TIME-VARYING VOLATILITY AND RETURNS ON ORDINARY SHARES:

AN EMPIRICAL INVESTIGATION

by

ENRIQUE SENTANA IVAÑEZ

London School of Economics and Political Science

Thesis submitted for the degree of
Doctor of Philosophy
at the
University of London

1991

To Pura

ABSTRACT

This research investigates various issues relating to the level and volatility of returns on ordinary shares. In particular, we have looked at the relation over time between volatility and risk premia, both at a univariate and multivariate levels.

We also look at the links between stock markets over the world, and whether they are integrated. We evaluate the role of measurable economic variables in explaining asset price (co-)movements over time. Our model combines an APT factor pricing approach with a GARCH-type parameterisation of the volatility of the factors. These can be "observable" (i.e. related to economic variables), "unobservable" and country-specific. Estimates of these factors and their time-varying variances are obtained using a Kalman Filter-based Full Information Maximum Likelihood method. Using monthly data on sixteen markets it is found that idiosyncratic risk is significantly priced, and that the "price of risk" is not common across countries, which rejects the null of global capital market integration. Another empirical finding is that most of the correlation between markets is accounted for by the "unobservables".

The econometric background to the conditionally heteroskedastic factor model employed is also analysed. We find that the matrix of factor loadings is unique under orthogonal transformations, and as a result, that it is possible to evaluate the separate contribution of the different factors to the risk premia if time-variation in the volatility of the factors is recognised. We also obtain a full characterisation of this model under the assumption that the conditional distribution is multivariate t, (the normal being a special case), and GARCH formulations for the conditional variances.

A fundamental question in Finance is whether the stock market satisfies the Efficient Market Hypothesis. In this regard, we explore whether lagged variables that help predict stock returns are merely proxying for mis-measured risk. Three different ways of measuring risk are employed (i.e. semi-parametric, GARCH and lagged squared returns). In an application to Japanese data, four key

predictor economic variables are shown to have non-trivial additional forecasting power irrespective of how risk is measured. Interestingly, unlike the US, the level of the lagged dividend yield is not positively correlated with returns in either Japan or South Korea. Moreover, there is no consistent relationship between expected volatility and excess returns.

Another interesting topic is the hypothesis that the degree of autocorrelation shown by high frequency stock returns may change with volatility. This may result from non-trading effects, feedback trading strategies or variable risk aversion. Results using a century of daily data suggest that when volatility is low there tends to be positive autocorrelation in returns, but this serial correlation can become negative during very volatile episodes. Our results also suggest that returns are more likely to exhibit negative serial correlation after price declines.

Finally, a new Quadratic ARCH model for the conditional variance of a time series is introduced, and interpreted as the quadratic projection of the square innovations on information. Since it nests the original ARCH model and several of its extensions, its statistical properties are very similar, while avoiding some of their criticisms. In an application to a century of daily US stock returns, QARCH models provide a better representation of the data by capturing the leverage effect (i.e. volatility is higher following price declines than after rises). QARCH models are also able to capture this asymmetry in a multivariate context: in a factor model for monthly excess returns on 26 industrial UK sectors, the common factor (which is highly correlated with the FTA500) also shows a significant leverage effect.

ACKNOWLEDGEMENTS

I am deeply grateful to several people who have helped me enormously throughout my research. Firstly, my supervisor Sushil Wadhwani has had a decisive influence on this thesis. He introduced me to the subject (and to empirical research in general) and has always had helpful suggestions of ways to solve the problems I encountered. This thesis owes very much to his detailed comments and invaluable suggestions. I wish to thank him more than I could possibly say.

I am also very much indebted to Mervyn King, from whom I learned much about the theory and practice of financial markets, and whose insights led me to tackle problems that I would not have otherwise explored. I also want to thank him for giving me the opportunity to work at the LSE Financial Markets Group. Very special thanks are due also to Andrew Harvey and Esther Ruiz, whose ideas have contributed very significantly to several sections of this thesis.

Special thanks are also due to John Campbell, who has read and helped improve many parts of this dissertation and Rob Engle, who also provided very useful comments on some sections of this research. Discussions with Antonis Demos, Mushtaq Shah and other colleagues at the Financial Markets Group, as well as comments from participants at LSE Capital Markets and Econometrics workshops have also been very important for several aspects of this work.

Fischer Black, Campbell Harvey, Xavier Freixas, Lars-Peter Hansen, Alessandro Penati, Bob Shiller, Guglielmo Weber, participants at the 1990 NBER Conference on Financial Markets, the World Congress of the Econometric Society, the 1990 Meeting of the European Economic Association, as well as seminar audiences at the Bank of Spain, Brown, Malinvaud Seminar in Paris, Nuffield College, LBS and Tilburg made useful comments on an earlier draft of Chapter 1. Manuel Arellano, Jan Magnus and Neil Shephard also read several sections from Chapter 2, for which I am most grateful. I also wish to thank Charlie Bean, Miguel Delgado, David Hendry, Peter Robinson, Larry Summers and participants at the 1989 LSE Financial Markets Group's Conference on the Econometric of Financial Markets, the 1989 European Meeting of the

Econometric Society and the 1990 Royal Economic Society Conference for helpful comments on an earlier draft of Chapter 3. I am also extremely grateful to Charles Goodhart, Ailsa Röell, G.W. Schwert, participants at the 1990 Paris Conference on ARCH Models and the 1991 Royal Economic Society Conference for providing useful comments to previous versions of Chapter 4. Similarly, Anil Bera, Russell Davidson, Frank Diebold, Steve Durlauf, Christian Gourieroux, Ludger Hentschel, Theo Nijman, Peter Robinson and participants at the 1991 Econometric Study Group Conference made helpful comments on the material in Chapter 5.

Financial support from the Bank of Spain, the British Council and the Spanish Ministry of Education and Science for carrying out this dissertation is also gratefully acknowledged. I also wish to thank for the use of their facilities the Centre for Labour Economics at the early stages of this work, the Financial Markets Group for most of this period, and the Department of Economics at LSE during the last part of my research. Thanks are due to June Jarman for most efficient secretarial assistance.

Last, but by no means least, I wish to thank my family, and especially my wife not only for her enormous patience and encouragement, but also for her help during the years I worked on this thesis.

PUBLICATIONS RELATED TO THIS THESIS

King, M.A., E. Sentana, and S.B. Wadhwani (1990): "A Heteroskedastic Factor Model of Asset Returns and Risk Premia with Time Varying Volatility: An Application to Sixteen World Stock Markets", LSE Financial Markets Group Discussion Paper 80.

Sentana, E. (1991a): "Identification of Multivariate Conditionally Heteroskedastic Latent Factor Models", mimeo, London School of Economics.

Harvey, A.C., Ruiz, E. and Sentana, E. (1990): "Unobservable Component Time Series Models with ARCH Disturbances", Journal of Econometrics, forthcoming.

Sentana, E. and S.B. Wadhwani (1991): "Semi-Parametric Estimation and the Predictability of Stock Returns: Some Lessons from Japan", Review of Economic Studies 58, 547-562.

Sentana, E. and Wadhwani, S.B. (1991): "Feedback Traders and Stock Return Autocorrelations: Evidence from a Century of Daily Data", Economic Journal, forthcoming March 1992.

Sentana, E. (1991b): "Quadratic ARCH Models: A Potential Re-interpretation of ARCH Models", LSE Financial Markets Group Discussion Paper 122.

CONTENTS

Abstract 3

Acknowledgements 5

Publications Related to this Thesis 7

Contents 8

List of Tables 12

List of Figures 13

Introduction 14

Chapter 1: Volatility and Links between National Stock Markets

1.1 Introduction 19

1.2 A Heteroskedastic Factor Model of Multivariate Asset Returns and Risk Premia with Time-varying Volatility 20

1.2.1 Intertemporal Asset Pricing 21

1.2.1.1 Theoretical Set-up 21

1.2.1.2 Modelling Risk Premia 23

1.2.1.3 Relationship with Other Models 24

1.2.2. Factor Specification 25

1.2.3 Identification 28

1.2.4 State-Space Representation and Kalman Filtering 29

1.2.5 Estimation 31

1.3 Empirical Application 32

1.3.1 Vector Autoregression for the Economic Variables 32

1.3.2 Estimation of the Joint Factor and Multivariate Returns Model 33

1.4 Links Between Stock Markets 37

1.4.1 Assessing Global Capital Market Integration 37

1.4.2 The Contribution of Observables to the Covariance Matrix of Returns 39

1.4.3 Changing Links between National Stock Markets 41

1.5 Conclusions 42

Endnotes 44

Data Appendix 1 47

Tables and Figures 48

Chapter 2: Identification and Estimation of Conditionally Heteroskedastic Latent Factor Models

2.1	Introduction	64
2.2	A Multivariate Conditionally Heteroskedastic Latent Factor Model	66
2.3.	Sufficiency Conditions for Identification	68
2.4.	Kalman Filter Equations	71
2.5	Maximum Likelihood Estimation	75
2.6	Introducing ARCH and GARCH-type Effects	75
2.7	Numerical Simulations	81
2.8	Conclusions	86
	Endnotes	89
	Appendix 2: Proofs	91
	Tables	93

Chapter 3: Semi-Parametric Estimation and the Predictability of Stock Market Returns: Some Lessons from Japan

3.1	Introduction	100
3.2	Theoretical Considerations	101
3.2.1	The Efficient Markets Model vs the "Fads" Model	101
3.2.2	Dividend Yields and Stock Returns	102
3.2.3	The Effect of Liquidity	103
3.2.4	Inflation and Stock Returns	103
3.2.5	Autocorrelation in Returns	104
3.3	Methodology	105
3.3.1	Alternative Hypothesis	105
3.3.2	Modelling the Variance	105
3.4	Empirical Results	108
3.4.1	OLS Estimates using the Proxy Method	108
3.4.2	GARCH-M Estimates	110
3.4.3	Semi-Parametric Estimates	111
3.5	Further Explorations	112
3.5.1	Extending the Sample	112
3.5.2	Evidence from South Korea	113
3.5.3	Time-Variation in the "Price of Risk"	114
3.6	Implications and Conclusions	114
	Endnotes	116

	Data Appendix 3	118
	Tables and Figures	119
Chapter 4:	**Feedback Traders and Stock Return Autocorrelations: Evidence from a Century of Daily US Returns**	
4.1	Introduction	130
4.2	Theoretical Arguments for Serially Correlated Returns	131
4.2.1	Feedback Traders and Serial Correlation	132
4.2.2	Non-Synchronous Trading and Serial Correlation	133
4.3	Empirical Evidence	134
4.3.1	Evidence Based on Hourly Data Around the October 1987 Crash	134
4.3.2	Evidence Using Daily Data from 1885-1988	136
4.4	Some Further Explorations	138
4.4.1	Is There More Positive Feedback Trading After Market Declines?	138
4.4.2	Margin Requirements and the Extent of Positive Feedback Trading	139
4.5	Conclusions	140
	Endnotes	142
	Tables	144
Chapter 5:	**Quadratic ARCH Models: A Potential Reinterpretation of ARCH Models**	
5.1	Introduction	147
5.2	The QARCH Model	150
5.2.1	Definition	150
5.2.2	QARCH Models and Quadratic Projections	153
5.2.3	QARCH Models and Second-Order Taylor Approximations	154
5.2.4	Estimation and Testing for QARCH Effects	155
5.2.5	GQARCH(p,q) models	156
5.3	Properties of the QARCH Model	157
5.3.1	Stationarity Conditions and Moments of the Unconditional Distribution	157
5.3.2	Dynamic Correlation Structure	159

5.4	An Empirical Application to Daily Stock Returns	161
5.5	Multivariate Extensions	164
5.6	Conclusions	168
	Endnotes	171
	Appendix 5: Proofs	178
	Data Appendix 5	181
	Tables and Figures	182

Conclusions 187

References 191

LIST OF TABLES

Table 1.1 Matrix of Observable Factor Loadings. 48

Table 1.2 Observable Factor Score Weights. 49

Table 1.3 Estimates of B and τ. 50

Table 1.4 Unobservable Factor Score Weights and Market Capitalization. 51

Table 1.5 Estimates of the Price of Idiosyncratic Risk. 52

Table 1.6 Likelihood Ratio Tests for Common Price of Risk. 53

Table 1.7 Proportion of Variance of Excess Returns Explained by Economic Factors. 54

Table 1.8 Granger-Causality Tests of Explanatory Variables for Dividends. 55

Table 2.1 Maximum Likelihood Estimates. 93

Table 2.2 Factor Scores Mean Square Errors. 94

Table 2.3 Correlations between Factor Scores Estimates. 95

Table 2.4 Sample Averages of Conditional Variances and Likelihood Function. 96

Table 2.5 Principal Components Based Estimates. 97

Table 2.6 Factor Scores Mean Square Errors and Correlations. 98

Table 3.1 Average Dividend Yield and Returns. 119

Table 3.2 Excess Returns Equations - Proxy Estimates. 120

Table 3.3 Excess Returns Equations - GARCH Estimates. 121

Table 3.4 Excess Returns Equations - Semi-Parametric Estimates. 122

Table 3.5 Excess Returns Equations - South Korea. 123

Table 4.1 Serial Correlation in Returns: Estimates using Hourly Data. 144

Table 4.2 EGARCH Model: Parameter Estimates. 145

Table 5.1 GQARCH(1,2) and GARCH(1,2) Parameter Estimates: US Daily Stock Returns 1885-1988. 182

Table 5.2 GQARCH(1,2) and GARCH(1,2) Parameter Estimates: US Daily Stock Returns 1962-1988. 183

Table 5.3 GQARCH(1,2) Conditional Moment Tests. 184

Table 5.4 GQARCH(1,1) and GARCH(1,1) Parameter Estimates: UK Monthly Excess Stock Returns. 185

LIST OF FIGURES

Figure 1.1	Equally-weighted Average of Individual Cross-Country Correlation Coefficients (over preceding 12 months).	56
Figure 1.2a	First Unobservable Factor.	57
Figure 1.2b	Second Unobservable Factor.	58
Figure 1.3a	Conditional Covariance between US and UK Markets.	59
Figure 1.3b	Conditional Covariance between US and Japanese Markets.	60
Figure 1.4	Decomposition of the Conditional Variance of US Stock Market.	61
Figure 1.5	Equally-weighted Average of Conditional Correlations between Markets.	62
Figure 3.1a	Coefficient on Dividend Yield.	124
Figure 3.1b	Coefficient on Interest Rate.	124
Figure 3.1c	Coefficient on Liquidity.	125
Figure 3.1d	Coefficient on Lagged Risk Premium.	125
Figure 3.2a	Volatility vs Changes in Dividend Yield.	126
Figure 3.2b	Volatility vs Changes in Interest Rate.	126
Figure 3.2c	Volatility vs Changes in Liquidity.	127
Figure 3.2d	Volatility vs Changes in Lagged Excess Returns.	127
Figure 3.3	Implicit Price of Risk.	128
Figure 5.1a	Conditional Variance of Common Factor: GQARCH(1,1) Gaussian Estimates.	186
Figure 5.1b	Conditional Variance of Common Factor: GARCH(1,1) Gaussian Estimates.	186

INTRODUCTION

That there are variations over time in the volatility of stock markets has been suggested for quite some time (see e.g. Fama (1965)) and, after the experience of the crash of October 1987, it is nowadays a widely recognised fact. Given that the concept of volatility and its relationship with risk is at the cornerstone of many financial theories, it is perhaps surprising that it has not been until fairly recently that applied researchers in finance have begun to incorporate time-variation in volatility in their work. The purpose of this research project is to investigate various issues relating to the level and volatility of returns on ordinary shares.

In recent years, a large family of statistical models for the variation over time in conditional variances has grown up, mostly, but by no means only, around Engle's ARCH (Autoregressive Conditional Heteroskedastic) model, and numerous applications to financial time series have already appeared (see Bollerslev, Chou and Kroner (1990) for a recent survey). By and large, though, most applied work on the stock market has been concerned with a single market: the New York Stock Exchange. Given that many issues in finance, and in particular, asset pricing theories, are related to the variances and covariances of many assets, it is of substantial practical importance to have manageable models for the time-variation in covariances between assets.

One such issue is the links between national stock markets, which have been attracting increasing attention recently. Casual observation indicates that there are periods when markets seem to move in unison - the 1987 crash being an obvious example - and others when the correlation between them is low. An empirical objective of the study is to account for the time-variation in the covariance matrix of returns on sixteen different markets, and in particular, to evaluate the role of measurable economic variables in explaining asset price (co-)movements over time. In order to identify the main sources of changing volatility, excess returns are assumed to depend both on innovations in observable economic variables, on common unobservable factors and on idiosyncratic (or country-specific) noise. This multivariate factor analytic formulation allows us to obtain a parsimonious representation of the conditional covariance matrix of returns in terms of the changing volatility of the underlying components. However, if idiosyncratic risk can be eliminated by holding a well diversified portfolio, the risk premium on an asset should only depend on the volatility of the common factors. Testing the APT-type cross equation restrictions enables us to examine the issue of whether stock markets over the world are integrated.

The econometric background to the multivariate conditionally heteroskedastic factor model employed is also the subject of this investigation. In particular the issues of identification, estimation, modelling of time-variation and distributional assumptions are considered. Identification is not only a fundamental statistical question, but also a particularly important one from the economic point of view in the context of Arbitrage Pricing Theory-type models, as it is of interest to evaluate the contribution of the different factors to asset risk premia. Similarly, since it is a generally recognised stylised fact that we observe more extreme returns on shares that a Gaussian model can generate, even after accounting for volatility clustering, the form of the distribution assumed is also of substantial practical importance.

A fundamental question in Finance is whether the stock market satisfies the Efficient Market Hypothesis (EMH), which states that current prices reflect all publicly available information. However, in recent years a growing number of researchers have found that stock market returns are predictable (see e.g. Fama and French (1988a,b), Poterba and Summers (1988)). These findings, though, do not necessarily point to the rejection of the EMH because they could be explained by an appeal to time-varying risk premia. In this regard, we attempt to explore whether lagged variables that help predict Japanese stock returns are merely proxying for mis-measured risk. For that reason, three different ways of measuring risk are employed - one that relies on semi-parametric methods, a second based on the GARCH-M model, and a third which just uses lagged squared returns. The first two specifications allow lagged variables to directly affect measured risk.

It has long been known that high frequency (e.g. daily, hourly) returns on shares may show mild positive autocorrelation due to the existence of non-synchronous trading. But even when the non-trading problem is not relevant, if one set of traders follows a feedback trading strategy and react to price changes, then returns will also exhibit serial correlation. In this respect, another interesting topic related to returns and volatility is the hypothesis that the degree of autocorrelation shown by high frequency stock returns may change with volatility. In a "noise" trader model, as volatility rises, "smart" money will allow the feedback traders to have a greater effect on the price, and therefore, the extent of serial correlation will rise. Further, if risk aversion declines with wealth, the extent of positive feedback trading may increase when volatility increases. This link between volatility and autocorrelations is researched using a century of daily US data.

Another well-known stylised fact concerning the relation between returns on shares and their volatility is the so-called leverage effect, i.e. volatility is higher following price declines than after rises of the same magnitude. However, standard versions of the ARCH model are unable to capture this dynamic asymmetry. A generalised, fully quadratic version of the ARCH formulation is introduced which can in principle capture this dynamic asymmetry in stock returns, but at the same time nests the original ARCH model and several of its extensions. To see if this new model represents the data significantly better than a standard ARCH does, two empirical applications are entertained: a univariate one to a century of daily US stock returns, the other multivariate for monthly excess returns on 26 industrial UK sectors.

Chapter 1

VOLATILITY AND LINKS BETWEEN NATIONAL STOCK MARKETS

1.1 Introduction

Attempts to explain the "excess volatility" of stock markets have focused in recent research on the issue of modelling time-varying volatility and the implied stochastic process for expected returns. A large family of statistical models for the variation over time in conditional variances has grown up, especially following Engle's (1982) work on ARCH processes (cf. sections 2.1 and 5.1), but these models do not enable us to disentangle the source of changes in volatility. In this chapter we try to identify those factors that are responsible for changes over time in stock market volatility. Another feature of the existing literature is that, by and large, it is concerned with explaining volatility only in the US stock market. When data on many stock markets are examined (as in section 5.5) an explicit multivariate model of the time-varying variance-covariance matrix of returns is required.

The links between national stock markets have been attracting increasing attention (Roll (1989) surveys the recent literature). There are certain periods - the 1987 stock market crash is a conspicuous example - when markets move in unison, and others when the correlation between them is low (see Figure 1.1). An empirical objective of this study is to account for the time-variation in the covariances between markets. We are especially interested in the role of measurable economic variables in explaining the changes in asset price co-movements over time. Understanding changes in conditional covariances is potentially useful in deciding on appropriate country weightings in global portfolio allocation. The common use of constant covariances between markets in "optimal" portfolio construction might lead to the use of sub-optimal portfolios.

We use data on sixteen national stock markets to estimate a multivariate factor model in which the volatility of returns is induced by changing volatility in the underlying factors. This allows us to obtain a parsimonious representation of the conditional variance-covariance matrix of excess returns as a function of the variances of a small number of factors. Excess returns are assumed to depend both on innovations in observable economic variables and on unobservable factors. We allow the conditional variances of the underlying factors to vary over time and parameterize this in terms of GARCH processes (Bollerslev (1986)). We use recent results from intertemporal asset pricing theory to model the risk premium on an asset as a linear combination of the volatility associated with the factors. Our theoretical model can therefore be understood as a dynamic version of the Arbitrage Pricing Theory. We estimate jointly the model that generates factors from observable economic variables and the equation for equilibrium excess returns.

A significant advantage in assuming that the conditional variances of the factors vary over time is that statistical identification of the factor model is less problematical than is the case in the usual static setting (see section 2.3 and Sentana (1991a)). In conventional factor analytic tests of the APT, the individual risk premia are only identifiable up to an orthogonal transformation (e.g. Dhrymes, Friend and Gultekin (1984)). In our model, the time-variation in the conditional variances enables us to identify the individual risk premia.

Testing the APT-type cross-equation restrictions implied by our intertemporal asset pricing model enables us to examine the issue of whether stock markets over the world are integrated. In particular, we test whether idiosyncratic risk is priced, and whether the "price of risk" associated with each factor is common across countries.

The chapter is organized as follows. In section 1.2, we discuss the theoretical asset pricing model and the estimation procedure. The results obtained by estimating the model using monthly data over the period 1970-1988 are presented in section 1.3. In section 1.4 we examine the issue of capital market integration, and look at the correlations between markets. Our conclusions are stated in section 1.5.

1.2 A Heteroskedastic Factor Model of Multivariate Asset Returns and Risk Premia with Time-varying Volatility

The model that we propose to estimate has four components.

The first is a conditional factor model for excess returns. The second is a dynamic model for the asset risk premia in terms of the changing volatility of the factors. The third is a method for generating factors related to measured economic variables that explain the behaviour of excess returns. The fourth is an econometric specification for the variation over time in the conditional variance of the factors.

1.2.1 Intertemporal Asset Pricing

1.2.1.1 Theoretical set-up

We shall base our theoretical analysis in a world with a countable (possibly infinite) collection of primitive assets, and begin by assuming that there is an underlying probability space of asset payoffs. Let R_{it} be the random (gross) return from a unit investment in asset i during period t, and let R_{ot} be the (gross) payoff on a riskless asset. In order to model the conditional distribution of time t asset returns, a specification of the conditioning information set, I_{t-1}, is required. We assume that this set contains (the sigma algebra generated by) the values of asset returns up to, and including, time t-1, as well as the values of other published statistics known by both agents and the econometrician[1]. We also assume that R_{ot}, which is determined in period t-1, is observed by the econometrician. In our empirical application, we shall work in terms of excess returns, r_{it}, which we measure as (real) excess returns above the riskless payoff, i.e. $r_{it}=R_{it}-R_{ot}$.

Let L_{2t} denote the collection of all random variables defined on the underlying probability space which are measurable with respect to I_{t-1} and have finite conditional variances. Nothing prevents, though, the unconditional variances from being unbounded. Define the conditional mean square inner product of two elements of L_{2t} as $E_{t-1}(p_t q_t)$ and the associated mean square norm $\|p\|_{t-1}=E^{1/2}_{t-1}(p^2)$. Hansen and Richard (1987) show that under mild regularity conditions, L_{2t} is the conditional analogue of a Hilbert space under the conditional mean square inner product.

Let R_{Nt} be a vector containing N (gross) asset returns (with

N possibly infinity), and let $\nu_{Nt}= E_{t-1}(R_{Nt})$, $\Sigma_{Nt}= V_{t-1}(R_{Nt})$ denote the conditional mean and covariance matrix of R_{Nt} (with $[\nu_{Nt}]_i=\nu_{it}$ and $[\Sigma_{Nt}]_{ij}=\sigma_{ijt}$). In terms of the vector of excess returns, r_{Nt}, we denote conditional means by $\mu_{Nt}= E_{t-1}(r_{Nt})= \nu_{Nt}-\nu_{ot}$ so that we can identify them with risk premia. Notice that $V_{t-1}(r_{Nt})= \Sigma_{Nt}$ remains unchanged. We assume that, conditional on I_{t-1}, means, variances and covariances of R_{it}, R_{jt} are bounded, so that they belong to L_{2t}. We assume also that Σ_{Nt} is positive definite for all N (and t), and restrict the stochastic structure of returns further by assuming that the unanticipated component of returns, $\eta_{Nt}=R_{Nt}-\nu_{Nt}=r_{Nt}-\mu_{Nt}$ has a conditional factor representation.

$$\eta_{Nt} = B_N f_t + v_{Nt} \tag{1}$$

where f_t are k (k<N) common factors which capture systematic risk affecting all assets, while v_{Nt} are idiosyncratic terms which reflect unsystematic risk (by construction, v_{Nt} is conditionally orthogonal to f_t). Notice that this is a statement about the cross-sectional dependence of asset returns, and essentially says that the "dimension" of undiversifiable risk is k. The matrix B_N is the associated matrix of (constant) factor loadings (i.e. $[B_N]_{ij}=\beta_{ij}$) which measure the sensitivity of the assets to the common factors. To guarantee $E_{t-1}(\eta_t)=0$ we assume that $E_{t-1}(f_t)=0$, $E_{t-1}(v_t)=0$. We also assume that $E_{t-1}(f_t f_t') = \Lambda_{t:t-1}$ represents a kxk diagonal positive definite matrix of (possibly) time-varying factor variances. Hence, the factors are assumed to be (conditionally) orthogonal. We finally assume that $E_{t-1}(v_{Nt}v_{Nt}') = \Omega_{Nt:t-1}$ is a NxN diagonal positive definite[2] matrix of (possibly) time-varying idiosyncratic variances, and so we are assuming an exact factor structure (cf. Chamberlain and Rothschild (1983)).

One of the attractions of the (conditional) factor model is that it provides a parsimonious representation for the NxN (conditional) covariance matrix of excess returns. In the absence of any structure, a multivariate model of time-varying volatility would consist of time series processes for each of the N(N+1)/2 distinct elements of the covariance matrix (see section 2.1). Here, by contrast, it follows from the above assumptions that:

$$V_{t-1}(r_{Nt}) = B_N \Lambda_{t:t-1} B_N' + \Omega_{Nt:t-1} \tag{2}$$

so that we need only model the (k_1+k_2+N) time-varying processes in the diagonal of $\Lambda_{t|t-1}$ and $\Omega_{Nt|t-1}$. In our empirical application, k=6, N = 16. The number of processes that must be estimated is reduced, therefore, from 136 to only 22, and even more gains could be achieved if $\Omega_{Nt|t-1}$ were constant.

1.2.1.2 Modelling Risk Premia

In order to model the conditional mean returns, or risk premia, we shall make use of recent developments in the intertemporal asset pricing theory literature. Let p_t be the random return of a linear combination or portfolio of the assets, and let $\bar{P}_t$ be the closure of the linear subspace consisting of all such combinations. Since $\bar{P}_t$ is a conditional linear subspace of L_{2t}, it is also a (conditional) Hilbert space under the mean square inner product. We shall make use of a linear functional defined on $\bar{P}_t$, namely the pricing or cost functional, C(), which maps elements of this space onto the information set. The basic mathematical tool that we shall use is a conditional version of the Riesz Representation theorem (see Hansen and Richard (1987)), which shows that there exists a portfolio $p_t^*\epsilon\bar{P}_t$ such that $C(p_t)=E_{t-1}(p_t^*p_t)$ for all $p_t\epsilon\bar{P}_t$, i.e. p_t^* represents C() in $\bar{P}_t$.

As a consequence, since we have normalised the assets so that they cost one unit, we have that

$$E_{t-1}(R_{it}p_t^*)=1 \quad (i=1,\dots) \tag{3a}$$

$$E_{t-1}(R_{ot}p_t^*)=1, \tag{3b}$$

Hence, $E_{t-1}(r_{it}p_t^*)=0$, and $E_{t-1}(p_t^*)=1/\nu_{ot}$ as $R_{ot}\epsilon I_{t-1}$.

In conformity with the linear factor structure, we assume that p_t^* can be represented as:

$$p_t^* = \nu_t^* + \beta_t^{*\prime}f_t + v_t^* \tag{4}$$

where $\nu_t^*=E_{t-1}(p_t^*)^3$. We also assume that $E_{t-1}(v_t^*v_{Nt})=0$, which implies that R_{it} and p_t^* are only correlated through the common factors. As a consequence, $\text{cov}_{t-1}(R_{it},p^*_t)=\beta_i'\Lambda_{t:t-1}\beta^*_t$, where β_i is the i-th row of B_N. If we substitute this into equation (3a) we obtain

$$\mu_{it}\nu^*_t + \beta_i'\Lambda_{t:t-1}\beta^*_t = 0 \tag{5}$$

and it follows that

$$\mu_{it} = -\beta_i'\Lambda_{t:t-1}\beta^*_t/\nu^*_t = -\beta_i'\Lambda_{t:t-1}\tau_t \tag{6}$$

Under the additional assumption that $\tau_t(=-\beta^*_t/\nu^*_t)$ is constant over time, we finally obtain our basic model for excess returns, which is given by:

$$r_{Nt} = B_N\Lambda_{t:t-1}\tau + B_N f_t + v_{Nt} \tag{7}$$

1.2.1.3 Relationship with Other Models

Given the conditional factor structure for risk, the arbitrage pricing theory of Ross (1976, 1977) is obviously related to our model. If investors can diversify away idiosyncratic risk, and if certain regularity conditions are satisfied[4], we can obtain a conditional version of the <u>exact</u> APT pricing relationship

$$\mu_{Nt} = B_N\pi_t \tag{8}$$

which is usually interpreted as saying that asset risk premia are linear combinations of k risk premia associated with k factor representing portfolios (i.e. portfolios with unit loading on only one factor and zero loadings on the other factors) with weights given by the factor loadings (see e.g. Admati and Pfleiderer (1985)). If we call $\pi_t=\Lambda_{t:t-1}\tau$, it is clear that our model is consistent with (8). But unlike our model, (8) alone is not a model of the time-variation in asset risk premia, only a restriction on the relative pricing of a subset of assets based on an arbitrage argument. Hence it does not provide a fully specified model of time-variation in risk premia, only a cross-sectional restriction.

In the case of only one factor, $\mu_{it} = \tau_1\ \beta_{i1}\lambda_{1t:t-1} = \tau_1\ \text{cov}_{t-1}(r_{it},f_{1t})$, and we can also relate our model to a standard CAPM restriction, with f_{1t} being interpreted as "market" risk. Besides, $\pi_{1t}= \tau_1\lambda_{1t|t-1}$, so the risk premium on the "factor representing portfolio", (which in the CAPM is the market portfolio) is proportional to the conditional variance of that portfolio (cf. equation 1 in section 3.2.1) and hence τ_1 is the "market" price of risk. When k>1, we obtain $\mu_{it}= \sum \tau_j cov_{t-1}(r_{it},f_{jt}) = \sum \beta_{ij}\pi_{jt}$, with $\pi_{jt}=\tau_j\lambda_{jt|t-1}$ which could be interpreted as saying that risk premia are determined by the covariance of the asset returns with the (orthogonal) systematic risk components, which is undiversifiable. A non-trivial advantage of having k factors is that the overall "price of risk" of each asset, i.e. the trade-off between risk and variance, $\rho_{jt}=\mu_{jt}/\sigma_{jjt}$, is not constant. This is line with recent research which has also relaxed the assumption of a constant price of risk (see e.g. Chou, Engle and Kane (1990)), and implies that the behaviour of conditional variances and asset risk premia is allowed to be rather different (see also section 3.5.3).

The above pricing formula can also be obtained (see Engle, Ng and Rothschild (1989, 1990)) by using the Consumption Capital Asset Pricing Model if we interpret p_t^* as the common intertemporal marginal rate of substitution in consumption, S_t. The advantage of the C-CAPM is that it is easier to interpret and can be used to price a complete set of asset payoffs, including derivative claims. One problem with our model is that p_t^* is by construction only guaranteed to price correctly returns with a linear factor representation, i.e. returns in $\bar{P}_t$. Unfortunately, this space does not necessarily contain the derivative claims on those securities, and hence there is no guarantee that S_t would belong to $\bar{P}_t$. However, Gallant, Hansen and Tauchen (1990) show that p_t^* can be interpreted as the (least square) projection of S_t on $\bar{P}_t$. As a consequence, the conditional variance of p_t^* provides a lower bound on the conditional variance of S_t (see also Hansen and Jaganathan (1991)).

1.2.2 Factor Specification

In our empirical application, the k common sources of systematic risk include k_1 "observable" factors, f_{1t}, which attempt to capture the correlation between the unanticipated innovations in observable descriptors of economic performance (e.g. industrial production, inflation, etc.) and stock returns, and a set of k_2 "unobservable" factors, f_{2t}, which are assumed to be only correlated with the returns process, and, therefore, orthogonal to innovations in the published economic variables used. The advantage of having "observable" factors is that the data may be more informative about their implied prices. In general, though, neither f_{1t} nor f_{2t} are literally observable, only partially revealed by the data, although we have superior information about f_{1t} because they are correlated with measured economic variables besides stock returns. Much of the existing literature assumes that the f_{1t}-s are wholly revealed by published economic data, and in that sense they are a special case of the framework employed here. However, for convenience, we shall maintain the terminology of "observable" and "unobservable". We believe that, in spirit, this distinction is meaningful in that while the f_{1t}-s summarize the influence of published economic variables for stock returns, the f_{2t}-s either represent the effect of fundamental influences on returns that are not captured by innovations in published statistics or proxy for changes in "fads" or investor sentiment normally associated with noise trader models (cf. section 3.2.1).

Notice that the orthogonality of the factors enables us to express the variance of asset returns and the risk premia into components attributable to "observable" factors, unobservable factors, and idiosyncratic terms, although in the case of the risk premia the latter should not appear according to our asset pricing equation. If we call B_{1N} and B_{2N}, respectively, the associated Nxk_1 and Nxk_2 full column rank matrices of factor loadings . $B_N = (B_{1N}, B_{2N})$, and $f'_t = (f'_{1t}, f'_{2t})$, we can re-write (2) and (7) as

$$r_{Nt}=B_{1N}\Lambda_{1t:t-1}\tau_1+B_{2N}\Lambda_{2t:t-1}\tau_2+B_{1N}f_{1t}+B_{2N}f_{2t}+v_{Nt} \tag{9}$$

and

$$V_{t-1}(r_{Nt}) = B_{1N}\Lambda_{1t:t-1}B'_{1N} + B_{2N}\Lambda_{2t:t-1}B'_{2N} + \Omega_{Nt:t-1} \tag{10}$$

Equation (9) encompasses the models of Engle, Ng and Rothschild (1989, 1990), who examined US sectoral stock returns, and US bond and aggregate stock returns respectively, and Diebold and

Nerlove (1989), who looked at exchange rates. The former assumed a linear factor model with only "unobservable" factors (k_1=0) and constant idiosyncratic variances ($\Omega_t=\Omega^5$), whereas the latter assumed only one unobservable factor and a zero risk premium (k_1=0, k_2=1, τ_2=0). In the case of stock returns it seems essential to relate the risk premium to the model of changes in volatility. Equation (9) also encompasses the model of Burmeister and McElroy (1988) who considered both "observable" and "unobservable" factors but with constant variances[6].

If we are to have a meaningful distinction between "observable" and "unobservable" factors then it is obviously important that we use a relatively comprehensive set of economic variables in order to generate the set of "observable" factors. However, the inclusion of each additional "observable" factor requires us to estimate an additional N+M+3 parameters, where N is the number of assets and M the number of economic variables. It is necessary, therefore, that the number of factors be kept to a manageable level. For that reason we use a factor analytic approach to extract a small number of factors from a larger number of publicly known economic variables. Since the factors represent unanticipated shocks to asset returns, we estimate a vector autoregressive process for the measured economic variables, and extract common "observable" factors from the innovations of these processes. An important feature of the model is that because we are only interested in those innovations to economic variables that also drive stock returns, we allow the choice of f_{1t} to be determined endogenously by estimating jointly the excess returns equation and the equation that relates the "observable" factors to innovations in the economic variables. The whole system is described by equation (9) together with the following equations

$$x_t = \sum_{j=1}^{p} A_j x_{t-j} + \varepsilon_t \tag{11a}$$

$$\varepsilon_t = C_1 f_{1t} + w_t \tag{11b}$$

where x_t is the (Mx1) vector of measured economic variables, ε_t their time t innovations, A_j is the matrix of coefficients on the jth lag in the VAR, C_1 is a (M$\times k_1$) full column rank matrix of factor sensitivities for the economic variables, and w_t is a (Mx1) vector of idiosyncratic error terms[7]. We assume that $E_{t-1}(w_t) = 0$, $E_{t-1}(f_t w_t') = 0$, $E_{t-1}(w_t v_t') = 0$ and $E_{t-1}(w_t w_t') = \Gamma_{t|t-1}$, a positive semidefinite diagonal matrix[8] (cf. equation (3) in section 2.2).

The estimation procedure consists of estimating the excess returns equations (9) jointly with the process for the economic variables (11) by maximum likelihood methods. Before turning to a description of the actual estimation procedure, we consider first the important issue of identification.

1.2.3 Identification

Since the scaling of the factors is irrelevant, then in the case of constant variance ($\Lambda_{t|t-1} = \Lambda$, $\forall t$) it is usual to impose the condition that the variance of each of the factors is unity, that is $\Lambda = I$. In that case, however, Bf_t, $C_1 f_t$, and $B\tau$ are indistinguishable from $B^0 f_t^0$, $C_1^0 f_t^0$ and $B^0 \tau^0$ where, for arbitrary orthogonal Q, $B^0 = BQ'$, $C_1^0 = C_1 Q'$, $f_t^0 = Qf_t$, and $\tau^0 = Q\tau$[9] (see section 2.3).

The advantage of modelling time-varying volatility is that identification is less problematic as in general the set of admissible orthogonal transformation matrices Q that preserve the diagonality of $\Lambda_{t|t-1}$ for all t is substantially reduced. In fact it can be shown (see Lemma 2 in section 2.3. and Sentana (1991a)) that if $\Lambda_{t|t-1}$ is diagonal (but not scalar) and $E(\Lambda_{t|t-1}) = I$, then these conditions are sufficient to ensure that B and τ' are always identifiable up to column sign changes. This result does not depend on any particular parameterization of the variances of the factors, only on their time-variation. The existence of identification when we relax the assumption of homoskedasticity may be apparently paradoxical. However, what actually drives the result is the fact that conditional orthogonality is a much stronger restriction on the correlation of the factors than the standard unconditional one (see the discussion in section 2.3).

In conventional factor analytic estimation of the APT, it is common for researchers to stress the identification problem and not report the non-identified individual risk premia (see Gultekin,

Gultekin and Penati (1989), Roll and Ross (1980) or Dhrymes, Friend and Gultekin (1984) for a detailed discussion). Since our model is, in general, identified, we do not face this problem, so, in this respect, our approach has a non-trivial advantage over the conventional approach.

1.2.4 State Space Representation and Kalman Filtering

In order to estimate of the unknown factors in our model, we use the Kalman filter, which is ideally suited to perform this task since it produces the best (in the Mean Square Error sense) estimates of the factors (see e.g. Harvey (1989) or Sentana (1991c)).

As in standard factor analysis, our model has a natural state-space representation (see section 2.4). Taking the common factors as the state, the measurement and transition equations are given, respectively, by:

$$\begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} B_1 & B_2 \\ C_1 & 0 \end{bmatrix} \begin{bmatrix} f_{1t} \\ f_{2t} \end{bmatrix} + \begin{bmatrix} v_t \\ w_t \end{bmatrix} \tag{12}$$

$$\begin{bmatrix} f_{1t} \\ f_{2t} \end{bmatrix} = \begin{bmatrix} \xi_{1t} \\ \xi_{2t} \end{bmatrix} \tag{13}$$

where:

$$\eta_t = r_t - B_1\Lambda_{1t:t-1}\tau_1 - B_2\Lambda_{2t:t-1}\tau_2 \tag{14a}$$

$$\varepsilon_t = (x_t - \sum_{j=1}^{p} A_j x_{t-j}) \tag{14b}$$

$$V_{t-1}\begin{bmatrix} v_t \\ w_t \end{bmatrix} = \begin{bmatrix} \Omega_{t:t-1} & 0 \\ 0 & \Gamma_{t:t-1} \end{bmatrix} \tag{15a}$$

$$V_{t-1}\begin{bmatrix} \xi_{1t} \\ \xi_{2t} \end{bmatrix} = \begin{bmatrix} \Lambda_{1t:t-1} & 0 \\ 0 & \Lambda_{2t:t-1} \end{bmatrix} \tag{15b}$$

The prediction equations for the state vector f_t, and its covariance matrix, $\Lambda_{t:t-1}$ are (see Diebold and Nerlove (1989))

$$f_{t|t-1} = 0 \tag{16a}$$

$$\Lambda_{t|t-1} = \Lambda_{t|t-1} \tag{16b}$$

whereas the updating equations are

$$f_{t|t} = \Lambda_{t|t-1}B^{+\prime}\Sigma^{-1}_{t|t-1}\zeta_t \tag{17a}$$

$$\Lambda_{t|t} = \Lambda_{t|t-1} - \Lambda_{t|t-1}B^{+\prime}\Sigma^{-1}_{t|t-1}B^{+}\Lambda_{t|t-1} \tag{17b}$$

where

$$B^{+} = \begin{bmatrix} B_1 & B_2 \\ C_1 & 0 \end{bmatrix} \tag{18}$$

$$\Sigma_{t|t-1} = \begin{bmatrix} B_1 & B_2 \\ C_1 & 0 \end{bmatrix}\begin{bmatrix} \Lambda_{1t|t-1} & 0 \\ 0 & \Lambda_{2t|t-1} \end{bmatrix}\begin{bmatrix} B_1' & C_1' \\ B_2' & 0 \end{bmatrix} + \\ + \begin{bmatrix} \Omega_{t|t-1} & 0 \\ 0 & \Gamma_{t|t-1} \end{bmatrix} \tag{19}$$

Since the transition equation is degenerate, the updated estimates $f_{t|t}$ coincide with the smoothing estimates $f_{t|T}$, (and equally $\Lambda_{t|T}=\Lambda_{t|t}$) which is analytically very convenient (see section 2.4. for details).

In order to complete the model we need to specify a particular parameterisation for the conditional variances of the factors. For practical purposes, we shall assume that the variances of f_t, v_t and w_t follow univariate GARCH(1,1)-type processes (Bollerslev (1986)). However, much care has to be exercised when dealing with conditional variances that depend on past squared values of these factors, as the true values of the factors do not belong to the information set (see section 2.6 and Harvey, Ruiz & Sentana (1990)). In particular, using estimates of the factors in place of the factor themselves induces errors in the estimates. One solution to this problem is to modify the GARCH framework so that the diagonal elements of $\Lambda_{t|t-1}$ depend only on known information (see section 2.6). This can be achieved by an equation of the form

$$\lambda_{it|t-1} = (1-\psi_{i1}-\phi_{i1})+\psi_{i1}E_{t-1}(f^2_{it-1})+\delta_{i1}\lambda_{it-1|t-2} \tag{20a}$$

where $\psi_{i1},\delta_{i1}\geq 0$, $\psi_{i1}+\delta_{i1}\leq 1$, whereas the diagonal elements of $\Omega_{t|t-1}$ are represented as

$$\omega_{jt|t-1} = \phi_{j0}+\phi_{j1}E_{t-1}(v^2_{jt-1})+\rho_{j1}\omega_{jt-1|t-2} \tag{20b}$$

where $\phi_{j0},\phi_{j1},\rho_{j1}\geq 0$, $\phi_{j1}+\rho_{j1}\leq 1$, with an analogous equation for the diagonal elements of $\Gamma_{t|t-1}$[10]. Notice that since $E_{t-1}(f^2_{it-1})= E^2_{t-1}(f_{it-1})+ V_{t-1}(f_{it-1})= f^2_{it-1|t-1}+\lambda_{it-1|t-1}$, (20a) incorporates a correction term in the usual GARCH variance which reflects the uncertainty of the factor estimates. This correction can be easily evaluated from the Kalman filter estimates (see section 2.6).

1.2.5 Estimation

If we assume that the factors and the idiosyncratic noise have a conditional normal distribution (which is a limiting case of the multivariate t discussed in section 2.5), we may write the log-likelihood of the sample (ignoring initial conditions) as:

$$L=-(M+N)T/2\ \ln(2\pi)-1/2\sum_{t=1}^{T}\ln|\Sigma_{t|t-1}|-1/2\sum_{t=1}^{T}\zeta_t'\Sigma^{-1}_{t|t-1}\zeta_t \tag{21}$$

(cf. equation (16) in section 2.5) where

$$\zeta_t = \begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix} \tag{22}$$

The computation of our estimates is considerably simplified if we first obtain consistent estimates of A_j (the coefficients of x_{t-j} in the vector autoregression) independently by using ordinary least squares. Still, the number of parameters to estimate by full information maximum likelihood is $k_1(N+M+3)+k_2(N+3)+3N+3M$. In order to obtain initial values for the maximisation of the global loglikelihood function in (21) we use a two-step procedure. Given the estimates of A_j obtained by OLS, we take the residuals ε_t, and, having standardised them, we use a principal components analysis to obtain estimates of the factor loadings C_1. Then we use the estimates of C and Γ obtained by this method as starting values for maximum likelihood estimation of the parameters in the sub-model given by equation (11). In this way we obtain Kalman filter based factors f_{1t} and variances $\Lambda_{1t|t-1}$. We next regress the actual excess returns on our estimated factors, f_{1t} and apply a principal components procedure on the residuals in order to extract estimates of the matrix of factor loadings, B_2. These estimates are then used as starting values for a maximum likelihood procedure to estimate the sub-model consisting of the excess returns equation (9) taking f_{1t} and $\Lambda_{1t|t-1}$ as fixed regressors.

This procedure gives us consistent estimates of all of the parameters of the model, and these are then used as initial values in the maximization of the global likelihood function given by equation (21)[11].

1.3 Empirical Application

We estimate the multivariate factor asset pricing model described above on monthly data for returns on sixteen national stock markets from 1970:1 to 1988:10. The sixteen countries are Australia, Austria, Belgium, Canada, Denmark, France, Germany, Italy, Japan, Netherlands, Norway, Spain, Sweden, Switzerland, the United Kingdom and the United States. Data on stock returns - the percentage change in the share price index plus the dividend yield - were obtained from the Morgan Stanley Capital International world indices. All stock returns are measured in US dollars, and excess returns were computed as the (real, gross) return on each market during the month minus the (real, gross) one-month US Treasury Bill yield at the beginning of the month. The US consumer price index was used as deflator. Consequently, r_t is the <u>real</u> return on stocks over and above the real return on cash.

1.3.1 Vector Autoregressions for the Economic Variables

Data on the ten macroeconomic variables that were available monthly, and might reasonably be expected to affect stock returns, were analysed. The variables are (i) short interest rates measured by the yield on US Treasury Bills, (ii) long interest rates (index of yields on long-term bonds weighted by GNP), (iii) the dollar-deutschmark exchange rate, (iv) the dollar-yen exchange rate, (v) industrial production for the G3 group of countries (an index weighted by GNP), (vi) inflation in G3 (consumer price index weighted by GNP), (vii) US trade deficit (% of GNP), (viii) real money supply in G3 (weighted by GNP), (ix) real oil price in US dollars, and (x) an index of real commodity prices. Details of the definitions and sources of these variables may be found in the data appendix 1 [12]. Because of lags in the publication of economic statistics the values of the variables for industrial production, the US trade deficit, real money supply, and inflation in period t were assumed to be the published values for month t-1. We report below the results of varying this assumption.

The first step is to estimate innovations in the economic variables by fitting vector autoregressions described by equation (8a). These VARs were estimated over the sample period 1970:8 to 1988:10 - a total of 219 observations. Given that we have only just over 200 monthly observations the dimensionality of the VAR has to be somewhat restricted. We experimented with different lag lengths in order to obtain a final specification with no detectable serial correlation in the residuals. The fitted equations also included seasonal dummies. To test for serial correlation we used the standard the Lagrange multiplier-based test against 12th order serial correlation (see Harvey (1981)) as well as a version robust to conditional heteroskedasticity (see Wooldridge (1988)). For eight of the ten variables considered, a specification with thirteen lags for the dependent variable and three lags for the other variables was sufficient to ensure that the null hypothesis of no serial correlation could not be rejected at the 5% level. But for two of the variables - inflation and real money supply - somewhat different lag lengths were necessary in order to ensure that the residuals were white-noise[13]

1.3.2 Estimation of the Joint Factor and Multivariate Returns Model

Standardized residuals from the VARs are used as data in order to estimate the joint factor and multivariate returns model described by equations (9) and (11b) by maximum likelihood methods. The general model permits both "observable" and "unobservable" factors. We used four "observable" factors, that is f_{1t} is a 4x1 vector. When estimating (11b) in order to generate initial values for the global maximum likelihood procedure, we tested for the appropriateness of four factors. On the one hand, a likelihood ratio test for the existence of a fifth factor yielded a value of 16.96 (with eleven degrees of freedom) for the difference in loglikelihoods, which is statistically insignificant at conventional levels. On the other hand, the omission of the fourth factor could be rejected with a LR value of 25.7, which is significant at the 1% level (see section 4 for a discussion of the effect of including additional "observable" factors)[14]. In addition to the four "observable" factors, we allowed for two unobservable factors. Attempts to allow for a third unobservable factor led to the idiosyncratic variance for the Netherlands market being driven to zero. Since this seemed an implausible outcome we used only two unobservable factors.

The matrix, C_1, of factor loadings for the economic variables that is obtained from estimation of the complete system is shown in Table 1.1. One way of interpreting the "observable" factors is to look at the weights attached to the innovation in economic variables when constructing the factor estimates by means of the Kalman filter updating equation (17a). But unless conditional variances are constant, these weights change from period to period. One possible way of obtaining "average weights" is to regress our "observable" factor estimates on the economic variables. If the weights were constant (as in standard factor analysis models), the corresponding R^2 would be 1. These results are reported in table 1.2. The first factor has relatively large coefficients for the innovations in interest rates, both short and long rates. It might be interpreted as an "interest rate factor". For the second factor the coefficients are high for the dollar-yen and dollar-DM exchange rates. This might be thought of as a "dollar exchange rate factor". The third factor also has substantial coefficients on the two exchange rate variables, but in this case the coefficient on the dollar-yen cross-rate has the opposite sign to that on the dollar-DM rate. It measures exchange rate shocks among currencies other than the dollar. The fourth factor reflects principally innovations in G3 real money supply and the inflation rate. Not surprisingly, this interpretation is consistent with the elements of the factor loading matrix C_1.

Estimates of the matrix of country factor loadings B and the vector τ of risk premia associated with each of the six factors are shown in Table 1.3. The coefficients of the first "observable" factor, f_{11}, are, with one exception, negative. If f_{11} is an interest rate factor, the implication is that unanticipated increases in interest rates reduce excess returns. The estimate of the risk premium associated with the interest rate factor is positive and statistically significant, which implies that an increase in the variance of innovations to interest rates is associated with a reduction in the required rate of return. This negative relation between risk premia and volatility may look surprising at first sight, but it is consistent with the empirical and theoretical results of Pagan and Hong (1991), Backus and Gregory (1989) and Glosten, Jaganathan and Runkle (1989) on the relation between conditional mean and variances in asset returns. An interpretation in terms of the consumption capital asset pricing model would suggest that this occurs because the rate of growth in consumption is positively correlated with innovations in interest rates.

As far as the dollar exchange rate factor is concerned, there is no unambiguous theoretical prediction about the sign of the coefficients. An appreciation of the dollar stimulates the exports of companies in other countries, which raises share prices in those markets. But it also may increase inflation in non-US economies and higher unanticipated inflation lowers share prices (see, for example, Fama and Schwert (1977)). In addition, for a given change in stock prices in local currency, a dollar appreciation implies a lower dollar return in non-US stock markets, although if currency hedging is available, this should not affect excess returns which should be invariant to the particular currency used[15]. Our estimates suggest that an unanticipated appreciation of the dollar leads to a fall in all non-North American stock markets and a rise in the US and Canadian markets.

Estimates for the third "observable" factor imply that an unanticipated appreciation of the yen and deutschmark local currency leads to increases in the stock market in Japan and Germany, respectively. As one might expect, the size of the coefficient is bigger for countries that belong to the "DM block", that is for Germany, Austria, Switzerland, Netherlands and Belgium. return in non-US stock markets. The "price of risk" terms associated with exchange rate volatility are statistically insignificant, suggesting that exchange rate volatility is not priced.

The coefficients on the real money supply factor imply that unanticipated increases in the money supply or unexpected falls in the inflation rate raise share prices in eleven of the sixteen national markets. The negative estimate of the associated risk premium means that an increase in the volatility of money innovations increases the equity risk premium, but again this is not significantly priced.

Figures 1.2a and 1.2b plot the estimated two factors over the sample period. The first unobservable factor is extremely volatile. For example, it exhibits a sharp spike in October 1987 which coincides with the worldwide stock market crash. Another sharp downward movement can be seen in September 1974, and there are upward surges in January 1975 and January 1987. The volatility persistence measure provided by the sum of the ARCH and GARCH coefficients for this factor is relatively high at 0.7899. This factor is uncorrelated with the innovations in the "observable" economic variables that we have analysed. The difficulty in explaining the 1987 stock market crash in terms of changes to observable fundamental variables is illustrated clearly in Figure 1.2a. The plot suggests that there are other episodes in the last twenty years when markets have moved together without any obvious explanation in terms of observable economic variables. The first unobservable factor has a substantial loading on most countries, the conspicuous exception being Austria which fell least in the 1987 crash. An increase in the volatility of this factor leads to a higher estimated equity risk premium, although it is not statistically significant. The second unobservable factor also exhibits high volatility (see Figure 1.2b), but the volatility persistence is much higher at 0.9797, being very close to the IGARCH region (cf. Engle and Bollerslev (1986), Nelson (1991))). Besides, the volatility associated with this factor seems to be significantly priced.

To "learn" more about the nature of the unobservable factors, we can regress the factor estimates on the set of stock returns. Although our model suggests that the "factor representing portfolios" would change from period to period (cf. Sentana (1991c)), we can understand the results reported in table 1.4 as providing "average" basis portfolios. It is striking that neither set of weights appear to be closely correlated with market capitalization (cf. the weightings on Japan with its market capitalization). Hence, not only is our model considerably more general than a simple market model in that we include two unobservable factors, but our results also suggests that including the "world" index as a market factor may induce a serious misspecification. The first unobservable factor has a significant positive association with stock returns in North America (US and Canada), and is negatively correlated with German returns. By contrast, the second unobservable factor is largely dominated by the German stock returns.

Notice that the estimates of the individual τ's are poorly determined. This may be due to collinearity in the variances of the factors. Therefore, we have computed joint test of the statistical significance for groups of τ coefficients. The LR test for the hypothesis that the risk associated with all 6 factors is not priced is 25.90, which is statistically significant. As for the pricing of "observable" and "unobservable" risk components, the corresponding test statistics yielded 19.56 ($\chi^2_{4,0.05}$=9.49) and 8.75 ($\chi^2_{2,0.05}$=5.99). Therefore, there is evidence that factor risk is priced, although the individual components are poorly determined.

1.4. Links Between Stock Markets

1.4.1 Assessing Global Capital Market Integration

The estimates of our model implicitly assume that stock markets are integrated. Specifically, we have assumed that:

A1. Idiosyncratic risk is not priced

A2. The "price of risk", τ_i, that is associated with the volatility of the underlying factors is assumed to be the same for each country.

Even if the theoretical restrictions implied by our mode may provide a valid representation of asset pricing in a domestic economy, it would be surprising if A1 and A2 were valid for the world stock markets given the existence of exchange controls (at least for part of the sample) and other barriers to investment. Of course, we may reject the intertemporal asset pricing model that we have estimated if any of the two hypotheses are invalid.

We attempted to test A1 by allowing the idiosyncratic volatility of each market to affect the corresponding risk premia. Our results are to be found in table 1.5. Note that idiosyncratic risk is significantly priced (at the 5% level) in 11 of the 16 countries. Moreover, it attracts the right sign (i.e. positive) in all 16 countries, and a joint test also rejects the null that country-specific risk is not priced at conventional levels. Hence, assumption A1 appears to be rejected.

Turning to A2, i.e. the assumption that the price of risk is common across countries, our results are presented in table 1.6. Here, in 2 of the 6 cases (i.e. for the two unobservable factors) we reject the null of the cross-equation restriction.

Since our model can be interpreted as a special case of the APT in that we make an additional assumption about the risk premia associated with the factors, it is obviously possible that we are rejecting this particular assumption and a different version of the APT might be valid in this context. However, the evidence suggests that our assumption of global capital market integration is probably unwarranted in the context of an asset pricing model.

It is, however, possible that our rejections stem from using the data from the seventies. Yet, the eighties have seen the removal of various barriers to foreign investment - e.g. the lifting of exchange controls in the UK in 1979, various liberalization measures in Japan including the revision of the Foreign Exchange Law in 1980, greater harmonization of regulation in the EEC, etc. Therefore, we repeated our test for A1 in the subsamples 1970:8-1979:12 and 1980:7 - 1988:10. The results, though, continue to reject the assumption that idiosyncratic risk is not priced.

1.4.2 The Contribution of Observables to the Covariance Matrix of Returns

One objective of this chapter is to attempt to link changes in co-movements between stock markets to variations in the economic variables that we consider. Therefore it is of interest to estimate the extent to which the covariance matrix of world stock markets can be explained in terms of innovations in observable economic variables. The conditional variance-covariance matrix of excess returns is given by equation (2). This is likely to be informative even though the mean restrictions are not satisfied.

Figure 1.3a shows the estimated conditional covariance, and the contribution to this of the variance of the "observable" factors, for the US and UK markets[16]. The picture is striking. Although the covariance between the two markets has ranged between 0.25 to values in excess of 0.6 during the sample period, the contribution of the "observable" factors has remained firmly in the range ± 0.05. The relative contribution of "observable" factors to changes in the covariance between the US and UK appears to have been minuscule. Figure 1.3b shows the estimated conditional covariance between the US and Japan. The relative contribution of "observable"s in this case is somewhat higher. The decline in the covariance in 1978-79 is clearly linked to changes in the volatility of "observables". However, during the rest of the period, the "observables" do not explain much of the variation in the conditional covariance.

Not only do "observable"s account for a small proportion of the covariance between markets, they also fail to explain a substantial fraction of the variance of the markets. Figures 4 shows the decomposition of the variance of the US market into three main components attributable to "observable" factors, "unobservable" factors, and the idiosyncratic term. For the US the contribution of the "observables" to the variance is negligible. It is clear that the contribution of the unobservables essentially mirrors the actual change in the variance. The contribution of the idiosyncratic component is constant for the US because the coefficients in the corresponding GARCH process were estimated to be zero[17].

The failure of our "observable" variables to account for changes in the correlations between markets is somewhat disappointing. It is therefore important to investigate whether our results can be explained by the fact that we only allowed ourselves four observable factors in order to economise on the number of parameters. The inclusion of one additional "observable" factor would raise the number of parameters to be estimated from 216 to 244. To allow as many factors as variables would entail the estimation of 374 parameters. In order to illustrate the effect of including additional economic variables, we regressed excess returns on all ten series of innovations from the VARs, computed the R^2, and compared them with the values obtained when we included only the four "observable" factors. The results are shown in Table 1.6. As measured by the average R^2, four factors enable us to capture over 75% of the explanatory power that can be obtained by using all ten innovations separately. This suggests that the use of four factors has not led us to underestimate significantly the contribution of "observable" factors[18].

It is possible that the "observable" factors are measured with error. For example, the information set used in our VARs may be too restricted. Under the efficient markets hypothesis other relevant information is incorporated in stock prices. Hence we re-estimated the VARs including in the information set three lags of the return in the world index as a proxy for variables that had been omitted. This produced new estimates of the ten innovations in the observable economic variables. We then regressed excess returns on these innovations. There was very little change in the explanatory power of the economic variables - the average R^2 actually fell from 0.136 to 0.128. Because of delays in the publication of economic statistics, we assumed a one month lag in the production of certain monthly series (see section 1.3.1). We experimented, however, with the current values of variables. Again this makes little difference - the mean R^2 rose marginally from 0.136 to 0.144.

The fact that our observable factors do not have substantial explanatory power to explain changes in co-movements between markets is disappointing, especially if we take into account that our observable variables do seem to Granger-cause dividends (see table 1.8).

1.4.3 Changing Links Between National Stock Markets

It is often asserted that the links between national stock markets have increased with improved electronic communications and the abolition of exchange controls in a number of countries. The 1988 Brady Commission, however, pointed out that there had been no trend increase in the (rolling) correlations between markets (computed by using a monthly window of the last twelve observations). Figure 1.5 plots the (cross-sectional) average estimated conditional correlation coefficient between the markets implied by our parameter estimates. When we regress this mean conditional correlation coefficient on a constant, a dummy variable for the 1980s and a dummy variable for the period after the 1987 stock market crash we obtain

$$\text{CORR} = 0.381 + 0.018 \ \text{D1980s} + 0.105 \ \text{Dcrash}$$

Hence the case for a trend increase in correlations between markets depends upon the weight that is attached to the observations surrounding the 1987 stock market crash. Those authors who argue that markets have become increasingly integrated from data in the period 1986-88 (for example, von Furstenberg and Nam Jeon (1989)) may be confusing a transitory with a permanent increase in correlations.

A stylised fact that has been noted before is that periods when markets are increasingly correlated are also times when markets are volatile (for example, King and Wadhwani (1990) and Roll (1989)). Indeed, King and Wadhwani argued that this might be because a rise in volatility caused by factors that are not closely related to "news", might lead agents to pay greater attention to other markets in an attempt to determine the change in the "taste for equity". In our model, periods when the volatility of the unobservable factors rises are also those when, ceteris paribus, markets appear to exhibit greater inter-correlation. To see this, consider a two-asset two-factor model

$$r_{1t} = \mu_{1t} + \beta_{11}f_{1t} - \beta_{12}f_{2t} + v_{1t} \tag{23a}$$

$$r_{2t} = \mu_{2t} + \beta_{21}f_{1t} + \beta_{22}f_{2t} + v_{2t} \tag{23b}$$

where $\beta_{ij} > 0 \ \forall i,j$. The conditional correlation coefficient between the two markets is given by

$$\rho_{12t}=(\beta_{11}\beta_{12}\lambda_{1t:t-1}-\beta_{12}\beta_{22}\lambda_{2t:t-1})/(\sigma^{1/2}_{11t:t-1}\sigma^{1/2}_{22t:t-1}) \tag{24a}$$

with

$$\sigma_{11t:t-1} = \beta^2_{11}\lambda_{1t:t-1}+\beta^2_{12}\lambda_{2t\dagger t-1}\omega_{1t:t-1} \tag{24b}$$

$$\sigma_{22t:t-1} = \beta^2_{21}\lambda_{1t:t-1}+\beta^2_{22}\lambda_{2t:t-1}+\omega_{2t:t-1} \tag{24c}$$

It follows that the (conditional) correlation coefficient is an increasing function of $\lambda_{1t:t-1}$ and a decreasing function of $\lambda_{2t:t-1}$ and the idiosyncratic variances. Hence an increase in the volatility of those factors that affect all stock markets with the same sign - the unobservable factors in our sample - will be associated with an increase in the correlation between markets. This is consistent with the observed rise in both volatility and inter-correlation around the time of the 1987 crash. Rises in the volatility of factors that move markets in different directions - exchange rates, for example - may be associated with falls in correlation coefficients. The positive link between volatility and correlation that has been noted by previous authors appears to reflect the fact that the unobservable factors have historically been more important in explaining stock returns than the "observable" factors. Our analysis provides some confirmation over a longer sample period for the stylised fact that conditional correlation is related to volatility. It does not, however, enable us to assess whether there is a causal relation between volatility and correlation.

1.5 Conclusions

Our results seems to suggest that global stock markets are not integrated. We were easily able to reject the null hypothesis that idiosyncratic risk is not priced. Moreover, the "price of risk" associated with the relevant factors is not always the same across countries. Also, although it is commonly accepted that globalisation has led to national stock markets moving more closely together, we were unable to find any trend increase in correlations.

Another of our empirical findings is that only a small proportion of the time-variation in the covariances between national stock markets can be accounted for by "observable" economic variables. Changes in correlations between markets are driven primarily by movements in unobservable variables, which may be interpreted as unobservable fundamental variables that we have ignored, or as variables related to the demand for equities by "noise" traders. Our results suggest that constructing models which help predict changes in covariances between markets on the basis of observed economic data is likely to prove difficult.

Our model is considerably more general than a simple market model in that we include two unobservable factors. It is of some interest that our unobservable factors do not seem well correlated with returns on a world index. (cf. the lack of correlation of the factors with Japanese returns and its market capitalization). Hence, our results also suggests that including the "world" index as a market factor may be misleading.

Further work is required in two main directions. First, it may be that the asset pricing restrictions of our model are more readily satisfied in other settings (e.g. on different industrial sectors in an integrated national stock market). The second is to explore alternative models for equity returns.

ENDNOTES

1 Relations between risk premia and conditional variances are sensitive to differential information between agents and econometricians (see Pagan and Ullah (1988) or Glosten, Jaganathan and Runkle (1989)). We shall therefore make the standard assumption in the applied literature that the relevant information is common to both.

2 We shall assume that $\Omega_{Nt|t-1}$ is actually a positive definite matrix for all N since otherwise we could form finite portfolios that contain only systematic risk.

3 Equation (4) can be understood as the orthogonal projection of $p_t^*-\nu_t^*$ on the space generated by the factors.

4 In our case, the implicit condition is that $E_{t-1}(v_t^* v_{Nt})=0$, which is equivalent to the condition in Chamberlain (1983) that the pricing functional is well diversified. This results in the APT pricing formula being exact, rather than approximate.

5 By contrast, Engle, Ng and Rothschild (1989, 1990) do not assume that Ω is a diagonal matrix.

6 If we define $f_t^*=\Lambda_{t|t-1}^{-1/2}f_t$, $B_t^*=B\Lambda_{t|t-1}^{1/2}$, $\tau_t=\Lambda_{t|t-1}^{1/2}\tau$, then our model can also be interpreted as a time-varying factor-betas model with homoskedastic factors in which the betas of different assets on a factor change proportionately (see Engle, Ng and Rothschild (1990)).

7 In order to make our analysis invariant to changes in scale, we divide the residuals from the vector autoregression by their own standard deviation - so that all the transformed values of ε_t have, by construction, unit variance.

8 The matrix $\Gamma_{t|t-1}$ is assumed to be positive semidefinite in order to allow for the possibility that the variance of one or some of the w_t-s is zero. This is related to the observability of the factors. If at least k_1 of the w_t-s are zero, the "observable" factors would be fully revealed by the set of economic variables, otherwise, they are only partially revealed. In the extreme case where $\Gamma_{t|t-1}=0$ $\forall t$, the model would be in the spirit of one-mode component analysis (see Magnus & Neudecker (1988)).

9 One approach in standard factor analysis models is to use some sufficiency conditions for identification, such as Dunn's (1973) zero-type restrictions on the factor loadings (see Lemma 1 in section 2.3). In the context of the present model these amount to imposing a zero restriction on the upper triangle of B_2 and C_1.

10 The restrictions on the coefficients of the GARCH processes are necessary to ensure stationarity and positivity of variances, while, for $\lambda_{1t|t-1}$, we also need to ensure that the unconditional variance is unity (see also sections 2.3 and 2.6).

11 Maximization of the log-likelihood function was carried out on the LSE VAX using the NAG library E04JBF routine. The block triangularity of the global factor loading matrix B^{*} and the special from of the covariance matrices is exploited by means of the Woodbury formula (Householder (1964)) so that the inversion of Σ_t (a 26x26 matrix) only involves the inversion of the 4x4 matrix $\Lambda_{1t}^{-1}+C_1'\Gamma_t^{-1}C_1$ and the 2x2 matrix $\Lambda_{2t}^{-1}+B_2'\Omega_t^{-1}B_2$ (for the case of Γ_t or Ω_t singular, see Sentana (1989)).

12 Aggregate variables with G3 GNP weights have been used as a rough way of reducing the number of observable economic variables.

13 See table 1 in King, Sentana and Wadhwani (1990) for details.

14 The appropriateness of four "observable" factors is also confirmed by the Akaike information criterion. These tests impose the parametric model for heteroskedasticity.

15 This is due to the fact that the difference between spot and forward exchange rates is equal to the difference between the two domestic interest rates under covered interest parity.

16 An obvious alternative is to look at conditional correlations. This has the advantage that they are easy to interpret. The problem is that we our model does not imply an additive decomposition of the conditional correlations into components attributable to observable and unobservable factors.

[17] This is not the case in the general. For instance, in the UK the contribution of the idiosyncratic component is important in capturing a sharp rise in volatility between December 1974 and February 1975 during which period there was a major fall in the market followed by a rapid recovery. This emphasises the need to allow for time-varying idiosyncratic variances.

[18] Our result that the contribution of "observable" factors to an explanation of returns is small is consistent with the findings of Cutler, Poterba & Summers (1989).

DATA APPENDIX 1

Details of the data series used are as follows (where appropriate the name of the series is followed by its Datastream code):

Stock prices and dividend yields:	from Morgan Stanley Capital International Perspectives.
Safe interest rate:	yield on 1 month US T Bills, beginning of period.
Short interest rate:	yield on 3 Month US T Bills, end of period (USOCTBL%).
Long Interest Rate:	US Yield on long-term government bonds, end of period (USOCLNG%) W. Germany: Yield on long-term government bonds,end of Period (BDOCLNG%) Japan: Yield on Central Government Bonds end of Period (JPOCLNG%)
Exchange rates:	US$-Yen, end of period (JPOCEXCH). US$-DM, end of period (BDOCEXCH).
Index of Industrial Production	US (USOCIPRDJ) West Germany (BDOCIPRDI) Japan (JPOCIPRDH)
Consumer Prices:	US - all items (USOCPCONF) W. Germany - total (BDOCPCONF) Japan - Tokyo, all items (JPOCCPTKF)
Trade Account:	US Foreign Trade Balance, US $million (USOCVBALA)
Money Supply:	US: M3, US $billion, current prices (USOCM3MNA) W. Germany: M3 DM million, current prices (BDTU0800A) Japan: M1 + Quasi Money, yen billion current prices (JPOCM1QSA)
Oil Price:	Saudi Arabian Light Oil Spot Price, US$ per barrel end of period (SAUDISPT) (up to Dec 72 from IMF Financial Statistics)
Commodity Prices:	The Economist World Commodity Price Index (last week of each month) (up to Dec 72 monthly average)
GNP Weights:	US GNP $billion at annual rates, current prices (USOCGNPDB) W. Germany GNP DM billion at annual rates, current prices (BDOCGNPDB) Japan GNP Yen billion at annual rates, current prices) (JPOCGNPDB)

Table 1.1

Matrix of Observable Factor Loadings

Variable	f11	f12	f13	f14
1. SHORT INTEREST RATE	0.633 (0.121)	-0.044 (0.061)	-0.067 (0.085)	-0.040 (0.066)
2. LONG INTEREST RATE	0.578 (0.117)	-0.008 (0.065)	-0.140 (0.081)	-0.065 (0.063)
3. DOLLAR/ YEN EXCH. RATE	-0.158 (0.065)	-0.843 (0.087)	0.464 (0.077)	-0.149 (0.074)
4. DOLLAR/ DM EXCH. RATE	-0.318 (0.065)	-0.852 (0.061)	-0.362 (0.075)	0.138 (0.071)
5. INDUST-RIAL PROD-UCTION	0.120 (0.067)	-0.053 (0.064)	-0.051 (0.070)	-0.184 (0.066)
6. INFLA-TION	0.020 (0.068)	-0.152 (0.069)	0.051 (0.072)	0.206 (0.067)
7. US TRADE ACCOUNT	-0.122 (0.072)	0.013 (0.069)	0.087 (0.067)	0.005 (0.064)
8. REAL MONEY SUPPLY	-0.041 (0.077)	-0.023 (0.066)	-0.318 (0.119)	-0.956 (0.630)
9. OIL PRICE	0.040 (0.040)	-0.071 (0.093)	-0.009 (0.039)	-0.020 (0.034)
10. COMMOD-ITY PRICES	-0.085 (0.064)	-0.321 (0.066)	-0.083 (0.070)	0.086 (0.067)

Note: Standard errors in brackets

Table 1.2

Observable Factor Score Weights

Variable	f11	f12	f13	f14
1. SHORT INTEREST RATE	0.601	-0.178	-0.073	0.038
2. LONG INTEREST RATE	0.360	-0.107	-0.044	0.022
3. DOLLAR/ YEN EXCH. RATE	0.034	-0.525	1.135	-0.217
4. DOLLAR/ DM EXCH. RATE	-0.142	-0.602	-1.008	0.348
5. INDUST-RIAL PROD-UCTION	0.043	-0.013	-0.005	0.003
6. INFLA-TION	-0.019	0.006	0.002	-0.001
7. US TRADE ACCOUNT	-0.071	0.021	0.009	-0.004
8. REAL MONEY SUPPLY	-0.068	-0.098	-0.259	-0.962
9. OIL PRICE	-0.015	0.005	0.002	0.000
10. COMMOD-ITY PRICES	0.014	-0.004	-0.002	0.000

Table 1.3

Estimates of B and τ

Country	f11	f12	f13	f14	f21	f22
AUSTRAL.	-0.105	-0.091	0.021	0.001	0.610	0.143
	(0.053)	(0.058)	(0.053)	(0.043)	(0.061)	(0.080)
AUSTRIA	-0.129	-0.284	-0.115	0.054	0.045	0.223
	(0.029)	(0.028)	(0.037)	(0.032)	(0.029)	(0.044)
BELGIUM	-0.150	-0.303	-0.064	-0.001	0.295	0.425
	(0.038)	(0.027)	(0.039)	(0.029)	(0.036)	(0.050)
CANADA	-0.093	0.036	-0.061	0.043	0.524	0.075
	(0.040)	(0.039)	(0.037)	(0.033)	(0.041)	(0.051)
DENMARK	-0.069	-0.197	-0.003	-0.022	0.217	0.202
	(0.039)	(0.034)	(0.041)	(0.028)	(0.041)	(0.052)
FRANCE	-0.164	-0.302	-0.054	0.029	0.367	0.456
	(0.047)	(0.045)	(0.053)	(0.045)	(0.049)	(0.062)
GERMANY	-0.170	-0.214	-0.061	0.014	0.108	0.583
	(0.040)	(0.032)	(0.042)	(0.044)	(0.043)	(0.072)
ITALY	0.021	-0.211	0.071	0.029	0.246	0.364
	(0.056)	(0.057)	(0.058)	(0.052)	(0.060)	(0.071)
JAPAN	-0.065	-0.295	0.146	-0.122	0.210	0.291
	(0.037)	(0.044)	(0.045)	(0.037)	(0.043)	(0.054)
NETHERL.	-0.115	-0.153	-0.086	0.033	0.347	0.406
	(0.037)	(0.034)	(0.045)	(0.032)	(0.037)	(0.056)
NORWAY	-0.136	-0.118	-0.087	0.065	0.514	0.332
	(0.058)	(0.060)	(0.061)	(0.048)	(0.065)	(0.083)
SPAIN	-0.077	-0.176	0.050	-0.038	0.212	0.231
	(0.040)	(0.047)	(0.048)	(0.034)	(0.049)	(0.055)
SWEDEN	-0.080	-0.179	0.017	-0.043	0.246	0.273
	(0.041)	(0.042)	(0.042)	(0.029)	(0.042)	(0.052)
SWITZER.	-0.175	-0.189	-0.074	0.016	0.274	0.469
	(0.040)	(0.029)	(0.050)	(0.033)	(0.037)	(0.061)
UK	-0.065	-0.114	-0.038	0.041	0.499	0.353
	(0.043)	(0.045)	(0.047)	(0.036)	(0.055)	(0.067)
USA	-0.052	0.052	-0.026	0.021	0.427	0.120
	(0.032)	(0.035)	(0.034)	(0.026)	(0.033)	(0.044)
τ	0.399	-0.028	-0.963	-1.803	0.223	0.175
	(0.143)	(0.059)	(0.836)	(1.161)	(0.144)	(0.062)

Note: Standard errors in brackets

Table 1.4

Unobservable Factor Score Weights and Market Capitalization

Country	f21 %	f22 %	World %
AUSTRAL.	12.83	-7.51	1.46
AUSTRIA	-4.91	-4.93	0.21
BELGIUM	-2.81	8.59	0.73
CANADA	33.66	-37.12	2.19
DENMARK	4.49	-11.89	0.42
FRANCE	3.51	-3.43	3.02
GERMANY	-27.21	89.85	3.54
ITALY	3.46	13.07	1.46
JAPAN	1.70	-2.58	32.74
NETHERL.	9.87	17.58	1.46
NORWAY	7.53	1.06	0.10
SPAIN	2.20	2.88	1.25
SWEDEN	1.08	-3.73	0.42
SWITZER.	1.82	23.24	1.56
UK	7.91	9.89	10.95
USA	44.63	5.03	38.48

Table 1.5

Estimates of the Price of Idiosyncratic Risk

Country	Price	t-Ratio
AUSTRAL.	0.188	1.370
AUSTRIA	0.466	3.152
BELGIUM	1.337	4.019
CANADA	0.654	2.002
DENMARK	0.276	1.555
FRANCE	0.654	3.507
GERMANY	1.487	2.458
ITALY	0.082	0.723
JAPAN	0.639	2.763
NETHERL.	1.062	2.650
NORWAY	0.286	2.312
SPAIN	0.133	1.117
SWEDEN	0.436	2.724
SWITZER.	1.255	2.440
UK	0.383	2.734
USA	0.534	1.214

Joint Test: LR = 29.48 ($\chi^2_{16,0.05}$ = 26.3)

Table 1.6

Likelihood Ratio tests for Common Price of Risk

Factor:	f11	f12	f13	f14	f21	f22
LR:	13.26	14.39	22.65	18.61	32.67	25.28

Note: $\chi^2_{15,0.05} = 25.0$

Table 1.7

Proportion of Variance of Excess Returns Explained by Economic Factors

Country	4 Factors	10 Innovations
AUSTRAL.	0.0142	0.0525
AUSTRIA	0.1936	0.2083
BELGIUM	0.1641	0.1781
CANADA	0.0619	0.0918
DENMARK	0.1149	0.1497
FRANCE	0.1471	0.1701
GERMANY	0.1486	0.2035
ITALY	0.0535	0.0896
JAPAN	0.1936	0.2204
NETHERL.	0.1175	0.1412
NORWAY	0.0654	0.1567
SPAIN	0.0583	0.0927
SWEDEN	0.0748	0.0961
SWITZER.	0.1852	0.2161
UK	0.0356	0.0604
USA	0.0380	0.0525
AVERAGE	0.1062	0.1362

Table 1.8

Granger-Causality Tests of Explanatory Variable for Dividends

Country	F-Test
AUSTRAL.	2.882
AUSTRIA	2.271
BELGIUM	8.055
CANADA	4.038
DENMARK	3.872
FRANCE	10.890
GERMANY	3.101
ITALY	3.569
JAPAN	1.659
NETHERL.	3.737
NORWAY	3.268
SPAIN	11.362
SWEDEN	3.753
SWITZER.	3.211
UK	14.140
USA	4.144

Notes: Estimated regression is

$$\Delta_{12}D_{jt} = \text{const.} + \alpha_1\Delta_{12}D_{jt-12} + \alpha_2\Delta_{12}D_{jt-24} + \sum_{i=1}^{10}\beta_i\Delta_{12}X_{it}$$

H_0: $\beta_i=0$ for i=1,..,10

F-Tests include Newey-West (1987) correction ($F(10,207)_{0.05} \simeq 1.83$)

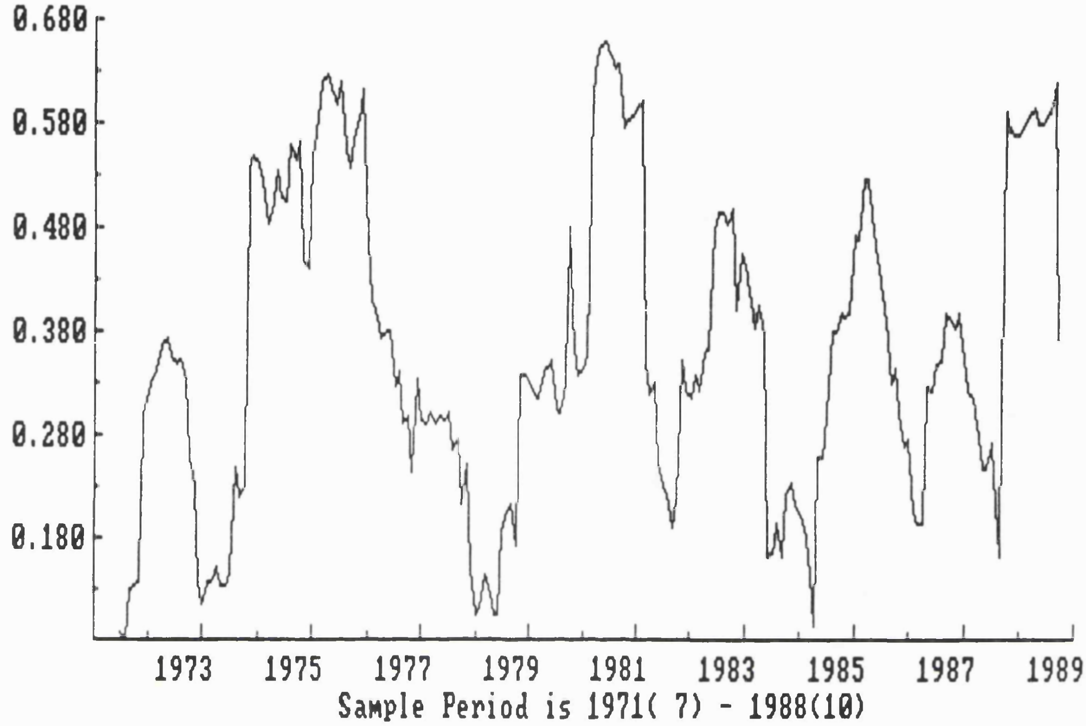


Figure 1.1

Equally-weighted average of individual cross-country correlation coefficients (over preceding 12 months)

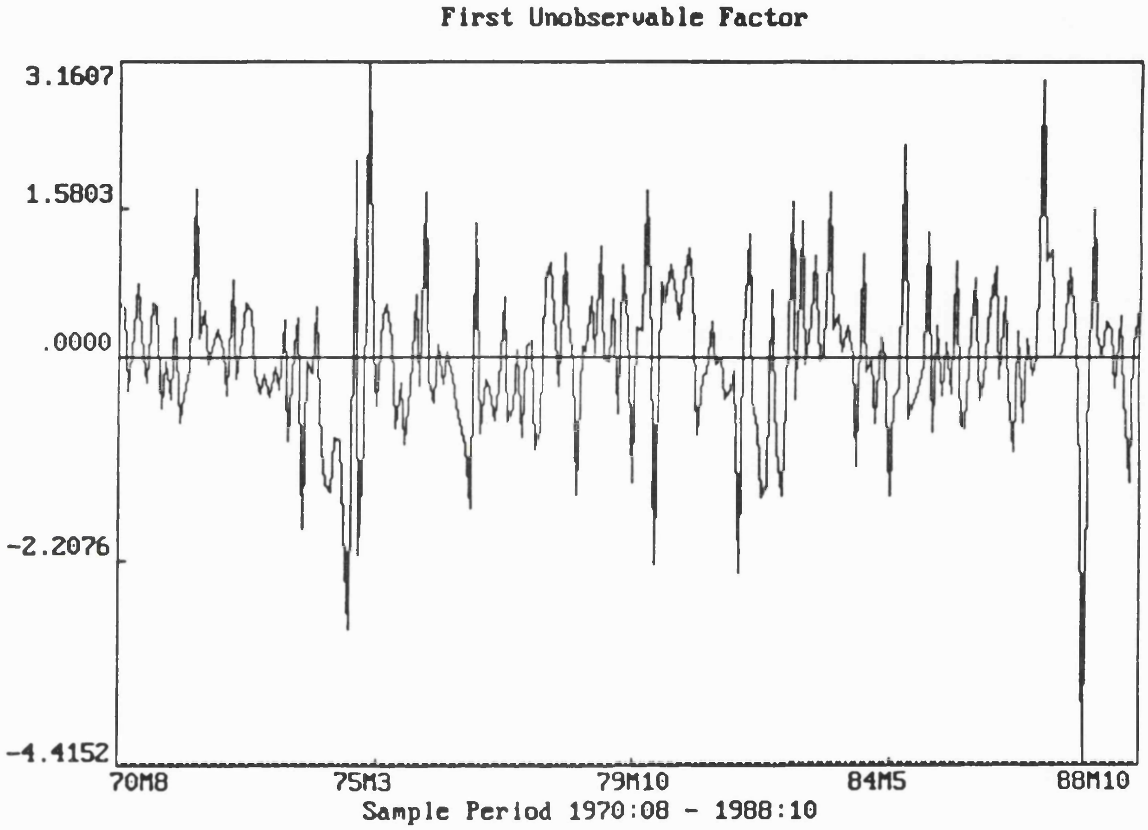

First Unobservable Factor
3.1607
1.5803
.0000
-2.2076
-4.4152
70M8
75M3
79M10
84M5
88M10
Sample Period 1970:08 - 1988:10

Figure 1.2a

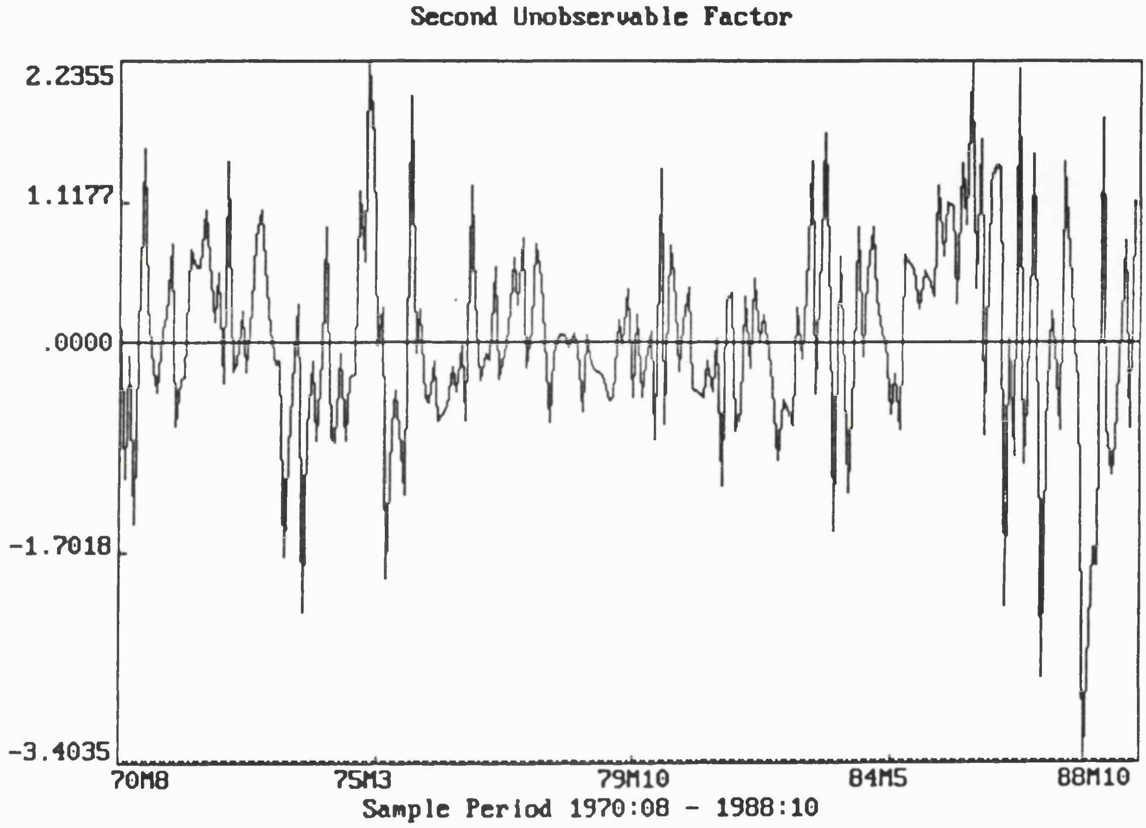

Second Unobservable Factor
2.2355
1.1177
.0000
-1.7018
-3.4035
70M8
75M3
79M10
84M5
88M10
Sample Period 1970:08 - 1988:10

Figure 1.2b

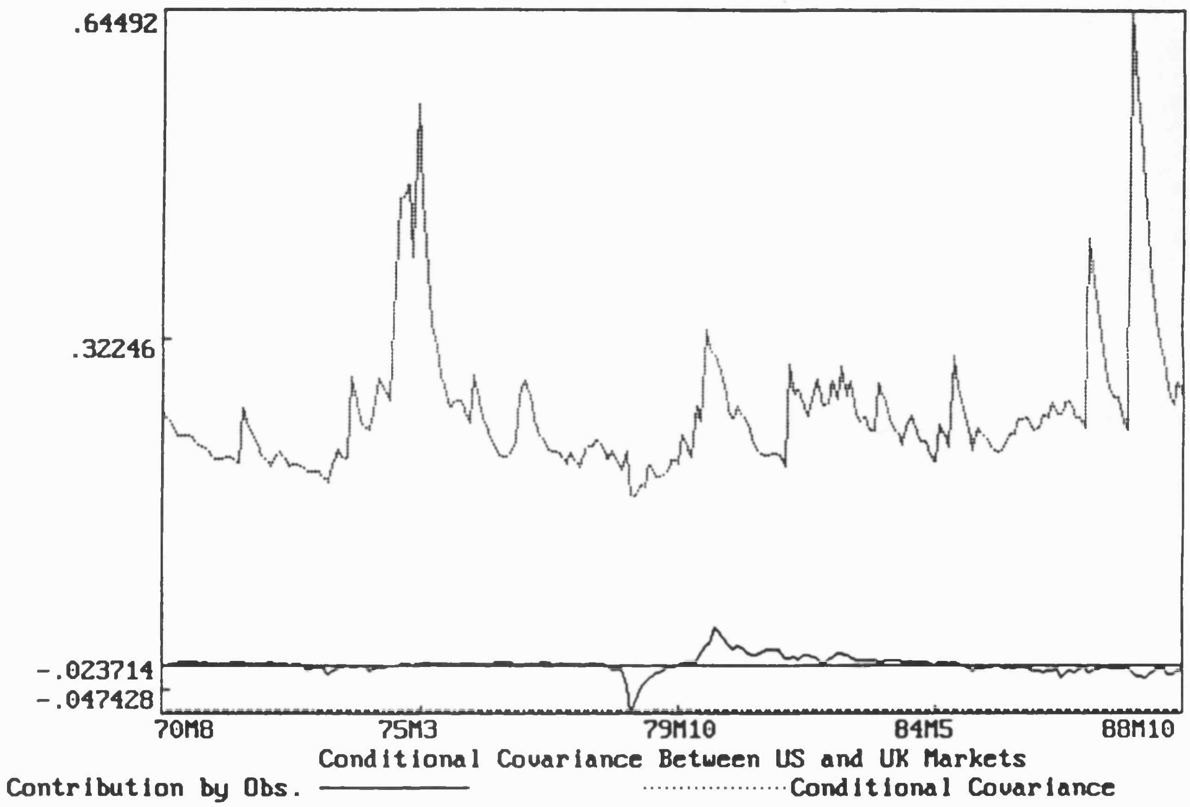

.64492
.32246
-.023714
-.047428
70M8
75M3
79M10
84M5
88M10
Conditional Covariance Between US and UK Markets
Contribution by Obs.
Conditional Covariance

Figure 1.3a

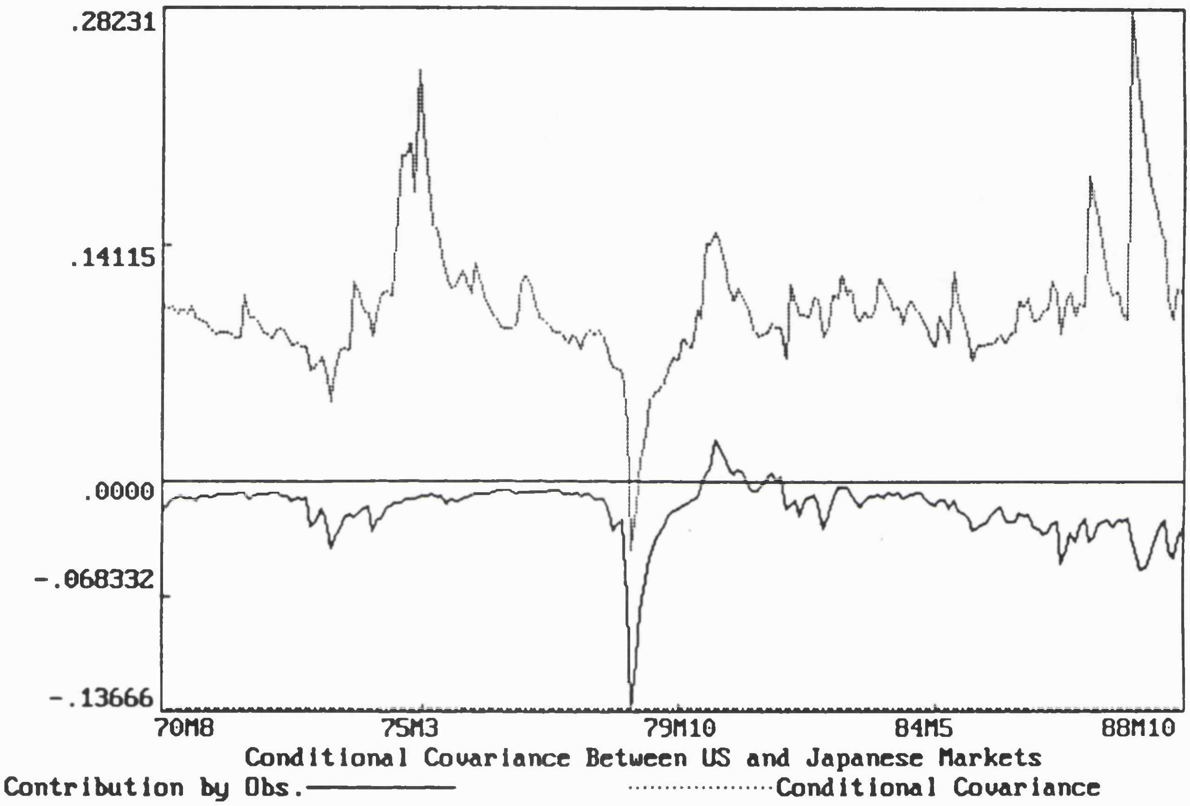

.28231
.14115
.0000
-.068332
-.13666
70M8
75M3
79M10
84M5
88M10
Conditional Covariance Between US and Japanese Markets
Contribution by Obs.
Conditional Covariance

Figure 1.3b

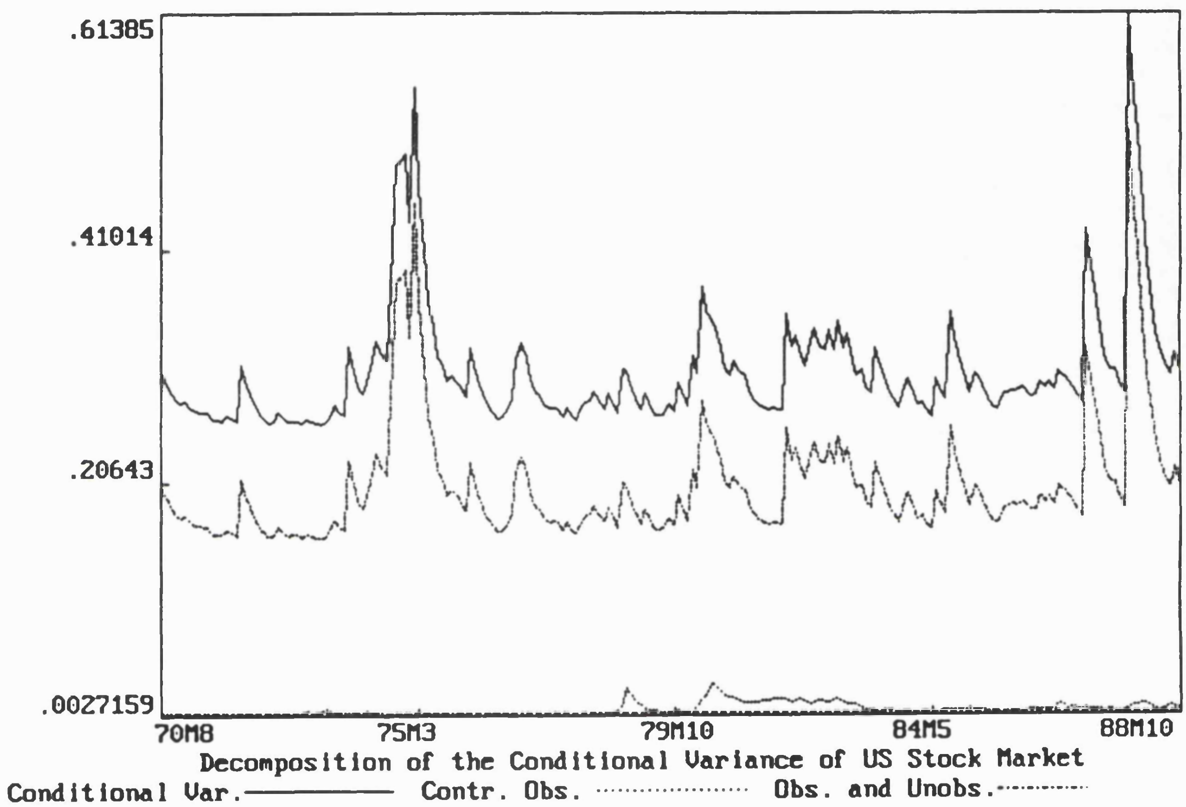
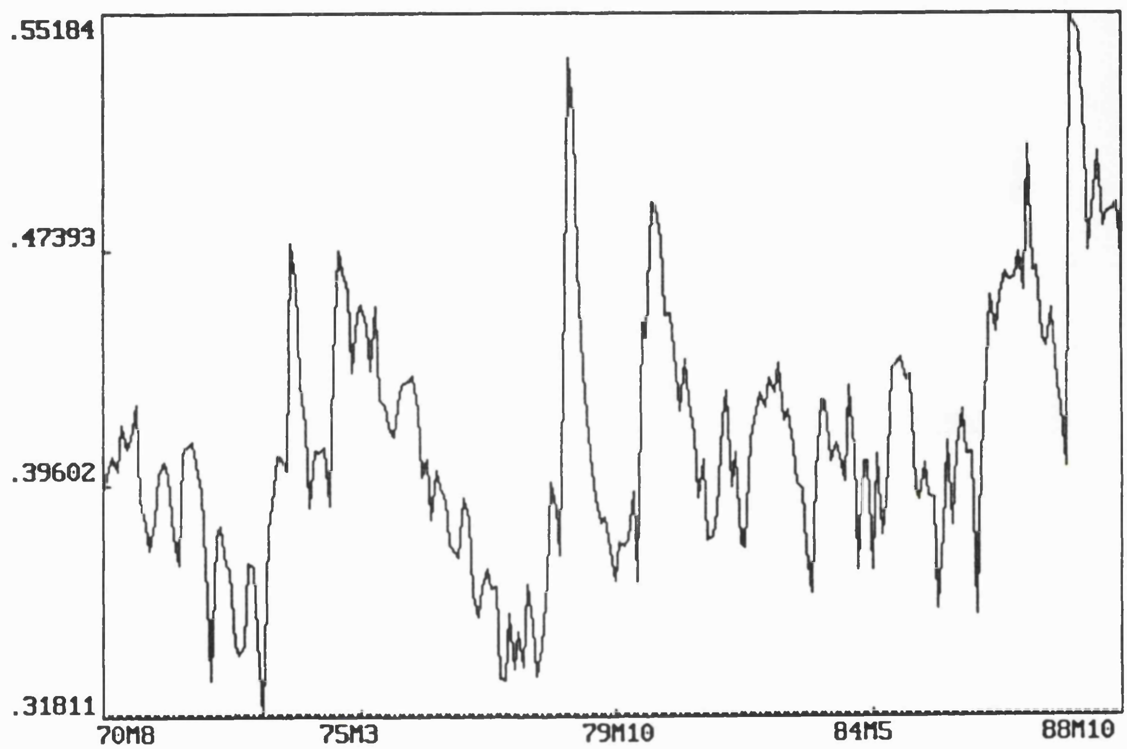

.61385
.41014
.20643
.0027159
70M8
75M3
79M10
84M5
88M10
Decomposition of the Conditional Variance of US Stock Market
Conditional Var.
Contr. Obs.
Obs. and Unobs.

Figure 1.4



Equally-weighted average of conditional correlations between markets

Figure 1.5

Chapter 2

IDENTIFICATION AND ESTIMATION OF CONDITIONALLY HETEROSKEDASTIC LATENT FACTOR MODELS

2.1 Introduction

In recent years increasing attention has been paid to modelling the behaviour over time of second moments, especially after the introduction of Engle's (1982) Autoregressive Conditional Heteroskedasticity (ARCH) model (see Bollerslev, Chou and Kroner (1990) for a recent survey). The ARCH(q) formulation assumes that the random variable ε_t is generated according to the following model:

$$\varepsilon_t = q_t^{1/2}\varepsilon_t^+, \quad \varepsilon_t^+ \sim \text{iid } (0,1), \; q_t = a_0 + \sum_{s=1}^{q} a_s \varepsilon_{t-s}^2 \tag{1}$$

This model has been extended to the multivariate case by Kraft & Engle (1983) as follows (cf. section 5.5):

$$y_t = H_t^{1/2} y_t^+, \quad y_t^+ \sim \text{iid } (0, I_m), \quad h_t = A_0 + \sum_{s=1}^{q} A_s \eta_{t-s} \tag{2}$$

where y_t is a mx1 vector, $h_t = v(H_t)$, $\eta_t = v(y_t y_t')$, A_0 is a $m(m+1)/2$ vector, the A_s are square matrices of order $m(m+1)/2$ and v() is the vector-lower triangle operator which stacks the lower triangular portion of a matrix (see Magnus (1988)).

However, the empirical application of dynamic conditional heteroskedasticity in a multivariate context has been hampered by the sheer number of parameters involved. For instance, without any restrictions, the number of parameters to be estimated in (2) is $O(qm^4)$ (cf. section 5.5), and for this reason in practice only two particular cases have been considered. The first one is the diagonal ARCH model which takes A_s (s=1,p) to be diagonal (see e.g. Attanasio & Edey (1987), Bollerslev, Engle & Wooldridge (1988) or Engle, Granger & Kraft (1984) as examples). The other one is the k-factor ARCH model of Engle (1987) in which k orthogonal linear combinations of the y_t-s follow univariate ARCH processes (see e.g. Engle, Granger & Kraft (1984), Engle (1987), Engle, Ng & Rothschild (1989, 1990), Kroner (1987), Lin (1991)).

Bollerslev (1990) has recently proposed a different parameterization for the time-varying variance-covariance matrix which holds the conditional correlation structure constant. Although this assumption simplifies the model considerably, in some instances the interest of the study may be the changing correlations themselves (as in chapter 1).

A new alternative approach to multivariate conditionally heteroskedastic models has been introduced by Diebold & Nerlove (1989) and extended in Chapter 1, and it is based on the same idea as traditional factor analysis models (see e.g. Johnson & Wichern (1982)). That is, it is assumed that each of m observed variables is a linear combination of k (k<m) common factors plus an idiosyncratic noise term, but allowing for dynamic conditional heteroskedastic-type effects in the underlying factors[1]. As in standard factor analysis it is in this way possible to obtain a parsimonious representation of the (conditional) second moments in terms of fewer processes.

In fact, Diebold & Nerlove (1989) propose this model as a natural way of capturing the co-movements in the variances of seven dollar exchange rates with a more parsimonious representation than the unrestricted multivariate ARCH model of (2), or indeed the diagonal version[2].

An additional advantage of this formulation is that in some cases (particularly in the context of Ross' (1976) Arbitrage Pricing Theory) it can be given a direct economic interpretation (see the discussion in section 1.2.1.3).

However its statistical properties have not been studied in detail. Several issues are particularly relevant. First the identification problems which affect factor analysis models (see section 1.2.3) have not been investigated for the case in which the factors show dynamic conditional heteroskedastic. This has important implications for empirical work related to the Arbitrage Pricing Theory, as the lack of identifiability of standard factor analysis models implies that the individual risk premia components associated with each factor are only identifiable up to an orthogonal transformation.

Another important issue that arises in the context of conditional heteroskedasticity is what distributional assumption should be made. Gaussianity is the most common one. However, although it can be proved that the associated unconditional distribution has thicker tails than the normal, conditional normality does not seem to capture completely the degree of leptokurtosis often observed in practice, especially in financial data. For this reason fat-tail conditional distributions have also been entertained in univariate models (see sections 4.3.2 and 5.4). Here we shall generalize Bollerslev's (1987) approach to the multivariate case by assuming that the joint conditional distribution of the factors is (proportional to) a multivariate t distribution, which includes the multivariate normal as a limiting case, but has generally fatter tails. This assumption ensures that the conditional distributions of the observed variables will be leptokurtic (see section 5.5 for an application).

Finally, the presence of unobservable variables makes statistical inference in this model somewhat complicated. In particular, it is of interest to discuss how ARCH-type effects can be handled since a standard ARCH model for the factors introduces unobservable components in the variance (see section 1.2.6 or Harvey, Ruiz & Sentana (1990)). For this reason, a Monte Carlo comparison of two related approaches to the Kalman-filter based maximum likelihood estimation procedure is carried out.

In section 2.2 the model is formally introduced. Identification is discussed in section 2.3, in which a generalization of sufficiency conditions for the constant-variance case is presented. The equations associated with the Kalman filter are derived in section 2.4. Maximum likelihood estimation is studied in section 2.5, whereas alternative ways to introduce ARCH and GARCH effects specifically are discussed in section 2.6. Some numerical simulations to compare the performance of these alternatives are carried out in section 2.7. Finally the conclusion are presented in section 2.8 together with some suggestions for further work.

2.2 A Multivariate Conditionally Heteroskedastic Latent Factor Model

Let's consider the following multivariate model

$$x_t = Cf_t + w_t \tag{3}$$

where

x_t is a mx1 vector of observed variables

f_t a kx1 vector of unobservable common factors

w_t a mx1 vector of unobservable idiosyncratic noises

C a mxk matrix of factor loadings, with m≥k and rank(C)=k, and both f_t and w_t are stochastic processes which show dynamic conditional heteroskedasticity.

In particular, we assume that given the information set available at time t-1, $X_{t-1}=\{x_{t-1},x_{t-2},\ldots\}$, $f_t=\Lambda^{1/2}_{t:t-1}f^+_t$ and $w_t=\Gamma^{1/2}_{t:t-1}w^+_t$, where:

$$\begin{bmatrix} f^+_t \\ w^+_t \end{bmatrix} \sim \text{iid}\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I_k & 0 \\ & I_m \end{bmatrix} \right] \tag{4}$$

$\Lambda_{t:t-1}$ is a kxk positive definite diagonal matrix and $\Gamma_{t:t-1}$ a mxm positive semidefinite diagonal matrix, with

$$\lambda_{it:t-1} = \lambda_i(X_{t-1}) \quad i=1,k \tag{5a}$$

$$\gamma_{jt:t-1} = \gamma_j(X_{t-1}) \quad j=1,m \tag{5b}$$

so that both $\Lambda_{t:t-1}$ and $\Gamma_{t:t-1}$ are measurable with respect to the information set (as in chapter 1)[3]. In order to retain full generality we shall not impose any restrictions on the exact functional forms in (5a) and (5b) (other than measurability with respect to X_{t-1}) until section 2.6. In any case, it is easy to see that the elements of f_t and w_t are serially uncorrelated with zero mean and a finite unconditional variance as in the homoskedastic case (provided that appropriate stationarity conditions are fulfilled), but in general they are not serially independent and their unconditional distributions are more leptokurtic than those of f^+_t and w^+_t (see section 5.3). Note also that the diagonality of $\Lambda_{t:t-1}$ and $\Gamma_{t:t-1}$, together with the contemporenous uncorrelatedness of $[f^{+\prime}_t, w^{+\prime}_t]$ implies that the factors are conditionally orthogonal. This, as we shall see, has important implications for the identifiability of the model.

This model is a straightforward generalization of Diebold & Nerlove (1989) who considered the case where k=1, the variance of w_t is constant (i.e. $\Gamma_t=\Gamma\ \forall t$) and the conditional variance of the common factor has an ARCH-type form. It is also an important special case of the model introduced in section 1.2.1, in which the variance of the common factors affects the mean of x_t (see also section 2.8 below), and the one discussed in Harvey, Ruiz and Sentana (1990), who allow for general dynamic in the mean. Besides if f_t and w_t are conditionally homoskedastic $\forall t$, it then reduces to the standard factor analysis model (e.g. Johnson & Wichern (1982)).

2.3 Sufficiency Conditions for Identification

Our assumptions imply that the distribution of x_t given X_{t-1} has conditional mean 0 and covariance matrix $\Sigma_{t|t-1}=V(x_t/X_{t-1})$ given by:

$$\Sigma_{t|t-1} = C\ \Lambda_{t|t-1}\ C' + \Gamma_{t|t-1} \tag{6}$$

It is clear than an arbitrary element of the model is the scaling of the factors. To remove this indeterminacy it is customary in the constant variance case to consider factors with unit variances. By analogy we shall require here $V(f_t)=I$[4].

Let's suppose that we were to ignore the time-variation in the conditional variances and base our estimation in the unconditional covariance matrix of x_t, Σ. Assuming that $V(w_t)=E(\Gamma_{t|t-1})=\Gamma$ exists, the unconditional variance is simply $\Sigma=CC'+\Gamma$.

As is well known from standard factor analysis models, it would then be possible in principle to generate an observationally equivalent model to (3) by premultiplying f_t by an arbitrary orthogonal kxk matrix Q and postmultiplying C by the transpose of that matrix since the covariance matrix Σ remains unchanged (see also the discussion in section 1.2.3).

Hence, some restrictions would be needed on C, and one way to impose them would be to use Dunn's (1973) set of sufficiency conditions for the homoskedastic factor model with orthogonal factors. These conditions are zero-type restrictions that guarantee that the only admissible orthogonal matrices Q above are I and its square roots (i.e. that C is locally identifiable up to column sign changes[5]). The conditions are stated in the following result:

Lemma 1:

Let $\Lambda=I$ and suppose that the columns of C are arranged so that for $s=1,2,\dots,k$ column s contains at least s-1 zeros. Let $C^{(s)}$ be any submatrix of C consisting of the s-1 left-most elements of any s-1 rows of C which have zeros in column s.

Then C is unique under orthogonal transformations (except for column sign) if for all $s=2,3,\dots,k$ there exists $C^{(s)}$ such that $|C^{(s)}|\neq 0$.

Proof: see Dunn (1973)

When C is otherwise unrestricted, imposing $c_{ij}=0$ for $j>i$ (i.e. C lower trapezoidal) implies that the condition above is satisfied. These restrictions mean that x_{1t} depends only on the first factor, x_{2t} on the first two, and so on until x_{kt}, $x_{k+1,t}$, ..., x_{mt} which depend on all k factors. Although this is clearly arbitrary (unless k=1), the factors can be orthogonally rotated to simplify their interpretation once the model has been estimated. In some other cases, identifiability can be achieved by imposing plausible a priori restrictions. For example, if in a two factor model it is believed that the second factor only affects a subset of the variables (say the first m_1, with $m_1<m$, so that $c_{i2}=0$ for $i=m_1+1,\dots,m$) the non-zero elements of C will always be identifiable.

Other alternative sets of sufficient local identifiability restrictions have been suggested, and for example Jennrich (1978) proves that when C is otherwise unrestricted, fixing (not necessarily to zero) the k(k-1)/2 supra-diagonal coefficients of (a permutation of) C also guarantees identifiability.

However, when time variation in $\Lambda_{t|t-1}$ is explicitly recognized in estimation, the set of admissible Q matrices is substantially reduced since the covariance matrix of the transformed factors $f_t^* = Qf_t$ has to remain diagonal $\forall t$.

Without loss of generality, let's divide the factors into two groups, the second of which, if it exists, is characterised for all t by a scalar covariance matrix (of at least dimension 2), i.e.:

$$\Lambda_{t|t-1} = \begin{bmatrix} \Lambda_{1t|t-1} & 0 \\ 0 & \lambda_{2t|t-1} I_{k_2} \end{bmatrix} \tag{8}$$

If we partition C accordingly, i.e.:

$$C = (C_1 \mid C_2) \tag{8}$$

the following result, which generalises Lemma 1, can be stated:

Lemma 2:

Let $\Lambda_{t|t-1}$ take the form of (7) and let $V(f_t)=I$.
Then C_1 is unique under orthogonal transformations (except for column sign) whereas the identifiability of C_2 can be established following Lemma 1

Proof: see appendix 2.

Notice the generality of Lemma 2 since it has been obtained without assuming any particular parameterization for the dynamic conditional heteroskedasticity, and hence relies only on time-variation of the conditional variances.

This result is apparently paradoxical, for relaxing the assumption of conditional homoskedasticity is what makes identification possible. The intuition, however, is as follows. Assume for simplicity that $\Lambda_{t|t-1}$ is not partially scalar (i.e. k_2=0). It is certainly true that for any t°, the orthogonally rotated factors $f_{t^\circ}^*$ =

$Q\Lambda^{-1/2}_{t^\circ:t^\circ-1}f_{t^\circ}$ and the rotated factor loading matrix $C^{*}_{t^\circ} = C\Lambda^{1/2}_{t^\circ:t^\circ-1}Q'$ generate the same conditional covariance matrix for x_{t°, for any orthogonal Q. Unlike in the homoskedastic case, though, different orthogonal rotations are required for different time periods. Hence the parameters in C are identifiable with respect to time-invariant orthogonal transformations.

As for the factors with common conditional variance, the particular parameterization chosen for $\lambda_i(X_{t-1})$ will imply more often than not that the only way two factors will always have the same variance is when this common variance is in fact constant. Lemma 2 could then be re-stated so that it would refer only to the relevant case when $\lambda_{2t:t-1}=1\ \forall t$. However in its present form it makes it clearer that the lack of identifiability comes from the factors having common, rather than constant, variances. Also, Lemma 2 could be modified by using an alternative set of identifiability conditions for C_2, such as the ones by Jennrich (1978) mentioned above.

The main message is the identifiability of C_1, so that even when C is unrestricted, identification problems only arise if the number of homoskedastic factors is at least 2. Therefore if none or only one of the factors is conditionally homoskedastic, the matrix C is locally identifiable under orthogonal transformations without additional restrictions, and the factors are uniquely defined. In this case, the imposition of unnecessary restrictions on C would produce totally misleading results. Nevertheless, the accuracy that can be achieved in estimating C depends on how much variability there is in $\Lambda_{t:t-1}$, for if the elements of this matrix are essentially constant for most of the time, identifiability problems will reappear.

2.4 Kalman Filter Equations

The multivariate conditionally heteroskedastic latent factor model (3) has a natural state-space representation. Taking f_t as the state it is clear that equation (3) can be interpreted as a measurement equation, with w_t being the measurement error, whereas the transition equation is trivial (Diebold & Nerlove (1989)):

$$f_t = \xi_t \tag{9}$$

where $V(\xi_t/X_{t-1})=\Lambda_{t|t-1}$.

Expressions (3) and (9) allow us to derive the Kalman filter prediction equations:

$$f_{t|t-1} = 0 \tag{10a}$$
$$x_{t|t-1} = 0 \tag{10b}$$
$$w_{t|t-1} = 0 \tag{10c}$$
$$\Omega_{t|t-1} = \Lambda_{t|t-1} \tag{10d}$$

with

$$\Omega_{t|t-1} = E[(f_t-f_{t|t-1})(f_t-f_{t|t-1})'/X_{t-1}] \tag{11}$$

(where the notation $t|t-1$ indicates the timing of expected value | information set).

In order to obtain minimum mean square error estimates of f_t (and w_t) given X_t, $f_{t|t}$, and the variances of the corresponding prediction errors, $\Omega_{t|t}$, we need the distributions of these variables given X_{t-1}, which ultimately depend on the distributions of f_t^+ and w_t^+. Normality is the usual assumption, and among other things it implies that the conditional distribution of x_t given X_{t-1} is normal. However, although the associated unconditional distribution will have thicker tails than the normal, conditional normality does not seem to capture completely the degree of leptokurtosis often observed in practice.

At the same time one must bear in mind that once a parametric distribution for $f_t,w_t/X_{t-1}$ is assumed, the linear form in (3) implies a particular distribution for x_t/X_{t-1} on which maximum likelihood estimation is based[6], and a conditional distribution for f_t/X_t from which the Kalman filter updating equations are derived. Fortunately, the special structure of the transition equation (9) allows for disturbances with fat-tailed conditional distributions.

In this respect, to allow for leptokurtosis while retaining tractability we shall assume that f_t^+ and w_t^+ follow a (standardized) multivariate t distribution (see Johnson & Kotz (1972) or Prucha & Kelejian (1984)), i.e.:

$$f_t^+ = [U_t/(\nu-2)]^{-1/2} f_t^* \tag{12a}$$

$$w_t^+ = [U_t/(\nu-2)]^{-1/2} w_t^* \tag{12b}$$

where f_t^*, w_t^* belong to a spherical normal distribution and U_t is an independent χ^2 with ν degrees of freedom (ν>2 but unknown)[7]. This implies that given X_{t-1} each factor (both common and idiosyncratic) is (proportional to) a Student's t distribution with ν degrees of freedom. Notice however that although the factors are conditionally orthogonal, they are not independent because of the common denominator.

If we now define the standardized variable $x_t^+ = \Sigma_{t:t-1}^{-1/2} x_t$ analogously, it is clear that x_t^+, f_t^+ (given X_{t-1}) follow a joint multivariate t distribution with variances I_k and I_m respectively, and covariance $\Sigma_{t:t-1}^{-1/2} C \Lambda_{t:t-1}^{1/2}$. As a consequence each x_{jt} (j=1,m) has a conditional distribution which is proportional to a standardized Student's t. Therefore by allowing $\nu<\infty$ we can have not only a leptokurtic unconditional distribution due to the ARCH effects, but also a fat tail conditional distribution with kurtosis coefficient $3(\nu-2)(\nu-4)^{-1}$ (provided ν>4) (see Bollerslev (1987) or Engle & Bollerslev (1986)).

Using the properties of the multivariate t distribution (see e.g. Zellner (1971)), the so-called updating equations, which give us the mean and variance of f_t conditional on X_t, would be:

$$f_{t:t} = \Lambda_{t:t-1}\, C' [C\, \Lambda_{t:t-1}\, C' + \Gamma_{t:t-1}]^{-1} x_t \tag{13a}$$

$$w_{t:t} = x_t - C f_{t:t} \tag{13b}$$

$$\begin{aligned} \Omega_{t:t} = {} & (\nu-2)(\nu+m-2)^{-1}[1+(\nu-2)^{-1} x_t' \Sigma_{t:t-1}^{-1} x_t] \cdot \\ & \cdot (\Lambda_{t:t-1} - \Lambda_{t:t-1} C' \Sigma_{t:t-1}^{-1} C\, \Lambda_{t:t-1}) \end{aligned} \tag{13c}$$

where $\Omega_{t:t}$ is not generally diagonal.

Notice that equations (13a) and (13b) coincide with the standard updating equation under normality (see Harvey (1989)), whereas a simple heteroskedastic scalar correction, which converges to 1 as $\nu\to\infty$, is introduced in (13c). Hence, no real extra complication arises by relaxing Gaussianity[8].

Another extremely convenient property of the model is that, given X_t, f_t and f_{t+u} are independent $\forall u\geq 0$ because of the degenerate nature of the transition equation. The smoothing equations then reduce to:

$$f_{t|T} = f_{t|t} \tag{14a}$$
$$\Omega_{t|T} = \Omega_{t|t} \tag{14b}$$

i.e. the smoothed estimates of f_t are the same as the filtered estimates (in fact $f_{t|t+u}=f_{t|t}$ and $\Omega_{t|t+u}=\Omega_{t|t}$ $\forall u\geq 0$). The intuition is that once x_t is observed, there is no extra information on f_t in future observations (see section 1.2.4). This property generally leads to substantial simplification in the way particular variance structures are handled (see sections 2.6 and 5.5 for examples)

Notice that for the standard (conditionally homoskedastic) orthogonal factor model, which is nested in equation (3) when $\lambda_{it|t-1}=\lambda_i$ $\forall i,t$ and $\gamma_{jt|t-1}=\gamma_j$ $\forall j,s$, the Kalman filter estimates:

$$f_{t|t}=C'(CC'+\Gamma)^{-1}x_t \tag{15}$$

correspond to what is known in the factor analysis literature as regression estimates of the factor scores, $\dot{f}_t$ (Johnson & Wichern (1982)). In fact, since given our assumptions, f_t and w_t are serially uncorrelated (but neither independent nor do they follow a Student's t) with zero mean and variances $\lambda_i=1$ $\gamma_j=E(\gamma_{jt|t-1})$, $\dot{f}_t$ are minimum mean square <u>linear</u> estimators even in the presence of conditionally heteroskedastic effects. This suggests that traditional methods, such us principal components based factor analysis, can also be applied to estimate the parameters in C and Γ. However in general not all the elements of C will be identifiable without restrictions using this procedure, as discussed in section 2.3.

2.5 Maximum Likelihood Estimation

The likelihood function of the sample can be constructed (ignoring initial conditions) as the product of the conditional distributions of the x_t-s given X_{t-1}. For the multivariate t distribution we have that:

$$L = -Tm/2 \ln[\pi(\nu-2)] + T\{\ln \Gamma[(\nu+m)/2]-\ln \Gamma(\nu/2\} -$$

$$- 1/2 \sum \{\ln |\Sigma_{t:t-1}| + (\nu+m) \ln[1+(\nu-2)^{-1}x_t'\Sigma_{t:t-1}^{-1}x_t]\} \qquad (16)$$

where $\Sigma_{t:t-1}$ is given by equations (6). This equation reduces to equation (25) in section 1.2.5 when $\nu\ \infty$, i.e., under normality. Maximum likelihood estimates of the parameters of interest are then obtained by numerical maximization of (16). Note that the non-normality assumption does not make the evaluation of the likelihood function more cumbersome, except for the gamma functions which do not depend on t.

Pre-sample values of $\lambda_{it:t-1}$ and $\gamma_{jt:t-1}$ could be set to their unconditional expectations (1 and γ_j respectively), and the corresponding values of $\Omega_{t:t}$ to 0. As a consequence $V(f_1/X_0)=I$, $V(w_1/X_0)=\Gamma$, their unconditional counterparts.

The fact that m is usually much bigger than k and the diagonality of $\Lambda_{t:t-1}$ and $\Gamma_{t:t-1}$ can be exploited by means of the Woodbury formula (Householder (1964)) so that the inversion of $\Sigma_{t:t-1}$ (a mxm matrix), only involves the inversion of the kxk matrix $[\Lambda_{t:t-1}^{-1}+C'\Gamma_{t:t-1}^{-1}C]$ (for the case of $\Gamma_{t:t-1}$ not having full rank see Sentana (1989)).

2.6 Introducing ARCH and GARCH-type effects

So far, we have deliberately left unspecified the functional forms of the conditional variances of the factors. In practice, though, a particular parameterisation is required, and undoubtedly, Engle's (1982) ARCH and Bollerslev's (1986) Generalised ARCH (or

GARCH) are the most popular ones. For simplicity, we shall start with ARCH(q) formulations (but see section 5.5 for an extension of this approach to a more general formulation).

There are basically three different ways of allowing for ARCH-type effects in f_t and w_t:

a) The most obvious one is to assume that they follow Engle's (1982) ARCH(q) processes, i.e.:

$$f_{it} = f^{*}_{it}\lambda^{1/2}_{it} = f^{*}_{it}(\psi_{io} + \sum_{s=1}^{p} \psi_{is} f^{2}_{it-s})^{1/2} \quad i=1,k \tag{17a}$$

$$w_{jt} = w^{*}_{jt}\gamma^{1/2}_{jt} = w^{*}_{jt}(\theta_{jo} + \sum_{S=1}^{p} \theta_{js} w^{2}_{jt-s})^{1/2} \quad j=1,m \tag{17b}$$

Positivity of λ_{it} and non-negativity of γ_{jt} is ensured by the parameter restrictions $\psi_{is} \geq 0$ $\forall i,s$, $\psi_{io} > 0$ $\forall i$, and $\theta_{js} \geq 0$ $\forall j,s$ $\theta_{jo} \geq 0$, $\forall j$.

The main advantage of this parameterization is that standard results apply: e.g. the elements of f_t and w_t are serially uncorrelated (but not independent) with zero mean and unconditional variance given by $\lambda_i = \psi_{io}/(1 - \sum_{s=1}^{q} \psi_{is})^9$ and $\gamma_j = \theta_{jo}/(1 - \sum_{s=1}^{q} \theta_{js})$ (provided that $\sum_{s=1}^{q} \psi_{is} < 1$ $\forall i$, and $\sum_{s=1}^{q} \theta_{js} < 1$ $\forall j$) and their unconditional distributions are leptokurtic. Besides, if fourth moments exist, the autocorrelation functions of f^{2}_{it} and w^{2}_{jt} behave like the autocorrelation function of q-th order autoregressive processes with coefficients ψ_{is} and θ_{js}.

However neither f_{t-s} nor w_{t-s} are measurable with respect to X_{t-1} and hence $\lambda_{it|t-1} = V(f_{it}/X_{t-1})$ and $\gamma_{jt|t-1} = V(w_{jt}/X_{t-1})$ do not generally coincide with $\psi_{io} + \sum_{s=1}^{p} \psi_{is} f^{2}_{it-s}$ and $\theta_{jo} + \sum_{S=1}^{p} \theta_{js} w^{2}_{jt-s}$ above. In fact:

$$\lambda_{it|t-1} = \psi_{io} + \sum_{s=1}^{q} \psi_{is} E(f^{2}_{it-s}/X_{t-1}) \quad i=1,k \tag{18a}$$

$$\gamma_{jt|t-1} = \theta_{jo} + \sum_{s=1}^{q} \theta_{js} E(w^2_{jt-s}/X_{t-1}) \qquad j=1,m \tag{18b}$$

where we have used the fact that the conditional expectation of f_t and w_t given X_{t-1} are 0. Besides it is not difficult to see that all covariances (given X_{t-1}) between the elements in both vectors are 0, so that $\Lambda_{t|t-1}$ and $\Gamma_{t|t-1}$ are diagonal matrices.

A more fundamental problem of using (17a) and (17b) is that the form of the conditional distribution of f_t (and w_t) given X_{t-1} is unknown, and e.g. there is no distribution for f^*_t which results in conditional normality for f_t/X_{t-1} (see also Harvey, Ruiz & Sentana (1990)). As a consequence, the Kalman filter is not optimal in the sense that it does not provide minimum mean square error estimates of the state variables f_t and w_t. However, as

$$f_{t-s} = f_{t-s|t-1} + (f_{t-s}-f_{t-s|t-1}) \tag{19}$$

(where the notation t-s|t-1 indicates the timing of the Kalman filter estimate | information set), if one proceeds using the Kalman filter "as if" f_t/X_{t-1} and w_t/X_{t-1} were t-distributed, then the approximate conditional variance for f_{it} will be given by:

$$\lambda_{it|t-1} \simeq (1-\sum_{s=1}^{q} \psi_{is}) + \sum_{s=1}^{q} \psi_{is} f^2_{it-s|t-1} + \sum_{s=1}^{q} \psi_{is} \omega_{iit-s|t-1} \tag{20}$$

where $\omega_{iit-s|t-1}$ is the i-th diagonal element of the mean square error matrix $\Omega_{t-s|t-1}$, and where from (14), |t-1 can be replaced by |t-s[10].

Comparing $\lambda_{it|t-1}$ with $\lambda_{it}=\psi_{io} + \sum_{s=1}^{p} \psi_{is} f^2_{it-s}$ it is clear that the former is the latter evaluated at the estimates of the factors as of time t-1 plus a term reflecting the uncertainty in those estimates.

Also, since:

$$\begin{aligned} w_{t-s} = y_{t-s}-Cf_{t-s} = (y_{t-s}-Cf_{t-s|t-1})+C(f_{t-s}-f_{t-s|t-1}) = \\ w_{t-s|t-1} + C(f_{t-s}-f_{t-s|t-1}) \end{aligned} \tag{21}$$

then we have

$$\gamma_{jt|t-1} \simeq \theta_{jo} + \sum_{s=1}^{p} \theta_{js} w^2_{jt-s|t-1} + \sum_{s=1}^{p} \theta_{js} c'_j \Omega_{t-s|t-1} c_j \tag{22}$$

where c'_j is the j-th row of C, and again the last term reflects the uncertainty of not knowing w_{jt-s}.

b) One possible way to get around these difficulties is to assume as an alternative formulation that $f_t = \Lambda^{1/2}_{t|t-1} f^*_t$ and $w_t = \Gamma^{1/2}_{t|t-1} w^*_t$, with the elements of $\Lambda_{t|t-1}$ and $\Gamma_{t|t-1}$ given by (18a) and (18b), so that the conditional variances are now measurable functions with respect to the information set X_{t-1}. As a consequence, (20) and (22) become exact and $f_{t-s|t-s}$ and $\Omega_{t-s|t-s}$ are indeed $E(f_{t-s}/X_{t-s})$ and $V(f_{t-s}/X_{t-s})$ respectively. This is the specification chosen in chapter 1.

The main problem with this approach is that the elements of f_t and w_t no longer follow Engle's (1982) ARCH(q) processes, and hence ARCH results cannot be used. Despite this fact, it turns out to be the case that the f_t-s and w_t-s so-generated are also uncorrelated with zero mean and the same unconditional variance, and their unconditional distribution are leptokurtic as well, although less than in the previous case. Besides, f^2_{it} and w^2_{jt} are also serially correlated, but their autocorrelation functions are bounded from above by the autocorrelation function of qth order autoregressive processes with coefficients ψ_{is} and θ_{js} (see the appendix in Harvey, Ruiz & Sentana (1990) for details). The intuition is that this second formulation implies smoother processes for f^2_{it} and w^2_{jt} (but not for f_{it} and w_{jt}) than (17) do.

c) The third possible way of introducing ARCH-type behaviour in f_t and w_t is to assume, as in b), that $f_t = \Lambda^{1/2}_{t|t-1} f^*_t$ and $w_t = \Gamma^{1/2}_{t|t-1} w^*_t$, but with the elements of $\Lambda_{t|t-1}$ and $\Gamma_{t|t-1}$ given by:

$$\lambda_{it|t-1} = (1 - \sum_{s=1}^{p} \psi_{is}) + \sum_{s=1}^{p} \psi_{is} E^2(f_{it-s}/X_{t-1}) \tag{23a}$$

and

$$\gamma_{jt|t-1} = \theta_{jo} + \sum_{s=1}^{p} \theta_{js} E^2(w^2_{t-s}/X_{t-1}) \tag{23b}$$

which, again, are measurable functions with respect to the information set so that the Kalman filter equations are exact. Notice that the square is now over the expectation, not over the unobservable components.

This is the actual specification used in Diebold & Nerlove (1989), and has the advantage that the recursion for $\Omega_{t|t}$ does not need to be carried out, as the elements of this matrix no longer affect any other estimate.

Once more, the problem with (23) is that standard ARCH results cannot be used. However it is possible to prove that f_{it} and w_{jt} are still uncorrelated with zero mean but with a smaller unconditional variance, and that their unconditional distributions are leptokurtic, although even less than in the previous case. Besides f^2_{it} and w^2_{jt} are also serially correlated, but their autocorrelation functions are now bounded from above by the autocorrelation function of the processes discussed in b) (see Harvey, Ruiz & Sentana (1990)). The intuition again is that (23) imply even smoother processes for f^2_{it} and w^2_{jt}.

Two likelihood functions can been used to estimate the parameters of the models discussed in a), b) and c). They are both based on (16) but differ in the precise form of the conditional variances. In the first case, i) say, equations (20) and (22) are used, whereas in the second case, ii) say, equations (23a) and (23b) are used instead. The validity of i) or ii) obviously depends on which one of the three data generation process a), b) or c) is the "true" one:

a) In this case i) is an approximation to the true likelihood function in the sense that neither is f_{it}/X_{t-1} t-distributed nor is (20) its conditional variance. In turn, ii) is an approximation to this approximation since it ignores the contribution of $\psi_{is}\omega_{t-s|t-1}$.

b) Here i) is the exact conditional distribution as f_{it}/X_{t-1} is indeed t-distributed with (20) as its exact conditional variance. However, ii) is still an approximation because it uses the wrong conditional variance (although the right functional form).

c) Now the reverse situation occurs since ii) becomes exact whereas i) approximate. This apparent paradox arises because b) and c) have in fact been devised so that i) and ii) yield the corresponding exact likelihood functions.

In practice, the choice between a), b) and c) cannot be decided completely on a priori grounds as, after all, our only concern must be the conditional distribution of the observed series x_t, and not the conditional distribution of the unobservable components. As a matter of fact, a) and b) are indistinguishable in the data as both are estimated using the same likelihood function.

In order to see how we could deal with Bollerslev's (1986) GARCH(p,q)-type effects in the factors let's consider a GARCH(1,1)-type process for simplicity.

a) In this case

$$\lambda_{it} = \psi_{io} + \psi_{i1} f^2_{it-1} + \phi_{i1}\lambda_{it-1} \qquad i=1,k \tag{24}$$

which, provided that $\phi_{i1}<1$, can be re-written as:

$$\lambda_{it} = \sum_{j=0}^{\infty} \phi^j_{i1}(\psi_{io}+\psi_{i1} f^2_{it-1-j}) \qquad i=1,k \tag{25}$$

i.e. as an ARCH(∞) with exponentially declining weights (see Bollerslev (1986)). As a consequence,

$$\lambda_{it|t-1} \simeq \sum_{j=0}^{\infty} \phi^j_{i1}[\psi_{io}+\psi_{i1}(f^2_{it-1-j|t-1}+\omega_{iit-1-j|t-1})] \tag{26}$$

Applying the so-called Koyck transformation we obtain:

$$\lambda_{it|t-1} \simeq \psi_{io}+\psi_{i1}(f^2_{it-1|t-1}+\omega_{iit-1|t-1})+\phi_{i1}\lambda_{it-1|t-2}+$$

$$+\psi_{i1}\sum_{j=1}^{\infty}\phi_{i1}^{j}[(f_{it-1-j|t-1}^{2}+\omega_{iit-1-j|t-1})-$$
$$-(f_{it-1-j|t-2}^{2}+\omega_{iit-1-j|t-2})]$$

But given that in this model conditioning on the future is equivalent to conditioning on the present (cf. the smoothing equations (13a) and (13b)), the last term vanishes and we are simply left with:

$$\lambda_{it|t-1} \simeq \psi_{io}+\psi_{i1}(f_{it-1|t-1}^{2}+\omega_{iit-1|t-1})+\phi_{i1}\lambda_{it-1|t-2} \tag{27}$$

which has the same structure as (24).

b) If we replace equation (26) by:

$$\lambda_{it|t-1} = \sum_{j=0}^{\infty}\phi_{i1}^{j}[\psi_{io}+\psi_{i1}E(f_{it-1-j}^{2}/X_{t-1})] \qquad i=1,k \tag{28}$$

an analogous transformation yields eventually:

$$\lambda_{it|t-1} = \psi_{io} + \psi_{i1}E(f_{it-1}^{2}/X_{t-1}) + \phi_{i1}\lambda_{it-1|t-2} \tag{29}$$

which has the same structure as (27) but it is now exact.

c) Proceeding as in a) and b) it is clear that we will have in this case

$$\lambda_{it|t-1} = \psi_{io} + \psi_{i1}E^{2}(f_{it-1}/X_{t-1}) + \phi_{i1}\lambda_{it-1|t-2} \tag{30}$$

as expected.

2.7 Numerical Simulations

Given that $\Omega_{t|t}$ does not have to be computed if we use a likelihood function based on (23a) and (23b), which may produce substantial savings in the computations, it is interesting to check the validity of such procedure when in fact the conditional variances are given by (20) and (22). Besides, it is also interesting to see the effects of estimating the model ignoring its ARCH structure. As in Diebold and Nerlove (1989) and chapter 1, conditional normality of f_t, w_t has been assumed.

To do so a small simulation study has been conducted in which five hundred samples of 200 observations each (plus another 100 for initialization) have been generated according to the following simple model using the NAG library G05DDF routine:

$$\begin{aligned} x_{1t} &= c_1 f_t + w_{1t} \\ x_{2t} &= c_2 f_t + w_{2t} \\ x_{3t} &= c_3 f_t + w_{3t} \end{aligned} \tag{31}$$

with $c_1=c_2=c_3=c$ and

$$\lambda_{t:t-1} = 1-\psi + \psi E(f^2_{t-1}/X_{t-1}) \tag{32}$$

$$\begin{aligned} \gamma_{1t:t-1} &= \theta_0 + \theta_1 E(w^2_{1t-1}/X_{t-1}) \\ \gamma_{2t:t-1} &= \theta_0 + \theta_1 E(w^2_{2t-1}/X_{t-1}) \\ \gamma_{3t:t-1} &= \theta_0 + \theta_1 E(w^2_{3t-1}/X_{t-1}) \end{aligned} \tag{33}$$

where for simplicity the factor loadings and ARCH parameters of the idiosyncratic variances are common[11]. Having only one factor implies that the c-s are identifiable even in the homoskedastic case.

In order to make the comparison of simulations simpler, the value of θ_0 is chosen so that the unconditional variance of the x-s is one, i.e.:

$$\theta_0 = (1-\theta_1)(1-c^2) \tag{34}$$

which implies that the unconditional correlation between the variables is simply c^2.

Two values of c have been selected, c=0.5, 0.9 representing low and high correlation (0.25, 0.81 respectively). Two values have also been selected for ψ and θ_1, namely 0.2 and 0.8 representing low and high persistence in variance. Therefore 8 parameter combinations in all have been considered.

Maximization of the log-likelihood function (16) with respect to the 6 parameters ($c_1, c_2, c_3, \psi, \theta_0$ and θ_1) was carried out on the LSE VAX using the NAG library E04JBF routine. Non-negativity and stationarity restrictions on ψ, θ_0 and θ_1 were imposed by reparameterizing in terms of $\sin^2(\psi^*)$, θ_0^{*2} and $\sin^2(\theta_1^*)$, with ψ^*, θ_0^* and θ_1^* unrestricted.

The results are presented in Tables 1 to 6. For simplicity of exposition, only averages are presented for all items which vary across equations, e.g. c_1=1/3 (c_1+c_2+c_3).

Table 2.1 shows the mean parameter estimates together with their biases, variances and mean square errors across replications for both exact and approximate ML estimates.

The most obvious result is that both estimates of c_i are generally upward biased, the latter substantially more than the former, especially for high ψ. Besides the persistence parameters of the variances, ψ and θ_1, are mostly underestimated, with the approximate and exact ML estimates being usually similar. As for the remaining parameter, θ_0, both methods yield upward biased estimates but the bias in the exact ML method is much smaller, almost negligible.

Although only eight parameter combinations are obviously not enough to carry out a detailed analysis of how these biases depend on the parameter values, it is perhaps worth mentioning that it seems that the biases in estimating c increase with ψ and slightly decrease with c itself, and also that the biases of the estimates of θ_1, and especially ψ, appear to increase with the value of the corresponding true parameter.

Table 2.2 contains the average of the square differences between the exact and approximate factor scores estimates and the true ones, i.e. the mean square errors of these factor scores estimates.

The first thing to notice is that despite the different parameter estimates, these statistics are remarkably close for the approximate and exact methods, with the ones based on the former being nearly identical for w_{jt} and slightly larger for f_t.

These mean square errors tend to decrease with c and θ_1, but while those referring to the common factor seem to increase with ψ, those related to the specific errors appear to be a decreasing function of ψ. However if we divide the reported mse by the corresponding unconditional variances, the estimates of w_{jt} do relatively better than those of f_t for c=0.5 and relatively worse for c=0.9. This is hardly surprising since the unconditional signal to noise variance ratio goes from 1/3 to 81/19 for f_t, and from 3 to 19/81 for w_{jt} respectively.

A similar, and unit free, measure of signal extraction performance is given by the correlations between the estimated and true factor scores presented in table 2.3.

For the common factor, the correlations vary from as low as 0.700 to as high as 0.979, whereas for the specific factors they go from 0.833 to 0.934. In fact, the pattern of correlations between factor scores across experiment designs is very similar to that of the (weighted) mean square errors in table 2.2, and again approximate and exact methods yield very similar results.

Table 2.4 presents the average sample means of the different estimates of the conditional variances of f_t and w_{jt}. In addition this table also contains the values of the (log) density function evaluated using the conditional variances of x_t given by the exact and approximate ML methods.

In order to analyze the results for the ML variance estimates, it is worth noticing that the sample averages of the true variances tend to be smaller than their theoretical counterparts (1 and 0.75 for c=0.5, 0.19 for c=0.9) when the associated persistence parameter is high. Recalling that in our simulations high persistence is associated with a smaller constant in the conditional variance to maintain the unconditional variance at the same level (cf. eqs. 26 and 28), the discussion in Hendry (1986) on the small persistence of simulated ARCH effects may rationalize this finding (see also Engle, Hendry and Trumble (1985)). Hence, it is not surprising that the averages of the time-varying variance of these factor scores are smaller the greater the corresponding ARCH effects, mimicking the behaviour of the true ones on a bigger scale.

The results also show that the approximate ML method produce substantially smaller variances of the common factor than the exact one, especially for high ψ and low c, as one would expect since the approximate method in (22) has one less term than (15). However the averages for the specific variances are very much the same, despite the difference between (23) and (18).

The picture that emerges from the above results is that although the approximate method produces more upward biased estimates of c and smoother common factor scores than the exact ML procedure, they balance in such a way that both methods do an equally good job in decomposing the observed series and their conditional variances into common and idiosyncratic components. This similarity is not too surprising as the latent factor model analysed here has no dynamics in the mean, only in the variance. Hence, the correction term has a one-off effect on the variance of the current observation and does not carry over to further observations.

If we recall that the (log) sample density functions (cf. (16)) differ in the way $V(x_t/X_{t-1})$ is computed, this conclusion is confirmed by the fact that in table 2.4 the average values of the likelihood function for both methods are almost identical, with a perfect correlation. Again, this is not unexpected: since the model has no conditional mean, the forecast error in (c_j times) f_t must cancel with the forecast error in w_{jt}.

The performance of the principal components based factor analysis is summarized in tables 5 and 6. For comparative purposes the sample variances and covariances of the generated observations x_{jt} are included in table 2.5.

In order to analyse the results it is worth remembering three facts about how these estimates are computed:

* they are based on the unconditional variance of x_t, and therefore should be less efficient that the ML based factor scores which exploit the conditional variances (although less so for ψ, θ_1 low).

* they use an approximation to the unconditional covariance matrix $\Sigma = CC' + \Gamma$ which assumes Γ small, and so should do badly for Γ high.

* the actual parameter estimates are essentially chosen so as to minimize the trace of the residual variance $\Sigma - \hat{C}\hat{C}'$, so one would expect $\hat{C}$ to be biased upwards while $\hat{\Gamma}$ downwards.

This is broadly speaking what happens. Indeed this method performs best for c high, ψ, θ_1 low (with factor scores almost as good as the ML based ones), and worst for c low, ψ, θ_1 high. Nevertheless, the results for the parameter estimates are distorted by the fact that the sample covariance matrix of x_t (on which these estimates are based) is generally downward biased when the ARCH effects are high, in line with the variances of the factors[12].

2.7 Conclusions

In this chapter the issues of identification and exact vs approximate estimation of multivariate conditionally heteroskedastic latent factor models have been discussed.

It turns out to be the case that the model only suffers from lack of identification in as much as the variances of some of the common factors are constant. In particular, if all but one common factor have time-varying variances the factor loading matrix C is (locally) identifiable under orthogonal transformations. Thus, there is a non-trivial advantage in explicitly recognising the existence of dynamic conditional heteroskedasticity when estimating factor analytic models.

With respect to the validity of the approximation used by Diebold and Nerlove (1989), our simulations suggest that the exact ML estimation procedure employed in chapter 1 performs significantly better in terms of the estimates of the factor loading matrix C and the time-varying variance of f_t/X_{t-1}, with the approximate method producing upward and downward biased estimates of this quantities respectively. However these biases appear to balance each other so that the decomposition of the observed series and their conditional variances into common and idiosyncratic components achieved by both methods are nearly identical. Unsurprisingly, the exact procedure involves a larger computational burden (about 45% slower on average in terms of CPU time).

The results also suggest that ignoring the ARCH structure of the factors imply that the variance of the specific terms may be seriously underestimated, although, as expected, this method works quite well when both the ARCH effects and the idiosyncratic variances are small. Besides, one would expect it to be more robust to specification errors. However in many economic applications of interest considering time-varying volatility seems crucial.

Details of the use of a multivariate t distribution with unknown degrees of freedom as an alternative to the multivariate normal are also presented. This distribution has thicker tails than the normal but converges to it as the degrees of freedom increase. This assumption allows for a higher degree of excess kurtosis in the unconditional distribution of the observed variables, and therefore should capture a higher proportion of the leptokurtosis exhibited by many financial data sets.

The model discussed in this chapter can be extended in several interesting ways. An important extension would include weakly exogenous or lagged dependent variables in the conditional mean of (3), e.g. x_t could be the innovations of a vector autoregressive process. Although this does not pose any theoretical difficulty and does not affect the analysis in section 2.4, in practice the maximization is not so simple since the number of additional unknown parameters would be generally large. Not surprisingly both Diebold & Nerlove (1989) and ourselves in Chapter 1 estimate first the autoregressions and carry out the maximum likelihood method on the residuals. Therefore finding efficient computational methods, constitutes an important remaining task.

The result presented here can also be applied to other closely related models, and in particular to the one in chapter 1 in which the conditional variance of the factors affects the mean:

$$x_t = D\Lambda_{t|t-1}\iota + Cf_t + w_t \tag{35}$$

where ι is a vector of ones. Note that the results of section 2.4 are perfectly valid since $\Lambda_{t|t-1}$ is by definition known at time t-1[13].

The identification results also apply to the model in Harvey, Ruiz and Sentana (1990) as well as to the *common trends* model (see e.g. Harvey (1989) or Stock & Watson (1988)):

$$x_t = Cy_t + w_t \tag{36a}$$

$$y_t = y_{t-1} + f_t \tag{36b}$$

where y_{t-1} is a kx1 vector of common trends and f_t, w_t are defined as in (2a), (2b). Again, time-variability in the conditional variances of the f_t-s will eliminate the lack of identifiability of the matrix C which arises when conditional homoskedasticity is assumed.

ENDNOTES

1 Despite the similarity in their names, the conditionally heteroskedastic latent factor model discussed here should not be confused with Engle's (1987) Factor ARCH model. The main difference is where the time-varying variances are introduced: on the unobservable orthogonal factors in the heteroskedastic factor model, on the linear combinations of the observed variables which proxy the factors in the Factor ARCH model. In many respects, though, both models are rather similar.

2 Since they only have one common factor and the idiosyncratic factors are homoskedastic, the conditional variances of the observed variables are perfectly correlated in their model. This would also be true in Engle's (1987) one-factor ARCH model.

3 In this respect, it is worth noticing that the information set available at time t-1, X_{t-1}, contains only lagged values of x_t. This has the fundamental implication that in general neither f_{t-1} nor w_{t-1} belong to the information set, and hence much care has to be exercised in specifying the conditional variances of f_t and w_t (see sections 1.2.4 and 2.6).

4 If the unconditional variance does not exist, other scaling assumptions could be made just as well, e.g. $\lambda_i(0)=1\ \forall i$.

5 The local identifiability can be trivially transformed into a global one by fixing arbitrarily the sign of one non-zero coefficient in each column of C.

6 Stable distributions, which are invariant under addition, would be potential candidates, were it not for the fact that the only member of this class with finite variance is the normal. Hence we must look for non-independent distributions.

7 The thickness of the tails of the multivariate t distribution depends on the degrees of freedom parameter ν. This distribution converges to a spherical normal distribution as ν tends to infinity.

8 These results can be further generalised if we assume that $[f_t^{+\prime}, w_t^{+\prime}]$ follow a joint elliptical distribution (see Fang, Kotz and Ng (1990)), of which the multivariate normal and the multivariate t are examples. In that case, $[f_t', x_t']$ is also jointly elliptical given X_{t-1}, as the elliptical class is closed under linear transformations. As a consequence, $f_{t|t}$ is still given by (13a) since the regression functions are linear for all elliptical distributions, and again $\Omega_{t|t} = g(x_t'\Sigma_{t|t-1}^{-1}x_t) \cdot (\Lambda_{t|t-1} - \Lambda_{t|t-1}C'\Sigma_{t|t-1}^{-1}C\,\Lambda_{t|t-1})$, where g() is a scalar function whose form depends on the particular member of the class used (see also Sentana (1991c)).

9 This implies that to fix the scale of the factors so that $V(f_t)=I$, we shall require $\psi_{i0}=1-\sum_{s=1}^{q}\psi_{is}$ (as in section 1.2.4).

10 If f_t/X_{t-1} and w_t/X_{t-1} were t-distributed, then $f_{t-s|t-s} = E(f_{t-s}/X_{t-s})$ and $\Omega_{t-s|t-s}=V(f_{t-s}/X_{t-s})$.

11 The fact that the true c-s are common is not exploited in estimation. However the restriction of common ARCH parameters in the idiosyncratic variances is imposed when maximizing the likelihood function.

12 Results similar to those presented in this section are obtained using $f_t=f_t^{+}\,[(1-\psi)+\psi f_{t-1}^2)]^{1/2}$ and $w_{it}=w_{it}^{+}\,(\theta_0+\theta_1 w_{it-1}^2)^{1/2}$ as the data generation process (cf. section 2.6 a).

13 If we were to assume that the data generation process is $f_t=f_t^{+}\Lambda_t^{1/2}$ (cf. section 2.6a), it would be tempting to think that equation (35) should be replaced by $x_t = D\Lambda_t\imath + Cf_t + w_t$, which would pose a problem as Λ_t is not in the econometrician's information set. However if the agents information set is indeed X_{t-1}, this formulation would be inappropriate because any underlying economic model would still imply equation (35). This potential problem is not shared by the ARCH-M model of Engle, Lillien & Robins (1987), in which $z_t= g(q_t) + \varepsilon_t$, with q_t,ε_t given by (1), and g() being an arbitrary function of the variance. To see why let's consider the "alternative" process $z_t = g[V(z_t/Z_{t-1})] + \varepsilon_t$. In this case ε_t is certainly in the information set at time t but then $V(z_t/Z_{t-1})=q_t$.

APPENDIX 2

To prove Lemma 2, let's call $\Lambda^{*}_{t|t-1}$ the covariance matrix of the transformed factors $f^{*}_{t}= Qf_{t}$, where Q is an arbitrary orthogonal matrix.

Let's partition Q as:

$$
Q = \begin{array}{cc}
\begin{array}{cc} k-k_2 & k_2 \end{array} & \\
\begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} & \begin{array}{c} k-k_2 \\ k_2 \end{array}
\end{array}
$$

in accordance with (7) and (8).

We shall show that the only admissible transformations are given by:

$$
\begin{array}{cc}
\begin{array}{cc} k-k_2 & k_2 \end{array} & \\
\begin{bmatrix} I^{1/2} & 0 \\ & Q_{22} \end{bmatrix} & \begin{array}{c} k-k_2 \\ k_2 \end{array}
\end{array}
$$

where Q_{22} is orthogonal and $I^{1/2}I^{1/2}= I$.

To see why let's partition $\Lambda^{*}_{t|t-1}=Q\Lambda_{t|t-1}Q'$ as:

$$
\Lambda^{*}_{t|t-1}=\begin{bmatrix} Q_{11}\Lambda_{1t|t-1}Q'_{11}+\lambda_{2t|t-1}Q_{12}Q'_{12} & Q_{11}\Lambda_{1t|t-1}Q'_{21}+\lambda_{2t|t-1}Q_{12}Q'_{22} \\ Q_{11}\Lambda_{1t|t-1}Q'_{21}+\lambda_{2t|t-1}Q_{12}Q'_{22} & Q_{21}\Lambda_{1t|t-1}Q'_{21}+\lambda_{2t|t-1}Q_{22}Q'_{22} \end{bmatrix}
$$

Given that $\Lambda_{1t|t-1}$ is time varying, for $\Lambda^{*}_{t|t-1}$ to preserve the form of (7) for all t the following conditions must all hold:

a) $Q_{11}\Lambda_{1t|t-1}Q'_{11}$ diagonal

b) $Q_{12}Q'_{12}$ diagonal

c) $Q_{11}\Lambda_{1t|t-1}Q'_{21}$ null

d) $Q_{12}Q'_{22}$ null

e) $Q_{21}\Lambda_{1t|t-1}Q'_{21}$ scalar

f) $Q_{22}Q'_{22}$ scalar

Let q_{21i} be the i-th column of Q_{21} and λ_{1itt-1} the i-th diagonal element of $\Lambda_{1t|t-1}$ ($i=1,k-k_2$). Then e) can be re-written as:

e') $\sum_{i=1}^{k-k_2} \lambda_{1it|t-1}q_{21i}q'_{21i}$

Now, since $\lambda_{1it|t-1}$ varies with both i and t, the expression in e) will be scalar if and only if $q_{21i}q'_{21i}$ is scalar for all i. But $q_{21i}q'_{21i}$ is scalar if and only if $q_{21i}=0$, so $Q_{21}=0$, and c) is also satisfied.

Besides f) transforms into:

f') $Q_{22}Q'_{22}=I$

so that Q_{22} must be orthogonal. But then d) is satisfied if and only if $Q_{12}=0$, and then b) is also satisfied.

Finally if q_{11i} is the i-th column of Q_{11} ($i=1,k-k_2$), a) can be re-stated as:

a') $\sum_{i=1}^{k-k_2} \lambda_{1it|t-1}q_{11i}q'_{11i}$diagonal

By a similar argument this condition will be satisfied if and only if each q_{11i} has a single non-zero element. Positive definiteness and the exclusion of mere permutations of the factors imply that Q_{11} must be (a square root of) the unit matrix. *q.e.d.*

Table 2.1

Maximum Likelihood Estimates

	c_1			ψ			θ_0			θ_1		
	T	E	A	T	E	A	T	E	A	T	E	A
Param	**0.5**	0.509	0.556	**0.2**	0.194	0.203	**0.600**	0.600	0.621	**0.2**	0.183	0.184
$bias^2$		0.000	0.003		0.000	0.000		0.000	0.000		0.000	0.000
var		0.012	0.028		0.048	0.053		0.005	0.004		0.007	0.008
MSE		0.013	0.031		0.048	0.053		0.005	0.005		0.008	0.008
Param	**0.5**	0.503	0.524	**0.2**	0.197	0.194	**0.150**	0.153	0.213	**0.8**	0.772	0.765
$bias^2$		0.000	0.001		0.000	0.000		0.000	0.004		0.001	0.001
var		0.003	0.004		0.024	0.024		0.001	0.001		0.010	0.010
MSE		0.003	0.005		0.024	0.024		0.001	0.005		0.011	0.011
Param	**0.5**	0.572	0.772	**0.8**	0.636	0.603	**0.600**	0.601	0.618	**0.2**	0.192	0.192
$bias^2$		0.005	0.074		0.027	0.039		0.000	0.000		0.000	0.000
var		0.082	0.134		0.094	0.092		0.004	0.003		0.006	0.006
MSE		0.087	0.208		0.121	0.131		0.004	0.004		0.006	0.006
Param	**0.5**	0.549	0.735	**0.8**	0.719	0.708	**0.150**	0.151	0.201	**0.8**	0.776	0.774
$bias^2$		0.002	0.055		0.007	0.008		0.000	0.003		0.001	0.001
var		0.041	0.082		0.042	0.043		0.001	0.001		0.008	0.007
MSE		0.043	0.137		0.049	0.052		0.001	0.003		0.008	0.008
Param	**0.9**	0.897	0.905	**0.2**	0.189	0.189	**0.152**	0.153	0.164	**0.2**	0.187	0.188
$bias^2$		0.000	0.000		0.000	0.000		0.000	0.000		0.000	0.000
var		0.005	0.005		0.014	0.014		0.000	0.000		0.011	0.011
MSE		0.005	0.005		0.014	0.014		0.000	0.000		0.011	0.011
Param	**0.9**	0.905	0.910	**0.2**	0.195	0.194	**0.038**	0.040	0.064	**0.8**	0.776	0.758
$bias^2$		0.000	0.000		0.000	0.000		0.000	0.001		0.001	0.002
var		0.004	0.004		0.012	0.012		0.000	0.000		0.012	0.012
MSE		0.004	0.004		0.012	0.012		0.000	0.001		0.013	0.013
Param	**0.9**	0.949	1.029	**0.8**	0.732	0.727	**0.152**	0.153	0.163	**0.2**	0.192	0.192
$bias^2$		0.002	0.017		0.005	0.005		0.000	0.000		0.000	0.000
var		0.154	0.161		0.026	0.025		0.000	0.000		0.010	0.010
MSE		0.156	0.177		0.031	0.031		0.000	0.000		0.010	0.010
Param	**0.9**	0.947	0.996	**0.8**	0.744	0.740	**0.038**	0.039	0.062	**0.8**	0.769	0.764
$bias^2$		0.002	0.009		0.003	0.004		0.000	0.001		0.001	0.001
var		0.196	0.194		0.021	0.021		0.000	0.000		0.011	0.011
MSE		0.198	0.203		0.024	0.025		0.000	0.001		0.012	0.012

T: True values, E: Exact ML estimates, A: Approximate ML estimates

Table 2.2

Factor Scores Mean Square Errors

			Common factor		Idiosyncratic factors	
c	ψ	θ_1	E	A	E	A
0.5	**0.2**	**0.2**	0.502	0.510	0.129	0.130
0.5	**0.2**	**0.8**	0.327	0.331	0.083	0.083
0.5	**0.8**	**0.2**	0.475	0.530	0.114	0.115
0.5	**0.8**	**0.8**	0.338	0.416	0.071	0.072
0.9	**0.2**	**0.2**	0.074	0.074	0.058	0.058
0.9	**0.2**	**0.8**	0.046	0.047	0.035	0.035
0.9	**0.8**	**0.2**	0.153	0.163	0.054	0.054
0.9	**0.8**	**0.8**	0.104	0.115	0.032	0.032

T: True values, E: Exact ML estimates, A: Approximate ML estimates

Table 2.3

Correlations between Factor Scores Estimates

			Common factor			Idiosyncratic factors		
c	ψ	θ_1	E vs T	A vs T	A vs E	E vs T	A vs T	A vs E
0.5	**0.2**	**0.2**	0.707	0.706	0.999	0.910	0.910	1.000
0.5	**0.2**	**0.8**	0.822	0.821	0.999	0.922	0.921	1.000
0.5	**0.8**	**0.2**	0.700	0.700	0.995	0.922	0.922	0.999
0.5	**0.8**	**0.8**	0.816	0.814	0.998	0.934	0.933	0.999
0.9	**0.2**	**0.2**	0.963	0.963	1.000	0.833	0.833	1.000
0.9	**0.2**	**0.8**	0.979	0.978	1.000	0.872	0.872	0.999
0.9	**0.8**	**0.2**	0.952	0.952	1.000	0.846	0.846	1.000
0.9	**0.8**	**0.8**	0.972	0.972	1.000	0.882	0.881	0.999

T: True values, E: Exact ML estimates, A: Approximate ML estimates

Table 2.4

Sample Averages of Conditional Variances and Likelihood Function

			Common factor			Specific factors			Likelihood function		
c	ψ	θ_1	T	E	A	T	E	A	E	A	E vs A
0.5	**0.2**	**0.2**	1.00	0.99	0.87	0.75	0.74	0.74	-273.0	-273.0	1.000
0.5	**0.2**	**0.8**	1.00	1.00	0.92	0.70	0.70	0.69	-141.9	-142.2	1.000
0.5	**0.8**	**0.2**	0.95	0.85	0.53	0.75	0.75	0.74	-265.9	-266.2	1.000
0.5	**0.8**	**0.8**	1.03	0.92	0.54	0.70	0.70	0.70	-115.7	-116.4	0.999
0.9	**0.2**	**0.2**	1.00	1.00	0.99	0.19	0.19	0.19	-55.7	-55.7	1.000
0.9	**0.2**	**0.8**	1.00	1.00	0.99	0.18	0.18	0.18	38.1	37.1	0.999
0.9	**0.8**	**0.2**	0.97	0.90	0.75	0.19	0.19	0.19	-0.6	-0.6	1.000
0.9	**0.8**	**0.8**	0.95	0.94	0.81	0.18	0.18	0.18	104.9	104.0	0.999

T: True values, E: Exact ML estimates, A: Approximate ML estimates

Table 2.5

Principal Components based Estimates

	c_i T	c_i P	γ T	γ P	σ_{ii} T	σ_{ii} P	σ_{ij} T	σ_{ij} P
Param	**0.5**	0.700	**0.75**	0.490	**1.0**	0.994	**0.25**	0.251
$bias^2$		0.040		0.068		0.000		0.000
var		0.014		0.008		0.014		0.006
MSE		0.054		0.076		0.014		0.006
Param	**0.5**	0.671	**0.75**	0.394	**1.0**	0.943	**0.25**	0.256
$bias^2$		0.029		0.127		0.004		0.000
var		0.099		0.033		0.578		0.005
MSE		0.128		0.160		0.581		0.005
Param	**0.5**	0.686	**0.75**	0.496	**1.0**	0.989	**0.25**	0.234
$bias^2$		0.035		0.064		0.000		0.000
var		0.021		0.011		0.024		0.015
MSE		0.056		0.076		0.024		0.015
Param	**0.5**	0.657	**0.75**	0.395	**1.0**	0.995	**0.25**	0.263
$bias^2$		0.025		0.126		0.004		0.001
var		0.127		0.041		0.655		0.116
MSE		0.152		0.167		0.658		0.116
Param	**0.9**	0.929	**0.19**	0.126	**1.0**	0.994	**0.81**	0.804
$bias^2$		0.001		0.004		0.000		0.000
var		0.004		0.000		0.015		0.012
MSE		0.005		0.004		0.015		0.012
Param	**0.9**	0.935	**0.19**	0.112	**1.0**	0.995	**0.81**	0.821
$bias^2$		0.001		0.006		0.000		0.000
var		0.008		0.003		0.048		0.013
MSE		0.010		0.009		0.048		0.013
Param	**0.9**	0.869	**0.19**	0.126	**1.0**	0.969	**0.81**	0.778
$bias^2$		0.001		0.004		0.001		0.001
var		0.088		0.000		0.977		0.976
MSE		0.089		0.004		0.978		0.977
Param	**0.9**	0.858	**0.19**	0.114	**1.0**	0.940	**0.81**	0.761
$bias^2$		0.002		0.006		0.004		0.002
var		0.089		0.003		1.096		1.065
MSE		0.091		0.009		1.100		1.067

T: True values, P: Principal Component based estimates

Table 2.6

Factor Scores Mean Square Errors and Correlations

			Common factor		Idiosyncratic factors	
c	ψ	θ_1	MSE	P vs T	MSE	P vs T
0.5	**0.2**	**0.2**	0.598	0.702	0.261	0.807
0.5	**0.2**	**0.8**	0.627	0.687	0.341	0.775
0.5	**0.8**	**0.2**	0.634	0.673	0.267	0.806
0.5	**0.8**	**0.8**	0.750	0.644	0.349	0.769
0.9	**0.2**	**0.2**	0.077	0.963	0.064	0.814
0.9	**0.2**	**0.8**	0.075	0.965	0.066	0.797
0.9	**0.8**	**0.2**	0.200	0.949	0.064	0.813
0.9	**0.8**	**0.8**	0.199	0.947	0.072	0.793

T: True values, P: Principal Component based estimates

Chapter 3

SEMI-PARAMETRIC ESTIMATION AND THE PREDICTABILITY OF STOCK MARKET RETURNS: SOME LESSONS FROM JAPAN

3.1 Introduction

There is growing evidence for the view that stock market returns are predictable (see, e.g. Fama & French (1988a,b), Poterba & Summers (1988)). Fama & French argue that these lagged variables which help predict returns are merely proxying for risk[1]. By contrast, other authors (e.g. Shiller (1984)) argue that stock returns are predictable because of the existence of "fads" or "noise traders". Since risk is not directly observable, its modelling is a crucial part of any attempt to discriminate between these two views.

For that reason, we attempt to model risk in three different ways - the standard procedure of using lagged squared excess returns as a proxy, a measure based on assuming a Generalised Autoregressive Conditional Heteroskedasticity in Mean specification (GARCH-M hereafter - see Engle et al. (1987)) and a measure of volatility based on recent advances in semi-parametric econometrics (see Pagan & Ullah (1988)).

We choose to study the Japanese stock market - in part because it is a market that puzzles outside observers (and investors!) and also because it is relatively under-researched and therefore most likely to yield evidence of a different character. We find that, in contrast to the evidence from Anglo-Saxon markets, the lagged dividend yield does not have a positive association with future returns. We go on to present some evidence from Korea, which suggests that there is a negative association between the dividend yield and stock returns.

Other predictor variables that we use include lagged returns and nominal interest rates, both of which normally have some predictive ability in the Anglo-Saxon markets. Further, in deference to much comment in the financial markets that the "weight of money" is important in Japan, we also include velocity.

The rest of the chapter is organised as follows. Section 3.2 reviews possible theoretical arguments for including the above variables. The econometric methodology is presented in section 3.3. Results are discussed in section 3.4, and we explore other possibilities in section 3.5. Finally, we conclude in section 3.6.

3.2 Theoretical Considerations

3.2.1 The Efficient Markets Model vs the "Fads" Model

The most commonly used model of stock returns makes the expected excess return a function of volatility (see, e.g. Merton (1980) and section 1.2.1.3), i.e.:

$$\mu_t = \rho\sigma_t^2 \tag{1}$$

where μ_t, σ_t^2 denote the conditional mean and variance of excess returns in period t (as of time t-1), $r_t = r_t^m - r_t^f$, with r_t^m the return on the market portfolio, r_t^f the risk-free rate, and ρ a measure of aggregate risk aversion, which we assume to be constant.

If we additionally assume Rational Expectations, then actual ex-post returns are given by

$$r_t = \rho\sigma_t^2 + \varepsilon_t \tag{2}$$

where $E_{t-1}(\varepsilon_t)=0$, $V_{t-1}(\varepsilon_t)=\sigma_t^2$, i.e., we have the "Efficient Markets" implication that information available at time t-1 cannot help us earn risk-adjusted excess returns.

A prominent alternative to the efficient markets model (EMM, hereafter) are those models where "fads" or "noise traders" influence share prices (see, e.g. Shiller (1984), De Long et al. (1987)). Typically, one has two kinds of agents: the "ordinary" investors who act exogenously and have some (proportionate) demand for shares, Y_{t-1} say (see section 4.2 for an example); and the "smart money" who are assumed to have a demand function for shares of the form:

$$Q_{t-1} = \frac{E_{t-1}(r_t)}{\mu(\sigma_t^2)} \tag{3}$$

where Q_{t-1} is the fraction of shares that they hold and $\mu(\sigma_t^2)$ is the risk premium needed to induce them to hold all the shares. The essential assumption in (3) is that $\mu'()>0$, i.e. that "smart" agents are risk averse so they do not take infinitely large positions when the expected return on the market deviates from the risk-free rate. Imposing market-clearing (i.e. $Q_{t-1} + Y_{t-1} = 1$) yields the result that:

$$E_{t-1}(r_t) = \mu(\sigma_t^2) - \mu(\sigma_t^2)Y_{t-1} \tag{4}$$

If we linearise $\mu()$ in (4), it is clear that the EMM in (2) is nested within (4). The "fads" model has the property that any variable dated t or earlier which are known to affect the demand of "ordinary" investors, Y_{t-1}, will also help predict future returns.

Therefore, in principle, if there are variables that help predict risk-adjusted excess returns, then this finding may be interpreted as evidence in favour of the "fads" model and against the EMM[2].

We next discuss some possible variables that might help predict risk-adjusted excess returns.

3.2.2 Dividend Yields and Stock Returns

The view that dividend yields help forecast returns is not new (see, e.g. Dow (1920)). Unsurprisingly, various possible explanations have been offered for this finding. One possibility is that the lagged dividend yield is just a proxy for risk (see, e.g. Fama & French (1988a)). Campbell & Shiller (1988) show that the log dividend-price ratio is (approximately) a function of the entire expected future time path of the difference between the return on equity (r_t) and dividend growth (g_t), i.e.

$$\ln(D_t/P_t) \simeq E_t \left[\sum_{j=0}^{\infty} \phi_1^j (r_{t+j} - g_{t+j}) + \phi_2 \right] \tag{5}$$

where ϕ_1, ϕ_2 are constants of linearisation. So if we mis-measure the expected value of r_t (by using an incorrect proxy for σ_t^2 say), then the dividend yield is likely to act as proxy for variations in expected returns. Therefore, when we estimate (2), we shall make σ_t^2 a function of the dividend yield. If the dividend yield continues to have an additional direct influence on expected returns, we shall argue that this fact makes an explanation based on risk less likely. Notice also that the risk-based explanation predicts a positive link between dividend yields and returns.

An alternative explanation of a link between dividend yields and returns may be in terms of "fads" (see Shiller (1984)). On this view, the dividend yield is a proxy for the optimism of ordinary investors, so, e.g. a period of excessive optimism is characterised by a dividend yield that is "too low" (relative to some steady-state value) and is likely to lead to a price fall (negative return). This, once again, results in a positive association between excess returns and the dividend yield.

However, casual inspection of the Japanese data suggests that a positive association between this two variables is unlikely. The dividend yield has been on a downward trend, its average value of 0.57% (on an annual basis) during 1985-89 was only a quarter of its average value during 1970-1974 (see Table 3.1)[3]. This fall in the dividend yield has been associated with a huge increase in five-year excess returns (200% in 1985-89 as compared to 29% in 1970-74). Hence, the relationship between dividend yields and stock returns in Japan may be quite different from that in the Anglo-Saxon countries.

3.2.3 The Effect of Liquidity

Practitioners commonly assert that the standard share valuation models should be disregarded in assessing the Japanese market, and that one should look at alternative factors[4]. Notable among these factors is the real money supply, which is a commonly recurring "explanatory" variable (see, e.g. Goldman, Sachs & Co., International Economic Analyst, March 1988). We shall therefore whether past (and known) levels of the money supply can help predict excess returns. Such a link could be rationalised in terms of a "fads" model of share price, if it were true that the demand for shares by "ordinary" investors, Y_t, were correlated with real money supply.

3.2.4 Inflation and Stock Returns

There is a great deal of evidence that higher inflation and/or nominal interest rates lead to lower stock returns (see, e.g. Fama & Schwert (1977), for the US, Cohn & Lessard (1981), for international evidence). One possible explanation is that higher inflation may lead to an increase in the variance of stockholders' returns either because higher inflation leads to greater relative price variability, or it is associated with more uncertain inflation (see, e.g. Pindyck (1984)). So, provided that we measure risk appropriately, there should be no link. Just as in the case of the dividend yield, we shall make σ_t^2 a function of inflation. Note, though, that Pindyck's model explain why unanticipated changes in inflation can lead to lower excess returns today, for an unanticipated rise in inflation leads to a rise in expected volatility which, this period, causes a fall in share prices. However, if inflation, and therefore volatility, are expected to remain higher thereafter, this then leads to higher excess stock returns because the required return is higher - so there is a positive association between the level of expected inflation and stock returns. Yet the evidence from Anglo-Saxon markets suggests a negative association.

Another popular explanation of the link between inflation and share prices is that advanced by Modigliani & Cohn (1979). They argue that investors commit valuation errors and, in particular, incorrectly compare the earnings yield with the nominal interest rate. We could embed such considerations into the "fads" model discussed above, where we may argue that the demand for shares by ordinary investor, Y_{t-1}, depends negatively on future expected inflation. However, in equilibrium, this yields a positive association between returns and expected inflation, which, again, is not consistent with existing evidence.

3.2.5 Autocorrelation in Returns

In recent years, the view that stock returns are uncorrelated has been increasingly questioned. Instead, there appears to be evidence for a mean-reverting, transitory component in stock prices, which tends to induce negative autocorrelation in returns, especially at longer horizons (see Fama & French (1988a), or Poterba & Summers (1988)). This transitory component could either result from variations in equilibrium expected returns, or could be consistent with the "fads" model, with Y_t, for example, representing long temporary swings away from the fundamental values. It is difficult to discriminate between these alternative hypotheses. In our case, since our baseline model does allow equilibrium expected returns to vary over time, if we did find that lagged excess returns help predict future returns, we might interpret that as supportive of the "fads" model.

3.3 Methodology

3.3.1 Alternative Hypothesis

Given our discussion in section 2, we shall consider the following alternative to (2)

$$\begin{aligned} r_t = &\rho\sigma_t^2 + \alpha_1(L)(D_{t-1}/P_{t-1}) + \alpha_2(L)Rf_{t-1} + \\ &\alpha_3(L)(M_{t-1}/P_{t-1}O_{t-1}) + \alpha_4(L)r_{t-1} + \varepsilon_t \end{aligned} \tag{6}$$

where the $\alpha_j(L)$ represent lag polynomials, D_{t-1}/P_{t-1} is the lagged dividend yield, Rf_{t-1} is the lagged nominal interest rate, and M_{t-1}/O_{t-1} is the ratio of (real) money supply to GDP. The EMM requires that H_0: $\alpha_j(L)=0$. Notice that we only include lagged values of variables, as the EMM could easily explain why contemporaneous variables might matter (i.e. they would be correlated with "news" as in section 1.2.2). The main econometric problem arises because σ_t^2 is unknown, so the finding that some variable helps predict excess return in (6) may be due to the fact that our proxy for risk is inappropriate. Therefore, we shall use three different ways of modelling risk.

3.3.2 Modelling the variance

A traditional way of proxying σ_t^2 is to use squared values of actual past returns (see Merton (1980)). A significant drawback of this method is that it measures the total variability of excess returns, and not the ex-ante uncertainty regarding them, and as a consequence leads to inconsistent estimates. For this reason, it has become more common to adopt a GARCH-M specification (see, e.g. Engle, Lilien & Robins (1987)) where σ_t^2 is essentially parameterised as a function of its past values, and where it is explicitly recognised that expected returns vary with σ_t^2. Specifically, equation (6) is estimated jointly with

$$\sigma_t^2 = \theta_o^2 + \sum_{j=1}^{q} \theta_j^2 \varepsilon_{t-j}^2 + \sum_{j=1}^{p} \pi_j^2 \sigma_{t-j}^2 + \sum_{j=1}^{k} \phi_j^2 z_{j,t-1}^2 \tag{7}$$

where $z_{j,t-1}$ represent variables dated time t-1 or earlier which may help predict volatility. In order to make use of maximum likelihood methods we assume that the conditional distribution of ε_t is Gaussian (cf. section 2.4).

The GARCH formulation is both simple and attractive, but the underlying theoretical model is not very informative about the appropriate specification of (7). We shall therefore use a specification test of the Newey (1985a) variety, where we test the requirement $E_{t-1}(\varepsilon_t^2) = \sigma_t^2$ by regressing the estimated value of $\varepsilon_t^2 - \sigma_t^2$ on a constant and σ_t^2, and checking the significance of the coefficient on σ_t^2, since, under the null, $E_{t-1}[(\varepsilon_t^2 - \sigma_t^2)\sigma_t^2] = 0.$, i.e. there should be no information in σ_t^2 (which is based in lagged information) which helps us predict ε_t^2 better. Since the GARCH-M model imposes some rather tight, parametric restrictions, it is as well to also experiment with alternative methods.

So for our next method, we use recent developments in semi-parametric econometrics (see Pagan & Ullah (1988) and Pagan & Hong (1991) for details). Here, we do not impose any specific functional form on $\sigma^2(I_{t-1})$ and make use of the fact that

$$\sigma_t^2 = m_{2t} - m_{1t}^2 \tag{8}$$

where $m_{it} = E_{t-1}(r_t^i)$ can be estimated using standard, non-parametric, regression techniques. We obtain two different estimates of m_{1t} and m_{2t} by using both a kernel estimator and a nearest-neighbour estimator. Assume that $z_t \in I_{t-1}$ is a vector of d variables such that $\sigma^2(I_{t-1})$ is actually known to only depend on z_t. A kernel estimate of the regression function $m_i(z)$, i=1,2 at $z=z_t$ is

$$\hat{m}_{1t} = \frac{\frac{1}{T\,\det(h)} \sum_{j=1}^{T} r_j^1 \, K[h^{-1}(z_j - z_t)]}{\frac{1}{T\,\det(h)} \sum_{j=1}^{T} K[h^{-1}(z_j - z_t)]} \tag{9}$$

where we take the "kernel" function K() to be the spherical multivariate normal density and the "bandwidth" matrix h proportional to $\mathrm{diag}(s_j)$, where s_j is the (sample) standard deviation of the jth explanatory variable (see, e.g. Robinson (1983)). We have experimented with four different values for the bandwidth[5]. We should also mention at this juncture that when estimating σ_t^2 we dropped the t-th observation from the formula for computing the nonparametric moments (see Pagan & Hong (1991), for discussion of this point).

Let $z_{t(r)}$ be the rth nearest neighbour of z_t, where we divide each explanatory variable by its sample standard deviation prior to computing the Euclidean distance and let $r_{t(r)}$ be the corresponding observations for r_t. For a given smoothing parameter k, let the weights c_r $(r=0,\ldots,T-1)$ be positive for $1 \le r \le k$, 0 otherwise, and add up to 1. Assuming for exposition that there are no tied observations, then the K-nearest neighbour (KNN) estimator of $m_1(z)$ at $z=z_t$ is defined as

$$\hat{m}_{1t} = \sum_{r=1}^{T} c_r r_{t(r)}^1 \tag{10}$$

(see, e.g. Robinson (1987)) where we have chosen uniform weights ($c_r = 1/k$ for $r=1,k$) and tried three different values for k (i.e. 7, 15 and 30)[6]. Notice that in this case the definition of (10) already "leaves out" the t-th observation.

Using either method a nonparametric estimate of $\hat{\sigma}_t^2 = \hat{m}_{2t} - \hat{m}_{1t}^2$ is obtained, as well as an estimate $\hat{\varepsilon}_t^2 = y_t - \hat{m}_{1t}$ of the residual. The natural next step is to replace σ_t^2 by $\hat{\sigma}_t^2$ in (2) and (6) and estimate by least squares, although this induces a "generated regressor" bias (see Pagan (1984)). However, we may avoid this problem by using an IV estimator, where $\hat{\sigma}_t^2$ is used as an instrument for $\hat{\varepsilon}_t^2$ (cf. Pagan & Ullah (1988)), though White's (1982) correction for the standard errors also needs to be used. Although the IV estimator will be consistent, it is likely to be fairly inefficient, especially if $\hat{\sigma}_t^2$ is not highly correlated with $\hat{\varepsilon}_t^2$. Therefore, we shall report both OLS and IV estimates.

Although the semi-parametric estimators are, conceptually, more general than a GARCH-M model, we should emphasize that, in practice, our sample size constrains us to use a relatively small number of conditioning variables. So, if the true model generating σ_t^2 were indeed GARCH, this would imply a dependence of σ_t^2 on an infinite number of lagged values of the risk premium and our nonparametric estimator would then be mis-specified (see Pagan & Hong (1991)).

3.4 Empirical Results

We use monthly data for Japan from January 1969 onwards, providing us with a sample of 236 observations. For the share price we used the end-of-month values of the Tokyo New Stock Exchange share price index, the safe rate is provided by the call money rate, while we used a broad money measure (M1+Quasi-money) scaled by GNP as our measure of liquidity. (Further details are in the Data Appendix 3)[7].

3.4.1 OLS Estimates using the Proxy Method

We first estimated equation (2) by using the lagged squared excess return as a proxy for σ_t^2. Our results are presented in column 1 of table 3.2. Notice that although $\hat{\rho}$ is positive, it is not statistically significant. We next include the level of the lagged dividend yield. It attracted a coefficient of -0.499 (with a t-ratio of -0.11) - so unlike the Anglo-Saxon countries, there is no significant positive association between the dividend yield and stock returns in Japan.[8]

We then went on to estimate (6), where we initially entered 3 lags of each variable, and, after simplification, obtained the equation reported in column 2 of table 3.2[9]. Given evidence of a January effect in US stock returns, we also included seasonals in our regression[10]. Notice that σ_t^2 has a coefficient with the "wrong" sign, although it is statistically insignificant. In addition, notice that the Efficient Market Model is easily rejected in that we have identified 4 economic variables which are statistically significant. Further, there seems to be substantial overall predictability of stock market returns, the R^2 of 0.16 for <u>monthly</u> returns is rather higher that what is typically obtained for the UK or US.

Our results suggest that <u>changes</u> in the dividend yield are <u>negatively</u> correlated with future excess returns - this is quite different from the more standard positive relationship between dividend yields and excess returns that is observed for the Anglo-Saxon countries. There is a more familiar negative relationship between changes in the nominal interest rate and excess returns. Further, those who believe that the Japanese market is partly driven by the "weight of money" will be pleased by the positive coefficient on the change in our proxy for liquidity. Finally, there appears to be positive autocorrelation in returns using Japanese monthly data. It is of some interest to note that the seasonal pattern of returns in Japan seems quite different from that in, say, the US. On including 12 seasonal dummies we found that the following seasonal coefficients had $|t|>1$[11]:

February	0.027 (2.27)	October	-0.018 (-1.72)
March	0.025 (2.48)	November	0.011 (1.11)
July	-0.013 (-1.28)	December	0.022 (2.15)

We should emphasis that the relationship that we have estimated appears to be reasonably stable. If we split the sample at October 1979, then a standard Chow test suggests that we cannot reject the null of parameter constancy. The change in the coefficients over time is displayed in Figures 3.1a though 3.1d. Notice that all the coefficients essentially stabilise after around 1976 (the fluctuations before that can probably be attributed to the small size of the sample). In addition, the ratio of the out-of-sample mean square error during 1979:10 - 1988:10 to the in-sample standard error is 0.74, so, the equation's performance during the eighties is actually <u>better</u> than what would have been predicted by its fit during the seventies. This is in contrast to some of the results in Campbell & Hamao (1989), whose particular set of predictor variables did worse in the 1980-s. We believe the result that the equation has constant coefficients to be fairly impressive for there are sound a priori grounds for anticipating parameter instability. It was not until 1978 that the authorities completely lifted restrictions on the short-term interest rate. Further, it was only in 1980 that restrictions on foreign capital flows were eased with the revision of Foreign Exchange Law.

In addition, note that the F-test for the inclusion of the extra explanatory variables is highly significant. This, combined with the stability of the coefficients, makes it highly unlikely that the significance of our regressors derives from spurious "data-mining". Further, a test of predictive failure for the period November 1987 - October 1988 also suggests that the model performs adequately[12].

These results cast serious doubt on the particular version of the EMM that we have been considering. Therefore, it is of some interest to consider whether these variables are significant because we have mis-measured risk.

3.4.2 GARCH-M Estimates

We initially jointly estimated equations (2) and (7), where we allowed (the squares of) our first three predictor variables from Table 3.2 to directly affect σ_t^2. We started with a GARCH(3,3) specification, which was then simplified to an ARCH(3) model (the relevant LR test yielded 2.8, $\chi^2_{3,0.05}$=7.81). Our results are reported in column 1 of table 3.3. Notice that while $\hat{\rho}$ is positive, it is still statistically insignificant. There is, though, some support for the view that an increase in the dividend yield increases risk (from the equation for σ_t^2). We next estimated equations (6) and (7) jointly -i.e., we included the predictor variables in the equation for the mean as well. Again, we initially started with a GARCH(3,3) specification but the simplified it to a GARCH(2,1) model (LR=2.94, $\chi^2_{3,0.05}$=7.81). Our results are reported in Table 3.3, column 2. Crucially the coefficients on the four predictor variables are largely unchanged, as is the fact that they are statistically significant. Further, the coefficient of σ_t^2 is negative and statistically insignificant. So, although there is some weak evidence that higher dividend yields increase risk, this does not affect the statistically significant association between lagged dividend yields and future excess returns.

Notice that there is some persistence in volatility (the sum of the coefficients is about 0.64) but the evidence does not point to any significant link with stock returns. However, there is significant evidence that our volatility series is mis-specified - the Newey test indicates that σ_t^2 can help predict $\varepsilon_t^2-\sigma_t^2$. This suggests that we may want to look for better measures of σ_t^2.

3.4.3 Semi-Parametric Estimates

Given our results above, we now calculated m_{1t} and m_{2t} as a function of the four variables that help predict stock returns. It is possible that a better measure of σ_t^2 would yield a positive relationship between excess returns and volatility, and that the predictor variables might lose their significance. However, our estimates of equation (6) are not noticeably different from our previous results. The coefficient on risk is <u>never</u> statistically significant, although there are some instances when it attracts a positive sign. Our four predictor variables continue to help us forecast excess returns despite the fact that we allow them to affect σ_t^2, hence pointing to a rejection of the EMM.

Notice that our results (see columns 1-4 of table 3.4) do not appear to depend on the method of estimation that happens to be adopted. There is little to chose between the IV and the OLS estimates. Further, the Nearest Neighbour and kernel-based estimates are also very similar. Also, our results were not sensitive to varying the bandwidth.

We may learn about the link between these variables and σ_t^2 by inspecting the plots of $\hat{\sigma}_t$ against each of these (kernel estimate c=4), which are presented in turn in Figures 3.2a-3.2d. In figure 3.2a a larger change in the dividend yield does appear to imply a higher σ_t^2. There is no indication of any systematic relationship between a change in the interest rate and σ_t^2 in Figure 3.2b. There is a suggestion of a positive association between a change in liquidity and σ_t^2 in figure 3.2c. Finally there is some evidence for the notion that a large negative excess return raises σ_t^2 more than a large positive excess return (Figure 3.2d), a fact that has been noticed before (e.g.

Black (1976) and chapter 5).

We also experimented with other measures of risk. First, varying the functional form of the return-risk relationship made little difference. Second including the conditional covariance between the Japanese index and the world did not significantly affect our results (further details in Sentana & Wadhwani (1989)). In deference to the Consumption CAPM, we included the expected change in consumption. Even though it attracted a positive coefficient, our predictor variables continued to significantly help predict returns.

3.5 Further Explorations

3.5.1 Extending the Sample

Sentana and Wadhwani (1989) estimated the model using data up to 1988:10 with the last 12 observations for forecasting purposes (as is true for the preceding sections). However, the Japanese market has fallen steeply during 1990, and it was of some interest to investigate how well the model has fared since. We, therefore, performed a predictive failure test until 1990:4. The value of the relevant test statistic is 40.14 ($\chi^2_{30,0.05}$=43.77), and, so, we were unable to reject the null of parameter constancy. It is rather impressive that the equation has continued to perform well in these turbulent times on the Japanese stock exchange.

Next we turn to the issue of why we began estimation in 1969. We also considered estimating the model over the sixties. However, it is worth recalling that there was virtually no free short-term interest rate before 1970, with the Gensaki market only beginning in 1969 (see e.g. Feldman (1986) or Suzuki (1987)). Also, note that the first issue of government bonds only occurred in 1966, and that there was no secondary market for them until 1977. Suzuki argues that, prior to the liberalisation of the market, "... short rates seemed not to reflect (the) demand and supply for funds... " (p.155). An indication of the change in the interest rate regime after around 1970 is provided by the fact that during 1960-1969 the interest rate change only in 45 out of a possible 120 months, while during

1970-1988 it changed in 210 out of 228 months. Given this fact we would be surprised if our estimated relationship between stock returns and the short interest rate were invariant to deregulation.

When we did re-estimate our model to include the sixties, we found that over 1960:6-1988:10:

$$
\begin{aligned}
r_t = &\underset{(-2.14)}{-64.07}\ \Delta \text{Div.Yield}_{t-1} + \underset{(2.61)}{76.14}\ \Delta \text{Liquidity}_{t-1} + \underset{(1.98)}{0.25} r_{t-1} + \\
&\underset{(-2.67)}{-23.00}\ \Delta Rf_{t-2} + \underset{(2.21)}{32.70}\ \Delta Rf_{t-2} {*} D60 - \underset{(-0.15)}{0.10}\ \hat{\sigma}^2_t + \text{seasonals}
\end{aligned}
$$

where D60 takes the value 1 during the sixties.

Notice that, with the exception of the interest rate, all the other variables continue to have a similar effect on stock returns. We are fairly encouraged by these results.

3.5.2 Evidence from South Korea

We have put much emphasis on the fact that the lagged dividend yield is not positively correlated with stock returns in Japan. However, even if it were true that dividend yields were truly positively correlated with returns it is possible to find a country with the opposite relationship. For this reason, we offer some evidence from another country in the same geographical region - South Korea. This is another market which has seen spectacular stock price gains, e.g. the index rose by a factor of 10 in the eighties.

We initially regressed excess returns (detailed data definitions and sources are in the appendix) on the lagged dividend yield, seasonals and a measure of volatility over the period 1975:6 - 1987:9. A selection of results are to be found in table 3.5. Notice that the level of the dividend yield attracts a statistically significant negative coefficient, as does $\hat{\sigma}^2_t$. Hence as in Japan, there is no evidence of a positive association between returns and the lagged dividend yield. However, as in Japan, unit root tests suggest that while we cannot reject the hypothesis that the dividend yield is I(1), returns seem to be I(0). Therefore, one should be appropriately cautious about these regressions.

We next examined the relationship between returns and the change in the dividend yield. As in Japan, this attracted a negative coefficient, as did $\hat{\sigma}_t^2$ (cf. columns 4-6, Table 3.5).

We view our results for South Korea as being broadly supportive of our work for Japan, in that, once again, we were unable to find any evidence for a positive association between returns and the lagged dividend yield in either country.

3.5.3 Time-Variation in the "Price of Risk"

In the basic model that we estimate (i.e. $\mu_t = \rho\sigma_t^2$), ρ is a measure of aggregate risk aversion which depends on the distribution of wealth and underlying preferences over risk. In practice, changes in wealth and its distribution, social provision for the poor, changes in the "spirit of the times", etc., can all lead to variations in ρ. The failure to allow ρ to vary over time implies that our estimation procedure may be incorrect. Equally, if ρ were allowed to vary over time in an entirely unrestricted fashion in (1), then there is no longer anything to test!. Therefore, the strategy that we followed was to compute an implicit price of risk series $\rho_t=\sigma_t^2/\mu_t$, where μ_t,σ_t^2 were estimated by using a kernel estimator with c=2. The results are presented in Figure 3.3. (Similar results were obtained with nearest neighbour estimates and different bandwidths). THey imply wide variations in ρ_t. For example, it rises from a value of about -5 to one over +25 and back to 0 all within one year!. In another two-year period (1980-1982), it fall from 0 to about -12, then rises to about +17, only to fall back to about zero. It is theoretically possible for our results to be consistent with a version of the EMM where the price of risk is allowed to vary (as in section 1.2.1.2). However, we would still need to explain why the price of risk varies as much as it does in figure 3.3.

3.6 Implications and Conclusions

Our substantive conclusions are:

i) Higher volatility does not appear to lead to significantly higher excess returns in Japan or South Korea. This finding is robust across three different ways of measuring risk (i.e. lagged squared excess returns, GARCH and semi-parametric methods)

ii) There is no significant positive association between returns and the dividend yield in either Japan or South Korea. If anything, there is a negative relationship in S.Korea. This would be difficult to explain in terms of risk considerations, which normally predicts a positive association (as normally found in Anglo-Saxon countries). This result is also at odds with standard mean reversion stories. Further, the lagged change in the dividend yield has a negative relationship with current excess returns.

iii) Lagged changes in the nominal interest rate also help predict excess returns (they are negatively related). This is difficult to explain. As already discussed a standard "noise" traders model cannot account for it.

iv) There is some support for the "weight of money" explanation of share price rises, in that lagged increases in velocity do lead to a lower excess return.

v) Excess returns are positively autocorrelated on a monthly basis

vi) The above results are robust to the use of different statistical measures of volatility, and to whether or not we use consumption risk, or market risk.

Our results suggest that you can predict Japanese monthly excess returns using lagged information even after controlling for risk, with an explanatory power of about 16%. The relationship between these predictor variables appears to be stable over our sample period.

It is possible that we have not modelled risk appropriately. It may well be possible to write down some rather general model of risk where the price of risk varies substantially (as in chapter 1) although we would still need to link this variation to our predictor variables. Our results suggest that it is these models of risk that we should be using - the standard models that we have experimented with above do not appear to fare particularly well.

ENDNOTES

1 Attanasio & Wadhwani (1990) have shown that in the U.S., the lagged dividend yield only helps predict returns because it also helps to predict future volatility.

2 It would, of course, only constitute evidence against our rather particular version of the EMM. Variants of the EMM with, say, time-varying price of risk could be consistent with our findings.

3 Despite recent falls in the market, the dividend yield had only recovered to 0.81% in October 1990.

4 A prominent Japanese securities house placed a full page advertisement in the Financial Times on January 30, 1989 describing P-E ratios as an outdated criterion.

5 The baseline bandwidth value was $O(T^{-1/(d+4)})$. We then halved it, doubled it and quadrupled it. We are grateful to Peter Robinson for helpful advice on this point.

6 We are also grateful to Peter Robinson for his helpful advice.

7 We are extremely grateful to Orazio Attanasio and Miguel Delgado for providing us with the Fortran code to estimate GARCH-M models and non-parametric kernel regressions respectively. The non-parametric nearest neighbour regression program is available on request. All these were run on the LSE VAX. The rest of the computations were performed using the regression packages DFIT and TSP.

8 We need to be appropriately cautious when interpreting this result, because we found that while standard tests did not reject the view that excess returns are I(0), we could not reject the hypothesis that the dividend yield is I(1).

9 We did perform unit root tests on all our predictor variables. In all cases we could not reject the null that these variables are I(0).

10 Note, though, that since we do not measure monthly dividends accurately, the seasonals may be proxying for a seasonal component in the measurement error.

[11] The value of the test for the significance of these dummies is 17.64 ($\chi^2_{6,0.05}$=12.59).

[12] While our variables are dated t-1 and earlier, it is possible that, because of announcement lags, our measure of liquidity is not actually known to market participants. Therefore, we re-estimated our equation, where lagged liquidity was now instrumented using earlier values. However the relevant coefficients were unchanged (the relevant Hausman (1978) tests yielded t=0.47).

DATA APPENDIX 3

Details of the data series used are as follows (where appropriate, the name of the series is followed by its Datastream code).

Japan:

Interest rate:	JPOCCAL (Call Money Interest Rate).
Share Prices:	JPTOKYO (Tokyo New Stock Exchange Share Price Index, End of Month Value, 1968=100).
Consumer Prices:	JPOCCPTKF (Consumer Prices, Tokyo: All Items, 1980=100).
Money:	JPOCMIQSA (Money Supply M1 plus Quasi-Money, Yen Billion, Current Prices).
Income:	JPOCGNPDB (GNP at annual rates, Yen Billion, Current Prices, Linearly Interpolated).
Dividend Yield:	Average Yield (End of Month, Dividend Paying Companies only, Economic Statistics Annual, Research & Statistics Department, Bank of Japan).
Consumption:	JPCONEXPP (Personal Consumption Expenditure, 1980 Prices).
World Stock Price Index and Dividend:	Morgan Stanley Capital International Index.
Exchange Rate:	JPOCEXCH ($-Yen Exchange Rate, End of Period).

South Korea:

Interest Rate:	KOI60 (Discount Rate).
Share Price Index:	KORCOMP (End of Month).
Dividends:	From "Stock" & "Securities Statistics Yearbook", Korean Stock Exchange.

Table 3.1

Average Dividend Yield and Returns
1970-89

Time Period	Average Dividend Yield	Five-Year Excess Return
1970-1974	2.30%	29%
1975-1979	1.53%	41%
1980-1984	1.22%	68%
1985-1989	0.57%	200%

Table 3.2

Excess Returns Equations - Proxy Estimates

1969:3-1987:10

Independent Variables	(1)	(2)
$\hat{\sigma}^2_t$	0.65	-0.33
	(0.88)	(-0.38)
ΔDividend Yield$_{t-1}$		-119.44
		(-2.26)
ΔInterest Rate$_{t-2}$		-25.84
		(-3.26)
ΔLiquidity$_{t-1}$		136.38
		(2.71)
Excess Return$_{t-1}$		0.46
		(2.29)
R^2	0.0026	0.160
Serial Correlation Test	7.91	8.75
	($\chi^2_{12,0.05}$=21.02)	($\chi^2_{12,0.05}$=21.02)
Predictive Failure Test	13.08	15.05
(forecast to 88:10)	($\chi^2_{12,0.05}$=21.02)	($\chi^2_{12,0.05}$=21.02)
ARCH test	12.94	15.93
	($\chi^2_{12,0.05}$=21.02)	($\chi^2_{12,0.05}$=21.02)
Parameter Stability Test	0.33	0.99
(split at 1979:9)	($F_{2,220,0.05}$=3.03)	($F_{17,190,0.05}$=1.68)
F-test for significance		4.74
of economic predictor variables		($F_{4,207,0.05}$=2.41)
s.e.	0.04427	0.04195

Notes: (i) (1) includes a constant, (2) seasonals

(ii) heteroskedastic-consistent t-ratios in parentheses (as in Eicker (1963) and White (1980))

Table 3.3

Excess Returns Equations - GARCH Estimates
1969:3-1988:10

Independent Variables	(1)	(2)
$\hat{\sigma}_t^2$	0.90 (0.53)	-0.61 (-0.25)
ΔDividend Yield_{t-1}		-119.19 (-2.46)
ΔInterest Rate_{t-2}		-15.71 (-2.53)
$\Delta\text{Liquidity}_{t-1}$		122.97 (3.12)
Excess Return_{t-1}		0.41 (2.52)
log L	637.29	656.55
Variables affecting σ_t^2		
ε_{t-1}^2	0.0	0.0
ε_{t-2}^2	0.39 (2.95)	0.35 (3.51)
ε_{t-3}^2	0.15 (1.91)	
σ_{t-1}^2		0.28 (2.96)
ΔDividend Yield_{t-1}	2.1×10^{-4} (1.64)	1.75×10^{-4} (0.95)
ΔInterest Rate_{t-2}	0.0	0.0
$\Delta\text{Liquidity}_{t-1}$	0.0	0.0
Newey Test	-3.73	-2.40
s.e.	0.04438	0.04135

Notes: (1) includes a constant in the mean and seasonals in the variance, whereas (2) additionally includes seasonals in the mean.

Table 3.4

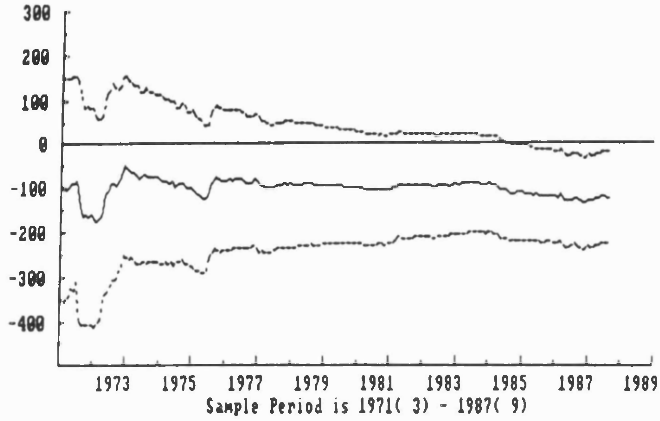
Excess Returns Equations - Semi-Parametric Estimates
1969:3-1987:10

Independent Variables	(1) OLS (Kernel)	(2) IV (Kernel)	(3) OLS (NN)	(4) IV (NN)
$\hat{\sigma}^2_t$	-8.20	3.63	3.99	-3.34
	(-0.85)	(0.80)	(0.76)	(-0.74)
ΔDividend Yield$_{t-1}$	-120.79	-141.39	-130.13	-113.27
	(-2.30)	(-2.35)	(-2.40)	(-2.21)
ΔInterest Rate$_{t-2}$	-25.11	-24.49	-25.59	-26.44
	(-3.24)	(-3.00)	(-3.23)	(-3.10)
ΔLiquidity$_{t-1}$	130.74	132.29	133.14	133.30
	(2.82)	(2.61)	(2.85)	(2.97)
Excess Return$_{t-1}$	0.42	0.42	0.44	0.46
	(2.37)	(2.26)	(2.43)	(2.47)
R^2	0.16		0.16	
Serial Correlation ($\chi^2_{12,0.05}$=21.02)	9.29		7.42	
Predictive Failure ($\chi^2_{12,0.05}$=21.02)	14.57		15.56	
Parameter Stability ($F_{17,190,0.05}$=1.68)	0.97		0.93	
F-test for significance of predictor variables ($F_{4,207,0.05}$=2.41)	4.59		5.01	
s.e.	0.04402	0.04571	0.04201	0.04173

Table 3.5

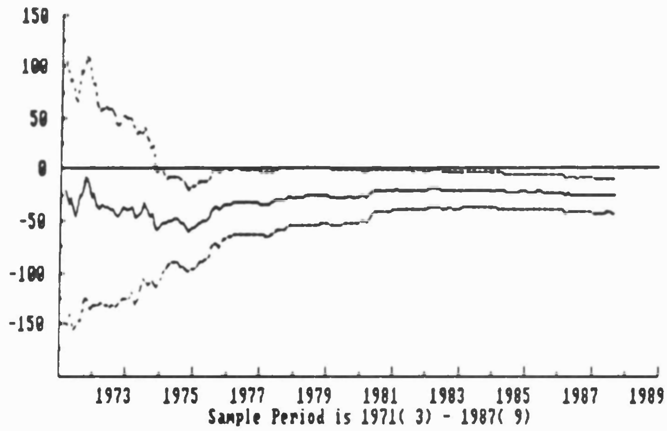
Excess Returns Equations - South Korea
1975:6-1987:9

Independent Variables	(1) OLS (Proxy)	(2) OLS (Ker)	(3) IV (Ker)	(4) OLS (Proxy)	(5) OLS (Ker)	(6) IV (Ker)
$\hat{\sigma}_t^2$	-1.28 (-1.89)	-1.08 (-1.19)	-3.19 (-0.97)	-1.21 (-1.76)	4.75 (1.14)	3.33 (1.25)
Dividend Yield_{t-1}	-2.39 (-2.51)	-2.15 (-2.28)	-2.03 (-1.76)			
Dividend Yield_{t-1}				-8.06 (-1.53)	-6.15 (-1.17)	-6.07 (-1.27)



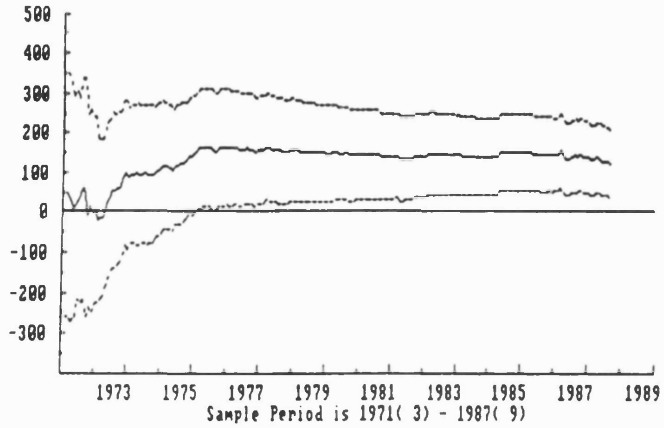
Coefficient on Dividend Yield ± 2 + S.E.

Figure 3.1a



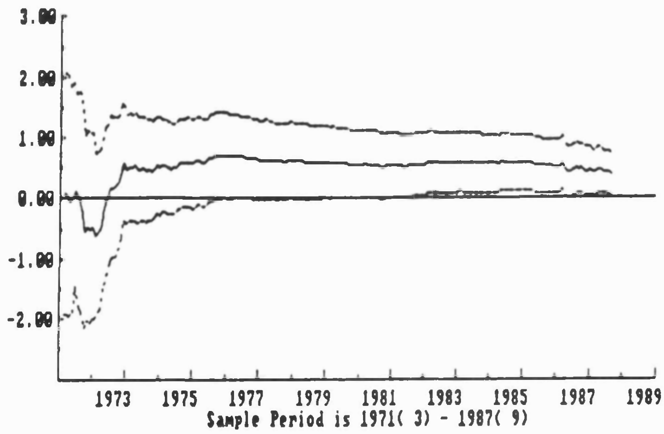
Coefficient on Interest Rate ± 2 + S.E.

Figure 3.1b



Coefficient on Liquidity ± 2 * S.E.

Figure 3.1c



Coefficient on Lagged Risk Premium ± 2 * S.E.

Figure 3.1d

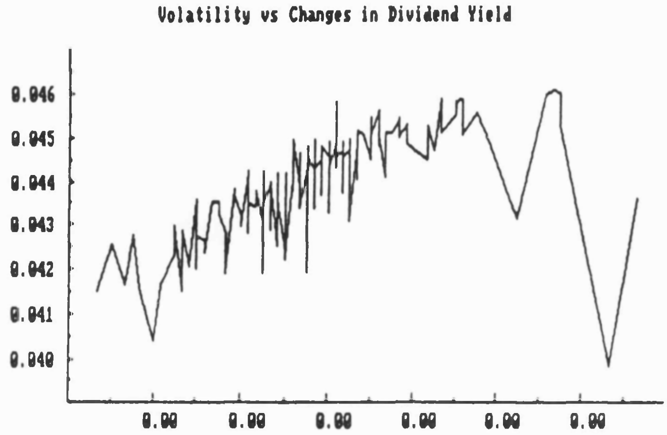


Figure 3.2a

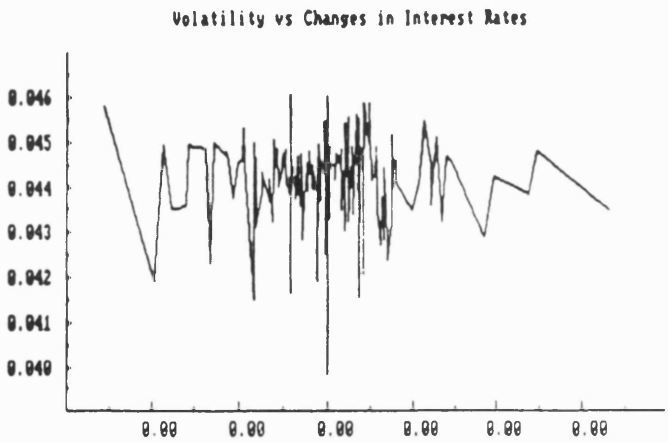


Figure 3.2b

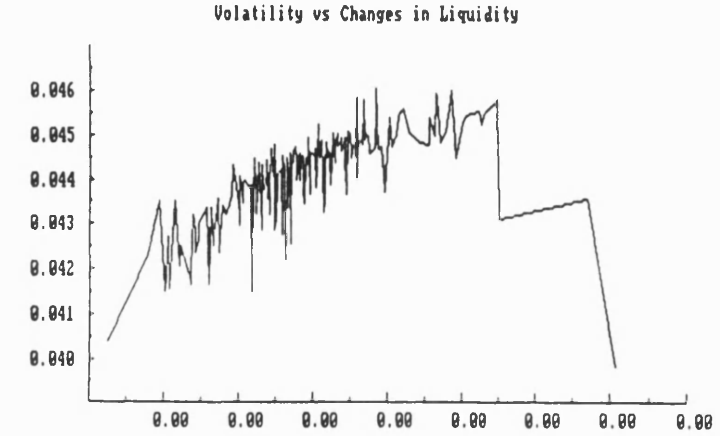


Figure 3.2c

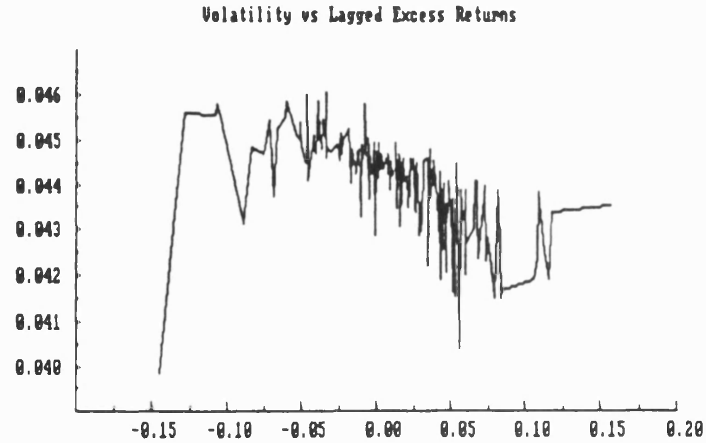


Figure 3.2d

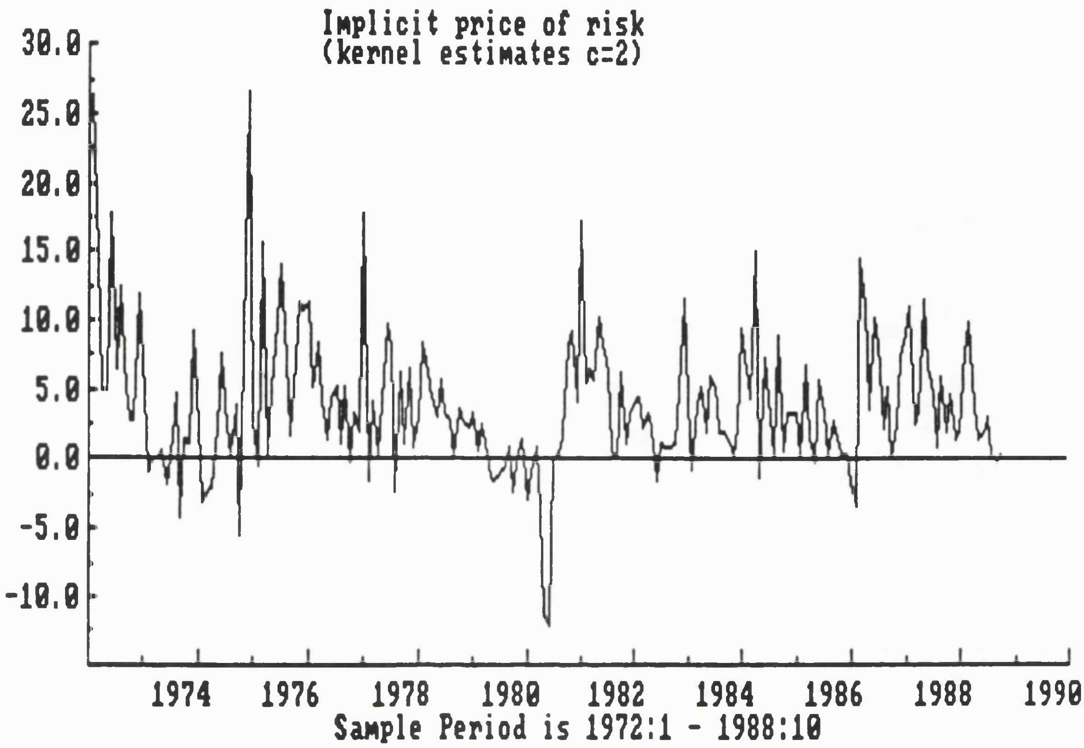

Implicit price of risk
(kernel estimates c=2)
30.0
25.0
20.0
15.0
10.0
5.0
0.0
-5.0
-10.0
1974
1976
1978
1980
1982
1984
1986
1988
1990
Sample Period is 1972:1 - 1988:10

Figure 3.3

Chapter 4

FEEDBACK TRADERS AND STOCK RETURNS AUTOCORRELATIONS: EVIDENCE FROM A CENTURY OF DAILY US DATA

4.1 Introduction

In recent years, increased attention has been devoted to models of share price determination that rely on the existence of heterogeneous investors - e.g. Shiller's (1984) "fads" model, or the "noise trader" model (see e.g. De Long, Shleifer, Summers & Waldman (1990)). There has also been renewed interest in the serial correlation properties of stock returns (see e.g. section 3.2.5). For example, Cutler, Poterba & Summers (1990) have argued that the characteristic serial correlation patterns in a variety of assets can be accounted for by models with feedback traders. We extend the logic of their analysis to look at the links between volatility and serial correlation.

If one set of traders follow a feedback trading strategy (i.e. they react to price changes), then returns will exhibit serial correlation. Further, as expected volatility rises, "smart" money will allow the feedback traders to have a greater effect on the price, and, therefore, the extent of serial correlation will rise. In this context, a greater degree of predictability of stock returns from higher (in absolute value) serial correlation is still compatible with equilibrium because high volatility makes it more risky for "smart" money to take advantage of the predictable patterns in stock returns. It is this link between volatility and return autocorrelations that we seek to test in this chapter.

Serial correlation in returns may also arise from a variety of reasons unconnected to the feedback traders model. One distinct possibility that we shall consider in interpreting the results, is that the serial correlation may arise from non-synchronous trading. It is also important to emphasize that if preferences exhibit a degree of risk aversion that declines with wealth, then a positive feedback trading strategy may be entirely rational (see e.g. Black (1988)). Such considerations lead to the additional testable implication that the extent of positive feedback trading rises with volatility.

We present evidence on the links between volatility and returns autocorrelations by using both hourly data around the period of the October 1987 crash and daily data for 1885-1988. We use alternative measures of volatility based on a standard GARCH specification, an exponential GARCH model (EGARCH), and, also, on non-parametric methods.

The rest of the chapter is organised as follows. The feedback traders model is summarized in section 4.2.1 and the consequences of non-synchronous trading are discussed in section 4.2.2. The empirical evidence is presented in section 4.3. We then explore some extensions to our basic framework in section 4.4 and finally some conclusions are to be found in section 4.5.

4.2 Theoretical arguments for serially correlated returns

4.2.1 Feedback traders and serial correlation

As in section 3.2.1, we consider here a simple model where agents follow different trading strategies[1]. The first group (smart money) are assumed to have a demand function for shares of the form:

$$Q_t = \frac{E_{t-1}(r_t^m) - \alpha}{\mu_t} \tag{1}$$

where Q_t is the fraction of shares that they hold, r_t^m is the ex-post return on the market in period t, E_{t-1} denotes the expectation operator using information available as of time t-1, α is the return at which the demand for shares by this group is zero and μ_t is the risk premium needed to induce them to hold all the shares. We shall assume again that:

$$\mu_t = \mu(\sigma_t^2) \tag{2}$$

with $\mu'()>0$, where σ_t^2 denotes the conditional variance of returns in period t (formed at time t-1), so that these investors are risk averse (cf. section 3.2.1). Hence a rise in expected volatility increases the risk premium needed to induce smart money to hold all the shares. Notice that if all investors had demand functions of the form (1), then market equilibrium (Q_t=1) would yield the familiar Capital Asset Pricing model:

$$E_{t-1}(r_t^m)-\alpha=\mu(\sigma_t^2) \tag{3}$$

with α set to the risk-free rate (see, e.g. Merton (1980) and section 3.2.1).

We shall assume initially that the second group of traders buy after price increases, i.e. positive feedback trading, so that their demand function is given by:

$$Y_t = \gamma r_{t-1}^m \tag{4}$$

with $\gamma>0$. Such behaviour is consistent with that of portfolio insurers and those who use stop-loss orders. It can also occur as a result of "distress" selling after significant market declines. In this respect, Shiller (1987) found that the single most important reason that prompted investors to sell shares in October 19th, 1987, was the fact that prices had fallen.

Market equilibrium requires that:

$$Q_t + Y_t = 1 \tag{5}$$

which yields:

$$E_{t-1}(r_t^m)-\alpha=\mu(\sigma_t^2)- \gamma\mu(\sigma_t^2)r_{t-1}^m \tag{6}$$

Hence, on comparing (6) with the standard CAPM model in equation (3), we have an additional term $(-\gamma\mu(\sigma_t^2)r_{t-1}^m)$ so that returns will exhibit negative serial correlation. Importantly, the extent to which the returns are serially correlated varies with volatility. Intuitively, as expected volatility rises, smart money needs a higher expected return, and this allows a larger deviation of the current price from its fundamental value, which, then, leads returns to exhibit stronger serial correlation. The above model suggest that stock price anomalies are larger when volatility is high.

So far, we have only allowed ourselves the possibility of positive feedback traders ($\gamma>0$). However, it is possible that some individuals adopt a negative feedback strategy ($\gamma<0$) - i.e. buying after price declines. This is consistent with so-called "buy low, sell high" strategies, and also with those that assign a constant share of wealth to a particular asset. From our previous discussion it is clear that this would induce positive autocorrelation in stock returns.

Combining both kinds of feedback investors it is indeed possible that (the cross-sectional average) γ varies over time with changes in volatility. In this regard, it is important to recognize that a portfolio insurance strategy can be entirely rational if preferences exhibit risk aversion that declines rapidly with wealth (see e.g. Black (1988, 1989) or Marcus (1989)). In such settings, an exogenous reduction in the values of the shares (the risky asset) can lead to an even larger reduction in the demand for the risky asset. These models also have the property that a given reduction in wealth induces <u>more</u> portfolio insurance selling the <u>lower</u> the initial level of wealth is[2]. Hence, in equation (4), γ depends on current wealth, or more correctly, current wealth, W, relative to current "subsistence" wealth, W_{min}. Note that a ceteris paribus rise in volatility, σ_t^2, lowers wealth and thereby raises γ. Therefore we shall set $\gamma=\gamma(\sigma_t^2)$ in equation (6), with the expectation that $\gamma'()>0$.

Allowing γ to depend on volatility enables us to generate a richer set of possible implications for the pattern of serial correlation from equation (6). For example, it is possible that, at low levels of volatility, negative feedback trading predominates (i.e. $\gamma<0$) and returns exhibit positive serial correlation. However, as volatility rises, this might increase the demand for portfolio insurance-type strategies (i.e. $\gamma>0$), which then leads to returns exhibiting negative serial correlation.

4.2.2 Non-synchronous trading and serial correlation

It is well-known that non-synchronous price quotes (i.e. the non-trading problem) induces serial correlation in returns (see e.g. Fisher (1966), Cohen et al (1980), Lo and Mackinlay (1990)). If the returns to two stocks A and B are independent, but B trades less frequently than A, then the price of A will respond more quickly when news affecting both stocks arrives. As a consequence, the return on B will appear to respond with a lag to the return on A, i.e. there will be positive cross-autocorrelation. If we now consider an index made up of a large number of securities (as is true of the indices used in our empirical work below) the positive serial cross-correlations will manifest themselves as positive autocorrelation in the index[3]. Lo and Mackinlay (1990) show that such positive autocorrelation in an index can be represented by an AR(1) specification for index returns. In the case of a portfolio of securities with the same non-trading probability, the AR(1) coefficient actually equals the non-trading probability.

In the light of our discussion of the model with feedback traders, it is important to establish whether and how we would expect any serial correlation that arises from non-synchronous trading to vary with the volatility of the market. If high trading volume on an index is an indicator of more trading, then the observed positive time-series correlations between volume and volatility (see Tauchen and Pitts (1983) or Schwert (1989b)) would suggest that periods of high volatility are also periods when the non-trading effect is small. On the other hand, one may recall that a period of exceptional volatility like the October 1987 crash was characterised by considerable non-trading of securities. Nevertheless, it is possible that the effect of higher volatility could be to reduce the index autocorrelation that is induced by non-trading. In this setting, the index autocorrelation could in principle fall from a positive number to zero, as the non-trading probability shrunk to zero, whereas in the model with positive feedback traders, the already negative autocorrelation rises in absolute value. Note that models of non-synchronous trading do not usually predict negative autocorrelation in index returns[4].

4.3. Empirical evidence

4.3.1 Evidence based on hourly data around the october 1987 crash

The October 1987 crash saw the largest single-day decline in New York this century. As a consequence there was an obvious rise in share price volatility. Indeed, if we compute the standard deviation of hourly returns for each week during July 1987 - February 1988, we find that the standard deviation of returns in New York was more than <u>seven</u> times as high as it had been in the period preceding the crash[5]. Therefore, the 1987 crash can be seen as providing something close to a natural experiment: it should enable us to assess the effects of volatility on the serial correlation pattern of returns without having to actually specify a measure of volatility as few would disagree with the proposition that the Crash was a period of higher volatility (see Pagan & Ullah (1988) for a discussion of some of the difficulties associated with measuring volatility).

We therefore estimated the following equation:

$$r^m_t = \alpha + (\gamma_0+\gamma_c \text{Crash}_t)r^m_{t-1} + \varepsilon_t \qquad (7)$$

where Crash_t is one during the crash week and 0 otherwise, using hourly data for both the US and the UK[6]. Our results are reported in table 4.1.

In the US we find that the coefficient of r^m_{t-1} is negative, and that it became significantly more negative during the crash week: the coefficient changed from -0.09 to -0.45. In the UK, the coefficient on r^m_{t-1} was positive, but it became negative during that week, moving from 0.12 to -0.05.

If we attempt to interpret our results in terms of the theoretical models discussed in section 4.2, they suggest that in the US, there are enough positive feedback traders to give us negative serial correlation, even though non-trading effects would tend to generate positive autocorrelation. The fact that there is even more negative serial correlation during the crash week would support the model outlined in section 4.2.1 above. Higher volatility made smart agents more cautious, and this allowed portfolio insurers and stop-loss traders to have a bigger effect on the price, which shows up as higher negative serial correlation in returns. The rise in the (absolute) value of the coefficient of r^m_{t-1} is also consistent with a rise in the extent of positive feedback trading (i.e. γ rising) or with a possible decline in the contribution from the non-trading effect.

The results for the UK suggest that for much of the period, returns exhibited positive serial correlation. This could arise from non-synchronous trading and/or negative feedback trading and the fact that at prevailing levels of volatility, the amount of positive feedback trading is not enough to offset the other effects. However, during the crash week, the coefficient turned negative, which is consistent with a decline in the non-trading effect, and with positive feedback traders having become more important.

The results for the UK are especially encouraging as the stock index used is based on mid-market prices. Hence, the negative serial correlation cannot attributed to prices bouncing between bid and ask.

We view our results as being consistent with the models of feedback traders that we outlined in section 4.2.1. However it could be argued that we should use more data, and that we should attempt to measure volatility. This is the subject of the next subsection.

4.3.2 Evidence using daily data from 1885-1988

We used the time series of daily data from 1885-1988[7]. This data set (29,137 observations) consists of Dow Jones returns between February 3, 1885 and January 3, 1928. From January 4, 1928 through July 2, 1962, we use the daily returns on the S&P composite portfolio. Finally, from July 3, 1962 through December 31, 1988, we use the CRSP value-weighted portfolio (see Schwert (1990), for further discussion of the data).

We next seek to estimate a linearised variant of (6) with $\gamma=\gamma(\sigma_t^2)$, i.e.:

$$r_t^m = \alpha + \rho\sigma_t^2 + (\gamma_0+\gamma_1\sigma_t^2)r_{t-1}^m + \varepsilon_t \tag{8}$$

There are various alternative methods of proxying for σ_t^2 (see section 3.3.2 for three such methods, and Pagan & Schwert (1990) for a comparison). We did use three different models for σ_t^2, the GARCH-M model (as in section 3.4.2), the Exponential GARCH model (see

Nelson (1991)) and a semi-parametric proxy (as in section 3.4.3). Since the results were quite similar, we only discuss the Exponential GARCH model here.

We estimated equation (8) jointly with

$$\ln \sigma_t^2 = \theta_{ot}^* + \sum_{i=1}^{p} \theta_i \ln \sigma_{t-i}^2 + g_{t-1} + \sum_{j=1}^{q} \phi_j g_{t-1-j} \tag{9}$$

$$\theta_{ot}^* = \theta_o + \pi_o N_t \tag{10}$$

$$g_t = \psi\, \xi_t + \delta\, (|\xi_t| - E|\xi_t|) \tag{11}$$

where N_t is the number of non-trading days (including holidays and weekends) between trading day t-1 and t[8].

Note that this model allows for sign (leverage) effects (cf. Black (1976), figure 3.2d and chapter 5) through $\psi\xi_t$[9].

The results obtained by estimating equations (8) and (9) are presented in table 4.2. Notice that there is strong evidence for leverage effects; ψ is negative, implying that volatility tends to rise more when a constant-modulus return surprise is negative than when it is positive. Our estimates also suggest a high degree of persistence in the variance (with the highest root being almost 1) as well as leptokurtosis of the conditional distribution (ν=1.4, significantly smaller than 2)

Crucially, we find that higher volatility is more likely to lead returns to exhibit negative serial correlation - for values of σ_t^2>5.84, returns will exhibit negative serial correlation. An illustration of the practical importance of this effect is provided in the context of the 1929 and 1987 crashes respectively. In 1929 the EGARCH measure of volatility peaked at nearly 42, implying a first-order autocorrelation coefficient of -0.71. The preceding week the implied coefficient had been 0.04. Similarly in October 87 the peak was even higher at over 56 with an implied serial correlation coefficient of -1 when the previous week it had been 0.08. This dramatically illustrates the extent to which the autocorrelation properties of stock returns can change with variations in volatility.

Hence our results about links between volatility and returns autocorrelations remain consistent with the notions discussed in section 4.2.

In the next section, we explore some extensions to our basic framework[10].

4.4 Some further explorations

4.4.1 Is there more positive feedback trading after market declines?

The evidence presented above is consistent with the notion that the extent of positive feedback trading rises with increases in volatility. It is also possible that large price <u>declines</u> lead to more positive feedback trading as compared with large price rises. This asymmetry could stem from the fact that those who trade on margin, and make large losses after price declines, often have no choice but to sell their holdings in order to meet their obligations (see section 4.4.3. below). It may also directly result from the already-mentioned possibility that if risk aversion declines rapidly with wealth, then a decline in wealth (caused by $r^m_{t-1}<0$) leads to an increase in positive feedback trading.

We explored the possibility that such an asymmetry exists by including an additional term, $\gamma_s|r^m_{t-1}|$, in the returns equation so that the coefficient on r^m_{t-1} would be:

$$\begin{rcases} \gamma_0+\gamma_s+\gamma_1\sigma^2_t \text{ if } r^m_{t-1}\geq 0 \\ \gamma_0-\gamma_s+\gamma_1\sigma^2_t \text{ if } r^m_{t-1}<0 \end{rcases} \qquad (12)$$

Ignoring for the moment the fact that σ^2_t may depend on the sign of r^m_{t-1}, equation (12) implies that for $\gamma_s>0$, market declines make it more likely that returns will exhibit negative serial correlation (as there will be more positive feedback trading). However, since in the EGARCH formulation we allowed σ^2_t to have an asymmetric response to price changes, it is possible that the effect that we seek to find here has already been captured.

On using the above formulation in the context of the EGARCH model we obtained the following parameter values (asymptotic standard errors in brackets):

$$\hat{\gamma}_0 = 0.101 \quad \hat{\gamma}_s = 0.054 \quad \hat{\gamma}_1 = -0.019$$
$$(0.004) \quad (0.004) \quad (0.001)$$

so the evidence seems to strongly suggest that there is more positive feedback trading after market declines than there is after market rises.

4.4.2 Margin requirements and the extent of positive feedback trading

It is sometimes argued that when stocks are purchased on margin, price declines cause positions to be liquidated. This selling causes further price declines, further liquidation of margin positions, and so on. This is usually called depyramiding (see e.g. Garbade (1982)).

This view predicts that a decrease in margin requirements increases the vulnerability of a market depyramiding. There has recently being renewed interest in the relation between margin requirements and stock market volatility (see e.g. Hardouvelis (1989), Hsieh & Miller (1990), Salinger (1989) and Schwert (1989b)). While Hardouvelis (1989) claimed to find that higher margin requirements depresses stock market volatility, much of the other research disagrees with this claim.

Here we look for evidence of depyramiding by examining the effects of margin requirements on the serial correlation pattern of returns. Since in our model positive feedback trading is associated with negative serial correlation in returns, depyramiding should increase the absolute value of this serial correlation.

We interacted r^m_{t-1} with (one minus) margin requirements as a proxy for margin credit, so that the additional component of the coefficient of r^m_{t-1} is β_1(1-margin$_t$). If higher margin requirements did reduce the extent of positive feedback trading, then, we would expect β_1<0

We re-estimated our model from 1934 onwards (when the Federal Reserve Board first set compulsory margin requirements) and obtained:

$$\hat{\beta}_1 = \underset{(0.0019)}{0.0020}$$

which is statistically insignificant, and in fact, assumes the opposite sign to what was predicted by the above analysis. There is no evidence hence for the view that variations in margin requirements significantly affect the serial correlation of returns.

One does, though, need to be cautious. For much of this period, margin credit has been a rather small proportion of the value of New York Stock Exchange stocks, e.g. between 1945-85, margin credit never exceeded 1 1/2% of the market value. It is perhaps necessary to look at periods when margin credit was more important (e.g. the 1917-1930 period when it was well over 10%) in order to examine the effects of variation in margin requirements. Unfortunately, we have little time series information on the average of broker-imposed margin requirements.

4.5 Conclusions

The empirical work reported here suggests that when volatility is low, stock returns at short horizons exhibit <u>positive</u> serial correlation, but when volatility is rather high, returns exhibit <u>negative</u> autocorrelation. This time-varying nature of the serial correlation pattern appears to be robust across different periods and different measures of volatility.

These results are consistent with a model where some traders follow feedback strategies, and where non-synchronous trading also contributes to the serial correlation of returns. As volatility increases, the positive feedback traders have a greater influence on the price (not only because smart traders are more cautious but also because there is then more positive than negative feedback trading), which then manifests itself in greater negative serial correlation in returns. The increase in the extent of positive feedback trading could also be consistent with rational behaviour if some investors have preferences with risk aversion that declines with wealth (see e.g. Black (1988) or Marcus (1989)). This, combined with a possibly diminishing contribution from non-synchronous trading can go some way towards explaining the switch in sign in the serial correlation of returns.

Although we know of no other empirical work that documents a volatility-induced switch in sign in the serial correlation of returns, there is other work on the foreign exchange market that is broadly consistent with ours. Ito & Roley (1986) and Goodhart and Giugale (1988) have demonstrated, using hourly data, that log price changes exhibit negative serial correlation, with the degree of serial correlation rising substantially after large "jumps" in exchange rates (which should correspond to periods of high volatility).

We also find some evidence which suggests that the extent of positive feedback trading is greater following price declines than it is after price rises. This asymmetry is consistent with both the possibility that risk aversion declines rapidly with wealth and with the existence of significant distress selling after price declines. Note, though, that we were unable to detect any effect of initial margin requirements on the serial correlation of returns.

An explanation of the time-varying pattern of serial correlation in returns that is based on feedback traders is of course not the only way in which one may interpret our results. It is likely that alternative explanations of the serial correlation in returns that rely on changes in the price of risk (see e.g. Black (1989), Marcus (1989)) could be modified in order to accommodate our empirical findings. Further, explanations based on market micro-structure might work[11], though note that neither non-synchronous trading nor the existence of bid-ask spreads can explain our findings.

ENDNOTES

1 The model here is a special case of Shiller's (1984) fads model - for recent examples of the use of this model, see, e.g. Cutler, Poterba & Summers (1990).

2 Note that the size of the current level of wealth, W, is to be assessed relative to some threshold value, W_{min}, which gradually increases with a rise in prosperity. So the model need not predict that the extent of portfolio insurance falls as investors get richer.

3 Non-trading may induce negative autocorrelation in the returns to an individual security. However, when considering an index, we would expect the positive serial cross-correlation to dominate the individual securities' autocorrelation.

4 Another possible reason for the existence of serial correlation in returns could be that transactions bounce randomly between bid and ask prices. We would, though, be surprised if this were a serious problem in the case of daily returns for the index, which is what we consider here. It is possible that, in exceptional periods, a large number of stocks may close at either the bid or ask at the same time. Note, though, that the stock index for the UK is based on mid-market prices, and does not suffer from this problem.

5 For New York we use hourly values of the Dow Jones Index, and for London, we use the Financial Times 30 Share Index.

6 To avoid weekend and overnight effects, we choose to drop the first observation on each day.

7 The data from 1885 up to 1962 was kindly provided to us by G.W.Schwert. The return series is expressed in percentage terms.

8 We also experimented with including a separate dummy variable in the variance for the trading halt between July 31 and December 12, 1914, with no significant effect on the results.

[9] To capture the degree of leptokurtosis exhibited by stock market data, we shall assume that the standardized innovation $\xi_t = \varepsilon_t / \sigma_t$ follows an i.i.d. Generalized Error distribution with 0 mean, unit variance and tail-thickness parameter ν (for $\nu=2$ this distribution reduces to the standard normal distribution, whereas for $\nu<2$ it becomes leptokurtic).

[10] We have also re-estimated our model over 3 sub-samples (1885-1916, 1917-1947 and 1948-1988) with very similar results. The only major difference was a striking increase in γ_0 in the last sub-period.

[11] While explanations based on market micro-structure can generate serial correlation (of either sign), we do not know of any model that can explain a <u>switch in sign</u> with changes in volatility.

Table 4.1

Serial Correlation in Returns
Estimates using Hourly Data (July 1987 - February 1988)

INDEPENDENT VARIABLE	U.S. Data	U.K. Data
r_{t-1}	-0.09	0.12
	(-1.79)	(2.75)
$Crash_t r_{t-1}$	-0.36	-0.17
	(-5.06)	(-2.77)
No. obs.	973	1138
R^2	0.083	0.018

Notes: (i) Equation estimated is $r_t = \alpha + (\gamma_0+\gamma_c Crash_t)r_{t-1} + \varepsilon_t$
(ii) Asymptotic t-ratios in brackets.

Table 4.2

EGARCH Model: Parameter Estimates
Daily data 1885-1988 (29,137 obs)

PARAMETER	ESTIMATE	T-RATIO
α	0.047	6.383
ρ	-0.002	-0.201
γ_0	0.111	16.493
γ_1	-0.019	-7.272
θ_0	-0.000	2.001
π_0	-0.000	-0.523
θ_1	1.891	176.864
θ_2	-0.891	83.806
π_1	-0.978	273.823
δ	0.215	21.442
ψ	-0.116	22.019
ν	1.403	-40.487

Note: t-ratio for ν tests normality (i.e. ν=2).

Chapter 5

QUADRATIC ARCH MODELS:
A POTENTIAL RE-INTERPRETATION OF ARCH MODELS

5.1 Introduction

The analysis of economic time series data usually involves the study of the first and possibly second conditional moments of the series (given past behaviour) in order to characterise the dependence of future observations on past values. As an example, consider the univariate stochastic process x_t, $t=1,2,\ldots$, whose first two conditional moments given the information set (generated by) $X_{t-1}=\{x_{t-1},x_{t-2},\ldots\}$ are:

$$\mu_t = E(x_t/X_{t-1}) = \mu(X_{t-1}) \tag{1a}$$

$$\sigma_t^2 = V(x_t/X_{t-1}) = \sigma^2(X_{t-1}) \tag{1b}$$

In this context, the first step often consists in the estimation of the conditional mean, μ_t, and variance, σ_t^2. In practice, though, this is not a simple task as both $\mu()$ and $\sigma^2()$ are generally unknown functions of the information set X_{t-1}. The most common approach employed is to assume a particular functional form for $\mu()$ and $\sigma^2()$ characterised by certain unknown parameters (which have to be estimated), although non-parametric techniques (which obtain estimates of μ_t and σ_t^2 directly) and mixed approaches (semi-parametric) have gained increased attention recently (see Robinson (1988) and Ullah (1988) for recent surveys).

Within the class of parametric functions, ARMA models are predominant for the conditional mean $\mu()$, while ARCH models enjoy a similar status, although to a lesser extent, with respect to the conditional variance $\sigma^2()$. While all models are by definition simplifications of reality, ARMA models are usually rationalized from a statistical point of view using the Wold Decomposition theorem. Potential justifications for ARCH are only beginning to emerge. Given that ARCH models closely resemble ARMA models for the squared innovations (see e.g. Bollerslev (1986) or Pantula (1986)) one obvious possibility is that Wold's theorem also applies to this process (see Brock, Hsieh and LeBaron (1990)), but this implicitly assumes that the non-negative squared innovations constitute a linear process. A less obvious justification is provided by Nelson (1990b), who shows that ARCH models can consistently approximate continuous time diffusion (or near diffusion) models when the sample frequency increases. This property is shared, though, with many other parameterizations for $\sigma^2()$.

But there are at least two other interpretations of ARMA models which can also be useful in modelling dynamic heteroskedasticity. For the sake of clarity, let's assume that $\mu(X_{t-1})$ is in fact a function of the past r values of x_t only, i.e. $\mu(X_{t-1}) = \mu(\mathbf{x}_{t-r})$, with $\mathbf{x}_{t-r} = (x_{t-1}, x_{t-2}, \ldots, x_{t-r})$. In this framework, it is always possible to understand AR(r) models from an analytical point of view as first-order Taylor approximations to the unknown conditional mean function $\mu()$[1]. That is, if we linearise $\mu(\mathbf{x}_{t-r})$ around some arbitrary point $\bar{\mathbf{x}}_{t-r}$ as:

$$\mu(\mathbf{x}_{t-r}) \approx \mu(\bar{\mathbf{x}}_{t-r}) + D\mu(\bar{\mathbf{x}}_{t-r})'(\mathbf{x}_{t-r}-\bar{\mathbf{x}}_{t-r}) = \alpha + \gamma'\mathbf{x}_{t-r} \tag{2}$$

where $D\mu(\bar{\mathbf{x}}_{t-r})$ denotes the gradient vector of $\mu()$ evaluated at $\bar{\mathbf{x}}_{t-r}$, and $\gamma = D\mu(\bar{\mathbf{x}}_{t-r})$, $\alpha = \mu(\bar{\mathbf{x}}_{t-r}) - D\mu(\bar{\mathbf{x}}_{t-r})'\bar{\mathbf{x}}_{t-r}$, we obtain an AR(r) process[2]. But from a forecasting point of view we can also think of AR(r) models as approximating the conditional mean $\mu(\mathbf{x}_{t-r})$ (which is the Minimum Mean Square Error predictor of x_t based on the relevant information set, $\mathbf{x}_{t-r}$) by the Minimum Mean Square Error Linear Predictor of x_t, say $BLP(x_t/\mathbf{x}_{t-r}) = \alpha + \gamma'\mathbf{x}_{t-r}$. This yields the familiar interpretation of γ_j as a (theoretical) regression coefficient[3].

The empirical success of ARMA models and the fact that there is no such universally accepted model for the conditional variance[4] suggests that it is perhaps worth extending these two interpretations of the conditional mean to the conditional variance function $\sigma^2()$. The main difference is that although $\sigma^2()$ will be positive (almost) everywhere, there is no guarantee that a Taylor approximation to $\sigma^2()$, or indeed some projection of the squared innovations on the information set, would[5]. In fact it is clear that a first-order polynomial cannot be always positive (unless it is trivially constant), and this is also true of any odd-order one. Therefore, an even-order polynomial is required.

The purpose of this chapter is to discuss how to model the conditional variance $\sigma^2()$ as a quadratic function of X_{t-1}. To stress this point we shall call this formulation the Quadratic ARCH (or QARCH) model. Despite the fact that we shall often treat QARCH as if it were the true data generation model, it should be clear from the above discussion that it might be more realistic to interpret it as the quadratic projection of the square innovation on the information set, or, alternatively, as a second-order Taylor approximation to the unknown function $\sigma^2()$.

Under certain conditions the QARCH model reduces to the recently proposed Augmented ARCH (AARCH) model of Bera and Lee (1990), which in turn encompasses Engle's (1982) original ARCH model. It also nests the linear "standard deviation" model discussed by Robinson (1991), and the asymmetric ARCH model in Engle (1990). This nesting has at least three non-trivial advantages. First, many theoretical results derived for these models still hold with minor modifications for the QARCH model, including estimation and testing, stationarity conditions and persistence properties, autocorrelation structure, temporal and contemporaneous aggregation, forecasting, etc. Second, QARCH conditional variances can also be easily integrated in economic models, just as traditional time series models for the conditional mean are (see Hentschel (1991) for an application to a stock returns model). Third, the QARCH model is capable of improving the empirical success of ARCH models, since it avoids some of its criticisms without departing significantly from the standard specification. In particular, it provides a very simple way of calibrating and testing for dynamic asymmetries in the conditional variance function of the kind postulated for some financial series (see e.g. Black (1976), Nelson (1991), and figure 3.2d).

In turn, the QARCH model may be nested into two nonparametric approaches to dynamic conditional heteroskedasticity. First, a quadratic polynomial constitutes the leading term in Gallant's (1981) Flexible Fourier Form approach, where extra trigonometric terms are added to the conditional variance function (see in this context Pagan and Hong (1991) and Pagan and Schwert (1990), or Andrews (1991) for other nonparametric polynomial series procedures which would also nest QARCH). Second, a piecewise quadratic spline approximation to the unknown conditional variance function would also encompass QARCH as a trivial smooth example, as well as the models of Glosten, Jaganathan and Runkle (1989) and Zakoian (1990). Hence, QARCH may also provide a useful benchmark to compare the relative performance of these estimators.

At the same time, given that many issues in finance, and in particular, asset pricing theories, are related to the variances and covariances of many assets, it is of the upmost practical importance in this context to be able to extend univariate models so as to capture time-variation in the (conditional) mean vector and covariance matrix. An additional advantage of the QARCH formulation is that it is very easy to generalize to multivariate models, either directly, or more conveniently, through the different covariance structures suggested in the literature. Hence, it can also capture potential dynamic asymmetries at a multiple assets level.

The chapter is organised as follows. Section 5.2 introduces the QARCH(q) model, and interprets it both as the quadratic projection of the square innovations on the information set, and as a quadratic Taylor approximation to the unknown conditional variance function. It also discusses its generalization to GQARCH(p,q) along the lines of Bollerslev (1986), and states under what parameter restrictions all the other proposed quadratic models can be obtained from it. Section 5.3 re-formulates the GQARCH model as a random coefficients model, and obtains the stationarity condition as well as an expression for the unconditional variance. Fourth moments, the autocorrelation function for the squares of the series and the covariances between these and lagged levels are also discussed. In section 5.4, an illustrative empirical application of the model to daily data on US stock returns is carried out. Potential generalizations to the multivariate case are entertained in section 5.5. In particular, a latent factor model with GQARCH factors is discussed in some detail, and estimated for monthly stock returns of twenty six UK sectors. Finally section 5.6 concludes.

5.2 The QARCH Model

5.2.1 Definition

Let's assume initially that $\mu(X_{t-1})=0$ and that $\sigma^2(X_{t-1})$ is a function of the past q values of x_t only, i.e. $\sigma^2(X_{t-1}) = \sigma^2(\mathbf{x}_{t-q})$, with $\mathbf{x}_{t-q} = (x_{t-1}, x_{t-2}, \ldots, x_{t-q})$. The following conditional variance parameterization:

$$\sigma^2(\mathbf{x}_{t-q}) = \theta + \psi'\mathbf{x}_{t-q} + \mathbf{x}_{t-q}'\mathbf{A}\mathbf{x}_{t-q} \tag{3}$$

is the most general quadratic version possible of the parametric ARCH variance function $h(\mathbf{x}_{t-q};\phi)$ considered in Engle (1982), and for that reason we shall call it the Quadratic ARCH (QARCH hereafter) model. It therefore encompasses all the examples of quadratic variance functions proposed in the literature: the Augmented ARCH model of Bera and Lee (1990), the standard ARCH model of Engle (1982), the linear "standard deviation" model considered by Robinson (1991), and the asymmetric ARCH model in Engle (1990) and Engle and Ng (1991). The AARCH model assumes that $\psi=0$, whereas Engle's ARCH assumes that, in addition, **A** is diagonal. The linear standard deviation model assumes that $\sigma_t^2 = (\rho + \varphi'\mathbf{x}_{t-q})^2$ which implies $\theta=\rho^2$, $\psi=2\rho\varphi$ and $A=\varphi\varphi'$, a rank 1 matrix[6]. An obvious fourth restricted parameterisation, of which the asymmetric ARCH model is a special case, is also encompassed: a model with **A** diagonal but $\psi\neq0$, which we shall term the diagonal QARCH process. In order to see more clearly the differences between the various specifications nested in the QARCH model, let's rewrite equation (3) as

$$\sigma_t^2 - \theta = \sum_{i=1}^{q} \psi_i x_{t-i} + \sum_{i=1}^{q} a_{ii} x_{t-i}^2 + \sum_{i\neq j} a_{ij} x_{t-i} x_{t-j} \tag{4}$$

Engle's (1982) ARCH model[7] would only include the second term in the right hand side of (4), Bera and Lee's (1990) AARCH both the second and third terms, whereas the diagonal QARCH, and therefore Engle's (1990) asymmetric ARCH, the first two. On the other hand, the linear "standard deviation" model of Robinson (1991) would include all three but with restrictions on the parameters.

The cross-product terms give an indication of the extra effect of the interaction of lagged values of x_t on the conditional variance. Hence, by allowing them to be non-zero, it is possible to account for the possibility that the occurrence of e.g. two successive large values of x_t of the same sign affects the conditional variance by more that a ARCH model would allow. But the substantive advantage of the QARCH formulation versus the AARCH and ARCH models which it nests, is that by allowing ψ to take any value, i.e. by not centring the quadratic polynomial for $\sigma^2()$ at 0, an dynamic asymmetric effect of positive and negative lagged values of x_t on σ_t^2 is allowed[8]. As an example, let's take the QARCH(1) model, i.e. $\sigma_t^2 = \theta + \psi_1 x_{t-1} + a_{11}x_{t-1}^2$. If ψ_1 is negative, the conditional variance will be higher when x_{t-1} is negative than when it is positive[9]. In the context of stock market volatility, this could capture the leverage effect noted by Black (1976). Hence, the QARCH model may provide an additive heteroskedastic alternative to the asymmetric multiplicative heteroskedastic EGARCH model of Nelson (1991) used in section 4.3.2. On the other hand, in ARCH and AARCH models (i.e. $\psi_1=0$), only the magnitude, not the sign, of x_{t-1} affect σ_t^2. By contrast, in the linear standard deviation model symmetry is only achievable if $\rho=0$ (in which case we have AARCH with $\sigma^2(0)=0$), or under homoskedasticity.

As we mentioned before, one of the reasons for using a quadratic polynomial is to ensure that our parameterization implies a positive variance everywhere. To see under what conditions the right hand side of (3) will be positive for any $\mathbf{x}_{t-q}$, let's consider the case when A has full rank. Then (3) can be re-written as:

$$\sigma^2(\mathbf{x}_{t-q}) = \theta - 1/4\ \psi'\mathbf{A}^{-1}\psi + (\mathbf{x}_{t-q} + 1/2\ \mathbf{A}^{-1}\psi)'\mathbf{A}(\mathbf{x}_{t-q} + 1/2\mathbf{A}^{-1}\psi) \tag{5}$$

This quadratic function will be positive if and only if $\mathbf{A}$ is positive definite and $\theta - 1/4\ \psi'\mathbf{A}^{-1}\psi > 0$[10]. But from the practical point of view we are interested in parameterising (3) in such a way that the estimated conditional variance is always non-negative. In view of the previous discussion one possible simple way of achieving this is as follows (see Schwallie (1982)):

$$\sigma^2{}_t = (1,\mathbf{x}'_{t-q})\mathbf{P}\mathbf{D}^2\mathbf{P}'(1,\mathbf{x}'_{t-q})' \tag{6}$$

where $\mathbf{P}$ a (q+1)x(q+1) unit lower triangular matrix and $\mathbf{D}$ a diagonal matrix of the same order. The relation between (3) and (6) is obtained by regarding $\mathbf{PD}^2\mathbf{P}'$ as the Cholesky decomposition of the (q+1) symmetric matrix $\begin{bmatrix} \theta & 1/2\ \psi' \\ 1/2\ \psi & \mathbf{A} \end{bmatrix}$. Unfortunately, this reparameterization is only one-to-one if **A** has full rank. For suppose that the last q_2 elements of **D** are zero (as in the linear "standard deviation" model, where q_2=q-1). It is then clear that the sub-diagonal elements in the last q_2 columns of **P** do not appear in the expression for the variance.

5.2.2 QARCH models and Quadratic Projections

Given that we are assuming that $\mu(X_{t-1})=0$, $\sigma^2(X_{t-1})=E(x_t^2/X_{t-1})$ can be regarded as the Minimum Mean Square Error (MMSE) Predictor of x_t^2 based on X_{t-1}. That is, $E[x_t^2-\sigma^2(X_{t-1})]^2 \leq E[x_t^2-P(x_t^2/X_{t-1})]^2$ for any predictor of x_t^2, $P(x_t^2/X_{t-1})$, which uses X_{t-1} as the information set. However, as $\sigma^2()$ is generally of unknown functional form, a natural generalization of the usual practice of replacing conditional means by linear (least squares) projections of information would be to approximate $\sigma^2()$ by the MMSE quadratic predictor of x_t^2 given X_{t-1}, say $BQP(x_t^2/X_{t-1})$.

Taking $\mathcal{X}_{t-1}$ as the space spanned by all possible linear combinations of the elements of the vector $\boldsymbol{x}_{t-1} \equiv \{1, x_{t-1}, x_{t-2}, \ldots, x_{t-1}^2, x_{t-2}^2, \ldots, x_{t-1}x_{t-2}, x_{t-1}x_{t-3}, \ldots\}$, we can define quadratic predictors of x_t^2 given X_{t-1} as linear predictors of x_t^2 based on the "quadratic" information set $\mathcal{X}_{t-1}$. That is, $QP(x_t^2/X_{t-1})=LP(x_t^2/\mathcal{X}_{t-1})$. Therefore, the MMSE (or best) quadratic predictor of x_t^2 given X_{t-1}, will simply be the least squares projection of x_t^2 on $\mathcal{X}_{t-1}$. For this reason, we shall refer to $BQP(x_t^2/X_{t-1})$ as the quadratic projection of x_t^2 on X_{t-1}.

As expected, the prediction error will be orthogonal to any linear combination of lag values of x_t, its squares or cross-products, but the significant result is that this projection will take the form given in (3), i.e. it will look like a QARCH model (with q possibly infinite). Hence, all QARCH parameters can be interpreted in this framework as the (theoretical) regression coefficients of x_t^2 on $\boldsymbol{x}_{t-1}$. For example, consider the QARCH(1) model where $BQP(x_t^2/X_{t-1}) = \theta + \psi_1 x_{t-1} + a_{11}x_{t-1}^2$. Since this can be re-written as $x_t^2 = \theta + \psi_1 x_{t-1} + a_{11}x_{t-1}^2 + \eta_t$, with $\eta_t=x_t^2-BQP(x_t^2/X_{t-1})$, under the additional assumptions that $E(x_t^3)=0$ and all necessary moments exist, $a_{11}=cor(x_t^2,x_{t-1}^2)$ and $\psi_1= cov(x_t^2,x_{t-1})/V(x_t)$. Hence, a_{11} is a measure of the correlation in the square series, while ψ_1 measures dynamic asymmetry[11].

An important result is that if the joint distribution of $x_t, x_{t-1}, \ldots$ is symmetric and the conditional variance function is also symmetric (i.e. $\sigma^2(X_{t-1})=\sigma^2(-X_{t-1})$ in an abuse of notation), then the quadratic projection will be symmetric too. The reason for this is that $\mathrm{cov}(x_t^2, x_{t-j})=E(\sigma^2(X_{t-1})x_{t-j})+E([x_t^2-\sigma^2(X_{t-1})]x_{t-j})=0$, so ψ_j must be 0 for all j. This actually suggests that $\mathrm{cov}(x_t^2, x_{t-j})$ could be a useful tool to identify potential dynamic asymmetries even if QARCH is not the true model, since when $\sigma^2(X_{t-1})\neq\sigma^2(-X_{t-1})$ this third moment may differ from 0[12]. This result also implies that when the data generation process is $x_t=x_t^*\sigma(\mathbf{x}_{t-q})$, with x_t^* iid N(0,1) and $\sigma^2(\mathbf{x}_{t-q})$ non-quadratic but symmetric (as in Higgins and Bera (1990), Schwert (1989b) or Engle and Bollerslev (1986)) the quadratic projection will simply be of the AARCH (or ARCH) forms.

This property is also consistent with the results of Drost and Nijman (1990) and Nijman and Sentana (1991) in the context of temporal and contemporaneous aggregation respectively of GARCH processes[13]. These papers show that if such terms do not appear in the quadratic projections of the underlying series, then they do not appear in the projection of the aggregated one either (but it is easy to prove that if one series shows dynamic asymmetry, then so will the aggregated one).

5.2.3 QARCH Models and Second-Order Taylor Approximations

If $\sigma^2(X_{t-1}) = \sigma^2(\mathbf{x}_{t-q})$, a quadratic Taylor approximation to $\sigma^2(X_{t-1})$ around some arbitrary point $\bar{\mathbf{x}}_{t-q}$ is given by,

$$\sigma^2(\mathbf{x}_{t-q}) \simeq \sigma^2(\bar{\mathbf{x}}_{t-q}) + D\sigma^2(\bar{\mathbf{x}}_{t-q})'(\mathbf{x}_{t-q}-\bar{\mathbf{x}}_{t-q})$$

$$+ 1/2\ (\mathbf{x}_{t-q}-\bar{\mathbf{x}}_{t-q})'\ D^2\sigma^2(\bar{\mathbf{x}}_{t-q})\ (\mathbf{x}_{t-q}-\bar{\mathbf{x}}_{t-q}) = \qquad (7)$$

where $D\sigma^2(\bar{\mathbf{x}}_{t-q})$ and $D^2\sigma^2(\bar{\mathbf{x}}_{t-q})$ denote respectively the gradient vector (i.e. $\partial\sigma^2/\partial\mathbf{x}_{t-q}$) and Hessian matrix (i.e. $\partial^2\sigma^2/\partial\mathbf{x}_{t-q}\partial\mathbf{x}'_{t-q}$) of $\sigma^2()$ evaluated at $\bar{\mathbf{x}}_{t-q}$. If we call $\theta = \sigma^2(\bar{\mathbf{x}}_{t-q}) - D\sigma^2(\bar{\mathbf{x}}_{t-q})'\bar{\mathbf{x}}_{t-q} + 1/2\ \bar{\mathbf{x}}_{t-q}'D^2\sigma^2(\bar{\mathbf{x}}_{t-q})\ \bar{\mathbf{x}}_{t-q}$, $\psi = D\sigma^2(\bar{\mathbf{x}}_{t-q}) - D^2\sigma^2(\bar{\mathbf{x}}_{t-q})\bar{\mathbf{x}}_{t-q}$, $\mathbf{A} = 1/2\ D^2\sigma^2(\bar{\mathbf{x}}_{t-q})$, the right hand side of (7) is again of the QARCH form.

From the discussion in section 5.2.1, an obvious example of a point, $\bar{x}_{t-q}$, for which the approximation in (7) will be positive everywhere, is the minimum of $\sigma^2()$, say x^{min}_{t-q}, since at the minimum the gradient is zero and the Hessian would be positive (semi) definite. However, there is no reason to believe that x^{min}_{t-q} is the only such value. For instance, if $\sigma^2()$ is itself quadratic any point will do just as well. Therefore, although is tempting to view (7) as $\sigma^2(x^{min}_{t-q}) + 1/2\ (x_{t-q} - x^{min}_{t-q})'\ D^2\sigma^2(x^{min}_{t-q})\ (x_{t-q} - x^{min}_{t-q})$, this interpretation cannot be taken for granted (cf. the discussion in footnote 2 about the mean). Notice also that the minimum of $\sigma^2()$, x^{min}_{t-q}, is not necessarily equal to the minimum of the approximation, $-1/2\ A^{-1}\psi$, unless we could choose $\bar{x}_{t-q} = x^{min}_{t-q}$[14].

5.2.4 Estimation and Testing for QARCH effects

Given the interpretation of QARCH models as quadratic projections of x_t^2 on x_{t-1}, a consistent method of estimating the parameters is provided by the OLS regression of x_t^2 on the relevant elements of x_{t-1} (cf. Weiss (1986)). If QARCH is the true model, though, this regression ignores information about the properties of the projection errors, η_t, resulting in inefficient parameter estimates. The preferred method of estimation for ARCH models has been maximum likelihood, but since this involves a nonlinear procedure, it is of some interest to have a simple preliminary test for the presence of QARCH effects. For the ARCH(q), Engle (1982) proposed an LM test which can be computed as TR^2 of the OLS regression of x_t^2 on a constant and its first q lags. This test is distributed as a χ^2_q under the null of no ARCH even if x_t is not conditionally Gaussian provided that the fourth moment of x_t is constant and finite (see Koenker (1981)). Bera, Higgins and Lee (1991) have extended this test to the case of AARCH(q) by including cross-product terms of the form $x_{t-i}x_{t-j}$ in the above regression, resulting in a $\chi^2_{q(q+1)/2}$ null distribution. It is straightforward to check that if we also add the first q lags of x_t as regressors, an LM test for QARCH(q) is obtained, which will be distributed as $\chi^2_{q(q+3)/2}$ under homoskedasticity. It is worth noticing that this test is the analogue of White's (1980a) general test for static heteroskedasticity, which suggests that it may also have good power against most other dynamic heteroskedastic alternatives. As

White's test, it can also be derived as a test for random coefficients (see section 5.3.1 below)

But as in the ARCH(q) model (see Demos and Sentana (1991a)), some of the true parameters lie at the boundary of the parameter space under the null of no QARCH. Although this does not affect the distribution of the LM test, intuition suggests that a more powerful test could be achieved by taking the (partially) one-sided alternative into account. For the sake of clarity let's consider testing against a QARCH(1) alternative. As we have seen, the two sided LM test is based on the regression of x_t^2 on a constant, x_{t-1}^2 and x_{t-1}. Notice that if x_t is symmetrically distributed, the regressors are orthogonal under the null, and we would expect each OLS coefficient to take any sign independently of the other. But under the alternative we would expect the coefficient of x_{t-1}^2 to be positive. Hence, a partially one-sided test of H_0: $\psi_1=a_{11}=0$ vs H_1: $\psi_1\neq0$, $a_{11}>0$ seems more appropriate. The test statistic will be the result of adding to the square of the t-ratio associated with x_{t-1}, the square t-ratio for x_{t-1}^2 when the coefficient is positive (cf. Yancey et al (1980)). Under the null this statistic is distributed as a 50:50 mixture of χ_1^2 and χ_2^2 [15].

5.2.5 GQARCH(p,q) models

So far we have maintained the assumption that the relevant information set contains only a finite number of lags, q. But this may be too restrictive (for instance, Nelson's (1990b) consistency results depend on q being unbounded). Besides, even if q were finite, estimating the model for large q will be difficult as the number of parameters in **A** is $O(q^2)$, and we would need to impose some structure on this matrix. For example, one could introduce a rank k (k<q) structure in A (e.g. if k=1, we would thus obtain the linear "standard deviation" model) but even in that case, the number of parameters could still be excessive for q large. Perhaps the most natural way to solve this problem is by introducing a declining structure on the coefficients. Just as in conditional mean models, this structure can be either rapidly decaying or more slowly so (as in Robinson (1991)). In particular, an exponentially declining lag structure can be obtained by including lagged values of σ_t^2 on the right-hand side of (3), as in Bollerslev (1986). That is:

$$\sigma^2_t = \theta + \psi' \mathbf{x}_{t-q} + \mathbf{x}'_{t-q} \mathbf{A} \mathbf{x}_{t-q} + \sum_{j=1}^{p} \delta_j \sigma^2_{t-j} \tag{8}$$

By analogy with Bollerslev's (1986) GARCH and Bera and Lee's (1990) GAARCH models, we shall term these models Generalized QARCH models of orders p and q, or GQARCH(p,q) for short. As in the case of ARMA models, these GQARCH models will generally result in longer memory models with a flexible lag structure, which, at the same time, could offer a more parsimonious approximation to the conditional variance function. For instance, the asymmetric GARCH(1,1) model in Engle (1990) and Engle and Ng (1991), a GQARCH(1,1) model, is equivalent for $\delta_1<1$ to a diagonal QARCH(∞) model with an infinite dimension matrix $\mathbf{A}_\infty$ whose diagonal elements decline exponentially at a rate δ_1, and an infinite dimension vector ψ_∞ whose elements also decline at the same rate. For higher order (invertible) models the pattern of coefficients of $\mathbf{A}_\infty$ and ψ_∞ will become more complex, but notice that the order of band-diagonality of $\mathbf{A}_\infty$ will always be 2q-1, where q is the order of the QARCH part[16].

5.3 Properties of the QARCH Model

5.3.1 Stationarity Conditions and Moments of the Unconditional Distribution

Let's now consider the following process:

$$x_t = \zeta_t + \sum_{i=1}^{q} \eta_{it} x_{t-i} + \sum_{j=1}^{p} \xi_{jt} \sigma_{t-j} \tag{9}$$

where ζ_t, the η_{it}-s and ξ_{jt}-s are random coefficients. If we assume that ζ_t is i.i.d. $(0,\theta)$, $\eta_t = (\eta_{1t},\ldots,\eta_{qt})$ i.i.d. $(0,\mathbf{A})$ with $\text{cov}(\eta_t,\zeta_t)=1/2\ \psi$, $\xi_t = (\xi_{1t},\ldots,\xi_{pt})$ i.i.d. $(0,\Delta)$, with Δ diagonal, and independent from ζ_t and η_t, we will then have that x_t is a stochastic process whose conditional mean is 0 and whose conditional variance is given by (8), i.e. a GQARCH(p,q) process[17].

This interpretation is not unexpected, since the GQARCH model encompasses the GAARCH model of Bera and Lee (1990), which is introduced as a random coefficients model[18]. Although either the quadratic projection or the second-order Taylor-approximation interpretations of QARCH (and hence AARCH) models are more intuitive, the random coefficients formulation allows us to derive more easily the (covariance) stationarity conditions[19]. In this respect, it is possible to prove (see appendix 5) that, provided that $\sum_{i=1}^{q} a_{ii} + \sum_{i=1}^{p} \delta_j$ is less than 1, the unconditional variance of x_t is given by

$$V(x_t) = \theta \,/\, (1 - \sum_{i=1}^{q} a_{ii} - \sum_{i=1}^{p} \delta_j) \tag{10}$$

Notice that the covariance stationarity of x_t does not depend at all on the linear term in the conditional variance, $\psi' \mathbf{x}_{t-q}$, only on the quadratic term associated with the matrix **A**. Loosely speaking, it is as if the quadratic term asymptotically dominates the linear one. Besides, the stationarity condition for GQARCH (and hence GAARCH) is the same no matter what the off-diagonal elements of **A** are, and so it coincides with that of the nested GARCH model. Notice also that the actual value of the unconditional variance does not depend on ψ. Besides, as (10) does not depend on the off-diagonal elements a_{ij}, the unconditional variance of a GAARCH(p,q) process equals the unconditional variance of the GARCH(p,q) process obtained from it by setting the off-diagonal elements of **A** to 0.

An heuristic proof of (10) may help us understand the reason for the similarity of the unconditional variances for GQARCH, GAARCH and GARCH models. Let's suppose that $E(x_t^2)$ exists and is equal to $\sigma^2 = E(\sigma_t^2)$. Then, if we use the parameterisation of the conditional variance given in (8) (i.e., $\sigma_t^2 = \theta + \sum_{i=1}^{q} \psi_i x_{t-i} + \sum_{i=1}^{q} a_{ii} x_{t-i}^2 + \sum_{i \neq j} a_{ij} x_{t-i} x_{t-j} + \sum_{j=1}^{p} \delta_j \sigma_{t-j}^2$), take expectations at both sides and solve for σ^2, we get exactly expression (10) as both the linear terms and the cross products vanish because x_t is a zero-mean uncorrelated process.

As in standard GARCH models, the sum $\sum_{i=1}^{q} a_{ii} + \sum_{j=1}^{p} \delta_j$ provides a measure of the persistence of shocks to the variance process (but see Nelson (1990a)). Hence, we can analogously define Integrated GQARCH process as those for which this sum is 1 (cf. Engle and Bollerslev (1986)). For instance, the IGQARCH(1,1) will be defined as $\sigma_t^2=\theta+\psi x_{t-1}+(1-\delta)x_{t-1}^2+\delta\sigma_{t-1}^2$, which shares with the IGARCH(1,1) the property that $E(\sigma_{t+j}^2/X_{t-1})=j\theta+\sigma_t^2$ [20].

In order to discuss higher moments, let's define x_t^* as the standardized variable associated with x_t (i.e. $x_t^*=x_t/\sigma_t$). So far we have mainly assumed that $E(x_t^*/X_{t-1})=0$, and $E(x_t^{*2}/X_{t-1})=1$, but if we assume that x_t^*/X_{t-1} is symmetric, so is the unconditional distribution of x_t. To obtain static fourth moments, assume for simplicity that x_t^* is i.i.d. with finite fourth moment μ_4. Provided that the appropriate moments exist, by Jensen inequality we will have that $E(x_t^4) = \mu_4 E(\sigma_t^4) \geq \mu_4 E^2(\sigma_t^2) = \mu_4 V^2(x_t)$, which shows that the degree of leptokurtosis of the unconditional distribution of any zero mean conditionally heteroskedastic model is higher than the degree of leptokurtosis of the assumed distribution for x_t^* (see also Clark (1973)). Thus, GQARCH models share this property with GAARCH and GARCH ones (or indeed with any other Conditionally Heteroskedastic model). In general, though, it is very difficult to make any general comparison of their leptokurtosis.

The GQARCH(1,1) process is an important exception. Assuming symmetry for x_t^*, it is possible to show that, provided $\mu_4 a_{11}^2+2a_{11}\delta_1+\delta_1^2<1$, the fourth moment of x_t is given by:

$$\frac{\mu_4\theta}{(1-\mu_4 a_{11}^2-2a_{11}\delta_1-\delta_1^2)(1-a_{11}-\delta_1)}\,[\theta(1+a_{11}+\delta_1)+\psi_1^2] \qquad (11)$$

which reduces to the expression in Bollerslev (1986) for $\psi_1=0$ and $\mu_4=3$ (i.e. normality). However, if $\psi_1\neq 0$, the GQARCH(1,1) process is more leptokurtic than the GARCH (and GAARCH) (1,1) model which it nests, although the condition for existence of fourth moments is the same (see Bollerslev (1986) for a plot of the appropriate region).

5.3.2 Dynamic Correlation Structure

QARCH processes are uncorrelated but certainly not serially independent. As an example, Bollerslev (1986) proves that in the GARCH(p,q) case the autocorrelation functions for x_t^2 corresponds to that of an ARMA[max(p,q),p] and suggests using the sample autocorrelations as a tool for tentatively selecting the orders p and q, the intuition being that GARCH models can always be re-written as ARMA processes for x_t^2. In the general GQARCH(p,q) model, though, the process for x_t^2 is more complicated. From (8):

$$x_t^2=\theta+\sum_1^q\psi_i x_{t-i}+\sum_{i\neq j} a_{ij}x_{t-i}x_{t-j}+\sum_1^q a_{ii}x_{t-i}^2+\sum_1^p\delta_j x_{t-j}^2+\upsilon_t-\sum_1^p\delta_j\upsilon_{t-j} \qquad (12)$$

with $\upsilon_t = x_t^2-\sigma_t^2$ and $E(\upsilon_t/X_{t-1})=0$.

Nevertheless, under the assumptions that x_t^* is symmetrically distributed and the fourth moment of x_t exists, it is easy to see that for k>max(p,q-1) the autocorrelations, ρ_k, follow the analogue of the Yule-Walker difference equations:

$$\rho_k = \sum_1^q a_{ii}\rho_{k-i}+ \sum_1^p \delta_j\rho_{k-j} \qquad (13)$$

whereas for k≤max(p,q-1), the autocorrelations will depend on all the parameters (θ,ψ,A,δ)[21].

In particular, the autocorrelations of the squares of the GQARCH(1,1) process are exactly the same as those of the GARCH(1,1) model. As a matter of fact, the similarity between GQARCH(1,1) and GARCH(1,1) is remarkable: they both have the same unconditional mean, variance and autocorrelation functions for both the series and its squares. Nevertheless, the GQARCH(1,1) has the advantage that with a single extra parameter (ψ_1) it allows for both an asymmetric effect on the conditional variance and higher kurtosis than the GARCH(1,1) process which it nests, and therefore, it goes in the right direction towards capturing some of the stylised facts characterising many financial time series.

As we saw earlier on, the asymmetry in the conditional variance can be captured more formally by dynamic third moments of the form $\kappa_j=Cov(x_t^2,x_{t-j})$. In the GQARCH(p,q) case, a look at (12) shows that κ_j also follow an analogue of the Yule-Walker equations for j>q. This apparently surprising results is again due to the quadratic structure of the conditional variance function, in which the square terms eventually dominate all others. These moments, though, have to be understood as unconditional, and therefore unable to capture conditional asymmetries. The AARCH(2) model provides an interesting example in which $\kappa_j=0$ for all j, but $E(x_t^2 x_{t-1}/X_{t-2})=2a_{12}x_{t-2}\sigma_{t-1}^2$.

QARCH processes may also allow for oscillatory behaviour in the conditional variance, which is ruled out in standard GARCH models (see Nelson (1991)). As a very simple example consider a stationary GQARCH(1,1) process under the additional assumption that the standardized innovation x_t^* is either 1 or -1 with probability 1/2. Since x_t^{*2} will be identically 1, we can write the variance process as:

$$\sigma_t^2-\sigma^2 = (a_{11}+\delta_1)(\sigma_{t-1}^2-\sigma^2) + \psi_1 x_{t-1}^* \sigma_{t-1} \tag{14}$$

For the sake of clarity, assume that ψ_1 is negative and $\sigma_0^2=\sigma^2$. Then a positive (negative) x_0^* will make σ_1^2 smaller (greater) than σ^2 by the amount $\psi_1\sigma$. Whatever value σ_2^2 takes will obviously depend on x_1^*, but its expectation, as of period 0, will be closer to σ^2 than σ_1^2 because $(a_{11}+\delta_1)$ is less than 1. In the GARCH case, though, $\psi_1=0$ and the conditional variance will always be equal to the unconditional variance σ^2. In general, the extra linear term means that the GQARCH(1,1) conditional variance will be oscillating around the GARCH(1,1) one, being below it when x_{t-1}^* is positive and above when negative.

5.4 An empirical application to daily stock returns

Econometric models of time-varying variances and covariances are particularly relevant in finance applications, and specifically in estimation and testing of asset pricing theories. Thus, it is hardly surprising that the analysis of financial time series have turned out to be the most fruitful application of conditionally heteroskedastic models (see Bollerslev, Chou and Kroner (1990) for a recent survey). Perhaps the best illustration of this is the sheer volume of research on different aspects of stock market return behaviour which assumes

ARCH-type formulations to model time-varying variances. With this background in mind, we shall use here a time series of daily US returns from 1885-1988[22] to investigate whether GQARCH models significantly improve the empirical success of standard GARCH models in representing the conditional variance of stock market returns.

In order to concentrate on the modelling of the conditional variance function $\sigma^2()$, the conditional mean of the returns series r_t is assumed to be simply a linear function of its conditional variance, plus a simple first order moving average term in an attempt to capture the small serial correlation present in the data, i.e.:

$$\mu_t = \alpha + \rho\sigma_t^2 + \pi\varepsilon_{t-1} \tag{15}$$

where ε_t is the innovation in r_t ($\varepsilon_t = r_t - \mu_t$), which we assume follows a GQARCH(p,q) process of the type introduced in (3)[23]. To make use of maximum likelihood methods we require to make an assumption on the conditional distribution of ε_t. Since conditional normality does not seem to capture completely the degree of leptokurtosis exhibited by financial data, fat-tail conditional distributions are often used (see e.g. Bollerslev (1987) or Nelson (1991)). Here we shall assume as in Nelson (1991) and section 4.3.2 that the standardized innovation $\varepsilon_t^* = \varepsilon_t/\sigma_t$ follows an i.i.d. Generalized Error distribution with 0 mean, unit variance and tail-thickness parameter ν.

Most applications of GARCH models to stock market returns have found that a GARCH(1,1) or GARCH(1,2) formulation provides a reasonable representation of the data (see Campbell and Hentschel (1991), Chou (1988), Engle and Ng (1991), French, Schwert and Stambaugh (1987), Pagan and Schwert (1990), etc.). Hence, we have taken a GARCH(1,2) model as our benchmark and nested it into a GQARCH(1,2)[24]. The number of free parameters is therefore 11. The estimates of the parameters are reported in table 5.1

The novel result that we obtain is that the vector ψ is significantly different from zero (see also Engle (1990), Engle and Ng (1991)). The Likelihood Ratio test for GAARCH(1,2) vs GQARCH(1,2) (distributed as a χ^2 with 2 degrees of freedom) yields a value of 289.7. In fact, a large proportion of the improved fit could be achieved with ψ_1 as the single extra parameter: when we fix ψ_2 to 0 (minus twice) the difference between the likelihoods is 253.6. There is also strong evidence of AARCH-type effects. The off-diagonal element of the matrix **A** is estimated to be significantly different from 0: the LR test for the null hypothesis of $a_{12}=0$ (i.e. a diagonal GQARCH model) equals 81.58. The sum of both effects is very important: the 3 restrictions which reduce the estimated GQARCH(1,2) model to a standard GARCH(1,2) model are rejected with a LR of 394.60. Nevertheless, the overall evidence against $\psi=0$ is more important than the evidence against $a_{12}=0$. The LR test for GARCH vs GAARCH is 104.9 whereas the corresponding LR test for GARCH vs diagonal QARCH is 313.02. Therefore, these results seem to provide some empirical support for the GQARCH formulation of (or rather approximation to) the (unknown) conditional variance function as a potential candidate to generalize GARCH models for time-varying variances.

As for the parameter values, the most important result is that the estimated ψ_1 is negative, which supports the common view that a decline in share prices increases volatility by more than a price increase of the same size does: the so-called "leverage effect" (see e.g. Black (1976), Nelson (1991), Campbell and Hentschel (1991))[25]. Besides, since the off-diagonal element of **A** is positive, two successive (say) negative return innovations (i.e. bad news) increase volatility (ceteris paribus) by more than a standard GARCH model would allow. As for the persistence of volatility, we find that the the denominator in the unconditional variance formula (i.e. the trace of **A** plus δ_1) is estimated to be 0.98579. This is again in accordance with earlier results. Note that the degree of leptokurtosis is also high, with an estimate of the tail-thickness parameter ν equal to 1.38500, which is significantly smaller than 2 (its value under normality), and indicates that the conditional distribution of returns has fat tails. Notice that the "price of risk" coefficient ρ is positive, although only marginally statistically significant, a fact which might be puzzling from the economic point of view but consistent with other empirical work in this area (see e.g. Pagan and Hong (1991), Nelson (1991) or the results in sections 3.4 and 4.3.2).

We have also estimated the model for the subperiod July 1962-October 1988. Although there are compelling reasons to believe that there may be a structural break in the model after the 1929 crash (see Pagan and Schwert (1990) or Schwert (1989b)), our specification seems to perform very similarly in the sub-sample. In fact, the major difference is the secular increase in the degree of serial correlation in the data consistent with the results in section 4.3.2. Although no claim is being made about the generality of the estimated specification, we have carried out several specification tests in this subsample based on the moment restrictions imposed by our assumptions on the "standardised" innovations ε_t^* (cf. Nelson (1991) and Braun, Nelson and Sunier (1990)). In particular, we have tested for zero mean, unit variance and static symmetry of ε_t^*, as well as serial independence of ε_t^* and $(\varepsilon_t^{*2}-1)$ for lags 1 to 5. Following Nelson (1991), we can compute conditional moment tests (cf. Newey (1985b)) by replacing these theoretical moments by their sample counterparts and checking if they are significantly different from 0. On the basis of the results in table 5.3 it seems that the model performs reasonably well. The most significant result is the fifth order serial correlation in ε_t^*, which is hardly unexpected given that no attempt has been made to capture the day-of-the-week seasonality in stock returns[26].

5.5 Multivariate Extensions

Let y_t be a multivariate stochastic process of dimension m whose conditional mean is $\mu_t = E(y_t/Y_{t-1}) = \mu(Y_{t-1})$ and whose conditional variance-covariance matrix is $\Sigma_t = V(y_t/Y_{t-1}) = \Sigma(Y_{t-1})$, where $Y_{t-1}=\{y_{t-1}, y_{t-2}, \ldots\}$. Again, for the sake of clarity, we shall deal initially with the case in which the dependence of the conditional moments on the past is limited to a finite number of lags of y_t. As in the univariate case, it is straightforward to see that multivariate AR(r) processes can be understood as first-order multivariate Taylor approximations to the unknown conditional mean function $\mu()$, or as the linear projection of y_t on the multivariate information set. A similar approach is available for the covariance matrix, but again it is clear that only an even-order polynomial can guarantee the positive (semi) definiteness of the conditional covariance matrix for all possible values of variables in the conditioning set.

Let $\mathbf{y}_{t-q} = \mathrm{vec}(y_{t-1}, y_{t-2}, \ldots, y_{t-q})$ be the mqx1 vector containing the values of the m series for those q lags, and let $\mathbf{s}_t = v(\Sigma_t) = v[\Sigma(Y_{t-1})] = \mathbf{s}(\mathbf{y}_{t-q})$, be the vector valued function which contains all the distinct elements of the conditional covariance matrix, where v() is the vector-lower triangle operator which stacks the lower triangular portion of a matrix (see Magnus (1988)). Since Σ_t contains m conditional variances and m(m-1)/2 different conditional covariances, the dimension of s_t is n=m(m+1)/2. For simplicity let's assume that $\mu(Y_{t-1}) = 0$ so that $\mathbf{s}_t = E_{t-1}[v(y_t y_t')]$. Again we can define a multivariate QARCH model as the quadratic projection of $v(y_t y_t')$ on Y_{t-1}, or as the quadratic Taylor approximation of $\mathbf{s}(Y_{t-1})$ around some arbitrary $\bar{\mathbf{y}}_{t-q}$, which is given by the expression:

$$\begin{aligned} \mathbf{s}(\mathbf{y}_{t-q}) &\simeq \mathbf{s}(\bar{\mathbf{y}}_{t-q}) + D\mathbf{s}(\bar{\mathbf{y}}_{t-q})(\mathbf{y}_{t-q}-\bar{\mathbf{y}}_{t-q}) + \\ &+ 1/2\,[I_n \otimes (\mathbf{y}_{t-q}-\bar{\mathbf{y}}_{t-q})']\, D^2\mathbf{s}(\bar{\mathbf{y}}_{t-q})\,(\mathbf{y}_{t-q}-\bar{\mathbf{y}}_{t-q}) = \\ &= \Theta + \Psi\mathbf{y}_{t-q} + (I_n \otimes \mathbf{y}_{t-q}')A\mathbf{y}_{t-q} \end{aligned} \tag{16}$$

where $D\mathbf{s}(\bar{\mathbf{y}}_{t-q})$ and $D^2s^2(\bar{\mathbf{y}}_{t-q})$ denote respectively the Jacobian (i.e. $\partial\mathbf{s}(\mathbf{y}_{t-q})/\partial\mathbf{y}_{t-q}'$) and Hessian (i.e. $\partial\mathrm{vec}'[\partial\mu(\mathbf{y}_{t-q})/\partial\mathbf{y}_{t-q}']/\partial\mathbf{y}_{t-q}'$) matrices of $\mathbf{s}()$ evaluated at $\bar{\mathbf{y}}_{t-q}$, and $\Theta = \mathbf{s}(\bar{\mathbf{y}}_{t-q}) - D\mathbf{s}(\bar{\mathbf{y}}_{t-q})\bar{\mathbf{y}}_{t-q} + 1/2\,(I_n \otimes \bar{\mathbf{y}}_{t-q}')D^2\mathbf{s}(\bar{\mathbf{y}}_{t-q})\,\bar{\mathbf{y}}_{t-q}$, $\Psi = D\mathbf{s}(\bar{\mathbf{y}}_{t-q}) - (I_n \otimes \bar{\mathbf{y}}_{t-q}')D^2\mathbf{s}(\bar{\mathbf{y}}_{t-q})$, $A = 1/2\, D^2\mathbf{s}(\bar{\mathbf{y}}_{t-q})$. By analogy, we shall call this formulation the Multivariate QARCH(q) model[27].

If we partition the nqm x qm matrix A in n square blocks of size qm, and if , in addition to Ψ being 0, each one of these n blocks is in turn a block diagonal matrix with q square blocks of size m, the right hand side of equation (16) reduces to the multivariate ARCH(q) model introduced by Kraft and Engle (1983) (see section 2.1):

$$s_t = B_0 + \sum_{i=1}^{q} B_i\omega_{t-i} \tag{17}$$

where B_0 is a nx1 vector and the B_i-s are nxn matrices and $\omega_t = v(\mathbf{y}_t\mathbf{y}_t')$. On the other hand, when Ψ=0 but the blocks of A are not block diagonal, equation (16) constitutes a multivariate generalization of the AARCH model.

Two problems affect the feasibility of estimating the multivariate QARCH process in (16). The first one is finding necessary and sufficient conditions to guarantee the positive (semi) definiteness of Σ_t. To date, this problem has not been solved completely even for the multivariate ARCH(q) model, although Baba et al. (1989) give a very general parameterization which ensures positive definiteness. The most important practical problem, though, is the sheer number of parameters involved, since Θ is n x 1, Ψ is n x mq and **A** is nmq x mq, with n=m(m+1)/2. Even the multivariate ARCH(q) model contains $n+qn^2$ parameters, and in practice further restrictions have been imposed (e.g. the diagonal ARCH model used in Attanasio and Edey (1987) or Bollerslev, Engle and Wooldridge (1988)).

But as in standard multivariate ARCH models, there are several alternatives to simplify the multivariate QARCH model above. One would be the a k-factor QARCH model (as in Engle (1987)) in which k linear combinations of the y_t-s follow univariate QARCH processes. Another would be a model in which the variances follow QARCH processes but the conditional correlation structure is held constant (as in Bollerslev (1990)). Finally, a conditionally heteroskedastic latent factor model of the type introduced by Diebold and Nerlove (1989), and extended in chapters 1 and 2 but with QARCH-type effects on the underlying factors can also be entertained.

However, since only information about lagged values of y_t is available, much care has to be exercised when dealing with QARCH-type effects in the unobservable factors f_t. An argument similar to that in section 2.6 shows that the conditional variances of each of the factors, $\lambda_{kt|t-1}=V(f_{kt}/Y_{t-1})$, will be given by:

$$\lambda_{kt|t-1} = \theta_k + \sum_{i=1}^{q} \psi_{ki} f_{kt-i|t-i} + \sum_{i\neq j} a_{kij} f_{kt-i|t-i} f_{kt-j|t-j}$$

$$+ \sum_{i=1}^{q} a_{kii}(f^2_{kt-i|t-i}+\omega_{kkt-i|t-i}) + \sum_{j=1}^{p} \delta_{kj}\lambda_{kt-j|t-j-1} \qquad (18)$$

where $f_{kt-i|t-j}=E(f_{kt-i}/Y_{t-j})$ and $\omega_{kkt-i|t-j}=V(f_{it-i}/Y_{t-j})$.

Notice that the only difference between expression (18) and a pure GQARCH(p,q) variance like (8) is the inclusion of a correction in the standard ARCH terms (which reflect the uncertainty in the Kalman filter estimates), but not in the AARCH or linear terms. Besides, the Kalman Filter prediction, updating and smoothing equations derived in section 2.4 for the latent factor model under the assumption that the conditional distribution of the factors is proportional to a (standardized) multivariate t with unknown degrees of freedom remain valid here since they do not actually depend on the particular functional form adopted for the conditional variances.

As a simple example to see whether the QARCH formulation is able to detect potential dynamic asymmetries at a multiple asset level, which could not be captured by multivariate ARCH models, we have estimated a conditionally heteroskedastic latent factor model with QARCH effects using monthly data from 1971:2 to 1990:10 on excess stock returns on 26 UK sectors (see the data appendix 5 for details). The fact that the first two eigenvalues of the unconditional covariance matrix are 20.46 and 0.73 suggest that one common unobservable factor is probably a reasonable initial assumption. Given the small number of observations, we have only considered GQARCH(1,1) parameterizations[28].

The most important result for our purposes is that there appears to be a significant leverage effect in the common factor, which is consistent with the results in Braun, Nelson and Sunier (1990). The LR statistic for $\psi=0$ under the assumption of conditional normality for the factors is 10.74. Figure 5.1 compares the two estimates of the conditional variances. The most noticeable difference is around January 1975, when there was a 51.6% surge in stock prices, which the GARCH(1,1) treats in an analogous manner as the 26.3% drop in October 1987 (despite the fact that the rest of 75 was not particularly volatile). The factor estimates, though, are remarkably close (r=0.999). In fact, the correlation of both the extracted factors with the (demeaned) FTA500 excess return series for the same period is also very high (=0.984). Not surprisingly, when we fit GARCH(1,1) and GQARCH(1,1) models to this series (see table 5.4), there is again evidence for a dynamic asymmetric effect (LR=17.67), with the corresponding variance estimates being very similar (r>0.99)[29].

We have also estimated the model under the alternative assumption of multivariate t conditionally distributed factors. In this respect, we find that the estimated degrees of freedom parameter is 9.73, with a LR test of 516.724! (29.418 for the FTA500 series)[30], which captures the typical excess kurtosis found in stock return data. Nevertheless, the QARCH effect is still significantly present in the results (LR test of 13.27 for the multivariate data, 9.832 for the FTA500).

5.6 Conclusions

In this chapter a general quadratic model for the conditional variance of a time series is introduced. The model can be interpreted as a second-order approximation to the unknown conditional variance function, or perhaps more interestingly, as a quadratic projection of the square innovations on the information set. This seems to be a natural generalization of the usual assumption that conditional means are linear projections of information[31]. It has also the advantage of yielding a class of models that, like ARMA models, is closed under temporal and contemporaneous aggregation.

It turns out that this model is the most general quadratic version possible of the class of Autoregressive Conditionally Heteroskedastic (ARCH) models introduced by Engle (1982) and for that reason we have called it Quadratic ARCH, or QARCH. It encompasses the Augmented ARCH model of Bera and Lee (1990), which allows interactions between different lags of the series to affect the conditional variance, and the standard ARCH model of Engle (1982). It also nests the "linear standard" deviation model in Robinson (1991) and the asymmetric ARCH model in Engle (1990). Its main distinctive feature is that it does not restrict the quadratic approximation to be centred at 0, and therefore allows an asymmetric effect of positive and negative lagged values of the series. It therefore provides an asymmetric "additive" heteroskedastic alternative to the "multiplicative" EGARCH model of Nelson (1991). To allow for infinite dependence on the past, lagged values of the conditional variance can be included as in Bollerslev (1986).

The fact that this new model nests both (G)AARCH and standard (G)ARCH models implies that the time-series properties of a (G)QARCH process are very similar to those of Bollerslev's (1986) GARCH. In particular, GQARCH(1,1) and GARCH(1,1) models are remarkably close: they both have the same mean, variance and autocorrelation functions for both the series and its squares, as well as the same forecasting recursion rule. Nevertheless, the GQARCH(1,1) has the advantage that by adding a single parameter, it can allow for both an asymmetric effect on the conditional variance and higher (unconditional) kurtosis, which goes in the right direction towards capturing some of the stylised facts characterising many financial time series. An additional advantage of quadratic ARCH models is that they are easy to integrate in economic models, just as linear models for the conditional mean are.

An empirical application to a century of daily US stock returns provides support for the fact that the GQARCH approximation to the unknown conditional variance function represents the data substantially better than a standard GARCH model, or even a GAARCH one does. The main reason is that it is able to capture the so-called leverage effect (i.e. price falls increase volatility by more than price increases) which the other two are ruling out a priori.

Multivariate extensions of the QARCH model are straightforward in theory, but as in multivariate ARCH models, difficult to estimate given the number of parameters involved. However, QARCH formulations are particularly easy to adapt to the context of conditionally heteroskedastic latent factor models, and do not entail a large computational burden. An application of such a model to 26 UK industrial sectors also shows empirical support for the GQARCH formulation versus the GARCH one. Again, this seems to be mainly due to the dynamic asymmetric effect.

An important extension of our model would be to allow for other variables to affect the conditional variance in an unrestricted quadratic manner. In standard GARCH models, if these variables can take both positive and negative values, only their squares may actually enter the variance, and with positive coefficients. Using a straightforward generalization of our approach we could allow for linear and quadratic effects of these variables, and for interactions both amongst themselves and with lagged innovations in x_t. This generalization would be closely related to traditional (additive) quadratic heteroskedastic models. In view of the high persistence found in conditional variances, we could also consider extending QARCH along the lines of long-memory models (cf. Robinson (1991)) as an alternative to the GARCH-type generalization. There is also scope for exploring alternative estimators. Since the quadratic projection interpretation of QARCH models stresses the relation between coefficients and moments, Hansen's (1982) Generalised Method of Moments is a clear candidate (see e.g. Rich et al. (1991)).

There is obviously no compelling theoretical reason why the conditional variance function should be literally quadratic, just as there is no reason why conditional means should be linear. If anything, one could expect non-linearities as well as "non-quadratities" to be the rule, rather than the exception, and not surprisingly, the literature on non-linear conditional mean and variance models is growing fast. At the same time, semi-parametric and non-parametric methods for mean and variances are becoming ever more popular. Hence no generality claim should be made about our approach. But equally, it seems sensible to use the quadratic conditional variance function discussed here as a first approximation (or benchmark), just as one would initially use linear models for conditional means.

ENDNOTES

1 Taylor expansions are only locally valid, and may not perform adequately outside a neighbourhood of the point of expansion (see e.g. White (1980b)). Nevertheless, in some instances they may still provide a reasonably good overall approximation.

2 Notice, though, that since $\bar{\mathbf{x}}_{t-r}$ is arbitrary, it is not possible to infer much about $\mu()$ from (2), as different functions could produce the same α and γ for different (or even the same) $\bar{\mathbf{x}}_{t-r}$-s.

3 In practice r can be big, and γ difficult to estimate accurately. This may be solved by imposing some structure on the elements of the vector γ. For instance, an invertible MA(1) process is an AR(∞) with exponentially declining coefficients, and far more complicated structures can be achieved by using mixed models.

4 See e.g. Engle's (1982) ARCH and Bollerslev's (1986) GARCH, Nelson's (1991) Exponential GARCH, Milhoj's (1987) Multiplicative ARCH, Gourieroux and Monfort's (1990) and Zakoian's (1990) versions of the Threshold ARCH, Bera and Lee (1990) Augmented ARCH, Higgins and Bera (1990) Nonlinear ARCH, Taylor (1986) and Schwert (1989b) Absolute Value ARCH.

5 By continuity, the Taylor approximation will have locally the same sign as the function. However we are interested here in approximating the unknown function everywhere.

6 It does not encompass, though, the models of Taylor (1986) and Schwert (1989b), in which the variance is quadratic in the absolute value of innovations.

7 Engle's (1982) original model is sometimes referred to as linear ARCH. However, the term linear is unfortunate since as (4) shows it is only linear in the squares of the past values, not in the information set X_{t-1}.

8 Both the (serially uncorrelated) AARCH model and the ARCH model are examples of processes for which "no news is good news" (as in Campbell and Hentschel (1991) since the minimum of the conditional variance function is achieved when $\mathbf{x}_{t-q}=0$. The fact that this property is not shared by the QARCH model (unless $\psi=0$) could be considered as a disadvantage. However, whether the shape of the unknown conditional variance function $\sigma^2()$ results in an estimated value of ψ different from 0 is, in fact, an empirical question which can be tested using the QARCH model, but not the AARCH or ARCH models.

9 Besides, when $\psi_1<0$ the absolute value of the derivative of $\sigma^2(x_{t-1})$ with respect to x_{t-1} $(=\psi_1+2a_{11}x_{t-1})$ is also higher for negative than positive x_{t-1}. Hence, the conditional variance function is not only asymmetric, but also steeper for x_{t-1} negative. The rate of growth of this derivative, though, is assumed to be the same $(=a_{11})$.

10 If $\mathrm{rank}[D^2\sigma^2(\bar{\mathbf{x}}_{t-q})]=\mathrm{rank}(\mathbf{A})=s<q$, the conditions for positivity of (3) are more complex. Let $\mathbf{A}=\mathbf{U}\Lambda\mathbf{U}'$ be the spectral decomposition of the symmetric matrix $\mathbf{A}$, with $\mathbf{U}=(\mathbf{U}_1,\mathbf{U}_2)$ and $\Lambda = \mathrm{diag}(\Lambda_1,0)$. Then $\theta + \psi'\mathbf{x}_{t-q} + \mathbf{x}_{t-q}'\mathbf{A}\mathbf{x}_{t-q} = \theta + \psi'\mathbf{U}\mathbf{U}'\mathbf{x}_{t-q} + \mathbf{x}_{t-q}'\mathbf{U}\Lambda\mathbf{U}'\mathbf{x}_{t-q} = \theta + \psi^{*\prime}\mathbf{x}^*_{t-q} + \mathbf{x}^{*\prime}_{t-q}\Lambda\mathbf{x}^*_{t-q}$. Given that the last q-s elements of Λ are zero, it is clear that positivity requires the last q-s elements of ψ^* to be zero as well. We are then left with $\theta + \psi_1^{*\prime}\mathbf{x}^*_{1t-q} + \mathbf{x}^{*\prime}_{1t-q}\Lambda_1\mathbf{x}^*_{1t-q} = \theta - 1/4\ \psi_1^{*\prime}\Lambda_1^{-1}\psi_1^* + (\mathbf{x}^*_{1t-q}+1/2\Lambda_1^{-1}\psi_1^*)'\Lambda_1(\mathbf{x}^*_{1t-q}+1/2\ \Lambda_1^{-1}\psi_1^*)$, which shows that to guarantee $\sigma_t^2>0$, in addition to $\psi^*_{q-s+1}=\ldots=\psi^*_q=0$, $\mathbf{A}$ has to be positive semi-definite and $\theta-1/4\ \psi_1^{*\prime}\Lambda_1^{-1}\psi_1^*>0$. It is worth mentioning that parameterising the log variance instead of the variance would eliminate the non-negative problem in the univariate case (cf. Nelson (1991)).

11 We can also consider MMSE quadratic predictors of x_{t+j}^2 given X_{t-1}, which coincide with the corresponding conditional variances if QARCH is the true model. In this respect, a useful property of QARCH(q) projections with q finite (and in some cases infinite, cf. equation (8)), is that they follow a simple recursion. As an example, suppose that $BQP(x_t^2/X_{t-1})$ is of the QARCH(2) form. Since x_{t+1}^2 can be expressed as $\theta + \psi_1 x_t + \psi_2 x_{t-1} + a_{11}x_t^2 + a_{22}x_{t-1}^2 + 2a_{12}x_t x_{t-1} + \eta_{t+1}$, with $\eta_{t+1} = x_{t+1}^2 - BQP(x_{t+1}^2/X_t)$, then $BQP(x_{t+1}^2/X_{t-1}) = BLP(x_{t+1}^2/X_{t-1}) = \theta + a_{11}BQP(x_t^2/X_{t-1}) + \psi_2 x_{t-1} + a_{22}x_{t-1}^2$, because η_{t+1} is (quadratically) unpredictable from X_t (and X_{t-1}) and $E(x_t/X_{t-1})=0$ by assumption. For $j \geq 2$, $E(x_{t+j-1}x_{t+j-2}/X_{t-1})=0$ and $BQP(x_{t+j}^2/X_{t-1}) = \theta + a_{11}BQP(x_{t+j-1}^2/X_{t-1}) + a_{22}BQP(x_{t+j-2}^2/X_{t-1})$, just as in the ARCH(2) case.

12 Campbell and Hentschel (1991) alternatively use the correlation coefficient between x_t^2 and x_{t-1} to measure what they call "predictive" (i.e. dynamic) asymmetries in stock returns. Of course, as with the usual skewness coefficient, it may be that in some special cases $cov(x_t^2, x_{t-j})=0$ but $\sigma^2(X_{t-1}) \neq \sigma^2(-X_{t-1})$.

13 Drost and Nijman (1990) Weak ARCH concept is a particular case of what we call quadratic projections, in which linear and cross product terms are not included.

14 Since the AARCH model is obtained from (3) when $\psi = 0$, it could be interpreted as a second-order Taylor approximation at a point $\bar{\mathbf{x}}_{t-q}$ where $D\sigma^2(\bar{\mathbf{x}}_{t-q}) = D^2\sigma^2(\bar{\mathbf{x}}_{t-q})\bar{\mathbf{x}}_{t-q}$ and the Hessian $D^2\sigma^2(\bar{\mathbf{x}}_{t-q})$ is positive (semi) definite. If $\mathbf{x}_{t-q}^{min} = 0$, the minimum could be such a point but again there might be others. If in addition the cross-derivatives are zero we would get the standard ARCH model.

15 It can be proved that the distribution of the corresponding LR and Wald tests is the same. To do so it is more convenient to work with the one-to-one reparameterization $\sigma_t^2 = \phi_{00} + 2\phi_{01}x_{t-1} + (\phi_{11} + \phi_{01}^2/\phi_{00})x_{t-1}^2$. If the distribution of x_t is not symmetric, though, the mixing weights in all three tests will change.

[16] Conditions for the positivity of the conditional variance in (8), can be obtained as a straightforward extension to θ_∞, ψ_∞ and $\mathbf{A}_\infty$ of the finite q results in 2.1. However, as Drost and Nijman (1990) and Nelson and Cao (1991) point out for the standard GARCH(p,q) model, requiring the positivity of the QARCH part plus $\delta_j \geq 0$ for all j is unduly restrictive. As in the GARCH(p,q) model, though, finding restrictions on the original parameters which guarantee nonnegative variances is not an easy task (see Nelson and Cao (1991) for some special GARCH cases and Demos and Sentana (1991b) for some GQARCH ones).

[17] If we take $(\zeta_t, \eta_t', \xi_t')$ as a (white noise) state vector, equation (9) can be interpreted as the state-space representation of the GQARCH(p,q) process. Strictly speaking, though, it is only valid for those cases when $\mathbf{A}$ and Δ are positive semi-definite, and $\theta - 1/4\psi'\mathbf{A}^{-1}\psi \geq 0$.

[18] Bera and Lee (1990) introduce autoregressive behaviour in x_t by allowing η_t to have a non-zero mean. This results, though, in a model which is different from an AR(r) process with AARCH disturbances, since the conditional variance function depends on the actual series directly, not on deviations from its conditional mean (see Bera, Higgins and Lee (1991)). Given that the conditional mean is linear, though, theirs will actually be an AR(r) process with QARCH disturbances. We shall not pursue this generalization here as our main interest is the conditional variance function, but it is worth noticing that for linear conditional mean models, a quadratic ARCH is obtained whether we use innovations or the actual series.

[19] The random coefficients interpretation also suggests potential generalizations of the GQARCH model. For example, one could allow for non-zero covariances between η_t and/or ζ_t with ξ_t, as well as non-zero covariances between different ξ_{it}-s. For instance, the asymmetric nonlinear GARCH model in Engle and Ng (1991) can be written as $\zeta_t + \eta_{1t}x_{t-1} + \xi_{1t}\sigma_{t-1}$, with $\mathrm{cov}(\eta_{1t}, \xi_{1t}) \neq 0$.

[20] IGARCH processes are clearly not covariance stationary, but they are strictly stationary and ergodic (cf. Nelson (1990a) and Bougerol and Picard (1990)). Given that the behaviour of GQARCH processes is dominated by the quadratic terms, one would expect a similar result to be true of IGQARCH. Unfortunately, the presence of linear and cross-products terms implies that the recursions usually employed to prove strict stationarity are not easy to solve, even in the simplest IGQARCH(1,1) case.

[21] In the GARCH(p,q) model, the autocorrelations follow equation (13) for k>p (see Bollerslev (1986)). Here, however, the existence of the other terms in (12) implies that the first q-1 autocorrelations do not follow equation (13) even when p=0.

[22] The data from 1885 up to 1962 was kindly provided by G.W.Schwert. See Schwert (1990) for further discussion of the data. It is worth noticing that we work with market returns, not excess returns. The results in Nelson (1991) suggests that this should make little difference as far as estimation of the conditional variance is concerned.

[23] It is unlikely that a simple MA(1) component will capture the type of serial correlation observed in high frequency returns (see Lo and MacKinlay (1990), LeBaron (1990) or Sentana and Wadhwani (1991)). We have also ignored other effects such as day-of-the-week or month-of-the-year seasonality (see Thaler (1987) or the the contribution of non-trading periods to market variance (see e.g. French and Roll (1986)).

[24] Although we recognize that a simple MA(1)-GQARCH(1,2)-M model may not provide a complete representation of a century of daily stock returns, our aim is merely to see if this offers an improvement over the standard GARCH(1,2) model. It is worth mentioning that our conclusions are rather robust to the particular values of p and q used.

25 The fact that ψ_2 is positive seems to contradict this claim. However, notice that the presence of lagged values of σ^2_t in the variance equation implies that the estimated values of $\psi_{2\infty}$ ($=\psi_2+\delta_1\psi_1$) in the QARCH(∞) representation of the model ($\sigma^2_t=\theta_\infty+\psi'_\infty \mathbf{x}_{t-\infty}+\mathbf{x}'_{t-\infty}A_\infty \mathbf{x}_{t-\infty}$) is -0.0405, consistent with the reported asymmetric effect.

26 Given the discussion in section 5.2.2, we have also tested the third moment restriction $E[(\varepsilon^{*2}_t-1)\varepsilon^*_{t-1}]=0$ in order to see if the estimated specification captures correctly dynamic asymmetries. The corresponding mean value and t statistic presented in Table 5.3 are -0.0336 and -0.8303 respectively (see Engle and Ng (1991) for alternative LM-type tests for dynamic asymmetries).

27 Infinite dependence on the past can again be achieved by including lagged values of s_t on the right hand side of (16) as in Bollerslev, Engle and Woodridge (1988).

28 To concentrate on the modelling of the conditional covariance matrix, the data has been demeaned prior to estimation (cf. Braun, Nelson and Sunier (1990))

29 In the GQARCH(1,1) case, positivity of the variance is achieved if $a_{11},\delta_1\geq 0$ and $\psi_1^2\leq 4a_{11}\theta$ (cf. Demos and Sentana (1991b)). To impose these restrictions, we have re-parameterised $\sigma^2()$ as $\sigma^2_t=c^2+l^2_{11}(\varepsilon_{t-1}-b_1)^2+d^2_1\sigma^2_{t-1}$ (cf. eq. A2). Notice that the fact that c is estimated to be close to 0 ($2.82\text{x}10^{-8}$) implies that the covariance matrix of the random coefficients ζ_t and η_{1t} is almost singular (cf. eq. 9). If d_1 were 0, it would correspond to a linear standard deviation model.

30 As in Bollerslev (1988), we have parameterised the degrees of freedom in terms of $1/\nu$, so that this parameter is 0 under the null. The distribution of the LR test is then a mixture of χ^2_0 (a constant equal to 0) and χ^2_1, with a 5% critical value of 2.7 (see Bollerslev (1988) or Demos and Sentana (1991a)).

31 Nelson and Foster (1991) have recently provided a different justification for a GQARCH(1,1)-type model. They show that if the true data generating model is given by the stochastic volatility model:

$$_hX_{(k+1)h} = {}_hX_{kh} + h\mu({}_hX_{kh}, {}_h\sigma_{kh}) + h^{1/2}\,{}_h\sigma_{kh}\;{}_hZ_{1,kh}$$

$$_h\sigma^2_{(k+1)h} = {}_h\sigma^2_{kh} + h\lambda({}_hX_{kh}, {}_h\sigma_{kh}) + h^{1/2}\Lambda({}_hX_{kh}, {}_h\sigma_{kh})\;{}_hZ_{2,kh}$$

with $({}_hZ_{1,kh}, {}_hZ_{2,kh})$ iid normal with zero mean, unit variances and correlation ρ, the optimal choice of ARCH functional form (in the sense of minimising the variance of the filtering error) is a linear combination of linear and quadratic past residuals (see Nelson and Foster (1991) for details).

APPENDIX 5

To prove covariance stationarity, it is more convenient to work with the following alternative re-parameterization:

$$x_t = \zeta_t + \sum_{i=1}^{q} \eta_{it}(x_{t-i}-b_i) + \sum_{j=1}^{p} \xi_{jt}\sigma_{t-j} \tag{A1}$$

where the b_i-s are constant parameters and ζ_t, the η_{it}-s and ξ_{jt}-s random coefficients. If we assume that $\eta_t = (\eta_{1t},\dots,\eta_{qt})$ is i.i.d. $(0,\mathbf{LL'})$, $\xi_t = (\xi_{1t},\dots,\xi_{pt})$ i.i.d. $(0,\Delta)$, with Δ diagonal, ζ_t also i.i.d. $(0,c)$, with η_t, ξ_t and ζ_t mutually independent, we will then have that x_t is a stochastic process whose conditional mean is 0 and whose conditional variance is that of a GQARCH(p,q) process:

$$\sigma^2_t = c + (\mathbf{x}_{t-q}-\mathbf{b})'\mathbf{LL'}(\mathbf{x}_{t-q}-\mathbf{b}) + \sum_{j=1}^{p} \delta_j\sigma^2_{t-j} \tag{A2}$$

Let $s=\max(p,q)$, and let $\mathbf{z}_t'=(x_t,x_{t-1},..,x_{t-s+1})$, $\mathbf{v}_t'=(\sigma_t,\sigma_{t-1},..,\sigma_{t-s+1})$, and $\mathbf{u}_t'=(\zeta_t,0,..,0)$ be sx1 vectors. Let's also define the following sxs matrices:

$$\Phi = \begin{bmatrix} \mathbf{0}_{1\times(s-1)} & 0_{1\times1} \\ \mathbf{I}_{s-1} & \mathbf{0}_{(s-1)\times1} \end{bmatrix}$$

$$\Psi_t = \begin{bmatrix} \eta_t' & \mathbf{0}_{1\times(s-q)} \\ \mathbf{0}_{(s-1)\times q} & \mathbf{0}_{(s-1)\times(s-q)} \end{bmatrix}$$

$$\Pi_t = \begin{bmatrix} \xi_t' & \mathbf{0}_{1\times(s-p)} \\ \mathbf{0}_{(s-1)\times p} & \mathbf{0}_{(s-1)\times(s-p)} \end{bmatrix}$$

so that the process can be represented as:

$$\mathbf{z}_t = (\Phi + \Psi_t)\mathbf{z}_{t-1} + \Pi_t\mathbf{v}_{t-1} + \mathbf{u}_t - \Psi_t\mathbf{b} \tag{A3}$$

Let $V_t = E(z_t z'_{t-1})$. Then, recalling the mutual and serial independence of η_t, ξ_t and ζ_t, it follows that:

$$V_t = E(\Phi z_{t-1} z'_{t-1} \Phi') + E(\Psi_t z_{t-1} z'_{t-1} \Psi'_t) +$$

$$+ E(\Pi_t v_{t-1} v'_{t-1} \Pi'_t) + (c + \mathbf{b}'\mathbf{LL}'\mathbf{b})G \tag{A4}$$

where:

$$G = \begin{bmatrix} 1 & 0_{1x(s-1)} \\ & \\ 0_{(s-1)x1} & 0_{(s-1)x(s-1)} \end{bmatrix}$$

Apart from the inclusion of $\mathbf{b}'\mathbf{LL}'\mathbf{b}$ in the constant term, this expression is the same as the one derived by Bera and Lee (1990) for the GAARCH model (i.e. when $\mathbf{b}=0$), and hence the stationarity condition for GQARCH models, stated in the following proposition, is the same as the stationarity condition for GAARCH models:

Let

$$L_s = \begin{bmatrix} L & 0_{qx(s-q)} \\ & \\ 0_{(s-q)xq} & 0_{(s-q)x(s-q)} \end{bmatrix}$$

be a sxs lower triangular matrix, and let

$$\Delta_s = \begin{bmatrix} \Delta & 0_{px(s-p)} \\ & \\ 0_{(s-p)xp} & 0_{(s-p)x(s-p)} \end{bmatrix}$$

be a sxs positive semi-definite diagonal matrix.

Then x_t as generated by equation (8) is covariance stationary if and only if all the eigenvalues of the sxs matrix $\mathbf{R}$ are less than 1 in absolute value, where:

$$\mathbf{R} = (\Phi \otimes \Phi) + \mathrm{vec}(G)\mathrm{vec}'(L_s L'_s + \Delta_s) \tag{A5}$$

Proof: see Bera and Lee (1990)

Provided that this condition is satisfied, the unconditional variance of z_t, $\mathbf{V}$, will be:

$$\mathrm{vec}(\mathbf{V}) = (c+\mathbf{b}'\mathbf{L}\mathbf{L}'\mathbf{b})\ [(\mathbf{I}_s \otimes \mathbf{I}_s)-\mathbf{R}]^{-1}\ \mathrm{vec}(\mathbf{G}) \tag{A6}$$

Given the shape of $\mathbf{G}$, the unconditional variance of x_t will be given by $(c+\mathbf{b}'\mathbf{L}\mathbf{L}'\mathbf{b})$ multiplied by the element 1,1 of $[(\mathbf{I}_s \otimes \mathbf{I}_s)-\mathbf{R}]^{-1}$, which is easily seen to be the reciprocal of the determinant of $[(\mathbf{I}_s \otimes \mathbf{I}_s)-\mathbf{R}]^{-1}$. Tedious algebra shows that this determinant is equal to the familiar expression $1-\sum_{i=1}^{q} a_{ii}-\sum_{j=1}^{p} \delta_j$, where a_{ii} is the ith diagonal element of the matrix $\mathbf{A}=\mathbf{L}\mathbf{L}'$. Therefore the unconditional variance of x_t is indeed given by equation (10)

DATA APPENDIX 5

The following list refers to the definition of the 26 Financial Times Actuaries Sector Indices as of December 31, 1990, and includes the DATASTREAM four-letter sector mnemonics (all starting with FTA).

BANK	Banks
BDIS	Brewers and Distillers
BMAT	Building Materials
CHEM	Chemicals
CONC	Contracting, construction
ELEC	Electricals
ENGG	Engineering General
FDMG	Food Manufacturing
FDRT	Food Retailing
INBR	Insurance (Brokers)
INCM	Insurance (Composite)
INLF	Insurance (life)
INVT	Investment Trusts
LEIS	Leisure
MERB	Merchant Banks
METL	Metals and Metal Forming
MISC	Miscellaneous
MISF	Other Financial
MTRS	Motors
NWSP	Publishing and Printing
OILS	Oil and Gas
PAPA	Packaging and Paper
PROP	Property
SHPT	Shipping and Transport
STOR	Stores
TEXT	Textiles

Of this sectors, seven (Banks, Life Insurance, Insurance General, Insurance Brokers, Merchant Banks, Property and Investment Trusts) are not in the FTA 500 share index (FTA500I). The beginning of the month 3-month Tbill rate was used as the safe interest rate (UKTRSBL%).

Table 5.1

GQARCH(1,2) and GARCH(1,2) Parameter Estimates
US Daily Stock Returns 1885-1988 (29,137 obs)

PARAMETER	GQARCH ESTIMATES (STD. ERROR)	GARCH ESTIMATES (STD. ERROR)
α	0.04725 (0.00594)	0.06364 (0.02591)
π	0.09878 (0.00637)	0.08400 (0.01775)
ρ	0.01685 (0.00847)	0.01121 (0.01438)
θ	0.01293 (0.00089)	0.01030 (0.00233)
ψ_1	-0.10533 (0.00843)	
ψ_2	0.05370 (0.00949)	
a_{11}	0.11336 (0.00558)	0.13319 (0.03832)
a_{21}	0.02475 (0.00278)	
a_{22}	-0.02212 (0.00636)	-0.04222 (0.07057)
δ_1	0.89455 (0.00322)	0.89912 (0.03110)
ν	1.38500 (0.01512)	1.34818 (0.01562)
Log-Likelihood	-34001.00	-34198.30

Table 5.2

GQARCH(1,2) and GARCH(1,2) Parameter Estimates
US Daily Stock Returns 1962-1988 (6,662 obs)

PARAMETER	GQARCH ESTIMATES (STD. ERROR)	GARCH ESTIMATES (STD. ERROR)
α	0.03728 (0.01409)	0.06190 (0.01030)
π	0.22902 (0.01500)	0.21659 (0.01263)
ρ	0.02729 (0.02086)	0.02000 (0.02053)
θ	0.00679 (0.00121)	0.00648 (0.00110)
ψ_1	-0.06489 (0.01646)	
ψ_2	0.01413 (0.01901)	
a_{11}	0.10774 (0.02033)	0.11616 (0.01780)
a_{21}	0.02124 (0.00421)	
a_{22}	-0.02114 (0.02433)	-0.02012 (0.02401)
δ_1	0.90649 (0.00876)	0.89819 (0.01081)
ν	1.48123 (0.03731)	1.43908 (0.03247)
Log-Likelihood	-6928.81	-6983.48

Table 5.3

GQARCH(1,2) Conditional Moment Tests
US Daily Stock Returns 1962-1988 (6,662 obs)

Orthogonality condition	Sample Average	t statistic
$E(\varepsilon_t^*)=0$	-0.0103	-0.8416
$E(\varepsilon_t^{*2}-1)=0$	0.0050	0.2044
$E(\varepsilon_t^*\|\varepsilon_t^*\|)=0$	-0.0354	-1.2927
$E(\varepsilon_t^*\varepsilon_{t-1}^*)=0$	0.0185	1.3946
$E(\varepsilon_t^*\varepsilon_{t-2}^*)=0$	0.0120	0.9850
$E(\varepsilon_t^*\varepsilon_{t-3}^*)=0$	0.0208	1.6005
$E(\varepsilon_t^*\varepsilon_{t-4}^*)=0$	0.0213	1.7029
$E(\varepsilon_t^*\varepsilon_{t-5}^*)=0$	0.0280	2.2559
$E[(\varepsilon_t^{*2}-1)(\varepsilon_{t-1}^{*2}-1)]=0$	0.1779	1.0157
$E[(\varepsilon_t^{*2}-1)(\varepsilon_{t-2}^{*2}-1)]=0$	-0.0095	-0.1919
$E[(\varepsilon_t^{*2}-1)(\varepsilon_{t-3}^{*2}-1)]=0$	0.1327	1.0408
$E[(\varepsilon_t^{*2}-1)(\varepsilon_{t-4}^{*2}-1)]=0$	0.0470	1.1319
$E[(\varepsilon_t^{*2}-1)(\varepsilon_{t-5}^{*2}-1)]=0$	0.0358	0.7862
$E[(\varepsilon_t^{*2}-1)\varepsilon_{t-1}^*]=0$	-0.0336	-0.8303

Table 5.4

GQARCH(1,1) and GARCH(1,1) Parameter Estimates
UK Monthly Excess Stock Returns 1971:2-1990:10 (237 obs)

PARAMETER	GQARCH ESTIMATES (STD. ERROR)	GARCH ESTIMATES (STD. ERROR)
c	0.00000	0.29589
	(0.20276)	(0.08809)
b_1	1.18553	
	(0.23073)	
l_{11}	0.31322	0.42521
	(0.05189)	(0.08903)
d_1	0.82814	0.84102
	(0.03099)	(0.06655)
Log-Likelihood	-269.487	-278.326
θ	0.13789	0.08755
ψ_1	-0.23261	
a_{11}	0.09810	0.18080
δ_1	0.68582	0.70732

Notes: Equation estimated:

$$\sigma_t^2 = c^2 + l_{11}^2(\varepsilon_{t-1} - b_1)^2 + d_1^2\sigma_{t-1}^2 = \theta + \psi_1\varepsilon_{t-1} + a_{11}\varepsilon_{t-1}^2 + \delta_1\sigma_{t-1}^2$$

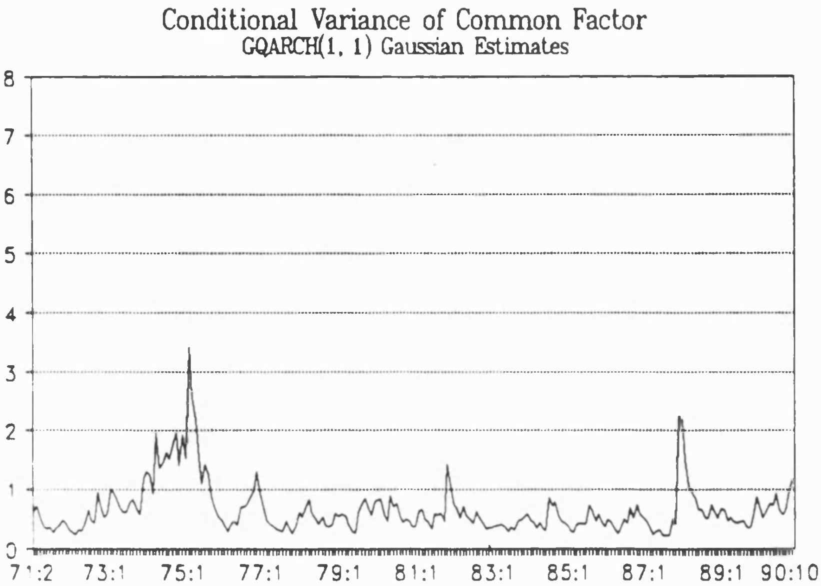


Figure 5.1a

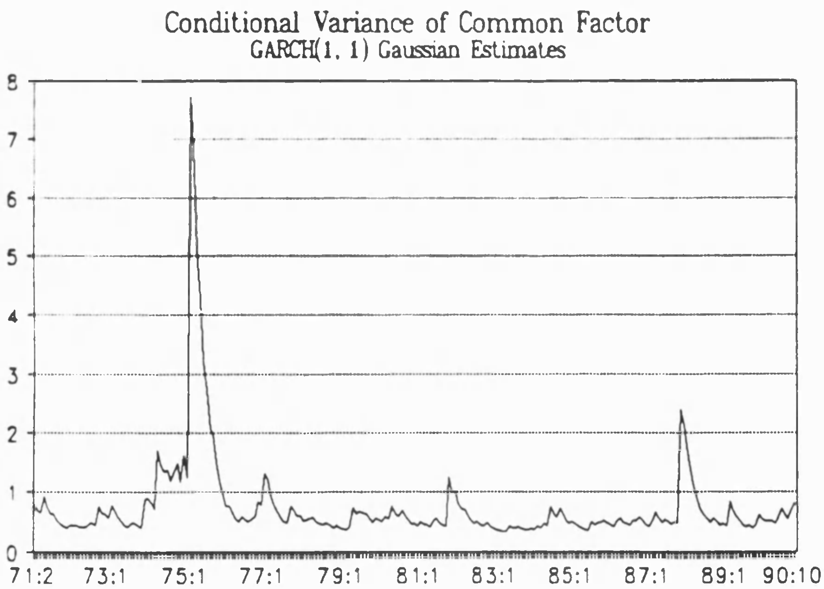


Figure 5.1b

CONCLUSIONS

In this research we have investigated various issues related to the level and volatility of returns on ordinary shares. In particular, we look at the relation over time between volatility and risk premia, both at a univariate and multivariate levels.

We have studied the behaviour of volatility for sixteen world stock markets, and tried to identify its major determinants. We have also attempted to assess the extent of capital market integration. Our results seem to suggest that global stock markets are not integrated in that idiosyncratic risk is significantly priced and the "price of risk" associated with the different factors is not common across countries. Another empirical finding is that only a small proportion of the covariances between national stock markets can be accounted for by "observable" economic variables. Changes in correlations between markets seem to be driven primarily by movements in unobservable variables.

An important result obtained is that it is possible to evaluate the separate contribution of the different factors to the risk premia in an APT-type model when time-variation in the volatility of the factors is taken into account in estimation. This result derives from our proof that in that case the matrix of factor loadings is unique under orthogonal transformations of the factors. We have also discussed the issues of Kalman filtering and estimation of this class of conditionally heteroskedastic latent factor models under the assumption that the conditional distributions are multivariate t (which includes the usual Gaussianity assumption as a special case) and GARCH formulations for the conditional variances of the factors. Both assumptions are particularly important in practice given the degree of leptokurtosis and volatility clustering of many financial time series.

We have also attempted to explore whether the Japanese stock market is efficient so that the lagged variables that help predict stock returns there are merely proxying for mis-measured risk. However, four key predictor economic variables are shown to have non-trivial additional forecasting power irrespective of how risk is measured. Interestingly, unlike the US, the level of the lagged dividend yield is not positively correlated with returns in either

Japan or South Korea. Moreover, there is no consistent relationship between volatility and excess returns, and a model with a constant "price of risk" is clearly rejected by the data. These findings are inconsistent with many existing models of the stock market, and they suggest either the existence of "noise" traders or the need for better models of risk.

We have also discussed whether the degree of autocorrelation shown by high frequency stock returns may change with volatility. Results using a century of daily data suggest that when volatility is low there tends to be positive autocorrelation in returns, but this serial correlation can become negative during very volatile episodes. Our results also suggest an important asymmetry: returns are more likely to exhibit negative serial correlation after price declines. This is consistent with price declines being more likely to induce positive feedback trading. We also find no significant relation between margin requirements and the serial correlation of returns.

Throughout this research we have encountered a well-known asymmetry in stock market volatility: the so-called leverage effect, i.e. volatility is higher following price declines than after rises of the same magnitude. To capture this dynamic asymmetry we have introduced a new, generalised, fully quadratic version of the ARCH formulation which nests the original ARCH model and several of its extensions. As a consequence, its statistical properties are very similar to those of standard models, but it avoids some of their criticisms. This model can also be interpreted as a second-order Taylor approximation to the unknown conditional variance function, or as the quadratic projection of the squared series on the information set. In an application to a century of daily US stock returns, QARCH models provide a closer approximation to the data because they are able to capture the so-called leverage effect. QARCH formulations are also easy to incorporate in multivariate models so as to capture dynamic asymmetric effects that GARCH rules out. In this respect, we have estimated a one-factor model for monthly excess returns on 26 industrial UK sectors, and again, found empirical support for the QARCH as the common factor (which is highly correlated with returns on the FTA500 index) shows a significant leverage effect.

Overall, we find that the link between volatility and risk premia seems to be rather weak and changing over time. This is true for different ways of modelling risk premia and different econometric measures of volatility. Our results are consistent with other empirical and theoretical results (see e.g. Pagan and Hong (1991), Backus and Gregory (1989), Glosten, Jagannathan and Runkle (1989)) on the relation between conditional means and variances in asset returns. They suggest the need to explore for alternative empirical models for equity returns.

REFERENCES

Admati, A.R. and Pfleiderer, P. (1985): "Interpreting the Factor Risk Premia in the Arbitrage Pricing Theory", Journal of Economic Theory 35, 191-195.

Andrews, D.W.K. (1991): "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models", Econometrica 59, 307-346.

Attanasio, O. and Edey M. (1987): "Non-constant Variance and Foreign Exchange Risks: An Empirical Study", LSE Centre for Labour Economics Discussion Paper 285.

Attanasio, O. and Wadhwani S. (1990): "Can the CAPM explain why the Lagged Dividend Yield helps predict Stock Returns?", LSE Financial Markets Group Discussion Paper 104.

Baba, Y., Engle, R.F., Kraft, D.K. and Kroner, K. (1989): "Multivariate Simultaneous Generalized ARCH", Department of Economics Discussion Paper 89-57, University of California at San Diego.

Backus, D.K. and Gregory, A.W. (1989): "Theoretical Relations between Risk Premiums and Conditional Variances", mimeo, Federal Reserve Bank of Minneapolis.

Bera, A.K. and S. Lee (1990): "On the Formulation of a General Structure for Conditional Heteroskedasticity", mimeo, University of Illinois at Urbana-Champaign.

Bera, A.K., Higgins, M.L. and Lee, S. (1991): "Interaction Between Autocorrelation and Conditional Heteroskedasticity: A Random Coefficients Approach", mimeo, University of Illinois at Urbana-Champaign.

Black, F. (1976): "Studies of Stock Market Volatility Changes", Proceedings of the 1976 Business Meeting of the Business and Economic Statistics Section, American Statistical Association, 177-181.

Black, F. (1988): "An Equilibrium Model of the Crash", NBER Macroeconomics Annual, 269-276.

Black, F. (1989): "Mean Reversion and Consumption Smoothing", NBER Working Paper 2946.

Bollerslev, T. (1986): "Generalised Autoregressive Conditional Heteroskedasticity", Journal of Econometrics 31, 307-327.

Bollerslev, T. (1987): "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return", Review of Economics and Statistics, 69, 542-547.

Bollerslev, T. (1990): "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Approach", Review of Economics and Statistics, forthcoming.

Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1990): "ARCH Modeling in Finance: A Selective Review of the Theory and Empirical Evidence", Journal of Econometrics, forthcoming.

Bollerslev, T., Engle, R.F. and Wooldridge, J.M. (1988): "A Capital Asset Pricing Model with Time Varying Covariances", Journal of Political Economy 95, 116-131.

Bougerol, P. and Picard, N. (1990): "Stationarity of GARCH Processes and Some Non-Negative Time Series", Journal of Econometrics, forthcoming.

Braun, P.A., Nelson, D.B. and Sunier, A.M. (1990): "Good News, Bad News, Volatilities and Betas", Graduate School of Business Working Paper 90-93, University of Chicago.

Brock, W.A., Hsieh, D.A. and LeBaron, B. (1990): "A Test for Nonlinear Dynamics and Chaos, MIT Press, forthcoming.

Burmeister, E. and McElroy, M.B. (1988): "Joint Estimation of Factor Sensitivities and Risk Premia for the Arbitrage Pricing Theory", Journal of Finance, 43, 721-735.

Campbell, J.Y. and Y. Hamao (1989): "Predictable Bond and Stock Returns in the United States and Japan. A Study of Long-Term Market Integration", LSE Financial Markets Group Discussion Paper 69.

Campbell, J. and Hentschel, L. (1991): "No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns", mimeo, Princeton University.

Campbell, J. and R.J. Shiller (1988): "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors", Review of Financial Studies, 1, 195-228.

Chamberlain, G. (1983): "Funds, Factors and Diversification in Arbitrage Pricing Models", Econometrica 51, 1305-1324.

Chamberlain, G. and Rothschild, M. (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets", Econometrica 51, 1281-1304.

Chou, R.Y. (1988): "Volatility Persistence and Stock Valuations: Some Empirical Evidence Using GARCH", Journal of Applied Econometrics 3, 279-294.

Chou, R.Y., Engle, R.F. and Kane, A. (1990): "On the Measurement of Risk Aversion with Time-varying Price of Volatility", Journal of Econometrics, forthcoming.

Clark, P.K. (1973): "A Subordinated Stochastic Process Model with Finite Variance with Speculative Prices", Econometrica 41, 135-155.

Cohen, K.J., Hawawini, G.A., Maier, S.F., Schwartz, R.A. and Whitcomb, D.K. (1980): "Implications of Microstructure Theory for Empirical Research on Stock Price Behaviour", Journal of Finance 35, 249-257.

Cohn, R.A. and D. Lessard (1981): "The Effect of Inflation on Stock Prices: International Evidence", Journal of Finance, 36, 277-289.

Cutler, D., Poterba, J.M. and Summers, L.H. (1989): "What Moves Stock Prices?", Journal of Portfolio Management Spring, 4-12.

Cutler, D., Poterba, J.M. and Summers, L.H. (1990): "Speculative Dynamics: International Comparisons", NBER Working Paper 3242

De Long, J.B., A. Shleifer, L.H. Summers and R. Waldman (1987): "The Economic Consequences of Noise Traders", NBER Working Paper 2395.

De Long, J.B., Shleifer, A., Summers, L.H. and Waldman, R. (1990): "Positive Feedback Investment Strategies and Destabilising Rational Speculation", Journal of Finance 45, 379-396.

Demos, A. and Sentana, E. (1991a): "Testing for GARCH Effects: A One-sided Approach", mimeo, London School of Economics.

Demos, A. and Sentana, E. (1991b): "Inequality Constraints on GQARCH Models", work in progress, London School of Economics.

Dhrymes, P.J., Friend, I. and Gultekin, N.B. (1984): "A Critical Reexamination of the Empirical Evidence on the Arbitrage Pricing Model", Journal of Finance 39, 323-346.

Diebold, F.X. (1986): "Modelling the Persistence of Conditional Variances: A Comment", Econometric Reviews, 5, 51-56.

Diebold, F.X. and Nerlove, M. (1989): "The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor ARCH Model", Journal of Applied Econometrics 4, 1-21

Dow, C.H. (1920): "Scientific Stock Speculation", The Magazine of Wall Street, New York.

Drost, F.C. and Nijman, T.E. (1990): "Temporal Aggregation of GARCH Processes", CENTER Discussion Paper 9066, Tilburg University.

Dunn, J.E. (1973): "A Note on a Sufficiency Condition for Uniqueness of a Restricted Factor Matrix", Psychometrika 38, 141-153.

Eicker, F. (1963): "Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions", The Annals of Mathematical Statistics 34, 447-456.

Engle, R.F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation", Econometrica 50, 987-1008.

Engle, R.F. (1987): "Multivariate ARCH with Factor Structures - Cointegration in Variance", mimeo, University of California at San Diego.

Engle, R.F. (1990): "Discussion: Stock Market Volatility and the Crash of 87", Review of Financial Studies 3, 103-106.

Engle, R.F. and Bollerslev, T. (1986): "Modelling the Persistence of Conditional Variances", Econometric Reviews 5, 1-50, 81-87.

Engle, R.F., Granger C.W.J. and Kraft, D.F. (1984): "Combining Competing Forecasts of Inflation using a Bivariate ARCH Model", Journal of Economic Dynamics and Control, 8, 151-165.

Engle, R.F., Hendry, D.F and Trumble, D. (1985): "Small-sample Properties of ARCH Estimators and Tests", Canadian Journal of Economics 18, 66-93.

Engle, R.F., Lilien, D.M. and Robins, R.P. (1987): "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model", Econometrica 55, 391-407.

Engle, R.F. and Ng, V.K. (1991): "Measuring and Testing the Impact of News on Volatility", Department of Economics Discussion Paper 91-12, University of California at San Diego .

Engle, R.F., Ng, V.M. and Rothschild, M. (1989): "A Factor ARCH-Model for Stock Returns", mimeo, Department of Economics, University of California, San Diego.

Engle, R.F., Ng, V.M. and Rothschild, M. (1990): "Asset Pricing with a Factor-ARCH Structure: Empirical Estimates for Treasury Bills", Journal of Econometrics 45, 213-237.

Fama, E.F. (1965): "The Behavior of Stock Market Prices", Journal of Business 38, 34-105.

Fama, E.F. and French, K.R. (1988a): "Permanent and Transitory Components of Stock Prices", Journal of Political Economy 96, 246-273.

Fama, E.F. and French, K.R. (1988b): "Dividend Yields and Expected Stock Returns", Journal of Financial Economics, 22, 3-26.

Fama, E.F. and G.W. Schwert (1977): "Asset Returns and Inflation", Journal of Financial Economics 11, 115-146.

Fang, K.T., Kotz, S. and Ng, K.W. (1990): Symmetric Multivariate and Related Distributions, Chapman and Hall, London.

Feldman, R.A. (1986): Japanese Financial Markets, MIT Press, Cambridge, Mass.

Fisher, L. (1966): "Some New Stock Market Indexes", Journal of Business 39, 191-225.

French, K.R. and Roll, R. (1986): "Stock Return Variances: The Arrival of Information and the Reaction of Traders" Journal of Financial Economics, 17, 5-26.

French, K., Schwert, G.W. and Stambaugh, R.F. (1987): "Expected Stock Returns and Volatility", Journal of Financial Economics, 19, 3-29.

Gallant, A.R. (1981): "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form", Journal of Econometrics 15, 211-244.

Gallant, R., Hansen, L.P. and Tauchen, G. (1990): "Using Conditional Moments of Asset Payoffs to Infer the Volatility of Intertemporal Marginal Rates of Substitution", Journal of Econometrics 45, 141-179.

Garbade, K.D. (1982): "Federal Reserve Margin Requirements: A Regulatory Initiative to Prevent Speculative Bubbles", in Crises in the Economic and Financial Structure (ed. P. Wachtel), Lexington Books, MA.

Glosten, L.R., Jagannathan, R. and Runkle, D. (1989): "Relationship between the Expected Value and the Volatility of Nominal Excess Returns on Stocks", Banking Research Center Working Paper 166, Northwestern University.

Goodhart, C.A.E. and Giugale, M. (1988): "From Hour to Hour in the Foreign Exchange Market", LSE Financial Markets Group Discussion Paper 33.

Gourieroux, C. and Monfort A. (1990): "Qualitative Threshold ARCH Models", Journal of Econometrics, forthcoming .

Gultekin, M.N., Gultekin, N.B. and Penati, A. (1989): "Capital Controls and International Capital Market Segmentation: The Evidence from the Japanese and American Stock Markets", Journal of Finance 44, 849-870.

Hansen, L.P. (1982): "Large Sample Properties of the Generalized Method of Moments Estimators", Econometrica 50, 1029-1054.

Hansen, L.P. and Jagannathan, R. (1991): "Implications of Security Market Data for Models of Dynamic Economies", Journal of Political Economy 99, 225-262.

Hansen, L.P. and Richard, S.F. (1987): "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models", Econometrica 55, 587-613.

Hardouvelis, G.A. (1989): "Margin Requirements, Volatility, and the Transitory Component of Stock Prices", American Economic Review 79, 736-62.

Harvey, A.C. (1981): The Econometric Analysis of Time Series, Phillip Allan, London.

Harvey, A.C. (1989): Forecasting, Structural Models and the Kalman Filter, Cambridge University Press, Cambridge.

Harvey, A.C., Ruiz, E. and Sentana, E. (1990): "Unobservable Component Time Series Models with ARCH Disturbances", Journal of Econometrics, forthcoming.

Hausman, J. (1978): "Specification Tests in Econometrics", Econometrica, 46, 1251-1272.

Hendry, D.F. (1986): "An Excursion into Conditional Varianceland", Econometric Reviews, 5: 63-69.

Hentschel, L. (1991): "Alternative Models of Asymmetric Volatility in Stock Returns", mimeo, Princeton University.

Higgins, M. and Bera, A. (1990): "A Class of Nonlinear ARCH Models", International Economic Review, forthcoming February 1992.

Householder, A.S. (1964): The Theory of Matrices in Numerical Analysis, Blaisdell Publishing Company, London.

Hsieh, D.A. and Miller, M.H. (1990): "Margin Requirements and Stock Market Volatility", Journal of Finance 45, 3-29.

Ito, T. and Roley, V. (1986): "News from the US and Japan: Which moves the Yen/Dollar Exchange Rate", NBER Working Paper 1853.

Jennrich, R.I. (1978): "Rotational Equivalence of Factor Loading Matrices with Specified Values", Psychometrika 43, 421-426.

Johnson, N.L. and Kotz, S. (1972): Distributions in Statistics: Continuous Multivariate Distributions, Wiley, New York.

Johnson, R.A. and Wichern, D.W. (1982): Applied Multivariate Statistical Analysis, Prentice-Hall, New Jersey.

King, M.A., E. Sentana, and S.B. Wadhwani (1990): "A Heteroskedastic Factor Model of Asset Returns and Risk Premia with Time Varying Volatility: An Application to Sixteen World Stock Markets", LSE Financial Markets Group Discussion Paper 80.

King, M.A. and Wadhwani, S.B. (1990): "Transmission of Volatility between Stock Markets", Review of Financial Studies 3, 5-33.

Koenker, R. (1981): "A Note on Studentizing a Test for Heteroskedasticity", Journal of Econometrics 17, 107-112.

Kraft, D. and Engle, R.F. (1983): "ARCH in Multiple Time-series Models", mimeo, University of California at San Diego.

Kroner, K.F. (1987): "Estimating and Testing for Factor ARCH", mimeo, University of California at San Diego.

LeBaron, B. (1990): "Some Relations Between Volatility and Serial Correlation in Stock Markets Returns", mimeo, University of Wisconsin, Madison.

Lin, W. L. (1991): "Alternative Estimators for Factor GARCH Models: A Monte Carlo Comparison", mimeo, University of Wisconsin, Madison.

Lo, A.W. and Mackinlay, A.C. (1990): "An Econometric Analysis of Non-synchronous Trading", Journal of Econometrics 45, 181-212.

Magnus, J.R. (1988): Linear Structures, Griffin's Statistical Monographs and Courses no 41, Charles Griffin, London.

Magnus, J. and Neudecker, H. (1988): Matrix Differential Calculus with Applications in Statistics and Econometrics, Wiley, Chichester.

Marcus, A.J. (1989): "An Equilibrium Theory of Excess Volatility and Mean Reversion in Stock Market Prices", NBER Working Paper 3106.

Merton, R.C. (1980): "On Estimating the Expected Returns on the Market", Journal of Financial Economics 8, 323-361.

Milhoj, A. (1987): "A Multiplicative Parameterization of ARCH Models", mimeo, University of Copenhagen.

Modigliani, F. and R. Cohn (1979): "Inflation, Rational Valuation and the Market", Financial Analysts Journal March/April, 24-44.

Nelson, D.B. (1990a): "Stationarity and Persistence in the GARCH(1,1) Model", Econometric Theory 6, 318-334.

Nelson, D.B. (1990b): "Filtering and Forecasting with Misspecified ARCH Models I: Getting the Right Variance with the Wrong Model", forthcoming, Journal of Econometrics

Nelson, D.B. (1991): "Conditional Heteroskedasticity in Asset Returns: A New Approach", Econometrica 59, 347-370.

Nelson, D.B. and Cao, C.Q. (1991): "A Note on Inequality Constraints in the Univariate GARCH Model", mimeo, University of Chicago.

Nelson, D.B. and Foster, D.P. (1991): "Estimating Conditional Variances with Misspecified ARCH Models: Asymptotic Theory", mimeo, Graduate School of Business, University of Chicago.

Newey, W.D. (1985a): "Maximum Likelihood Specification Testing and Conditional Moments", Econometrica 53, 1047-1070.

Newey, W.K. (1985b): "Generalized Method of Moments Specification Testing", Journal of Econometrics 29, 229-256.

Newey, W.K. and West, K.D. (1987): "A Simple, Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", Econometrica 55, 703-708.

Nijman, T. and Sentana, E. (1991): "Contemporaneous Aggregation of GARCH Processes", work in progress, Tilburg University and London School of Economics.

Pagan, A. (1984): "Econometric Issues in the Analysis of Regressions with Generated Regressors", International Economic Review 25, 221-247.

Pagan, A.R. and Hong, Y. (1991): "Non-Parametric Estimation and the Risk Premium", in Semiparametric and Nonparametric Methods in Econometrics and Statistics (W. Barnett, J. Powell and G. Tauchen, eds.), Cambridge University Press, Cambridge.

Pagan, A.R. and Schwert, G.W. (1990): "Alternative Models for Conditional Stock Volatility", Journal of Econometrics 45, 267-290.

Pagan, A. and Ullah, A. (1988): "The Econometric Analysis of Models with Risk Terms", Journal of Applied Econometrics 3, 87-105.

Pantula, S.G. (1986): "Modeling the Persistence of Conditional Variances: A Comment", Econometric Reviews 5, 71-73.

Pindyck, R.S. (1984): "Risk, Inflation and the Stock Market", American Economic Review, 74, 335-351.

Poterba, J.M. and L.H. Summers (1988): "Mean Reversion in Stock Prices; Evidence and Implications", Journal of Financial Economics, 22, 27-60.

Prucha, I.R. and Kelejian, H.H. (1984): "The Structure of Simultaneous Equation Estimators: A Generalization towards Nonnormal Disturbances", Econometrica 52, 721-736.

Rich, R.W., Raymond, J. and J.S. Butler (1991): "Generalised Instrumental Variables Estimation of ARCH Models", Economic Letters 35, 179-185.

Robinson, P.M. (1983): "Non-parametric Estimators for Time Series", Journal of Time Series Analysis 4, 185-207.

Robinson, P.M. (1987): "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form", Econometrica 55, 857-889.

Robinson, P.M. (1988): "Semiparametric Econometrics: A Survey", Journal of Applied Econometrics 3, 35-51.

Robinson, P.M. (1991): "Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression", Journal of Econometrics 47, 67-84.

Roll, R. (1989): "Price Volatility, International Market Links and Implications for Regulatory Policies", Journal of Financial Services Research 3, 211-246.

Roll, R. and Ross, S.A. (1980): "An Empirical Investigation of the Arbitrage Pricing Theory", Journal of Finance 35, 1073-1103.

Ross, S. (1976): "The Arbitrage Theory of Capital Asset Pricing", Journal of Economic Theory, 13, 341-360.

Ross, S. (1977): "Return, Risk and Arbitrage", in I.Friend and J.L. Bicksler, eds, Risk and Return in Finance I, 189-218, Cambridge Mass, Ballinger.

Salinger, M.A. (1989): "Stock Market Margin Requirements and Volatility: Implications for Regulations of Stock Index Futures", Journal of Financial Services Research, 3, 121-138.

Schwallie, D.P. (1982): "Unconstrained Maximum Likelihood Estimation of Contemporaneous Covariances", Economic Letters 9, 359-364.

Schwert, G.W. (1989a): "Business Cycles, Financial Crises and Stock Volatility", Carnegie-Rochester Conference on Public Policy 31, 83-126.

Schwert, G.W. (1989b): "Why does Stock Market Volatility Change over Time?", Journal of Finance, 44, 1115-1153.

Schwert, G.W. (1990): "Indexes of United States Stock Prices", Journal of Business, 63, 399-426.

Sentana, E. (1989): "The Likelihood Function of a Heteroskedastic Latent Factor Model with Heywood Cases", mimeo, London School of Economics.

Sentana, E. (1991a): "Identification of Multivariate Conditionally Heteroskedastic Latent Factor Models", mimeo, London School of Economics.

Sentana, E. (1991b): "Quadratic ARCH Models: A Potential Re-interpretation of ARCH Models", LSE Financial Markets Group Discussion Paper 122.

Sentana, E. (1991c): "Factor Representing Portfolios on Large Asset Markets", mimeo, London School of Economics.

Sentana, E. and S.B. Wadhwani (1989): "Semi-Parametric Estimation and the Predictability of Stock Returns: Some Lessons from Japan", LSE Financial Markets Group Discussion Paper 62.

Sentana, E. and S.B. Wadhwani (1991): "Semi-Parametric Estimation and the Predictability of Stock Returns: Some Lessons from Japan", Review of Economic Studies 58, 547-562.

Sentana, E. and Wadhwani, S.B. (1990): "Feedback Traders and Stock Return Autocorrelations: Evidence from a Century of Daily Data", LSE Financial Markets Group Discussion Paper 90.

Sentana, E. and Wadhwani, S.B. (1991): "Feedback Traders and Stock Return Autocorrelations: Evidence from a Century of Daily Data", Economic Journal, forthcoming March 1992.

Shiller, R.J. (1984): "Stock Prices and Social Dynamics", Brookings Papers on Economic Activity 2, 457-498.

Shiller, R.J. (1987): "Investor Behavior around the October 1987 Stock Market Crash: Survey Evidence", NBER Working Paper 2446.

Stock, J.H. and Watson, M.W. (1988): "Testing for Common Trends", Journal of the American Statistical Association 83, 1097-1107.

Suzuki, Y. (1987) (ed.): The Japanese Financial System, Clarendon Press, Oxford.

Tauchen, G. and Pitts, M. (1983): "The Price Variability-Volume Relationship on Speculative Markets", Econometrica 51, 485-505.

Taylor, S. (1986): Modelling Financial Time Series, John Wiley and Sons, New York.

Thaler, R. (1987): "Anomalies: Weekend, Holiday, Turn of the Month, and Intraday Effects", Journal of Economic Perspectives 1, 169-178.

Ullah, A. (1988): "Non-parametric Estimation of Econometric Functionals", Canadian Journal of Economics 21, 625-658.

Von Furstenberg, G.M. and Nam Jeon, B. (1989): "International Stock Price Movements: Links and Messages", Brookings Papers on Economic Activity I, 125-180.

Weiss, A. (1986): "Asymptotic Theory for ARCH Models: Estimation and Testing", Econometric Theory 2, 107-131.

White, H. (1980a): "A Heteroskedastic-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", Econometrica 48, 817-838.

White, H. (1980b): "Using Least Squares to Approximate Unknown Regression Functions", International Economic Review 21, 149-170.

White, H. (1982): "Instrumental Variables Regression with Independent Observations", Econometrica 50, 483-499.

Wooldridge, J.M. (1988): "A Unified Approach to Robust Regression Based Specification Tests", mimeo, MIT Department of Economics.

Yancey, T.A., Judge, G.G. and Bock, M.E. (1981): "Testing Multiple Equality and Inequality Hypothesis in Economics", Economic Letters 7, 249-255.

Zakoian, J.M. (1990): "Threshold Heteroskedastic Model", Discussion Paper - INSEE.

Zellner, A. (1971): An Introduction to Bayesian Inference in Econometrics, Wiley, New York.