. D_{2t} to D_{18t} are the interval dummies for the trading day. R^2 (bar) is the R-bar-squared measure of fit. Refer to table 5.8 for the definitions of the other variables and additional details.

223

Table 5.11: Results from the Simultaneous Equation Estimation Where the Volume of Trade of the FTSE 100 September 1996 Contract is the Dependent Variable.

Variable	Equation 1	Equation 2	Equation 3
Constant	9.667	0.413	1.662
Y_{1t}	-0.196**	-0.101*	-0.062*
X_{3t}	0.315**	0.149**	0.099**
X_{4t}	0.003	0.898E-03	-0.728E-03
X_{5t}	-0.207	0.017	-0.006
Y_{2t-1}	0.347**	0.696**	-
D_{2t}	-0.784**	-0.514**	-0.294**
D_{3t}	-0.877**	-0.635**	-0.357**
D_{4t}	-0.858**	-0.602**	-0.353**
D_{5t}	-0.804**	-0.553**	-0.281**
D_{6t}	-0.919**	-0.593**	-0.319**
D_{7t}	-0.800**	-0.551**	-0.227**
D_{8t}	-1.025**	-0.631**	-0.359**
D_{9t}	-0.864**	-0.493**	-0.353**
D_{10t}	-0.689**	-0.333**	-0.340**
D_{11t}	-0.640**	-0.320**	-0.245**
D_{12t}	-0.737**	-0.367**	-0.359**
D_{13t}	-0.638**	-0.323**	-0.343**
D_{14t}	-0.564**	-0.356**	-0.265**
D_{15t}	-0.711**	-0.521**	-0.283**
D_{16t}	-0.686**	-0.583**	-0.243**
D_{17t}	-1.204**	-0.753**	-0.511**
D_{18t}	-1.006**	-0.444**	-0.590**
R^2 (bar)	0.554	0.757	0.393

Note: All three equations are estimated by OLS. D_{2t} to D_{18t} are the interval dummies for the trading day. R^2 (bar) is the R-bar-squared measure of fit. Y_{2t-1} = lagged total volume in equation 1 and the lagged expected component of volume in equation 2. Refer to table 5.9 for the definitions of the other variables and additional details.

Table 5.12: Results from the Simultaneous Equation Estimation Where the Bid-Ask Spread of the Long Gilt September 1986 Contract is the Dependent Variable.

Variable	Equation 1	Equation 2	Equation 3
Constant	-17.447	40.087	-110.071*
Y_{2t}	4.478**	6.4867**	11.000**
X_{1t}	-1.375	-15.858	23.106*
X_{2t}	-4.647**	-3.779**	-4.999**
D_{2t}	2.196**	0.8315*	4.044**
D_{3t}	3.404**	1.925**	5.109**
D_{4t}	3.707**	3.093**	3.837**
D_{5t}	4.294**	4.026**	3.692**
D_{6t}	5.491**	4.718**	5.476**
D_{7t}	6.232**	5.318**	6.179**
D_{8t}	6.914**	6.287**	6.406**
D_{9t}	6.783**	5.895**	6.677**
D_{10t}	4.042**	3.396**	4.171**
D_{11t}	1.351**	1.665**	0.509
D_{12t}	2.425**	1.130**	4.041**
D_{13t}	2.014**	0.589	3.903**
D_{14t}	2.075**	0.782	3.819**
D_{15t}	2.174**	-0.240	5.655**
GR^2 (bar)	0.719	0.574	0.810

Note: All three equations are estimated by 2SLS. D_{2t} to D_{15t} are the interval dummies for the trading day. GR^2 (bar) is the generalised R-bar-squared measure of fit proposed by Pesaran and Smith (1994). Refer to table 5.8 for the definitions of the other variables and additional details.

Table 5.13: Results from the Simultaneous Equation Estimation Where the Volume of Trade of the Long Gilt September 1986 Contract is the Dependent Variable.

Variable	Equation 1	Equation 2	Equation 3
Constant	12.154**	9.539**	2.615**
Y_{1t}	-5.416**	-3.448**	-1.968**
X_{3t}	2.139**	1.355**	0.784**
X_{4t}	-	-	-
X_{5t}	-	-	-
D_{2t}	0.468	0.491	-0.224
D_{3t}	0.235	0.347	-0.112
D_{4t}	0.190	0.176	0.014
D_{5t}	0.786	0.495	0.290
D_{6t}	1.045	0.718	0.326
D_{7t}	0.666	0.487	0.179
D_{8t}	1.581	1.035	0.546
D_{9t}	1.174	0.814	0.360
D_{10t}	0.535	0.398	0.137
D_{11t}	-0.339	-0.279	-0.060
D_{12t}	0.128	0.262	-0.134
D_{13t}	-0.138	0.114	-0.251
D_{14t}	0.152	0.279	-0.127
D_{15t}	0.258	0.505	-0.247
GR^2 (bar)	0.471	0.461	0.332

Note: All three equations are estimated by 2SLS. D_{2t} to D_{15t} are the interval dummies for the trading day. GR^2 (bar) is the generalised R-bar-squared measure of fit proposed by Pesaran and Smith (1994). Refer to table 5.9 for the definitions of the other variables and additional details.

Table 5.14: Results from the Simultaneous Equation Estimation Where the Bid-Ask Spread of the Long Gilt September 1996 Contract is the Dependent Variable.

Variable	Equation 1	Equation 2	Equation 3
Constant	-344.834**	-321.914**	-337.064**
Y_{2t}	5.726**	9.249**	14.775**
X_{1t}	68.378**	57.559**	84.798**
X_{2t}	-6.222**	-4.607**	-8.697**
D_{2t}	1.877**	-0.991*	6.419**
D_{3t}	2.237**	-1.264**	7.784**
D_{4t}	1.967**	-1.168**	6.932**
D_{5t}	3.044**	-0.246	8.234**
D_{6t}	3.704**	0.345	8.993**
D_{7t}	4.168**	1.123	8.950**
D_{8t}	3.632**	0.573	8.440**
D_{9t}	4.671**	1.233	10.072**
D_{10t}	4.323**	0.797	9.865**
D_{11t}	3.690**	0.333	8.9764**
D_{12t}	1.331**	-1.195*	5.336**
D_{13t}	2.176**	-1.450**	7.925**
D_{14t}	1.847**	-1.998**	7.950**
D_{15t}	0.721	-2.378**	5.652**
D_{16t}	1.108**	-1.504**	5.257**
D_{17t}	3.227**	0.753	7.118**
D_{18t}	0.057	-1.509**	2.535**
D_{19t}	2.597**	0.059	6.583**
D_{20t}	4.705**	0.598	11.165**
GR^2 (bar)	0.794	0.798	0.776

Note: All three equations are estimated by 2SLS. D_{2t} to D_{20t} are the interval dummies for the trading day. GR^2 (bar) is the generalised R-bar-squared measure of fit proposed by Pesaran and Smith (1994). Refer to table 5.8 for the definitions of the other variables and additional details.

Table 5.15: Results from the Simultaneous Equation Estimation Where the Volume of Trade of the Long Gilt September 1996 Contract is the Dependent Variable.

Variable	Equation 1	Equation 2	Equation 3
Constant	176.026	13.529	162.496
Y_{1t}	-2.036**	-0.943**	-1.093**
X_{3t}	0.948**	0.451**	0.4973**
X_{4t}	-36.117	-0.958	-35.159
X_{5t}	-0.019	-0.010	-0.009
D_{2t}	-0.123	0.202*	-0.325**
D_{3t}	0.050	0.345**	-0.296**
D_{4t}	-0.171	0.207**	-0.377**
D_{5t}	-0.175	0.194*	-0.369**
D_{6t}	0.021	0.277**	-0.257
D_{7t}	0.091	0.264	-0.173
D_{8t}	-0.234	0.128	-0.362**
D_{9t}	-0.096	0.212	-0.307*
D_{10t}	-0.164	0.192	-0.355**
D_{11t}	0.020	0.279*	-0.259
D_{12t}	-0.338	0.080	-0.418**
D_{13t}	0.081*	0.375**	-0.294**
D_{14t}	0.112	0.421**	-0.309**
D_{15t}	0.052	0.340**	-0.289**
D_{16t}	0.105	0.303**	-0.198
D_{17t}	0.184	0.273*	-0.089
D_{18t}	-1.444**	-0.503**	-0.940**
D_{19t}	-1.476**	-0.479**	-0.997**
D_{20t}	-1.399**	-0.323**	-1.076**
GR^2 (bar)	0.380	0.268	0.356

Note: All three equations are estimated by 2SLS. D_{2t} to D_{20t} are the interval dummies for the trading day. GR^2 (bar) is the generalised R-bar-squared measure of fit proposed by Pesaran and Smith (1994). Refer to table 5.9 for the definitions of the other variables and additional details.

5.5.5.1 Bid-Ask Spread Equation Analysis

The regression results for Equation 1 for each contract show the determination of the bid-ask spread by total volume, average price, average volume and a series of interval dummies. For all of the contracts, with the exception of the 1986 Long Gilt contract, average price is a significant determinant of the spread. The signs on this variable are, however, not entirely as expected. The average price of the 1986 FTSE 100 contract is positively related to the spread which is in line with the arguments of Demsetz (1968). For the other two contracts, however, the sign is negative. One possible explanation is that this occurs to encourage investors to continue trading. If prices rise margins will often rise accordingly. Costs due to the spread may therefore be lowered as a form of compensation. Alternatively, it may reflect the fact that for these contracts exploitable opportunities exist for investors posting limit orders.

For all of the contracts average volume is a significant, but negative, determinant of the spread. The expectation is that if larger bundles of contracts tend to be traded by informed investors then average volume would put upward pressure on the spread. The advantage of higher volume in this instance appears to outweigh such costs.

The total volume variable is also significant for all four contracts. The fact that the impact is positive is a very revealing result. This suggests that the information costs dominate the inventory costs. The results from chapter 4 indicated that both the FTSE 100 and the Long Gilt markets are dominated by informed traders. These statistics provide further confirmation of that discovery. The use of logarithmic variables allows us to state explicitly how this impact varies across the different contracts. For the 1986 FTSE 100 contract a 1% increase in total volume leads to a 2.84%[24] increase in the spread. Similarly a 1% increase in total volume results in an increase in the spread of 0.14%, 4.49% and 5.73% for the 1996 FTSE 100, the 1986 Long Gilt and the 1996 Long Gilt contracts respectively. A more detailed analysis of these results will be possible when the different components of volume are considered.

---

[24] All percentages are written to 2 decimal places.

The significance of some or all of the dummy variables for each contract reveals that the specification of the bid-ask spread equation is not able to completely explain the variation in the spread that occurs across the trading day. The dummies in the 1996 FTSE 100 equation reveal that a U-shape in the spread exists even when taking the key explanatory variables into consideration. It has already been suggested that in a mature market market-makers are able to set the spread based on well-developed expectations. It might be possible to capture some of these expectations using a variable such as lagged open interest. High open interest levels would suggest that more trades are likely. Unfortunately this data was not available for this study. The dummy variables for the other three contracts show some evidence of an inverted U-shape. Since this is not evident from the plots of the data, this suggests that this may be the result of the absence of a variable that pushes down costs. One possibility is the number of market-makers in the market. This would measure the effect of competition on the spread.

The regression results for Equation 2 reveal the impact of a similar set of variables on the spread as Equation 1, although total volume is replaced by the expected component of volume. The impacts of average price and average volume are very similar to those in Equation 1. The impact of the dummy variables has, however, changed for the 1986 FTSE 100 and the 1996 Long Gilt contracts. In both cases a large number of the dummy variables have become insignificant. This suggests that a large proportion of the intra-day variation in the spread can be attributed to the expected component of volume.

The expected component of volume has, like total volume, a positive impact on the spread. In the case of the 1986 FTSE 100 contract a 1% increase in expected volume causes a 6.37% increase in the bid-ask spread. A similar increase in expected volume results in increases of 0.14%, 6.49%, and 9.25% for the 1996 FTSE 100, and the 1986 and 1996 Long Gilt contracts respectively. Therefore, like total volume, increases in expected volume also lead to increased information costs. At the end of section 5.2 it was suggested that although market-makers may not know who is informed or what that information is, they can predict when informed traders are likely to enter the market. This conjecture appears to be supported by these results since expected

volume is positively related to the spread. This suggests that expected volume has an informed component that dominates any benefits of increased trading in terms of reduced inventory costs.

In Equation 3 the expected component of volume is replaced by the unexpected component of volume. Once again the signs on average price and average volume are similar to those for Equations 1 and 2. The one change is for the 1986 FTSE 100 contract where average price no longer has a significant impact on the spread. The impact of the dummy variables is much the same as in Equation 1. This supports the suggestion that the intra-day variation in the spread for the 1986 FTSE 100 and 1996 Long Gilt contracts can be partly explained by the variation in expected volume. For the other two contracts it appears that variables other than those in the model are still required.

The impact on unexpected volume for all four contracts is positive. This might be expected since unexpected trading is most likely to be driven by investors holding information. The impact on the bid-ask spread of the 1986 FTSE 100 contract of a 1% increase in unexpected volume is an increase of 4.70%. A similar increase in the unexpected components of volume for the 1996 FTSE 100, the 1986 Long Gilt and the 1996 Long Gilt contracts is 0.15%, 11.00% and 14.77% respectively.

It is interesting to note that the percentage variation in the bid-ask spread due to the variation in the different components is much lower for the 1996 FTSE 100 contract than the other three contracts. One possible explanation is the relative amounts of informed and noise investors trading in each contract. Chapter 4 indicated, (albeit over the period 1992-1996, rather than 1996 explicitly), that the relative proportions of noise traders to informed traders is greater for the FTSE 100 contract than the Long Gilt contract. Therefore, for the FTSE 100 contract, the probability of informed investors exploiting the market-maker is lower. However, even the unexpected component of volume, which we believe is information driven, has a relatively smaller impact on the spread, so this argument is difficult to defend vigorously. Another possible explanation is that the benefits of reduced inventory costs due to increases in volume are greatest for the 1996 FTSE 100 contract. A reduction in the time that a

market-maker has to hold unwanted assets puts downward pressure on the spread. For the other contracts the trade-off between reduced inventory costs and increased information costs is tipped in favour of the latter effect.

While the elasticities in Equations 2 and 3 are similar, when changes in the components of volume are considered in terms of the actual number of contracts traded, the results are quite dramatic. Suppose, for example that the mean expected and unexpected components of volume increase by one contract. What is the percentage impact on the spread of such an increase? Table 5.16 shows this increase as a percentage change in mean volume and the consequent percentage change in the spread.

These results clearly reveal that in real terms changes in unexpected volume have a much bigger impact on the spread than changes in expected volume. This suggests that while market-makers are relatively comfortable with variations in expected volume they appear to be very sensitive to any investors arriving at the market 'unexpectedly'. It is interesting to note how changes in unexpected volume of the Long Gilts contract have a much greater impact in 1996 than in 1986.

This suggests that as a market becomes more established, and trading follows more predictable patterns throughout the day, market-makers form relatively conservative expectations and so the shock of unexpected trading is that much more dramatic. This situation is reversed for the FTSE 100 contract with the impact being greater in 1986 than in 1996. This is less easy to explain. One possibility, if one also considers the impact of the expected component of volume, is that since this contract is two years younger than the Long Gilt contract in 1986, market-makers are still finding their way in terms of judging when investors will enter the market.

Table 5.16: The Relative Impacts on the Spread of Changes in Expected and
Unexpected Volume

Contract	FTSE 100		Long Gilts	
Year	1986	1996	1986	1996
Mean Expected Volume	36.583	319.402	302.596	1220.800
Mean Unexpected Volume	1.150	1.039	1.092	1.088
1) % Increase in Expected Volume*	2.735	0.313	0.331	0.082
2) % Increase in Unexpected Volume*	86.987	96.246	91.609	91.895
% Change in the Spread due to 1)	17.416	0.045	2.144	0.758
% Change in the Spread due to 2)	416.815	14.285	1007.695	1357.710

Note: * is the percentage change in volume due to the increase in trade of 1 contract.

Their spread setting is therefore likely to be very cautious as the higher mean levels of the spread in 1986 indicate. Further evidence of the uncertainty of the market-makers is also provided by the relatively high levels of volatility of the spread during this period.

The different magnitudes of the impact of the unexpected components of volume between the FTSE 100 and the Long Gilt contracts may also be attributed to the differences in the ability of market-makers to set spreads that can absorb variations in trade. Market-makers trading in the FTSE 100 contract may simply be more skilled at setting accommodating bid and ask prices. Another possible explanation is that in the high volume Long Gilt market there is greater competition between market-makers simply because there are likely to be more agents acting as scalpers. This will have the effect of driving down the spread, preventing the sort of flexible price setting that appears to exist in the FTSE 100 market.

The fact that unexpected trading has such a dramatic impact, particularly for the Long Gilt contract, should be interpreted as evidence that on the whole individuals follow predictable patterns in terms of their investment behaviour.

Ultimately, this analysis of the bid-ask spread equations has produced two very important results. The first is that the information costs of dealing with well-informed investors outweighs the benefits of high frequency transactions for all of the contracts considered here. This is contrary to much of the theoretical work which argues that the opposite is likely to be true. It helps to provide an explanation of the coincidence of high volume and high costs in intra-day trading that is not based on arguments of the inelastic demand of non-discretionary traders. These so-called 'noise' traders continue to play a very important role in the facilitation of trading but, as chapter 4 has already suggested, they should not be regarded as the driving force in these futures markets.

The second important result is that while the majority of trading in these contracts has a large element of predictability, unexpected levels of investment have a very significant impact on the market.

These two issues together should be considered seriously by both market-makers and market-regulators. The results above suggest that spreads are primarily determined by informed investors. If market-makers react strongly to increases in trading, particularly unexpected trading, the danger is that they may set spreads prohibitively wide[25]. This will have serious implications for the market. A parallel can be drawn with the overnight break between the closing and opening of the market. It is clear that the opening of the market represents a period of very heavy trading. This break in overnight trading can be effectively viewed as a trading halt.

---

[25] See Glosten and Milgrom (1985) for a similar argument.

234

Trading halts are imposed if prices move beyond predetermined limits within a given period. This policy is based on the assumption that 'excessive' price movements should be avoided. However, as pointed out in chapter 4, if these price movements merely reflect information flows then any artificial halt will reduce the price discovery role of the futures market. The evidence above suggests that information is accumulated during the break in trading and then exploited as soon as the market re-opens. A similar situation is likely to occur with an artificially imposed halt with the result that the movement that it was designed to suppress will have an even greater impact once the market again starts trading. It is therefore clear that market-makers and regulators need to be careful to avoid occasions which prohibit investors from achieving their demand objectives.

5.5.5.2 Volume Equation Analysis

The regression results for Equation 1 for each contract show the determination of the total volume of trade, by the bid-ask spread, own price volatility, and the price volatility and average price of the Short Gilt contract, as well as series of interval dummies. The first thing to notice is that for all of the contracts considered the Short Gilt variables do not have a significant impact on volume. This suggests that the Short Gilt contract is not a suitable substitute for either the FTSE 100 or the Long Gilt contracts. It also suggests that information is market specific. Although there may be market-wide information that affects more than one market, it does not have the same impact as news unique to a particular contract.

A particular mention should be made of the 1986 Long Gilt contract. When Equation 1 was initially estimated, although the GR-bar-squared value was relatively high, all of the variables appeared to be statistically insignificant. These results suggested that the equation had a multicollinearity problem. Therefore a Wald restriction test was carried out on the two Short Gilt variables under the null hypothesis that their impact is statistically negligible. The economic basis for this test was a suspicion that, in the early stages of the formation of the 1986 Long Gilt market, the behaviour patterns of investors and market-makers are somehow mirrored in each contract because both parties are, to some extent, feeling their way in the market. One might expect that such links are more likely to exist between two Gilt contracts rather than between a

Gilt contract and a contract on the FTSE 100. The joint test of zero restrictions produced a chi-square Wald statistic of 0.51. The critical value with two degrees of freedom is 5.99 at the 5% significance level. The null hypothesis therefore cannot be rejected. The results in table 5.13 represent the regression of Equation 1 with these two variables excluded.

The results also show that for all four contracts volatility is a significant determinant of volume. Interestingly, the link between these two variables is strongest for the 1986 contracts. This supports earlier suggestions that when a contract is relatively new the proportion of informed traders in the market is greater than those who are uninformed.

For the majority of contracts the dummy variables are generally insignificant which suggests that the specification of the equation is good in terms of explaining the intra-day variation in trading volume. It should be noted, however, that for the 1996 contracts the dummy variables that coincide with the period of APT trading remain significant. This suggests that there are other factors that have not been considered that explain the trading behaviour in this period. A more detailed investigation into the operation of automated exchanges would appear to be necessary.

It is also worth noting that, unlike the other two contracts, the dummy variables for the 1996 FTSE 100 contract and some of those for the 1986 contract remain significant. Equation 1 appears to explain the U-shape in the trading of the 1996 FTSE 100 contract that has already been identified, since the impact of the dummies is now relatively constant. The sign of the dummy variables shows that total volume is lower in every period relative to the opening of trade. The special nature of this period has already been discussed, but it is unclear what factor may cause this particularly even pattern in subsequent intervals. It is also unclear what might explain the significance of certain dummy variables in Equation 1 of the 1986 FTSE 100 contract.

The 1996 FTSE 100 contract is also unusual in that the diagnostics for Equation 1 revealed that a dynamic element should be considered in the specification of the model. The introduction of lagged volume appeared to solve this problem[26]. The significance of this variable reveals that volume in one period has a positive impact on volume in the next period. This is suggestive of persistent feedback effects in investor behaviour.

The impact of the spread on trading volume is, as expected, negative for all four contracts. A 1% increase in the spread results in a 1.70% fall in the total volume of trading of the 1986 FTSE 100 contract. A similar increase results in falls in total volume of 0.20%, 5.42% and 2.04% for the 1996 FTSE 100, the 1986 Long Gilt and the 1996 Long Gilt contracts respectively. This illustrates the important role that costs play in determining the volume of trade; a point that will be returned to later when the issue of regulation and costs is again considered.

The regression results of Equation 2 demonstrate the impact of our set of key variables on the expected component of volume. The two issues of insignificant variables and dynamic behaviour, relating to the 1986 Long Gilt and the 1996 FTSE 100[27] contracts respectively, also apply here. The results reveal that for all four contracts volatility is an important determinant of expected volume. As in Equation 1, the Short Gilt variables do not help to explain the variations in volume. There have, however, been some changes with regard to the dummy variables. The lower levels of volume during the periods of APT trading in 1996 are still not explained by the specification of this model. The intra-day variation in expected volume of the 1986 Long Gilt contract appears to be fully described by the changes in the spread and price volatility. The pattern in the dummies of Equation 1 of the 1996 FTSE 100 contract, also remains when the regressand is expected volume. The main changes occur for the other two contracts. A lot more of the dummy variables for the 1996 Long Gilt contract are now significant and also some of those for the 1986 FTSE 100 contract. The interesting

---

[26] The Hausman (1978) specification test reveals that this variable does not alter the original result of no simultaneity.
[27] The Hausman (1978) specification test reveals that this variable does not alter the original result of no simultaneity.

feature is that their signs are positive. This suggests that there is some unidentified factor that results in some part of expected volume, on the open-outcry exchange, being higher than during the opening period of trading.

The impact of the spread on the expected component of volume is significantly positive. An increase of 1% in the spread set by market-makers trading the 1986 FTSE 100 contract results in a 0.65% fall in expected volume. A similar increase results in falls in expected volume of 0.10%, 3.44% and 0.94% for the 1996 FTSE 100, the 1986 Long Gilt and the 1996 Long Gilt contracts respectively. These results suggest that expected volume is less responsive to changes in the spread than total volume. This supports the argument put forward by Brock and Kleidon (1992) that there are benefits to trading at particular times of the day that outweigh the costs imposed by the bid-ask spread.

The determination of the unexpected component of volume is described by the regression results of Equation 3. The dynamic variable issue relating to the 1996 FTSE 100 contract does not arise in Equation 3. This is an expected result since by definition unexpected volume in one period is unlikely to affect unexpected volume in another period. However, the multicollinearity in the specification of the volume equation for the 1986 Long Gilts contract is still an issue.

All four contracts show that a positive relationship exists between volatility and unexpected volume, but the impact of the Short Gilt variables is again insignificant. It should also be noted that for all of the contracts, with the exception of the 1986 Long Gilt contract, a large number of the dummy variables are significant. This characterises all of the volume specifications, but unlike Equation 2 the dummy variables are negatively signed. The periods of APT trading clearly require further research to understand the determinants of trading volume. However, the same could be said of the periods when contracts are traded by open-outcry. The variables in each model help to explain the U-shaped pattern of intra-day investor behaviour, but they cannot account for all of the differences between intervals.

The impact of the spread on unexpected volume is negative for all of the contracts. A 1% increase in the spread of the 1986 FTSE 100 contract results in a 1.05% fall in the level of unexpected volume. A similar increase in the spread leads to falls in unexpected volume of 0.06%, 1.97%, and 1.09% for the 1996 FTSE 100, the 1986 Long Gilt and the 1996 Long Gilt contracts respectively. The comparison of these values with those of the falls in expected volume due to increases in the spread, suggest that unexpected volume is less sensitive to costs. One possible explanation is that the rewards from holding news outweigh the costs of carrying out a transaction.

This set of results provides information on some very important issues. It supports the work carried out in chapters 2 and 4 that volume and volatility are related. It also raises an interesting point with regard to regulation. All three equation specifications for each contract reveal that the costs imposed on the individual due to the bid-ask spread are a significant determinant of the volume of trade. In fact, any increases in these costs will reduce the number of contracts traded in the market. Regulators could use the elasticities provided here to judge the impact of additional costs, for example the increase in transaction fees, on volume. This assumes of course that investors react to costs such as transaction fees in the same way that they react to the costs due to the spread. This does not seem an unreasonable assumption since brokers will normally quote a single commission fee to an investor rather than break it up into its various components. Table 5.17 uses the elasticities of total volume with respect to the spread for the 1996 contracts to show how a £0.20 increase in transaction costs, imposed by a market regulator, will affect the volume of trade. LIFFE typically charge a fee per contract exchanged to those who are not members of the exchange. At present this stands at £0.25. Investors acting through brokers will then bear the burden of any increases as part of the commission fee.

Information on commission fees is not easy to obtain. It is even more difficult to obtain this information for 1996. A brief survey revealed charges, in 1999, of between £15 and £25 pounds. Let us assume that the average cost in 1996 was £15.

Table 5.17: Estimates of the Impact of Increased Transaction Costs on Volume

Contract	Mean Total Volume	Increase as a % of Total Costs	Elasticity	Change in Total Volume
FTSE 100	358.771	1.33	-0.196	0.935
Long Gilt	1420.300	1.33	-2.036	38.460

Table 5.17 shows that an increase of £0.20 in commission fees would result in a fall in total volume of just less than one FTSE 100 contract and just over 38 Long Gilt contracts per 30-minute interval. Regulators can use this information to balance losses in revenue due to falls in volume with gains in revenue due to the imposition of increased charges. It should be noted, however, that the burden of charges is not equal among investors on a futures exchange, so these calculations would be more complicated than those in this illustrative example. This should not detract from the importance of these results in allowing regulators to see that the benefits of tighter regulation, in terms of increased revenues, may have serious consequences in terms of the impact on volume. The success of any contract is dependent on the amount of trade that it generates. As London's status as a financial centre, and in particular the position of LIFFE, comes under pressure from the increasingly competitive European markets, holding onto and attracting investors becomes of crucial importance. The issue of costs in ensuring that business is not lost could not be more relevant.

5.6 CONCLUSION

Chapter 4 investigated the relationship between the volume of trade and return price volatility that had been discovered in chapters 2 and 3. The specification of the model of the Mixture of Distributions Hypothesis exploited in that chapter, uses as its basis a microstructure model based on the relationship between investors and market-makers. The discovery that volume and volatility are linked by a common directing variable and that the majority of trading is driven by information raised some important issues. In particular, if the difference between the bid and the ask price set by the market-maker represents part of the cost of trading, what is the role of volume in the determination of these costs?

With this question in mind, in this chapter, an extensive investigation has been carried out into the relationship between the volume of trade and the bid-ask spread. The key empirical points have been the use of transactions data for two UK futures market, analysis of intra-day trading patterns, the reaction of market-makers to unexpected levels of trading, and the impact of the spread on volume as well as of volume on the spread.

This in-depth analysis has allowed us to make some very interesting discoveries. The intra-day plots of the data suggest that there is a U-shape in both volume and the spread during normal trading hours. They also suggest that the periods of APT trading are unlike the rest of the day; characterised by low costs and low volume.

This positive relationship between our two key variables was supported by the regression analysis. In the markets for the two assets investigated in this study, the market-makers appear to regard the increased probability of trading with better informed traders as the most important factor (with regard to volume) in the determination of their prices. This rejects the commonly held view that it is the reduced inventory costs of increased volume that are the major determinant of prices.

Unsurprisingly, investors are also sensitive to costs. Analysis of the impact of the spread on volume shows that the two variables are negatively related. If costs rise, as proxied by the spread, fewer contracts are traded.

The results also suggest that as the market for a contract matures, patterns of trade become more established. The distinction between the expected and unexpected components of volume allowed us to show that market-makers are very sensitive to unexpected levels of trading; their sensitivity increasing with time.

These results also have very important implications with regard to the successful operation of UK futures markets. Section 5.5 has already discussed the dangers of market-makers who are overly sensitive to unexpected levels of trading. Restricting volume by imposing artificial trading halts is only likely to reduce the efficient

241

functioning of the market. Section 5.5 also shows how it is possible for market practitioners to judge the impact of increasing costs.

These practical issues have particular relevance in the increasingly competitive derivatives markets. LIFFE has been accused of arrogance in assuming that it could maintain its position within Europe as the number one futures and options market[28]. Recent events, particularly concerning German treasury bond futures, have revealed that LIFFE cannot afford to be complacent. It needs to continue to attract investors. The issue of cost is, therefore, of vital importance.

This study also has interesting implications with regard to research issues. As already mentioned, it questions the bias towards inventory cost models that prevails in this field. In line with the work of Chapter 4, it reinforces the movement in the microstructure literature towards models based on information costs. It also suggests that there are a number of areas that demand further investigation. The period of APT trading is clearly different to the rest of the trading day. This issue is important as more markets become fully automated. Indeed, the patterns of trading may alter significantly from those documented in this chapter. It will be interesting to see if the theoretical issues discussed by O'Hara (1997) still apply in this environment.

Identifying all of the patterns of trading is not a simple exercise. In a number of cases in this study the continued significance of the dummy variables reveals that we have not been able to account for all aspects of the trading process. The data for UK markets is only gradually revealing the sort of detail that would allow us to explore issues of trading behaviour that have already been possible for US markets.

---

[28] In 1990 LIFFE was the biggest market (as measured by volume) for futures and options in the world outside the US.

For example, measures of direct costs and data on the numbers of market-makers, would both have improved this study. We also know relatively little about the activities of scalpers in UK markets. Our knowledge is derived from US-based work. It may be incorrect to assume that they are directly related.

Ultimately, however, the achievement of this chapter has been to investigate, in detail, issues that have not previously been studied in any great depth. This insight into the operation of futures markets should be of interest to the academic as well as the practitioner and the regulator.

CHAPTER SIX: CONCLUSION

6.1 OVERVIEW OF THE THESIS

A measure of the success of a futures contract is often taken to be the amount of trade that it attracts. The assumption that an asset can be judged in this way is made with little appreciation of what the volume of trade actually represents. The main motivation for the work carried out in this thesis, therefore, is to obtain a better understanding of the role and impact of volume. The four empirical chapters included here use UK futures markets as a basis to investigate the relationship between volume and price volatility, the links between volume and the cost of trading, and the role of volume in describing the precision and dispersion of information.

Chapter 2 used two well-established techniques to carry out a preliminary examination of the relationship between price volatility and the volume of trade. The term preliminary is used to reflect the fact that although the results are supportive of the underlying theories, we do not set up an hypothesis to test explicitly their credibility. This is in contrast to the large number of empirical studies in this field that do not appear to view this as a problem.

The principal findings of chapter 2 are the discovery of a contemporaneous relation between volume and volatility and the use of volume to account for the non-normality in futures price returns. The results also suggest that it is important to account for trends in the data, particularly those due to contract expiration and the exogenous growth in the popularity of derivatives trading. The underlying feature, although not proven, is the role of information in defining the volume-volatility relation.

Chapter 3 looked at volume from a slightly different angle and investigated the role of volume in determining the precision and dispersion of information. The simple scatter plot technique revealed that it is very difficult to model volume data in terms of information precision. However, assuming a given level of precision, it was possible to show that the dispersion levels for all of the five UK futures contracts considered

are very high. These results suggested that the majority of investors carry information which is contrary to the popular view of markets dominated by feedback traders.

Although these first two empirical chapters represent an interesting beginning they do not, particularly with reference to chapter 2, tell us why the link between volume and volatility exists. This is a failing of the majority of empirical studies in this field. Chapter 4, therefore, carried out a direct test of one of the underlying theories of this relation: the Mixture of Distributions Hypothesis. The results were quite striking and revealed that the driving force behind the volume-volatility relationship is the flow of information, thus supporting the tentative conclusions of chapter 2. It was also possible to identify the noise and informed components of volume which indicated that for the three contracts considered the latter effect dominates, in line with the results of chapter 3. Chapter 4 also provided further evidence of the importance of accounting for trends in the data.

Chapter 5 represents an amalgamation of the concepts considered in the earlier chapters and uses as its foundation the idea of a symbiotic relationship between market-makers and investors. More specifically, it investigated the role of volume in the determination of transaction costs as measured by the bid-ask spread. Unlike the majority of studies in this field, it also considered how changes in the spread affect trading decisions. Using high-frequency transaction data for two UK financial futures contracts, a number of interesting discoveries were made. It was found that, at the intra-day level, there is evidence of a U-shape pattern in both volume and the spread. Both variables appeared to be at their peak at the open and close of the normal trading day. The period of computer based trading is unique and is characterised by small spreads and low levels of volume.

The regression results revealed that increases in volume have a positive impact on the spread. This rejects the commonly held view that it is the reduced inventory costs of increased volume that are the major determinants of bid and ask prices. In terms of the impact of the spread on levels of trading, it was found that investors are sensitive to costs with the two variables being negatively related. The results also showed that patterns of trade become more established with the length of time that a contract has

been traded. Time is also a factor in the response of market-makers to unexpected levels of trading. By distinguishing between the expected and unexpected components of volume it was possible to show that market-makers are very sensitive to changes in the latter variable.

6.2 THE IMPORTANCE OF THESE RESULTS

These four empirical chapters together represent a very important set of results. They have addressed a number of weaknesses in the existing literature and provided an insight into the role and impact of volume in UK futures markets that has previously not been available.

This is the first study to establish, using a direct test of the theory, that in UK futures markets it is the flow of information that drives the relationship between price volatility and volume. This is important because it allows us to accurately interpret the distribution of price returns. In addition, it allows us to discriminate between the various theories of market structure. If the majority of investors are informed, as the results suggest, then there needs to be a reconsideration of the view that futures markets are home to a casino culture. Market regulators need to be aware that any artificial restrictions imposed on volume, or price movements, in the naive belief that they must have a destabilising influence on the market, may simply serve to limit its ability to fulfil its role in terms of price discovery. Although uncertainty is crucial to the existence of futures markets, increasing the element of risk may only serve to encourage the sort of gambling behaviour that regulators wish to avoid.

Although this study is by no means the first to discover the existence of non-normality in returns series, it adds to those suggesting cautious use of the central limit theorem. The inability to exploit this econometric tool has widespread implications for empirical work.

The discovery that there is information inherent in the volume statistic is important, not only because of what it tells us about the balance between relatively informed and uninformed traders, and hence the trading process. In the spirit of the Blume et al. (1994) study, it also indicates that those involved in technical analysis who use

246

patterns in the volume of trade to form their demand schedules, are at a distinct advantage in comparison to those who consider prices in isolation.

The analysis of the relationship between volume and the bid-ask spread is particularly revealing in that the information costs of increased volume appear to dominate the reduced inventory costs. The majority of studies that consider these specific issues argue in favour of the latter effect dominating. This result, in addition to the finding that market-makers are very sensitive to unexpected levels of trading, also has implications with regard to artificially imposed trading halts. The likelihood is that they will only result in the market failing to function efficiently.

The results of the impact of costs on the volume of trade have important policy implications at a time when futures exchanges are operating in an increasingly competitive environment. If the aim is to provide liquidity at low cost market monitors need to be aware that there is a trade-off. Chapter 5 gives some clear guidance as to how volume varies with changes in costs. Practitioners could use these figures to help them design a cost structure that minimises the loss of investment that LIFFE can ill-afford.

6.3 RESEARCH ISSUES

The work in this thesis has raised a number of research issues, some of which may provide the impetus for future work. The movement in the microstructure literature towards the development of information, rather than inventory, based models appears to be well-founded in the context of this study. Where the analysis of chapter 4 fails is in being unable to describe the dynamic nature of the trading process. Our understanding of the volume-volatility relationship is based on static models. In particular, it would be interesting to investigate how the informed and uninformed components of volume vary over time.

The importance of accounting for trends in the data and the successful use of the Holmes-Rougier (1997) roll-over adjustment suggests that it may be worth revisiting the work carried out in chapters 2 and 3 to incorporate this technique.

247

It would be nice to carry out this investigation over a greater selection of futures contracts. Transaction data has only recently become available for UK markets and in some cases the incompatibility of the data has made it impossible to use as many contracts as we would have liked. This study could also be improved if data on, for example, the number of market-makers and direct transaction costs was available.

Another possible extension would be to consider volume linkages across exchanges. This study provided some evidence, in chapter 4, that information is common to more than one market. It would be interesting to look at the patterns of trading across markets and whether the impact of volume differs, particularly where contracts are quoted on more than one exchange. There is also a need to investigate how these results translate to the underlying spot markets. The different nature of the trading process in equities would allow insightful comparisons to be made.

The apparently idiosyncratic nature of the period of automated trading also deserves further investigation, particularly as more exchanges abandon the traditional open-outcry system.

Ultimately, the achievement of this thesis is an in-depth understanding of the role of the volume of trade and its impact on UK futures markets that should be of interest to the academic and the practitioner. Weaknesses in the existing literature have been addressed and new issues raised in what deserves to be an important area of research.

BIBLIOGRAPHY

Admati, A. and Pfleiderer, P. (1988) "A Theory of Intraday Patterns: Volume and Price Variability", The Review of Financial Studies, 1, 3-40.

Admati, A. and Pfleiderer, P. (1992) " Volume in Capital Markets", in The New Palgrave Dictionary of Money and Finance, P. Newman, M. Milgate and J Eatwell eds., London: Macmillan.

Affleck-Graves, J., Hedge, S.P. and Miller, R.E. (1994) "Trading Mechanisms and the Components of the Bid-Ask Spread", Journal of Finance, 49, 1471-1488.

Anderson, T. (1996) "Return Volatility and Trading Volume in Financial Markets: An Information Flow Interpretation of Stochastic Volatility", Journal of Finance, 51, 169-204.

Andrews, D.W.K. (1991) "Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation", Econometrica, 59, 817-858.

Antoniou, A. and Holmes, P. (1995) "Futures Trading, Information and Spot Price Volatility: Evidence for the FTSE-100 Stock Index Futures Contract Using GARCH", Journal of Banking and Finance, 19, 117-29.

Bachelier, L. (1900) "Theorie de la speculation", Annales de l'Ecole Normale Superieure, 17, 21-86.

Bagehot, W. (1971) "The Only Game in Town", Financial Analysts Journal, 2.

Barclay, M.J., Litzenberger, R.H. and Warner, J.B. (1990) "Private Information, Trading Volume, and Stock-Return Variances", Review of Financial Studies, 3, 233-253.

Barr, D.G. (1997) Econometrics for Finance: A Users' Manual, Brunel University: unpublished.

Bera, A. and Higgins, M.L. (1993) "Arch Models: Properties, Estimation, and Testing", Journal of Economic Surveys, 7, 305-366.

Bessembinder, H. and Seguin, P.J. (1992) "Futures-Trading Activity and Stock Price Volatility", Journal of Finance, 47, 2015-2034.

Bessembinder, H. and Seguin, P.J. (1993) "Price Volatility, Trading Volume, and Market Depth: Evidence from Futures Markets", Journal of Financial and Quantitative Analysis, 28, 21-39.

Bhattacharya, M. (1983) "Transaction Data Tests of Efficiency of the Chicago Board of Options Exchange", Journal of Financial Economics, 12, 161-185.

Blume, L., Easley, D. and O'Hara, M. (1994) "Market Statistics and Technical Analysis: The Role of Volume", Journal of Finance, 49, 153-181.

Bollerslev, T. (1986) "Generalised Autoregressive Conditional Heteoscedasticity", Journal of Econometrics, 31, 307-327.

Box, G.E.P. and Pierce, D.A. (1970) "Distribution of Residual Autocorrelations in Autoregressive-Integrated-Moving Average Time Series Models", Journal of the American Statistical Association, 65, 1509-26.

Brock, W.A. and Kleidon, A.W. (1992) "Periodic Market Closure and Trading Volume", Journal of Economic Dynamics and Control, 16, 451-489.

Brockwell, P.J. and Davis R.A. (1987) Time Series: Theory and Methods, New York: Springer-Verlag.

Brooks, C. (1998) "Predicting Stock Index Volatility: Can Market Volume Help?", Journal of Forecasting, 17, 59-80.

Brown, D.P. and Jennings, R.H. (1989) "On Technical Analysis", Review of Financial Studies, 2, 527-552.

Carlton, D.W. (1984) "Futures Markets: Their Purpose, Their History, Their Growth, Their Successes and Failures", Journal of Futures Markets, 4, 237-271.

Chan, K., Chung, P. and Johnson, H. (1995) "The Intraday Behaviour of Bid-Ask Spreads for NYSE Stocks and CBOE Options", Journal of Financial and Quantitative Analysis, 30, 329-346.

Charemza, W.W. and Deadman, D.F. (1997) New Directions In Econometric Practice, Cheltenham: Edward Elgar, 2nd Edition.

Choi, H. and Subrahmanyam, A. (1994) "Using Intraday Data to Test for Effects of Index Futures on the Underlying Stock Markets", Journal of Futures Markets, 14, 293-322.

Clark, P. (1973) "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices", Econometrica, 41, 135-55.

Copeland, T.E. (1976) "A Model of Asset Trading under the Assumption of Sequential Information Arrival", Journal of Finance, 31, 1149-1168.

Copeland, T.E. and Galai, D. (1983) "Information Effects on the Bid-Ask Spread", Journal of Finance, 38, 1457-1469.

Daigler, R.T. (1997) "Intraday Futures Volatility and Theories of Market Behaviour", Journal of Futures Markets, 17, 45-74.

Darnell, A.C. and Evans, J.L. (1990) The Limits of Econometrics, Aldershot: Edward Elgar.

De Bondt, W.F.M. and Thaler, R.H. (1985) "Does the Stock Market Overreact?", Journal of Finance, 40, 793-805.

De Long, J.B., Shleifer, A., Summers, L.H. and Waldmann, R.J. (1990) "Positive Feedback Investment Strategies and Destabilising Rational Speculation", Journal of Finance, 45, 379-395.

Demos, A.A. and Goodhart, C.A.E. (1996) "The Interaction Between the Frequency of Market Quotations, Spread and Volatility in the Foreign Exchange Market", Applied Economics, 28, 377-386.

Demsetz, H. (1968) "The Cost of Transacting", Quarterly Journal of Economics, 82, 33-53.

Dickey, D.A. and Fuller, W.A. (1979) "Comparing Alternative Tests of Causality in Temporal Systems", Journal of Econometrics, 21, 161-194.

Easley, D. and O'Hara, M. (1987) "Price, Trade Size and Information in Securities Markets", Journal of Financial Economics, 19, 69-90.

Edwards F.R. and Ma C.W. (1992) Futures and Options, Singapore: McGraw-Hill.

Ekman, P.D. (1992) "Intraday Patterns in the S&P 500 Index Futures Market", Journal of Futures Markets, 12, 365-381.

Engle, R.F. (1982) "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", Econometrica, 50, 987-1007.

Engle, R.F. and Bollerslev, T. (1986) "Modelling the Persistence of Conditional Variances", Econometric Reviews, 5, 1-50.

Epps, T.W. (1975) "Security Price Changes and Transaction Volumes: Theory and Evidence", American Economic Review, 65, 586-597.

Epps, T.W. and Epps, M.L. (1976) "The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implication for the Mixture-of-Distributions Hypothesis", Econometrica, 44, 305-321.

Fama, E.F. and French, K.R. (1988) "Permanent and Temporary Components of the Stock Market", Journal of Political Economy, 96, 246-273.

Feller, W. (1971) <u>An Introduction to Probability Theory and its Applications, Volume II</u>, Toronto: John Wiley and Sons Inc, 2nd Edition.

Followill, R.A. and Rodriguez, A.J. (1991) "The Estimation and Determinants of Bid-Ask Spreads in Futures Markets", <u>Review of Futures Markets</u>, 10, 339-352.

Foster, A. (1996) "Volume-Volatility Relationships for Crude-Oil Futures Markets", <u>Journal of Futures Markets</u>, 15, 929-951.

Foster, F.D. and Viswanathan, S. (1990) "A Theory of the Inter-Day Variations in Volume, Variance, and Trading Costs in Securities Markets", <u>Review of Financial Studies</u>, 3, 593-624.

Foster, F.D. and Viswanathan, S. (1993) "Variations in Trading Volume, and Trading Costs: Evidence on Recent Price Formation Models", <u>Journal of Finance</u>, 48, 187-211.

Foster, F.D. and Viswanathan, S. (1994) "Strategic Trading with Asymmetrically Informed Traders and Long-Lived Information", <u>Journal of Financial and Quantitative Analysis</u>, 29, 499-518.

Foster, F.D. and Viswanathan, S. (1995) "Can Speculative Trading Explain the Volume-Volatility Relation?", <u>Journal of Business and Economic Statistics</u>, 13, 379-396.

French, K.R. and Roll, R. (1986) "Stock Return Variances: The Arrival of Information and the Reaction of Traders", <u>Journal of Financial Economics</u>, 17, 5-26.

Fujihara, R.A. and Mougoué, M. (1997) "An Examination of Linear and Nonlinear Causal Relationships Between Price Variability and Volume in Petroleum Futures Markets", <u>Journal of Futures Markets</u>, 17, 385-416.

Gallant, A.R., Rossini, P.E., and Tauchen, G. (1992) "Stock Prices and Volume", <u>Review of Financial Studies</u>, 5, 199-242.

Gallant, A.R. and Tauchen, G. (1989) "Semi-nonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications", <u>Economica</u>, 57, 1091-1120.

Garman, M. and Klass, M. (1980) "On the Estimation of Security Price Volatilities from Historical Data", <u>Journal of Business</u>, 53, 67-78.

George, T.J. and Longstaff, F.A. (1993) "Bid-Ask Spreads and Trading Activity in the S&P 100 Index Options Market", <u>Journal of Financial and Quantitative Analysis</u>, 28, 381-397.

Gerber, S. (1996) "A Measure of the Bid-Ask Spread in the Secondary Market", Applied Economics, 28, 363-370.

Geweke, J., Meese, R., and Dent, W. (1983) "Comparing Alternative Tests of Causality in Temporal Systems", Journal of Econometrics, 21, 161-94.

Glosten, L.R. and Harris, L.E. (1988) "Estimating the Components of the Bid/Ask Spread", Journal of Financial Economics, 21, 123-142.

Glosten, L.R. and Milgrom, P.R. (1985) "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders", Journal of Financial Economics, 14, 71-100.

Goss, B.A. and Avsar, S.G. (1998) "Increasing Returns to Liquidity in Futures Markets", Applied Economic Letters, 5, 105-109.

Grammatikos, T. and Saunders, A. (1986) "Futures Price Variability: A Test of Maturity and Volume Effects", Journal of Business, 59, 319-330.

Granger, C.W.L. (1969) "Investigating Causal Relations by Econometric Models and Cross Spectral Methods", Econometrica, 38, 424-438.

Grossman, S.J. and Stiglitz, N. (1980) "On the Impossibility of Informationally Efficient Markets", American Economic Review, 70, 393-408.

Grundy, B.D. and McNichols, M. (1989) "Trade and Revelation of Infomation Through Prices and Direct Disclosure", Review of Financial Studies, 2, 495-526.

Gujarati, D.M. (1995) Basic Econometrics, Singapore: McGraw-Hill.

Gwilym, O.A.P. and Buckle, M. (1996) "An Analysis of Bid-Ask Spreads on American- and European-Style Index Options", Applied Economics Letters, 3, 445-449.

Hansen, L.P. (1982) "Large Sample Properties of Generalised Method of Moments Estimators", Econometrica, 50, 1029-1054.

Harris, L. (1986) "Cross-Security Tests of the Mixture of Distributions Hypothesis", Journal of Financial and Quantitative Analysis, 21, 39-46.

Harris, L. (1987) "Transaction Data Tests of the Mixture of Distributions Hypothesis", Journal of Financial and Quantitative Analysis, 22, 127-141.

Harris, L. and Gurel, E. (1986) "Price and Volume Effects Associated with Changes in the S&P 500 List: New Evidence for the Existence of Price Pressures", Journal of Finance, 41, 815-829.

Harvey, A.C. (1981a) The Econometric Analysis of Time Series, Oxford: Phillip Allan.

Harvey, A.C. (1981b) Time Series Models, Oxford: Philip Allan.

Harvey, A.C. (1994) Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge: Cambridge University Press.

Harvey, A.C. and Peters, S. (1990) "Estimation Procedures for Structural Time Series Models", Journal of Forecasting, 9, 89-108.

Harvey, A.C. and Todd, P.H.J. (1983) "Forecasting Economic Time Series With Structural and Box-Jenkins Models: A Case Study", Journal of Business and Economic Statistics, 1, 299-315.

Hasbrouck, J. (1988) "Trades, Quotes, Inventories, and Information", Journal of Financial Economics, 22, 229-252.

Hasbrouck, J. (1991) "Measuring the Information Content of Stock Trades", Journal of Finance, 46, 179-207.

Hausman, J.A. (1978) "Specification Tests In Econometrics", Econometrica, 46, 1251-1271.

Hendry, D.F. (1979) "Predictive Failure and Econometric Modelling in Macroeconomics", in Ormerod, P. (ed.) Economic Modelling, London: Heinemann.

Hendry , D.F. and Mizon, G.E. (1978) "Serial Correlation as a Convenient Simplification Not a Nuisance", Economic Journal, 88, 549-563.

Hiemstra, C. and Jones, J.D. (1994) "Testing for Linear and Non-Linear Granger Causality in the Stock Price-Volume Relation", Journal of Finance, 49, 1639-1664.

Hillmer, S.C. and Tiao, G.C. (1982) "An ARIMA-Model-Based Approach to Seasonal Adjustment", Journal of the American Statistical Association, 77, 63-70.

Ho, T. and Stoll, H.R. (1981) "Optimal Dealer Pricing Under Transactions and Return Uncertainty", Journal of Financial Economics, 9, 47-73.

Holden, C.W. and Subrahmanyam, A. (1992) "Long-Lived Private Information and Imperfect Competition", Journal of Finance, 47, 247-270.

Holmes, P. and Rougier, J. (1997) "Trading Volume and Contract Rollover in Stock Index Futures", Discussion Paper, Durham University.

Huang, R.D. and Stoll, H.R. (1997) "The Components of the Bid-Ask Spread: A General Approach", Review of Financial Studies, 10, 995-1034.

Intriligator, M.D. (1978) Econometric Models, Techniques and Applications, New Jersey: Prentice-Hall.

Jain, P.C. and Joh, G.H. (1988) "The Dependence Between Hourly Prices and Trading Volume", Journal of Financial and Quantitative Analysis, 23, 269-283.

Jennings, R.H. and Barry, C. (1983) "Information Dissemination and Portfolio Choice", Journal of Financial and Quantitative Analysis, 18, 1-19.

Jennings, R.H., Starks, L.T. and Fellingham, J.C. (1981) "An Equilibrium Model of Asset Trading with Sequential Information Arrival", Journal of Finance, 36, 143-161.

Jones, C.M., Kaul, G. and Lipson, M.L. (1994) "Information, Trading, and Volatility", Journal of Financial Economics, 36, 127-154.

Jordan, J.V., Seale, W.E., Dinehart, W.E. and Kenyon, D.E. (1988) "The Intra-Day Variability of Soybean Futures Prices: Information and Trading Effects", Review of Futures Markets, 1, 97-108.

Kalman, R.E. (1960) "A New Approach to Linear Filtering and Prediction Theory", Journal of Basic Engineering, Transactions ASME, Series D 82, 35-45.

Kalman, R.E. and Bucy, R.S. (1961) "New Results in Linear Filtering and Prediction Theory", Journal of Basic Engineering, Transactions ASME, Series D 83, 95-108.

Karpoff, J. (1986) "A Theory of Trading Volume", Journal of Finance, 41, 1069-1087.

Karpoff, J. (1987) "The Relation Between Price Changes and Trading Volume: A Survey", Journal of Financial and Quantitative Analysis, 22, 109-26.

Kim, O. and Verrecchia, R.E. (1991) "Market Reactions to Anticipated Announcements", Journal of Financial Economics, 30, 273-310.

Kolb, R.W. (1988) Understanding Futures Markets, Illinois: Scott, Foresman and Company, 2nd Edition.

Krinsky, I. and Lee, J. (1996) "Earnings Announcements and the Components of the Bid-Ask Spread", Journal of Finance, 51, 1523-1535.

Kuserk, G.J. and Locke, P.R. (1993) "Scalper Behaviour in Futures Markets: An Empirical Investigation", Journal of Futures Markets, 13, 409-431.

Kyle, A. (1985) "Continuous Auctions and Insider Trading", Econometrica, 53, 1315-35.

Lakonishok, J., and Smidt, S. (1989) "Past Price Changes and Trading Volume", The Journal of Portfolio Management, 15, 18-24.

Lamoureux, C. and Lastrapes, W. (1990) "Heteroscedasticity in Stock Return Data: Volume Versus GARCH Effects", Journal of Finance, 45, 221-229.

Lamoureux, C. and Lastrapes, W. (1994) "Endogenous Trading Volume and Momentum in Stock-Return Volatility", Journal of Business and Economic Statistics, 12, 253-260.

Lang, L.H.P., Litzenberger, R.H. and Madrigal, V. (1992) "Testing Financial Market Equilibrium Under Asymmetric Information", Journal of Political Economy, 100, 2, 317-348.

Lapper, R. (1995a) "Survey of International Capital Markets: Caution Prevails-Derivatives", Financial Times Article, 1 June.

Lapper, R. (1995b) "Survey of Derivatives: Evolution in the Shadow of Disaster", Financial Times Article, 16 November.

Lapper, R and Morse L. (1995) "International Capital Markets: Record Trading Volume for LIFFE and Matif in 1994-Derivatives", Financial Times Article, 5 January.

Laux, P.A. (1995) "Dealer Market Structure, Outside Competition, and the Bid-Ask Spread", Journal of Economic Dynamics and Control, 19, 683-710.

Laux, P.A. and Senchack, A.J. (1992) "Bid-Ask Spreads in Financial Futures", Journal of Futures Markets, 12, 621-634.

Leng, H. (1996) "Announcement Versus Nonannouncement: A Study of Intraday Transaction Price Paths of Deutsche Mark and Japanese Yen Futures", Journal of Futures Markets, 16, 829-857.

Ljung, G.M. and Box G.E.P. (1978) "On a Measure of Lack of Fit in Time Series Models", Biometrika, 65, 297-303.

Locke, P.R. and Sayers, C.L. (1993) "Intraday Futures Price Volatility - Information Effects and Variance Persistence", Journal of Applied Econometrics, 8, 15-30.

Locke, P.R. and Venkatesh, P.C. (1997) "Futures Market Transaction Costs", Journal of Futures Markets, 17, 229-245.

Lütkepohl, H. (1993) <u>Introduction to Multiple Time Series Analysis</u>, Berlin: Springer-Verlag, 2nd Edition.

Ma, C.K., Peterson, R.L. and Sears, R.S. (1992) "Trading Noise, Adverse Selection, and Intraday Bid-Ask Spreads in Futures Markets", <u>Journal of Futures Markets</u>, 12, 519-538.

Madhavan, A. (1992) "Trading Mechanisms in Securities Markets", <u>Journal of Finance</u>, 47, 607-641.

Malliaris, A.G. and Urrutia, J.L. (1998) "Volume and Price Relationships: Hypothesis Testing for Agricultural Futures", <u>Journal of Futures Markets</u>, 18, 53-72.

Mandelbrot, B. (1963) "The Variation of Certain Speculative Prices", <u>Journal of Business</u>, 36, 394-419.

Martell, T.F. and Wolf, A.S. (1987) "Determinants of Trading Volume in Futures Markets", <u>Journal of Futures Markets</u>, 7, 233-244.

McCarthy, J. and Najand, M. (1993) "State Space Modelling of Price and Volume Dependence: Evidence from Currency Futures", <u>Journal of Futures Markets</u>, 13, 335-344.

Merton, R.C. (1971) "Optimum Consumption and Portfolio Rules in a Continuous-Time Model", <u>Journal of Economic Theory</u>, 3, 373-413.

Miller, M. (1991) <u>Financial Innovations and Market Volatility</u>, Oxford: Blackwell.

Morgan, I.G. (1976) "Stock Prices and Heteroscedasticity", <u>Journal of Business</u>, 49, 496-508.

Najand, M. and Yung, K. (1991) "A GARCH Examination of the Relationship Between Volume and Price Variability in Futures Markets", <u>Journal of Futures Markets</u>, 11, 613-621.

Newey, W.K. and West, K.D. (1987) "A Simple, Positive Semi-Definite, Heteroscedasticity Consistent Covariance Matrix", <u>Econometrica</u>, 55, 703-708.

O'Hara, M. (1997) <u>Market Microstructure Theory</u>, Massachusets: Blackwell.

Pesaran, M.H. and Smith, R.P. (1994) "A Generalised R-Bar-Squared Criterion for Regression Models Estimated by the Instrumental Variables Method", <u>Econometrica</u>, 62, 705-710.

Pfleiderer, P. (1984) "The Volume of Trade and the Variability of Prices: a Framework for Analysis in Noisy Rational Expectations Equilibria", Working Paper, Stanford Univ.

Pindyck, R.S. and Rubenfeld, D.L. (1991) Econometric Models and Econometric Forecasts, New York: McGraw-Hill, 3rd Edition.

Richardson, M. and Smith, T. (1994) "A Direct Test of the Mixture of Distributions Hypothesis: Measuring the Daily Flow of Information", Journal of Financial and Quantitative Analysis, 29, 101-116.

Roll, R. (1984a) "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market", Journal of Finance, 39, 1127-1139.

Roll, R. (1984b) "Orange Juice and Weather", American Economic Review, 74, 861-880.

Rougier, J. (1996) "An Optimal Price Index for Stock Index Futures Contracts", Journal of Futures Markets, 16, 189-199.

Schwert, G.W. (1989) "Why Does Stock Market Volatility Change Over Time?", Journal of Finance, 44, 1115-1151.

Schwert, G.W. (1990) "Stock Volatility and the Crash of '87", Review of Financial Studies, 3, 77-102.

Siber, W.L. (1984) "Market-Maker Behaviour in an Auction Market: An Analysis of Scalpers in Futures Markets", Journal of Finance, 39, 937-953.

Smirlock, M. and Starks, L. (1985) "A Further Examination of Stock Price Changes and Transaction Volume", Journal of Financial Research, 8, 217-225.

Smith, T. and Whaley, R.E. (1994) "Estimating the Effective Bid/Ask Spread from Time and Sales Data", Journal of Futures Markets, 14, 437-455.

Stewart, J. (1991) Econometrics, Hemel Hempstead: Phillip Allan.

Stoll, H.R. (1978) "The Supply of Dealer Services in Securities Markets", Journal of Finance, 33, 4, 1133-1151.

Stoll, H.R. (1989) "Inferring the Components of the Bid-Ask Spread: Theory and Empirical Tests", Journal of Finance, 44, 115-134.

Subrahmanyam, A. (1991) "Risk Aversion, Market Liquidity, and Price Efficiency", Review of Financial Studies, 4, 417-441.

Sutcliffe, C.M.S. (1993) Stock Index Futures: Theories and International Evidence, London: Chapman and Hall.

Tauchen, G. and Pitts, M. (1983) "The Price Variability-Volume Relationship on Speculative Prices", Econometrica, 51, 485-505.

Taylor, S.J. (1985) "The Behaviour of Futures Prices Over Time", Applied Economics, 17, 713-734.

Thompson, S.R. and Waller, M.L. (1988) "Determinants of Liquidity Costs in Commodity Futures Markets", Review of Futures Markets, 7, 111-126.

Tinic, S.M. and West, R.R. (1972) "Competition and the Pricing of Dealer Services in the Over-the-Counter Stock Market", Journal of Financial and Quantitative Analysis, 7, 1707-1728.

Upton D.E. and Shannon, D.S. (1979) "The Stable Paretian Distribution, Subordinated Stochastic Processes, and Asymptotic Lognormality: An Empirical Investigation", Journal of Finance, 34, 1031-1039.

Wang, G.H.K., Michalski, R.J., Jordan, J.V. and Moriarty, E.J. (1994) "An Intraday Analysis of Bid-Ask Spreads and Price Volatility in the S&P 500 Index Futures Market", Journal of Futures Markets, 14, 837-859.

Wang, G.H.K., Yau, J. and Baptiste, T. (1997) "Trading Volume and Transaction Costs in Futures Markets", Journal of Futures Markets, 17, 757-780.

Wang, J. (1994) "A Model of Competitive Stock Trading Volume", Journal of Political Economy, 102, 127-168.

Westerfield, R. (1977) "The Distribution of Common Stock Price Changes: An Application of Transactions Time and Subordinated Stochastic Models", Journal of Financial and Quantitative Analysis, 12, 743-765.

Wood, R.A., McInish, T.H. and Ord, J.K. (1985) "An Investigation of Transactions Data for NYSE Stocks", Journal of Finance, 40, 723-741.

Zellner, A. (1984) Basic Issues in Econometrics, Chicago: Chicago University Press.