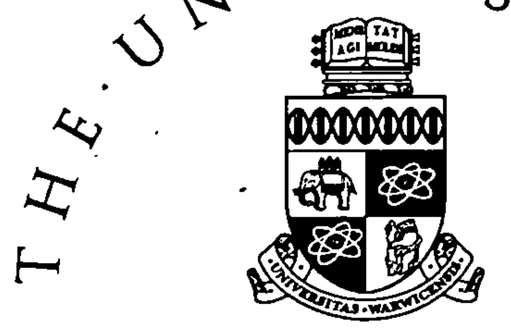


Density Forecasting in Financial Risk Modelling

by

Mascia Bedendo

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

FORC - Warwick Business School

September 2003

Contents

List of Tables vi

List of Figures viii

Acknowledgments xi

Declarations xii

Abstract xiii

Abbreviations xiv

Chapter 1 Introduction 1

Chapter 2 The Estimation of Option-Implied Distributions 5

2.1 Introduction 5

2.2 Related Literature 8

2.2.1 Parametric techniques 10

2.2.2 Non-parametric techniques 11

2.2.3 Comparative studies 13

2.3 The Estimation Techniques 14

2.3.1 Cubic B-splines for the volatility curve 14

2.3.2 Mixture of two lognormal densities 16

2.3.3 Normal Inverse Gaussian 18
2.4 The Data Set 20
2.5 Implied RND Estimates: a Discussion of the Results 22
2.5.1 Pricing errors and implied volatility fit 22
2.5.2 Implied RND shapes and summary statistics 24
2.6 Conclusions 27

Chapter 3 Are Option-Implied Distributions Unbiased Forecasts? 36

3.1 Introduction 36
3.2 Related Literature 38
3.3 The Forecasting Ability of Implied RNDs: Testing Techniques 44
3.3.1 Testing the uniformity of z_t series 46
3.3.2 Testing the normality of x_t series 48
3.3.3 Testing the tails of the implied RNDs 50
3.4 Are Implied Densities Unbiased Predictors? A Discussion of the Results 51
3.4.1 RNDs from currency futures options 52
3.4.2 RNDs from index futures options 53
3.4.3 A summary of the results 57
3.5 Assessing the Relative Forecasting Performance of Implied Densities: a Comparison with Historical-Based Density Forecasts 59
3.6 Conclusions and Further Work 64

Chapter 4 The Dynamics of the Volatility Skew: a Kalman Filter Approach 82

4.1 Introduction 82
4.2 A Review of the Literature 85
4.3 The Data Set 90
4.4 The Kalman Filter Model for the Dynamics of the Volatility Curve . 91

4.5 Two Alternative Models for the Dynamics of the Volatility Curve . . 94
4.5.1 The sticky-delta model 94
4.5.2 The vega-gamma model 95
4.6 The Estimation of the Models 97
4.7 The Option Portfolios 100
4.8 Density Forecasts of the Changes in Option Portfolios' Value 101
4.8.1 Derivation of the density forecasts 102
4.8.2 Assessment of the density forecasts 103
4.9 Conclusions and Further Research 111

Chapter 5 A Parsimonious Continuous Time Model of Equity Index Returns (Inferred from High Frequency Data) 132

5.1 Introduction 132
5.2 The Informative Content of High Frequency Data 135
5.3 The Data Set 139
5.4 Data Analysis and Derivation of the Model 141
5.4.1 The seasonal component 142
5.4.2 The stochastic volatility component 145
5.4.3 The estimation of the model at an intraday level 157
5.5 The Appraisal of Intraday Volatility and Density Estimates 166
5.6 A Monte Carlo Simulation Exercise 169
5.7 Conclusions and Further Work 172

Chapter 6 Multivariate Density Tests for Risk Management: an Empirical Characteristic Function Approach 196

6.1 Introduction 196
6.2 A Review of the Related Literature 199
6.3 The Empirical Characteristic Function Goodness-of-Fit Techniques . 202

6.3.1 The continuous "Weight Function" (WF) approach 202
6.3.2 The discrete "Fixed Grid Points" (FGP) approach 205
6.4 Size and Power of the ECF Tests: Some Monte Carlo Experiments . 208
6.4.1 General framework 208
6.4.2 Size of the tests 210
6.4.3 Bias in the first moment 210
6.4.4 Bias in the second moment 211
6.4.5 Misspecification of the form of the density forecast 213
6.4.6 Bias in higher moments 214
6.5 Conclusions 215

Chapter 7 Conclusions and Directions for Future Research 224

Appendix A Intraday Volatility Estimates 228

A.1 Decomposition of the variance of high frequency volatility innovations 228
A.2 Transition probabilities for the high frequency volatility process . . . 229
A.2.1 Diffusion model 230
A.2.2 Jump diffusion model 230

Appendix B Multivariate Continuous ECF Test Statistics 232

List of Tables

2.1 Summary statistics for implied RND parameters. 29

2.2 Pricing errors for implied RND estimates. 29

2.3 Summary statistics for actual continuously compounded log returns and implied distributions. 30

3.1 Tails Tests. 66

3.2 Tails Tests (cont'd). 67

3.3 Uniform distributional tests - $H_0 : z \sim$ i.i.d. $U(0,1)$. 68

3.4 Uniform distributional tests (cont'd) - $H_0 : z \sim$ i.i.d. $U(0,1)$. 69

3.5 Normal distributional tests - $H_0 : z \sim$ i.i.d. $N(0,1)$. 70

3.6 Normal distributional tests (cont'd) - $H_0 : z \sim$ i.i.d. $N(0,1)$. 71

3.7 Average parameter values for estimated GARCH distributions. . . . 72

3.8 Tails Tests for GARCH-based distributional forecasts. 73

3.9 Uniform distributional tests for GARCH-based distributional forecasts. 74

3.10 Normal distributional tests for GARCH-based distributional forecasts. 75

4.1 Estimates of A_t and μ_t - Kalman filter model. 113

4.2 Estimates of correlation matrix and standard deviations of system equation errors - Kalman filter model. 114

4.3 Estimates of measurement error volatility per buckets of log moneyness - Kalman filter model. 115
4.4 Estimates of EGARCH(1,1) model with normal errors. 116
4.5 Estimates of GARCH(1,1) model with normal errors. 117
4.6 Density forecasts tests - $H_0 : x_t \sim$ i.i.d. $N(0,1)$. 118
4.7 Summary statistics for actual profits and losses and density forecasts. 119
4.8 Percentage of VaR breaks. 120

5.1 Summary sample statistics for intraday returns. 175
5.2 First order serial correlation of intraday returns. 175
5.3 Tests for equality of seasonal volatility patterns. 176
5.4 Coefficient estimates for ARFIMA(p,d,q) models. 176
5.5 Coefficient estimates for two-factor AR(1) model. 177
5.6 Intraday volatility estimates evaluation. 177
5.7 Distributional forecast evaluation. 178
5.8 Estimates from simulated samples. 179

6.1 Critical values for continuous Weight Function testing procedures. . 218
6.2 Size of the tests - $H_0 : N(0, I_d)$ vs $H_1 : N(0, I_d)$. 219
6.3 Bias in the mean - $H_0 : N(0, I_d)$ vs $H_1 : N(\mu, I_d)$. 220
6.4 Bias in the covariance - $H_0 : N(0, I_d)$ vs $H_1 : N(0, \Sigma)$. 221
6.5 Misspecification of the density forecast - $H_0 : N(0, I_d)$. 222
6.6 Bias in higher moments. 223

List of Figures

2.1 Implied volatility curves from currency futures options: December 93 contract. 31

2.2 Implied volatility curves from index futures options: December 93 contract. 32

2.3 Implied RNDs from currency futures options. 33

2.4 Implied RNDs from index futures options (pre- and post-crash). . . 34

2.5 Time series of higher moments for parametric implied RNDs - currency (left) and index (right) futures contracts. 35

3.1 Graphical assessment of goodness-of-fit of risk-neutral NIG densities implied from currency futures options: entire (top) and truncated (bottom). 76

3.2 Graphical assessment of goodness-of-fit of risk-neutral NIG densities implied from index futures options: entire (top) and truncated (bottom). 77

3.3 Graphical assessment of goodness-of-fit of mean-shifted risk-neutral NIG densities implied from index futures options: entire (top) and truncated (bottom). 78

3.4 Graphical assessment of goodness-of-fit of risk-adjusted NIG densities implied from index futures options: entire (top) and truncated (bottom). 79
3.5 Graphical assessment of goodness-of-fit of historical-based density forecasts for currency futures: entire (top) and truncated (bottom). . 80
3.6 Graphical assessment of goodness-of-fit of historical-based density forecasts for index futures: entire (top) and truncated (bottom). . . 81

4.1 Cubic fit to the implied volatility curve for three expiries of the March 99 contract. 121
4.2 Time series of the volatility skew coefficients (Jan.98 - Dec.00). . . . 122
4.3 Autocorrelation functions of the volatility skew coefficients (Jan.98 - Dec.00). 123
4.4 Histograms of density forecasts on 20/8/2001 - short straddle. 124
4.5 Histograms of density forecasts on 20/8/2001 - long risk reversal. . . 125
4.6 Histograms of density forecasts on 20/8/2001 - long butterfly spread. 126
4.7 Histograms of density forecasts on 20/8/2001 - long Mexican hat. . . 127
4.8 CDF plots of goodness-of-fit - short straddle. 128
4.9 CDF plots of goodness-of-fit - long risk reversal. 129
4.10 CDF plots of goodness-of-fit - long butterfly spread. 130
4.11 CDF plots of goodness-of-fit - long Mexican hat. 131

5.1 Intraday patterns in returns and volatility. 180
5.2 Seasonal (normalised) volatility patterns for days of the week and different subsamples. 181
5.3 Smoothed B-spline estimation of seasonal coefficients. 182
5.4 ACF of intraday absolute returns and daily average absolute returns. 183
5.5 Coefficient of variation for the daily volatility measure - Time Series. 184

5.6 ACF of residuals from AR(1) and two-factor AR(1) specifications. . 185
5.7 Sample ACF vs. theoretical ARFIMA ACF. 186
5.8 Histograms of residuals from state and measurement equations. . . . 187
5.9 Standard deviation and excess kurtosis of conditional returns by intraday intervals using daily updated (left) and 5-minute updated (right) volatility estimates. 188
5.10 QQ plots of conditional returns for some intraday intervals. 189
5.11 Intraday and interday serial correlation of absolute returns. 190
5.12 Persistence of jumps' impact on volatility measured in 5-minute intraday intervals. 191
5.13 Normal QQ plots - return density forecasts using changing intraday volatility without (left) and with (right) jumps in volatility. 192
5.14 Normal QQ plots - return density forecasts using constant intraday volatility. 193
5.15 Skewness and excess kurtosis for intraday simulated and observed returns. 194
5.16 Scatter plot and time series of leverage measure against log volatility proxy - Market data (left) and simulated sample (right). 195

Acknowledgments

I want to express all my gratitude to my supervisor, Prof. Stewart D. Hodges, for both his invaluable advice and his continuous encouragement and great enthusiasm that have stimulated me throughout the different stages of my work.

I would like to thank Peter Weigel, Russell Grimwood and Euan Sinclair for many insightful discussions. Thanks are also due to all my colleagues at FORC, as well as to all my friends who made my permanence at Warwick enjoyable. Financial support from Warwick Business School and FORC Corporate Members is gratefully acknowledged. I am also thankful to Prof. Robert Tompkins for providing me with option data.

For their patience in coping with their "lost sheep", I am deeply grateful to my parents. Last but not least, a special thank goes to "QQ" Lorenzo Liesch, who has borne with me during the preparation of this thesis.

Declarations

I declare that the content of the present thesis is entirely my own work and has not been submitted for a degree at another university.

The research work carried out in Chapters 2 and 3 is part of a collaborative research project, co-authored with Iliana Anagnou, Stewart D. Hodges and Robert Tompkins. The main results are reported in: Anagnou, Bedendo, Hodges and Tompkins *The Relation between Implied and Realised Probability Density Functions*, FORC Pre-print 03/123. Only my individual contribution to the research (estimation of option-implied densities with three approaches and assessment of their unbiasedness) has been included in this thesis.

The content of Chapter 5 has recently been submitted for publication to the International Journal of Theoretical and Applied Finance.

The research work in Chapter 6 led to another FORC working paper: Bedendo M. *Multivariate Distributional Tests based on the Empirical Characteristic Function Approach: a Comparison*, FORC Pre-print 00/112.

Abstract

As a result of an increasingly stringent regulation aimed at monitoring financial risk exposures, nowadays the risk measurement systems play a crucial role in all banks. In this thesis we tackle a variety of problems, related to density forecasting, which are fundamental to market risk managers. The computation of risk measures (e.g. Value-at-Risk) for any portfolio of financial assets requires the generation of density forecasts for the driving risk factors. Appropriate testing procedures must then be identified for an accurate appraisal of these forecasts.

We start our research by assessing whether option-implied densities, which constitute the most obvious density forecasts of the distribution of the underlying asset at expiry, do actually represent unbiased forecasts. We first extract densities from options on currency and equity index futures, by means of both traditional and original specifications. We then appraise them, via rigorous density forecast evaluation tools, and we find evidence of the presence of biases.

In the second part of the thesis we focus on modelling the dynamics of the volatility curve, in order to measure the vega risk exposure for various delta-hedged option portfolios. We propose to use a linear Kalman filter approach, which gives more precise forecasts of the vega risk exposure than alternative, well-established models.

In the third part we derive a continuous time model for the dynamics of equity index returns from a data set of 5-minute returns. A model inferred from high-frequency data will produce more precise density forecasts over the short time horizons typical of risk measures calculations.

The last part of our work deals with evaluating density forecasts of the joint distribution of the risk factors. We find that, given certain specifications for the multivariate density forecast, a goodness-of-fit procedure based on the Empirical Characteristic Function displays good statistical properties in detecting misspecifications of different nature in the forecasts.

Abbreviations

In order of appearance:

RDN Risk-neutral density

ATM At-the-money

ITM In-the-money

OTM Out-of-the-money

NIG Normal Inverse Gaussian

PIT Probability Integral Transform

LR Likelihood ratio

ACF Autocorrelation function

VaR Value-at-Risk

ECF Empirical Characteristic Function

WF Weight Function

FGP Fixed Grid Points

Chapter 1

Introduction

In the last decade, financial risk management has become a core activity for all banks, given both the high level of instability observed in the financial environment, and the imposition of regulatory capital requirements to cover financial risk exposures. A lot of attention has been dedicated, by both academics and practitioners, to the issue of deriving accurate risk measures for various kinds of portfolios of financial assets.

The main concern of a risk manager is to compute risk measures for any conceivable portfolio of financial assets and derivatives. For this purpose, he needs to: 1) specify models for the dynamics of the financial assets; 2) generate density forecasts for the driving risk factors of those models. Given the non-normality of the returns on the risk factors, the estimation of only the first two moments of the distribution is clearly insufficient, and a complete characterisation of the density of the future outcomes is needed. For this reason, as Tay and Wallis [2000] state in their excellent survey, density forecasting constitutes a crucial issue in risk management.

A density forecast is an estimate of the density function that, according to the forecaster, will characterise the future evolution of a random variable. Therefore, unlike point or interval forecasts, which restrict the attention to a single point or

percentile of the forecasted distribution, a density forecast provides an accurate and complete representation of the uncertainty associated with the changes in the values of the variable of interest.

For risk management purposes, the assessment of these density forecasts by means of appropriate testing techniques is equally crucial. In particular, the evaluation techniques should be able to detect misspefications of different nature in the density forecasts, in order to provide useful directions for their improvement.

Our research investigates a variety of problems, related to the creation and/or the appraisal of density forecasts, which are fundamental to market risk managers.

Chapters 2 and 3 deal with the very popular issue of deriving accurate density forecasts for the price of the underlying asset at expiry, from traded option prices. As we will see, the related literature has mainly focussed on the identification of the most appropriate estimation techniques. The assessment of the density forecasts has often been carried out on the basis of descriptive accounts of their behaviour around specific events of interest. In the first part of our analysis we estimate implied densities from both futures currency and futures index options according to various specifications. We suggest a non-traditional parametric form, the Normal Inverse Gaussian, as a good candidate for fitting implied densities. In the second part, we carefully assess the unbiasedness of our risk-neutral density forecasts by means of a complete battery of distributional forecast tests. For the futures equity index options, where a risk-premium exists, the evaluation extends to the corresponding risk-adjusted densities. A comparison is also drawn with historical-based density forecasts obtained from very simple GARCH-type specifications.

The modelling of the dynamics of the implied volatility curve, which is essential for an accurate measurement of the vega risk exposure for option portfolios, is the object of Chapter 4. As we will see, many recent related contributions agree in claiming that more than one factor (i.e. the underlying asset) are needed to correctly

model the dynamics of the volatility curve. In our work we suggest, for the first time, to model the evolution of the coefficients of the (cubic) volatility skew according to a linear Kalman filter technique. This approach presents the advantages of being easy to implement and capable of accounting for movements in level, slope, curvature and skewness of the volatility curve. We produce density forecasts of the vega risk exposure for four option portfolios sensitive to shifts of different nature in the volatility curve. These distributional forecasts are then carefully assessed, in comparison with both the actual changes in the portfolios' value, and the corresponding density forecasts obtained from two benchmark models, Derman's [1999] sticky-delta and a vega-gamma approximation.

In Chapter 5 we propose a parsimonious specification for a continuous time model capable of describing the dynamics of equity index returns at different time frequencies. As we will discuss more in detail later on, several related works in the literature prefer to specify a model in all its components *a priori*, and then use the data (usually daily) to estimate and test it. Instead, we attempt to infer the features of our model from the empirical analysis of a data set of 5-minute returns on the S&P500 futures contract. Since risk measures are calculated over short horizons,[1] we believe that the density forecasts obtained from a model estimated on high frequency data are more precise than those generated from a model estimated on daily data. Our model is first "roughly" derived on a daily basis, and then refined at a high frequency level by means of a simple non-linear filtering technique. Through this Bayesian technique, we also generate 5-minutes ahead volatility estimates and density forecasts of the distribution of the intraday returns, whose accuracy is carefully assessed via, respectively, point and density forecast tests.

The issue of assessing multivariate density forecasts is addressed in Chapter 6. Despite its importance for risk management purposes, where joint density

[1]The standard temporal horizon for the Value-at-Risk measure is 10 days.

forecasts of the different risk factors need to be specified and evaluated, not much attention has been dedicated to the solution of this problem. Most research studies in this area either resort to multidimensional versions of univariate density forecast evaluation tests, or focus on reducing the dimensionality of the problem. In our work, we investigate whether multivariate goodness-of-fit tests based on the Empirical Characteristic Function could represent good candidates for the appraisal of multidimensional density forecasts within a risk management context. We assess the statistical power of both a continuous and a discrete version of the test, at detecting misspecifications of different kind in the density forecasts, for different dimensions, when the distributional assumptions are not too far from gaussian.

The various applications of density forecasting to risk management explored in this work are quite different from each other. Consequently, the corresponding related literatures have very little (if anything) in common. For clarity and consistency of exposition, we then prefer to review the relevant literatures separately within each chapter, instead of summarising all the existing contributions of interest in a single chapter.

Chapter 2

The Estimation of Option-Implied Distributions

2.1 Introduction

The activity of risk management entails a continuous and attentive measurement and monitoring of financial risks. In this context, the availability of reliable forecasts of the future distribution of financial returns turns out to be of crucial importance for the implementation of accurate risk management strategies. Similarly, financial investors and policy-makers need information regarding future states of the world, the former in order to refine their portfolio allocation strategies, the latter to produce sensible forecasts of key financial and monetary variables (e.g. inflation) and, when appropriate, to promptly intervene in financial markets to avert market turbulence.

Whenever predictive information on future outcomes is necessary, the individual's perception of the future states of the world is compared to the market's one. In finance, the most natural source of forecasts of future return distributions is the options market: option prices reflect forward-looking risk-neutral distributions of financial asset prices and, for options traded on the market, such distributions

mirror the market's consensus on the probabilistic distributions of the underlying assets at expiry. When market prices are available for a sufficiently large set of European options on the same underlying asset, with the same time to maturity, but with different strike prices, it is possible to extract the risk-neutral probability density function (RND hereafter) of the underlying at expiry.

In the last decade, growing attention has been devoted by academics, practitioners and regulators to the identification of the most appropriate techniques to extract RNDs from option prices, as well as to the investigation of the information content of such densities. Besides conveying markets' expectations of future events, RNDs implied from traded option prices can be usefully employed to price any payoff of illiquid options with the same time to expiration.

A large number of authors have concentrated on the estimation issue, in an attempt to find methods for recovering the RNDs which achieve good flexibility in fitting (without overfitting) the estimated density to option prices, and which are easy to implement and stable for a realistic range of available prices. The process of assessment of implied RNDs is naturally driven by the particular application at hand. Most papers have investigated the changes in the shape of RNDs around specific events, such as the announcement of economic news or financial crashes. Only very few recent studies have focussed on a rigorous assessment of the forecasting power of the implied RNDs at predicting the future distribution of asset returns.

In this chapter and in the following one we give our contribution in this respect to the existing literature. We develop a thorough analysis of two series of non-overlapping distributions extracted quarterly from the prices of two very liquid futures option contracts, the US dollar/British pound and the S&P 500, over a period of 16 years. The present chapter deals with the estimation of the implied RNDs from currency and index futures option prices, which is conducted with one non-parametric and two parametric techniques. Two of the specifications chosen, the

non-parametric cubic B-spline for the volatility curve and the parametric mixture of lognormals, are fairly standard in the literature, whereas the parametric Normal Inverse Gaussian (NIG) distribution represents a non-traditional specification for the implied distribution of log returns on the underlying at expiry. A comparison amongst implied RNDs estimated according to the three techniques indicates that the NIG specification represents a very good candidate for fitting RNDs in terms of both goodness-of-fit and flexibility properties, as well as parsimony of the number of parameters to be recovered.

The relevance of the present chapter is twofold. A parametric NIG form is compared to alternative long-established specifications, as a valid choice for modelling implied RNDs. The density forecasts produced in this part constitute the input for the analysis carried out in the following chapter, aimed at performing a complete and statistically rigorous appraisal of the forecasting properties of implied densities,[1] also in comparison with the corresponding properties of historical density forecasts, by way of a whole battery of distributional tests.

The chapter is organised as follows. Section 2.2 presents the relevant literature on alternative approaches for estimating implied RNDs. Section 2.3 illustrates the parametric and non-parametric specifications chosen to model the RNDs. Section 2.4 describes the data set used in the present study. Section 2.5 discusses and compares the densities extracted from currency and index futures options. Section 2.6 summarises the main findings and concludes.

[1] A comparison of the summary statistics of the implied RNDs with those of the actual returns on the underlying is introduced already in this chapter, in order to guarantee a more meaningful evaluation of the different implied RND techniques.

2.2 Related Literature

Since the late 1980s, financial researchers have become increasingly interested and sophisticated in analysing the market expectations incorporated in option prices. Moving beyond the study of the nature of "volatility smiles" and "skews" (whose existence has been widely documented in the option literature, especially for index options after the stock market crash of 1987), many studies have investigated how to back out either the risk-neutral distribution of the underlying at expiry from options with a given time to maturity, or the entire stochastic process followed by the asset across all times from a set of options with various times to expiration.[2]

The extraction of risk-neutral probability distributions of the underlying asset at expiry from option prices is based on the result by Cox and Ross [1976], according to which the price of a European call option with strike K at time $t = 0$ can be obtained as the discounted value of its expected payoff at maturity T, computed under the risk-neutral probability density $f(S_T)$ as follows:

$$C(K) = \mathrm{e}^{-rT} \int_K^\infty (S_T - K) f(S_T)\, \mathrm{d}S_T \tag{2.1}$$

In a complete market one can deduce the entire risk-neutral probability distribution (or, equivalently, density) from a set of European option prices with the same time to maturity and this distribution turns out to be unique.

Breeden and Litzenberger [1978] derived these risk-neutral distribution and density functions of the asset at expiry by taking, respectively, the first and the second partial derivatives of the European call pricing formula in Eq. (2.1) with respect to the exercise price K:[3]

---

[2]The present study does not extend to the recovery of the entire asset process; the interested reader should refer to the articles by Rubinstein [1994], Derman and Kani [1994a, 1998], Jackwerth [1997], Barle and Cakici [1998], Britten-Jones and Neuberger [2000] for implied tree models and to the papers by Bakshi, Cao and Chen [1997], Malz [1996], Bates [1991] for parametric models. It is worth noticing that when the focus is on the whole stochastic process, the implied RNDs are obtained as a by-product of the exercise.

[3]An alternative, iterative, algorithm for the derivation of the state prices was suggested by Banz

$$
\begin{aligned}
F(S_T) &= 1 + \mathrm{e}^{rT} \frac{\partial C(K)}{\partial K} \\
f(S_T) &= \mathrm{e}^{rT} \frac{\partial^2 C(K)}{\partial K^2}
\end{aligned}
\tag{2.2}
$$

Given that option prices for a continuum of strikes cannot be observed in practice on the markets, the implementation of the result in Eq. (2.2) raises the need for interpolation techniques, to be applied either to the option prices or to the implied volatilities.[4]

Any of three methods can be chosen to recover RNDs from market option prices. A first approach involves the assumption of a particular parametric specification for the RND $f(S_T)$ in Eq. (2.1), then the resulting theoretical option prices are fitted to the market option prices over the available range of strikes by means of a non-linear least squares optimisation routine. The main issue of this approach regards the choice of an appropriate functional form, which should be both flexible enough to capture a sufficient range of plausible shapes, and parsimonious in the number of parameters to be estimated.

For the second method, a function of option prices across strikes is calibrated to the observed option prices. The RND is then derived via Eq. (2.2). This method is very hard to implement in practice given that option prices are bounded by no-arbitrage conditions, and that the value of the function approaches zero for very high exercise prices.

The third method is similar in spirit to the second one, but the function of option prices across strikes is derived after calibrating the implied volatility curve (as a deterministic function of the strikes) to the observed implied volatilities. Again,

---

and Miller [1978].

[4]Instead of resorting to interpolation, some studies approximate the partial derivatives in Eq. (2.2) with finite difference methods. However, this approach is not recommended since it does not allow either to smooth out the irregularities coming from the market prices or to model the tails of the implied RNDs.

the model in Eq. (2.2) is then applied to recover the RND. This approach is widely used in practice since approximating the volatility curve turns out to be numerically easy and stable and to produce very sensible estimates of the RNDs. Its main drawback is that it requires a large amount of good quality data, which is difficult to obtain, particularly for out-of-the-money options.

Following the classification suggested in the paper by Jackwerth [1999], we now proceed to review in detail the numerous parametric and non-parametric methods which have been proposed to extract implied RNDs from option prices.

2.2.1 Parametric techniques

One class of parametric techniques is represented by expansion methods, where series expansions around the normal or the lognormal distribution are employed in order to achieve a more flexible and richer shape for the RND, which is then fitted to the market option prices. This approach guarantees good flexibility but the non-negativity of RND values must be continuously checked as it can be easily violated for certain choices of the relevant parameters. Several types of series expansions have been proposed in the literature to produce appropriate correction terms capable of accounting for the skewness and the excess kurtosis in the implied RNDs. Jarrow and Rudd [1982], Corrado and Su [1996], Longstaff [1995], and Rubinstein [1998] all used Edgeworth expansions and then matched the first four cumulants. Abadir and Rockinger [1997] suggested a Kummer function adjustment to the normal density. Hermite polynomial expansions were investigated by Madan and Milne [1994] and Abken, Madan and Ramamurtie [1996a, 1996b]. Jondeau and Rockinger [1998] and Corrado and Su [1997] proposed the class of Gram-Charlier expansions.

As an alternative to applying corrections to two-parameter distributions, generalised distribution methods directly assume more general functional forms for the implied RNDs, with additional parameters to capture higher moments of the

distributions. Sherrick, Garcia and Tirupattur [1996] modelled implied RNDs with a Burr III distribution. Sherrick, Irwin and Forster [1992, 1996] proposed a Burr XII distribution for S&P 500 futures options. Aparicio and Hodges [1998] used a generalised beta function of the second kind, which is a very general specification including many popular distributions as special cases. Posner and Milevsky [1998] suggested the Johnson family of densities.

A third class of parametric techniques is given by mixture methods, which model implied risk-neutral distributions as mixtures of standard two-parameter distributions. A mixture of lognormals is the most common specification of this kind employed in the existing literature. Moreover, this specification turns out to be the functional form for implied RNDs preferred by the policy-makers of a large number of industrialised countries. Bahra [1996, 1997], Melick and Thomas [1997], Söderlind and Svensson [1997], Gemmill and Saflekos [2000] represent the most relevant examples of application of mixtures of two or three lognormal distributions. Ritchey [1990] also proposed to model the RND of log returns as a mixture of normal densities. The main advantage of the mixture approach is its high flexibility in modelling a good variety of shapes for the implied distributions. However, greater flexibility is achieved for specifications with a large number of parameters; as a result, when only option prices across a small range of strikes are available, the mixture model overfits the data.

2.2.2 Non-parametric techniques

Non-parametric methods seek to attain high flexibility in fitting the RNDs to observed option prices without having to specify a parametric form and by employing, instead, much more general functions. One strand of non-parametric techniques proposes extracting the RND from option prices by means of kernel regressions. The main limitations to the use of this method are the large amount of good quality data

required and the unsatisfactory treatment of missing data, where the resulting gaps are filled and modelled in a non-intuitive manner, instead of applying a more appropriate smooth interpolation across them. Rookley [1997] used a bivariate kernel estimator, whereas kernel estimators in five variables were suggested by Ait-Sahalia and Lo [1998].

In line with the Bayesian framework, maximum-entropy methods attempt to identify a "posterior" non-parametric distribution which is as close as possible (in the sense of maximisation of the cross-entropy measure) to a "prior" distribution while satisfying a number of constraints such as positivity of the probabilities and correct pricing of the options and the underlying. Relevant applications of these methods can be found in Buchen and Kelly [1996] and Stutzer [1996].

Curve-fitting methods are probably the broadest and most popular class of non-parametric techniques. In this approach flexible and smooth functions such as the class of polynomials or the more sophisticated and flexible splines are used to approximate either the function of implied volatilities against strikes or the RND itself. Fitting the volatility smile tends to be easier and more robust than fitting the entire RND directly, given that the latter is much more non-linear and irregular than the former. In his pioneering paper, Shimko [1993] fitted a quadratic polynomial to the volatility smile. He then derived firstly the option prices and secondly the implied RND according to Eq. (2.2), to which he appended lognormal tails. Malz [1997] also used a quadratic polynomial to fit implied volatilities across call option deltas instead of strikes. Rosenberg and Engle [1997] chose to fit a polynomial to the log of the smile. Campa, Chang and Reider [1998] employed cubic splines, whereas Aparicio and Hodges [1998] worked with cubic B-splines. The direct approximation of the risk-neutral distribution through minimisation of the distance from a prior distribution was investigated by Rubinstein [1994] and Jackwerth and Rubinstein [1996].

2.2.3 Comparative studies

Given the plethora of methods for recovering implied RNDs from market option prices, a number of papers in the literature have carried out comparisons amongst different techniques in order to verify whether some of them can be considered superior to others in terms of both flexibility in fitting the RNDs to option prices and robustness of the results also for data sets with a narrow range of option prices. Jackwerth and Rubinstein [1996] compared different curve-fitting methods aimed at approximating directly the implied RNDs. Jondeau and Rockinger [2000] compared the fit of two expansion methods and a mixture of lognormals for options on exchange rates. Campa, Chang and Reider [1998] contrasted three methods (mixture of lognormals, cubic splines and Rubinstein's [1994] minimal distance approach) for the purpose of fitting the volatility curve of options on exchange rates. Coutant, Jondeau and Rockinger [2001] employed three techniques (mixture of lognormals, Hermite polynomial expansion and maximum entropy) to fit options on bond futures. McManus [1999] compared the fit to Eurodollar options obtained from mixture of lognormals, Hermite polynomial approximations, maximum entropy methods for the implied RNDs and a jump diffusion specification for the entire stochastic process.

All these comparative studies agree in asserting that any reasonable parametric or non-parametric approach would yield sensible estimates for the RNDs, which in fact turn out to be fairly similar across the different techniques, unless option prices are available only for a very small range of strikes. In this case, parametric methods provide more accurate and robust estimates and should be preferred to non-parametric alternatives.

2.3 The Estimation Techniques

In the present work we use three methods to extract RNDs from both currency and index futures option prices. The choice of employing a variety of estimation techniques has been driven by two considerations. First, in order to support the Normal Inverse Gaussian distribution as an appropriate candidate for fitting implied RNDs, we need to draw a comparison with alternative, more standard specifications that will be used as benchmarks. Second, in the following chapter we will be interested in assessing whether the forecasting performance of our implied distributions depends upon the particular functional form assumed to model them. Each of the three estimation approaches will now be discussed in detail.

2.3.1 Cubic B-splines for the volatility curve

Following Shimko [1993], the non-parametric method we adopt consists in modelling the implied volatility curve first and then recovering the RND as the second derivative of the option pricing formula with respect to the strike price. However, here we choose to fit the volatility curve in a more flexible way, by means of a smoothed linear combination of cubic B-splines functions. As discussed in Aparicio and Hodges [1998], this specification guarantees both an excellent fit to the observed volatility smile and the degree of smoothness necessary to calculate the derivatives and, therefore, to recover the implied RNDs.

For our purposes, we look for a spline function which is as smooth as possible and still lies within a given tolerance level with respect to the actual data. Using the smoothing spline curve fitting algorithm proposed by Dierckx [1995], we estimate the cubic smoothing spline $s(K)$ that minimises the following function:

$$\varepsilon = \eta + \rho\delta \tag{2.3}$$

where:

$$\eta = \sum_{j=1}^{J} (s^{(3)}(\lambda_j^+) - s^{(3)}(\lambda_j^-))^2 \quad \text{and} \quad \delta = \sum_{i=1}^{n+m} (o(K_i) - O(s(K_i)))^2$$

η is a measure of smoothness defined as the size of the discontinuity jump of the third derivative of $s(K)$ at the interpolation knots λ_j. δ is an indicator of the fit to the $n+m$ (n calls and m puts) market option prices, denoted by $o(K_i)$, and ρ represents the penalty imposed on the roughness of the approximation.

An optimal ρ must be identified as a result of the minimisation routine such that δ is smaller than or equal to a pre-specified degree of tolerance. As the level of tolerance varies, the spline changes within the two extreme cases of the least-squares third order polynomial ($\rho = 0$) and the natural least-squares spline ($\rho = +\infty$). In our case, tolerance levels between 0.0001 and 0.0005 turn out to work particularly well.

A drawback of the spline interpolation technique (and, more in general, of non-parametric approaches) is that the tails of the RND outside the range of available exercise prices cannot be directly recovered.[5] The only information on the tails that can be extracted from option prices concerns the probability mass (i.e. the area under the density curve in each tail). A number of solutions have been proposed in the literature to approximate the tails by extrapolating either the volatility curve, the call price curve or the truncated RND itself. Since we believe that the attempt of fitting the tails is more an art than a science, we prefer not to extrapolate them. As we will see in the next chapter, we choose to perform the estimation and the appraisal of both the truncated density and the probability mass in each tail.

[5] In fact, the tails implicitly obtained from fitting a parametric specification are also approximated. However, if the family of the functional form chosen for modelling the RND is not too rich, there is normally a unique fit for the tails (see Aparicio and Hodges [1998]).

2.3.2 Mixture of two lognormal densities

Many studies favour the mixture of lognomal distributions for the parametric modelling of the distribution of the underlying asset at expiry, since its flexible specification allows us to approximate a wide range of shapes. We think that such a distributional choice is not particularly appropriate in this context, given the large number of parameters to be estimated and the consequent risk of overfitting the data. However, we have included it in our analysis for completeness and to provide a fairly standard, widely used parametric term of comparison for the Normal Inverse Gaussian.

We recall that the prices of European call and put options with strike K and time to maturity $\tau = T - t$, calculated at time t, can be written as follows:

$$C(K_i) = \mathrm{e}^{-r\tau} \int_{K_i}^{\infty} (S_T - K_i) f(S_T)\, \mathrm{d}S_T \tag{2.4}$$

$$P(K_j) = \mathrm{e}^{-r\tau} \int_{0}^{K_j} (K_j - S_T) f(S_T)\, \mathrm{d}S_T \tag{2.5}$$

for $i = 1, \ldots, n$ calls and $j = 1, \ldots, m$ puts. Here we assume that the density function of the underlying at expiry $f(S_T)$ is given by a mixture of two lognormal densities, that is:

$$f(S_T) = \sum_{p=1}^{2} [\theta_p \, L(S_T; \alpha_p, \beta_p)] \tag{2.6}$$

where $L(S_T; \alpha_p, \beta_p)$ is the p^{th} lognormal density with parameters α_p and β_p:

$$\alpha_p = \ln F_t + \left(\mu_p - \frac{1}{2}\sigma_p^2\right)\tau \quad \text{and} \quad \beta_p = \sigma_p \sqrt{\tau} \quad \text{for} \quad p = 1, 2.$$

Given the two-lognormal mixture assumption made for $f(S_T)$, Bahra [1997] derives closed-form solutions to the Eqs. (2.4) and (2.5), which can then be expressed as follows:

$$\begin{aligned} C(K_i) &= \theta[\mathrm{e}^{\alpha_1 + 0.5\beta_1^2} N(d_1) - K_i N(d_2)] + \\ &= +(1-\theta)[\mathrm{e}^{\alpha_2 + 0.5\beta_2^2} N(d_3) - K_i N(d_4)] \end{aligned} \tag{2.7}$$

$$
\begin{aligned}
P(K_j) &= \theta[-\mathrm{e}^{\alpha_1+0.5\beta_1^2}N(-d_1) + K_j N(-d_2)] + \\
&= +(1-\theta)[-\mathrm{e}^{\alpha_2+0.5\beta_2^2}N(-d_3) + K_j N(-d_4)] \qquad (2.8)
\end{aligned}
$$

where:

$$
\begin{aligned}
d_1 &= \frac{-\ln K + \alpha_1 + \beta_1^2}{\beta_1}, \qquad d_2 = d_1 - \beta_1 \\
d_3 &= \frac{-\ln K + \alpha_2 + \beta_2^2}{\beta_2}, \qquad d_4 = d_3 - \beta_2
\end{aligned}
$$

and $N(\cdot)$ denotes the standard normal cumulative probability distribution function.

The existence of closed-form expressions for the option pricing formulas represents a very nice feature of the mixture of lognormal densities, as it obviates the need for numerical integration in these equations and therefore, it ensures higher computational tractability and greater accuracy in the estimation of the parameters of the implied RNDs.

The estimates for the five parameters α_1, α_2, β_1, β_2, θ, are obtained by minimising the deviation of the theoretical option prices given by Eqs. (2.7) and (2.8) from the market prices (c_i and p_j), across the range of available strikes, under the constraint that the mean of the distribution equals the forward price of the underlying.[6] The total sum of squared errors for both call and put options is minimised by means of a non-linear least squares optimisation routine.[7] The minimisation problem then becomes:

$$
\min_{\alpha_1,\alpha 2,\beta_1,\beta_2,\theta} \sum_{i=1}^{n}[C(K_i) - c_i]^2 + \sum_{j=1}^{m}[P(K_j) - p_j]^2 \qquad (2.9)
$$

subject to:

$$
\theta \mathrm{e}^{\alpha_1+0.5\beta_1^2} + (1-\theta)\mathrm{e}^{\alpha_2+0.5\beta_2^2} = F_t
$$

for β_1, $\beta_2 > 0$ and $0 \leq \theta \leq 1$.

[6] In the present work the forward price is approximated by the current futures price F_t.

[7] The Levenberg-Marquardt algorithm is particularly appropriate for this kind of problem. As initial guess for the set of parameters we use the estimates of the parameters obtained for the previous quarter.

2.3.3 Normal Inverse Gaussian

Originally introduced by Barndorff-Nielsen [1977, 1978], the Normal Inverse Gaussian distribution (NIG hereafter) has achieved large popularity in the last few years. Many recent studies (see Barndorff-Nielsen [1997], Barndorff-Nielsen and Jiang [1998], Prause [1997], Rydberg [1997]) highlighted its ability to provide a very good fit for the distribution of financial log returns.[8]

The NIG constitutes a special case of the class of generalised hyperbolic densities and its density function is given by:

$$NIG(x;\alpha,\beta,\mu,\delta) = \pi^{-1}\alpha e^{\delta\sqrt{\alpha^2-\beta^2}-\beta\mu} q\left(\frac{x-\mu}{\delta}\right)^{-1} K_1\left[\delta\alpha q\left(\frac{x-\mu}{\delta}\right)\right] e^{\beta x} \quad (2.10)$$

where $q(x) = \sqrt{1+x^2}$, $\mu \in \mathrm{R}$, $\delta > 0$, $0 \leq |\beta| \leq \alpha$ and K_1 is the modified Bessel function of third order and index 1. The moment generating function of a NIG possesses a nice and neat expression:

$$M(u;\alpha,\beta,\mu,\delta) = \exp\left\{\delta\left[\sqrt{\alpha^2-\beta^2} - \sqrt{\alpha^2-(\beta+u)^2}\right] + \mu u\right\}$$

which allows a convenient way of mapping between the parameters of the distribution and the first four moments. As an intuitive explanation of the parameters, α represents the steepness, β is an asymmetry parameter, δ describes the scale and μ the location of the distribution. A further advantage of adopting a NIG specification lies in its property of being closed under convolution, unique within the generalised hyperbolic class.

Given its ability to account for crucial features of the distribution of financial asset returns, such as semi-heavy tails and asymmetry, and the parsimony in the number of parameters to be estimated, the NIG seems a particularly appropriate parametric specification for the implied RND of the log returns on the underlying at expiry. Some recent studies have proposed option pricing models where the

[8] It is worth emphasising that, unlike the mixture of lognormals, which is a distribution for the price of the underlying S_T, the NIG represents a specification for the log returns on the underlying $X_\tau = \ln(S_T) - \ln(S_t)$, given that its argument takes values on the entire real axis.

price of the underlying is driven by a Lévy process generated by the NIG distribution or other generalised hyperbolic distributions (see Eberlein and Prause [2002]). The resulting option prices and volatilities were then compared to the standard Black-Scholes' ones. However, a comparison between the results obtained from the estimation of implied RNDs via NIG, and the corresponding findings from alternative specifications (more realistic than the lognormal one dictated by Black-Scholes) has not been attempted so far.

Similarly to the mixture of lognormals case, the parameters of the implied NIG RND are estimated by minimising the sum of squared pricing errors for both call and put options, while constraining the mean of the distribution of the underlying at expiry to be equal to the forward price. For the NIG specification, the theoretical prices for call and put futures options are given by:

$$C(K_i) = \mathrm{e}^{-r\tau} \int_{\ln(K_i/F_t)}^{\infty} (F_t \mathrm{e}^{X_\tau} - K_i) NIG(X_\tau; \alpha, \beta, \mu, \delta)\, \mathrm{d}X_\tau \qquad (2.11)$$

$$P(K_j) = \mathrm{e}^{-r\tau} \int_{-\infty}^{\ln(K_j/F_t)} (K_j - F_t \mathrm{e}^{X_\tau}) NIG(X_\tau; \alpha, \beta, \mu, \delta)\, \mathrm{d}X_\tau \qquad (2.12)$$

for $i = 1, \ldots, n$ calls and $j = 1, \ldots, m$ puts. Unfortunately, no closed-form solution is available for the expressions in (2.11) and (2.12) and we have to resort to numerical integration techniques to calculate the model option prices to provide as input to the minimisation algorithm[9] that can then be represented as follows:

$$\min_{\alpha,\beta,\mu,\delta} \sum_{i=1}^{n} [C(K_i) - c_i]^2 + \sum_{j=1}^{m} [P(K_j) - p_j]^2 \qquad (2.13)$$

subject to:

$$F_t \exp\left\{\delta\left[\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta+1)^2}\right] + \mu 1\right\} = F_t$$

and the inequality constraints on the parameters as specified above.[10]

[9] As before, we use the Levenberg-Marquardt algorithm to solve the minimisation problem.

[10] The constraint on the mean of the distribution of the underlying at expiry can be derived in a straightforward manner from the MGF, since $E[F_t \mathrm{e}^{X_\tau}] = F_t[E\mathrm{e}^{X_\tau}] = F_t MGF(1; \alpha, \beta, \mu, \delta)$.

2.4 The Data Set

Estimates of the RNDs are extracted quarterly from the prices of call and put futures options on:

- Standard & Poor's 500 index;
- US dollar/British pound;

traded at the Chicago Mercantile Exchange over the period March 1986 - September 2001. Both contracts are highly liquid and traded across a wide range of strike prices. We choose to work with quarterly data and to derive non-overlapping RND estimates (one at the beginning of every quarter, for the contract expiring at the end of the same quarter), in order to ensure consistency with the requirement of independent observations in the Probability Integral Transform approach, which constitutes the building block of the testing techniques implemented in the following chapter. We identify all the expiration dates for the quarterly expiration cycle during the entire sample period (March, June, September and December). On each expiration date, we record the settlement levels of the futures contract expiring on that day, the futures contract with exactly three months to expiration and all available call and put option prices on this latter contract.

Since the options are American-style, the Barone-Adesi and Whaley [1987] approximation is used to recover the implied volatilities, which are then plugged back into the Black's formula to calculate the pseudo-European option prices needed to derive the RNDs.

All options which are traded at the minimum price level for the relevant market are excluded from the data set, as well as those which violate the put-call parity, the standard no-arbitrage conditions for futures options:

$$c_i \quad > \quad F_t - K_i$$

$$p_j \geq K_j - F_t$$

or the butterfly no-arbitrage relationships (for options with equally spaced strike prices):

$$c_i \leq (c_{i-1} + c_{i+1})/2$$

$$p_j \leq (p_{j-1} + p_{j+1})/2$$

Furthermore, much larger variations occur in the implied volatilities for in-the-money options relative to out-of-the-money options, which translate into more distorted and jagged smiles. This feature is attributable to a greater impact on in-the-money options of non-synchronicity errors which arise when prices for the options and the underlying futures are not recorded simultaneously. To reduce this problem we only consider out-of-the-money call and put options. Thus, put options are chosen for strikes smaller than, or equal to, the underlying futures price and call options for strikes higher than the futures price. The risk-free interest rate is the US Dollar Treasury Bill rate provided by the Federal Reserve Bank of New York.

RNDs are then estimated once a quarter on a set of out-of-the-money call and put options with three months to expiration. At expiry, such distributional forecasts will be compared to the actual realisations for the price of the underlying futures contract. Given that we have examined nearly 16 years of data on a quarterly basis, all together we are able to estimate and test 63 implied RND forecasts. In Table 2.1 we report average and dispersion values for the time series of parameters of implied parametric RNDs, for both currency and index futures options.

2.5 Implied RND Estimates: a Discussion of the Results

In this section we provide a comparison amongst the implied RNDs estimated with the two parametric and one non-parametric models described above, for both currency and index futures options, in order to: 1) identify which specifications seem more appropriate than others in terms of in-sample goodness-of-fit; 2) assess whether the NIG could represent a valid model for the implied densities. First, the models are compared according to their pricing errors (i.e. the differences between the theoretical and the observed option prices) and their fit to the implied volatility curve. Second, the two parametric specifications are contrasted by examining the shapes of the implied RNDs and by means of standard summary statistics. It is worth emphasising that here we do not attempt to compare the forecasting properties of the implied RNDs obtained with the different methods, which instead will be the object of the following chapter.

2.5.1 Pricing errors and implied volatility fit

As mentioned above, the raw pricing error is given by the difference between the theoretical option price from the model and the option price observed on the market. Given that we consider several options across a range of strikes on each quarter, and that our sample consists of 63 quarters, rather than examining the individual pricing errors, it is more meaningful to combine the raw pricing errors into a single quantity that measures the accuracy of fit for each model on each contract. The measures of fit we employ are the mean squared error (MSE) and the mean squared percentage pricing error (MSPE):

$$MSE = \frac{1}{n+m-k}\sum_{i=1}^{n}[c_i - C(K_i)]^2 + \frac{1}{n+m-k}\sum_{j=1}^{m}[p_j - P(K_j)]^2 \quad (2.14)$$

$$MSPE = \frac{1}{n+m-k}\sum_{i=1}^{n}\left[\frac{c_i - C(K_i)}{c_i}\right]^2 + \frac{1}{n+m-k}\sum_{j=1}^{m}\left[\frac{p_j - P(K_j)}{p_j}\right]^2 \quad (2.15)$$

where k is the number of independent parameters for the model under examination. The MSE places more weight on larger errors. The MSPE is a dimensionless measure and therefore, facilitates the comparison across different models and different contracts. MSE and MSPE are computed every quarter for each model, for both currency and index futures options. Average values of these measures are then calculated across the 63 quarters in the sample, to provide an overall indicator of accuracy of fit.

The results, displayed in Table 2.2, indicate that for both option contracts, the non-parametric cubic B-spline method ranks first, with an average MSPE of 0.0085 for currency options and 0.0047 for index options.[11] Between the two parametric methods, the NIG approach shows a better fit, with an average MSPE of 0.0272 for currency options and 0.0224 for index options, against average values of, respectively, 0.0606 and 0.0449 computed for the mixture of lognormals specification.

As expected, the visual inspection of the fit of the theoretical implied volatility curves to the implied volatilities extracted from market prices leads to very similar conclusions. Fig. 2.1 refers to the implied volatility fit for the December 93 currency futures option contract, and Fig. 2.2 displays the implied volatility curves for the December 93 contract on index futures options. No clear preference can be expressed for the currency futures options, given that the theoretical implied volatilities for all of the three models exhibit a very similar fit to the market values. Instead, the superiority of the implied volatility fit provided by the cubic B-spline and the NIG models over the mixture of lognormal specification is obvious for the index futures options.

[11] The cubic B-spline method directly estimates the volatility curve and not the option prices. The theoretical option prices for the model have been obtained simply by plugging the estimated implied volatilities into Black's pricing formula.

The analysis of pricing errors and implied volatility fit suggests that the non-parametric cubic B-spline method provides the best in-sample fit to the observed market values amongst our three specifications. This result is not too surprising, given the great flexibility that characterises non-parametric methods which is, however, usually accompanied by a concrete danger of overfitting the data. To investigate this point, Aparicio and Hodges [1998] calculated time series of the probability of several strike intervals for all dates available on a specific contract. The negative and significant autocorrelation found in the innovations of the time series obtained from the non-parametric approach (for all strike intervals and all contracts) was taken as indication of overfitting the market data. Due to the structure of our data set (one quarterly observation available for each contract), we cannot replicate their analysis here, however it is reasonable to believe that this problem affects our non-parametric estimates as well. For this reason, not much attention is dedicated to the cubic B-splines implied densities in this work, and the main role they play is to provide a matter of comparison in terms of in-sample goodness-of-fit for the alternative (parametric) specifications.

More interestingly, our findings support the NIG specification as an appropriate method for estimating implied RNDs, given its very good fit to the data, clearly superior to the one provided by the alternative parametric specification.

2.5.2 Implied RND shapes and summary statistics

In this section we carry on the comparison between our two parametric specifications, to ascertain whether the preference given to the NIG method on the basis of the accuracy of fit is actually supported by the estimation of implied RNDs with realistic shapes and sensible values for some basic summary statistics.

In Fig. 2.3 we plot the implied RNDs for the December 93 currency futures

contract, according to both the NIG and the mixture approach.[12] In this case, which can be extended to the generality of the currency futures, the shapes of the two implied densities are practically indistinguishable, almost symmetric and only slightly fat-tailed.

Almost identical shapes for the implied NIG and mixture densities are also observed for index futures before the 1987 crash (see Fig. 2.4, top), characterised by little negative skewness and very moderate fat-tailness. After the stock market crash, the form of the implied RNDs changes substantially, becoming remarkably negatively skewed and heavily fat-tailed. Also the shapes of estimated implied densities differ significantly according to the particular parametric form adopted (Fig. 2.4, bottom).

Differences in the shape of the implied RNDs are obviously reflected in their summary statistics. Table 2.3 contains summary statistics for the time series of the continuously compounded actual log returns over one quarter and for the forecasted densities of the log returns both extracted from option prices and simulated from GARCH-type models for the volatility (which will be discussed in the next chapter). The first part of the table refers to the currency contract and the second half to the index contract. For this latter contract, the statistics have been computed for both unadjusted RNDs and RNDs adjusted through a utility based transformation to account for the risk premium in the equity market, whose discussion is postponed to the next chapter. To provide a consistent comparison between the moments of the observed point realisations and those of the density forecasts, we proceed as follows. Sample conventional measures of location, dispersion, skewness and excess kurtosis are computed on realised log returns. Equivalent statistics are then calculated for the distribution obtained as an equally weighted mixture of the single

[12] To permit the comparison between the two methods, the implied RNDs refer to the log returns on the underlying at expiry. Therefore, the mixture of lognormal specification for the price translates in a mixture of normal densities for the log return.

density forecasts estimated for each quarter. This method seems more appropriate than simply averaging the summary statistics of the individual densities across time. However, simple averages of the time series of the statistics have been computed as well, and the average values for skewness and excess kurtosis are reported in the last two columns of the table.[13]

The RNDs from the currency futures options display very similar moments, whatever specific functional assumption is made. The distributions exhibit slightly negative skewness (around -0.10) and moderate excess kurtosis (around 2). As expected, these moments underestimate the sample moments of the observations when the September 1992 (EMU ejection of the British Pound) is included in the sample, which record a skewness of -0.40 and an excess kurtosis of 3.4. If this event is excluded, the implied kurtosis closely matches the empirical one, whereas the forecasted and realised skewness exhibit opposite sign.

For the S&P 500 futures options, implied NIG densities turn out to be more negatively skewed (skewness of -2.57) and much more leptokurtic (excess kurtosis of around 24) than the corresponding implied mixture of normal densities (skewness of -1.50 and excess kurtosis of 6). For either parametric specification, the mean of unadjusted RNDs is smaller than the sample mean, since it does not account for the risk premium. The implied dispersion measure closely matches the empirical one. The values for the implied higher moments instead always exceed the sample counterparts (skewness of -1.20 and excess kurtosis around 3), supporting the generally accepted conclusion that, for equity options, these moments are strongly model-dependent and tend to overestimate the skewness and the fat-tailness observed in the markets.

[13]The higher absolute figures obtained for skewness and fat-tailness for the mixture can be explained with Jensen's inequality: since higher statistics are computed as ratios of moments, the average of those ratios (method of time series) is smaller than the ratio of those average moments (mixture method).

To further investigate how pronounced and volatile higher moments are for the different specifications, we also report in Fig. 2.5 time series of the skewness and excess kurtosis for both currency and index RNDs. For implied RNDs on the index, higher moments become much more volatile after the 1987 crash, especially for the NIG specification. For the currency options, the two parametric models display very similar values for the higher moments, which become a bit more volatile after September 1992, but still remain much less pronounced and dispersed through time than those recorded for the index options.

The inspection of the implied RNDs shapes for both currency and index futures contracts provides additional evidence that the NIG represents a particularly appropriate specification for modelling density functions for log returns extracted from option prices.

2.6 Conclusions

In this chapter we have estimated and analysed time series of quarterly implied RNDs from both currency and index futures options, over nearly 16 years, according to both one non-parametric (cubic B-splines) and two parametric (mixture of lognormals and Normal Inverse Gaussian) specifications. We have drawn an extensive comparison amongst the different methods in terms of goodness-of-fit to the market values for option prices and volatilities, not only to identify the approach which provides the best fit, but also, and more importantly, to assess how the NIG density performs relative to alternative specifications and, thus, whether it could represent a valid functional form for modelling implied RNDs.

As expected, our empirical results indicate that the best in-sample fit is given by the cubic B-spline method, which guarantees great flexibility and an excellent fit, without imposing distributional assumptions. However, non-parametric approaches

present several drawbacks as they require a large amount of very good quality market input and they can easily overfit the data.

Amongst the parametric specifications, the NIG largely outperforms the mixture of lognormal densities in terms of goodness-of-fit to the observed option prices and volatilities. The empirical evidence of a very good fit to the data, together with its flexible specification and the small number of parameters to be estimated, support our conclusion that the NIG represents an excellent candidate for the parametric fitting of implied RNDs.

It is worth reminding that in this part we have compared alternative specifications in terms of their in-sample goodness-of-fit to the market values and not in terms of their out-of-sample ability to correctly forecast the actual distribution of the underlying at expiry. This aspect will be thoroughly investigated in the following chapter.

Motivated by our choice of not extrapolating the tails of the implied distributions, further work might explore how *ad hoc* this extrapolation can be, and quantify his distortive impact on the results.

Table 2.1: Summary statistics for implied RND parameters.

		Mixture of lognormals			**NIG**	
		Mean	**St.Dev.**		**Mean**	**St.Dev.**
US$/British Pound						
Overall	μ_1	7.3631	0.1043	α	46.0802	40.0090
	σ_1	0.0496	0.0314	β	-4.7392	15.2142
	μ_2	7.3591	0.0951	μ	0.1599	0.2473
	σ_2	0.0798	0.0272	δ	0.0142	0.0679
	θ	0.7104	0.1828			
S&P 500						
Pre-Crash '87	μ_1	5.5941	0.1098	α	20.8044	9.6114
	σ_1	0.0818	0.0183	β	-3.3220	2.6088
	μ_2	5.5264	0.1317	μ	0.1992	0.0831
	σ_2	0.1434	0.0506	δ	0.0326	0.0331
	θ	0.6847	0.0789			
Post-Crash '87	μ_1	6.4558	0.5372	α	21.6629	7.6972
	σ_1	0.0567	0.0164	β	-14.1291	6.7670
	μ_2	6.3105	0.5118	μ	0.0854	0.0345
	σ_2	0.1242	0.0385	δ	0.0688	0.0372
	θ	0.7860	0.0587			
Overall	μ_1	6.3029	0.5905	α	21.5105	7.9885
	σ_1	0.0612	0.0192	β	-12.2117	7.4817
	μ_2	6.1714	0.5556	μ	0.1056	0.0635
	σ_2	0.1276	0.0411	δ	0.0624	0.0388
	θ	0.7680	0.0732			

Table 2.2: Pricing errors for implied RND estimates.

	Currency options		**Index options**	
	MSE	MSPE	MSE	MSPE
Cubic B-spline	0.0992	0.0085	0.0358	0.0047
Mixture of lognormals	1.0533	0.0606	0.2366	0.0449
NIG	0.1406	0.0272	0.1751	0.0224

Table 2.3: Summary statistics for actual continuously compounded log returns and implied distributions.

			Mean	Std. Dev.	Skew	Exc. Kurt.	Sk.*	Exc. Kur.*
US$/UK£								
Sample (full)			0.0101	0.060	-0.402	3.393		
Sample (ex 9-92)			0.0123	0.052	0.349	1.240		
Implied RNDs		Mixture	-0.0017	0.059	-0.139	2.548	-0.091	1.214
		NIG	-0.0017	0.058	-0.095	1.839	-0.077	1.979
Simulated		GARCH n	3.02E-05	0.056	-0.001	0.768	0.002	0.358
		GARCH t	1.13E-05	0.057	-0.018	2.228	-0.010	1.857
S&P 500								
Sample (full)			0.0180	0.079	-1.206	3.241		
Implied RNDs	Unadj.	Mixture	-0.0053	0.105	-1.516	6.156	-1.125	2.371
		NIG	-0.0056	0.110	-2.571	24.029	-1.561	6.493
	Adj.	Mixture	0.0001	0.100	-1.334	5.768	-1.088	2.136
	$\gamma = 1$	NIG	-0.0002	0.105	-2.041	16.554	-1.493	5.521
	$\gamma = 2$	Mixture	0.0052	0.098	-1.203	5.152	-1.070	2.431
		NIG	0.0052	0.099	-1.553	9.922	-1.338	5.007
Simulated		GARCH n	0.0160	0.107	0.057	8.936	-0.005	2.124
		GARCH t	0.0159	0.095	0.069	10.848	-0.012	3.981
		AGARCH	0.0159	0.092	-0.317	9.091	-0.340	2.787

* Average values of skewness and excess kurtosis across the time series of density forecasts.

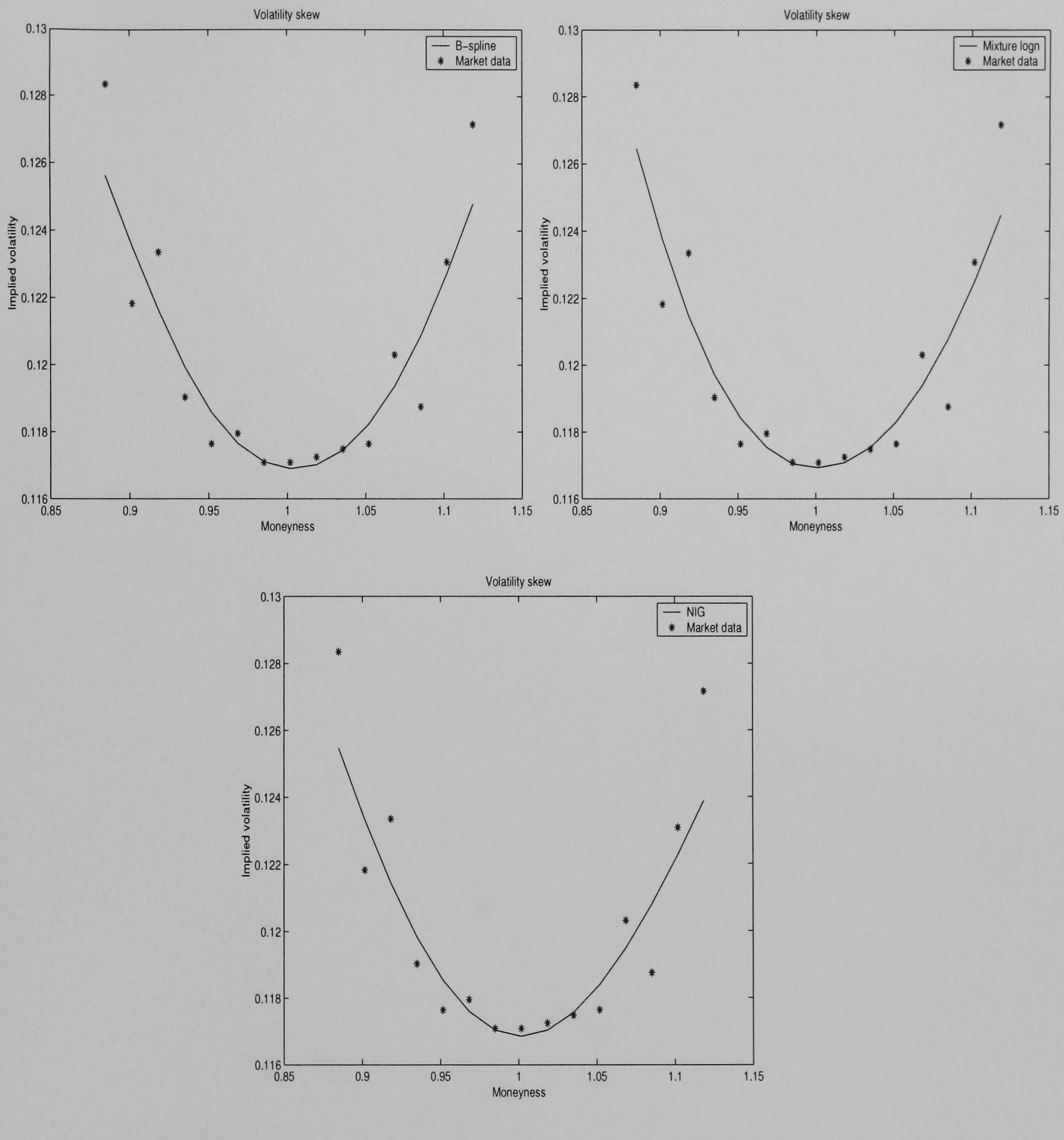
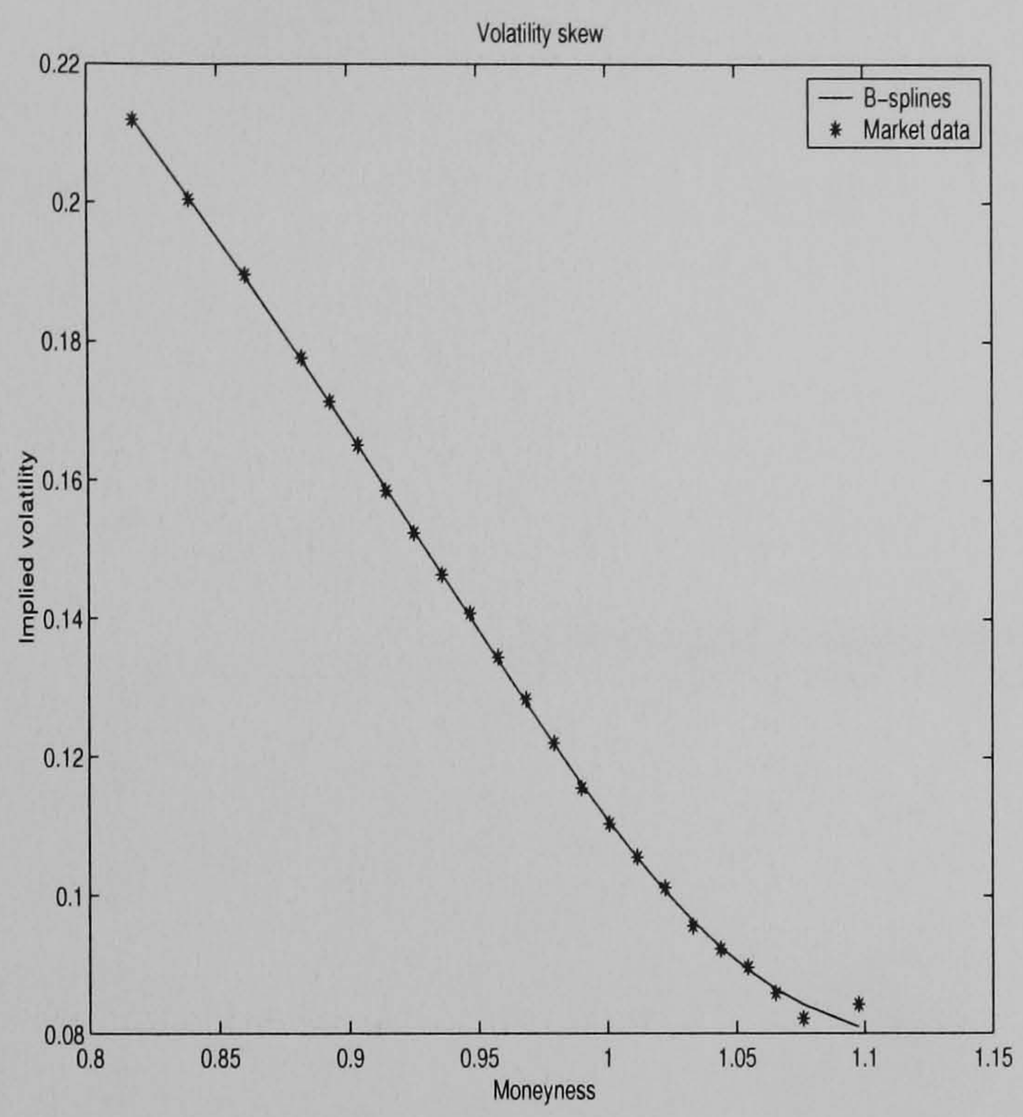
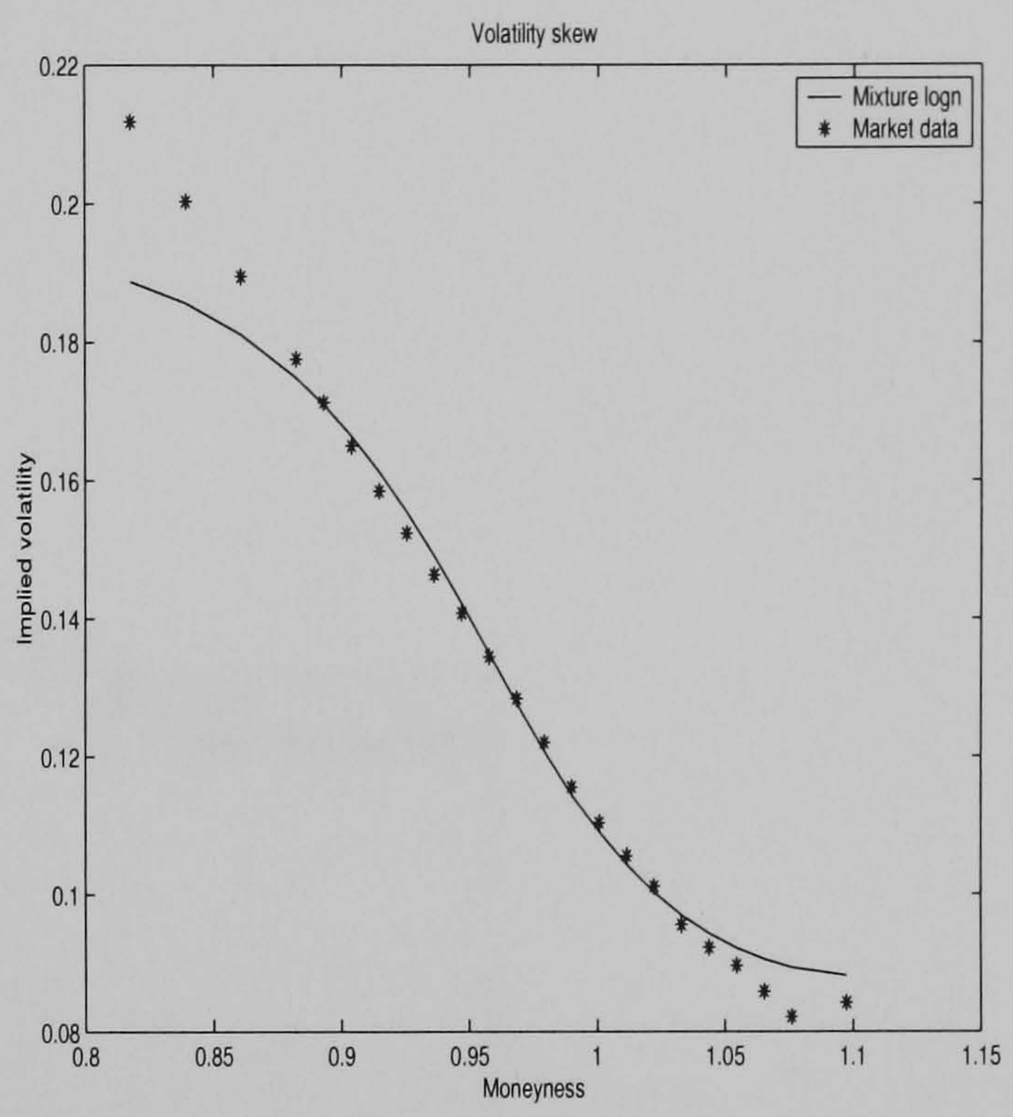


Figure 2.1: Implied volatility curves from currency futures options: December 93 contract.





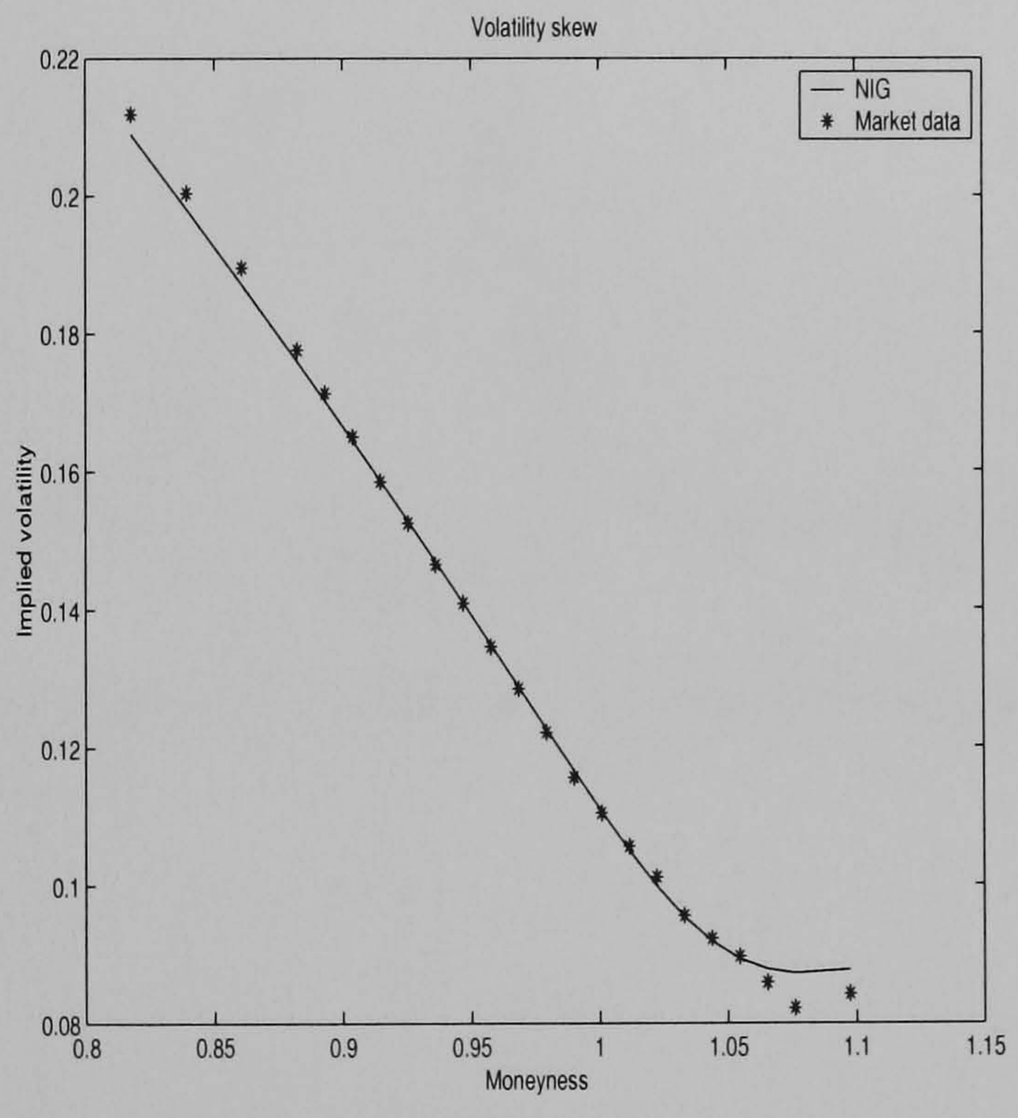


Figure 2.2: Implied volatility curves from index futures options: December 93 contract.

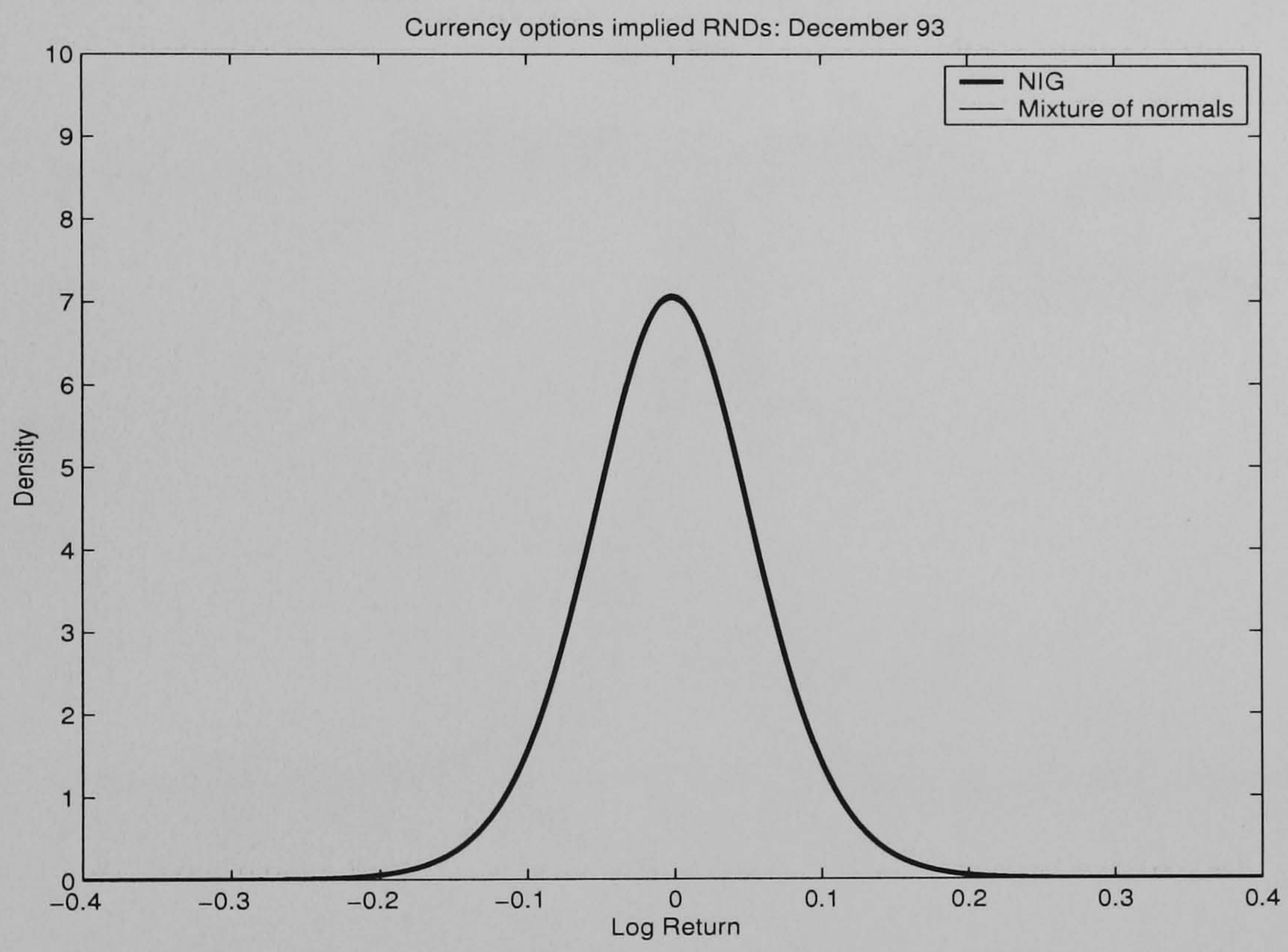


Figure 2.3: Implied RNDs from currency futures options.

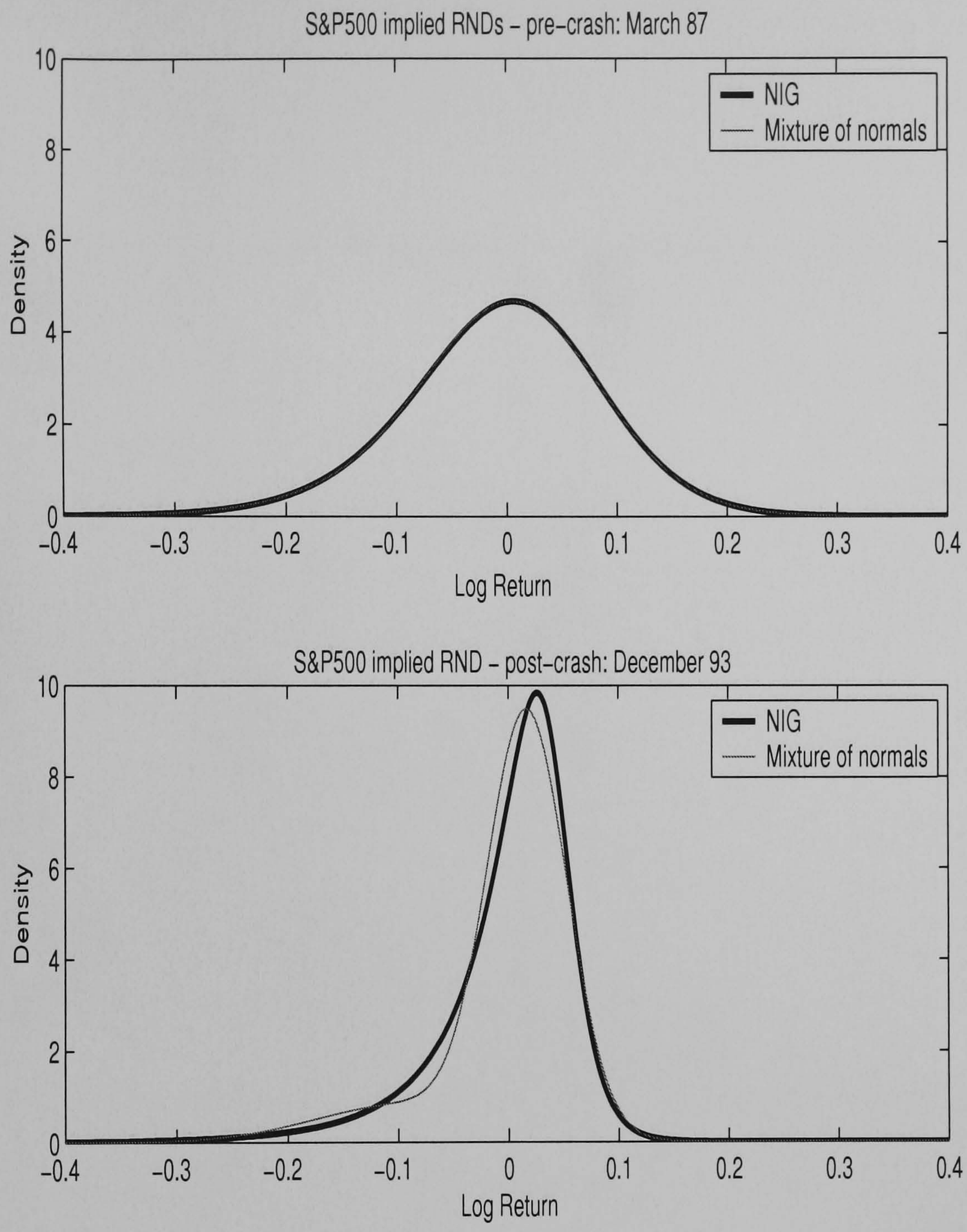
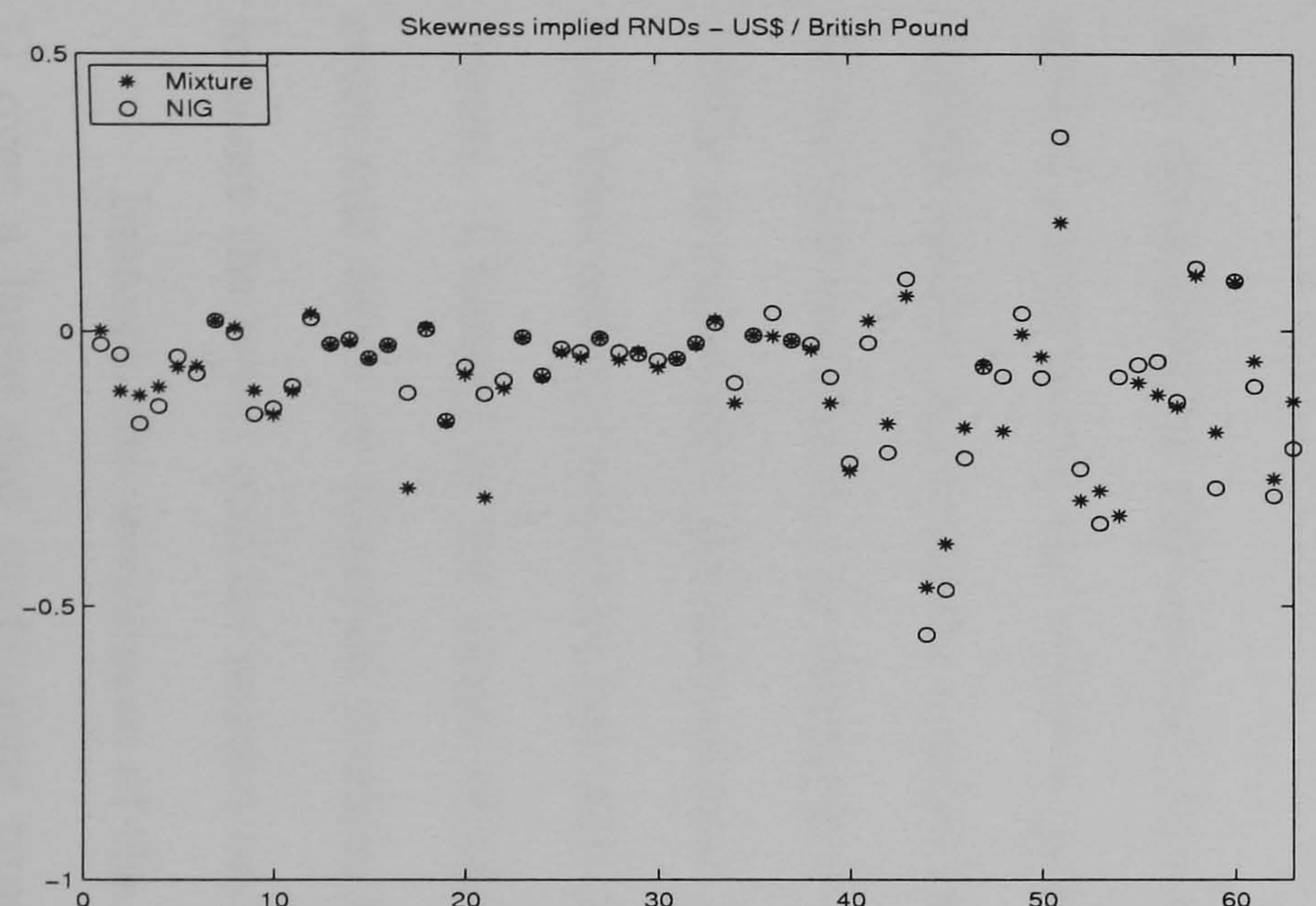
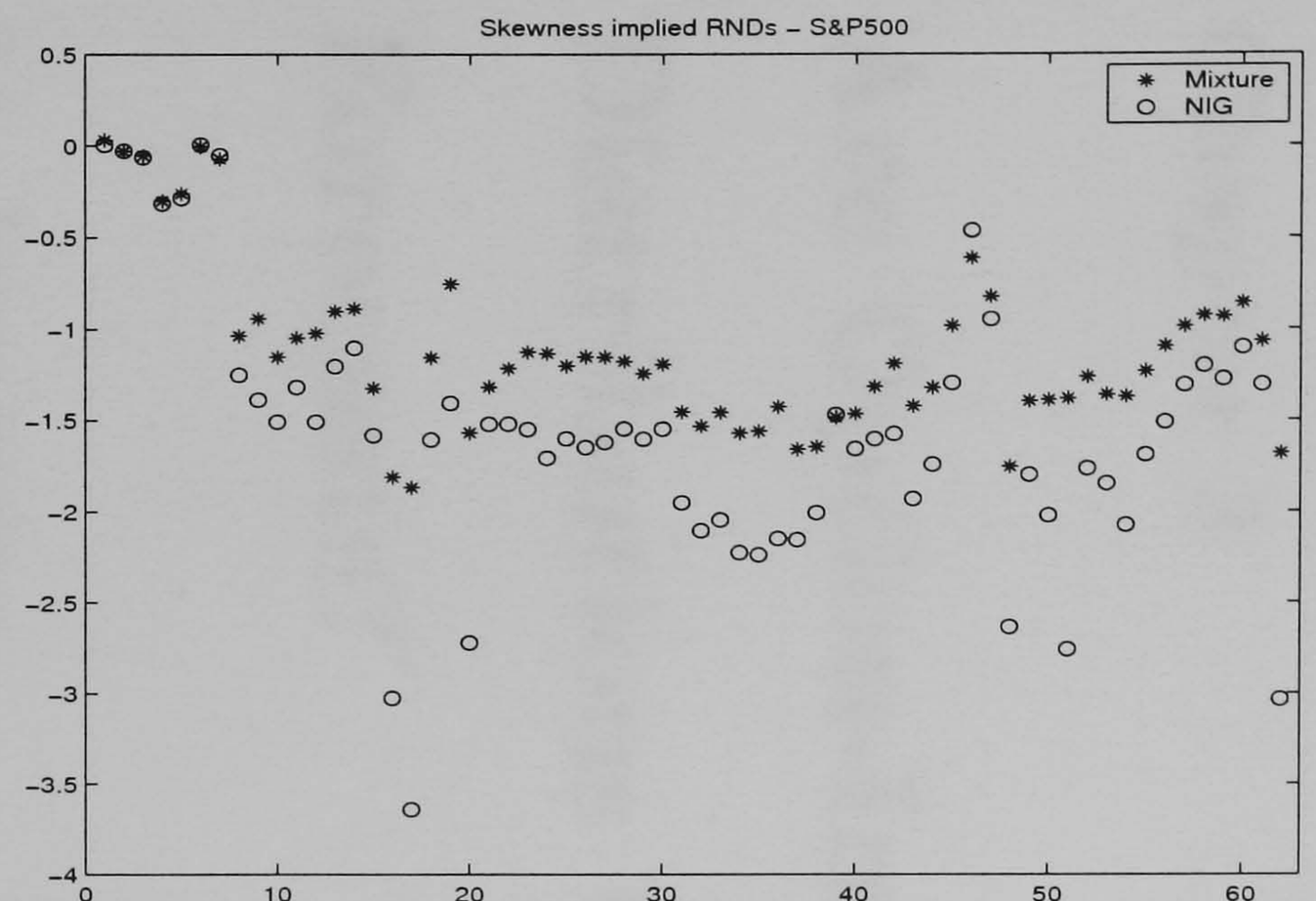
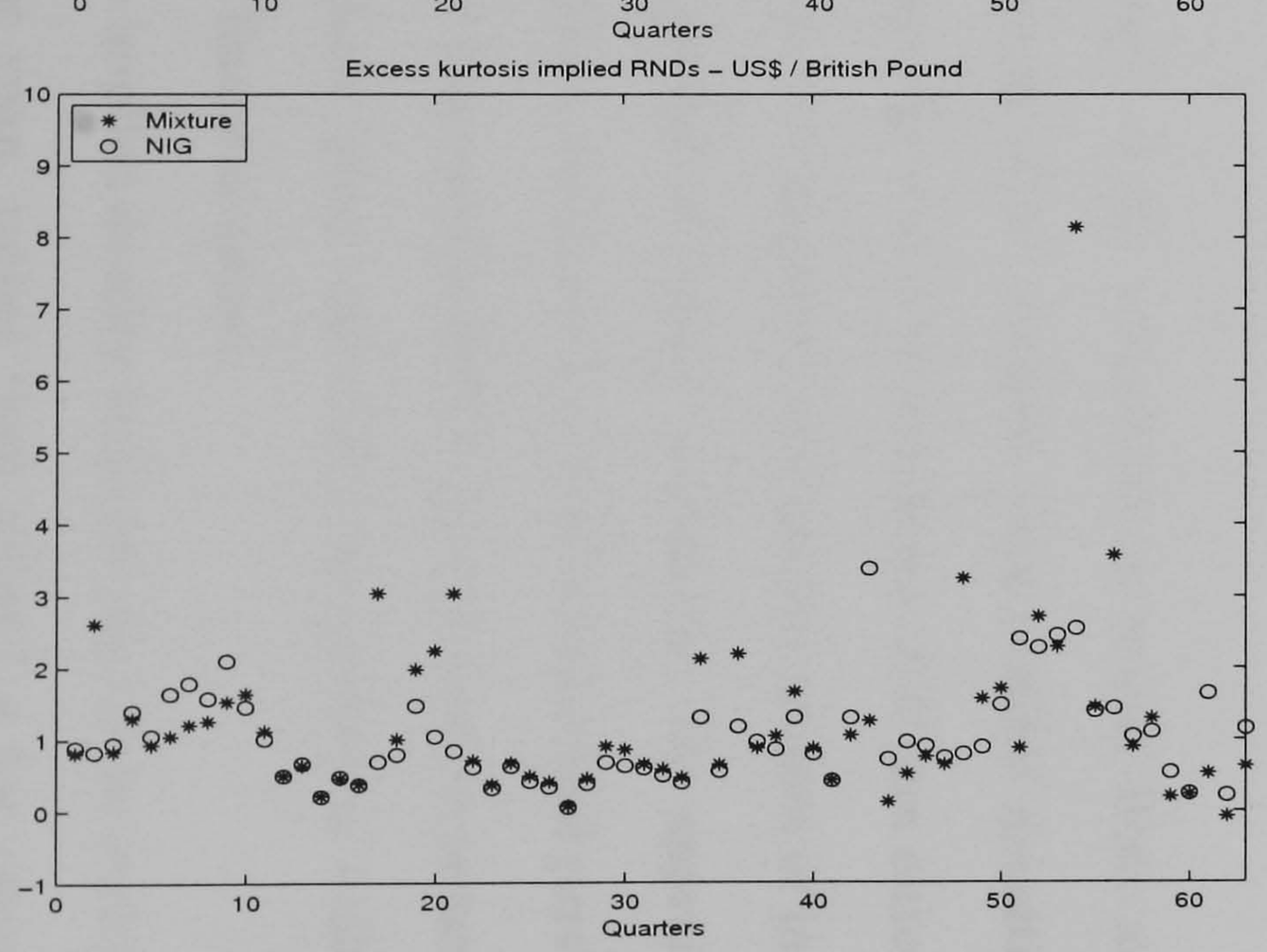


Figure 2.4: Implied RNDs from index futures options (pre- and post-crash).







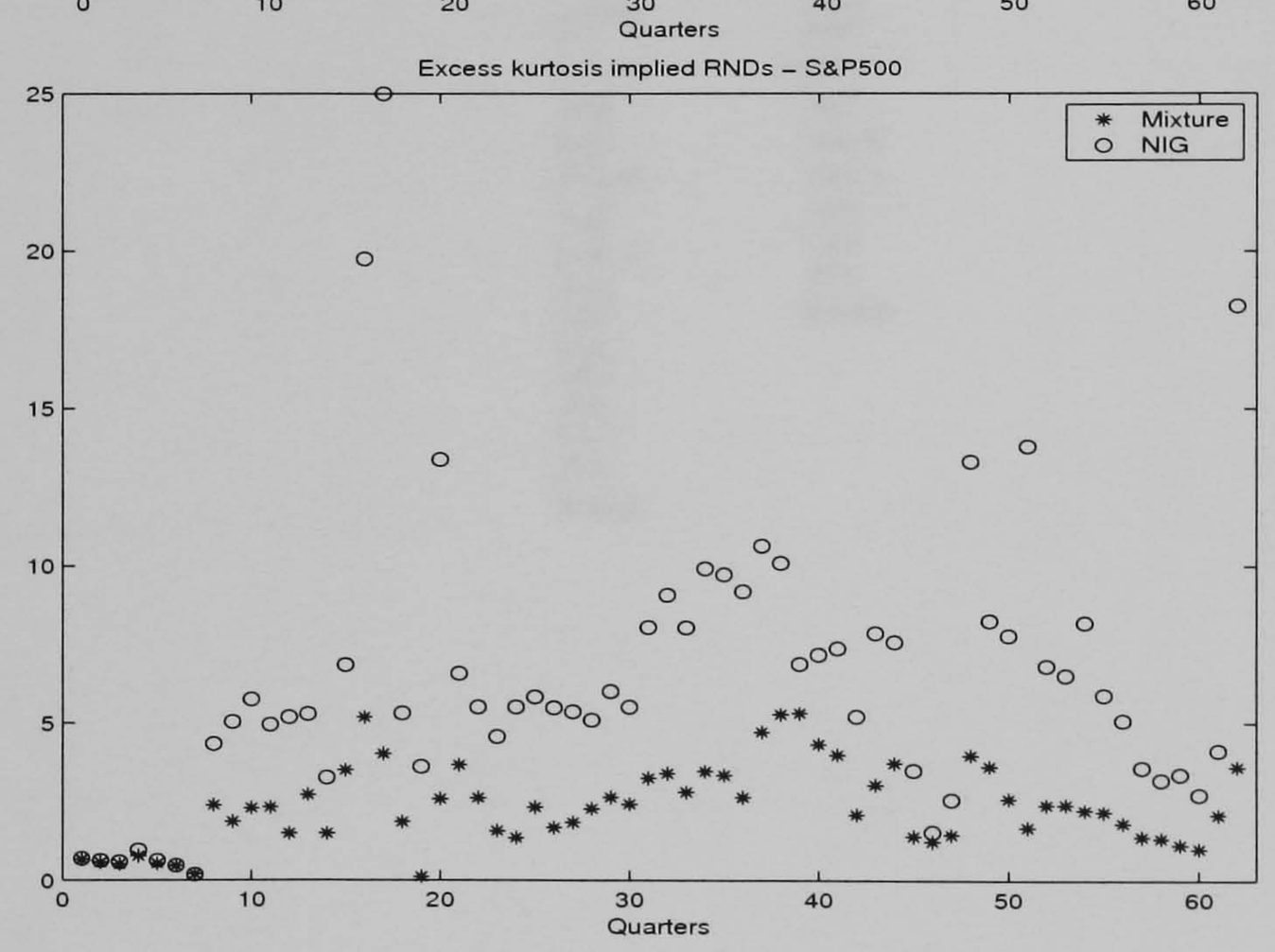


Figure 2.5: Time series of higher moments for parametric implied RNDs - currency (left) and index (right) futures contracts.

Chapter 3

Are Option-Implied Distributions Unbiased Forecasts?

3.1 Introduction

The derivation of risk-neutral densities of the underlying at expiry from market option prices raises the obvious question as to whether these implied densities do provide useful information concerning the actual probabilities of future outcomes. As we will see shortly, by claiming that the potential information content of implied RNDs is especially relevant around periods of market turbulence, the appraisal of their forecasting properties has often been restricted to a few dates around particular events. Changes in the shape of the risk-neutral PDFs around announcements of economic news or financial crashes have been commonly interpreted as ability to forecast the event and its impact on financial prices.

Instead, the assessment of the implied density forecasts should be performed: 1) over a large and continuous time span, rather than around a few occasional

events; 2) by means of statistically accurate density forecast evaluation tools, rather than through intuitive but informal visual inspections. Only very recently, research studies that focus on developing a rigorous and complete evaluation of the forecasting performance of the option-based densities have begun to appear in the literature.

The present work fits in this context by investigating, in a fully systematic and formal way, whether the implied densities derived in the previous chapter constitute unbiased forecasts of the future distributions of asset returns. We recall that our RNDs have been extracted from quarterly options on both the S&P 500 future and the US dollar/British pound future, for a period of nearly 16 years. The unbiasedness of the implied density forecasts is thoroughly investigated by means of density forecast evaluation techniques. An entire battery of distributional tests of goodness-of-fit is implemented in an attempt to obtain satisfactory statistical power for our small sample, as well as to distinguish between different sources of misspecification. We find that risk-neutral densities implied from either currency or index futures options give biased forecasts of the actual future returns. For the currency case, we record a misspecification of the tails, whereas for the S&P 500 index the bias is in the mean of the distribution.

Following some recent studies (see Bliss and Panigirtzoglou [2004]), in order to obtain more meaningful density forecasts of the future realisations of the equity index, capable of accounting for the presence of a risk premium, we also derive risk-adjusted densities via a simple utility based transformation. We assume that investors' preferences are well represented by a power utility function. The evaluation of the forecasting properties of these subjective densities reveals that the bias in the mean is removed, and we can no longer reject the hypothesis that the risk-adjusted implied densities are unbiased predictors of the distribution of realised returns on the S&P 500 future.

It is important to emphasise that our purpose here is to test for unbiasedness

of the option-implied density forecasts. We do not attempt to test the efficiency and/or the power of our forecasts which, as we will discuss more in detail later on in the chapter, is a much more complicated task that would require the availability of a richer data set.

Finally, we provide an original contribution to the existing related literature by drawing a comparison between option-based and historical-based density forecasts.[1] The latter are obtained via simulation of three different GARCH-type models. Our results indicate similar forecasting properties, even for very simple GARCH specifications of the volatility model.

The chapter is structured as follows. Section 3.2 summarises the relevant literature. Section 3.3 describes the testing techniques. The results from the appraisal of the implied risk-neutral and risk-adjusted densities are discussed in Section 3.4. Section 3.5 presents the comparison with the historical-based density forecasts. Section 3.6 concludes.

3.2 Related Literature

To date, most of the literature directed at evaluating the out-of-sample forecasting properties of risk-neutral densities extracted from option prices, has concentrated on the investigation of the behaviour of the implied densities around specific events of particular interest. Jackwerth and Rubinstein [1996] compared implied RNDs before and after the October 1987 stock market crash. Melick and Thomas [1997] focussed on the impact of the Persian Gulf crisis of 1990-91 on the shape of the risk-neutral distributions. Campa and Chang [1996, 1998], and Campa, Chang and Refalo [1999] studied implied RNDs around exchange rate crises and realignments, to assess the credibility of commitments to exchange rate target zones. Malz [1996] and Mizrach

[1]The only other study including a similar comparison we are aware of is by Liu, Shackleton, Taylor and Xu [2002], which is, however, subsequent to our research.

[1996] analysed how RNDs changed around the crisis in the European exchange rate mechanism in September 1992, and Söderlind [2000] extended the analysis to include the announcement of a new monetary policy in the UK. McCauley and Melick [1996] and Malz [1997] tried to infer general market sentiments in foreign exchange markets from alterations of the shape of implied RNDs. Bahra [1997] and McManus [1999] investigated the change in the shape of implied RNDs around announcements of economic news and shifts in the interest rates. Coutant, Jondeau and Rockinger [2001], and Jondeau and Rockinger [2000] focussed on the RNDs around the French 1997 snap elections. Leahy and Thomas [1996] examined the RNDs of the US dollar/Canadian dollar exchange rate around the time of the Quebec independence referendum in 1995. Gemmill and Saflekos [2000] investigated the shapes of implied RNDs before and after some important market breaks (October 1987, October 1989, September 1992 and October 1997), as well as the British general elections of May 1987, April 1992 and May 1997.

The evidence reported in these contributions is contradictory. Some authors claim that significant changes in the shape of the implied RNDs as the event approaches are indicative of some ability to forecast the event of interest, whereas other authors sustain that the most relevant and sensible changes in the RNDs tend to follow, rather than precede, the event.

A weakness of the event-based approach is the lack of statistical rigor that usually accompanies the appraisal of the risk-neutral distributions (see Jackwerth [1999], and Tay and Wallis [2000] on this criticism). In most cases, the comparison amongst RNDs extracted at various moments around the event is based on the visual inspection of their graphs and on the computation of some basic summary statistics. Therefore, there is a danger of reading too much information into few simple plots and figures.

Also, a complete and precise evaluation of the forecasting ability of implied

RNDs should not be restricted to a few episodes of market turbulence. Instead, the forecasting performance of risk-neutral densities must be assessed over a continuous and sufficiently long time span, and by means of robust statistical techniques. Fackler and King [1990] made an early attempt in this direction. They evaluated the fit between RNDs implied from various agricultural commodity options and the actual returns, over the period following the initiation of trading in commodity option futures in 1984. The evaluation tool they suggested, the empirical calibration function, consists in plotting the sequence of cumulative probabilities of the realised future values for the various estimated distribution functions. Silva and Kahl [1993] updated their results by using the same graphical testing approach. Dumas, Fleming and Whaley [1998] devised a new statistical measure, the hedge portfolio error, to assess the forecasting performance of different option pricing techniques over a period of 5 years. The hedge portfolio error is the difference between the change in the market option price and the change in the theoretical option price from the model. Dumas, Fleming and Whaley examined a series of hedge errors computed on a weekly basis for S&P 500 index options. Gemmill and Saflekos [2000] also used the hedge portfolio error (although calculated on a daily basis) to evaluate and compare the out-of-sample forecasting performance of various implied models for FTSE 100 index options over the 1987-97 period.

A study closely related to ours, in terms of both data set analysed and testing techniques employed, was carried out by Weinberg [2001]. He investigated the fit between risk-neutral densities extracted from option prices and actual returns over the period 1988-1999, for S&P 500 futures, US dollar/Japanese yen futures and US dollar/Deutsche mark futures. The goodness-of-fit was measured with distributional tests based on the Empirical Distribution Function, as well as graphically, by means of the empirical calibration function. His findings indicated good forecasting properties for the RNDs implied from currency options, and a significant bias in

the mean for the RNDs extracted from index options, due to the existence of a risk premium. After shifting the distributions by the average excess return of the index futures, the unbiasedness of the forecasts was no longer rejected.

Another research study closely related to our work is by Bliss and Panigirtzoglou [2004], who examined the forecasting performance of densities implied from S&P 500 and FTSE 100 options for a period of, respectively, 18 and 9 years. Given that risk-neutral PDFs extracted from equity index options differ from market expectations due to the existence of a risk premium, they derived subjective PDFs by adjusting the RNDs through a stationary utility function. On the basis of the results obtained from a set of density forecast evaluation techniques, they concluded that the subjective PDFs correctly forecast the actual returns at expiry. Besides testing for unbiasedness of the implied density forecasts, they also suggested possible ways to assess their efficiency, by working on subsamples of data.

Two main approaches have been proposed in the literature to investigate the relationship between risk-neutral densities, risk-adjusted densities and risk aversion functions. The first separately estimates implied RNDs from option prices and risk-adjusted distributions, generally from historical data, and then draws a comparison amongst them to infer the risk preferences of market participants. Relevant examples in this respect are the papers by Jackwerth [2000], Ait-Sahalia and Lo [2000], Rosenberg and Engle [2002], Brown and Jackwerth [2002]. Comparisons of implied RNDs with the risk-adjusted stochastic process of the asset, modelled as a one-factor diffusion, have been proposed by Ait-Sahalia, Wang and Yared [2001], and Coutant [2001]. The main drawback of this method is that, by not imposing restrictions on the structure of the risk preferences, the resulting implied risk aversion functions are often inconsistent with the theory. The alternative approach, which is the one followed in this work, consists in specifying a given parametric form for the risk aversion function, and combining this with the implied RNDs to derive the

subjective density forecasts: see Bartunek and Chowdury [1997], Bakshi, Kapadia and Madan [2003], Bliss and Panigirtzoglou [2004], Liu, Shackleton, Taylor and Xu [2002].

As stressed above, only the most recent contributions (e.g. Weinberg [2001], Bliss and Panigirtzoglou [2004]) have implemented rigorous techniques for density forecast evaluation, in order to assess the unbiasedness of PDFs implied from option prices.

The key device in the area of density forecast evaluation is the Probability Integral Transform (PIT) approach, which dates back to Fisher [1930] and Rosenblatt [1952], and was subsequently adopted in several expositions, such as Dawid [1984] and Cooke [1991]. Given a sequence of n one-step-ahead density forecasts $p_t(y_t)$ of the density $f_t(y_t)$, the PIT of the realisation of the process y_t, taken with respect to the density forecast produced in $t-1$, is given by:

$$\begin{aligned} z_t &= \int_{-\infty}^{y_t} p_t(u)\,\mathrm{d}u \\ &= P_t(y_t) \end{aligned} \tag{3.1}$$

for $t = 1, \ldots, n$. If the forecasts and the true densities coincide, assuming that the Jacobian of the transformations is non-zero over the support of the realisations, with continuous partial derivatives, then the sequence of the PITs z_t is distributed as i.i.d. $U(0,1)$. The evaluation of the density forecasts thus translates into assessing whether the PITs are independent uniform $U(0,1)$ variates.

Diebold, Gunther and Tay [1998] were the first authors who rigorously investigated the appropriateness of the PIT approach for the statistical evaluation of density forecasts in finance. The PIT approach holds regardless of the specific distributional form of the realisations y_t, and of the way the forecasts have been obtained. This latter aspect makes this approach particularly suitable to our problem of evaluating the forecasting performance of implied RNDs, since RND functions

are extracted from option prices and may not be generated from a specific model.

Deviations from uniform i.i.d. indicate that the forecasts are not correct. Inaccuracies in modelling the dynamics induce serial correlation in the z_t sequence and/or in its powers, whereas non-uniformity may be the result of improper distributional assumptions and/or poorly modelled dynamics. To test for uniform i.i.d., Diebold, Gunther and Tay [1998] suggested a visual approach, claiming that more formal joint goodness-of-fit tests of i.i.d. $U(0,1)$, as well as separate tests of i.i.d. and uniformity, cannot reveal the nature of the misspecification when rejection of the null hypothesis occurs. Therefore, they used histograms to evaluate unconditional uniformity, and correlograms of both the levels and the powers of the z_t series to detect inaccuracies in modelling the linear and non-linear dynamics of the true process.

Plots of the empirical distribution function against the 45° line have also been widely used to graphically assess the uniformity of the PIT series. More formal goodness-of-fit testing procedures, such as the Kolmogorov-Smirnov statistic, are also very popular in the relevant literature, even though their distribution theory usually rests on an assumption of random sampling, which is part of the joint hypothesis to be tested.

Stressing the inadequacy of the tests commonly associated with the PIT approach for realistic sample sizes, Berkowitz [2001] suggested to apply a transformation to normality to the series of PITs z_t as follows:

$$x_t = \Phi^{-1}(P_t(y_t)) \tag{3.2}$$

If the sequence of z_t is i.i.d. $U(0,1)$, that of x_t must be i.i.d. $N(0,1)$. At this point, conventional testing techniques for normally distributed data, whose statistical properties are well documented also for small sample sizes, can be employed. Berkowitz proposed to use log likelihood ratio tests, which, however, implicitly rest

on the assumption of normality of the transformed series, which should be tested on its own.

3.3 The Forecasting Ability of Implied RNDs: Testing Techniques

The aim of the present section is to provide a detailed description of the battery of goodness-of-fit techniques we adopt in our work to assess whether the risk-neutral densities implied from market option prices produce correct forecasts of the prices of the underlying observed at expiry.

As stated above, the PIT approach seems to represent an appropriate method to test this hypothesis. In our notation, the density forecasts $p_t(y_t)$ are represented by the quarterly implied risk-neutral densities $f(S_T)$ of the underlying at expiry S_T or, equivalently, by $f(x_\tau)$ of the log return until expiry x_τ. Since our non-overlapping implied RNDs are extracted on a quarterly basis immediately after the expiry of the previous contract, and they refer to the contract expiring exactly in three months time, here $t = 1, \ldots, 63$ quarters.

An important issue arises in the application of the PIT device to implied RND forecasts. The set of option prices and strikes across which the RND is derived, is discrete and, for some contracts, quite narrow. When we choose a parametric approach to estimate the RND, the tails of the resulting distribution outside the range of available strikes are implicitly obtained. Therefore, in the presence of an entirely specified density function, the series of PIT z_t can be computed following exactly the procedure described in the previous section.

On the contrary, the implementation of non-parametric techniques only consents to recover the implied RND within the range of available exercise prices. The resulting density is then truncated on both tails, and we only possess information on

the probability mass in each tail. In this case, the applicability of the PIT approach is limited to the truncated density, and the assessment of the forecasting properties is confined to the body of the distribution. Since the PITs are obtained from a truncated, rather than an entire, density, a modification of the formula in Eq. (3.1) becomes necessary. We devise a truncated version of the PIT as follows:[2]

$$z_t^* = \frac{P_t(y_t) - P_t(K_{min,t-1})}{P_t(K_{max,t-1}) - P_t(K_{min,t-1})}, \qquad K_{min,t-1} \leq y_t \leq K_{max,t-1} \tag{3.3}$$

where $K_{min,t-1}$ and $K_{max,t-1}$ denote, respectively, the minimum and the maximum strike available on the contract when the forecast was made, and $P_t(\cdot)$ is the value of the cumulative distribution associated with the density forecast. If the estimate of the truncated RND function coincides with the actual process followed by the underlying at expiry within the range of available strikes, the series of z_t^* is i.i.d. $U(0,1)$.[3]

In the present work we implement various alternative test statistics to assess both the truncated and the entire probability integral transform series. Since we restrict our attention to non-overlapping quarterly implied density forecasts, the sample size of our PIT series is very small, equal to $n = 63$ observations only. Several studies[4] have shown that for such a small sample size, most test statistics generally display very little statistical power to reject the density forecast under the null hypothesis, when it is false. Therefore, we believe that comparing findings from several alternative testing procedures might be more informative on the robustness of the results themselves. Also, the implementation of various tests may be useful in order to distinguish amongst different sources of misspecification, when the null hypothesis of unbiasedness is rejected.

---

[2]The analogy with the formulation for the truncated version of a density function is evident.

[3]The modified version of the PIT, although not strictly necessary for the computation of the Kolmogorov-Smirnov and the Kuiper statistics, becomes essential for the calculation of those tests which involve a nonlinear transformation of the z_t.

[4]See, amongst the others, Berkowitz [2001], and Noceti, Smith and Hodges [2003].

A remark is appropriate here. As stressed already, in our work we only test the unbiasedness of the implied RND forecasts, not their efficiency and/or power. If the production of efficient density forecasts is not an easy task, the identification of appropriate tests of efficiency is equally hard. Following the approach suggested by Bliss and Panigirtzoglou [2004], in principle we could implement tests of efficiency of the implied density forecasts by building and testing single partitions of the data. However, this approach has the drawback of not being exaustive, since there are several different criteria according to which the data can be partitioned. Also, in our case, the implementation of this technique becomes unfeasible since the overall sample size is already too small, and goodness-of-fit tests performed on single subsamples certainly would not show enough power.

3.3.1 Testing the uniformity of z_t series

We start by computing Empirical Distribution Function (EDF) test statistics, based on the vertical difference between the EDF and the theoretical distribution function of the null. Let $z_{(1)} < z_{(2)} < \ldots < z_{(n)}$ be the values of the z_t series[5] arranged in ascending order. Some statistical measures are computed as a function of the supremum of that vertical difference. The most famous amongst these measures is the Kolmogorov-Smirnov statistic, introduced by Kolmogorov [1933]:

$$D = \max(D^+, D^-) \tag{3.4}$$

where $D^+ = \max_i(i/n - z_{(i)})$ and $D^- = \max_i(z_{(i)} - (i-1)/n)$. A similar measure is the Kuiper statistic (Kuiper [1962]), defined as:

$$V = D^+ + D^- \tag{3.5}$$

Several other tests are based on a quadratic measure of the difference between the EDF and the theoretical distribution function under the null. We employ two of

[5]Here z_t refers to both the truncated and the non truncated series.

the most common statistics belonging to this class, the Watson U^2 statistic, introduced by Watson [1961], and the Anderson-Darling [1952] A^2 test, whose respective expressions are given by:

$$U^2 = \sum_i \left[z_{(i)} - \frac{2i-1}{2n} \right]^2 + \frac{1}{12n} - n(\bar{z} - 0.5)^2 \tag{3.6}$$

(where $\bar{z} = \sum_i \frac{z_i}{n}$)

$$A^2 = -n - (1/n) \sum_i (2i-1) \log z_{(i)} + \log[1 - z_{(n+1-i)}] \tag{3.7}$$

Comparative studies of the statistical performance of various EDF tests (D'Agostino and Stephens [1986], Noceti, Smith and Hodges [2003]) suggested that D and A^2 are particularly appropriate for the identification of errors in the mean, U^2 and V are more adequate for misspecifications in the variance, and A^2 is the most powerful when the forecasted distribution departs from the true distribution in the tails. Since our aim is to detect generic deviations from the null hypothesis of uniformity of the z_t series, we also need to focus on test statistics that possess good power against a wide range of alternatives. D'Agostino and Stephens [1986] identified Watson's U^2 test as the best amongst the EDF statistics for this purpose.

We also evaluate the uniformity of the z_t series by plotting the empirical cumulative distribution of the PITs against their theoretical cumulative function (45° line). The purpose is to provide a visual and more intuitive tool that facilitates understanding of how the estimated implied RNDs differ from the actual distributions of the underlying, when rejection of the null hypothesis occurs.

The absence of both linear and non-linear dependence in the series of PITs has been ascertained by plotting the autocorrelograms of the z_t and z_t^2 series.[6] In general there is no one-to-one correspondence between the dependence pattern exhibited by the z_t series and that of the underlying, nevertheless the analysis of

[6]Since the correlograms do not exhibit statistically significant autocorrelation, they have been omitted for clarity of exposition.

the correlograms can still provide some intuition on the potential misspecification in the dynamics of the forecasts. Also, the study of the dependence in z_t is important because the distribution theory for many of the conventional goodness-of-fit tests rests on an assumption of independence of the variable, which instead should be tested.

3.3.2 Testing the normality of x_t series

Berkowitz [2001] claimed that most of the testing procedures outlined above are not powerful enough for small samples. In order to obtain more robust results, the inverse probability transformation in Eq. (3.2) can be applied to the z_t series and the resulting x_t series can be tested for i.i.d. $N(0,1)$. Both graphical methods and formal goodness-of-fit tests can be implemented to test such hypothesis.

Following Berkowitz, we test the null hypothesis against a first order autoregressive alternative given by:

$$x_t - \mu = \rho(x_{t-1} - \mu) + \varepsilon_t \tag{3.8}$$

such that the null of i.i.d. $(0,1)$ translates into $\mu = 0$, $\rho = 0$, and $var(\varepsilon_t) = \sigma^2 = 1$. We then use log likelihood ratio (LR) tests. The formula for the log likelihood function associated with the model in Eq. (3.8) is the following:

$$\begin{aligned} L(\mu, \sigma^2, \rho) \quad = \quad & -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log[\sigma^2/(1-\rho^2)] - \frac{(x_1 - \mu/(1-\rho))^2}{2\sigma^2/(1-\rho^2)} + \\ & -\frac{n-1}{2}\log(2\pi) - \frac{n-1}{2}\log(\sigma^2) - \sum_{t=2}^{n}\left(\frac{(x_t - \mu(1-\rho) - \rho x_{t-1})^2}{2\sigma^2}\right) \end{aligned}$$

In order to test the hypothesis of independence across the observations, the log likelihood ratio test can be expressed as:

$$LR1 = -2(L(\hat{\mu}, \hat{\sigma}^2, 0) - L(\hat{\mu}, \hat{\sigma}^2, \hat{\rho})) \sim \chi^2(1) \tag{3.9}$$

Similarly, the log likelihood ratio test for the joint hypothesis of independent observations with zero mean and unit variance can be formulated as:

$$LR2 = -2(L(0,1,0) - L(\hat{\mu}, \hat{\sigma}^2, \hat{\rho})) \sim \chi^2(3) \tag{3.10}$$

Although LR tests seem adequate for our problem since they usually possess good statistical power against general alternatives also for small sample sizes, they are based on the assumption of normality of the x_t series, which is part of the joint hypothesis to be tested.

The normality of the transformed probability series is assessed via the Doornik and Hansen [1994] (DH) test, which should be more powerful than the traditional Jarque-Bera test in the case of small samples. The test statistic is based on transformed measures of skewness and kurtosis, and it is distributed as a $\chi^2(2)$.

For completeness of our testing experiment, we also carry out diagnostic tests on the single parameters of the autoregressive model in Eq. (3.8), which should be more informative on the nature of the violations of the assumptions made, when rejection of the joint null hypothesis occurs. First we estimate the parameters of the model using Ordinary Least Squares, then we apply t-statistics[7] to test the individual hypotheses $\mu = 0$ and $\rho = 0$, as well as a χ^2 statistic to test $\sigma^2 = 1$.

At this point it is worth reminding that the PIT technique provides a useful instrument for judging the unbiasedness of non-parametric implied RNDs only within the body of the distribution. To guarantee a complete evaluation of the forecasting ability of non-parametric density estimates extracted from traded option prices, we need to integrate the PIT analysis with testing procedures applicable to the tails of the distribution.

[7]The use of t-tests in the context under examination can still be justified asymptotically, even though the t-statistic is not exact for autoregressive specifications, or for non-Gaussian error terms.

3.3.3 Testing the tails of the implied RNDs

Given that we only possess information about the probability mass below the minimum strike and above the maximum strike, the natural testing techniques for the tails of the distribution seem to be the ones used for evaluating probability forecasts.[8]

The most popular measures of accuracy for probability forecasts are the so-called "scoring rules", based on the distance between the probability forecast P_t^f formulated at time $t-1$ for an event at time t, and a binary variable R_t which assumes value of 1 if the event occurs, and value of 0 otherwise. In our case, P_t^f coincides with the estimated probability mass in a given tail, and $R_t = 1$ if the actual realisation of the underlying falls in the tail, whereas $R_t = 0$ otherwise.

Amongst the various "scoring rules", we choose to work with Brier's [1950] quadratic probability score:

$$B_n = \frac{1}{n} \sum_t 2(P_t^f - R_t)^2 \tag{3.11}$$

Clearly, B_n assumes values between 0 and 2, and more accurate probability forecasts are reflected in smaller values of the score. To assess whether B_n departs significantly from its expected value $\sum_t P_t^f(1 - P_t^f)$, we adopt the test statistic proposed by Seillier-Moiseiwitsch and Dawid [1993]:

$$Y_n^B = \sum_t (1 - 2P_t^f)(R_t - P_t^f) / \left[\sum_t (1 - 2P_t^f)^2 P_t^f (1 - P_t^f) \right]^{1/2} \tag{3.12}$$

which is asymptotically distributed as a standard normal.[9]

In the following section, the combination of techniques for probability forecast evaluation and PIT approach is adopted to assess not only the quality of RNDs extracted via non-parametric methods (B-spline), but also the performance of implied densities estimated according to a parametric functional form (mixture of

[8] For a good review, see Diebold and Lopez [1996].

[9] Since an accurate forecast produces a small value for B_n, a one-sided test is more appropriate in our case.

lognormals and NIG). In the latter case, this combined analysis is additional to the implementation of the PIT on the entire RND, and consents a more immediate comparison, in terms of forecasting performance, with the non-parametric specification.

3.4 Are Implied Densities Unbiased Predictors? A Discussion of the Results

The tests on the tails, as well as the uniformity and the normality tests on the PIT series, are performed on the implied RNDs derived from both the currency and the index futures options. For the latter, the assessment is carried out on both the original implied RNDs and the adjusted ones, which, as we will explain later, are corrected to account for the presence of a risk premium.

Tables 3.1 and 3.2 report the results of the tests conducted on the tails outside the range of available strikes. For the right tail, the left tail and the combination of both tails, we compare the frequency with which actual observations fall in those areas with the probability mass assigned by the B-splines and the truncated versions of the parametric distributions to the tails. We also compute the test statistic Y_n^B at 5% and 1% confidence level.

Tables 3.3 and 3.4 display the results of the goodness-of-fit tests for uniformity of the z_t series, computed on parametric and non-parametric implied RND specifications.[10] The critical values (at 5% and 1% confidence level) for the tests are taken from D'Agostino and Stephens [1986].

The results of the tests on the "normalised" x_t series for both parametric (entire and truncated) and non-parametric implied RNDs are shown in Tables 3.5 and 3.6. The first two columns report the t-test on the parameters of the model in Eq. (3.8), with p-values in brackets. The third column displays the χ^2 test on the

[10]For the parametric RNDs, test statistics are derived on both the entire and the truncated distribution.

variance. The normality statistic and the two log likelihood ratio tests follow.

Plots of the empirical distribution function for the PIT series under the parametric NIG specification against the theoretical 45° line are reported in Fig. 3.1-3.4.[11]

3.4.1 RNDs from currency futures options

We start by investigating the results of the tests on the tails. The probability forecast for the right tail is rejected. More specifically, the figures in Table 3.1 indicate that the forecast probability mass in the right tail is a downward biased estimate of the frequency with which the observed values for the underlying exceed the maximum strike. Instead, the goodness of the probability forecast for the left tail cannot be rejected, as well as the combined probability forecast for both tails.

The hypothesis that the body of the truncated distributions estimated with parametric methods well represents the corresponding portion of the actual distribution cannot be rejected by either uniformity or normality tests. On the contrary, a rejection is recorded for the B-spline, as suggested by the values of the U^2 and V statistics. This may be the result of a bias in the variance, given the low value of the estimated variance (0.6776).

The unbiasedness of implied mixture of lognormals and NIG non-truncated RNDs can never be rejected according to our distributional tests for uniformity. However, when the transformation in Eq. (3.2) is applied, the normality of the resulting series is rejected by the Doornik-Hansen test, as a consequence of the misspecification in the tails discussed above. The non-uniformity of the probability transforms obtained from the (entire) density forecasts is also detectable, for the NIG specification, from the visual inspection of the plot in Fig. 3.1. The distance

[11] Analogous plots have been drawn for the z_t series derived from the B-spline and the mixture of lognormal specifications. Since they closely resemble the plots under the NIG specification, their inclusion would have been redundant.

between the empirical and the theoretical cumulative densities is not negligible, and wider than the corresponding distance observed for the truncated NIG RNDs. Our findings suggest that the failure to reject of the tests for uniformity was probably due to the low statistical power possessed by these test statistics, rather than to the good fit of the implied density forecast to the actual density.

3.4.2 RNDs from index futures options

The battery of tests conducted on the RNDs implied from the S&P 500 futures options seems, at first sight, to lead to quite different results from the ones described above. The tests on the tails, for both the B-spline and the truncated versions of the parametric densities, seem to indicate that the probability mass left in each tail represents a good forecast of the actual frequency with which realisations of the underlying at expiry fall into that region. No rejection occurs in either the left or the right tail.

On the contrary, the values of the D and A^2 statistics lead to a rejection of the hypothesis of correct forecast for the body of the estimated RNDs, for both parametric and non-parametric specifications.[12] A deeper investigation via diagnostic tests on the coefficients of the AR(1) model in Eq. (3.8), confirms the intuition that the principal reason for rejecting is represented by a bias in the mean, with the addition of a statistically significant downward bias in the variance for the B-spline. The misspecification in the mean is also evident from the plot in Fig. 3.2, as the empirical distribution of the PIT series consistently lies below the theoretical 45° line. The presence of a bias in the mean is not surprising, given the historical evidence

[12] According to some test statistics, the rejection occurs more often for the truncated version than for the entire parametric distribution. This could be explained by the fact that these tests place more weight on the tails of the distribution, which are the original ones for the entire distribution, whereas they have been somehow "redetermined" for the truncated one. If the outcomes from the tail tests suggested that the original tails are not misspecified, it is less likely that the full density specification is rejected by the distributional tests, than that a rejection occurs for the truncated one.

of the existence of a risk premium in equity indices which implies that the mean of the risk-neutral distribution always understates the mean of the actual distribution. Here we propose two measures of adjustment of our risk-neutral density forecasts, in an attempt to correct for the risk premium effect.

A simple mean-adjustment for the risk premium

The first type of correction we apply consists of a simple adjustment of the actual realisations. Let:

$$AR = \frac{1}{n} \sum_{t=1}^{n} \frac{y_t}{F_t} \tag{3.13}$$

be the average value of the ratios of the actual observation y_t (in three months time) to the forward price F_t observed at the beginning of the quarter, when the RND was estimated. The adjusted realisations, corrected for the risk premium, are computed as $y_t^* = y_t \cdot \frac{1}{AR}$. This type of adjustment provides a very simple way of centering the mean of the actual distribution on the forward price.

The PIT series is recomputed on the adjusted realisations of the underlying, while maintaining the same implied RND forecasts as before. The findings from our testing techniques are now much more in line with those derived for the currency futures contract. The t-test on μ coefficient reveals that the bias in the mean has been removed. The right tail is misspecified, as indicated by the tests on the tails for the B-spline and the truncated distributions.[13] More specifically, the probability assigned to the right tail overestimates the actual frequency of observations in this tail. This evidence is also reinforced by the rejection of the null hypothesis in some of the uniformity tests and in the normality test of the PITs computed on the entire NIG and mixture of lognormals specifications. The upper plot in Fig. 3.3 also highlights an excessive positive departure of the empirical distribution from the

[13]Since the actual observations have been adjusted, whilst the truncation points (corresponding to the minimum and the maximum strikes) have been left unchanged, only the tail frequencies are affected, and the probability forecasts for the tails stay the same.

theoretical one in the right tail. A bias in the variance is also reported for both entire and truncated versions of our parametric density forecasts, with an estimated variance significantly smaller than one at a 5% confidence level.

Although this simple mean-adjustment would be correct under the assumptions of normally distributed returns and power utility function, it has no obvious theoretical justification under our non-Gaussian specifications. Therefore, we also implement a more rigorous and general utility based transformation.

A utility based adjustment

Following Bliss and Panigirtzoglou [2004], the second type of adjustment we adopt is a utility based transformation, and requires specific assumptions on the representative investor's utility function $U(S_T)$. It is well known that when markets are complete and frictionless, and a single risky asset is traded, the subjective density function $q(S_T)$ can be related to the risk-neutral density function $f(S_T)$ through:[14]

$$q(S_T) = \frac{\frac{f(S_T)}{U'(S_T)}}{\int \frac{f(x)}{U'(x)}\,\mathrm{d}x} \tag{3.14}$$

Common choices for $U(S_T)$ are the power utility function $U(S_T) = \frac{S_T^{1-\gamma}-1}{1-\gamma}$ and the exponential utility function $U(S_T) = -\frac{e^{-\gamma S_T}}{\gamma}$. Both utility functions depend on a single parameter γ. The power specification constitutes a sensible choice to make given its analytical tractability and its constant relative risk aversion (RRA), measured by the parameter γ. A utility function of the exponential type is the specification assumed in the application of the popular Esscher transform. However, it has constant absolute risk aversion, whilst the relative risk aversion is time varying, depending on both γ and the realisations for S_T. Given that the specific parametric forms we employ to model the RNDs do not support an exponential transformation,

[14]The denominator represents a normalisation for the subjective density function to integrate to one.

and that we give preference to constant RRA utility functions, we choose to work with the power utility function.

Therefore, the transformation in Eq. (3.14) is applied to the time series of the RNDs obtained under the mixture of lognormals and the NIG specifications, for different choices of the parameter γ. Summary statistics for the resulting risk-adjusted density functions for values of γ equal to 1 and 2 are displayed in Table 2.3 (at the end of the previous chapter).[15] As expected, the adjustment has the effect of pushing the mean of the implied densities towards the observed one, as well as reducing the gap between implied and sample higher moments, which remain, however, overestimated.

We then recompute the PIT series according to the new risk-adjusted density forecasts, and we subsequently re-run all of the tails and distributional tests, whose outcomes are reported in Tables 3.2, 3.4, and 3.6. Our findings consistently indicate that, for choices of $\gamma = 1$ or $\gamma = 2$, we can no longer reject, at 5% confidence level, the hypothesis that the risk-adjusted density functions represent unbiased forecasts of the actual distribution of the underlying at expiry. The only exception is given by the truncated version of the adjusted mixture of lognormal density, with $\gamma = 1$, whose unbiasedness is rejected by both $LR1$ and $LR2$ tests.

These results clearly denote an improvement in the quality of the density forecasts, consequent upon an appropriate adjustment of the original RNDs to account for the risk premium. However, we suspect that the inability to reject is largely due to the low statistical power of our testing procedures for the small sample size considered, rather than to the unbiasedness of our implied density forecasts.

Our skepticism is also reinforced by the findings of Bliss and Panigirtzoglou [2004],

[15] The risk adjustments and their corresponding tests have been performed across the range of integer values for γ between 1 and 6. However, since the outcomes from the test statistics do not differ substantially, we only report the results for $\gamma = 1, 2$, which are, in fact, the most sensible values to assume for the constant RRA parameter.

who reported the non-stationarity of the utility function, whose parameters seem to depend on both the at-the-money volatility level and the time to maturity. Since the stationarity of the utility function is a fundamental assumption in this approach, and its violation induces misspecification in the tails, in order to obtain more reliable and informative results we would need to split our sample and re-run the test statistics on the subsamples. Unfortunately this is clearly unfeasible, given the already small size of our complete sample.

It is worth noticing that, at least conceptually, the simple mean-adjustment and the utility based transformation are not essentially different. In fact, under the assumptions of normal returns and power utility function, they are equivalent. Although the assumptions made for our density forecasts are not too far from these, the significant difference in the results can be explained with the different treatment of the realised observations in the two adjustment methods.

3.4.3 A summary of the results

Our main results can be summarised as follows. For both currency and index futures contracts, risk-neutral densities extracted from market option prices do not represent unbiased forecasts of the actual distribution of the underlying at expiry, whatever functional form (either parametric or non-parametric) is chosen to model them.

As expected, the main reason for rejecting the null hypothesis of unbiasedness is given by the misspecification of the tails outside the range of available strikes, on which very little information is available. The second reason (however, not common to all specifications) seems to be a bias in the variance, i.e. an underestimation of the variance of the actual process. Not too surprisingly, the worst specification, in terms of forecasting performance, turns out to be the cubic B-spline, probably as a consequence of the overfit of non-parametric methods to the actual option prices. The mixture of lognormals and the NIG exhibit very similar forecasting

performances, only slightly better for the latter for the majority of the test statistics implemented in our work. The superior out-of-sample goodness-of-fit recorded for the parametric methods over the non-parametric specification seems to suggest that parametric specifications are more appropriate when implied RNDs are employed with predictive purposes.

The rejection of the unbiasedness of the (unadjusted) implied RND forecasts is stronger for the index future. This is not altogether surprising, given the existence of a significant risk premium in this market. However, things change when an adjustment is introduced to account for the risk premium. If a simple adjustment applied to the actual realisations has the main effect of shifting the bias from the mean to the tails, a more rigorous adjustment to the mean of the original risk-neutral density, translates into a sensible improvement of the goodness-of-fit of the resulting risk-adjusted implied density. This indicates not only that unadjusted RNDs should never be used to forecast the actual density of the underlying at expiry, when the underlying is a stock index, but also that particular attention should be paid to the specific adjustment chosen to correct for the risk premium.

3.5 Assessing the Relative Forecasting Performance of Implied Densities: a Comparison with Historical-Based Density Forecasts

So far we have carried out an evaluation of the goodness-of-fit of density forecasts extracted from option prices *per se*. At this stage, we believe that a comparison with the corresponding forecasting properties of density forecasts made on an entirely different principle would be of interest. For this purpose, we propose to use GARCH-type models to produce distributional forecasts of quarterly returns, whose

predictive performance is then assessed by means of the same testing techniques as before, and contrasted to the performance exhibited by implied densities. It is worth noticing that, since historical-based density forecasts for the index futures contract include an estimate for the risk premium, the appropriate term of comparison is given by the risk-adjusted implied densities.

As customary for GARCH models, we assume that the evolution of the (demeaned) log returns X_t follows the conditional process:

$$X_t = \sigma_t \varepsilon_t \tag{3.15}$$

where ε_t is an i.i.d. Gaussian process with zero mean and unit variance, and the conditional volatility σ_t is a time varying, positive and measurable function of the information set at time $t-1$. As for the specific process that models the dynamics of the conditional variance σ_t^2, we first investigate a simple GARCH(1,1) specification, as suggested by Bollerslev [1986]:

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{3.16}$$

where $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$, and $\alpha_1 + \beta_1 < 1$ to ensure stationarity. A vast literature (see, for example, Hsieh [1989], Baillie and Bollerslev [1989]) found strong evidence that the normal specification for the conditional distribution of the returns cannot account for the significant fat-tailness exhibited by the empirical data. Therefore, conditional returns in Eq. (3.15) are also modelled according to a Student's t distribution.

When modelling the returns on the stock index, the leverage effect observed in the data (for which positive and negative shocks have a different impact on the volatility) must also be considered. Several variants of GARCH models that introduce asymmetry have been proposed by the relevant literature (TARCH, EGARCH, GJR, etc.) and many studies have provided evidence of their superior performance

at describing the dynamics of equity returns.[16] After attempting to estimate different models on our data, we have opted for a very simple asymmetric GARCH specification that seems to guarantee robust estimates and reliable forecasts.[17] The conditional variance is modelled according to the following process:

$$\sigma_t^2 = \alpha_0 + \alpha_1(X_{t-1} - \kappa_1)^2 + \beta_1\sigma_{t-1}^2 \tag{3.17}$$

where κ_1 represents the asymmetry parameter that, for positive values, models the leverage effect.

To estimate the GARCH specifications we proceed as follows. At the beginning of the quarter corresponding to each of the 63 three month futures contracts of interest, we estimate the model in Eqs. (3.15) and (3.16) with both Gaussian and Student's t errors, over the past series of daily log returns, for both the currency and the index futures contracts. For the latter, we also estimate the asymmetric model with conditional variance given by Eq. (3.17) and Student's t errors. The data set available for the estimation consists of daily log returns on both the currency and the index futures quarterly contracts[18] from April 1982 to September 2001. Only the data available at the time the estimates are made is used to produce them. The series of past log returns employed for the estimation include daily returns from April 1982 to the beginning of the quarter under examination. Therefore the length of the estimation window increases as we consider more recent quarterly contracts.[19]

Average values of the coefficient estimates, together with their dispersion figures, are displayed in Table 3.7. In line with the general findings for these models, the sum of the AR and MA parameters is close to one, and the GARCH lag coeffi-

[16] See, amongst the others, Brailsford and Faff [1996], Loudon, Watt and Yadav [2000].

[17] In particular, we abandoned the EGARCH specification because, even though we could obtain stable estimates, the resulting density forecasts were far too volatile.

[18] For contracts expiring within one week, daily log returns on the following contract are considered.

[19] We choose to fit our models over long return series in order to minimise the impact of extraordinary high or low volatility periods on the estimation process and, therefore, to obtain robust and stable estimates for the coefficients which constitute an essential basis for reliable forecasts.

cients are large, indicating that shocks to conditional variance die out very slowly. The average figures for the degrees of freedom suggest that conditional fat-tailed distributions are very close to the boundary between finite and infinite kurtosis (which characterises Student's t with four or less degrees of freedom). As expected, the sign of the asymmetry coefficient is positive and, therefore, consistent with the leverage effect observed in equity markets. Perhaps with the exception of the number of degrees of freedom, there is little variation amongst the estimates.

At the beginning of every quarter, the estimated coefficients are then used to produce recursive daily forecasts of the log returns over the following three months according to:

$$\widehat{X}_{t+1} = \mu + \hat{\sigma}_{t+1}\varepsilon_{t+1} \tag{3.18}$$

with:

$$\hat{\sigma}^2_{t+1} = \hat{\alpha}_0 + \hat{\alpha}_1 X_t^2 + \hat{\beta}_1 \sigma_t^2$$

or:

$$\hat{\sigma}^2_{t+1} = \hat{\alpha}_0 + \hat{\alpha}_1 (X_t - \hat{\kappa}_1)^2 + \hat{\beta}_1 \sigma_t^2$$

where ε_{t+1} are random numbers generated from a standard normal (for the GARCH model only) and from a standardised Student's t (for both the GARCH and the asymmetric GARCH models) with the estimated degrees of freedom.[20] The daily log returns are then added up together to compute the three-month log return, and the process is repeated 10,000 times to obtain a whole distribution for the simulated returns.

Summary statistics for equally weighted mixtures of the quarterly return density forecasts are shown in Table 2.3 (at the end of the previous chapter), and contrasted to the corresponding statistics for mixtures of implied density forecasts,

[20] Somehow arbitrarily, we assign values of $\mu = 0$ for the currency and $\mu = 0.07$, on an annual basis, for the index. We prefer doing so rather than deriving imprecise estimates for μ which could be heavily distorted by sampling error.

as well as for the observed continuously compounded quarterly log returns. For the currency futures contract, the GARCH specification with fat-tailed errors produces moments similar to the implied RNDs' ones. However, statistics closer to the empirical moments are obtained from the GARCH model with normal errors. For the S&P 500 futures contract, the most striking feature is the gap between the excess kurtosis of the mixture and the average excess kurtosis from the time series of the individual distributional forecasts. In fact, mixtures of historical-based density forecasts exhibit, for any of the GARCH specifications, higher fat-tailness than both the empirical sample and the mixtures of implied density forecasts (with the exception of the NIG specification). Only the asymmetric GARCH specification with t errors enables us to produce a significant negative skewness. However, this degree of asymmetry is smaller than both the asymmetry of the implied density estimates, and the skewness observed in the data.

In order to test how closely our simulated distributions for the log returns on quarterly contracts fit the actual distributions, we run the entire battery of tail and distributional tests on the density forecasts from the various GARCH models. The results are reported in Tables 3.8 - 3.10. The most appropriate specification for the currency contract seems to be the GARCH(1,1) with fat-tailed errors. Similarly to what we observed for the implied density forecasts, we record a misspecification of the right tail, due to an underestimate of the probability of the observations falling beyond the maximum strike. However, such misspecification is not captured by either uniformity or normality tests when we evaluate the entire (i.e. not split into tails and body) version of the simulated distributions, for which the null hypothesis of correct forecasts can never be rejected. This result can be again ascribed to a lack of statistical power of our testing techniques. Nevertheless, our findings suggest that the density forecasts obtained by simulating from a very simple GARCH(1,1) process with Student's t errors do a similar (and perhaps slightly better) job in predicting

the actual distribution of returns, as do the distributional forecasts extrapolated from currency futures option prices. This conclusion is also supported by the visual inspection of the CDF plots of the PIT series from the simulated distributions (Fig. 3.5), which lie closer to the theoretical 45° line than their option-implied counterparts.

Amongst our GARCH alternatives, the best choice we can make in order to forecast log returns on the index futures contract is the asymmetric GARCH with fat-tailed conditional distribution. For this specification, when the distributional forecasts are split into tails and body, and each component tested separately, the goodness-of-fit is never rejected, even though a careful look at the figures from the tail tests might suggest a misspecification in the right tail. When the unbiasedness of the entire density forecasts is assessed, a rejection occurs for the Watson's U^2 statistic and for the normality test on the transformed PIT series. A comparison between these findings and those from the evaluation of the implied density forecasts reveals that historical-based density forecasts are superior, in terms of accuracy of the forecasts, to risk-neutral densities extracted from option prices,[21] but not as good as risk-adjusted density forecasts.[22] Again, the same evidence can be found by contrasting the CDF plots for the GARCH-based density forecasts (Fig. 3.6) with the corresponding plots for the risk-neutral (Fig. 3.2) and the risk-adjusted (Fig. 3.4) implied densities.

3.6 Conclusions and Further Work

A significant recent literature has investigated whether risk-neutral densities implied from market option prices provide useful information about future distributions of

[21] The smaller measures of dispersion recorded for the historical-based density forecasts (see Table 2.3) also suggest that they are more efficient than the implied RNDs.

[22] As pointed out before, implied density forecasts adjusted to account for the risk premium are the sensible matter of comparison in this context.

the underlying assets. The evidence is mixed and mainly based on the investigation around single extreme events.

In this research we have tested whether RNDs implied from quarterly options on the S&P 500 index future and on the US dollar/British pound future, do provide unbiased forecasts of the actual distributions of future returns. The option-implied densities have been estimated by using a variety of methods, over a period of 16 years. In accordance with Weinberg [2001] and Bliss and Panigirtzoglou [2004], we find that simple implied RNDs constitute biased forecasts of the actual densities. Also, our results do not differ significantly according to the various specifications employed to model the implied densities.

The implied RNDs on the currency future exhibit a misspecification in the tails. As expected, the major source of bias for the implied RNDs on the index future is the value of the mean. Once we correct for the risk premium via a utility based transformation, we are unable to reject the unbiasedness of the resulting risk-adjusted densities. However, this kind of adjustment introduces the problem of the choice of a utility function which is stationary through time. Given this additional issue, and the low power to reject wrong specifications displayed by our testing techniques for small sample sizes, we prefer to interpret our findings as inability to reject the hypothesis of correct forecasts, rather than as acceptance of such hypothesis.

Despite having to work with a very small sample (which precluded the possibility to implement tests of efficiency/power of our forecasts), we have then found some evidence of biasedness of our option-implied density forecasts.

A comparison between option-implied and historical-based density forecasts reveals that, in some cases, even density estimates derived from very basic GARCH-type specifications display better forecasting properties than more involved implied densities.

In the light of our results, we can conclude that the information content of densities implied from option prices should not be overestimated. In particular, the use that some financial regulators make of them, as indicators of future market turbulence, can be dangerous and counterproductive.

Future work on this research will mainly involve the analysis of implied densities from monthly option contracts, which would increase the sample size and consent not only to obtain much more reliable results from the evaluation process, but also to implement tests of efficiency and power of the density forecasts. Alternative risk-adjustment methods could also be investigated.

Table 3.1: Tails Tests.

Test statistics		**Right tail**			**Left tail**			**Both tails**		
		Freq.	Prob. forec.	Y_n^B	Freq.	Prob. forec.	Y_n^B	Freq.	Prob. forec.	Y_n^B
5% conf. level				1.6449			1.6449			1.6449
1% conf. level				2.3263			2.3263			2.3263
US$/UK£										
	B-spline	0.1194	0.0566	*2.1281	0.0299	0.0431	-0.3084	0.1493	0.0977	1.1111
	Mixture	0.1194	0.0549	*2.2102	0.0299	0.0430	-0.3302	0.1493	0.0980	1.1484
	NIG	0.1194	0.0548	*2.2222	0.0299	0.0433	-0.3437	0.1493	0.0982	1.1419
S&P 500										
(unadjusted)	B-spline	0.0645	0.0777	-0.0583	0.0161	0.0343	-0.6143	0.0806	0.1121	-0.4217
	Mixture	0.0645	0.0780	-0.2070	0.0161	0.0364	-0.6840	0.0806	0.1144	-0.5862
	NIG	0.0645	0.0757	-0.1234	0.0161	0.0337	-0.6075	0.0806	0.1094	-0.4800

*rejected at 5% confidence level.

Table 3.2: Tails Tests (cont'd).

Test statistics			**Right tail**			**Left tail**			**Both tails**	
		Freq.	Prob. forec.	Y_n^B	Freq.	Prob. forec.	Y_n^B	Freq.	Prob. forec.	Y_n^B
5% conf. level				1.6449			1.6449			1.6449
1% conf. level				2.3263			2.3263			2.3263
S&P 500										
(Mean adj.)	B-spline	0.0168	0.0777	*-1.7875	0.0161	0.0343	-0.6143	0.0287	0.1121	*-1.9654
	Mixture	0.0168	0.0780	*-1.7916	0.0161	0.0364	-0.6840	0.0287	0.1144	*-1.9775
	NIG	0.0168	0.0757	*-1.7438	0.0161	0.0337	-0.6075	0.0287	0.1094	*-1.9215
S&P 500										
(Utility adj.)	Mixture									
	$\gamma = 1$	0.0645	0.0883	-0.4540	0.0161	0.0281	-0.4091	0.0806	0.1164	-0.6212
	$\gamma = 2$	0.0645	0.0995	-0.6825	0.0161	0.0218	-0.1474	0.0806	0.1213	-0.6965
	NIG									
	$\gamma = 1$	0.0645	0.0858	-0.3681	0.0161	0.0254	-0.3045	0.0806	0.1112	-0.4961
	$\gamma = 2$	0.0645	0.0966	-0.5938	0.0161	0.0195	-0.0268	0.0806	0.1161	-0.5676

*rejected at 5% confidence level.

Table 3.3: Uniform distributional tests - $H_0 : z \sim$ i.i.d. $U(0,1)$.

Test statistics			**D**	$\mathbf{A^2}$	$\mathbf{U^2}$	**V**
5% conf. level			1.3580	2.4920	0.1870	1.7470
1% conf. level			1.6280	3.8570	0.2680	2.0010
US$/UK£						
	B-spline		1.1704	1.3305	*0.2101	*1.8189
	Mixture lognormals	truncated	1.0711	0.9918	0.1537	1.5762
		entire	1.2019	1.5066	0.0899	1.3311
	NIG	truncated	1.0421	0.9780	0.1489	1.5492
		entire	1.1865	1.5177	0.0849	1.3159
S&P 500						
(unadjusted)	B-spline		*1.4018	*2.9740	*0.1997	*1.8423
	Mixture lognormals	truncated	*1.4222	*3.4280	0.1684	1.7401
		entire	*1.3364	*2.5544	*0.1923	1.6642
	NIG	truncated	1.3413	*3.1801	0.1474	1.6025
		entire	1.3317	2.3951	0.1705	1.6101

*rejected at 5% confidence level.

Table 3.4: Uniform distributional tests (cont'd) - $H_0 : z \sim$ i.i.d. $U(0,1)$.

Test statistics				**D**	**A²**	**U²**	**V**
5% conf. level				1.3580	2.4920	0.1870	1.7470
1% conf. level				1.6280	3.8570	0.2680	2.0010
S&P 500							
(Mean adj.)		Mixture lognormals	truncated	0.7816	0.9685	0.1205	1.5038
			entire	1.0112	1.0023	*0.2023	1.6845
		NIG	truncated	0.8439	1.1201	0.1528	1.5872
			entire	1.0195	1.7389	*0.2279	*1.7473
S&P 500							
(Utility adj.)	$\gamma = 1$	Mixture lognormals	truncated	1.2257	2.2117	0.1224	1.5940
			entire	1.1319	1.4176	0.1444	1.5019
		NIG	truncated	1.1028	2.0225	0.1048	1.4079
			entire	1.0089	1.2966	0.1259	1.3257
	$\gamma = 2$	Mixture lognormals	truncated	1.0328	1.3215	0.0875	1.4547
			entire	0.8366	0.7662	0.1079	1.2507
		NIG	truncated	0.9011	1.1786	0.0725	1.2531
			entire	0.7867	0.6865	0.0945	1.1553

*rejected at 5% confidence level.

Table 3.5: Normal distributional tests - $H_0 : z \sim$ i.i.d. $N(0,1)$.

Test statistics			μ	ρ	σ^2	**D-H**	**LR1**	**LR2**
5% conf. level			(> 0.05)	(> 0.05)		5.9915	3.8415	7.8147
1% conf. level			(> 0.01)	(> 0.01)		9.2103	6.6349	11.3449
US$/UK£								
	B-spline		0.0678 (0.4667)	-0.1946 (0.1517)	*0.6776	2.9851	2.0746	6.4250
	Mixture	truncated	0.0642 (0.5213)	-0.1732 (0.2027)	0.7572	2.5221	1.6315	4.2031
		entire	0.1450 (0.2328)	-0.0316 (0.7959)	0.9859	*7.1201	0.0528	2.0036
	NIG	truncated	0.0659 (0.5119)	-0.1732 (0.2028)	0.7635	2.5294	1.6305	4.0930
		entire	0.1487 (0.2235)	-0.0324 (0.7911)	0.9958	*7.2972	0.0555	2.0916
S&P 500								
(unadjusted)	B-spline		0.2361 (0.0684)	0.1094 (0.4198)	*0.6567	1.2221	0.7357	*9.1551
	Mixture	truncated	*0.2881 (0.0386)	0.1014 (0.4549)	0.7470	0.7546	0.6375	*8.4560
		entire	0.2153 (0.0733)	0.0189 (0.8858)	0.7738	3.5481	0.0218	5.4918
	NIG	truncated	*0.2808 (0.0483)	0.1130 (0.4049)	0.7681	0.6162	0.7801	*7.9044
		entire	0.2134 (0.0833)	0.0310 (0.8134)	0.7969	3.4374	0.0573	5.0050

*rejected at 5% confidence level.

Table 3.6: Normal distributional tests (cont'd) - $H_0 : z \sim$ i.i.d. $N(0,1)$.

Test statistics				μ	ρ	σ^2	**D-H**	**LR1**	**LR2**
5% conf. level				(> 0.05)	(> 0.05)		5.9915	3.8415	7.8147
1% conf. level				(> 0.01)	(> 0.01)		9.2103	6.6349	11.3449
S&P 500									
(Mean adj.)		Mixture	truncated	-0.0167 (0.8821)	0.1741 (0.2164)	*0.6054	0.0716	1.7302	*7.9393
			entire	-0.0924 (0.3862)	0.0063 (0.9631)	*0.6437	*5.9983	0.0173	5.3069
		NIG	truncated	-0.0215 (0.8465)	0.1792 (0.1979)	*0.5983	0.7503	1.7544	6.8452
			entire	-0.0931 (0.3922)	0.0171 (0.8978)	*0.6428	*6.8825	0.0267	5.8162
S&P 500									
(Utility adj.)	$\gamma = 1$	Mixture	truncated	0.2173 (0.1175)	0.1127 (0.4055)	0.7717	0.7333	*3.8415	*7.8148
			entire	0.1274 (0.2894)	0.0339 (0.7962)	0.7940	3.5017	0.0675	3.2305
		NIG	truncated	0.2103 (0.1383)	0.1245 (0.3581)	0.7914	0.6074	0.9215	5.5187
			entire	0.1256 (0.3088)	0.0454 (0.7294)	0.8177	3.3902	0.1204	2.8174
	$\gamma = 2$	Mixture	truncated	0.1485 (0.2876)	0.1238 (0.3605)	0.7968	0.7199	0.8953	4.0339
			entire	0.0426 (0.7273)	0.0480 (0.7142)	0.8146	3.4655	0.1337	2.0007
		NIG	truncated	0.1417 (0.3210)	0.1357 (0.3160)	0.8152	0.6039	1.0699	3.7808
			entire	0.0412 (0.7422)	0.0587 (0.6540)	0.8387	3.3537	0.2001	1.6650

*rejected at 5% confidence level.

Table 3.7: Average parameter values for estimated GARCH distributions.

	α_0	α_1	Asymmetry coeff.	β_1	Degrees of freedom
US$/UK£					
GARCH(1,1) - normal errors	8.99E-07	0.0471		0.9376	-
	(3.74E-07)	(0.0073)		(0.0122)	-
GARCH(1,1) - t errors	6.27E-07	0.0472		0.9449	5.4744
	(3.32E-07)	(0.0052)		(0.0098)	(1.1048)
S&P 500					
GARCH(1,1) - normal errors	5.6E-06	0.1421	-	0.8311	-
	(3.6E-06)	(0.0467)	-	(0.0603)	-
GARCH(1,1) - t errors	1.6E-06	0.0527	-	0.9379	4.8635
	(1.1E-06)	(0.0087)	-	(0.0129)	(1.0452)
Asymmetric GARCH(1,1) - t errors	1.1E-06	0.0560	0.0038	0.9287	5.0194
	(1.1E-06)	(0.0099)	(0.0010)	(0.0149)	(0.9931)

Standard deviation values in brackets.

Table 3.8: Tails Tests for GARCH-based distributional forecasts.

Test statistics		**Right tail**			**Left tail**			**Both tails**		
		Freq.	Pr. forec.	Y_n^B	Freq.	Pr. forec.	Y_n^B	Freq.	Pr. forec.	Y_n^B
5% conf. level				1.6449			1.6449			1.6449
1% conf. level				2.3263			2.3263			2.3263
US$/UK£										
	GARCH(1,1) normal errors	0.1194	0.0554	*2.1591	0.0299	0.0377	-0.1668	0.1493	0.0931	1.3230
	GARCH(1,1) t errors	0.1194	0.0563	*2.0590	0.0299	0.0391	-0.2373	0.1493	0.0954	1.2139
S&P 500										
	GARCH(1,1) normal errors	0.0645	0.1582	*-1.8923	0.0161	0.0199	0.0026	0.0806	0.1782	*-1.6454
	GARCH(1,1) t errors	0.0645	0.1266	-1.4186	0.0161	0.0130	0.5191	0.0806	0.1396	-1.1298
	Asymm. GARCH(1,1) t errors	0.0645	0.1229	-1.3156	0.0161	0.0141	0.4046	0.0806	0.1370	-1.0322

*rejected at 5% confidence level.

Table 3.9: Uniform distributional tests for GARCH-based distributional forecasts.

Test statistics			**D**	**A^2**	**U^2**	**V**
5% conf. level			1.3580	2.4920	0.1870	1.7470
1% conf. level			1.6280	3.8570	0.2680	2.0010
US$/UK£						
	GARCH(1,1)					
	normal errors	truncated	1.0996	1.0934	0.1780	1.7395
		entire	1.0899	1.3629	0.0833	1.2173
	GARCH(1,1)					
	t errors	truncated	1.0201	0.9827	0.1586	1.6576
		entire	1.1384	1.3391	0.0733	1.2594
S&P 500						
	GARCH(1,1)					
	normal errors	truncated	**1.8050	**4.2127	**0.3170	**2.2270
		entire	1.2752	*2.6595	**0.4430	**2.4478
	GARCH(1,1)					
	t errors	truncated	*1.4613	*2.9512	*0.2087	*1.7678
		entire	1.0303	1.7241	*0.2568	*1.8188
	Asymm. GARCH(1,1)					
	t errors	truncated	1.2243	2.1189	0.1572	1.5821
		entire	0.9289	1.3345	*0.2034	1.6901

*rejected at 5% confidence level.
**rejected at 1% confidence level.

Table 3.10: Normal distributional tests for GARCH-based distributional forecasts.

Test statistics			μ	ρ	σ^2	**D-H**	**LR1**	**LR2**
5% conf. level			(> 0.05)	(> 0.05)		5.9915	3.8415	7.8147
1% conf. level			(> 0.01)	(> 0.01)		9.2103	6.6349	11.3449
US$/UK£								
	GARCH(1,1)							
	normal errors	truncated	0.0551 (0.5600)	-0.2068 (0.1269)	0.7177	3.1117	2.3630	5.5006
		entire	0.1341 (0.2778)	-0.0093 (0.9385)	0.9793	*7.4266	0.0002	1.8084
	GARCH(1,1)							
	t errors	truncated	0.0590 (0.5434)	-0.1993 (0.1417)	0.7463	2.5794	2.1872	4.7786
		entire	0.1422 (0.2427)	-0.0176 (0.8842)	0.9620	3.7750	0.0113	1.9532
S&P 500								
	GARCH(1,1)							
	normal errors	truncated	*0.2816 (0.0409)	0.1273 (0.3499)	*0.6896	4.3461	0.9462	*9.7887
		entire	0.0164 (0.8713)	-0.0518 (0.6933)	*0.6872	**13.8051	0.1542	4.9204
	GARCH(1,1)							
	t errors	truncated	0.2331 (0.0864)	0.1807 (0.2807)	0.7842	3.6286	0.3996	5.8067
		entire	0.0309 (0.7812)	-0.0302 (0.8182)	0.7960	**10.5957	0.0521	2.2154
	AGARCH(1,1)							
	t errors	truncated	0.1858 (0.1651)	0.0902 (0.5085)	0.7725	2.7779	0.4698	4.8811
		entire	0.0086 (0.9388)	-0.0151 (0.9084)	0.7900	*7.8341	0.0131	2.2365

*rejected at 5% confidence level.
**rejected at 1% confidence level.

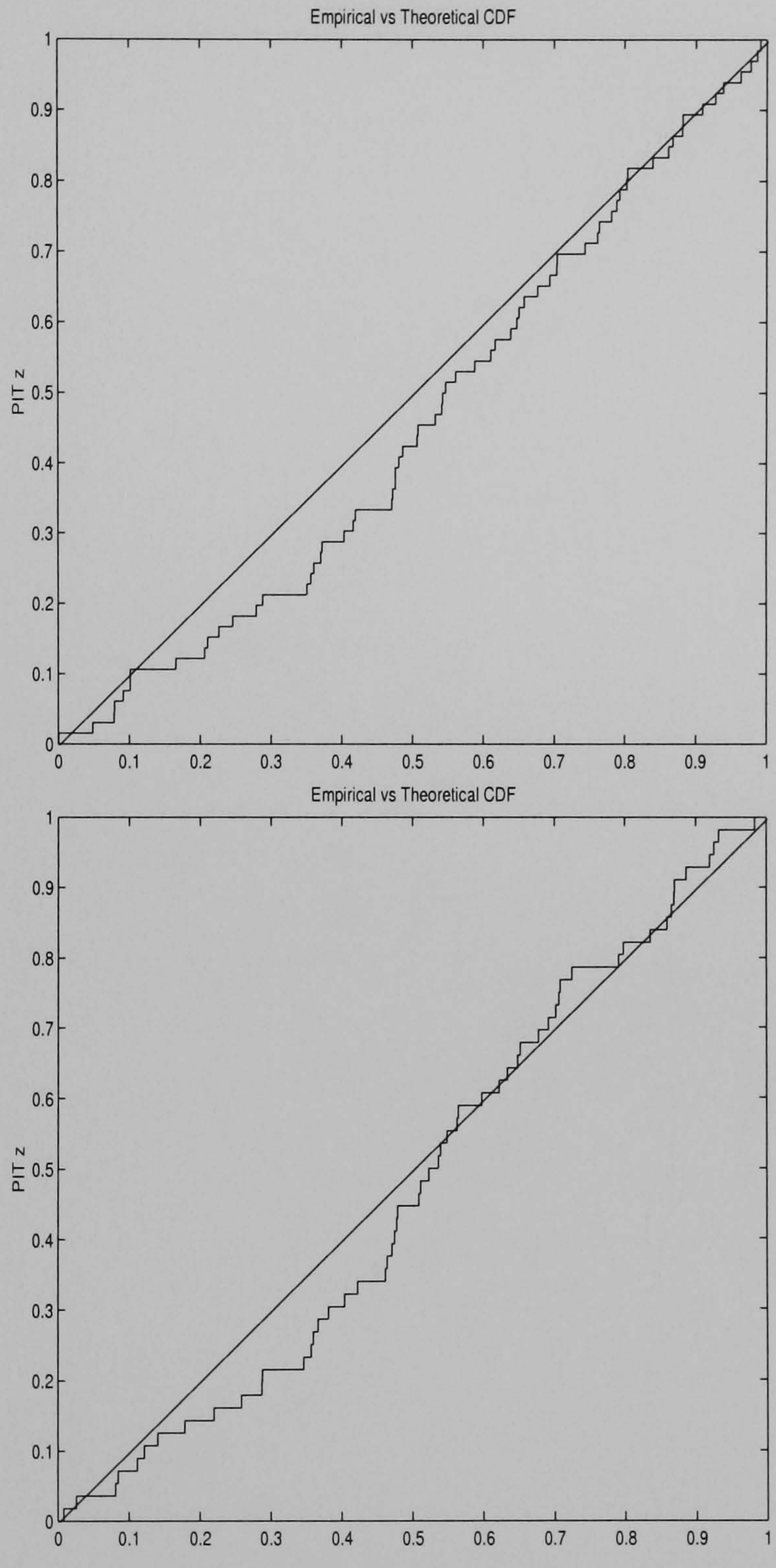


Figure 3.1: Graphical assessment of goodness-of-fit of risk-neutral NIG densities implied from currency futures options: entire (top) and truncated (bottom).

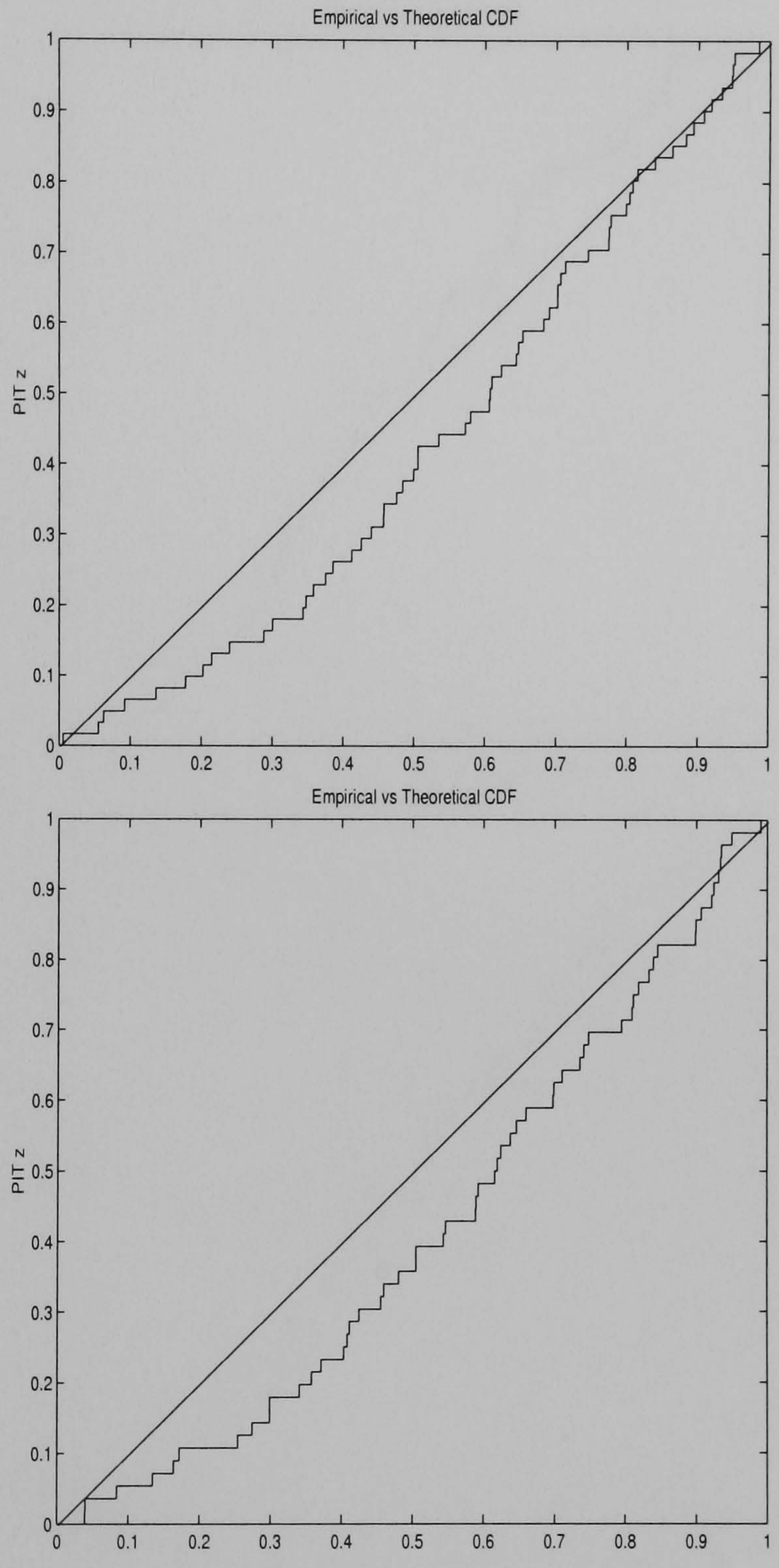


Figure 3.2: Graphical assessment of goodness-of-fit of risk-neutral NIG densities implied from index futures options: entire (top) and truncated (bottom).

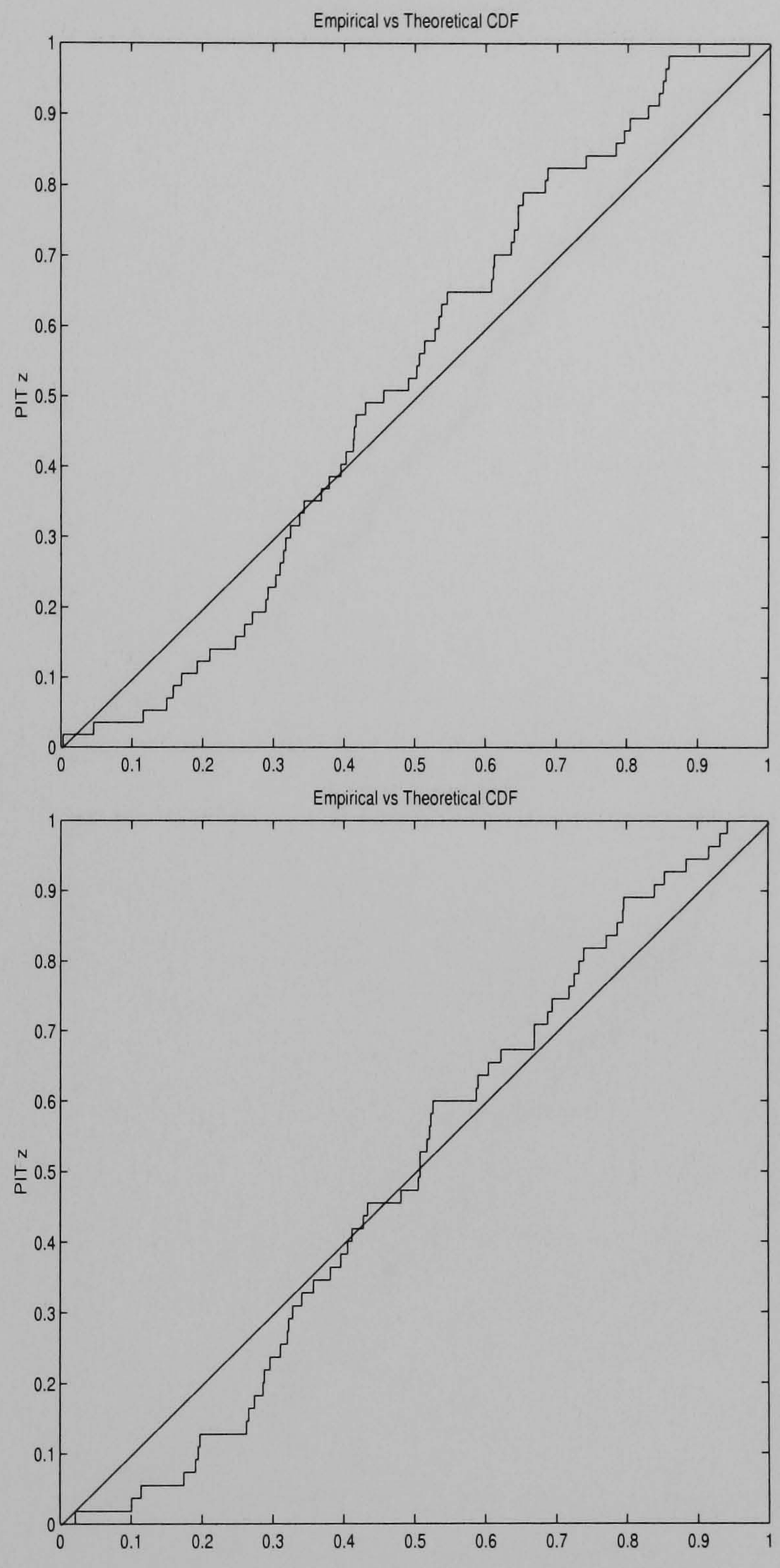


Figure 3.3: Graphical assessment of goodness-of-fit of mean-shifted risk-neutral NIG densities implied from index futures options: entire (top) and truncated (bottom).

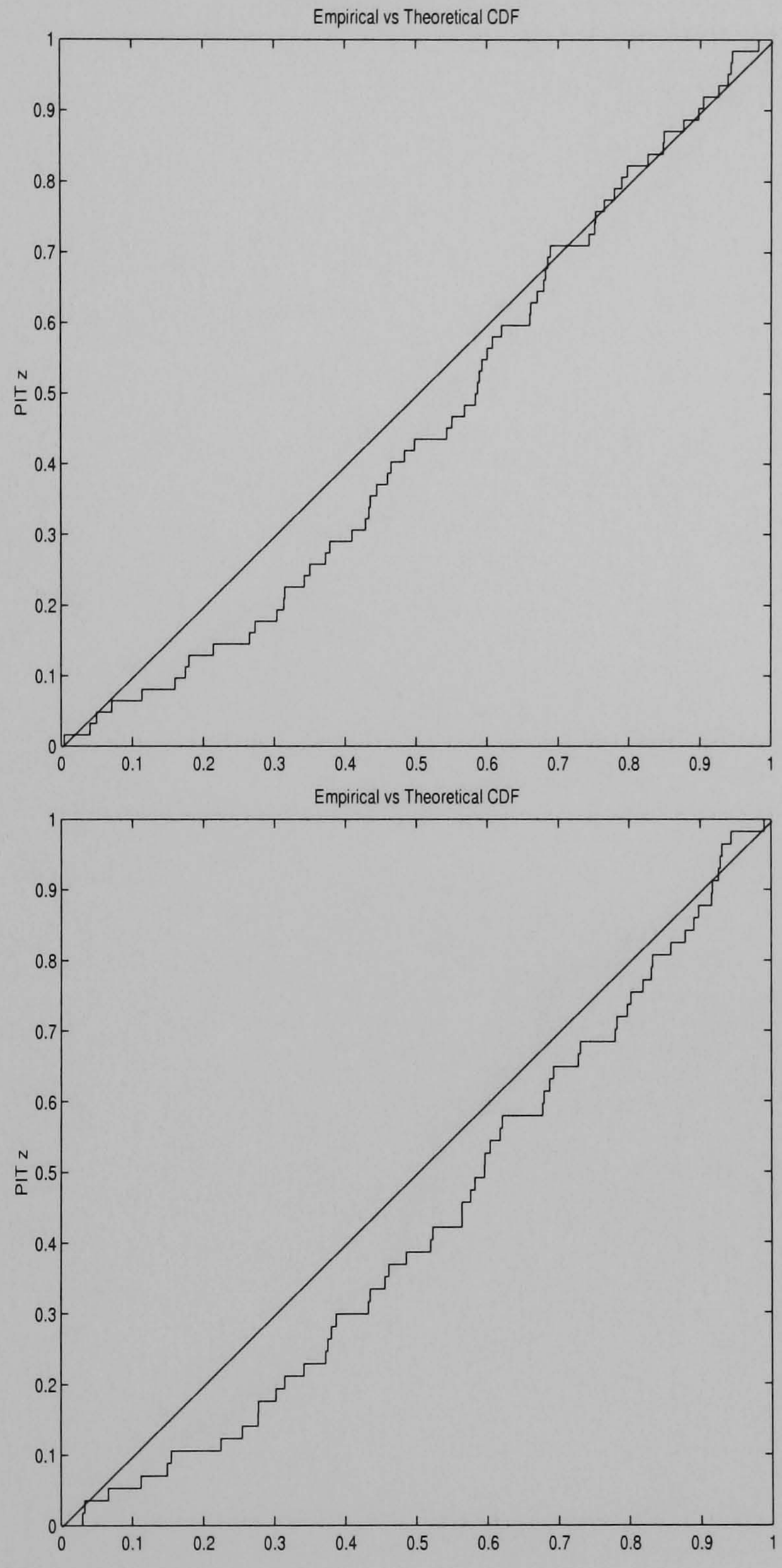


Figure 3.4: Graphical assessment of goodness-of-fit of risk-adjusted NIG densities implied from index futures options: entire (top) and truncated (bottom).

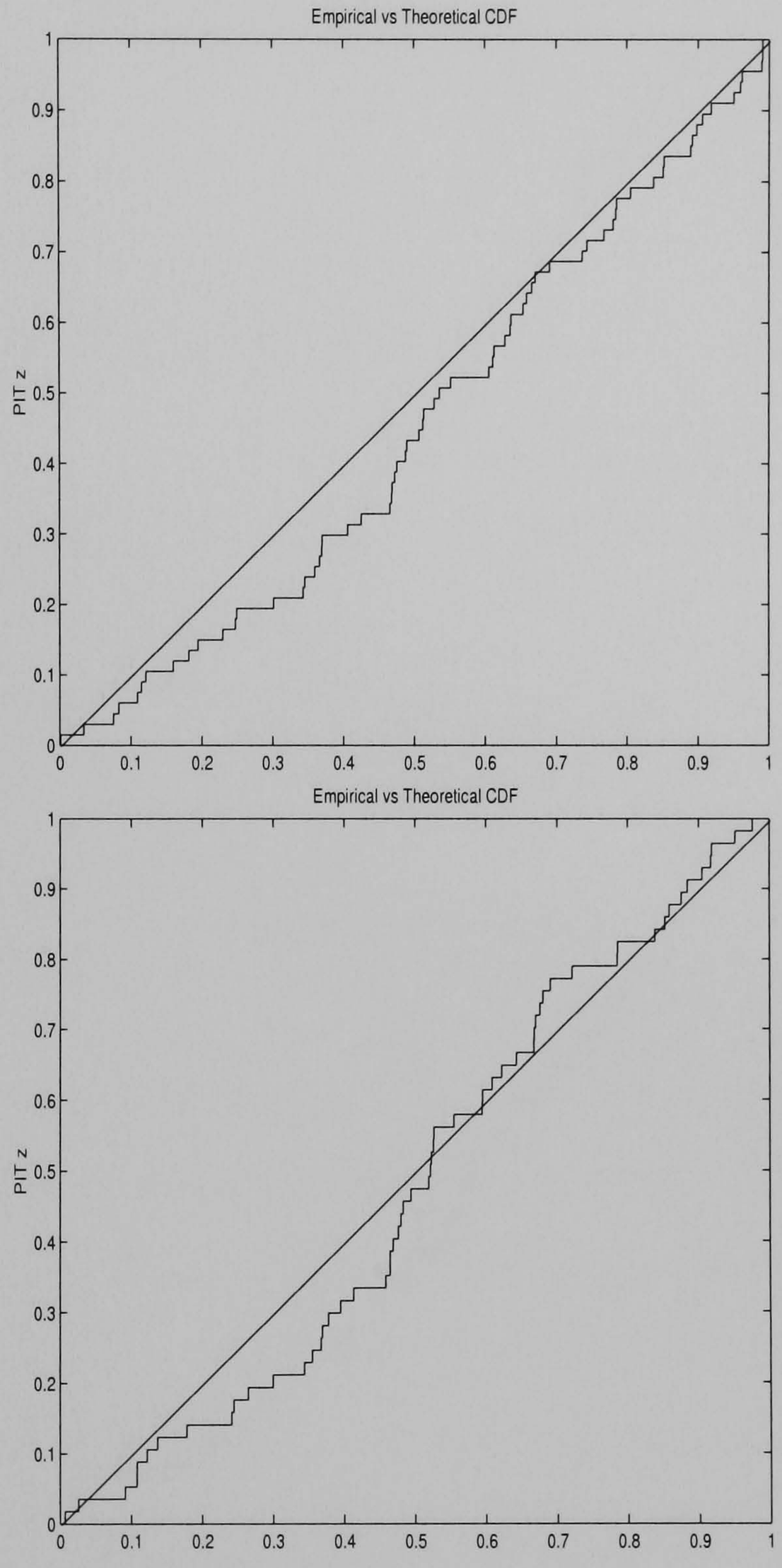


Figure 3.5: Graphical assessment of goodness-of-fit of historical-based density forecasts for currency futures: entire (top) and truncated (bottom).

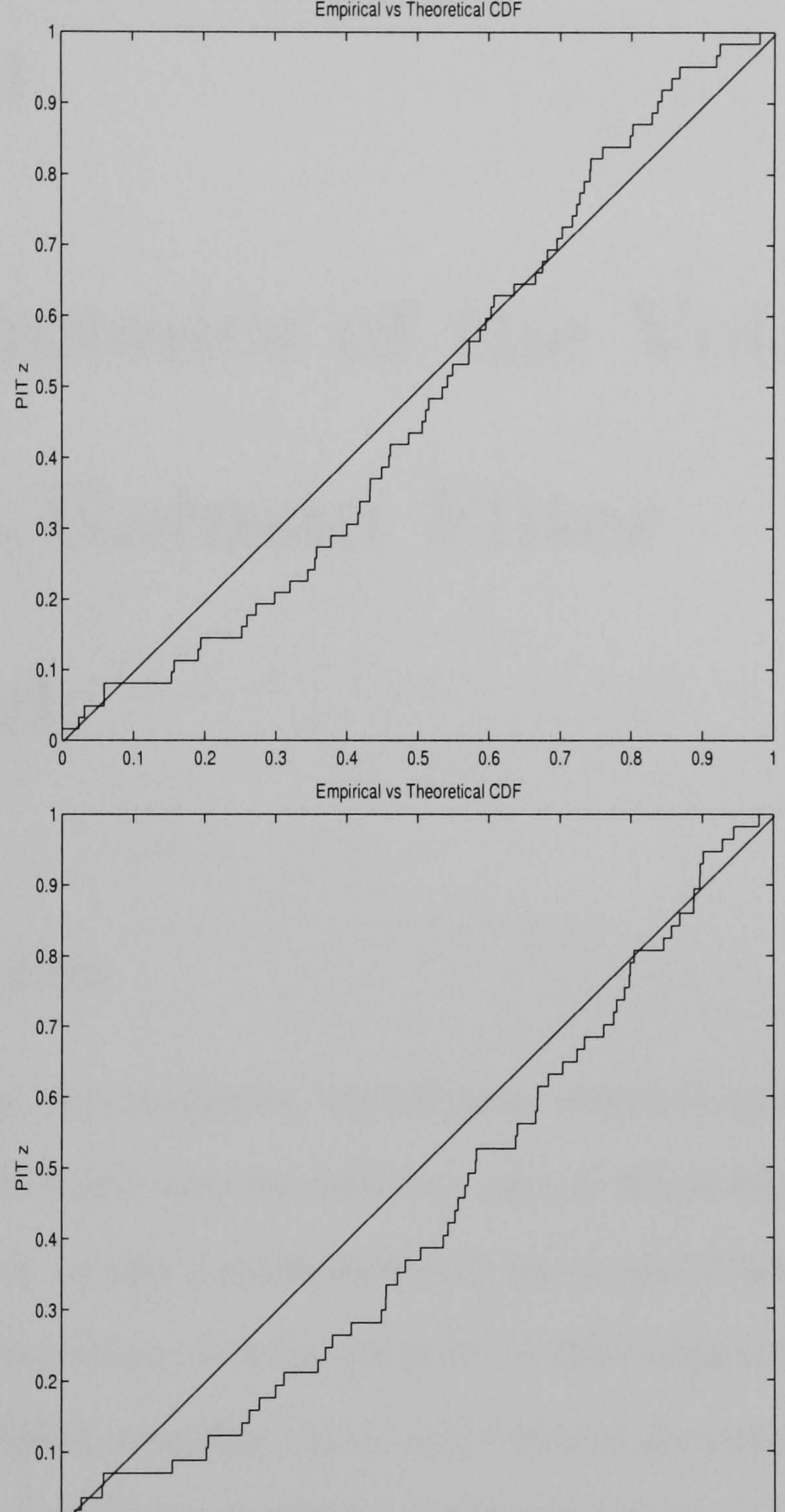


Figure 3.6: Graphical assessment of goodness-of-fit of historical-based density forecasts for index futures: entire (top) and truncated (bottom).

Chapter 4

The Dynamics of the Volatility Skew: a Kalman Filter Approach

4.1 Introduction

It is well known that the volatilities implied from observed option prices are not constant across strikes and time to maturity, as the Black-Scholes model would predict. Instead, they exhibit a smile/skew pattern across strikes for a given time to maturity, which extends to an entire volatility surface when different expiries are examined. These implied volatility curves and surfaces also change through time, raising the need for an accurate modelling of their dynamics, which is essential for the purposes of option pricing, trading and risk management.

The pricing of most volatility derivatives, such as options on implied volatility indices (VIX, VDAX and VXN are some examples) and forward start options[1] crucially depends on an accurate prediction of the implied volatility dynamics.

[1] But not volatility swaps.

Knowing how the volatility curve/surface is going to evolve as trades take place, is critical to traders, especially in fixed income and foreign exchange desks, who normally hold large exposures across a wide range of strikes.

From a risk management perspective, the risk arising from changes in implied volatility (vega risk) can represent the largest risk component of a portfolio containing options. Therefore, a precise modelling of the dynamics of implied volatility skews/surfaces (often neglected by risk managers) is of primary importance in order to obtain accurate risk measures.

In the present work, we suggest to model the temporal evolution of the volatility skew by means of a linear Kalman filter approach. We first fit cubic polynomials to the empirical volatility curves. The cubic's coefficients evolve through time according to a gaussian Ornstein-Uhlenbeck process correlated with the stock price process. The estimates for the skew coefficients are then promptly adjusted as information from new option trades becomes available.

Even though the Kalman filter technique seems a natural tool for financial problems of this kind, our study constitutes, to our knowledge, the first application of this powerful and robust econometric technique to the updating of the volatility skew. The approach we follow is intuitively simple and easy to implement. From an academic perspective, we are interested in appraising its performance at modelling the dynamics of the volatility curve, in order to decide whether it could represent a valid alternative to other approaches proposed by the relevant literature. For risk management purposes, our model is appealing because it accounts for changes not only in the level, but also in the slope, in the curvature and in the skewness of the implied volatility curve. For trading purposes, an important advantage of this technique is that the volatility curve can be accurately updated after observing few new option trades, instead of a complete set of new option prices for the entire range of strikes.

Since we do not possess data on a trade-by-trade basis, in this study we investigate a simple application of the technique to the risk management of portfolios of futures options on the S&P500. By implementing a standard Monte Carlo technique, we produce density forecasts of the daily changes in the marked-to-market value of four option portfolios, sensitive to shifts of different nature in the volatility curve. The distributions of the changes in value predicted by our model are then compared with the actual daily profits and losses on the option portfolios, across one year.

A comparison is also drawn with the changes in the marked-to-market value of the portfolios estimated according to two methods which are widely used in practice to incorporate volatility risk in a risk management analysis. The first benchmark model is the sticky-delta model by Derman [1999]. The second benchmark model applies a Taylor expansion of Black-Scholes option prices to account for first and second order changes in the underlying (delta and gamma), and first order changes in the volatility level (vega).

As we will see, producing accurate density forecasts of the vega risk exposures is a particularly hard task, and none of the models under examination proves to be successful for all of the four option portfolios. As a whole, the linear Kalman filter performs better than the two alternative methods, in terms of both the entire distributional forecast and the left tail only, which is relevant for Value-at-Risk (VaR) computations.

The chapter is structured as follows. Section 4.2 reviews the existing literature. Section 4.3 presents the data set that we use. In Section 4.4 we describe the Kalman filter model for the dynamics of the volatility skew, whereas the two alternative methods investigated in this study are presented in Section 4.5. Section 4.6 deals with the estimation of our model and the two benchmarks. Section 4.7 illustrates the criteria for building the option portfolios. Section 4.8 discusses and

evaluates the density forecasts obtained from the Monte Carlo exercise. Section 4.9 concludes.

4.2 A Review of the Literature

Given its high tractability, the Black-Scholes formula is still widely used in practice to price and hedge European options. The Black-Scholes model states a one-to-one relation between the price of a European option and its volatility. Therefore, option prices are often quoted by stating their implied volatility, which is the unique value of the volatility parameter that, placed into Black-Scholes formula, yields the option price observed in the market.

According to this model, implied volatility is constant across strikes and time to maturity. In reality, the profile of implied volatility across strikes for a given maturity is not flat, displaying instead a skew or a smile pattern (see Rubinstein [1985], Sheikh [1991]). Also, the level of implied volatility varies with time, changing the shape of the entire implied volatility surface (see Heynen [1993], Cont and da Fonseca [2001]).

In order to explain these empirical deviations from the Black-Scholes model, various attempts have been made in the literature, which can be broadly grouped along three directions.

One possible reason for the presence of a non-flat pattern is given by market frictions, which imply the existence of an entire band of arbitrage-free option prices. The role played by transaction costs, illiquidity and other trading restrictions in explaining the smile/skew was investigated and supported by Figlewski [1989a, 1989b], and Longstaff [1995]. In a related study, Constantinides [1998] pointed out that market frictions can account only partially for the existence of a volatility curve.

Alternative option pricing models in which the dynamics of the underlying

follows a process with jumps have been suggested, amongst the others, by Bates [1996], Carr, Geman, Madan and Yor [2002]. Since in the long term, positive and negative jumps tend to compensate each other, the introduction of jumps in the process seems especially appropriate for modelling the implied volatility smile/skew for short maturities.

A third class of models explains the existence of the volatility curve with the variability of the volatility over time, which can be either stochastic or deterministic. In stochastic volatility models (see, for example, Hull and White [1987], Scott [1987], Heston [1993]), the instantaneous (or local) volatility itself follows a stochastic process. If the volatility process is uncorrelated with the process for the stock price, a true smile occurs. A negative correlation between the two processes yields a skew profile, whereas an opposite pattern is the result of a positive correlation.

The main drawback of the majority of these models, also common to jump models, is the impossibility to express directly the shape of the implied volatility curve (or, equivalently, surface) in terms of the model parameters. Therefore, the calibration of the model parameters to a set of market option prices becomes very difficult, and unrealistic parameters are often required to generate volatility curves or surfaces that are consistent with those implied by observed option prices (see, e.g. Bakshi, Cao and Chen [1997], Andersen and Andreasen [2000], Das and Sundaram [1999]).[2]

The introduction of deterministic local volatility models, where the instantaneous volatility is modelled as a deterministic function of time and stock price, constituted a valid attempt to overcome this problem. By preserving market completeness, these local volatility models are self-consistent, arbitrage-free, and can be easily calibrated to match observed market volatility surfaces and curves. In the

[2]For stochastic volatility models, the failure to correctly reproduce empirical implied volatility skews/surfaces is particularly relevant for options with a short time to maturity.

non-parametric implied tree approach suggested by Dupire [1994, 1997], Derman and Kani [1994a, 1994b], Rubinstein [1994] and Derman, Kani and Chriss [1996], the focus is on extracting the binomial or trinomial tree that achieves an exact cross-sectional fit of the market option prices. An extension of the implied trinomial lattice to incorporate stochastic volatility, without making assumptions on the option pricing model and on the underlying process, was proposed by Britten-Jones and Neuberger [2000]. In order to obtain local volatilities that are smoother and more stable over time than those obtained by direct calibration, a number of semi-parametric and parametric specifications for deterministic local volatility models have been recently proposed in the literature (see, amongst others, McIntyre [2001], Beaglehole and Chebanier [2002], and the Normal Mixture Diffusion model of Brigo and Mercurio [2000, 2002] and Brigo, Mercurio and Sartorelli [2003]).

Local volatility models perform very well at a static level, by attaining an exact (for non-parametric specifications) or almost exact (for semi-parametric and parametric forms) calibration to the implied volatility surfaces observed on a given day. However, as documented by Dumas, Fleming and Whaley [1998], the dynamic behaviour of implied volatilities predicted by these models is inconsistent with the dynamics observed in the option markets. Besides raising the need for constant re-calibration of the model, this drawback leads, from a risk management perspective, to inaccurate and often unstable hedges of the option portfolios.

Motivated by the increasing evidence that option markets have become progressively more autonomous, showing movements in option prices driven not only by the underlying dynamics but also by specific sources of randomness, a new stream of literature has developed in the last few years. This innovative branch focusses on the identification of these extra sources of randomness in the option markets and consequently, on the investigation of their dynamics, also in relation to the dynamics of the underlying, in order to explain and capture the evolution of the empirical

implied volatilities.

The implied volatility then becomes a financial state variable by itself. Its dynamic properties have been studied mainly by focussing on either the term structure of at-the-money (ATM) implied volatilities or the volatility skews for a given maturity. Investigations of the dynamics followed by the entire volatility surface have begun to appear recently.

The most common approach to study the volatility dynamics consists in identifying the number and shapes of the shocks in the implied volatility via Principal Component Analysis (PCA). Avellaneda and Zhu [1997], Härdle and Schmidt [2000], Sylla and Villa [2000] applied PCA to the term structure of ATM implied volatilities. Alexander [2001] performed a PCA on the changes in the volatility smile for a given maturity, whereas Skiadopoulos, Hodges and Clewlow [1999] investigated three different maturity buckets. Recently, the PCA approach (Derman and Kamal [1997]) and some of its extensions, such as the Common Principal Component Analysis (Fengler, Härdle and Villa [2001]), the Independent Component Analysis (Ané and Labidi [2001]), and the Karhunen-Loève decomposition (Cont and da Fonseca [2002]) have been proposed to model the dynamics of the entire volatility surface.

Other empirical studies of the dynamics of the implied volatility curves (not based on PCA) include Heynen, Kemma and Vorst [1994] on the term structure of ATM volatilities, and Das and Sundaram [1999] on the term structure of implied skewness and kurtosis. Also, Peña, Rubio and Serna [1999] studied the dynamics of skews for the Spanish IBEX-35 options and Haffner and Wallmeier [2000] performed a similar analysis on DAX options. Tompkins [2001] searched for consistencies in the dynamics of volatility surfaces between different markets.

The most recent contributions involve the specification of a deterministic or stochastic model for the implied volatility smile or surface, which fully describes their evolution through time. The deterministic implied volatility models introduced by

Derman [1999] assume that either the per-delta or the per-strike implied volatility surface has a deterministic evolution. Due to their simplicity and tractability, the sticky-delta and the sticky-strike models are currently largely used by practitioners.

A stochastic evolution of the entire smile surface characterises the stochastic implied volatility models by Schönbucher [1999], Ledoit and Santa-Clara [1998] and Cont, da Fonseca and Durrleman [2002]. In these studies, the prices of liquid vanilla options across a set of strikes and maturities are taken as given, in order to derive the initial volatility surface. The dynamics of the implied volatility is then modelled as a joint diffusion process with the underlying. In the first two models, constraints are imposed on the drift processes followed by the implied volatilities to ensure absence of arbitrage. An arbitrage-free specification of future volatility smiles, when the process for the underlying is unknown, has been investigated by Rebonato and Joshi [2003]. In a much simpler approach, Rosenberg [2000] proposed a stochastic evolution for the ATM implied volatility only, while keeping the shape of the curve fixed.

Our research study fits in this latter class of stochastic implied volatility models, since we assume that the coefficients that identify the implied volatility curve evolve through time according to an Ornstein-Uhlenbeck process. However, we do not impose no-arbitrage conditions, in order to avoid the strong restrictions that would be necessary to satisfy them.

To our knowledge, the Kalman filter approach has never been used so far to model the dynamics of the implied volatility smile. The main application of this powerful and simple econometric technique in finance is the estimation of dynamic models of interest rates (see, for some recent examples, Duan and Simonato [1999], Ball and Torous [1996], Babbs and Nowman [2001], Cortazar, Schwartz and Naranjo [2003]).

From the risk management perspective, very little attention has been dedi-

cated to the effects of the dynamics of the volatility skew/surface on the vega risk of a portfolio of options, and on the possible interaction with other risk factors. Most studies simply adjust the traditional Value-at-Risk formulation to account for the Black-Scholes vega, without worrying about the existence of a skew/surface. Malz [2001] investigated a linear delta-vega VaR, Cárdenas *et al.* [1997] derived a closed formula for a delta-vega-gamma VaR, Glasserman, Heidelberger and Shahabuddin [2001] included vega risk in an efficient Monte Carlo simulation exercise. The most notable exception to this oversimplified treatment of the volatility risk is given by Malz [2001], who incorporated smile effects on VaR by letting implied volatilities of options with a different delta vary in correlated fashion. Also, Cont, da Fonseca and Durrleman [2002] and Fengler, Härdle and Villa [2001] hinted at the possibility of performing scenario simulations, for VaR purposes, of the joint evolution of the option portfolio and the underlying, on the basis of their principal components.

4.3 The Data Set

Our data set consists of daily data on quarterly[3] futures options on the S&P 500 index traded at the Chicago Mercantile Exchange over the period 1998-2001. The first three years are employed to estimate the parameters of the models, whereas the assessment of their out-of-sample forecasting properties is carried out over the last year.

We only use closing option prices on the quarterly contract closest to expiry, except for the days within two weeks to expiration, when we roll on to the next contract.

All of the no-arbitrage restrictions described in Section 2.4 apply here as well. Again, we use the Barone-Adesi and Whaley [1987] approximation for American

[3]March, June, September and December expiries.

options.

As before, we exclude in-the-money (ITM) options, which are more sensitive to non-synchronicity pricing errors.

We also filter out some options with extreme strikes, which may have very low liquidity. In particular, we eliminate the options traded at the minimum price (tick), as well as those options for which a change in the premium equal to the tick size yields a change in the corresponding implied volatility larger than 15%.

Data on the daily closing prices of the index futures are employed to estimate the model for the underlying.

4.4 The Kalman Filter Model for the Dynamics of the Volatility Curve

The first step of our analysis is the choice of a method for fitting the volatility curve as a function of the level of moneyness. We opt for a cubic polynomial, such that:

$$\sigma_K = \alpha_0 + \alpha_1 * m + \alpha_2 * m^2 + \alpha_3 * m^3 + \varepsilon \tag{4.1}$$

m denotes the measure of moneyness which, following Natenberg [1994], is expressed as $\frac{ln(K/F)}{\sqrt{\tau}}$, i.e. the natural log of the ratio between strike and underlying futures price, normalised by the square root of time to maturity.[4] Since we do not work with fixed time to maturity options, the normalisation corrects for the effects of τ shrinking over time, and yields more meaningful estimates for the polynomial coefficients.

In Eq. (4.1), σ_K denotes the market implied volatility for an option with strike K, α_0 is the estimated level of the ATM implied volatility (with log moneyness

[4]The time to maturity is computed as the number of calendar days to expiry divided by the total number of days in the calendar year.

equal to 0), and the coefficients α_1, α_2 and α_3 capture, respectively, the slope, curvature and skewness of the volatility skew.

The cubic is easy to implement, has only four coefficients of immediate interpretation, and provides a good fit to the observed volatility curves. In Fig. 4.1 we report, as an illustrative example, plots of both the market and the fitted implied volatilities for the March 1999 contract, with time to expiration of one, two and three months.[5]

The updating of the skew coefficients is performed by means of a discrete time linear Kalman filter (see Harvey [1989]). The 4×1 state vector of the coefficients, denoted as $\underline{x}_t$ evolves under the system equation:

$$\underline{x}_t = \underline{a}_t + U_t \underline{x}_{t-1} + C_t \underline{u}_t \tag{4.2}$$

$\underline{a}_t = A_t * \underline{\mu}_t$, where A_t represents the 4×4 diagonal matrix of the mean reversion coefficients and $\underline{\mu}_t$ is the vector of the long run means. $U_t = I - A_t$, and $C_t C_t' = V_t$ is the covariance matrix of the error terms of the process. $\underline{u}_t \sim N(0,1)$ are serially uncorrelated disturbances, independent of each other and of $\underline{x}_{t-1}$.

The skew coefficients are not directly observable. Instead, we observe an $n \times 1$ vector of market implied volatilities $\underline{\sigma}_{K,t}$, which is related to $\underline{x}_t$ by the observation equation:

$$\underline{\sigma}_{K,t} = G_t \underline{x}_t + D_t \underline{v}_t \tag{4.3}$$

where G_t is the $n \times 4$ matrix with elements $[G]_{i,j} = m_i^{j-1}$, $D_t D_t' = R_t$ represents the $n \times n$ measurement noise covariance matrix, and $\underline{v}_t \sim N(0,1)$ are again serially uncorrelated disturbances, independent of each other and of $\underline{x}_t$. Since the number of observed implied volatilities can be small (even one only), this method is particularly suitable for practitioners, who need to update the volatility skew as soon as new trades occur.

[5]More sophisticated methods, such as the cubic B-splines, would provide a better fit, but the identification and the updating of the curve coefficients become much more involved.

We use $\underline{\hat{x}}_t$ and S_t to denote, respectively, $E_{t-1}[\underline{x}_t]$ and $\mathrm{Var}_{t-1}[\underline{x}_t]$, before observing $\underline{\sigma}_{K,t}$. Finally, we assume that the initial distribution of $\underline{x}_1$ is multivariate Normal with known values for $\underline{\hat{x}}_1$ and S_1.

According to this state space model, our optimal forecast $\underline{\hat{x}}_t$ for the skew coefficients $\underline{x}_t$ (and their covariances) is first adjusted according to the observed $\underline{\sigma}_{K,t}$. This updated quantity then evolves under the system equation, to produce the new optimal forecast $\underline{\hat{x}}_{t+1}$ for the next period. The updating equations for both expected value and covariance are given by:

$$\underline{\hat{x}}_{t+1} = U_{t+1}\left[I - S_t G_t' T_t^{-1} G_t\right]\underline{\hat{x}}_t + K_t \underline{\sigma}_{K,t} + \underline{a}_{t+1} \tag{4.4}$$

$$S_{t+1} = U_{t+1}\left[S_t - S_t G_t' T_t^{-1} G_t S_t\right] U_{t+1}' + C_{t+1} C_{t+1}' \tag{4.5}$$

where $T_t = G_t S_t G_t' + D_t D_t'$, and $K_t = U_{t+1} S_t G_t' T_t^{-1}$.

At each time t, after observing the market volatility skew, we produce one-day-ahead forecasts for the expected value of the coefficient vector $\underline{\hat{x}}_{t+1}$ and the covariance matrix S_{t+1}. In fact, at each step we produce complete density forecasts of the evolution of the skew, since $\underline{x}_{t+1} \sim N(\underline{\hat{x}}_{t+1}, S_{t+1})$.

In order to translate the density forecasts of the skew (expressed in terms of log moneyness) into forecasts for the distribution of the daily changes in the value of our option portfolios (expressed in terms of strikes), we also need to produce one-day-ahead forecasts of the distribution of the log returns on the underlying futures.

To keep our analysis simple, we choose to model the dynamics of the underlying as an EGARCH(1,1) model with normally distributed errors.[6]

According to the EGARCH model, introduced by Nelson [1991] to account for the presence of a leverage effect in equity markets, the specification for the log

[6] We also attempted to use an EGARCH model with Student's t errors, but the resulting forecasts for returns and volatilities turned out to be too volatile.

returns and the conditional variance is given by:

$$r_t = \sigma_t \epsilon_t \tag{4.6}$$

$$\ln(\sigma_t^2) = \omega + \alpha \left(\left| \frac{r_{t-1}}{\sigma_{t-1}} \right| - \sqrt{\frac{2}{\pi}} \right) + \gamma \frac{r_{t-1}}{\sigma_{t-1}} + \beta \ln(\sigma_{t-1}^2) \tag{4.7}$$

where $\epsilon_t \sim N(0,1)$, and can be correlated with the disturbances in Eq. (4.2). Values of $\gamma < 0$ denote the existence of a leverage effect, since the impact on the volatility of negative past returns is larger than that of positive returns.

4.5 Two Alternative Models for the Dynamics of the Volatility Curve

The performance of the Kalman filter method at describing the dynamics of the volatility skew is compared with that of two alternative models widely used in practice.

4.5.1 The sticky-delta model

The sticky-delta model was first discussed by Derman [1999], who described it as the method commonly used by traders to predict the evolution of the volatility curve in a situation of stable trending markets.

Following the sticky-delta model, the volatility curve for a fixed maturity is parametrised as:

$$\sigma_K = \sigma_{ATM} - b(K - F) \tag{4.8}$$

where $b > 0$ yields the volatility skew. The fixed strike volatilities increase with the futures level, whereas σ_{ATM} is independent of the underlying and fixed. Therefore, the volatility curve changes if measured in terms of strikes, but it is fixed if measured with respect to moneyness (or, equivalently, delta).

The only factor that affects the volatility skew in this very popular model is the evolution of the underlying asset. As before, we model the dynamics of the equity futures with an EGARCH specification with normal errors.

4.5.2 The vega-gamma model

One of the standard approaches adopted by risk managers in order to measure the risk of an option portfolio, involves applying a Taylor series expansion of Black-Scholes option prices around the risk factors. The variables that affect option prices are the changes in the value of underlying asset and volatility (which are stochastic), and the changes in the time to expiry (which are deterministic).

According to this technique, the change in value of the option portfolio $\Delta\Pi$ is approximated by:[7]

$$\begin{aligned}\Delta\Pi &= \frac{\partial\Pi}{\partial F}\Delta F + \frac{\partial\Pi}{\partial\tau}\Delta\tau + \frac{\partial\Pi}{\partial\sigma}\Delta\sigma + \frac{1}{2}\frac{\partial^2\Pi}{\partial F^2}(\Delta F)^2 \\ &= \delta\Delta F + \Theta\Delta\tau + V\Delta\sigma + \frac{1}{2}\Gamma(\Delta F)^2\end{aligned}$$

where δ (delta), Θ (theta), V (vega) and Γ (gamma) represent the sensitivities of the portfolio's value to variations in the risk factors.

Since, in our case, we deal with delta-neutral portfolios and we consider profits and losses on the portfolios at a daily level, the approximation can be simplified to:[8]

$$\Delta\Pi = V\Delta\sigma + \frac{1}{2}\Gamma(\Delta F)^2 \tag{4.9}$$

The computation of both V and Γ relies on specific assumptions on the option valuation model. Given that we work with futures options, we use Black's model for option pricing. Call and put option prices are given by:

$$C(K) = \mathrm{e}^{-r\tau}[F_t N(d_1) - K N(d_2)] \tag{4.10}$$

[7]The remaining higher order terms of the series expansion are usually neglected.

[8]We have also performed the calculations including the effect of changing time to maturity, but the difference in the results was negligible.

$$P(K) = \mathrm{e}^{-r\tau}[KN(-d_2) - F_t N(-d_1)] \tag{4.11}$$

where $N(\cdot)$ denotes the gaussian cumulative distribution function and d_1, d_2 are computed as:

$$d_1 = \frac{\ln(F_t/K) + (\sigma^2\tau)/2}{\sigma\sqrt{\tau}}$$

$$d_2 = d_1 - \sigma\sqrt{\tau}$$

The calculation of the appropriate partial derivatives yields the following formulations for V and Γ of both call and put options:

$$\Gamma = \frac{N'(d_1)\mathrm{e}^{-r\tau}}{F_t\sigma\sqrt{\tau}} \tag{4.12}$$

$$V = F_t\sqrt{\tau}N'(d_1)\mathrm{e}^{-r\tau} \tag{4.13}$$

where $N'(\cdot)$ is the first derivative of the cumulative distribution function (i.e. the normal probability density).

The vega (gamma) exposure of a single option position is computed as the option vega (gamma) times the number of options in the position. The vega (gamma) exposure of a portfolio of options on the same underlying, which is our case, is simply the sum of the vega (gamma) exposures of the single options.

In the vega-gamma model there are two sources of risk, whose dynamics need to be modelled: the underlying and the implied volatility. Only parallel changes in the volatility curve are allowed, with the implied volatilities of all options changing by the same amount $\Delta\sigma$.

Following a conventional approach, we assume that changes in the log of the implied volatility are normally distributed with zero mean and standard deviation $\sigma_{vol,t}$. We let $\sigma^2_{vol,t}$ evolve through time according to a simple GARCH(1,1) model with normal errors:

$$\sigma^2_{vol,t} = \omega + \alpha(\Delta \ln \sigma_{t-1})^2 + \beta\sigma^2_{vol,t-1} \tag{4.14}$$

For consistency with the other two methods, the evolution of the underlying is modelled by means of an EGARCH model with normal errors, correlated with the disturbances in the volatility process.

It is worth noticing that both benchmark models can be interpreted as special cases of the more general Kalman filter specification. The sticky-delta model represents an oversimplified version, where all the volatility skew coefficients are constant, and the only factor which evolves is the underlying asset. The vega-gamma model constitutes a special case, where both the ATM level α_0 and the underlying evolve stochastically (and in a correlated fashion). The remaining coefficients of the volatility curve are instead constant. As an alternative to the approach followed in this work (i.e. individual estimation and assessment of the single models), the validity of the different specifications could therefore be assessed by testing the appropriate restrictions in the more general model.

4.6 The Estimation of the Models

In this section we present the results from the estimation of the parameters of the Kalman filter, the sticky-delta and the vega-gamma models, over the period 1998-2000.

The parameters of the Kalman filter model can be constant, or they can vary at each time step. In our application, the first assumption would probably be too strong, since we test our model over one year, and the second one would be unnecessarily computationally intensive, given that we do not expect the model parameters to change significantly on a daily basis. Therefore, we choose to re-estimate the parameters $A_t, \underline{\mu_t}$ and V_t every quarter, over a rolling window of three years of data.[9] The measurement error matrix D_t, instead, is kept constant, since

[9]The four estimation windows are: Jan.98 - Dec.00, Mar.98 - Feb.01, Jun.98 - May01, Sep.98 -

its quarterly estimates turned out to be not significantly different from each other.

For the estimation of the Kalman filter model we first calculate the skew coefficients for each day of the estimation window. Plots of the time series of the four coefficients are reported in Fig. 4.2, and their autocorrelation functions are displayed in Fig. 4.3.[10] For all the coefficients, the autocorrelation function shows an exponential decay which can be conveniently represented by an AR(1)/Ornstein-Uhlenbeck process. Therefore, the assumption made for the system equation in (4.2) seems appropriate.[11]

The skew coefficients are then employed to estimate A_t, $\underline{\mu_t}$ and V_t from the system equation, by means of a Vector Auto-Regressive (VAR) model. The estimates for A_t and $\underline{\mu_t}$, computed on each of the four estimation windows, are reported in Table 4.1. Table 4.2 refers to the correlation matrices associated with V_t, which are more intelligible than the corresponding covariance matrices.

As already suggested by the inspection of the autocorrelation functions, the slope coefficient reverts towards its long run mean very slowly. A slightly faster mean reversion characterises the ATM volatility level, whereas considerably higher (although slightly decreasing through time) mean reversion coefficients are estimated for both curvature and skewness. The analysis of the correlation matrix of the error terms of the process reveals that the highest correlation occurs between curvature and skewness ($\simeq 0.8$). A correlation of around 0.45 relates curvature and slope. Smaller correlations are found between ATM level and curvature, and ATM level and slope. In both cases, however, the correlation coefficients increase (in absolute terms) through time, from a value of, respectively, -0.20 and -0.27 to a level of

---

Aug.01.

[10]The plots refer to the first estimation period, but the patterns stay the same if we consider the other three windows.

[11]The spikes in both time series and autocorrelation function for the slope coefficient are due to the effect of the change in time to maturity when we roll on to the following contract, not completely captured by the normalisation of the log moneyness measure.

-0.36 and -0.34. Very small correlations are found between slope and skewness, and ATM level and skewness.

Considering that the fit of the cubic polynomial to the market implied volatility skew is better the closer we are to the ATM level, and less accurate as far OTM we move, the estimate for the measurement error covariance matrix is based on a grouping in buckets of log moneyness. We select ten buckets of log moneyness, in consideration of the ranges of log moneyness in our data set, and for each of them we calculate the standard deviation of the measurement error across the estimation period. The results, displayed in Table 4.3, suggest that there is a significant variability in the estimated measurement error volatility which, for deep OTM puts and calls, turns out to be, respectively, four and five times its value for ATM options (0.2%).

The estimates of the coefficients of the EGARCH model for the dynamics of the underlying are presented in Table 4.4 (with standard errors in brackets). As expected, there is a statistically significant leverage effect in our series, captured by the coefficient γ.

In order to investigate whether the dynamics of underlying asset and skew coefficients are related, we calculate empirical correlation coefficients between the futures daily log returns and the daily changes in each of the skew coefficients, over the four estimation windows. Only the correlation between log returns and ATM volatility level turns out to be statistically significant, and basically constant around the value of -0.82 for all the four estimation periods.

For the purposes of the sticky-delta model, only the estimates of the EGARCH model are needed, since the underlying is the only stochastic variable.

For the vega-gamma model, the estimates of the EGARCH specification are relevant for modelling the dynamics of the underlying. As discussed in the previous paragraph, we assume that the log changes in the implied volatility follow a normal

distribution with zero mean and standard deviation $\sigma_{vol,t}$, which evolves according to a GARCH(1,1) model. The estimates for the GARCH model on the log changes in the ATM implied volatility are reported in Table 4.5.[12] To account for the correlation between the two risk factors, we use again a correlation coefficient of -0.82.

4.7 The Option Portfolios

In order to assess the goodness of the linear Kalman filter technique at modelling the dynamics of the volatility curve for risk management purposes, we test how well this method predicts the actual daily variations in the marked-to-market value of option portfolios sensitive to changes of different nature in the volatility curve.

We consider the following four option portfolios:

- A short straddle (short one call and one put ATM), which is sensitive to changes in the level of the ATM implied volatility. The portfolio's value decreases (increases) when the volatility level goes up (down).

- A long risk reversal (short one OTM put and long one OTM call), sensitive to changes in the slope of the volatility smile. A loss (profit) occurs when the slope increases (decreases).

- A long butterfly spread (long one call and one put OTM, short one call and one put ATM), sensitive to changes in the curvature of the volatility smile. The portfolio loses (gains) value when the curvature decreases (increases).

- A long "Mexican hat" (long two calls and two puts OTM, short one call and one put ATM), which is vega-neutral.

[12] Standard errors in brackets.

Each portfolio is made delta-neutral by assuming the appropriate position in the underlying future.[13]

For practical implementation, the ATM options are those with strike price closest to the current level of the underlying. The choice of the OTM options for our portfolios is based on two levels of moneyness (defined, for this purpose, as the ratio between the strike and the underlying futures price), one for the OTM puts and one for the OTM calls, equidistant from the ATM level. A trade-off exists between choosing OTM options too close to the ATM, which are highly liquid also for a short time to expiry but not very sensitive to non-parallel changes in the volatility curve, and selecting too far OTM options, which are very sensitive to different sources of changes in volatility, but become too illiquid as time to maturity approaches. In our case, a ratio K/F of 0.92 for OTM puts and of 1.08 for OTM calls seems to represent a satisfactory compromise.

Each day of the testing period (year 2001) we build the four delta-neutral portfolios from our data set, and we calculate their marked-to-market value on both that day and the following day, in order to compute the actual change in value.

4.8 Density Forecasts of the Changes in Option Portfolios' Value

For each of the three models, parametrised according to the estimates obtained in Section 4.6, we produce daily forecasts of the changes in value of the four option portfolios, over the year 2001.[14] In order to assess these distributional forecasts not only in relative, but also in absolute terms, a comparison is drawn with the actual

[13]The analysis of the corresponding portfolios of opposite sign (long straddle, short risk reversal, short butterfly spread, short "Mexican hat") would be redundant. For completeness, we will report the upper percentiles of the distributions of the changes in value for the original portfolios, which correspond to the lower percentiles of the distributions for the portfolios with opposite sign.

[14]It is worth emphasising that in our work we only consider out-of-sample density forecasts.

daily changes in the portfolios' value across the year. Particular attention is paid to the tails of the density forecasts, which are especially relevant for risk management purposes.

4.8.1 Derivation of the density forecasts

The density forecasts of the changes in option portfolios' value are derived by means of a simple Monte Carlo simulation exercise.

In the Kalman filter model, we start with initial estimates $\underline{\hat{x}}_1$ and S_1, at the beginning of the testing period. For simplicity, the coefficients of the cubic fitted to the empirical volatility curve on the last day of the estimation period are chosen as $\underline{\hat{x}}_1$. S_1 is set equal to the covariance matrix of the stationary distribution of the multivariate Ornstein-Uhlenbeck process in Eq. (4.2), $S_1 = C(A + A')^{-1}C'$. Subsequent estimates $\underline{\hat{x}}_t$ and S_t are obtained through the updating Eqs. (4.4) and (4.5).

Each day t we draw $5,000$ correlated samples from $N(\underline{\hat{x}}_t, S_t)$ for the skew coefficients, and $N(0, \sigma_t)$ for the log returns on the underlying future. Once we possess forecasts for the underlying level at time t, we can calculate the log moneyness (again, at t) for the strikes of the options included in the four portfolios at time $t-1$. We then obtain forecasts for the corresponding implied volatility levels, via the cubic polynomial in Eq. (4.1), with coefficients equal to the coefficient forecasts for time t. The resulting implied volatilities are placed into Black's formula to derive the forecasts for the marked-to-market value of the options in the portfolios.[15] Since we repeat this procedure $5,000$ times, we obtain an entire density forecast of the changes in the marked-to-market portfolios' value.

In order to produce density forecasts from the sticky-delta model, each day

[15]For a call (put) which was ATM at time $t-1$, and becomes ITM at t, we place the implied volatility for that strike into the formula for the corresponding OTM put (call), and then derive the call (put) price via put-call parity.

t we draw $5,000$ samples from $N(0, \sigma_t)$, the distribution of the log returns on the underlying. For each sample we compute the forecasted value of the underlying at t, and the new levels of log moneyness for all the options included in the four portfolios at time $t-1$. We then move along the volatility curve observed at $t-1$. The implied volatilities corresponding to the updated levels of log moneyness are placed into Black's formula to obtain forecasts of the new marked-to-market option prices.[16]

In the vega-gamma model, each day we draw $5,000$ correlated samples from $N(0, \sigma_{vol,t})$ for the log changes in the volatility level, and $N(0, \sigma_t)$ for the log returns on the underlying future. We then obtain $5,000$ pairs $(\Delta\sigma, (\Delta F)^2)$ which, combined with the vega and gamma exposures (V and Γ) computed for the four option portfolios at $t-1$, yield density forecasts of the changes in portfolios' value.

For illustrative purposes, we report in Figs. 4.4 - 4.7 histograms of the density forecasts of changes in the marked-to-market value of our four portfolios on the 20th August 2001, for each method. As expected, the various models produce significantly different distributional forecasts. In general, the sticky-delta model generates highly skewed distributions, with a very thick left or right tail. Less skewed densities are obtained from both the Kalman filter and the vega-gamma models. For two portfolios (delta-hedged risk reversal and "Mexican hat") this latter method also predicts much smaller changes in value than the competing models.

4.8.2 Assessment of the density forecasts

At this stage of the research, we want to assess whether the density forecasts produced by the Kalman filter method represent unbiased forecasts of the actual daily changes in the value of the four option portfolios. These density forecasts are also compared to the corresponding forecasts obtained from alternative specifications

[16]Former ATM options that become ITM are treated as described in the previous footnote.

which only model parallel changes in the volatility curve. First we discuss the testing techniques that will be employed, and then we appraise the performance of the different models for each portfolio.

As explained in Section 3.2, if a sequence of one-step-ahead density forecasts is correct, then the series of Probability Integral Transforms $z_t \sim$ i.i.d. $U(0,1)$ or, equivalently, the series of the normalised transforms $x_t \sim$ i.i.d. $N(0,1)$. In Figs. 4.8 - 4.11 we plot the empirical cumulative distribution functions (CDF) of the z_t series obtained from each model, against the theoretical 45° line, for all option portfolios. The visual inspection of the plots immediately indicates that none of the three models under investigation correctly forecasts the daily profits/losses for each portfolio. In fact, almost all our density forecasts appear to be affected by some kind of bias, which seems to suggest that forecasting daily changes in value of option portfolios is not a straightforward task.

In order to be more precise about the different nature of such biases, the graphical approach is integrated with a more formal analysis. Again, we use log likelihood ratio (LR) tests on the transformed x_t series. The LR tests for independence ($LR1$), and for the joint hypothesis of i.i.d.$(0,1)$ ($LR2$) were reported in Eqs. (3.9) and (3.10). Similarly, the LR tests for zero mean ($LR3$), and for unit variance ($LR4$) are given by:

$$LR3 = -2(L(0,\hat{\sigma}^2,\hat{\rho}) - L(\hat{\mu},\hat{\sigma}^2,\hat{\rho})) \sim \chi^2(1) \tag{4.15}$$

$$LR4 = -2(L(\hat{\mu},1,\hat{\rho}) - L(\hat{\mu},\hat{\sigma}^2,\hat{\rho})) \sim \chi^2(1) \tag{4.16}$$

The results from the implementation of the LR testing techniques are displayed in Table 4.6. The values for the $LR1$ test indicate that the independence of the PIT series is never a problem. Instead, the joint hypothesis of i.i.d.$(0,1)$ is rejected in all cases but for the density forecasts of changes in value of the butterfly spread produced by the Kalman filter model. This evidence highlights the existence

of biases in the first two moments of our distributional forecasts. Since the rejection of the null hypothesis in LR tests is only based on the first two moments, we also compute a Jarque-Bera test for normality. The outcomes of the Jarque-Bera test reveal that the normality of the PIT series is always rejected. In the light of these results, detecting potential misspecifications in higher moments of the distributional forecasts becomes essential.

For a better comparison between actual and forecasted daily profits and losses, we then compute basic summary statistics for both the time series of the actual daily changes in the values of the portfolios, and the three equally-weighted mixtures (one for each model) of the single density forecasts derived for each day of the testing period.[17] The summary statistics (mean, standard deviation, variance, skewness and excess kurtosis) for the four option portfolios are reported in Table 4.7.

The analysis of the tails of the density forecasts is also relevant, both as part of a more general and complete assessment of the goodness of our forecasts, and for VaR computations. Following Barone-Adesi, Giannopoulos and Vosper [2002], in order to evaluate whether our density forecasts are appropriate for the calculation of conventional risk measures, we proceed as follows. We compute the 1-day VaR at both 99% and 95% confidence levels as, respectively, the 1^{st} and the 5^{th} percentile of the forecasted distribution of the changes in value of the option portfolios. We then record the number of breaks over the entire testing period, which occur when the actual loss is larger (in absolute value) than the estimated VaR. We also compute the 95^{th} and the 99^{th} percentiles, together with the corresponding number of breaks, which are relevant in terms of VaR calculations for the option portfolios of opposite sign. The number of breaks recorded for both lower and upper percentiles of the

[17] This method yields more accurate results than the simple averaging of the summary statistics of the individual density forecasts across time.

density forecasts obtained from the different models, as a percentage of the total number of days in the testing period, are displayed in Table 4.8, for each of the option portfolios. If the VaR forecasts were correct, we would expect a percentage of breaks of around 1% for both the 1^{st} and the 99^{th} percentiles, and of 5% for both the 5^{th} and the 95^{th} percentiles. Again, none of the three models produces exact forecasts of both lower and upper tails for each of the option portfolios of interest.

Having described the evaluation techniques, we can now focus on the appraisal of the density forecasts of the changes in value in each of our four option portfolios.

The Short Straddle

The summary statistics in Table 4.7 indicate that the time series of the actual changes in the value of the short straddle position is highly skewed to the right (skewness of -2.31) and leptokurtic (excess kurtosis of 8.70). The density forecasts which seem to best reproduce these features are those generated by the sticky-delta model, which however present a significant downward bias in the mean (-0.25 against a realised mean of 0.24), as confirmed by the value of the $LR3$ statistic. The consequent shift of the density forecast to the left induces the misspecification in the right tail reported in Table 4.8, and also evident from the CDF plot.

The mixture of density forecasts obtained from the Kalman filter model predicts the mean correctly, but does not seem to exhibit enough skewness (-1.25) and fat-tailness (3.35), compared to the time series of the actual changes in the portfolio's value. The variance is also significantly biased upwards (a value of 17 against an empirical sample variance of 8), as indicated by both the $LR4$ statistic and the shape of the CDF plot. The high volatility explains the low percentages of VaR breaks in the tails.

The vega-gamma model produces density forecasts which are even more sym-

metric (skewness of -0.29) and less fat-tailed (excess kurtosis of 2.07) than the Kalman filter's ones. Also, all our evaluation tools highlight the existence of significant misspecifications in both the mean (underestimated at -1.30) and the variance (overestimated at 25) of the mixture of vega-gamma forecasts. Again, the small number of breaks in the tails is attributable to the overestimation of the variance.

The Long Risk Reversal

The distribution of the realised changes in the value of the long risk reversal portfolio over time turns out to be slightly negatively skewed (skewness of -0.80) and leptokurtic (excess kurtosis of 0.74). The results from the LR tests, the summary statistics, and the plots of the PIT series, show that the density forecasts derived from both the Kalman filter and the sticky-delta model are affected by significant upward biases in the mean and in the variance, which are larger for Derman's specification.[18] Both models capture the skewness quite closely, whereas the sticky-delta model produces too fat-tailed density forecasts. Due to the upward bias in the mean, the distributions of the changes in the portfolio's value are shifted to the right with respect to the realised changes, causing a high percentage of VaR breaks in the left tail (up to almost 20% at 95% confidence level for the sticky-delta model).

The mixture of density forecasts from the vega-gamma model is correctly centered on the actual mean, but it displays a very large downward bias in the variance (estimated variance of 0.05 against a sample variance of 0.27), also evident from the visual inspection of the CDF plot. Moreover, the density forecasts are essentially symmetric and much more fat-tailed than the distribution of the actual changes over time. The large underestimation of the dispersion measure explains the severe misspecification of both tails, with percentages of VaR breaks (at 95%

[18]The mean (variance) of the mixture of density forecasts from the Kalman filter and the sticky-delta model are, respectively, 0.32 (0.43) and 0.41 (0.50), against an empirical sample mean (variance) of 0.07 (0.27).

confidence level) up to 29% in the left tail and 23% in the right tail.

The Long Butterfly Spread

The sample summary statistics of the time series of the actual changes in the value of the long butterfly spread portfolio reveal the presence of a moderate negative skewness (−0.60) and a small excess kurtosis (0.92). Similar higher moments are reported for the mixture of density forecasts generated from the Kalman filter approach, which, however, are affected by a small downward bias in the mean (−0.01 against an empirical sample mean of 0.14) and an upward bias in the variance (2.30 against a variance of the actual profits and losses of 1.50).

Much less accurate density forecasts are obtained from either the sticky-delta or the vega-gamma model. A much more asymmetric (skewness of −1.81) and leptokurtic (excess kurtosis of 10) shape characterises the mixture of distributional forecasts from the first benchmark model. The upward bias in the dispersion measure is smaller than for the Kalman filter model, but the downward bias in the mean is larger. The misspecifications in both the mean and the shape of the density forecasts explicate the high percentage of VaR breaks (around 15% at 95% confidence level) in the right tail.

The density forecasts from the vega-gamma model seem to capture the asymmetry of the actual distribution quite accurately, but they are too fat-tailed (excess kurtosis of 6.13) and clearly misspecified in the first two moments. In fact, as immediately suggested by a simple comparison of the CDF plots, they exhibit the worst biases in both the mean (−0.49) and the variance (2.92), amongst our three specifications. The higher number of VaR breaks recorded in the right tail than in the left one, is mainly due to the downward bias in the mean, which has the effect of shifting the distribution to the left.

The Long "Mexican Hat"

The time series of the actual daily changes in the value of the long "Mexican hat" vega-neutral position displays a sample mean very close to zero (0.04), a sample variance of 2, and very pronounced higher moments (skewness of 2.87 and excess kurtosis of 14.89). Similar values for the higher moments can be observed only for the mixture of density forecasts generated from the sticky-delta model, which, however, exhibits a significant downward bias in the variance (1.72), easily detectable from the CDF plot. The joint effect of the bias in the variance and a small (not statistically significant) downward bias in the mean, induces the misspecification in the tails reported in Table 4.8.

The mixtures of density forecasts generated from both the Kalman filter and the vega-gamma model are instead significantly more symmetric (skewness coefficients around 0.75) and less leptokurtic (excess kurtosis around $4-5$). The Kalman filter model predicts the variance correctly, but it significantly underestimates the mean (-0.14). On the contrary, the density forecasts from the vega-gamma model display an obvious upward misspecification of the variance (2.55). For both models, the relatively high percentage of VaR breaks in the right tail (around 9% at 95% confidence level) is a consequence of the failure to correctly capture the strong positive asymmetry observed in the actual stream of profits and losses.

As already anticipated, none of the three models investigated here generates unbiased density forecasts of the daily changes in the value for each of the four option portfolios. Therefore, in absolute terms, a winning model cannot be identified.

In relative terms, we express a preference for the discrete Kalman filter as a model for the dynamics of the volatility skew. According to our results, this method outperforms[19] the alternative specifications at producing daily forecasts of

[19]Both in the body and in the tails of the distribution.

the changes in the marked-to-market value of those option portfolios sensitive to non-parallel movements in the volatility curve (risk reversal and butterfly spread).

As for the other portfolios, the sticky-delta model seems to perform better at modelling the distribution of the daily profits and losses for both the short straddle position and the vega-neutral "Mexican hat". However, the goodness of these density forecasts at replicating the highly asymmetric and leptokurtic shape of the actual distributions, is almost offset by the presence of severe misspecifications in the first two moments. These biases often undermine the relative accuracy of the sticky-delta density forecasts in comparison with the competing ones generated by the Kalman filter model. Also, they always introduce misspecifications in the tails of the sticky-delta density forecasts, not present in the Kalman filter distributional forecasts.

Of our three models, the vega-gamma specification seems to be the most inappropriate one, given the strong biases that affect its distributional forecasts for all the option portfolios of interest.

In the light of our outcomes, we can conclude that, on the whole, the Kalman filter approach constitutes a promising method for modelling the evolution of the volatility curve, given its ability to account for changes in level, slope, curvature and skewness. This is consistent with the findings of some recent studies on the dynamics of the volatility curve via component analysis,[20] according to which more than one factor are needed to describe the movements of the curve, and these factors are not perfectly correlated with the underlying asset.[21]

---

[20] See, for example, Skiadopoulos, Hodges and Clewlow [1999], Alexander [2001], Cont and da Fonseca [2002].

[21] In this respect, it could be interesting to re-express our correlated skew coefficients in terms of these orthogonal factors.

4.9 Conclusions and Further Research

In consideration of the importance of predicting the evolution of the volatility skews/surfaces for pricing, trading and risk management purposes, we have applied here for the first time a discrete Kalman filter model to describe the dynamics of the volatility skew.

From a risk management point of view, we have assessed whether this model is capable of generating unbiased density forecasts of the daily changes in the marked-to-market value of a number of option portfolios exposed to variations of different nature in the volatility curve. The results have also been contrasted with those based on density forecasts obtained from two widely used benchmark models, which only allow for parallel shifts in the volatility curve.

Two main conclusions can be derived from our analysis. First, producing good forecasts of daily changes in the marked-to-market value of option portfolios is much harder than most risk managers believe. None of the methods investigated here yields unbiased forecasts for all the four option portfolios. Second, on the whole, the Kalman filter method seems to perform better than the alternative models, whose density forecasts for portfolios sensitive to non-parallel volatility changes turned out to be heavily biased.

In order to obtain better density forecasts, possible improvements to our work may include the relaxation of some of our assumptions, such as the intertemporal independence of the shocks in the system equation, and/or the cross-sectional independence of the measurement errors. We could also attempt a more sophisticated specification for the dynamics of the underlying process in the model, and a refinement of the estimation techniques.

Finally, it would be interesting to compare the Kalman filter updating of the volatility skew against richer alternatives than the ones explored in this work.

For risk management purposes, a meaningful comparison could be drawn with the filtered historical simulation approach devised by Barone-Adesi, Giannopoulos and Vosper [2002].

Table 4.1: Estimates of A_t and μ_t - Kalman filter model.

	1st estimation period		*2nd estimation period*	
	Diag. A_t	μ_t	Diag. A_t	μ_t
Level	0.0541	0.2155	0.0573	0.2192
Slope	0.0169	-0.1229	0.0195	-0.1316
Curvature	0.1625	0.1112	0.1578	0.1111
Skewness	0.1782	0.1033	0.1761	0.1005
	3rd estimation period		*4th estimation period*	
	Diag. A_t	μ_t	Diag. A_t	μ_t
Level	0.0593	0.2218	0.0508	0.2163
Slope	0.0189	-0.1222	0.0189	-0.1200
Curvature	0.1320	0.1142	0.1208	0.1176
Skewness	0.1699	0.0999	0.1672	0.0992

Table 4.2: Estimates of correlation matrix and standard deviations of system equation errors - Kalman filter model.

1st estimation period

	Level	Slope	Curvature	Skewness
Level	1	-0.2673	-0.1969	-0.0016
Slope		1	0.4875	0.1144
Curvature			1	0.8346
Skewness				1
Std. dev.	0.0132	0.0094	0.0303	0.0319

2nd estimation period

	Level	Slope	Curvature	Skewness
Level	1	-0.2698	-0.1994	0.0001
Slope		1	0.4756	0.0761
Curvature			1	0.8290
Skewness				1
Std. dev.	0.0133	0.0094	0.0281	0.0300

3rd estimation period

	Level	Slope	Curvature	Skewness
Level	1	-0.2966	-0.2670	-0.0384
Slope		1	0.4382	-0.0212
Curvature			1	0.7849
Skewness				1
Std. dev.	0.0133	0.0092	0.0252	0.0271

4th estimation period

	Level	Slope	Curvature	Skewness
Level	1	-0.3366	-0.3570	-0.1131
Slope		1	0.4491	-0.0399
Curvature			1	0.7579
Skewness				1
Std. dev.	0.0110	0.0087	0.0239	0.0251

Table 4.3: Estimates of measurement error volatility per buckets of log moneyness - Kalman filter model.

Buckets log moneyness	Std. dev. measurement error
≤ -0.50	0.0097
$> -0.50, \leq -0.40$	0.0086
$> -0.40, \leq -0.30$	0.0059
$> -0.30, \leq -0.20$	0.0043
$> -0.20, \leq -0.10$	0.0030
$> -0.10, \leq -0.01$	0.0028
$> -0.01, \leq 0.01$	0.0023
$> 0.01, \leq 0.10$	0.0037
$> 0.10, \leq 0.20$	0.0072
> 0.20	0.0123

Table 4.4: Estimates of EGARCH(1,1) model with normal errors.

	1st est. period		*2nd est. period*		*3rd est. period*		*4th est. period*	
ω	-0.4201	(0.0971)	-0.4727	(0.1057)	-0.4766	(0.1054)	-0.3693	(0.1035)
α	0.1023	(0.0307)	0.1111	(0.0323)	0.0936	(0.0320)	0.0884	(0.0317)
γ	-0.1779	(0.0177)	-0.1771	(0.0197)	-0.1714	(0.0196)	-0.1240	(0.0188)
β	0.9520	(0.0102)	0.9459	(0.0113)	0.9451	(0.0112)	0.9579	(0.0105)

Standard errors in brackets.

Table 4.5: Estimates of GARCH(1,1) model with normal errors.

	1st est. period		*2nd est. period*		*3rd est. period*		*4th est. period*	
ω	0.0002	(3.50E-05)	0.0002	(3.58E-05)	0.0001	(3.31E-05)	0.0002	(5.97E-05)
α	0.1438	(0.0278)	0.1643	(0.0302)	0.1628	(0.0298)	0.1511	(0.0340)
β	0.8099	(0.0284)	0.7936	(0.0281)	0.8001	(0.0277)	0.7766	(0.0283)

Standard errors in brackets.

Table 4.6: Density forecasts tests - $H_0 : x_t \sim$ i.i.d. $N(0,1)$.

		Straddle			*Risk-reversal*		
		K.F.	Sticky-Δ	$V-\Gamma$	K.F.	Sticky-Δ	V-Γ
$LR1$ (independence)	(3.84)	0.10	0.09	0.07	3.73	1.93	1.50
$LR2$ (i.i.d.$(0,1)$)	(7.81)	*20.01	*270.61	*73.72	*52.45	*188.63	*727.39
$LR3$ (zero mean)	(3.84)	0.95	*20.72	*48.44	*31.52	*79.48	2.08
$LR4$ (unit variance)	(3.84)	*19.29	*211.73	*50.34	*22.55	*59.14	*705.92
Jarque-Bera	(5.99)	*38.36	*54.61	*45.27	*7.60	*39.99	*19.71

		Butterfly spread			*Mexican hat*		
		K.F.	Sticky-Δ	V-Γ	K.F.	Sticky-Δ	V-Γ
$LR1$ (independence)	(3.84)	1.07	0.34	3.14	2.48	0.24	0.07
$LR2$ (i.i.d.$(0,1)$)	(7.81)	3.49	*112.24	*44.19	*18.82	*129.90	*39.02
$LR3$ (zero mean)	(3.84)	1.12	*18.41	*55.30	*10.98	1.83	0.65
$LR4$ (unit variance)	(3.84)	1.32	*76.99	*6.13	0.99	*125.24	*37.85
Jarque-Bera	(5.99)	*32.70	*28.12	*77.65	*19.58	*6.33	*60.73

Critical values in brackets.
* rejected at 5% confidence level.

Table 4.7: Summary statistics for actual profits and losses and density forecasts.

	Mean	**Std.dev.**	**Variance**	**Skew**	**Exc. Kurt.**
Straddle					
Sample	0.240	2.878	8.283	-2.313	8.703
Kalman filter	-0.009	4.167	17.365	-1.251	3.351
Sticky-Δ	-0.251	3.190	10.174	-1.888	6.861
Vega-gamma	-1.301	5.038	25.383	-0.286	2.070
Risk-reversal					
Sample	0.071	0.518	0.268	-0.793	0.741
Kalman filter	0.318	0.659	0.434	-0.920	0.835
Sticky-Δ	0.411	0.692	0.479	-0.786	3.142
Vega-gamma	0.058	0.223	0.050	-0.036	4.822
Butterfly spread					
Sample	0.139	1.229	1.509	-0.602	0.918
Kalman filter	-0.015	1.518	2.304	-0.636	1.875
Sticky-Δ	-0.110	1.404	1.970	-1.810	10.177
Vega-gamma	-0.494	1.710	2.922	-0.700	6.128
Mexican hat					
Sample	0.039	2.007	4.029	2.866	14.389
Kalman filter	-0.140	2.021	4.086	0.734	3.865
Sticky-Δ	-0.069	1.716	2.945	2.045	14.300
Vega-gamma	0.113	2.551	6.507	0.773	5.176

Table 4.8: Percentage of VaR breaks.

	Lower percentiles		**Upper percentiles**	
	1%	5%	99%	95%
Straddle				
Kalman filter	0.79%	2.38%	0.40%	1.98%
Sticky- Δ	0.79%	3.57%	17.46%	21.83%
Vega-gamma	0.40%	0.40%	0.79%	3.17%
Risk-reversal				
Kalman filter	1.59%	5.16%	0.00%	0.40%
Sticky- Δ	14.29%	19.05%	0.00%	0.40%
Vega-gamma	27.38%	28.97%	13.10%	23.41%
Butterfly spread				
Kalman filter	1.98%	3.17%	1.19%	3.17%
Sticky- Δ	1.98%	2.38%	10.32%	14.29%
Vega-gamma	0.79%	1.59%	2.38%	5.95%
Mexican hat				
Kalman filter	0.00%	1.59%	3.17%	9.52%
Sticky- Δ	6.35%	9.92%	5.56%	13.49%
Vega-gamma	1.98%	3.57%	7.14%	9.13%

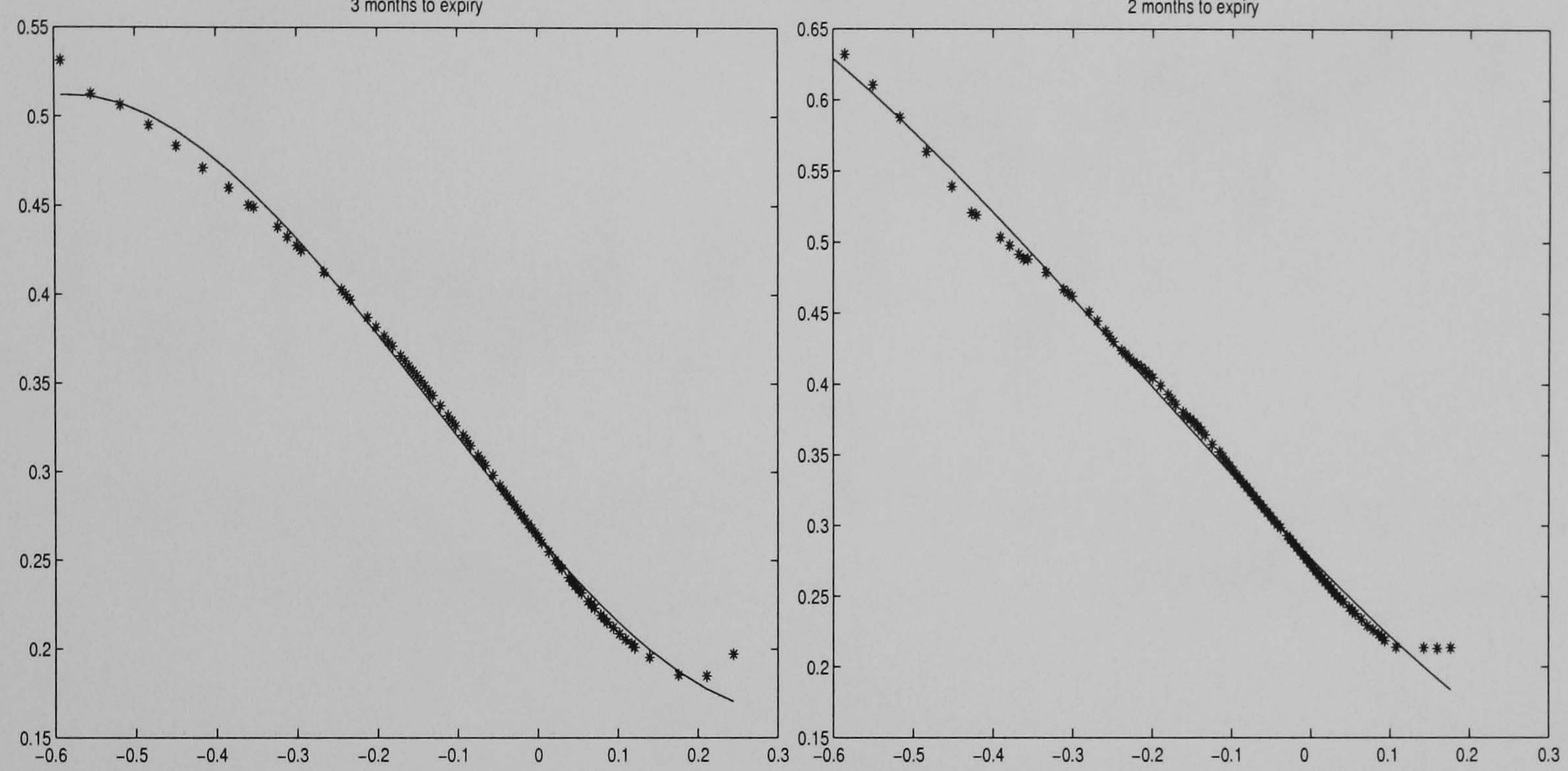


Figure 4.1: Cubic fit to the implied volatility curve for three expiries of the March 99 contract.

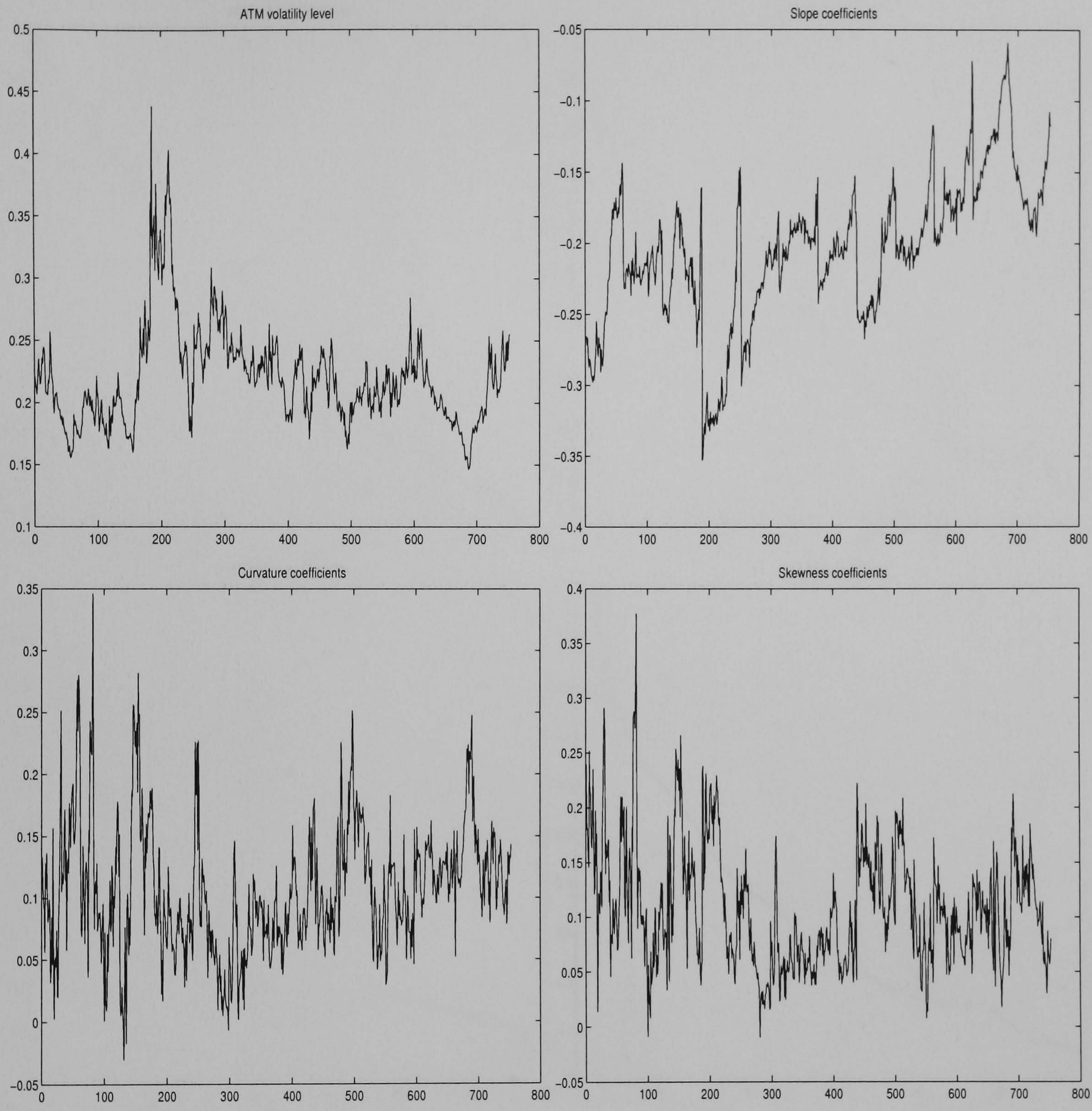


Figure 4.2: Time series of the volatility skew coefficients (Jan.98 - Dec.00).

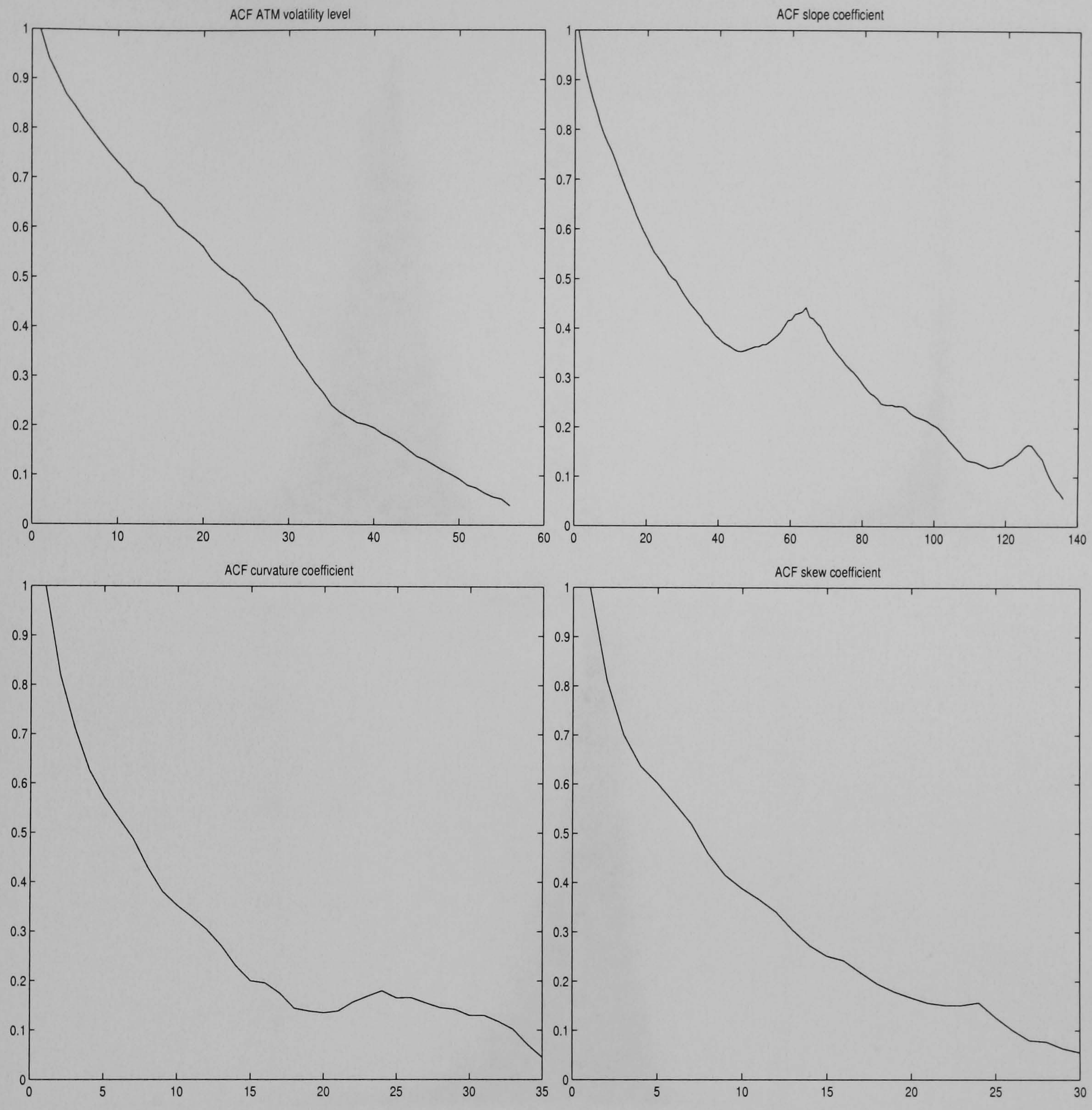


Figure 4.3: Autocorrelation functions of the volatility skew coefficients (Jan.98 - Dec.00).

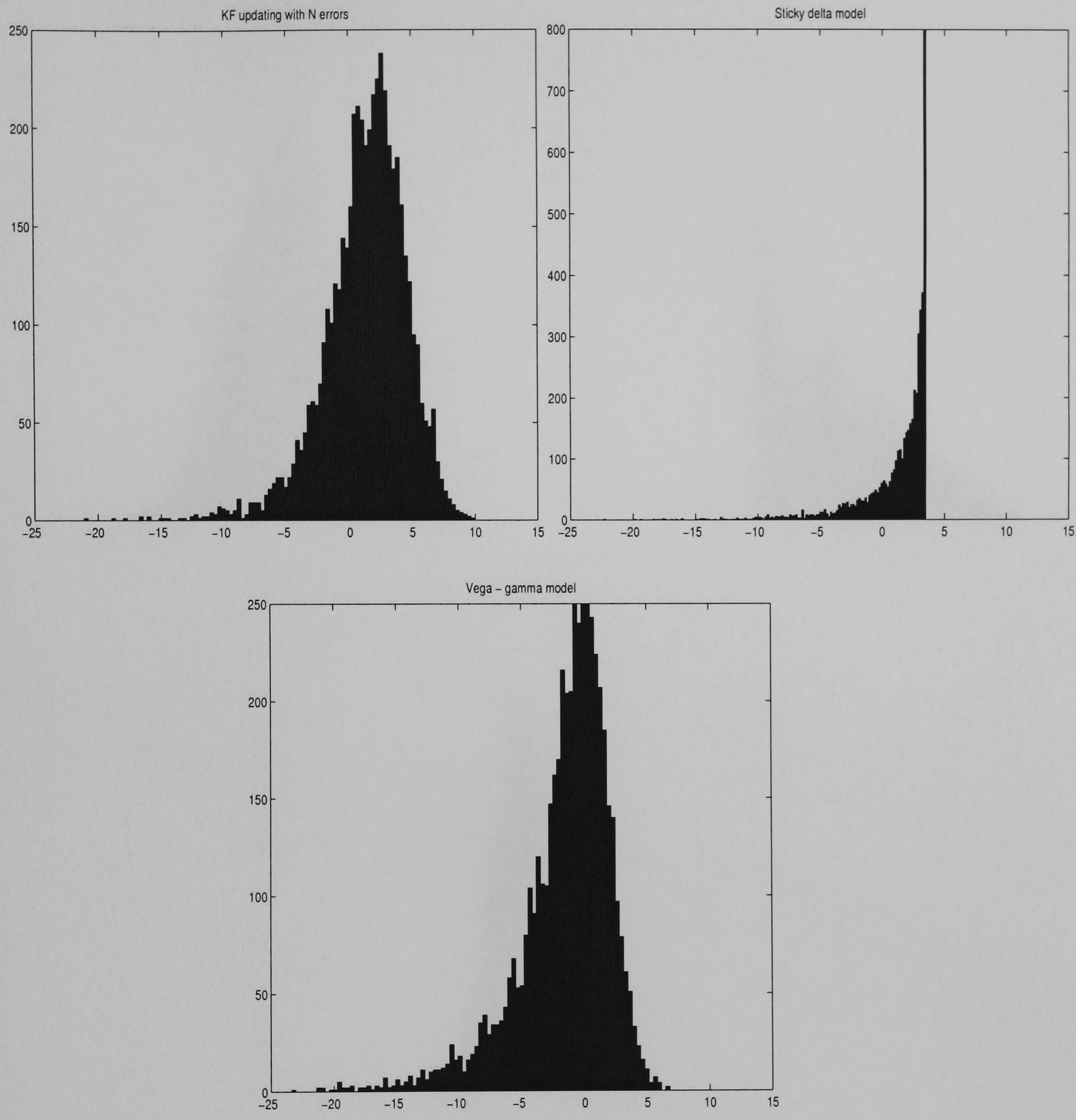


Figure 4.4: Histograms of density forecasts on 20/8/2001 - short straddle.

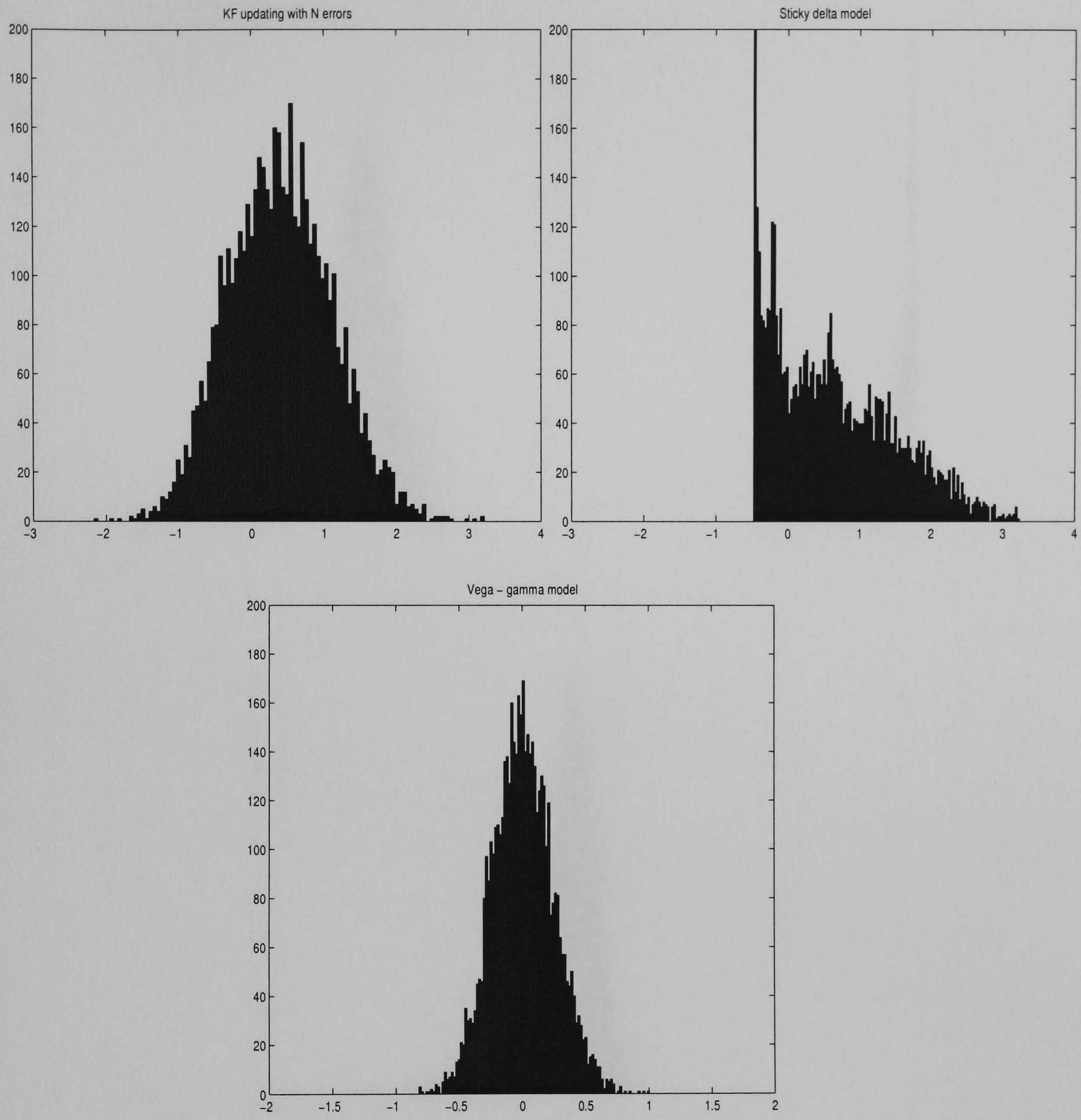


Figure 4.5: Histograms of density forecasts on 20/8/2001 - long risk reversal.

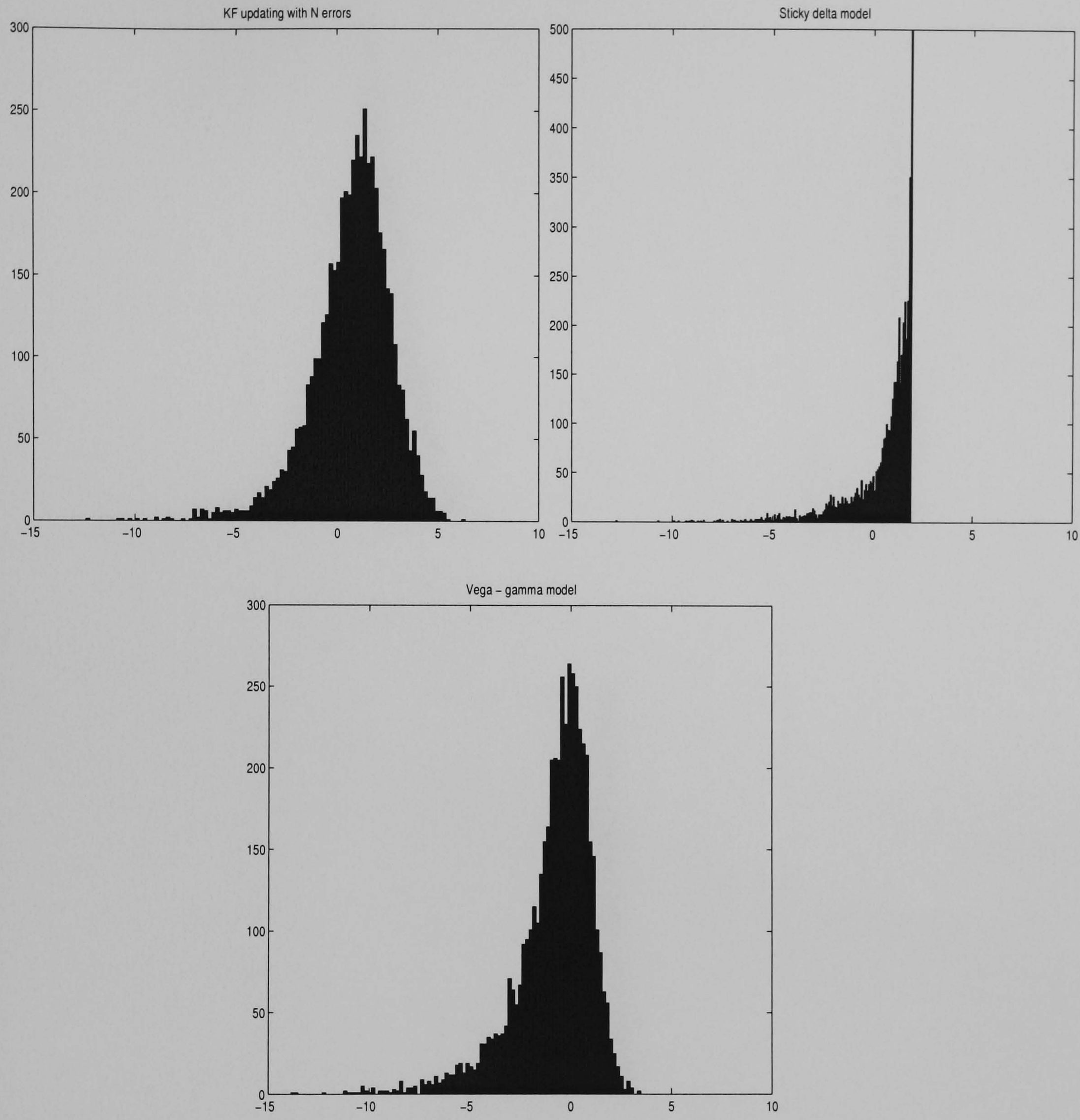
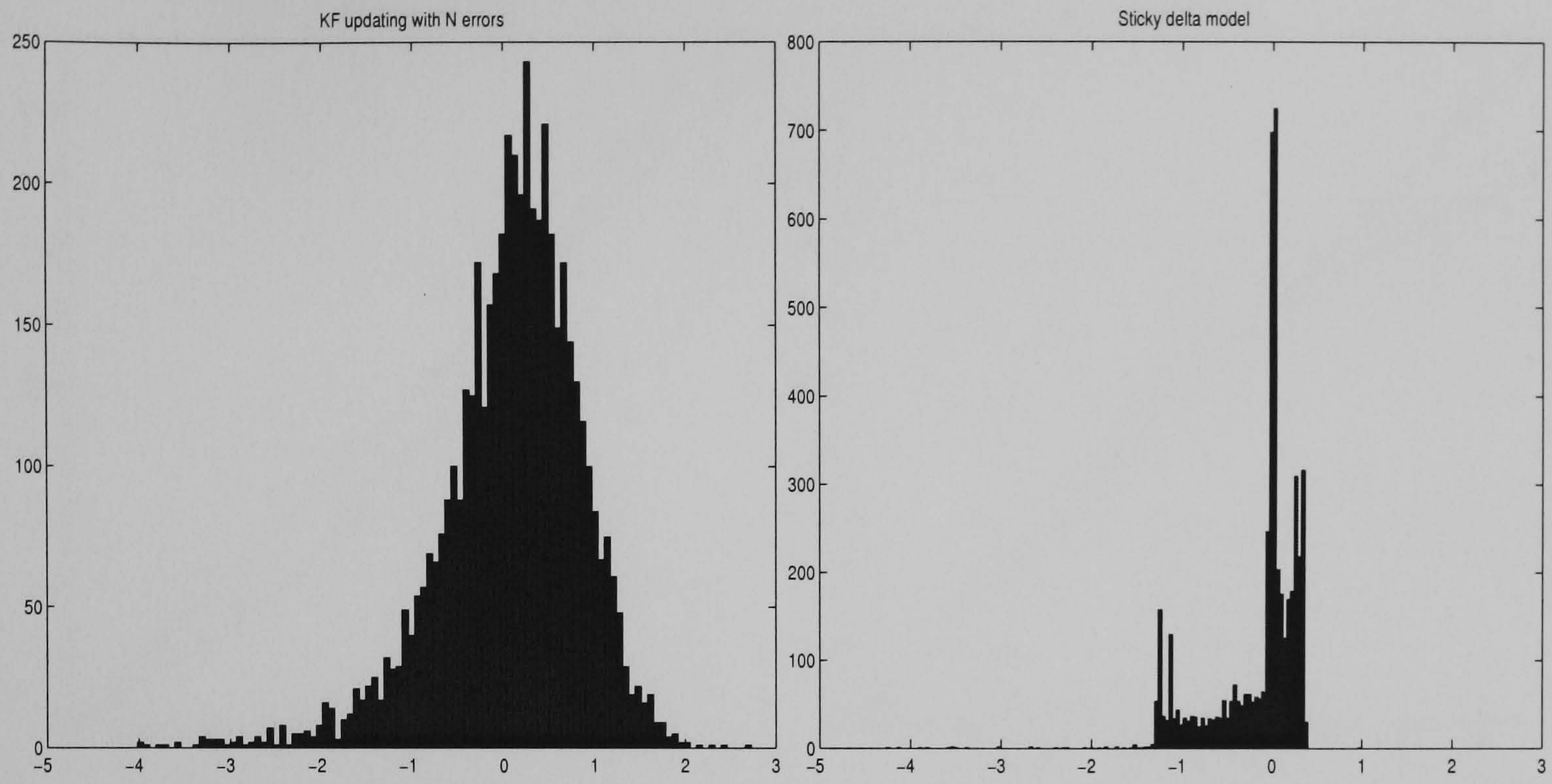


Figure 4.6: Histograms of density forecasts on 20/8/2001 - long butterfly spread.



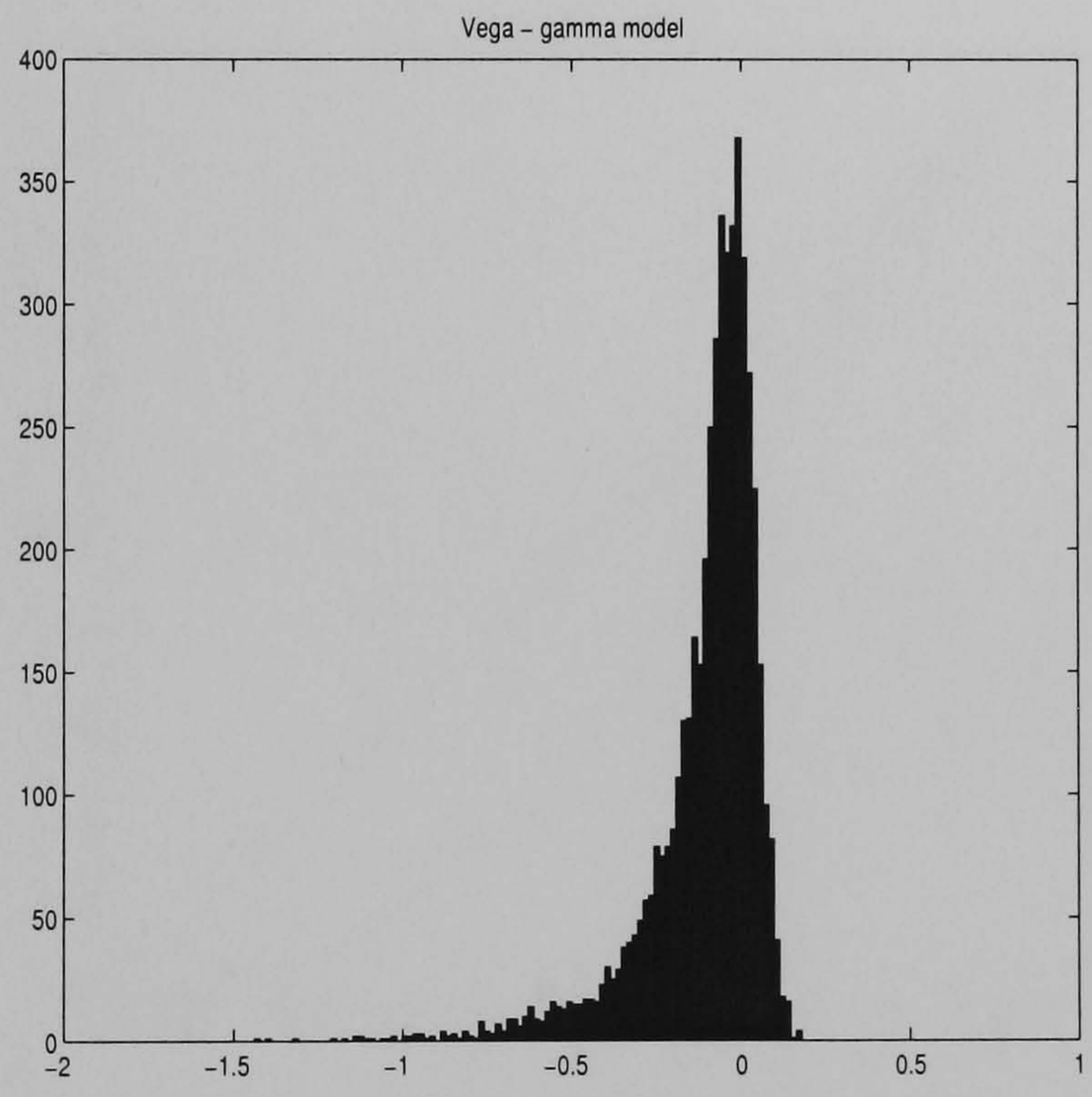
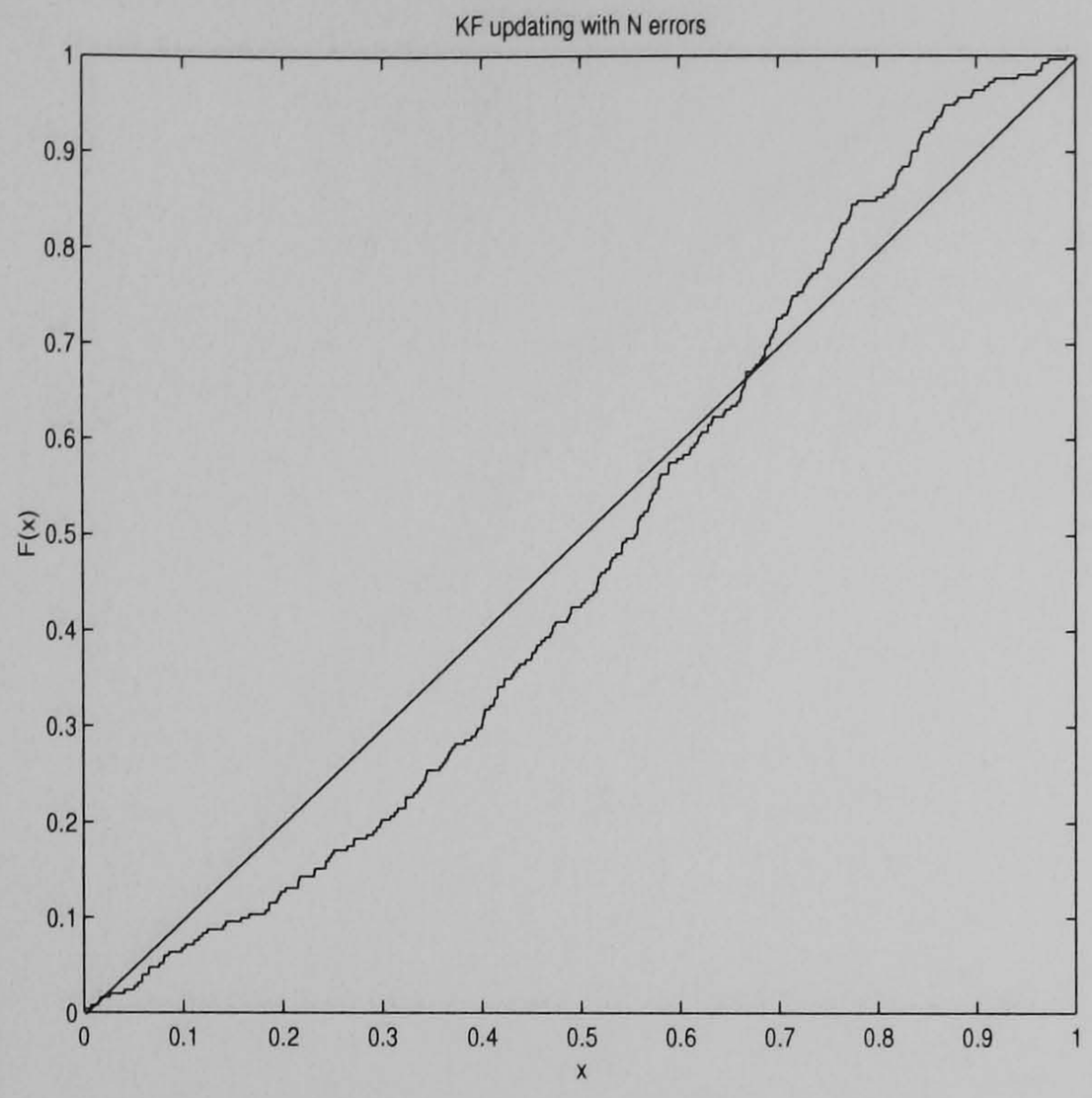
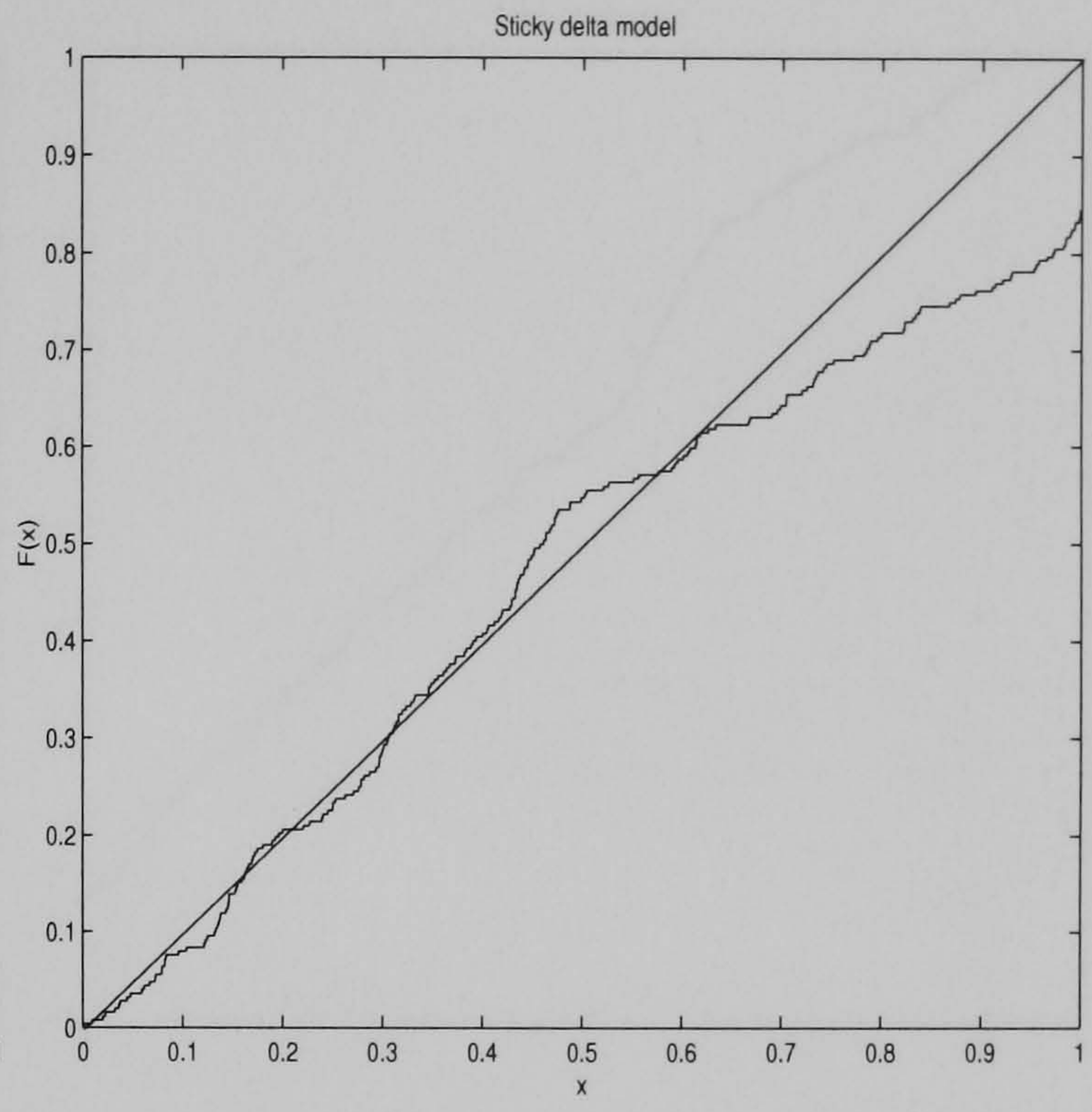


Figure 4.7: Histograms of density forecasts on 20/8/2001 - long Mexican hat.





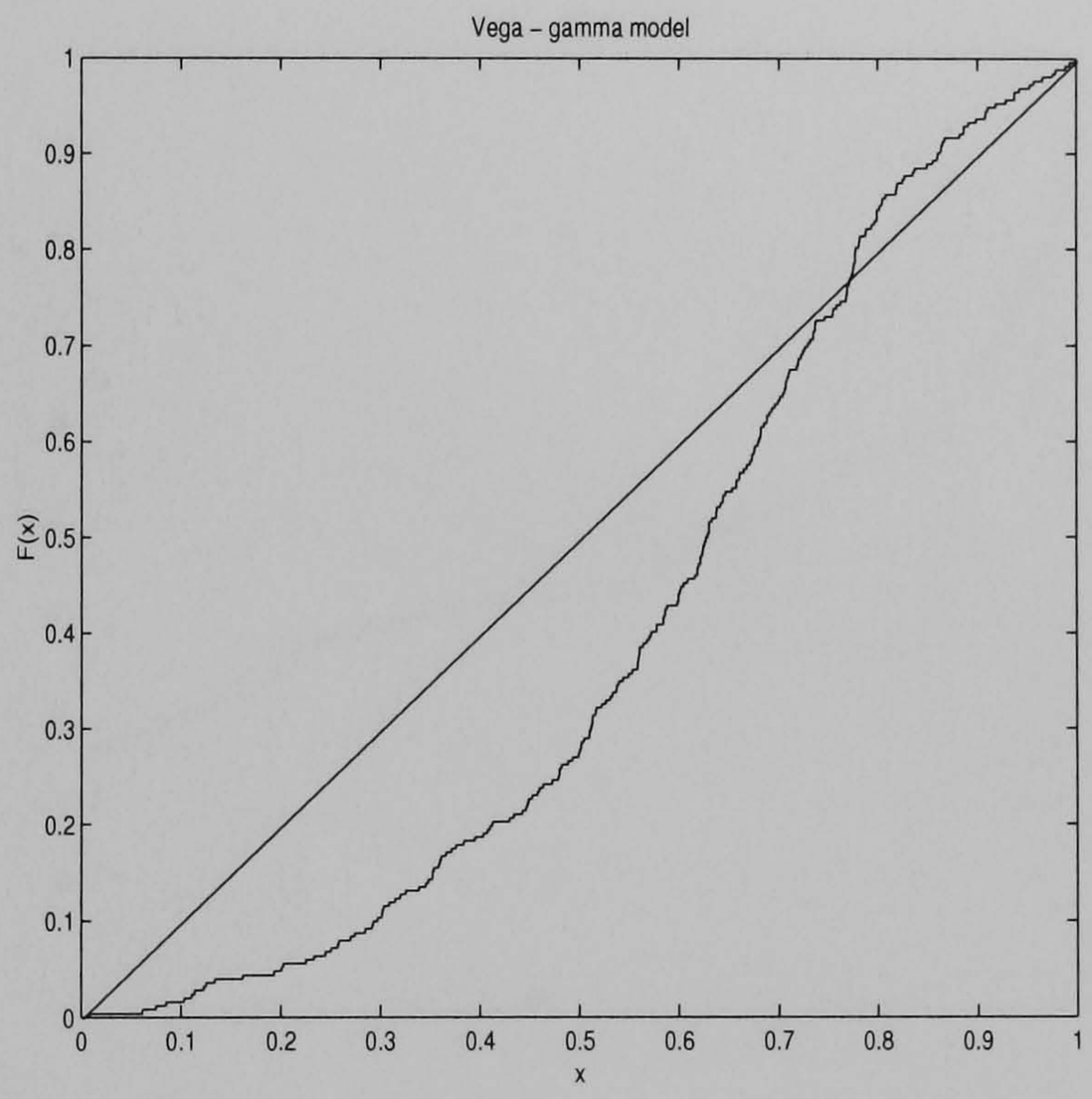


Figure 4.8: CDF plots of goodness-of-fit - short straddle.

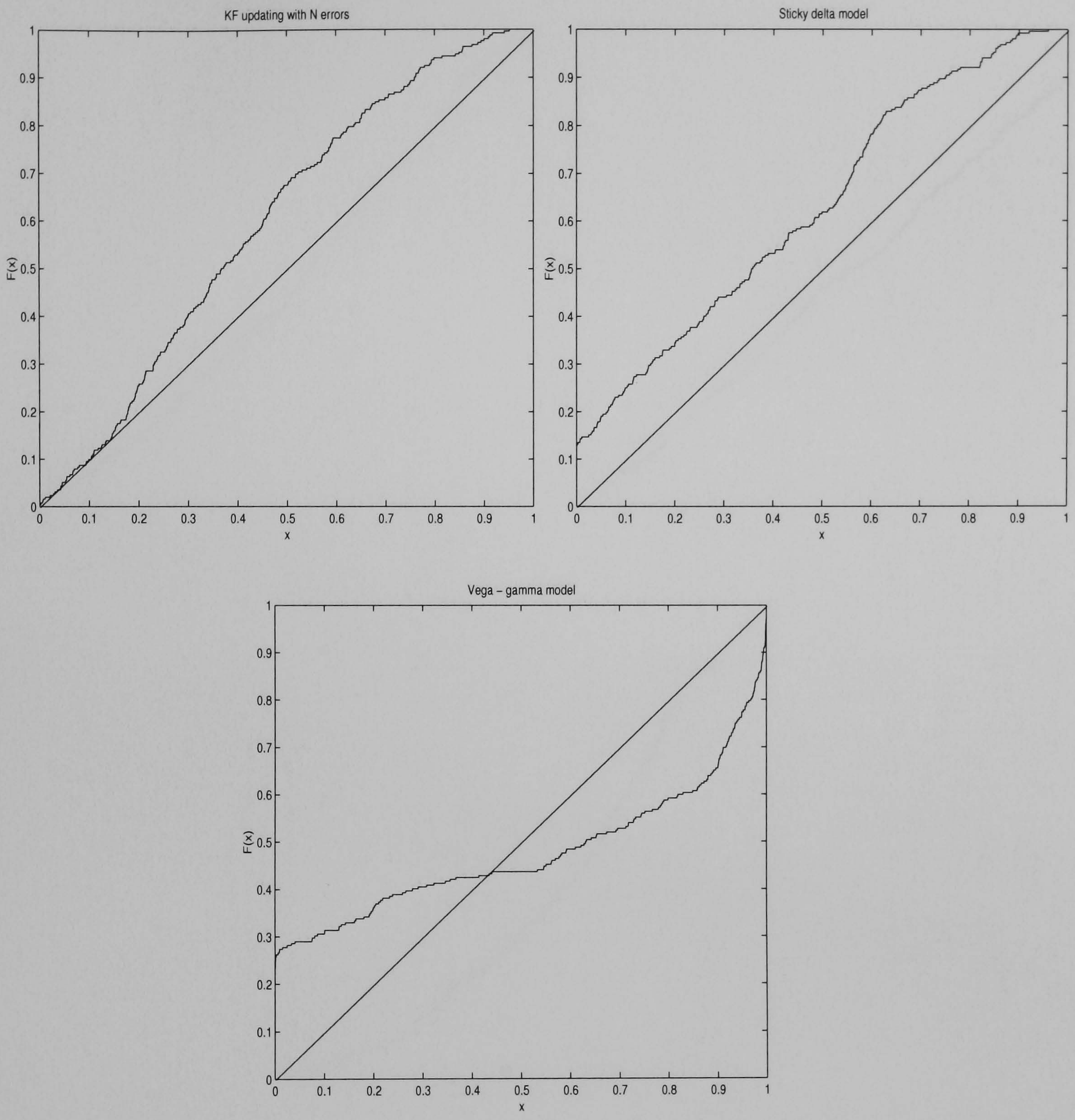
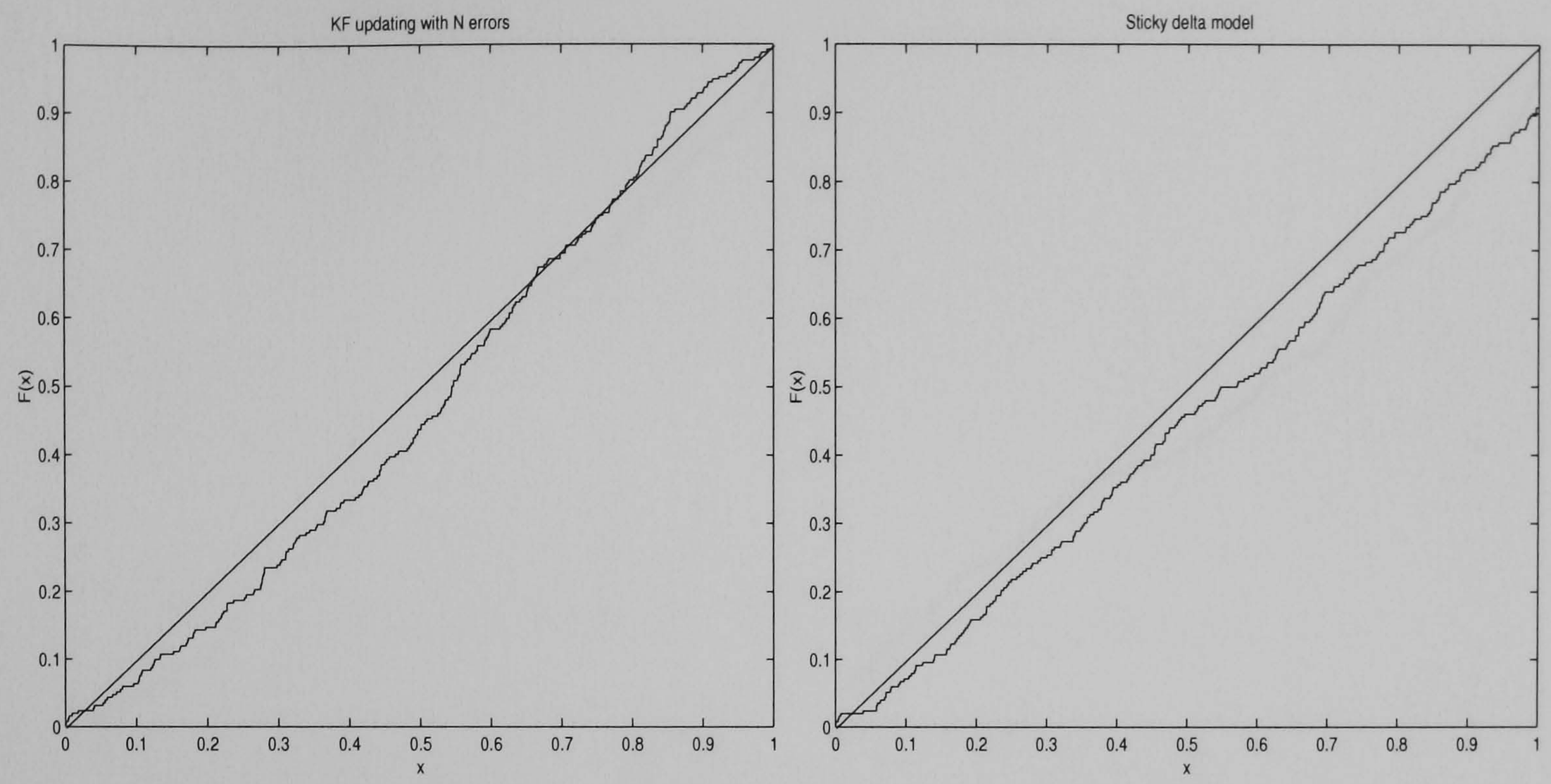


Figure 4.9: CDF plots of goodness-of-fit - long risk reversal.



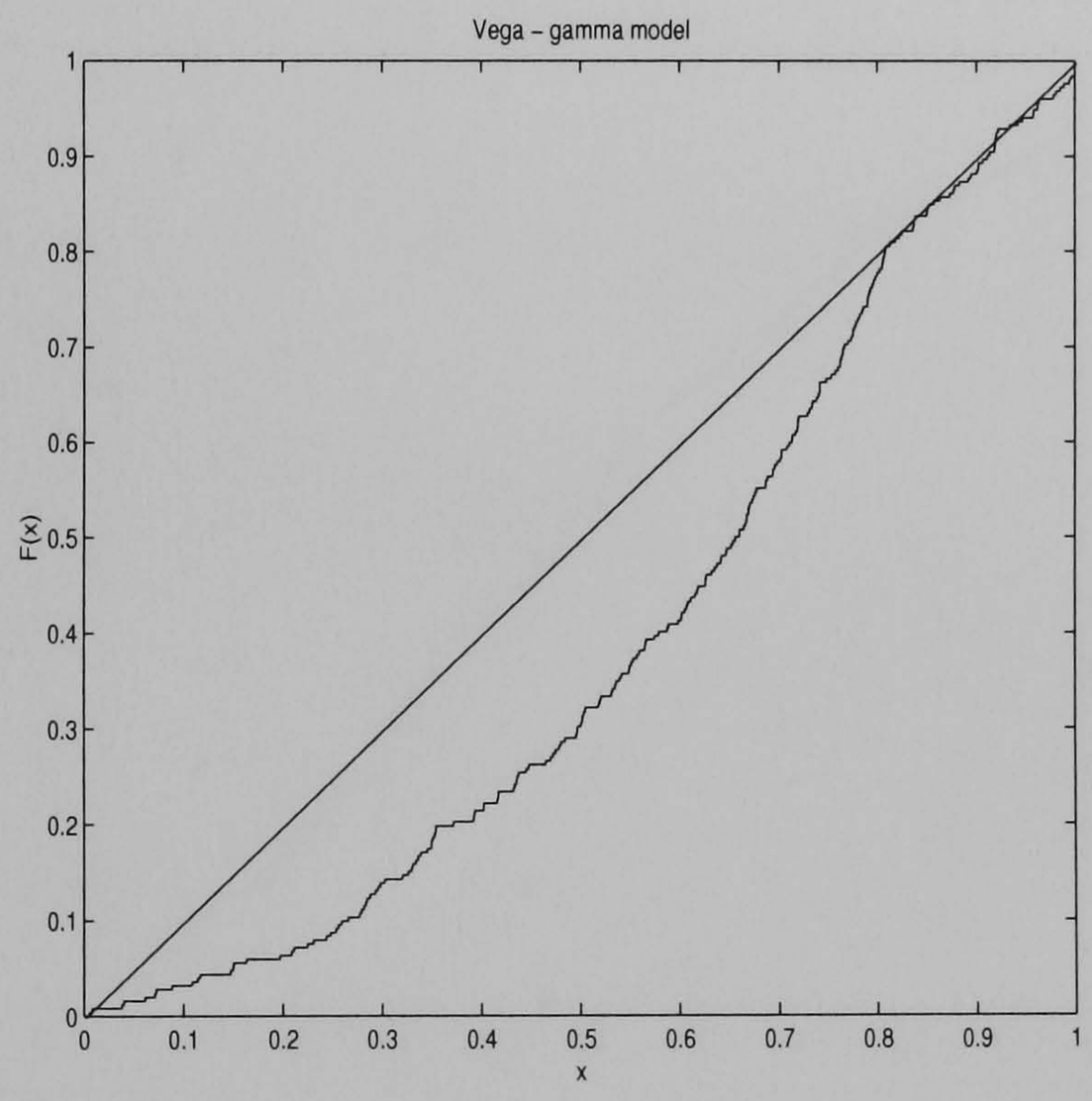


Figure 4.10: CDF plots of goodness-of-fit - long butterfly spread.

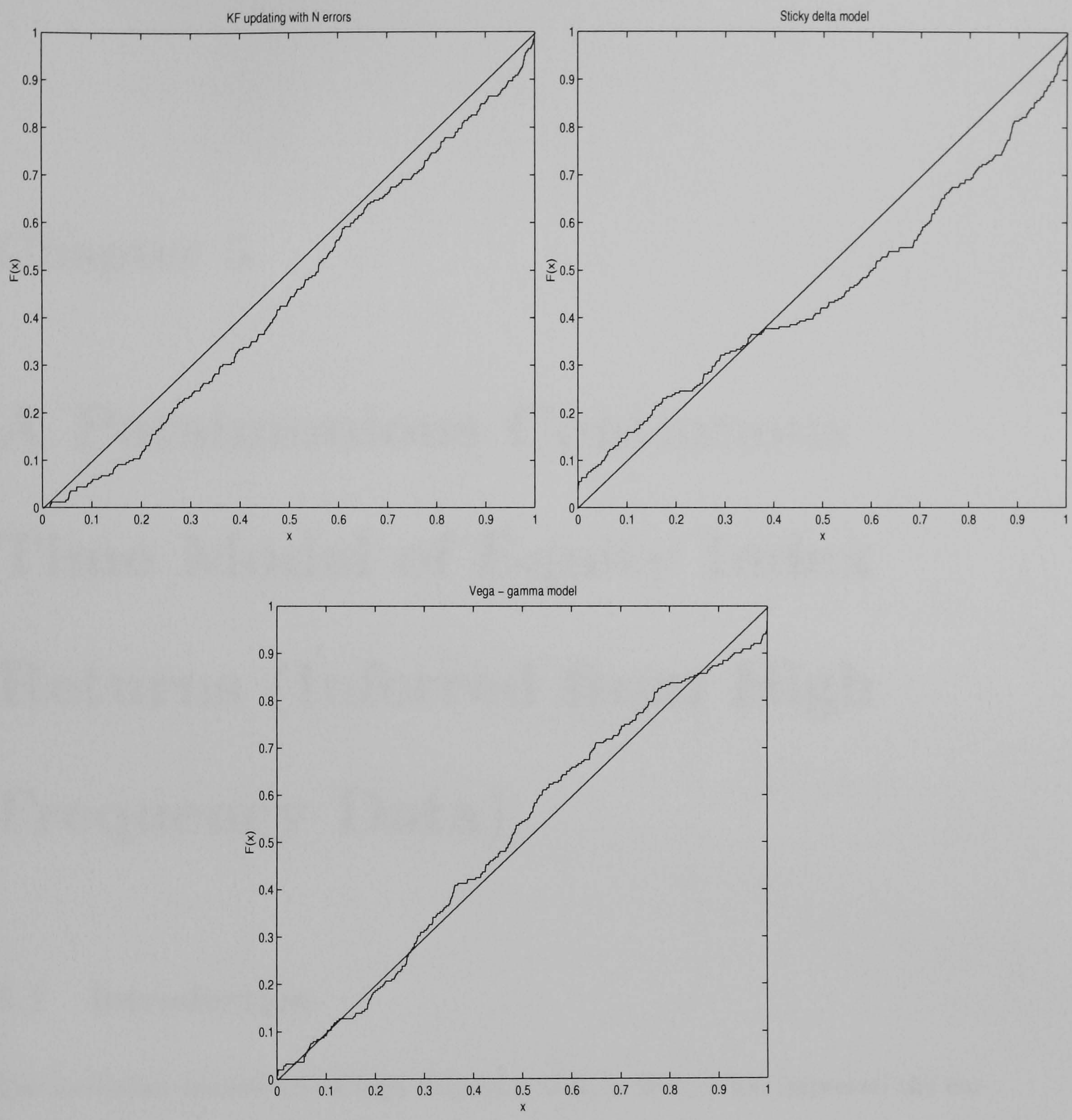


Figure 4.11: CDF plots of goodness-of-fit - long Mexican hat.

Chapter 5

A Parsimonious Continuous Time Model of Equity Index Returns (Inferred from High Frequency Data)

5.1 Introduction

The increasing availability of high frequency data in finance has improved the empirical analysis of financial asset returns in several respects. In the first place, it has enabled the investigation of the dynamics of intraday volatility and returns *per se*, in consideration of the various market microstructure effects that characterise high frequency financial data. Secondly, and perhaps more importantly, it has enriched the information set available to develop and test continuous time models, which are able to explain and replicate the dynamics of financial returns observed in the markets in a consistent manner across different time horizons. Traditionally, continuous time

models in finance have been estimated and tested on moderate frequency (normally daily) financial data. In most cases, however, the asset returns generated from those models manage to capture the dynamics of daily or weekly returns fairly accurately, but fail to mirror the behaviour of high frequency financial returns. Therefore, intraday data can be usefully employed to derive a more consistent specification for a continuous time model. From a risk management perspective, where the temporal horizon of interest for the computation of the risk measures is short, a model estimated on high frequency data is expected to produce better short-term density forecasts of the changes in value of the relevant risk factors. Also, the need for a model that can accurately describe the dynamics of high frequency returns is especially evident in view of the computation of intraday Value-at-Risk (or alternative risk measures) and the implementation of intraday hedging strategies.

The aim of the present work is to identify the simplest possible model which is both congruent with the specifications commonly adopted in this context, and capable of replicating the essential features that characterise the evolution of intraday returns and volatility in the market. A continuous time specification will turn out to be the most convenient and appropriate one for such purpose.

A distinctive aspect of our study, which we consider a significant contribution to the related literature, is that we adopt a parsimonious approach. Throughout the different steps, we let the data suggest the model as much as possible, rather than imposing a model ourselves. The standard approach commonly followed by the literature consists of assuming from the beginning a particular specification for the model in all its components, and using the data to estimate and test it. Instead, we believe that a model for financial data should originate from the data itself, therefore here we avoid specifying a model *a priori*. We start with a very general model structure and we perform a careful step by step analysis of the data, recording the relevant features to be modelled, whose peculiar characteristics will

actually drive the choice among different specifications. At each step we also look carefully for possible specification errors. Throughout the entire paper we try to keep the modelling assumptions to a minimum, while retaining an adequate level of structure. Our approach is also parsimonious in terms of the statistical and econometric techniques employed to estimate the resulting model. Our main interest here is in assessing whether the data-driven, step-by-step criteria we propose for selecting the model and subsequently refining it on the basis of intraday returns, enables us to derive a valid specification that adequately explains the empirical features. Producing the most precise estimates for the parameters of our model is not our main concern, especially since this would require the implementation of very complicated econometric tools that would introduce a lot of complexity to the analysis without contributing significantly to the main results. Therefore we use simple techniques that still produce reasonably accurate estimates.

At the conclusion of our analysis we propose a relatively simple specification, able to capture and model most of the aspects observed in equity index futures markets, namely: seasonality in intraday volatility, stochastic volatility, presence of jumps and leverage effect. By means of a simple Bayesian filtering technique we also generate 5-minutes ahead volatility estimates and density estimates for the distribution of intraday returns, whose accuracy is thoroughly assessed via both point and distributional forecast tests.

The chapter is structured as follows. Section 5.2 introduces the related literature. Section 5.3 describes the data set. Section 5.4 details in its subsections the various steps of the data analysis and the modelling of each component, up to the identification and estimation of a simple and accurate model in continuous time. The assessment of volatility and density intraday estimates produced by our model is carried out in Section 5.5. A Monte Carlo simulation exercise of the complete model is performed in Section 5.6. Section 5.7 summarises the main conclusions

and a few suggestions for further research.

5.2 The Informative Content of High Frequency Data

During the last few years, the availability of high frequency data on financial assets has stimulated the production of a very rich literature.

One stream of literature (not immediately related to the present work) has focused on deriving tailored models for intraday returns and volatility, capable of capturing their distinctive features. For data recorded at 5-minute or less frequent intervals and, therefore, scarcely affected by market microstructure effects, traditional volatility models (ARCH, GARCH) are still widely employed to model the dynamics of the process, after correcting for the seasonal component. The first attempt in this direction was made by Andersen and Bollerslev [1997]. A similar two-step approach was followed by the same authors in a later paper (Andersen and Bollerslev [1998a]) and by Martens, Chang and Taylor [2002], who also compared different seasonality adjustment methods. Beltratti and Morana [1999] introduced a stochastic seasonal component to estimate volatility models for risk measurement purposes. A one-step periodic GARCH model was suggested by Bollerslev and Ghysels [1996]. Andersen, Bollerslev and Das [2001] investigated how changes in intraday volatility patterns affect high frequency volatility models.

When the frequency at which observations are recorded is very high, market microstructure effects such as bid/ask spreads and non-synchronous trading induce a substantial degree of serial correlation in returns, which makes the conventional models inappropriate. Price discreteness also represents an issue when trade-by-trade data is used.[1] In this context, models for the dynamics of transaction prices were suggested, amongst the others, by Rydberg and Shephard [2000, 2003] and

[1]For a detailed discussion of market microstructure effects, see Campbell, Lo and MacKinlay [1997], Madhavan [2000].

Rogers and Zane [1998]. Giot [2000] proposed a high-frequency duration approach to characterise the volatility process of non-regularly spaced transaction data. Oomen [2002] chose a continuous time model based on a generalisation of a compound Poisson process that accounts for the peculiar features of high frequency data.

A second stream of literature has exploited the informative content of intraday data to obtain more accurate measures of the volatility of financial returns. Most of these studies approximate the volatility over a certain period, such as a day, with the sum of intraday squared, or absolute, returns, a measure called realised volatility (see Taylor and Xu [1997] and Andersen and Bollerslev [1998b] among the authors who first proposed this measure). The theoretical justification for this approximation[2] is to be found in the theory of quadratic variation (see Karatzas and Shreve [1988]): as the length of the intraday interval tends to zero, the sum of uncorrelated intraday squared returns tends in probability to the quadratic variation of the underlying diffusion process which, in absence of jumps, corresponds to the integrated volatility. A complete asymptotic theory of the convergence of the realised volatility to the integrated volatility was derived by Barndorff-Nielsen and Shephard [2001b, 2002, 2003], under the assumptions that conditional returns are normally distributed and volatility follows either a diffusion specification or a Lévy process. They also considered extensions to account for the presence of a leverage effect. Bai, Russell and Tiao [2001] and Andreou and Ghysels [2002] discussed the distortions induced in the realised volatility measure by microstructure effects when returns are measured at a very high frequency. Meddahi [2002] provided both qualitative and quantitative measures of the precision of the approximation to the integrated volatility for different specifications, including the impact of a leverage effect.

[2]See Andersen, Bollerslev, Diebold and Labys [2001, 2003], Barndorff-Nielsen and Shephard [2001a].

An impressive number of papers have appeared in the last couple of years in this area, proposing various possible applications for the informative content of intraday data via the realised volatility measure. Here we review the most relevant ones; for a survey of this literature, the interested reader can consult Andersen, Bollerslev and Diebold [2003], Barndorff-Nielsen, Nicolato and Shephard [2002] and Dacorogna *et al.* [2001]. Andersen and Bollerslev [1998b] employed realised volatility as a measure against which to compare daily volatility forecasts of exchange rates produced with a GARCH model, which turned out to be far more accurate than expected. Blair, Poon and Taylor [2001] found similar results for stock index data. Some authors investigated whether out-of-sample volatility forecasts could be improved by using intraday data. Andersen, Bollerslev and Lange [1999] found that the modelling of intraday returns contributed to produce better daily and longer horizons forecasts. Blair, Poon and Taylor [2001] suggested instead to account for the incremental information given by high-frequency data by incorporating the realised volatility as an explanatory variable in the variance equation of a daily GARCH model. Martens [2001] contrasted the two approaches for exchange rate data using a GARCH specification and concluded that they produce very similar results. A similar comparison, but on extensions of stochastic volatility models, was made by Hol and Koopman [2002].

The distribution and the time series properties of realised volatility have also been studied. Examples in this context are given by Andersen, Bollerslev, Diebold and Labys [2001, 2003] for exchange rates, Ebens [1999] for the Dow Jones Industrial Average index, Andersen, Bollerslev, Diebold and Ebens [2001] for both the index and its constituent stocks, Areal and Taylor [2002] for FTSE-100 index futures. All of these studies agree on two points: 1) the time series of realised volatility exhibits a long memory feature well described by an ARFIMA type of model; 2) the distribution of the returns conditioned on the realised volatility strictly

resembles a Gaussian. Barndorff-Nielsen and Shephard [2002] fitted a continuous time superposition of Ornstein-Uhlenbeck processes to the time series of realised volatility.

A third stream of literature employs intraday data in order to estimate and test continuous time models in which financial returns are described by a time-changed Brownian motion or Lévy process, where the stochastic time change is given by a measure of the intraday economic activity (e.g. trading volumes, proxy of integrated stochastic volatility). The theoretical justification for such an approach is that all arbitrage-free processes defining asset returns can be represented as time-changed Brownian motions,[3] where the time change (or business time) must account for information arrival and market activity. This stream of research originates from the pioneering paper by Clark [1973], who showed how, once re-specified in the new business time (expressed in terms of the cumulative volume of activity), financial returns are virtually distributed according to a Gaussian law. Amongst the most relevant contributions in this field, we recall Andersen [1996], who investigated the returns specification when the intraday information flow is modelled as a stochastic volatility process, and Ané and Geman [2000], who extended Clark's results by using the cumulative number of trades as a stochastic clock. Very recent studies include Geman, Madan and Yor [2001], Carr, Geman, Madan and Yor [2003], Carr and Wu [2003], where various kinds of Lévy processes are evaluated at a time change given by the integrated volatility, generally modelled as a square root process.

Our work directly relates to the second and the third streams of research. At an intermediate stage of our analysis, we derive a daily realised volatility quantity from intraday data in order to obtain an almost noise-free measure which can provide reliable insights on the stochastic volatility dynamics and on the shape of

[3]See Monroe [1978] for the proof that any semimartingale is a time-changed Brownian motion.

the conditional distributions. Similarly to the literature on stochastic time changes, our purpose is to estimate a valid continuous time specification from high frequency data. However, contrary to the contributions listed above, we will find that, after correcting for the seasonality in volatility, no stochastic time change is necessary and the model can be set in calendar time.

5.3 The Data Set

Our data set consists of 5-minute frequency intraday prices on the S&P 500 stock index futures contract from September 15, 1997, to July 26, 2001. All prices are for the futures contract closest to maturity, except for the days within one week to expiration, when the next contract is considered, in order to always refer to the contract with the highest trading volume. Days that recorded transactions only for part of the entire trading day[4] are excluded from the data set. We also eliminate days which exhibit very large returns on some intraday interval immediately followed by equally large returns of the opposite sign, which could be indicative of mistakes in recording the price. Some other days were originally missing from the data set. All in all, our final sample consists of intraday prices for 960 days.

The full trading day in the futures market at the Chicago Mercantile Exchange starts at 8:30 a.m. and ends at 3:15 p.m. Chicago time. Intraday log returns are computed on the consecutive logarithmic closing prices for each of the 81 5-minute intervals that constitute a trading day. Since in modelling the intraday dynamics of returns and volatility it is important to take into account the close-to-open returns and their volatility, we also analyse overnight log returns, calculated as the difference between the logarithm of the open price and the logarithm of the closing price for the previous day. For the same reason, unlike other works on high

[4]Most of those days are those immediately preceding holidays, such as December 24 and December 31.

frequency data (see, for example, Andersen and Bollerslev [1997]), we retain the return on the first interval of the trading day, which mainly reflects the information accumulated overnight and shows a high volatility.

In the top half of Table 5.1 we report some sample statistics for the 5-minute and the overnight returns, which we consider separately, given the different nature and characteristics of the two series. The intraday returns display an almost zero sample mean, a sample standard deviation of 0.121%, positive sample skewness of 0.88 and a strong sample excess kurtosis of 35.48. As expected, the standard deviation of the overnight returns is considerably larger, as it refers to a longer temporal horizon. The higher moments are closer to those of a normal distribution, with a sample skewness of -0.38 and an excess kurtosis of 3.04, by effect of the aggregation process which takes place over a longer time horizon.

Table 5.2 displays the values of first order autocorrelation coefficients in the series of high frequency returns for each year under analysis, together with the percentage bid-ask spread, estimated following Roll [1984].[5] Although statistically significant for the first two and a half years, the serial correlation in the intraday futures returns always seems to be economically negligible. To ascertain that, we compute the bias in the variance induced by ignoring first and second order serial correlation, which turns out to be, respectively, -0.318% and -0.319% of the correct variance.[6] Therefore, it does not make any substantial difference if we remove the autocorrelation from our series or not. The percentage bid-ask spread is consistently small, around 0.06%. Our findings seem to suggest that here we do not need to worry about market microstructure issues such as the bid-ask bounce, which would bring a strong negative serial dependence and complicate the analysis further, by

[5]Roll simply defines a measure for the bid-ask spread in percentage of the geometric average of the average bid and ask prices as: $s_r \equiv 2\sqrt{-\mathrm{Cov}[R_{t-1}, R_t]}$.

[6]Our measure for the bias in the variance is obtained by comparing the variance of the returns with the variance of the residuals resulting from fitting, respectively, an AR(1) and an AR(2) process to the high frequency returns.

introducing a serious bias in the volatility measures.

5.4 Data Analysis and Derivation of the Model

Throughout the present section we develop a careful step by step analysis of the data, aimed at isolating its main distinctive features and their nature, and, therefore, at providing directions for plausible model specifications.

We start by postulating a very general structure for our model of the dynamics of intraday returns, represented as follows:

$$r_{it} = s_{it}\sigma_{it}\varepsilon_{it} \tag{5.1}$$

for $i = 1, \ldots, 82$, $t = 1, \ldots, T$, where r_{it} represents the unconditional intraday (or overnight) log return for interval i at day t;[7] s_{it} identifies the volatility for sub-interval i at day t attributable to the seasonal pattern in intraday volatility; σ_{it} stands for the stochastic volatility component, independent of the seasonal component;[8] ε_{it} symbolises the conditional intraday log return, with zero mean, independent of both the seasonal and the stochastic volatility parts. Once both the seasonal and stochastic volatility components have been correctly modelled, this latter component should translate into a series which is independent[9] across the intraday intervals.

An important consideration needs to be made here. The structure in Eq. (5.1) is indeed a very rich specification which admits an infinite variety of models as special cases. Nevertheless, even these very general assumptions may easily be too strong. However, in this work we do not take our assumptions for granted, but

[7] r_{it} is de-meaned by the sample mean on the corresponding sub-interval which, however, is not significantly different from zero.

[8] The choice of such a specification for the volatility seemed natural, since the empirical evidence indicates that both a periodic pattern and a stochastic volatility component exist in intraday volatility and need to be modelled.

[9] The assumption of independence is, in fact, justified by the absence of a significant autocorrelation structure, as ascertained in the previous section.

we attempt to test their validity as much as possible, as part of our data analysis. Finally, it is worth emphasising that we do not attempt to model the risk premium, given that four years of data would not be a sufficiently long time span to obtain reliable estimates for such purpose.

A note on the terminology that will be used in the present work and on the scaling of the model in (5.1). After ascertaining the deterministic nature of the seasonal volatility component, we choose to work with de-seasonalised returns, which involves scaling our model so that $E[\sigma_{it}] = 1$ and $E[|\varepsilon_{it}|] = 1$.[10] The use we make here of the term volatility to denote σ_{it} is therefore quite unorthodox and, to be precise, we should refer to σ_{it} as "relative volatility".

In the following subsections we proceed to investigate the features of the data that characterise the nature of the components in Eq. (5.1) and then, to propose and test accurate model specifications for each component.

5.4.1 The seasonal component

Fig. 5.1 displays the intraday patterns in average returns (plotted with 95% confidence intervals around the zero mean) and average absolute returns, computed across the time series of the single 5-minute intervals.[11] No clear predictable pattern is discernible for the average returns, whereas an obvious U-pattern characterises the intraday volatility. The presence of a U-pattern in intraday volatility of stock returns was first documented by Wood, McInish and Ord [1985] and Harris [1986], whereas the impact of this seasonal component on the volatility dynamics was first investigated by Andersen and Bollerslev [1997].

---

[10]Throughout the paper, we prefer to use absolute, rather than squared, returns, to measure volatility. As largely documented in the existing literature (see, for an exhaustive discussion, Barndorff-Nielsen and Shephard [2003]), absolute returns are less sensitive to large outliers and more reliable when the fourth moment of the distribution of returns is not finite.

[11]Overnight returns are not included in the plots. Their average and average absolute values of, respectively, 0.02% and 0.41% are clearly not in line with the rest of the intraday data and their inclusion would have distorted the analysis.

The average absolute returns start out at about 0.1% at the market opening, keep on increasing for the first half an hour up to 0.12%, decline smoothly to the lowest level of 0.062% before noon and then increase again to 0.113% at the closure of the cash index market. The final spike in the last 15 minutes of the trading day is attributable to the post cash market trading.

In principle, the model in Eq. (5.1) allows for a seasonal volatility component that changes through time t. To test for the stability of the seasonal pattern across different moments, we analyse the shape of the intraday periodicity in volatility on subsamples computed on single days of the week, on the first and the second half of the entire sample period, and on the 50% highest and the 50% lowest volatility days (Fig. 5.2).[12] The visual inspection of the plots reveals the presence of very similar and almost indistinguishable patterns in intraday volatility. We also conduct more formal tests of equality between intraday volatility patterns for, respectively, high and low volatility days, first and second half of the sample, each trading day of the week and the overall sample. We first perform, for each intraday interval, a two sample Student's t-test for mean equality (at 95% confidence level) on the average normalised absolute returns of the two subsamples we want to compare. The percentage of sub-intervals on which the null hypothesis of equal means is rejected (on the total of 82 intervals) is displayed in the second column of Table 5.3. We then derive the series of intraday ratios computed on the average normalised absolute returns of the two subsamples of interest; average values and standard deviations for these series are also reported in Table 5.3. Both the small percentages of rejections for the mean equality test (ranging from 0 to 12%) and the low dispersion of the ratios of intraday volatility coefficients around the average level of one (with values for the standard deviation between 0.055 and 0.075), seem to support the stability

[12]The comparison amongst subsets of data with different levels of volatility has been made possible by using absolute returns normalised by their average value across the day, taken as a volatility proxy for the day.

of the seasonal pattern. Therefore, our results suggest that over the time period spanned by our data we can safely assume constant deterministic intraday seasonal pattern, which can then be appropriately represented by s_i.[13]

Since the intraday periodicity in volatility has a strong impact on the dynamic properties of intraday returns, it is essential to correct for this component in order to reveal and model the stochastic volatility dynamics present in our series. The average absolute returns for the individual sub-intervals constitute simple estimates of the intraday seasonal component in volatility s_i, both for the 5-minute intervals and for the overnight returns.[14] Different approaches have been proposed in the literature to obtain smoothed estimates of these seasonal coefficients. In a recent comparative paper, Martens, Chang and Taylor [2002] recommended the Flexible Fourier Functions approach, first suggested by Andersen and Bollerslev [1997]. According to this technique, smoothed estimates of the seasonal components (only for the 5-minute subperiods) are produced by fitting Flexible Fourier Functions (which consist of combinations of linear, quadratic, trigonometric functions and dummy variables) to multiplicative coefficients based on the average absolute returns. In Fig. 5.3 we show how good smoothed estimates can be obtained much more easily by fitting a set of cubic B-splines to the average absolute returns for the 5-minute intervals.[15] Cubic B-splines present several advantages over the FFF approach: they are easier to fit, more flexible, more general and they do not rely on *ad hoc* specifications for the inclusion of dummy variables.

Having obtained accurate smoothed estimates for the deterministic intraday

---

[13]There is a chance that different seasonal patterns arise during daylight saving periods, which we have not investigated. However, we expect the impact of a correction in this direction to be much less important for equity data than it would be for exchange rate data.

[14]Taylor and Xu [1997] proposed a similar adjustment for seasonality, based on averages of squared returns.

[15]Once again, the overnight period has been excluded from the analysis. As estimate of the overnight seasonal volatility component we use the average absolute overnight returns. Also note the presence of a spike in our cubic B-splines curve, due to a knot placed to capture the drop-and-rise movement typical of the futures contract.

pattern in volatility,[16] we proceed to derive the time series of de-seasonalised unconditional intraday returns, by dividing the unconditional returns r_{it} by its seasonal coefficient estimate s_i. In the same way we compute the time series of de-seasonalised unconditional overnight returns. As expected, the sample higher moments for the series of de-seasonalised unconditional returns (both 5-minute and overnight), displayed in bottom half of Table 5.1, are very similar to the corresponding sample moments of the original returns series.

5.4.2 The stochastic volatility component

The analysis of daily volatility estimates

Once we adjust for the intraday periodicity in volatility, the model in (5.1) translates into a mixture process, such that each de-meaned and de-seasonalised intraday return is a combination of independent realisations from a stochastic volatility process and from a conditional density. Therefore, the next step is to identify an appropriate stochastic volatility process capable of generating good intraday volatility estimates.

To investigate the presence and the nature of the stochastic volatility component, we start by plotting the autocorrelogram of the absolute de-seasonalised 5-minute[17] returns for $4,100$ lags, corresponding to 50 days (Fig. 5.4, top half). The highly significant serial correlation in absolute intraday returns over many lags reveals an important stochastic component in volatility. The slow decay of the autocorrelation coefficients through time indicates the persistence of such component.

Our target is to model stochastic volatility at a high frequency level. However, any single 5-minute absolute return obviously gives a very poor estimate of volatility, as confirmed by the strongly irregular pattern of the ACF. We can elimi-

[16] To ensure that no relevant information has been lost as an effect of the smoothing procedure, we have also computed de-seasonalised high frequency returns on the average intraday absolute returns, and re-estimated our model on this series. The changes in the results were negligible.

[17] In our notation, 5-minute should not be interpreted literally since it also refers to the overnight interval.

nate most of this noise by working initially with daily averages of the intraday absolute returns, in order to understand the low frequency component of the volatility dynamics. The daily average of the 82 5-minute absolute unconditional returns is computed as follows:

$$\hat{\sigma}_t = \frac{1}{m} \sum_{i=1}^{m} \left| \frac{r_{it}}{s_i} \right| \tag{5.2}$$

The measure in Eq. (5.2) directly relates to the realised volatility measures mentioned earlier, which have been recently recommended by several authors as very accurate estimates of the latent daily volatility.

The model estimated on these volatility proxies will yield good estimates of daily volatility. Only if volatility were constant across the 5-minute subintervals of a same day, such estimates would also be accurate at an intraday level. We know that this is not the case, therefore at a high frequency level such estimates will show inaccuracies due to both some measurement error and intraday changes in the volatility. Instead, our daily volatility estimates will prove to be very useful at revealing most of the structure that our stochastic volatility model should possess in order to capture the essential features of the volatility dynamics, including the impact of the leverage component. Moreover, the daily estimates can be considered good enough for the purpose of normalising the series of unconditional intraday returns and, consequently, providing more precise information on the shape of the conditional distributions. This information will prove to be essential in order to re-estimate the stochastic volatility model at an intraday level.

The autocorrelogram for the daily average of absolute high frequency returns up to lag 50 is displayed at the bottom of Fig. 5.4 to provide a comparison with the one at the top and, therefore, to ensure that our measure for the volatility at a daily level reproduces the basic characteristics displayed by the volatility estimates at an intraday level (hopefully highlighted by the reduction in the noise). As expected,

the elimination of most of the noise produces an overall increase in the level of serial correlation for the daily volatility, which is around four times as much as the intraday level. Also the averaging process has the obvious effect of drastically reducing the very high autocorrelation recorded in intraday volatility for the first 150-200 lags. The visual inspection of the serial correlation in absolute returns at both daily and 5-minute level reveals some common features. First the stochastic volatility factor seems to be the result of two components: a) a fast mean reverting component; b) a more persistent component, which seems to decline almost linearly in time. Secondly, the irregularities in the ACF are still present, since averaging the absolute returns across a day reduces the noise component, but does not reduce the sampling variation in the ACF.[18]

To assess whether the hypothesis of a constant level of stochastic volatility throughout a day would be too unrealistic, in Fig. 5.5 we plot the time series of the coefficients of variation for our measure of daily volatility, obtained by dividing the standard deviation of the series of absolute intraday returns across one day by the corresponding daily average. This useful dimensionless statistic expresses both the sampling variation across a day and, most importantly, the occasional fat-tailness in the distribution of conditional returns, or the occasional spikes in the stochastic volatility occurring within the day. For most days the coefficient of variation does not vary too much around its average level of 0.8 and therefore, the daily average of absolute returns represents a reasonably accurate measure of the level of the volatility within that day. However, we observe a few spikes in the time series of the coefficients of variation that can only be explained with the presence of different distributions for conditional returns within one single day, or with sudden pulses in

[18]It is worth noticing that the ACF is informative when the underlying model is linear and the variables are Gaussian, which is not the case in our context. When we deal with non-linearities in the model and heavy tailed distributions, the ACF might suggest the presence of spurious long memory effects (see Davis and Mikosch [2000]).

the volatility during that day. The nature of these spikes, that must be taken into account in order to obtain a valid model for high frequency data, will be investigated later on in the chapter.

The statistical techniques adopted to estimate our stochastic volatility model provide more accurate results the closer the series to be modelled is to a normal. Therefore, we have chosen to work with the time series of the logarithm of the daily volatility proxy $\ln(\hat{\sigma}_t)$, whose skewness of 0.40 and excess kurtosis of 0.31 are much closer to the corresponding moments of a Gaussian than those of the volatility proxy itself (equal to, respectively, 1.94 and 6.63).

In the following subsections, we first investigate the impact of the leverage effect and suggest a model to account for the changes in volatility induced by this component. We then explore two alternative ways of modelling the (daily) dynamics of the volatility in such a way to capture the features described above.

The leverage effect

The leverage effect was first discussed by Black [1976], who observed how the amplitude of the volatility of a stock tends to increase when its price drops. A simple plot of the time series of the two variables[19] reveals the presence of a leverage effect in our sample. However, a direct comparison between volatility and stock prices is not possible, since the first series is stationary and the second one is not.[20] In order to investigate presence and magnitude of the leverage effect, we propose a specification which is very easy to estimate and well supported by our data.

The distance between the current level of the index and its moving average quite naturally represents a new, stationary, variable, which is intuitively related to the volatility of the index. The average index level is computed as an exponentially

[19]Omitted for clarity of exposition.

[20]Here, we refer to stationarity in a technical sense, i.e. a stationary series is one that reverts back to a constant mean over sufficiently long periods of time.

weighted moving average M of closing log prices for the S&P 500 stock index futures: $M_t = (1 - \theta dt)M_{t-1} + \theta \ln(S_{t-1})dt$. The new stationary series is derived as $\ln(S_t) - M_t$. A measure of the leverage effect is then given by the correlation between $\ln(S_t) - M_t$ and the log volatility for the following period $\ln(\sigma_{t+1})$.

To quantify the leverage effect for variables measured at a daily level we proceed as follows. The initial value M_0 is set equal to the initial log price and we choose $\theta = 0.03$ (corresponding to a half life of 23 days), which is the value that maximises (in absolute terms) the correlation between the series of daily log price movements and daily log volatility proxy. For this parametrisation, we obtain a correlation of $\rho = -0.545$ between the two series, which confirms the existence of a strong leverage effect, whose impact on the volatility dynamics needs to be adequately modelled.

In order to separate the changes in volatility induced by the leverage effect from those arising from the dynamics of the stochastic volatility component, we propose the following specification:

$$\ln(\hat{\sigma}_t) = \kappa(\ln(S_{t-1}) - M_{t-1}) + \upsilon_t \tag{5.3}$$

The regression in Eq. (5.3), performed on daily measures, provides us with an estimate for κ of -4.34 (standard error 0.26) and with time series of daily residuals υ_t, whose evolution thus mirrors the dynamics of the (ex-leverage) stochastic volatility. Given that the ACF inspection carried out in the previous section suggests the presence of both a transient and a more permanent component in the volatility process, and that the leverage effect turns out to be quite persistent, we start by assessing whether the volatility expressed by the residuals could be adequately modelled by means of an AR(1) specification.[21] Unfortunately, this simple and appealing specification is immediately ruled out, as indicated by the ACF of

[21] The continuous time equivalent of an AR(1) model is a standard Ornstein-Uhlenbeck process.

the residuals from the AR(1) process (Fig. 5.6, top), which clearly highlights the existence of a more persistent dynamics ignored by our model.

More complete specifications capable of taking into account this feature are then needed in order to achieve a satisfactory model for the stochastic volatility component.

A long memory model

In an attempt to find directions for modelling the stochastic volatility after correcting for the leverage effect, we look back at the ACF of our volatility measure, which seems to suggest a slow, hyperbolic-decaying pattern in the serial correlation, typical of a long memory volatility process. We then try to fit Autoregressive Fractionally Integrated Moving Average models ARFIMA(p,d,q) of different kinds to our data.[22] We recall that the ARFIMA(p,d,q) model for a process x_t is given by:

$$\phi(L)(1-L)^d(x_t-\mu)=\theta(L)\varepsilon_t \tag{5.4}$$

where d is the order of fractal integration, $\phi(L)$ is the polynomial of degree p for the autoregressive component, $\theta(L)$ is the polynomial of degree q for the moving average component, μ represents the expectation of x_t and ε_t is a zero-mean white noise process with variance ϵ^2. For $0<d<1$ the process has long memory and it is covariance stationary if $0<d<0.5$.

Exact Maximum Likelihood (EML) and Modified Profile Likelihood (MPL) techniques have been applied to estimate various ARFIMA models on the time series of the residuals from Eq. (5.3).[23] Since the results from the two estimation methods are very similar, we only report the findings from the MPL technique in Table 5.4, which indicate the ARFIMA(0,d,0) model with a fractal integration

[22] See Baillie [1996] for an excellent review of long memory processes.

[23] In our analysis μ has been replaced by the sample mean of the series.

parameter $d = 0.485$ as the best candidate for our long memory stochastic volatility model.

In order to check how appropriate the estimated process is to describe the dynamics of the volatility for our data, we compare the theoretical ACF for the ARFIMA(0,$\hat{d}$,0) model with the ACF of our daily residual series, up to 50 days lag (Fig. 5.7). Clearly the estimated long memory model does not constitute a representation of the volatility process consistent with our entire sample. To get alternative estimates for d we then run an experiment in which we assume that the correct specification for the volatility is a long memory model, but we obtain the fractal integration parameter by minimising the distance between the theoretical ACF and the sample ACF. For $H = 30$ and $H = 50$ we get estimates of $d = 0.34$ and $d = 0.22$.

Since the implementation of different procedures produces very different estimates for the parameter d (hence, for the most appropriate long memory model) and consequently the preference towards one or another would be too *ad hoc*, we can conclude that the choice of a long memory process for explaining the dynamics of the stochastic volatility component is not fully supported by our data set.

A short memory model

The slow, almost hyperbolic, decay in the sample autocorrelogram for the stochastic volatility, which suggests the presence of long memory effects, could actually have an alternative explanation. Gallant, Hsu and Tauchen [1999] showed how, for an appropriate choice of parameters, the sum of two AR(1) processes exhibits long memory features. Modelling the stochastic volatility as a sum of two AR(1) or equivalently, in continuous time framework, with a superposition of Ornstein-Uhlenbeck processes, provides a sufficiently accurate description of the empirical results, while maintaining the nice properties of a short memory process.

We therefore explore the use of a model similar to Alizadeh, Brandt and Diebold [2002], and represent log volatility in continuous time as the sum of two independent Ornstein-Uhlenbeck processes, each of them mean reverting towards the long run level of zero.[24]

Since we employ the residuals from Eq. (5.3) as our (ex-leverage) log volatility measure, the discrete time version of the model for the dynamics of the log volatility becomes:

$$\begin{aligned} \upsilon_t &= \ln(\sigma_t) + \xi_t \\ \ln(\sigma_t) &= \ln(\sigma_{s,t}) + \ln(\sigma_{l,t}) \\ \ln(\sigma_{s,t}) &= \rho_s \ln(\sigma_{s,t-1}) + \beta_s \sqrt{\Delta t}\omega_{s,t} \\ \ln(\sigma_{l,t}) &= \rho_l \ln(\sigma_{l,t-1}) + \beta_l \sqrt{\Delta t}\omega_{l,t} \end{aligned} \tag{5.5}$$

The two log volatility components follow a Gaussian first-order autoregressive process with mean zero, autoregressive parameter $\rho_j = 1 - \alpha_j \Delta t$ (where α_j denotes the mean reversion parameter) and variance $\beta_j^2 \Delta t$. $\Delta t = 1$,[25] since we estimate the model on daily volatility proxies, whose measurement error ξ_t is small and can be easily bounded.

The variance of the measurement error associated with our log volatility proxy should not be very large, since intraday sample path information is embedded in the proxy, contributing to the reduction of the noise. In fact, the distribution of the residuals from the measurement equation includes both the noise component and the sampling variation from the conditional distribution of the log volatility proxy, since $\ln(\hat{\sigma}_r) = \ln(\sigma) + \ln(\hat{\sigma}_\varepsilon)$ and, therefore, $\xi = E[\ln(\hat{\sigma}_\varepsilon)] + \epsilon$. Ideally we

[24] In our multiplicative model, the expected value for the stochastic volatility is one, therefore the level at which the log volatility must revert is zero. In fact, statistical estimates of the long run mean turned out to be not significantly different from zero.

[25] The empirical issue of the choice of *dt* at intraday level, in view of the overnight market closure, will be discussed later on in the paper.

should be able to separate the two effects but in practice this turns out to be very difficult. Since the variance of the log volatility proxy from unconditional returns is equal to the sum of the variance of the true log volatility and the variance of the error term, this latter quantity represents an upper bound to the percentage of the total variance attributable to measurement error. In our example, the variance of the residuals, equal to 0.052 accounts for 38.5% of the total variance of the log volatility measure (equal to 0.135). A lower bound on the variance explained by measurement error is obtained by calculating what the variance of the log volatility proxy from conditional returns would be if the conditional distribution of the returns was normal, i.e. $N(0, \sqrt{\pi/2})$.[26] The variance of the log of the daily average of absolute intraday conditional returns can then be computed numerically: in our case it is equal to 0.0072, which corresponds to 5.3% of the total variance for the log volatility proxy on the unconditional returns.

The estimation is carried out by applying a Kalman filter algorithm to the state space system in Eq. (5.5). If the measurement equation errors ξ_t were normally distributed, we could obtain exact maximum likelihood estimates of the model, and the linear projections produced by the Kalman filter procedure would represent conditional expectation. As we will discuss later, in our case the measurement errors turn out not to be Gaussian. However, quasi-maximum likelihood procedures can still produce consistent estimates of the parameters.[27]

The estimated parameters, with standard errors in brackets, are displayed in Table 5.5. We can clearly identify a transient volatility component, with $\alpha_s = 0.734$ corresponding to a half life of 0.94 days and a more persistent one, with $\alpha_l = 0.018$ and half life of approximately 37.5 days. Most of the short-run variance of the model

[26] Under the assumption of normally distributed conditional returns, their variance must be equal to $\pi/2$ in order to satisfy the condition $E[|\varepsilon_{it}|] = 1$.

[27] The model has also been estimated by maximising the spectral log-likelihood function for the sum of two AR(1) processes. The two estimation methods produce essentially the same results.

can be attributed to the transient component, as the values of the β coefficients suggest, whereas 52% of the unconditional long-run variance is explained by the more persistent component. The ACF of the residuals from the two factor AR(1) specification (Fig. 5.6, bottom) reveals how all the dynamics of the volatility has now been correctly captured.

The distributions of the residuals from both the state equations of the two components and the measurement equation have been analysed (Fig. 5.8). All the residual series exhibit positive skewness and excess kurtosis, which lead to a rejection of their normality. However, for all the distributions, skewness and fat-tailness are not too pronounced. This "approximate" Gaussianity should ensure a reasonable efficiency of both the Gaussian quasi-maximum likelihood estimates and the consequent inferences about the latent volatility process.

Some insights on the conditional return densities

The estimation of the stochastic volatility model on a daily basis provides us with both a structure for the dynamics of the stochastic volatility component, and estimates of the (log) volatility level, adjusted daily according to the new value for the log volatility proxy $\ln(\hat{\sigma})$. Although inaccurate as 5-minute volatility estimates, these constant intraday volatility measures can be usefully employed to extract information on the distribution of conditional returns, as a necessary preliminary step to perform in view of refining the estimates of our model at a 5-minute level.

The time series of conditional returns is obtained by normalising the unconditional de-seasonalised return, r_{it}/s_i, by the volatility estimate for day t made at the end of the previous day.[28]

[28] A remark is appropriate here: in order to avoid tautological results, when deriving the series of conditional returns we always condition on the information available before the actual return was observed. In other words, the return r_{it} observed on interval i of day t will be conditioned on the volatility estimate produced the day before (for daily-updated volatility forecasts) or the 5-minutes before (for 5-minute updated volatility forecasts).

Once the volatility dynamic has been accurately modelled, if the conditional return distribution is Gaussian and independent from the volatility process, then the conditional intraday returns will be identically distributed across all intervals of the day and no changes in the shape of their density (i.e. more fat-tailed in intervals of higher activity and less fat-tailed when there are less transactions on the market) should be discernible.[29]

In order to empirically assess such hypotheses, we start by computing summary sample statistics of the time series of conditional returns for each of the 82 intraday intervals. First we investigate how these sample statistics relate to the theoretical ones from a normal distribution, and then we discuss their stability across the 5-minute subintervals. We plot in Fig. 5.9 (top-left) the standard deviation of the time series of the conditional returns for the individual intervals, together with a straight horizontal line at $\sqrt{\pi/2}$, which represents the theoretical level of standard deviation under the assumption of normality for the distribution of conditional returns.[30] We can clearly detect a few spikes for some intraday intervals, that seem to suggest the fat-tailed nature of the conditional distribution. However, looking more carefully at our data, we can see that the spikes are mainly attributable to a very small number of outliers (around 15 for the whole dataset, i.e. less than 0.020% of the total observations) that distort the tails of the distributions over some intervals (not necessarily the busiest ones). Disregarding the spikes, the standard deviation of the conditional distribution of intraday returns turns out to be fairly close to the theoretical $\sqrt{\pi/2}$ level.

However, the fact that the volatility of the empirical distribution is persistently higher than the theoretical normal one seems to suggest that the conditional

[29]Similarly, given our choice of scaling, if we refer to the distribution of absolute conditional intraday returns, the following properties should hold: $E[|\varepsilon_{it}|] = 1$ and $Var[|\varepsilon_{it}|]$ constant for $i = 1, \ldots, m$.

[30]Under the assumption of normally distributed conditional returns, their variance must be equal to $\pi/2$ in order to satisfy the condition $E[|\varepsilon_{it}|] = 1$.

distribution is more fat-tailed than a Gaussian. The plot showing the average excess kurtosis of the conditional returns for the individual intraday periods is reported in Fig. 5.9 (bottom-left). Again, we observe a high level of excess kurtosis over some intervals of the day, which is mainly due to the presence of few sporadic outliers. If we removed these outliers, the excess kurtosis for the overall conditional distributions would be around 2, pushing the distributions much closer to a Gaussian.[31]

Once we find evidence of near-Gaussian 5-minute conditional return densities, we turn to the investigation of the stability of the shape of such distributions across the intraday intervals. The comparison amongst the values of standard deviation and excess kurtosis recorded for the individual subintervals reveals that the main source of variability in the shapes occurs as an effect of the few outliers already discussed above. Except for that, on average the values of the summary sample statistics computed for the intraday conditional distributions prove to be quite flat across the 5-minute intervals of the day.

To further support the hypothesis of conditional intraday return distributions having a similar shape across different periods of the day we also compare the Q-Q plots drawn by contrasting the time series of conditional returns for each of the 82 intraday intervals with the overall series of 5-minute conditional returns. A sample of plots for the overnight period, two busy and one quiet intervals are displayed in Fig. 5.10. The plots for the different intervals are not clearly distinguishable from each other, which suggests that the shape of the conditional distributions does not differ much across different moments of the day.

To summarise, the distribution of conditional returns computed by normalising upon constant intraday volatility forecasts turns out to be surprisingly close to Gaussian and virtually the same across the different 5-minute intervals of the day.

[31] The average skewness of the conditional returns for each of the intraday intervals has also been computed. However, its analysis is not particularly informative, given that, apart from very few exceptions, skewness coefficients do not depart significantly from zero.

However, it exhibits a small degree of fat-tailness that could be explained by the changes in volatility across the day that our simplified estimates do not capture. If that was the case, the derivation of intraday volatility forecasts updated every 5-minutes should correct for the fat-tailness, and conditioning upon them should give conditional returns distributed according to a Gaussian. This will be the object of the following subsection.

5.4.3 The estimation of the model at an intraday level

As previously stated, the model calibrated on a daily basis cannot accurately describe the actual dynamics of high frequency data. For this purpose, the estimation of our continuous time specification must be refined by exploiting the information content of the 5-minute return series. The aim of the present subsection is to derive improved estimates for our continuous time model by means of a simple non-linear filtering technique in which the update occurs every 5 minutes, based on observed intraday market returns.

The assumption of a continuous time specification for our model seems to be supported by the inspection of the serial correlation $\rho_{t,t+k}$ for 5-minute absolute returns within the same day (t and $t+k$ belong to the same day, $k \leq 80$) and between one day and the following (t and $t+k$ belong to adjacent days, $k \leq 161$), reported in Fig. 5.11. The fact that the two segments for $k \leq 80$ are different, with a higher serial correlation within the same day, indicates that the overnight period has a significant impact on the mean reversion of the model, an evidence in favour of a continuous time specification.

In deriving high frequency estimates, we start by dismissing the persistent component of the stochastic volatility process in Eq. (5.5), since we expect the contribution of the fast mean-reverting part to be predominant for such purpose. We also ignore the impact of the leverage effect at intraday level, given that this

component is quite persistent and, therefore, its effect should be better investigated and modelled at a lower frequency level. To justify our choices, we have computed the proportions of the variance of log volatility innovations at 5-minute frequency attributable to each component. The results, derived in Appendix A.1, indicate that the proportions of variance explained by the leverage effect and the persistent volatility component are fairly small (respectively, less than 5% and 4% of the total variance). Hence, these components can be safely disregarded for the purpose of improving the high frequency volatility process.

In what follows, we first consider a standard diffusion model for the 5-minute volatility process, of the kind described in Eq. (5.5) for the transient component. We then introduce jumps in our volatility specification which will significantly improve our return density forecasts.

A simple diffusion process for intraday volatility

In order to describe the intraday volatility dynamics, we maintain the standard Gaussian Ornstein-Uhlenbeck specification employed in the daily model to characterise the evolution of the transient component, and we make use of the information available on high frequency returns to obtain improved estimates for the parameters of the process.

To avoid imposing strong structural assumptions on the derivation of high frequency volatility estimates, the latter are obtained and updated through a simple non-linear filtering technique based on observed intraday market returns. A range of possible discrete values for the log volatility $\ln(\sigma_j)$ for $j = 1, \ldots, N$ is specified, together with the corresponding set of initial probabilities P_j assigned to each value. These initial probabilities are then combined with the transition probabilities $P^a_{i,j}$ between log volatility values j and i to produce a discrete set of prior probabilities

P_i^* for $i = 1, \ldots, N$ as follows:[32]

$$P_{i,t}^* \approx \sum_{j=1}^{N} P_{(i,t),(j,t-1)}^a P_{j,t-1} \tag{5.6}$$

which will then be applied to the corresponding volatility values in the range in order to return the intraday variance estimate $\sigma^{*^2} = \sum_{i=1}^{N} P_{i,t}^* \sigma_i^2$. Within this framework, the discretisation of our continuous time volatility process is achieved by evolving the analogous discrete mean reverting process on a trinomial grid structure. The resulting transition probabilities, which we assume constant, are derived in Appendix A.2.1.[33]

Under the assumption of normality for the conditional returns, justified on the basis of the results derived earlier, a density forecast for the unconditional de-seasonalised returns r_t^f is represented by a mixture of normal densities, where each component is a normal with zero mean and standard deviation equal to one of the volatility values in the range multiplied by $\sqrt{\pi/2}$, and the mixing probabilities are given by the prior probabilities for the individual values in the volatility range:

$$r_t^f \sim \sum_{i=1}^{N} P_{i,t}^* N(0, \sigma_i \sqrt{\pi/2}) \tag{5.7}$$

Once the 5-minute unconditional de-seasonalised return r_t/s_t is observed, the probability $P_{i,t}^r$ that such return represents an observation from each of the Gaussian components of the mixture is computed and a Bayesian probability update is applied to the set of prior probabilities. This will produce a corresponding set of posterior probabilities $P_{i,t}^p$:

---

[32]For simplicity of exposition, here t denotes the intraday moment previously indicated as it, hence $t = 1, \ldots, 82T$.

[33]For the practical implementation of the model, we chose a log volatility range between -1.5 and 1.5, with step size equal to 0.1, roughly corresponding to three times the estimated volatility of the mean reverting process. Alternative similar choices for the volatility range and the step size have been investigated, and the results do not differ too significantly. For simplicity, the initial probabilities P_j are assumed to be equal.

$$P^p_{i,t} \approx \frac{P^r_{i,t} P^*_{i,t}}{\sum_{j=1}^{N} P^r_{j,t} P^*_{j,t}} \tag{5.8}$$

which will replace the initial probabilities P_i in order to re-start the process. Unlike in the previous case, here volatility and return density forecasts are updated every 5 minutes on the basis of the actual evolution of returns observed in the financial market.

An important empirical issue concerning the implementation of our continuous time specification is the choice of the time step Δt. Our data seems to support a time step equal to 1/106 for 5-minutes intervals and 25/106 for the overnight period.[34] Since the variance of the unconditional overnight returns is about 25 times the variance of the corresponding 5-minute returns, this specification is equivalent to expressing time in calendar terms during the trading day and in volatility terms overnight. This hybrid choice seems clearly unorthodox, compared to the specifications commonly employed in the relevant literature. Traditionally, models are derived entirely in calendar time or in volatility time, with the introduction, in the latter case, of a stochastic time change associated with a measure of intraday trading activity. The attempts of estimating our model according to these conventional specifications have been rejected by our data, which instead supports a model where a time change is needed only for the overnight period. The most plausible explanation we can provide for such evidence is that, contrary to our work, most related studies explicitly exclude overnight returns from the dataset, since their peculiar properties may introduce a significant complication in the analysis of the volatility dynamics. The inclusion of overnight returns in our data has probably induced such an unconventional choice of the time step.

At this point, we need estimates of both the mean reverting coefficient α_s

[34] It is worth noticing that during the 25 steps of the overnight period, the process evolves only on the grid and the Bayesian update of the probability does not occur until the opening price for the day is known.

and the volatility parameter β_s such that the likelihood that the observed returns are realisations of our non-linear filtering model, given by:

$$L(r_T) \approx \prod_{t=1}^{T} \left(\sum_{j=1}^{N} P_{j,t}^{r} P_{j,t}^{*} \right) \tag{5.9}$$

is maximised and that, on average, the volatility of the intraday changes in the log volatility estimates is equal to the volatility parameter β_s of the process.[35] Working on a grid of possible values for α_s and β_s^2 (spaced at a step of, respectively, 0.05 and 0.01, which turns out to be a good compromise between complexity and accuracy) we find that estimated values of $\alpha_s = 0.6$ and $\beta_s = 0.28$ meet these requirements.

In the present work, we employ very simple filtering and estimation techniques to produce step-by-step volatility and return density forecasts and to obtain estimates of the relevant parameters. Much more sophisticated econometric methods have been developed recently in the literature: auxiliary particle filtering techniques have been suggested for volatility filtering in continuous and discrete time models by Pitt and Shephard [1999], Durham and Gallant [2002], Chib, Nardari and Shephard [2002], Johannes, Polson and Stroud [2002]. Markov Chain Monte Carlo (MCMC) methods provide very precise parameter estimates for a variety of diffusion and jump-diffusion models (see Elerian, Shephard and Chib [2001], Eraker [2001], Eraker, Johannes and Polson [2003]). Estimation strategies based on GMM procedures have been proposed, amongst the others, by Singleton [2001] and Bollerslev and Zhou [2002]. The implementation of such techniques would certainly improve the accuracy of our results, but at the cost of an increased complexity which would not be justified in our context, given that the best possible estimation accuracy is not our main concern. Therefore we choose to use simple techniques that still

[35] The likelihood function for mixture models is known to be unbounded at some points on the edge of the parameter space (see Kiefer [1978]). In our case, however, we do not attempt to maximise the likelihood *per se*, and we only use it to discriminate between various set of parameters that satisfy the volatility constraint.

produce reasonably accurate estimates, as will be shown later.

Once the model is fully parametrised, 5-minute volatility estimates and return density estimates can be extracted. The time series of conditional returns is now obtained by normalising unconditional returns r_{it} upon the intraday volatility forecasts σ^*_{it} computed 5-minutes earlier.

The analysis aimed at investigating shape and constancy of the conditional return distribution across the different subintervals of the day is replicated. Again we compute summary sample statistics for the time series of conditional returns for each of the 82 intraday intervals and we report plots of the standard deviation (top-right Fig. 5.9) and the excess kurtosis (bottom-right Fig. 5.9) across the individual subintervals.[36] A few spikes due to the presence of very large outliers, rather than to the effect of some external source of information not captured by our model, can still be easily detected. If we ignore these outliers, we observe values for both the standard deviation and the excess kurtosis very close to the values we would have for normally distributed conditional returns, with a standard deviation closely oscillating around the value of $\sqrt{\pi/2}$ and an average excess kurtosis of 0.9. In accordance with the hypothesis of a Gaussian specification, the summary sample statistics for the conditional return densities also exhibit very similar values across all the intervals of the trading day.[37]

These findings are entirely in line with our expectations: volatility estimates which are updated at a high frequency level can account for most of the fat-tailness left in the conditional return density after normalising with respect to volatility estimates which remain constant across the day. The assumption of a Gaussian distribution to describe returns computed by conditioning on accurate 5-minute volatility

[36] As before, the inspection of the skewness does not add much value to our analysis, since almost all of the coefficients display values not significantly different from zero.

[37] As before, the visual inspection of the Q-Q plots reveals an indistinguishable shape for the conditional return distribution across the single intervals. The results have been omitted for clarity of exposition.

estimates turns out to work surprisingly well, actually better than expected.

At this stage, in order to correct for the presence of the outliers and for the residual fat-tailness, while maintaining a specification valid in continuous time, we propose to introduce jumps in the model.

A process with jumps in intraday volatility

The introduction of jumps can take place in the return process, in the volatility process, or in both. To avoid arbitrary assumptions, we analyse the nature of the outliers (identified with all conditional returns larger, in absolute value, than $3\sqrt{\pi/2}$) to decide whether they are more likely to represent jumps in returns or in volatility.

In order to investigate if the increased volatility consequent to a jump is persistent, the regression $(\mid r_{t+1}/s_{t+1} \mid -\sigma^*_{t+1}) = a + b(\mid r_t/s_t \mid - \mid \bar{r/s} \mid)$ has been run at a 5-minute level on both the entire sample and the subsample where r_t are all outliers. The estimated coefficients of $a = -0.014$ (s.e. 0.010) and $b = 0.027$ (s.e. 0.014) for the entire sample, and $a = 0.03$ (s.e. 0.058) and $b = 0.109$ (s.e. 0.017) for the outliers, suggest that the impact of the jumps seems to persist and not to die out immediately, as the nature of jumps in returns would predict. The inspection of the temporal distribution of the outliers[38] highlights a significant clustering in the incidence of jumps, which again contradicts the i.i.d. assumption which characterises jumps in returns. Our empirical results indicate that the outliers exhibit more the features of jumps in volatility than those of jumps in returns. This is in line with some recent findings which pointed out how models with diffusive stochastic volatility and jumps in returns are incapable of capturing the empirical features of equity returns (see Pan [2002], Bates [2000], Duffie, Singleton and Pan [2000], Eraker, Johannes and Polson [2003]).[39]

[38] Not reported here for clarity of exposition.

[39] A more rigorous specification would also allow for jumps in returns. For simplicity, here we restrict our attention to jumps in volatility, which still yields good results.

We then need to ascertain how persistent the impact of jumps on the intraday volatility turns out to be, with the aim of deciding to which volatility process (transient or permanent) the jumps should be added. In Fig. 5.12 we plot the average (computed across all outliers) difference between post-jump volatility levels and the average 5-minute volatility level across the 82 subintervals preceding the jump, for several post-jump intervals, ranging from five minutes to three day after the outlier has occurred.[40] The rapid decay in the volatility difference suggests that the inclusion of the jumps can be safely restricted to the transient volatility component.

The continuous time process for the dynamics of intraday volatility then becomes:

$$d\ln(\sigma_{s,t}) = -\alpha_s dt \ln(\sigma_{s,t}) + \beta_s \sqrt{dt} dW_{s,t} + \sum_{i=1}^{N(t)} Y_i - \lambda dt E[Y] \qquad (5.10)$$

where $N(t)$ denotes the total number of jumps in dt (arrivals of a Poisson process with intensity λ) and Y_i are i.i.d. random variables corresponding to the Poisson jump magnitudes.[41] In order to fit the discrete version of this stochastic volatility model into our grid structure, we need to work with jumps of discrete size (i.e. Y_i will have a discrete distribution), expressed as a multiple of our step size Δ_r. To simplify the analysis, jump sizes and intensities are assumed to be constant.

At this stage, having identified the essential features of the jump component, i.e. its non i.i.d. nature and its fast-decaying impact on the intraday volatility, we can allow some flexibility in specifying the details of the model as, for instance, the jump magnitudes and intensities. The discrete values for the jump magnitudes and the respective intensities are obtained via calibration with the empirical features of the outliers. Our outliers can be roughly grouped into three categories: small (up to 6 standard deviations), medium (between 6 and 12 standard deviations) and large (above 12 standard deviations). This suggests three possible kinds of jumps in

[40] As usual, the intraday volatility is approximated by 5-minute absolute returns.

[41] A compensated jump process is chosen to maintain the mean of the volatility process unchanged.

volatility: a small jump, of size equal to $3\Delta_r$ which, given our log volatility model, is roughly equivalent to an outlier in conditional returns of 4 standard deviations; a medium jump of size $10\Delta_r$, corresponding to 8 standard deviations; a large jump of size $20\Delta_r$ roughly equivalent to 22 standard deviations. A simple investigation of the frequency of the outliers reveals that, on average, a jump occurs every other week, therefore we choose $\lambda = 1/10$ as our overall jump intensity expressed on a daily basis. A more detailed analysis of the frequency of jump incidence for each group (given by the number of jumps occurred within that group divided by the total number of days in the sample) returns frequencies of 0.00729 (= 7/960) for large jumps, 0.027 (= 26/960) for medium jumps and 0.0708 (= 68/960) for small jumps. On the basis of these empirical frequencies, properly re-scaled by the overall jump intensity $\lambda = 1/10$, we derive probabilities of jump incidence for each group equal to $p^l = 8\%$ for the biggest jumps, $p^m = 27\%$ for intermediate jumps and $p^s = 65\%$ for small jumps.[42]

Once the jump sizes and intensities have been specified, the Bayesian filtering procedure illustrated in the previous section can be entirely replicated here, with the only difference that the log volatility process evolving on the grid is now the mean reverting model augmented by the jumps component. Therefore, the transition probabilities must be recomputed, following Amin [1993] (details in Appendix A.2.2). The values for log volatility range, initial probabilities, step size and time step are the same as before. The estimates for the remaining parameters of the volatility process produced by our methods are equal to $\alpha_s = 0.65$ and $\beta_s = 0.24$.

As before, we obtain 5-minutes ahead volatility and return density forecasts, whose accuracy in both absolute and relative terms needs to be adequately assessed.

[42]To ensure that our choices of jumps sizes and intensities are not too *ad hoc*, we have replicated the analysis using alternative values obtained from different groupings of the outliers. As long as we allow for at least three kinds of jumps in volatility, the results do not differ too much.

5.5 The Appraisal of Intraday Volatility and Density Estimates

The present section focusses on the assessment of our high frequency volatility and return density estimates, through the implementation of statistical techniques borrowed from both point and density forecast evaluation practice.

Point forecast evaluation techniques are used to assess the 5-minute volatility estimates, through a comparison with the absolute value of de-seasonalised high frequency unconditional returns, taken as a proxy of the actual intraday volatility level. In line with the existing literature on volatility forecasts evaluation,[43] we first regress the absolute return on the volatility prediction, $|r_t/s_t| = \alpha + \beta\sigma_t^* + \epsilon_t$. The forecast is unbiased only if $\alpha = 0$ and $\beta = 1$ and, what is most important for a good prediction, has got small forecast errors if R^2 is large. However, the presence of a strong noisy component in our volatility proxy induces very small R^2 coefficients: the regression performed on returns simulated from exact volatility forecasts yields an $R^2 = 0.2244$, which therefore represents the best we could expect to achieve.

We also report a measure P of the proportion of volatility explained by the forecasts, first introduced by Blair, Poon and Taylor [2001], which compares the amount of variations in the forecast errors with that in actual volatility:

$$P = 1 - \frac{\sum \left(|r_t/s_t| - \sigma_t^*\right)^2}{\sum (|r_t/s_t| - E[|r_t/s_t|])^2} \tag{5.11}$$

Values closer to one indicate better forecasts, with a small variation in the forecast errors. Finally, a standard Mean Absolute Deviation measure, determined as simple average of the absolute deviations of the volatility forecasts from the volatility proxies, is computed.

The findings from the point forecast evaluation of constant and changing 5-minute volatility estimates are displayed in Table 5.6. As expected, the results

[43] See, for a comprehensive review, Granger and Poon [2003].

indicate a very poor forecasting performance in all cases, but we should keep in mind that they are heavily distorted by the noise in the high frequency absolute returns. In a relative comparison, the forecasts updated on an intraday basis (with and without jumps) perform significantly better than the ones updated on a daily basis, as suggested by a higher R^2 (0.16 against 0.09) of the regression, as well as by a higher value of the P statistic (0.14 vs. 0.09) and a slightly smaller MAD (0.635 against 0.64).

Density forecast evaluation techniques are employed to evaluate the intraday density forecasts for the returns. Following a standard procedure, from the sequence of 5-minutes ahead density forecasts $f_t(r)$, we derive the series of the probability integral transforms of the realised intraday unconditional returns taken with respect to the corresponding density forecasts as follows:

$$z_t = \int_{-\infty}^{r_t/s_t} f_t(r)\,\mathrm{d}r \tag{5.12}$$

If the forecasts and the true densities coincide, then the sequence of PITs is distributed as i.i.d. $U(0,1)$. Equivalently, the sequence of transformed PITs, where a transformation to normality is applied to the PITs series (see Berkowitz [2001]), follows an i.i.d. $N(0,1)$.

To guarantee more robust results against possible misspecifications of different type in the forecasted distributions, we implement several goodness-of-fit techniques. The popular Kolmogorov-Smirnov, Anderson-Darling and Watson statistics have been chosen to test for uniformity. The normality is assessed via Jarque-Bera and Doornik-Hansen tests, as well as via normal Q-Q plots. Two likelihood ratio tests are performed to test for independence ($LR1$) and for the joint hypothesis of independent observations with zero mean and unit variance ($LR2$). Since the size of our sample is huge (82 observations for 960 days), virtually any distributional fore-

cast, even a very good one, can be easily rejected. To overcome, at least partially, this problem, we sort our sample in four subsamples according to the level of the volatility forecast.

Before discussing the results, we briefly recall what our intraday density forecasts for the returns look like, under the assumption of normally distributed conditional returns. When the volatility forecasts stay the same across the day, the 5-minutes ahead density forecast for the unconditional returns on each of the intraday intervals is given by a Gaussian with zero mean and standard deviation equal to the forecasted volatility for that day multiplied by $\sqrt{\pi/2}$. For changing intraday volatility forecasts, the density forecast for the returns is represented by the mixture of normal densities derived in the previous section.

The results from the density forecasts tests on both the entire sample and the four groups are reported in Table 5.7. The null hypothesis that the return density forecasts represent accurate predictions of the actual distribution of unconditional returns is generally rejected by all our goodness-of-fit statistics, on the entire sample as well as on the four subsamples, for both constant and changing intraday volatility estimates. However, a substantial improvement is recorded in the production of density forecasts when volatility estimates are updated every 5 minutes, which becomes even more striking when jumps are introduced in the volatility process. The values of the goodness-of-fit statistics (especially of the normality tests) are now much closer to their critical values and the normal Q-Q plots for the case of changing volatility estimates (Fig. 5.13) indicate that empirical quantiles are much closer to the normal ones than for the case of constant volatility estimates (Fig. 5.14). In particular, the fat-tailness induced by the jumps in volatility corrects for most of the misspecification in the tails recorded for both the daily updating method and the intraday method without jumps.

To summarise, in the light of our findings, we can conclude that in order to

correctly model the dynamics at an intraday frequency, we need to rely on a non-linear filtering technique, where the forecasts are updated every 5 minutes on the basis of the observed unconditional returns. Also, a simple diffusion process for the intraday volatility is not appropriate, and a specification which allows for jumps is to be preferred.

A stochastic volatility model of the kind in Eq. (5.5), which works well at both high and lower frequency level, can be obtained by combining the permanent component, whose parameters are estimated on a daily basis, with the transient component (with jumps), whose parameters are estimated on a non-linear intraday filtering model.

5.6 A Monte Carlo Simulation Exercise

Throughout the previous sections the relevant features in the evolution of the observed returns have been carefully isolated, studied and modelled: all the individual components have then been assembled together to produce a complete continuous time model as follows:[44]

$$\begin{aligned}
r_t &= s_i \sigma_t \varepsilon_t \\
dM_t &= \theta(\ln(S_{t-1}) - M_{t-1})dt \\
\ln(\sigma_t) &= \kappa(\ln(S_{t-1}) - M_{t-1}) + \ln(\sigma_{s,t}) + \ln(\sigma_{l,t}) \\
d\ln(\sigma_{s,t}) &= -\alpha_s dt \ln(\sigma_{s,t}) + \beta_s \sqrt{dt} dW_{s,t} + \sum_{i=1}^{N(t)} Y_i - \lambda dt E[Y] \\
d\ln(\sigma_{l,t}) &= -\alpha_l dt \ln(\sigma_{l,t}) + \beta_l \sqrt{dt} dW_{l,t}
\end{aligned} \tag{5.13}$$

In order to: 1) test whether the dynamics of unconditional high frequency returns generated from our model in Eq. (5.13) does actually mirror the empirical

[44] Again, t denotes time on a 5-minute, and not daily, basis, and dt indicates the intraday interval.

one; 2) assess whether our simple estimation procedure produces reliable estimates; we perform a simple Monte Carlo simulation experiment.

A number of 82 intraday unconditional returns is generated each day for a total of 960 days according to the model in Eq. (5.13) where:

- The deterministic seasonal coefficients are given by their smoothed estimates.
- The parameters for the leverage component are those estimated at a daily level in Section 5.4.2.
- Realisations of the stochastic volatility are generated from the two-factor model, where both components have Gaussian innovations, and the parameters are estimated on a daily basis for the permanent component and on an intraday basis with jumps for the transient component.
- Realisations of the conditional returns are obtained by sampling from the Gaussian density $N(0, \sqrt{\pi/2})$.

For simplicity, we only simulate one full sample whose properties will be compared to the empirical ones, with the purpose of verifying whether our data set could actually represent a random sample generated from the model. Four more samples are simulated to assess the estimation technique.

We start by looking at the plots of higher moments, skewness and excess kurtosis, computed across the time series of high frequency returns for each of the intraday intervals, which indicate very similar values for both simulated and observed returns (Fig. 5.15). We then aggregate the simulated high frequency values to derive daily log volatility proxies as averages of absolute de-seasonalised returns, and daily measures of leverage. The time series of these daily simulated variables are contrasted with their daily empirical counterparts (Fig. 5.16, bottom) to check for possible significant differences in the evolution of simulated and observed volatility

proxy and leverage measure. The dependence between log volatility and leverage component from simulated data has also been investigated and compared to the empirical one, via scatter plot (Fig. 5.16, top) and computation of the correlation coefficient, equal to $\rho = -0.533$. The results are very encouraging, since both the temporal evolution of simulated volatility proxy and leverage measure, and their correlation structure closely resemble the empirical ones. These findings at high frequency, as well as daily, level, suggest that the model in Eq. (5.13) seems capable of capturing and replicating the most significant features observed in futures equity returns.

To evaluate the adequacy of the estimation techniques employed so far, we derive estimates of our model from each of the five simulated samples, and compare the resulting parameters with the actual parameters of the data generating process. Following the steps of our data analysis, we start by investigating the seasonal component, whose pattern (not reported here) is indistinguishable from the one shown by the market data for all five simulated samples. Daily measures of log volatility and leverage, computed on simulated data, are then used to obtain estimates for the leverage model through the regression in Eq. (5.3), as well as for the two-factor stochastic volatility model via Kalman filter on the residuals from the previous regression. The estimates, displayed in Table 5.8, are in all cases very close to the original parameters of the process from which the samples have been simulated, and only the mean reversion parameter of the transient volatility component is slightly underestimated in all samples. In relative terms, the larger (but still quite small in absolute terms) dispersion can be observed for the estimates of the parameters of the permanent volatility component.

As before, the non-linear filtering technique with intraday updating of volatility and return density estimates is implemented in order to refine the high frequency volatility process. First we produce estimates of the volatility specification without

jumps and we employ the resulting volatility forecasts to obtain a series of conditional returns. Again the inspection of the outliers provides us with information on the characteristics of the jumps. For simplicity, we maintain the discrete magnitudes of the three kinds of jumps unchanged (equal to $3\Delta_r$, $10\Delta_r$ and $20\Delta_r$, with $\Delta_r = 0.1$) and we only re-estimate the overall and the individual jump intensities on the basis of their frequency of incidence. Finally, we re-estimate the parameters of mean reversion and volatility of volatility on the grid. The estimates for the log volatility process with jumps, shown in the bottom part of Table 5.8, are fairly satisfactory since they turn out to be quite close to the actual parameters of the data generating process. However, we can detect an underestimate of both the overall incidence of jumps, with jumps occurring every 3 or 4 weeks for four out of five simulated samples, and the intensity of the biggest jumps, since no evidence of the presence of large jumps can be found for two samples. The mean reversion and volatility parameters of the diffusion component seem also to be a little underestimated. On the whole, our findings suggest that the estimates produced by applying our simple techniques are quite reliable and adequate for our purposes.

5.7 Conclusions and Further Work

In the present work we have attempted to build a simple and accurate continuous time model capable of describing and replicating the dynamics of both high and moderate frequency index returns. In our approach we have performed a careful analysis of a set of intraday data, aimed at: 1. identifying the relevant features that need to be modelled; 2. investigating the best possible model specification, without imposing too much structure *a priori*, and by testing step by step the assumptions made.

Let us briefly recap the stages that led us to the identification of the complete

model. We have started with specifying a very general multiplicative structure for the model of 5-minute unconditional returns, as a function of a seasonal and a stochastic volatility component, and of intraday conditional returns. We have then examined the nature of the seasonal component in intraday volatility, which has proved to be deterministic. As for the analysis of the stochastic volatility component, given the large amount of noise present in high frequency data, we have derived much less noisy daily average measures of volatility, which represent a considerably more useful starting point for studying the volatility dynamics. On the basis of these daily volatility proxies we have first investigated the presence of a leverage effect in our data and devised a simple specification for its modelling. Following the evidence of the existence of both a transient and a persistent feature in the volatility, we have then explored how to model the ex-leverage volatility dynamics. A two-factor short memory volatility model has been successfully estimated on the daily volatility measures. The volatility estimates obtained from the daily model have also been employed to provide insights on the distribution of the conditional returns, which has turned out to be very close to a Gaussian and fairly stable across the various subintervals of the day. At this stage, the estimation of our model has been refined at an intraday level by exploiting the information content of the 5-minute return series, so as to obtain a specification capable of describing the dynamics of high frequency data. The fine-tuning of the model has been performed by means of a simple non-linear filtering technique, in which the estimates are updated every five minutes, following the observed intraday market returns. Finally, in order to account for the presence of some outliers and for the residual fat-tailness in the model, we have introduced jumps in our volatility specification.

An attractive feature of our work that would deserve further investigation is the possibility to obtain simplified versions of our general model specification, which will possess the correct properties for various specific time horizons of interest. It

would also be interesting to replicate the analysis on data sets of high frequency returns of other financial assets (e.g. currencies, individual stocks).

Table 5.1: Summary sample statistics for intraday returns.

Unconditional Intraday Returns		
	5-minute	Overnight
Mean	-9.95E-09	0.0002
Std. Dev.	0.121%	0.575%
Skewness	0.883	-0.378
Excess Kurtosis	35.476	3.044
Unconditional De-seasonalised Intraday Returns		
	5-minute	Overnight
Mean	-1.15E-05	1.9E-17
Std. Dev.	1.409%	1.408%
Skewness	0.908	-0.378
Excess Kurtosis	35.622	3.044

Table 5.2: First order serial correlation of intraday returns.

	1997-98	1999	2000	2001
Serial correlation	-0.0793*	-0.0901*	-0.0491	-0.0364
Critical value	-0.0372	-0.0423	-0.0493	-0.0570
Estimated BA spread (%)	0.064%	0.064%	0.059%	0.053%

*statistically significant at 5% confidence level.

Table 5.3: Tests for equality of seasonal volatility patterns.

Subsamples	% of rejection mean equality test	Average value ratio intraday coeff.	Std. dev. ratio intraday coeff.
Monday	0.0%	1.002	0.057
Tuesday	12.20%	0.998	0.075
Wednesday	2.44%	0.998	0.056
Thursday	0.0%	1.002	0.055
Friday	4.88%	1.001	0.064
First-second half	12.20%	1.004	0.074
High-low volatility	7.32%	1.000	0.064

Table 5.4: Coefficient estimates for ARFIMA(p,d,q) models.

	Log-likelihood	Coefficient	Std. Error	t-statistic
ARFIMA(0,d,0)	12.55			
d		0.485	0.015	30.5
ARFIMA(1,d,0)	14.01			
d		0.455	0.033	13.6
AR(1)		0.075	0.048	1.58
$\chi^2 = 2.50(0.11)$				
ARFIMA(1,d,1)	14.09			
d		0.458	0.032	14.0
AR(1)		-0.074	0.398	-0.18
MA(1)		0.146	0.381	-0.38
$\chi^2 = 2.45(0.12)$				
ARFIMA(2,d,0)	14.15			
d		0.464	0.035	13.4
AR(1)		0.067	0.049	1.40
AR(2)		-0.019	0.036	-0.54
$\chi^2 = 2.54(0.11)$				

χ^2 is a test for exclusion of the AR(p) and/or the MA(q) terms.
p-values in brackets.

Table 5.5: Coefficient estimates for two-factor AR(1) model.

	ρ	$\alpha(=1-\rho)$	β^2	Half life in days	Unconditional var. $\beta^2/(1-\rho^2)$
Transient	0.261	0.739	0.044	0.94	0.046
	(0.095)		(0.570)		
Permanent	0.982	0.018	0.0019	37.50	0.051
	(0.008)		(0.321)		

Standard errors in brackets.

Table 5.6: Intraday volatility estimates evaluation.

	Daily update	5-minute update	5-minute with jumps
Regression			
α	-0.0164*	-0.0487	-0.0421
	(0.0139)	(0.0083)	(0.0097)
β	1.0505	1.0416	1.0252
	(0.0158)	(0.0173)	(0.0207)
R^2	0.0895	0.1596	0.1677
P-statistic	0.0882	0.1387	0.1429
MAD	0.6433	0.6364	0.6343

* not significantly different from zero at 5% level.
Standard errors in brackets.

Table 5.7: Distributional forecast evaluation.

	Uniformity tests			**Normality tests**		**LR tests**	
	K-S (1.36)	A-D (2.49)	Watson (0.19)	J-B (5.99)	D-H (5.99)	LR1 (3.84)	LR2 (7.81)
Entire sample (78720 obs)							
Daily	3.76*	65.66*	3.32*	13,995*	8,255*	249.18*	1,418.0*
5-minute	3.67*	31.66*	4.66*	2,271*	1,733*	2.36	4.36
5-m. jumps	3.90*	41.17*	6.23*	895*	751*	1.23	5.29
4 Subsamples							
Daily							
Low vol.	2.08*	19.99*	1.09*	3,452*	1,948*	51.78*	348.61*
Medium low	2.19*	17.87*	1.32*	2,724*	1,698*	62.80*	277.21*
Medium high	2.48*	15.83*	0.75*	2,322*	1,503*	68.76*	359.43*
High vol.	2.62*	15.83*	0.67*	5,844*	3,127*	67.04*	445.56*
5-minute							
Low vol.	2.79*	18.59*	2.69*	975*	693*	2.37	14.58*
Medium low	2.40*	13.18*	1.96*	686*	497*	1.06	5.16
Medium high	2.08*	7.64*	1.04*	312*	252*	2.11	5.54
High vol.	2.88*	3.52*	0.50*	425*	345*	1.41	11.91*
5-m. jumps							
Low vol.	3.61*	24.34*	3.99*	480*	373*	2.58	33.04*
Medium low	2.42*	13.89*	2.07*	244*	196*	0.55	6.68
Medium high	2.07*	5.84*	0.80*	152*	133*	0.20	2.60
High vol.	2.92*	3.65*	0.59*	120*	108*	0.71	8.51*

* rejected at 5% level.

Table 5.8: Estimates from simulated samples.

Samples	1	2	3	4	5	Avg.	Std.Dev.	Data
Daily model								
κ	-4.74	-4.02	-4.94	-3.89	-4.15	-4.348	0.464	-4.34
α_s	0.63	0.61	0.68	0.65	0.67	0.648	0.029	0.73
α_l	0.02	0.02	0.03	0.01	0.01	0.019	0.009	0.02
β_s	0.19	0.22	0.20	0.23	0.23	0.214	0.018	0.21
β_l	0.04	0.03	0.06	0.05	0.05	0.046	0.011	0.04
ρ	-0.55	-0.45	-0.57	-0.45	-0.48	-0.500	0.057	-0.54
Intraday model								
α_s	0.70	0.60	0.60	0.60	0.70	0.640	0.055	0.65
β_s	0.20	0.23	0.19	0.24	0.20	0.212	0.021	0.24
p^s	0.70	0.60	0.70	0.65	0.76	0.682	0.060	0.65
p^m	0.30	0.32	0.27	0.35	0.20	0.288	0.057	0.27
p^l	0.00	0.08	0.03	0.00	0.04	0.030	0.033	0.08
λ	0.07	0.05	0.10	0.07	0.05	0.066	0.020	0.10

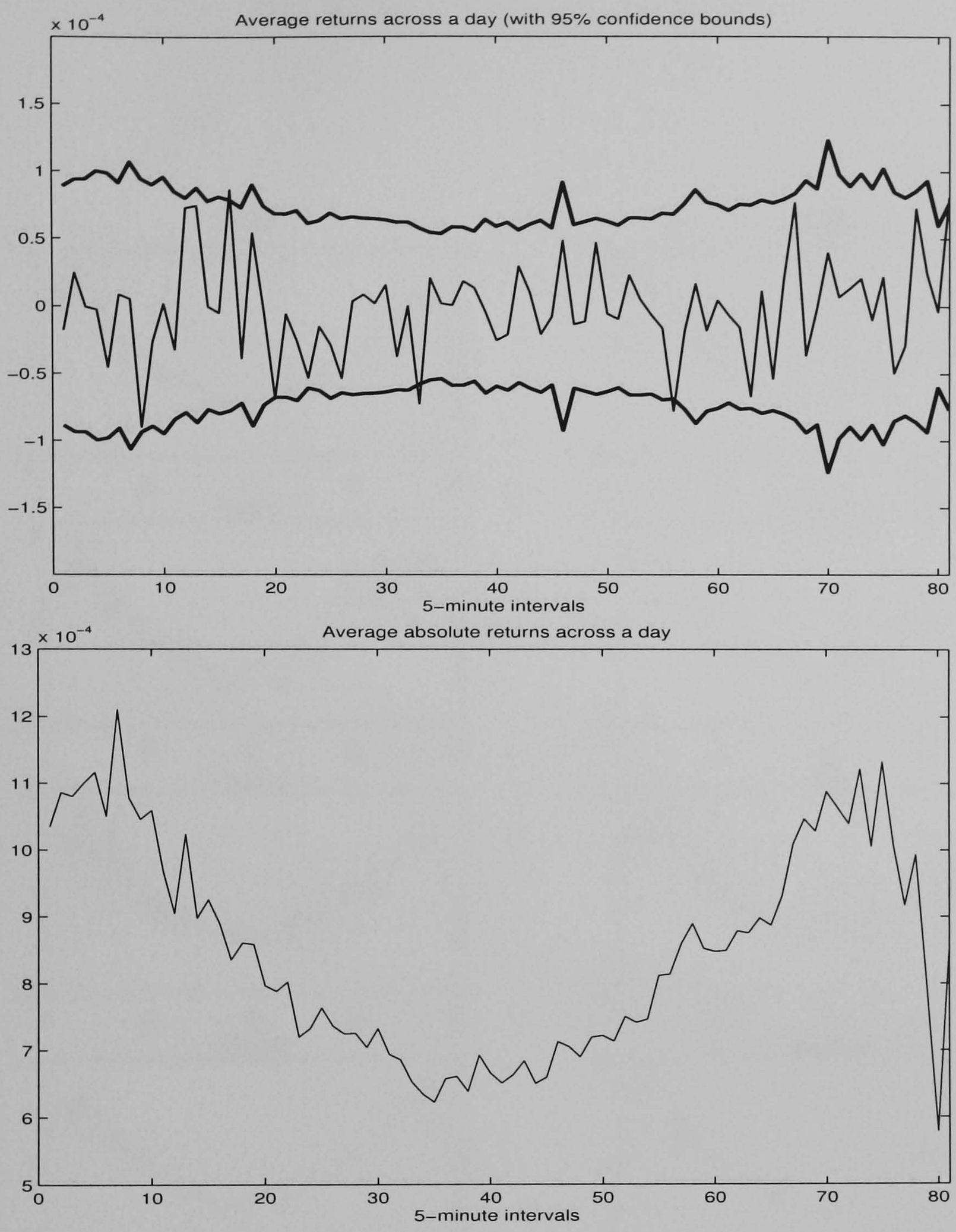


Figure 5.1: Intraday patterns in returns and volatility.

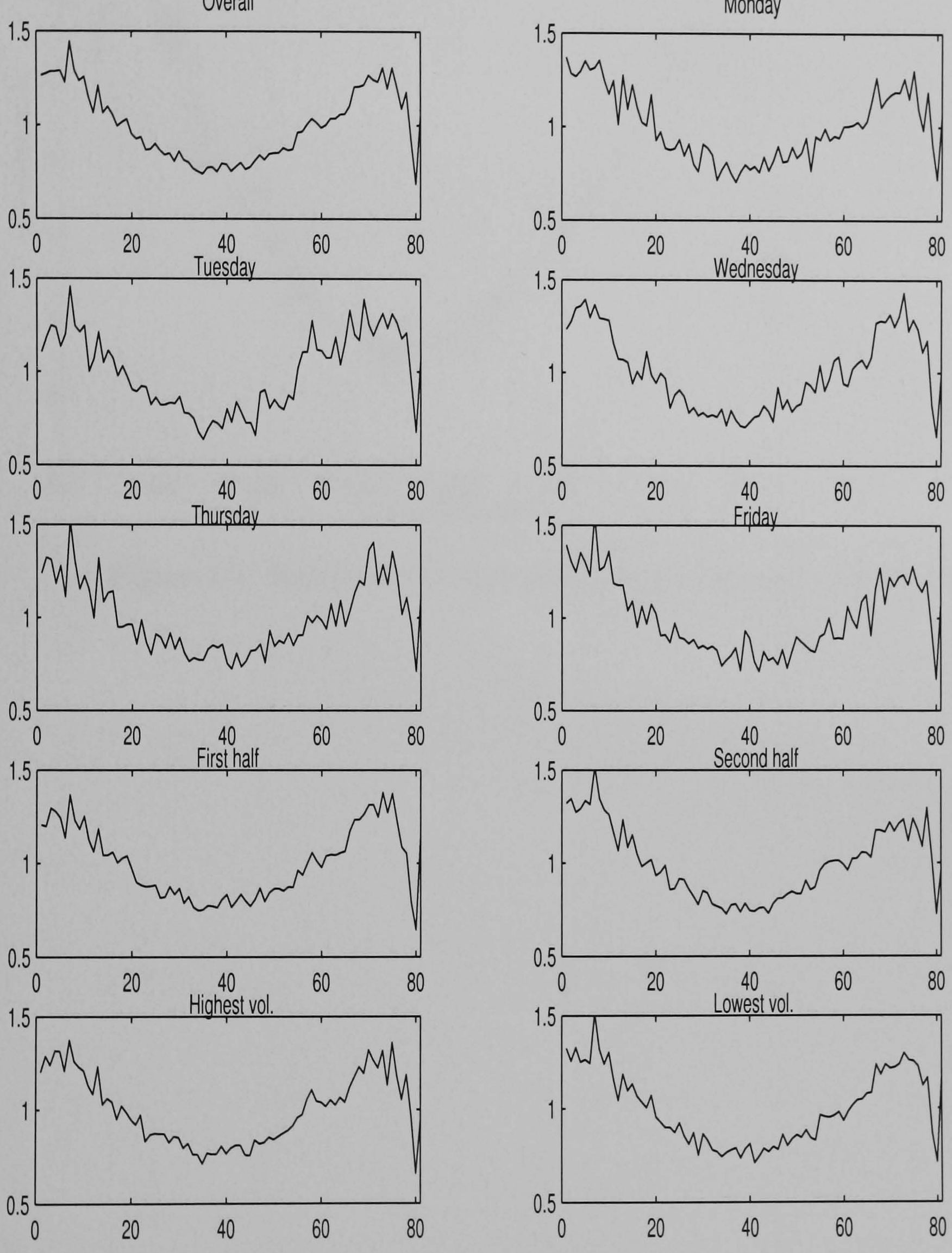


Figure 5.2: Seasonal (normalised) volatility patterns for days of the week and different subsamples.

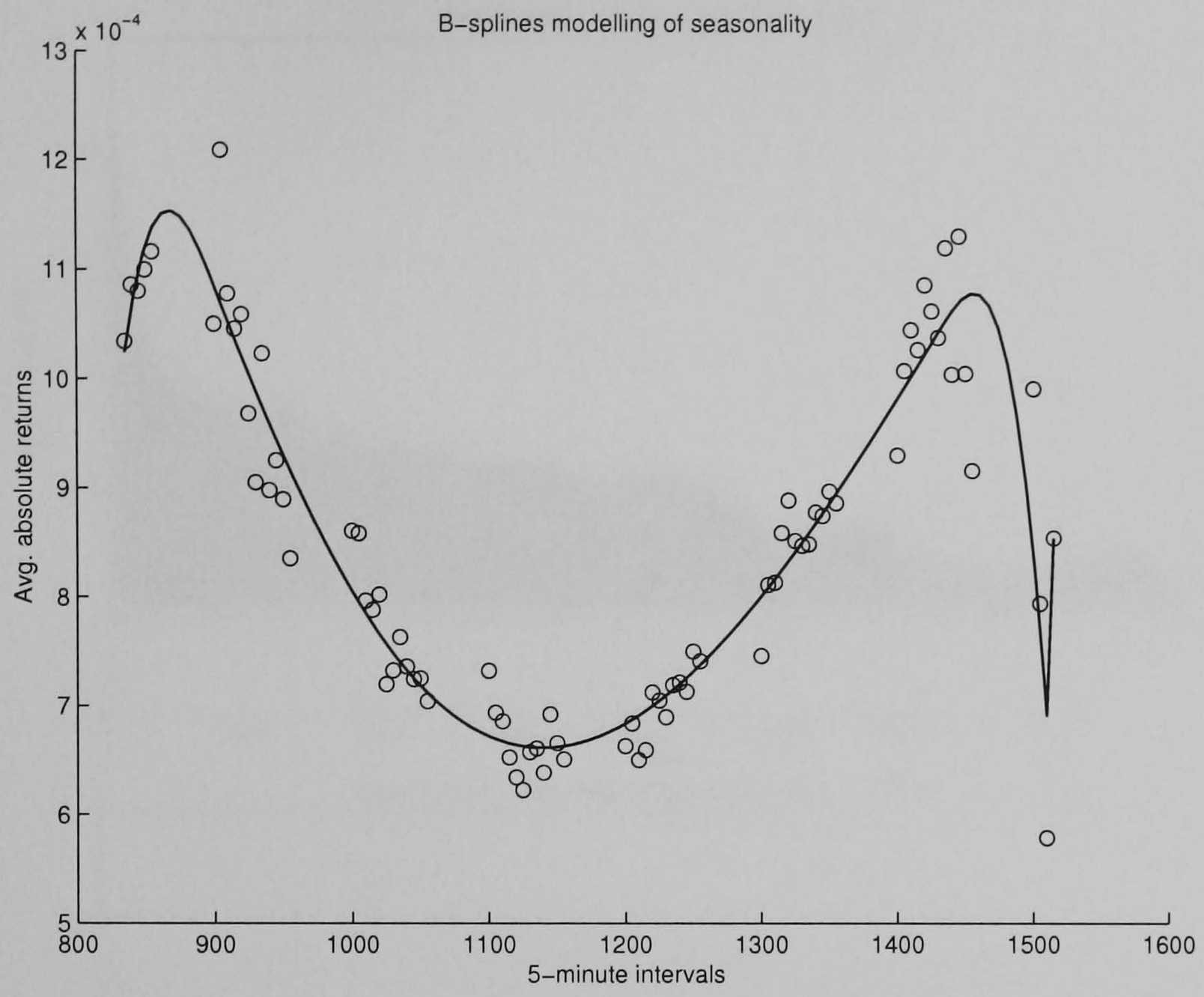


Figure 5.3: Smoothed B-spline estimation of seasonal coefficients.

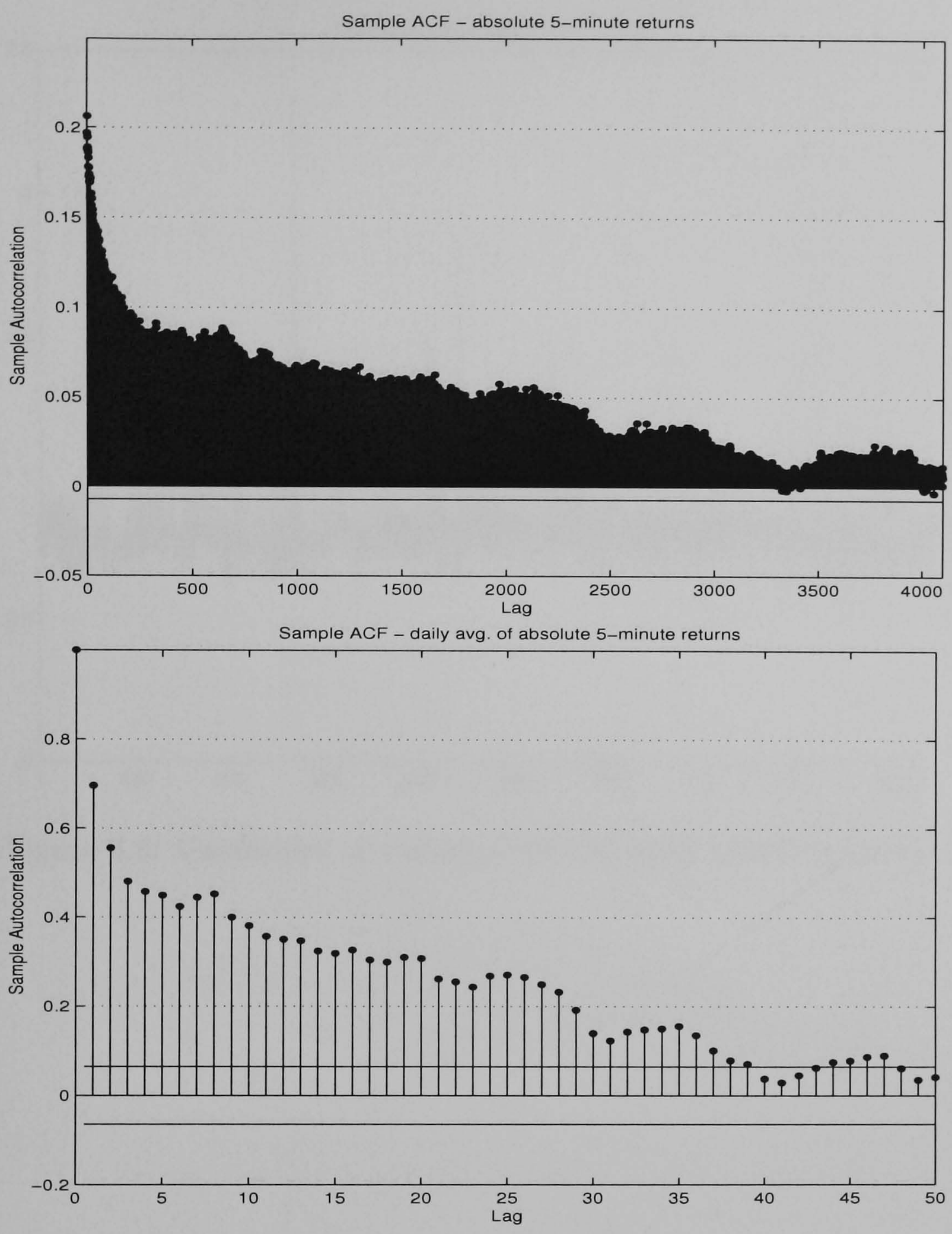


Figure 5.4: ACF of intraday absolute returns and daily average absolute returns.

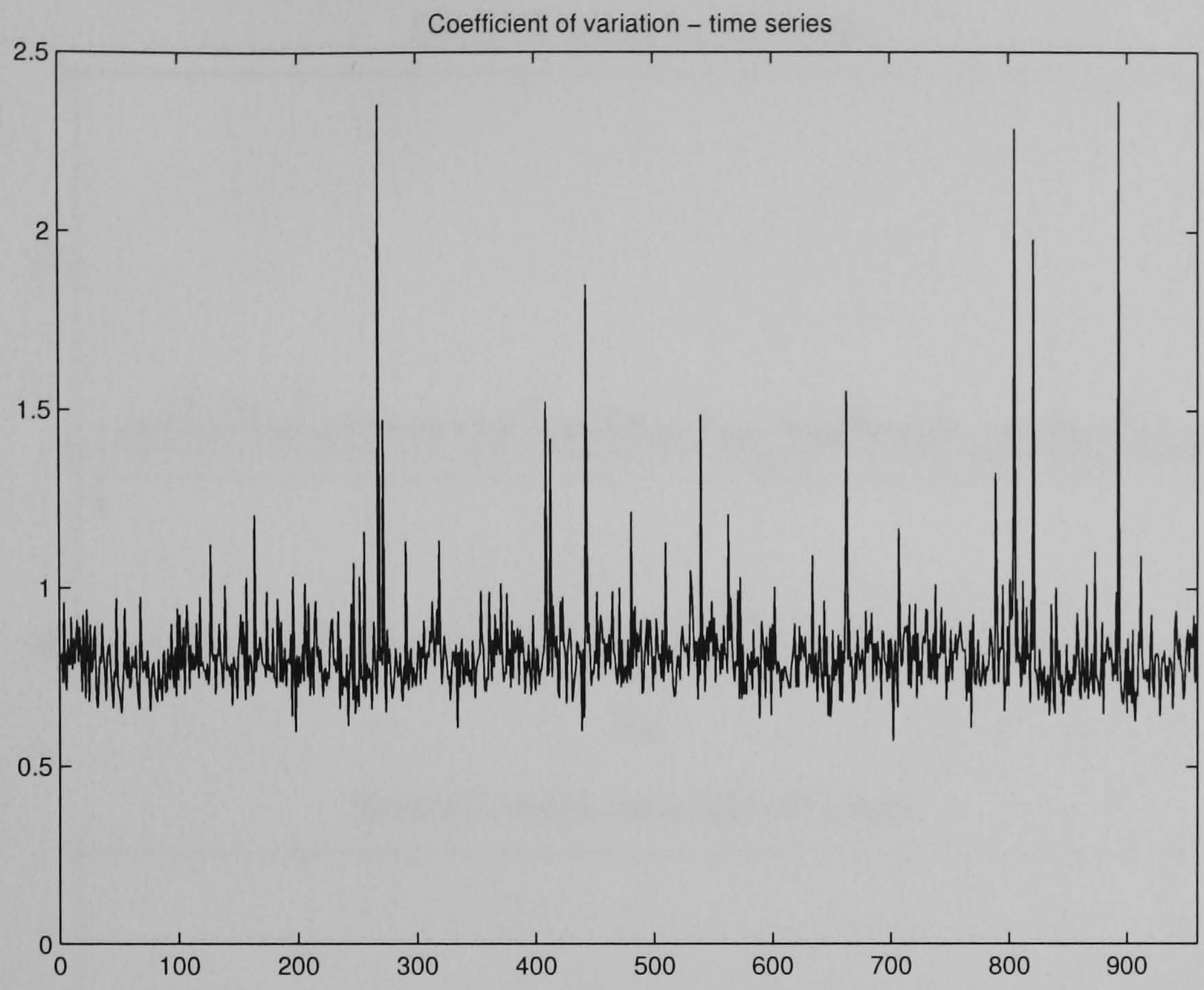
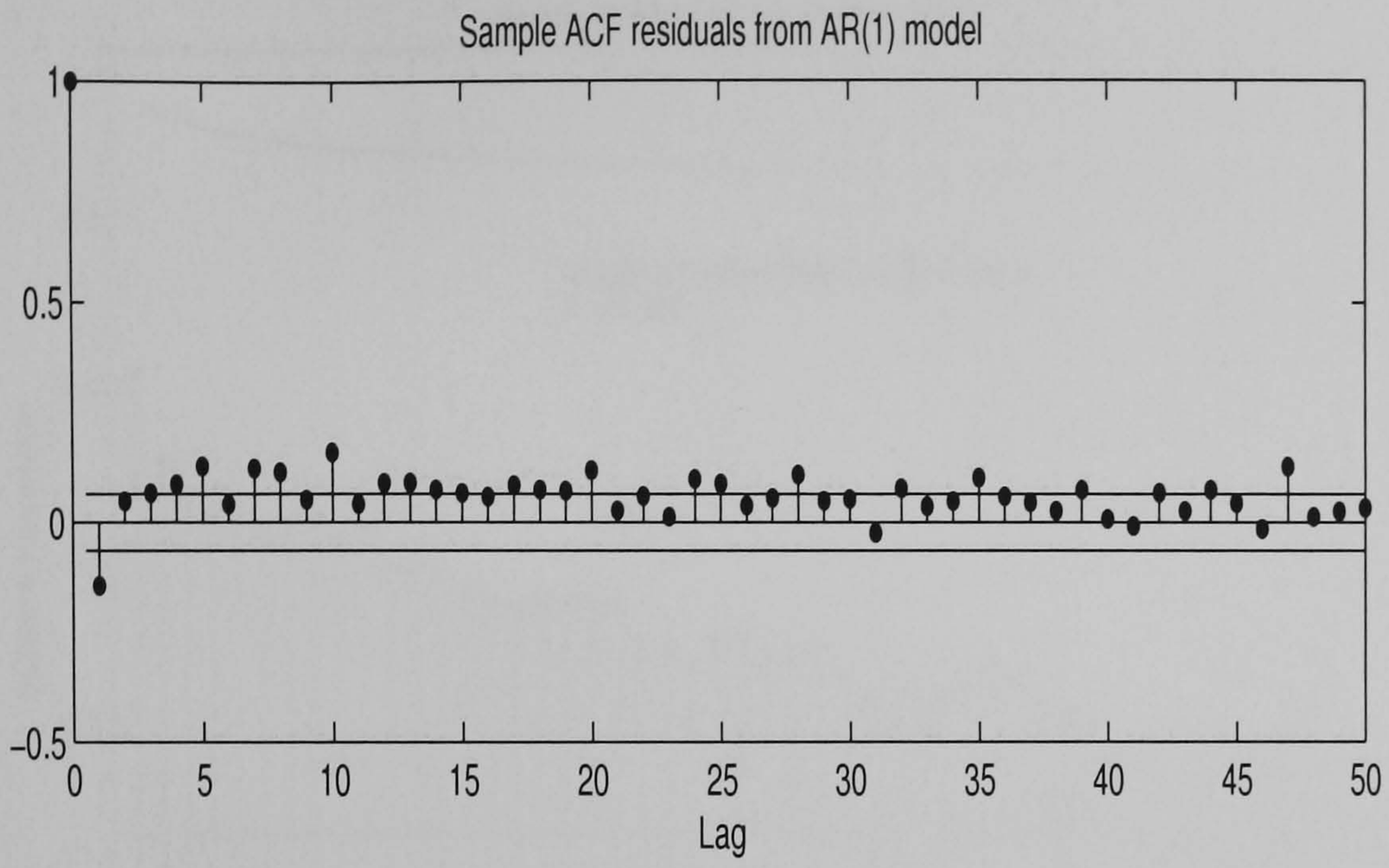


Figure 5.5: Coefficient of variation for the daily volatility measure - Time Series.



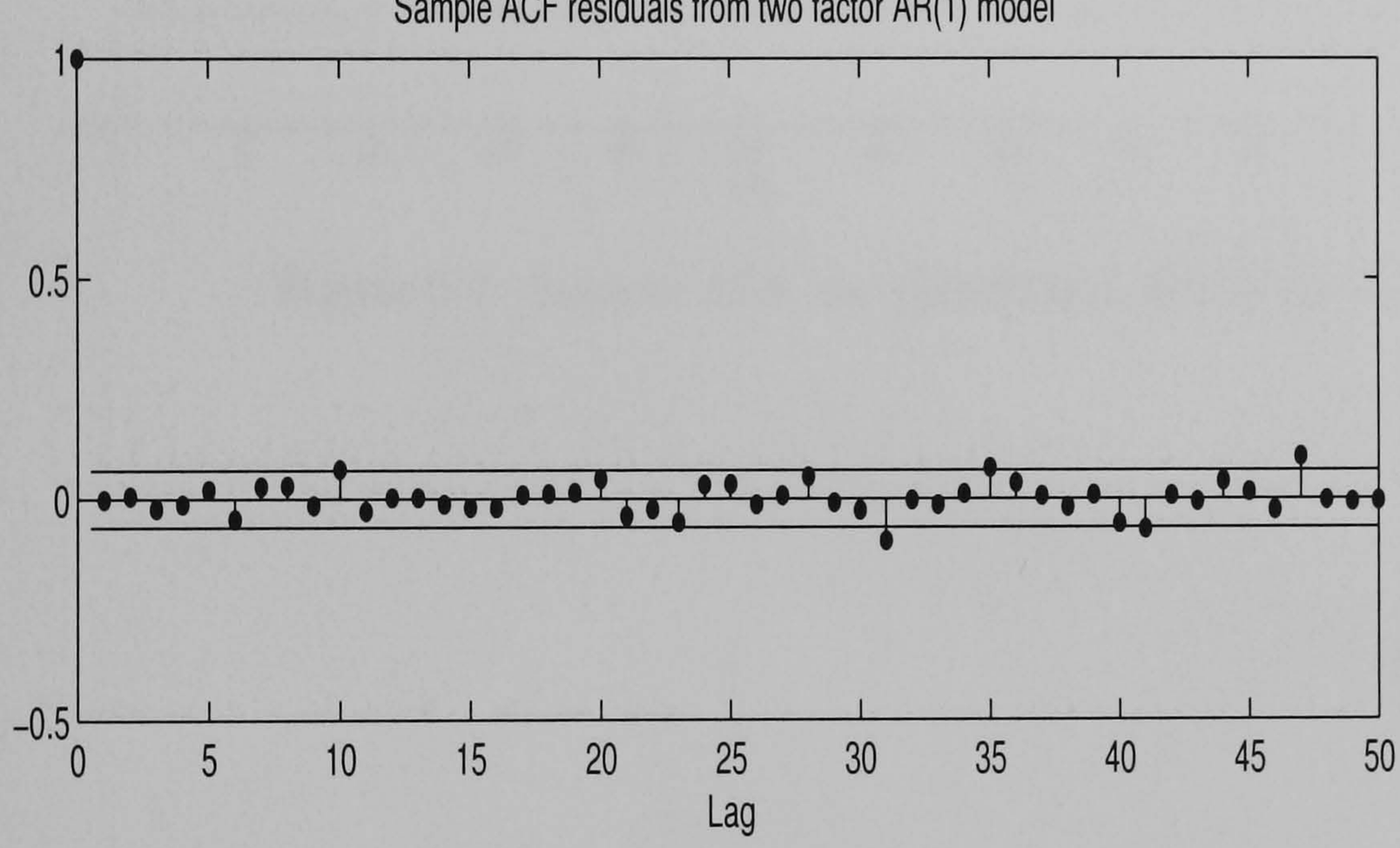


Figure 5.6: ACF of residuals from AR(1) and two-factor AR(1) specifications.

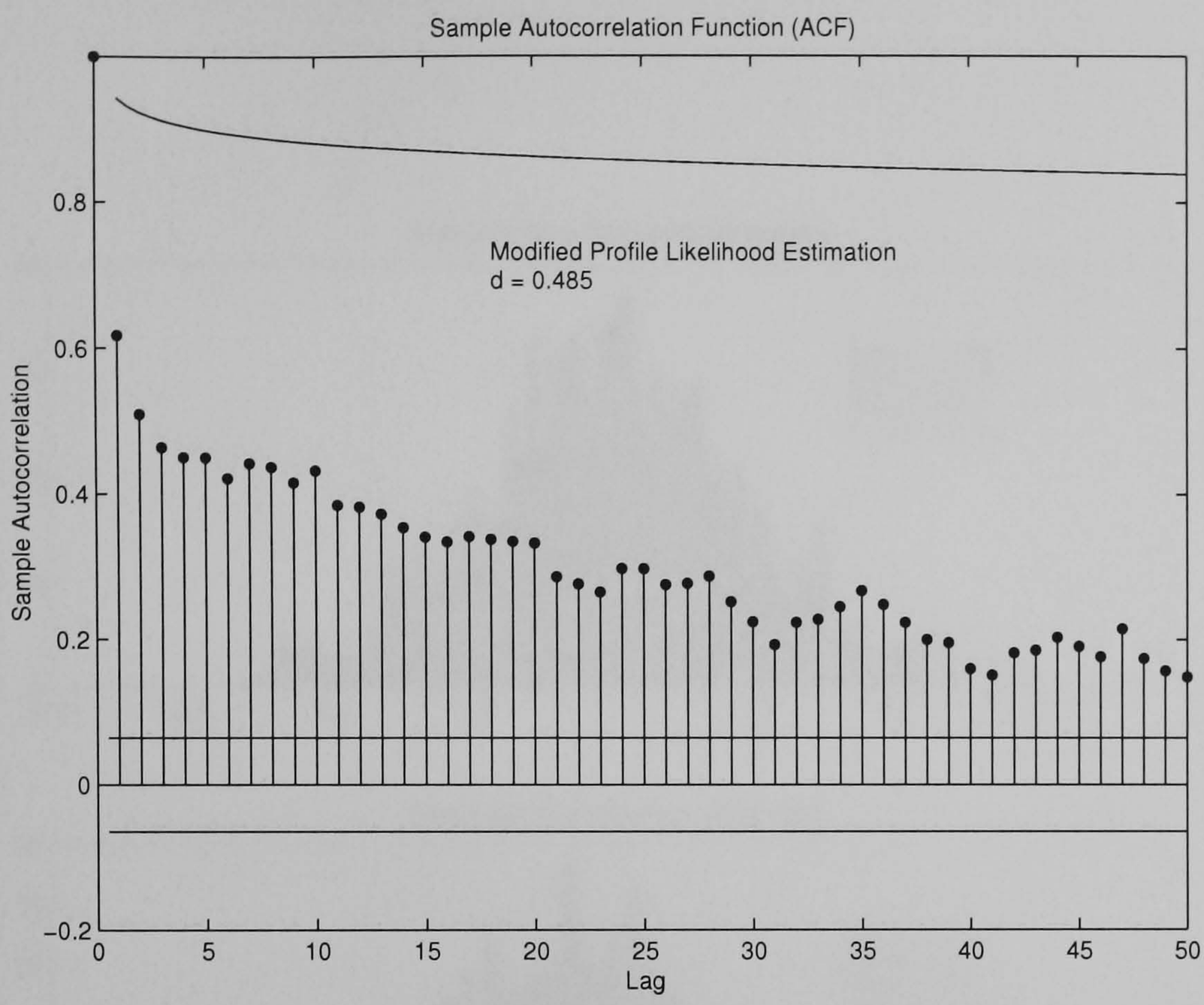


Figure 5.7: Sample ACF vs. theoretical ARFIMA ACF.

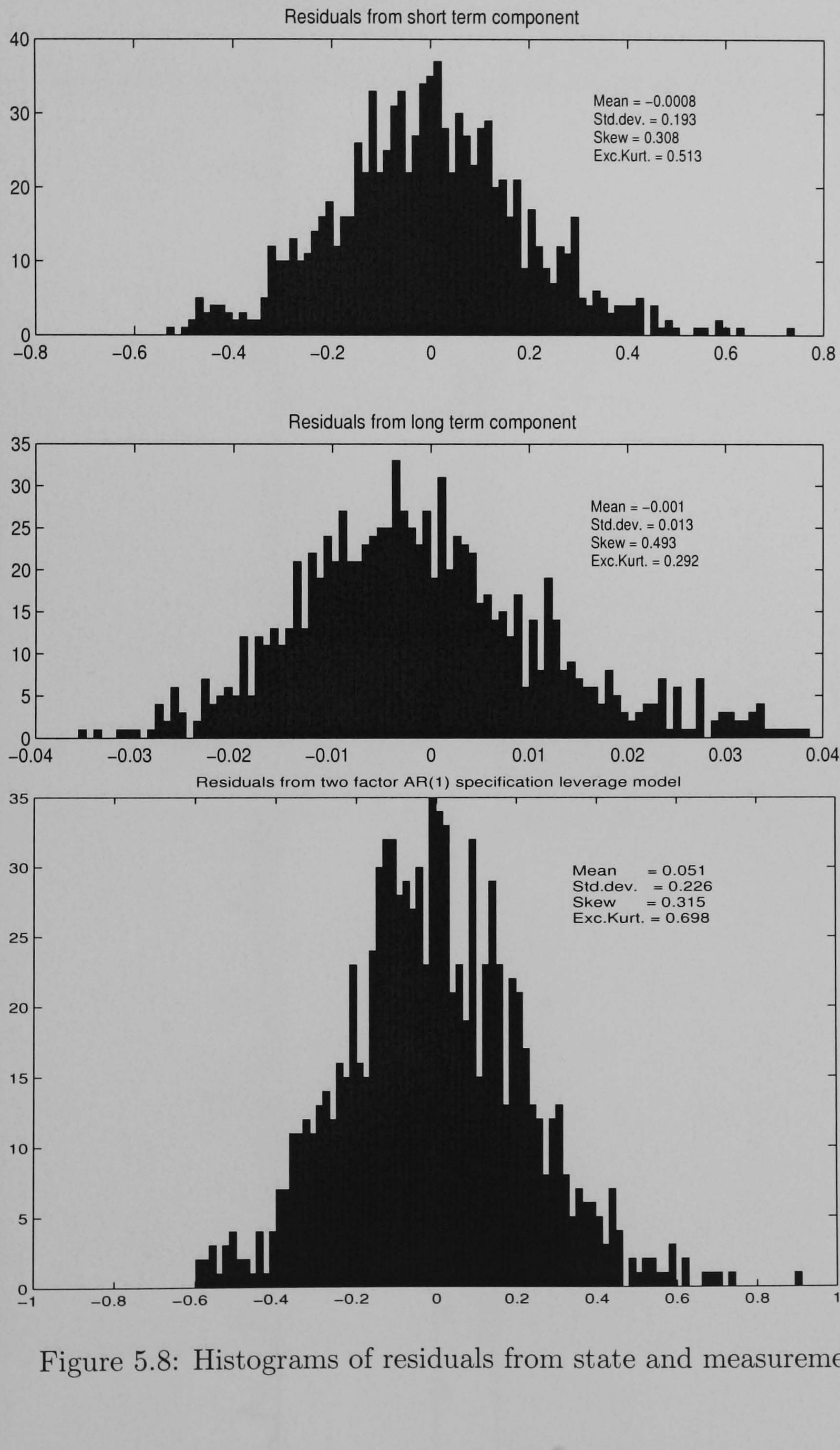
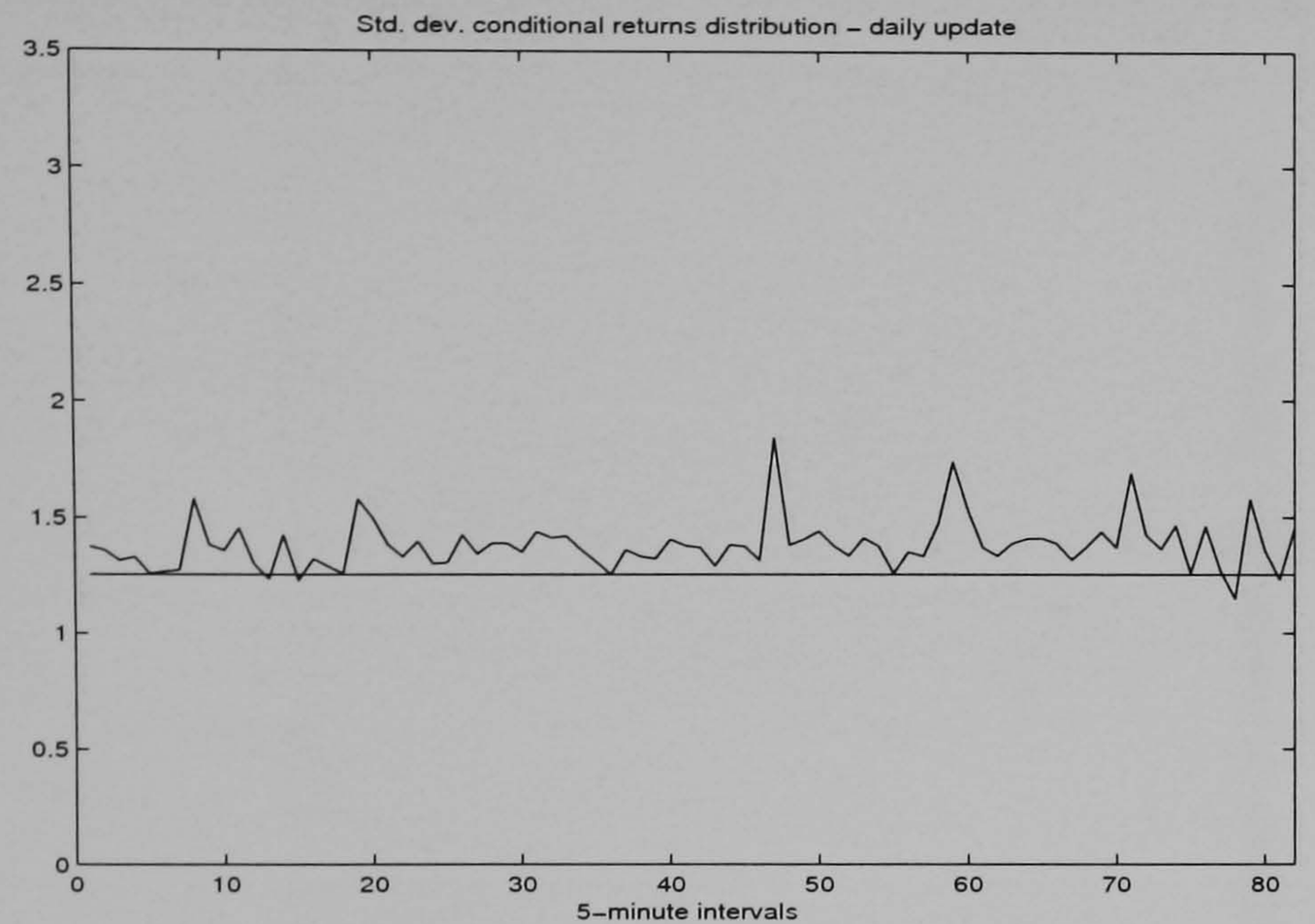
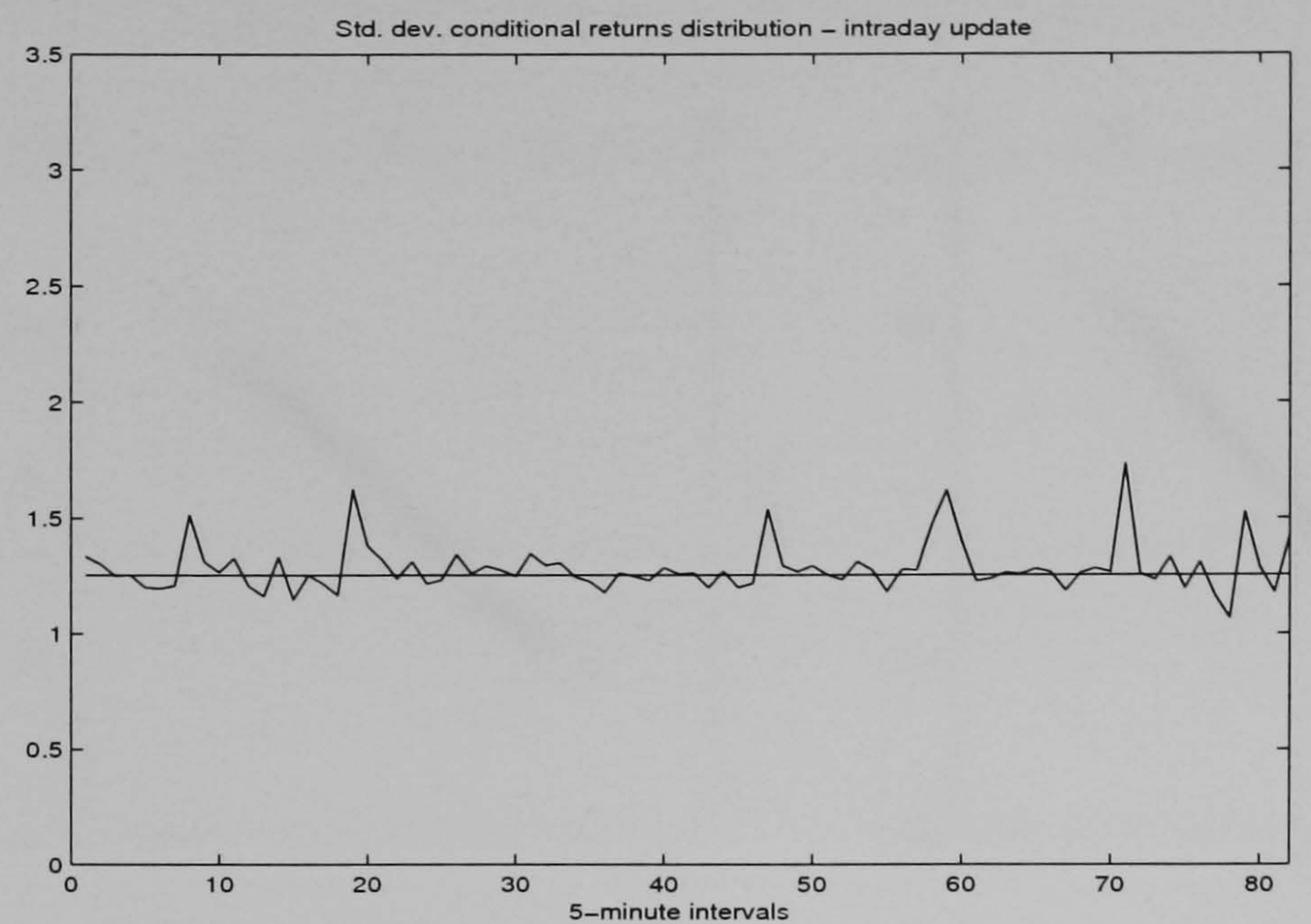
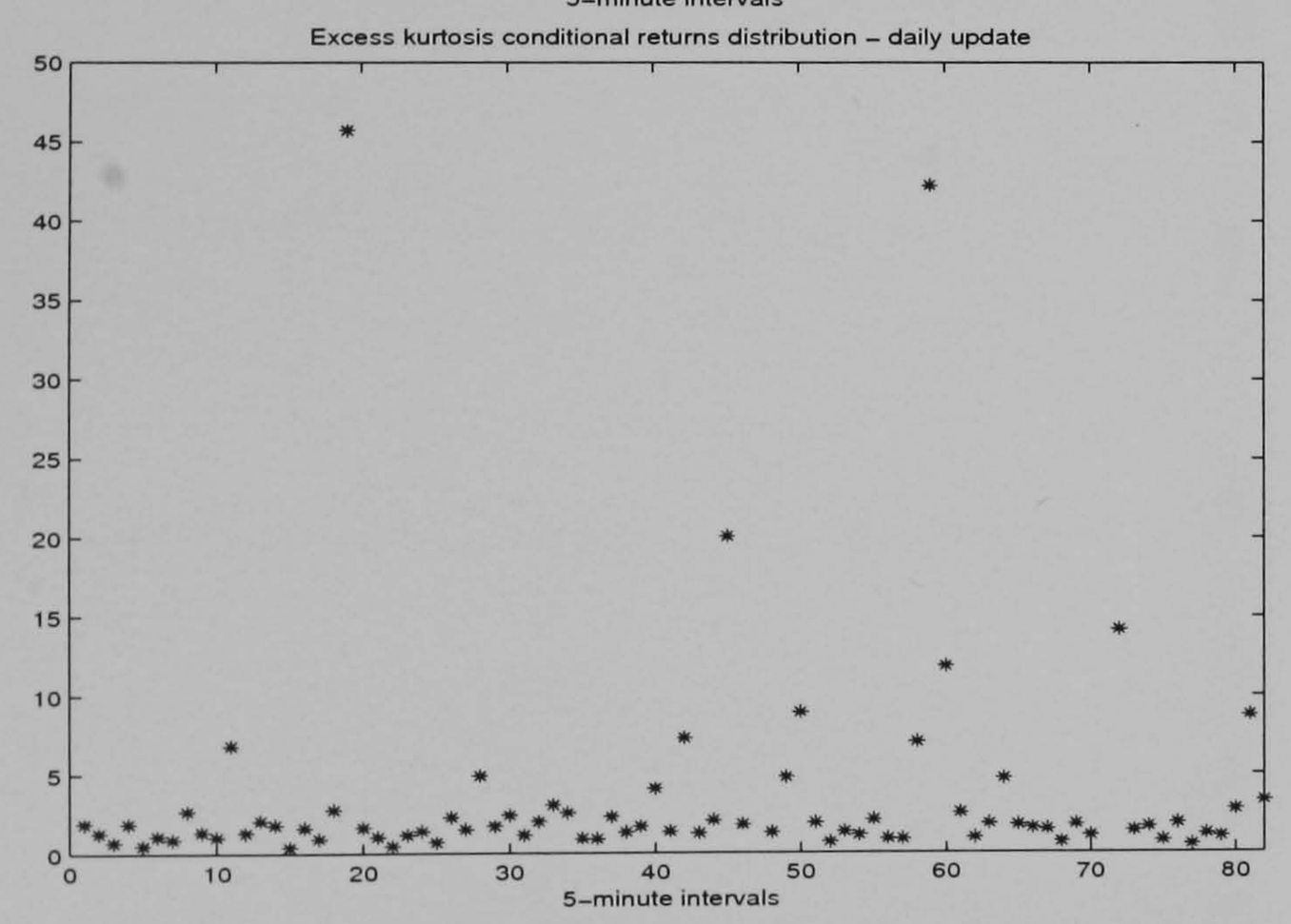


Figure 5.8: Histograms of residuals from state and measurement equations.







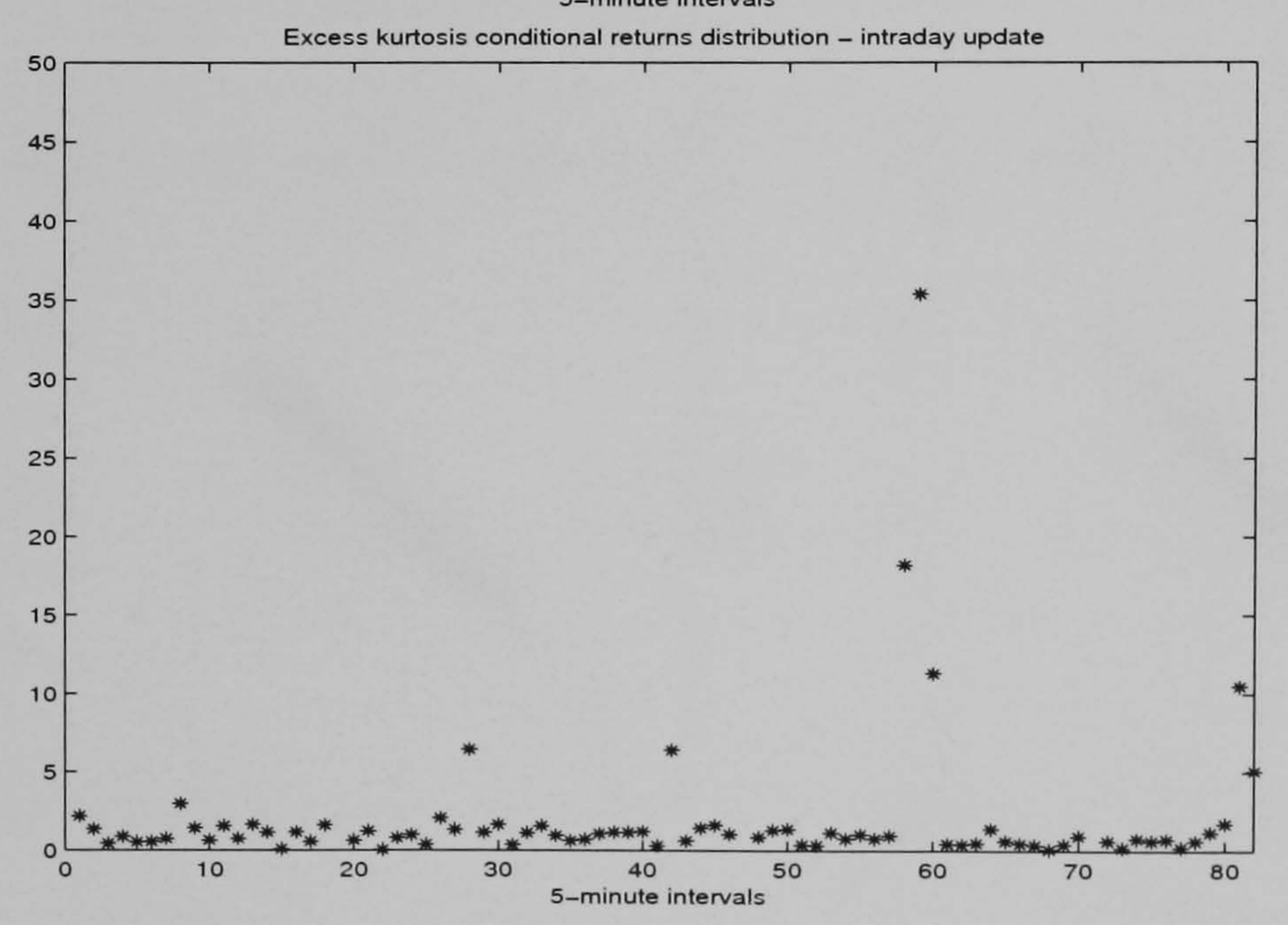


Figure 5.9: Standard deviation and excess kurtosis of conditional returns by intraday intervals using daily updated (left) and 5-minute updated (right) volatility estimates.

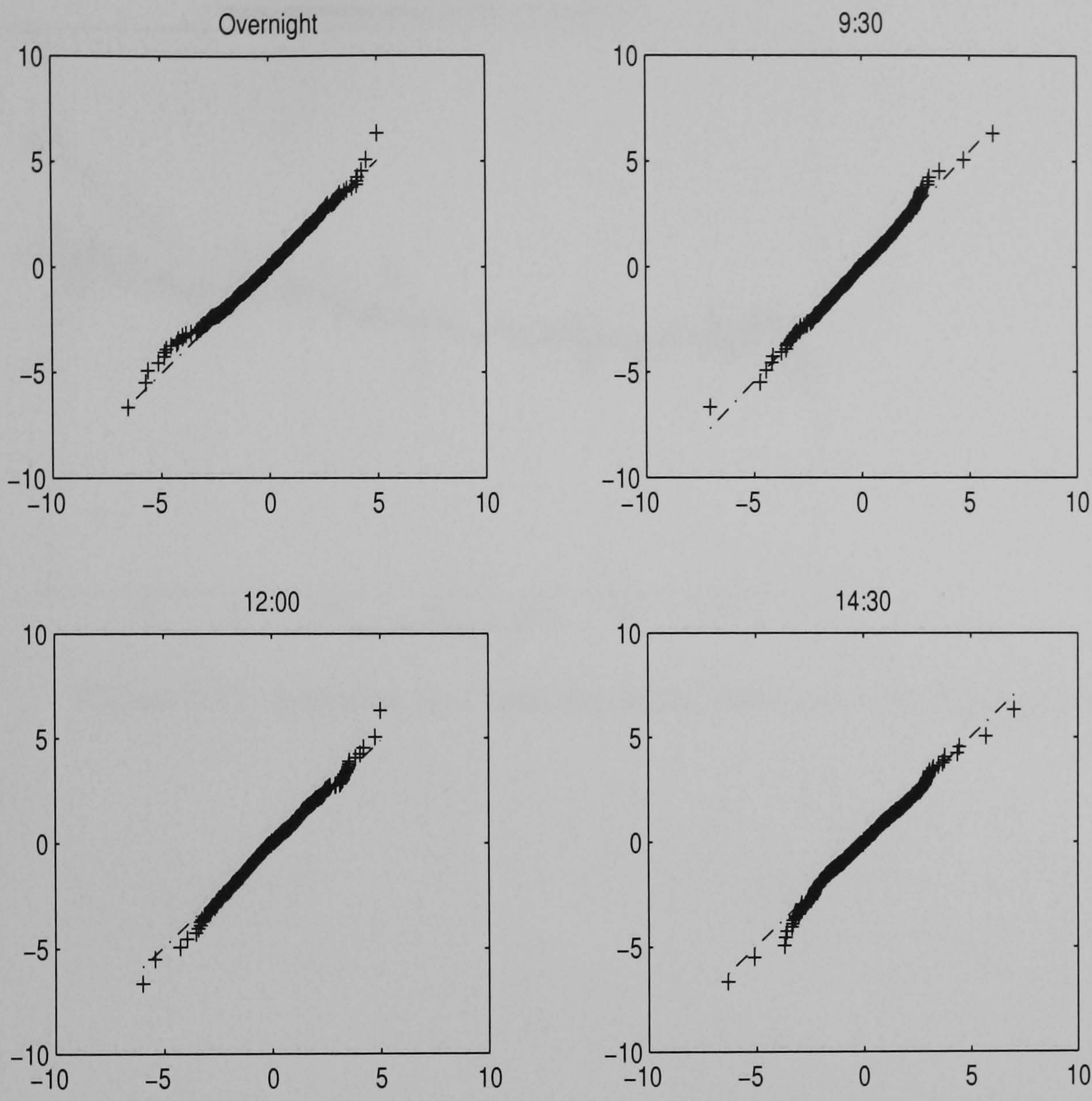


Figure 5.10: QQ plots of conditional returns for some intraday intervals.

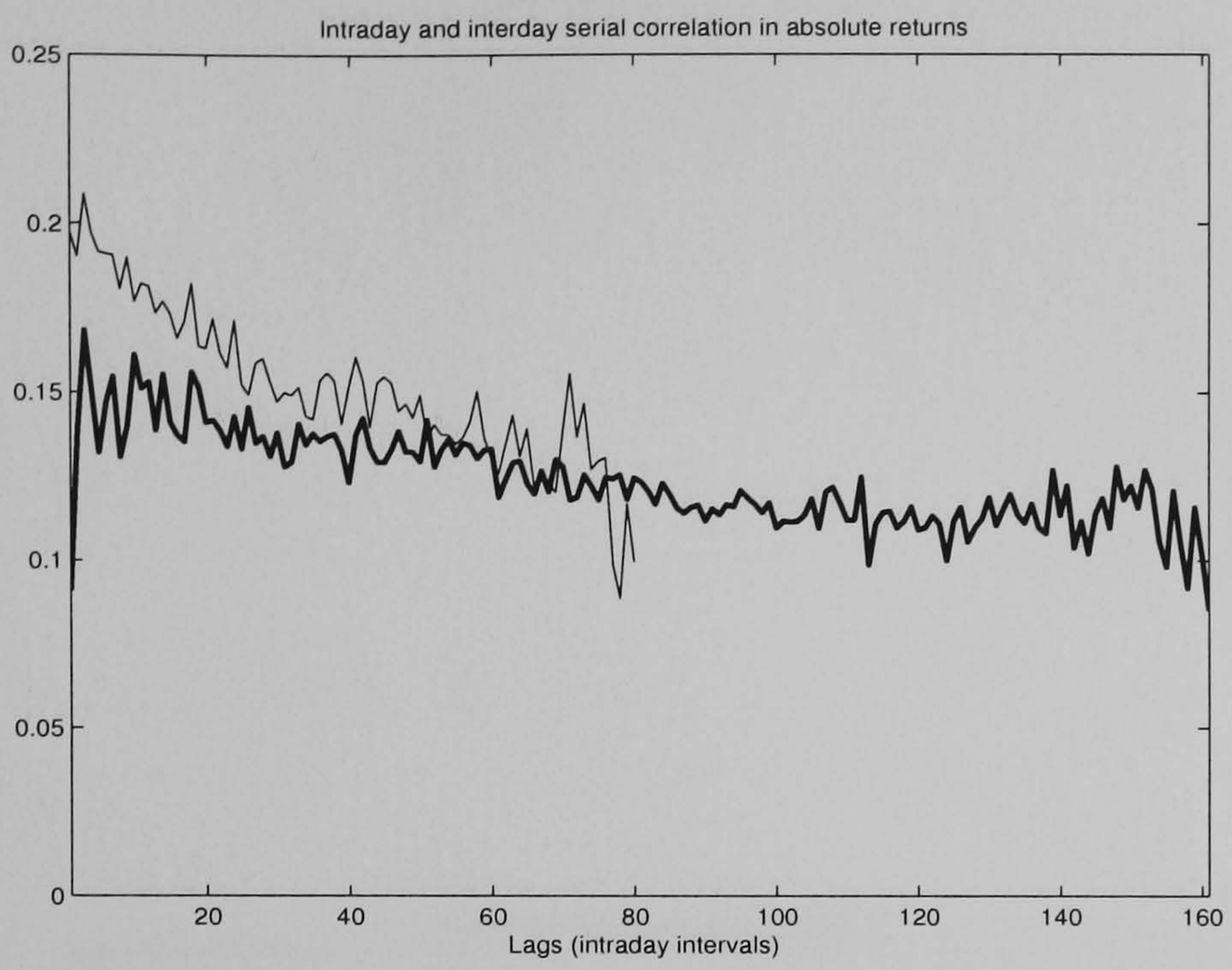


Figure 5.11: Intraday and interday serial correlation of absolute returns.

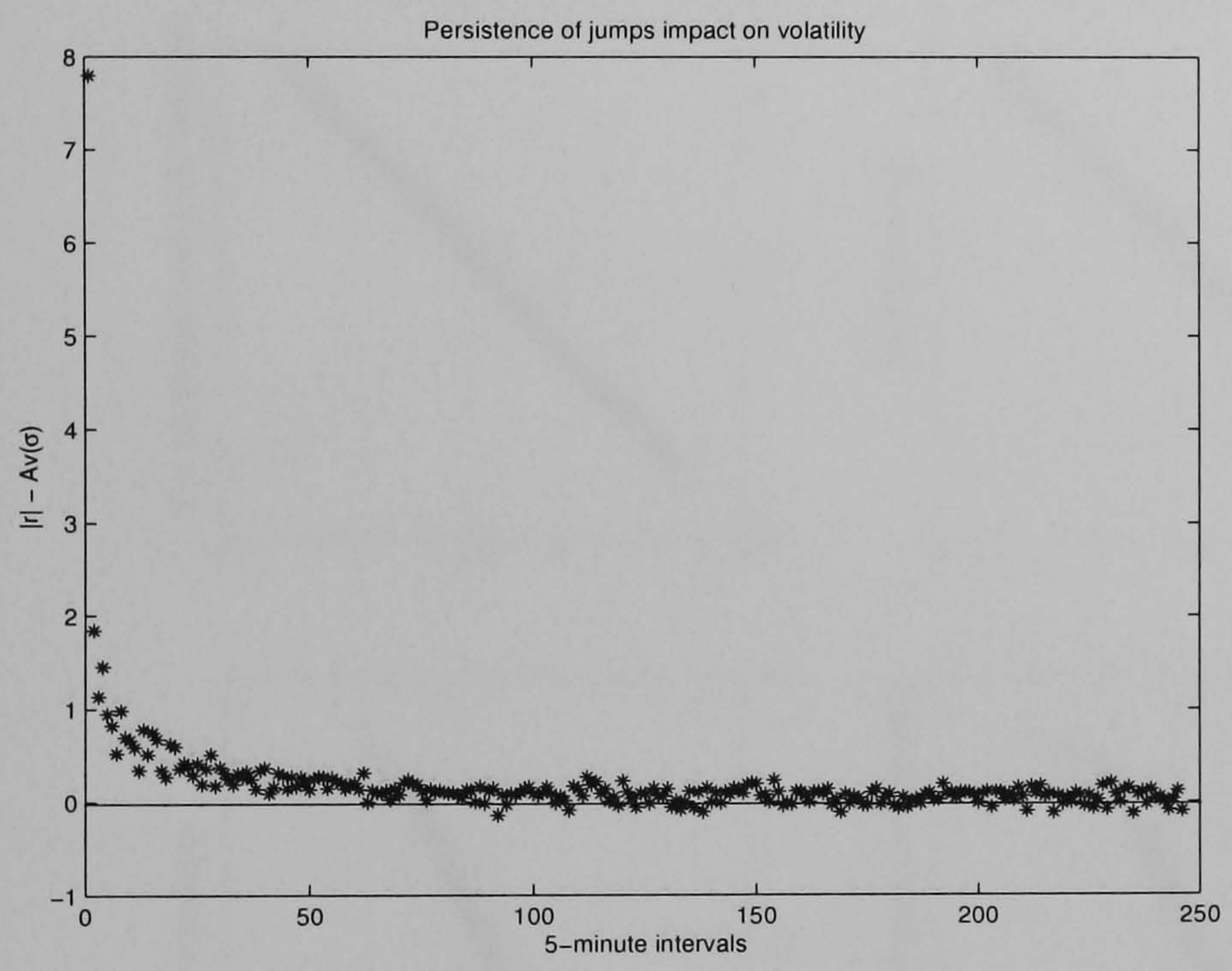


Figure 5.12: Persistence of jumps' impact on volatility measured in 5-minute intraday intervals.

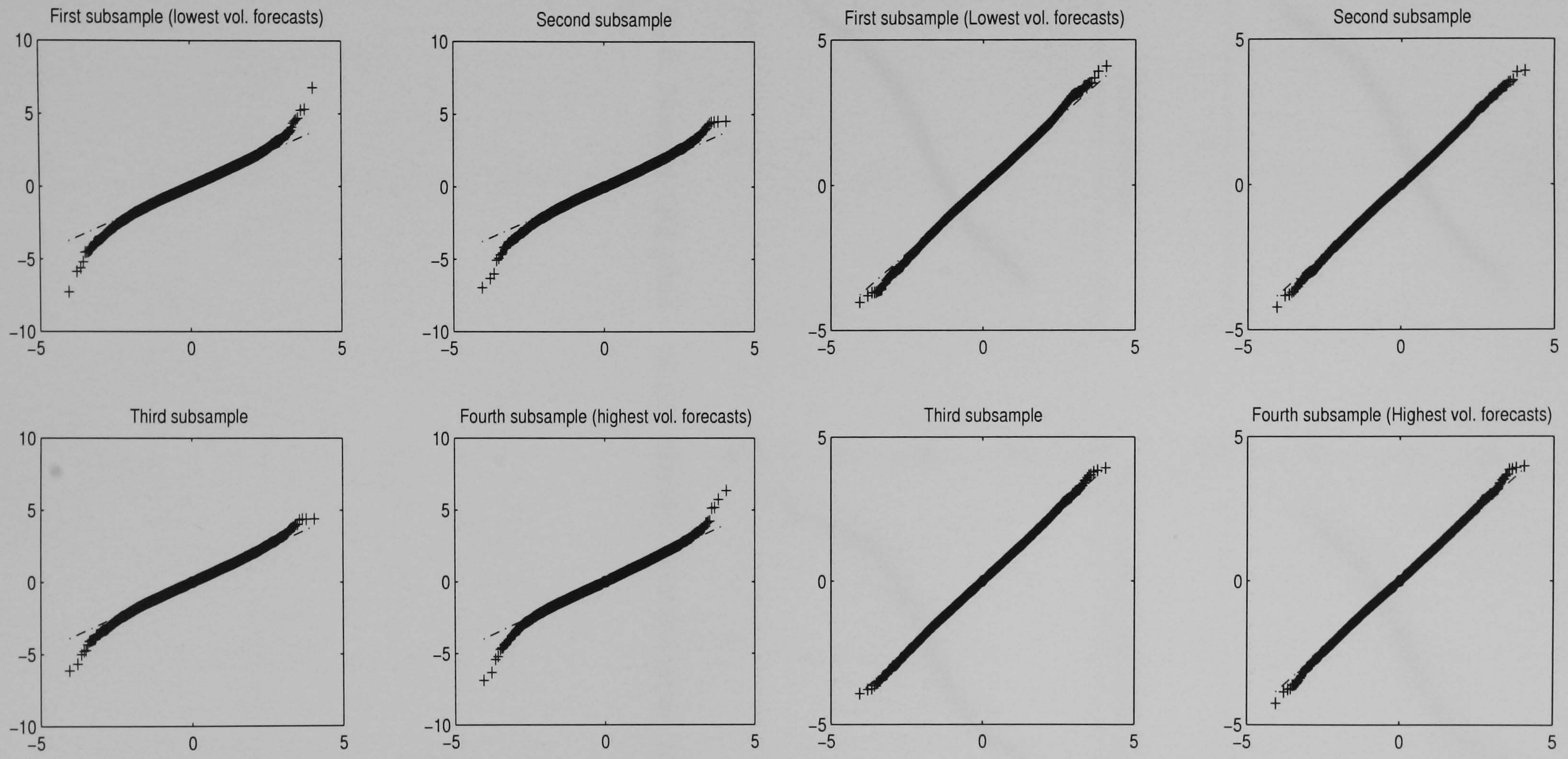


Figure 5.13: Normal QQ plots - return density forecasts using changing intraday volatility without (left) and with (right) jumps in volatility.

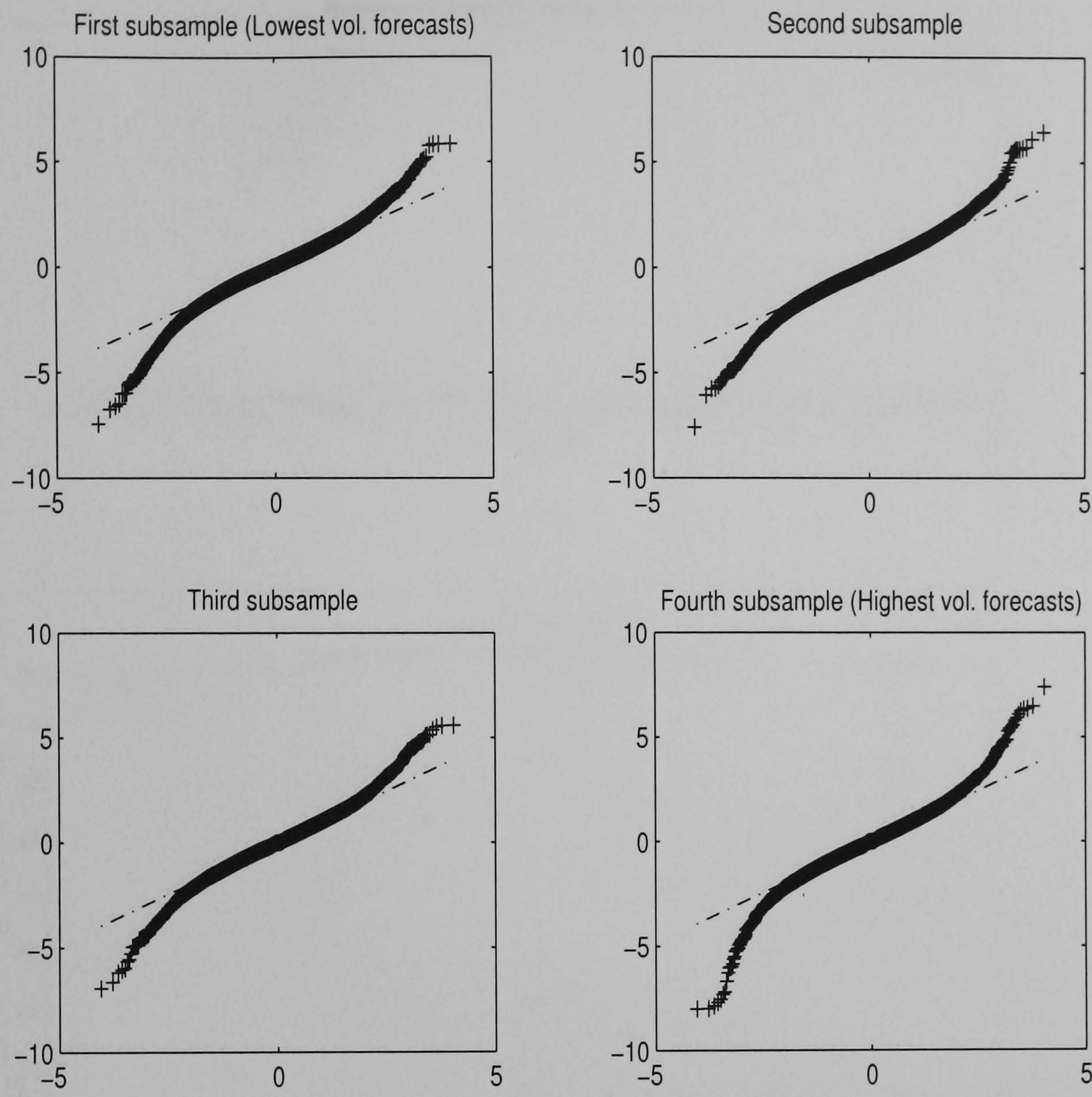


Figure 5.14: Normal QQ plots - return density forecasts using constant intraday volatility.

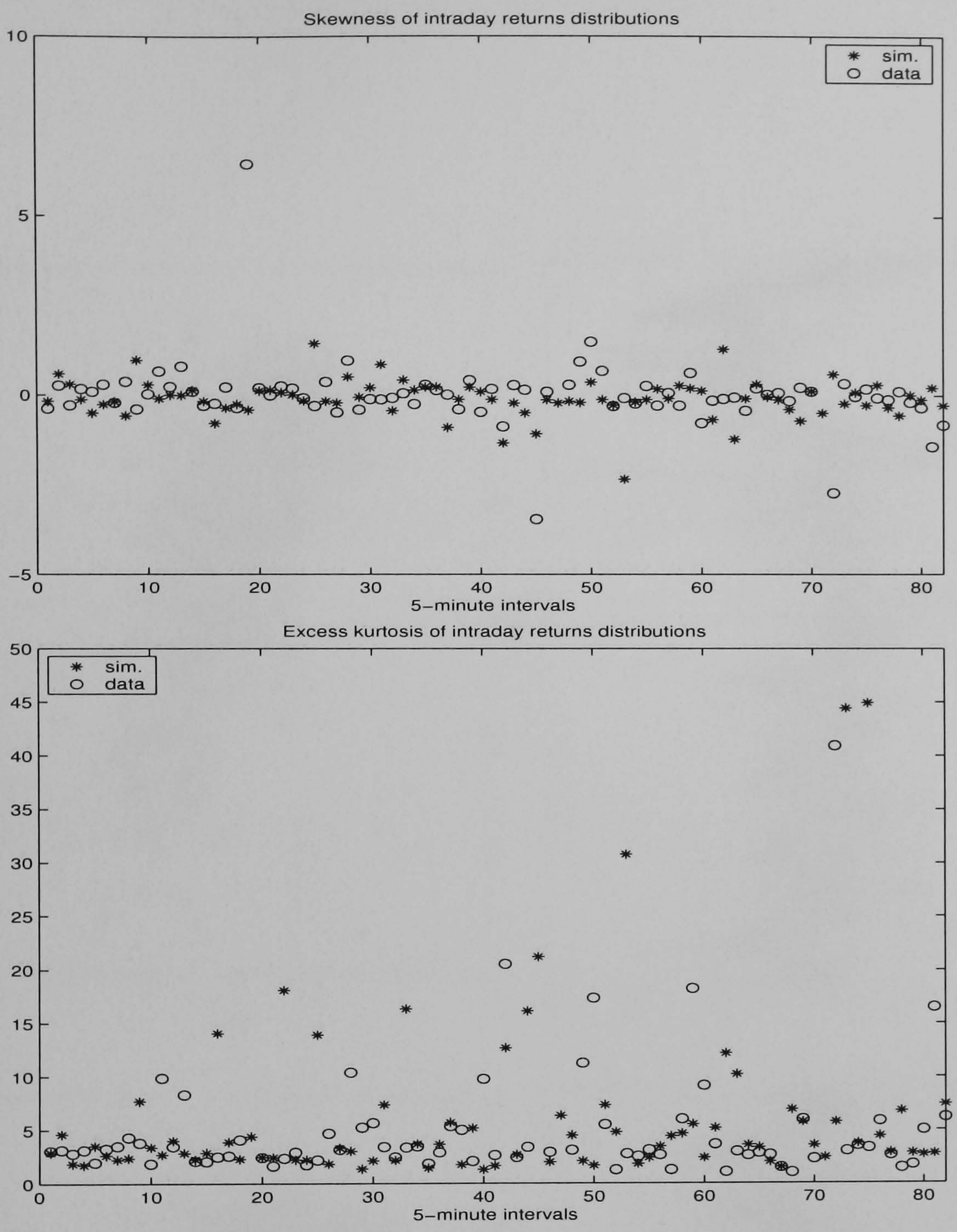
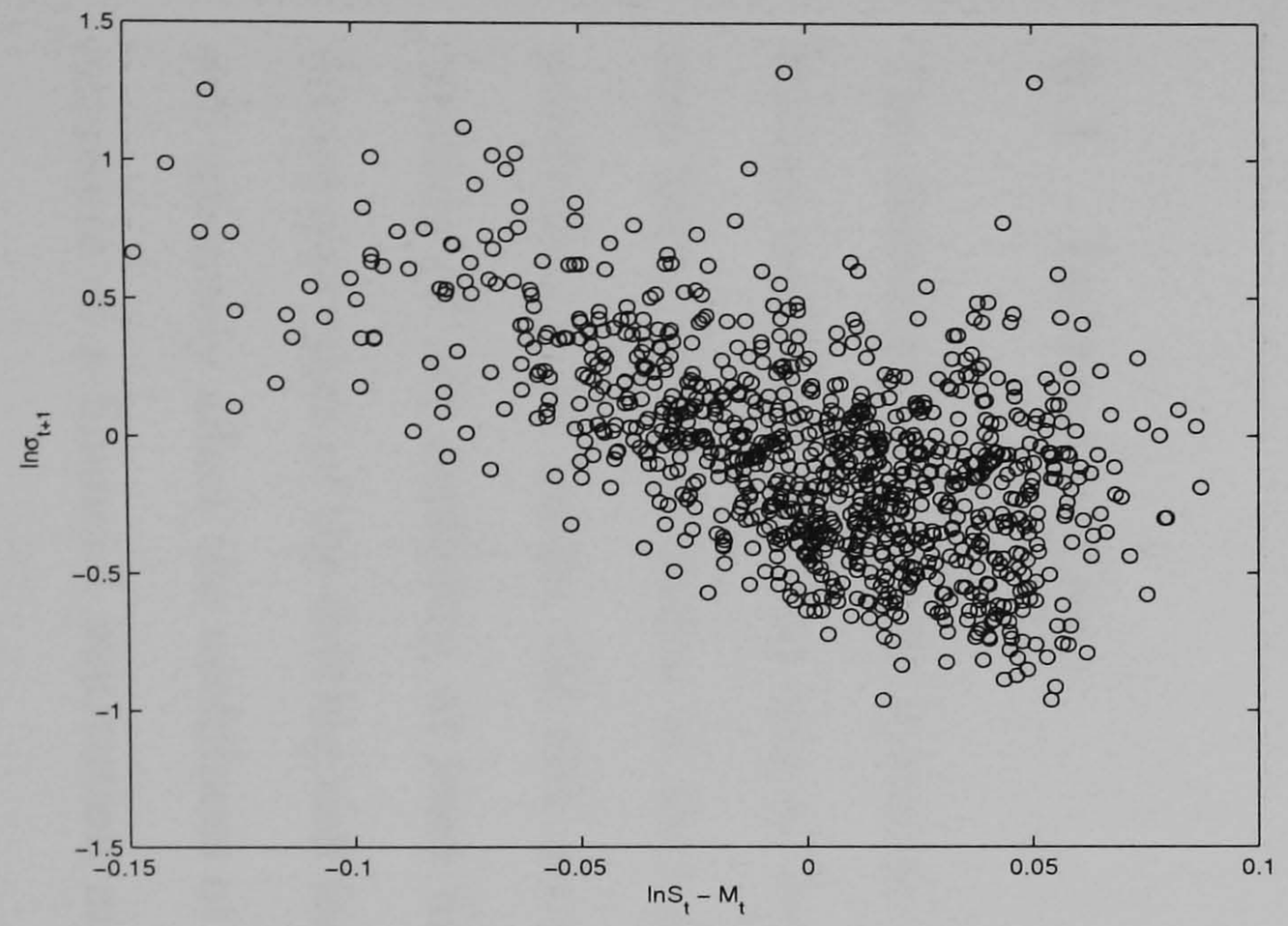
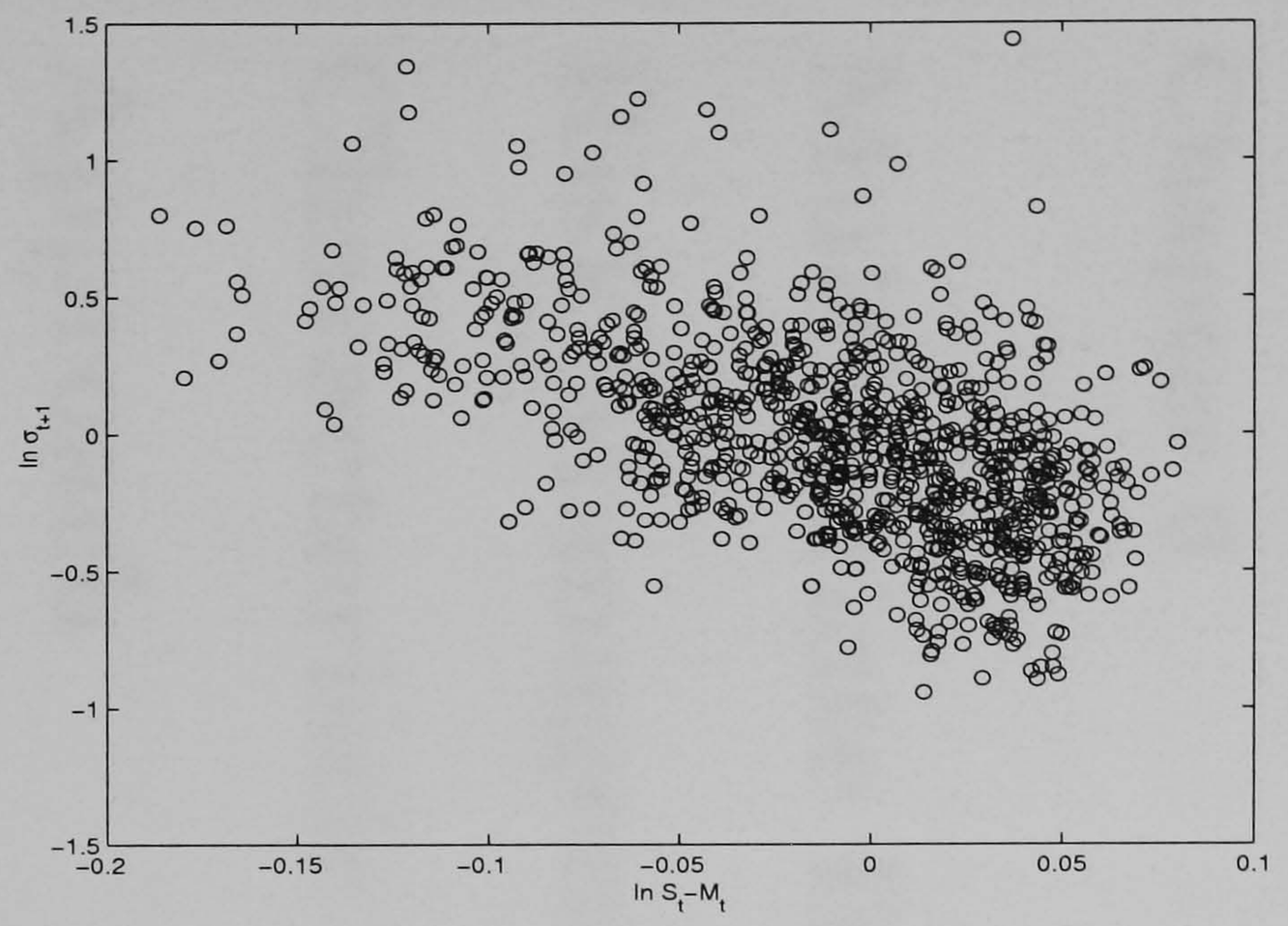
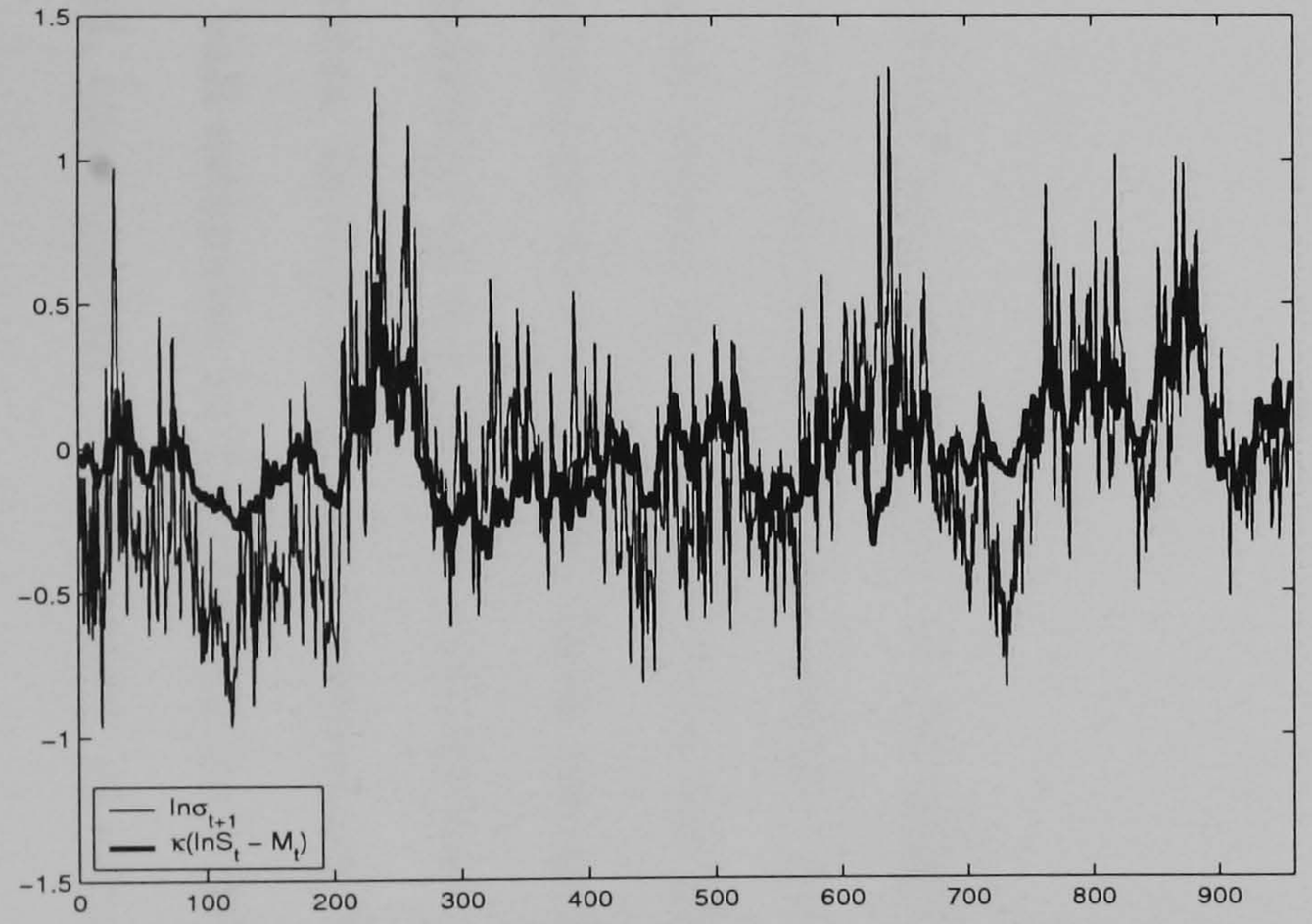


Figure 5.15: Skewness and excess kurtosis for intraday simulated and observed returns.







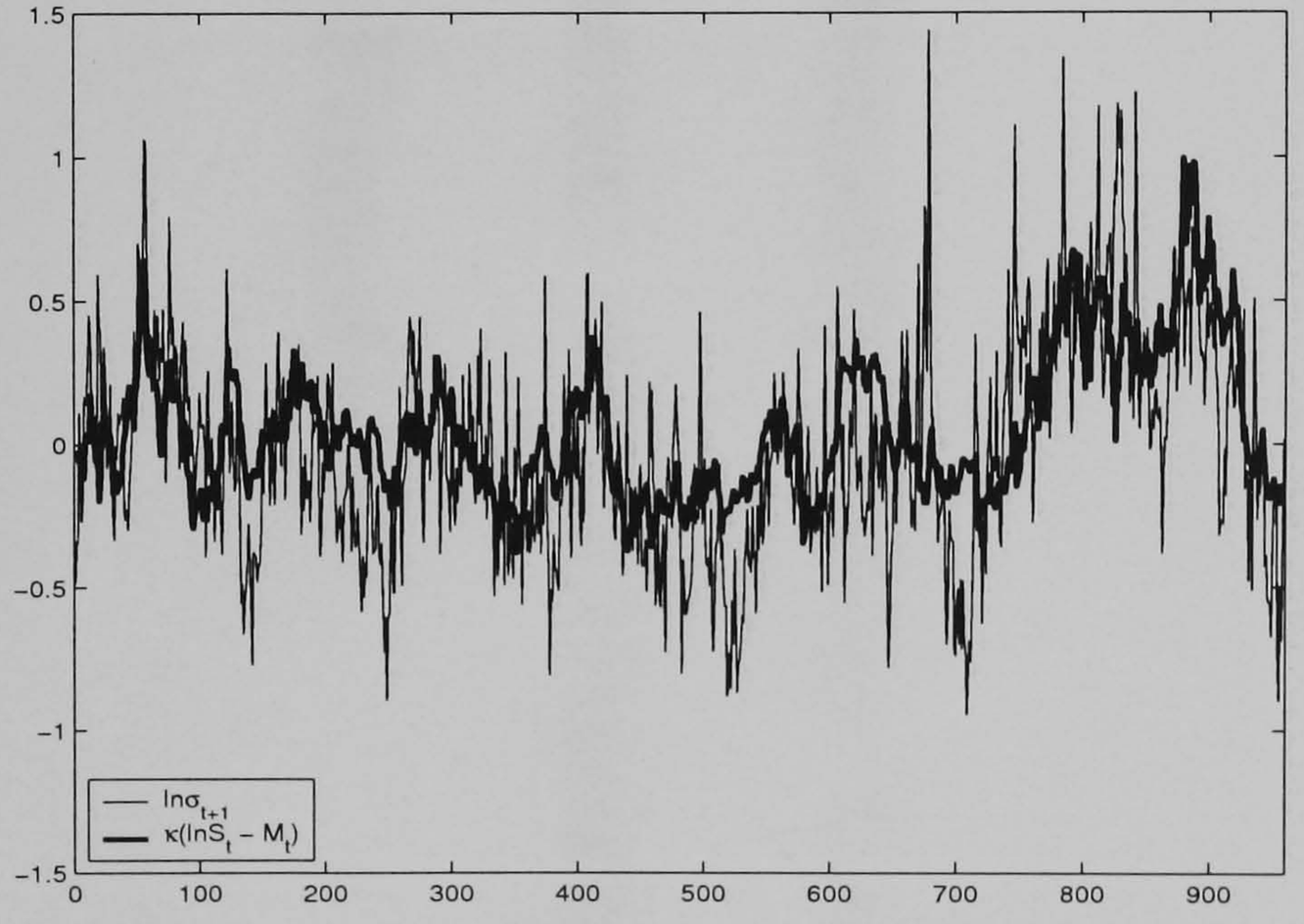


Figure 5.16: Scatter plot and time series of leverage measure against log volatility proxy - Market data (left) and simulated sample (right).

Chapter 6

Multivariate Density Tests for Risk Management: an Empirical Characteristic Function Approach

6.1 Introduction

The derivation of density forecasts of the joint distribution of the financial risk factors represents the first step in the computation of the Value-at-Risk, the second step being the translation of that joint density into a distribution for the bank portfolios. In all banks, the risk management then constantly faces the problem of creating (if not explicitly, at least implicitly) multivariate density forecasts of the future outcomes of the driving risk factors. In this context an inappropriate forecast will adversely affect the usefulness of VaR estimates for quantifying the market risk exposure of a financial institution and, consequently, the adequacy of the amount

of capital addressed to meet the capital requirements for market risk. Therefore, besides the creation of the distributional forecast itself, the *ex post* assessment of the "goodness" of such forecast is of crucial importance in risk management for providing directions in which the forecasting process could be improved.

However, little attention has been dedicated to the study of this second aspect, in particular for high-dimensional problems such as those that frequently occur in finance. The analysis generally becomes very complicated in a multidimensional context. The main issue here is the "curse of dimensionality": since a high-dimensional space is mostly empty, it is particularly difficult to implement accurate testing procedures that are able to pick up small deviations from the null hypothesis, unless the sample size is gigantic. Moreover, the concept of ordering used in many univariate goodness-of-fit tests is not obvious in a multidimensional space.

For all these reasons, many testing procedures commonly adopted in the univariate case do not easily generalise to the multivariate case, unless the dimensionality is very small.[1] In the univariate case, the common approach consists of applying a probability integral transformation to the density forecast, and testing the resulting probability transform for uniformity. In the multivariate case, deriving the probability integral transforms of the marginal distributions would not be sufficient, and all of the conditional densities should be considered in order to retrieve the dependence structure. Methods other than the ones adopted in the unidimensional framework of the previous chapters need therefore to be identified for an appropriate appraisal of multivariate density forecasts.

The present study constitutes an attempt in this direction, by investigating whether multivariate goodness-of-fit tests based on the empirical characteristic

[1]This is clearly not the case for VaR models, where the number of the relevant risk factors to be modelled could be around 200-300, and the sample size is not large enough to avoid the "curse of dimensionality" problem.

function (ECF) might represent good candidates for the assessment of multidimensional distributional forecasts. As we will see shortly, most of the literature on the ECF approach to multivariate hypotheses testing has focused on developing tests of specific distributional assumptions (i.e. normality, uniformity, symmetry), and deriving their theoretical properties. Very few papers have investigated more general goodness-of-fit procedures capable of spotting different sources of misspecification in the density forecasts. Even less attention has been dedicated to the practical application of these tests through sampling experiments, especially for a truly "high" number of dimensions. Here we implement both a continuous and a discrete version of multivariate ECF goodness-of-fit tests. Their relative performances in detecting deviations of various nature in the forecasted distribution from the actual distribution of the data, are compared in different dimensions, up to 16. The purpose of our analysis is to identify whether these ECF tests might be appropriate for an accurate, yet computationally feasible, assessment of multivariate density forecasts developed in the context of VaR models.

Not surprisingly, we find that the discrete version displays both a substantial loss in power and an increase in computational complexity, as the dimensionality of the problem grows. Instead, the continuous version generally exhibits satisfactory statistical properties and, therefore, seems to constitute a promising approach for these kind of evaluation problems.

For simplicity, in this work we do not deal with the estimation of multivariate density forecasts. This issue, crucial to risk managers, becomes very complicated in a high-dimensional context, where an enormous amount of data is needed, unless we resort to fully parametrised specifications.

The chapter is structured as follows. Section 6.2 briefly reviews the existing relevant literature. Section 6.3 outlines features and properties of the two testing procedures based on the ECF approach contrasted in the present work. In Section

6.4 we provide and discuss the outcomes from some Monte Carlo experiments carried out to evaluate and compare the relative statistical performance of the two tests. Section 6.5 concludes the chapter.

6.2 A Review of the Related Literature

The existing literature on the evaluation of density forecasts, which has been reviewed in Chapter 3, is rather recent and has focused mainly on univariate distributions. Much less research has been devoted to the assessment of multivariate density forecasts. As stated earlier, the univariate distributional testing procedures cannot usually be extended to the multivariate case.

Most studies in the multidimensional framework have devised test statistics of specific distributional hypotheses such as normality, uniformity, symmetry.[2] Only few contributions have concentrated on more general goodness-of-fit testing procedures. Some attempts to develop multivariate versions of Kolmogorov-Smirnov and Cramér-von Mises statistics were made by, for example, Kotz and Johnson [1985], Cabaña and Cabaña [1994], Justel, Peña and Zamar [1997]. Multivariate goodness-of-fit statistics based on generalised L^p discrepancies (i.e. measures of distance in the L^p space) were recently suggested by Hickernell [1999], and Liang, Fang and Hickernell [2001]. The most popular approach to multivariate density forecast evaluation is a multidimensional application of the probability integral transform method, which consists of a decomposition of multivariate forecasts into univariate conditionals. Amongst the most relevant examples, see Diebold, Hahn and Tay [2000], and Clements and Smith [2000, 2002]. Unfortunately, all these test statistics present serious drawbacks when the dimensionality of the problem increases: the computational burden becomes unsustainable and/or the properties of the test statistics

[2]For a review of these techniques, see Bilodeau and Brenner [1999].

themselves are unknown. Such problems rule out the applicability of these tests in a risk management context, where a large number of risk factors is involved.

The difficulty in the practical implementation for truly high-dimensional problems also constitutes the main drawback of a recent, promising approach aimed at modelling and evaluating the dependence structure between variables, based on the use of copula functions.[3] The attractive feature of this approach, that has contributed to its increasing popularity amongst the researchers, is the possibility to model more general concepts of dependence than the basic linear correlation, which is an adequate measure of dependence only for distributions belonging to the elliptical family. In these studies, copula functions of different kinds are fitted either to the entire multivariate data sample, or to the extreme values only, when the analysis is carried on for risk management purposes. The choice between the various copulas is normally based on the values of the Akaike's information criterion, as well as on the results from the multidimensional probability integral transform test. Relevant contributions in this area include Klugman and Parsa [1999], Embrechts, McNeil and Straumann [2002], Embrechts, Lindskog and McNeil [2003], Embrechts, Hoeing and Juri [2003]. Unfortunately, since the formulations commonly proposed for the copula functions have very few parameters, they do not seem suitable for describing satisfactorily the complicated dependence structure amongst a high number of risk factors.

Projection-pursuit type testing techniques (Huber [1985]) have also been suggested as multivariate goodness-of-fit tests. The underlying idea is to re-map multivariate data set into a unidimensional framework along interesting low-dimensional linear projections (see Zhu, Fang and Zhang [1995] and Zhu, Fang and Bhatti [1997], for some applications). The disadvantage of this method is the large number of projections that have to be generated to guarantee good statistical power.

[3]For a general survey on copula functions, see Nelsen [1998].

An alternative approach for assessing multidimensional distributional forecasts involves the implementation of multivariate goodness-of-fit tests based on the empirical characteristic function (ECF). The area of statistical inference where the ECF has been more extensively used is parameter estimation. Examples of employment of this technique in univariate hypotheses testing can be found, among others, in Heathcote [1972], Feigin and Heathcote [1976], Feuerverger and Mureika [1977], Epps and Pulley [1983], Feuerverger [1993]. Extensions to the multivariate framework include the tests for normality of Baringhaus and Henze [1988], Henze and Zirkler [1990], Henze and Wagner [1997], the tests for independence of Csörgő [1985] and Bilodeau and Lafaye de Micheaux [2002], the test of affine equivalence for elliptically symmetric distributions by Gupta, Henze and Klar [2003]. A more general goodness-of-fit test for both simple and composite hypotheses was proposed by Fan [1997].

The use of the ECF in statistical inference is justified by the asymptotic convergence of the ECF to the theoretical one, and the one-to-one correspondence between distribution and characteristic function of a variable (Heathcote [1972]). Employing the characteristic function in testing procedures has important advantages over the use of the distribution function. Since the former is the Fourier transformation of the latter, it allows the retention of all the information contained in the sample, while presenting a more convenient formulation for computation, as it behaves friendly under shifts, scale changes, and summation of independent variables (Feuerverger and Mureika [1977]). The asymptotic properties of the ECF have been largely investigated by Feuerverger and Mureika [1977], Feuerverger and McDunnough [1981a, b], Csörgő [1981].

Goodness-of-fit tests based on the ECF are usually computed from the integrated squared difference between the ECF $C_n(t)$ of the data and the characteristic function $C_0(t)$ of the multivariate null distribution, evaluated:

- at a single fixed value of t (Heathcote [1972]). In this case the advantage of a simplified analysis is generally offset by the lack of statistical consistency of the testing procedure;

- at a set of values $t_1, t_2, \ldots, t_m$. This approach was adopted by Fan [1997] to develop a general multivariate goodness-of-fit test;

- over a continuous range of t described by a weight function. Epps and Pulley [1983] suggested this procedure to test for univariate normality. The multivariate case was investigated by Baringhaus and Henze [1988], Henze and Zirkler [1990], Henze and Wagner [1997].

In the present work we implement and investigate testing procedures belonging to the second and the third groups.

6.3 The Empirical Characteristic Function Goodness-of-Fit Techniques

In this section we describe the features of the two multivariate goodness-of-fit testing procedures, based on the ECF, that are contrasted in the present study.

6.3.1 The continuous "Weight Function" (WF) approach

The first of the tests we consider is based on the integrated squared difference between the empirical $C_n(t)$ and the theoretical $C_0(t)$ characteristic functions, weighted by the function $G(t)$, which is itself a distribution function, whose corresponding density $g(t)$ is symmetric around the origin:

$$D_{n,\beta} = \int |C_n(t) - C_0(t)|^2 \, \mathrm{d}G(t) \tag{6.1}$$

where $\mathrm{d}G(t) = g(t)\,\mathrm{d}t$. Let $X_1, X_2, \ldots, X_n$ represent a sample of independent observations from a random d-dimensional vector, with sample mean $\overline{X}$ and sample

covariance matrix S. The ECF $C_n(t)$ is defined as follows:

$$C_n(t) = \frac{1}{n}\sum_{j=1}^{n}\exp(it'X_j) = \left[\frac{1}{n}\sum_{j=1}^{n}\cos(t'X_j)\right] + i\left[\frac{1}{n}\sum_{j=1}^{n}\sin(t'X_j)\right] \tag{6.2}$$

and the theoretical characteristic function $C_0(t)$ as:

$$C_0(t) = E[\exp(it'X)] \tag{6.3}$$

This approach was originally suggested by Baringhaus and Henze [1988] to test for multivariate normality. This test statistic presents some appealing properties of consistency and good power against any fixed alternative to the null distribution. However, it cannot be easily extended to test any kind of multivariate null distribution, as that would require the evaluation of a d-dimensional integral.

The main issue here regards the choice of the weight function. According to Epps and Pulley [1983], the following requirements should be met:

- high weight must be assigned where the distance between the characteristic functions of the alternative and of the null hypotheses is larger. For many continuous distributions, this applies for values of t between 0 and 3;

- more weight should be put in some interval around the origin, where most information conveyed by the characteristic function is contained, and the precision of the sample characteristic function is greater. In particular, the length of the interval should be inversely related to the dispersion of the data;

- the weight function must present a convenient formulation, such that the integral in Eq. (6.1) has a closed form, to ensure the computational feasibility of the test statistic. This point represents the main limitation to extend the application of this approach to a wide range of distributions.

In our analysis, we follow the common practice of selecting as weight function

the density of a multivariate normal distribution $N_d(0, \beta^2 I_d)$:

$$g(t) = (2\pi\beta^2)^{-d/2} \exp\left(-\frac{\| t \|^2}{2\beta^2}\right), \qquad t \in \mathbf{R}^d \tag{6.4}$$

that, for an appropriate choice of β, satisfies the above-mentioned requirements. This choice is particularly attractive, for both its statistical properties and its high tractability, but suitable for few specification forms for the null distribution, close to the multivariate normal. In the present work this is not a problem, since we restrict the analysis precisely to distributional forecasts of this type, i.e. multivariate normal and mixtures of multivariate normal distributions.

As for the first specification, we are aware of the fact that the joint distribution of financial returns clearly deviates from multivariate normality. However, it is a matter of fact that in practice many forecasts are still based only on estimates of the mean vector and the covariance matrix of asset returns and, hence, are specified as a multivariate Gaussian.[4] Besides that, some studies (see Sornette, Simonetti and Andersen [1999]) indicate that most of the plausible specifications chosen to describe the distribution of financial asset returns, support a transformation to normality that maintains the fat-tailed nature of the marginal distributions, and the existence of nonlinear dependence between the assets. Therefore, the assumption of multivariate normality can still represent a sensible choice for the distribution of those transformed variables.

The second specification represents a more plausible assumption for the multivariate distribution of financial asset returns. Such an assumption implies that the each single variable follows itself a mixture of normal distributions. This functional form has found vast support in the relevant literature (see, amongst the others, Ball and Torous [1983], Kon [1984], Hull and White [1998], Zangari [1996], Venkataraman [1997]) as an appropriate distribution for financial asset returns. First, unlike

[4]The well known J.P.Morgan RiskMetrics approach constitutes a relevant example in this respect.

the Gaussian density, it can account for both the asymmetry and the fat-tailness exhibited by financial variables. Second, it provides a more immediate interpretation than alternative specifications, given that the various components of the mixture can be easily related to different states of the market (i.e. high and low volatility markets).

The test statistic is given by:

$$T_{n,\beta}\left(X_1,\ldots,X_n\right) = n\left(4\,I\{S\text{singular}\} + D_{n,\beta}\,I\{S\text{nonsingular}\}\right) \tag{6.5}$$

For the purposes of our study, an exact expression for $D_{n,\beta}$ is derived under the null of 1) multivariate normality on the original variables; 2) multivariate normality on standardised variables; 3) mixture of multivariate normal distributions, on standardised variables. For clarity of exposition, the formulas have been included in Appendix B. The empirical critical values for the tests, computed at 95% confidence level, are reported in Table 6.1.

6.3.2 The discrete "Fixed Grid Points" (FGP) approach

The second testing procedure we adopt is a simplified version of the goodness-of-fit procedure suggested by Fan [1997]. The test is expressed as a quadratic measure of the difference between real and imaginary components of the ECF and of the characteristic function under the null distribution, evaluated at m fixed grid points $t_1, t_2, \ldots, t_m$ (where m is a positive integer and $\to \infty$ as $n \to \infty$ to ensure consistency of the test).

As before, $X_1, X_2, \ldots, X_n$ represents a sample of d-dimensional independent observations. The test statistic introduced by Fan [1997] has the following formulation:

$$T_n = [Z_n(t_m) - Z_0(t_m)]'\Omega_0^{-1/2}(t_m)W(t_m)\Omega_0^{-1/2}(t_m)[Z_n(t_m) - Z_0(t_m)]$$

where $Z_n(t_m)$ is the vector of real and imaginary components of the ECF, and $Z_0(t_m)$ is the corresponding vector for the hypothetical characteristic function:

$$Z_n(t_m) = \begin{bmatrix} \mathrm{Re}C_n(t_1) \\ \mathrm{Re}C_n(t_2) \\ \vdots \\ \mathrm{Re}C_n(t_m) \\ \mathrm{Im}C_n(t_1) \\ \mathrm{Im}C_n(t_2) \\ \vdots \\ \mathrm{Im}C_n(t_m) \end{bmatrix} \qquad Z_0(t_m) = \begin{bmatrix} \mathrm{Re}C_0(t_1) \\ \mathrm{Re}C_0(t_2) \\ \vdots \\ \mathrm{Re}C_0(t_m) \\ \mathrm{Im}C_0(t_1) \\ \mathrm{Im}C_0(t_2) \\ \vdots \\ \mathrm{Im}C_0(t_m) \end{bmatrix}$$

$$Z_n(t_m) - Z_0(t_m) = \frac{1}{n}\sum_{j=1}^{n} \begin{bmatrix} \cos(t_1' X_j) - E[\cos(t_1' X_j)] \\ \vdots \\ \cos(t_m' X_j) - E[\cos(t_m' X_j)] \\ \sin(t_1' X_j) - E[\sin(t_1' X_j)] \\ \vdots \\ \sin(t_m' X_j) - E[\sin(t_m' X_j)] \end{bmatrix} \equiv \frac{1}{n}\sum_{j=1}^{n} K_j(t_m)$$

$\Omega_0(t_m) = Var\left[K_j(t_m)\right]$ represents the $(2m \times 2m)$ covariance matrix of $K_j(t_m)$, whose components are computed as specified in Fan's paper. W is a diagonal weight matrix aimed at directing the power of the test statistic towards different frequencies. In our analysis, we assume $W = I_{2m}$, which assures both computational convenience and more powerful tests against fat-tailed alternatives. The formula of the test statistic then simplifies to:

$$T_n = [Z_n(t_m) - Z_0(t_m)]' \Omega_0^{-1}(t_m)[Z_n(t_m) - Z_0(t_m)] \tag{6.6}$$

In the general case, the asymptotic distribution of the test statistic under the null is a weighted sum of independent χ^2 with one degree of freedom. Given our

specific choice for the matrix W, the distribution of the test in Eq. (6.6) simplifies to a χ^2 with $2m$ degrees of freedom.

In principle, the discrete ECF approach can be widely used for testing the goodness of any type of distributional forecast in any dimension, under both simple and composite null hypotheses, provided that consistent estimators of the unknown parameters are given. As it will be shown later on in the analysis, the statistical properties of this testing procedure are strongly affected by the choice of the grid points. Therefore, the main issues in this context regard the criteria that drive the choice of m and, consequently, of the different t, and their location.

Since the WF approach can be interpreted as the continuous version of the FGP one, it presents some obvious advantages. First of all, by using a continuous weight function, all the moments of the empirical and the theoretical characteristic functions are matched continuously and, hence, more information from the sample is exploited. Secondly, in the continuous case there is no need to worry about the selection of the different t, as they are simply integrated out, provided that the weight function is chosen appropriately.

Intuitively, as the dimensionality of the problem increases, we do not expect the discrete version of the ECF test to perform well, given that the choice of the grid points becomes both extremely important and very difficult in such context. However, we believe that the investigation of its statistical size and power is useful in order to understand for which dimensionality the discrete version of the test can still be employed.

6.4 Size and Power of the ECF Tests: Some Monte Carlo Experiments

6.4.1 General framework

The main purpose of the present analysis is to assess whether the ECF testing procedures outlined in the previous section possess a satisfactory power to detect deviations of the actual distribution of the data from the forecasted one, while maintaining a good level of tractability in a high-dimensional context. Since, for finite samples, theoretical results regarding the relative power of these two classes of tests are not available, we provide and discuss the evidence obtained from some Monte Carlo simulations. We investigate how the two testing procedures compare, in terms of power, to each other (for the purpose of ranking between them) and, when appropriate, to alternative techniques whose good statistical properties are widely recognised (for the purpose of a more general appraisal of the performance of the ECF tests).

In the attempt of identifying a nested sequence of tests for the density forecasts, that would enable us to diagnose the nature of the misspecification, controlled experiments are designed to test separately for biases of various type in the forecast. Samples are generated according to a distribution different, in some aspects, from the forecasted one. Various choices for the possible specifications for the densities under the null and the alternative hypotheses are contemplated, according to the specific nature of the bias in the forecast we are interested in detecting. More specifically, we test for the presence of biases: 1) in the mean; 2) in the covariance structure; 3) in higher moments when the forecasted distribution is a multinormal, which then provides a test for misspecification of the form of the distribution; 4) in higher moments when the forecasted distribution is fat-tailed.

The power of the two testing procedures is expressed as percentage of rejections out of $5,000$ replications. In order to investigate how the performance of the tests changes with the dimensionality, we consider problems in dimensions $d = 2, 4$ and 16. The sample size is $n = 100$ for $d = 2, 4$, and $n = 200$ for $d = 16$. The tests of type 1) and 2) are computed on the original variables, whereas scaled residuals are used for tests 3) and 4), to guarantee invariance of the testing procedures to transformations of location and scale.

A further aim of our experiments is to gain useful insights on the extent to which the value of the parameters of the two tests (β and grid points) affects their power and, consequently, on the most appropriate choice that would ensure the best statistical properties in both cases. For this purpose, each analysis is replicated for various choices of the relevant parameters. The values selected for β in the WF approach are $\beta = 0.5, 1, 2$, in line with those suggested by Henze and Zirkler [1990]. Since the tail behaviour of a probability distribution is reflected by the behaviour of the characteristic function at the origin, a small value for β that concentrates the mass of the weight function near the origin, should improve the sensitivity of the test against alternative distributions with fatter tails. The simulations for the FGP test are carried on by using $5, 10$, and 20 grid points for problems in 2 and 4 variables and 20, 30 and 50 points in a 16-dimensional context. The choice for the grid points is made by sampling from a standard multivariate normal with low-discrepancy Faure sequences, in order to ensure that the points are concentrated around the origin.[5]

[5]Alternative choices for the number of grid points and the criteria for their selection (such as equally spaced points on an interval around the origin) were attempted, but they proved to be less powerful than the ones adopted here.

6.4.2 Size of the tests

The size of the two tests under comparison is evaluated by computing the frequency of rejection (at the 95% confidence interval) out of 10,000 replications when the forecasted distribution (assumed to be, for simplicity, $N(0, I_d)$) coincides with the true distribution of the data. The results, reported in Table 6.2, seem to indicate that for both approaches the size increases as the dimensionality of the problem increases. In particular, for $d = 16$ the WF tests displays percentages of rejection above 5% whatever choice is made for the parameter β. This suggests a tendency to over-reject the null in high dimensions that should be taken into account when assessing the power of this procedure.[6] The discrete version generally exhibits better size for most of the possible choices in terms of number of grid points.

6.4.3 Bias in the first moment

We start by comparing the power to detect a bias in the mean of the distribution displayed by the two ECF testing procedures under investigation, as well as by the Hotelling-T^2 statistic, which represents a good benchmark for testing this type of misspecification under the null hypothesis of multivariate normality.[7]

We assume that the distributional forecast coincides with a standard multivariate normal $N(0, I_d)$, when the data follow a multivariate $N(\mu, I_d)$. The mean vector μ of the true distribution is set equal to the mean of the process under the null (a vector of zeros in our case) augmented by a vector of equal components, which correspond to the contours of the multivariate normal for a given choice for the constant $c > 0$:[8]

$$(x - \mu)'\Sigma^{-1}(x - \mu) = c^2$$

[6] Similar figures for the size of this test were reported by Henze and Zirkler [1990].

[7] The expression and the distribution of the Hotelling-T^2 statistic can be found, for example, in Bilodeau and Brenner [1999].

[8] A contour consists of the set of points of a multivariate normal having equal probability density.

In order to explore a wide range of alternatives, our choices for c are $c = 2, 4, 6, 8$, plus the value for c that defines the contours relative to the 95% of probability density.

The results, in terms of percentage of rejections of the null hypothesis, are shown in Table 6.3. As expected, all the tests exhibit a loss in power as the dimensionality of the problem increases. Hotelling-T^2 statistic turns out to be the most powerful test: a satisfactory percentage of rejections of the null is reached for biases in the mean when $c \geq 4$ (for $d = 2, 4$) or $c \geq 5$ (for $d = 16$). The performance of the two tests based on the ECF is quite similar, with the weight function approach performing slightly better in low dimensions and slightly worse in higher dimensions. For this testing procedure, the highest power is obtained for a small β ($\beta = 0.5$), and deviations from the forecasted mean relative to a choice of $c \geq 4$ (for $d = 2$) or $c \geq 6$ (for $d = 4, 16$). The loss in power consequent to a wrong choice for β is important, especially when the problem involves many variables: for $d = 16$, the loss from choosing $\beta = 2$ instead of $\beta = 0.5$ is around 70-80 percentage points for the two largest biases. For the test based on fixed grid points, the number of points associated with better statistical properties seems to be around $m = 5$ when the dimensionality of the problem is low and 20 for a 16-dimensional problem. Again, an inappropriate selection of the parameters of the test (i.e. number of grid points) would substantially reduce the power of the test statistic, in particular as the dimensionality increases.

6.4.4 Bias in the second moment

Here we investigate the power of the tests based on the ECF to detect a bias in the covariance matrix when the forecasted distribution is $N(0, I_d)$, while the true data generating process is $N(0, \Sigma)$. The following measure of the "distance" between the two matrices Σ and I_d:

$$\delta = \text{trace}[(\Sigma^{-1} - I_d)(\Sigma^{-1} - I_d)]$$

is employed to quantify how different the true and the hypothesised covariance structures are. The second moments of the alternative, real data generating process are selected such that a similar range of distances ($\delta \simeq 0.03, 0.5, 1.5$) is replicated for all the dimensions considered in the analysis. As a benchmark, we propose a procedure suggested by Nagao [1973] to test for the equality of a covariance matrix to a given matrix, based on the eigenvalues of a combination of the two matrices.

Table 6.4 displays the outcomes of the testing experiment. As expected, given the ability of the eigenvalues to capture the covariance structure of a distribution, Nagao's test is the most powerful in detecting a bias in the second moment of the forecasted distribution. At a 95% confidence interval, a good level of power is attained for biases corresponding approximately to $\delta \simeq 0.5$.

Between the testing criteria based on the ECF, the FGP approach is preferable for low dimensions, but the order of preference reverts as the dimensionality increases, when the WF approach exhibits a level of power very close to the one displayed by Nagao's test. The difference in power between the two testing procedures rapidly increases along with the dimensionality of the problem, rising from few percentage points for the largest bias in $d = 2, 4$ to 15-20 points in 16 dimensions. In this latter case, a percentage of rejections of the null above 95% cannot be attained for the discrete version of the test even when the difference between the true and the hypothesised covariance matrix is the largest amongst the alternatives investigated, and the choice made for the number of grid points is the one that shows the highest power ($m = 30$). This seems to support the intuition that the test based on the WF is more appropriate for problems involving several variables. For this testing procedure, small values for β again seem to guarantee better power, especially in

a high-dimensional context. As before, an inappropriate selection for β or for the number of grid points strongly affects the power of the tests, though less drastically than when a bias occurs in the mean.

6.4.5 Misspecification of the form of the density forecast

An experiment is designed to assess the performance of the two ECF tests in detecting a misspecification in the form of the distributional forecast, when the first two moments are correctly specified. The null hypothesis is that the data follow a standard multivariate normal. The alternative specifications for the true distribution are: 1) two mixtures of multivariate normal distributions with same mean and covariance matrix as the null, but different degrees of kurtosis; 2) a multivariate standard t with 5 degrees of freedom; 3) a multivariate standard t with 10 degrees of freedom. For this experiment, the power of the ECF tests is compared to that of the well-known test for multivariate normality proposed by Mardia [1970], based on a measure of multivariate kurtosis.

The outcomes of the analysis are reported in Table 6.5. The counterintuitive increase in the power to spot a misspecification in the form of the distribution that all the testing procedures seem to exhibit as d increases, is due to the particular choice made for the distributions used for the sample generation. The WF approach clearly outperforms the FGP approach for all the alternatives under investigation, and in each dimension. At the 95% confidence level, the WF approach (for $\beta = 0.5$ or $\beta = 1$) in a 16-dimensional context, represents a powerful test against all the alternatives we specified. For the case of 2 and 4 variables, this test possesses good power only against few alternative specifications, namely the mixture with the highest level of excess kurtosis, and the multivariate t with 5 degrees of freedom. On the other hand, the FGP version of the test hardly shows a reasonable power against any of the alternatives considered, in any dimension, and for whatever choice of the

number of grid points. The gap in the power with respect to the continuous version increases dramatically along with the dimensionality of the problem: for $d = 16$ the gap varies from 30 to 80 percentage points, according to the different alternative distributions.

Again, the power of the WF test against distributions that are more fat-tailed than the null proves to be higher for small values of β, and the loss in power consequent to a wrong choice for this parameter becomes more relevant in higher dimensions. Our results show a substantial loss in power from choosing $\beta = 2$ when $d = 16$, that reaches a maximum of 50 percentage points when the true process is the mixture with the highest kurtosis. Quite surprisingly, the WF procedure seems also to outperform Mardia's test when the alternative hypotheses are represented by the two mixtures of normal distributions. However, when the data are distributed according to a multivariate t, the test suggested by Mardia displays the best statistical performance.

6.4.6 Bias in higher moments

To assess the statistical performance of the ECF testing procedures in detecting biases in higher moments of the forecasted distribution, we also implement our tests in a context where both the null and the true distributions have fat tails, but the latter displays higher kurtosis. As null hypothesis we choose, from the previous analysis, the mixture of two multivariate normal distributions with less excess kurtosis. As alternative distributions we consider the other, more leptokurtic, mixture of normals, and the multivariate t distributions, with 5 and 10 degrees of freedom. The choice of a mixture of multivariate normal distributions as a possible specification for the joint distributional forecast of financial asset returns guarantees a high tractability, and seems a plausible assumption to make given its ability to model the fat-tailness exhibited by financial returns. Under this particular assumption for

the null, the WF statistic consists of an adaptation of the original test suggested by Henze and Zirkler [1990] to the case when the null hypothesis is a standardised mixture of multivariate normal distributions (see Appendix B).

The results of the experiment, displayed in Table 6.6, recall those from the previous analysis on the misspecification in the type of density forecast. Again the WF approach seems to be the most powerful test, provided that an appropriate, small value is selected for the parameter β. The performance exhibited by the continuous version of the test overcomes the one shown by the discrete version for any choice of the alternative distribution and any dimension. However, the gap in power between the two procedures, especially for high-dimensional problems, is narrower than the one observed in the previous case. As expected, small biases in the excess kurtosis cannot be easily detected in a multivariate context. None of the testing procedures seems to display a satisfactory power against two out of the three alternative specifications chosen for the true data generating process. A percentage of rejections of the null above 95% can be obtained only from the WF approach, for $d = 16$, and for a highly leptokurtic (multivariate t with 5 degrees of freedom) alternative, whereas much less power has been observed when the bias in the excess kurtosis is less significant.

6.5 Conclusions

In the present work we have investigated the statistical size and power of two multivariate goodness-of-fit tests based on the ECF, to detect misspecifications in density forecasts which are close to a multivariate normal specification. The outcomes from the overall analysis suggest that the continuous version (WF) of the test statistic maintains good power to spot misspecifications of different nature in the null distribution also when the dimensionality of the problem increases. Therefore, this test

statistic seems to represent a good candidate for the evaluation of the multivariate distributional forecasts developed within VaR problems, where several financial variables are involved.

As expected, the performance of the discrete version (FGP) of the ECF test worsens as the dimensionality grows. Insufficient statistical power to detect biases in the second moment is observed for $d = 16$. Failure to spot misspecifications in the shape of the density forecasts is already a problem in the four-dimensional case. We should acknowledge the fact that the particular choice made for the grid points in the FGP test may not be the most appropriate, and that a different one could have produced better power. However, in any case, the excessive degree of arbitrariness involved in the process of selection of number and magnitude of the grid points would seriously affect the reliability and usefulness of the FGP approach, especially in high-dimensional problems. Much less arbitrariness is involved in the WF approach. Here the power of the test strongly depends on the constant β, but the results from all our experiments agree in recommending a small value for β, particularly in a multidimensional context.

As already pointed out, the WF testing procedure presents a serious drawback: its applicability is limited to few specifications for the null distribution, whose characteristic function possesses a tractable form that consents a closed form solution to the integral in Eq. (6.1). In the present analysis we have assumed only two forms for the density forecast, both satisfying this requirement: 1) a multivariate normal distribution; 2) a mixture of multivariate normal distributions. We have seen that in many cases these distributional assumptions (in particular the second one) are not too unrealistic for modelling the joint density of the outcomes of the financial risk factors.

However, in some cases it might be necessary to consider more general specifications. In such circumstances, the implementation of the WF statistic can be-

come highly problematic, or even impossible. Future research might throw light on whether the WF approach can be extended and applied to other classes of multivariate distributions, provided that different, appropriate choices are made for the weight function.

Alternative testing procedures might also be explored. A promising approach is based on investigating the multidimensional sample along few (unidimensional) projections of interest. Particular attention must be paid to the choice of the projections, which in fact represent specific portfolios of the risk factors. A comparison between the implementation of projection-based techniques and ECF tests on a real multidimensional data set constitutes interesting material for future research.

Table 6.1: Critical values for continuous Weight Function testing procedures.

CV for WF tests on standardised variables
H_0: multivariate normal

	$\beta = 0.5$	$\beta = 1$	$\beta = 2$
$d = 2$	0.0790	0.5542	1.0965
$d = 4$	0.1670	0.7793	1.0557
$d = 16$	0.7304	0.9999	1

CV for WF tests on original variables
H_0: multivariate normal

	$\beta = 0.5$	$\beta = 1$	$\beta = 2$
$d = 2$	0.8440	1.3427	1.4189
$d = 4$	1.0371	1.2751	1.1253
$d = 16$	1.0622	1.0037	1

CV for WF tests on standardised variables
H_0: mixture of multivariate normal distributions

	$\beta = 0.5$	$\beta = 1$	$\beta = 2$
$d = 2$	0.0971	0.5965	1.1375
$d = 4$	0.1899	0.8121	1.0897
$d = 16$	0.7937	1.0061	1.0001

Table 6.2: Size of the tests - $H_0 : N(0, I_d)$ vs $H_1 : N(0, I_d)$.

d	*Weight Function approach*			*Fixed Grid Points approach*				
	$\beta = 0.5$	$\beta = 1$	$\beta = 2$	$m = 5$	$m = 10$	$m = 20$	$m = 30$	$m = 50$
2	4.9	4.5	4.4	3.9	5.2	2.7	-	-
4	3.4	4.7	4.9	1.7	2.6	3.7	-	-
16	6.0	5.7	5.4	-	-	5.7	5.8	3.7

Table 6.3: Bias in the mean - $H_0 : N(0, I_d)$ vs $H_1 : N(\mu, I_d)$.

d	c	μ	*Weight Function approach*			*Fixed Grid Points approach*					*Hotelling*
			$\beta = 0.5$	$\beta = 1$	$\beta = 2$	$m = 5$	$m = 10$	$m = 20$	$m = 30$	$m = 50$	T^2
2	2	0.141	40.5	28.0	17.3	25.3	20.4	7.2	-	-	41.8
	*2.45	0.173	54.0	42.0	25.6	36.5	24.9	8.8	-	-	56.6
	4	0.283	94.7	86.2	62.1	79.0	63.5	27.0	-	-	94.5
	6	0.424	100.0	99.7	96.9	99.4	97.1	76.3	-	-	100.0
	8	0.566	100.0	100.0	100.0	100.0	100.0	98.9	-	-	100.0
4	2	0.100	27.7	17.9	8.1	22.2	22.5	10.5	-	-	29.2
	*3.08	0.154	62.8	39.3	15.4	51.6	42.8	19.0	-	-	68.7
	4	0.200	87.1	67.9	27.7	80.0	68.4	33.0	-	-	93.2
	6	0.300	99.8	96.4	66.3	99.4	98.4	81.6	-	-	100.0
	8	0.400	100.0	100.0	93.6	100.0	100.0	99.7	-	-	100.0
16	2	0.035	11.7	5.7	18.3	-	-	14.6	15.3	6.3	17.4
	4	0.071	49.8	9.2	20.0	-	-	51.6	36.9	9.1	70.2
	*5.13	0.091	76.0	15.0	19.1	-	-	81.0	60.4	19.4	92.6
	6	0.106	91.0	21.3	20.5	-	-	92.0	78.3	29.5	98.3
	8	0.141	99.8	39.5	20.9	-	-	99.9	98.6	76.0	100.0

Table 6.4: Bias in the covariance - $H_0 : N(0, I_d)$ vs $H_1 : N(0, \Sigma)$.

d	C	δ	*Weight Function approach*			*Fixed Grid Points approach*					*Nagao's test*
			$\beta = 0.5$	$\beta = 1$	$\beta = 2$	$m = 5$	$m = 10$	$m = 20$	$m = 30$	$m = 50$	
2	$C = \begin{pmatrix} 2 & 0.9 \\ 0.9 & 3 \end{pmatrix}$	1.519	100.0	100.0	100.0	100.0	100.0	100.0	-	-	100.0
	$C = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$	0.526	18.5	40.6	33.4	66.2	66.5	39.2	-	-	93.7
	$C = \begin{pmatrix} 1 & 0.1 \\ 0.1 & 1 \end{pmatrix}$	0.021	5.8	6.8	5.9	18.1	16.8	8.3	-	-	12.6
4	$C = \begin{pmatrix} 1 & 0.4 & 0.4 & 0.4 \\ 0.4 & 1 & 0.4 & 0.4 \\ 0.4 & 0.4 & 1 & 0.4 \\ 0.4 & 0.4 & 0.4 & 1 \end{pmatrix}$	1.631	93.0	95.9	72.0	99.7	99.4	99.8	-	-	100.0
	$C = \begin{pmatrix} 1 & 0.3 & 0.2 & 0.1 \\ 0.3 & 1 & 0 & 0 \\ 0.2 & 0 & 1 & 0 \\ 0.1 & 0 & 0 & 1 \end{pmatrix}$	0.432	13.8	19.1	13.6	38.3	39.8	36.0	-	-	64.8
	$C = \begin{pmatrix} 1.05 & 0 & 0 & 0 \\ 0 & 1.05 & 0 & 0 \\ 0 & 0 & 1.05 & 0 \\ 0 & 0 & 0 & 1.05 \end{pmatrix}$	0.033	10.3	10.7	6.8	10.5	13.4	11.2	-	-	21.6
16	$C = \begin{pmatrix} 1.2 & 0 & \cdots & 0 \\ 0 & 1.2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1.2 \end{pmatrix}$	1.494	100.0	100.0	100.0	-	-	83.7	85.8	62.1	100.0
	$C = \begin{pmatrix} 1.1 & 0 & \cdots & 0 \\ 0 & 1.1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1.1 \end{pmatrix}$	0.482	100.0	83.3	73.9	-	-	40.2	44.9	21.8	99.8
	$C = \begin{pmatrix} 1.025 & 0 & \cdots & 0 \\ 0 & 1.025 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1.025 \end{pmatrix}$	0.036	17.0	6.4	19.1	-	-	12.6	16.2	7.3	23.5

C correlation matrix of Σ.

Table 6.5: Misspecification of the density forecast - $H_0 : N(0, I_d)$.

d	H_1	*Weight Function approach* $\beta = 0.5$	$\beta = 1$	$\beta = 2$	*Fixed Grid Points approach* $m = 5$	$m = 10$	$m = 20$	$m = 30$	$m = 50$	*Mardia test*
2	$0.3N\left(0, C = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}\right) + 0.7N\left(0, C = \begin{pmatrix} 1 & -0.3 \\ -0.3 & 1 \end{pmatrix}\right)$	12.9	12.9	11.3	19.7	22.3	15.1	-	-	16.5
	$0.5N\left(0, C = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}\right) + 0.5N\left(0, C = \begin{pmatrix} 1 & -0.9 \\ -0.9 & 1 \end{pmatrix}\right)$	83.4	99.7	99.9	97.0	95.4	72.0	-	-	75.9
	t_5	79.4	73.4	61.2	51.7	56.3	49.1	-	-	87.7
	t_{10}	32.3	27.0	16.4	21.5	27.2	17.6	-	-	47.9
4	$0.3N\left(0, C = \begin{pmatrix} 1 & 0.7 & 0.7 & 0 \\ 0.7 & 1 & 0 & 0 \\ 0.7 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\right) + 0.7N\left(0, C = \begin{pmatrix} 1 & -0.3 & -0.3 & 0 \\ -0.3 & 1 & 0 & 0 \\ -0.3 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\right)$	18.4	23.1	22.6	4.8	6.4	16.7	-	-	18.7
	$0.5N\left(0, C = \begin{pmatrix} 1 & 0.7 & 0.7 & 0 \\ 0.7 & 1 & 0 & 0 \\ 0.7 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\right) + 0.5N\left(0, C = \begin{pmatrix} 1 & -0.7 & -0.7 & 0 \\ -0.7 & 1 & 0 & 0 \\ -0.7 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\right)$	74.5	99.6	98.5	30.0	28.1	41.0	-	-	52.5
	t_5	97.7	95.8	83.9	14.0	16.9	49.3	-	-	98.5
	t_{10}	56.0	46.4	24.2	4.0	8.1	24.7	-	-	73.8
16	$0.7N\left(0, C = \begin{pmatrix} 1.1 & 0 & \cdots & 0 \\ 0 & 1.1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1.1 \end{pmatrix}\right) + 0.3N\left(0, C = \begin{pmatrix} 0.71 & 0 & \cdots & 0 \\ 0 & 0.71 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 0.71 \end{pmatrix}\right)$	100.0	99.1	60.1	-	-	6.8	20.1	9.0	99.1
	$0.6N\left(0, C = \begin{pmatrix} 0.79 & 0 & \cdots & 0 \\ 0 & 0.79 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 0.79 \end{pmatrix}\right) + 0.4N\left(0, C = \begin{pmatrix} 1.25 & 0 & \cdots & 0 \\ 0 & 1.25 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1.25 \end{pmatrix}\right)$	100.0	99.9	49.3	-	-	12.1	36.7	19.4	100.0
	t_5	100.0	100.0	99.9	-	-	21.2	71.4	58.4	100.0
	t_{10}	100.0	99.7	66.7			10.2	42.6	26.4	100.0

Table 6.6: Bias in higher moments.

d	H_0 and H_1	*Weight Function approach* $\beta=0.5$	$\beta=1$	$\beta=2$	*Fixed Grid Points approach* $m=5$	$m=10$	$m=20$	$m=30$	$m=50$
2	$\mathbf{H_0}: 0.3N\left(0, C=\begin{pmatrix}1 & 0.7\\ 0.7 & 1\end{pmatrix}\right)+0.7N\left(0, C=\begin{pmatrix}1 & -0.3\\ -0.3 & 1\end{pmatrix}\right)$								
	$0.5N\left(0, C=\begin{pmatrix}1 & 0.9\\ 0.9 & 1\end{pmatrix}\right)+0.5N\left(0, C=\begin{pmatrix}1 & -0.9\\ -0.9 & 1\end{pmatrix}\right)$	42.8	83.8	94.2	90.6	83.4	35.3	-	-
	t_5	58.2	56.0	40.9	16.4	26.1	23.4	-	-
	t_{10}	18.3	14.0	8.9	4.4	9.5	6.8	-	-
4	$\mathbf{H_0}: 0.3N\left(0, C=\begin{pmatrix}1 & 0.7 & 0.7 & 0\\ 0.7 & 1 & 0 & 0\\ 0.7 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}\right)+0.7N\left(0, C=\begin{pmatrix}1 & -0.3 & -0.3 & 0\\ -0.3 & 1 & 0 & 0\\ -0.3 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}\right)$								
	$0.5N\left(0, C=\begin{pmatrix}1 & 0.7 & 0.7 & 0\\ 0.7 & 1 & 0 & 0\\ 0.7 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}\right)+0.5N\left(0, C=\begin{pmatrix}1 & -0.7 & -0.7 & 0\\ -0.7 & 1 & 0 & 0\\ -0.7 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}\right)$	22.8	57.4	43.6	13.4	12.1	10.6	-	-
	t_5	90.8	89.2	73.1	5.2	5.4	28.4	-	-
	t_{10}	37.5	33.4	22.2	2.0	1.7	8.1	-	-
16	$\mathbf{H_0}: 0.7N\left(0, C=\begin{pmatrix}1.1 & 0 & \cdots & 0\\ 0 & 1.1 & & \vdots\\ \vdots & & \ddots & 0\\ 0 & \cdots & 0 & 1.1\end{pmatrix}\right)+0.3N\left(0, C=\begin{pmatrix}0.71 & 0 & \cdots & 0\\ 0 & 0.71 & & \vdots\\ \vdots & & \ddots & 0\\ 0 & \cdots & 0 & 0.71\end{pmatrix}\right)$								
	$0.6N\left(0, C=\begin{pmatrix}0.79 & 0 & \cdots & 0\\ 0 & 0.79 & & \vdots\\ \vdots & & \ddots & 0\\ 0 & \cdots & 0 & 0.79\end{pmatrix}\right)+0.4N\left(0, C=\begin{pmatrix}1.25 & 0 & \cdots & 0\\ 0 & 1.25 & & \vdots\\ \vdots & & \ddots & 0\\ 0 & \cdots & 0 & 1.25\end{pmatrix}\right)$	22.5	6.0	3.2	-	-	4.4	20.2	8.7
	t_5	100.0	99.2	57.1	-	-	8.1	57.4	42.6
	t_{10}	36.9	6.4	4.0			2.8	24.1	16.4

Chapter 7

Conclusions and Directions for Future Research

This thesis has explored a variety of density forecasting problems of crucial importance to market risk managers. Such issues are related to the creation and/or the appraisal of density forecasts for the future outcomes of the relevant driving risk factors.

As expected, our analysis has highlighted how rich and complicated this fundamental component of the risk measurement systems turns out to be. Despite the increasing research activity in this area, there is still a lot of scope for improving the techniques aimed at both the generation and the appraisal of the relevant density forecasts. Our work has made a contribution in this respect. On the particular findings relative to our specific applications we have already extensively commented at the conclusion of each chapter. It is worth summarising the key points here.

The study of the density forecasts implied from option prices has produced three main results. On the estimation side, the Normal Inverse Gaussian specification has turned out to be a particularly appropriate parametric form for fitting implied densities, given both its simple parametrisation and its very good in-sample

fit. On the evaluation side, despite the small size of our sample, we have found some evidence of the existence of biases in the implied density forecasts. This indicates that the widespread use of option-implied densities as correct forecasts of the distribution of the underlying asset at expiry should perhaps be reconsidered or, at least, be preceded by an adequate appraisal of the density forecasts. We have also illustrated the importance of performing a separate and careful analysis of the tails of the implied densities, to avoid the distortions that are often introduced by *ad hoc* extrapolation techniques.

The analysis of the volatility curve has revealed how non-trivial it is to accurately model its dynamics and, consequently, to generate correct density forecasts of the changes in value of option portfolios. The comparison between vega risk exposures predicted by the Kalman filter model and by two benchmarks, for four option portfolios, has highlighted the importance of relying on models (such as the Kalman filter) which can account for non-parallel shifts in the volatility curve. This evidence, entirely in line with several recent findings in the literature, is especially relevant to risk managers, who often make use of oversimplified one-factor models, unable to fully capture the evolution of the volatility curve.

In Chapter 5 we have illustrated how to "build" successfully a continuous time model of equity index returns from a data set of 5-minute returns on an equity index future. The specification we have derived is relatively simple, and able to capture and model most of the aspects observed in equity index markets, namely: seasonality in intraday volatility, stochastic volatility, presence of jumps and leverage effect. Instead of imposing a model *a priori*, we have shown how the information content of intraday data can be effectively exploited to derive a "bottom-up" model capable of generating realistic density forecasts of the returns on an equity index.

In Chapter 6 we have dealt with the issue of evaluating multivariate distributional forecasts. As expected, the main conclusion that can be drawn from our

analysis is that assessing multivariate density forecasts for truly highly dimensional problems is indeed a very hard task. Our results suggest that a continuous version of a multivariate goodness-of-fit test based on the Empirical Characteristic Function (ECF) may represent a promising technique, when the density forecasts do not depart too significantly from a multivariate gaussian.

Future research on density forecasting for market risk management can proceed along several directions. Other applications are awaited. An aspect which has not been dedicated enough attention by either the existing literature or our work, is the optimal revision (or "tuning") of the density forecasts. Once evidence of the existence of some sources of misspecification in the density forecasts is found, a formal and rigorous treatment of the optimal way of performing a recalibration is desirable.

Some ideas for further research on the specific issues explored in this thesis were given at the end of each chapter. They are now recapped here for clarity.

Future work on the estimation of option-implied distributions may include the investigation of the distortions produced by the common practice of extrapolating the tails outside the truncation points (i.e. the extremes of the range of available strikes). As for the appraisal of option-implied densities, the availability of a larger data set will allow the implementation of tests of power/efficiency, precluded by the small size of our quarterly sample. Alternative risk-adjustment methods could also be attempted.

The choice of more robust and sophisticated benchmark models constitutes the most immediate direction for improvement of our vega risk analysis. Given the high variability of our density forecasts across the four options portfolios, it would be very interesting, for risk management purposes, to assess the robustness and the generality of our results across a very rich class of portfolios.

Further research on the continuous time model inferred from high frequency data may involve: 1) the replication of the analysis for different kinds of financial assets; 2) the derivation and the assessment of simplified versions of the model, relative to specific time horizons of interest.

The exercise on the appraisal of multivariate density forecasts would be more informative if performed on real data. Also, it would be worth comparing the results from the ECF tests with those from alternative testing procedures. Techniques based on the projection of the multivariate data along few directions might constitute a valid benchmark. In this case, the projections, which in fact represent specific portfolios of the risk factors, should be chosen with care, to ensure generality and robustness of the results.

Appendix A

Intraday Volatility Estimates

A.1 Decomposition of the variance of high frequency volatility innovations

Given our specification for the continuous time model:

$$\begin{aligned}
r_t &= s_t \sigma_t \varepsilon_t \\
dM_t &= \theta(\ln(S_{t-1}) - M_{t-1})dt \\
\ln(\sigma_t) &= \kappa(\ln(S_{t-1}) - M_{t-1}) + \upsilon_t \\
\upsilon_t &= \ln(\sigma_{s,t}) + \ln(\sigma_{l,t}) \\
d\ln(\sigma_{s,t}) &= -\alpha_s dt \ln(\sigma_{s,t}) + \beta_s \sqrt{dt} dW_{s,t} \\
d\ln(\sigma_{l,t}) &= -\alpha_l dt \ln(\sigma_{l,t}) + \beta_l \sqrt{dt} dW_{l,t}
\end{aligned}$$

the innovations in log volatility are expressed as follows:

$$\begin{aligned}
d\ln(\sigma_t) &= \kappa(\ln(S_{t-1}) - M_{t-1}) - \kappa(\ln(S_{t-2}) - M_{t-2}) + d\ln(\sigma_{s,t}) + d\ln(\sigma_{l,t}) \\
&= [\kappa r_{t-1} - \kappa(M_{t-1} - M_{t-2})] + d\ln(\sigma_{s,t}) + d\ln(\sigma_{l,t})
\end{aligned}$$

Since the three components on the right hand side are independent, the total variance of the innovations $var[d\ln(\sigma_t)]$ is given by the sum of the variances of the

individual components:

$$var[d\ln(\sigma_t)] = var[\kappa r_{t-1} - \kappa(M_{t-1} - M_{t-2})] + var[d\ln(\sigma_{s,t})] + var[d\ln(\sigma_{l,t})]$$

where:

$$\begin{aligned} var[d\ln(\sigma_{s,t})] &= \beta_s^2\, dt \\ var[d\ln(\sigma_{l,t})] &= \beta_l^2\, dt \\ var[\kappa r_{t-1} - \kappa(M_{t-1} - M_{t-2})] &\approx \kappa^2 var[r_{t-1}] \end{aligned}$$

given that, at 5-minute level, we expect the variable M not to change significantly and, therefore, $var[(M_{t-1} - M_{t-2})] \approx 0$.

We recall that $E[\sigma_t] = 1$ and $var[\varepsilon] = \pi/2$ hence, on average, $var[r_t] = s_t^2(\pi/2)$. For our estimates of the parameters of the model, $\kappa = -4.34$, $\beta_s^2 = 0.044$, $\beta_l^2 = 0.0019$, and given an average value for the seasonal coefficient of $\overline{s} = 0.00085$, we obtain $var[d\ln(\sigma_t)] = 0.000454$, $var[d\ln(\sigma_{s,t})] = 0.000415$, $var[d\ln(\sigma_{l,t})] = 0.00001792$ and $var[\kappa r_{t-1}] = 0.00002137$. Therefore, 91.35% of the total variance is explained by the transient component, 3.94% by the persistent component and the remaining 4.71% is attributable to the leverage effect.

A.2 Transition probabilities for the high frequency volatility process

In what follows we derive the transition probabilities associated with the discretised versions of the two continuous time processes (without and with jumps) that we specified in Section 5.4.3 to model the dynamics of intraday volatility.

A.2.1 Diffusion model

The discrete analogous, obtained via Euler discretisation,[1] of the continuous time model proposed to describe the evolution of 5-minute volatility, is given by:

$$\Delta \ln(\sigma_{s,t}) = -\alpha_s \Delta t \ln(\sigma_{s,t}) + \beta_s \sqrt{\Delta t}\omega_s$$

with $\omega_s \sim N(0,1)$. In order to fit this structure into our non-linear filtering technique, we let each log volatility value in our range $\ln(\sigma_j)$ for $j = 2, \ldots, N-1$, evolve according to a trinomial tree. Between t and $t+\Delta t$ the log volatility (equal to $\ln(\sigma_i)$ in t) can go up by the amount Δ_r (step size) to level $\ln(\sigma_{i-1})$ with probability p_u, down by $-\Delta_r$ to level $\ln(\sigma_{i+1})$ with probability p_d or stay at level $\ln(\sigma_i)$ with probability p_e. Here we assume that both the transition probabilities p_u, p_d, p_e and the step size are constant. Following a standard procedure, the transition probabilities are obtained by equating the first two moments of the discrete time process to the corresponding moments of the continuous time model:

$$\begin{aligned}
p_u\Delta_r + p_d(-\Delta_r) &= E[\Delta \ln(\sigma_i)] = -\alpha_s \Delta t \ln(\sigma_i) \\
p_u(\Delta_r)^2 + p_d(-\Delta_r)^2 &= var[\Delta \ln(\sigma_i)] + E^2[\Delta \ln(\sigma_i)] = \beta_s^2 \Delta t + [-\alpha_s \Delta t \ln(\sigma_i)]^2 \\
p_u + p_d + p_e &= 1
\end{aligned}$$

for $i = 2, \ldots, N-1$. For $i = 1$, $p_e^{(1)} = p_e$ and $p_d^{(1)} = 1 - p_e^{(1)}$ and, similarly, for $i = N$, $p_e^{(N)} = p_e$ and $p_u^{(N)} = 1 - p_e^{(N)}$. Δ_r should be chosen close to three times the standard deviation of the continuous model ($\beta_s\sqrt{dt}$) to ensure an adequate representation of the process in a discrete framework.

A.2.2 Jump diffusion model

The discrete time version of our process for intraday stochastic volatility with jumps is given by:

[1] In this context, the Euler discretisation should not introduce a significant bias, since we work with high frequency data, which are frequently spaced.

$$\Delta \ln(\sigma_{s,t}) = -\alpha_s \Delta t \ln(\sigma_{s,t}) + \beta_s \sqrt{\Delta t}\omega_s + \sum_{i=1}^{N(\Delta t)} Y_i - \lambda \Delta t E[Y]$$

The probability of a generic jump occurring in a time step Δt is equal to $\lambda \Delta t$ and we assume that multiple jumps cannot happen in a single time step. As before, the step size is equal to Δ_r. Let p^l denote the probability of a large jump of size $20\Delta_r$, p^m the probability of an intermediate jump of $10\Delta_r$ and $p^s = 1 - p^l - p^m$ the probability of a small jump of size $3\Delta_r$.

Between t and $t + \Delta t$ the log volatility (equal to $\ln(\sigma_i)$ in t) can evolve according to the trinomial structure analysed before, with probability $1 - \lambda \Delta t$ (no jumps occurring), or it can jump up by $3\Delta_r$, $10\Delta_r$ or $20\Delta_r$ with probabilities $\lambda \Delta t p^s$, $\lambda \Delta t p^m$, $\lambda \Delta t p^l$, respectively. Again, the transition probabilities are obtained by equating the first two moments of the discrete and the continuous time process:

$$(1 - \lambda \Delta t)(p_u \Delta_r + p_d(-\Delta_r)) + \lambda \Delta t(p^s 3\Delta_r + p^m 10\Delta_r + p^l 20\Delta_r) =$$
$$-\alpha_s \Delta t \ln(\sigma_i)$$

$$(1 - \lambda \Delta t)(p_u(\Delta_r)^2 + p_d(-\Delta_r)^2) + \lambda \Delta t(p^s(3\Delta_r)^2 + p^m(10\Delta_r)^2 + p^l(20\Delta_r)^2) =$$
$$\beta_s^2 \Delta t + \lambda \Delta t(p^s(3\Delta_r)^2 + p^m(10\Delta_r)^2 + p^l(20\Delta_r)^2) + [-\alpha_s \Delta t \ln(\sigma_i)]^2$$
$$p_u + p_d + p_e = 1$$

with the following restrictions dictated by the grid structure:

- $\lambda = 0$, $p_e^{(1)} = p_e$, $p_u = 0$ and $p_d^{(1)} = 1 - p_e^{(1)}$ for $i = 1$ (highest volatility value, no possibility of upward movements).
- $\lambda = 0$ for $i = 2, 3$ (no possibility of jumps).
- $p^s = 1$, $p^m = p^l = 0$ for $i = 4, \ldots, 10$ (only small jumps possible).
- $p^m = 1 - p^s$, $p^l = 0$ for $i = 11, \ldots, 20$ (only small and intermediate jumps possible).

Appendix B

Multivariate Continuous ECF Test Statistics

In this Appendix we include the complete formulations for the test statistics, based on the continuous version of the ECF multivariate goodness-of-fit test, that have been employed in our analysis.

The original statistic suggested by Henze and Zirkler [1990] is computed on the scaled residuals in place of the original variables:

$$Y_j = S^{-1/2}(X_j - \overline{X})$$

and under the null hypothesis of multivariate normality.[1] The formula is as follows:

$$\begin{aligned} D_{n,\beta} &= \frac{1}{n^2}\sum_{j,k=1}^{n}\exp\left(-\frac{\beta^2}{2}\parallel Y_j - Y_k \parallel^2\right) - 2(1+\beta^2)^{-d/2}\frac{1}{n}\cdot \\ &\cdot\sum_{j=1}^{n}\exp\left(-\frac{\beta^2}{2(1+\beta^2)}\parallel Y_j \parallel^2\right) + (1+2\beta^2)^{-d/2} \end{aligned} \tag{B.1}$$

and it is used in our study to test for misspecifications of the form of the density forecast, when the forecasted density is a standard multivariate Gaussian, and the true distribution is fat-tailed.

[1] $\overline{X}$ denotes the sample mean and S the sample covariance matrix.

When the test statistic is computed on the original variables, the formulation becomes:

$$
\begin{aligned}
D_{n,\beta} \;=\; & \frac{1}{n^2}\sum_{j,k=1}^{n}\exp\left(-\frac{\beta^2}{2}\parallel X_j - X_k \parallel^2\right) + \beta^{-d}\left|2\Sigma + \frac{1}{\beta^2}I\right|^{-1/2} - \\
& -2\beta^{-d}\frac{1}{n}|S|^{-1/2}\sum_{j=1}^{n}\exp\left(-\frac{1}{2}\|S^{-1/2}(X_j-\mu)\|^2\right) \qquad \text{(B.2)}
\end{aligned}
$$

This statistic is used to test for biases in the mean and in the covariance matrix, when the true distribution is a multivariate Gaussian with mean μ and covariance matrix Σ, and the null is a standard multivariate normal.

To test for biases in higher moments when the density forecast assumes the form of a mixture of multivariate normal distributions $pN(\mu_1,\Sigma_1)+(1-p)N(\mu_2,\Sigma_2)$, we derive an extended version of the first statistic, computed on standardised variables. Its expression is the following:

$$
\begin{aligned}
D_{n,\beta} \;=\; & \frac{1}{n^2}\sum_{j,k=1}^{n}\exp\left(-\frac{\beta^2}{2}\parallel Y_j - Y_k \parallel^2\right) + p^2\beta^{-d}\left|2\Sigma_1^* + \frac{1}{\beta^2}I\right|^{-1/2} + \\
& +(1-p)^2\beta^{-d}\left|2\Sigma_2^* + \frac{1}{\beta^2}I\right|^{-1/2} - 2\frac{p}{n}\beta^{-d}|S_1|^{-1/2}\cdot \\
& \cdot\sum_{j=1}^{n}\exp\left(-\frac{1}{2}\left\|S_1^{-1/2}Y_j\right\|^2\right) - 2\frac{1-p}{n}\beta^{-d}|S_2|^{-1/2}\sum_{j=1}^{n}\exp\left(-\frac{1}{2}\left\|S_2^{-1/2}Y_j\right\|^2\right) + \\
& +2p(1-p)\beta^{-d}\left|\Sigma_1^* + \Sigma_2^* + \frac{1}{\beta^2}I\right|^{-1/2} \qquad \text{(B.3)}
\end{aligned}
$$

where $S_j = \Sigma_j^* + \frac{1}{\beta^2}I$, $\Sigma_{mix} = p\Sigma_1 + (1-p)\Sigma_2$ represents the covariance structure of the mixture,[2] and $\Sigma_j^* = \Sigma_{mix}^{-1/2}\Sigma_j(\Sigma_{mix}^{-1/2})'$ is the standardised covariance matrix of the j-th component, such that the mixture based on the scaled components has zero mean and identity covariance matrix.

[2]Since the mean is the same for both components of the mixture ($\mu_1 = \mu_2$), all the higher moments around the mean are obtained as average of the moments of the single components, weighted by their respective proportions (p and $1-p$) in the mixture.

Bibliography

[1] K. Abadir and M. Rockinger. Density-embedding functions. Working paper, HEC, 1997.

[2] P. Abken, D.B. Madan, and S. Ramamurtie. Estimation of risk-neutral and statistical densities by Hermite polynomial approximation: with an application to Eurodollar futures options. Working paper 96-5, Federal Reserve Bank of Atlanta, 1996a.

[3] P. Abken, D.B. Madan, and S. Ramamurtie. Pricing S&P500 index options using a Hilbert space basis. Working paper 96-21, Federal Reserve Bank of Atlanta, 1996b.

[4] Y. Ait-Sahalia and A.W. Lo. Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance*, 53:499–547, 1998.

[5] Y. Ait-Sahalia and A.W. Lo. Nonparametric risk management and implied risk aversion. *Journal of Econometrics*, 94:9–51, 2000.

[6] Y. Ait-Sahalia, Y. Wang, and F. Yared. Do option markets correctly price the probabilities of movement of the underlying asset? *Journal of Econometrics*, 102:67–110, 2001.

[7] C. Alexander. Principles of the skew. *Risk*, 14:29–32, 2001.

[8] S. Alizadeh, M.W. Brandt, and F.X. Diebold. Range-based estimation of stochastic volatility models. *Journal of Finance*, 57:1047–1091, 2002.

[9] K.I. Amin. Jump diffusion option valuation in discrete time. *Journal of Finance*, 48:1833–1863, 1993.

[10] L. Andersen and J. Andreasen. Jump-diffusion processes: volatility smile fitting and numerical methods for pricing. *Review of Derivatives Research*, 4:231–262, 2000.

[11] T.G. Andersen. Return volatility and trading volume: an information flow interpretation of stochastic volatility. *Journal of Finance*, 51:169–204, 1996.

[12] T.G. Andersen and T. Bollerslev. Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4:115–158, 1997.

[13] T.G. Andersen and T. Bollerslev. DM-Dollar volatility: intraday activity patterns, macroeconomic announcements and longer run dependencies. *Journal of Finance*, 53:219–265, 1998a.

[14] T.G. Andersen and T. Bollerslev. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39:885–905, 1998b.

[15] T.G. Andersen, T. Bollerslev, and A. Das. Variance-Ratio statistics and high frequency data: testing for changes in intraday volatility patterns. *Journal of Finance*, 56:305–327, 2001.

[16] T.G. Andersen, T. Bollerslev, and F.X. Diebold. Parametric and nonparametric measurements of volatility. In Y. Ait-Sahalia and L.P. Hansen, editors, *Handbook of Financial Econometrics*. 2003. Forthcoming.

[17] T.G. Andersen, T. Bollerslev, F.X. Diebold, and H. Ebens. The distribution of realized stock return volatility. *Journal of Financial Economics*, 61:43–76, 2001.

[18] T.G. Andersen, T. Bollerslev, F.X. Diebold, and P. Labys. The distribution of exchange rate volatility. *Journal of the American Statistical Association*, 96:42–55, 2001.

[19] T.G. Andersen, T. Bollerslev, F.X. Diebold, and P. Labys. Modeling and forecasting realized volatility. *Econometrica*, 71:579–626, 2003.

[20] T.G. Andersen, T. Bollerslev, and S. Lange. Forecasting financial market volatility: sample frequency vis-à-vis forecast horizon. *Journal of Empirical Finance*, 6:457–477, 1999.

[21] T.W. Anderson and D.A. Darling. Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. *Annals of Mathematical Statistics*, 23:193–212, 1952.

[22] E. Andreou and E. Ghysels. Rolling-sampling volatility estimators: some new theoretical, simulation and empirical results. *Journal of Business and Economic Statistics*, 20:363–376, 2002.

[23] T. Ané and H. Geman. Order flow, transaction clock and normality of asset returns. *Journal of Finance*, 55:2259–2284, 2000.

[24] T. Ané and C. Labidi. Implied volatility surfaces and market activity over time. *Journal of Economics and Finance*, 25:259–275, 2001.

[25] S.D. Aparicio and S.D. Hodges. Implied risk-neutral distributions: A comparison of estimation methods. FORC Preprint 98/95, University of Warwick, 1998.

[26] N. Areal and S.J. Taylor. The realised volatility of FTSE-100 futures prices. *Journal of Futures Markets*, 22:627–648, 2002.

[27] M. Avellaneda and Y. Zhu. An E-ARCH model for the term structure of implied volatility of FX options. *Applied Mathematical Finance*, 4:81–100, 1997.

[28] S.H. Babbs and K.B. Nowman. Kalman filtering of generalized Vasicek term structure models. *Journal of Financial and Quantitative Analysis*, 34:115–130, 2001.

[29] B. Bahra. Probability distributions of future asset prices implied by option prices. *Bank of England Quarterly Bulletin*, pages 299–311, August 1996.

[30] B. Bahra. Implied risk-neutral probability density functions from option prices: Theory and application. Working Paper, Bank of England, 1997.

[31] X. Bai, J. Russell, and G.C. Tiao. Beyond Merton's utopia: effects of non-normality and dependence on the precision of variance estimates using high frequency financial data. Working Paper, University of Chicago, 2001.

[32] R. Baillie and T. Bollerslev. The message in daily exchange rates: a conditional variance tale. *Journal of Business and Economic Statistics*, 7:297–305, 1989.

[33] R.T. Baillie. Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73:5–59, 1996.

[34] G. Bakshi, C. Cao, and Z. Chen. Empirical performance of alternative option pricing models. *Journal of Finance*, 52:2003–2049, 1997.

[35] G. Bakshi, N. Kapadia, and D.B. Madan. Stock return characteristics, skew laws, and differential pricing of individual equity options. *Review of Financial Studies*, 16:101–143, 2003.

[36] C. Ball and W. Torous. Unit roots and the estimation of interest rate dynamics. *Journal of Empirical Finance*, 3:215–238, 1996.

[37] C.A. Ball and W.N. Torous. A simplified jump process for common stock returns. *Journal of Financial and Quantitative Analysis*, 18:53–65, 1983.

[38] R.W. Banz and M.H. Miller. Prices for state-contingent claims: some estimates and applications. *Journal of Business*, 51:653–672, 1978.

[39] L. Baringhaus and N. Henze. A consistent test for multivariate normality based on the empirical characteristic function. *Metrika*, 35:339–348, 1988.

[40] S. Barle and N. Cakici. How to grow a smiling tree. *Journal of Financial Engineering*, 7:127–146, 1998.

[41] O.E. Barndorff-Nielsen. Exponentially decreasing distributions for the logarithm of particle size. *Proceedings Royal Society London*, A353:401–419, 1977.

[42] O.E. Barndorff-Nielsen. Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of Statistics*, 5:151–157, 1978.

[43] O.E. Barndorff-Nielsen. Normal Inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*, 24:1–13, 1997.

[44] O.E. Barndorff-Nielsen and W. Jiang. An initial analysis of some German stock price series. Working paper 15, Centre for Analytical Finance, 1998.

[45] O.E. Barndorff-Nielsen, E. Nicolato, and N. Shephard. Some recent developments in stochastic volatility modelling. *Quantitative Finance*, 2:11–23, 2002.

[46] O.E. Barndorff-Nielsen and N. Shephard. Non-Gaussian OU based models and some of their uses in financial economics. *Journal of the Royal Statistical Society B*, 63:167–241, 2001a.

[47] O.E. Barndorff-Nielsen and N. Shephard. Estimating quadratic variation using realised variance. *Journal of Applied Econometrics*, 17:457–477, 2001b.

[48] O.E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society B*, 64:253–286, 2002.

[49] O.E. Barndorff-Nielsen and N. Shephard. Realised power variation and stochastic volatility. *Bernoulli*, 9:243–265, 2003.

[50] G. Barone-Adesi, K. Giannopoulos, and L. Vosper. Backtesting derivative portfolios with Filtered Historical Simulation (FHS). *European Financial Management*, 8:31–58, 2002.

[51] G. Barone-Adesi and R.E. Whaley. Efficient analytical approximation of American option values. *Journal of Finance*, 42:301–320, 1987.

[52] K.S. Bartunek and M. Chowdhury. Implied risk aversion parameter from option prices. *The Financial Review*, 32:107–124, 1997.

[53] D. Bates. The crash of '87: Was it expected? the evidence from option markets. *Journal of Finance*, 46:1009–1044, 1991.

[54] D. Bates. Dollar-jump fears, 1984-1992: distributional abnormalities implicit in currency futures options. *Journal of International Money and Finance*, 15:65–93, 1996.

[55] D. Bates. Post-87 crash fears in the S&P500 futures options. *Journal of Econometrics*, 94:181–238, 2000.

[56] D. Beaglehole and A. Chebanier. Mean reversion with a smile. *Risk*, 15:95–98, 2002.

[57] A. Beltratti and C. Morana. Computing value-at-risk with high frequency data. *Journal of Empirical Finance*, 6:431–455, 1999.

[58] J. Berkowitz. Testing density forecasts with applications to risk management. *Journal of Business and Economic Statistics*, 19:465–474, 2001.

[59] M. Bilodeau and D. Brenner. *Theory of multivariate statistics.* Springer Verlag, 1999.

[60] M. Bilodeau and P. Lafaye de Micheaux. A multivariate empirical characteristic function test of independence with normal marginals. Working paper, University of Montreal, 2002.

[61] F. Black. Studies of stock price volatility changes. In *Proceedings of the 1976 American Statistical Association*, Alexandria, VA, 1976. Business and Economical Statistics Section, American Statistical Association.

[62] B. Blair, S. Poon, and S.J. Taylor. Forecasting S&P100 volatility: the incremental information content of implied volatilities and high frequency returns. *Journal of Econometrics*, 105:5–26, 2001.

[63] R. Bliss and N. Panigirtzoglou. Option-implied risk aversion estimates. *Journal of Finance*, February 2004. Forthcoming.

[64] T. Bollerslev. Generalised autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[65] T. Bollerslev and E. Ghysels. Periodic autoregressive conditional heteroscedasticity. *Journal of Business and Economic Statistics*, 14:139–151, 1996.

[66] T. Bollerslev and H. Zhou. Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics*, 109:33–65, 2002.

[67] T. Brailsford and R. Faff. An evaluation of volatility forecasting techniques. *Journal of Banking and Finance*, 20:419–438, 1996.

[68] D. Breeden and R. Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51:621–651, 1978.

[69] G.W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78:1–3, 1950.

[70] D. Brigo and F. Mercurio. A mixed up smile. *Risk*, 13:123–126, 2000.

[71] D. Brigo and F. Mercurio. Lognormal-mixture dynamics and calibration to market volatility smiles. *International Journal of Theoretical and Applied Finance*, 5:427–446, 2002.

[72] D. Brigo, F. Mercurio, and G. Sartorelli. Alternative asset-price dynamics and volatility smile. *Quantitative Finance*, 3:173–183, 2003.

[73] M. Britten-Jones and A. Neuberger. Option prices, implied price processes and stocastic volatility. *Journal of Finance*, 55:839–866, 2000.

[74] P. Brown and J.C. Jackwerth. The pricing kernel puzzle: reconciling index option data and economic theory. Working paper, University of Wisconsin, 2002.

[75] P.W. Buchen and M. Kelly. The maximum entropy distribution of an asset inferred from option prices. *Journal of Financial and Quantitative Analysis*, 31:143–159, 1996.

[76] E.M. Cabaña and A. Cabaña. Goodness-of-fit and two-sample comparison tests of the Kolmogorov-Smirnov type for bivariate populations. *Annals of Statistics*, 22:3–13, 1994.

[77] J. Campa and K. Chang. Arbitrage based tests of target zone credibility: evidence from ERM cross-rate options. *American Economic Review*, 86:726–740, 1996.

[78] J. Campa and K. Chang. ERM realignment risk and its economic determinants as reflected in cross-rate options. *Economic Journal*, 108:1046–1066, 1998.

[79] J. Campa, K. Chang, and J. Refalo. An option-based analysis of emerging market exchange rate expectations: Brazil's real plan, 1994-1999. Working paper, New York University, 1999.

[80] J. Campa, K. Chang, and R. Reider. Implied exchange rate distributions: Evidence from OTC option markets. *Journal of International Money and Finance*, 17:117–160, 1998.

[81] J.Y. Campbell, A.W. Lo, and A.C. MacKinlay. *The econometrics of financial markets*. Princeton University Press, Princeton, N.J., 1997.

[82] J. Cárdenas, E. Fruchard, E. Koehler, C. Michel, and I. Thomazeau. VAR: one step beyond. *Risk*, 10:72–75, 1997.

[83] P. Carr, H. Geman, D. Madan, and M. Yor. The fine structure of asset returns: and empirical investigation. *Journal of Business*, 75:305–332, 2002.

[84] P. Carr, H. Geman, D. Madan, and M. Yor. Stochastic volatility for Lévy processes. *Mathematical Finance*, 13:345–382, 2003.

[85] P. Carr and L. Wu. Time-changed Lévy processes and option pricing. *Journal of Financial Economics*, 2003. Forthcoming.

[86] S. Chib, F. Nardari, and N. Shephard. Markov Chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, 108:281–316, 2002.

[87] P.K. Clark. A subordinated stochastic process with finite variance for speculative prices. *Econometrica*, 41:135–156, 1973.

[88] M.P. Clements and J. Smith. Evaluating the forecast densities of linear and non-linear models: Applications to output growth and unemployment. *Journal of Forecasting*, 19:255–276, 2000.

[89] M.P. Clements and J. Smith. Evaluating multivariate forecast densities: A comparison of two approaches. *International Journal of Forecasting*, 18:397–407, 2002.

[90] G.M. Constantinides. Transaction costs and the volatility implied by option prices. Working Paper, University of Chicago, 1998.

[91] R. Cont and J. da Fonseca. Deformation of implied volatility surfaces: an empirical analysis. In Takayasu, editor, *Empirical Approaches to Financial Fluctuations*. Springer, Tokyo, 2001.

[92] R. Cont and J. da Fonseca. Dynamics of implied volatility surfaces. *Quantitative Finance*, 2:45–60, 2002.

[93] R. Cont, J. da Fonseca, and V. Durrleman. Stochastic models of implied volatility surfaces. *Economic Notes*, 31:361–377, 2002.

[94] R.M. Cooke. *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford University Press, Oxford, 1991.

[95] C. Corrado and T. Su. Skewness and kurtosis in S&P500 index returns implied by option prices. *Journal of Financial Research*, 19:175–192, 1996.

[96] C. Corrado and T. Su. Implied volatility skews and stock index skewness and kurtosis implied by the S&P500 index option prices. *Journal of Derivatives*, 4:8–19, 1997.

[97] G. Cortazar, E.S. Schwartz, and L. Naranjo. Term structure estimation in low-frequency transaction markets: a Kalman filter approach with incomplete panel-data. Working Paper, UCLA, 2003.

[98] S. Coutant. *Implied Risk Aversion in Option Prices.* PhD thesis, University of Paris IX Dauphine, 2001.

[99] S. Coutant, E. Jondeau, and M. Rockinger. Reading interest rate and bond futures options' smiles around the 1997 French snap election. *Journal of Banking and Finance*, 25:1957–1987, 2001.

[100] J.C. Cox and S.A. Ross. The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 3:145–166, 1976.

[101] S. Csörgő. Limit behaviour of the empirical characteristic function. *Annals of Probability*, 9:130–1444, 1981.

[102] S. Csörgő. Testing for independence by the empirical characteristic function. *Journal of Multivariate Analysis*, 16:290–299, 1985.

[103] M.M. Dacorogna, R. Gençay, U. Muller, R.B. Olsen, and O.V. Pictet. *An Introduction to High-Frequency Finance.* Academic Press: New York, 2001.

[104] R.B. D'Agostino and M.A. Stephens. *Goodness-of-Fit Techniques.* Marcel Dekker Inc., New York, 1986.

[105] S. Das and R. Sundaram. Of smiles and smirks: a term-structure perspective. *Journal of Financial and Quantitative Analysis*, 34:211–240, 1999.

[106] R.A. Davis and T. Mikosch. The sample autocorrelations of financial time series models. In A.T. Walden W.J. Fitzgerald, R.L. Smith and P.C. Young, editors, *Nonlinear and Nonstationary Signal Processing*, pages 247–274. Isaac Newton Institute and Cambridge University Press, Cambridge, 2000.

[107] A.P. Dawid. Statistical theory: the prequential approach. *Journal of the Royal Statistical Society, ser.A*, 147:278–292, 1984.

[108] E. Derman. Regimes of volatility. *Risk*, 12:55–59, 1999.

[109] E. Derman and M. Kamal. The patterns of change in implied index volatilities. Quantitative Strategies Research Notes, Goldman Sachs, 1997.

[110] E. Derman and I. Kani. Riding on a smile. *Risk*, (2):32–39, 1994a.

[111] E. Derman and I. Kani. The volatility smile and its implied tree. Quantitative Strategies Research Notes, Goldman Sachs, 1994b.

[112] E. Derman and I. Kani. Stochastic implied trees: Arbitrage pricing with stochastic term and strike structure of volatility. *International Journal of Theoretical and Applied Finance*, 1:7–22, 1998.

[113] E. Derman, I. Kani, and N. Chriss. Implied trinomial trees of the volatility smile. *Journal of Derivatives*, 3:7–22, 1996.

[114] F.X. Diebold, T.A. Gunther, and A.S. Tay. Evaluating density forecasts, with applications to financial risk management. *International Economic Review*, 39:863–883, 1998.

[115] F.X. Diebold, J. Hahn, and A.S. Tay. Multivariate density forecast evaluation and calibration in financial risk management: high-frequency returns on foreign exchange. *Review of Economics and Statistics*, 81:661–673, 2000.

[116] F.X. Diebold and J.A. Lopez. Forecast evaluation and combination. In G.S. Maddala and C.R. Rao, editors, *Handbook of Statistics*. North-Holland, Amsterdam, 1996.

[117] P. Dierckx. *Curve and Surface Fittings with Splines*. Oxford Science Publications, Clarendon Press, 1995.

[118] J. Doornik and H. Hansen. An omnibus test for univariate and multivariate normality. Working paper, Nuffield College, Oxford University, 1994.

[119] J.C. Duan and J.C. Simonato. Estimating exponential-affine term structure models by Kalman filter. *Review of Quantitative Finance and Accounting*, 13:111–135, 1999.

[120] D. Duffie, K. Singleton, and J. Pan. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68:1343–1376, 2000.

[121] B. Dumas, J. Fleming, and R. Whaley. Implied volatility functions: empirical tests. *Journal of Finance*, 53:2059–2106, 1998.

[122] B. Dupire. Pricing with a smile. *Risk*, 7:18–20, 1994.

[123] B. Dupire. Pricing and hedging with smiles. In M.A.H. Dempster and S.R. Pliska, editors, *Mathematics of Derivatives Securities*, pages 103–111. Cambridge University Press, Cambridge, 1997.

[124] G. Durham and R. Gallant. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business and Economic Statistics*, 20:279–316, 2002.

[125] H. Ebens. Realized stock volatility. Working Paper, John Hopkins University, 1999.

[126] E. Eberlein and K. Prause. The generalized hyperbolic model: financial derivatives and risk measures. In H. Geman, D. Madan, S. Pliska, and T. Vorst, editors, *Mathematical Finance - Bachelier Congress 2000*, pages 245–267. Springer Verlag, 2002.

[127] O. Elerian, S. Chib, and N. Shephard. Likelihood inference for discretely observed non-linear diffusions. *Econometrica*, 69:959–993, 2001.

[128] P. Embrechts, A. Hoeing, and A. Juri. Using copulae to bound the Value-at-Risk for functions of dependent risk. *Finance and Stochastics*, 7:145–167, 2003.

[129] P. Embrechts, F. Lindskog, and A. McNeil. Modelling dependence with copulas and applications to risk management. In S. Rachev, editor, *Handbook of heavy tailed distributions in finance*, chapter 8, pages 329–384. Elsevier, 2003.

[130] P. Embrechts, A. McNeil, and D. Straumann. Correlation and dependence in risk management: properties and pitfalls. In M.A.H. Dempster, editor, *Risk management: Value at Risk and beyond*, pages 176–223. Cambridge University Press, Cambridge, 2002.

[131] T.W. Epps and L.B. Pulley. A test for normality based on the empirical characteristic function. *Biometrika*, 70:723–726, 1983.

[132] B. Eraker. MCMC analysis of diffusion models with applications to finance. *Journal of Business and Economic Statistics*, 19:177–191, 2001.

[133] B. Eraker, M. Johannes, and N. Polson. The impact of jumps in volatility and returns. *Journal of Finance*, 58:1269–1300, 2003.

[134] P.L. Fackler and R.P. King. Calibration of option-based probability assessments in agricultural commodity markets. *American Journal of Agricultural Economics*, 72:73–83, 1990.

[135] Y. Fan. Goodness-of-fit tests for a multivariate distribution by the empirical characteristic function. *Journal of Multivariate Analysis*, 62:36–63, 1997.

[136] P.D. Feigin and C.R. Heatcote. The empirical characteristic function and the Cramér-von Mises statistic. *Sankhya Ser. A*, 38:309–325, 1976.

[137] M. Fengler, W. Härdle, and C. Villa. The dynamics of implied volatilities: a common principal component approach. Discussion Paper 38/01, Humboldt University, Berlin, 2001.

[138] A. Feuerverger. A consistent test for bivariate dependence. Unpublished manuscript, University of Toronto, 1993.

[139] A. Feuerverger and P. McDunnough. On some Fourier methods for inference. *Journal of American Statistical Association*, 76:379–386, 1981a.

[140] A. Feuerverger and P. McDunnough. On the efficiency of empirical characteristic function procedures. *Journal of the Royal Statistical Society B*, 43:147–156, 1981b.

[141] A. Feuerverger and R.A. Mureika. The empirical characteristic function and its applications. *Annals of Statistics*, 5:88–97, 1977.

[142] S. Figlewski. Options arbitrage in imperfect markets. *Journal of Finance*, 44:1289–1311, 1989a.

[143] S. Figlewski. What does an option pricing model tell us about option prices? *Financial Analysts Journal*, 45:12–15, 1989b.

[144] R.A. Fisher. Inverse probability. *Proceedings of the Cambridge Philosophical Society*, 36:528–535, 1930.

[145] R. Gallant, C.T. Hsu, and G. Tauchen. Using daily range data to calibrate volatility diffusions and extract the forward integrated variance. *The Review of Economics and Statistics*, 81:617–631, 1999.

[146] H. Geman, D. Madan, and M. Yor. Time changes for Lévy processes. *Mathematical Finance*, 11:79–96, 2001.

[147] G. Gemmill and A. Saflekos. How useful are implied distributions? Evidence from stock index options. *Journal of Derivatives*, 7:83–98, 2000.

[148] P. Giot. Time transformations, intraday data and volatility models. *Journal of Computational Finance*, 4:31–62, 2000.

[149] P. Glasserman, P. Heidelberger, and P. Shahabuddin. Efficient Monte Carlo methods for Value-at-Risk. In *Mastering Risk*, volume 2. Financial Times-Prentice Hall, 2001.

[150] C.W.J. Granger and S. Poon. Forecasting volatility in financial markets. *Journal of Economic Literature*, 41:478–539, 2003.

[151] A.K. Gupta, N. Henze, and B. Klar. Testing for affine equivalence of elliptically symmetric distributions. *Journal of Multivariate Analysis*, 2003. Forthcoming.

[152] W. Härdle and P. Schmidt. Common factors governing VDAX movements and the maximum loss. Discussion Paper 97/00, Humboldt University, Berlin, 2000.

[153] L. Harris. A transaction data study of weekly and intradaily patterns in stock returns. *Journal of Financial Economics*, 16:99–117, 1986.

[154] A.C. Harvey. *Forecasting, structural time series models and the Kalman filter.* Cambridge University Press, 1989.

[155] C.R. Heatcote. A test of goodness-of-fit for symmetric random variables. *Australian Journal of Statistics*, 14:172–181, 1972.

[156] N. Henze and T. Wagner. A new approach to the BHEP tests for multivariate normality. *Journal of Multivariate Analysis*, 62:1–23, 1997.

[157] N. Henze and B. Zirkler. A class of invariant consistent tests for multivariate normality. *Communications in Statistics. Theory and Methods*, 19:3595–3617, 1990.

[158] S. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–344, 1993.

[159] R. Heynen. An empirical investigation of observed smile patterns. *Review of Futures Markets*, 13:317–353, 1993.

[160] R. Heynen, A. Kemma, and T. Vorst. Analysis of the term structure of implied volatilities. *Journal of Financial and Quantitative Analysis*, 29:31–56, 1994.

[161] F.J. Hickernell. Goodness-of-fit statistics, discrepancies and robust designs. *Statistics and Probability Letters*, 44:73–78, 1999.

[162] E. Hol and S.J. Koopman. Stock index volatility forecasting with high frequency data. Tinbergen Institute Discussion Papers, 2002.

[163] D.A. Hsieh. Modelling heteroskedasticity in daily exchange rates. *Journal of Business and Economic Statistics*, 7:307–317, 1989.

[164] P. Huber. Projection pursuit (with discussion). *Annals of Statistics*, 13:435–535, 1985.

[165] J. Hull and A. White. The pricing of options on assets with stochastic volatility. *Journal of Finance*, 42:281–300, 1987.

[166] J.C. Hull and A. White. Value-at-risk when daily changes in market variables are not normally distributed. *Journal of Derivatives*, 5:9–19, 1998.

[167] J. Jackwerth and M. Rubinstein. Recovering probability distributions from option prices. *Journal of Finance*, 51:1611–1631, 1996.

[168] J.C. Jackwerth. Generalized binomial trees. *Journal of Derivatives*, 5:7–17, 1997.

[169] J.C. Jackwerth. Option-implied risk-neutral distributions and implied binomial trees: A literature review. *Journal of Derivatives*, 7:66–82, 1999.

[170] J.C. Jackwerth. Recovering risk aversion from option prices and realised returns. *Review of Financial Studies*, 13:433–451, 2000.

[171] R. Jarrow and A. Rudd. Approximate valuation for arbitrary stochastic processes. *Journal of Financial Economics*, 10:347–369, 1982.

[172] M. Johannes, N. Polson, and J. Stroud. Nonlinear filtering of stochastic differential equations with jumps. Working paper, Columbia University, 2002.

[173] E. Jondeau and M. Rockinger. Estimating Gram-Charlier expansions under positivity constraints. Working paper, HEC, 1998.

[174] E. Jondeau and M. Rockinger. Reading the smile: the message conveyed by methods which infer risk neutral densities. *Journal of International Money and Finance*, 19:885–915, 2000.

[175] A. Justel, D. Peña, and R. Zamar. A multivariate Kolmogorov-Smirnov test of goodness-of-fit. *Statistics and Probability Letters*, 35:251–259, 1997.

[176] I. Karatzas and S.E. Shreve. *Brownian motion and stochastic calculus.* Springer Verlag, New York, 1988.

[177] N.M. Kiefer. Discrete parameter variation: efficient estimation of a switching regression model. *Econometrica*, 46:427–434, 1978.

[178] S.A. Klugman and R. Parsa. Fitting bivariate loss distributions with copulas. *Insurance: Mathematics and Economics*, 24:139–148, 1999.

[179] A.N. Kolmogorov. Sulla determinazione empirica di una legge di distribuzione. *Giornale Istituto Italiano Attuari*, 4:83–91, 1933.

[180] S.J. Kon. Models of stock returns: a comparison. *Journal of Finance*, 39:147–165, 1984.

[181] S. Kotz and N.L. Johnson. *Encyclopedia of Statistical Sciences*. Wiley, New York, 1985.

[182] N.H. Kuiper. Tests concerning random points on a circle. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen*, 63:38–47, 1962.

[183] M. Leahy and C. Thomas. The sovereignty option: the Quebec referendum and market views on the Canadian dollar. Working paper, Federal Reserve Board, 1996.

[184] O. Ledoit and P. Santa-Clara. Relative pricing of options with stochastic volatility. Working paper, UCLA, 1998.

[185] J. Liang, K.T. Fang, and F.J. Hickernell. Testing multivariate uniformity and its applications. *Mathematics of Computation*, 70:337–355, 2001.

[186] X. Liu, M.B. Shackleton, S.J. Taylor, and X. Xu. Closed-form transformations from risk-neutral to real-world distributions. Working paper, Lancaster University, 2002.

[187] F. Longstaff. Option pricing and the martingale restriction. *Review of Financial Studies*, 8:1091–1124, 1995.

[188] G. Loudon, W. Watt, and P.K. Yadav. An empirical analysis of alternative parametric ARCH models. *Journal of Applied Econometrics*, 15:117–136, 2000.

[189] D.B. Madan and F. Milne. Contingent claims valued and hedged by pricing and investing in a basis. *Mathematical Finance*, 4:223–245, 1994.

[190] A.N. Madhavan. Market microstructure: a survey. *Journal of Financial Markets*, 3:205–258, 2000.

[191] A. Malz. Using option prices to estimate realignment probabilities in the European Monetary System: the case of Sterling-Mark. *Journal of International Money and Finance*, 15:717–748, 1996.

[192] A. Malz. Estimating the probability distribution of future excange rates from option prices. *Journal of Derivatives*, 5:18–36, 1997.

[193] A.M. Malz. Vega risk and the smile. *Journal of Risk*, 3, 2001.

[194] K.V. Mardia. Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57:519–530, 1970.

[195] M. Martens. Forecasting daily exchange rate volatility using intraday returns. *Journal of International Money and Finance*, 20:1–23, 2001.

[196] M. Martens, Y.C. Chang, and S.J. Taylor. A comparison of seasonal adjustment methods when forecasting intraday volatility. *The Journal of Financial Research*, 25:283–299, 2002.

[197] R. McCauley and W. Melick. Risk reversal risk. *Risk*, 9:54–57, 1996.

[198] M. McIntyre. Performance of Dupire's implied diffusion approach under sparse and incomplete data. *Journal of Computational Finance*, 4:33–84, 2001.

[199] D.J. McManus. The information content of interest rate futures options. Working paper, Bank of Canada, 1999.

[200] N. Meddahi. A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics*, 17:479–508, 2002.

[201] W. Melick and C. Thomas. Recovering an asset's implied PDF from option prices: an application to crude oil during the gulf crisis. *Journal of Financial and Quantitative Analysis*, 32:91–115, 1997.

[202] B. Mizrach. Did option prices predict the ERM crisis? Working paper, Rutgers University, 1996.

[203] I. Monroe. Processes that can be embedded in Brownian motion. *Annals of Probability*, 6:42–56, 1978.

[204] H. Nagao. On some test criteria for covariance matrix. *Annals of Statistics*, 1:700–709, 1973.

[205] S. Natenberg. *Option Volatility and Pricing: Advanced Trading Strategies and Techniques.* Probus Publishing, Chicago, 1994.

[206] R.B. Nelsen. *An introduction to copulas.* Lecture Notes in Statistics 139. Springer Verlag, 1998.

[207] D. B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59:347–370, 1991.

[208] P. Noceti, J. Smith, and S.D. Hodges. An evaluation of tests of distributional forecasts. *Journal of Forecasting*, 2003. Forthcoming.

[209] R. Oomen. Statistical models for high frequency security prices. Working paper, University of Warwick, 2002.

[210] J. Pan. The jump-risk premia implicit in options: evidence from an integrated time series study. *Journal of Financial Economics*, 63:3–50, 2002.

[211] I. Peña, G. Rubio, and G. Serna. Why do we smile? On the determinants of the implied volatility function. *Journal of Banking and Finance*, 23:1151–1179, 1999.

[212] M.K. Pitt and N. Shephard. Filtering via simulation: auxiliary particle filter. *Journal of the American Statistical Association*, 94:590–599, 1999.

[213] S. Posner and M. Milevsky. Valuing exotic options by approximating the SPD with higher moments. *Journal of Financial Engineering*, 7, 1998.

[214] K. Prause. Modelling financial data using generalised hyperbolic distributions. FDM Preprint 48, University of Freiburg, 1997.

[215] R. Rebonato and M. Joshi. Assigning future smile surfaces: Conditions for uniqueness and absence of arbitrage. QUARC Working Paper, Royal Bank of Scotland, 2003.

[216] R.J. Ritchey. Call option valuation for discrete normal mixtures. *Journal of Financial Research*, 13:285–296, 1990.

[217] L. Rogers and O. Zane. Designing and estimating models of high-frequency data. Working paper, University of Bath, 1998.

[218] R. Roll. A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, 39:1127–1140, 1984.

[219] C. Rookley. Fully exploiting the information content of intra day option quotes: Applications in option pricing and risk management. Working paper, University of Arizona, 1997.

[220] J. Rosenberg and R. Engle. Option hedging using empirical pricing kernels. Working paper, University of California, San Diego, 1997.

[221] J. Rosenberg and R. Engle. Empirical pricing kernels. *Journal of Financial Economics*, 64:341–372, 2002.

[222] J.V. Rosenberg. Implied volatility functions: a reprise. *Journal of Derivatives*, 7:51–64, 2000.

[223] M. Rosenblatt. Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23:470–472, 1952.

[224] M. Rubinstein. Non-parametric tests of alternative option pricing models. *Journal of Finance*, 40:455–480, 1985.

[225] M. Rubinstein. Implied binomial trees. *Journal of Finance*, 49:771–818, 1994.

[226] M. Rubinstein. Edgeworth binomial trees. *Journal of Derivatives*, 5:20–27, 1998.

[227] T.H. Rydberg. The Normal Inverse Gaussian Lévy process: simulation and approximation. *Communications in Statistics: Stochastic Methods*, 13:887–910, 1997.

[228] T.H. Rydberg and N. Shephard. A modelling framework for the prices and times of trades made on the NYSE. In A.T. Walden W.J. Fitzgerald, R.L. Smith and P.C. Young, editors, *Nonlinear and Nonstationary Signal Processing*, pages 217–246. Isaac Newton Institute and Cambridge University Press, Cambridge, 2000.

[229] T.H. Rydberg and N. Shephard. Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics*, 1:2–25, 2003.

[230] P.J. Schönbucher. A market model for stochastic implied volatility. *Philosophical Transactions of the Royal Society, Ser. A*, 357:2071–2092, 1999.

[231] L. Scott. Option pricing when the variance changes randomly: theory, estimation and an application. *Journal of Financial and Quantitative Analysis*, 22:419–437, 1987.

[232] F. Seillier-Moiseiwitsch and A.P. Dawid. On testing the validity of sequential probability forecasts. *Journal of the American Statistical Association*, 88:355–359, 1993.

[233] A. Sheik. Transaction data tests of S&P 100 call option pricing. *Journal of Financial and Quantitative Analysis*, 26:459–475, 1991.

[234] B. Sherrick, S. Irwin, and D. Forster. Option-based evidence of the nonstationarity of expected S&P500 futures price distributions. *Journal of Futures Markets*, 12:275–290, 1992.

[235] B. Sherrick, S. Irwin, and D. Forster. An examination of option-implied S&P500 futures price distributions. *Financial Review*, 31:667–694, 1996.

[236] B.J. Sherrick, P. Garcia, and V. Tirupattur. Recovering probabilistic information from options markets: Tests of distributional assumptions. *Journal of Futures Markets*, 16:545–560, 1996.

[237] D. Shimko. Bounds of probability. *Risk*, 6:33–37, 1993.

[238] E. Silva and K. Kahl. Reliability of soybean and corn option-based probability assessments. *Journal of Futures Markets*, 13:765–779, 1993.

[239] K. Singleton. Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics*, 102:111–141, 2001.

[240] G. Skiadopoulos, S.D. Hodges, and L. Clewlow. Dynamics of the S&P500 implied volatility surface. *Review of Derivatives Research*, 3:263–282, 1999.

[241] P. Söderlind. Market expectations in the UK before and after the ERM crisis. *Economica*, 67:1–18, 2000.

[242] P. Söderlind and L. Svensson. New techniques to extract market expectations from financial instruments. *Journal of Monetary Economics*, 40:383–429, 1997.

[243] D. Sornette, P. Simonetti, and J.V. Andersen. "nonlinear" covariance matrix and portfolio theory for non-Gaussian multivariate distributions. Working paper, UCLA, 1999.

[244] M. Stutzer. A simple nonparametric approach to derivative security valuation. *Journal of Finance*, 51:1633–1652, 1996.

[245] A. Sylla and C. Villa. Measuring implied volatility surface risk using principal component analysis. In *Measuring Risk in Complex Stochastic Systems*, number 147 in Lecture Notes in Statistics. Springer Verlag, 2000.

[246] A.S. Tay and K.F. Wallis. Density forecasting: a survey. *Journal of Forecasting*, 19:235–254, 2000.

[247] S.J. Taylor and X. Xu. The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance*, 4:317–340, 1997.

[248] R.G. Tompkins. Implied volatility surfaces: uncovering regularities for options on financial futures. *The European Journal of Finance*, 7:198–230, 2001.

[249] S. Venkataraman. Value-at-risk for a mixture of normal distributions: the use of quasi-Bayesian estimation techniques. *Economic Perspective*, pages 2–13, March-April 1997. Federal Reserve Bank of Chicago.

[250] G.S. Watson. Goodness-of-fit tests on a circle. *Biometrika*, 48:109–114, 1961.

[251] S.A. Weinberg. Interpreting the volatility smile: an examination of the informational content of option prices. International Finance Discussion Papers, No. 706, Federal Reserve Board, Washington D.C., 2001.

[252] R.A. Wood, T.H. McInish, and J.K. Ord. An investigation of transaction data for NYSE stocks. *Journal of Finance*, 25:723–739, 1985.

[253] P. Zangari. An improved methodology for measuring VaR. Riskmetrics Monitor, 1996.

[254] L.X. Zhu, K.T. Fang, and M.I. Bhatti. On estimated Projection Pursuit type Cramér-von Mises statistics. *Journal of Multivariate Analysis*, 63:1–14, 1997.

[255] L.X. Zhu, K.T. Fang, and J.T. Zhang. A projection NT-type test for spherical symmetry of a multivariate distribution. In E.M. Tiit, T. Kollo, and H. Niemi, editors, *New trends in probability and statistics: multivariate statistics and matrices in statistics*, volume 3, pages 109–122. NSP Utrecht, 1995.