



Durham E-Theses

Price change and trading volume in a speculative market

Rougier, Jonathan

How to cite:

Rougier, Jonathan (1996) *Price change and trading volume in a speculative market*, Durham theses, Durham University. Available at Durham E-Theses Online: <http://etheses.dur.ac.uk/5347/>

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full Durham E-Theses policy](#) for further details.

**Price Change and Trading Volume
in a Speculative Market**

by

Jonathan Rougier

A thesis submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

Economics

The University of Durham

1996

The copyright of this thesis rests
with the author. No quotation
from it should be published without
the written consent of the author
and information derived from it
should be acknowledged.

Copyright © 1996 by Jonathan Rougier.

The copyright of this thesis rests with the author. No quotation from it should be published without the author's prior written consent and information derived from it should be acknowledged.



29 MAY 1997

Abstract

This thesis is concerned with the daily dynamics of price change and trading volume in a speculative market. The first part examines the news-driven model of Tauchen and Pitts (1983), and develops this model to the point where it is directly testable. In order to implement the test a new method for creating a price index from futures contracts is proposed. It is found that news effects can explain some but not all of the structure of the daily price/volume relationship. An alternative explanation is presented, in which the model of Tauchen and Pitts is generalized in a non-linear fashion.

In the second part of the thesis, the presence of a small amount of positive autocorrelation in daily returns is exploited through the development of a timing rule. This timing rule applies to investors who are committed to a purchase but flexible about the precise timing. The computation of the timing rule is discussed in detail. In practice it is found that this timing rule is unlikely to generate sufficiently large returns to be of interest to investors in a typical stock market, supporting the hypothesis of market efficiency. However, the incorporation of extra information regarding price/volume dynamics, as suggested by the analysis of Part I, might lead to a much improved rule.

Acknowledgements

First, I would like to thank my Supervisors, Professor Denis O'Brien of the Department of Economics and Dr. Allan Seheult of the Department of Mathematical Sciences, not just for their many insightful comments concerning the direction and content of this thesis, but also for their unfailing support over the time I have been at Durham.

Second, I would like to acknowledge gratefully the help and support of my colleagues in the Department of Economics, particularly Dr. Phil Holmes, Dr. Simon Parker, John Ashworth, Brian Hillier and Cheryl Conway.

Third, I would like to thank the following for their assistance in various parts of the thesis: Professor Michael Goldstein of the Department of Mathematical Sciences, for discussions regarding the nature of belief and for suggestions regarding the optimal timing problem of Chapter 8; Dr. Matthew Penrose and Dr. Peter Craig, also of the Department of Mathematical Sciences, for their help with various stochastic processes; Professor Mark Powers and two anonymous referees of the Journal of Futures Markets; and Professor Hal Varian, now at the University of California, Berkeley, for discussions regarding General Equilibrium securities prices.

Finally, I would like to absolve all of the above from any errors that remain, which are entirely my own responsibility.

Contents

Contents	ii
List of Tables	vi
List of Figures	vii
1 Precursors	1
1.1 Introduction	1
1.2 Outline of Chapters	2
1.3 Some Clarifications	5
1.4 An Overview of the Literature	10
1.4.1 Descriptive Models	10
1.4.2 Homogeneous Investors	13
1.4.3 Pseudo-Homogeneous Investors	15
1.4.4 Heterogeneous Investor Models	17
1.5 General Equilibrium	18
1.5.1 The Arrow-Debreu Model	19
1.5.2 Varian's Analysis	20
1.5.3 An Asymptotic Generalization	22
I Return Leptokursis and News Arrival	27
2 The Role of News	28

2.1	Leptokurtosis and Clark's Model	29
2.2	The Epps and Epps Model	32
2.3	The Tauchen and Pitts Model	36
2.3.1	Intra-Day Model	36
2.3.2	Inter-Day Model	40
2.4	The Implied Price-Volume Relationship	41
2.5	A Critical Evaluation of the TP Model	46
2.6	The News Arrival Process	50
2.7	Summary and Conclusion	55
3	What Is News?	56
4	An Optimal Price Index	66
4.1	Introduction	66
4.2	'Optimality' in Futures Price Indices	68
4.3	Clark's Price Index	74
4.4	Practical differences	76
4.5	Conclusion	80
5	A Short Empirical Study	82
5.1	Introduction	82
5.2	Prices	84
5.3	Trading Volume	91
5.4	Symmetry	96
5.5	Testing the Tauchen and Pitts Model	100
5.6	Summary and Conclusion	106
6	A Market-Clearing Model	108
6.1	Introduction	108

6.2	Market Microstructure	110
6.3	Tauchen and Pitts Again	112
6.4	The Vector of Reservation Prices	116
6.5	Monte Carlo Simulations	122
6.5.1	The Cross-Sectional Relationship	123
6.5.2	The Time-Series Relationship	127
6.6	Autocorrelation in Price Changes	130
6.7	Summary and Conclusion	136
II	Return Autocorrelation	139
7	The Role of Market Microstructure	140
7.1	The Martingale Theorem	141
7.2	Risk Aversion	143
7.3	Heterogeneous Investors	145
7.4	Trading Rules	149
7.5	Summary	152
8	The Theory of Optimal Timing	154
8.1	Introduction	154
8.2	The Timing Problem	156
8.3	Optimal Timing When Returns are AR(1)	161
8.4	Sufficient Conditions	166
8.5	The Solution	169
8.6	Summary	173
9	Computing the Reward Function	174
9.1	One Simple Approach	175

CONTENTS

9.2	Computation of the Reward Function	176
9.3	Computation Results	182
9.4	Discussion	188
10	Summary and Conclusion	190
	Bibliography	201

List of Tables

4.1	Regression results, 1985–1994	79
5.1	Distribution of Daily Returns (r_t , %)	87
5.2	Autocorrelation and ARCH in Returns and Volume	89
5.3	Distribution of Trading Volume (v_t^*)	96
5.4	WLS Regression for Symmetry	99
5.5	χ^2 Test on Volume-adjusted Squared Returns	103
6.1	Standard Correlations for Price Change and Volume	125
6.2	Partial Correlations for Price Change and Volume	126
6.3	VAR Results for Price Change and Volume	129
6.4	χ^2 Tests of Run Length	135
9.1	Values of r^* at Different Parameter Settings	183
9.2	Values of $\mathcal{E}[\tau]$ at Different Parameter Settings	184

List of Figures

1.1	A Taxonomy of Models of Speculative Markets	10
4.1	The Distribution of FTSE Contracts by Time to Expiry	77
4.2	The Notional Time to Expiry During a Quarter	78
5.1	Sample of Original Data from LIFFE	83
5.2	Optimal Price Index (p_t)	85
5.3	Daily Returns (r_t)	86
5.4	Trading Volume (v_t)	88
5.5	Detrended Trading Volume (\hat{v}_t)	90
5.6	Contract Rollover Proxy (l_t)	93
5.7	Detrended Trading Volume adjusted for Rollover, (v_t^*)	94
5.8	Squared Returns and Trading Volume (r_t^2, v_t^*)	97
5.9	Volume-adjusted Returns (r_t^2/\hat{v}_t^* , rescaled)	101
5.10	Observed and Actual Distributions of r_t^2/v_t^* by Quartile	104
6.1	The Function $f(x; \alpha)$, for $x \geq 0$	119
6.2	Price Changes from the First 250 Periods, $\alpha = 1.0$	131
6.3	XY Plot of Price Changes from the First 250 Periods, $\alpha = 1.0$	131
6.4	The Distribution of Run Lengths	134
9.1	Marginal Effects on the Reward Function $f^*(r)$	186

Chapter 1

Precursors

1.1 Introduction

This thesis is concerned with the day to day behaviour of a speculative asset trading in a secondary market. The primary concern is to develop our understanding of the price/volume relationship, and its implications. Traditionally, our understanding of the daily price/volume relationship has been summarized by two stylized facts: (i) the tendency of large price changes to follow each other, and to be accompanied by a large amount of trading volume; (ii) the tendency for daily price changes to display a small amount of positive autocorrelation. In the same way, this thesis is divided into two parts. Part I considers 'news' explanations of price/volume dynamics without explicit consideration of the endogeneity of the mean daily return. Part II considers the endogeneity of the mean daily return, but over a short time-period in which other factors may be taken to be roughly constant. Linking the two parts is a model in which both stylized facts can occur simultaneously within a market-clearing framework; again, this reflects more recent work in which both stylized facts emerge from considerations of market structure.

The benefit of increased understanding in this area is the more efficient pricing



of speculative assets. This arises through a better understanding of the process underlying prices, giving rise to better (i.e. less biased, more efficient) estimators of the parameters governing the process. However, this thesis also contains a more immediate pay-off. The second part of the thesis develops a new type of timing rule which will enable certain investors to exploit the positive autocorrelation in daily price changes, by making the timing of their purchases depend upon the price history.

1.2 Outline of Chapters

Following this outline, Section 1.3 clarifies the notion of a speculative asset in general and futures contracts in particular, the operation of the market in futures contracts, and investor rationality. Section 1.4 provides an overview of the price/volume literature. Finally, Section 1.5 examines the General Equilibrium picture for complete markets using the Arrow-Debreu model proposed by Varian (1985).

Part I: Return Leptokursis and News Arrival

The analysis of markets in Part I is from the bottom up. This requires both the static analysis of market-clearing and the dynamic analysis of the way in which investors' beliefs about the future are revised through time. Fundamental to a bottom-up approach is the notion that investors differ, and hence can be satisfied with different positions in a speculative asset at the same price. Following a revision to beliefs an optimizing investor will often want to adjust his¹ portfolio, and this leads to trading volume and a possible change in the market-clearing price.

On this basis the daily price/volume relationship can be disaggregated into the

¹The phrase 'he or she' will be shortened to 'he' in this thesis, likewise 'his or her' to 'his'; this reflects the fact that, at the time of writing, investors in speculative markets are predominantly male.

price/volume relationship per item of news, and the distribution of the number of items of news per day. This disaggregation is found in the seminal price/volume model of Tauchen and Pitts (1983), which plays an important part in the thesis and serves as a starting point for the analysis of Part I.

Chapter 2 describes the background to the Tauchen and Pitts (1983) model, before going on to describe the model itself and its shortcomings, particularly in the light of subsequent time-series work on prices and trading volumes. This work has suggested that the amount of news per day is not independently and identically distributed, as was originally assumed, but appears to be positively autocorrelated from day to day. This chapter also develops the theoretical properties of the model to derive a new statistical test of the model which is unaffected by the properties of the news arrival process.

Chapter 3 is a short chapter to clarify what we mean by news and related issues such as the quantity of news per period. This chapter suggests that it is important to consider the costs associated with the process of updating beliefs following the arrival of new information, and that the number of transactions may be a better indicator of the quantity of news than the number of news stories, as is typically used.

Chapters 4 and 5 test the Tauchen and Pitts model using the results developed in Chapter 2 and data from the London International Financial Futures Exchange (LIFFE), 1985–94. In order to use futures prices, which have several advantages in this context over spot prices, it is necessary to combine data from individual futures contracts into a single price index. Chapter 4 defines the notion of ‘optimality’ in price indices for futures contracts; and shows how optimal price indices can be created. Using the data from LIFFE and from the London Stock Exchange, the optimal price index is shown to outperform its alternatives, including the index used by Tauchen and Pitts in their estimation. In Chapter 5 it is found necessary

to transform the trading volume data to allow for the growth in the market over the period and also to remove the effects of contract rollovers, which introduce a complicated pattern of quarterly seasonality into the volume data. A method is proposed in which contract rollovers are proxied using the data on daily open interest.

Finally in Part I, Chapter 6 is an analysis of market-clearing which retains the linear demand functions of Tauchen and Pitts, but which generalizes their model for updating investors' forecasts of future prices. This leads to a relationship between the cross-sectional distribution of forecasts at the end of each day and the joint distribution of price change and trading volume over the following days. Since the cross-sectional distribution of forecasts is likely to change only slowly through time, this introduces a time-dependent element into price changes and trading volume even in the absence of time-dependency in the news-arrival process. This model is significant in the context of the previous chapters, where the main conclusion is that news effects alone cannot completely explain price/volume dynamics.

Part II: Return Autocorrelation

The main purpose of the second part of the thesis is to develop and analyse a timing rule for investors which exploits the small amount of predictability inherent in the positive autocorrelation of daily returns. Chapter 7 explains why it is that returns have this autocorrelation, with particular reference to the Martingale theorems of Samuelson (1965, 1973) which suggest that there should be no relationship between future and past returns. The rewards from trading strategies based upon return autocorrelation are also discussed, along with the transactions costs which make them unprofitable in operation.

Chapters 8 and 9 develop and analyse a timing rule based around the return autocorrelation. With a timing rule, the decision to purchase has already been taken,

and the only unresolved issue is when exactly to make the purchase. Consequently the transactions costs which ruin trading rules are effectively forgone. Chapter 8 solves the problem of optimal timing to give the 'optimal timing rule', which bases the decision to purchase (as opposed to delay) upon the return of the previous day. The function describing the performance of this rule is solved as the fixed point of a non-linear operator in a Hilbert space, expressed as an infinite power series in an integral operator. Chapter 9 tackles the computation of this solution, and analyses the returns that might be generated by following the optimal timing rule in practice in typical stock markets.

Finally, Chapter 10 concludes the thesis with a summary of the main results and some conclusions pointing towards areas for future research.

1.3 Some Clarifications

Speculative Assets

This thesis is about the demand for speculative assets. The primary common feature of all speculative assets is uncertainty about the outcome for the holder, which is contingent upon the state of nature. Perhaps the simplest speculative asset of all would be an agreement to pay or receive a certain amount of money contingent on the result of a single toss of a fair coin to be made in the near future. This agreement could be bought or sold at a given price, depending upon the pay-off and the participants' attitudes to risk.

The simplicity in this example arises from the head or tail dichotomy of the future, the known fixed probability of 0.5 for each of the two outcomes, and the absence of a time dimension in which the needs of the participants might change. In contrast, the speculative assets in financial markets tend to be categorized by complex pay-offs spanning a large number of states of nature, subjective and poorly

understood probabilities, and a significant time dimension in which the assessments and needs of the participants might change significantly. Of course, this is what makes these assets so interesting.

This thesis will concentrate on stock index futures contracts. This is a contract to pay or be paid the money-equivalent of the difference in price between the contract when it is taken out and the underlying basket of stocks at a fixed time in the future. An investor at time t taking a position of q contracts for expiry at time k ($k > t$), receives a pay-off at k of

$$f(t, k) = qc(p_k - p_{tk}) \quad (1.1)$$

where p_{tk} is the contract price at time t , p_k the stock index price at expiry at time k , and c the conversion factor from index points to currency.² At any point in time prior to the expiry of the contract, the holder has the option of closing his position by an offsetting transaction (e.g. buying one contract to close out a short position of one contract). In this case his pay-off is simply the money-equivalent difference in the contract price over the period:

$$f(t, s) = qc(p_{sk} - p_{tk}), \quad t < s < k. \quad (1.2)$$

These two formulas are equivalent at the time $s = k$ since institutional arrangements ensure that a contract held to expiry closes at the value of the underlying index, i.e. $p_{kk} = p_k$.

The payout of a futures contract is uncertain at the point at which it is bought. The only way in which the payout can be made certain ahead of k is if the holder has an exactly offsetting position in the stock index. The application of the standard arbitrage framework then gives rise to the *fair value* relationship, in which the spot

²For the FTSE-100 Futures contract, for example, the price c is £25 per index point.

and futures prices are linked via the risk-free interest rate, the dividend yield and the time to expiry. Imperfections in the two markets, particularly sizable round-trip transactions costs in the spot market, mean that in practice the difference between the spot price and the futures price varies around its fair value.³

In relation to the discussion above, futures contracts have a simple pay-off in relation to the underlying stock price index, but this underlying index has a very complicated pay-off in relation to the various possible states of nature at time k . Therefore futures have a complicated pay-off over the states of nature. In the interval (t, k) the arrival of new information might cause investors to reassess their need for futures contracts (possibly by changing their attitude towards risk) and/or their subjective assessment of the probability of various states of nature. Even an investor who is unaffected in these ways may still be affected by the resulting change in the market-clearing price, and want to change his holding accordingly. The relation between the desired position size and the market-clearing price is completely described by an investor's *demand function*.

Demand Functions

Each investor is presumed to have a demand function which relates his demand for stock index futures contracts to the contract price. This demand function arises as the result of some explicit or implicit optimization problem: in this thesis I will be concerned with explicit problems, and the demand functions will arise as a consequence of the problem. The presence of variables other than the price in the demand function depends on the type of investor. For example, an arbitrageur might have the current spot price, interest rate and dividend yield, and various transactions costs; a private investor using futures as a way of getting quick short-

³For more details of the theoretical properties of a futures contract's fair value, see Cox *et al.* (1981); see Yadav and Pope (1990, 1992, 1994) for empirical studies regarding the size and variability of the premium that futures prices tend to command over spot prices and the relationship of this premium to fair value.

term exposure to the stock market might have wealth and parameters describing his belief about the likely course of prices.

For any differentiable demand function there will be some prices at which the investor wants to be neither short nor long of the market. If the investor's utility function is strictly concave, there will be exactly one price at which this is true (Samuelson, 1983, Mathematical Appendix C), and this is termed the *reservation price*.

Definition 1.1 (Reservation Price) *The reservation price of an investor for a speculative asset is that unique market price at which the investor desires to be neither short nor long of the asset.*

Samuelson showed that without any restrictions on the utility function beyond concavity the demand function may change direction repeatedly (e.g., as determined by the investor's beliefs about the future) but it can cross the x-axis only once, and with negative gradient. This does not rule out the demand functions being sometimes increasing in market prices, as was suggested by Tobin (1958, 1969).

Reservation prices play a crucial role in models of market-clearing. It is possible to recast an investor's demand function in terms of the deviation between the market-clearing price and the reservation price. When these demand functions are combined into a market-clearing condition, which in the case of futures would say that the sum of all positions must be zero (short positions are negative), the result is that the market-clearing price is seen to be some function of the set of reservation prices. Consequently changes in the market-clearing price are in general a function of the two sets of reservation prices: those from before the change, and those from after. In exceptional circumstances it is possible to examine changes in the market-clearing price entirely in terms of a symmetric function of the set of changes in the reservation prices. This thesis considers such a model in detail, that of Tauchen and

Pitts (1983). But it also considers the general case.⁴

Rational Expectations

Tied up with the form of the demand functions and market-clearing is the issue of rational expectations (RE). Only certain types of demand function will satisfy the RE criterion: "... a model that would generate price from the [distribution of reservation prices] and that would simultaneously characterize agents' [reservation prices] in a manner that is consistent with agents' expectations that price will be determined in a similar fashion when the next period arrives" (adapted from LeRoy, 1989, p. 1604). A combination of demand functions and a model for updating reservation prices in the light of news gives rise to a stochastic process for prices. For RE to hold, this stochastic process should be part of the problem which gives rise to the demand functions.

To take a simple example, suppose that the demand functions are linear and that reservation prices follow a random walk with normally-distributed increments (this is in fact the model of Tauchen and Pitts, 1983). In this case the market-clearing price is also a random walk with normally-distributed increments. Then the critical question for RE is "Can linear demand functions arise when investors expect prices to be a random walk with normally-distributed increments?" Interestingly, the answer is "Yes", as will be shown in Chapter 2. Unfortunately, this particular model also has some substantial weaknesses, and in eliminating these weaknesses the RE consistency is lost, as will be discussed in Chapter 6.

⁴This analysis forms the basis of Chapter 6.

Figure 1.1: A Taxonomy of Models of Speculative Markets

Descriptive Models No explicit optimizing behaviour by investors.

Homogeneous Investors All investors solve an identical problem, and are therefore indistinguishable.

Pseudo-Homogeneous Investors All investors solve the same generic problem, although this problem may not be parametrized in the same way for all investors.

Heterogeneous Investors Investors solve different problems.

1.4 An Overview of the Literature

Before commencing this overview of the price/volume literature it is helpful to introduce a taxonomy of models of speculative asset price determination. The models considered in this thesis may be distinguished by the assumptions they make about investor heterogeneity. The taxonomy is given in Figure 1.1.

This taxonomy may be applied to the development of the theory of the distribution of daily prices, and, more recently, the joint distribution of daily price changes and trading volume. Parts of the following overview will be covered in much more detail in later chapters. More general surveys of the theoretical and empirical literature on speculative prices and the price/volume relationships can be found in Fama (1970, 1991), Karpoff (1987), West (1988) and LeRoy (1989).

1.4.1 Descriptive Models

Initially the emphasis in stock market research was on modeling prices directly, rather than deriving from theory implications about prices that are testable. The earliest extant empirical suggestion for speculative price changes is that of Bachelier (1900). As Samuelson (1972) noted, Bachelier suggested that speculative prices

p_t, p_{t+1}, \dots followed a stochastic process satisfying

$$\Pr\{p_{t+s} \leq p \mid p_t, p_{t-1}, \dots\} = F(p - p_t, s) \quad s, t = 1, 2, \dots \quad (1.3)$$

In words, the distribution of the price change over some period is independent of the price history preceding the period, and independent of the length of the period excepting changes of scale. Bachelier asserted that eq. (1.3) implied that $F(\cdot)$ was the normal cumulative distribution function, in which case prices would conform to a random walk with normally distributed increments.

It is not clear on what grounds Bachelier derived the relation eq. (1.3). His motivation was to state formally that investors should not be able to profit from the study of past prices, but eq. (1.3) goes much further by specifying the invariance of the distribution of increments, which makes prices a random walk, and that these increments have a certain scaling property (which will be discussed further below). However, the notion that prices followed a random walk subsequently received strong empirical support. Cowles (1933) showed that professional stockmarket analysts were unable consistently to outperform the market, and this led to studies confirming that price changes appeared to be similar to random walks (e.g., Working, 1934; Kendall, 1953; Granger and Morgenstern, 1963; Godfrey *et al.*, 1964; Fama, 1965). The consensus gradually shifted from a random walk in levels to a random walk in the logarithm of levels in order to respect the property of limited liability which prevents stock prices from going negative (Osborne, 1959). This idea was formalised by Samuelson (1965) who proposed the alternative stochastic process

$$\Pr\{p_{t+s} < p \mid p_t, p_{t-1}, \dots\} = F\left(\frac{p}{p_t}, s\right), \quad s, t = 1, 2, \dots \quad (1.4)$$

which has as a solution the lognormal distribution for prices, or equivalently the normal distribution for log-returns. Samuelson termed this process *economic brownian*

motion.

Return Autocorrelation

With the exception of Working (1934),⁵ the above studies of random walks focused on the serial correlation in the returns r_t, r_{t+1}, \dots , i.e. the significance of ρ in the regression

$$r_t = \mu + \rho r_{t-\tau} + u_t, \quad (1.5)$$

where τ is the differencing period and u_t is a white-noise disturbance term. In general it was found for one period lags that there were fewer stocks with negative autocorrelations than with positive ones. For example, Fama (1965) reports 8 negative and 22 positive for the Dow Jones Industrial average portfolio over 5 years to 1962. Nine of the 22 positive autocorrelations were significant, compared to only two of the negative ones.⁶ Assessing this and other evidence for autocorrelation, Fama declared that it was "... probably insignificant from an economic point of view" (Fama, 1970, p. 394). The small positive autocorrelation is now an accepted fact in daily returns, with studies tending to confirm that, after allowing for trading costs, there is no risk-adjusted profit to be made from trading on its basis, above that of a buy-and-hold strategy (see, e.g., Conrad and Kaul, 1988; Lo and MacKinlay, 1988; Brock *et al.*, 1992; Corrado and Lee, 1992).⁷ Since the autocorrelation is so small, much of the theory of speculative prices proceeds on the basis that daily returns are effectively independent of the price history.

⁵This prescient study showed that a random walk looked very much like the evolution of a speculative price.

⁶As Fama (1970) notes, however, the precise significance is hard to assess given the apparent leptokurtosis of the disturbance term, u_t .

⁷A more detailed discussion of trading rules can be found in Section 7.4.

1.4.2 Homogeneous Investors

The first true homogeneous investor model was that of Samuelson (1965, 1973). This was not just an empirical specification, but a direct implication of a model of optimizing behaviour by investors. In Samuelson's model all investors have identical opinions about the future distribution of speculative prices and are risk-neutral. In this case the expected return on every asset will be bid down to the return on the risk-free asset, and then the expected change in the price of any speculative asset will be zero.⁸ Thus the model implies that prices have the property

$$\mathcal{E} [p_{t+s} \mid p_t, p_{t-1}, \dots] = p_t, \quad s, t = 1, 2, \dots \quad (1.6)$$

known as a *martingale process*. This implies that returns will have zero expectation, and so be a *fair game* (see, e.g., LeRoy, 1989). Samuelson's martingale model had an attractive basis in investor behaviour, but it also offered a palliative to the random walk model. If daily prices were a random walk then it was generally presumed that the increments would have a normal distribution, by the Central Limit Theorem. However, evidence of leptokurtosis in the return distribution (Osborne, 1959; Alexander, 1961) suggested otherwise.

One response was to recast the random walk in the more general form of a *stable non-normal distribution*. Mandelbrot (1963a,b) noted that eq. (1.3) was necessary but not sufficient for $F(\cdot)$ to be the normal distribution. It actually defines a class of distributions of which the normal is a special, limiting, case. In general members of this class are leptokurtic. However, the evidence of Fama (1965), which showed that absolute returns seemed to clump into periods of large and small, supported an alternative explanation, that of a mixture of distributions. It had been noted by

⁸Technically, the expected *excess* return should be zero, to allow for the risk-free rate. This makes prices a *sub-martingale*. The risk-free rate is therefore presumed to be zero for simplicity. Samuelson's model is examined in more detail in Section 7.1.

Osborne (1959) that a mixture of normal return distributions with similar means but dissimilar variances would also be leptokurtic. Mandelbrot's explanation lacked the time-dimension implicit in the clumping which could be attributed to time-series properties in the mixing process.

The notion of a mixture of distributions was formalized by Clark (1973) as a *subordinated stochastic process*. The price change on day t is determined by the amount of news arrival during the period $(t - 1, t]$, denoted n_t :

$$p_t = p_{t-1} + \sum_{i=1}^{n_t} \Delta p_i, \quad (1.7)$$

where Δp_i is independently and identically distributed, with zero mean to preserve the martingale property. Clark suggested that each increment be a normal random variable, and the amount of news on day t be an independent lognormal random variable. This would give rise to a daily return distribution with zero mean, finite variance and leptokurtosis.⁹ There have also been empirical models in which the distribution is a mixture of normals (Kon, 1984), or a more complicated Poisson jump-diffusion process (Akgiray and Booth, 1986, 1987). These models are not random walks, but *are* martingales on the imposition of zero expected return.

Up to this point the role of trading volume in the distribution of daily price changes was poorly defined. This is not surprising, since a voluntary transaction must indicate a difference of opinion or a difference in circumstances between the buyer and the seller. There can be no such differences when investors are homogenous. The empirical analysis of Granger and Morgenstern (1963) found no relationship between price changes and trading volume, but they later found that there

⁹Clark's model of is examined in detail in Section 2.1. An alternative suggestion to Clark's was that the amount of news had a stable non-normal distribution with infinite mean (Mandelbrot and Taylor, 1967; Blattberg and Gonedes, 1974), which would give rise to a return distribution with infinite variance. In practice it is still hard to tell these two models apart (see, e.g., Hall *et al.*, 1989) although the consensus, spurred by Occam's razor, favours finite variance—see also footnote 2 on page 30.

was a correlation between squared price change and trading volume (Granger and Morgenstern, 1970). This suggests that the relationship between price change and trading volume is symmetric with respect to price falls and rises.¹⁰ Clark (1973) had only moderate success in using volume as a proxy for news arrival; this rather ad hoc inclusion of trading volume prompted the sardonic rejoinder “[Clark’s experiments] have shown that daily increments of local time are like daily volume to the power 2.13. This *empirical* discovery seems very interesting and deserves careful thought” (Mandelbrot, 1973, p. 159, my emphasis).

1.4.3 Pseudo-Homogeneous Investors

In a pseudo-homogeneous investor model all agents solve the same problem, although certain of the parameters in the problem might vary across investors. Consider, for example, the simple case of deciding how many futures contracts to hold to expiry. An investor determines the optimal number of contracts by maximizing the expected utility of his pay-off over the various states of nature. His pay-off includes the pay-off from holding the contracts, as in eq. (1.1), but also the interest from the wealth he must make available to the exchange in order to trade.¹¹ Denoting the investor’s wealth as w and the risk-free interest rate as i , the optimization problem is¹²

$$\max_q \int_{-\infty}^{\infty} u(q(p_k - p_{tk}) + w i; \eta) f(p_k; p_{tk}, \theta) dp_k, \quad (1.8)$$

where $u(\cdot; \eta)$ is a strictly concave utility function parametrized by η (which might represent the investor’s degree of risk-aversion), and $f(\cdot; p_{t,k}, \theta)$ is a probability

¹⁰Karpoff (1987) surveys the empirical evidence on the price/volume relationship. He makes an interesting distinction between spot and futures markets regarding asymmetry of the costs of short and long positions, and suggests that only in futures markets, where the costs are symmetric, will there appear no relationship between price change and trading volume. This symmetry is discussed in more detail in Section 5.4.

¹¹In practice the maximum size of the position q is constrained by the amount of wealth and the exchange’s margin requirement.

¹²For simplicity, the price of the index, c from eq. (1.1), will be set to 1 in the rest of this thesis.

density function representing future states of nature, and parametrized by the current futures price and θ (which might reflect the investor's expectation of p_k at time t). In a pseudo-homogeneous investor model the investors are distinguished from each other by the value of the parameters η , θ and w ; consequently they are likely to have different reservation prices. But all these investors are buying or selling futures contracts as their only speculative asset, and operating regardless of the behaviour of other investors, as described in eq. (1.8).

One of the first pseudo-homogeneous investor models was that of Epps and Epps (1976).¹³ In this model investors' demands are derived from portfolio theory in such a way that the resulting demand functions are linear in price. Investors differ only in their expectation of future speculative asset prices, so that they are all identically risk-averse and share a common belief about the covariance matrix of future asset prices. New information causes investors to modify the expectations they attach to different assets, which causes them to adjust their portfolios. This adjustment generates trading volume, as well as the potential for a change in the market-clearing price.¹⁴ Tauchen and Pitts (1983) introduce a stochastic process for reservation prices into a framework similar to that of Epps and Epps, but which avoided some of the latter's rather arbitrary assumptions.¹⁵

The attraction of these two models is that they unify price change and trading volume within a model of rational investor behaviour. Therefore they should in principle provide testable hypotheses for the bivariate distribution of price changes and trading volume. Unfortunately, an incomplete understanding of the daily distribution of news prevented these models from being fully exploited. At the same

¹³Another possible contender is the sequential information arrival model of Copeland (1976). In this model effectively identical investors differ in the order in which they receive public information, and price changes and trading volume are generated as this information disperses across the market. In this thesis it is assumed that public information is revealed to all investors at the same time, as tends to be the case in financial markets.

¹⁴The Epps and Epps (1976) model is discussed in more detail in Section 2.2.

¹⁵The Tauchen and Pitts (1983) model is discussed in more detail in Sections 2.3, 2.4 and 2.5.

time evidence started to accumulate which suggested that the amount of news per day might be autocorrelated.¹⁶ This possibility was in fact foreseen by Tauchen and Pitts and suggested by them as a possible reason for the poor fit of their model. It is important to note that the Tauchen and Pitts model takes the news-arrival process as given. However, if the amount of news per day is not independently and identically distributed, the fitting of their model by maximum likelihood would be extremely hard given the high dimensionality of the resulting integral expression.

1.4.4 Heterogeneous Investor Models

In pseudo-homogeneous models trading volume is generated by news arrival. Investors are essentially competitive, pitting their interpretation of commonly available information against each other. Economists questioned whether this could be entirely rational (Milgrom and Stokey, 1982; Tirole, 1982). Why should investors on both sides of a transaction believe that they have the advantage? Risk-averse investors avoid zero sum games, so the existence of a large amount of trading volume implies either that many investors are not risk-averse, or, even worse, that some investors are simply irrational. Either way, investors can be distinguished by more than just the parameters of their utility functions and their subjective probability assessments.

This heterogeneity now goes by the name of 'noise' (Black, 1986). In the simplest noise model, uninformed or 'noise' traders tend to chase price trends while informed 'fundamental' traders tend to trade only when price has been pushed a long way from value.¹⁷ The consequence is that while much trading volume will be related

¹⁶This evidence is discussed in Section 2.6.

¹⁷More generally, the designation 'noise trader' might be replaced by 'noninformational trader'—any investor buying or selling for reasons other than superior information. For example, Admati and Pfleiderer (1988, 1989) discuss the possibility of 'liquidity traders' during the course of the day and around the weekend, and Rougier (1993) examines the impact of margin traders under the Account settlement system in the London Stock Exchange, which injects liquidity into the market on a fortnightly basis.

to news, some of it will not, being related instead to the strategic exploitation of noise traders by fundamental traders. Even if the news arrival process was independently and identically distributed from day to day, the time series of trading volume might display bursts of high volume which may or may not be associated with large movements in price.¹⁸

The problem with heterogeneous investor models is that they are in some sense too good. It is almost as though any facet of the price/volume relationship can be 'explained' by an appropriately chosen noise trader; this is not surprising given that a feature of noise models is the presence of investors trading irrationally. For this reason, recent empirical studies of the price/volume relationship have tended to be data-based. In their empirical analysis of the daily price/volume relationship over 60 years of New York Stock Exchange data, Gallant *et al.* observe: "Existing models, however, do not confront the data in its full complexity and have not evolved sufficiently to guide the specification of an empirical model of daily stock market data" (Gallant *et al.*, 1992, p. 202). Ironically, then, we appear to have come full circle back to empirical models, but supported by recent theoretical developments.

1.5 General Equilibrium

The models of the following chapters are partial equilibrium models in which financial markets are taken in isolation. This ignores the broader picture of rational agents allocating resources between current and future consumption by the use of financial markets. It also ignores the simultaneous determination of prices across financial markets. These might be termed the dynamic and the static general equilibrium (GE) problems, respectively. The scale of these problems makes it very hard to incorporate any degree of investor heterogeneity. Thus dynamic GE with ratio-

¹⁸Noise trading models are discussed in more detail in Section 7.3.

nal expectations operates at the level of homogeneous investors (see, e.g., Lucas, 1978; Harrison and Kreps, 1979; Cox *et al.*, 1985).¹⁹ However, there is an analysis of static GE which incorporates pseudo-homogeneous investors due to Varian (1985, 1992), based on the Arrow-Debreu model of complete markets (Arrow, 1964; Debreu, 1959). Varian considers how differences in opinion regarding probabilities might affect prices in GE.

1.5.1 The Arrow-Debreu Model

In the Arrow-Debreu model there are n investors and S states of nature. Each investor assigns a probability to each state of nature, and these probabilities are collected in the $n \times S$ matrix Π , such that π_{is} is investor i 's subjective probability of state s . In what follows it will also be useful to use Π_i to denote the i^{th} row of Π , and Π_s the s^{th} column. The market is complete in that there is a security that pays off for every possible state of nature; each pay-off is one consumption unit if that state occurs, and nothing otherwise.

Investors each start with some endowment of securities, the $n \times S$ matrix \widehat{C} where \widehat{c}_{is} is investor i 's initial endowment of the security which pays off in state s ; as with the probabilities, \widehat{C}_i denotes the i^{th} row of \widehat{C} , and in addition the total number of each security will be denoted \widehat{c}_s . Each investor acts as a price-taker in maximizing expected utility over the security bundle $C_i = (c_1, \dots, c_S)$, subject to the budget constraint imposed by his initial endowment \widehat{C}_i and given prices $P = (p_1, \dots, p_S)$.

The Lagrangian function of investor i is

$$\mathcal{L}(C_i, \lambda_i; P, \Pi_i, \widehat{C}_i) = \max_{C_i} \sum_{s=1}^S \pi_{is} u_i(c_{is}) - \lambda_i \sum_{i=1}^S p_s (c_{is} - \widehat{c}_{is}), \quad (1.9)$$

where $u_i(\cdot)$ is investor i 's utility function, assumed to be strictly concave: $u'_i > 0$,

¹⁹An excellent and rigorous exposition of these models and many related issues can be found in Duffie (1996).

$u_i'' < 0$. The first order conditions may be written

$$G_i(C_i, \lambda_i; P, \Pi_i, \widehat{C}_i) = [0] \quad (1.10)$$

where

$$G_i^s(C_i, \lambda_i; P, \Pi_i, \widehat{C}_i) \stackrel{\text{def}}{=} \pi_{is} u_i'(c_{is}) - \lambda_i p_s \quad s = 1, \dots, S \quad (1.11)$$

$$G_i^\lambda(C_i, \lambda_i; P, \Pi_i, \widehat{C}_i) \stackrel{\text{def}}{=} - \sum_{s=1}^S p_s (c_{is} - \widehat{c}_{is}). \quad (1.12)$$

By the strict concavity of the utility function the Jacobian of G is non-singular, and eq. (1.10) can be solved for the demand functions and the Lagrangian function:

$$c_{is}^* = c_{is}(P, \Pi_i, \widehat{C}_i) \quad s = 1, \dots, S \quad (1.13)$$

$$\lambda_i^* = \lambda_i(P, \Pi_i, \widehat{C}_i). \quad (1.14)$$

1.5.2 Varian's Analysis²⁰

Varian considers the identical satisfaction of the first-order condition for investor i in security s :

$$\pi_{is} u_i'(c_{is}^*) - \lambda_i^* p_s \equiv 0. \quad (1.15)$$

Since the utility function is strictly concave, this can be inverted to give

$$c_{is}^* = f_i \left(\frac{\lambda_i^* p_s}{\pi_{is}} \right). \quad (1.16)$$

The function $f_i(\cdot)$ must be strictly decreasing. Varian then applies the market-

²⁰I would like to thank Professor Hal Varian for several very helpful discussions regarding the nature and implications of his analysis. The presentation here is my own, to avoid some confusion which arises regarding his notation and exposition.

clearing condition for security s to endogenize the price, p_s :

$$\hat{c}_s \equiv \sum_{i=1}^n c_{is}^* = \sum_{i=1}^n f_i \left(\frac{\lambda_i^* p_s}{\pi_{is}} \right). \quad (1.17)$$

He considers two securities within a given equilibrium, s and t say. In a given equilibrium we may take λ_i as fixed, in which case the righthand side of eq. (1.17) is decreasing in p_s (since each f_i is decreasing in p_s), and the market-clearing condition may be expressed as, in general,

$$p_s = F(\Pi_s, \hat{c}_s \mid \Pi, \hat{C}) \quad s = 1, \dots, S. \quad (1.18)$$

This says that a necessary condition for $p_s \neq p_t$ is $(\Pi_s, \hat{c}_s) \neq (\Pi_t, \hat{c}_t)$. In other words, if every investor agrees that two states are equally likely (although they may disagree about how likely) and there is the same initial endowment of security in each state, then the price of the securities paying off in those two states must be the same. It is also clear from eq. (1.17) that

$$p_s <> p_t \quad \text{accordingly as} \quad \begin{cases} \hat{c}_s >> \hat{c}_t & \Pi_s = \Pi_t \\ \Pi_s <> \Pi_t & \hat{c}_s = \hat{c}_t \end{cases} \quad (1.19)$$

Thus, all other things being equal, greater supply decreases the equilibrium price, as does lower probability: the securities with the highest price are those where the aggregate initial endowment is small and the probability generally agreed to be large.

On its own, this result might be considered a classic example of the use of complex abstract reasoning to reach a completely obvious conclusion. However, Varian goes on to relate the dispersion of probabilities to the equilibrium price, under the condition that all investors have identical utility functions. Varian shows that if investors are sufficiently risk-averse then $f_i(\cdot) = f(\cdot)$ ($i = 1, \dots, n$) is a concave

function of π_{is} . In this case the sum of the n terms in eq. (1.17) will be lower for security s than for security t if Π_s has the same mean as Π_t , but a larger dispersion (this is, effectively, Jensen's inequality). Thus if the endowment of s and t is the same, then $p_s < p_t$. In this way Varian establishes that diversity of opinion tends to depress a security's price within a GE framework. As Varian notes, this result has empirical support from Cragg and Malkiel (1982), who show that the relationship between ex post return and risk is clearest when risk is proxied by a measure of diversity of opinion among investors about future prices.

1.5.3 An Asymptotic Generalization

It is possible that Varian's results may be extended by considering the market-clearing condition eq. (1.17) when the number of investors (n) becomes large. Asymptotic limits are useful since they allow vectors such as Π_s to be replaced by a consideration of the distribution of the individual components, π_{is} ($s = 1, \dots, S$). Statements about the distribution parameters are much more general than the qualified inequalities of eq. (1.19),²¹ and an explicit relationship about the trade-off between, say, mean probability in state s and the variance of the probabilities in state s , may be derived.

Consider the case where each investor has a probability vector Π_i and an initial endowment \hat{C}_i allocated randomly from the same underlying distributions; each investor has the same utility function. The result of the individual optimization exercise is a value for λ_i^* which is also a random drawing for each investor from the same underlying distribution (this would not be the case were utility functions to vary across investors). Denote the mean probability in state s as $\bar{\pi}_s$ ($s = 1, \dots, S$) and the mean of the resulting multipliers as $\bar{\lambda}$ —clearly these two means will be

²¹For example, if the mean probability in state s is denoted $\bar{\pi}_s$ ($s = 1, \dots, S$) then the set of probabilities for which $\Pi_s > \Pi_t$ is a strict subset of that for which $\bar{\pi}_s > \bar{\pi}_t$.

related in some manner.

Now consider expanding each function in the summation of eq. (1.17) around the means, $\bar{\lambda}$ and $\bar{\pi}_s$. To simplify the notation, set

$$g(\lambda, \pi) \stackrel{\text{def}}{=} f\left(\frac{\lambda p_s}{\pi}\right) \quad (1.20)$$

(once again, the i subscript on f is no longer required since all utility functions are the same). Using the approximation to second order,

$$\begin{aligned} g(\lambda_i, \pi_{is}) \approx & g(\bar{\lambda}, \bar{\pi}_s) + \begin{bmatrix} \lambda_i - \bar{\lambda} \\ \pi_{is} - \bar{\pi}_s \end{bmatrix}' \begin{bmatrix} g_\lambda \\ g_\pi \end{bmatrix}_{\bar{\lambda}, \bar{\pi}_s} \\ & + \frac{1}{2} \begin{bmatrix} \lambda_i - \bar{\lambda} \\ \pi_{is} - \bar{\pi}_s \end{bmatrix}' \begin{bmatrix} g_{\lambda\lambda} & g_{\lambda\pi} \\ g_{\pi\lambda} & g_{\pi\pi} \end{bmatrix}_{\bar{\lambda}, \bar{\pi}_s} \begin{bmatrix} \lambda_i - \bar{\lambda} \\ \pi_{is} - \bar{\pi}_s \end{bmatrix}. \end{aligned} \quad (1.21)$$

Using this approximation in eq. (1.17) gives

$$\begin{aligned} \frac{\hat{c}_s}{n} & \approx \frac{1}{n} \sum_{i=1}^n g(\lambda_i, \pi_{is}) \\ & = g(\bar{\lambda}, \bar{\pi}_s) + (\bar{\lambda}^{(n)} - \bar{\lambda}) g_\lambda + (\bar{\pi}_s^{(n)} - \bar{\pi}_s) g_\pi \\ & \quad + \frac{1}{2} \left\{ \sigma_\lambda^{2(n)} g_{\lambda\lambda} + \sigma_\pi^{2(n)} g_{\pi\pi} + 2\sigma_{\lambda\pi}^{(n)} g_{\lambda\pi} \right\} \end{aligned} \quad (1.22)$$

where the superscript (n) indicates the n -term sample statistic, σ denotes the variance and covariance, and all expressions in g are evaluated at $(\bar{\lambda}, \bar{\pi}_s)$. The sample means approximate the population mean to order $O(n^{-0.5})$, while we can assume that the sample variances and covariances approximate to $O(n^{-1})$. Therefore eq. (1.22)

can be written (remembering that the lefthand side also tends to a limit)

$$\begin{aligned} \bar{c}_s + O(n^{-0.5}) &\approx g(\bar{\lambda}, \bar{\pi}_s) + O(n^{-0.5}) g_\lambda + O(n^{-0.5}) g_\pi \\ &+ \frac{1}{2} \left\{ (\sigma_\lambda^2 + O(n^{-1})) g_{\lambda\lambda} + (\sigma_\pi^2 + O(n^{-1})) g_{\pi\pi} \right. \\ &\left. + 2 (\sigma_{\lambda\pi} + O(n^{-1})) g_{\lambda\pi} \right\}, \end{aligned} \quad (1.23)$$

where \bar{c}_s is the mean endowment of security s for each investor. In the limit $n \rightarrow \infty$ this gives

$$\bar{c}_s \underset{n \rightarrow \infty}{\approx} g(\bar{\lambda}, \bar{\pi}_s) + \frac{1}{2} \left\{ \sigma_\lambda^2 g_{\lambda\lambda} + \sigma_\pi^2 g_{\pi\pi} + 2 \sigma_{\lambda\pi} g_{\lambda\pi} \right\}. \quad (1.24)$$

This expression can be simplified further on the realization that $\sigma_{\lambda\pi} = 0$. Since the probabilities sum to 1, a larger-than-average probability in state s will be matched by a smaller-than-average probability in some other state. Therefore each investor will have a collection of probabilities with both positive and negative deviations and the unconditional covariance between any probability and the multiplier λ_i will be zero.

From the definition of $g(\cdot)$, the partials can be found as

$$g_{\lambda\lambda} = f'' \left(\frac{p_s}{\pi} \right)^2, \quad g_{\pi\pi} = f'' \left(\frac{\lambda p_s}{\pi^2} \right)^2 + 2f' \left(\frac{\lambda p_s}{\pi^3} \right). \quad (1.25)$$

Furthermore, from the sufficient condition,

$$f' = \frac{1}{u''}, \quad f'' = \frac{-u'''}{(u'')^3} \quad (1.26)$$

(noted by Varian, 1985, p. 315). Making these substitutions and rearranging gives

the final expression

$$\bar{c}_s \underset{n \rightarrow \infty}{\approx} f\left(\frac{\bar{\lambda} p_s}{\bar{\pi}_s}\right) + \frac{1}{2} \left(\frac{\bar{\lambda} p_s}{\bar{\pi}_s}\right)^2 \times \left\{ \left(\frac{-u'''}{(u'')^3}\right) \left(\frac{\sigma_{\bar{\lambda}}^2}{\bar{\lambda}^2} + \frac{\sigma_{\bar{\pi}_s}^2}{\bar{\pi}_s^2}\right) + \left(\frac{1}{u''}\right) \left(\frac{2\sigma_{\bar{\pi}_s}^2}{\bar{\pi}_s \bar{\lambda} p_s}\right) \right\}. \quad (1.27)$$

What does eq. (1.27) say about equilibrium asset prices? The first term in the curly bracket is positive (presuming $u''' > 0$) and the second negative. Thus the sign of the second term in eq. (1.27) is, strictly speaking, indeterminate. This indeterminance may be inherent in the problem or it may be a consequence of the truncation of the Taylor Series after the second term—there is no way to distinguish. Proceeding on the basis that the approximation is a good one (i.e. the indeterminance is inherent) we may use Varian's results to 'configure' the sign of the second term.

Varian showed that $\Pi_s < \Pi_t \Leftrightarrow p_s < p_t$, all other things being equal (see eq. (1.19)). This implies that the curly bracket *probably* has negative sign with its second term dominating the first, and in the first term the squared coefficient of variation of λ_i dominating that of π_{is} .²² In this way, a higher $\bar{\pi}_s$ raises the first term in eq. (1.27), since $f' < 0$, and makes the second less negative. This can be offset by a higher p_s , which lowers the first term and makes the second more negative. The negative sign on the curly bracket is also consistent with a higher variance of probabilities, $\sigma_{\bar{\pi}_s}^2$, lowering prices. The higher variance makes the curly bracket more negative, and this is offset by a lower price which makes the curly bracket less negative and also makes the first term more positive.²³

²²To expand a little on this, consider the case where there are just a few probable states, so that we might take $\bar{\pi}_s \approx 0.3$ and $\sigma_{\bar{\pi}_s} \approx 0.1$. In this case the term $(\sigma_{\bar{\pi}_s}/\bar{\pi}_s)^2 \approx 0.1$. The values of the mean and variance of λ_i^* are completely unknown except for their sign—positive. But it seems plausible to assert that, in the absence of any other evidence, $(\sigma_{\bar{\lambda}}/\bar{\lambda})^2 \gg 0.1$, and hence the λ term dominates the π term.

²³I appreciate that this type of 'backward induction' is at best heuristic, and at worst illogical. However, it seems a worthwhile exercise in 'probabilistic' reasoning, relative to the alternative of leaving the implications of eq. (1.27) completely unresolved.

Therefore eq. (1.27) is capable of being consistent with Varian's results, and, possibly, transcending them in generality. It also provides a platform for consideration of more explicit utility functions, although this must take place jointly with an analysis of the functional dependencies between the four distributional parameters that arise from within their definitions and from the optimization process.

As for the impact of these results on the partial equilibrium models of the following chapters, it must be born in mind that considerations of equilibrium pricing are made here on a strictly cross-sectional basis, i.e. comparisons are between two security prices in the same GE equilibrium. The concern of the following chapters is the evolution of equilibrium prices through time. It is *not* an implication of the GE analysis that a security in which the dispersion of reservation prices increases between one equilibrium and the next will *necessarily* suffer a drop in price within a GE framework. To examine this case would require a dynamic heterogeneous GE model. However, insofar as the cross-sectional case has any relevance, the probability of finding such an effect in dynamic heterogeneous GE is increased. But this effect will be ingored in the forthcoming chapters, where the market for a financial asset will be considered in isolation from that of other assets.

Part I

Return Leptokursis and News

Arrival

Chapter 2

The Role of News

An overview of the literature on price/volume models has already been given in Section 1.4. This chapter centres around the seminal price-volume model of Tauchen and Pitts (1983). This model serves as a focal point for Part I of this thesis, with subsequent chapters being concerned with the issues raised by the explicit assumptions concerning investor behaviour and investor homogeneity, the implicit assumptions about the news arrival process, and the data requirements.

The outline of the chapter is as follows. The first two sections review the models of Clark (1973) and Epps and Epps (1976), respectively. Section 2.3 presents the model of Tauchen and Pitts (1983), a complete model of price/volume dynamics in which the trading activities of pseudo-homogeneous investors are driven by the news-arrival process. Section 2.4 is a digression in which the implications (particularly statistical) of the Tauchen and Pitts model are developed much further than in the original paper. Section 2.5 is a theoretical critique of the model, while Section 2.6 considers the more recent evidence on the news-arrival process, and its implications for price/volume dynamics within the Tauchen and Pitts model. Section 2.7 concludes.

2.1 Leptokursis and Clark's Model

As was discussed in Subsection 1.4.2, Clark (1973) generalized the mixture of distributions model proposed by Fama (1965) to a *subordinated stochastic process*. In Clark's model, prices evolve according to the amount of news arrival (or 'market' time), rather than according to the passing of calendar time. If the information on day t is represented by Ω_t , then the amount of news is, crudely,

$$n_t \stackrel{\text{def}}{=} \text{number } \{ \Omega_t - \Omega_{t-1} \} \quad (2.1)$$

(i.e. the number of bits by which the information stock has increased). The daily price series is not really p_1, p_2, \dots but $p_{\Omega_1}, p_{\Omega_2}, \dots$. Price evolution in market time is said to be subordinate to that in calendar time, and the information stock Ω_t is known as the directing process.

In Clark's model the daily price change, $\Delta p_t \stackrel{\text{def}}{=} p_t - p_{t-1}$, is the sum of a random number of independent random shocks:

$$\Delta p_t = \sum_{i=1}^{n_t} \Delta p_{it}, \quad (2.2)$$

where Δp_{it} is i.i.n. $(0, \sigma_p^2)$ for all i and t . The condition $\mathcal{E}[\Delta p_{it}] = 0$ imposes the martingale property. In market time, as measured by increments in the information stock, prices are still a random walk since each shock is drawn from the same (normal) distribution. But in calendar time, as measured by the passing of trading days, prices are only a martingale since the expected price change is zero but the variance depends upon the amount of news. As it stands eq. (2.2) can describe a wide range of different processes, depending crucially upon whether n_t has time-varying properties. However, initially it was supposed that n_t was independently and identically distributed through time.

The distribution of daily price changes (Δp_t) depends upon the distribution of the individual shocks (Δp_{it}) and the nature of the directing process, Ω_t . Interestingly, prior to Clark's model, Mandelbrot and Taylor (1967) had considered a subordinated stochastic process as one explanation of price changes having a stable non-normal distribution.¹ In their model the increments were normal but the directing process was a stable non-normal variable with infinite mean. Clark showed that any stochastic directing process would generate leptokurtosis in the return distribution, without the conceptual difficulties of infinite variance that were present in the Mandelbrot and Taylor model.²

The unconditional moments of Δp_t can be found by conditioning on n_t , which is assumed to have finite mean and variance, μ_n and σ_n^2 respectively for all t . The mean and the variance of Δp_t are straightforward:

$$\mathcal{E} [\Delta p_t] = \mathcal{E} [\mathcal{E} [\Delta p_t | n_t]] = \mathcal{E} [n_t \mathcal{E} [\Delta p_{it}]] = 0, \quad (2.3)$$

$$\mathcal{V} [\Delta p_t] = \mathcal{E} [\Delta p_t^2] = \mathcal{E} [\mathcal{E} [\Delta p_t^2 | n_t]] = \mathcal{E} [n_t \sigma_p^2] = \mu_n \sigma_p^2. \quad (2.4)$$

The skewness of Δp_t is clearly zero since Δp_{it} is symmetric with zero mean. To find the kurtosis, note that given n_t , Δp_t is normally distributed and so must have kurtosis equal to 3, i.e.

$$\mathcal{E} [(\Delta p_t)^4 / (n_t^2 \sigma_p^4) | n_t] = 3. \quad (2.5)$$

Taking the unconditional expectation of Δp_t^4 gives

$$\mathcal{E} [\Delta p_t^4] = \mathcal{E} [3 (n_t^2 \sigma_p^4)] = 3 \sigma_p^4 (\sigma_n^2 + \mu_n^2). \quad (2.6)$$

¹See Subsection 1.4.2.

²In defense of his model, Mandelbrot noted in another paper "An added virtue of the Gaussian is that its moments are finite, but after all, moments are an acquired taste" (Mandelbrot, 1973, p. 158).

Dividing by the squared unconditional variance gives the result

$$\text{Leptokursis} = 3 \left(\frac{\mu_n^2 + \sigma_n^2}{\mu_n^2} \right) = 3(1 + \tau_n^2), \quad (2.7)$$

where $\tau_n \stackrel{\text{def}}{=} \sigma_n / \mu_n$, the coefficient of variation of n_t .

From eq. (2.7) it is clear that the degree of leptokursis in the price differences will be determined by the coefficient of variation of the news process. There will be substantial leptokursis if the standard deviation of news is high relative to its mean. Using price differences from cotton futures, Clark finds values for kurtosis of 19.45 (1957–1950) and 20.49 (1951–1955). This suggests that if his model is correct the coefficient of variation of news is about 2.4. A priori, this seems rather high for a non-negative random quantity such as news. Consider, for example, the poisson distribution, which is often used to represent arrival processes. This has a coefficient of variation of $\lambda^{-0.5}$, where λ is the arrival rate (i.e. mean number of arrivals per period). If the figure of 2.4 were correct, this would imply that the arrival rate was about 0.2, and the probability of a no-news day was about 0.8. Were this to be true, four days in every five would have zero price change—clearly not consistent with the evidence. Therefore, this very high degree of leptokursis casts doubts on the constancy of the parameters during these two periods, or more generally on the model itself.

Clark did not make this observation about the implied coefficient of variance, but went on to test his model using trading volume to proxy the amount of news arriving in a day. Were trading volume (v_t) and news (n_t) to be perfectly correlated (i.e. $v_t = a + b n_t$), then $\mathcal{E} [\Delta p_t^2 \mid v_t]$ would be linear in v_t since $\mathcal{E} [\Delta p_t^2 \mid n_t] = \sigma_p^2 n_t$. In fact, Clark finds that a highly convex relationship fits far better than a linear one. This is not that surprising in the light of the comment above about the high coefficient of variation: the volume data need to be ‘beefed up’ in order to fulfill the

role of a proxy for news.

2.2 The Epps and Epps Model

The Clark (1973) model provided a very intuitive rationale for leptokurtosis in daily stock price changes: the amount of news per day is not fixed. However, trading volume appeared only as a rather imperfect proxy for the amount of news arrival, rather than an endogenous outcome of market clearing: in the taxonomy of Section 1.4 this is a descriptive model, or at best a homogeneous investor model. Epps and Epps (1976), hereafter EE, incorporated volume directly by considering the decisions of individual investors: this is a pseudo-homogeneous model, with a basis in optimizing investor behaviour. In this way EE were able to model the influence of news on the amount of stock demanded by any investor, and the subsequent volume generated as each investor altered his position in the light of changes in the market-clearing price.

In the EE model, each investor maximizes utility over the blend of assets in his portfolio, subject to a wealth constraint. The end of period asset values are unknown, but utility is assumed to be a function of a portfolio's expected value and its variance, and so the problem is

$$\max_{q_0, Q} u(q_0 x_0 + Q'X, Q'SQ) \quad \text{subject to} \quad q_0 p_0 + Q'P = w. \quad (2.8)$$

In eq. (2.8), Q is the vector of quantities of risky assets, X their expected end-of-period values (including coupons, dividends, etc.), S the covariance matrix of these values, P the current prices, and w the investor's current wealth; the riskless asset, (asset 0) has been separated out.

It is interesting to note the implicit assumption that the general portfolio problem can be written as a problem involving the maximization of a function over portfolio

mean and variance. This is only the case when the distribution of asset prices at the end of the period is jointly normal. In other cases the problem of the maximization of expected utility does not collapse in this convenient fashion (see, e.g., Copeland and Weston, 1988, pp. 96–9). This has a bearing on the issue of rational expectations, discussed briefly in Section 1.3, which will be returned to in Subsection 2.3.2.

The Lagrangian for this problem is

$$\mathcal{L}(\lambda, q_0, Q) = u(q_0 x_0 + Q'X, Q'SQ) - \lambda (q_0 p_0 + Q'P - w) \quad (2.9)$$

from which the first order conditions are

$$\mathcal{L}_\lambda = -(q_0 p_0 + Q'P - w) = 0, \quad (2.10)$$

$$\mathcal{L}_{q_0} = u_1(\cdot) x_0 - \lambda p_0 = 0, \quad (2.11)$$

$$\mathcal{L}_Q = u_1(\cdot) X + u_2(\cdot) (2SQ) - \lambda P = \{0\}, \quad (2.12)$$

where the subscripts on $u(\cdot)$ indicate partial derivatives. Substituting for λ from eq. (2.11) into eq. (2.12) and dividing through by $u_1(\cdot)$ gives

$$X + 2(u_2/u_1) SQ - (x_0/p_0) P = \{0\}. \quad (2.13)$$

Now EE assume that the investor has constant absolute risk aversion, $\beta \stackrel{\text{def}}{=} -u_2/u_1$. This suggests that the investor trades off an increased expectation (μ) against an increased variance (σ^2) in the linear fashion $u = \mu - \beta \sigma^2$; risk-aversion, i.e. $\beta > 0$, is a sufficient condition for utility maximization (for further details see Varian, 1992, pp. 189–90). Denoting the riskless rate of interest $i \stackrel{\text{def}}{=} x_0/p_0 - 1$, this gives the solution for Q from eq. (2.13) as

$$Q = (2\beta S)^{-1} (X - (1+i)P). \quad (2.14)$$

Concentrating on the investor's demand for any one particular risky asset, say k , eq. (2.14) can be written

$$\begin{aligned} q_k &= \frac{1}{2\beta} \sum_{j=1}^n (S^{-1})_{kj} (x_j - (1+i)p_j) \\ &\equiv c(p_k^* - p_k) \quad k = 1, \dots, n \end{aligned} \quad (2.15)$$

where

$$c \stackrel{\text{def}}{=} \frac{(S^{-1})_{kk} (1+i)}{2\beta}, \quad (2.16)$$

$$p_k^* \stackrel{\text{def}}{=} \frac{x_k}{1+i} + \sum_{j \neq k} \frac{(S^{-1})_{kj}}{(S^{-1})_{kk}} \left(\frac{x_j}{1+i} - p_j \right). \quad (2.17)$$

By arranging the demand functions in this fashion it is possible to interpret p_k^* as the *reservation price* of asset k , such that at prices greater than p_k^* the investor is a seller of asset k , and at prices less than p_k^* a buyer (see Definition 1.1 regarding reservation prices). This demand function will be crucial in the model of Tauchen and Pitts (1983) to be examined below. The strength of the EE model is that the demand functions of the form in eq. (2.15) have their basis in portfolio optimization, and it is clear how the reservation prices in these demand functions depend on a number of factors including all other prices and all expected end of period values.

Assumptions in the EE model

Up to this point, the only major assumption of EE is that of constant absolute risk aversion. Additionally, they will require that every investor has the same degree of risk aversion and the same assessment of the covariance matrix, which is sufficient to ensure that c in eq. (2.16) is equal across investors. From here they go on to relate the change in the market clearing price to the changes in the reservation prices of investors, and they identify trading volume as arising from each investor's rebalancing of his portfolio in the light of a change in the difference between the

reservation price and the market clearing price.

However, in the development of their model to the point at which it yields a relationship between price change variance and trading volume, EE require a further succession of unlikely assumptions. First, they require that at successive market clearing prices there will always be an equal number of buyers and sellers (p. 307). Second, they require that the summation term in p_j in eq. (2.17) be of negligible size, which denies the realism of the portfolio theory approach (p. 308). Third, they impose a specific functional form on the relationship between the size of the change in investors' reservation prices and the extent to which investors disagree (p. 309, eq. 15). Consequently EE are able to show that the expected logarithm of the variance of price changes is a linear function of the logarithm of volume.

The benefit of this very highly-structured model is that no specific distribution need be assumed for changes in the reservation prices. However, to estimate their model by maximum likelihood, EE are required to specify a distribution for the change in market clearing price, and so this hard-won benefit is of little practical use. The estimated model gave a concave relationship between Δp_t^2 and v_t , rather than the convex relationship which was found by Clark (1973). However, it is quite possible, due to the large number of restrictions imposed prior to deriving a testable model, that EE's results are misspecified. It is also possible that EE's results and Clark's results are not directly comparable, since Clark's sample covered two four year periods while EE use data from one month. The kurtosis in the EE data averages 3.48, compared with Clark's measure of about 20. By the same calculations used for Clark's data, this equates to an arrival rate for news of 6.25, and a probability of a no-news day of 0.002—much more reasonable.

One explanation for the difference in the two sets of results is that the number of investors active in the market has an impact upon price/volume dynamics, and that Clark's sample period was long enough to permit substantial variation in this

quantity. The size of the market is an important part of the Tauchen and Pitts (1983) model, to be examined in Section 2.3. Another explanation is that there is some low-frequency variation in the mean amount of news per day, which might only show up over periods longer than a month. This was one of the conclusions of Tauchen and Pitts (1983), and will be examined in more detail in Section 2.5.

Despite the assumptions and the ambiguity of the empirical results, the EE model is important because it pioneered the study of the price/volume relationship as the aggregation of individual investors' changing demands, and provided a rationale for the 'linear' demand functions without which aggregation is extremely complicated.

2.3 The Tauchen and Pitts Model

To recapitulate on the two models described above, Clark (1973) developed the idea that variations in the amount of news arrival would cause variation in the variance of daily price changes, and identified empirically that trading volume picked up some of these effects. Epps and Epps (1976) showed how linear demand functions might be used to examine the price/volume relationship as arising from successive portfolio rebalancings in a model of pseudo-homogeneous investors. Tauchen and Pitts (1983), hereafter TP, combined these two models with a striking simplification of the determination of reservation prices. From this they derive a specification for the joint distribution of daily price change variance and daily trading volume, parametrized by the size of the market as measured by the number of active investors.

2.3.1 Intra-Day Model

The first stage in the development of the TP model is a framework that describes the evolution of the market clearing price and the generation of trading volume between successive market equilibria, i.e. on a 'per news-item' basis. TP start with

the demand function given in eq. (2.15). These demand functions are now expressed for investor j with regard to some particular speculative asset at the i^{th} within-day equilibrium:

$$q_{ji} = c(p_{ji}^* - p_i), \quad j = 1, \dots, J \quad (2.18)$$

where, as in the EE model, the parameter c is assumed unchanging over time and across investors. For the moment, consider J to be fixed.

To clear the market requires that $\sum_j q_{ji} = 0$, implying that the market-clearing price at equilibrium i is simply the mean of investors' reservation prices, and the market-clearing price change between $i-1$ and i the mean of the change in investors' reservation prices is

$$\Delta p_i = J^{-1} \sum_{j=1}^J \Delta p_{ji}^* \quad (2.19)$$

The trading volume generated by investor j between $i-1$ and i is simply $|q_{ji} - q_{j,i-1}|$ (positive whether the investor increases or decreases his position), but since every investor's sale is another investor's purchase, the total trading volume between $i-1$ and i is half of the sum across investors,

$$v_i = \frac{1}{2} \sum_{j=1}^J |q_{ji} - q_{j,i-1}| = \frac{c}{2} \sum_{j=1}^J |\Delta p_{ji}^* - \Delta p_i|, \quad (2.20)$$

by eq. (2.18), where Δp_i is itself a function of the reservation prices by eq. (2.19). Therefore, both Δp_i and v_i are functions of the change in investors' reservation prices. The specification of the joint distribution of these changes across investors leads directly to the specification of the joint distribution of price changes and trading volume, per market equilibrium. This very general result has arisen from the linear demand functions in eq. (2.18) and the invariance of the scaling coefficient

c across investors at each point in time.

TP propose a simple variance-components model for the change in investors' reservation prices:

$$\Delta p_{ji}^* = \phi_i + \psi_{ji} \quad \text{where} \quad \phi_i \sim \text{i.i.n.}(0, \sigma_\phi^2), \psi_{ij} \sim \text{i.i.n.}(0, \sigma_\psi^2). \quad (2.21)$$

Thus each investor's reservation price changes as the sum of a common component ϕ_i and a unique component ψ_{ji} . The martingale condition is imposed on the model by requiring the expectations of both terms to be zero.

Digression on Rational Expectations

It is interesting to note that in choosing the normal distribution for the shocks ϕ and ψ , TP satisfy rational expectations in their model. The reasoning is straightforward. If reservation prices are a random walk with normally-distributed increments, then each market-clearing price will also be a random walk with normally-distributed increments, *because the demand functions are linear*. Conversely, linear demand functions arise from the belief that the distribution of prices at the end of the period is jointly normal. Investors' expectations are consistent with the outcome of the model, and therefore they are rational.

The normally-distributed increments to reservation prices may be justified by the Central Limit Theorem. From the reservation price expression eq. (2.17) the reservation price is a linear combination of expected prices, with the weights being determined by the covariance matrix. Therefore if a piece of news causes each investor to alter each expectation by a random amount, the weighted sum of these adjustments will be asymptotically normally distributed, as long as the adjustments are drawn at random from a distribution with finite second moment.

Returning to the model, the incorporation of the variance decomposition model,

eq. (2.21), into the price change and volume expressions, eq. (2.19) and eq. (2.20), gives

$$\Delta p_i = \phi_i + J^{-1} \bar{\psi}_i \quad (2.22)$$

$$v_i = \frac{1}{2} \sum_{j=1}^J |\psi_{ji} - \bar{\psi}_i| \quad (2.23)$$

where $\bar{\psi}_i \stackrel{\text{def}}{=} \sum_j \psi_{ji}$, from eq. (2.19) and eq. (2.20). The absence of ϕ_i in eq. (2.23) indicates that large price changes can occur without generating large volume. This is a consequence of the linear demand functions. Were all reservation prices to jump by the same amount following a piece of news, the market clearing price would also need to jump by that amount, in which case no investor would want to change his position and no volume would be generated. Clearly, trading volume is generated in this model as a consequence of disagreements between pseudo-homogeneous investors regarding the interpretation of commonly-available news.

TP show that both distributions are asymptotically normal:

$$\Pr\{\Delta p_i < x\} = \Phi(x; 0, \sigma_p^2) \quad (2.24)$$

$$\lim_{J \rightarrow \infty} \Pr\{v_i < x\} = \Phi(x; \mu_v, \sigma_v^2) \quad (2.25)$$

where $\Phi(\cdot)$ is the cumulative normal distribution, and

$$\sigma_p^2 \stackrel{\text{def}}{=} \sigma_\phi^2 + \frac{\sigma_\psi^2}{J} \quad (2.26)$$

$$\mu_v \stackrel{\text{def}}{=} \left(\frac{c\sigma_\psi}{2}\right) \sqrt{\frac{2}{\pi}} J \quad (2.27)$$

$$\sigma_v^2 \stackrel{\text{def}}{=} \left(\frac{c\sigma_\psi}{2}\right)^2 \left(1 - \frac{2}{\pi}\right) J + o(J) \quad (2.28)$$

Additionally they note that Δp_i and v_i are stochastically independent given J . Thus the distribution of price changes is determined for large J primarily by the variance of the common component ϕ_i , while the distribution of volume is dependent upon

the variance of the investor-specific components ψ_{ji} .

To summarize the intra-day model, the market is initially in equilibrium and then a piece of news arrives which causes investors to reassess their reservation prices. Following this reassessment, investors move to a new equilibrium in which trading volume is generated. This movement to a new equilibrium presupposes the existence of a Walrasian auctioneer or some similar market-clearing device. If the number of investors is large and fixed, the distribution of the change in price and the distribution of the amount of trading volume generated are both normal, and stochastically independent.

2.3.2 Inter-Day Model

TP go on to incorporate this inter-equilibrium model within a model similar to that of Clark (1973), in which the number of distinct equilibria during a trading day is itself a random variable.³ Hence

$$\Delta p_t = \sum_{i=1}^{n_t} \Delta p_{it} \quad (2.29)$$

$$v_t = \sum_{i=1}^{n_t} v_{it} \quad (2.30)$$

where the t subscript has been appended to Δp_i and v_i to indicate the day of the intra-day equilibria. The expression for Δp_t given in eq. (2.29) is as in Clark's model (given as eq. (2.2)). Additionally, however, the TP model also gives an explicit outcome for volume, eq. (2.30). The inter-day shocks Δp_{it} and v_{it} have the distributions given in eq. (2.24) and eq. (2.25), respectively. The unconditional moments of Δp_t have been given previously in eq. (2.3)–eq. (2.7). Once again, by

³Harris (1987) also derives a model of this kind for price changes, and he suggests using the number of *transactions* per day as a proxy for news, rather than trading volume. However, this modification is not appropriate when demand functions are strictly monotonic, since it is then almost always the case that all investors transact every period, even if it is by a very small amount.

conditioning on n_t , it can also be shown that

$$\mathcal{E}[v_t] = \mu_v \mu_n, \quad (2.31)$$

$$\mathcal{V}[v_t] = \mu_v^2 \sigma_n^2 + \sigma_v^2 \mu_n, \quad (2.32)$$

$$\mathcal{C}[(\Delta p_t)^2, v_t] = \sigma_p^2 \mu_v \sigma_n^2. \quad (2.33)$$

Therefore TP implies at the very least that there will be a positive relationship between squared price changes and daily trading volume whenever the amount of news per day is non-constant.

However, TP do not solve eq. (2.29) and eq. (2.30) for any further implications of their model for the price/volume relationship. Instead, having estimated the parameters on the assumption that news has a log-normal distribution, they show by numerical integration that $\mathcal{E}[(\Delta p_t)^2 | v_t]$ has the following properties: (i) $\mathcal{E}[(\Delta p_t)^2 | v_t = 0] > 0$; (ii) $\mathcal{E}[(\Delta p_t)^2 | v_t]$ is increasing in v_t ; (iii) $\mathcal{E}[(\Delta p_t)^2 | v_t]$ appears to asymptote to a ray through the origin (all inferred from Figure 1, p. 502).

The next section generalizes and extends these results.

2.4 The Implied Price-Volume Relationship

Tauchen and Pitts fail to extract the structure of the price-volume relationship from their inter-day model, beyond identifying that there will be a positive covariance between squared price changes and volume and considering one special case, mentioned above. In this section the inter-day model will be restated for generality and convenience as

$$S = x_1 + x_2 + \cdots + x_N \quad (2.34)$$

$$T = y_1 + y_2 + \cdots + y_N \quad (2.35)$$

where x_i i.i.d. $(0, \sigma^2)$ with support on the real line, y_i i.i.d. (μ, τ^2) with support on the non-negative integers, where in particular $\Pr\{y_i = 0\} > 0$, and N is a random quantity with support on the non-negative integers. Thus it is not required that x_i or y_i be normal, merely that they have the appropriate means and finite second moments, and the specification of N is unstated (although it is likely to be of Poisson-type). Two sets of results will be derived. The first concerns the behaviour of $\mathcal{E}[S^2 | T]$, which corresponds to $\mathcal{E}[(\Delta p_t)^2 | v_t]$ in the Tauchen and Pitts inter-day model.⁴

Proposition 2.1 *Given the relations eq. (2.34) and eq. (2.35)*

$$\mathcal{E}[S^2 | T] = \sigma^2 \mathcal{E}[N | T], \quad (2.36)$$

and $\mathcal{E}[N | T]$ has the following properties:

1. $\mathcal{E}[N | T = 0] > 0$,
2. $\mathcal{C}[\mathcal{E}[N | T], T] = \mu \mathcal{V}[N] > 0$, but
3. $\mathcal{E}[N | T]$ is not necessarily increasing in T .

Proof: The first part may be shown by conditioning the expectation of S^2 on N and T :

$$\begin{aligned} \mathcal{E}[S^2 | T] &= \mathcal{E}[\mathcal{E}[S^2 | N, T] | T] \\ &= \mathcal{E}[N \sigma^2 | T] \\ &= \sigma^2 \mathcal{E}[N | T]. \end{aligned} \quad (2.37)$$

In other words, the properties of $\mathcal{E}[S^2 | T]$ are directly proportional to the properties of $\mathcal{E}[N | T]$.

⁴I would like to thank Dr. Matthew Penrose and Dr. Peter Craig for their suggestions regarding the stochastic process described in eq. (2.34) and eq. (2.35).

A heuristic explanation will suffice for (1), although a more formal argument can be found in Section 5.5. For $T = 0$ it is possible that $N = 0$, or that $N > 0$ and each $y_i = 0$. Therefore the probability-weighted (i.e. expected) value of N given that $T = 0$ must be greater than zero as long as $\Pr\{y_i = 0\} > 0$, which is assumed.

To prove part (2), consider the conditional covariance identity:

$$\mathcal{C}[N, T] = \mathcal{E}[\mathcal{C}[N, T | N]] + \mathcal{C}[\mathcal{E}[N | N], \mathcal{E}[T | N]]. \quad (2.38)$$

Hence by conditioning on N , $\mathcal{C}[N, T] = 0 + \mu \mathcal{V}[N]$. However, the same result may be found by conditioning on T , which gives $\mathcal{C}[N, T] = 0 + \mathcal{C}[\mathcal{E}[N | T], T]$. Equating these two proves part (2), and the covariance must be positive since y_i has support on the non-negative integers, which implies $\mu > 0$ providing only that $\Pr\{y_i > 0\} > 0$.

Part (3) may be proved by example. Consider the case where y_i can take only two values, 1 and 10 with equal probability. This choice for y_i violates the condition $\Pr\{y_i = 0\} > 0$ in the interests of simplicity, but it should be clear that the result also holds for the case where this probability is positive but vanishingly small. Clearly, $\mathcal{E}[N | T = 9] = 9$, since the only possible outcome giving $T=9$ is $y_1 = y_2 = \dots = y_9 = 1$. But for $T = 10$ the expected value for $N | T$ is only a little greater than 1, since the outcome $y_1 = 10$ is far more likely than the alternative $y_1 = y_2 = \dots = y_{10} = 1$. This could be shown formally using Bayes Theorem. ■

The interpretation of Proposition 2.1 is that $\mathcal{E}[(\Delta p_t)^2 | v_t]$ is generally increasing in v_t but not necessarily increasing everywhere. In particular, the example in the proof of point (3) suggests that when v_t is very small, the fact that it is discrete can cause 'lumpiness' in $\mathcal{E}[(\Delta p_t)^2 | v_t]$. However, if v_t tends to be large, this problem does not arise. This asymptotic property is formalized in the following proposition.⁵

⁵Actually, this Proposition is technically a Conjecture, the conjecture appearing as eq. (2.40). This statement may be true only for a certain class of compound processes, i.e. only for N and y_i

Proposition 2.2 *In the limit as $t \rightarrow \infty$, $(\mu S^2)/(\sigma^2 T) \mid T = t$ has a χ_1^2 distribution.*

Proof: The conditions under which the Central Limit Theorem may be applied to the sum of the terms x_1, \dots, x_N are satisfied (notably x_i ($i = 1, \dots, N$) independently and identically distributed with finite second moment). The Central Limit Theorem states that $S \mid N = n$ will have an asymptotically normal distribution with mean 0 and variance $\sigma^2 n$. Hence in the limit as $n \rightarrow \infty$, $(S^2)/(\sigma^2 N) \mid N = n$ has a χ_1^2 distribution, by definition.

From eq. (2.35), note that

$$\lim_{n \rightarrow \infty} \{(T/N) \mid N = n\} = \mu, \quad (2.39)$$

by the strong law of large numbers. From this I conjecture

$$\lim_{n \rightarrow \infty} f(N) \mid N = n \iff \lim_{t \rightarrow \infty} f(T/\mu) \mid T = t, \quad (2.40)$$

where f is any continuous function. In particular, if $f(N) \stackrel{\text{def}}{=} (S^2)/(\sigma^2 N)$, then

$$\lim_{n \rightarrow \infty} \frac{S^2}{\sigma^2 N} \mid N = n \iff \lim_{t \rightarrow \infty} \frac{\mu S^2}{\sigma^2 T} \mid T = t. \quad (2.41)$$

Since the first of these has been shown to have a χ_1^2 distribution, so the second must also have a χ_1^2 distribution, as was to be shown. ■

The following two corollaries follow directly from Propositions 2.1 and 2.2.

taking specific types of distribution. In the trivial case where $\Pr\{y_i = 1\} \rightarrow 1$ it is clearly true, since then $T = N$. More work is required to establish its generality. The simplest non-trivial case is probably the Poisson/Poisson. In this case it can be shown that $\mathcal{E}[N \mid T = t] = \mu'_{t+1}/\mu'_t$, where μ'_t is the moment of order t about zero of a Poisson distribution with mean $\lambda e^{-\mu}$, where $\lambda = \mathcal{E}[N]$. The analysis of this ratio requires the derivation of a recursive expression for the moments about zero of a Poisson distribution. Currently the only known recursion concerns the moments about the mean (see, e.g., Johnson and Kotz, 1969, p. 91).

Corollary 2.3 *Asymptotically in T ,*

$$\mathcal{E}[S^2 | T] = \frac{\sigma^2}{\mu} T \quad (2.42)$$

$$\mathcal{V}[S^2 | T] = 2 \left(\frac{\sigma^2}{\mu} \right)^2 T^2 \quad (2.43)$$

Proof: This corollary follows directly from the property of the χ^2 family of distributions, that the expectation is the number of degrees of freedom, and the variance twice the number of degrees of freedom. ■

Corollary 2.4 $\mathcal{E}[S^2 | T]$ *is nonlinear in T .*

Proof: Proposition 2.1 showed that $\mathcal{E}[S^2 | T]$ has a positive intercept. However, Proposition 2.2 implies that $\mathcal{E}[S^2 | T]$ asymptotes along a ray through the origin. The simultaneous satisfaction of these two conditions implies that $\mathcal{E}[S^2 | T]$ is nonlinear in T ; at the very least it is convex. ■

Therefore the findings of TP as summarized above and displayed in their Figure 1 are (with the exception of the monotonic increasing property) perfectly general, and do not depend upon their specific choice of the normal distribution for the price change and trading volume increments and the log-normal distribution for the amount of news per day, nor upon their chosen parameter values. In addition, the gradient of the ray along which $\mathcal{E}[(\Delta p_t)^2 | v_t]$ asymptotes is found as σ^2/μ , and $\mathcal{V}[(\Delta p_t)^2 | v_t]$ is shown to be asymptotically quadratic in v_t .

These results allow the inter-day TP model to be statistically tested (asymptotically), rather than estimated in conjunction with further assumptions. When TP attempt to estimate their model by maximum likelihood using the structural relations eq. (2.29) and eq. (2.30) they must assume not only that the news-arrival process is time-invariant but also a specific form for the distribution of the amount of news per day. Following their estimations, TP find misspecification in the form of

residual autocorrelation in both prices and trading volume. They attribute this to “nondeterministic low-frequency noise in both the number of traders and the rate at which new information flows to the market” (p. 500). Section 2.5 critically examines the TP model in the light of this misspecification, while Section 2.6 presents evidence which suggests that, as they conjectured, the news-arrival process is not time-invariant.

2.5 A Critical Evaluation of the TP Model

The equivocal results from estimation suggest that the TP model has some shortcomings as a specification for the relationship between price change and trading volume over time. These will be examined in this section.

Parameter Constancy

In the TP intra-day model pseudo-homogeneous investors are permitted to disagree about their reservation prices but not about the coefficient c in eq. (2.15). This permits disagreement between investors about expected prices but not about the terms in the covariance matrix.

The term c comprises three parts: the appropriate diagonal entry of the inverse of the covariance matrix, $(S^{-1})_{kk}$; the risk-free interest rate, i ; and the coefficient of risk-aversion, β . While none of these is strictly constant through time, it can be argued that both i and β are at least stationary: neither the interest rate nor risk-aversion are likely to be moving systematically in one direction. Therefore comparisons spanning a long period of time will probably be over different values of β and i , but not widely different. Parameter constancy is therefore a valid first-order approximation.

For $(S^{-1})_{kk}$, however, the situation is different. While it may be true that on

a day-to-day basis the mean of the price change distribution is not distinguishable from zero, over long periods it is quite clear that speculative prices trend upwards, and this would be expected for a risky asset in a market of risk-averse investors. Utility is defined on the value of the portfolio, and with a positive trend in prices this value must be expected to rise over long periods of time. Therefore both the vector of expected prices X and the covariance matrix S will have a time dimension. Comparisons spanning a long period of time will be over systematically different values for X and S , and so constancy in these parameters is not valid even as a first-order approximation. Furthermore, since the weights on the expected prices in the reservation price expression eq. (2.17) are derived from the covariance matrix, there might also be a time element in the disturbances ψ and ϕ .

Therefore, even if the TP intra-day model is taken to be an excellent description of the behaviour of the market at a point in time, it will not function as part of a time-series description of the market because it does not account for these systematic changes in c .

The Variance Decomposition Model

The second problem relates to the choice of joint distribution for the change in reservation prices. If reservation prices are not set with some reference to the market clearing price, then the dispersion of reservation prices around the market-clearing price tends to increase without limit for a non-trivial joint distribution of reservation price changes. This is exactly what happens in the TP model given in eq. (2.21).

Proposition 2.5 *If investors' reservation prices evolve according to the process in eq. (2.21), then the expected-second moment of the distribution of reservation prices around the market-clearing price increases linearly in the number of items of news.*

Proof: Consider a situation in which all reservation prices are the same—say unity for simplicity. The market-clearing price will also be unity, since the TP model implies that the market-clearing price is the mean of the reservation prices. After τ pieces of news, the dispersion of reservation prices about the market-clearing price can be measured by the expected second moment:

$$\begin{aligned} \mathcal{E} \left[J^{-1} \sum_{j=1}^J (p_{j,t+\tau}^* - p_{t+\tau})^2 \mid P_t^* = \{1\} \right] = \\ \mathcal{E} \left[J^{-1} \sum_{j=1}^J p_{j,t+\tau}^{*2} - p_{t+\tau}^2 \mid P_t^* = \{1\} \right], \end{aligned} \quad (2.44)$$

where $p_{j,t+\tau}^*$ is the reservation price of investor j at the τ^{th} equilibrium after t (the k subscript denoting the asset now being taken as given), and $p_{t+\tau}$ is the market-clearing price at the same time, which is equal to the mean reservation price at that time. Taking the first term in eq. (2.44),

$$\begin{aligned} J^{-1} \mathcal{E} \left[\sum_{j=1}^J p_{j,t+\tau}^{*2} \mid P_t^* = \{1\} \right] &= J^{-1} \sum_{j=1}^J \mathcal{E} \left[p_{j,t+\tau}^{*2} \mid P_t^* = \{1\} \right] \\ &= \mathcal{E} [p_{j,t+\tau}^{*2} \mid p_{j,t}^* = 1] \\ &= \mathcal{E} \left[(1 + \sum_{i=1}^{\tau} (\phi_i + \psi_{ji}))^2 \right] \\ &= (1 + \tau (\sigma_{\phi}^2 + \sigma_{\psi}^2)). \end{aligned} \quad (2.45)$$

where the distributions of ϕ_i and ψ_{ji} were given in eq. (2.21). Taking the second term in eq. (2.44),

$$\begin{aligned}
\mathcal{E} [p_{t+\tau}^2 \mid P_t^* = \{1\}] &= J^{-2} \mathcal{E} \left[\left(\sum_{j=1}^J [1 + \sum_{i=1}^{\tau} (\phi_i + \psi_{ji})] \right)^2 \right] \\
&= J^{-2} \mathcal{E} \left[\left(J (1 + \sum_{i=1}^{\tau} \phi_i) + \sum_{j=1}^J \sum_{i=1}^{\tau} \psi_{ji} \right)^2 \right] \\
&= J^{-2} \mathcal{E} \left[J^2 (1 + \sum_{i=1}^{\tau} \phi_i)^2 \right] \\
&\quad + J^{-2} \mathcal{E} \left[\left(\sum_{j=1}^J \sum_{i=1}^{\tau} \psi_{ji} \right)^2 \right] \\
&= 1 + \tau \sigma_{\phi}^2 + J^{-1} \tau \sigma_{\psi}^2.
\end{aligned} \tag{2.46}$$

Substituting eq. (2.45) and eq. (2.46) into eq. (2.44) gives the result

$$\mathcal{E} \left[J^{-1} \sum_{j=1}^J (p_{j,t+\tau}^* - p_{t+\tau})^2 \mid P_t^* = \{1\} \right] = \frac{J-1}{J} \tau \sigma_{\psi}^2, \tag{2.47}$$

which is linear in τ , the number of days forward from t . ■

The implication of Proposition 2.5 is that over time investors will become more and more extreme in their views, and they will hold bigger and bigger positions, both positive and negative, without limit. This effect does not show up in the variance of price changes precisely because the market-clearing price is the mean. Were the market-clearing price to be some measure on the reservation prices other than the mean this time-increasing dispersion would be reflected in a time-increasing price change variance.

For these two reasons, the intra-day TP model does not seem at all satisfactory as a time-series description of the way in which market-clearing prices respond to news. The role of the intra-day model is to establish functional dependencies between the moments of price change and volume, in relation to the number of active investors, J . Therefore this parameterization must be suspect. However, the failure of the intra-day model does not rule out the inter-day model as a description of the joint

distribution of price change and volume. It simply relieves the inter-day model of a certain amount of parameter inter-dependency.

2.6 The News Arrival Process

Suppose then that we join the TP model halfway through, at the stage where they propose the inter-day model given in eq. (2.29) and eq. (2.30). In other words we simply assert that this is an empirical specification which may explain aspects of the price/volume relationship, but without the functional inter-dependency of the moments of the individual shock terms (Δp_{it} and v_{it}) caused by the explicit consideration of the number of traders, as was described in eq. (2.26)–eq. (2.28). The crucial determinant of the time-series properties of price change and trading volume is now the news arrival process (n_t).

In the early work on news arrival, the amount of news per day was taken to be independently and identically distributed (Osborne, 1959; Mandelbrot and Taylor, 1967; Clark, 1973), and this was the model adopted by TP. They considered two distributions for the amount of news per day: the Poisson and the lognormal (where, notionally, the lognormal was used to approximate a discrete distribution). The Poisson has the superior claim on theoretical grounds, being the outcome of events that happen independently in time.⁶ However, the lognormal fitted much better and was preferred by TP for this reason and for consistency with Clark (1973). In fact the lognormal is almost certain to fit better, since it has an extra parameter which permits the variance to be determined independently of the expectation. The size of the improvement could be explained as misspecification, perhaps in presuming parameter constancy over the period.

⁶If arrivals follow a Poisson distribution, the time between arrivals is exponentially distributed (and *vice versa*). The exponential has the 'memoryless' property that the probability distribution is independent of how much waiting time has already elapsed (see, e.g., Ross, 1988, pp. 174–5).

ARCH Models

The TP model was derived just prior to the publication of Engle (1982), who described a new type of stochastic process for univariate time-series, that of Autoregressive Conditional Heteroskedasticity (ARCH). In a univariate ARCH process the variance of the price change distribution at time t is determined by lagged values of the squared price changes. Consequently, ARCH models display volatility persistence since one large price change can feed into the variance equation and generate a larger variance for the next few periods. This mixing of distributions with differing variances generates leptokurtosis in exactly the manner described by Clark (1973) and derived in eq. (2.7). Various generalizations of ARCH processes (particularly the GARCH process of Bollerslev, 1986, 1987) are now ubiquitous in finance (see, e.g., Bollerslev *et al.*, 1992; Bera and Higgins, 1993; Kim and Kon, 1994).

One interpretation of ARCH effects in speculative prices is that the distribution of the quantity of news is not independently and identically distributed, but positively autocorrelated (Diebold, 1986; Stock, 1988). This interpretation harks back to the clustering of volatilities observed by Fama (1965). There are, however, other explanations which are not news-related. Bera and Higgins (1993) suggest (i) random coefficients and/or (ii) non-linear autocorrelation, in the return process. Alternatively, heterogeneous investors models might generate ARCH effects: for example, noise traders are only sporadically exploited by informed traders.

Heat Waves and Meteor Showers

The news interpretation of ARCH effects in speculative prices has been investigated by Engle, Ito and Lin (Engle *et al.*, 1990; Ito *et al.*, 1992). Engle *et al.* (1990) categorize volatility persistence as being either country-specific (the so-called 'heat wave' model) or time-specific (the 'meteor shower'). Since news is a global phenomenon, volatility generated by news should follow the meteor-shower model and transmit

from one trading centre to another. But if volatility persistence is a consequence of intra-market dynamics such as noise trading then volatility will not transmit across market boundaries. Engle *et al.* also note an alternative cause of meteor-shower effects: stochastic policy coordination. For example, a change in US domestic monetary policy might affect the US market directly, but might also increase uncertainty about monetary policy in Japan. Engle *et al.* find that the data rejects the heat-wave model, but since their sample covered the period October 1985 to September 1986, a time of international monetary policy coordination, they were unable to distinguish between the news and policy coordination interpretations.

Ito *et al.* (1992) investigate the policy coordination interpretation directly, by using foreign exchange data from three different periods. The first is prior to the lifting of capital controls by Japan on 1 December 1980. The second and third are either side of the Plaza Accord on 22 September 1985, which heralded the period of closer monetary policy coordination. They find that the heat-wave model is best in the first period, and the meteor-shower in the second and third. This rejects the policy coordination interpretation, since the lack of policy coordination prior to the Plaza agreement should have ruled out the meteor-shower.⁷ The lifting of capital controls in Japan had the effect of integrating the Japanese markets into the world market, and its attendant news-process. However, Ito *et al.* also find that a certain amount of the volatility persistence is geographic, and does not transmit. Hence although their results support the news interpretation, they do not rule out market-microstructure effects.

Volume and Return Variance

Further evidence for positive autocorrelation in the quantity of news comes from Lamoureux and Lastrapes (1990). Lamoureux and Lastrapes model volatility per-

⁷Note, however, the observation of Professor O'Brien that this interpretation is conditional upon the Plaza Agreement being perceived by investors at the time as substantive, which is questionable.

sistence in stock returns as a simple GARCH process, and then investigate the effect of introducing trading volume linearly as an independent variable in the variance equation. Without volume there is strong evidence of ARCH effects. With volume included the ARCH effects are much weaker. However, there is a problem with this analysis, as the authors acknowledge. If both price change and volume are driven by the same stochastic process (e.g. news arrival) then their simultaneous inclusion will cause bias in the estimators (the problem of 'errors in variables'). The usual approach to this problem is to find a proxy for one of the variables. However, they find the use of v_{t-1} as a proxy for v_t is not successful.

Volume and News

Mitchell and Mulherin (1994) study the cross-sectional relationship between a proxy for news and trading volume and absolute price change. Their proxy was the number of items per day appearing on the Dow Jones newswire service and in the Wall Street Journal, following the earlier work of Thompson *et al.* (1987). Not surprisingly, they find quarterly seasonality relating to financial reporting, and intra-weekly seasonality with a build up of news through the week to Thursday, and then a fall-off on Friday. The correlation between news and trading volume is strong (0.37), but that between news and absolute returns much less so (0.06 for the index, and 0.11 for summed absolute returns of individual stocks). Mitchell and Mulherin also examine the importance of news, by counting the number of category codes assigned to each story (importance here being measured by width of impact). They find little change in their results, and note that the correlation between the quantity of news and the importance of news is high (0.88). In a separate but related study, Berry and Howe (1994) find evidence of intra-day relation between trading volume and information.

Unfortunately, none of these papers documents the time-series properties of their news proxies. Recently, however, Moschetti (1996) has performed a Vector Autore-

gression (VAR) analysis of the relation between information and trading volume on a daily basis. Moschetti's first finding is that "informational aggregates have quite a long influence on the market" (Moschetti, 1996, p. 17). The lag lengths for the different types of information vary between two and four days, demonstrating a time-series element in information which is reflected in trading volumes.

Moschetti finds a symmetric causal (in the sense of Granger, 1969) relation between information and volume, i.e. there is evidence of trading volume 'causing' information. The main explanation for this is feedback from reports of the performance of the market the previous day. Considering the long lag length and causality together, Moschetti (1996) concludes "... the market does not entirely process all public information immediately ... This could provide evidence for theories according to which it takes a certain period to traders [sic] to elaborate strategies based on public information" (Moschetti, 1996, p. 19). This interesting and slightly heterodox conclusion (see, e.g. Patel and Wolfson, 1984) is in accordance with the model of news assimilation proposed in Chapter 3.

Implications for the TP Model

The tentative conclusion from these different strands is that the amount of news per day is not independently and identically distributed, but rather shows positive autocorrelation. However, there are alternative explanations which mean that it is unlikely that allowing for news will completely remove the volatility clustering in daily returns.

Bearing this point in mind, it is interesting to note that the properties of the TP model derived in Section 2.4 are quite general of the news-arrival process providing only that the quantity of news per day be independently and identically distributed, and the asymptotic results do not require even this. Proposition 2.2 in particular may be used to formulate a statistical test of the joint hypothesis that the TP inter-

day model is appropriate and its parameters constant, despite the possibility of a time-varying news arrival rate. This test will be performed as part of the empirical analysis of Chapter 5.

2.7 Summary and Conclusion

This Chapter has examined in detail the Tauchen and Pitts (1983) model, which relates price change and trading volume via the news arrival process. In its complete formulation this model can also incorporate the size of the market to introduce functional dependencies between the moments of the price change and trading volume distributions. However, problems with the requirement for parameter constancy and with the implications of the model for updating beliefs (reservation prices) make this complete formulation untenable. This leaves the pair of relations eq. (2.29) and eq. (2.30) as a possible descriptive model of the price/volume relationship.

To estimate this pair of relations requires the specification of a process for news arrival. There is strong but not conclusive evidence (e.g. ARCH effects in returns) that this process is not independently and identically distributed through time but displays positive autocorrelation, at least when mirrored in the time-series behaviour of price changes and trading volume. This makes the descriptive model extremely hard to estimate. However, Proposition 2.2 shows that the descriptive model can be tested without being estimated, even in the presence of an autocorrelated news arrival process. This test will be performed in Chapter 5. A clear failure to reject the model would suggest that the news process is indeed the dominant influence on the price/volume relationship. A more equivocal result would indicate that other factors, including noise trading and other aspects of market-microstructure, might have an important role to play.

Chapter 3

What Is News?

This chapter essays a short formal definition of news and its related concepts. The need for this definition is an imprecise understanding of exactly what news is. The objective is to tie in, so far as it is possible, our operational understanding of news with a formal treatment of beliefs about the prevailing state of nature. The result is a model of a decision-making in financial markets (characterized by non-trivial information-processing costs) in which the quantity of news per day should be proxied by the number of transactions, rather than the number of news 'bits'.

Beliefs

Suppose that it is possible to conceive of the present and future as an exclusive and exhaustive collection of states of nature, S in total, indexed by $s = 1, \dots, S$. For simplicity suppose also that all investors agree on what these states are, but that they may disagree on how likely they are. The source of disagreement is assumed to be the information stock available to each investor (although it could equally well be each investor's ability to analyse a common information stock). Denoting the information stock of investor i at time t as Ω_{it} , the probability attached by investor i at time t to state s is $\Pr\{s \mid \Omega_{it}\}$. Finally suppose that all investors are rational,

in the sense that their beliefs, as described by these probabilities, are *coherent* (i.e. they satisfy the standard properties of probabilities and may be updated according to Bayes Theorem, see, e.g., Lindley, 1985).

According to this framework, an investor's beliefs at any time are entirely encapsulated by his vector of probabilities. This statement is not uncontentious, since many people would assert that probabilities themselves can be held with various degrees of belief: i.e. there exists a well-defined notion of beliefs about beliefs. So, for example, Fellner notes "A good many reasonable decision makers—though by no means all—seem to act differently depending on whether they act under the influence of shaky degrees of belief, i.e., of probabilities the numerical values of which are highly unstable in their minds, or act under the guidance of firm and stable degrees of belief." (Fellner, 1965, p. 4).

Bayesians assert that the notion of 'beliefs about beliefs' is misguided in the context of a single-decision problem. This may be illustrated by an example. Suppose I think that the probability of being run over as I cross the road is either 0.01 or 0.10, and in the former case I would cross the road, but in the latter I would regard this as too risky. The question is, under what conditions would it be correct for me to believe that I can distinguish these two cases, and that this distinction was meaningful in my attempt to cross the road.

The obvious answer to the first part is that the distinction is the result of some process which I only partially understand. So, for example, there might be a set of traffic lights further up the road which are controlling the flow of traffic, and thus the probability of me being knocked over. I cannot see the traffic lights, hence both outcomes are possible: I need to attach probabilities to each outcome, red or green. Say the probabilities are 0.5 each for red and green. But then I find directly that the probability of being run over is not either 0.01 or 0.10, but is in fact 0.055. Therefore although I can distinguish between the two cases in this context, the distinction is

not meaningful in my decision to cross the road, except insofar as it assists me in determining the probability of being run over.

This is clearly a rejection of the separation of risk and uncertainty (Knight, 1920). According to Knight 'risk' is a probability vector, while 'uncertainty' is the degree of belief attached to the probability vector. Bayesians reject this distinction, using probability to represent subjective uncertainty, however it arises (see, e.g., Cyert and DeGroot, 1987). There is a case, however, when the complete structure of 'beliefs about beliefs' should be preserved rather than be collapsed into a single probability vector: learning. Each probability vector represents an alternative scenario, and by repeated experience I can improve my understanding of the probability attached to each scenario. Thus if I was only ever to cross the road once, I would not need anything more than the 'collapsed' probability vector in order to make my decision. However, if I am to cross the road repeatedly I can use my experience, up until the moment that I am terminally run over, to adapt the probabilities attached to the traffic lights being red or green, which represent alternative scenarios. Consequently the complete belief structure is necessary for learning, but at any decision point only the collapsed probability vector is required.

A Definition of News

Having clarified the sufficiency of a single probability vector for representing beliefs, I turn now to the definition of news: in particular, under what conditions is information bit ω 'news' to investor i at time t , presuming that $\omega \notin \Omega_{it}$?

Definition 3.1 *Information bit ω is news to investor i at time t if and only if $\Pr\{s \mid \omega, \Omega_{it}\} \neq \Pr\{s \mid \Omega_{it}\}$ for some s .*

To take the negation of this definition, ω is not news to investor i at time t if it fails to alter his beliefs. The following proposition is a direct consequence of the coherence of each investor's beliefs.

Proposition 3.1 *Information bit ω is not news to investor i at time t if and only if it has an equal probability of occurring in all states of nature.*

Proof: The proposition follows directly from Bayes Theorem, which in this context states that

$$\Pr\{s \mid \omega, \Omega_{it}\} = \frac{\Pr\{\omega \mid s, \Omega_{it}\} \Pr\{s \mid \Omega_{it}\}}{\Pr\{\omega \mid \Omega_{it}\}} \quad (3.1)$$

For $\Pr\{s \mid \omega, \Omega_{it}\} = \Pr\{s \mid \Omega_{it}\}$ to hold across all s , it is necessary and sufficient that $\Pr\{\omega \mid s, \Omega_{it}\} = \Pr\{\omega \mid \Omega_{it}\}$ for all s , i.e. ω has an equal probability of occurring in any state of nature. ■

This proposition gives a simple way of determining whether or not any information bit is news: we simply ask “Was this information equally likely to have arisen in any state of nature?” If the answer is ‘Yes’, then the information bit is not news and as such is irrelevant to the decision process since it will not cause an optimal plan to be altered. Note that the arrival of a piece of news is a necessary but not sufficient for the alteration of such a plan.

It is interesting to digress for a moment on the relationship between the number of states of nature on the one hand and the definition of news on the other. In a world of unlimited information processing resources, in which it is possible to envisage a huge number of subtly-differentiated states of nature, almost all information bits will be news. Somewhere in the range of possible states of nature there will be found one (or more than one) which impinges upon the information bit we observe. Another way of expressing this is to say that, were we capable of the analysis, we would be able to inter-relate everything.

There are two ways of escaping from this conundrum. One is to assert that such a level of sophistication will always be beyond us. In this case, the limited number of states of nature that we can envisage force us to make a distinction between relevant

and irrelevant information, and this engenders a distinction between information and news. The other, which I favour, is to switch attention away from news per se, to consider the magnitude of news.

Magnitude of News

Having defined news in such a way that we can say whether or not a bit of information is news to an agent at a certain time, I turn now to the question of how 'big' is the news? It is important at this point to distinguish between the news and its impact. The magnitude of news per se is not the same thing as the impact the news might have on an optimal plan, although the two might be closely related. In fact, it is apparent from some simple examples that we tend to consider 'big' news not from our own needs but according to some other criterion. So, for example, the sinking of the Titanic was a big piece of news, but its impact on many people, in terms of the alterations they made in their optimal plans, was negligible. Likewise the Moon Landing, and perhaps the fall of the Berlin Wall. In these cases it is clear that the probabilities on various states of nature have changed dramatically, but that our optimal plan remains unchanged. This must be because the pay-offs in the states of nature in which the probabilities have changed are roughly the same.

Therefore, it is in keeping with the popular use of the term 'big' as applied to news to relate the magnitude of news to the amount by which the vector of probabilities which describe beliefs has altered. The metric I favour is the sum of the number of states of nature in which the probabilities change.

Definition 3.2 (Magnitude of News) *The magnitude of the news generated by information bit ω for investor i at time t is*

$$m_{it}(\omega) \stackrel{\text{def}}{=} \text{number } \{s : \Pr\{s \mid \omega, \Omega_{it}\} \neq \Pr\{s \mid \Omega_{it}\}\}. \quad (3.2)$$

As in Proposition 3.1, this may be interpreted, using Bayes Theorem, as the number of states of nature in which the probability of observing ω is different from the unconditional probability of observing ω . By this metric, the magnitude of a piece of news is equated with its breadth—news of large magnitude impacts upon the probabilities of a large number of states of nature.

Alternative definitions of magnitude might use the sum of the absolute values of the probability differences or the log-ratios. They are not directly related to breadth. Furthermore, these alternatives result in an ordinal measure with little intuitive appeal. A magnitude such as $m_{it}(\omega) = 3$ has a clear interpretation if the number of states of nature, S , is known. Were the same number to represent the sum of the absolute differences, or the sum of the absolute log-ratios, the meaning is far from clear unless in comparison with another. But the conclusive reason for favouring $m_{it}(\omega)$ is that it does not require explicit updating of the probabilities. Using the Bayesian formulation, it is necessary only to know whether or not each state has an impact upon the probability of the news bit—the size of the impact (i.e. the updating of the probability of the state) is not needed to determine magnitude. This will be an important consideration below, in the discussion of the *quantity* of news per period.

In the case discussed in the previous subsection, where S is very large and almost all information bits constitute news, the magnitude of news by this definition can be used to establish a hierarchy of information sources. In this way our failure to distinguish news from information is unimportant, because we can instead distinguish large-magnitude news from small-magnitude news.

Quantity of News

So far, I have formalized two familiar concepts—the notion of news itself and its magnitude. I now turn to a more difficult concept: the quantity of news that arrives

during a given period. As seen in the previous chapter, this is an important quantity in the formulation of price/volume models in speculative markets.

In one sense the quantity of news is a simple concept. Consider, for example, a typical day in a speculative market, where stories are posted as they break, on a Reuters terminal. It should be simple enough to distinguish those information bits which are news (e.g., ignore the cricket scores, usually), and then add them up over the day to find the quantity of news in that day. This is in fact the standard procedure for empirical studies regarding news (see Section 2.6). Unfortunately, however, this method tends to double-count small news bits whenever there is a cost associated with analysing the news for its implications for the probabilities, or in implementing the updated probabilities in determining the optimal plan.

Consider an investor who has just received some information. He must choose between the costly process of updating his beliefs and his optimal plan, or saving the information until the arrival of the next piece of information and then making the same choice again. This is an optimal stopping problem, and these are, in general, quite hard to solve.¹ To make the analysis simpler, consider the case where institutional requirements will force the probabilities to be updated after two pieces of information have arrived, but the investor has the option of updating the probabilities after the first piece as well, should he so choose. In this case he has a straight trade-off between the cost of updating, which is known, and the sub-optimality of the plan up until the time at which the second piece of information arrives, which is uncertain. The expected sub-optimality given his current information set will be increasing in the magnitude of the information which has just arrived (which, for the purposes of this example can be ascertained at zero cost), in which case his decision will be determined according to some critical magnitude.

Generalizing this argument to the optimal stopping problem, the investor will

¹A similar type of problem is the concern of Chapters 8 and 9.

update his probabilities when the magnitude of the set of information which he has saved (but not analysed beyond magnitude) since his last update exceeds some critical value. This critical value will, in general, depend upon his information set at his last update. The time at which the investor updates his probabilities will be referred to as an 'update point'. A simple tally of the number of information arrivals between update points cannot account for the way in which earlier information can be completely or partially negated by later information—hence the possibility of double-counting. For this reason a much better measure of the quantity of news when several information bits are considered together is the joint magnitude. This gives rise to the following definition.

Definition 3.3 (Quantity of News per Investor) *The quantity of news between update points $t - 1$ and t , during which period the k information bits $\omega_1, \dots, \omega_k$ are made known to investor i is*

$$q_{it}(\omega_1, \dots, \omega_k) \stackrel{\text{def}}{=} m_{i,t-1} \left(\bigcup_{j=1}^k \omega_j \right). \quad (3.3)$$

The effect of this definition is to highlight the importance of the update points in determining the quantity of news. The reason that these points do not occur concurrently with the arrival of news is the cost of the updating process, and the availability of magnitude as a cheap indicator of the impact of news on probabilities and the optimal plan.

News in Speculative Markets

Up to this point the discussion has been entirely subjective. Different investors will treat the same information in different ways, depending upon their current information stock, the costs they face in updating their probabilities and their optimal plans, and the trade-off they perceive between these costs and the loss of expected utility

from operating with a sub-optimal plan. There are obvious problems in aggregating over these subjective and unobservable issues to derive a quantity of news which is applicable to the whole market. There is, however, one measure which is observable and which fulfills precisely the requirements of models such as that of Tauchen and Pitts (1983).

Definition 3.4 (Quantity of news per day) *The quantity of news per day in a speculative market is the number of investors' update points which occur during the day.*

The first advantage of this definition is that it related to a process which is at least partly measurable. For example, in a speculative market the number of calls to stockbrokers would provide a useful lower bound on updates, since an update is a necessary (but not sufficient) condition for considering the purchase or sale of an asset. Second, in the extreme case where there are no costs in updating probabilities it collapses to the number of news bits, which is the traditional definition. Third, and most important, it generalizes to the case where there are updating costs, and explains the way in which these costs can affect the appearance of news in financial aggregates like price changes and trading volume—a rise in costs would cause a fall in the amount of news. Fourth, it relates the quantity of news per day to both the flow of information to the market and the magnitude of that information. Taking the critical magnitude at which investors update as fixed, this threshold will tend to be breached sooner if either the rate of information flow rises or the mean magnitude of an information bit rises. This will lead to more frequent updates and a higher density of update points within a given period, i.e. a greater quantity of news.

There is one implication of this definition that needs further consideration, which is that it relates the amount of news to the number of investors in the market. In some respects this is a good thing, since increases in trading volume in newly

established financial markets are often related to the entry of new investors. The empirical analysis of Chapter 5 gives a good example of this, since the data covers a market from inception. The trading volume shows an initial period of strong growth, which then flattens out gradually. This would be an implication of the Tauchen and Pitts (1983) model for trading volume, eq. (2.30), with Definition 3.4. However, the variance of price changes is stable over the same period, which contradicts the model for price change, eq. (2.29), if the same definition of news is used. Therefore the appropriate measure in markets where the number of investors is changing is more likely to be the mean amount of news per investor, with separate consideration given to the way in which market size can affect the price/volume relationship.²

Summary

I have attempted here to provide formal definitions of 'news', 'news magnitude' and 'news quantity'. Using these definitions I have provided a descriptive model of decision-making, in which it is quite rational to accumulate information and re-optimize sporadically in the presence of information-processing costs. This model has been used to suggest that a measure such as the number of transactions per day is a better proxy for the quantity of news than the more typical proxy—the number of stories per day. In the special case of zero costs these two will be the same. But in the general case where it is expensive to update beliefs, re-optimize and implement a new plan, the accumulation of information causes a proxy related to the number of stories to over-estimate the propensity for investors to re-optimize, since during accumulation later news bits can negate earlier ones.

²As discussed in Chapter 2, Tauchen and Pitts explicitly consider the case where the number of investors can vary, but their model has some uncomfortable assumptions (Section 2.5).

Chapter 4

An Optimal Price Index

This chapter defines the notion of 'optimality' for creating price indices out of futures contract prices. As will be discussed below, futures price data have several advantages over spot price data, and one glaring disadvantage—every futures contract expires. Both Clark (1973) and Tauchen and Pitts (1983) used futures data to fit their models, using the method proposed by Clark to join the individual contract prices together. The optimal index proposed here is shown to be superior to that of Clark both in theory and in practice. This chapter has appeared in a slightly different form as Rougier (1996).¹

4.1 Introduction

One of the features of the Tauchen and Pitts (1983) model analysed in Chapter 2 is the symmetry of costs with respect to short and long positions; in fact, traders in this model have no transactions costs at all. In practice, it is more expensive to go short than to go long in spot markets, since to go short often involves either

¹I would like to thank the Editor of the Journal of Futures Markets, Professor Mark Powers, and the two anonymous referees for their very helpful comments on an earlier version of this paper.

having to borrow stock or having to purchase the right to borrow stock.² In futures markets costs are symmetric with respect to short and long positions.

Karpoff (1987) notes that the price-volume relationship appears to be affected by the asymmetry of costs between short and long positions. In spot markets the price-volume relationship for negative price changes is not reflectionally symmetric with that of positive changes, while in futures markets the relationship is reflectionally symmetric. Consequently, spot markets display a significant correlation between price change and volume while futures markets do not. Both markets display a correlation between absolute price change and volume. Therefore empirical work on price-volume models without transactions costs should use futures rather than spot data.

Unfortunately, however, every financial futures contract expires. Therefore time series analysis of futures markets will often be limited by the short period in which any particular contract exists. For stock index futures this is a period of about half a year.³ An obvious alternative to using a single contract is to use a price index constructed from one or more of the outstanding contracts at each point in time. There are a number of ways in which such an index may be constructed.

This chapter makes a theoretical case for a new type of index, known as an 'optimal index', and demonstrates that such an index outperforms the two currently-accepted indices. Section 4.2 defines optimality, and shows how optimal indices may be derived for different numbers of available futures contracts. Interestingly, where there are more than two contracts available at any one time, there is more than one optimal index. Section 4.2 also contrasts the optimal index with the widely-practised method of simply taking the near contract and ignoring the others, here

²For the situation in the UK, see *Stock Borrowing and Lending*, available from the London Stock Exchange, and 'Some Old Peculiar Practices in the City of London' (*The Economist*, February 18, 1995).

³Sometimes even this period will be inappropriately long, given the finding of Yadav and Pope (1990) that mispricing in stock index futures tends to increase with time to expiry.

referred to as ‘splicing’.

Section 4.3 discusses the other alternative, the ‘Clark index’ first described in Clark (1973) and used subsequently by Tauchen and Pitts (1983). This was originally proposed as an improvement on splicing, and it is shown in this section that the Clark index represents a compromise in performance between splicing and optimality. Section 4.4 examines the practical differences between the three indices using data on the FTSE 100 stock index futures contract traded on the London International Financial Futures Exchange (LIFFE). To anticipate the conclusion, the optimal index as proposed in this chapter should be unambiguously preferred to either the spliced index or Clark’s index. The former is demonstrably sub-optimal, while the latter is much harder to calculate and potentially unreliable.

4.2 ‘Optimality’ in Futures Price Indices

The objective is to create an optimal price index from the prices of the available futures contracts. This begs the question “What do we mean by optimal?”

Definition 4.1 (Optimal Futures Price Index) *An optimal futures price index preserves a constant proportionality between the change in the price index and the change in the underlying spot price, in the absence of market imperfections.*

In the absence of market imperfections, each futures contract trades at a ‘fair value’ relative to the spot price. This fair value is maintained by arbitrage. In the case of stock index futures,

$$F_k = P(1 + (r - y)(k - t)), \quad (4.1)$$

where F_k is the price at t of the futures contract expiring at time k (where $t \leq k$), P the spot price at t , r the interest rate and y the dividend yield. All periods are

measured in years, and capitals are used to denote price variables assumed to be time-varying.⁴

The simplest index of futures contract prices would be the price of the near contract, F_k . This is the generally-accepted method of creating a contiguous series.⁵ Notionally, the price series of the near contracts are spliced together so that the daily return of the resulting price index is equal to the daily return of the near contract. The problem with the spliced index is that the change in the index depends upon both the change in the spot price and in the time to expiry of the near contract:

$$dF_k = (1 + (r - y)(k - t)) dP - P(r - y) dt. \quad (4.2)$$

This expression shows that the ratio dF_k/dP is seasonal, regardless of whether dP is seasonal. This seasonality is introduced by the declining interval between t , the current time, and k , the fixed point in the future at which the contract expires. It is this problem of seasonality which motivates the definition of optimality given above. An optimal futures price index should have no seasonal variations other than those in the spot price.

The optimal futures price index, denoted F^* , will be restricted to convex combinations of the prices of the currently available futures contracts:

$$F^* = \sum_{i=0}^{n-1} \lambda_i(t) F_{k+iv}, \quad (4.3)$$

where there are n contracts available and the time between contract maturities is v (i.e. $k - v \leq t \leq k$). The weights in eq. (4.3) are written explicitly as functions

⁴It is assumed that the 'cost of carry' $r - y$ is constant over the period for which there are futures contracts available. For most of the time this is a good approximation. Occasionally, however, there will be some marked seasonality (e.g. the interest rate will jump upwards at the end of each quarter if the yields curve is upward sloping). The empirical results of Section 4.4 demonstrate that the impact of this factor, and likewise the impact of transactions costs, is small.

⁵See, e.g., Buckle *et al.* (1994).

of time, since without time-variation in the weights there will always be seasonality similar to that of the spliced index. The weights satisfy the usual conditions

$$\lambda_i(t) \in [0, 1] \quad i = 0, \dots, n-1 \quad (4.4)$$

$$\sum_{i=0}^{n-1} \lambda_i(t) = 1 \quad (4.5)$$

$$\sum_{i=0}^{n-1} \lambda'_i(t) = 0 \quad (4.6)$$

for all t , where $\lambda'_i(t)$ denotes the derivative (the third of these conditions is actually a direct consequence of the second). The spliced index is the simple case $\lambda_0(t) = 1$, and $\lambda_i(t) = 0$ for $i = 1, \dots, n-1$.

An optimal index by Definition 4.1 will have weights which satisfy the following Proposition.

Proposition 4.1 *For F^* in eq. (4.3) to be an optimal futures price index by Definition 4.1 it is sufficient that the weights satisfy eq. (4.4), eq. (4.5) and eq. (4.6) and the additional condition*

$$v \sum_{i=0}^{n-1} i \lambda'_i(t) = 1, \quad (4.7)$$

for all t .

Proof: The definition of optimality requires that

$$\frac{dF^*}{dP} = c \quad (4.8)$$

for all t , where c is some constant yet to be determined.⁶ Condition eq. (4.8) holds if and only if

$$\frac{d}{dt} \left(\frac{dF^*}{dP} \right) = 0. \quad (4.9)$$

From eq. (4.1) and eq. (4.3),

$$\begin{aligned} \frac{dF^*}{dP} &= \sum_{i=0}^{n-1} \lambda_i(t) \{1 + (r - y)(k + iv - t)\} \\ &= 1 + (r - y)(k - t) + v(r - y) \sum_{i=0}^{n-1} i \lambda_i(t) \end{aligned} \quad (4.10)$$

remembering that the weights sum to one, eq. (4.5). Differentiating with respect to t and setting the result to zero gives

$$\begin{aligned} -(r - y) + v(r - y) \sum_{i=0}^{n-1} i \lambda'_i(t) &= 0 \\ \Rightarrow v \sum_{i=0}^{n-1} i \lambda'_i(t) &= 1 \end{aligned} \quad (4.11)$$

which is condition eq. (4.7). ■

It can easily be confirmed that the spliced index is not optimal, since eq. (4.7) cannot be satisfied when $\lambda'_i(t) = 0$ for $i = 1, \dots, n-1$. Trivially, no index comprising just one futures price can be optimal. Proposition 4.1 can now be used to derive the optimal weights for different numbers of contracts.

⁶This condition may also be written

$$\frac{dF^*}{dt} = c \frac{dP}{dt}.$$

In other words, the rate of change of the futures price index and the spot price should be in constant proportion no matter what the relation between the current date and the expiry pattern of the futures contracts.

Corollary 4.2 For the case of only two futures contracts, the optimal weights are

$$\lambda_0(t) = \frac{k-t}{v} \quad \lambda_1(t) = \frac{v-(k-t)}{v}. \quad (4.12)$$

Proof: Since $n = 2$, eq. (4.7) implies that $\lambda_1'(t) = 1/v$, from which eq. (4.6) implies $\lambda_0'(t) = -1/v$. Integrating these two expressions with respect to t gives

$$\lambda_0(t) = \frac{-t}{v} + C_0 \quad \lambda_1(t) = \frac{t}{v} + C_1. \quad (4.13)$$

Each weight is constrained to lie in the range $[0, 1]$ over the admissible values of t , which are $t \in [k-v, k]$. Therefore the constants must satisfy certain inequalities.

Taking C_0 at the two extreme values of t ,

$$\left. \begin{array}{l} t = k - v : \frac{-(k-v)}{v} + C_0 \leq 1 \\ t = k : \frac{-k}{v} + C_0 \geq 0 \end{array} \right\} \implies C_0 = \frac{k}{v}. \quad (4.14)$$

To ensure that $\lambda_0(t) + \lambda_1(t) = 1$ for all t , the constants must sum to one, and this gives

$$C_1 = 1 - \frac{k}{v}. \quad (4.15)$$

Substituting these two values into eq. (4.13) completes the proof. ■

These two weights when substituted into eq. (4.3) give

$$F^* = P(1 + (r - y)v). \quad (4.16)$$

The optimal price index behaves like a notional forward contract which has v years until expiry. There is only one optimal price index using just two contracts.

More Than Two Contracts

For stock index futures there are often only two contracts available, and so optimal indices with $n > 2$ are not required. However, there are other futures markets where there are many more than two contracts available (e.g. the market for cotton futures examined by Clark (1973), in which there were generally eight contracts available). Therefore the optimal indices for three contracts will be derived, in order to demonstrate the approach and illustrate the properties of the resulting indices.

Generally, solutions can be found by using the sufficient condition for eq. (4.7),

$$\lambda'_i(t) = \frac{1}{iv(n-1)} \text{ for } i = 1, \dots, n-1, \quad (4.17)$$

finding $\lambda'_0(t)$ by eq. (4.6), and then working back to the weights in exactly the manner demonstrated in the proof to Corollary 4.2. For $n = 3$, there are three sets of weights found by this method:

$$\begin{aligned} \lambda_0 &= \frac{v+3(k-t)}{4v} & \lambda_1 &= \frac{2v-2(k-t)}{4v} & \lambda_2 &= \frac{v-(k-t)}{4v} \\ \lambda_0 &= \frac{3(k-t)}{4v} & \lambda_1 &= \frac{3v-2(k-t)}{4v} & \lambda_2 &= \frac{v-(k-t)}{4v} \\ \lambda_0 &= \frac{3(k-t)}{4v} & \lambda_1 &= \frac{2v-2(k-t)}{4v} & \lambda_2 &= \frac{2v-(k-t)}{4v} \end{aligned} \quad (4.18)$$

These correspond to notional forward contracts which have v , $(5/4)v$ and $(3/2)v$ years to expiry, respectively. Interestingly, the first set of weights can be generalized further since none of the individual weights ever fall to zero.⁷ The generalisation is

$$\lambda_0 = \frac{v+3(k-t-a)}{4v} \quad \lambda_1 = \frac{2v-2(k-t-a)}{4v} \quad \lambda_2 = \frac{v-(k-t-a)}{4v} \quad (4.19)$$

where $0 \leq a \leq v/3$. The notional time to expiry of this contract is $v+a$. In other

⁷I am grateful to an anonymous referee of the Journal of Futures Markets for pointing this out to me.

words, the investigator can choose any notional time to expiry between v and $(4/3)v$. This includes $(5/4)v$, the second of the indices in eq. (4.18), which would require $a = v/4$. Therefore two different sets of weights can give the same notional time to expiry. Investigators should be aware of this possibility when creating indices with three or more contracts.

When there is more than one optimal index available, a general rule would be to choose that set of weights which gives a notional time to expiry closest to the mean time to expiry of all contracts in the market over the period in question.

4.3 Clark's Price Index

The deficiency of the splicing method, which takes just the near contract, has already been highlighted. Clark (1973) noted this, and proposed an alternative solution which shares many of the properties of the optimal price indices described above. Clark's method is asymptotically optimal in the number of available contracts, but not so attractive where there are only a small number of contracts available. As has been mentioned above, this is the case with stock index futures where two contracts is usual. In Clark's defence, however, his method was applied to cotton futures, where the number of contracts available at any time was eight.

Clark's index is a convex combination of futures contract prices, as in eq. (4.3) and eq. (4.4). However, the weights are not derived from the theoretical properties of futures contracts. Rather, they are derived from the empirical distribution of the time until expiry across all contracts in the sample period. This distribution is denoted $W(\cdot)$, such that $W(d)$ shows the proportion of all contracts in the sample period which have d time in years until expiry. Clark's index is

$$F^c \stackrel{\text{def}}{=} \sum_{i=0}^{n-1} w_i(t) F_{k+iv}, \quad (4.20)$$

where

$$w_i(t) \stackrel{\text{def}}{=} \frac{W(k-t+iv)}{\sum_{i=0}^{n-1} W(k-t+iv)}. \quad (4.21)$$

Thus each contract is weighted according to the popularity (i.e. frequency of occurrence) of its particular time to expiry. Clearly the weights in eq. (4.20) do change with time and sum to 1. However, they do not necessarily satisfy eq. (4.7).

If Clark's index is not optimal, it will contain some seasonality in addition to that in the underlying spot price. The source of this seasonality is variation in the *de facto* time to expiry. This is $\mathcal{E}[W]$ on average, but varies on a cycle of v years because the contracts available at any given time provide just a small subset of the total range of possible times to expiry. In fact, the *de facto* time to expiry of the Clark index at time t is

$$\begin{aligned} \bar{w}_t &\stackrel{\text{def}}{=} \sum_{i=0}^{n-1} w_i(t) (k-t+iv) \\ &= k-t+v \sum_{i=0}^{n-1} i w_i(t) \end{aligned} \quad (4.22)$$

and the Clark index might be written

$$F^c = P(1 + (r - y) \bar{w}_t). \quad (4.23)$$

From eq. (4.21) and eq. (4.22) it can be seen that \bar{w}_t tends to $\mathcal{E}[W]$, a constant, as n becomes large and $v \rightarrow \Delta t$ (which also implies $k \rightarrow \Delta t$), where Δt is the resolution of W , typically 1 day. This confirms the point made above that Clark's index is less sub-optimal where there are a large number of different times to expiry available.⁸

⁸The process by which the *de facto* time to expiry at t tends to the expectation of W as the number of available contracts (n) becomes large is exactly the same as that in which a numerical integration approaches the true value in the limit as the number of points over which the integrand is evaluated becomes large. The smoothness of W is a crucial factor in deciding how quickly the limit is approached. As will be seen below, for FTSE futures W is very smooth and so the

However, even where there are a large number of contracts available Clark's index still has disadvantages over the optimal index. First, it is much harder to calculate, requiring an extra dataset (the daily open interest on each contract) and the estimation of the distribution W . Second, it is sample-dependent since W will depend upon the period of calculation. This sample-dependence would probably not be a problem in a mature market since W would be stable. In a new or evolving market, the 'seasoning' of the contract could make W vary, in which case the Clark index calculated over a period might be substantially different from its values in a sub-period.

4.4 Practical differences

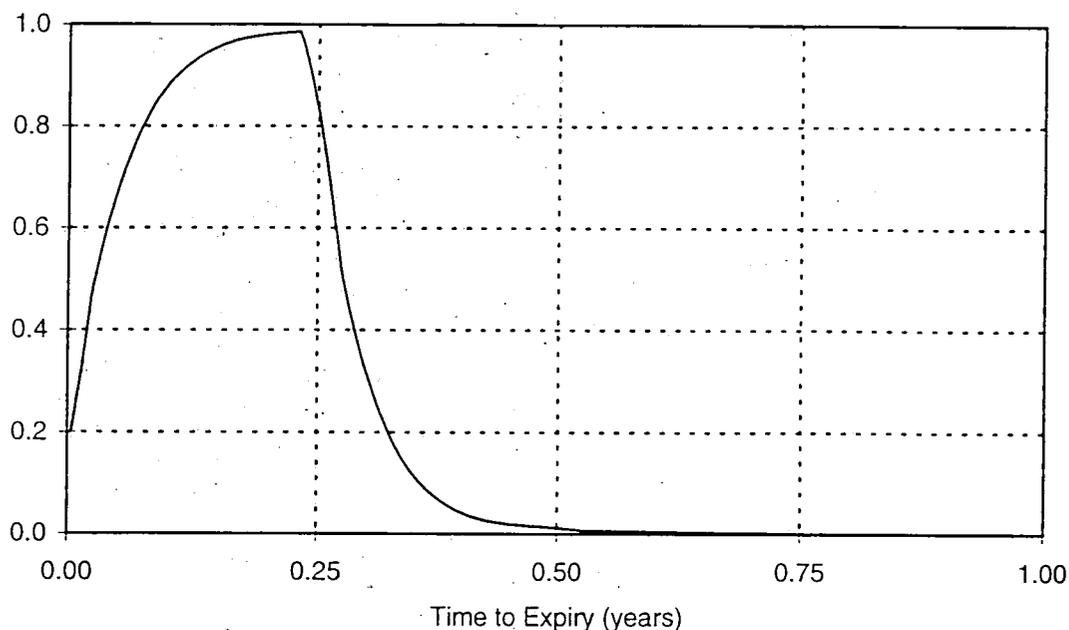
The notion of optimality depends upon there being no market imperfections, since the fair value relationship between spot and futures prices, eq. (4.1), is maintained by arbitrage. Since there are imperfections in actual markets, most notably transactions costs, but also sources of delay which prevent the concurrent execution of spot and futures transactions, it is important to examine whether there are empirical as well as theoretical differences between the three indices. There is also the issue of ease of use to be considered. This section examines these questions using the FTSE 100 contracts traded on LIFFE over the period 1985–1994.

The optimal index is the two-contract specification eq. (4.12), since there are sometimes no more than two futures contracts available.⁹ The Clark index requires the estimation of the time to expiry distribution, $W(\cdot)$, prior to calculating the index values. This is shown in Figure 4.1. The sharp fall-off in $W(t)$ just before $t=0.25$

convergence is fast, and few contracts are required in order for the Clark index to behave as if it has a fixed amount of time until expiry.

⁹Indeed, sometimes at the start of a new quarter there is only one contract available. In this case the weights are set to $\lambda_0(t) = 1$ and $\lambda_1(t) = 0$; these are close to the theoretical values since $(k - t) \approx v$.

Figure 4.1: The Distribution of FTSE Contracts by Time to Expiry



indicates that most of the interest in FTSE 100 futures is in the near contract. The *de facto* times to expiry for the three indices are shown in Figure 4.2. The mean time to expiry for the Clark index is 0.169 years (about 43 days), but the availability of just two contracts causes the *de facto* time to expiry to vary from 0.253 years (64 days) to 0.123 years (31 days) over the course of one quarter.

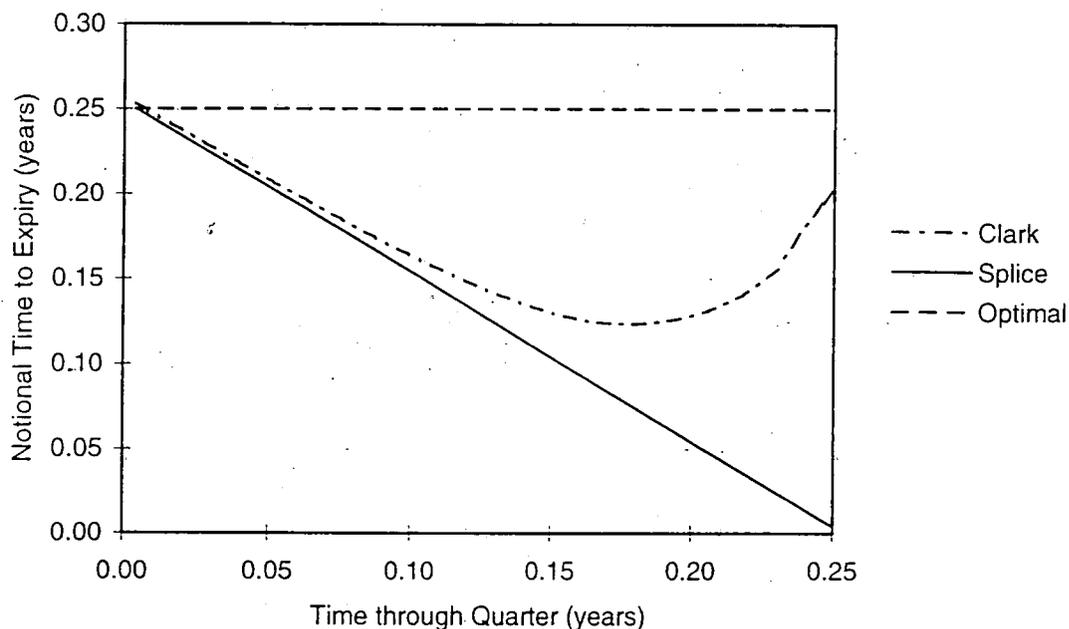
From the properties of the three indices

$$\frac{dF_k}{dt} = -(r - y)P + \{1 + (r - y)(k - t)\} \frac{dP}{dt}, \quad (4.24)$$

$$\frac{dF^*}{dt} = \{1 + (r - y)v\} \frac{dP}{dt}, \quad (4.25)$$

$$\frac{dF^c}{dt} = (r - y)P \frac{d\bar{w}_t}{dt} + \{1 + (r - y)w_t\} \frac{dP}{dt}. \quad (4.26)$$

Figure 4.2: The Notional Time to Expiry During a Quarter



This suggests a bivariate regression of the form

$$\Delta F = \alpha + \beta \Delta P + u_t \quad (4.27)$$

where the daily change in each of the indices, denoted by Δ , is used to approximate the time derivatives, and u_t is some disturbance term.¹⁰ Regression eq. (4.27) should only be correctly specified for the optimal index F^* , in which case $\alpha = 0$ and $\beta > 1$, and the disturbance should be white noise. In the case of the spliced index significant time-variation in β should cause the regression to be mis-specified. It is hard to say, a priori, whether the variation in \hat{w}_t over the course of a quarter is sufficient to have a significant impact on the specification of the Clark index regression.

On running the regressions, it was immediately apparent that all three models

¹⁰The change in the spot price, ΔP , is calculated from close to close of trading in the futures contracts at LIFFE (4.10 p.m.) rather than close to close in the spot market, since the two are not synchronous.

Table 4.1: Regression results, 1985–1994

	Dependent variable		
	ΔF_k	ΔF^*	ΔF^c
$\hat{\alpha}$	-0.307* (0.112)	-0.041 (0.131)	-0.039 (0.131)
$\hat{\beta}$	0.762 (0.006)	1.072+ (0.008)	1.069+ (0.008)
$\hat{\rho}$	-0.285* (0.019)	-0.379* (0.018)	-0.375* (0.018)
R ²	0.839	0.867	0.866
DW	2.102	2.174	2.172

The regression is $\Delta F = \alpha + \beta \Delta P + e_t$, where $e_t = \rho e_{t-1} + u_t$, and u_t is a white noise disturbance term.

* significantly less than 0.0 at 5 percent; + significantly greater than 1.0 at 5 percent. The critical values for the Durbin-Watson (DW) statistic are about 1.77, 2.23 at 5%. The hypothesis that the three parameters are the same in the optimal and Clark regressions has a test statistic of 1.30, which is comfortably below the critical value (χ^2_3 at 5 percent) of 7.82.

were characterised by strong negative first order autocorrelation in the residuals. This may be presumed to have arisen as a consequence of the arbitrage process. If a futures contract price was sufficiently in excess of its fair value, then arbitrage would cause $\Delta F_k < 0$ and $\Delta P > 0$ in the absence of news, and the disturbance in eq. (4.27) would be negative. If arbitrage next occurred in the other direction the disturbance would be positive. Therefore, negative autocorrelation in the disturbances of eq. (4.27) suggests some form of 'overshooting' by arbitrageurs.

The regressions were re-estimated allowing for first-order autocorrelation in the disturbances. The results of these regressions are given in Table 4.1. These show that the spliced index performs very differently from the optimal index and the Clark index. The optimal regression has precisely the properties predicted by the theory: a zero intercept and a gradient slightly but significantly in excess of 1.0. These properties are shared by the Clark index where, additionally, the $\hat{\beta}$ is slightly

less than that of the optimal regression in accordance with the shorter mean time to expiry. This contrasts with the spliced index regression, where the gradient is not significantly greater than 1.0, contradicting the theory. This suggests that the spliced index regression is mis-specified.

The high R^2 values for the optimal and Clark regressions indicate that nearly 90 percent of the variation in the futures price can be explained by the fair value relationship. This leaves just over 10 percent to be explained by the sporadic actions of arbitrageurs, and slack introduced by transactions costs and variation in the interest rate and the dividend yield.

4.5 Conclusion

The need for a contiguous price index for futures contracts is well-established. This chapter has considered three such indices. The spliced index is the standard and consists simply of the returns of each near contract in turn, notionally joined together into a price series. The Clark index is a convex combination of futures contract prices, where the weights are derived from the empirical distribution of time to expiry. By the definition provided in this chapter neither of these two indices is optimal, since they both introduce seasonality in the futures price index in addition to that present in the spot price.

The third index is optimal by the same criterion. The optimal index preserves a constant proportionality between spot price change and futures price index change. It is shown mathematically that there are several optimal indices. The investigator should choose that index with the most appropriate notional time to expiry, since it is in this respect that the indices differ from one another.

The empirical evidence suggests that the spliced index should be avoided, but that Clark's index and the optimal index perform almost identically. This suggests

that market imperfections are sufficient to cloak the Clark index's sub-optimality. However, given the much greater cost of calculating the Clark index and the uncertainty engendered by its sample-dependence, the optimal index should be unambiguously preferred when creating a price index for stock index futures contracts. Similar optimal indices can also be calculated for other futures markets.

Chapter 5

A Short Empirical Study

5.1 Introduction

This Chapter investigates the price/volume relationship using 10 years of daily data from the FTSE-100 futures contracts traded on the London International Financial Futures Exchange (LIFFE). Section 5.2 describes the price data, and Section 5.3 the trading volume data, which needs some adjustment to account for the growth of the market and for non-news-related trading. The following sections examine two aspects of the price/volume relationship: symmetry following Karpoff (1987) (Section 5.4), and the χ_1^2 test for the Tauchen and Pitts (1983) inter-day model developed in Section 2.5 (Section 5.5). Section 5.6 concludes.

The original data used in this study were kindly supplied by LIFFE, and consisted of date, contract code, delivery month, index settlement price, index opening range (two values), daily high, daily low, volume and open interest, for each contract for each trading day from the middle of 1984, when the FTSE-100 future was first traded. The period under study covers the ten years 1985–1994. The original format of the data is shown in Figure 5.1.

The attraction of futures data for empirical work has been briefly discussed

Figure 5.1: Sample of Original Data from LIFFE

2	JAN	1985	X	MAR	85	1216.0	1229.5	1227.0	1229.5	1215.0	194	744
2	JAN	1985	X	JUN	85	1218.0	0	0	0	0	0	48
3	JAN	1985	X	MAR	85	1204.5	1201.0	1202.5	1208.0	1199.5	231	748
3	JAN	1985	X	JUN	85	1206.5	1204.0	1204.0	1204.0	1204.0	2	46
4	JAN	1985	X	MAR	85	1213.0	1206.0	1207.5	1217.5	1206.0	155	741
4	JAN	1985	X	JUN	85	1214.0	0	0	0	0	0	46
7	JAN	1985	X	MAR	85	1228.5	1211.0	1213.0	1229.0	1211.0	236	763
7	JAN	1985	X	JUN	85	1229.5	0	0	0	0	0	46
8	JAN	1985	X	MAR	85	1238.0	1237.5	1237.0	1242.0	1230.5	276	776
8	JAN	1985	X	JUN	85	1239.0	0	0	0	0	0	46
8	JAN	1985	X	SEP	85	1240.0	1240.0	1240.0	1240.0	1240.0	5	5

Thus, there are two contracts available on 2 Jan 1985, one expiring in March 1985 and the other in June 1985 (both at the end of the month). The settlement (i.e. closing) price of the first of these was 1216.0, the opening range 1229.5–1227.0, the high 1229.5, the low 1215.0, the volume 194 and the open interest 744; the data for the second contract is not updated since there was no trading on the day. When trading is very light, the data can be ‘stale’—this must be taken into account when creating price indices.

in the introduction to Chapter 4. In addition to the arguments advanced there, which centred on the low and symmetric transactions costs, an index is preferable to individual stock prices since the diversification in the index lowers the standard deviation of returns and makes it easier to perceive the time series structure of the mean returns (Lo and MacKinlay, 1988; Conrad and Kaul, 1988).

One potential problem with an index is asynchronous trading bias (Fisher, 1966; Cohen *et al.*, 1980). The presence of less-traded stocks in an index means that at the close of the trading day some of the prices will be ‘stale’, i.e. not updated with the latest news. These prices adjust on the following day (at the next trade) and so introduce positive autocorrelation into the return series. This is unlikely to be a problem with an index of highly traded stocks such as the FTSE-100.

Another problem with *spot price* indices is that they are never directly traded. In this case many of the psychological effects linked to the impact of technical analysis (e.g. support and resistance levels) are not present in the index in the same way as

in underlying stocks. However, a futures contract on the index *is* traded, and so it does not suffer from this problem.

5.2 Prices

The price series was constructed from the prices of the existing contracts on each trading day using the two-contract Optimal Price Index described in Chapter 4. The resulting index is displayed in Figure 5.2. The daily return series was created using log price differences:

$$r_t \stackrel{\text{def}}{=} \log p_t - \log p_{t-1}, \quad t = 1, \dots, 2529. \quad (5.1)$$

This series is shown in Figure 5.3. It is clear from these two Figures that the behaviour of prices during the year of the stock market 'crash', 1987, was atypical but not exceptionally so. It is a pleasing coincidence that the bounds of the year also define very closely the bounds of the 'bubble', the crash and the recovery. But it is easy to identify other one-year periods in which the behaviour of prices is equally atypical: 1988 was strikingly flat, 1989q3–1990q2 with its 'triple top', or 1993q3–1994q2 with its interesting 'single top'. There are certainly no grounds for excluding 1987 from the sample *at this stage*, although there might be a case in subsequent empirical work for masking the one day fall of 16.6% which occurred on October 19, 1987.

Descriptive statistics for the return series are given in Table 5.1. Although the trend in prices is clearly upwards, the mean daily return is less than one twentieth of 1%, compared to a standard deviation of over 1%.

Six out of the ten years show negative skewness but only two of these are significantly negative compared to a normal distribution; there are also two significantly positively-skewed years. From the quartiles it is clear that if there is negative skew-

Figure 5.2: Optimal Price Index (p_t)

FTSE-100 Futures, 1985-94

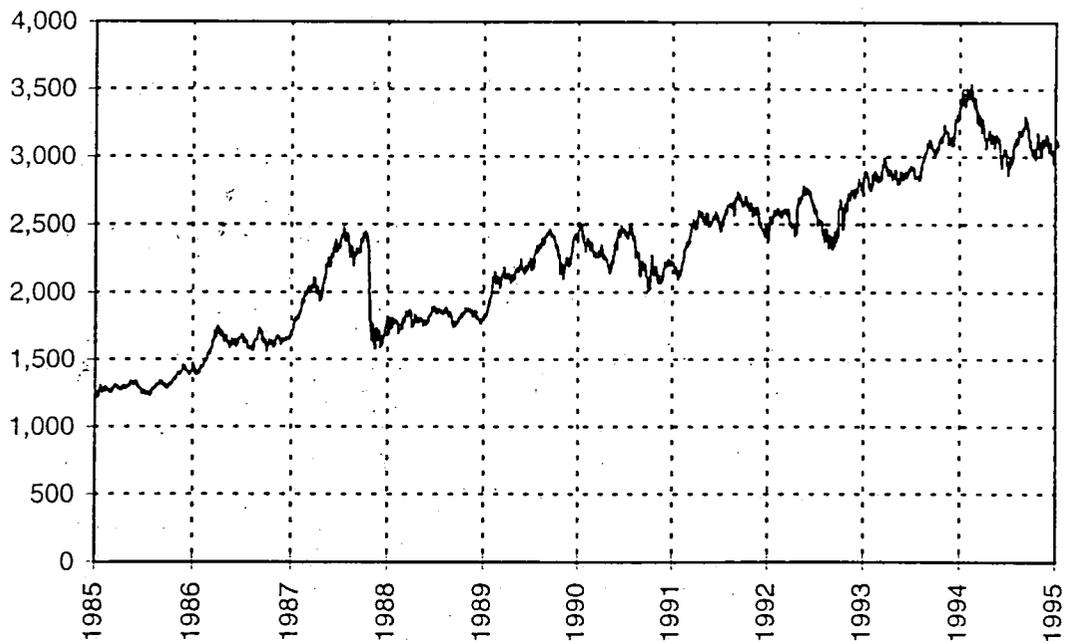


Figure 5.3: Daily Returns (r_t)

FTSE-100 Futures, 1985-94

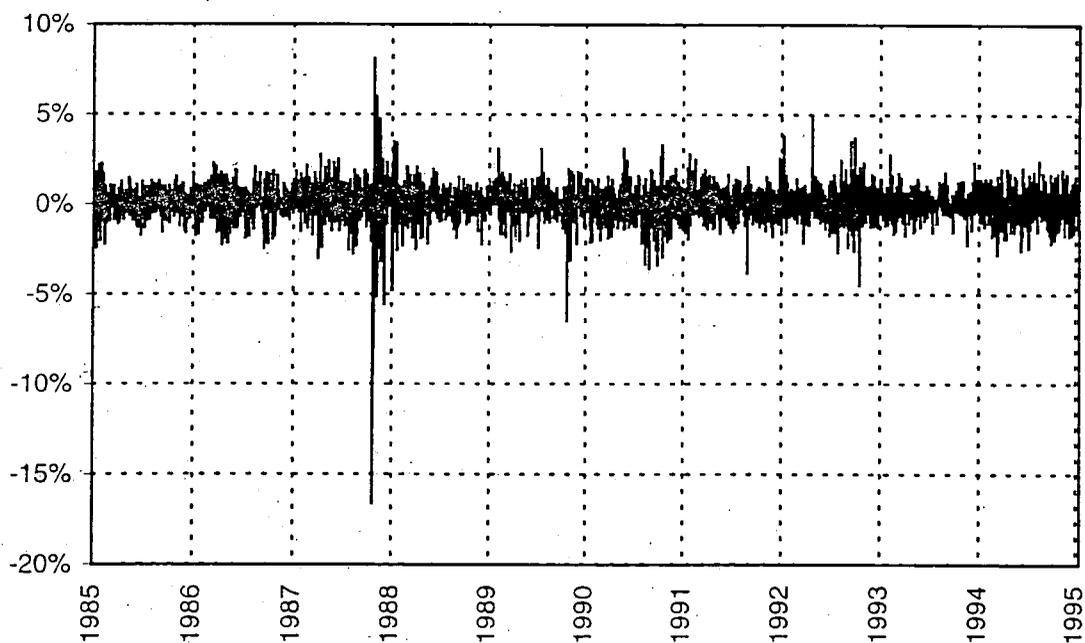


Table 5.1: Distribution of Daily Returns (r_t , %)

FTSE-100 Futures, 1985-94

	Mean	Std.	Skew.	Kurt.	Q1	Med.	Q3	Min.	Max.
1985-94	0.0	1.1	-1.7†	25.0†	-0.6	0.0	0.7	-16.6	8.1
1985	0.0	0.8	-0.1	0.2	-0.5	0.1	0.7	-2.4	2.3
1986	0.1	0.9	-0.2	-0.2	-0.5	0.1	0.7	-2.4	2.3
1987	0.0	2.0	-2.8†	21.5†	-0.7	0.2	1.1	-16.6	8.1
1988	0.0	0.9	0.2	1.6†	-0.5	0.0	0.6	-2.5	3.5
1989	0.1	1.0	-1.1†	6.5†	-0.4	0.2	0.8	-6.5	3.1
1990	-0.1	1.1	-0.2	0.5	-0.8	0.0	0.7	-3.6	3.3
1991	0.0	0.9	0.3†	1.4†	-0.6	0.0	0.7	-3.8	3.8
1992	0.0	1.1	0.4†	2.5†	-0.7	0.0	0.8	-4.5	5.0
1993	0.0	0.7	0.2	0.8†	-0.4	0.0	0.5	-2.2	2.8
1994	0.0	1.0	-0.1	-0.6†	-0.9	0.0	0.7	-2.8	2.4

† Significantly different from the normal distribution at a Type I error of 5%. The standard error for skewness from a normal parent is $\sqrt{6n}$, for kurtosis, $\sqrt{24n}$, where n is the sample size (see, e.g., Kendall and Stuart, 1969, p. 243).

ness in daily returns it occurs primarily in the tails, since the first and third quartiles are roughly symmetric around the median. But the minimum and maximum values also appear to be symmetric about the median. The conclusion is that there is no strong evidence that the return distribution is not symmetric.

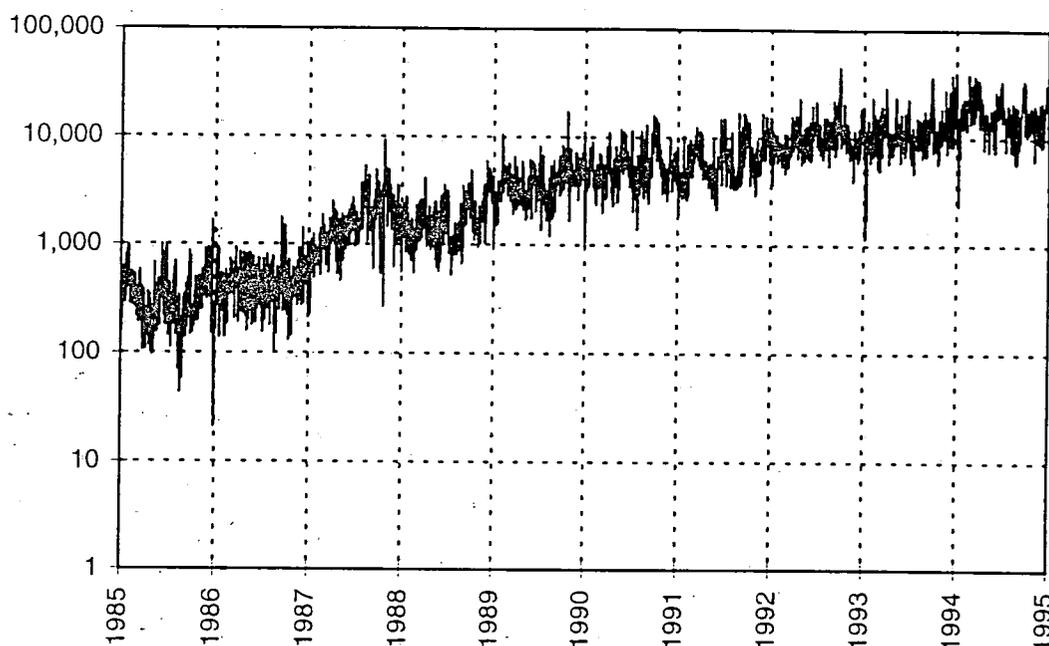
The leptokurtosis is much more unambiguous: eight out of the ten years show leptokurtosis, six of them significant, against one significant platykurtic year. Not surprisingly, the leptokurtosis of the complete dataset is much more pronounced than that of the individual years, since the period 1985-94 spans years with marked differences in the standard deviations.

Autocorrelation

The first-order autocorrelation characteristics of returns are shown in the lefthand panel of Table 5.2. While none of the coefficients is more than two standard errors

Figure 5.4: Trading Volume (v_t)

FTSE-100 Futures, 1985-94



from zero, there are six positive values versus four negative ones. The Durbin h statistics for residual autocorrelation, used here as a broad indicator of misspecification in the AR(1) model, are generally insignificant; the ARCH statistics show that ARCH effects appear to be sporadic. The very large value for the ARCH statistic for the full sample, when compared with the statistics for the individual years, suggests that the frequency of the variance process be quite low. This agrees with the findings of Table 5.1.

Overall the suggestion from the data of Tables 5.1 and 5.2 is that the determinants of the stochastic process of r_t are quite unstable. This is shown by the instability of the statistics for individual years, and the substantial difference, particularly with respect to dispersion, between the full-sample statistics and those of the sub-periods.

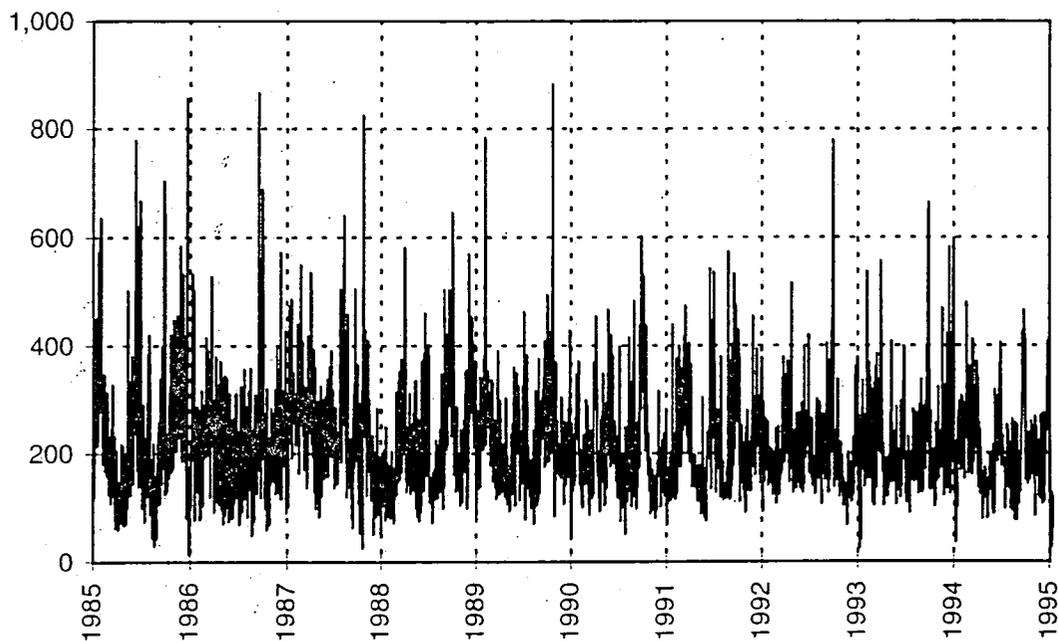
Table 5.2: Autocorrelation and ARCH in Returns and Volume

FTSE-100 Futures, 1985-94

	Returns, r_t			Volume, \hat{v}_t^*		
	$\hat{\rho}$	Durbin h	ARCH	$\hat{\rho}$	Durbin h	ARCH
1985-94	0.013	1.438	29.858	0.441**	-7.701	47.599
1985	-0.076	-2.391**	0.059	0.544**	-3.305	0.130
1986	0.040	0.128	4.018	0.206**	-1.177	1.362
1987	0.055	0.431	24.747**	0.430**	-2.091*	16.987**
1988	0.048	4.552**	10.254**	0.360**	-2.538**	0.178
1989	0.022	1.244	0.001	0.519**	-1.360	0.568
1990	-0.033	1.991*	3.519	0.532**	-3.694**	0.087
1991	-0.064	-1.244	0.362	0.365**	-1.960	0.016
1992	0.042	-0.057	10.590**	0.491**	-2.008*	3.589
1993	0.054	0.602	0.685	0.307**	0.102	0.911
1994	-0.109	0.070	1.387	0.429**	-3.433**	1.042

* Significant at a Type I error of 5% (** 1%). The standard errors for the autocorrelation coefficient are $(n-1)^{-1/2}$ where n is the sample size (see, e.g., Kendall and Stuart, 1969, p. 396); by this criterion none of the correlation coefficients on returns is significant at 5%. 'Durbin h' is Durbin's test for first order autocorrelation in the residuals; 'ARCH' is from a regression of squared residuals on a constant and one period lagged squared residuals.

Figure 5.5: Detrended Trading Volume (\hat{v}_t)
FTSE-100 Futures, 1985-94



5.3 Trading Volume

The trading volume for each day was calculated by adding the volume on each of the contracts available. It is shown in Figure 5.4. Clearly there has been a huge increase in trading volume over the period which probably reflects the gradual acceptance of futures contracts by professional investors. Interestingly, the growth now shows some sign of flattening out. Models such as Tauchen and Pitts (1983) can account for the increase in traders on the amount of trading volume generated per news bit. However, in the light of the criticism of this model advanced in Section 2.5, the simpler expedient of detrending the volume data is used. The trend was fitted as a 63 day moving average over log-volume. This period, which is the mean number of trading days between contract expiries, was chosen since any seasonality is likely to be related to contract expiry. Detrending by this moving average will tend to remove cycles of an order greater than one quarter but preserve higher-order cycles (including such things as day-of-the-week effects, although these are not allowed for here). The usual alternative, the fitting of an exponential trend (as in Gallant *et al.*, 1992), would clearly introduce low-order cyclicity since volume v_t does not follow such a trend (Figure 5.4); it seems better to use the more conservative technique.

Detrended volume, v'_t , is scaled to be of the same magnitude as the first observation of 1985:

$$v'_t \stackrel{\text{def}}{=} v_1 \times \exp(\log v_t - \text{m.a.} \{ \log v_t, 63 \}) \quad t = 1, \dots, 2529. \quad (5.2)$$

This series is displayed in Figure 5.5. At this point it is quite clear that volume displays some further seasonality relating to contract expiry, which shows up as a repeating quarterly pattern.

Contract Rollovers

The cause of this seasonality is contract rollovers (see, e.g., Yadav and Pope, 1990; Holmes, 1993), and this also provides for a remedy. As a contract nears expiry an investor wanting to maintain a position will close the near contract and simultaneously open an identical position in the far contract. This activity generates two lots of trading volume which are entirely un-news-related (i.e. systematic). Were it possible to estimate the number of contracts rolled over each day, then twice this number could be subtracted from the trading volume figure for the day to leave news-related volume only.

It is not possible to estimate the number of rollovers from the data supplied by LIFFE. However, it is possible to proxy the number using the data on open interest.¹ If I roll-over n contracts then the open interest in the near contract will fall by n , and the open interest in the far contract will rise by n . Therefore a sign of rolling over is a fall in near open interest and a corresponding rise in far open interest. The amount of rolling over is taken to be the smaller of these two numbers, since this ensures that for every rolled over contract there is both a closure and an opening. This gives rise to the proxy for rolled over contracts, l_t ,

$$l_t \stackrel{\text{def}}{=} \min \left\{ \max \{0, -(o'_t - o'_{t-1})\}, \max \{0, (o''_t - o''_{t-1})\} \right\}, \quad (5.3)$$

where o_t is the open interest on day t and a single prime denotes the near contract and a double prime the far contract. The maximizations in eq. (5.3) are to ensure that only a fall in the near contract and a rise in the far contract open interests will count, while the minimization is to ensure that there is matched closing and opening.

If the reasoning is correct, then l_t should show marked seasonality, being near

¹I am grateful to Phil Holmes for this insight on the use of open interest data.

Figure 5.6: Contract Rollover Proxy (l_t)

FTSE-100 Futures, 1985-94

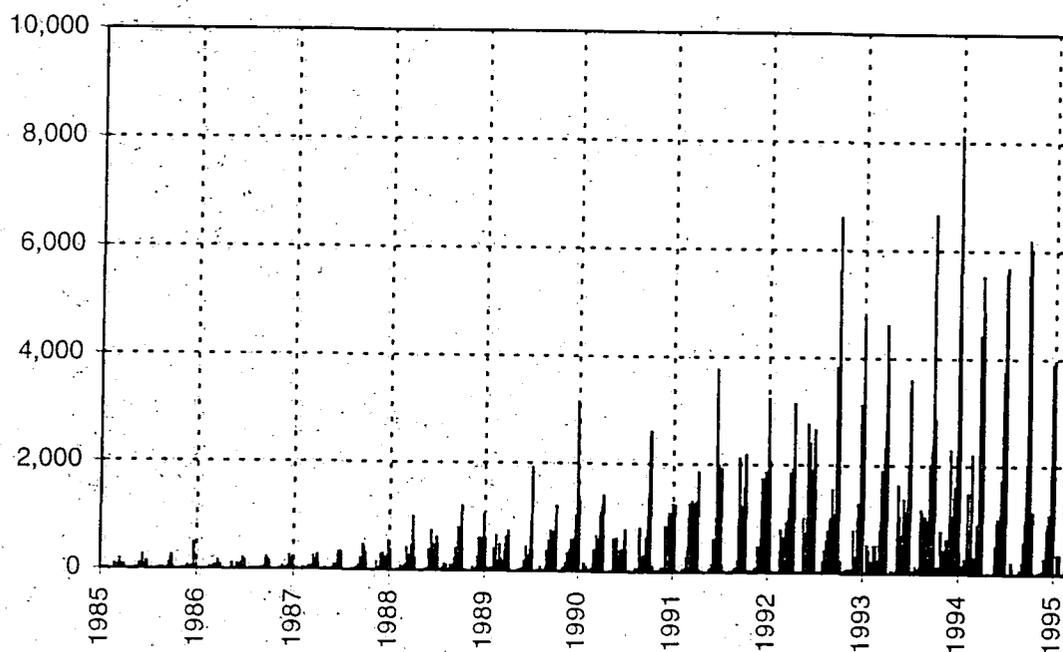
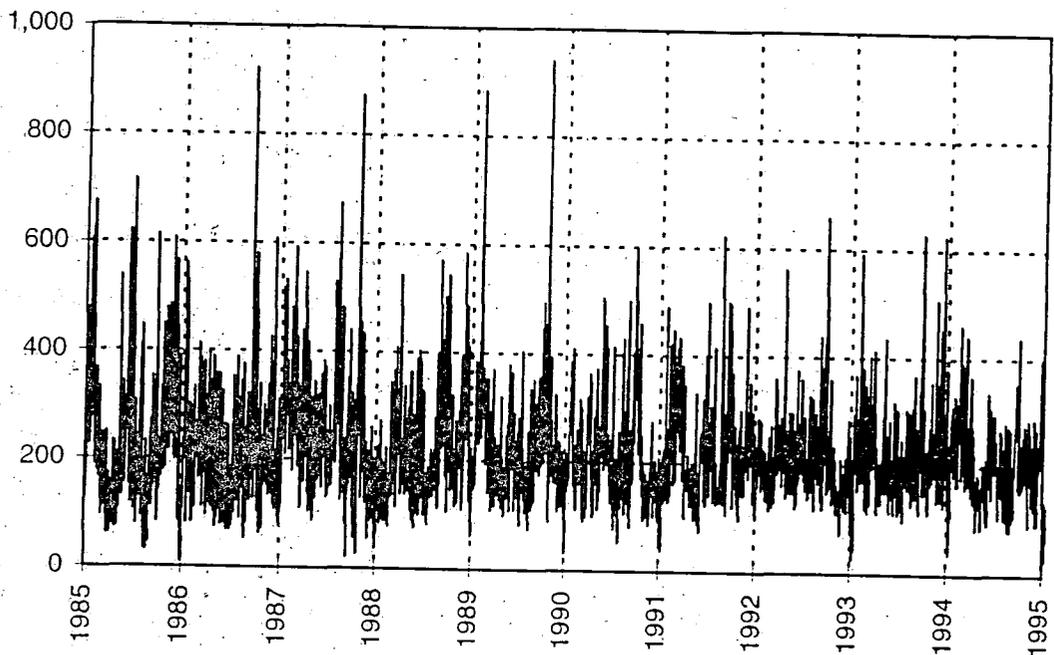


Figure 5.7: Detrended Trading Volume adjusted for Rollover, (v_t^*)

FTSE-100 Futures, 1985-94



zero early in the quarter, and only large in the final few days. The graph of l_t is given in Figure 5.6, and happily it has approximately this form. The detrended rollover-adjusted volume series is denoted v_t^* , where

$$v_t^* \stackrel{\text{def}}{=} v_t \times \exp(\log(v_t - 2l_t) - \text{m.a.} \{\log(v_t - 2l_t), 63\}). \quad (5.4)$$

This is shown in Figure 5.7. The graph is much less 'spiky', but there is still some quarterly seasonality in v_t^* . This is to be expected since the series l_t is only a proxy for rollovers. It is quite possible, for example, for both amounts of open interest to fall and for there still to be some rolling over of contracts. This might occur if there just happened to be a large number of closing positions taken in the far contract. In this case the correct value for rollovers should be positive, but l_t would give zero. Hence the adjustment for rollovers using l_t is not perfect (it is probably biased downwards) but it is simple and intuitive.

Descriptive statistics for the volume series v_t^* are given in Table 5.3. As they stand these figures show that the distribution of trading volume is clearly not normal. Both the positive skewness and the leptokurtosis are significant in all cases (and highly significant in almost all). However, these statistics are biased estimators of their population equivalents because of the high degree of dependence between successive observations apparent from Figure 5.7.

The most notable feature of these descriptive statistics is the non-normality of the volume data. This non-normality is consistent with the 'mixture of distributions' model of Tauchen and Pitts (1983) discussed in Section 2.3, since in this model volume is normal conditional upon the number of pieces of news per day. As the amount of news varies from day to day so do both the mean and the variance of volume.

Table 5.3: Distribution of Trading Volume (v_t^*)

FTSE-100 Futures, 1985–94

	Mean	Std.	Skew.	Kurt.	Q1	Med.	Q3	Min.	Max.
1985–94	225	103	1.6†	5.1†	157	204	271	10	941
1985	234	135	1.0†	1.0†	139	219	311	10	715
1986	226	108	1.7†	7.0†	156	211	275	53	922
1987	260	127	1.3†	3.7†	170	238	325	20	873
1988	218	99	1.2†	1.5†	143	198	264	62	584
1989	230	109	2.5†	11.7†	158	208	278	33	941
1990	213	93	1.0†	1.1†	145	197	260	38	596
1991	226	89	1.2†	2.0†	160	213	267	74	621
1992	204	78	2.1†	8.1†	160	191	232	27	658
1993	222	85	1.7†	4.9†	168	207	267	38	628
1994	207	73	0.6†	0.6†	157	195	255	28	461

† Significantly different from the normal distribution at a Type I error of 5%. See note to Table 5.1.

Autocorrelation

The first-order time series properties of the volume data are given in the righthand panel of Table 5.2. It is clear that there is highly significant positive autocorrelation in trading volume. The Durbin h statistics show evidence of misspecification from a simple AR(1) process, but there is little evidence of ARCH effects. In general, a comparison of the return data and the volume data in this Table suggests that the stochastic process governing volume is considerably more complicated than that governing returns.

5.4 Symmetry

I now turn to the bivariate properties of the data. It was argued by Karpoff (1987) that the return/volume relationship in futures should be symmetric with respect to price falls and rises because of the symmetry of the cost of short and long positions.

Figure 5.8: Squared Returns and Trading Volume (r_t^2, v_t^*)

FTSE-100 Futures, 1985-94

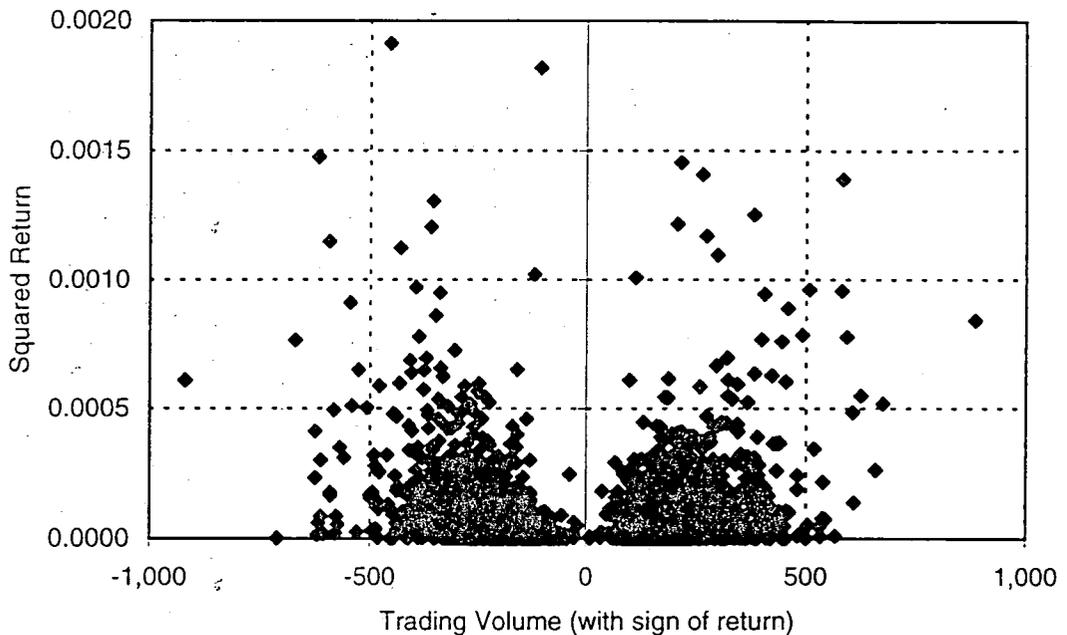


Figure 5.8 shows the relation between squared returns, r_t^2 and trading volume, \hat{v}_t^* , separated according to whether the return is positive or negative. Visually there is nothing to differentiate the positive returns from the negative returns, supporting Karpoff's proposition.²

It would be helpful to support this visual evidence with fitted lines. The analysis of Section 2.4 provides some guidance for the functional form and the specification of the disturbance term, on the assumption that the Tauchen and Pitts (1983) inter-day model is broadly correct. The simple exponential relationship (writing v_t for trading volume on day t for simplicity)

$$(\Delta p_t)^2 = \exp\{\alpha + \beta v_t\} u_t, \quad (5.5)$$

²In fact, Figure 5.8 is rather reminiscent of the folded inkblot butterflies I made as a child.

where u_t has support on the non-negative real line with $\mathcal{E}[u_t] = 1$, $\mathcal{V}[u_t | v_t] = v_t^2 \sigma^2$, has the following appropriate properties:

1. $(\Delta p_t)^2 \geq 0$ (required by definition);
2. $\mathcal{E}[(\Delta p_t)^2 | v_t = 0] > 0$ (Proposition 2.1);
3. $\mathcal{E}[(\Delta p_t)^2 | v_t]$ increasing in v_t providing $\beta > 0$ (in sympathy with Proposition 2.1, although note the discussion that in theory $\mathcal{E}[(\Delta p_t)^2 | v_t]$ does not have to be increasing for all $v_t > 0$);
4. $\mathcal{E}[(\Delta p_t)^2 | v_t]$ non-linear in v_t (Corollary 2.4);
5. $\mathcal{V}[(\Delta p_t)^2 | v_t]$ asymptotically quadratic in v_t (Corollary 2.3).

The only major problem with this functional form is that it does not asymptote to a ray through the origin. This is partly addressed by estimating the regression using Weighted Least Squares (WLS), which is necessary to accommodate the heteroskedasticity in u_t . Taking the logarithm of both sides of eq. (5.5) gives

$$\ln \{(\Delta p_t)^2\} = \alpha + \beta v_t + w_t, \quad (5.6)$$

where $w_t \stackrel{\text{def}}{=} \ln u_t$. To a first order approximation, $\mathcal{E}[w_t] = 0$, $\mathcal{V}[w_t | v_t] = v_t^2 \sigma^2$.³ In order to offset the heteroskedasticity, the regression must be estimated using the weights v_t^{-1} . This has the effect of down-weighting the large volume observations, and so mitigating the failure of the functional form to asymptote to a ray.⁴

There is also a minor problem with eq. (5.5), which is that asymptotically the distribution of the disturbance term is not χ_1^2 , as suggested by Proposition 2.2.

³If $g(x)$ is some function of x , then to a first order approximation, $\mathcal{E}[g(x)] = g(\mathcal{E}[x])$ and $\mathcal{V}[g(x)] = (g'(\mathcal{E}[x]))^2 \mathcal{V}[x]$ (see, e.g., Kendall and Stuart, 1969, pp. 231–2). Hence if $\mathcal{E}[u] = 1$, $\mathcal{V}[u | v] = v^2 \sigma^2$ and $w \stackrel{\text{def}}{=} \ln u$, then $\mathcal{E}[w] \approx 0$ and $\mathcal{V}[w | v] \approx (\frac{1}{1})^2 v^2 \sigma^2 = v^2 \sigma^2$.

⁴It also down-weights the period surrounding the stock market Crash, 1987q4–1988q1, since this was a period of high trading volume—see Figure 5.7. This period is discussed further in Section 5.5.

Table 5.4: WLS Regression for Symmetry

FTSE-100 Futures, 1985-94

	n	$\hat{\alpha}$	$\hat{\beta}$	R^2	DW	SSR
All r_t	2,524	-12.2691** (0.0700)	0.0078** (0.0005)	0.0943	1.8597	11,148.4
$r_t > 0$	1,344	-12.5514** (0.0885)	0.0095** (0.0006)	0.1372	1.3782	5,684.8
$r_t < 0$	1,180	-11.6695** (0.1157)	0.0045** (0.0007)	0.0312	1.3718	5,369.1

** Significantly different from zero at a Type I error of 1%. The critical value for the Durbin Watson (DW) statistic is 1.778, indicating that there may be some misspecification. Proceeding on the basis that this is not serious, a Chow test for stability across positive and negative returns using the sum of squared residuals (SSR) has a test statistic of 10.77, well in excess of the critical value at a Type I error level of 1% (5.57).

However, if a normal distribution is assumed for w_t in eq. (5.6), then u_t has a log-normal distribution which shares many of the characteristics of the χ_1^2 distribution, notably having support on the non-negative real line, a similar mean and positive skewness.

The results of estimating eq. (5.6) over the sample are given in Table 5.4. The coefficient $\hat{\beta}$ has the correct sign and both parameters are highly significant. However, there is some evidence of misspecification from the Durbin-Watson (DW) statistics, which is to be expected since the functional form eq. (5.5) satisfies only certain necessary conditions for representing $(\Delta p_t)^2 | v_t$ within the Tauchen and Pitts model. Proceeding with the Chow test for symmetry with respect to positive and negative returns, the null hypothesis is strongly rejected. This contradicts the findings of the review of price/volume studies conducted by Karpoff (1987), where a symmetric relationship is inferred from the absence of any correlation between Δp_t and v_t in futures markets. It also contradicts the Tauchen and Pitts (1983) inter-day model in which there is no differentiation between good and bad news.

On the basis of Table 5.4, negative price changes seem to have an initially higher but flatter $\mathcal{E} [(\Delta p_t)^2 | v_t]$ than positive ones. The critical amount of volume at which the two lines intersect is $v_t^* \approx 177$. This corresponds to a daily return of $\pm 0.4\%$, implying that about 30% of all days have a higher negative than positive relationship. Recent studies have emphasised the asymmetry of price variability with respect to good and bad news (see, e.g., Engle and Ng, 1993). It is generally found that bad news increases subsequent price variability. If it is presumed that the response of trading volume to news is unaffected by whether the news be good or bad, then this would be the case in the above data only for days in which there were a small amount of news.

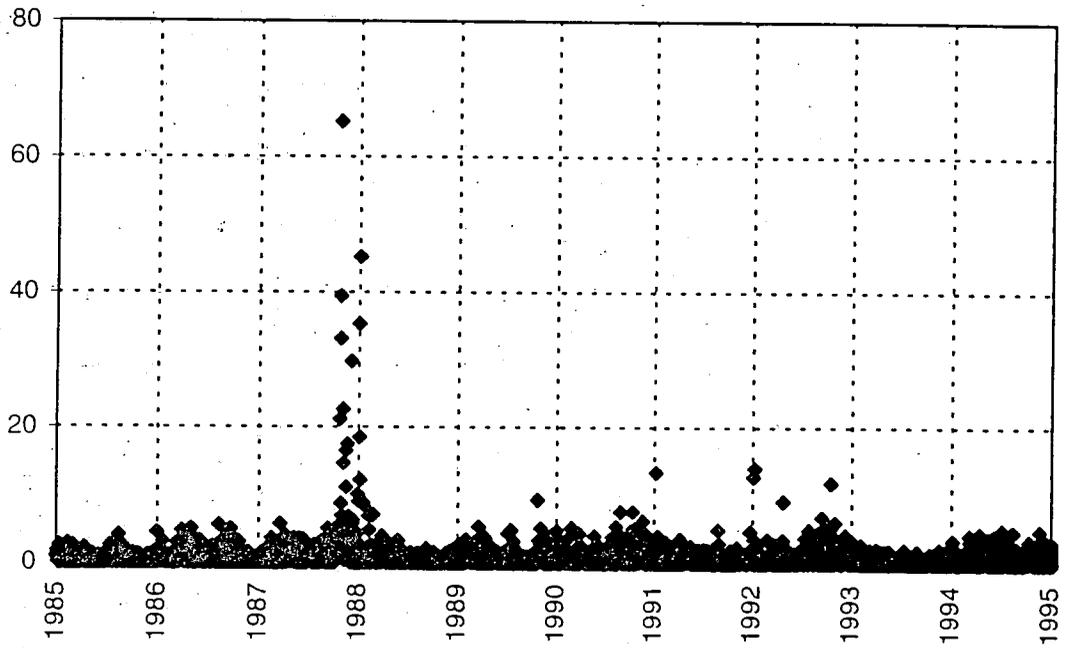
This inconsistency, the parameter instability and the evidence of misspecification from the DW test statistics suggest that an interpretation of the price and volume data within the Tauchen and Pitts inter-day framework is incomplete, i.e. there is more going on in the price/volume relationship than simple news-dynamics. The next section considers an alternative test of this framework, using the asymptotic results of Proposition 2.2.

5.5 Testing the Tauchen and Pitts Model

Rather than using the findings of Section 2.4 indirectly in an attempt at line-fitting, Proposition 2.2 may be used directly in a test of the model. If the Tauchen and Pitts (1983) inter-day model is correct, then the squared return values adjusted by volume should have approximately a scaled χ_1^2 distribution for large volumes.

Figure 5.9 shows the time-series of volume-adjusted squared returns, r_t^2/v_t^* , rescaled to have a mean of 1 in order to match the χ_1^2 distribution. This figure shows the unusual nature of the 'stock market Crash' and its repercussions in a way that neither squared price change nor trading volume can do on their own. The

Figure 5.9: Volume-adjusted Returns (r_t^2/\hat{v}_t^* , rescaled)
FTSE-100 Futures, 1985-94



most unusual aspect of the Crash period was the very high ratio of price volatility to trading volume. As well as the very strong time-dependency around this period, there also appears to be a lesser degree of time-dependency throughout the sample period, in the way that the data tends to cluster. This time-dependency is quite consistent with the Tauchen and Pitts model, except in the asymptotic case where the amount of trading volume (equivalently, the amount of news) becomes large.

For an example of how this time-dependency might arise, consider the case of positive autocorrelation in the mean of the news-arrival process, as might be inferred from the success of ARCH models in modeling returns (see Section 2.6). This will affect the value of $\mathcal{E}[N | T]$ in the notation of Section 2.4, which denotes the expectation of the amount of news which has arrived given an observation on the amount of trading volume. As a simple example of this, consider the expectation of the amount of news given that no volume has been observed, on the presumption that the amount of news per day is Poisson with arrival rate λ . By Bayes Theorem

$$\mathcal{E}[N | T = 0] = \sum_{n=0}^{\infty} n \frac{\Pr\{T = 0 | N = n\} \Pr\{N = n\}}{\Pr\{T = 0\}}. \quad (5.7)$$

Writing $\Pr\{y_i = 0\} = q$ it follows that $\Pr\{T = 0 | N = n\} = q^n$. By the Poisson distribution, $\Pr\{N = n\} = e^{-\lambda}(\lambda^n/n!)$. Making these substitutions,

$$\begin{aligned} \mathcal{E}[N | T = 0] &= \frac{\sum_{n=0}^{\infty} n q^n e^{-\lambda}(\lambda^n/n!)}{\sum_{n=0}^{\infty} q^n e^{-\lambda}(\lambda^n/n!)} \\ &= \frac{e^{-\lambda} q \lambda \sum_{n=1}^{\infty} \frac{(q\lambda)^{n-1}}{(n-1)!}}{e^{-\lambda} \sum_{n=0}^{\infty} \frac{(q\lambda)^n}{n!}} = \frac{q\lambda e^{q\lambda}}{e^{q\lambda}} = q\lambda. \end{aligned} \quad (5.8)$$

Table 5.5: χ^2 Test on Volume-adjusted Squared Returns

FTSE-100 Futures, 1985–94

	Quartiles			
	Q1	Q2	Q3	Q4
Full Sample	19.201 [0.024]	16.241 [0.062]	9.753 [0.371]	32.312 [0.000]
Less 87q4–88q1	12.706 [0.176]	13.209 [0.153]	12.871 [0.169]	8.779 [0.458]

The values in brackets represent the area of the χ^2 distribution (9 degrees of freedom) lying above the value of the test statistic.

Consequently, dynamic behaviour in λ leads to dynamic behaviour in $\mathcal{E}[N | T = 0]$, and, by extension, $\mathcal{E}[N | T = t]$ ($t > 0$). As was shown in Section 2.4, this in turn leads to dynamic behaviour in the expected price change squared conditional upon the amount of trading volume, i.e. $\mathcal{E}[(\Delta p_t)^2 | v_t]$.

Proposition 2.2 shows that the distribution of squared price changes scaled by volume should become more and more like a scaled χ_1^2 distribution as volume increases. To investigate this, the complete sample is divided into its four quartiles by trading volume (as given in Table 5.3, p. 96), both with and without the unusual Crash period, 1987q4–1988q1. If the Tauchen and Pitts model is correct, the fit between the volume-adjusted squared returns and the χ_1^2 should be at least as good for samples from high volume days as from low volume days. The observed and expected distributions for the four quartiles are given in Figure 5.10, and the results of the χ^2 tests in Table 5.5. It can be seen that the full sample does not fit the χ_1^2 distribution in the largest quartile by volume at all well (although the fit in quartile 3 is quite good). The exclusion of the Crash period, however, substantially improves the fit to the point where it is entirely consistent with the Tauchen and Pitts model.

⁵In fact, it is easy to see that $N | T = 0$ is Poisson with arrival rate $q\lambda$, from which this expectation follows.

Figure 5.10: Observed and Actual Distributions of r_t^2/v_t^* by Quartile

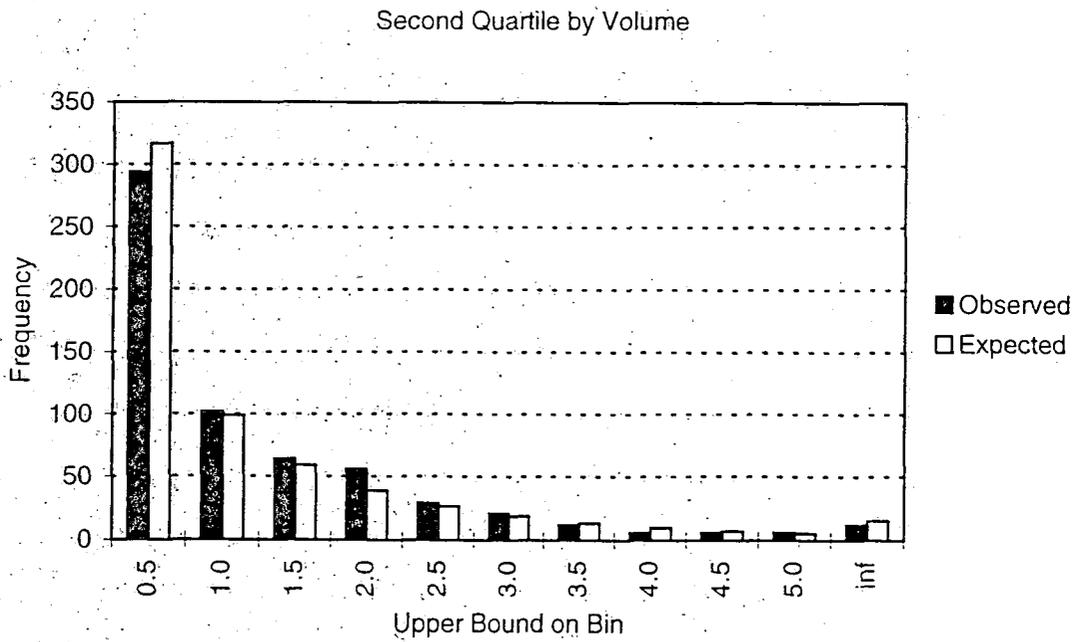
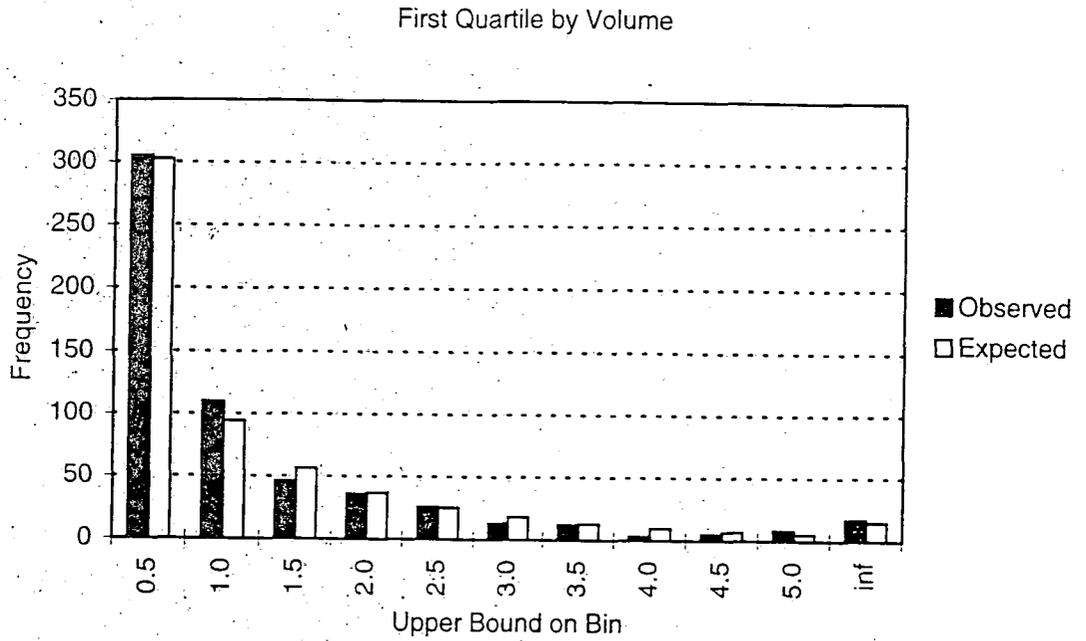
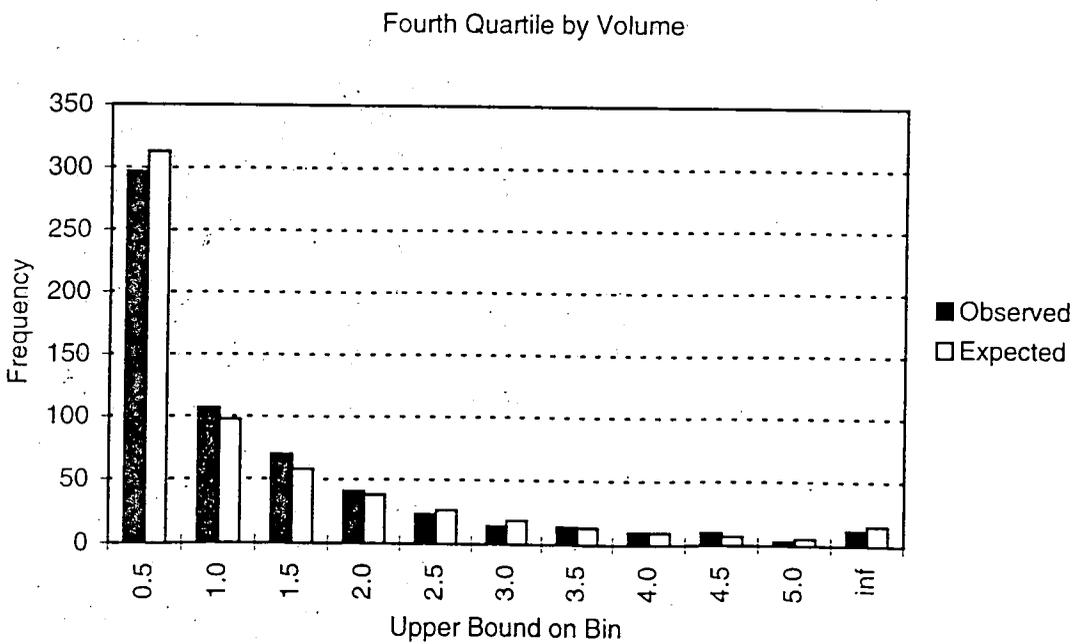
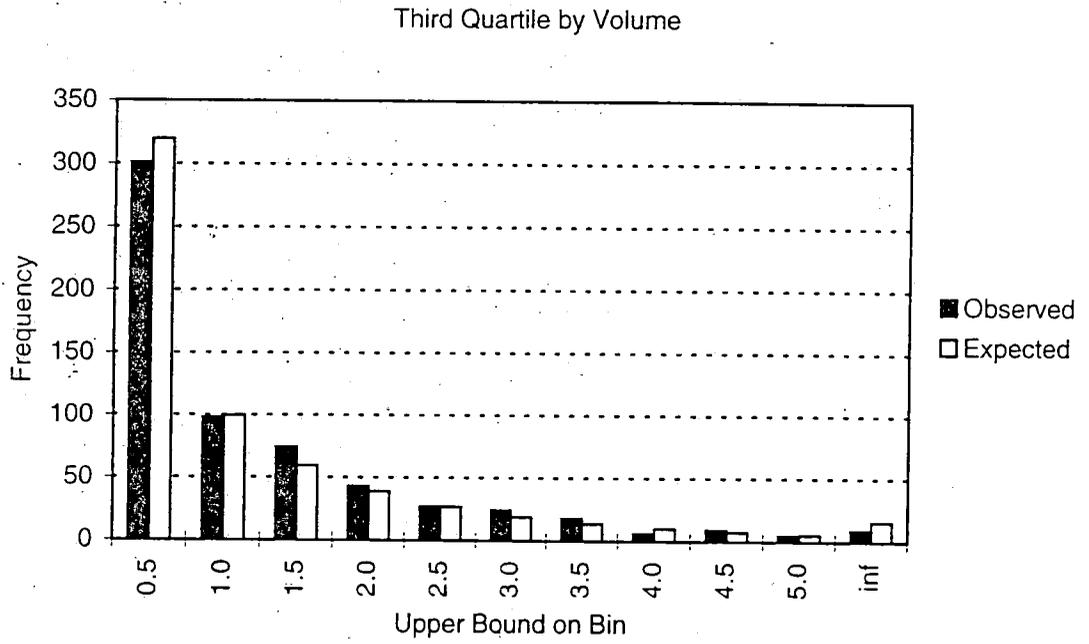


Figure 5.10: Observed and Actual Distributions of r_t^2/v_t^* by Quartile (cont)



5.6 Summary and Conclusion

This chapter has examined the Tauchen and Pitts (1983) inter-day model of price change and trading volume, using the analysis of that model presented in Section 2.4 and data from the London International Financial Futures Exchange. Prior to performing the analysis it was necessary to create a price index for futures contracts using the 'optimal' index presented in Chapter 4, and to adjust the trading volume for an upward trend, reflecting the gradual acceptance of the benefits of futures contracts by investors, and for the seasonal effects of contract roll-overs, as discussed in Section 5.3.

In the Tauchen and Pitts model both price change and trading volume are driven by the amount of news arriving in the day, with the result that the variance of price change and the expectation and variance of trading volume are all linear in the amount of news per day. One of the problems with this model is the difficulty of estimation unless it is assumed that the amount of news per day is independently and identically distributed. Unfortunately, there is a substantial amount of evidence to suggest that the amount of news per day displays some positive autocorrelation, as was discussed in Section 2.6 and as can be inferred from the autocorrelation of trading volume shown in Table 5.2 and the time-series properties of volume-adjusted squared price changes displayed in Figure 5.9. However, Section 2.4 derived testable implications of the Tauchen and Pitts model; some of which are robust to the behaviour of the news-arrival process.

Proposition 2.1 proposed some necessary conditions which should be satisfied by any Tauchen and Pitts-like model. In Section 5.4 of this chapter these conditions are used to select a functional form for the relationship between squared price change and trading volume, which is estimated over the full sample period of 1985–94. Although the model performs fairly well, the parameters are shown to be unstable with respect

to price rises and falls. This contradicts the suggestion of Karpoff (1987), which is that the price/volume relationship should be symmetric with respect to rises and falls in markets where the transactions costs are symmetric (such as futures markets). It also contradicts the Tauchen and Pitts model, in which there is no differentiation between good and bad news. More recent empirical analysis has suggested that the price response to good and bad news is not symmetric, which is consistent with this type of instability of the parameters.

Proposition 2.2 provided an asymptotic distribution for volume-adjusted squared price changes, which is independent of the news arrival process (since it matters only that the amount of news, and similarly the amount of volume, be large). Section 5.5 used this as a basis of a goodness of fit test. This test suggested, once due allowance had been made for the very unusual events of the stock market Crash in 1987q4 and 1988q1, that the data was consistent with the Tauchen and Pitts model.

Taking the evidence of Sections 5.4 and 5.5 together, the conclusion of this chapter is that the Tauchen and Pitts model is at least partly correct, although due allowance must be made for a time-varying news arrival rate. However, the counter side to this conclusion is that news effects cannot explain all of the price/volume relationship. In other words, even were the amount of news per day to be completely fixed, there would still be price dynamics and volume dynamics. Exactly how this might be the case is the subject of Chapter 6.

Chapter 6

A Market-Clearing Model

6.1 Introduction

This chapter considers the price/volume dynamics that arise from market-clearing pseudo-homogeneous investor models, under a general (non-linear) model of the way in which reservation prices are updated. The analysis highlights the way in which price/volume dynamics are sensitive to the manner in which reservation prices are updated from period to period. This sensitivity is demonstrated in Monte Carlo simulations.

The evidence of Chapters 2 and 5 suggests that while news effects, in particular positive autocorrelation in the amount of news per day, can explain much of the price/volume relationship they cannot explain it all. Or to put this another way, the Tauchen and Pitts (1983) interday model analysed in Section 2.4 could explain the price/volume relationship but only under the condition that the parameters of the model themselves varied in some fashion through time. This explanation is simply a *reductio ad absurdum*, since it transfers our interest from the non-constancy of the news process to the non-constancy of the parameters describing the impact of the news process, which is not very satisfactory.

The alternative explanations have been mentioned in Section 2.6. Bera and Higgins (1993) consider that the ARCH effects observed in daily price changes may also be caused by random coefficients (i.e. non-constant parameters again), or by a non-linear autocorrelation structure. The difficulty with these explanations is that there is, as yet, little understanding of how they might come about within a model of the behaviour of optimizing agents. They therefore fall within the purview of 'descriptive models' in the taxonomy of Section 1.4 (Figure 1.1, page 10). In contrast, the 'heterogeneous investors' models of the same taxonomy can too easily explain price/volume dynamics as the interaction of different types of agent, some possibly irrational. The problem is that in their generality these models provide little structure within which the price/volume relationship may be estimated, a point made by Gallant *et al.* (1992) in their data-based (i.e. descriptive) analysis.

This chapter provides an explanation of aspects of the price/volume relationship based around the 'pseudo-homogeneous investor' model of Tauchen and Pitts (1983), but which is not news-related. In the Tauchen and Pitts model optimizing investors differ in the way in which they interpret the stock of public information, and dynamics arise as a result of the dynamics of the news-arrival process. In the model of this chapter, the amount of news per period can be constant, and yet the result will be that there is the possibility of autocorrelation in price changes, absolute price changes and trading volume. It is also the case that the strength of these effects will vary, so it will appear as though the price/volume relationship has random coefficients and/or non-linear autocorrelation.

The outline of the chapter is as follows. Section 6.2 describes the general framework for considering market-clearing models, and Section 6.3 considers the Tauchen and Pitts (1983) model within this framework. Section 6.4 considers the cross-sectional reservation price distribution and how it might be important, and proposes a model for updating reservation prices which incorporates the model of Tauchen

and Pitts as a special case. Section 6.5 performs a Monte Carlo analysis using the general model, and Section 6.6 examines the implications for autocorrelation in price changes in particular. Section 6.7 concludes.

6.2 Market Microstructure

In this Chapter all investors are assumed to solve the same problem, first given in eq. (1.8)

$$\max_q \int_{-\infty}^{\infty} u(q(p_k - p_{t,k}) + w i; \eta) f(p_k; \theta, p_{t,k}) dp_k, \quad (6.1)$$

subject to constraints on the largest position that can be held. The utility function $u(\cdot)$ is parametrized by η , and the probability distribution $f(\cdot)$ by the current futures price $p_{t,k}$ and the parameter(s) θ ; p_k is the underlying spot price at k .

The first order condition for this problem, ignoring the constraints, is

$$\int_{-\infty}^{\infty} u'(q(p_k - p_{t,k}) + w i; \eta) (p_k - p_{t,k}) f(p_k; \theta, p_{t,k}) dp_k = 0. \quad (6.2)$$

Following the work of Tobin (1958, 1969), Samuelson (1983) considered the impact on demand functions arising from eq. (6.2) of changes in the expectation of p_k . Samuelson found that there was a unique value for the expectation at which demand was zero (Samuelson, 1983, Mathematical Appendix C). Analogously, there will be a unique value for the futures price at which demand will be zero, and this is termed the *reservation price*. For values below the reservation price, the investor desires to be long of contracts (i.e. $q > 0$), and for values above, short. The reservation price, denoted $p_{t,k}^*$, is thus a function of the parameters w , i , η and θ .

Reservation Prices and Market-Clearing

The demand function for any investor can be expanded around the reservation price using a Taylor Series (for convenience the k subscript denoting the time of contract expiry is suppressed)

$$q(p_t; p_t^*) = (p_t - p_t^*) q'(p_t^*) + \frac{1}{2} (p_t - p_t^*)^2 q''(p_t^*) + \frac{1}{3!} (p_t - p_t^*)^3 q'''(p_t^*) + \dots \quad (6.3)$$

where the first term in the series is identically zero and has been dropped. In this way the demand function of any investor is simply a relationship between the reservation price p_t^* and the market-clearing price p_t . This is very convenient for models of market-clearing price and volume dynamics. Once the process by which reservation prices are updated has been specified, the full dynamics drop out from the market-clearing condition

$$\sum_{j=1}^J q_j(p_t; p_{j,t}^*) = 0 \implies p_t = p(P_t^*), \quad (6.4)$$

where there are J investors in total, indexed by j , and P_t^* is the vector of reservation prices at time t . So, for example, if investors are pseudo-homogeneous and assumed to differ only in their reservation prices, and the coefficients in eq. (6.3) are known, then knowledge of the set of reservation prices at t and at $t + 1$ will be sufficient to determine the price change and the trading volume over the period $(t, t + 1]$.¹

¹Strictly, as in Tauchen and Pitts (1983), t should index trading rather than calendar time. However, since this chapter will concentrate entirely on trading time it is convenient to envisage a day as consisting of a starting equilibrium, the revelation of a fixed amount of new information, and a *tâtonnement* process in which equilibrium is restored at the day's close.

6.3 Tauchen and Pitts Again

The Tauchen and Pitts (1983) model slots neatly into this general framework. Their demand functions will be re-derived here, since the futures framework in eq. (6.1) is different from the general portfolio problem approach discussed in Sections 2.2 and 2.3. First, Tauchen and Pitts assume² that p_k is normally distributed with a mean \bar{p}_k and a variance σ^2 . In this case the expected utility expression can be rewritten as a utility function expressed on the mean and variance (see, e.g., Copeland and Weston, 1988, pp. 96–99):

$$\max_q u\left(q(\bar{p}_k - p_t) + w_i, q^2\sigma^2\right). \quad (6.5)$$

The first order condition for this problem is

$$u_1(\cdot)(\bar{p}_k - p_t) + u_2(\cdot)2q\sigma^2 = 0. \quad (6.6)$$

Second, Tauchen and Pitts assume that all investors trade-off expectation linearly against variance at a rate $\beta \stackrel{\text{def}}{=} -u_2/u_1 \geq 0$, which is the same at all inputs for the utility function. Dividing the first order condition through by u_1 , substituting β and rearranging gives the demand function

$$q(p_t) = \frac{1}{2\beta\sigma^2}(\bar{p}_k - p_t). \quad (6.7)$$

From this demand function it is clear that the reservation price in the Tauchen and Pitts model represents the expectation, \bar{p}_k . Since the demand function is already linear, all terms in eq. (6.3), bar the second, are zero, and the relation between

²It should be noted that the assumptions ascribed to Tauchen and Pitts are implicit, since in the paper they start with the linear demand functions given in eq. (2.18), rather than derive them within an optimizing framework.

market-clearing price and reservation price is simply

$$q(p_t; p_t^*) = c(p_t - p_t^*), \quad (6.8)$$

where $c \stackrel{\text{def}}{=} -1/2\beta\sigma^2$. The absence of constraints in this demand function would only be the case if there were no margin requirement.³

Finally, Tauchen and Pitts assume that investors are identical in all respects bar the reservation price. From the market-clearing condition this gives

$$\sum_{j=1}^J c(p_t - p_{j,t}^*) = 0, \quad (6.9)$$

which implies that the market-clearing price is the mean of investors' reservation prices, and the change in the market-clearing price is the mean of the change in investors' reservation prices. This latter result implies that we do not need to know the distribution of reservation prices at the beginning and the end of a period in order to infer the change in market-clearing price over the period. In fact, the distribution of reservation prices is entirely immaterial: the process by which reservation prices are updated is the only relevant information. The same is true of trading volumes: the trading volume on day t is determined entirely by the way in which reservation prices change over the period $(t-1, t]$ (see eq. (2.20)).

Updating Reservation Prices

Tauchen and Pitts propose a variance decomposition model in which all investors update their reservation prices independently of the current level of their reserva-

³It is worth noting that everything that follows holds *a fortiori* with the addition of margin requirements. The margin requirements appear as horizontal extremes to the demand functions, and the resulting non-linearity of these functions means that the market-clearing price is no longer the arithmetic mean of reservation prices. Consequently the change in the market-clearing price cannot be expressed as a function of the *changes* in the reservation prices, but must be a function of the reservation prices themselves, before and after.

tion price, and independently of the prevailing market-clearing price. This model has already been criticised in Section 2.5. One property of the model is that the dispersion of reservation prices increases linearly in time (strictly, trading time—see footnote 1), resulting in a greater and greater disagreement among investors, and consequently larger and larger positions. It seems far more natural that the dispersion of investors' expectations and the mean size of their positions be stationary over time.

The cause of this non-stationarity is the egoism of Tauchen and Pitts's investors, who, in completely disregarding the market-clearing price when updating their reservation prices, are each asserting that there is no extra information to be gained by considering the reservation prices of other investors. So, for example, an investor with a reservation price well above the market-clearing price is as happy to revise his reservation price up as he is to revise it down, and by the same amount. This behaviour is fundamentally at odds with the implications of heterogeneous investors models (Grossman, 1976; Grossman and Stiglitz, 1976). To *précis* models of this type, uncertainty among investors about the quality of their information and their analysis should make them unwilling to stray too far from the consensus, as reflected in the market-clearing price.

One Possible Generalization

Interestingly, a generalization of Tauchen and Pitts's updating process, in which investors are aware of the market-clearing price when determining their reservation prices, preserves the irrelevance of the reservation price distribution. Consider the case where the updating rule is linear:

$$p_{j,t+1}^* = p_t + (1 - \alpha)(p_{jt}^* - p_t) + u_{j,t+1}, \quad \alpha \in [0, 1], \quad (6.10)$$

where $u_{j,t+1}$ is identically and independently distributed with zero expectation. In other words the updated reservation price will drift towards the old market-clearing price by a proportion α : In the Tauchen and Pitts model, $\alpha = 0$.

Proposition 6.1 *If investors' reservation prices update according to eq. (6.10), then the expected change in market-clearing price between t and $t + 1$ is zero, i.e. entirely independent of the distribution of reservation prices at t .*

Proof: The proof is straightforward. From the market-clearing condition

$$\Delta p_{t+1} = J^{-1} \sum_{j=1}^J \Delta p_{j,t+1}^* \quad (6.11)$$

Substituting from eq. (6.10) and taking expectations,

$$\begin{aligned} \mathcal{E}[\Delta p_{t+1}] &= J^{-1} \sum_{j=1}^J \{p_t + (1 - \alpha)(p_{j,t}^* - p_t) - p_{j,t}^*\} \\ &= p_t + (1 - \alpha)(p_t - p_t) - p_t \\ &= 0, \end{aligned} \quad (6.12)$$

where the market-clearing condition is again used to equate p_t with the mean reservation price at t . ■

Perhaps the most interesting point about this proposition is that the result will clearly not hold for any expectations model which is non-linear in $p_{j,t}^* - p_t$. For example, were the drift to be a non-linear expression in the deviation of reservation price from market price, or were the variance to fall with the absolute size of this deviation, then we would have, in general,

$$\begin{aligned} \Delta p_{t+1} &= f(p_{1t}^*, \dots, p_{Jt}^*, p_{1,t+1}^*, \dots, p_{J,t+1}^*) \\ &\neq f(\Delta p_{1,t+1}^*, \dots, \Delta p_{J,t+1}^*) \end{aligned} \quad (6.13)$$

and similarly for trading volume. At time t the probability distribution for Δp_{t+1} and for volume v_{t+1} will be conditional on the reservation price vector at time t ,

$$P_t^* = \{p_{1t}^*, \dots, p_{Jt}^*\}.$$

6.4 The Vector of Reservation Prices

Why is the potential presence of the vector of current reservation prices in Δp_{t+1} important? The reason is that since each reservation price updates locally (i.e. $p_{j,t+1}^*$ is generally quite close to $p_{j,t}^*$) the reservation price distribution shows a certain amount of inertia: it tends to change slowly over time. If there is a relationship between aspects of the distribution of reservation prices and the price change and trading volume distributions, then the inertia of the reservation price distribution will introduce a time-series pattern into these two variables, *quite independently of any pattern which may be caused by news*, or, indeed, of any pattern which may arise from heterogeneous investor models. This would be the simplest explanation of the time-series properties in Δp_t and v_t and is worth further investigation.

Consider the following example. The reservation price distribution P_t^* is initially symmetric. But in the updating of individual reservation prices suppose there is, purely by chance, a tendency for negative changes to be large while positive ones are small. Consequently the distribution P_{t+1}^* displays pronounced negative skewness. After another round of updating, this time far more typical, the severity of the skewness has been blunted by a layer of noise, but the skewness itself is unlikely to have been completely eradicated. Therefore P_{t+2}^* , P_{t+3}^* , ... will have diminishing traces of that single large fluctuation. In the meantime, of course, another wild fluctuation could have introduced an alternative feature, say unusual kurtosis. Again, the presence of this will die away only slowly.

To give a physical analogy, the reservation price distribution is like the arrange-

ment of gas molecules in a box. Each molecule pursues its own brownian motion, and the consequence is that the density of molecules at any point in the box is not constant but is always smoothly changing. In the gas analogy, the number of molecules is very large, and the fluctuations in density small. But in financial markets the number of active investors, although sizable, is relatively small. Moreover, this number may vary through time. In the case of a new market, the 'seasoning' period may involve substantial changes in the behaviour of prices changes and trading volumes as the number of active investors grows and the magnitude of the fluctuations in the reservation price distribution shrinks.

A Model for Updating

In the light of the preceding discussion, an acceptable model for updating reservation prices should have the following features.

1. Symmetry with respect to reservation prices above and below the market-clearing price.
2. A drift towards the market-clearing price in which strength increases non-linearly with the deviation between market-clearing price and reservation price. Reservation prices are a symmetric random walk only when the reservation price and the market-clearing price are the same.
3. A mechanism for preventing the distribution of reservation prices from becoming more and more dispersed over time (i.e. for imposing stationarity around the market-clearing price).

The Tauchen and Pitts model satisfies the first requirement but not the other two. The requirement of non-linearity is to make the model 'interesting', since, as has been shown in Proposition 6.1, a linear drift makes price a martingale. An interesting model should have the potential (which may not be realized in practice) for prices

to be a random walk/martingale for some of the time only, depending upon the distribution of reservation prices. The requirement for stationarity prevents the increasing discord implied by the Tauchen and Pitts model.

The following model is a simple implementation of these three features:

$$p_{j,t+1}^* = p_t + f(p_{j,t}^* - p_t) + \sigma z_{j,t+1}, \quad (6.14)$$

where

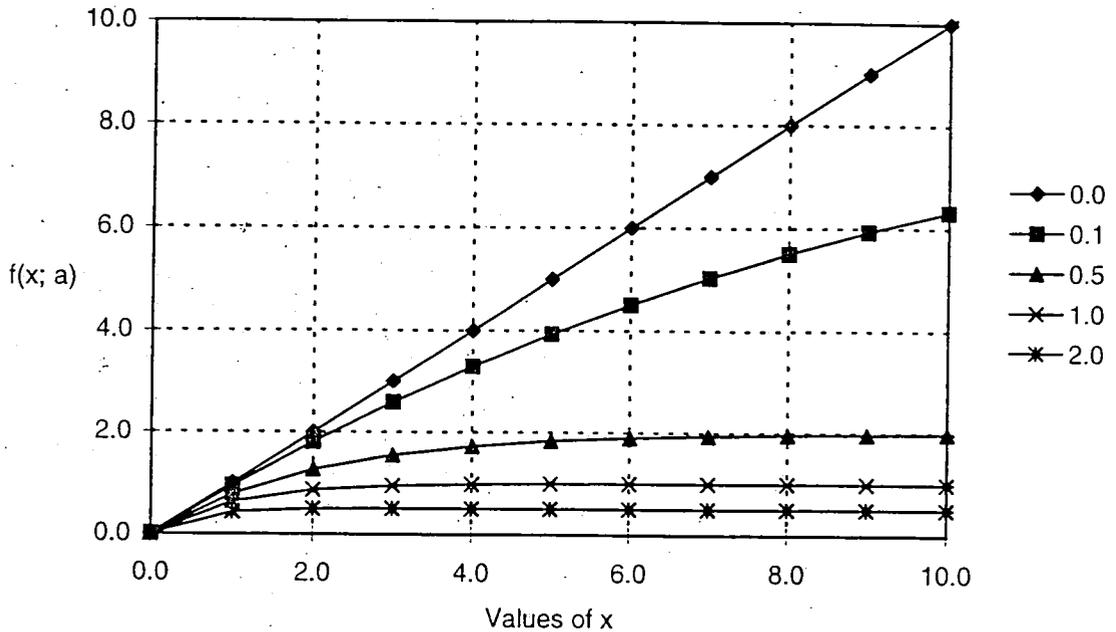
$$f(x) \stackrel{\text{def}}{=} \text{sign } \{x\} \times \frac{1}{\alpha} (1 - e^{-\alpha|x|}), \quad \alpha > 0 \quad (6.15)$$

and $z_{j,t}$ is independently and identically distributed as a unit normal for all j and t . The function $p_t + f(\cdot)$ represents the expectation at t of the updated reservation price at $t+1$; $f(x)$ has been constructed so that it is concave in $|x|$ with gradient 1 at the origin, and bounded by $\pm 1/\alpha$. For reservation prices very near to p_t , $f'(0) = 1$ ensures that the expectation of the updated reservation price is close to the current reservation price. For reservation prices further from p_t , the concavity of $f(\cdot)$ ensures that the expectation of the updated reservation price is nearer to p_t than the current reservation price.

The parameter α in eq. (6.15) determines the bounds of the expectation of the updated reservation price, and also the degree of concavity of f , i.e. the strength of the drift. The model has the attractive property that in the limit as $\alpha \rightarrow 0$ it becomes the random walk of Tauchen and Pitts:

$$\lim_{\alpha \rightarrow 0} \frac{1}{\alpha} (1 - e^{-\alpha|x|}) = \lim_{\alpha \rightarrow 0} \frac{|x| e^{-\alpha|x|}}{1} = |x|, \quad (6.16)$$

using L'Hôpital's rule. The parameter α is therefore an index of 'non-Tauchen and Pitts-ness': the larger it gets, the tighter the bounds for reservation prices around

Figure 6.1: The Function $f(x; \alpha)$, for $x \geq 0$ 

the market-clearing price and the stronger the drift back towards the market-clearing price. The function $f(x; \alpha)$ is illustrated for positive x in Figure 6.1.

The Role of Skewness

Using this model for reservation price updating it is now shown how the reservation price distribution enters into the expectation (and, a fortiori, the higher moments) of the price change distribution. The following result relates the range of possible values for the expected price change to the ratio of bullish to bearish investors.

Proposition 6.2 *Using the model for reservation price updating given in eq. (6.14), the expected price change over the period $(t, t + 1]$ lies in the range*

$$-\frac{1}{\alpha} \frac{J_t^-}{J} < \mathcal{E} [\Delta p_{t+1} \mid P_t^*] < \frac{1}{\alpha} \frac{J_t^+}{J}, \quad (6.17)$$

where J_t^- and J_t^+ represent, respectively, the number of investors with reservation prices at time t below and above the market clearing price at t (i.e. the bearish and bullish investors, respectively).

Proof: The expected price change conditional upon the prevailing reservation prices is, from eq. (6.14)

$$\begin{aligned}
 \mathcal{E} [\Delta p_{t+1} \mid P_t^*] &= J^{-1} \sum_{j=1}^J \mathcal{E} [\Delta p_{j,t+1}^* \mid P_t^*] \\
 &= J^{-1} \sum_{j=1}^J \{p_t + f(p_{j,t}^* - p_t) - p_{j,t}^*\} \\
 &= J^{-1} \sum_{j=1}^J f(p_{j,t}^* - p_t).
 \end{aligned} \tag{6.18}$$

Writing $x_j \stackrel{\text{def}}{=} p_{j,t}^* - p_t$, and using eq. (6.15),

$$\begin{aligned}
 \alpha J \mathcal{E} [\Delta p_{t+1} \mid P_t^*] &= \sum_{j=1}^J \text{sign}\{x_j\} (1 - e^{-\alpha|x_j|}) \\
 &= \sum_{j \in J_t^+} (1 - e^{-\alpha x_j}) - \sum_{j \in J_t^-} (1 - e^{\alpha x_j}) \\
 &= S_t^+ - S_t^-
 \end{aligned} \tag{6.19}$$

where S_t^+ and S_t^- are defined as the first and second terms in eq. (6.19), respectively, and J_t^+ and J_t^- double as the sets of bullish and bearing investors.

From the definition of S_t^+ and S_t^- we have the inequalities

$$0 < S_t^+ < J_t^+ \quad \text{and} \quad 0 < S_t^- < J_t^- \tag{6.20}$$

Multiplying the second inequality through by -1 and adding the two inequalities together gives

$$-J_t^- < S_t^+ - S_t^- < J_t^+$$

$$\Leftrightarrow -J_t^- < J \alpha \mathcal{E} [\Delta p_t \mid P_t^*] < J_t^+ \quad (6.21)$$

$$\Leftrightarrow -\frac{1}{\alpha} \frac{J_t^-}{J} < \mathcal{E} [\Delta p_{t+1} \mid P_t^*] < \frac{1}{\alpha} \frac{J_t^+}{J} \quad (6.22)$$

which completes the proof. ■

The interpretation of Proposition 6.2 is presented as the following corollary regarding the skewness of the reservation prices.

Corollary 6.3 *The covariance between the expected price change and the skewness of the reservation prices is positive, where skewness is defined to be proportional to the difference in the numbers of bullish and bearish investors.*

Proof: Skewness is measured in terms of the difference in the number of bullish and bearish investors, e.g. $(J_t^+ - J_t^-)/J$. In the absence of further information regarding the reservation prices, the expected price change may be taken to lie at the centre of the range defined in Proposition 6.2. This centre point will be zero if and only if the skewness is zero. Positive skewness will raise the lower bound towards zero and the upper bound away from zero, i.e. raise the centre point above zero; likewise negative skewness will decrease the centre point below zero. Therefore, for random drawings of reservation price vectors, skewness and the expected price change are positively related. ■

This corollary is deliberately stated in terms of covariances, rather than making the stronger assertion that positive skewness implies a positive expected price change. This is a consequence of expressing the expected price change as a range, rather than as an explicit function of the reservation prices. While the stronger

statement may be true, it is not necessary in order to justify one of the main contentions of this chapter, which is that knowledge of the reservation prices is an asset in making forecasts of price changes. In this case it has been shown that a simple count of bullish and bearish investors is informative. Further, as the local nature of the reservation price updating process implies a positive covariance between J_{t-1}^+ and J_t^+ , the two positive covariances together suggest that the expected price change might display sign-dependence.

6.5 Monte Carlo Simulations

The previous section has shown how the vector of reservation prices might determine aspects of the price/volume relationship, and presented a model for updating reservation prices which generalizes the random walk specification of Tauchen and Pitts. Proposition 6.2 and Corollary 6.3 show how this model will cause asymmetries in the reservation prices to affect the expected price change. In this section, the more general properties of this model are demonstrated by simulation.

The two parameters which need to be determined prior to any simulations are J , the number of active investors, and σ , the standard deviation of the disturbance term for updating reservation prices. I will set J to 103 and σ to 1.⁴ Simulations will be performed over values of $\alpha \in \{0.0, 0.1, 0.5, 1.0, 2.0\}$, remembering that α is a coefficient of the deviation of eq. (6.14) from the simple Tauchen and Pitts specification. This combination of σ and α covers a range of models from linear with high inertia ($\alpha = 0$) to non-linear with low inertia ($\alpha = 2$). In the latter case the bounds on the expected reservation prices at time $t + 1$ are $p_t \pm 0.5$, and the low inertia comes about because σ is large relative to the dispersion of reservation prices. If reservation prices were gas molecules in a box, $\alpha = 0.1$ represents a weak

⁴There is also the scale parameter c which appears in the trading volume expression, which is set to 1; the choice of $J = 103$ is justified below eq. (6.23).

shake between successive periods, and $\alpha = 2.0$ a strong one.

In order to describe the distribution of reservation prices at any time I will use the interquartile range (IQR) as a measure of dispersion and Bowley's coefficient of skewness:

$$\text{Bowley's coefficient} = \frac{2(Q3 + Q1 - 2M)}{Q3 - Q1}, \quad (6.23)$$

(see, e.g., Hoyle and Ingram, 1991, pp. 209–10), where $Q1$ and $Q3$ are the first and third quartiles, and M the median; the choice of $J = 103$ (as opposed, for example, to 100) is simply to make these fall on the observations rather than between them. There are clearly many other ways in which the reservation prices could be described, including the simpler definition of skewness given in the proof of Corollary 6.3. However, since the conclusions of the previous sections should be general with respect to descriptions of the reservation prices, these two robust and familiar measures will be used.

The objective of the simulation is two-fold. First, to examine the cross-sectional relationship between dispersion and skewness on the one hand and price change and trading volume on the other. Second, to examine the time-series relationship between the variables in the light of the cross-sectional evidence.

6.5.1 The Cross-Sectional Relationship

The important part about generating data on the cross-sectional relationship is to ensure that each observation is independently drawn. This rules out using a single time-series since it is expected that there will be features of the reservation price distribution which change only slowly through time, compromising independence. Therefore the following experimental design was chosen for each α : (i) run through 100 periods in order to 'season' the reservation price distribution; (ii) from the re-

sulting distribution generate 1,000 sets of reservation prices, each advanced another 20 periods; (iii) for each of these 1,000 sets, generate the next period's reservation prices; (iv) for each pair of consecutive periods, find the change in the market-clearing price, the trading volume, the first period dispersion and the first period skewness. The 20 period advance from the common base is to ensure that the resulting reservation price distributions are more-or-less independent. This gives a total of 1,000 observations on four variables, augmented to five by also including the absolute price change.

To analyse the resulting data, the standard and the partial correlation matrices of the various market quantities and descriptive statistics are calculated. The partial correlation coefficient identifies the unique relationship between two variables, as opposed to the standard correlation matrix which only shows the gross relationship (see, e.g., Mardia *et al.*, 1979; Whittaker, 1990). If S is the covariance matrix, the partial correlation matrix is found by scaling $-S^{-1}$ to have 1's in the leading diagonal.⁵ The standard and partial correlations are shown in Tables 6.1 and 6.2 for the different values of α . Only those values at least one standard error from 0 are shown, to give a visual key to the structure of the covariance matrix.

The first noticeable feature of the two tables is that there is little difference between the standard and the partial correlation coefficients. This indicates a very simple covariance structure consisting of separate bivariate relationships. The second noticeable feature is that increasing the non-linearity of the model (i.e. a larger α) causes the emergence of structure in the covariance matrix. This was anticipated in the previous discussion. Compare the first panel in Table 6.1 with the following ones. The first panel is the Tauchen and Pitts linear model. There is little or no relationship between any of the variables. In the next panel, a small increase in α from 0.0 to 0.1, corresponding to a small amount of non-linearity, causes a strong

⁵The application of partial correlation analysis in a portfolio framework similar to that of Epps and Epps (1976) [see Section 2.2], is discussed in Rougier (1995).

Table 6.1: Standard Correlations for Price Change and Volume

	Δp	$ \Delta p $	v	disp.
$\alpha = 0.0$				
$ \Delta p $	0.036			
v				
Disp.				
Skew.	-0.034			0.037
$\alpha = 0.1$				
$ \Delta p $	0.036			
v				
Disp.			0.271†	
Skew.				
$\alpha = 0.5$				
$ \Delta p $	0.033			
v		0.033		
Disp.		-0.065†	0.239†	
Skew.			0.057	0.038
$\alpha = 1.0$				
$ \Delta p $				
v		0.049		
Disp.		-0.045	0.227†	
Skew.	-0.095†		0.046	0.044
$\alpha = 2.0$				
$ \Delta p $				
v		0.044		
Disp.		-0.036	0.254†	
Skew.	-0.153†	-0.057		

† At least 2 standard errors from zero (only correlations at least 1 standard error from zero are shown). The standard error is $(n-1)^{0.5}$ under the null hypothesis that the correlation is zero (see, e.g. Kendall and Stuart, 1969, p. 396).

Table 6.2: Partial Correlations for Price Change and Volume

	Δp	$ \Delta p $	v	disp.
$\alpha = 0.0$				
$ \Delta p $	0.036			
v				
Disp.				
Skew.	-0.035			0.037
$\alpha = 0.1$				
$ \Delta p $	0.036			
v		0.031		
Disp.			0.272†	
Skew.				
$\alpha = 0.5$				
$ \Delta p $	0.032			
v		0.049		
Disp.		-0.075†	0.240†	
Skew.			0.048	
$\alpha = 1.0$				
$ \Delta p $				
v		0.061		
Disp.		-0.057	0.229†	
Skew.	-0.094†		-0.040	0.031
$\alpha = 2.0$				
$ \Delta p $				
v		0.054		
Disp.		-0.048	0.256†	
Skew.	-0.152†	-0.054		

See note to Table 6.1.

positive relationship between dispersion and volume. For $\alpha = 0.5$ there is additionally a negative relationship between dispersion and absolute price change, while at larger values of α there is additionally a strong negative relationship between skewness and price change.⁶

Overall, the two most striking features which arise from non-linearity are the positive relationship between dispersion and trading volume and, with greater non-linearity, the negative relationship between skewness and price change. These two correlations are of a different magnitude to the others. This pair of bivariate relationships, dispersion \implies expected trading volume and skewness \implies expected price change, accounts for the similarity of the standard and partial correlations.

At this point we can conjecture about the results from the time-series analysis that follows. Inertia in the distribution of reservation prices implies that the measures of dispersion and skewness will show some persistence. In this case, both trading volume and price change will have a small amount of positive autocorrelation which will be related to dispersion and skewness, respectively. The amount of autocorrelation will depend on the degree of non-linearity, as governed by α .

6.5.2 The Time-Series Relationship

The experimental design for this simulation is very straightforward: generate a single time-series of 1,000 periods of data after 100 periods of seasoning, recording price change, volume, dispersion and skewness for each period.

The resulting data can be analysed using a Vector Autoregression (VAR). The VAR approach was proposed by Simms (1980) as an alternative to structural time-

⁶This negative correlation between skewness and price change is clearly contrary to the positive skewness suggested in Corollary 6.3. The explanation is that Bowley's coefficient of skewness incorporates data on the reservation price vector beyond a simple count of bearish and bullish investors, giving particular weight to reservation prices that are closer to the median (and, therefore, typically closer to the mean). The general conclusion is that the relationship between skewness and price change is a complex one, and therefore sensitive to the precise way in which skewness is measured.

series modeling of the type promoted by the Cowles Commission. A VAR is effectively a reduced form estimated by ordinary least squares (OLS) in which the righthand side variables are contemporaneously uncorrelated with the disturbance. Typically these variables are exogenous variables and lags up to some order of the endogenous variables.⁷ In the VAR, $-\Delta p$, $|\Delta p_t|$, v_t , dispersion and skewness are all included as endogenous variables. The lag length was chosen according to the Schwartz Information Criterion. In all cases the result was a one-period VAR, with 30 coefficients (5 equations \times a constant plus 5 lagged variables). This is not surprising given that the updating model is a Markov process. The results are shown in Table 6.3.⁸

The first point to note is that when $\alpha = 0.0$ the VAR is misspecified. This is clear from the own-lag coefficients of dispersion and skewness, which are close to 1, and entirely unlike the coefficients in the other VARs, including $\alpha = 0.1$. This is not surprising, given that $\alpha = 0.0$ represents the Tauchen and Pitts model, and this model has already been shown to have a non-stationary reservation price distribution around the market-clearing price. In the non-linear model (i.e. $\alpha > 0.0$) the bounds on the expectations ensure that the reservation price distribution is stationary around the market-clearing price.

The salient features of the non-linear models in Table 6.3 are (i) a negative relationship between price change and lagged skewness, which becomes stronger as the non-linearity increases; (ii) a positive relationship between absolute price change and lagged dispersion at all levels of non-linearity; (iii) positive autocorrelation in volume at higher levels of non-linearity and a strong positive relationship between volume and lagged dispersion at all levels of non-linearity; (iv) positive autocorrelation in both dispersion and skewness at high levels of non-linearity; (v) exogeneity of dispersion and skewness.

⁷For a discussion of some of the methodological issues raised in VAR modeling, see Darnell and Evans (1990).

⁸It should be noted that tests for possible non-stationarity among the variables, as would usually accompany a VAR analysis, are not necessary since by construction the data are stationary.

Table 6.3: VAR Results for Price Change and Volume

	Const.	Δp_{t-1}	$ \Delta p_{t-1} $	v_{t-1}	Disp_{t-1}	Skew_{t-1}	Exog.
$\alpha = 0.0$							
Δp_t	0.086†			-0.002			1.467
$ \Delta p_t $			-0.036	0.002†	0.000		4.194**
v_t	41.114	1.358	5.924†		-0.010		3.990**
Disp_t	0.715	0.473		-0.014	0.996†		1.284
Skew_t			0.074		0.001	0.854†	1.421
$\alpha = 0.1$							
Δp_t	0.144†			-0.002†			2.706*
$ \Delta p_t $	0.060†		-0.032		0.035†		2.070
v_t	43.212†		3.356		7.858†		23.541**
Disp_t	1.381†	0.060	0.219†	-0.002	0.052		2.534*
Skew_t		-0.251†	0.192				2.855*
$\alpha = 0.5$							
Δp_t	0.128†			-0.002†			2.292
$ \Delta p_t $			-0.034		0.044†		3.896**
v_t	42.078†		4.505†		4.655†	-0.573	12.408**
Disp_t	1.413†		0.153	-0.002	0.075†		1.318
Skew_t	0.188				-0.068		0.706
$\alpha = 1.0$							
Δp_t	0.102†			-0.002†		-0.033†	3.556**
$ \Delta p_t $		-0.026			0.041†		4.639**
v_t	39.386†		-0.227†	4.830†	0.033	4.028†	12.540**
Disp_t	1.429†	0.090	0.146		0.093†		1.186
Skew_t	0.243			-0.003	-0.046	0.037	1.088
$\alpha = 2.0$							
Δp_t	0.081			-0.002†		-0.072†	11.070**
$ \Delta p_t $		-0.025			0.028†		3.145**
v_t	35.874†		3.700†	0.075†	4.430		17.629**
Disp_t	1.577†	0.091			0.138†		1.990
Skew_t		0.210†	-0.164			0.069†	0.548

† At least 2 standard deviations from 0 (only coefficients at least 1 standard deviation from zero are shown).

* Significant at a Type I error of 5% (** 1%). The test statistic is $F(4,994)$ under the null hypothesis that the variable is exogenous (i.e. not determined by lagged values of the other variables).

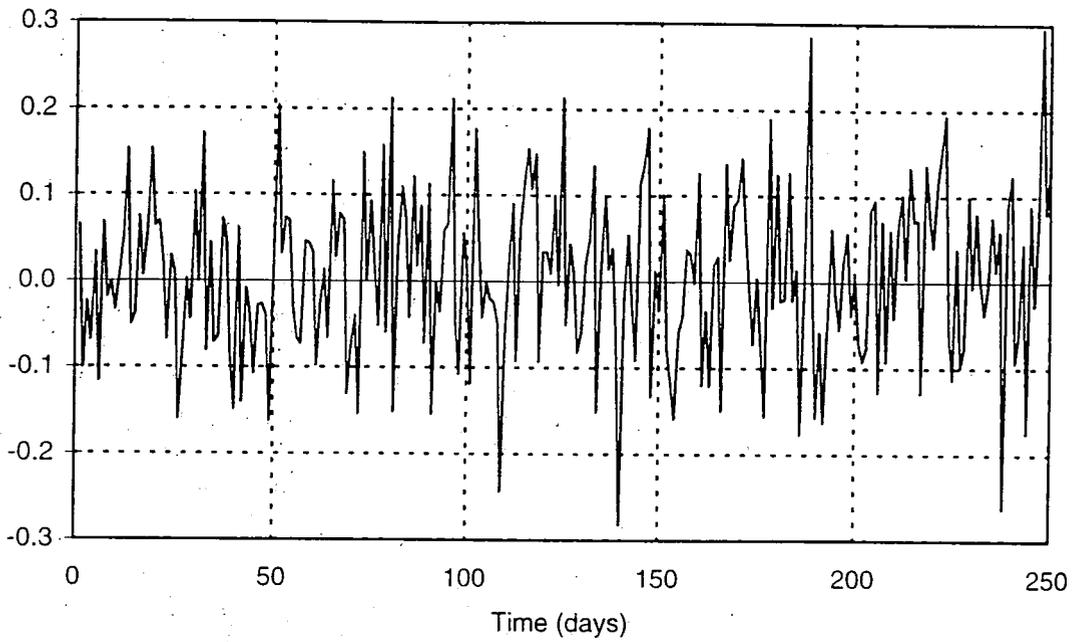
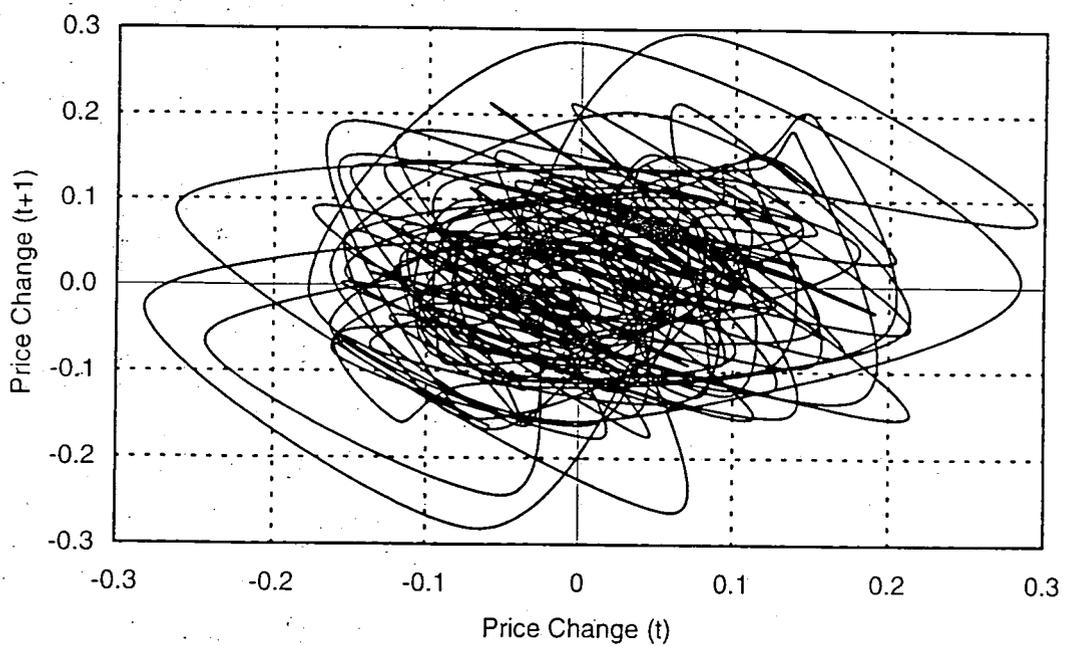
Finding (iv) confirms the conjecture that the local way in which reservation prices update causes inertia in the reservation price distribution, demonstrated here by persistence in both dispersion and skewness. Finding (iii) was conjectured in the previous subsection: inertia in dispersion causing autocorrelation in trading volume. Finding (i) has the negative relationship between skewness and price change, but not the autocorrelation that was a possible consequence of inertia in skewness. Finding (ii) was not anticipated, although there is some evidence from the cross-sectional analysis.⁹ One possibility is that although the correlation between dispersion and absolute price change is weak it is also robust, and so shows over time as forcefully as other relationships which are stronger but less robust. Finally, finding (v) identifies that the causality runs from the reservation price distribution to the price/volume relationship, as would be expected in a model of this type.

The main feature missing from Table 6.3 is the autocorrelation in price change and absolute price change implied by the autocorrelation in skewness and dispersion, respectively. In the light of Proposition 6.2, one explanation for this, as suggested above, is that the relationship between price change and skewness, although strong, is not robust. In other words the correlation can vary widely depending upon the reservation price distribution, and a value of 0.153 (from $\alpha = 2.0$) is only a midpoint of a range which might stretch down into negative values at certain times. This is investigated further in the next section.

6.6 Autocorrelation in Price Changes

The possibility is that the autocorrelation coefficient on price changes is not robust, i.e. it changes over time according to the reservation prices. As an initial check

⁹It is also interesting to note that there is both empirical evidence (Frankel and Froot, 1990) and theoretical evidence (Shalen, 1993) supporting the positive relationship between the dispersion of reservation prices and price volatility and trading volume, just as is illustrated here in points (ii) and (iii).

Figure 6.2: Price Changes from the First 250 Periods, $\alpha = 1.0$ Figure 6.3: XY Plot of Price Changes from the First 250 Periods, $\alpha = 1.0$ 

on this possibility, Figure 6.2 shows the first 250 periods from the simulation with $\alpha = 1.0$, which represents a midpoint in the trade-off between non-linearity and inertia, while Figure 6.3 gives the joined-up XY plot of the same data.¹⁰ These two Figures suggest that there is definitely something odd going on: the price change series seems to display conditional sign dependence which appears as a swirling pattern in the XY plot.¹¹

One frequently-used test for sign-dependence is the runs test, which compares the actual number of runs (i.e. sign changes, also known as 'reversals') with the expected number under the null hypothesis that at any point in time either a positive or a negative change is equally likely. Typically speculative prices display a higher than expected number of runs; one explanation attributes this to the behaviour of market makers (Niederhoffer and Osborne, 1966). The problem with runs tests in this context is that they do not capture the time-varying element. To a runs test, a period consisting of, say, a run of length 15 and then six runs of length 1 (i.e. seven runs in 21 periods) is the same as seven runs of length 3. Yet it appears that it might be the former pattern which is the more typical of the simulation data. Therefore I will use a test which is sensitive to this difference, by considering the distribution of run length.

Suppose the probability of a non-reversal is s ,

$$\Pr\{\text{sign}\{\Delta p_t\} = \text{sign}\{\Delta p_{t-1}\}\} = s, \quad s \in [0, 1), \quad (6.24)$$

¹⁰The XY plot is smoothed to be easier on the eye.

¹¹My thanks to Denis O'Brien for pointing out to me that this pattern appears to be chaotic about an attractor. Strictly, chaos is a feature of deterministic systems, and its presence in a stochastic system such as the model presented here would be hard to identify empirically (although see Brock, 1986). However, the two necessary conditions for chaotic behaviour, non-linearity and feedback, are both present in the model. In this respect it is reminiscent of the exchange rate models of De Grauwe and Vansanten (1990) and De Grauwe and Dewachter (1990), although in these models the authors make an explicit attempt to introduce the necessary conditions for chaos, rather than noting chaotic behaviour as a possible implication.

and consider the geometric random variable r which represents the length of a run,

$$\Pr\{r = i\} = s^{i-1} (1 - s), \quad i = 1, 2, \dots \quad (6.25)$$

The expected length of a run can then be found as $1/(1 - s)$.¹² Since the unconditional expectation of Δp_t is zero (i.e. in the absence of information regarding the reservation price distribution at $t - 1$), so $s = 0.5$ and we would expect the mean length of a run to be close to 2.0.

The observed and expected distribution of run lengths is shown in Figure 6.4a, for the complete sample of 1,000 periods using $\alpha = 1.0$. There are 489 runs in total giving a mean run length is 2.045. The expected distribution is under the condition $s = 0.5$. From the graph the observed distribution appears slightly leptokurtic relative to the expected: there are relatively more runs of length 1 and length 6 or more.

What about for runs which start with a large price change? To investigate this, the sample is divided into two halves around the quartiles, so that large changes are those which fall into the first or fourth quartiles. The two observed and expected distributions for the run lengths are shown in Figures 6.4b and 6.4c. The most striking feature of these two graphs is the excess number of short run lengths for the large price changes, and symmetrically the excess number of long run lengths for the small price changes. This suggests that there might be small amounts of negative autocorrelation following large price changes and positive autocorrelation following small ones.

¹²Proof:

$$\mathcal{E}[r] = \sum_{i=1}^{\infty} i s^{i-1} (1 - s) = (1 - s) \sum_{i=1}^{\infty} \frac{d}{ds} (s^i) = (1 - s) \frac{d}{ds} \left(\frac{s}{1 - s} \right) = \frac{1}{1 - s}.$$

■

Figure 6.4: The Distribution of Run Lengths

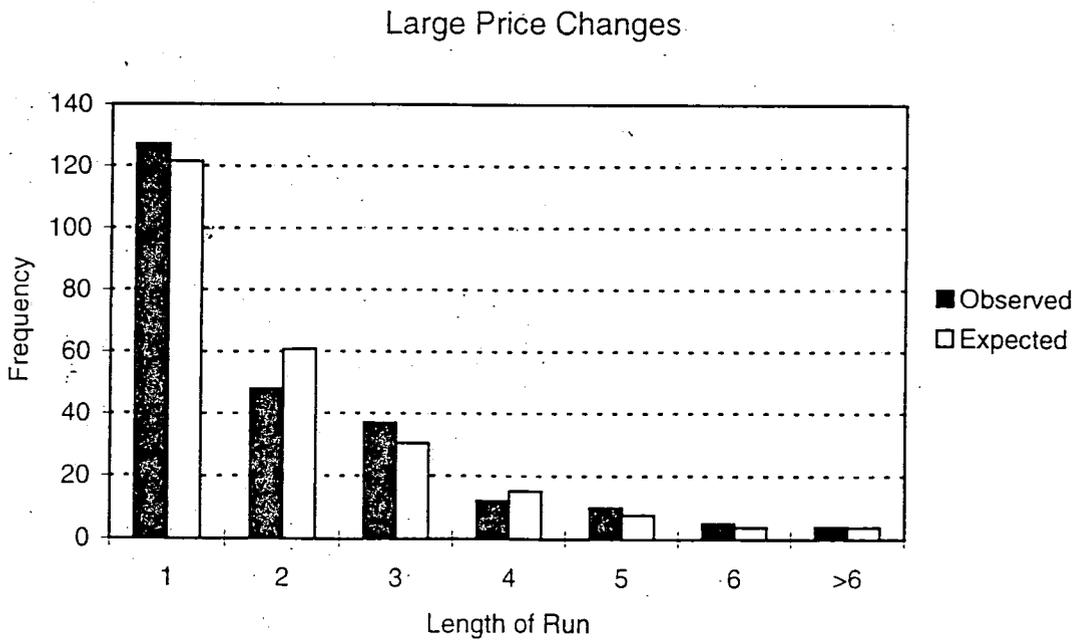
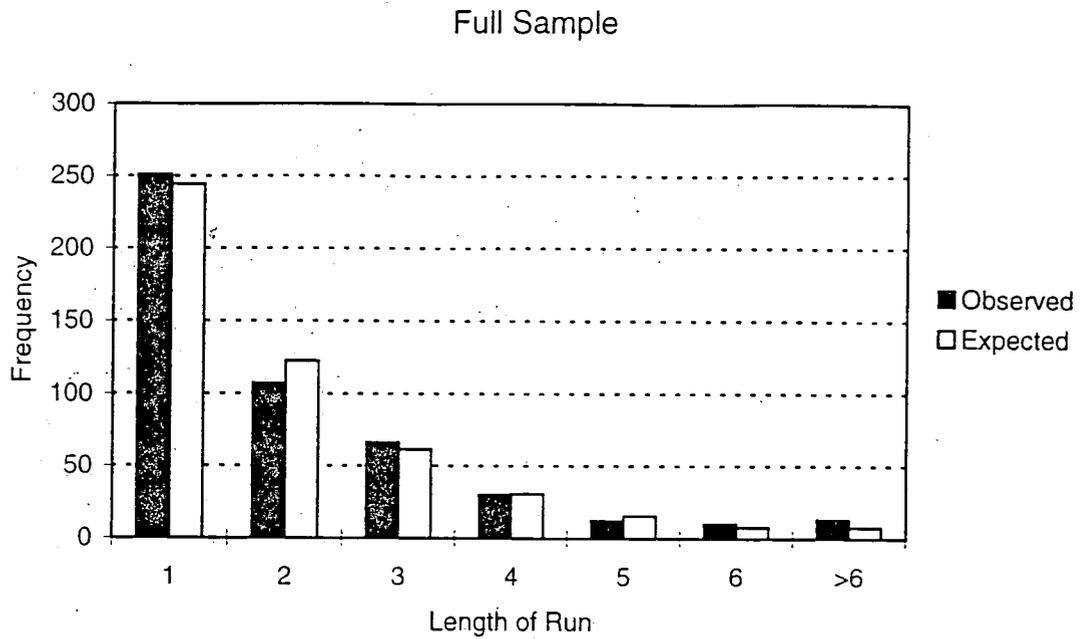


Figure 6.4: The Distribution of Run Lengths (cont)

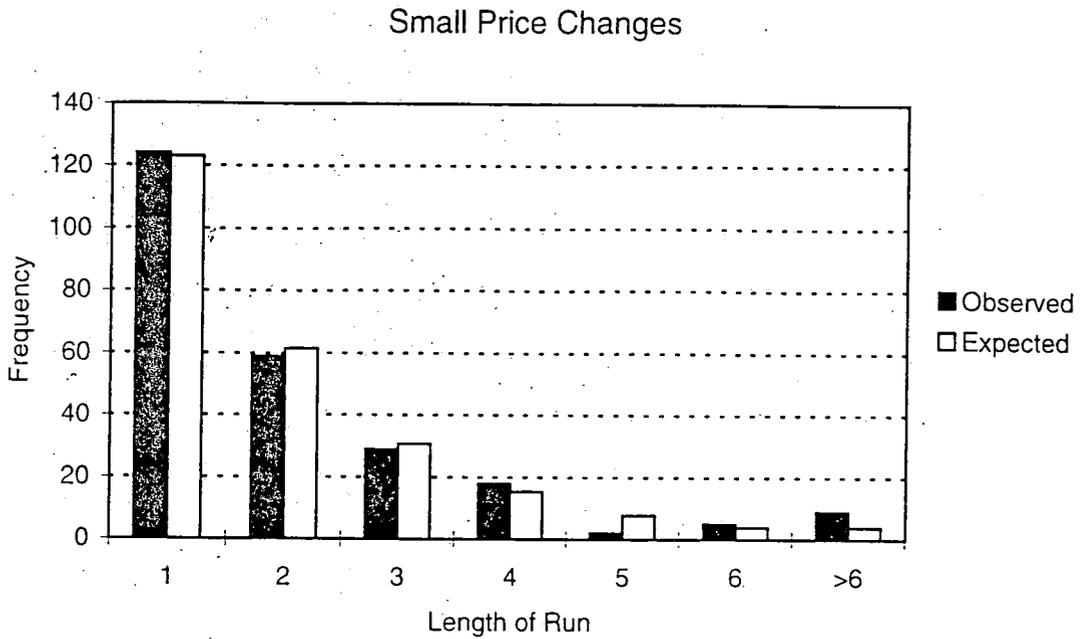


Table 6.4: χ^2 Tests of Run Length

	All Data	Large	Small	Independence
Simulation Data (1,000 observations)				
No. Runs	489	243	246	
Mean Length	2.045	2.012	1.988	
χ^2	7.667	6.193	12.130	10.575
p-value	[0.264]	[0.402]	[0.059]	[0.102]
FTSE-100, 1985-94 (2,528 observations)				
No. Runs	1,275	631	644	
Mean Length	1.983	2.021	1.946	
χ^2	2.139	5.155	4.406	7.672
p-value	[0.906]	[0.524]	[0.622]	[0.263]

For the three tests labeled 'All Data', 'Large' and 'Small', the null hypothesis is that the probability of reversal (s) is 0.5 (giving a mean run length of 2.0). For the independence test, the null hypothesis is that the run length distributions for initial small and large price changes are drawn from the same population. The p-values show the area to the right of the test statistic under the null hypothesis.

Table 6.4 gives the χ^2 statistics of the hypothesis $s = 0.5$, i.e. the reservation price distribution has little effect on the probability of reversal, and also the χ^2 test of independence between the distribution of run length and the size of the initial price change. None of the χ^2 tests quite breaches the classical threshold of significance at a Type I error of 5%. However, given the cost of a Type II error in this test, i.e. of asserting erroneously that there is effectively no time-variation in the probability of a reversal, a more generous Type I error might be appropriate, in which case the evidence does not rule out the possibility of some time variation.

Out of interest, Table 6.4 also reports the results of the same tests using the FTSE-100 data described in Chapter 5. The results from the FTSE-100 are in all cases more supportive of the null hypothesis than those of the simulation (i.e. the values of the test statistics are lower), but this is to be expected given that the futures market is likely to have several hundred active participants. As with the box of gas molecules, larger numbers lowers the magnitude of random fluctuations in density.

Finally it should be stressed that the simulations were performed completely without revision of any kind, and that the model chosen for updating reservation prices was simply the a straightforward generalization of the existing linear model. It is quite possible, therefore, that other specifications might generate much more extreme results. The purpose of the simulations is to demonstrate that such effects are possible, not to quantify the effects or to match them in any way to particular properties of speculative price data.

6.7 Summary and Conclusion

This chapter took as its starting point the inappropriate property of the Tauchen and Pitts model for updating reservation prices, the linearity of which caused the distri-

bution of reservation prices around the market-clearing price to be non-stationary. One implication of this model is that investors would disagree more and more through time, and hold larger and larger positions as a consequence.

The Tauchen and Pitts model was replaced by a generalized non-linear specification which included the linear expression as a limiting case. This general model has the property of being a random walk only when the investor's reservation price is equal to the market-clearing price. In other cases there is a tendency for reservation prices to be revised towards the market-clearing price. Further, in the generalized model the deviation of the expected reservation price in the next period from the current market-clearing price is bounded, ensuring that the distribution of reservation prices around the market-clearing price is stationary. The generalized model implies, through the mechanism of market-clearing, that the cross-sectional distribution of reservation prices at time t will enter into the distribution of market-clearing price change and trading volume over the period $(t, t + 1]$. It is shown, for example, that the expected price change is determined by the skewness of reservation prices. It was suggested that inertia in this distribution might introduce a time-series element into the price change and trading volume distributions, in accordance with observation.

In order to examine this conjecture Monte Carlo simulations were used at different degrees of non-linearity. In general, there was unambiguous evidence of a relationship between the skewness of the reservation price distribution and price change, and between the dispersion of the reservation price distribution and trading volume. Skewness and dispersion were also shown to have time-series properties, but the simulations did not show that these properties fed through to generate time-series properties in price changes. Trading volume, on the other hand, showed strong positive autocorrelation. A further investigation of one set of price change data suggested graphically and statistically that the hypothesis that the probability

of a sign reversal was equal to 0.5 irrespective of time or the size of the change (and, by extension, irrespective of the distribution of reservation prices) was not unambiguously accepted.

The conclusion of this analysis may be stated as a syllogism: (1) if reservation prices update linearly, they have no impact on the price change and trading volume time-series; (2) reservation prices are unlikely to update linearly; (3) simple simulations with non-linear updating generate complex price/volume dynamics; (4) therefore a possible cause of complex price/volume dynamics is non-linear updating of reservation prices.

Finally it should be stressed that the model as it stands, even if augmented by a dynamic news-arrival process, cannot be a completely satisfactory description of the price/volume relationship for the reasons discussed in Subsection 1.4.4: it is irrational for risk-averse investors to participate in zero-sum games, and yet this is the nature of investing if all investors work from a public information stock. The model is an explanation of the way in which the rich dynamic behaviour of the price/volume relationship can be generated within a simple optimizing model.¹³ Insofar as the model encompasses the explanations of ARCH effects advanced by Bera and Higgins (1993) it rescues ARCH from being purely descriptive, but it does not provide the kind of structural framework desired by Gallant *et al.* (1992). Rather the model shows that, with the constantly shifting reservation price distribution, we should not expect to be able to pin down the price/volume relationship to a process with fixed coefficients.

¹³As an interesting aside, this model has much in common with rational expectations real business cycle models. In such models the optimizing behaviour of heterogeneous agents transforms a white-noise input into a non-white-noise output (see, e.g., Hillier and Rougier, 1996). A much earlier business cycle model of a similar nature was Schumpeter's celebrated 'ticking clock on a wobbly table'.

Part II

Return Autocorrelation

Chapter 7

The Role of Market

Microstructure

This chapter explains how market microstructure is the likely cause of one of the main stylized facts about the daily return distribution. Market microstructure means seeing the market not as an organic whole, but rather as a forum within which a large number of not-necessarily-identical agents interact: "... market microstructure treats the interplay between market participants, trading mechanisms, and the dynamic behaviour of security prices in a regime where friction impedes the trading process" (Cohen *et al.*, 1980, p. 249).

The stylized fact is mean reversion in prices. Mean reversion shows up as positive low-order autocorrelation and negative high-order autocorrelation in returns (see, e.g., Poterba and Summers, 1988). Positive first-order autocorrelation in returns has been known to be a feature of daily returns since at least Fama (1965), and recent evidence has been summarized in Fama (1991). Corrado and Lee (1992) calculate the mean daily autocorrelation coefficient over 120 large stocks, each calculated over the period 1963–1989, to be 0.059, with a t-statistic of 4.63. Similar magnitudes for large stocks are found in French and Roll (1986); Brock *et al.* (1992), larger magnitudes

are found by Campbell *et al.* (1993), who also note that the autocorrelation may not be independent of the day of the week.

The evidence for high-order negative autocorrelation in returns is much weaker, unsurprisingly given that most financial time series only go back, at best, to the first part of the century. Fama and French (1988) find some evidence for values of about -0.25 for three year lags, but these estimates are particularly dependent upon the pre-second world war period. The evidence from Poterba and Summers (1988) is similarly ambiguous. An alternative approach is to examine the relative performance of winner and loser portfolios. deBondt and Thaler (1985, 1987) find that stocks identified as losers over a three year period subsequently outperform the market, while winners subsequently underperform. However, this result could be explained by a rise in the risk premium demanded by investors holding underperforming stocks (Zarowin, 1989).

Therefore although we may treat mean-reversion as a stylised fact, it is the positive first-order autocorrelation in returns which is the more pervasive aspect.

7.1 The Martingale Theorem

One of the interesting things about the positive first order autocorrelation in daily returns is that under two simple conditions it should not exist. This is the implication of the Martingale Theorem of Samuelson (1965, 1973). These two conditions are:

1. All investors are identical (i.e. share a common information set which they each interpret in the same way, and have a common rate of time-preference);
2. All investors are risk-neutral.

Under these two conditions the price of each asset will be bid to the point at which the expected return on the asset is the same as the risk-free interest rate. This

interest rate will be the same as the (common) rate of time preference, hence every asset satisfies the property that

$$\mathcal{E} [r_{t+1} \mid \Omega_t] = \delta, \quad (7.1)$$

where δ is the rate of time preference, Ω_t is the information set at time t , and the return is defined

$$r_t \stackrel{\text{def}}{=} \frac{p_t + d_t}{p_{t-1}} - 1, \quad (7.2)$$

where p_t is the price at time t and d_t is the dividend for period t , which is received at the end of the period. Samuelson shows that eq. (7.1) and eq. (7.2) imply that price at time t is the discounted value of the expectation of price plus dividend at time $t + 1$:

$$p_t = (1 + \delta)^{-1} \mathcal{E} [p_{t+1} + d_{t+1} \mid \Omega_t]. \quad (7.3)$$

The appellation ‘Martingale Theorem’ arose from the 1965 paper, in which Samuelson considered the special case of eq. (7.3) where $\delta = 0$ and $d_{t+1} = 0$. In this case, $\mathcal{E} [p_{t+1} \mid p_t, p_{t-1}, \dots] = p_t$, making prices a Martingale process, and returns a ‘fair game’.

By forward substitution for p_{t+1} , eq. (7.3) is shown to be equivalent to the discounted cash flow expression

$$p_t = \sum_{i=1}^{\infty} (1 + \delta)^{-i} \mathcal{E} [d_{t+i} \mid \Omega_t], \quad (7.4)$$

provided that $\lim_{n \rightarrow \infty} (1 + \delta)^{-n} \mathcal{E} [d_{t+n} \mid \Omega_t] = 0$. A sufficient condition for this convergence is that dividends are expected to grow in the long term at a rate less

than δ .¹

One feature of Samuelson's model is that it would be very easy to test eq. (7.1) using ANOVA across the ex post returns of different asset classes. In fact we know that this test would show systematic differences in the ex post returns of the assets both across asset classes (e.g. equities return more than the risk-free rate) and within asset classes (e.g. small company stocks return more than large company stocks, see Banz, 1981). Therefore since this implication of the Martingale Theorem does not hold, at least one of Samuelson's assumptions must be wrong.

7.2 Risk Aversion

Samuelson (1965) mistakenly believed that the risk-neutrality assumption was not critical, and could be accommodated by a risk premium. However, the possibility that this premium might be time-varying destroys the Martingale property (LeRoy, 1989). Consider the case where the variance of returns is positively autocorrelated over time, so that a large absolute return in one period makes it more likely that the return in the next period will also be large. In this case risk-averse investors will bid down asset prices following a large absolute return, so that in future periods the expected return is higher. The opposite pattern occurs following a small absolute return. These effects lead to a pattern in returns which looks a bit like positive first order autocorrelation with spikes.² This example is relevant because of the strong

¹The practice of valuing stock using a discounted cash flow method such as eq. (7.4) is one of the tenets of *fundamental analysis* (see, e.g., Williams, 1938). The number of inputs to the model is usually much reduced by presuming a constant growth rate for dividends, often related to the dividend yield and the payout ratio (Gordon and Shapiro, 1956; Gordon, 1962). Therefore in a sense the discovery of Samuelson's pricing model 'legitimized' the existing valuation model.

²These spikes, representing the one-off adjustment to the new regime, are sometimes known as the *discount rate effect*. Since they operate in the opposite direction to the change in the required return, it can be hard to pin down the time-varying properties of the ex ante mean return. This may be one explanation for the lack of success of time series techniques such as GARCH-M (see, e.g., Bera and Higgins, 1993), where the volatility appears as a variable in the mean daily return. Fama (1991) advances a similar argument about our inability to distinguish irrational bubbles from time-varying expected returns.

evidence suggesting that variances are positively autocorrelated over time, dating from Fama (1965) and previously discussed in Chapter 2, particularly Section 2.6.

One enduring model that explains systematic differences in returns across and within asset classes is the Capital Asset Pricing Model (CAPM) originating with Sharpe (1964). This model takes as axiomatic that investors are risk-averse. Although the testability of the CAPM has been extensively questioned (see, e.g., Roll, 1977), its widespread use in portfolio management suggests at least tacit acceptance of the axiom of risk aversion among practitioners. There is also direct evidence of risk-aversion from attempts to fit utility functions consistent with investors' choices (see, e.g., Friend and Blume, 1975; Blake, 1996).³ Therefore the CAPM may be used to tie in the failure of prices to be a Martingale with the fact of investors' risk aversion.

General Equilibrium

A more direct route can be found in the general equilibrium analysis of Lucas (1978). In Lucas's model the equilibrium asset price turns out to be the same as that of the Martingale model, eq. (7.3), *only* in the case of risk-neutrality (Lucas, 1978, p. 1434, eq. 6). LeRoy (1989) notes as an implication of Lucas's model that in production economies (as opposed to exchange economies) the possibility of corner solutions will affect the simple relation between risk-neutrality and the Martingale property. Therefore the conclusion from general equilibrium analysis is that, even with identical investors, risk-neutrality is probably necessary (but not sufficient) for asset prices to have the Martingale property.

Therefore it seems that if we rule out risk-neutrality we also rule out the Martingale property of asset prices. This raises an interesting question: What is being

³There are also, of course, the various paradoxes of human behaviour that are usually solved by positing risk-aversion. For a fascinating discussion of perhaps the most famous of these, the *St. Petersburg Game* of Nicolas Bernoulli, see Fellner (1965), particularly the Appendix to Chapter 3.

tested in a regression of r_{t+1} on r_t ? In other words, what is our conclusion when we find, as we do, a significant autocorrelation demonstrating a violation of the Martingale property? Fama (1991) calls this the *Joint Hypothesis Problem*. Any hypothesis test is a joint test of a valuation model and the assumption that the behaviour underlying the valuation model can be implemented at zero cost by the marginal market participants. Even were the Martingale assumptions to hold, the lagged return might be significant on account of transaction and information costs. But since investors are unlikely to be risk-neutral and risk-neutrality is necessary (but not sufficient) for the Martingale property, we should be completely unsurprised by the existence of significant autocorrelation.

7.3 Heterogeneous Investors

To summarize the previous section, we do not expect prices to be a Martingale because of the existence of transaction and information costs, and even in the absence of these costs we would not expect prices to be a Martingale because investors are typically risk-averse. This non-Martingale conclusion is true even when all investors are supposed to be identical. In this section, this heterogeneity condition is relaxed as well:

Noise

The starting point is Black's observation "Noise makes financial markets possible, but it also makes them imperfect." (Black, 1986, p. 530). Black contrasts 'noise' with information, and classifies market participants at any time as 'noise traders' and 'information traders'. Over time most investors will have traded on both information and noise; sometimes they will have traded on noise in the belief that it is information, at other times they will have traded simply because they enjoy

trading.⁴ Without noise speculative markets would not function, because the desire to trade would then be a product of superior information only and there would be no one to take the other side of any trade (see, e.g., Milgrom and Stokey, 1982; Tirole, 1982). However, with noise the price of an asset may differ from its value, since trades motivated by noise will still affect prices. It is this difference between price and value that motivates investors to seek out costly information, which they can then use to 'put one over' the noise trader at the other end of the trade.

When noise is seen as an integral part of a financial market, the notion of *market efficiency* must be reconsidered. Black chooses to define an efficient market as one in which price is within a factor 2 of value; he asserts that this would mean that markets are efficient in practice about 90% of the time. In Black's model the relation between price and value is like a piece of elastic of a certain length (this is my analogy, not Black's). When price and value are close the elastic is slack, and there is no tendency for the two to move any closer: "All estimates of value are noisy" (Black, 1986, p. 533). Hence within the slack region there will be noise traders on both sides of each trade and no net effect on price relative to value, so price will be a simple diffusion process. If price and value are further apart there is a pressure for price to move towards value which is crudely proportional to their separation. As price moves into the elastic region the more aggressive information traders start to join in but all on the same side. Whether they initially prevail against the noise traders depends upon the price's momentum as it enters the elastic region, and this in turn depends upon the precise way in which the noise traders form their demands. To complicate the issue further, value itself is not constant in time, and so any reversion that takes place is towards a continually shifting target.

Black also makes the point, alluded to above, that few investors can know for sure that they are trading on information rather than on noise (the exception might

⁴To expand slightly on this latter reason, fundmanagers may be obliged to trade by the prerequisites that they consume, supplied by stockbrokers.

be an insider with exclusive information, although he or she would ultimately be spotted). To use the standard portfolio management cliché, the good investors probably trade on information 55% of the time.⁵ Formal models of noise trading tend to dichotomize investors as either noise-based or information-based. These models may be interpreted as modeling states of behaviour rather than investors, i.e. a good investor has a 55% chance of behaving in an information-based manner.⁶

Mimetic Contagion

One recent model of heterogeneous investors which is able to explain the stylized facts described in the introduction is that of Lux (1995), which is based upon the notion of *mimetic contagion* among noise traders. In this model a noise trader becomes more willing to buy (respectively, sell) if he sees other traders buying (selling). This behaviour is not necessarily irrational, although this is certainly one explanation. First, a trader without information may be under the impression (correct or erroneous) that the market is generally slow to assimilate fully new information.⁷ Second, even traders with superior information may choose to follow the majority since in doing this their reputation cannot suffer.⁸ Lux also introduces a class of information traders with demand proportional to the difference between price and value and a marketmaker who causes prices to move in the same direction as net excess demand.

⁵This means that their identity will be very hard to infer from the performance of their portfolios. It also means that the bad investors will take a very long time to go bust and withdraw from the market. Therefore evolutionary explanations of market behaviour are not appropriate within the timescale of institutional stability.

⁶For examples of models of this type see, e.g., Shiller (1984); Grossman and Miller (1988); deLong *et al.* (1989, 1990); Campbell and Kyle (1993).

⁷This is sometimes known, for obvious reasons, as the 'greater fool' theory/fallacy.

⁸This is certainly the case in the highly competitive world of fund management, with its quarterly appraisals and its independently assessed quartile rankings. In the absence of any information portfolios are typically adjusted to a benchmark representative of the median fund, this being the position of least risk to the managers' reputations.

Lux shows that this model gives rise to a rich variety of dynamic behaviour. One possibility is two stable equilibria in which the positive (respectively, negative) impact of bullish noise traders is exactly offset by the negative (positive) impact of information traders. Another is cyclicity where all initial trajectories converge on a periodic orbit in which the net sentiment of the noise traders alternates between bullish and bearish. When Lux goes on to introduce an endogenous sentiment factor based on actual returns compared with expected returns (assumed to be constant) he finds that his model also generates crashes: "Once infection has reached the overwhelming majority of speculative traders, a change in basic sentiment occurs because *the exhaustion of the pool of potential buyers causes price increases to diminish*" (Lux, 1995, p. 893, original emphasis).

This model explains mean reversion as speculative overshooting. A small deviation from value can become amplified by the mimetic contagion of noise traders. The correction induced by a change in sentiment tends to push prices too far in the opposite direction.⁹ Lux stresses that the process of mimetic contagion is widespread in the social and natural sciences; in other words he is importing into finance a general paradigm which happens to explain stylized facts in financial markets, rather than proposing a model which is sufficient for the same.

In conclusion of this section, the heterogeneity of investors appears to be an important part of the functioning of a financial market. Moreover, models with heterogeneous investors, such as that of Lux, are capable of generating many of the price dynamics which we observe (which is, of course, only a necessary condition for validity). This then is a second blow to the Martingale model: not only are investors risk-averse, but they are also heterogeneous. Consequently, non-Martingale behaviour in prices should be established from theory, a fact which is confirmed by empirical studies.

⁹Likewise *excess volatility*, another stylized fact (although not one discussed in this thesis), arises because of the cyclicity of price around value.

7.4 Trading Rules

In the two previous sections it has been shown that, since investors are neither risk-neutral nor homogeneous, there is no reason to suppose that speculative asset prices follow a Martingale process, as proposed in Samuelson's original Martingale Theorem. On the contrary, there is a great deal of evidence that returns may be in part predictable from the prevailing information set. One instance of this has already been presented as a stylised fact: the mean reversion of prices which appears as positive low order autocorrelation in daily returns. There is also a large and developing literature on 'anomalies', which are instances where elements in the information set appear to have significant predictive power. Popular contenders include calendar events, dividend yields and price earnings ratios, and money supply growth (see, e.g., Fama, 1991).

In the microstructure approach, returns are in part predictable because at the very least there is no reason why they should not be. There is no class of investor in a position to 'arbitrage away' the predictive power of, say, past returns. One reason for this lack is that in practice such trading would not be arbitrage (i.e. riskless) since the predictive power of these anomalies is typically small, and investors are not risk-neutral. A second is the presence of transactions costs, both direct and indirect. Direct costs comprise the stockbroker's commission, the bid-ask spread (the 'touch') and in some countries the purchase tax (Stamp Duty in the UK). The indirect costs (sometimes known as 'implicit costs') are the opportunity cost of funds and the market impact, which for large transactions pushes the price at which assets can be bought or sold in the non-profitable direction. Hibbert (1995) puts the total cost of a round trip in the UK at around 1.5% plus opportunity cost, for large institutions in highly liquid stocks (i.e. FTSE-100). For smaller but still liquid stocks (i.e. FTSE mid 250) this figure would be more like 10% plus opportunity cost since the touch

alone can be 5% or more.

What of an investor attempting to profit from exploiting the non-Martingale behaviour of returns? Consider the case of using historic return data alone as the basis for a trading rule. This is analagous to the practice of 'technical analysis', which remains much in vogue among professional investors. An investor using such a rule is clearly a noise trader, since he has no informational advantage over other traders. Therefore we would expect that in the long run he will lose his money since he will, sporadically but inevitably, come up against information traders. The problem faced by the investor is that he does not know in what relation price currently stands to value. Therefore he must implement his strategy both in the slack and in the elastic regions discussed above. In the slack region he neither gains nor loses money (on average) on his buying and selling price, since he trades only with other noise traders. When transactions costs are incorporated he loses at least 1.5% per round trip, according to the figures given above.

Suppose the investor is lucky enough to be long when one of Lux's positive mimetic contagions catches on and pulls him into the elastic region in which information traders are the sellers. He will have a long run of small positive gains but ultimately he will be wiped out by the crash which according to Lux's model will take prices through the slack region and into the elastic region on the other side of value. If he is still trading by the same rule he will now stay short for another run of small positive gains (with information traders being the buyers) but then be wiped out again, and so on. Since these crashes are relatively rare, the median return from the trading rule may well be positive, even allowing for transactions costs, but the mean return will be negative. Unfortunately for the investor, his long run performance is determined by the mean return, not the median. When transactions costs are incorporated, the long run performance will be more negative.

The Filter Rule

This negative skewness which can arise in the returns (i.e. the discrepancy between the mean and the median) must be borne in mind when analysing the performance of trading rules based on noise. The classic trading strategy is the *filter rule* (Alexander, 1961, 1964). In this rule the trader who is in the market stays in until the price has fallen by a proportion x from its high since he has been in, and a trader who is out of the market (either short or invested in the risk-free asset) stays out until the price has risen a proportion x from its low since he has been out. The usual way to implement this rule would be to determine the amount x historically as the value which maximizes trading profit.

In the absence of transactions costs, filters of size $0.0025 \leq x \leq 0.0050$ appear to be able to generate profits through a very large amount of trading. Sweeney (1988) found that at their typical level of transactions costs, floor traders in DOW Jones industrials might profit from such rules.¹⁰ Generally, however, the incorporation of transactions costs totally swamps the excess return that might be achieved through high levels of trading. Corrado and Lee (1992) find that transactions costs of as little as 12 basis points (0.12%) eliminate the difference between a filter rule and the performance of a buy-and-hold portfolio. Costs of the order 1.5%, as mentioned above, would make this filter rule ruinous in operation.

A final problem with the assessment of filter rules in practice is that the samples over which the filters are tested are often too small to have a representative from each of the tails of the return distribution. Hence the mean of the performance of the rule over a sample has more in common with the population median than the population mean. As mentioned above, the population has a large amount of negative skewness so that the median, and consequently the sample, will tend to

¹⁰Sweeney (1986) presents similar evidence from the foreign exchange market.

overstate the profitability of the rule.¹¹

7.5 Summary

In Chapter 6 it was suggested that a pseudo-homogeneous investor model could generate interesting price/volume dynamics. This chapter has adopted a more classical approach in showing that the conditions under which we would expect prices to be a Martingale are not fulfilled.

These findings agree with the empirical evidence of generally positive first-order autocorrelation in daily returns. But the smallness of the autocorrelation coefficient (i.e. the small proportion of return variance explained by past returns) led Fama to brand this effect “insignificant from an economic viewpoint” (Fama, 1970, p. 394). Subsequent work has confirmed that it is very hard to make a trading profit on the basis of filter rules, just as it is very hard to make a profit exploiting other types of anomaly. Often, where a trading profit does appear it can be explained by a change in riskiness or an insufficiently large sample.

However, this does not mean that the autocorrelation can have no useful role to play in investors’ behaviour. In the next chapter it will be shown how autocorrelation can be used as the basis of a timing rule for investors already committed to buying an asset and holding it for a substantial period. Usually, the final (generally insurmountable) hurdle to profiting from the daily return autocorrelation are the transactions costs. But in the case where the only issue undecided is the timing of a purchase to which the investor is already committed, transactions costs are immaterial. It is likely that the rewards from a timing strategy are small on a

¹¹I might also mention an acknowledged problem in the literature concerning trading rules and financial markets—the preference of journal editors for studies refuting the hypothesis of market efficiency. This has the unfortunate effect of burying a large number of studies, such as those that show trading rules producing insignificant or negative returns in out of sample periods. My own experience and the experience of students that I have supervised suggests that this is at least as likely as finding a positive return.

per-transaction basis. But like gravity they will accrue cumulatively and so could make a substantial contribution to portfolio performance over the course of a year. Following Chapter 8, which develops the theory of optimal timing to a successful conclusion, Chapter 9 considers the magnitude of the reward from following such a rule, in order to determine whether in this context the non-Martingale property of prices is still *economically* insignificant.

Chapter 8

The Theory of Optimal Timing

8.1 Introduction

The starting point of this chapter is the presence in daily speculative asset returns of a small but significant amount of generally positive first order autocorrelation. This autocorrelation has persisted since being widely-publicised in Fama's influential paper on efficient capital markets (Fama, 1970); in his follow-up twenty-one years later, Fama notes " ... research is able to show confidently that that daily and weekly returns are predictable from past returns" (Fama, 1991, p. 1580).

In Chapter 6 it was shown that this autocorrelation might arise naturally within pseudo-homogeneous investor models, as a result of the way in which beliefs were updated with reference to the prevailing market-clearing price. The alternative explanation was examined in Chapter 7: heterogeneous investors and 'noise trading'. Crudely, 'noise' traders chase prices away from value, to the point at which it becomes profitable for 'fundamental' traders to step in. Therefore daily returns tend to move in an autocorrelated manner while prices remain within a range determined by the transactions costs of the lowest cost fundamental trader. As prices move beyond this point the pressure to reverse increases as more and more fundamental

traders step in. The situation can be complicated by the fundamental traders having different notions of value.

An often-asked question is whether a third group of traders can profit by designing a strategy around this pattern of autocorrelation. Generally, empirical tests of such strategies confirm their profitability (relative to a buy-and-hold strategy), but also show that the presence of transactions costs usually more than offsets the benefits from trading. In Fama's phrase, the autocorrelation is *insignificant from an economic point of view*.

This chapter and the following one challenge this view by considering the needs of investors with stock-holding horizons of a year or more. Because of this long horizon, these investors are flexible, to a degree, about the *precise purchase date* once the decision has been taken to commit a certain amount of capital to a certain stock. They can use the autocorrelation to squeeze an extra few shares out of their allocated capital by sometimes waiting to see if prices fall. The transactions costs that stymie attempts to trade directly on the autocorrelation are not a factor in this case, since the decision to trade has already been taken; the only issue remaining is 'when?'

This is the problem of 'optimal timing'. The general optimal timing problem is described in Section 8.2, and for the case of AR(1) returns in Section 8.3. Section 8.4 derives sufficient conditions for a solution to the optimal timing problem, and Section 8.5 uses these to solve the problem for a function describing the rewards from following the optimal timing strategy. This function turns out to be extremely complicated and its computation is discussed in Chapter 9.

8.2 The Timing Problem

An investor has made the decision to buy a stock and has been assigned a given amount of capital for the purpose. In other words, everything has been done except actually purchase the stock, including the assessment of beliefs and the portfolio optimization. In fact, it is appropriate to think of the investor in this scenario as a specialist trader, as indeed there are such specialists in the large fund management firms. Their job is to take the needs of the individual portfolio managers, satisfy them internally if possible (i.e. by the simple transfer of stock from one portfolio to another) and then make the net purchases and sales through the stockbroker that provides the most cost-effective service.¹

Optimal Stopping

The investor's simplest option is to make the purchase at the first opportunity, and this may be explicit in his role. Suppose, however, that he has some flexibility about the precise timing of the purchase. As well as an immediate purchase, he has the alternative of waiting to try and get a better (i.e. lower) price at a later date, while picking up some interest on the capital in the meantime. This is an *optimal stopping problem* because the investor is constantly having to reaffirm his decision to wait, right up until the moment that he purchases the stock. Therefore in choosing whether or not to purchase today, he is comparing the known reward of immediate action with the value of the opportunity to take the same decision tomorrow.

Optimal stopping problems have a wide, if fairly recent, provenance in Economics. For example, one way of analysing voluntary unemployment is by job-search models, in which the agent compares a known and immediately available job with

¹As an interesting aside, the task of these specialists is to assign deals across stockbrokers not just by cost but also to reflect the institution's consumption of each stockbroker's research services (and, inevitably, each stockbroker's provision of perquisites). Unless the trading of stocks is centralized (i.e. taken away from the fundmanagers), monitoring these activities is extremely hard and the system is open to abuse.

his expectation of the reward from waiting for a period to see what comes up. In this way the agent is constantly reaffirming his decision to remain unemployed up until the point at which taking the offered job seems the most attractive option (see, e.g., Sargeant, 1987, for some simple models of this type).

Another example is the analysis of investment opportunities under uncertainty. Since investments are often irreversible, a mis-timed investment can be very expensive. The agent here is constantly reaffirming the decision to delay the investment up until the point at which the cost of delaying (lost profits and the danger of a competitor entering the market first) exceeds the benefit of waiting (mainly extra information leading to reduced uncertainty). This type of analysis is used extensively by Dixit and Pindyck (1994), who consider an investment opportunity to have an option value which is killed at the point at which the investment goes ahead. Consequently Net Present Value (NPV) calculations should include this loss of option value among the outgoings of the first year of the project. If this option value is ignored it appears as if businesses are using too high a discount rate in project appraisal, which is in fact a commonly-held view (among economists). Dixit and Pindyck also suggest this as a reason for many businesses choosing the payback period method of project appraisal rather than the theoretically superior NPV approach.

The Reward From Optimal Timing

Since the future is uncertain, the reward from following a timing rule based on optimal stopping analysis, known here as an *Optimal Timing Rule* (OTR), is uncertain at the point of implementation. Therefore the reward is a random variable, defined as follows.

Definition 8.1 (Reward From Optimal Timing) *The reward from implementing an Optimal Timing Rule (OTR) is the random variable $\tau \in R_{++}$, where R_{++} denotes the positive real line, representing the factor by which the number of shares*

purchased is increased over the alternative of immediate purchase. The function $f : \Omega_t \mapsto R_{++}$, where

$$f(\Omega_t) \stackrel{\text{def}}{=} \mathcal{E} [\tau \mid \Omega_t] \quad (8.1)$$

and Ω_t is the information stock at time t , will be known as the reward function, and it represents the expected reward from implementing the OTR at time t .

To illustrate Definition 8.1, suppose the investor has K capital and the price of the target stock is currently p_t . If the investor purchases immediately, he receives K/p_t shares. If he decides to implement the OTR he receives $(K/p_t) \times \tau$ shares at some point in the future; the expected reward at time t is $(K/p_t) \times f(\Omega_t)$.² In general, i.e. without reference to a particular point in time, the unconditional expectation of the reward from the optimal timing rule can be found, in principle, by integrating the reward function over the distribution of Ω_t , by the relation

$$\mathcal{E} [\tau] = \mathcal{E} [\mathcal{E} [\tau \mid \Omega_t]] = \mathcal{E} [f(\Omega_t)]. \quad (8.2)$$

The Benefit of Delaying

Consider initially that the investor has the options of either purchasing immediately, at the end of period t , or delaying and making the purchase at the end of period $t+1$. This is not an optimal stopping problem since there is only one decision point, at the end of time t . By purchasing right away, the investor assures himself a reward of 1. If the investor delays for a day he benefits by one day of interest, denoted i , tempered by his impatience, denoted δ , where both i and δ are unitless continuously

²It will be assumed in this Chapter that $p_t \in \Omega_t$, that the investor is a price-taker (i.e. can purchase at the price p_t), and that the OTR is reviewed once a day, at the close of trading. These assumptions are made for simplicity and are not crucial to the analysis.

compounded daily proportions.³ While δ could be set to zero and ignored, it will be seen below that impatience is a crucial factor in the realism of the OTR, since otherwise the OTR can imply that the investor will often wait for very long periods of time, even years. In optimal stopping problems generally, impatience tends to take the place of utility. With impatience built in, the alternatives can be compared by their expectations. Those alternatives which imply stopping a long way into the future are penalized by the impatience term, in much the same manner as a concave utility function would penalize them according to their larger dispersion.

If the investor chooses to delay, he is also exposed to a change in the price of the stock. As in Chapter 5, define the daily logarithmic return r_t as

$$r_t \stackrel{\text{def}}{=} \ln(p_t/p_{t-1}). \quad (8.3)$$

If at t the investor waits until $t + 1$, he takes a chance on the return r_{t+1} being positive (resp. negative), and so his capital buys less (resp. more) stock than before. Putting these together, the investor is faced with a straightforward comparison of rewards, denoted τ_1 for the single decision point:

$$\tau_1 = \begin{cases} 1 & \text{Buy at } t \\ e^{i-\delta} e^{-r_{t+1}} & \text{Wait until } t + 1 \end{cases} \quad (8.4)$$

In order to compare these two alternatives he takes expectations conditional upon his information at time t , and chooses which ever alternative has the larger expectation,

³It might be suggested that in not purchasing the stock the investor misses out on its dividends, and so i should be defined as interest less dividends. However, the investor is compensated for the missed dividend by the ex-dividend fall in the stock price. Therefore if in the market the investor gets the dividend, while if out he gets the benefit of the ex-dividend fall in price. In the absence of tax complications, these two effects offset and the role of dividends can be ignored. However, if there are different marginal tax rates on dividend income and capital gains, these should be incorporated into an operational analysis (see, e.g., Elton and Gruber, 1970).

giving the reward function in this simple case as

$$f_1(\Omega_t) = \max \{1, e^{i-\delta} \mathcal{E} [e^{-r_{t+1}} \mid \Omega_t]\}, \quad (8.5)$$

where, once again, the 1 subscript indicates the single decision point.

Consider now the slightly more general case, which is an optimal stopping problem, where the investor is permitted to delay at most 2 days, i.e. he must be invested by $t + 2$. There are now two decision points: at t and, conditional upon not purchasing at time t , at $t + 1$. Having not bought at t , the problem at $t + 1$ is identical in structure to the problem at time t , giving the rewards

$$\tau_2 = \begin{cases} 1 & \text{Buy at } t \\ e^{i-\delta} e^{-r_{t+1}} & \text{Wait until } t + 1 \\ e^{2(i-\delta)} e^{-(r_{t+1}+r_{t+2})} & \text{Wait until } t + 2 \end{cases} \quad (8.6)$$

and the resulting reward function is

$$f_2(\Omega_t) = \max \{1, e^{i-\delta} \mathcal{E} [e^{-r_{t+1}} f_1(\Omega_{t+1}) \mid \Omega_t]\}. \quad (8.7)$$

In general, the n decision problem (i.e. permitting a delay of up to n days) has f_n on the lefthand side of the reward expression, and f_{n-1} on the right. In the limit as $n \rightarrow \infty$ the investor is permitted to delay for as long as he sees fit, and the two functions f_n and f_{n-1} converge. This gives the following recursive expression for the reward function of the OTR, known as its *Bellman equation* (see, e.g., Dixit, 1990; Dixit and Pindyck, 1994).

Definition 8.2 (Bellman Equation) *The reward function of the Optimal Timing*

Rule satisfies the Bellman equation

$$f(\Omega_t) = \max \{1, e^{i-\delta} \mathcal{E} [e^{-r_{t+1}} f(\Omega_{t+1}) \mid \Omega_t]\}. \quad (8.8)$$

The reward function satisfying eq. (8.8) partitions the information space Ω_t into 'buy' regions where $f(\Omega_t) = 1$, and 'wait' regions where $f(\Omega_t) > 1$.

Expressed in words, eq. (8.8) states that at the end of period t the investor chooses either to purchase, in which case there is no more waiting, no more interest, and no more price fluctuations (and so the relative reward is 1), or he chooses to wait until the end of period $t + 1$ and then take the decision again in the light of the information he has gained by waiting. At t he cannot know what information he will gain by waiting or the return r_{t+1} : his best guess is the expectation conditional on his current information. The benefit of waiting includes interest on his capital but is tempered by his impatience, represented in the discount factor $\exp(i - \delta)$. Having found f the investor considers his particular information set, Ω_t . If $f(\Omega_t) > 1$ the optimal decision is to wait, otherwise (i.e. when $f(\Omega_t) = 1$) it is to purchase immediately.

8.3 Optimal Timing When Returns are AR(1)

In this chapter it will be assumed that returns follow the positive AR(1) process

$$r_{t+1} = \mu + \rho r_t + \sigma z_{t+1}, \quad \rho \in (0, 1), \quad (8.9)$$

where z_t is a gaussian white noise disturbance process. In this case daily returns are Markov, and the information set Ω_t boils down to the return over the previous period, r_t . From eq. (8.9), the conditional distribution $r_{t+1} \mid r_t$ is normal with mean $\mu + \rho r_t$ and variance σ^2 . The unconditional distribution of r_{t+1} is also normal with

mean μ_u and variance σ_u^2 , where

$$\mu_u \stackrel{\text{def}}{=} \mu/(1 - \rho), \quad \sigma_u^2 \stackrel{\text{def}}{=} \sigma^2/(1 - \rho^2) \quad (8.10)$$

(see, e.g., Johnson, 1984, pp. 289-90). From Chapter 5, typical values are $\mu = 0.0003$ [7% annualized], $\sigma = 0.0095$ [15% annualized] and $\rho = 0.05$. The small value for the autocorrelation coefficient causes there to be little difference between the conditional and unconditional moments: the unconditional mean is 7.4% annualized and the unconditional standard deviation is 15.04% annualized.

Incorporating eq. (8.9) into the Bellman equation eq. (8.8) gives the complete problem, whose domain is the mathematical field of functional analysis and operator theory.⁴

Definition 8.3 (AR(1) Optimal Timing Problem) *The optimal timing problem is to solve, if possible, for the fixed point $f^* = Af^*$, where $A(f(r), r)$ is the operator*

$$A(f(r), r) \stackrel{\text{def}}{=} \max \left\{ 1, e^{i-\delta} \int_R e^{-r'} f(r') \phi(r'; \mu + \rho r, \sigma^2) dr' \right\}. \quad (8.11)$$

The Optimal Timing Rule is then to buy immediately if $f^(r_t) = 1$, otherwise to delay and take the decision again on the basis of r_{t+1} .*

The integral in eq. (8.11) is the expectation, with r' playing the role of r_{t+1} , since there is no need for an explicit time-dimension. The expression $f^* = Af^*$ is then the Bellman equation of eq. (8.8), where f^* might be thought of as the root of $f - Af = 0$.

In general it is not possible to say whether the Optimal Timing Problem in

⁴For an excellent exposition of this field see Hutson and Pym (1980), upon which much of the following notation is based. An applied introduction in economics can be found in Stokey *et al.* (1989).

Definition 8.3 has a solution, let alone how that solution might be found. However, if f is located within a class of functions with certain 'spatial' properties, then a solution is guaranteed and sometimes can be found, conditional upon operator A also having certain basic properties when used on functions from the class.

Banach Spaces

The necessary spacial properties of f are encapsulated in a *Banach Space*. A Banach space is a complete normed vector space. In a vector space the operations of addition and scalar multiplication are defined; a normed vector space is a vector space in which a distance metric is defined; a complete normed vector space is a normed vector space in which all Cauchy sequences in the space converge within the space.

The simplest Banach spaces are the Euclidean spaces, R^n . Addition is defined for $X, Y \in R^n$ as $Z = X + Y$ where $z_i \stackrel{\text{def}}{=} x_i + y_i$ ($i = 1, \dots, n$); similarly multiplication by the scalar c as $Z = cX$ where $z_i \stackrel{\text{def}}{=} cx_i$ ($i = 1, \dots, n$). There are many distance metrics, of which the most familiar is the Euclidean distance:

$$\|X\| = \sqrt{x_1^2 + \dots + x_n^2}, \quad (8.12)$$

where $\|\cdot\|$ denotes the norm of the space. A sequence X_1, X_2, \dots is Cauchy if

$$\lim_{m,n \rightarrow \infty} \|X_m - X_n\| = 0. \quad (8.13)$$

In R^n all Cauchy sequences are convergent to a point in R^n , and so R^n is a Banach space when equipped with the Euclidean norm or, indeed, with many other norms (Hutson and Pym, 1980, pp. 17-19).

More generally, consider the space of bounded continuous functions on R , denoted \mathcal{C} , where addition is defined for $f, g \in \mathcal{C}$ as $(f + g)(x) \stackrel{\text{def}}{=} f(x) + g(x)$ and multiplication by the scalar c as $(cf)(x) \stackrel{\text{def}}{=} cf(x)$. The usual norm in this space is

the *sup norm*

$$\|f\| = \sup_{x \in R} |f(x)|. \quad (8.14)$$

The space \mathcal{C} is a Banach space when equipped with the sup norm (Hutson and Pym, 1980, pp. 21–22).

Unfortunately the simple Banach space \mathcal{C} with the sup norm is not appropriate for the problem in Definition 8.3. The reason is that the requirement that f^* be bounded is over-restrictive. By the definition of the problem, a hugely negative return will cause f^* to be hugely positive due to the $\exp(-r')$ term, and this gives rise to the conjecture $\lim_{r \rightarrow -\infty} f^*(r) = \infty$. The use of the Banach space \mathcal{C} would therefore impose inappropriate structure.

The right Banach Space for this problem turns out to be a more specialized space known as a *Hilbert space*. A Hilbert space is a Banach space equipped with an *inner product*, which is a function defined on any two members of the space satisfying certain properties, which can give rise to generalized notions of orthogonality. To give my own completely heuristic comparison of Banach spaces and Hilbert spaces, a Banach space (with the sup norm) is flat, while a Hilbert space with the appropriate inner product (which can act as a norm) is a ‘curved’ space in which the curvature can reflect the domain of the underlying problem. In the case of the problem in Definition 8.3, the Banach space \mathcal{C} with the sup norm is flat along the domain of $r \in R$, and so the problem in this space would fail to distinguish between those values of r which were likely and those values which were highly unlikely. But a Hilbert space can be made to ‘curve’ in R to reflect the fact that the unconditional distribution of r implies that certain values of $r \in R$ are highly unlikely and should not be accorded much weight.

Therefore the inner product of the appropriate Hilbert space is defined over the

unconditional distribution of r :

$$(f, g) \stackrel{\text{def}}{=} \int_R f(r) g(r) \phi(r; \mu_u, \sigma_u^2) dr \quad (8.15)$$

where (\cdot, \cdot) denotes the inner product, and the Hilbert space of squared summable continuous functions on R with this inner product is denoted $L^2(R, \phi)$. The norm of this space is derived directly from the inner product:

$$\|f\| = (f, f)^{0.5}, \quad f \in L^2(R, \phi). \quad (8.16)$$

The choice of $L^2(R, \phi)$ also has an extremely useful algebraic side-effect, without which the subsequent analysis would be far more complicated. This is described in the following lemma.⁵

Lemma 8.1 *For any constant $k \in R$, any function $f \in L^2(R, \phi)$ and any parameters $m \in R$, $s^2 \in R_{++}$*

$$\int_a^b e^{kr} \phi(r; m, s^2) dr = e^{mk+0.5s^2k^2} \int_a^b \phi(r; ks^2 + m, s^2) dr.$$

Proof: The proof is straightforward. Combining the term $\exp(kr)$ with the exponential expression in the normal distribution gives an exponential term in the integrand of

$$\begin{aligned} & \exp\left(kr - \frac{(r-m)^2}{2s^2}\right) \\ &= \exp\left(\frac{2krs^2 - r^2 + 2rm - m^2}{2s^2}\right) \\ &= \exp\left(0.5(ks)^2 + km - \frac{(r - (ks^2 + m))^2}{2s^2}\right) \end{aligned} \quad (8.17)$$

⁵When $a = -\infty$ and $b = \infty$, this lemma will be familiar to mathematicians as the derivation of the moment generating function of an arbitrary normal distribution, i.e. $\mathcal{E}[e^{kX}]$ where X is normal with mean m and variance s^2 .

after completing the square in r . The first term in eq. (8.17) can then go outside the integral, while the second term represents the exponential expression from the normal distribution not with mean m but with mean $ks^2 + m$; everything else, including the variance and the limits, remains unchanged. ■

8.4 Sufficient Conditions

Identifying $L^2(R, \phi)$ as an appropriate space for f in the Optimal Timing Problem is only half the battle for solving for the fixed point f^* of A . The other half is ensuring that the parameters of the problem, μ , σ , ρ , i and δ , are such that the operator A has appropriate properties for functions such as f in $L^2(R, \phi)$. There are several fixed point theorems applicable to operators on Banach spaces, but one of them stands out as being not only sufficient for the existence of a unique fixed point, but also providing an algorithm by which this point can be found. This is the *Contraction Mapping Principle* (CMP).⁶

Theorem 8.1 (Contraction Mapping Principle) *Suppose that the operator A maps the closed subset \mathcal{D} of the Banach space \mathcal{B} into \mathcal{D} and is a contraction. Then A has exactly one fixed point, f^* say, in \mathcal{D} . Further, for any initial guess $f_0 \in \mathcal{D}$, the successive approximations $f_n = A f_{n-1}$ ($n > 0$) converge to f^* . (Hutson and Pym, 1980, p. 116)*

In order to employ the CMP on the optimal timing problem, two points must be established. First, there must be a closed subset $\mathcal{D} \subset L^2(R, \phi)$ in which the operator A maps members back into \mathcal{D} . Second, A must be a *contraction* on \mathcal{D} . The first of these points is easy to establish.

⁶The condition for operator A to be a contraction is given in eq. (8.20). Crudely, an operator is a contraction on some set if it shrinks the distance between any two members of that set.

Proposition 8.1 Define $\mathcal{D} \subset L^2(R, \phi)$ as those members of $L^2(R, \phi)$ which are non-increasing and nowhere less than 1; then \mathcal{D} is closed and $f \in \mathcal{D}$ implies $Af \in \mathcal{D}$.

Proof: It is clear that \mathcal{D} is a closed subset of $L^2(R, \phi)$ (crudely, \mathcal{D} contains its own endpoints). Now consider $f \in \mathcal{D}$ and $f' \stackrel{\text{def}}{=} Af$; it must be shown that $f' \in \mathcal{D}$ where

$$f'(r) = \max \left\{ 1, e^{i-\delta} \int_R e^{-r'} f(r') \phi(r'; \mu + \rho r, \sigma^2) dr' \right\}. \quad (8.18)$$

To prove that f' is non-increasing in r , remember that $\rho > 0$. As r increases, the p.d.f. ϕ in the second term of eq. (8.18) shifts upwards, i.e. higher weight in the expectation is attached to larger values of r' . Since the first term in the expectation, $e^{-r'}$, is decreasing in r' and positive, and the second, $f(r')$, is non-increasing and positive by construction, their product must be decreasing in r' , and thus the value of the expectation must fall as r increases. Therefore f' is non-increasing. Furthermore, since the product falls to zero in the limit $r' \rightarrow \infty$, so at some finite point the function f' will take the value 1 and become horizontal. ■

The second point, that A is a contraction on \mathcal{D} , is harder to establish, and requires a condition on the parameters.

Proposition 8.2 The operator A is a contraction on \mathcal{D} providing that the following condition holds:

$$i - \delta + 0.5 \sigma^2 \left(\frac{1 + \rho^2}{1 - \rho^2} \right) - \mu \left(\frac{1}{1 - \rho} \right) < 0. \quad (8.19)$$

Proof: For A to be a contraction on \mathcal{D} it must be shown that

$$\|Af - Ag\| < \|f - g\| \quad (f, g \in \mathcal{D}). \quad (8.20)$$

Consider some function f , and $g \stackrel{\text{def}}{=} f + k$ for some positive constant k . It is easy to see that

$$\|f - g\| = \left\{ \int_R (-k)^2 \phi(r; \mu_u, \sigma_u^2) dr \right\}^{0.5} = k. \quad (8.21)$$

Now consider using the operator on f and g , using $\gamma \stackrel{\text{def}}{=} \exp(i-\delta)$ for convenience.

For Af ,

$$Af = \max \left\{ 1, \gamma \int_R e^{-r'} f(r') \phi(r'; \mu + \rho r, \sigma^2) dr' \right\}. \quad (8.22)$$

For Ag ,

$$\begin{aligned} Ag &= \max \left\{ 1, \gamma \int_R e^{-r'} [f(r') + k] \phi(r'; \mu + \rho r, \sigma^2) dr' \right\} \\ &= \max \left\{ 1, \gamma \int_R e^{-r'} f(r') \phi(r'; \mu + \rho r, \sigma^2) dr' \right. \\ &\quad \left. + \gamma k \int_R e^{-r'} \phi(r'; \mu + \rho r, \sigma^2) dr' \right\}. \end{aligned} \quad (8.23)$$

The absolute difference between Af and Ag is always less than or equal to the second integral term in eq. (8.23). Using Lemma 8.1, this term can be written

$$\begin{aligned} &\gamma k \int_R e^{-r'} \phi(r'; \mu + \rho r, \sigma^2) dr' \\ &= \gamma k e^{-(0.5\sigma^2 + \mu + \rho r)} \int_R \phi(r'; -\sigma^2 + \mu + \rho r, \sigma^2) dr' \\ &= \gamma k e^{0.5\sigma^2 - (\mu + \rho r)}. \end{aligned} \quad (8.24)$$

Hence,

$$\begin{aligned}
\|Af - Ag\| &= \left\{ \int_R (Af - Ag)^2 \phi(r; \mu_u, \sigma_u^2) dr \right\}^{0.5} \\
&\leq \left\{ \int_R \left[\gamma k e^{0.5\sigma^2 - (\mu + \rho r)} \right]^2 \phi(r; \mu_u, \sigma_u^2) dr \right\}^{0.5} \\
&= \gamma k \left\{ e^{\sigma^2 - 2\mu} \int_R e^{-2\rho r} \phi(r; \mu_u, \sigma_u^2) dr \right\}^{0.5} \\
&= \gamma k e^{0.5\sigma^2 - \mu} \left\{ e^{-2\rho(0.5(-2\rho)\sigma_u^2 + \mu_u)} \right\}^{0.5} \\
&= k \exp(i - \delta + 0.5\sigma^2 - \mu + \rho^2\sigma_u^2 - \rho\mu_u), \tag{8.25}
\end{aligned}$$

using Lemma 8.1 once again in the penultimate line.

For the contraction to hold, it is sufficient that eq. (8.25) be less than eq. (8.21), i.e. less than k . This requires that the exponent in eq. (8.25) be less than 0. Expressing μ_u and σ_u^2 according to their definitions in eq. (8.10) and rearranging gives the contraction condition of the proposition. ■

Before going any further, it is worth asking whether the contraction condition from Proposition 8.2 is likely to hold for typical values of the five parameters. The first point to note is that there is always a value for δ which will ensure that any valid combination of the other parameters will satisfy the inequality. Typical (stylized) values of the return parameters were given after eq. (8.10) (p. 162). Taking $i = 0.0002$ [6% annualized] and setting $\delta = 0$ gives a just-admissible value for the contraction condition, eq. (8.19), of -0.0001 . Therefore since $\delta \geq 0$, the contraction condition is likely to hold for typical parameter values.

8.5 The Solution

By Propositions 8.1 and 8.2 the optimal timing problem defined in Definition 8.3 satisfies the CMP, and so there is a unique solution which may be found by iteration.

This solution is now found in two stages. First, a general expression for $f_n \in \mathcal{D}$ can be derived from the definition of the operator A . Second, the facts of uniqueness and convergence can be used to solve for the reward function f^* in the limit as $n \rightarrow \infty$.

The general expression for f_n is given in the following proposition.

Proposition 8.3. *Let the n^{th} iteration of f , f_n , be written as*

$$f_n(r) = \begin{cases} f'_n(r) & r < r^{(n)} \\ 1 & r \geq r^{(n)} \end{cases}, \quad (8.26)$$

where $r^{(n)}$ denotes the 'elbow' in f_n , i.e. the root of $f'_n(r) - 1 = 0$. Then, for any initial choice $f_0 \in \mathcal{D}$,

$$f'_n(r) = \prod_{j=1}^n B_{n-j} f'_0(r) + \sum_{j=0}^{n-2} \prod_{i=1}^{n-(j+1)} B_{n-i} g_j(r) + g_{n-1}(r), \quad (8.27)$$

where

$$B_n(f(r), r) \stackrel{\text{def}}{=} e^{i-\delta} \int_{-\infty}^{r^{(n)}} e^{-r'} f(r') \phi(r'; \mu + \rho r, \sigma^2) dr', \quad (8.28)$$

$$g_n(r) \stackrel{\text{def}}{=} e^{i-\delta} \int_{r^{(n)}}^{\infty} e^{-r'} \phi(r'; \mu + \rho r, \sigma^2) dr'. \quad (8.29)$$

Proof: Consider the iteration from f_{n-1} to f_n , and write $\phi(r' | r) \stackrel{\text{def}}{=} \phi(r'; \mu + \rho r, \sigma^2)$ for simplicity.

$$\begin{aligned}
 f_n(r) &= A f_{n-1}(r) \\
 &= \max \left\{ 1, e^{i-\delta} \int_R e^{-r'} f_{n-1}(r') \phi(r' | r) dr \right\} \\
 &= \max \left\{ 1, e^{i-\delta} \int_{-\infty}^{r^{(n-1)}} e^{-r'} f'_{n-1}(r') \phi(r' | r) dr \right. \\
 &\quad \left. + e^{i-\delta} \int_{r^{(n-1)}}^{\infty} e^{-r'} \phi(r' | r) dr \right\} \\
 &= \max \{ 1, B_{n-1} f'_{n-1}(r) + g_{n-1}(r) \}
 \end{aligned} \tag{8.30}$$

by the definition of the operator B_n and the function g_n in eq. (8.28) and eq. (8.29) respectively. This shows the iterative relationship between f'_n and f'_{n-1} :

$$f'_n(r) = B_{n-1} f'_{n-1}(r) + g_{n-1}(r). \tag{8.31}$$

By back-substitution of $f'_{n-1}(r)$, ignoring the functional argument for simplicity

$$f'_n = B_{n-1} (B_{n-2} f'_{n-2} + g_{n-2}) + g_{n-1}. \tag{8.32}$$

Since B_n is a linear operator it is distributive across the parenthesis, hence

$$\begin{aligned}
 f'_n &= B_{n-1} B_{n-2} f_{n-2} + B_{n-1} g_{n-2} + g_{n-1} \\
 &= B_{n-1} B_{n-2} B_{n-3} f_{n-3} + B_{n-1} B_{n-2} g_{n-3} + B_{n-1} g_{n-2} + g_{n-1} \\
 &\vdots
 \end{aligned}$$

Continuing to make these back-substitutions leads directly to the expression in the proposition.⁷ ■

⁷Proposition 8.3 may also be proved by induction, but I prefer this proof because it is constructive.

For the second step in finding the solution, the expression for $f'_n(r)$ is found as the limit $n \rightarrow \infty$.

Proposition 8.4 *The solution to the optimal timing problem given in Definition 8.3 is*

$$f^*(r) = \max \left\{ 1, \sum_{j=0}^{\infty} B^j g(r) \right\}, \quad (8.33)$$

where

$$B(h(r), r) \stackrel{\text{def}}{=} e^{i-\delta} \int_{-\infty}^{r^*} e^{-r'} h(r') \phi(r'; \mu + \rho r, \sigma^2) dr' \quad (8.34)$$

$$g(r) \stackrel{\text{def}}{=} e^{i-\delta} \int_{r^*}^{\infty} e^{-r'} \phi(r'; \mu + \rho r, \sigma^2) dr' \quad (8.35)$$

and $r = r^*$ solves $\sum_{j=0}^{\infty} B^j g(r) = 1$.

Proof: It has already been established by the CMP (given as Theorem 8.1) that there is a unique fixed point which represents the solution f^* , and that successive iterations will converge upon that fixed point, i.e. $\lim_{n \rightarrow \infty} f_n = f^*$. At convergence $f_n = f_{n+1} = \dots = f^*$, and this implies that $r^{(n)} = r^{(n+1)} = \dots = r^*$, $B_n = B_{n+1} = \dots = B$ (defined in eq. (8.34)) and $g_n = g_{n+1} = \dots = g$ (defined in eq. (8.35)). Presuming n is already arbitrarily large (i.e. we are already arbitrarily close to convergence),

$$\begin{aligned} \lim_{n \rightarrow \infty} f'_n(r) &= \lim_{n \rightarrow \infty} \left\{ B^n f'_0(r) + \sum_{j=0}^{n-2} B^{n-(j+1)} g(r) + g(r) \right\} \\ &= \lim_{n \rightarrow \infty} \left\{ B^n f'_0(r) + \sum_{j=0}^{n-1} B^j g(r) \right\}. \end{aligned} \quad (8.36)$$

The infinite series must be convergent, which in turn implies that $\lim_{n \rightarrow \infty} B^n f'_0(r) = 0$, and the result follows directly. ■

Finally, it should be noted that the solution of the Optimal Timing Problem given in Proposition 8.4 appears deceptively simple. Expanding the infinite power series in the operator B for the first few terms gives:

$$\begin{aligned}
 & g(r) \\
 & + e^{i-\delta} \int_{r^*}^{\infty} e^{-r'} g(r') \phi(r' | r) dr' \\
 & + e^{2(i-\delta)} \int_{-\infty}^{r^*} e^{-r'} \left\{ \int_{r^*}^{\infty} e^{-r''} g(r'') \phi(r'' | r') dr'' \right\} \phi(r' | r) dr' \\
 & \vdots
 \end{aligned}$$

where $g(r)$ is itself an integral expression. The calculation of this function to the point of convergence is a major endeavour, and is therefore the subject of the next chapter.

8.6 Summary

The main result of this chapter is a function, the 'Reward Function' given in Definition 8.1, and solved in Proposition 8.4. This function shows the expected benefit for an investor implementing an optimal timing rule for a speculative asset for which the daily return shows positive first order autocorrelation. The OTR in this case is to buy the asset immediately should its return over the period just ending be greater than or equal to some threshold value, denoted r^* , otherwise hold on for one period and take the same decision again at the end of the next period. The attraction of this rule is its extreme simplicity in implementation. This is in stark contrast to both the solution of the Optimal Timing Problem for the reward function expression, and the calculation of this expression for a given set of parameter values. It is to this latter problem that I turn in the next chapter.

Chapter 9

Computing the Reward Function

The last chapter posed and solved the problem of purchase timing in a market with positive autocorrelation in daily returns. Briefly, an investor intends to purchase a given stock with a given amount of capital. Since his investment horizon is quite long (say, a year) he is relaxed about the precise timing of the purchase. He can sometimes take advantage of the small amounts of positive autocorrelation in daily returns by delaying his purchase. When daily returns follow an AR(1) process, the decision about whether or not to purchase is made by a comparison between the return over the period just ending, r_t , and a threshold value, denoted r^* . The Optimal Timing Rule (OTR) is to purchase immediately (i.e. at the end of period t) whenever $r_t \geq r^*$; otherwise to wait until the end of period $t + 1$ and then apply the same rule again.¹

One question left unanswered in Chapter 8 was the magnitude of the reward from following the OTR in a stock market. In this chapter the reward function f^* is computed over a range of likely values for the parameters to identify the threshold value r^* and the unconditional expectation of the OTR. Section 9.1 describes the usual approach to solving optimal stopping problems, and its failure in this case.

¹The notation in this chapter is the same as that of Chapter 8.

Section 9.2 describes in overview and in detail the computation of the reward function, and Section 9.3 the results over ranges of typical parameter values. Section 9.4 discusses the results.

9.1 One Simple Approach

The threshold value r^* can be thought of as the ‘elbow’ in the reward function f^* , i.e. the smallest value of r for which $f^*(r) = 1$. Therefore knowledge of f^* is sufficient for knowledge of r^* , and any method which leads to f^* will also yield r^* . In the last chapter the optimal timing problem was solved explicitly, expressing the reward function in terms of the parameters, μ , σ , ρ , i and δ , as given in Proposition 8.4.

In general, however, it is extremely difficult to find solutions to non-trivial optimal stopping problems, and so another method is used. This method makes use of the Contraction Mapping Principle (CMP). As long as it can be shown that the problem satisfies the CMP conditions (given in Theorem 8.1), then iterations of the form $f_n = Af_{n-1}$ from an appropriate starting-point f_0 are bound to converge on the true reward function f^* . (For a simple exposition, see Dixit and Pindyck, 1994, Appendix to Chapter 3.) To implement this method, it is necessary to be able to describe the functions f_0, f_1, \dots in a consistent and flexible manner.

In my first attempts to find f^* , made before finding the solution described in Section 8.5, I used a cubic spline to describe the successive iterated functions.² I found that convergence was extremely slow and extremely unreliable (i.e. sensitive to the parameter values). However the fact of convergence, in whatever fashion, was enough to suggest that the problem was well-defined and solvable, and so encouraged further theoretical investigation.

I subsequently tried more parsimonious representations for the iterated functions

²For an introduction to the theory and application of splines, see Press *et al.* (1992).

which were fitted by OLS following each iteration. It was clear from the cubic splines that the function f^* was roughly piecewise linear around r^* , suggesting

$$f(r) = \begin{cases} e^{a+br} & r < r^* \\ 1 & r \geq r^* \end{cases} \quad (9.1)$$

where $r^* \stackrel{\text{def}}{=} -a/b$ and $b < 0$. However, these failed to improve in any meaningful way over the splines. Convergence was still slow and sensitive to the parameter values, and the notional errors about r^* were much too large in relation to the standard deviation of daily returns to make the implementation of a strategy based around the estimate viable.

9.2 Computation of the Reward Function

The explicit solution found in Proposition 8.4 replaces one set of problems with another. Clearly the iterative method was not working very well, but attempting to calculate an infinite power-series in operators was likely to be equally, if not more, tricky. The problem is that there is no way of knowing, a priori, how quickly the series will converge. Therefore the integration routines must be capable of going to an arbitrary number of dimensions, to the point in the series at which some convergence criterion holds. This would be quite out of the question for standard integration routines, where the number of function calls would increase by a factor n for each extra dimension, where n is the number of points used in the integration.³

³So, for example, the tenth term in the series would require about 3×10^8 function calls with just seven points in the integration. Each one of these function calls introduces round-off error, not to mention the large amount of truncation error from using only seven points. A 'quick and dirty' calculation using a recursive algorithm established the infeasibility of this approach. One promising alternative for high dimensional integrals is that of lattice integration (see, e.g., Sloan, 1992), but unfortunately a generic technique for integrals of greater than two dimensions has yet to be developed. For a general overview of the techniques of numerical integration, see Davis and Rabinowitz (1984).

However, it is possible to exploit the particular structure of the power series to find the reward function f^* to an accuracy close to machine accuracy.

Overview of the Method

The reward function has to be computed numerically, since it is not possible to resolve the terms in the operator power series into elementary functions of the parameters. Consequently, each term in the series will be evaluated not over the whole of $r \in (-\infty, r^*]$, but over a finite collection of points, $r_0, \dots, r_n \in R$, where $r_0 < r_1 < r_2 < \dots < r_n$ and $r_0 = -\infty$ and $r_n = r^*$. Initially, assume that r^* is known. For each of these points bar the first (which requires special handling, as described below), the first term in the series, $g(r_k)$ ($k = 1, \dots, n$), can be calculated directly using Lemma 8.1. After this term has been found the next term in the series, $Bg(r_k)$, can be approximated at each r_k ($k = 1, \dots, n$) by numerical integration over the points $r_{k'}$ ($k' = 0, \dots, n$):

$$Bg(r_k) = e^{i-\delta} \int_{-\infty}^{r^*} e^{-r'} g(r') \phi(r' | r_k) dr' \quad (9.2)$$

$$\approx e^{i-\delta} \sum_{k'=0}^n w(k', n) e^{-r_{k'}} g(r_{k'}) \phi(r_{k'} | r_k) \Delta r \quad (9.3)$$

where $w(k', n)$ is a weight function determined by the particular method of numerical integration and Δr is the interval width. Since the integral has an infinite lower limit, the weights and the interval width both need to be chosen carefully, as discussed further below.

Progressing in this way, the matrix F may be built up row by row, where

$$F = \begin{bmatrix} g(r_1) & g(r_2) & \cdots & g(r_n) \\ Bg(r_1) & Bg(r_2) & \cdots & Bg(r_n) \\ B^2g(r_1) & B^2g(r_2) & \cdots & B^2g(r_n) \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix} \quad (9.4)$$

If $F(j, k)$ denotes the value of the term in row j column k , i.e. $B^jg(r_k)$, the general rule for constructing F is

$$F(j, k) = e^{i-\delta} \sum_{k'=0}^n w(k', n) e^{-r_{k'}} F(j-1, k') \phi(r_{k'} | r_k) \Delta r \quad (9.5)$$

where $k = 1, \dots, n$ and $j = 1, 2, \dots$. The matrix continues to be built up row by row until the column sums converge. At this point we know the values of f^* at the points r_1, \dots, r_n , and the general picture of f^* for $r \in R$ can be found by interpolation and extrapolation.

There is a slight flaw in this method in that it presupposes that we know r^* , when in fact r^* can only be known once f^* has been found. However, the value of r^* can be found iteratively by successive application of the above method. An initial value is chosen, say r_1^* , and then F is computed until convergence. If r_1^* is actually the true value r^* then the sum of column n will be exactly 1, since $f^*(r^*) = 1$. If the column sum is not 1, a new (hopefully better) value is selected, r_2^* , and F is recalculated, and so on. Therefore r^* can be found iteratively and simultaneously with f^* as the root of the equation $f^*(r^*) - 1 = 0$, or in terms of F , as the root of $\sum_{j=0}^{\infty} F(j, n) - 1 = 0$.

What about the values of r_k ?

Numerical integration routines fall, crudely, into two camps. First, there are those which presume equally-spaced points and are robust across a wide range of smooth functions (e.g. the standard methods such as Simpson's rule); second, those in which the spacing of the points can be optimized for particular types of function ('quadrature'). Since the nature of the functions $B^k g(r)$ are not well-known, the first method is applicable. However, it would take an infinite number of equally-spaced points to fill the range $(\infty, r^*]$. Therefore the integral must be transformed by a change of variable so that the range of integration is finite.

The transformation used is $y = 1/x$ which gives rise to the new integral

$$\int_{-\infty}^c f(x) dx = - \int_0^{c^{-1}} f(y^{-1}) y^{-2} dy \quad (9.6)$$

providing that $c < 0$.⁴ In application it cannot be guaranteed that $r^* < 0$ which necessitates a split in the range of integration, and so the general form of the integral becomes

$$\int_{-\infty}^{r^*} f(x) dx = - \int_0^{c^{-1}} f(y^{-1}) y^{-2} dy + \int_c^{r^*} f(x) dx, \quad (9.7)$$

where c is chosen as some value less than zero. The resulting values for r_1, \dots, r_n

⁴It should be noted that several other transformations are available, in particular $y = e^x$. However, the transformation must be chosen to preserve, or enhance if possible, the smoothness of the function (i.e. ensure a low rate of change in the gradient), in order that the numerical approximation be as accurate as possible. The main determinant of this smoothness in the operator B is the normal distribution, which has a well-defined domain on the real line representing the return r . It was found by experiment that $y = e^x$ compressed the middle part of this domain too much, resulting in a rapid change in gradient of the normal distribution and inaccuracy in the numerical integrations. In contrast, $y = 1/x$ tended to compress mainly the tails of the normal distribution, where the gradient was almost zero, and so the smoothness was not compromised.

and Δr in the two halves are then

$$r_k \stackrel{\text{def}}{=} \begin{cases} \frac{cn_l}{k} \\ c + \frac{(r^* - c)(k - n_l)}{n_r} \end{cases}, \Delta r \stackrel{\text{def}}{=} \begin{cases} \frac{1}{cn_l} & k = 1, \dots, n_l - 1 \\ \frac{r^* - c}{n_r} & k = n_l, \dots, n_l + n_r \end{cases} \quad (9.8)$$

where there are n_l intervals in the lefthand integral and n_r in the righthand, $n_l + n_r = n$.

The Weight Functions

The second integral, on the closed range $[c, r^*]$, presents no problems and a standard weight set such as that of Simpson's rule can be used:

$$w^c(k, n) = \begin{cases} \frac{1}{3} & k = 0, n \\ \frac{4}{3} & k = 1, 3, 5, \dots, n - 1 \\ \frac{2}{3} & k = 2, 4, 6, \dots, n - 2 \end{cases} \quad (9.9)$$

(see, e.g., Press *et al.*, 1992, p. 134), where the superscript 'c' on the weight function indicates that it applies to a closed interval. The number of intervals, n , should be even.

The first integral is on the semi-open interval $(0, c]$. For this a combination of open-interval and closed interval weights can be used:

$$w^{so}(k, n) = \begin{cases} 0 & k = 0, 2 \\ \frac{27}{12} & k = 1 \\ \frac{13}{12} & k = 3 \\ w^c(k - 1, n) & k = 4, \dots, n - 1 \\ w^c(n, n) & k = n \end{cases} \quad (9.10)$$

(see, e.g., Press *et al.*, 1992, pp. 134–136), where the superscript ‘so’ indicates a semi-open interval. The number of intervals in this case should be odd.⁵

Accelerated Convergence

Finally in this section, I turn to the question of how many rows of F need to be calculated before the sum of the excluded terms becomes negligible. To find this out requires the sum of the series of terms for each column to be extrapolated from the terms already available. When these extrapolated sums converge across all r_k ($k = 1, \dots, n$) then no more rows of F are required.

By observing the evolution of the rows of F it was clear that each column was ultimately converging to its limit geometrically (sometimes known as *linear* convergence). This suggests that the column sums can be extrapolated using Aitken’s Δ^2 method (see, e.g., Davis and Rabinowitz, 1984, pp. 43–44). By this method, the sum of the series may be extrapolated as

$$s'_n \approx \frac{s_n s_{n+2} - s_{n+1}^2}{s_{n+2} - 2s_{n+1} + s_n}, \quad (9.11)$$

where s_n is the sum up to and including the n^{th} term of the sequence. Using s'_{nk} to denote the extrapolated sum of the k^{th} row at the n^{th} term, the convergence criterion can be written

$$\sup_{k=1, \dots, n} |s'_{n+1,k} - s'_{n,k}| < \epsilon \quad (9.12)$$

where ϵ is determined externally, for example by machine accuracy.

⁵The point is not explicitly made in Press *et al.* (1992), but it is clear that the weights in all cases should sum to the number of intervals. Hence in the closed formula there must be an even number of intervals (e.g. $1/3 + 4/3 + 2/3 + 4/3 + 1/3 = 4$). This determines that the number of intervals in the semi-open weights be odd (e.g. $0 + 27/12 + 0 + 13/12 + 4/3 + 2/3 + 4/3 + 1/3 = 7$). Hence the $4/3$ factors fall on the odd terms in the closed weights (remembering the numbering starts at zero) and the even terms in the semi-open weights.

9.3 Computation Results

The above methods were implemented in C++. Various numbers of intervals in the two integrations were tried, and the results were found to be stable with anything at or above 13 intervals in the range $(-\infty, -3\sigma_u]$, and 10 intervals in the range $[-3\sigma_u, r^*]$, so these were used. Convergence down the rows of F was extremely quick, generally taking not more than about five iterations. The simultaneous determination of r^* and f^* , which involves finding the root of $f^*(r^*) - 1 = 0$, was performed by bracketing and bisection (see, e.g., Press *et al.*, 1992, Ch. 9). The convergence in both the column sums (ϵ) and the bisection was to six decimal places. With these settings the time taken to find the reward function f^* and the threshold r^* was typically less than 1 second, indicating that the optimal timing rule can be updated in real time should there be changes in the parameters.⁶

Results

As expected from the results of the original spline fitting, the reward function is roughly linear to the left of r^* , as approximated by eq. (9.1). However, attempts to solve the Bellman equation, eq. (8.8), on the presumption that the reward function was linear or log-linear were not successful, and suggested strongly that there is some non-linearity present, perhaps up in the lefthand tail.⁷

The reward function was computed for different sets of values of σ , ρ and δ : $\sigma \in \{0.10, 0.15, 0.20\}$ % annualized, $\rho \in \{0.05, 0.10, 0.15\}$ and $\delta \in \{0, i, 2i\}$. The two other parameters μ and i were set to 7 % annualized and 6 % annualized,

⁶This also indicates that the optimal timing rule could be implemented for durations considerably less than one day, perhaps to take advantage of the intra-day return autocorrelation structure.

⁷It is worth mentioning here that there is another possible method of computation for the infinite power series, which I was holding in reserve should the method described above not have been effective. As in a power series of scalars it is sometimes possible to express an infinite power series in operators in the form (crudely) $(1 - B)^{-1}g(r)$. The possible or near linearity of f^* is suggestive of a neat solution to this problem. However, this approach was not required since the convergent rows of F method was so effective.

Table 9.1: Values of r^* at Different Parameter Settings

		Values of σ (annualized)		
		$\sigma = 0.10$	$\sigma = 0.15$	$\sigma = 0.20$
Actual Values of r^*				
$\rho = 0.05$				
δ	0	-0.0113	-0.0169	-0.0225
	i	-0.0127	-0.0182	-0.0237
	$2i$	-0.0143	-0.0196	-0.0251
$\rho = 0.10$				
δ	0	-0.0095	-0.0143	-0.0190
	i	-0.0104	-0.0151	-0.0198
	$2i$	-0.0113	-0.0160	-0.0207
$\rho = 0.15$				
δ	0	-0.0083	-0.0126	-0.0167
	i	-0.0091	-0.0132	-0.0174
	$2i$	-0.0098	-0.0139	-0.0181
Normalized Values of r^*				
$\rho = 0.05$				
δ	0	-1.84	-1.81	-1.80
	i	-2.04	-1.95	-1.90
	$2i$	-2.30	-2.09	-2.00
$\rho = 0.10$				
δ	0	-1.55	-1.53	-1.52
	i	-1.68	-1.62	-1.58
	$2i$	-1.82	-1.71	-1.65
$\rho = 0.15$				
δ	0	-1.36	-1.34	-1.33
	i	-1.47	-1.41	-1.39
	$2i$	-1.58	-1.49	-1.44

The two other parameters took constant values: $\mu = 7\%$ annualized and $i = 6\%$ annualized. The normalized values are found as $(r^* - \mu_u)/\sigma_u$, where μ_u is the unconditional mean and σ_u^2 the unconditional variance of the AR(1) process for returns—see eq. (8.10).

Table 9.2: Values of $\mathcal{E}[\tau]$ at Different Parameter Settings

		Values of σ (annualized)		
		$\sigma = 0.10$	$\sigma = 0.15$	$\sigma = 0.20$
		$\rho = 0.05$		
δ	0	1.0 ⁴ 03	1.0 ⁴ 05	1.0 ⁴ 07
	i	1.0 ⁴ 02	1.0 ⁴ 04	1.0 ⁴ 06
	$2i$	1.0 ⁴ 01	1.0 ⁴ 03	1.0 ⁴ 05
		$\rho = 0.10$		
δ	0	1.0 ⁴ 13	1.0 ⁴ 20	1.0 ⁴ 27
	i	1.0 ⁴ 10	1.0 ⁴ 17	1.0 ⁴ 24
	$2i$	1.0 ⁴ 07	1.0 ⁴ 14	1.0 ⁴ 21
		$\rho = 0.15$		
δ	0	1.0 ⁴ 28	1.0 ⁴ 45	1.0 ⁴ 62
	i	1.0 ⁴ 23	1.0 ⁴ 39	1.0 ⁴ 55
	$2i$	1.0 ⁴ 18	1.0 ⁴ 33	1.0 ⁴ 49

See note to Table 9.1.

respectively. Variation in the latter two parameters was captured through variation in δ . The resulting values for r^* are given in Table 9.1. Also given in Table 9.1 are the normalized values for each r^* , i.e. $(r^* - \mu_u)/\sigma_u$. Expressions for the unconditional mean and variance (μ_u and σ_u^2) for the AR(1) process were given in eq. (8.10). Likewise, the unconditional expectations of the reward, found using eq. (8.2) as

$$\mathcal{E}[\tau] = \int_{-\infty}^{\infty} f^*(r) \phi(r; \mu_u, \sigma_u^2) dr, \quad (9.13)$$

where $\phi(\cdot)$ is the normal density function, are given in Table 9.2. The marginal effect on the reward function of changes in these three parameters is displayed in Figure 9.1. The base case in each of the three parts of this Figure is $\sigma = 0.10$, $\rho = 0.05$ and $\delta = 0$. For each graph the range is from $\mu_u - 4\sigma_u$ to $\mu_u + 2\sigma_u$, giving some indication of natural scale.

The first point to note from Table 9.1 is that the threshold values, r^* , are all substantially negative and as such lie in the lefthand tail of the unconditional distribution of r . At the point $(0.10, 0.05, 2i)$ [order σ, ρ, δ], $r^* = -0.0143$, which is -2.30 unconditional standard deviations from the unconditional mean. An investor implementing this strategy would find himself delaying on only 0.0107 of all purchases (i.e. about 1 in 100); where there was a delay it would almost always be for one day only. At the opposite end of the table, at the point $(0.20, 0.15, 0)$ $r^* = -0.0167$, which is -1.33 standard deviations from the mean. In this case the investor delays on about 10 transactions in every 100. In the 'typical' case $(0.15, 0.05, i)$, the investor delays on about 3 in every 100.

The structure of the entries in Table 9.1, and also in Table 9.2, is highly regular and almost linear. This is confirmed in Figure 9.1. A rise in σ has the effect of lowering r^* but at the same time raising the probability of delay, since the nor-

Figure 9.1: Marginal Effects on the Reward Function $f^*(r)$

Figure 9.1a: Changes in σ

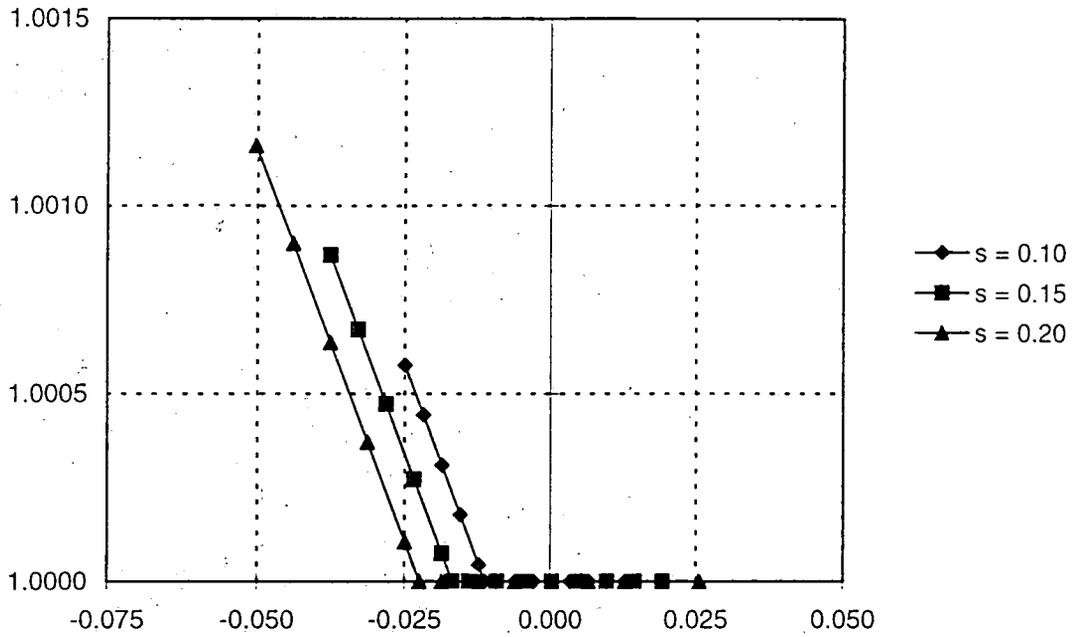


Figure 9.1b: Changes in ρ

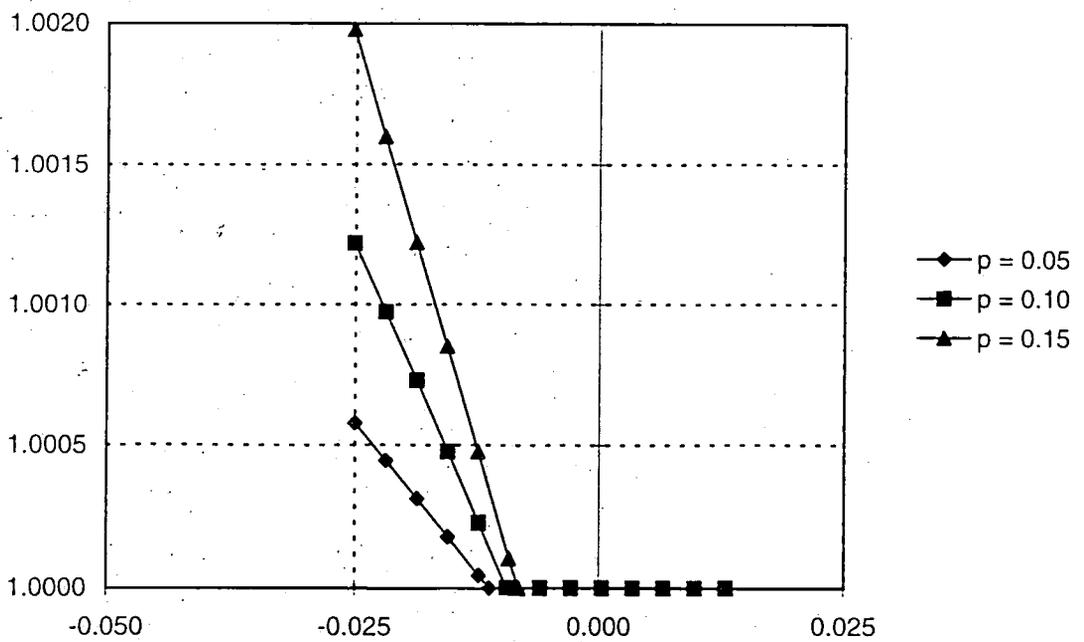
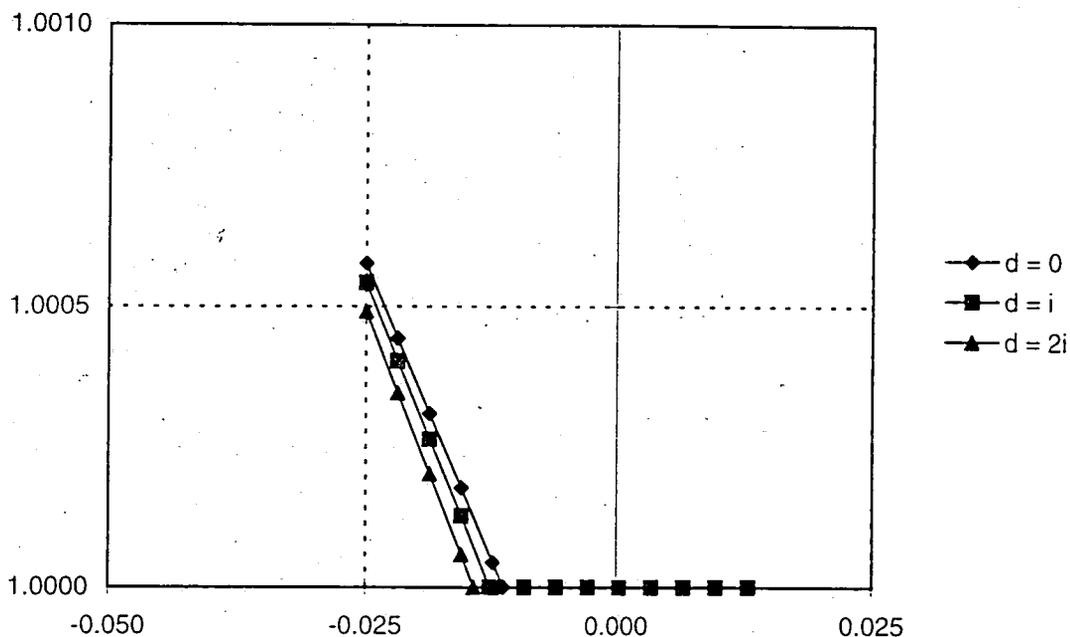


Figure 9.1: Marginal Effects on the Reward Function $f^*(r)$ (cont)Figure 9.1c: Changes in δ 

malized value rises towards the mean; the gradient of f^* below r^* remains almost unchanged. A rise in δ lowers both r^* and the probability of delay; again, the gradient below r^* remains almost unchanged. A rise in ρ has the effect of raising r^* and raising the probability of delay, but it also of raising the gradient of f^* below r^* . Consequently the reward from operating the OTR is most sensitive to changes in the autocorrelation coefficient. This can be seen also in a comparison of the three panels in Table 9.2.

One consequence of this near-linearity is that it would take substantial changes in the standard deviation σ and/or the autocorrelation ρ in order to drive r^* up to the point at which an investor following the OTR is regularly delaying his purchases. Changes of the necessary magnitude would take the value of those two parameters well outside their typical ranges for stockmarkets. Figure 9.1 and Table 9.2 also show that the expected benefit from following the OTR is not very large—not surprisingly

given the infrequency with which purchases are likely to be delayed. The best outcome, $(0.20, 0.15, 0)$, generates an average of about 6 thousandths of a percent per purchase. In this case it would take about 160 purchases to achieve 1 percent outperformance of the OTR over the alternative of always buying immediately. In the typical case it takes about 2,500 purchases. Therefore it must be conceded that the OTR is unlikely to make much difference to the performance of an investor in a market similar in character to a typical stock market, at least if operated over a period of one day.

9.4 Discussion

It is undoubtedly disappointing that, after all the efforts directed at solving the optimal timing problem and computing the solution, it transpires that at typical parameter values there is very little incentive for investors to adopt the OTR. The rewards generated by the rule are simply too small to be of interest to a busy investor (although they might be delegated to a computer). In one sense this is the correct result, since it indicates that the autocorrelation is indeed *economically insignificant*, as Fama (1970) puts it.

It is interesting to speculate on what would have happened had the expected reward been a little larger. In this case it might have been worthwhile for some investors to implement the OTR. The result would be that after falls in the stock price there would be a lower-than-usual demand for stock as investors delayed for a day. On the other side of the transaction, the symmetry of the problem suggests that sellers of stock will sometimes delay when the price has risen by a large amount. The consequence of these two activities would be non-linear autocorrelation in the daily stock returns: a large fall might trigger another fall the next day, a large rise might trigger another rise, but a medium-sized change in either direction would not

cause either buyers or sellers to delay.

This non-linear autocorrelation is exactly the kind of process which would generate ARCH effects in the daily returns (Bera and Higgins, 1993). Therefore it is tempting to speculate that some investors are actually operating optimal timing rules based on threshold values already. This might be because some investors believe that daily returns are not exactly AR(1), leading to a different timing rule which may be more profitable. Alternatively, it might be that expanding the information set to include other variables such as trading volume leads to a more powerful rule which is profitable. While there is no direct evidence for a better trading rule, it stands to reason that the correlation between trading volume and absolute price change and the time-series properties of trading volume will together combine to give a smaller standard error on the one-period-ahead daily return forecast. This is something to be investigated in the future.

Chapter 10

Summary and Conclusion

The theme of this thesis has been daily price change and trading volume dynamics in a speculative asset. In reviewing the preceding chapters it is clear that the material presented has been relevant to this theme in three ways. First, there has been the theoretical analysis of the price/volume relationship, mainly but not exclusively through the critical appraisal and extension of an existing model. Second, there has been the development of tools to facilitate empirical analysis of this model and of price dynamics more generally. Third, the empirical analysis itself has been conducted.

Theoretical Analysis

The starting point for the analysis of the price/volume relationship is the news-driven model of Tauchen and Pitts (1983). In this model both price change and trading volume per day are driven by the amount of news arriving during the day. Section 1.4 gives an overview of the location of this model in the literature. This overview is structured around a new taxonomy which identifies a class of models such as that of Tauchen and Pitts, known as *pseudo-homogeneous investor* models,

which stand between descriptive and homogeneous investor models on the one side (e.g. the descriptive model of Bachelier (1900) and the homogeneous investor model of Samuelson (1965)) and heterogeneous 'noise' models (e.g. the liquidity trader models of Admati and Pfleiderer, 1988, 1989) on the other.

Members of the class of pseudo-homogeneous models are sufficiently complicated to permit the modeling of disagreement among optimizing investors, while at the same time being sufficiently structured to provide a framework within which to analyse price/volume dynamics. This contrasts with descriptive or homogeneous investor models; in these there may be behavioural optimization, but there can be no disagreement between investors. Heterogeneous investor models, however, are not structured enough to provide much guidance concerning the daily price/volume relationship (Gallant *et al.*, 1992). While it is clear in practice that investors are heterogeneous, it is an empirical issue as to whether this heterogeneity is sufficiently large to invalidate the 'first-order' pseudo-homogeneous approximation: that investors are broadly similar in their aims and disagree only in their interpretation of the public information stock.

Tauchen and Pitts

Chapter 2 considers the pseudo-homogeneous model of Tauchen and Pitts (1983), and its direct antecedents, in detail. Part of the Tauchen and Pitts model, the 'intra-day model', concerns the response of price change and trading volume to a single item of news, in a way that also incorporates the size of the market. Section 2.5 suggests that there are major deficiencies in the intra-day model, some of which are addressed in Chapter 6. Tauchen and Pitts also provide a model for the way in which both price change and trading volume are generated by news on a day-to-day basis, the 'inter-day' model. Their attempts to estimate this model were not very successful, probably due, as they suggested, to the failure of their assumption that

the amount of news per day was independently and identically distributed. The time-dependence of the news-arrival process has been confirmed, at least in part, by studies made subsequently, some of which are discussed in Section 2.6.

Section 2.4 develops the inter-day model to examine in more detail the relationship between squared price change and trading volume. Tauchen and Pitts consider the case where both price change and trading volume have normal distributions and the amount of news per day is log-normal, and they present their results graphically following numerical integration. Section 2.4 shows what can be inferred from general specifications for the random quantities concerned. In particular it is interesting to note that the expected squared price change will tend to be increasing in trading volume, but not necessarily so, particularly at small volumes. An explicit functional form can be found asymptotically in the amount of trading volume: expected squared price change is shown to be linear in trading volume with a χ_1^2 disturbance term. This result, which is independent of the news process, is used to test the Tauchen and Pitts model in Chapter 5.

News

Chapter 3 is an attempt to define news, and on the basis of this definition to consider what is meant in financial markets by 'the quantity of news per day', which is a crucial variable in models such as that of Tauchen and Pitts. A definition of news is proposed which implies that a piece of new information is *not* news to an investor if and only if that investor considers it equally likely have occurred in every possible state of nature. The notion of news 'magnitude' is also discussed.

Chapter 3 goes on to consider a model of information assimilation in which the process of updating beliefs and re-optimizing plans is costly. These costs can cause investors to aggregate information and only sporadically update, according to a trade-off between the cost of updating and the loss of expected utility from having a

sub-optimal plan. The result is that simply counting the amount of new information bits which arrive will tend to over-represent news, since later information bits can partially or totally negate earlier bits if information is aggregated. Consequently it is proposed that the quantity of news per period should be measured by the number of investors who re-optimize their plans during the period. This has the advantage that it is easily proxied, and it simplifies to the number of news bits per period in the case where the updating costs are zero.

Market Microstructure

The final chapter centering on the theoretical analysis of the price/volume relationship is Chapter 6. In this chapter the dynamics of price change and trading volume are considered on a per-news-item basis, i.e. at the level of the intra-day model of Tauchen and Pitts (1983). A model for updating beliefs is proposed which generalizes the Tauchen and Pitts model in order to eliminate the deficiencies highlighted in Section 2.5. This has the effect of liberating the Tauchen and Pitts intra-day model from its very strict independence of price change and trading volume both contemporaneously and from news item to news item.

Crucially, the distribution of investors' beliefs about future prices is found to play an important role in determining both the price change and trading volume subsequent to the arrival of new information. Since this belief distribution changes only slowly through time (because each investor's belief updates locally, i.e. an investor is unlikely to go from being strongly bullish to strongly bearish on the receipt of a single piece of news), the result is a contemporaneous correlation between absolute price change and trading volume which also has time-series properties. A combination of theoretical results and Monte Carlo simulations suggests that the skewness of investors' beliefs impacts mainly on the price change, while dispersion of beliefs impacts mainly on the absolute price change and the trading volume. This

latter result already has some theoretical and empirical support.

New Tools

In the course of this thesis it has been necessary to devise and implement tools for processing price and trading volume data, and for describing optimal investor behaviour under certain circumstances.

Optimal Price Index

A new index for futures prices is described in Chapter 4. The need for a futures price index, which joins together in some fashion the prices of individual contracts, has long been recognized. The simple practice of merely joining together the return series of the near contract ('splicing'), while it is still in widespread use, has the unfortunate side-effect of introducing seasonality into the futures price in addition to that which was already in the spot price. Chapter 4 defines the notion of *optimality* in the futures price index to be the complete absence of any seasonality unrelated to that in the spot price, in frictionless markets. From this definition a condition for the nature of the index weights (as functions of time and expiry dates) follows. This condition is solved in the special cases where there are two and three available contracts, to find the precise expressions for the weights.

Chapter 4 also considers the alternative index proposed by Clark (1973). It contrasts the three methods, using data from the London Stock Exchange and the London International Financial Futures Exchange. In theory Clark's index is shown to lie between the spliced index and the optimal index in its lack of seasonality. In practice, the Clark index and the optimal index perform perfectly and almost identically, with the spliced index clearly showing signs of mis-specification. Other aspects of the Clark index, however, such as its sample-dependence, its requirement

of an extra data item for each contract in each period (the open interest), and its difficulty of calculation, made the optimal index clearly superior.

Contract Rollovers

A second tool for processing the data is a method for removing some of the effects of contract rollover from trading volume in futures markets, described in Section 5.3. An investor's desired holding period will often extend beyond the expiry date of the near contract in a futures market, and yet he may prefer to hold only the near contract because of its superior liquidity. To maintain his position past expiry, the investor has to roll over his contracts by simultaneously closing in the near contract and opening in the next-to-near contract. These contract rollovers, while an integral part of the functioning of the market, are entirely unrelated to news arrival. Therefore they should be removed, if possible, from the trading volume series if that series is to be used in investigating the news arrival process. The simple method of deseasonalization according to the quarterly expiry pattern is not particularly helpful, since rollovers can occur well in advance of the expiry date.

As a proxy for the number of contract rollovers, Section 5.3 proposes a measure calculated from the change in the amount of open interest on the near and the next-to-near contracts, given in eq. (5.3). In the absence of any other effects the rollover of one contract will cause the open interest in the near contract to fall by one, and the open interest in the far contract to rise by one. The proxy counts the number of matched changes in open interest following this pattern. The resulting measure has the seasonal pattern expected, and diminishes appreciably the amount of seasonality in the trading volume series, although it does not eliminate it.

Optimal Timing Rule

The optimal timing rule of Chapters 8 and 9 is a tool which permits an investor to exploit positive first-order autocorrelation in daily returns in a manner which maximizes his expected reward. The first of these chapters describes the timing problem and the solution in the case where speculative returns follow an AR(1) process. In Chapter 7 the evidence, both theoretical and empirical, for the persistence of this process in returns is discussed. The optimal timing rule is applicable to an investor who is committed to buying a speculative asset, but has not yet implemented the decision. It consists of buying immediately if the daily return of the period just ending is greater than or equal to a certain threshold, otherwise delaying and taking the same decision in one period's time. To find the threshold is a complicated business, and much of Chapter 8 is taken up with the mathematics of the solution, which is presented in Proposition 8.4.

Following on from Chapter 8, Chapter 9 discusses the computation of the solution to the optimal timing problem and the magnitudes of rewards which would accrue to an investor following the optimal timing rule. Since the solution is an infinite power-series in an integral operator, its computation is itself a major challenge. Two sections of the chapter are devoted to a discussion of the failure of traditional methods and the complete description of the methods used in their place to find the solution at close to machine accuracy. The chapter then turns to calculating the decision threshold and the expected reward from following the optimal timing rule. It is found that, for typical values of the parameters, the threshold was one or more standard deviations below the mean return, often two or more. This implies that the rule in operation will cause very few delays, and this is confirmed by expected rewards very little in excess of that from buying immediately. It was disappointing to find that, after all the effort of solution and computation, the rule as it stood had little commercial value. On the other hand, this could be interpreted as evidence in

support of efficient markets.

Empirical Analysis

The empirical analysis in this thesis is concentrated in Chapter 5, although Chapters 8 and 9 use descriptive statistics in order to assess the feasibility and performance of the Optimal Timing Rule in practice.

Chapter 5 examines the univariate properties of price change and trading volume, and also the bivariate properties of squared price change and trading volume. It uses two of the tools discussed above, the optimal price index for futures contracts and the method of proxying contract rollovers using changes in open interest. The analysis of the Tauchen and Pitts inter-day model given in Section 2.4 is used to specify the structure of the bivariate relationship.

Overall, the empirical evidence from FTSE-100 contracts traded on the London International Financial Futures Exchange over the period 1985–1994 gives qualified support for the Tauchen and Pitts inter-day model. In particular, the asymptotic distribution for squared price change divided by trading volume does appear to be χ_1^2 , as suggested by Proposition 2.2, although it is necessary to exclude the period of the stock market Crash (1987q4–1988q1). The price/volume dynamics during this period are clearly different from those of the rest of the period, as is made clear in Figure 5.9 (page 101).

However, the relationship between squared price change and trading volume appears to be unstable with respect to price direction, which contradicts the symmetry of the Tauchen and Pitts model regarding good and bad news. This instability also contradicts the survey of Karpoff (1987), who suggests that generally the relationship is stable in futures markets, but not in spot markets (due to the latter's asymmetry of transactions costs for short and long positions). However, the instability in

the price/volume relationship is consistent with recent findings using asymmetric Autoregressive Conditional Heteroskedasticity models, where future price volatility appears to be affected by the direction of the price change. In the light of these conflicting findings, it is not possible to say whether the Tauchen and Pitts model itself is misspecified, or whether the problem lies instead in the precise functional form chosen.

Conclusion

The conclusion of this thesis should be distinguished from its contribution. The contribution has been summarized in the previous sections, and covers theoretical developments, new analytical tools, and new empirical results. This conclusion summarizes what I have learned about the price/volume relationship while researching and writing this thesis.

The primary conclusion of this thesis is that news alone cannot explain daily price/volume dynamics. The main reason for this is that, while price change and trading volume may well be related to the flow of news to the market, as in the model of Tauchen and Pitts (1983), the parameters which govern this relationship are inherently unstable. Therefore to explain the price/volume relationship also requires an explanation of this parameter instability.

Two explanations of parameter instability are given. One represents the currently prevailing orthodoxy—investors are heterogeneous. In this case the parameters fluctuate in response to the weight and inclinations of the various different groups in the market. The other explanation is presented in Chapter 6. It has the advantage that it does not require investors to be heterogeneous and is, in this sense, the simpler explanation.

Chapter 6 suggests that investors update their beliefs about the future price

of a speculative asset in a non-linear fashion, and with reference to the prevailing market-clearing price. The implication is that the joint distribution of price change and trading volume per item of news is parametrized by the cross-sectional distribution of investors' beliefs, as summarized in their reservation prices. Hence, as the distribution of reservation prices changes through time so does the price/volume relationship.

One implication of this model is that there will be time-dependency in both price change and trading volume. It is suggested that price change relates to the skewness of the reservation price distribution, and absolute price change and trading volume to the dispersion. Since the reservation price distribution will change only slowly through time (i.e. a bullish investor will typically remain bullish in the next period), so we should expect to see positive autocorrelation in price changes, absolute price changes and trading volume, irrespective of the characteristics of the news-arrival process.

This has implications for empirical work in speculative markets and, in consequence, for the behaviour of investors. One important task is to separate out news effects from micro-structure effects, and to do this we need a series proxying the quantity of news per day. Chapter 3 provides a good reason for using the number of transactions recorded by stockbrokers, in preference to more direct measures such as the number of stories carried by information services. But we also need a prior specification for the price/volume relationship which incorporates parameter instability correlated across price change and trading volume.

Finally, it is possible that, following the development of these more general models, the range of variation of the parameters of the return process becomes clearer, and we become better able to identify, in advance, periods in which the autocorrelation coefficient and the return mean and variance are all likely to be unusually large. This would be, in terms of the reservation price model, at times when the

distribution of investors' reservation prices was substantially skewed and dispersed. At these times the reward from following the optimal timing rule of Chapters 8 and 9 might be substantial, although this reward would accrue most surely only to those who implement first, since the gradual adoption of timing strategies will lead in time to a change in the return dynamics.

THE END.

Bibliography

- Admati, A. and Pfleiderer, P. (1988). A theory of intraday patterns: Volume and price variability. *Review of Financial Studies*, **1**, 3–10.
- Admati, A. and Pfleiderer, P. (1989). Divide and conquer: A theory of intraday and day of the week mean effects. *Review of Financial Studies*, **2**, 189–223.
- Akgiray, V. and Booth, G. (1986). Stock price processes with discontinuous time paths: An empirical examination. *Financial Review*, **21**, 163–84.
- Akgiray, V. and Booth, G. (1987). Compound distribution models of stock returns: An empirical comparison. *Journal of Financial Research*, **10**, 269–79.
- Alexander, S. (1961). Price movements in speculative markets: Trends or random walks. *Industrial Management Review*, **2**, 7–26.
- Alexander, S. (1964). Price movements in speculative markets: Trends or random walks, no. 2. *Industrial Management Review*, **5**, 25–46.
- Arrow, K. (1964). The role of securities in the optimal allocation of risk-bearing. *Review of Economic Studies*, **31**, 91–6.
- Bachelier, L. (1900). *Théorie de la Speculation*. Paris: Gauthier-Villars. Reprinted in Cootner (1964).

- Banz, R. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, **9**, 3–18.
- Bera, A. and Higgins, M. (1993). Arch models: Properties, estimation and testing. *Journal of Economic Surveys*, **7**, 305–66.
- Berry, T. and Howe, K. (1994). Public information arrival. *Journal of Finance*, **49**, 1331–46.
- Black, F. (1986). Noise. *Journal of Finance*, **41**, 529–43.
- Blake, D. (1996). Efficiency, risk aversion and portfolio insurance: An analysis of financial asset portfolios held by investors in the united kingdom. *Economic Journal*, **106**, 1175–92.
- Blattberg, R. and Gonedes, N. (1974). A comparison of the stable and student distribution as statistical models for stock prices. *Journal of Business*, **47**, 244–80.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, 307–27.
- Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, **69**, 542–7.
- Bollerslev, T., Chou, R., and Kroner, K. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*, **52**, 5–59.
- Brock, W. (1986). Distinguishing random and deterministic systems: Abridged version. *Journal of Economic Theory*, **40**, 168–95.
- Brock, W., Lakonishok, J., and LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance*, **47**, 1731–64.

- Buckle, M., Clare, A., and Thomas, S. (1994). Predicting the returns from stock index futures. Discussion Paper 94-04, CERF, Brunel, The University of West London.
- Campbell, J. and Kyle, A. (1993). Smart money, noise trading, and stock price behaviour. *Review of Economic Studies*, **60**, 1-34.
- Campbell, J., Grossman, S., and Wang, J. (1993). Trading volume and serial correlation in stock returns. *Quarterly Journal of Economics*, **108**, 905-39.
- Clark, P. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, **41**, 135-159.
- Cohen, K., Hawawini, G., Maier, S., Schwartz, R., and Whitcomb, D. (1980). Implications of microstructure theory for empirical research on stock price behaviour. *Journal of Finance*, **35**, 249-257.
- Conrad, J. and Kaul, G. (1988). Time-variation in expected returns. *Journal of Business*, **61**, 409-25.
- Cootner, P. (1964). *The Random Character of Stock Market Prices*. Cambridge: MIT Press.
- Copeland, T. (1976). A model of asset trading under the assumption of sequential information arrival. *Journal of Finance*, **31**, 1149-68.
- Copeland, T. and Weston, J. (1988). *Financial Theory and Corporate Policy*. Reading, MA: Addison-Wesley Publishing Company, 3rd edition. Reprinted with corrections, 1992.
- Corrado, C. and Lee, S.-H. (1992). Filter rule tests of the economic significance of serial dependencies in daily stock returns. *Journal of Financial Research*, **15**, 369-87.

- Cowles, A. (1933). Can stock market forecasters forecast? *Econometrica*, **1**, 309–24.
- Cox, J., Ingersoll, J., and Ross, S. (1981). The relationship between forward and futures prices. *Journal of Financial Economics*, **9**, 321–46.
- Cox, J., Ingersoll, J., and Ross, S. (1985). A theory of the term structure of interest rates. *Econometrica*, **53**, 385–407.
- Cragg, J. and Malkiel, B. (1982). *Expectations and the Structure of Share Prices*. Chicago: University of Chicago Press.
- Cyert, R. and DeGroot, M. (1987). *Bayesian Analysis and Uncertainty in Economic Theory*. London: Chapman and Hall Ltd.
- Darnell, A. and Evans, J. (1990). *The Limits of Econometrics*. Aldershot, England: Edward Elgar Publishing Ltd.
- Davis, P. and Rabinowitz, P. (1984). *Methods of Numerical Integration*. Orlando, Florida: Academic Press Inc., 2nd edition.
- De Grauwe, P. and Dewachter, H. (1990). A chaotic monetary model of the exchange rate. Discussion Paper 466, Centre for Economic Policy Research.
- De Grauwe, P. and Vansanten, K. (1990). Deterministic chaos in the foreign exchange market. Discussion Paper 370, Centre for Economic Policy Research.
- deBondt, W. and Thaler, R. (1985). Does the stock market overreact? *Journal of Finance*, **40**, 793–805.
- deBondt, W. and Thaler, R. (1987). Further evidence on investor overreaction and stock market seasonality. *Journal of Finance*, **42**, 557–81.
- Debreu, G. (1959). *The Theory of Value*. New York: John Wiley & Sons, Ltd.

- deLong, J., Shleifer, A., Summers, L., and Waldmann, R. (1989). The size and incidence of the losses from noise trading. *Journal of Finance*, **44**, 681–96.
- deLong, J., Shleifer, A., Summers, L., and Waldmann, R. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, **98**, 703–38.
- Diebold, F. (1986). Comment on modelling the persistence of conditional variance. *Economics Reviews*, **5**, 51–6.
- Dixit, A. (1990). *Optimization in Economic Theory*. Oxford: Oxford University Press, 2nd edition.
- Dixit, A. and Pindyck, R. (1994). *Investment Under Uncertainty*. Princeton, New Jersey: Princeton University Press.
- Duffie, D. (1996). *Dynamic Asset Pricing Theory*. Princeton, New Jersey: Princeton University Press, 2nd edition.
- Elton, E. and Gruber, M. (1970). Marginal stockholder tax rates and the clientele effect. *Review of Economics and Statistics*, **52**, 68–74.
- Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1007.
- Engle, R. and Ng, V. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, **48**, 1749–78.
- Engle, R., Ito, T., and Lin, W. (1990). Meteor-showers or heat waves: Heteroskedastic intradaily volatility in the foreign-exchange market. *Econometrica*, **58**, 525–42.
- Epps, T. and Epps, M. (1976). The stochastic dependence of security price changes and transaction volumes: Implications for the Mixture-of Distributions hypothesis. *Econometrica*, **44**, 305–321.

- Fama, E. (1965). The behaviour of stock market prices. *Journal of Business*, **38**, 34-105.
- Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, **25**, 383-420.
- Fama, E. (1991). Efficient capital markets: II. *Journal of Finance*, **46**, 1575-617.
- Fama, E. and French, K. (1988). Permanent and temporary components of stock prices. *Journal of Political Economy*, **96**, 246-73.
- Fellner, W. (1965). *Probability and Profit*. Homewood, Illinois: Richard D. Irwin, Inc.
- Fisher, L. (1966). Some new stock-market indices. *Journal of Business*, **39**, 191-225.
- Frankel, J. and Froot, K. (1990). Chartists, Fundamentalists and trading in the foreign exchange market. *American Economic Review*, **80**, 181-5.
- French, K. and Roll, R. (1986). Stock return variances: The arrival of information and reaction of traders. *Journal of Financial Economics*, **17**, 5-26.
- Friend, I. and Blume, M. (1975). The demand for risky assets. *American Economic Review*, **65**, 900-22.
- Gallant, A., Rossi, P., and Tauchen, G. (1992). Stock prices and volume. *Review of Financial Studies*, **5**, 199-242.
- Godfrey, M., Granger, C., and Morgenstern, O. (1964). The random walk hypothesis of stock market behaviour. *Kylos*, **17**, 1-30.
- Gordon, M. (1962). The savings investment and valuation of a corporation. *Review of Economics and Statistics*, **44**, 37-51.

- Gordon, M. and Shapiro, E. (1956). Capital equipment analysis: The required rate of profit. *Management Science*, **3**, 102–10.
- Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37**, 424–38.
- Granger, C. and Morgenstern, O. (1963). Spectral analysis of New York stock market prices. *Kylos*, **16**, 1–27.
- Granger, C. and Morgenstern, O. (1970). *Predictability of Stock Market Prices*. Lexington MA: Heath Lexington Books.
- Grossman, S. (1976). On the efficiency of competitive stock markets where traders have diverse information. *Journal of Finance*, **31**, 573–86.
- Grossman, S. and Miller, M. (1988). Liquidity and market structure. *Journal of Finance*, **43**, 617–33.
- Grossman, S. and Stiglitz, J. (1976). Information and competitive price systems. *American Economic Review*, **66**, 246–53.
- Hall, J., Brorsen, B., and Irwin, S. (1989). The distribution of futures prices: A test of the Stable Paretian and Mixture of Normals hypotheses. *Journal of Financial and Quantitative Analysis*, **24**, 105–16.
- Harris, L. (1987). Transaction data tests of the Mixture of Distributions Hypothesis. *Journal of Financial and Quantitative Analysis*, **22**, 127–41.
- Harrison, M. and Kreps, D. (1979). Martingales and arbitrage in multi-period securities markets. *Journal of Economic Theory*, **20**, 381–408.
- Hibbert, J. (1995). UK equity trading costs. Copyright Barrie & Hibbert Ltd., confidential.

- Hillier, B. and Rougier, J. (1996). Real business cycles, investment finance, and multiple equilibria. Discussion Paper 164, Department of Economics, University of Durham.
- Holmes, P. (1993). *The Economics of Stock Index Futures: Theory and Evidence*. Ph.D. thesis, Brunel, The University of West London.
- Hoyle, K. and Ingram, R. (1991). *Statistics for Business*. Oxford: Butterworth-Heinemann Ltd.
- Hutson, V. and Pym, J. (1980). *Applications of Functional Analysis and Operator Theory*. London: Academic Press Inc. Ltd.
- Ito, T., Engle, R., and Lin, W. (1992). Where does the meteor shower come from: The role of stochastic policy coordination. *Journal of International Economics*, **32**, 221–40.
- Johnson, J. (1984). *Econometric Methods*. New York: McGraw-Hill, Inc., 3rd edition.
- Johnson, N. and Kotz, S. (1969). *Discrete Distributions*. New York: John Wiley & Sons.
- Karpoff, J. (1987). The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis*, **22**, 109–26.
- Kendall, M. (1953). The analysis of economic time-series, Part I: Prices. *Journal of the Royal Statistical Society*, **96**, 11–25.
- Kendall, M. and Stuart, A. (1969). *The Advanced Theory of Statistics*, volume 1. London: Charles Griffin & Co. Ltd., 3rd edition.
- Kim, D. and Kon, S. (1994). Alternative models for the conditional heteroscedasticity of stock returns. *Journal of Business*, **67**, 563–98.

- Knight, F. (1920). *Risk, Uncertainty and Profit*. New York: Harper & Row.
- Kon, S. (1984). Models of stock returns: A comparison. *Journal of Finance*, **39**, 147–65.
- Lamoureux, C. and Lastrapes, W. (1990). Heteroskedasticity in stock return data: Volume versus GARCH effects. *Journal of Finance*, **45**, 221–9.
- LeRoy, S. (1989). Efficient capital markets and Martingales. *Journal of Economic Literature*, **27**, 1583–1621.
- Lindley, D. (1985). *Making Decisions*. London: John Wiley & Sons, 2nd edition.
- Lo, A. and MacKinlay, A. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, **1**, 41–66.
- Lucas, Jr., R. (1978). Asset prices in an exchange economy. *Econometrica*, **46**, 1429–45.
- Lux, T. (1995). Herd behaviour, bubbles and crashes. *Economic Journal*, **105**, 881–97.
- Mandelbrot, B. (1963a). New methods in Statistical Economics. *Journal of Political Economy*, **71**, 421–40.
- Mandelbrot, B. (1963b). The variation of certain speculative prices. *Journal of Business*, **36**, 394–419.
- Mandelbrot, B. (1973). Comments on: "A subordinated stochastic process model with finite variance for speculative prices," by Peter K. Clark. *Econometrica*, **41**, 157–9.
- Mandelbrot, B. and Taylor, H. (1967). On the distribution of stock price differences. *Operations Research*, **15**, 1057–62.

- Mardia, K., Kent, J., and Bibby, J. (1979). *Multivariate Analysis*. London: Harcourt Brace & Co.
- Milgrom, P. and Stokey, N. (1982). Information, trade and common knowledge. *Journal of Economic Theory*, **26**, 17–27.
- Mitchell, M. and Mulherin, J. (1994). The impact of public information on the stock market. *Journal of Finance*, **49**, 923–50.
- Moschetti, G. (1996). Public information and trading volume. does trading generate trading? Discussion Paper 9629, Department of Economics, University of Southampton.
- Niederhoffer, V. and Osborne, M. (1966). Market making and reversal on the stock exchange. *Journal of the American Statistical Association*, **61**, 897–916.
- Osborne, M. (1959). Brownian motion in the stock market. *Operations Research*, **7**, 145–73.
- Patel, J. and Wolfson, M. (1984). The intraday speed of adjustment of stock prices to earning and dividend announcements. *Journal of Financial Economics*, **13**, 23–52.
- Poterba, J. and Summers, L. (1988). Mean reversion in stock prices: Evidence and implications. *Journal of Financial Economics*, **22**, 27–59.
- Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (1992). *Numerical Recipes in C*. Cambridge: Cambridge University Press, 2nd edition.
- Roll, R. (1977). A critique of the Asset Pricing Theory's tests. *Journal of Financial Economics*, **6**, 129–76.
- Ross, S. (1988). *A First Course in Probability*. New York: Macmillan, 3rd edition.

- Rougier, J. (1993). The impact of margin-traders on the distribution of daily stock returns: The London Stock Exchange. *Applied Financial Economics*, **3**, 325–8.
- Rougier, J. (1995). Understanding portfolio composition. Discussion Paper 146, Department of Economics, University of Durham.
- Rougier, J. (1996). An optimal price index for stock index futures contracts. *Journal of Futures Markets*, **16**, 189–99.
- Samuelson, P. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, **6**, 41–9.
- Samuelson, P. (1972). Mathematics of speculative price. In R. Day and S. Robinson, editors, *Mathematical Topics in Economic Theory and Computation*, pages 1–42. Philadelphia: SIAM.
- Samuelson, P. (1973). Proof that properly discounted present value of assets vibrate randomly. *Bell Journal of Economics*, **4**, 369–74.
- Samuelson, P. (1983). *Foundations of Economic Analysis, Enlarged Edition*. Cambridge: Harvard University Press.
- Sargeant, T. (1987). *Dynamic Macroeconomic Theory*. Cambridge, MA: Harvard University Press.
- Shalen, C. (1993). Volume, volatility, and the dispersion of beliefs. *Review of Financial Studies*, **6**, 405–34.
- Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, **19**, 425–42.
- Shiller, R. (1984). Stock prices and social dynamics. *Brookings Papers on Economic Activity*, pages 457–98.

- Simms, C. (1980). Macroeconomics and reality. *Econometrica*, **48**, 1–47.
- Sloan, I. (1992). Numerical integration in high dimensions—the Lattice Rule approach. In T. Espelid and A. Genz, editors, *Numerical Integration: Recent Developments, Software and Applications*, volume 357 of *NATO ASI, Series C*, pages 55–69. Dordrecht, Neth: Kluwer Academic Publishers.
- Stock, J. (1988). Estimating continuous time processes subject to time deformation. *Journal of the American Statistical Association*, **83**, 77–85.
- Stokey, N., Lucas, R., and Prescott, E. (1989). *Recursive Models in Economic Dynamics*. Cambridge MA: Harvard University Press.
- Sweeney, R. (1986). Beating the foreign exchange market. *Journal of Finance*, **41**, 163–82.
- Sweeney, R. (1988). Some new filter rule tests: Methods and results. *Journal of Financial and Quantitative Analysis*, **23**, 285–300.
- Tauchen, G. and Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica*, **51**, 485–505.
- Thompson, R., Olsen, C., and Dietrich, J. (1987). Attributes of news about firms: An analysis of firm-specific news reported in the Wall Street Journal Index. *Journal of Accounting Research*, **25**, 245–74.
- Tirole, J. (1982). On the possibility of speculation under rational expectations. *Econometrica*, **50**, 1163–81.
- Tobin, J. (1958). Liquidity preference as behaviour towards risk. *Review of Economic Studies*, **25**, 65–86.
- Tobin, J. (1969). Comment on Borch and Feldstein. *Review of Economic Studies*, **36**, 13–14.

- Varian, H. (1985). Divergence of opinion in complete markets: A note. *Journal of Finance*, **40**, 309-17.
- Varian, H. (1992). *Microeconomic Analysis*. New York: W.W. Norton & Co. Inc., 3rd edition.
- West, K. (1988). Bubbles, fads and stock price volatility tests: A partial evaluation. *Journal of Finance*, **43**, 639-60.
- Whittaker, J. (1990). *Graphical Methods in Applied Mathematical Multivariate Statistics*. London: John Wiley & Sons, Ltd.
- Williams, J. (1938). *The Theory of Investment Value*. Cambridge, MA: Harvard University Press.
- Working, H. (1934). A random difference series for use in the analysis of time series. *Journal of the American Statistical Association*, **29**, 11-24.
- Yadav, P. and Pope, P. (1990). Stock index futures arbitrage: International evidence. *Journal of Futures Markets*, **10**, 573-603.
- Yadav, P. and Pope, P. (1992). Intraday and intraweek seasonalities in stock market risk premia: Cash and futures. *Journal of Banking and Finance*, **16**, 233-70.
- Yadav, P. and Pope, P. (1994). Stock index futures mispricing: Profit opportunities or risk premia? *Journal of Banking and Finance*, **18**, 921-53.
- Zarowin, P. (1989). Does the stock market overreact to corporate earnings information? *Journal of Finance*, **44**, 1385-99.

