H4

MODELLING THE DYNAMICS OF IMPLIED VOLATILITY SMILES AND SURFACES

Georgios Skiadopoulos

Thesis submitted in fulfilment of the requirements for the Degree of

Doctor of Philosophy in Finance

University of Warwick,

Warwick Business School

July 1999

Contents

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I gratefully acknowledge the guidance, patience, time, and support of my supervisor Professor Stewart Hodges on writing this thesis. He introduced me to the area of options in the Finance literature, and he showed to me what is the meaning of being at the cutting edge of research. His unique insights into the world of finance have been immeasurably influential in this thesis, as well as in my development as a student of the financial markets.

I am grateful to Dr Les Clewlow who taught me some of the secrets of computer programming and shared part of his expertise with me. I am also grateful to Dr Abhay Abhyankar, for reading carefully and commenting on several parts of this study, and to Wojtek Krzanowski, Joao Pedro Nunes, Juan Carlos Mehia Perez, and Alessandro Rossi for the long discussions that helped me clarifying certain ambiguities. Further, I have benefited from the comments of David Bates, Peter Honore, Jens Jackwerth, Kostas Milas, Pablo Nocetti, Michael Rockinger, Chris Strickland, Norman Strong, Robert Tompkins, Mark Wong, participants at the various conferences and seminars where parts of this study were presented, and of two anonymous referees of the REDR.

I owe special thanks to my close friends, to Shelley Deane, Yianni Sardi, and to George and Joyce Mouza for encouraging and advising me at difficult moments. To my late close friend Niko Siakantari, I want him to know that I will always remember him. To all of my teachers from the primary school until now, thank you for contributing to my education.

This thesis and effort is dedicated to my father, mother, grandmother and brother. It started and it has been accomplished because of the way that they have brought me up. My parents have always believed in the power of knowledge and they have sacrificed many things from themselves, in order to educate their sons. My family taught me not to give up, to be modest, and to keep my dignity and kindness in difficult circumstances. It has respected my decisions, and it has supported me continuously. Dedicating this last degree to them, is the least that I can do to express my love.

DECLARATION

Three papers have been produced based on the results from this thesis. All of them are co-authored with Stewart Hodges and Les Clewlow. Two more papers are going to be prepared for publication. A paper based upon Chapters 3 and 4 has been presented at the 1998 European and French Finance Associations meetings. It has appeared at the Financial Options Research Centre's Pre-Print series, and the *Proceedings of the French Finance Association Meetings, Lille 1998*, with the title "The Dynamics of Smiles". A paper based upon Chapters 4 and 5 has been presented at the 1998 Financial Options Research Centre's Annual Conference, and it will be presented at the 1999 Decision Sciences Institute International Conference. It has appeared at the Financial Options Research Centre's Pre-Print series, and the *Proceedings of the Decision Sciences Institute, Athens 1999* (forthcoming) with the title "The Dynamics of Implied Volatility Surfaces". The third paper based on Chapters 3, 4, and 5 has been presented at the 1999 Boston University's Derivatives Securities Conference, and it is forthcoming in the *Review of Derivatives Research*. The two new papers are based on Chapters 2 and 6, respectively (the second is co-authored with Stewart Hodges).

SUMMARY

"Smile-consistent" no-arbitrage stochastic volatility models take today's option prices as given, and they let them to evolve stochastically in such a way as to preclude arbitrage. This allows standard options to be priced correctly, and enables exotic options to be valued and hedged relative to them. We study how to model the dynamics of implied volatilities, since this is a necessary prerequisite for the implementation of these models. First, we investigate the number and shape of shocks that move implied volatility smiles, by applying Principal Components Analysis. The technique is applied to two different metrics: the strike, and the moneyness. Three distinct criteria are used to determine the number of components to retain. Subsequently, we construct a "Procrustes" type rotation in order to interpret them. Second, we use the same methodology to identify the number and shape of shocks that move implied volatility surfaces. In both cases, we find that the number of shocks is the same (two), in both metrics. Their interpretation is a shift for the first one, and a Z-shaped for the second. The results have implications for both option pricing and hedging, and for the economics of option pricing. Finally, we propose a new and general method for constructing a "smile-consistent" no-arbitrage stochastic volatility model: the simulation of the implied risk-neutral distribution. An algorithm for the simulation is developed when the first two moments change over time. It can be implemented easily, and it is based on the idea of mixture of distributions. It can also be generalized to cases where more complicated forms for the mixture are assumed.

ABBREVIATIONS

BS:	Black-Scholes
BJ:	Bera-Jarque
BAW:	Barone-Adesi, Whaley
CME:	Chicago Mercantile Exchange
CDF	Cumulative Distribution Function
HJM:	Heath, Jarrow, Morton
IFV	Instantaneous Forward Volatility
ITM:	In-the-Money
OTC:	Over-the-Counter
OTM:	Out-of-the-Money
NR:	Newton-Raphson
OLS:	Ordinary Least Squares
PCA:	Principal Components Analysis
PC:	Principal Component
PDE:	Partial Differential Equation
PDF:	Probability Density Function
S&P 500:	Standard and Poor 500 Index

Chapter 1

Introduction

The growing literature on "smile consistent" no-arbitrage stochastic volatility models (Dupire [59], [61], Derman and Kani [54], Ledoit and Santa-Clara [106]) has been motivated by the need to price and hedge vanilla and exotic options consistently. The first objective of this thesis is to model the dynamics of implied volatility and smiles, since this is a prerequisite for the implementation of these models. The second objective, is to provide a new and general method for developing a "smile consistent" stochastic volatility model.

The Black-Scholes model [20] is widely used to price and hedge standard and exotic options. Its popularity among practitioners arises from its tractability. However, the empirical evidence (see among others Gemmill [73], Jackwerth and Rubinstein [94], Rubinstein ([125], [126]), Derman and Kani [54]) contradicts its prediction of a constant implied volatility. Implied volatilities vary across: (a) different *strikes* for options at the same point in time with the same time-to-expiration (smiles or skews), (b) different *times-to-expiration* for options at the same point in time and the same strike price (term structure), (c) different *points in time* for options with the same expiration date and the same ratio of strike price to the underlying asset price (dynamics of implied volatilities). These results suggest that the implied volatilities of options with different strikes and expirations, form a two-dimensional surface which has certain dynamics.

The evolution of the implied volatility surface undermines the use of the Black-

Scholes formula as a model for the correct pricing and hedging of standard, and especially, of exotic options (see for example, Davydov and Linetsky [48]). In order to cope with this issue, a number of option pricing models have been proposed which give rise to smiles or skews, and to a term structure of implied volatilities, roughly similar to what is observed empirically. These models provide for stochastic volatility (see Hull and White [90], Johnson and Shanno [101], Scott [131], Wiggins [147]), or jump models (see Bates [10] Merton [113]), or both (see Bates [11], [12], Scott [132]). However, none of these models fit observed implied volatility patterns well (see Clewlow and Xu [36], Das and Sundaram [47], Taylor and Xu [141]).

These problems have led to the recent literature on "smile consistent" no-arbitrage stochastic volatility models. Rather than specifying the underlying asset's process in advance, they use the European market option prices to infer information about the underlying asset process. They do this by taking today's standard option prices as given, and letting them evolve stochastically in such a way as to preclude arbitrage. This ensures the correct pricing of standard options, and is relevant to the pricing of exotic options (see Carr, Ellis, and Gupta [31], Davydov and Linetsky [48], Derman, Ergener, and Kani [50]).

In order to implement this type of models, we need to understand the dynamics of the implied volatility surface. In this study, we first investigate the dynamics of implied volatility smiles, and then we investigate those of the more complex implied volatility surface. We answer the three questions related to this: (1) how many factors are needed to explain the dynamics of the implied volatility smiles and surfaces?, (2) what do these factors look like?, and (3) how are these factors correlated with the innovation in the underlying asset's process? The results from the implied volatility smiles analysis, are used as a check of the robustness of those from the surface analysis.

The technique that we use in order to answer the three questions is Principal Components Analysis (PCA). It is applied to the changes in implied volatilities over time. The changes in implied volatilities are indexed in two different ways

(metrics): the strike level, and the moneyness level. This is because a deterministic volatility model predicts that the dynamics of implied volatilities should be measured under the strike metric (Derman, Kani and Zou [51]). On the other hand, if a stochastic volatility model is the correct model, then the moneyness metric is appropriate (Taylor and Xu [140], [141]). In addition, we group the data in distinct ranges of days to expiry, so as to control for the time to expiry; various models (such as Stein [137]), and studies (Bates and Clewlow [16], Hsieh [89], and Taylor and Xu [149]) have documented that the variation of the implied volatilities is a function of the time to expiry. The dynamics of individual smiles are analyzed by applying PCA separately to the expiry buckets. We then analyze the dynamics of the whole implied volatility surface by pooling the buckets together, and applying the PCA to the whole data set.

The empirical results are reported for daily data on futures options on the Standard and Poor 500 index from the Chicago Mercantile Exchange for the years 1992-95. First, we screen the data carefully for any errors and noise. Then, we use a variety of criteria, in order to decide on the number of shocks which drive the implied volatilities dynamics. Two factors are identified. These explain on average 53% of the variance of the implied volatility surface in the strike metric, and 60% of the variance in the moneyness metric. Subsequently, we construct a "Procrustes" rotation, in order to interpret the retained components. The first factor is interpreted as a parallel shift, and the second has a Z-shape. The results are similar in both the smile, and the surface analysis, under both metrics. Moreover, they are remarkably consistent across years. We conclude that to implement a "smile-consistent" no-arbitrage stochastic volatility model for the pricing and hedging of futures options on the S&P 500, we need three factors. One is required for the underlying asset, and two more for the implied volatility.

Finally, we propose a new and general method for constructing a "smile-consistent" no-arbitrage stochastic volatility model. The method simulates the evolution through time of the implied risk-neutral distribution, starting from

today's implied distribution. The simulation of the implied distribution is a natural tool for option pricing and hedging within this framework of models, because of the relation of the mean of the risk-neutral distribution with the current asset price. However, even though the way to extract the risk-neutral distribution from European option prices has been studied extensively (see for a survey Bahra [5], and Mayhew [111]), no research, to our knowledge, has been undertaken on its simulation, so far.

We present the method for the simulation, when the first two moments of the distribution change over time. It is constructed by considering a mixture of distributions as a model for today's implied distribution. Then, a mapping between today's and tomorrow's cumulative probability, for a given asset value, is established. Our algorithm can be implemented easily, and it can be extended to the cases where complex forms for the mixture are assumed. It can be used for pricing purposes and for assessing the performance of hedges. On the other hand, it can not be used for the valuation of American type products.

The thesis is organized as follows. We first review the developing literature on "smile-consistent" no-arbitrage stochastic volatility models, in Chapter 2. The historical development of this literature is narrated, and we provide a taxonomy of the various models. The main ideas behind them are highlighted, and their advantages and limitations are outlined. Practical issues in implementing the models are also discussed. The data, and the criteria we use to reduce the noise, are described in Chapter 3. In Chapters 4 and 5, we analyze the dynamics of implied volatility smiles, and surfaces, respectively. The simulation of the implied risk-neutral distribution is presented in Chapter 6. Chapter 7 concludes, and suggests issues for future research.

Chapter 2

Literature Review

2.1 Introduction

Exotic options are often hedged with European options. This hedging strategy is called static replication (see Carr, Ellis, and Gupta [31], Derman, Ergener, and Kani [50]). To improve the hedging performance, exotic and standard options need to be valued consistently. This is done by first assuming a stochastic process which describes the underlying asset price dynamics. The process is calibrated to the observed market prices of exchange-traded options. Then, the resulting (implied) process is used to price over the counter options. The need to price and hedge exotic options consistently with the prices of standard European options has led to the quite recent literature of "smile-consistent" no-arbitrage models. The aim of this chapter is to survey concisely this developing literature which is of particular importance to both academics and practitioners.

The behavior of implied volatilities derived from inverting Black-Scholes [20] (BS) model, questions the validity of the model. The empirical evidence shows (see among others Derman and Kani [54], Rubinstein [125], [126] and for an extensive literature survey Mayhew [110]) that implied volatilities vary across different strikes (smiles or skews), and different times-to-expiration (term structure), while the BS model does not predict any such variation. These results suggest that implied volatilities could be viewed as a two-dimensional non-flat surface.

In addition, the functional form of this surface changes over time (Gemmill [73], Jackwerth and Rubinstein [94]).

The non-flat implied volatility surface is roughly explained by either stochastic volatility (see Hull and White [90], Johnson and Shanno [101], Scott [131], Wiggins [147]), or jump models (see Bates [10] Merton [113]), or both (see Bates [11], [12], Scott [132]). The approach taken by these models consists of specifying the parameters of the processes for the underlying traded and non-traded securities (stochastic volatility and, or jump). The market price of risk of the non-traded sources of risk has also to be specified[1]. Then, option prices are derived as a function of the parameters of the processes and the prices of the underlying securities.

However, these models do not fit observed implied volatility patterns well (see Clewlow and Xu [36], Das and Sundaram [47], Taylor and Xu [141]), making it difficult to use them in practice to price and hedge exotic options. These problems have motivated the recent literature on "smile consistent" no-arbitrage models. "Smile-consistent" models reverse the approach taken by the conventional stochastic volatility, or jump models. The prices of standard European options are taken as given, and they are used to infer information about the underlying price processes.

In this chapter, we survey the "smile-consistent" no-arbitrage literature by classifying the two stages through which it has been developed. First, deterministic volatility models which fit the observed European option prices were introduced (Andersen [2], Andreasen [4], Barle and Cakici [7], Derman and Kani [49], Derman, Kani and Chriss [53], Dupire [60], [62], Jackwerth [96], Rubinstein [126]). Next, stochastic volatility models were provided which allowed for smile-consistent option pricing under the no-arbitrage evolution of the volatility surface (Britten-Jones and Neuberger [27], Derman and Kani [54], Dupire [59], [61], [63],

[1]The market price of risk is specified by invoking equilibrium arguments. In this sense these models allow for equilibrium, and not for arbitrage pricing. For an intuitive, clarification of the difference between equilibrium and arbitrage pricing, see Dupire [61].

Ledoit and Santa-Clara [106]). The second class of models is more general and it nests the first class. The various models are developed in continuous, or discrete time, or both. We describe them by commenting on the key ideas behind them and we outline their advantages and limitations. Furthermore, some practical issues in implementing these models are addressed.

The remainder of the chapter is structured as follows. In the second and third section, we discuss the smile consistent deterministic volatility models in continuous and discrete time, respectively. We compare them, and we discuss some practical issues in implementing them. The empirical results from the research on the validity of the smile-consistent no-arbitrage deterministic volatility models are presented in section four. Section five describes the smile-consistent no-arbitrage stochastic volatility models, and brings together the different definitions of the key concept of the forward variance. The investigation of the dynamics of volatilities, as a prerequisite for the implementation of this class of models, is also pointed out. The last section concludes.

2.2 Smile Consistent Deterministic Volatility Models in Continuous Time

Option prices calculated from the BS model deviate from the market option prices, especially after the market crash on October 19, 1987. An alternative way of stating this is by describing the stylized characteristics of implied volatilities (see for instance Jackwerth and Rubinstein [94], Rubinstein [126]). Depending on the underlying asset and the sample period under scrutiny, implied volatilities have a term structure which is upward, or downward sloping (Derman and Kani [49]). In addition, in contrast to the BS prediction of a constant implied volatility across strikes, there is an implied volatility bias; they vary across strikes giving rise to smiles, or skews. Black [18] finds that implied volatilities decline as the strike price goes down; Macbeth and Merville [108] find that call options implied volatilities tend to be higher when the strike price declines. Rubinstein [125] finds

that the implied volatility bias changes direction depending on the sample period under scrutiny. Shastri and Wethyavivorn [133] find an implied volatility smile (see also Bates [12] for an extensive survey).

A simple way of explaining the implied volatility skew which appears in Index and Futures options markets (Bates [14]), is by resorting to standard deterministic volatility models (see Cox [41], Cox and Ross [42], Emanuel and MacBeth [65], MacBeth and Merville [109]) which allow for an inverse relationship between the price of the underlying security, and the variance of the rate of return. These models specify exogenously the instantaneous volatility σ as a deterministic function of the price of the underlying asset S_t and time t, i.e.

$$\frac{dS_t}{S_t} = a(S_t, t)dt + \sigma(S_t, t)dW_t \tag{2.1}$$

In contrast to standard deterministic volatility models, smile consistent deterministic volatility models do not specify $\sigma(S_t, t)$ in advance, but endogenously from the European option prices. Therefore, they preserve the "pricing by no-arbitrage" property of the BS model, and the option's payoff can be synthesized from a portfolio of existing assets, i.e. the markets are complete (see Dothan [57], and for a concise description Sundaram [139]). In addition, they provide us with a method for specifying $\sigma(S_t, t)$ from the market option prices, i.e. they deliver to us an implied process.

The knowledge of the process allows for the pricing and hedging of path-dependent options (Monte-Carlo methods) and American options (by dynamic programming). The hedging will be effective throughout the life of the option if the asset price behaves according to the inferred process.

2.2.1 Theoretical Justifications for "Smile-Consistent" Deterministic Volatility Models

There are three ways, to our knowledge, of explaining the systematic relationship of volatility with the underlying asset. A first informal approach is the one

suggested by Derman, Kani and Zou [51]. They call $\sigma(S_t, t)$ as the local volatility. It is the volatility which prevails at the asset level S_t at time t. They think of the implied volatility as an average of local volatilities across the state space (rather than the time domain). Assuming that the local volatility varies linearly with the asset price, they show that the local volatility varies with the asset level about twice as rapidly, as implied volatility varies with the strike. Therefore, according to them, the smile can be explained by the variation of local volatility with the asset price and time, and other effects such as stochastic volatility and jumps are less important.

The second approach, uses the negative correlation between $\sigma(S_t, t)$ and the asset price which was first observed by Black [18]. This negative relationship can be explained either as a leverage effect (Christie [33]), or by the portfolio insurance strategies that investors use (Grossman and Zhou [77]). The idea behind Christie's model is that as the asset price increases, the riskiness of the outstanding debt of a levered firm decreases, and hence the volatility declines. The intuition in Grossman and Zhou is that the use of an insurance strategy increases the sensitivity of an individual risk aversion to changes in his wealth. As a result, when the price falls, the risk aversion of the agents increases and this increases volatility. Similarly, when the price rises, their risk aversion decreases, and this decreases volatility.

The third approach, invokes Platen and Schweizer [129] model's which is similar in spirit to Grossman and Zhou's. They start from a microeconomic equilibrium model, where part of the demand for the underlying asset is induced by a hedging strategy. The limit of their model is a deterministic volatility diffusion, where the volatility coefficient is derived endogenously from assumptions about agents' trading behavior.

2.2.2 Dupire (1993 and 1994)

Let $C(K, T)$ be a European call option of exercise (strike) price K and maturity T. Assume that the continuum of all $(C(K, T))_{K,T}$ are traded and that their

prices today are consistent with no arbitrage. Breeden and Litzenberger [25] have shown that the observed European call option prices deliver to us the conditional terminal risk-neutral density as a function of K, i.e.

$$\Phi_T(K) = e^{-r(T-t)} \frac{\partial^2 C(K,T)}{\partial K^2} \tag{2.2}$$

where $\Phi_T(K)$ is the terminal risk-neutral density of S_T conditional on the information at current time t, and r is the interest rate. In general, the converse is not true. From the terminal implied risk-neutral density we can not recover uniquely the asset process which generates today's option prices (see Dupire [60], Melick and Thomas [112]). However, Dupire proves that there is an exception. Under some technical regularity conditions, we can recover a *unique* diffusion process from the terminal risk-neutral implied density, if we restrict ourselves to risk-neutral deterministic volatility diffusions (see Dupire [60], [62]). This is proved by means of the forward Kolmogorov equation. Given the process

$$dx = a(x,t)dt + b(x,t)dW \tag{2.3}$$

the forward Kolmogorov equation is given by

$$\frac{1}{2}\frac{\partial^2(b^2 f)}{\partial x^2} - \frac{\partial(af)}{\partial x} = \frac{\partial f}{\partial T} \tag{2.4}$$

where $f(x,T) \equiv \Phi_T(x)$. In other words, in general the forward Kolmogorov equation is used for deriving conditional densities (or distributions) starting from a given process[2]. However, in our case we cope with the converse problem: *f is known and b is the unknown.*

Restricting ourselves to risk-neutral densities (and assuming without loss of

[2]For a further discussion on the forward Kolmogorov equation and its use, see Cox and Miller [40].

generality that the interest rate is zero) equation (2.4) becomes[3]

$$\frac{1}{2}\frac{\partial^2(b^2 f)}{\partial x^2} = \frac{\partial f}{\partial T} \tag{2.5}$$

As f can be written as $\frac{\partial^2 C}{\partial x^2}$ (x denotes the strike price), he shows that (Dupire [60])

$$\frac{1}{2}b^2\frac{\partial^2 C}{\partial x^2} = \frac{\partial C}{\partial T} \tag{2.6}$$

Both derivatives are positive by arbitrage. Hence[4],

$$b(x,T) = \sqrt{\frac{2\frac{\partial C(x,T)}{\partial T}}{\frac{\partial^2 C(x,T)}{\partial x^2}}} \tag{2.7}$$

Combining equations (2.1) and (2.3), we obtain the instantaneous volatility by $\sigma(S,T) = \frac{b(S,T)}{S}$.

The forward equation (2.6) presents the option pricing problem in a different way than the BS partial differential equation (PDE) does. It turns the option pricing problem into a problem in strikes and maturities with fixed spot and time, rather than in a problem in spot and time with fixed strike and maturity. Andreasen [4] shows that in general, the forward equations for the option prices imply a duality: the problem of pricing and hedging of European options can be solved in a dual economy, where the spot is the strike, the strike is the spot, the call is the put, the interest rate is the dividend yield, and the dividend yield is the interest rate. However, the BS PDE applies to any contingent claim, while

[3]By restricting himself to the risk-neutral environment, he has one equation (the forward equation) and one unknown (the instantaneous volatility). Otherwise, he would have two unknowns, the drift and the instantaneous volatility and he would not be able to determine uniquely the instantaneous volatility.

[4]Equation (2.7) holds for every system of call prices, provided that the time derivative of the European call vanishes as the strike goes to infinity and a slow growth condition is satisfied (see Dupire [60]). For instance, the BS model for a European call on a no dividend asset with interest rate equal to zero, satisfies (2.7) (see Bick and Reisman [22]). Bick and Reisman also derive equation (2.7), independently. However, they make no distinction between the concept of local and implied volatility.

equation (2.6) holds only because the intrinsic value of a call happens to be the second integral of a Dirac function.

2.3 Smile Consistent Deterministic Volatility Models in Discrete Time

The implementation of a smile-consistent deterministic volatility model, for pricing and hedging purposes, is done in a discrete time framework. The tools used are either binomial, or trinomial implied trees, or implicit finite difference schemes. The former discretizes the asset price process, while the latter discretizes the BS type fundamental PDE (see Geske and Shastri [75]).

Binomial (or trinomial) trees are built from the known prices of European options. Such trees are called *implied trees* because they are consistent with or implied by the volatility smile. Their continuous time limit is a deterministic volatility process (see Nelson and Ramaswamy [115]). In the standard Cox, Ross, Rubinstein [43] (CRR) tree the size of the upper and down move of the underlying asset, and the respective probabilities of such moves are constant (because they depend on the volatility which is assumed to be constant). This is not any longer the case with implied trees.

In general, in order to construct a tree we need to know the way that the underlying asset price evolves and the transition probabilities corresponding to the links of the tree. The traditional way of calculating these two, is by establishing conditions under which a sequence of processes converges in distribution to the given diffusion (Nelson and Ramaswamy [115])[5]. In the case of implied trees, there is also the additional constraint that they must correctly reproduce the volatility smile. Once the tree has been built, backward induction (see Cox, Ross, Rubinstein [43]) is applied for the evaluation of the option. Alternatively, the constructed tree delivers the local volatility surface which can be used for the

[5]For example, for the lognormal diffusion (1) we match the first two moments of the continuous process and of the discretised process (see Cox, Ross, Rubinstein [43]).

pricing and hedging of options via Monte Carlo simulation (see Derman, Kani, and Zou [51], and Zou and Derman [150]).

There are three different approaches to the construction of implied trees. First, binomial implied trees are constructed by using both backward and forward induction (Derman and Kani [49], Barle and Cakici [7])[6]. They fit implied volatilities in both the maturity and strike dimension. Second, binomial implied trees are constructed by using only backward induction. They fit either the strike dependence of implied volatilities (Rubinstein [126]), or both the strike and the term dependence (Jackwerth [96]). Third, trinomial trees are built by using simultaneously forward and backward induction (Dupire [60], [62], Derman, Kani and Chriss [53]). They fit both the strike and the term structure of implied volatilities. Their main difference with implied binomial trees is that the state space is fixed in advance, and the construction of the tree is reduced to the calculation of transition probabilities.

Finally, implicit finite difference schemes (Andersen [2], Andreasen [4]) are proposed in order to solve some of the problems encountered with implied trees.

2.3.1 Derman and Kani (1994)

Derman and Kani [49] build a recombining binomial implied tree by using forward and backward induction simultaneously.

Their tree has uniformly spaced levels which are Δt apart. In order to construct it, they assume that they have already implied the tree's nodes and the transition probabilities out to level n. The known price at node i and level n $S_{i,n}$ can evolve into an "up" node with price $S_{i+1,n+1}$, or into a "down" node with

[6]Backward and forward induction are the discrete analogues of the Kolmogorov backward and forward equations, respectively. The binomial backward equation states that the price at any period n is the discounted value of the average of the prices at the two up and down nodes in the next period $n+1$. The binomial forward is the "dual" or the "adjoint" of the binomial backward equation. It states that the price of an Arrow-Debreu primitive security of any maturity $(n+1)$ is the average of the discounted at the previous two up and down nodes of the Arrow-Debreu security of maturity n (for a further description and application of the technique in the context of interest rate models see Hull and White [92], Jamshidian [97], and Rebonato [122]).

price $S_{i,n+1}$ at level $(n+1)$. The (unknown) probability of making a transition into the "up" node is denoted by p_i. The aim is to determine the nodes of the $(n+1)^{th}$ level at time t_{n+1} and the corresponding transition probabilities. In total, there are $2n+1$ parameters that define the transition from the n to the $(n+1)$ level of the tree. These parameters are the $n+1$ stock prices $S_{i,n+1}$, and the n transition probabilities p_i. Derman and Kani determine them by using the smile.

They find the distribution of $S_{i,n+1}$ and the transition probabilities p_i by using the theoretical values of n forwards and n *European* options, all expiring at time t_{n+1}. They require that these theoretical values match the (interpolated) market values. This provides $2n$ equations for these $2n+1$, parameters and it ensures that they fit today's smile. They use the one remaining degree of freedom to make the centre of their tree to coincide with the centre of the standard CRR tree that has constant local volatility[7].

The above can be expressed formally as follows: The martingale condition delivers the forward price $F_{i,n}$ of the stock as

$$F_{i,n} = p_i S_{i+1,n+1} + (1-p_i) S_{i,n+1} \tag{2.8}$$

Let $C(S_{i,n}, t_{n+1})$ and $P(S_{i,n}, t_{n+1})$, respectively, be the known market values for a European call and put, struck today at $K = S_{i,n}$ and expiring at t_{n+1}. The values of each of these calls and puts are known from interpolating the smile curve implied from options expiring at time t_{n+1}. The theoretical binomial value of a European call struck at K and expiring at t_{n+1} in a complete market is given by:

$$C(K, t_{n+1}) = e^{-r\Delta t} \sum_{j=1}^{n} \{Q_{n,j} p_j + Q_{n,j+1}(1-p_{j+1})\} \max(S_{j+1} - K, 0) \tag{2.9}$$

[7]For a different choice of the "centrering condition", the constructed tree would have been different. However, in the continuous time limit, where there are an infinite number of nodes at each time step, the choice of the "centrering condition" is not important (see Derman, Kani and Chriss [53]).

where the sum is taken over all nodes j at the $(n+1)$ level and $Q_{n,j}$ is the price of an Arrow-Debreu security expiring at t_{n+1}. From equations (2.8) and (2.9) they get that:

$$S_{i+1,n+1} = f(r, \Delta t, S_{i,n+1}, C(S_{i,n}, t_{n+1}), \Sigma_c, Q_{i,n}, S_{i,n}, F_{i,n}) \tag{2.10}$$

$$p_i = f(F_{i,n}, S_{i,n+1}, S_{i+1,n+1}) \tag{2.11}$$

where $\Sigma_c = \sum_{j=i+1}^{n} Q_{j,n}(F_{j,n} - S_{i,n})$. The Arrow-Debreu prices $Q_{i,n}$ have been calculated by applying forward induction.

For all the nodes above the centre of the tree we can find iteratively $S_{i+1,n+1}$ and p_i, from equations (2.10) and (2.11) if we know $S_{i,n+1}$ at one initial node. "Centering conditions" are imposed so that to calculate $S_{i,n+1}$. If the number of nodes at the $(n+1)^{th}$ level is odd they choose the central node $S_{i,n+1}$ (for $i = \frac{n}{2} + 1$) to be today's spot price, as in the CRR tree. If the number of nodes at the $(n+1)^{th}$ level is even, they start instead by identifying as initial $S_{i,n+1}$ and $S_{i,n}$, the nodes just below and above the center of the level (i.e. $i = \frac{n+1}{2}$). This is done by making the average of the natural logarithms of the two central nodes' stock prices equal to the logarithm of today's spot price. Substituting this condition in equation (2.10) gives the formula for the upper of the two central nodes for even levels

$$S_{i+1,n+1} = f(r, \Delta t, S_{i,n}, C(S_{i,n}, t_{n+1}), \Sigma_c, Q_{i,n}, F_{i,n}) \tag{2.12}$$

for $i = \frac{n}{2} + 1$. Once we have this initial node's stock, we can continue to fix higher nodes from equation (2.10).

Similarly, the asset value for the nodes below the central node at level n , are calculated by using known put prices[8]. The analogous formula that determines

[8]Non-synchronous trading and the bid-ask bounce create noise in the observed option prices (see Harvey and Whaley [81], and Roll [124], respectively). By using out-of-the money calls and puts, they minimize the effect of noisy option prices on the construction of their tree. This is because the delta for these options is low.

a lower stock price from a known upper one is

$$S_{i,n+1} = f(r, \Delta t, S_{i+1,n+1}, P(S_{i,n}, t_{n+1}), \Sigma_p, Q_{i,n}, F_{i,n}) \tag{2.13}$$

where $\Sigma_p = \sum_{j=1}^{i-1} Q_{j,n}(S_{i,n} - F_{j,n})$.

Applying equations (2.10), (2.11) and (2.13) for every level and for small enough time steps between successive levels, completes the construction of the tree. These equations reveal the idea behind the "implied trees" methodology. From the current option prices, we can back out a discrete approximation to the risk-neutral stock process, and the risk-neutral transition probabilities.

The advantage of Derman and Kani's algorithm is that it provides the asset price evolution, and the transition probabilities by capturing both the term and the strike structure of implied volatilities (they interpolate across option prices for each time level). On the other hand, Barle and Cakici [7] find that Derman and Kani's algorithm fails to reproduce the smile accurately if the interest rate is high. In the next section, we demonstrate how Barle and Cakici extend Derman's and Kani's algorithm.

2.3.2 Barle and Cakici (1995)

In order to ensure that transition probabilities remain in the interval [0,1], Derman and Kani require that $F_{i,n} < S_{i+1,n+1} < F_{i+1,n}$. If the stock price $S_{i+1,n+1}$ violates this inequality, then they override the option price that produced it. The missing stock price is replaced by the one which keeps $\ln S_{i+1,n+1} - \ln S_{i,n+1} = \ln S_{i+1,n} - \ln S_{i,n}$. However, Barle and Cakici [7] note that Derman and Kani's algorithm fails to reproduce the smile accurately when the interest rate is high. The reason is that with higher interest rate, negative probabilities are more frequently encountered, leading to overriding the corresponding option prices. Hence, the constructed tree does not fully incorporate the information from the smile. In order to correct for this problem they propose three modifications to Derman and Kani's method.

First, they choose the option to be struck at $K = F_{i,n}$. Second, rather than fixing the center of the tree at the current stock price, they allow it to follow the evolution of the mean of the risk-neutral distribution by setting it to $Se^{rt_{n+1}}$. Third, when there is a missing stock price due to the violation of the arbitrage condition, they set $S_{i+1,n+1} = \frac{F_{i,n} + F_{i+1,n}}{2}$. Barle and Cakici's modifications are equivalent to working with the futures rather than the spot price. In a standard binomial tree this trick guarantees non-negative transition probabilities (see Hull [91]).

Even though, their modified method fits the smile accurately for very high interest rates (e.g. $r = 40\%$), it fails to do so for increasing interest rates and smile slope. Negative probabilities occur even with this modification. "These weaknesses are a consequence of the strict requirements that continuous diffusion can be modelled as a binomial process and on a recombining tree", as they state in their conclusions.

2.3.3 Rubinstein (1994)

Rubinstein [126] constructs an implied binomial tree which has T levels, by using only backward and not forward induction. In this sense his tree is an extension of the CRR tree. The key input to his algorithm is the terminal total (nodal, as opposed to the one period) risk-neutral implied probabilities of the underlying asset. He extracts them from the observed prices of European options which mature at time T, by using a nonlinear minimization method[9].

His method consists of establishing a prior guess of the terminal risk-neutral distribution; his guess is the log-normal one. Then, the implied posterior risk-neutral probabilities are those which are, in the least-squares sense, closest to the lognormal[10]. The minimization is performed subject to some constraints.

[9] In general, there is a number of approaches for estimating risk-neutral functions from option prices. For a survey of these methods see Bahra [5], and Mayhew [111].

[10] Jackwerth and Rubinstein [94] examine alternative specifications of the minimization criterion using historically observed option prices. All of the specifications, including the quadratic one, produce similar posterior distributions (for near-the-money) which seem to be independent

The probabilities must add up to one and be non-negative. Moreover, they are calculated so that the present value of the underlying assets and all the European options calculated with these probabilities to fall between their respective bid and ask prices.

In order to proceed further, he imposes a number of assumptions: (a) binomial evolution of the asset price, (b) recombining nodes, (c) ending nodal values organized from lowest to highest, (d) constant interest rate, and (e) all paths leading to the same ending node have the same risk-neutral probability. Then, the tree is constructed through four very simple steps.

1. For every node j, calculate the terminal path probabilities corresponding to the T level, from the terminal nodal probabilities.

2. From the path probabilities of the T level, calculate the path probabilities for the $T-1$ level.

3. From the path probabilities of the $T-1$ level, calculate the transition probabilities of a transition from level $T-1$ to level T.

4. Uses the transition probabilities to calculate the return for the jth node at the $T-1$ level.

This exercise is repeated for every time level and completes the construction of the tree. Then, the value and the hedging parameters of any derivative instrument maturing with or before the European options can be calculated. However, the constructed tree fits only the European options with maturity T (in the sense that the model price falls within the bid and ask observed prices). This is because it uses as input only the options maturing at the T level. It does not capture the term structure of implied volatilities something which can be considered as a limitation of the technique.

---

of the assumed prior distribution.

2.3.4 Jackwerth (1997)

Rubinstein's implied binomial tree fits only the European options which expire at the terminal level of the tree. On the other hand, Derman and Kani's model fits intermediate maturity options, but the construction of the tree depends on the chosen interpolation and extrapolation method. Moreover, negative transition probabilities are frequently encountered. As a solution to these problems, Jackwerth [96] develops a generalized implied binomial tree. It is 'generalized' in the sense that the simplicity of Rubinstein's implied binomial tree is preserved, but it relaxes the assumption that all the paths which lead up to the same ending node are equally probable. This allows him to fit intermediate maturity European option prices. In addition, the transition probabilities are constrained by construction to lie within 0 and 1, as it was the case with Rubinstein's implied tree.

Let $i = 0, 1, \ldots, n$ be the time step, and $j = 0, 1, \ldots, i$, be the nodes at each time step starting with the lowest stock price at the bottom of the step. In order to fit the intermediate maturity options, Jackwerth works with nodal, rather than path probabilities, and he uses a weight function $w_{i,j}$ which has a particular interpretation. $w_{i,j}$ can be interpreted as the portion of nodal probability at the upper node going into the preceding node at the previous time step (down weight). For a standard binomial tree, $w_{i,j} = \frac{j}{i}$ (linear function). In a generalized implied binomial tree, $w_{i,j}$ is an arbitrary function; it is determined so that to fit the intermediate maturity options. Jackwerth reports that concave weight functions explain the observed European index on the S&P 500 option prices better than either linear, or convex weight functions. A concave weight function implies that a path going first down and then coming up, is more likely to be taken than a path going first up and then coming down.

Given the nodal probabilities and stock prices at time i, he can solve for the nodal probability and stock price at the preceding node in three steps:

1. $P^{nodal}_{i-1,j} = (1 - w_{i,j}) P^{nodal}_{i,j} + w_{i,j+1} P^{nodal}_{i,j+1}$

2. $p = P_{i-1,j} = w_{i,j+1}\frac{P^{nodal}_{i,j+1}}{P^{nodal}_{i-1,j}}$

3. $S_{i-1,j} = [(1 - P_{i-1,j})S_{i,j} + P_{i-1,j}S_{i,j+1}]/(r/\delta)$,

where r and δ are the interest rate and dividend yield per step, and p is the transition probability. Note, that as long as the weights are between 0 and 1, the transition probabilities will also be between 0 and 1.

Jackwerth's technique recognizes that the evolution of nodal probabilities throughout any tree (standard, or implied binomial or trinomial tree) is governed by a transition probability weight. Changing the functional form for this weight, changes the nodal probabilities, i.e. changes the transition probabilities and the local volatilities. By implying the functional form of this weight from European option prices, one can change the nodal probabilities so that to price shorter term options consistently with the shorter and longer term European option prices.

2.3.5 Trinomial Trees

Implied trinomial trees are proposed as a solution to the problem of not-acceptable transition probabilities occurring in Derman and Kani's implied binomial tree. Moreover, trinomial trees provide a much better approximation to the continuous time process than the binomial tree for the same number of steps. This is because there are three possible future movements over each time rather than two (see Clewlow and Strickland [35]).

Trinomial trees have more parameters than binomial trees. The constraints remain the same, i.e. matching the moments of the continuous process and of the discretized process, and matching the model's forward and option prices with the market's. Inevitably, in order to match the parameters with the constraints that we have, we have to select the state space in advance (see Derman, Kani, and Chriss [53]). Then, the transition probabilities between the nodes can be easily calculated from the constraints. The "freedom" to fix the state space, enables us to come up with acceptable transition probabilities. On the other hand, we need

to be careful choosing the state space so that to fit the current smile.

Derman, Kani and Chriss [53] discuss the issue of constructing the state space when volatility varies significantly with time to expiration and strike, producing a skew. In such a case, the nodal spacing has to change significantly with time and stock level. The method that they propose for constructing the state space, consists of two steps. In the first step, they assume that interest rates and dividends are zero, and they build the state space which corresponds to a trinomial tree with constant volatility (this is also suggested by Dupire [60], [62]). Then, they modify the time spacing and subsequently the nodal spacing so that to capture the basic term and skew structures of local volatility in the market. In the second step, if there are any forward price violations in any of the nodes, they multiply all node prices by the growth factor $e^{r(r-\delta)t_i}$. This is equivalent to working with the futures price, rather than the asset price (as Barle and Cakici [7] proposed), and it will remove all forward price violations.

Having set the state space in advance, the problem of the construction of the implied tree is reduced to the calculation of consistent with the smile transition probabilities (implied transition probabilities). Dupire [61], [62] sketches a way for calculating them. His technique can be summarized as follows. The state prices (Arrow-Debreu prices) are implied by the market prices of European calls and puts (implied Arrow-Debreu prices). Then, the implied transition probabilities are calculated from fitting the smile by using simultaneously backward and forward induction.

To make Dupire's description concrete, assume that we observe the market prices of European calls and puts for any strike and maturity. Then, the theoretical price of a European call with strike price K and maturity date $n\Delta t$, in a complete market is given by

$$C(n\Delta t, K) = \sum_{j=-n}^{n} Q_{n,,j} \max(S_{n,j} - K, 0) \tag{2.14}$$

By taking $C(n\Delta t, K)$ as observed from the market, we have to invert somehow

equation (2.14), so that to get $Q_{n,j}$ (implied state prices). Assume that we wish to compute the implied state prices for time step n. We start at the top node n at time step n. The price of a European call with strike price $S_{n,n-1}$ and maturity date $n\Delta t$ is

$$C(n\Delta t, S_{n,n-1}) = Q_{n,n}(S_{n,n} - S_{n,n-1}) \tag{2.15}$$

which can be rearranged to give the state price at node (n, n). We can compute the state prices for the nodes down to the middle of the tree in a similar way, i.e. for node (n, k) we compute $Q_{n,k}$ by choosing $K = S_{n,k-1}$.

We then compute the state prices for the lower half of the tree by starting from the bottom node of the tree and using puts. This is repeated for every time step in the tree and completes the calculation of $Q_{n,,j}$ for every (n, j). Notice that for the, consistent with the smile, evaluation of European options we do not need the transition risk-neutral probabilities, but only the implied state prices. However, transition probabilities are necessary for the evaluation of more complex options. We are going to show how to calculate them by using the already derived implied state prices.

Imagine that we are at node (n, j) and we have computed the transition probabilities and the state prices for all nodes above (n, j). We want to calculate the transition probabilities $p_{u_{n,j}}, p_{m_{n,j}}, p_{d_{n,j}}$ from node (n, j) to the upward, middle, and downward node at the next time level $n + 1$. We have three unknowns, and therefore we need three equations. The three equations are given by

(a) the forward induction of the state prices:

$$Q_{n+1,j+1} = e^{-r\Delta t}(p_{d_{n,j+2}}Q_{n,j+2} + p_{m_{n,j+1}}Q_{n,j+1} + p_{u_{n,j}}Q_{n,j}) \tag{2.16}$$

(b) the backward induction for the price of the asset:

$$S_{n,j} = e^{-r\Delta t}(p_{d_{n,j}}S_{n+1,j-1} + p_{m_{n,j}}S_{n+1,j} + p_{u_{n,j}}S_{n+1,j+1}) \tag{2.17}$$

(c) the forward price of a one period bond:

$$1 = (p_{d_{n,j}} + p_{m_{n,j}} + p_{u_{n,j}}) \tag{2.18}$$

The first equation can be rearranged to give $p_{u_{n,j}}$ directly. We then solve the second and third equations for $p_{m_{n,j}}$ and $p_{d_{n,j}}$. The above procedure is repeated for every time step in the tree. Transition probabilities are calculated by using simultaneously forward and backward induction. Hence, the constructed trinomial tree fits the observed smile.

Derman, Kani and Chriss [53] use equations which are very similar to the ones that Derman and Kani [49] use for the construction of their implied binomial tree. They calculate the transition probabilities from the known option prices, asset prices and Arrow Debreu prices. These equations have been derived by applying backward and forward induction in a way which is similar to the one that Dupire sketched.

2.3.6 Implicit Finite Difference Schemes

Andersen [2], and Andreasen [4] construct implicit and semi-implicit (Crack-Nicolson) schemes, which are consistent with the equity option volatility smile. They employ these schemes so that to overcome the problem of negative transition probabilities that may be encountered with binomial and trinomial trees. Avoiding negative transition probabilities is feasible within implicit finite difference schemes because these schemes have better properties, in terms of stability and convergence, than binomial and trinomial trees (see Clewlow and Strickland [35]. Andersen also shows that stability is equivalent to having acceptable probabilities). This is expected because the negative probabilities arise in the binomial and trinomial tree from pushing the asset price process into a certain kind of evolution. On the other hand, the implicit schemes discretize the BS type fundamental PDE (see Geske and Shastri [75]).

Andersen incorporates the forward induction technique into the implicit finite

scheme. This enables him to estimate the risk-neutral distribution from a set of option prices. Then, he backs out the local volatilities from the risk-neutral densities of the different maturities by solving a constrained quadratic problem. Andreasen, contrary to Andersen, directly extracts the local volatilities from the input implied volatilities. He achieves this by deriving an explicit formula that relates the surface of implied volatilities to the surface of local volatilities. Moreover, in addition to Dupire's forward equation that the option price satisfies, he derives forward equations that the "Greek" hedging ratios (i.e. delta, gamma, vega, theta) must satisfy, as well.

2.4 Testing the Validity of the Deterministic Volatility Assumption

Smile consistent deterministic volatility models have theoretical and practical advantages. They are a simple extension of the BS model preserving its arbitrage pricing property. In addition, they achieve a cross-sectional fit of the observed for the different strikes and maturities option prices. Finally, they are easy to implement. However, the validity of the deterministic volatility assumption has to be investigated empirically, before concluding that these models are appropriate for option pricing and hedging purposes. We are aware of three papers which address this issue: Dumas, Fleming and Whaley's [58], Jackwerth and Rubinstein's [95], and Buraschi and Jackwerth[28] [11].

Dumas, Fleming and Whaley [58] assess the stability of the deterministic volatility function for the S&P 500 Index, by examining how well it predicts future option prices. They estimate, every week, various specifications of the volatility function. Their estimation is performed by minimizing the sum of squared deviations of theoretical option prices from the observed market option prices. Then, they examine the price deviations from theoretical values one week later. The

[11] Related empirical studies by Bates [14], and Bakshi, Cao, and Chen [6] compare different option models, without looking explicitly at smile-consistent deterministic volatility models.

deterministic volatility model can always fit the cross-section of observed option prices, as long as the volatility function is complex enough. However, their out-of-sample results indicate that the instantaneous volatility function is not stable over time.

Jackwerth and Rubinstein [95] compare the out-of-sample empirical performance of alternative models, including Jackwerth's [96] generalized implied binomial tree, in terms of the pricing and hedging errors. They find that both generalized binomial trees and stochastic volatility models outperform the BS model. However, the size of the standard deviation of the pricing errors makes it difficult to conclude whether generalized binomial trees are better, or worse than stochastic volatility models.

Buraschi and Jackwerth [28] rather than exploring the empirical performance of deterministic volatility models being based on the size in dollars of the pricing and hedging errors, they provide a general statistical test of deterministic volatility models versus stochastic volatility ones. They test directly the implication of deterministic volatility models that options are redundant securities. Their null hypothesis is that the payoff of any asset can be replicated with a dynamic trading strategy that involves two primitive assets, such as the underlying asset. They construct their tests from the properties of the implied risk-neutral density. Their tests reject the null hypothesis. The results suggest that the returns of the in-and out-of-the-money options are needed for spanning purposes. This finding is even stronger in the postcrash period.

The above-mentioned empirical studies indicate that the instantaneous volatility is not a deterministic function of the asset price and of time. Therefore, the asset price does not behave according to the inferred, from the implied deterministic volatility models, process. In this case the hedging will not be effective. As a solution to this Dupire [60] proposes a method of hedging which is robust, no matter what the dynamics of the asset price look like. He hedges a claim against movements in the volatility by using a portfolio of European options. This portfolio is rebalanced periodically, so that a change in the volatility surface

will change the value of both the targeted claim and of the portfolio by the same amount. Bates [15] also provides a simple non-parametric method for inferring the deltas and gammas from the implied volatility patterns. His method is based on the assumption that the underlying asset price follows a stochastic process with constant returns to scale, so that option prices are homogeneous of degree one in the underlying asset price and strike.

Despite these suggestions, the empirical evidence (Gemmill [73], Jackwerth and Rubinstein [94]) implies that deterministic volatility models have to recalibrated every day, so that to fit the smile. Therefore, they do not offer a unified theory of volatility which can be used for the pricing and hedging of exotic options. S*mile-consistent stochastic volatility models* have been developed in order to provide such a theory.

2.5 Smile Consistent No Arbitrage Stochastic Volatility Models

The development of the stochastic volatility literature is similar to the evolution of the interest-rate literature. In the latter, there was a transition from equilibrium considerations (e.g. Vasicek [144], Cox, Ingersoll, Ross [45]) to arbitrage arguments (e.g. Ho and Lee [86], Heath, Jarrow, Morton [84], (HJM)) (for a survey of the continuous time interest rate literature, see Strickland [138]). In that context, the aim is not to explain the yield curve, but taking it for granted and along with evolution assumptions to obtain arbitrage prices for derivative securities. In the stochastic volatility literature, first models which start by assuming a stochastic process for the evolution of the instantaneous volatility were introduced. Then, models which start from today's European option prices and achieve option pricing and hedging of more complicated claims under a consistent with no-arbitrage evolution of the volatility surface, were developed.

2.5.1 Dupire (1992)

Dupire's [59] approach is similar to HJM's. He starts from today's European option prices, and he derives the process of the instantaneous volatility "endogenously" from the process of the "forward" volatility (see below for his definition of forward volatility).

Let the financial instrument f_T delivering $f(S_T)$ at time T. Without loss of generality, he assumes the interest rate to be zero at all times and he also assumes the absence of any arbitrage opportunities in the market. Let (Ω, I, I_t, P), be a filtered probability space, where (I_t) is a right continuous filtration, and P is the objective probability measure. Let now P_T be the P-equivalent probability measure. Then, the value at time 0 of f_T is given by (see Dothan [57])

$$f_T(0) = \int f(S_T)\Phi_T(S_T)dS_T \equiv E^{P_T}[f(S_T) \mid I_0] \tag{2.19}$$

where $\Phi_T(S_T)$ has been extracted from the known prices of European call options (equation (2.2)) and the expectation is taken with respect to P_T (risk-neutral measure). Hence, in order to price any instrument, we have to get the risk-neutral process of the underlying asset. In order to get the risk-neutral process for the spot he assumes that

$$\frac{dS_t}{S_t} = \mu_t dt + \sigma_t dW_{1,t} \tag{2.20}$$

where W_1 is a P-Brownian motion adapted to F_t and μ_t and σ_t can be measurable processes themselves adapted to F_t. By defining $d\overline{W_{1,t}} \equiv dW_{1,t} + \frac{\mu_t}{\sigma_t}dt$, Dupire gets

$$\frac{dS_t}{S_t} = \sigma_t d\overline{W_{1,t}} \tag{2.21}$$

where $\overline{W_1}$ is a Brownian motion under P_T, obtained by Girsanov's theorem (see Oksendal [117] for a description of Girsanov's theorem). Equation (2.21) is the risk-neutral process for the Spot. We can use it to price options by Monte-Carlo simulation (see Boyle [24]). However, in order to price options consistently

with the market, the process for instantaneous volatility σ_t must incorporate information from today's option prices. Moreover, it should respect the evolution of the implied volatility surface. For pricing purposes, we need the risk-neutral process for such an evolution of the volatility (this point is explained in Hull [91]). Then, we will be able to plug it into equation (2.21) and achieve smile-consistent pricing.

Applying Ito's lemma to $lnSt$, using equation (2.20), integrating the result between T_1 and T_2, and rearranging terms we get

$$\int_{T_1}^{T_2} \sigma_t^2 dt = 2 \int_{T_1}^{T_2} \frac{dS_t}{S_t} - 2(\ln S_{T_2} - \ln S_{T_1}) \tag{2.22}$$

The left hand side (LHS) of equation (2.22) can be thought of as being the payoff at time T_2 of a forward contract which is traded at time t. Assuming that $T_2 - T_1 = \varepsilon$, Dupire defines as V_T this forward contract which delivers the instantaneous variance to be observed at time T[12]. By interpreting the stochastic integral on the right hand side of equation (2.22) as the gain or loss from a self financing strategy, results in the integral to vanish[13]. Let L_T be a claim delivering the logarithm of S_T at time T (log-contract). Its price at time 0 is[14]

$$L_T(0) = E^{P_T}[\ln S_T \mid I_0] \tag{2.23}$$

So, the payoff of a contract traded on the future volatility, is equal to the payoff of a self-financing strategy of buying and selling today's log-contracts of maturities T_1 and $T_1 + \varepsilon$. Then, their prices must be the same, if no arbitrage is to exist. By dividing both sides by ε, and taking the limits, so that $\varepsilon \to 0$, we

[12]The assumption of the existence of a forward contract which is traded on volatility is not unrealistic. Over-the-counter futures and options contracts on foreign-currency and interest-rate volatility indexes, are currently being developed by a number of investment banking firms in the U.S. and Europe (see Gruenbichler and Longstaff [78]).

[13]For the definition of a self-financing strategy, see Dothan [57].

[14]Notice that in order to price the log-contract consistently with the market, we need the implied risk-neutral density function. This explains why Dupire needs to take the observed option prices as granted.

get by arbitrage pricing that the value of the forward contract V_T at any time $t < T$ is

$$V_T(t) = -2\frac{\partial L_T(t)}{\partial T} \tag{2.24}$$

This is the point where Dupire defines implicitly as $V_T(t)$ the instantaneous forward variance (IFV) to be observed at time T. The IFV $V_T(t)$, is defined as the value at time t of a forward contract which will deliver the instantaneous variance to be observed at some time T in the future[15].

From the values $(L_T(0))_T$, we can deduce the initial instantaneous forward variance curve as

$$V_T(0) = -2\frac{\partial L_T(0)}{\partial T} \tag{2.25}$$

Next, he models the forward variance which automatically ensures compatibility with $(L_T(0))_t$ as equation (2.25) shows. Consequently, this will ensure compatibility with the current volatility surface, as equation (2.23) shows. He makes among other possible choices, the assumption that $V_T(t)$ is lognormal i.e.

$$\frac{dV_T(t)}{V_T(t)} = adt + bdW_{2,t} \tag{2.26}$$

where a and b are constant or a deterministic function of time and W_2 is another Brownian motion adapted to F_t, *possibly correlated with* W_1. Defining $d\overline{W_{2,t}} = dW_{2,t} + \frac{a}{b}dt$, equation (2.26) can be rewritten as

$$\frac{dV_T(t)}{V_T(t)} = bd\overline{W_{2,t}} \tag{2.27}$$

where $\overline{W_2}$ is a Brownian motion under Q_2, the P-equivalent probability measure obtained by Girsanov's theorem. Equation (2.27) is the risk-neutral process for

[15]Notice that by definition, $V_T(t) = V_T(t, S_t)$. This is because

$$V_T(t) = E^{Q_2}[-2(\ln S_{T+\varepsilon} - \ln S_T) \mid I_t] =$$

$$E^{Q_2}[-2(\ln S_{T+\varepsilon} - \ln S_T)]$$

since the increments of a Brownian motion are orthogonal and this holds true for the increments of any function of the Brownian motion.

the instantaneous forward variance. Using equation (2.27), he derives the risk-neutral process for the instantaneous volatility. That is

$$\frac{d\sigma_t}{\sigma_t} = (\frac{1}{2}\frac{\partial \ln V_t(0)}{\partial t} - \frac{b^2}{8})dt + \frac{b}{2}d\overline{W_{2,t}} \tag{2.28}$$

Equation (2.28) is probably the most important contribution of this paper. The first term of the drift of equation (2.28) can be estimated from the prices of the log-contracts, as equation (2.25) shows, and the second term can be estimated from the initial forward volatility surface. Therefore, the drift of (2.28) ensures consistency with the initial (implied) volatility term structure. In addition, $\overline{W_{2,t}}$ moves stochastically the volatility surface. In this way he hopes that his model will fit better the smile every day[16]. Solving equation (2.28) and plugging it into equation (2.21) delivers the risk-neutral process for the spot.

As a final step, he assumes that the filtration associated with $\overline{W_1}$ and $\overline{W_2}$ conveys the same information, and therefore is the same. This is a reasonable assumption because the filtration associated with $\overline{W_1}$ is used for the pricing of the log-contract; the log-contract is used for the construction of the forward variance and hence of $\overline{W_2}$. Define Q (through Girsanov's Theorem) as the P-equivalent measure, under which $\overline{W_1}$ and $\overline{W_2}$ are Q-Brownian motions. Then, the price at time t, $h(t)$ of an instrument h that delivers at time T a payoff $h(S_T)$, is

$$h(t) = E^Q[h(S_T) \mid F_t] \tag{2.29}$$

where F_t is the natural augmented filtration associated with W_1 and W_2. The expectation in equation (2.29) can be evaluated consistently with the smile by "joint" Monte Carlo (see Boyle [24]) of the risk-neutral process of the spot (equation (2.21)) and of the instantaneous volatility (equation (2.28)). Monte Carlo simulation allows us also to compute the hedging parameters through a small

[16]This is analogous to the form of the drift of the short-term interest rate that Heath, Jarrow, and Morton [84] derived by starting from the process for the instantaneous forward variance. That drift depends among other things and on the initial instantaneous forward rate which reflects the current yield curve.

shift of the paths.

To summarize, Dupire starts by assuming that European Calls of all strikes and maturities are traded, and that their market prices are consistent with no arbitrage. From these prices, he deduces the arbitrage price $L_T(t)$ of contingent claims (log-contract) that promise lnS_T at date T (the log-contract is not traded in the listed market, but it can be approximated by a combination of standard listed options (see Derman et al. [52]). These prices $L_T(t)$ permit to synthesize the value of a forward market on the instantaneous variances to be observed at any maturity, as equation (2.24) shows[17]. Next, he assumes one factor model for the forward variance and he derives the risk neutral process for the instantaneous variance. Using the risk-neutral processes for the spot and the instantaneous volatility, he obtains arbitrage free prices that do not depend on any risk premia, nor on a volatility drift. They do depend on the term structure of the IFV, on the correlation between spot and IFV, and on the volatility of the latter, as equation (2.21) combined with equation (2.28) show.

A limitation of the model is that it recognizes the strike and term structure of implied volatilities for only the current time, and not for future times, as the volatility surface evolves stochastically. The two dimensional information $(C_{K,T}(0))_{K,T}$ has been compacted at each time into a 1-dimensional $(L_T(0))_T$ in the process. Equivalently stated, only the first moment of the asset price distribution has been used for the construction of the forward variance, as equation (2.24) combined with equation (2.23) shows.

2.5.2 The Definition of the Forward Variance Revisited

Even though Dupire [59] introduced the idea of the forward variance (some people call it local), a formal treatment of the forward volatility concept was not done

[17]Neuberger [116] shows that the log-contract is very useful in order to hedge against volatility. He does this by delta-hedging log-contracts on futures. Then, he regresses the percentage hedging error on the squared outcome volatility, and he gets a correlation coefficient of 99.99%. That is to say that by trading a portfolio of log-contracts, we can replicate (and therefore eliminate) the variance.

until 1996 in Kani, Derman and Kamal's [103] paper. There, they assume that the risk-neutral index evolution is governed by the following equation

$$\frac{dS_t}{S_t} = (r_t - \delta)dt + \sigma_t d\overline{Z_t} \tag{2.30}$$

where r_t is the riskless rate of return at time t, assumed to be a deterministic function of time, δ is a continuous compounded dividend yield, and σ_t is the instantaneous index volatility at time t, assumed to follow some unspecified stochastic process.

The forward (local) variance $\sigma^2_{K,T}$, corresponding to level K, and maturity T, is defined as the conditional expectation of the instantaneous variance of index return at the future time T, contingent on index level S_T being equal to K, i.e.

$$\sigma^2_{K,T}(t,S) = E^*_t(\sigma^2_T \mid S_T = K) \tag{2.31}$$

where the expectation has been taken with respect to the risk-neutral measure and the subscript t indicates that the expectation is based on information at time t.

The local volatility $\sigma_{K,T}$, is defined as the square root of the local variance, $\sigma_{K,T} = (\sigma^2_{K,T})^{\frac{1}{2}}$. Hence, the local volatility is defined as the forecast (estimate) of index volatility at a particular future time and market level, so that to make current option prices fair. It is worth noting that the concept of the local volatility is different to the concept of the implied volatility $\Sigma_{K,T}$ which can be thought of as the market's estimate of the expected average future index volatility during the life of the option[18].

Note, that in the case where the instantaneous index volatility is assumed to

[18]The statement that implied volatilities are the estimates of the (expected) average future index volatility is rather loose. It can be proved formally only for the case that (a) we are in a Hull-White world of stochastic volatility, and (b) we deal with at-the-money options. To make things more concrete, Hull and White [90] show that when the volatility risk is not priced and the correlation between the underlying asset and the volatility is zero, then the price P of an option is given by

$$P = \int BS(\overline{V_t})h(\overline{V_t} \mid I_t)d\overline{V} = E[BS(\overline{V_t}) \mid I_t]$$

be deterministic as a function of the index level and time (deterministic volatility case) i.e. $\sigma_T = \sigma(S_T, T)$, equation (2.31) becomes

$$\sigma^2_{K,T}(t,S) = E_t(\sigma^2_T \mid S_T = K) = E_t[\sigma^2(S_T,T) \mid S_T = K] = \sigma^2(K,T) \qquad (2.32)$$

Equation (2.32) shows that the forward variance equals the instantaneous variance, when the latter is assumed to be deterministic. That is to say that the smile-consistent stochastic volatility case, nests the deterministic volatility one. In this case we are left with a static local volatility surface, whose shape remains unchanged as time evolves since the right hand side is independent of t and S. This is analogous to the result that in a deterministic interest rates economy, the forward rates do not change over time (see Rebonato [122]).

Derman and Kani [54] show that if the asset price evolves according to equation (2.30), then the forward variance is given by

$$\sigma^2_{K,T}(t,S) = \frac{2\{\frac{\partial C_{K,T}}{\partial T} + (r-\delta)K\frac{\partial C_{K,T}}{\partial K} + \delta C_{K,T}\}}{K^2\frac{\partial^2 C_{K,T}}{\partial K^2}} \qquad (2.33)$$

Equation (2.33) shows that the local volatility can be locked in by trading portfolios of currently available options. Hence, it can be thought as the market price of volatility we can lock in today, in order to obtain volatility exposure over some specific range of future times and market levels. Local volatility is the volatility analog of the forward rate. The forward rate is the future rate prevailing in a time interval, so that the current yields to maturity are justified, and it can be locked in by trading current bond portfolios[19].

---

where $BS(.)$ is the BS pricing formula, $\overline{V} = \frac{1}{T-t}\int_t^T V_s ds$. $h(\overline{V_t} \mid I_t)$ is the density of $\overline{V_t}$ conditional on the the information set at time t I_t, and T is the expiration date of the option. Moreover, Cox and Rubinstein [44] show that the BS formula is approximately a linear function of the standard deviation for at-the-money options, so that $E[BS(\overline{V_t}) \mid I_t] = BS[E(\overline{V_t}) \mid I_t]$.

[19] Dupire [63] has also derived equation (2.33) independently, for the case that $r = \delta = 0$. He defines the forward variance $V_{K,T}(S_0, t_0)$ as the price of a forward contract $V_{K,T}$ introduced at time t_0. The difference with his earlier definition (Dupire [59]) is that the contract will exchange at time T, the instantaneous variance observed at time T $v(S_T, T)$, against an agreed

The existence of the risk neutral measure is used for the expectation definition of local volatility, and it also allows for an alternative definition of the local volatility. In order to show this, Derman, Kamal and Kani [52] assume that under the risk neutral measure, the local volatility evolves according to the following process

$$\frac{d\sigma^2_{K,T}}{\sigma^2_{K,T}} = a_{K,T}dt + \beta_{K,T}d\overline{Z_t} + \sum_i \vartheta^i_{K,T}d\overline{W^i_t} \tag{2.34}$$

where the instantaneous volatility σ_t is equal to the instantaneous local volatility at time t and level S_t, i.e. $\sigma_t = \sigma_{S,t}(t,S)$. Notice also that $\sigma^2_{K,T}$ depends on the shock of the asset process, i.e. the stochastic variations of the local volatility surface may depend on the prevailing level.

Let $P_{K,T} = P(t,S_t,K,T)$ be the total probability that the index level at time T arrives at $S_T = K$ given that the stock price at time t is S. Both the stock price and volatility are assumed to evolve stochastically. Knowing that under the risk neutral measure $P_{K,T}$ evolves as a martingale, i.e.

$$\frac{dP_{K,T}}{P_{K,T}} = \phi_{K,T}d\overline{Z_t} + \sum_i X^i_{K,T}d\overline{W^i_t} \tag{2.35}$$

they show that

$$\frac{d\sigma^2_{K,T}}{\sigma^2_{K,T}} = \beta_{K,T}d\widehat{Z_t} + \sum_i \vartheta^i_{K,T}d\widehat{W^i_t} \tag{2.36}$$

where $d\widehat{Z_t} = d\overline{Z_t} - \phi_{K,T}dt$ and $d\widehat{W^i_t} = d\overline{W^i} - X^i_{K,T}dt$ are the Brownian motions with respect to the new measures. In other words, under these measures, the local variance is a martingale. They call the new measure *the K-level, T-maturity forward risk-adjusted measure*. Letting $E^{K,T}_t(.)$ to denote expectations with respect to this measure, conditional on the information at time t, equation (2.31) can be rewritten as

$$\sigma^2_{K,T} = E^{K,T}_t(\sigma^2_T) \tag{2.37}$$

Hence, in the $K-T$ forward risk-adjusted measure, the local variance $\sigma^2_{K,T}$

at time t_0 amount $V_{K,T}(S_0,t_0)$ if and only if $S_T = K$. If $S_T \neq K$, no exchange takes place. His definition is an alternative way of interpreting equation (2.31).

is the conditional expectation of the future instantaneous variance σ_T^2[20].

2.5.3 Derman and Kani (1998)

Derman and Kani [54] use the concept of local volatility, as this is defined by Kani, Derman and Kamal [103], to present a method for option pricing and hedging based on it. They follow a methodology similar to Dupire's [59], by starting from the initial set of index option prices and their associated local volatility surface. They assume that the asset price and the forward volatility evolve according to equations (2.30) and (2.34), respectively.

The drift coefficients $a_{K,T}(t,S)$ in equation (2.34) must satisfy mild measurability and integrability conditions, as the factor volatility $\vartheta^i_{K,T}$. Moreover, they must also be restricted in such a way, so that the stochastic theory described by equations (2.30) and (2.34), precludes any arbitrage opportunities among the standard options, forwards and their underlying stock. This is similar to the HJM approach, where the drift of the instantaneous forward rate had to be constrained by a no-arbitrage condition, so that forward rates to evolve in a no-arbitrage fashion.

In order to derive this no-arbitrage restriction, they work with the total transition probability $P_{K,T}(t,S)$[21]. They show that the drift functions $a_{K,T}(t,S)$ have to satisfy the following no-arbitrage condition

$$\begin{aligned} a_{K,T}(t,S) \;=\; & -\sum_{i=0}^{n} \vartheta^i_{K,T}(t,S)\{\frac{1}{P(t,S,T,K)} \\ & \int_t^T \int_0^\infty \vartheta^i_{K',T'}(t,S) P(t,S,T',K') K'^2 \frac{\partial^2}{\partial K'^2} P(T',K',T,K) dK' dT' - \Pi^i\} \end{aligned} \tag{2.38}$$

where the quantities Π^i denote the market prices of risk associated with the volatility risk factors W^i, $i = 1,\ldots,n$ and $W^0 = Z$. The double integral of

[20] This is analogous to the relationship between the forward and the future short rate. The forward rate f_T is the T-maturity forward risk-adjusted expectation of the future short rate at time T (see Jamshidian [98]).

[21] Working with total transition probabilities is equivalent to working with option prices, as equation (2.2) shows.

equation (2.38) reflects the two dimensional dependence of the local volatility on K and T, and it makes the continuous time implementation of their model very difficult. Therefore, they prefer implementing their model in discrete time by building a trinomial *stochastic implied tree.*

A stochastic implied trinomial tree is an extension of the implied trinomial tree. The local volatilities and the transition probabilities which correspond to the future nodes, vary stochastically as time elapses and the index level moves. The stochastic tree is the output of an algorithm which combines a standard trinomial implied tree and Monte Carlo simulation. Four steps are necessary for the construction of the stochastic tree. First an implied trinomial tree, like Derman, Kani and Chriss's [53] is built. This provides the initial inputs to the algorithm. Second, the extracted from the implied trinomial tree initial local volatility surface is perturbed. This is done by simulating a discretized version of equation (2.34) which can be written as

$$\Delta\sigma^2_{m,n}(i,j) = \sigma^2_{m,n}(i,j)[a_{m,n}(i,j)\Delta t_i + \sum_{l=0}^{n} \vartheta^l_{m,n}(i,j)\Delta W^l_i] \qquad (2.39)$$

where (i,j) corresponds to the node (t_i, S_j), describing the current location of the stock at the ith step of the simulation. (n,m) labels the future node (t_n, S_m) corresponding to the future time and level (T,K). The evolution of the local volatility surface has to be consistent with no-arbitrage. This is achieved by determining the drift parameters $a_{m,n}(i,j)$ from a "martingale condition" (the volatility parameters $\vartheta_{m,n}(i,j)$ are pre-specified by applying for example Principal Components Analysis). This condition is that the total probabilities $P_{m,n}(i,j)$ of arriving at the future node (n,m) from the (fixed) initial node (i,j) must be jointly martingales for all future nodes (n,m). Once $a_{m,n}(i,j)$ are determined by $P_{m,n}(i,j)$ a random vector $(\Delta W^0_i, \Delta W^1_i, \ldots, \Delta W^n_i)$ is drawn from the distribution of the increments of W^i at time t_i. This vector is used to arrive at new values for all future local volatilities $\sigma^2_{m,n}(i+1,j)$.

In the third step, the $\sigma^2_{m,n}(i+1,j)$ are used for the calculation of the one period

transition probabilities. These probabilities describe the transition from node (i, j) to the up, middle and down nodes at time t_{i+1}. The transition probabilities are calculated from the condition that they must add up to one, and from the conditions which match the first two moments of the discretized process with those of the continuous process. In the fourth step, a new location S_j for the asset price at time t_{i+1}, is determined. This is done by comparing the random draw to the calculated from the third step one period transition probabilities. The third and fourth steps establish the dependence of the evolution of the asset price on the evolution of the local volatilities.

Therefore, the key idea for the construction of the stochastic tree, is to perturb the local volatility surface, so that to exclude arbitrage. Once one of the possible positions of the surface is determined, the stochastic volatility problem is reduced to a deterministic one. Then, the consistent with the new surface, asset price can be traced. Notice that since the state space is fixed for every step of the simulation, it may be the case that large simulated local volatilities produce unacceptable transition probabilities at certain nodes. In this case, the unacceptable probabilities are overwritten and current option prices may not be fitted exactly, as discussed in Derman and Kani [49], and Derman, Kani and Chriss [53]. However, Derman and Kani find that for their stochastic implied tree the overwriting is rarely encountered.

In general, the stochastic implied tree model is a very promising one, and future research should (a) explore the empirical performance of the model, and (b) investigate the number and the nature of the shocks appearing in the local volatility process.

2.5.4 Extracting the Local Volatility Surface

The local volatility surface implied from today's option prices is necessary for initializing Derman and Kani's algorithm. They extract it by building an implied trinomial tree. In this section we are going to present alternative methods for inferring the local volatilities.

Equation (2.33) shows that in principle, we can extract the local volatilities from the option prices. This requires a smooth surface of option prices, so that the partial derivatives appearing in equation (2.33) to be evaluated. In other words, we need to interpolate between the observed strikes and maturities option prices, and to extrapolate beyond them. However, it is always easier to interpolate in the space of implied volatilities, rather in the space of option prices, because the former is smoother than the latter. One way that this can be carried out is by using Shimko's [135] method. The method can be described as follows:

1. Convert the observed option prices to implied volatilities.
2. Use a least squares regression to estimate a quadratic volatility smile.
3. Convert the estimated implied volatilities back into option prices.

To be more precise, Shimko's method is an estimation, rather than an interpolation and the estimated function of option prices does not necessarily go through the original prices. Hence, the extracted local volatilities may not be consistent with today's smile.

Andersen [2] and Andreasen [4] interpolate and extrapolate in the space of implied volatilities, as well, by using a different method from Shimko's. Andersen derives the local volatility surface by using the implicit finite difference scheme that we have already described. Andreasen derives a formula which explicitly expresses the local volatilities in terms of the strike and maturity partial derivatives of implied volatilities. Then, he approximates these derivatives by performing a bicubic spline procedure, so that to achieve smoothness in both directions of the implied volatility surface.

Derman, Kani and Zou [51], and Zou and Derman [150] point out that the calculation of the second derivative is very sensitive to the interpolation method (see also Andreasen [4]). Rather than interpolating, they recognize the property that the second derivative of a European call option with respect to the strike price is a probability density function, and they approximate it with an Edgeworth expansion. This is similar to Rubinstein [128], who approximates the expiration-date risk neutral distribution by means of an Edgeworth expansion in

a discrete-space framework, assuming that the investor knows its skewness and kurtosis. The idea behind an Edgeworth expansion is that the probability distribution function can be expressed in terms of another known distribution plus its derivatives, provided that all moments of both distributions exist (it is similar to the Taylor expansion for analytical functions)[22]. This expansion extracts the asset risk-neutral probability distribution from prices of options that expire at a fixed time. The time derivative is calculated by interpolating the volatility term structure between options' expiration dates by means of a cubic spline. However, it is not certain that even their method is robust because the Edgeworth expansion does not always represent a proper probability density function; there are many intervals for which it could take negative values (see Johnson and Kotz [100]).

2.5.5 Britten-Jones and Neuberger (1998)

A limitation of Derman and Kani's [54] algorithm is that it is quite computer intensive. As a solution to this, Britten-Jones and Neuberger [27] provide a trinomial lattice procedure which is simple and fast. Their aim remains the same: smile consistent arbitrage pricing under stochastic volatility. In order to construct their trinomial tree they establish a sufficient and necessary condition for a process to be consistent with a set of initial option prices.

To make the above more concrete, let time $t = 0, h, 2h, \ldots, T$. They define a grid consisting of nodes of time-price events (t, K) where $K = \{k : k = S_0 u^i, i = 0, \pm 1, \pm 2, \ldots, \pm T/h\}$, and $u > 1$. The single risky underlying asset has initial price S_0, and can move on this grid. Moreover, define

$$\lambda(k,t) \equiv \frac{1}{h} \frac{(1+u)(C(t+h,k) - C(t,k))}{C(t,ku) - (1+u)C(t,k) + uC(t,k/u)}$$

[22] Jarrow and Rudd [99] were the first to apply the Edgeworth expansion method to the problem of option valuation.

and

$$\pi(k,t) \equiv \frac{C(t,ku) - (1+u)C(t,k) + uC(t,k/u)}{k(u-1)}$$

Their first proposition states that the terminal probability that the asset price takes the value k, is given by

$$\Pr(S_t = k) = \pi(k,t)$$

for $0 \leq t \leq T$, $k \in K$. Their second proposition states that if the transition probability for a continuous martingale process is given by

$$\Pr[S_{t+h} \neq S_t \mid S_t = k] = h\lambda(K,t)$$

then this is a sufficient and necessary condition for this process to fit today's option prices. Furthermore, they show that their second proposition is equivalent to extracting the local volatility from equation (2.33).

Consider now a martingale process in which the probability of stock price moves depends not only on time and the stock price, but also on the volatility state Z, where $z = 1, 2, \ldots, N$. Z is assumed to evolve as a time homogeneous Markov process with transition matrix $P = p_{jk}$, where $\Pr[Z_{t+h} = j \mid Z_t = k, F_t] = p_{jk}$. The transition probabilities are chosen exogenously to reflect the type of volatility behavior we think appropriate. Define then the asset transition probabilities as

$$\Pr[S_{t+h} \neq S_t \mid S_t = k, Z_t = z] \equiv h\lambda(k,z,t)$$

We need to calculate $\lambda(k,z,t)$ in order to apply the backward induction for option pricing and hedging (recall that we do not have to worry about the determination of the state space, because it is set exogenously in a trinomial tree). In order to ensure that the model's option prices fit the market option prices, $\lambda(k,z,t)$ has to be calculated consistently with the European option market prices, i.e. their

second proposition has to be satisfied. They show that the proposition holds true if the $\lambda(k,z,t)$ satisfy

$$\lambda(k,t)\pi(k,t) = \sum_{z=1}^{N} \lambda(k,z,t)\pi(k,z,t) \tag{2.40}$$

where $\pi(k,z,t) \equiv \Pr(S_t = k \text{ and } Z_t = z)$ is the joint probability of a particular stock price and volatility state. Notice that the left-hand-side of equation (2.40) can be calculated from today's option prices, while the right-hand side is unobserved.

They calculate $\pi(k,z,t)$ by using the forward Kolmogorov equation. Then, in order to calculate $\lambda(k,z,t)$ they assume that

$$\lambda(k,z,t) = q(t,k)v(z)$$

This is a separability assumption, where the function $v(z)$ is chosen exogenously. The form of the volatility process is defined by the transition probabilities p_{jk}, and the $v(z)$ function which serves for moving local volatilities in a parallel fashion. Then,

$$q(t,k) = \frac{\lambda(k,t)\pi(k,t)}{\sum_{z=1}^{N} v(z)\pi(k,z,t)}$$

This completes the calculation of $\lambda(k,z,t)$.

Britten-Jones and Neuberger's model provides a simple and fast implementable algorithm for smile-consistent pricing under stochastic volatility. It does not have to specify a priori any process for the forward volatility. On the other hand, the separability assumption is crucial for the development of the algorithm, and its implications should be explored further. Finally, the determination of the transition probabilities p_{jk} and of the $v(z)$ function may pose problems to the practitioner.

2.5.6 Ledoit and Santa-Clara (1998)

In general, option pricing and hedging can be done by Monte Carlo simulation if we have the risk-neutral processes for the underlying (traded and non-traded) assets. In Derman and Kani's model [54] this was not possible because of the very complicated no-arbitrage condition that the drift of the local volatility process had to satisfy. Ledoit and Santa- Clara's [106] model overcomes this problem by modeling implied rather than local volatilities.

They start from the observed at time t implied volatilities $V(t,T,X)$ corresponding to an option with time to maturity $T \equiv s-t$, and moneyness $X \equiv \frac{S_t}{K}$, and they allow them to evolve stochastically. In order to simulate jointly the processes for the underlying asset and the implied volatility, they establish a relationship between the stock's instantaneous volatility σ_t and the implied volatility V. Under the assumption that the stock price evolves according to equation (2.30), they show that the Black-Scholes implied volatility of an at-the-money call option converges to the stock price volatility when the time to maturity goes to zero, i.e. $\sigma_t = V(t,0,1)$. Then, they deduce the dynamics of implied volatilities and of the call option price, by applying Ito's lemma on $V(t,T,X)$ and $C(t,S_t,V)$. The evolution of the call option prices has to preclude any arbitrage opportunities, i.e. call option prices have to evolve as martingales. Imposing this martingale constraint they get the risk-neutral process for the implied volatility, i.e.

$$\begin{aligned} dV(t,0,1) = [&\sigma_{V1}(t,0,1)(r-q-\frac{1}{2}V(t,0,1)^2) - \frac{\partial \sigma_{V1}}{\partial X}(t,0,1)V(t,0,1)^2 \\ &+2\frac{\partial V}{\partial T}(t,0,1) - \frac{\partial V}{\partial X}(t,0,1)V(t,0,1)(\sigma_{V1}(t,0,1)+\frac{1}{2}V(t,0,1)) \\ &-\frac{1}{2}\frac{\partial^2 V}{\partial X^2}(t,0,1)V(t,0,1)^2]dt \\ &+V(t,0,1)\sigma_{V1}(t,0,1)d\overline{W_{1,t}} + V(t,0,1)\sigma_{V2}(t,0,1)d\overline{W_{2,t}} \end{aligned} \tag{2.41}$$

where σ_{V1}, σ_{V2} are the volatilities of the implied volatilities.

Even though the risk-neutral process for the implied volatility seems to be relatively simple for a continuous time implementation, there are three issues that needs to be addressed in order to implement their model. First, they assume that the implied volatility process is driven by two shocks. This is an issue that it has to be explored empirically. Second, the coefficients $\sigma_{V1}(t,0,1)$ and $\sigma_{V2}(t,0,1)$ have to be estimated for every t, so that to be used in every step of the simulation. Third, in order to calculate the partial derivatives which appear in equation (2.41), we have to interpolate the estimated volatilities of implied volatilities and the implied volatilities themselves, across moneyness and the time to maturity. Given that the functional form of the implied volatility surface and (probably of the shocks, as well) changes unpredictably over time, the choice of an appropriate interpolation scheme is a difficult task.

2.6 Concluding Remarks

This chapter has reviewed the fast developing literature of smile consistent models. We have recognized the two classes of models which constitute this literature, so far. This allows several issues and models to be presented, compared and discussed. We classify the various smile consistent models in deterministic, and stochastic volatility ones. We remind that the former is nested within the latter, and we bring together the several definitions of the key concept of forward volatility.

Both classes of models start from the observed European option prices, but they differ on the assumption about how the underlying asset evolves. Part of the popularity of deterministic volatility models, can be explained by their theoretical and practical advantages. They preserve the Black-Scholes arbitrage-pricing property and they are tractable enough for implementation purposes. However, the empirical evidence undermines the deterministic volatility assumption. On the other hand, stochastic volatility models allow for the evolution of the (implied or forward) volatility surface, so that to preclude any arbitrage opportunities.

Further empirical work needs to be done to explore the out-of-sample pricing and hedging performance of implied stochastic volatility models, as this is compared to the deterministic volatility one. It is very probable that a jump component should be included (see Das and Sundaram [47]), and researchers should look at how to develop a smile-consistent stochastic volatility-jump model (see also Section 7.2).

Chapter 3

Description and Screening of the Data Set

3.1 Introduction

In this Chapter, we first describe the data set that we will use in our study. Then, implied volatilities are calculated, and a range of screening criteria is applied to reduce the noise in our estimates. The data are daily closing prices on futures options on the Standard and Poor 500 Index, traded in the Chicago Mercantile Exchange for the years 1992-95. Implied volatilities are calculated by the Barone-Adesi, Whaley quadratic approximation. Given that implied volatilities are notorious for the measurement errors in their calculation due to the non-synchronous trading and the bid-ask spreads, we apply a variety of screening criteria, in order to reduce the noise. We find that the screening is satisfactory.

3.2 The Data Set

We use daily data on futures options on the Standard and Poor Index (S&P 500) from the Chicago Mercantile Exchange (CME) for the years 1992-95. CME introduced futures on the S&P 500 Stock Index in 1982. The S&P 500 is a capitalization-weighted index of the 500 listed firms. Each component stock's

price is multiplied by the number of common shares outstanding and the resulting market values are totalled. The total value is then calculated relative to the baseline period (1941-43≡10) to derive the index value. This index has long been the benchmark by which professionals measure portfolio performance, since the market value of the 500 firms is equal to about 80% of the value of all stocks listed on the New York Stock Exchange.

The primary source database for this study is the transaction report "Stats Database", compiled daily by CME. This database contains the following daily information for each option traded: the date, the style (call or put), the options and futures expiration month, the exercise price, the number of contracts traded, the opening, closing, low and high future's and option's price, the opening, closing, low and high bid-ask future's and option's price, and the settlement price. We extract from this, for the purposes of our study, the *closing* options and futures prices.

The futures contracts have maturities every March, June, September and December. Their last trading day is the business day prior to the third Friday of the contract month. Options for long dated expiries, (up to a year) are traded with expiries of March, June, September, December. For the intermediate "serial months", just short dated options, (up to three months) are traded. For the serial months expiry options, the underlying instrument is the futures with the nearest expiry, e.g. for option expiring in 9302, the underlying instrument is the futures expiring in 9303. The last day of trading for the long expiry options is the same day as the underlying futures contract. We calculate the days to maturity by taking into account these specifications.

3.3 The Calculation of Implied Volatilities

Futures options prices on the S&P 500 are functions of five parameters (for the option valuation purposes of a futures option, the cost of carry is zero): the futures price F, the exercise price of the option K, the time to expiration of the

option τ, the riskless rate of interest r, and the standard deviation of the future's price σ, that is

$$C = f(F, K, \tau, r, \sigma) \tag{3.1}$$

$$P = g(F, K, \tau, r, \sigma) \tag{3.2}$$

where C, (P) represents the theoretical prices of a futures call (put) option.

The calculation of the calls and puts implied volatilities $\sigma_{imp,c}$ and $\sigma_{imp,p}$, is obtained by equating the observed (market) calls and puts prices c_M and p_M with the model (theoretical) prices, and inverting equations (3.1) and (3.2) i.e.[1]:

$$\sigma_{imp,c} = f^{-1}(F, K, \tau, r, c_M) \tag{3.3}$$

$$\sigma_{imp,p} = g^{-1}(F, K, \tau, r, p_M) \tag{3.4}$$

Equations (3.3) and (3.4) indicate that the calculation of calls and puts implied volatilities is subject to some sources of errors. These can occur because of (a) the mis-specification of the functions f and g, and (b) measurement errors in r, F, and the market option prices.

The specification of the functions f and g is correct as long as the theoretical model that we use to calculate the implied volatilities, is the same with the one that the market uses to value the options. Given that the S&P 500 futures options are American-style options, we need to use a model which explicitly takes into account the early exercise premium. Therefore, for the calculation of implied volatilities we use the algorithm of Barone-Adesi, Whaley [8] (BAW)[2] [3].

[1]This method of calculation represents a general method known as calibration. In contrast to estimation, calibration is not based on any statistical theory. It just fits the market prices to the model prices, so as to back out some of the parameters of the model under consideration.

[2]Other researchers have used binomial, or finite-difference methods for the calculation of σ_{imp}. For example Diz and Finucane [55] use the binomial American approximation of Cox, Ross, Rubinstein [43]. Canina and Figlewski [30], and Harvey and Whaley [82] use a modification of Cox, Ross, Rubinstein for discrete paid dividends. However, Barone-Adesi, Whaley find that their approximation is accurate, and considerably more efficient than finite-difference or binomial methods.

[3]Notice that there is an inevitable inconsistency here because we aim to investigate the dynamics of stochastic implied volatilities by using implied volatilities extracted from a model

Since option values are non-linear functions of volatility, we have to use an iterative algorithm for the calculation of implied volatilities. The algorithm that we use is the Newton-Raphson (NR) iterative method (see Chapra and Canale [32]). The iterations are performed through the following formula

$$x = x_0 + \frac{f(x) - f(x_0)}{f'(x) \mid_{x=x_0}} \tag{3.5}$$

where

x is the implied volatility and x_0 is the initial guess for the implied volatility,

$f(x)$ is the market option price, $f(x_0)$ is the BAW option price evaluated at x_0, and $f'(x) \mid_{x=x_0}$ is the first derivative of the model with respect to the volatility evaluated at $x = x_0$.

We let maximum 100 iterations for the method to converge (by convergence we mean $f(x) - f(x_o) < 0.00001$). If it does not converge within 100 iterations, then we set the implied volatility equal to zero. This is because NR is a method which converges very quickly. If the convergence has not occurred within 100 iterations, then our model does not support the observed market option price, and therefore those prices should be excluded[4].

Another issue is whether the interest rate should be selected from Treasury bills, commercial paper, or Eurodollars. However, it seems that there is a small impact on the estimation of implied volatilities from using the wrong interest rate (see Bates [12]). We are using London Euro-Currency interest rates (middle-rates) on the US dollar, obtained from Datastream to proxy for the riskfree rate, as Clewlow and Xu [37] do[5].

The London Euro-Currency interest rates were chosen because they consist of

which assumes constant variance. However, "examining volatilities inferred under the Black-Scholes model is a reasonable and informative initial diagnostic of volatility dynamics", as Bates [12] notes.

[4] In most of the cases that no convergence had occured, we had violation of the lower arbitrage boundary conditions. Therefore, those market option prices could not be supported, not only by BAW model, but by no model.

[5] Middle rates are the midpoints between the bid and offered rate. We use middle rates because they are a kind of an "average" of the offer and the bid rate.

the maximum number of different maturities that we could use, in order to make the interest rates used in the calculation of implied volatilities, as accurate as possible. Daily interest rates for 7-days, one-month, three months, six months, and one year were used, while for the other maturities were obtained by linear interpolation[6]. In this way we use a comparable-maturity to the option yield.

In order to do the interpolation we convert the interest rates to continuously compounded by using the formula

$$r_2 = \frac{365}{n} \ln(1 + \frac{r_1 * n}{360}) \tag{3.6}$$

where r_1 is the rate for a period n[7], r_2 is the resulting continuously compounded rate, 360 is the days in a year according to the convention for the US dollar, and 365 is the days in a year.

3.4 Screening the Data

In this section, we describe how we screen our data. In the first stage, the raw data are screened for data errors. In a second stage, we exclude data likely to introduce errors into our volatility estimates.

3.4.1 Screening the raw Data

We eliminate data where the option price is less than, or equal to its intrinsic value. Otherwise, a riskless arbitrage could arise[8]. We also exclude options having

[6]The linear interpolation is done as follows: let r_t be the interest rate for $t \in A$ where $A = (\ 7$ days, 1 month, 3 months, 6 months, 1 year) and let τ be the required time to maturity where $a < \tau < b$, with $a, b \in A$ and a, b consecutive elements of A. Then the interpolated rate $R(\tau)$ which corresponds to time to maturity τ is $R(\tau) = r_a + \frac{(\tau - a)(r_b - r_a)}{b - a}$

In order that the options and interest rate data sets have the same days, we excluded from the interest rates data sets the redundant days.

[7]The Datastream Interest Rate convention is that 1 month interest rate, for example, means that if we lend money we are going to get it back after 30 days. The meaning for the other maturities' interest rates is analogous.

[8]It should be the case that we have already taken into account the violation of the arbitrage boundary conditions in the routine which calculated the implied volatilities.

a price of less than 10 cents (see Clewlow and Xu [37], and Xu and Taylor [149]). This is because for options with prices smaller than 10 cents, transaction costs including the bid-ask spread, and liquidity premia are large relative to the option price. Finally, we eliminate short term options with less than 10 days to maturity. This is because as the time to maturity gets shorter, vega becomes smaller for most strikes for these options, and consequently implied volatilities become very sensitive to small errors in the option price[9]. In addition, Day and Lewis [46] find that significant increases in implied volatility occur during the week prior to maturity.

3.4.2 Screening Implied Volatilities

One of the sources of measurement errors in the calculation of σ_{imp} is the non-synchronous trading. Non-synchroneity means that the option market does not close at the same time with the underlying asset's market. Harvey and Whaley [81] find that in the presence of non-synchroneity, spurious negative serial correlation is induced in the volatility changes if closing option prices are used[10].

Another source of noise in the calculation of implied volatilities is the bid-ask spread on both the option and the underlying asset. The bid-ask spread effect refers to the fact that either the option's or the underlying asset's (or both) closing price, is quoted at a bid, or an ask level. Roll [124] shows that the bid-ask spread produces significant spurious negative first-order serial covariance in price

[9]Other researchers have used different cut-off points. For example, Canina and Figlewski [30] eliminate options with less than 7 and more than 127 days to maturity. Harvey and Whaley [80] eliminate options with less than 21 days to maturity. Our choice is not crucial because we deal with the noisy data with additional constraints that we describe later.

[10]Even if the option and the underlying asset's markets close at the same time, the non-synchroneity effect can occur due to a reported option price which is based on a stale futures price. Let C_{t_1} be the options price based on F_{t_1} at time t_1. The calculated true implied volatility would be σ. Non-synchronous reporting might have a stale C_{t_1} matched with a later up-to-date F_{t_2} $(t_2 > t_1)$. If the market went up $(F_{t_2} > F_{t_1})$, C_{t_1} is too low and we would calculate an implied volatility too low, as well (lower than σ). The next day, when both markets open, we are going to have synchronous data and the implied volatility will go up to the correct level. The error in the calculation of the implied volatility is greater for ITM options where the error in C_{t_1} is bigger by an amount $\Delta \times \Delta F_t$; Δ is the delta of an option, and ΔF_t is the movement in the futures price.

changes series[11].

The trading hours for the S&P 500 Index and for the S&P 500 futures Options are 8.30 am. till 3.15 p.m. (Chicago time). Even though, the closing times for the option's and the underlying asset's market coincide, deep in-the money (ITM), and out-of the money (OTM) options are rather illiquid and they trade less frequently than those nearer to the-money. Consequently, for these options non-synchroneity, and bid-ask spreads is a larger problem.

A common way of eliminating noisy data is to exclude deep ITM, and OTM options. However, the way that the range of the OTM and the ITM options to be eliminated is chosen, differs between researchers and seems to be somewhat subjective[12]. This may have as a consequence that either a significant amount of noise still remains in the data, or that valuable information has been lost.

We use criteria which minimize this trade-off. We construct our smiles by using OTM puts for low strikes, and OTM calls for the high ones (see also Kamal and Derman [102], and Zou and Derman [150]). The data on ITM calls and puts are not used because they have high deltas; their prices and their implied volatilities are therefore very sensitive to the non-synchroneity problem. Provided the put-call parity relationship holds (there is a simple arbitrage if it does not), puts and calls must have identical implied volatilities (in the absence of measurement error problems). Our procedure reduces the effects of errors due to non-synchronous data, without introducing any sources of bias.

[11] If the recorded option price comes from a transaction executed at the market's ask price, it will appear to be relatively expensive and its implied volatility will be high. If the option price was the bid, implied volatility will be lower. Similarly, if the contemporaneous underlying asset price which is used in the calculation of implied volatility comes from a trade at the market's ask price, a call will appear to be deeper in-the-money than it really is. Hence, the implied volatility will be artificially increased.

[12] Canina and Figlewski [30] eliminate options that are more than 20 points in or out-of-the money. Rubinstein [125], and Sheikh [134] eliminate options with a ratio (S/K) less than 0.75 (S denotes the asset price). Xu and Taylor [149] eliminate options when $K < 0.8S$, or $K > 1.2S$.

ΔC	$\Delta\sigma_{imp}$	v
0.05	$\rightarrow \infty$	0
0.05	0.05	1
0.05	0.025	2
0.05	0.0125	4
0.05	0.00625	8
0.05	0.003125	16

Table 3.1: Upper bound in the Error in Implied Volatilities, set by Different Values of Vega.

3.4.3 The Vega Constraint

Finally, we decided to exclude implied volatilities calculated from options having a vega less than eight (vega constraint)[13]. For a given measurement error ΔC in the option price (e.g. arising from the bid-ask spread), the resulting error in the implied volatility $\Delta\sigma_{imp}$ is approximately $\Delta\sigma_{imp} = \frac{\Delta C}{V}$. Hence, to constrain $\Delta\sigma_{imp}$, vega has to be greater than a certain cut-off point. Table 3.1 shows the upper bound of error in the calculated implied volatilities, that is set by several values of the vega constraint. The calculations are done by assuming that the measurement error in the option prices is 0.05. Clearly, the constraint on $\Delta\sigma_{imp}$ is tighter as vega increases.

The advantages of the vega-constraint over the constraint on $\frac{F}{K}$ that other researchers have used so far, are that (a) the cut-off point is not determined arbitrarily, but it has a sensible reasoning, and (b) the vega-constraint takes explicitly into account the days to expiry of the option, while the ratio $\frac{F}{K}$ does not.

The choice of the vega cut-off point is not obvious because there is a trade-off between accuracy, and the number of observations that we exclude. Figure 3.1 shows the variation of vega against the strike, for a European option with 10, 50, and 200 days to expiry, when the futures price is 420, the volatility is 13%, and the interest rate is 4.167%. As vega increases, we constrain the errors $\Delta\sigma_{imp}$, but

[13]The vega V of an option is defined as $Vega = \frac{\partial C}{\partial \sigma_{imp}}$.

we exclude a larger amount of data, especially in the shorter expiries.

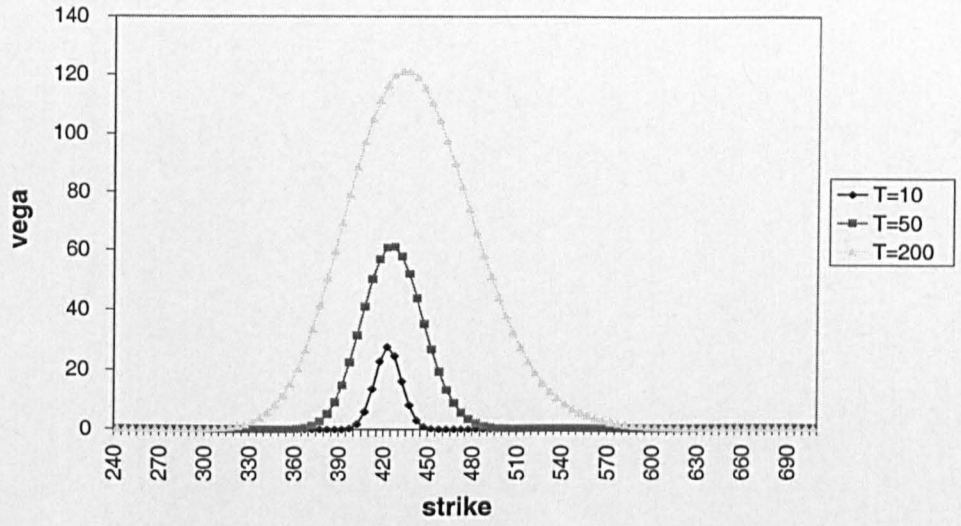


Figure 3.1: Variation of Vega as a function of Strike, for European Option with 10, 50, and 200 days to maturity.

The choice of eight for the cutoff was made after checking the number of observations that we exclude, and the amount of noise in the remaining data, for different cut-off points. Tables 3.2, and 3.3 show the percentage of observations that we exclude over the years 1992-95, from call and put contracts, respectively, maturing in May and September. The second column shows the percentage of observations excluded after eliminating the ITM options and after applying the screening criteria described in Section 3.4.1. The remaining columns show the additional percentage of observations excluded for each vega cut-off point[14]. Setting the vega constraint to eight retains a satisfactory number of observations (52%-30% of the observations for calls, and 90-50% of the observations for puts).

[14]Without loss of accuracy, we use the vega for European-style futures options, as calculated from Black's [19] model, i.e.

$$v = e^{-r\tau} F \sqrt{\tau} n(d_1) \tag{3.7}$$

where $n(x)$ is the normal density function of a random variable x, and $d_1 = \ln(\frac{F}{K}) + \frac{r + \frac{\sqrt{\sigma}}{2}\tau}{\sigma\sqrt{\tau}}$ (v is the same for calls and puts). This is because the early exercise privilege of American futures options contributes significantly to the value of only ITM futures options (Whaley [146]).

Call	Other Criteria and ITM	v = 1	v = 2	v = 4	v = 8	v = 16
9205	45.01%	0%	0%	0.2%	3.46%	9.78%
9209	46.95%	0%	0%	0%	0.51%	2.58%
9305	48.86%	0%	0%	2.28%	9.61%	10.88%
9309	56.29%	0%	0%	0%	4.78%	4.78%
9405	52.19%	0%	0%	1.29%	6.7%	11.86%
9409	48.63%	0%	0%	0.11%	4.11%	9.19%
9505	62.64%	0%	0%	0.22%	2.46%	5.26%
9509	68.73%	0%	0%	0.05%	0.58%	1.3%

Table 3.2: Percentage of Excluded Observations for Call Contracts after eliminating the ITM options, applying the Different Screening Criteria and imposing the Vega Constraint for Different Vega Cut-off Points.

Put	Other Criteria and ITM	v = 1	v = 2	v = 4	v = 8	v = 16
9205	34.27%	0%	0.99%	3.98%	9.94%	21.36%
9209	28.98%	0%	0.55%	1.84%	4.65%	9.97%
9305	44.22%	0%	0.02%	3.11%	10.36%	21.45%
9309	32.37%	0%	1.87%	3.66%	8.32%	15.96%
9405	46.84%	0%	0%	1.22%	3.68%	10.48%
9409	44.76%	0%	0.34%	0.91%	2.85%	5.69%
9505	37.09%	0%	3.94%	8.25%	17.04%	27.86%
9509	6.73%	0%	0.66%	1.37%	2.7%	4.7%

Table 3.3: Percentage of Excluded Observations for Put Contracts after eliminating the ITM options, applying the Different Screening Criteria and imposing the Vega Constraint for Different Vega Cut-off Points.

In order to check whether the retained data are noisy, we consider the implied volatilities of option contracts within the interval 30-27 days to expiry[15]. If the data are clean, then the implied volatilities for each strike should not differ a lot over such a short time interval, i.e. the standard deviation of implied volatilities for each strike should be small.

For each strike, we calculate the average and the standard deviation of these four implied volatilities. Then, we plot the average and plus, minus one standard deviation. We call these graphs the "accuracy graphs". We construct such graphs for all the years, for both the call and put contracts maturing in the May and September. Figure 3.2 shows such a graph for a call contract which matures in September 1993 (9309). We can see that the implied volatilities of ITM calls have very big standard deviations, i.e. they are very noisy, as expected.

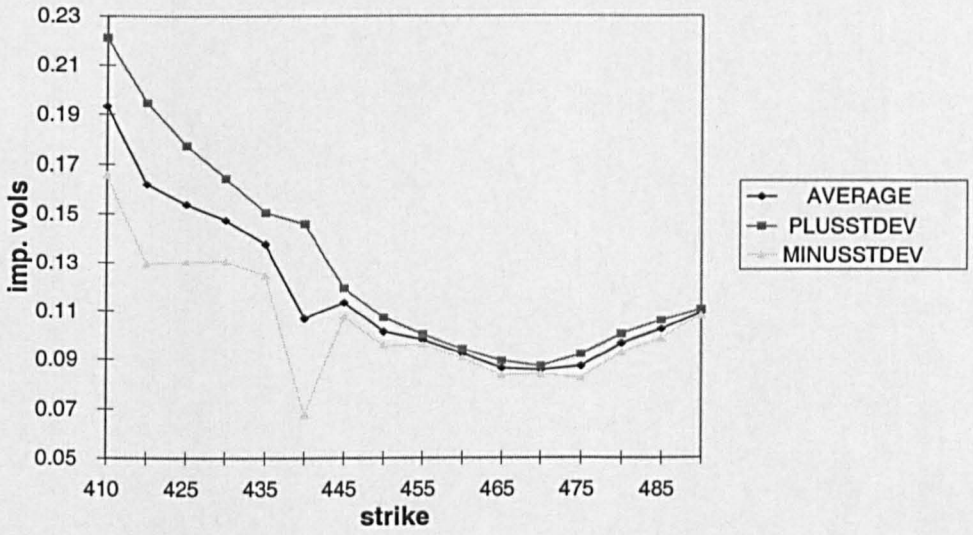


Figure 3.2: Raw Implied Volatilities for a Call Contract with Maturity 9309.

Then, we construct similar graphs when we impose *only* the vega constraint, for $v = 1, 2, 4, 8, 16$ (notice that the various graphs are drawn to different scales).

Figure 3.3 shows the screened implied volatilities of a call contract which matures in 9309 when we have imposed on it the vega constraint with a value of

[15]We choose this range of days to maturity, because options are actively traded within it. Hence, we have information for a wide range of strikes, something which enables us to check the amount of noise in our data.

8. We can see that, compared to Figure 3.2, the effect of the vega constraint is to trim the extreme strikes, i.e. the deep ITM and OTM calls. However, the big amount of noise for the ITM calls remains, and therefore the elimination of them needs to be carried out, as already explained.

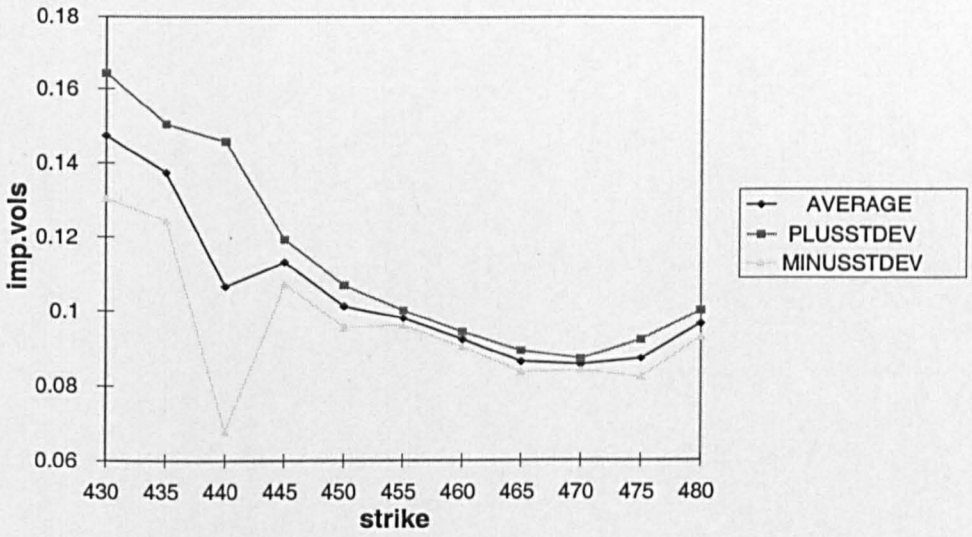


Figure 3.3: Screened Implied Volatilities Only with the Vega Constraint for v=8, for a Call Contract with Maturity 9309.

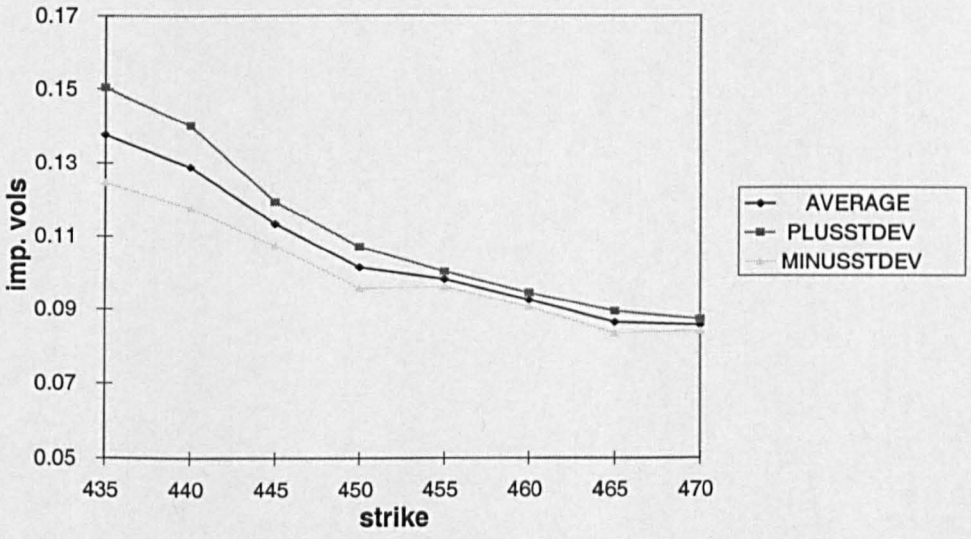


Figure 3.4: Screened Implied Volatilities Only with the Vega Constraint v=16, for a Call Contract with Maturity 9309.

Figure 3.4 shows the screened implied volatilities of a call contract which matures in 9309 when we have imposed only the vega constraint, with a value

of 16. There is of course a further reduction of noise, but given the amount of information that we exclude as vega increases (see Table 3.2), and given that the noise for ITM calls remains, we choose the cutoff point for our vega constraint to be 8.

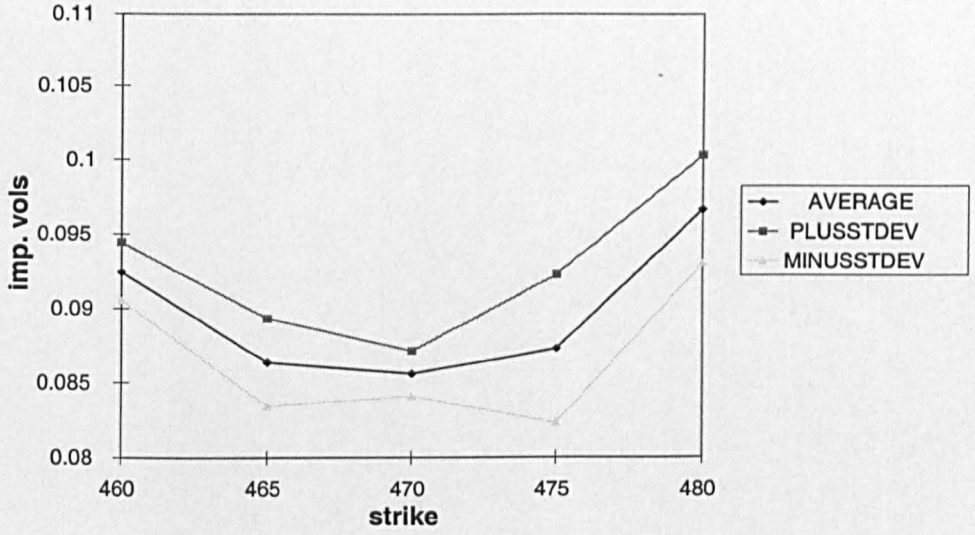


Figure 3.5: Screened Implied Volatilities for a Call Contract with Maturity 9309.

Figure 3.5 shows what the screened implied volatilities of a call contract with maturity 9309 look like, when we have screened the raw data, we have eliminated the ITM calls, and we have imposed the vega constraint with a value of 8. We can see that the amount of noise has been decreased considerably, compared to Figure 3.2.

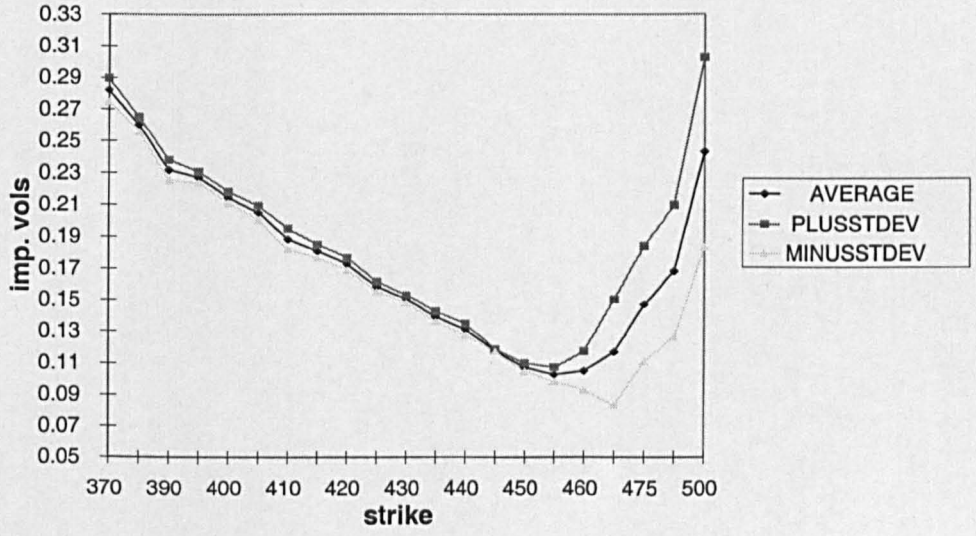


Figure 3.6: Raw Implied Volatilities of a Put Contract with Maturity 9309.

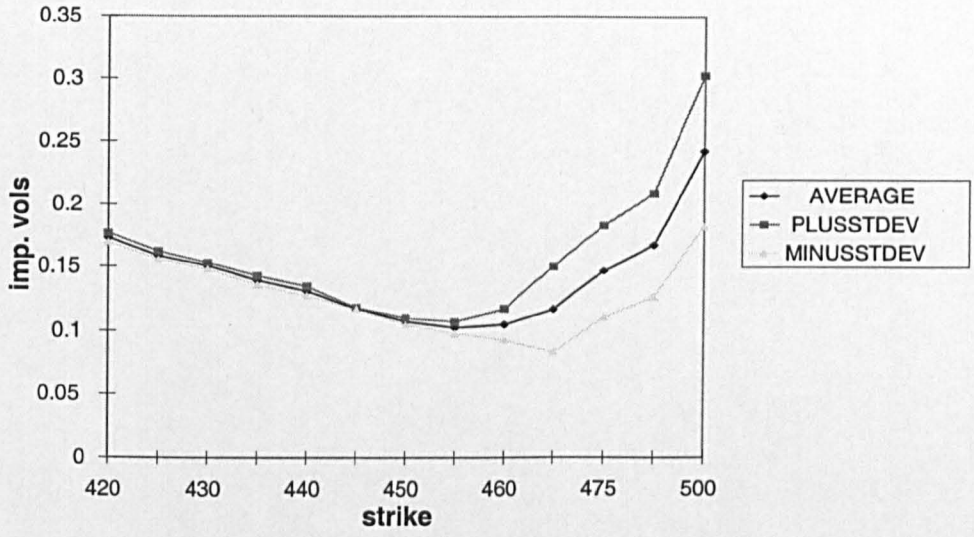


Figure 3.7: Screened Implied Volatilities Only with the Vega Constraint for v=8, of a Put Contract with Maturity 9309.

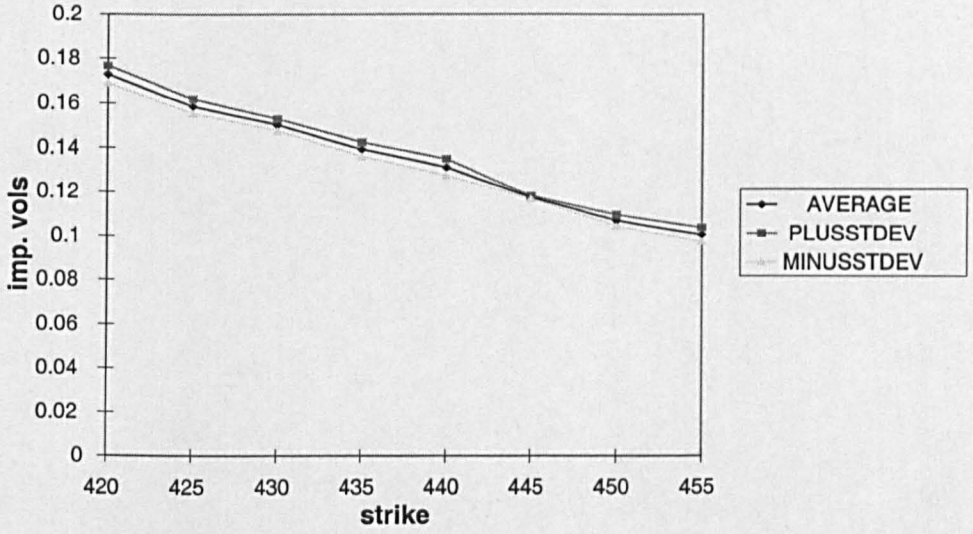


Figure 3.8: Screened Implied Volatilities of a Put Contract with Maturity 9309.

Similarly, we have constructed accuracy graphs for a put contract which matures in 9309. Figure 3.6 shows that the ITM raw implied volatilities are very noisy. Figure 3.7 reveals that the imposition of the vega constraint with a value of 8 trims the OTM puts, but it does not reduce the noise for the ITM puts. Therefore, the elimination of the ITM puts is necessary. Figure 3.8 shows that

the put implied volatilities of the put contract do not exhibit noise after applying all the screening criteria.

Our screening procedure ensures that our data are sufficiently smoothed, so as to get reliable results from our subsequent analysis. This is confirmed by Figure 3.9 where we show the recovered from OTM calls and OTM puts implied volatility skew.

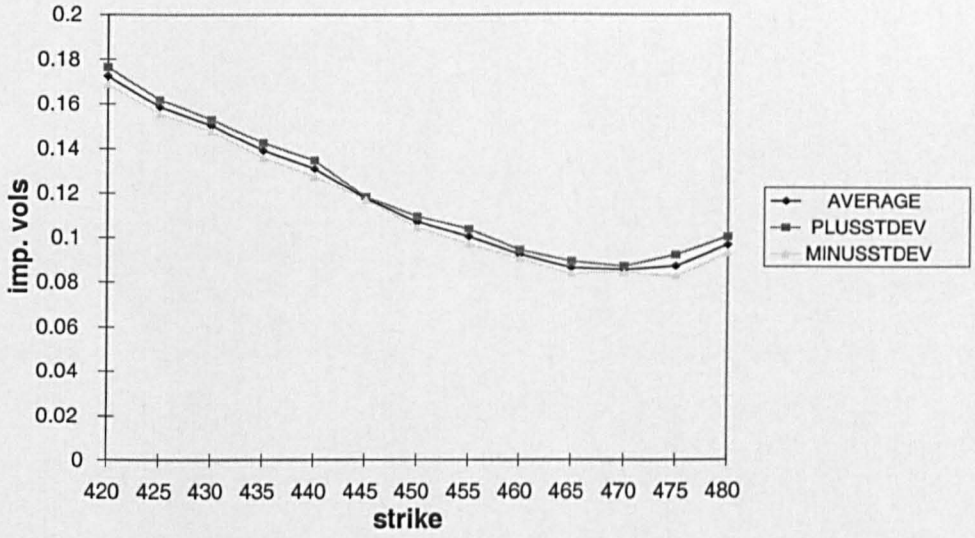


Figure 3.9: Recovered Smile from OTM Calls and OTM Puts with Maturity 9309.

3.5 Summary

We use closing prices on futures options on the S&P 500 for the years 1992-95. Our goal is to investigate the dynamics of their implied volatilities. To achieve this, we need to exclude the noise in the data, without losing any valuable information. We dealt with the possible sources of measurement errors, and noise by applying a range of criteria. We checked their efficacy, and we found that they are successful in delivering credible data to work with.

Chapter 4

The Dynamics of Smiles

4.1 Introduction

The developing literature on "smile consistent" no-arbitrage stochastic volatility models (Dupire [59], [61], Derman and Kani [54], Ledoit and Santa-Clara [106]) has been motivated by the need to price and hedge exotic options consistently with the prices of standard European options. The objective of this chapter is to investigate the dynamics of implied volatility smiles, since this is a necessary prerequisite for the implementation of these models.

In recent years considerable interest has focused on the behavior of the implied volatilities of options contracts, derived from inverting Black-Scholes [20] (BS) model. The empirical evidence shows (see Rubinstein ([125], [126]), Derman and Kani [54]) that implied volatilities vary across different strikes (smiles or skews), and different times-to-expiration (term structure) for options at the same point in time. In addition, implied volatilities also vary, in a stochastic way, across different points in time for a given option (Dumas, Fleming, and Whaley [58], Gemmill [73], Jackwerth and Rubinstein [94]). These results suggest that implied volatilities could be viewed as a two-dimensional surface which evolves over time.

A number of option pricing models have been proposed which give rise to smiles or skews, and to a term structure of implied volatilities, roughly similar to what is observed empirically. However, none of these models fits observed implied

volatility patterns well, making it difficult to use them in practice to price and hedge exotic options (see Section 2.1). These problems have led to the recent literature on "smile consistent" no-arbitrage stochastic volatility models.

This "evolutionary" approach is similar to the Heath, Jarrow, Morton (HJM, [84]), and Ho and Lee [86] methodology for stochastic interest rates, and was originally inspired by it (Dupire [59], [61]). The models take today's option prices (or equivalently the implied volatility surface) as given and they let them evolve stochastically in such a way as to preclude arbitrage. This allows for correct pricing of standard options and is relevant to the pricing of exotic options. For example, Derman and Kani [54] start from today's option prices and postulate a stochastic process for the forward volatility. They then find the no-arbitrage condition that its drift must satisfy. Ledoit and Santa-Clara [106] propose a model where the implied rather than the local volatilities are used to obtain a simpler than Derman and Kani's no-arbitrage condition that the drift of the implied volatility process must satisfy.

Practical implementation of this class of models requires us to understand the dynamics of the implied volatility smile, and surface. In this chapter, we explore the three questions related to the former: (1) how many factors are needed to explain the dynamics of the volatility smile? (2) what do these factors look like? and (3) how are these factors correlated with the innovation in the underlying asset's process?

In order to answer these questions we apply Principal Components Analysis (PCA) to the first differences of implied volatilities, as these are measured in two different metrics: the strike metric and the moneyness metric.

After considering a variety of criteria, we find two components driving the dynamics of implied volatility smiles, in both metrics. We construct a Procrustes rotation method, so as to obtain a simple interpretation for them. The first rotated component is interpreted as a shift, and the second has a Z-shape. The two components explain between 56% and 85% of the variance of implied volatilities in the strike metric, and between 65% to 92% of the variance in the moneyness

metric.

This chapter consists of eight sections. In the second section, we describe the PCA technique, including why it was chosen, and how it will be applied to analyze the data. In the third and fourth section we apply PCA on the strike and the moneyness metric. In the fifth section we check whether the technique is robust to the presence of any serial correlation in the data. Section six compares the results obtained from the analysis in the two metrics. In section seven, we calculate the correlations between the changes of the principal components and the underlying asset price. Concluding remarks are presented in the last Section.

4.2 Principal Components Analysis and Implied Volatilities

4.2.1 Description of the Principal Components Analysis

In this section, we outline the Principal Components Analysis (PCA) methodology used in this (and the next) chapter. PCA is used to explain the systematic behavior of observed variables, by means of a smaller set of unobserved latent random variables. Its purpose is to transform p correlated variables to an orthogonal set which reproduces the original variance-covariance structure.

We apply PCA to decompose the variance-covariance structure of first differences of implied volatilities. To achieve this, we measure the daily differences of implied volatilities across different strikes (or different levels of moneyness) and different ranges of days to expiry (expiry buckets). For example, one of our variables provides a time series of the first differences of implied volatilities which correspond to a moneyness level of (-1.5%) and expiry range 10-30. Typically, for each expiry bucket we have 7-10 levels of strike (moneyness) in each year, with 100-225 observations for each level.

In general, denote time by $t = 1, \ldots, T$ and let p be the number of variables. Such a variable is a $T \times 1$ vector $\mathbf{x}$. The purpose of the PCA is to construct

Principal Components (PCs hereafter) as linear combinations of the vectors $\mathbf{x}$, orthogonal to each other, which reproduce the original variance-covariance structure. The first PC explains most of the overall variance of the p variables (maximization problem). The second PC explains the most of the remaining variance of the p variables, and so on. The coefficients with which these linear combinations are formed are called the *loadings*. In matrix notation

$$\mathbf{Z} = \mathbf{XP} \tag{4.1}$$

where $\mathbf{X}$ is a $(T \times p)$ matrix, $\mathbf{Z}$ is a $(T \times p)$ matrix of PCs, and $\mathbf{P}$ is a $(p \times p)$ matrix of loadings. The first order condition of this maximization problem results to

$$(\mathbf{X}'\mathbf{X} - l\mathbf{I})\mathbf{P} = \mathbf{0} \tag{4.2}$$

where l_i are the Lagrange multipliers. From equation (4.2) it is evident that the PCA is simply the calculation of the eigenvalues l_i and the eigenvectors $\mathbf{P}$ of the variance-covariance matrix $\mathbf{S} = \mathbf{X}'\mathbf{X}$. Furthermore, the variance of the *ith* PC is given by the *ith* eigenvalue, and the sum of the variances of the PCs equals the sum of the variances of the $\mathbf{X}$ variables.

Equation (4.1) delivers $\mathbf{Z}$ as a $(T \times p)$ matrix whose columns consist of unstandardized components. Standardizing them to unit length we get a new matrix $\mathbf{Z}^*$ which is $\mathbf{Z}^* = \mathbf{ZL}^{-\frac{1}{2}} = \mathbf{XPL}^{-\frac{1}{2}}$ where now $\mathbf{Z}^{*'}\mathbf{Z}^* = \mathbf{I}$ and $\mathbf{L}$ is a $(p \times p)$ diagonal matrix having on the diagonal the eigenvalues l_i. Hence, $\mathbf{X} = \mathbf{Z}^*(\mathbf{PL}^{-\frac{1}{2}})^{-1} = \mathbf{Z}^*\mathbf{L}^{\frac{1}{2}}\mathbf{P}'$ $= \mathbf{Z}^*\mathbf{A}'$ i.e.

$$\mathbf{X} = \mathbf{Z}^*\mathbf{A}' \tag{4.3}$$

where

$$\mathbf{A}' = \mathbf{L}^{\frac{1}{2}}\mathbf{P}' \tag{4.4}$$

When both variables and components are standardized to unit length, the elements of $\mathbf{A}'$ are correlations between the variables and PCs, and they are called *correlation loadings* (see Basilevsky [9] for more details). If we retain $r < p$ PCs

then

$$\mathbf{X} = \mathbf{Z}_{(r)}\mathbf{A}'_{(r)} + \boldsymbol{\varepsilon}_{(r)} \tag{4.5}$$

where $\boldsymbol{\varepsilon}_{(r)}$ is a $(T \times p)$ matrix of residuals and the other matrices are defined as before having r rather than p columns. The percentage of variance of $\mathbf{x}$ which is explained by the retained PCs (*communality* of $\mathbf{x}$) is calculated from the correlation loadings. After retaining $r < p$ components, we look at equation (4.5) to examine the size of the communalities, and the meaning of the retained components.

PCA is a natural and parsimonious technique to identify the number, and the interpretation of stochastic shocks that move the implied volatilities. It enables us to simplify the complex dynamics of the volatility surface, by identifying its most important components, without imposing any prior structure. This contrasts, for example, with the alternative regression analysis approach of estimating a specific function of time and moneyness (see Taylor and Xu [141]). PCA is preferred to factor analysis, for our analysis, because of its variance maximization property (for a comparison of the two techniques see Wilson [148]). Efficient market considerations suggest that serial correlation effects in implied volatility innovations are likely to be weak, and we seek to understand the nature of *contemporaneous* shocks across the whole volatility structure. Thus, vector autoregressions (Hamilton [79]) are not suitable for our study.

4.2.2 The Metrics

We investigate the dynamics of implied volatilities by performing PCA on the first differences of implied volatilities. Over short time horizons, implied volatilities can be viewed as forward volatilities, and therefore we expect them to follow approximately a random walk. Derman and Kani [54] for example, show that forward volatilities are martingales under the appropriate measure. If this was not the case, then it would be possible to attain large profits by trading volatility-

i.e. there would be gross violations of market efficiency. Therefore, they are non-stationary and they need to be differenced once, because PCA is misleading when applied to non-stationary variables (Frachot, Jansi and Lacoste [68])[1].

The variables $\mathbf{x}$ of changes of implied volatilities to which we apply the PCA will be indexed in two different ways (*metrics*): (a) the strike level (strike metric), and (b) the moneyness level $\frac{K - F_t}{F_t} * 100$ (moneyness metric). For a fixed day, the smiles (or skews) are going to look the same in both the strike and moneyness metric. However, the dynamics of implied volatilities will be different across the two metrics (see Section 4.7, Proposition 1).

The strike metric is chosen because the arguments of Derman, Kani and Zou [51] (see Section 2.2.1) suggest that it is a natural metric to examine the dynamics of deterministic volatility models. This is because the variation of local volatilities with the underlying asset price, implies a variation of the implied volatility with the strike price. However, we need to bear in mind that their arguments assume implicitly an instantaneous change in the asset price. This is not the case in the framework of our analysis since we deal with changes in the underlying asset over yearly periods. This may pose difficulties in comparing the results obtained from applying the PCA for each year separately. This is due to the fact that the options which are traded over the years are not the same; since the futures price is drifting upwards, many options are becoming far away from the money and new options with higher strikes have to be introduced.

The moneyness metric is chosen because there are theoretical reasons (see Heynen [85], Taylor and Xu [140] and [141]) suggesting that the dynamics of smiles are a function of moneyness. Heynen [85] shows this by deriving necessary conditions for the implied volatilities of options with different exercise prices and

[1]We confirmed the non-stationarity of implied volatilities by testing for unit roots using different specifications of the augmented Dickey-Fuller test (Greene [76]). The null hypothesis of a unit root was accepted. Our results do not conflict with the results of Poterba and Summers [119] and Hsieh [88] who find that over long periods, a time series of implieds with fixed expiry, e.g. three months (rather than a time series of implieds with fixed expiry date, as in our case) is mean reverting.

the same time to maturity, to exhibit a U-shaped pattern[2]. Taylor and Xu show that implied volatilities are approximately a quadratic function of moneyness when volatility is stochastic, volatility risk is not priced, and asset price and volatility differentials are correlated (see Taylor and Xu [141]), or uncorrelated (see Taylor and Xu [140]). Moreover, Taylor and Xu's approach implies that the moneyness metric is the appropriate metric for the investigation of the dynamics of a stochastic volatility model. If a deterministic volatility model is the true model, then we expect to find "small" variation of implied volatilities in the strike metric, and "significant" variation in the moneyness metric. The reverse will happen if a stochastic volatility model is the correct model[3].

4.2.3 The Determination of the Expiry Buckets

Next, we group the data into different buckets for distinct ranges of days to expiry. We need to control for the time to expiry because we expect the implied volatilities of the shorter dated options to vary more than those of the longer dated ones. This is a feature of various models (such as Stein [137]) and earlier empirical studies e.g. Bates and Clewlow [16], Hsieh [89], and Taylor and Xu [149][4] [5].

The expiry buckets were chosen, so as to cope with two constraints: (a) getting

[2]Heynen concludes that the only pricing density functions that generate the smile pattern are those that are higher peaked than the lognormal density function. This result contrasts the usual conception that pricing density functions that have fatter tails than the lognormal distribution, should generate a U-shaped implied volatility pattern. In addition, it shows that we can get a clear smile picture even when a limited range of strikes is available, and we have no information about the precise shape of the tails of the risk-neutral distribution.

[3]However, we need to point out that we do not have a yardstick to measure the magnitude of the variation of the empirical implied volatilities in either metric. To do that, we should compare the variation of empirical implieds with the simulated, from a deterministic, or a stochastic volatility model, implieds in both metrics.

[4]Stein shows that under the assumption that volatility follows a mean-reverting process with a constant long-run mean, and a constant coefficient of mean reversion, shorter expiry implied volatilities are more volatile than longer expiry ones.

[5]There are two other possible methods for accommodating the effect of the time to expiry on the variance of changes in implied volatilities. We could interpolate to get series for fixed expiries, or we could use a parametric model which uses the expiry date as an explanatory variable. We prefer grouping the data into buckets because the expiries in our data are too thinly spaced, and because we do not want to impose any prior structure.

a sufficient amount of data for each range in order to perform the PCA, and (b) treating the missing observations that occur due to the screening criteria that we have applied.

Two broad strategies are often adopted when data are randomly missing: (a) deletion, i.e. delete the missing cases and get estimates for them from the complete sample and (b) estimation e.g. replace the missing values with the means of the variables[6]. We do not replace the missing values, because this may bias the results from our analysis. Instead, we are going to apply *listwise deletion* i.e. we will delete the whole day for which at least the observation for one variable is missing[7]. Listwise deletion is applied after taking the first differences, so as differences in implieds are 1-day differences.

Table 4.1 shows the number of variables, and of observations, after applying the listwise deletion, for several candidate expiry buckets in year 1992. We expect that in the other years more observations will fall in these expiry buckets because the liquidity of the option markets increases. Since we would like to have expiry buckets as fine as possible, we fix six intervals of days to expiry: 30-10, 60-30, 90-60, 150-90, 240-150 and 360-240 days[8]. The dynamics of individual smiles are analyzed by applying PCA separately to the expiry buckets. Then, in Chapter 5 we analyze the dynamics of the whole implied volatility surface by pooling the buckets together, and performing the PCA on the whole data set.

To summarize, in our framework, each variable $\mathbf{x}$ is a time series collection of differences in implied volatilities for a given strike (moneyness) level, and within a certain expiry range. The first PC is the linear combination of these variables

[6]The literature on treating missing observations is vast, but as Anderson et al. [3] note "The only real cure for missing data is to not have any".

[7]Listwise deletion is one of the three methods which belong to the category of the so-called zero-order methods. The other two are (a) pairwise deletion of individuals for covariance calculations for whom one variable in the pair is missing, and (b) replacing missing values by mean values of the variables.

[8]In some cases, for a given variable there are two contracts traded on the same day, which fall in the same expiry bucket, but they have different days to expiry. For example, in the range 60-30 there is one contract with 30 days to maturity, and another one with 59. We keep the contract which ensures that our variable is a time series, i.e. if in the previous day there was a contract with 31 days to expiry, we keep the 30 days to expiry contract.

Expiry Buckets	30-10	60-30	90-60	120-90	150-90	240-150	270-150	360-240
No of Variables	8	12	9	18	16	11	11	6
No of Obs.	100	206	121	78	161	215	210	212

Table 4.1: Determining the Expiry Buckets: Checking the Number of Variables and Observations for Different Candidate Expiry Buckets in 1992.

which contributes the most to explaining the overall variance of the changes in the implied volatility surface. The second PC is the linear combination (of the same variables) orthogonal to the previous one, which explains the most of the remaining variance, and so on.

4.3 PCA on the Strike Metric

4.3.1 Choosing the Strike Variables

The application of listwise deletion, creates a trade-off between the number of variables on which we perform the PCA, and the total number of observations. For each year, we choose the variables within a given expiry range so as to get (a) a sufficient number of strikes (say not less than 7 ideally) i.e. a wide range of the smiles[9], (b) a sufficient amount of observations (say no less than 100), and (c) a satisfactory correlation between them; if the covariance matrix of the variables is diagonal, then there is no gain from the PCA (see Basilevsky [9]).

Table 4.2 shows the number of variables on which we are going to perform the PCA, the number of observations (after the listwise deletion), and the Kaiser-Meyer-Olkin (KMO) measure of the correlations between the variables[10]. This

[9]Exercise prices are set at either 5.00 or 10.00 points intervals depending on the contract month (see CME regulations at http://www.productive.com/folio.p).

[10]The KMO measure is an index which compares the magnitudes of the observed correlation coefficients to the magnitudes of the partial correlation coefficients. It is computed as

$$KMO = \frac{\sum\sum_{i \neq j} r_{ij}^2}{\sum\sum_{i \neq j} r_{ij}^2 + \sum\sum_{i \neq j} a_{ij}^2}$$

where r_{ij} is the correlation coefficient between variables i and j, and a_{ij} is the partial correlation coefficient between variables i and j. The KMO measure is close to 1 if the sum of partial correlations is close to zero, i.e. the correlations between pairs of variables can be only explained

information is reported across expiries and years.

We can see that in general we have a quite wide range of strikes (7-18), the number of observations is bigger than 100, and the KMO coefficients are between 0.70, and 0.90[11].

4.3.2 Some Descriptives for the chosen Variables

Figures 4.1-4.4 show the means of the implied volatilities for the levels of the strike variables that we have chosen over the expiry buckets and years. The average implied volatility in year 1992, varies from 0.2 to 0.12 across the different expiries. In 1993, it varies from 0.18 to 0.09, and in 1994 from 0.2 to 0.09. In 1995, it ranges from 0.2 to 0.12. Figures 4.5- 4.8 show the standard deviations of the first differences of implied volatilities of the chosen strike levels. It is not clear that the shorter expiry options implied volatilities are more volatile than the longer ones.

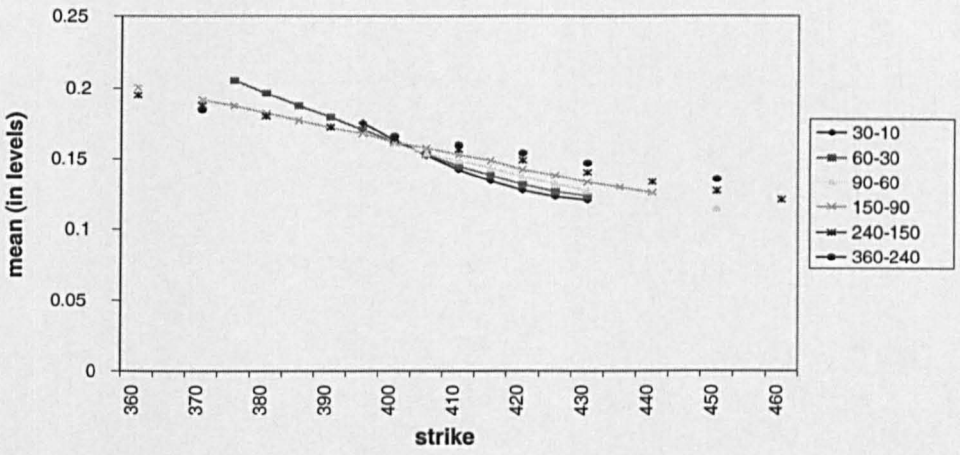


Figure 4.1: Smile Analysis on the Strike Metric: Average Implied Volatilities (levels) in 1992.

by the other variables. In the SPSS manual [136], measures in the 0.90's are characterized as marvelous, in the 0.80's as meritorious, in the 0.70's as middling, in the 0.60's as mediocre, in the 0.50's as miserable, and below 0.50 as unacceptable.

[11]The fact that correlations are highy, is not only encouraging for the application of the PCA, but it is also supportive for the selection of a linear PCA model. When there are nonlinearities, a set of highly related random variables can exhibit low correlation, unless nonlinearities are taken explicitly into account. See Basilevsky [9], page 162.

Range	Year	Number of Variables	Number of Observations	KMO
30-10	92	8	100	0.87841
	93	7	114	0.85267
	94	8	107	0.84658
	95	8	61	0.8449
60-30	92	12	206	0.83963
	93	11	219	0.81259
	94	13	206	0.74957
	95	6	141	0.88004
90-60	92	9	121	0.79977
	93	9	149	0.7913
	94	7	174	0.73537
	95	7	92	0.6993
150-90	92	16	161	0.90531
	93	14	168	0.84778
	94	18	167	0.87356
	95	12	142	0.86919
240-150	92	11	215	0.91266
	93	12	208	0.90134
	94	12	237	0.90269
	95	9	203	0.76594
360-240	92	6	212	0.80947
	93	8	134	0.88847
	94	8	135	0.87225
	95	5	117	0.71903

Table 4.2: Smile Analysis on the Strike Metric: Number of Variables, Number of Observations, and the KMO measure.

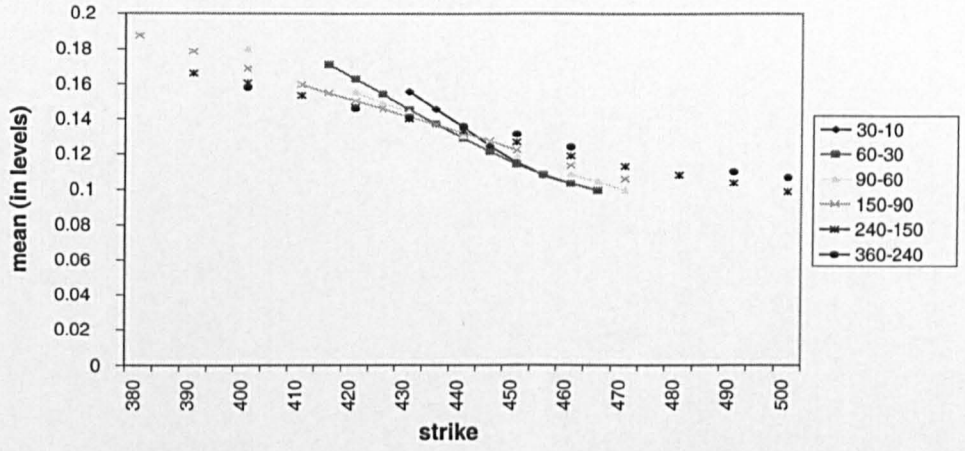


Figure 4.2: Smile Analysis on the Strike Metric: Average Implied Volatilities (levels) in 1993.

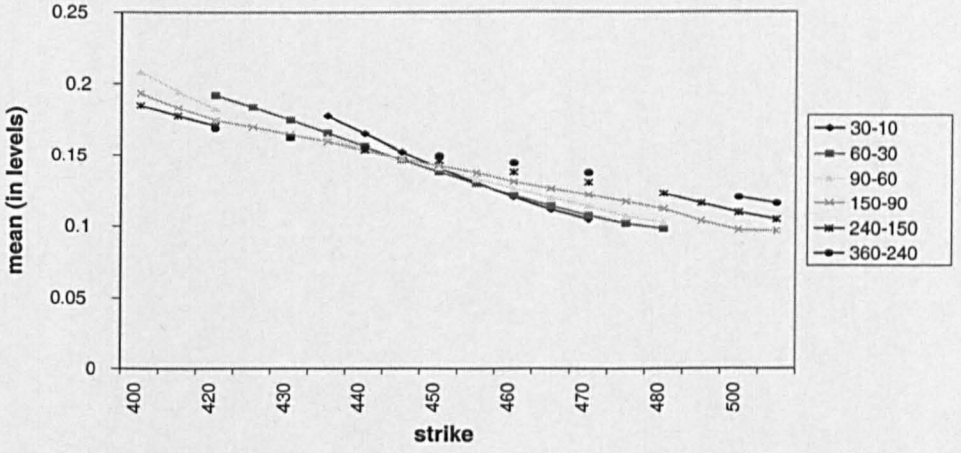


Figure 4.3: Smile Analysis on the Strike Metric: Average Implied Volatilities (levels) in 1994.

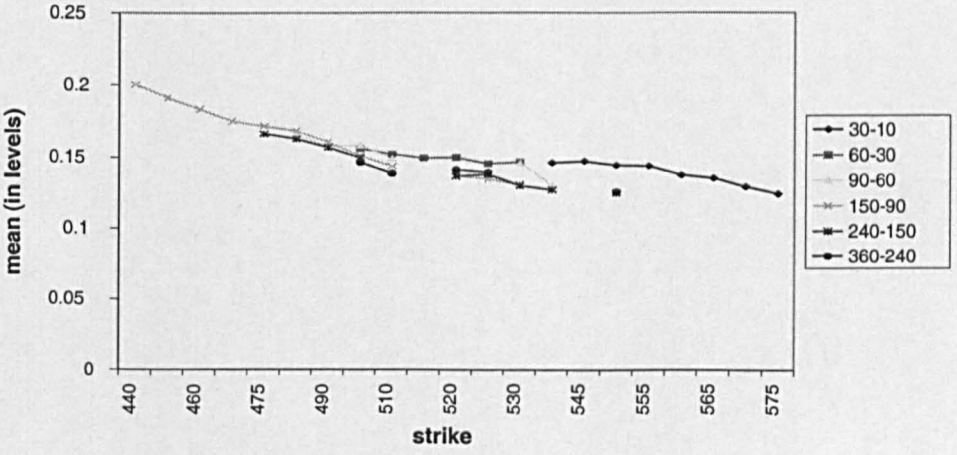


Figure 4.4: Smile Analysis on the Strike Metric: Average Implied Volatilities (levels) in 1995.

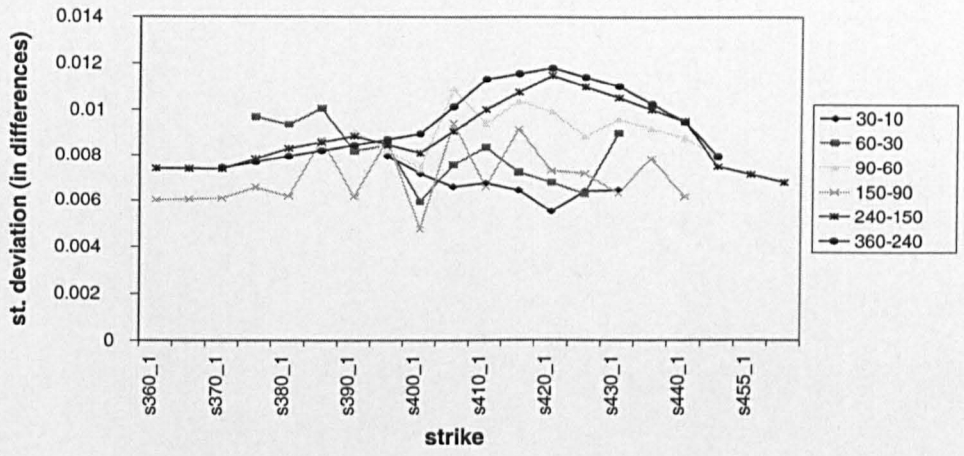


Figure 4.5: Smile Analysis on the Strike Metric: Standard Deviation of Implied Volatilities (in differences) in 1992.

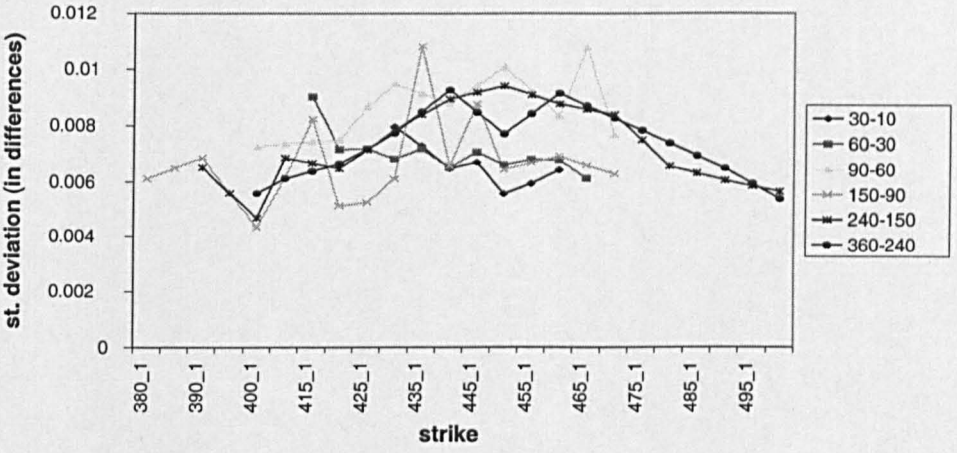


Figure 4.6: Smile Analysis on the Strike Metric: Standard Deviation of Implied Volatilities (in differences) in 1993.

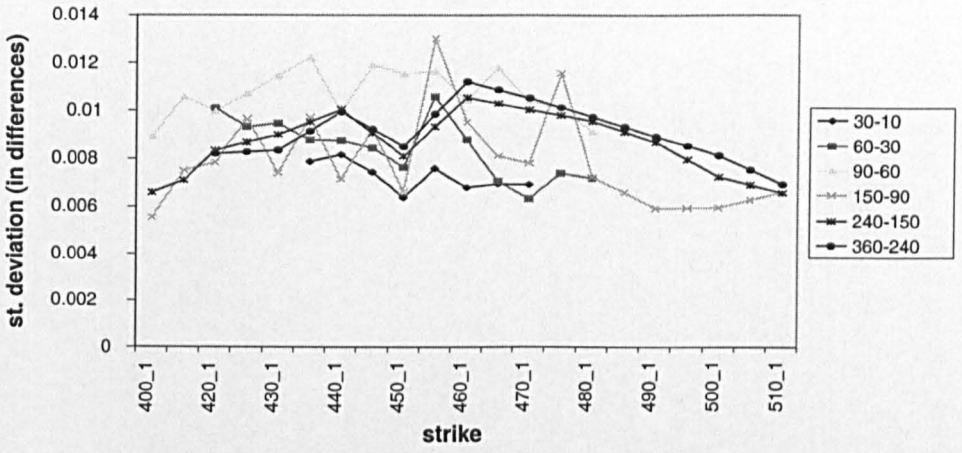


Figure 4.7: Smile Analysis on the Strike Metric: Standard Deviation of Implied Volatilities (in differences) in 1994.

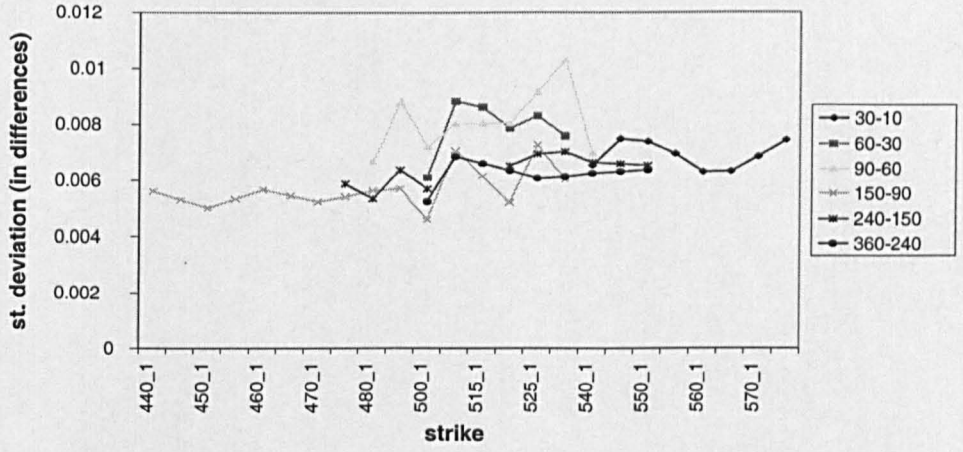


Figure 4.8: Smile Analysis on the Strike Metric: Standard Deviation of Implied Volatilities (in differences) in 1995.

4.3.3 Number of Retained Principal Components and a First Interpretation

Testing for Normality

In this section, we decide on the number of Components to be retained. Earlier researchers have used a variety of rules of thumb to determine the number of components to be retained. For example, they keep the components which have

eigenvalues bigger than the mean of all the eigenvalues (mean eigenvalue rule of thumb), or they keep the components which explain 90% of the total variance[12]. As Basilevsky [9] notes "such practice is statistically arbitrary, and seems to be prompted more by intuitive concepts of practicality and "parsimony", than by probabilistic requirements of sample-population inference."

We determine the number of components to be retained by looking at a variety of criteria in an even handed way. First, we apply Velicer's [145] non-parametric criterion. Next, working with components retained under this criterion, we look at the communalities. Finally, we look at the interpretation of the PCs.

We now explain why we have to resort to a non-parametric test like Velicer's. The formal tests that are usually employed look at the eigenvalues, and at the loadings. They test the equality of all the eigenvalues, and subsets of the eigenvalues, or alternatively they look at the residuals of the PC model. In order to interpret the shape of the components, researchers have developed tests for the entire eigenvector and for elements of the eigenvector[13]. However, all of these tests are based on the assumption that the variables are multivariate normaly distributed. We check whether this assumption is supported by our data. It suffices to test for univariate normality. If at least one variable from the whole set of variables is not normally distributed, then the whole set is not multivariately normal. This holds because multivariate normality implies univariate normality.

We test for univariate normality by using the Bera-Jarque test (BJ) (see Harvey [83]), and quantile-quantile (Q-Q) plots[14]. The idea behind the BJ test

[12]Litterman and Scheinkman [107] apply PCA to the yield curve, and they retain three components because these are explaining about 98% of the total variance. However, as Jackson [93] (page 44) notes "this procedure is not recommended. There is nothing sacred about any fixed proportion". For a description and discussion of the several rules of thumb, see Jackson [93].

[13]For a review of these tests see Basilevsky [9].

[14]The idea behind a Q-Q plot is the following: let $g_X(x)$ be the empirical density function of a variable X. We want to test whether it is the same as the normal density function $f_X(x)$. The probability mass that X accumulates under the points i and q for the two densities respectively, is: $P(X \leq i) = \int_{-\infty}^{i} g(x)dx$, and $P(X \leq q) = \int_{-\infty}^{q} f(x)dx$. $P(X \leq i)$ has been calculated by ranking the data and using a formula for the calculation of the probability (for the different formulae, see SPSS manual [136]). We require that the accumulated probabilities under the

is that the empirical distribution is compared to the normal one, by comparing the skewness β_1 to zero and its kurtosis β_2 to three. When the observations are normally and independently distributed with constant mean and variance, the test statistic is distributed as chi-squared with two degrees of freedom, and it is given by the expression

$$W = n(\frac{b_1}{6} + \frac{(b_2 - 3)^2}{24}) \sim X^2(2) \tag{4.6}$$

where n is the sample size, b_1 is the sample estimator of β_1, and b_2 is the sample estimator of β_2. These are the standardized third and fourth moments of the observations y_t about the mean, namely $\sqrt{b_1} = \hat{\sigma}_y^{-3} \sum_{t=1}^{n} (y_t - \overline{y})^3/n$, and $b_2 = \hat{\sigma}_y^{-4} \sum_{t=1}^{n} (y_t - \overline{y})^4/n$, where the standardization is done with the sample estimator of the standard deviation $\hat{\sigma}_y$.

Table 4.3 shows the results from the BJ test. The null hypothesis of univariate normality is rejected in all the expiries; the only exception is in the range 30-10 in year 1995. We accompany the BJ test's results with Q-Q plots. This is because the BJ test assumes that observations are distributed independently. The Q-Q plots agree with the BJ test's results in all the cases, apart from the range 30-10 in year 1995[15]. This is depicted in Figure 4.9 where the departure from the 45 degrees line is evident[16]. Therefore, we conclude that we need to use a non-parametric test.

---

two points are the same. Then, if the two densities are the same, it has to be $i = q$. Hence, we solve the equation $P(X \leq i) = \int_{-\infty}^{q} f(x)dx$ for q . The Q-Q plot is a plot of q against i. If the empirical density is the normal one, then the Q-Q plot will be a 45 degrees line.

[15]This discrepancy between the two tests may be attributed to the erratic behaviour of the BJ test when the sample size is small (less than 100) (see Harvey [83]). It is also known that failing to reject normality under the BJ test, does not necessarily confirm it because the test is only a test of symmetry and kurtosis (see Greene [76]).

[16]The probabilities accumulated for the emirical density are calculated by using Rankit's method (see SPSS manual [136]), i.e. $\frac{r - \frac{1}{2}}{n}$; n is the number of observations, and r is the rank, ranging from to 1 to n.

Range	Year	Variable	n	Skewness	Ex. Kurtosis	Test Value	P-Value
30-10	92	415_1	137	0.463	1.068	11.4058	0.0033
	93	450_1	146	0.733	0.029	13.0791	0.0014
	94	465_1	143	0.613	2.569	48.2794	0.0000
	95	540_1	81	-0.361	0.49	2.5697	0.2767
60-30	92	400_1	238	-0.536	4.258	191.1908	0.0000
	93	425_1	235	0.766	3.432	138.3136	0.0000
	94	465_1	232	-0.079	3.908	147.8751	0.0000
	95	520_1	180	-1.055	5.322	245.8184	0.0000
90-60	92	450_1	179	0.481	1.603	26.0673	0.0000
	93	420_1	188	0.753	4.699	190.7310	0.0000
	94	420_1	191	0.552	2.719	68.5354	0.0000
	95	490_1	159	-0.309	10.101	678.4803	0.0000
150-90	92	390_1	168	0.089	1.471	15.3687	0.0005
	93	410_1	168	0.647	1.512	27.7241	0.0000
	94	480_1	167	-0.219	3.333	78.6343	0.0000
	95	480_1	165	-0.229	2.474	43.5218	0.0000
240-150	92	400_1	245	0.692	6.158	406.6635	0.0000
	93	440_1	244	0.435	7.694	609.5378	0.0000
	94	440_1	242	0.286	4.607	217.3123	0.0000
	95	475_1	238	0.684	2.519	81.4831	0.0000
360-240	92	400_1	244	0.187	4.493	206.6571	0.0000
	93	440_1	208	0.614	3.327	108.9999	0.0000
	94	440_1	204	0.459	3.053	86.39	0.0000
	95	530_1	172	0.566	1.123	18.2216	0.0001

Table 4.3: Smile Analysis on the Strike Metric: Bera-Jarque Test for Univariate Normality. X-1 denotes the differenced once implied volatilities corresponding to strike level X.

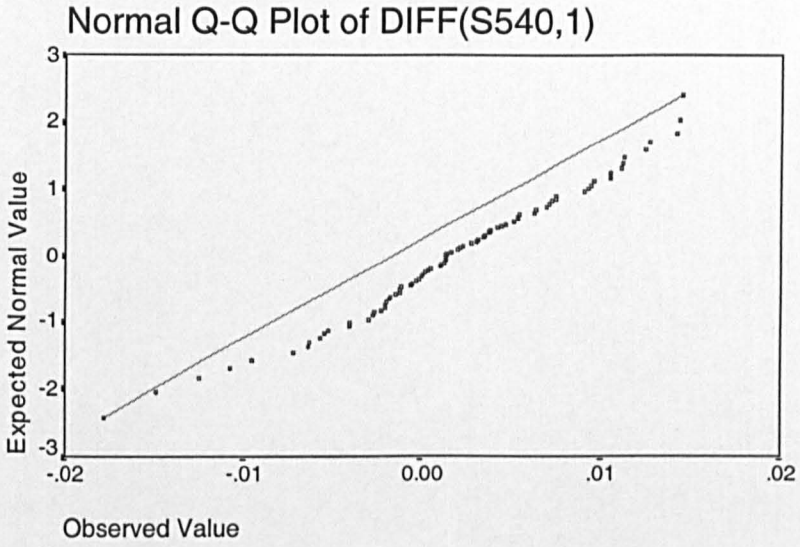


Figure 4.9: Smile Analysis on the Strike Metric: Q-Q Plot for the First Differences of Implied Volatilities in the 30-10 expiry, for the strike 540 for 1995.

Velicer's Criterion

Velicer [145] proposes a non-parametric method for selecting nontrivial PCs, i.e. components which have not arisen as a result of random sampling, measurement error, or individual variation[17]. His method is based on the partial correlations of the residuals of the PCs model, after $r < p$ components have been extracted. The criterion can be described as follows: Basilevsky [9] shows (theorem 3.13, page 132) that $\mathbf{X}'\mathbf{X} = \mathbf{A}\mathbf{A}'$. Hence, the variance-covariance matrix of the residuals $\varepsilon_{(r)}$ in equation (4.5) is given by

$$\varepsilon'_{(r)}\varepsilon_{(r)} = \mathbf{X}'\mathbf{X} - \mathbf{A}_{(r)}\mathbf{A}'_{(r)} \tag{4.7}$$

Let $\mathbf{D} = \mathbf{diag}(\varepsilon'_{(r)}\varepsilon_{(r)})$. Then, $\mathbf{R}^* = \mathbf{D}^{-\frac{1}{2}}\varepsilon'_{(r)}\varepsilon_{(r)}\mathbf{D}^{-\frac{1}{2}}$ is the matrix of partial correlations of the residuals[18]. If r^*_{ij} represents the ith row, jth column element of

[17] The only non-parametric tests for choosing the number of retained components, are Velicer's test and the cross-validation method [64]. However, there is no evidence of which one of the two performs better (see Jackson [93]).

[18] We deal with partial correlations because we have kept out of the analysis $p-r$ components.

R^*, then the Velicer statistic is given by

$$f_r = \sum\sum_{i \neq j} \frac{r_{ij}^{2*}}{p(p-1)} = \sum\sum_{i=j} \frac{r_{ij}^{2*} - p}{p(p-1)} \tag{4.8}$$

and lies in the interval 0 to 1. The behavior of f_r is that it is decreasing until a number r^* and then it increases again.

Velicer suggests that $r = r^*$ should be the number of components to be retained. This is because as long as f_r is declining, there is still space for the additional components to capture part of the covariance of the residuals. The formula of the partial correlation between the residuals shows that f_r decreases if the partial covariances are declining faster than the residual variances. Hence, Velicer's procedure will terminate when, on the average, additional PCs would explain more of the residual variances than their covariances.

Table 4.4 shows the number of components retained under Velicer's criterion, and under the mean eigenvalue rule of thumb[19]. It also shows the percentage of the variance explained by the first three components. These results come from estimating the PCs separately for six distinct expiry ranges, and in four distinct one year periods. We can see that under Velicer's criterion, we should keep either one, or two PCs[20]. This is more conservative than the eigenvalue rule of thumb which in most expiries retains one or two more PCs than Velicer's. This is consistent with Zwick and Velicer's [93] results. Using simulated data, they conclude that the mean eigenvalue rule generally retains too many components, comparing to Velicer's criterion. Note that $f_0 > f_1$, showing that at least one component can be extracted[21].

[19] Krzanowski [105] performed Monte Carlo simulations and he finds that the mean eigenvalue test tends to keep the same amount of components as the cross validation method. However, in a discussion that we had with him, he said that before reaching a definite conclusion, more research needs to be carried out on it. Therefore, we continue treating the mean-eigenvalue rule as an ad-hoc procedure.

[20] These results do not change, even if we apply Velicer's criterion to a smaller number of variables.

[21] Reddon [123] evaluates the type-I error (reject the null while it is true) rates of this criterion. He considers that the null hypothesis of Velicer's criterion is the reduction in dimensionality. Then, he carries out simulations using data generated from populations having unit variances

Range	Year	f_o	f_1	f_2	f_3	r*	l	1st PC	2nd PC	3rd PC
30-10	92	0.2863	0.2835	0.2843	0.2847	1	2	57.60	16.70	5.80
	93	0.3156	0.3131	0.3138	0.3143	1	1	61.30	16.30	6.30
	94	0.2586	0.2557	0.2563	0.2569	1	2	55.90	14.00	8.70
	95	0.3125	0.3074	0.3088	0.3096	1	1	60.00	17.30	6.40
60-30	92	0.1383	0.1377	0.1377	0.1377	2	3	39.30	18.20	9.30
	93	0.1210	0.1205	0.1204	0.1204	2	2	34.90	24.00	7.70
	94	0.0999	0.0995	0.0994	0.0994	2	3	28.00	23.40	14.60
	95	0.3323	0.3297	0.3304	0.3310	1	1	64.40	11.40	8.10
90-60	92	0.1736	0.1718	0.1720	0.1723	1	2	46.40	15.80	9.90
	93	0.1529	0.1516	0.1516	0.1518	2	2	43.20	18.80	10.70
	94	0.1266	0.1253	0.1252	0.1254	2	2	36.90	17.70	10.60
	95	0.1044	0.1024	0.1022	0.1027	2	2	39.20	21.10	11.30
150-90	92	0.2296	0.2283	0.2285	0.2286	1	4	50.30	8.60	7.30
	93	0.1528	0.1519	0.1519	0.1520	2	3	41.20	14.70	8.80
	94	0.1845	0.1835	0.1835	0.1836	2	4	44.00	14.00	7.60
	95	0.1910	0.1896	0.1898	0.1900	1	2	46.60	15.50	7.00
240-150	92	0.3452	0.3441	0.3442	0.3444	1	2	61.80	9.70	7.50
	93	0.3063	0.3053	0.3054	0.3055	1	2	58.10	12.10	6.50
	94	0.2886	0.2877	0.2878	0.2879	1	2	56.70	10.80	6.90
	95	0.1551	0.1543	0.1544	0.1545	1	2	42.30	21.40	10.60
360-240	92	0.3177	0.3168	0.3169	0.3172	1	2	61.90	18.70	9.60
	93	0.3712	0.3696	0.3698	0.3701	1	2	64.90	13.20	8.40
	94	0.4062	0.4054	0.4053	0.4055	2	2	66.30	18.50	7.00
	95	0.2234	0.2217	0.2222	0.2226	1	2	56.30	20.10	11.50

Table 4.4: Smile Analysis on the Strike Metric: r* = number of components retained under Velicer's criterion (minimum f0,...,f3), l= number of components retained under rule of thumb, with percentage variance explained by components 1-3.

The first component explains between 28% and 66% of the variance, depending on which year and which expiry we consider. The second component explains between 10% and 24%. The two components combined together, explain between 56% and 85% of the variance[22]. The highest proportions are explained for the shortest and longest expiry options.

Communalities

We now look at the communalities that the retained, according to Velicer's criterion, PCs explain. We would like the number of retained components to explain a sufficient amount of the variability of each one of the variables. In Appendix A, we show in Tables A.1 and A.2, the communalities (expressed in percentage terms) explained by retaining one and two PCs for year 1994.

Table A.1 shows that keeping one PC, as Velicer's criterion does in most of the ranges, results in low communalities for the middle strikes. Table A.2 shows that the inclusion of the 2nd PC improves the communalities. This picture is repeated in the other years and ranges where Velicer's criterion retains one PC. The addition of the second PC is necessary for the improvement of the explained communalities (we do not report the results for the other years because of space limitations)[23].

---

and zero covariances. He finds that Velicer's test makes excessive type-I errors in the cases where the number of observations does not exceed two times the number of variables. As the number of observations increases beyond two times the number of variables, the type-I error rates rapidly becomes zero. Since the number of observations that we use is far more than twice the number of variables, the results from comparing f_o and f_1 are reliable.

[22]These results are very different from those found in the interest rate literature. Litterman and Scheinkman [107] find that on average the first PC explained 89.5% of the total variance, the second PC explained 8.5% and the third 2%. In total, the three retained PCs explained 98.4% on average, across the different yield maturities.

[23]We also looked at the communalities explained by three PCs. The inclusion of the third PC does not increase the explained communalities, significantly.

Interpretation of the Components

Next, we look at the interpretation of the first three PCs. If the third PC appears to be just noise, then we prefer to reject it[24].

The interpretation of the principal components can be achieved by looking at the correlation loadings $\mathbf{A}'$ (see equation (4.3)). The first three columns of $\mathbf{A}'$ show how the first three PCs, respectively, affect the implied volatilities. Figures 4.10-4.15, show the correlation loadings of the first PC, for each expiry over the different years. For the sake of clarity, we have interpolated across the missing variables in these graphs.

The figures show that the first PC in the expiry range 30-10 is like a parallel shift with a slight attenuation at the edges, in every year[25]. In the range 60-30, it still moves the implied volatilities to the same direction in all the years, but 1992 and 1994. For all the other ranges, it has a mixture of both negative and positive correlation loadings in the years 1992, 1993, and 1994, while it is a shift in 1995.

[24] Wilson [148] also acts similarly in a study of the number of shocks which affect the term structure of interest rates. Even though the third PC explains 8-10% of the total variance of the changes of yields, he decides not to retain it because it is just noise.

[25] We use the terminology "shift" and "slope" for the interpretation of the PCs, as Litterman and Scheinkman [107] have already established. By "shift" they mean that the PC moves the implied volatilities of all the strikes (moneyness) to the same direction. By "slope" they mean that the PC moves the low strikes implied volatilities upwards, and the high strikes ones downwards (or vice versa).

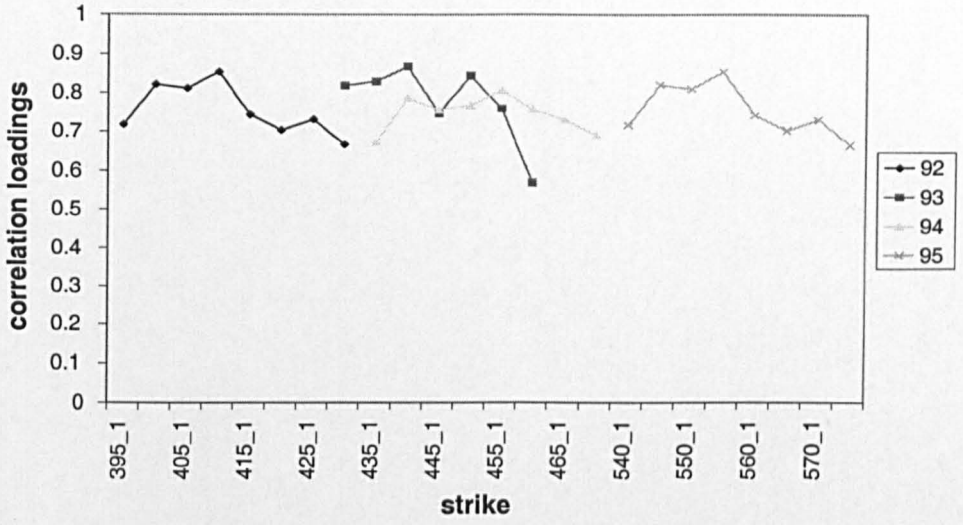


Figure 4.10: Smile Analysis on the Strike Metric: Interpretation of the First PC for 30-10.

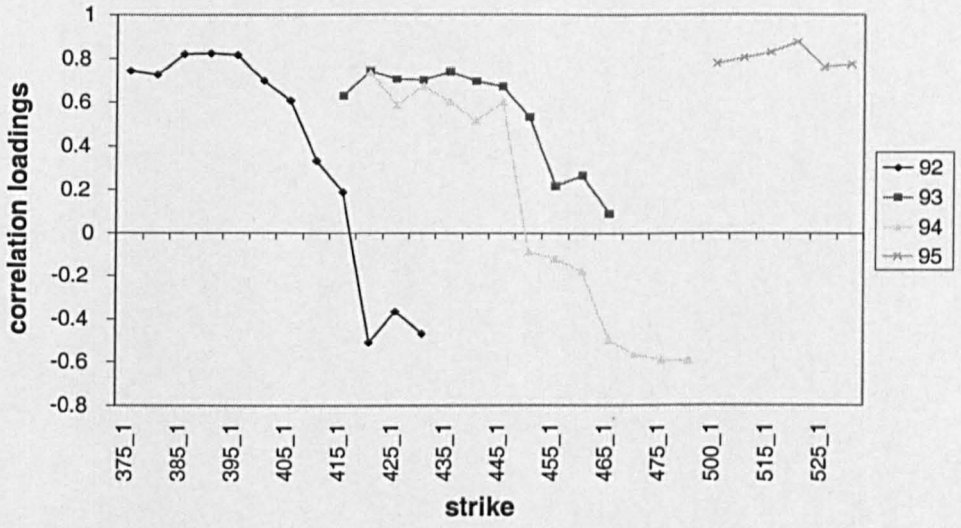


Figure 4.11: Smile Analysis on the Strike Metric: Interpretation of the First PC for 60-30.

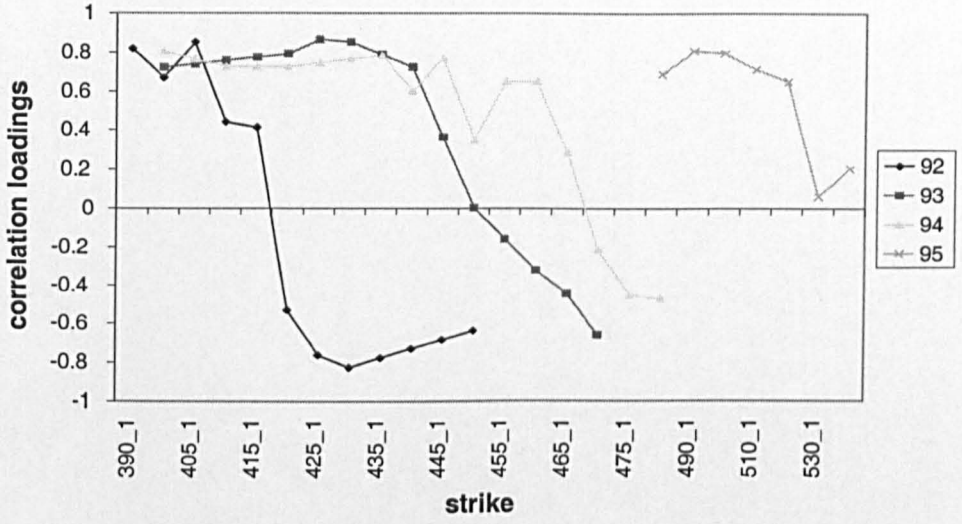


Figure 4.12: Smile Analysis on the Strike Metric: Interpretation of the First PC for 90-60.

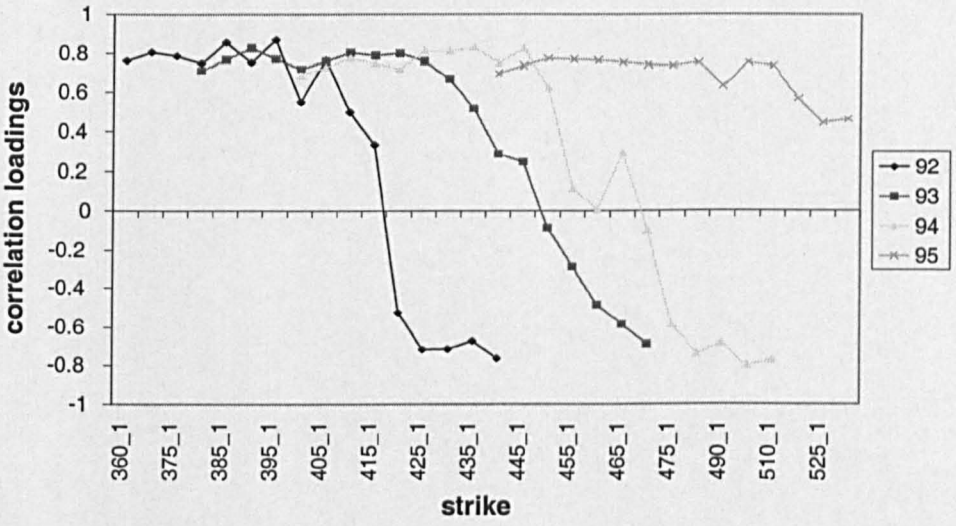


Figure 4.13: Smile Analysis on the Strike Metric: Interpretation of the First PC for 150-90.

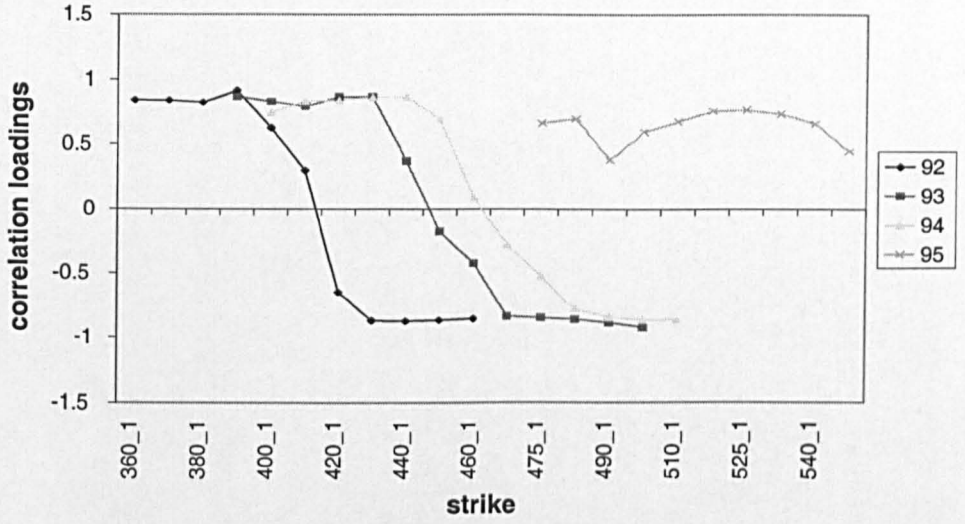


Figure 4.14: Smile Analysis on the Strike Metric: Interpretation of the First PC for 240-150.

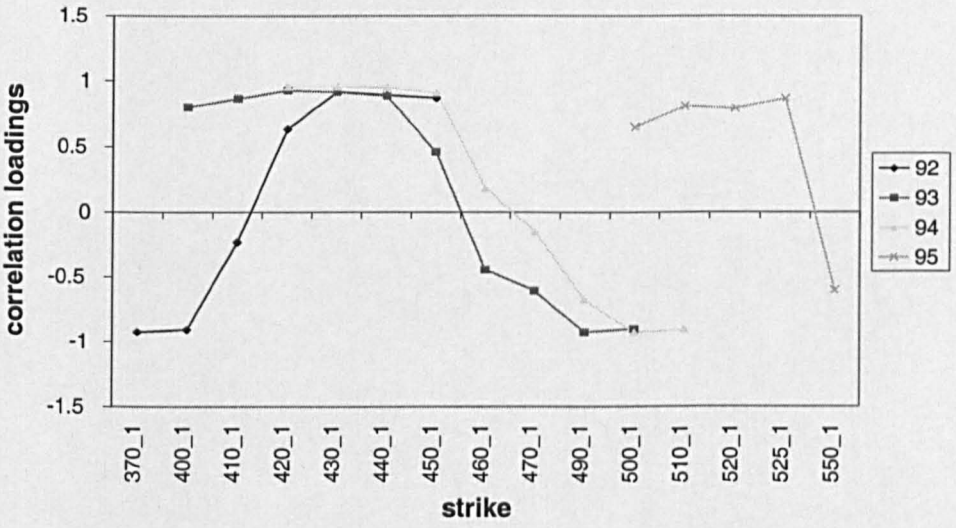


Figure 4.15: Smile Analysis on the Strike Metric: Interpretation of the First PC for 360-240.

Figures 4.16-4.21 show the correlation loadings of the second PC for each expiry over the years. The graphs show that the second PC has the interpretation of a slope (it is like a Z-shape, i.e. a slope with attenuation at the edges) in the expiry range 30-10 in every year. In the range 60-30 the same is true with the exception of year 1994. In the other ranges, we see a triangular shape with

negative correlation loadings at the edges in the years 1992, 1993, and 1994. However, in all the expiries, the second PC seems to have the slope interpretation in the year 1995.

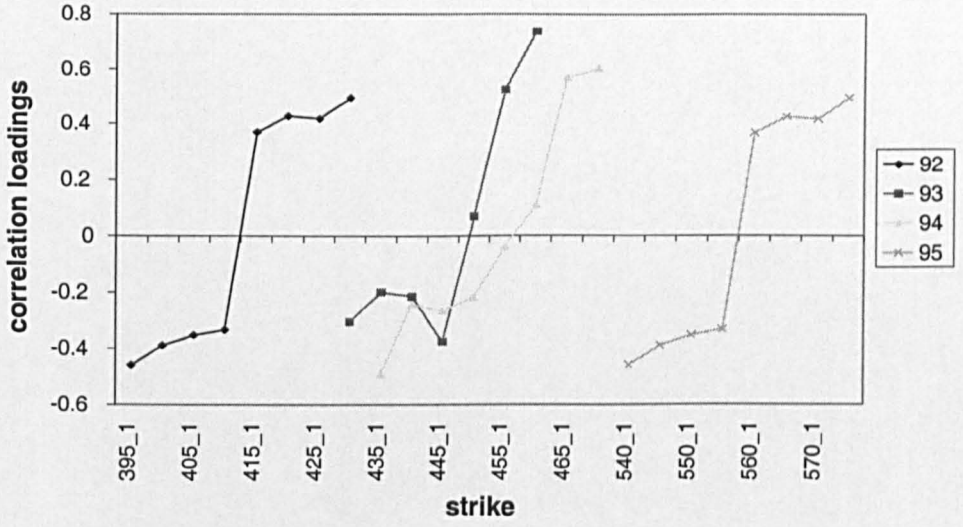


Figure 4.16: Smile Analysis on the Strike Metric: Interpretation of the Second PC for 30-10.

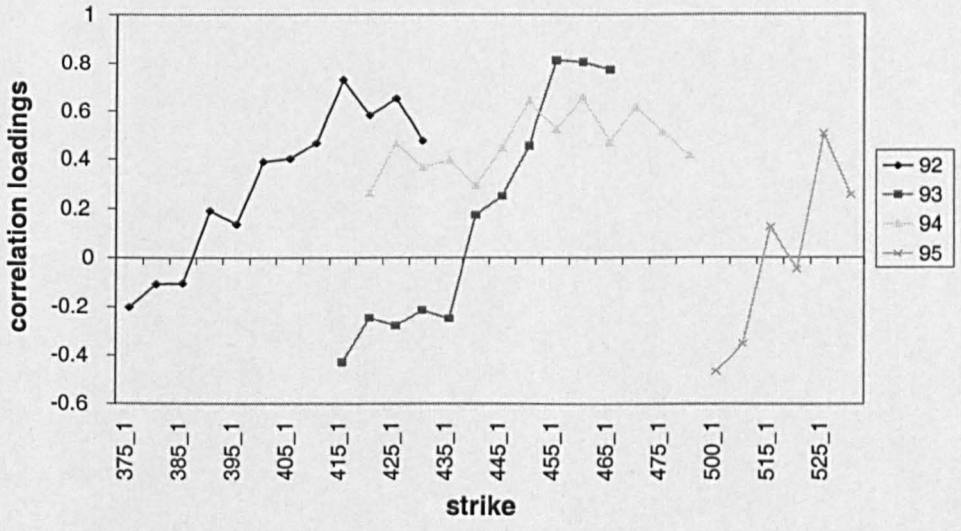


Figure 4.17: Smile Analysis on the Strike Metric: Interpretation of the Second PC for 60-30.

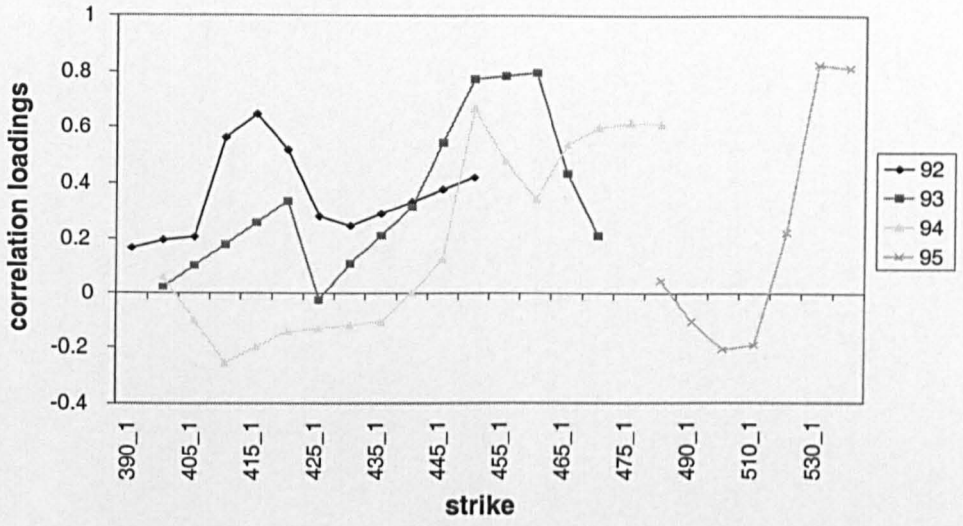


Figure 4.18: Smile Analysis on the Strike Metric: Interpretation of the Second PC for 90-60.

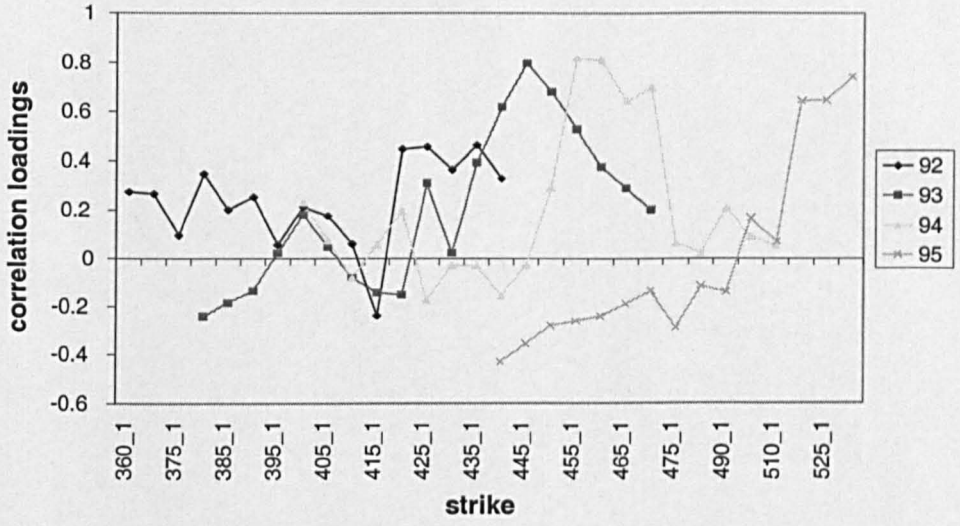


Figure 4.19: Smile Analysis on the Strike Metric: Interpretation of the Second PC for 150-90.

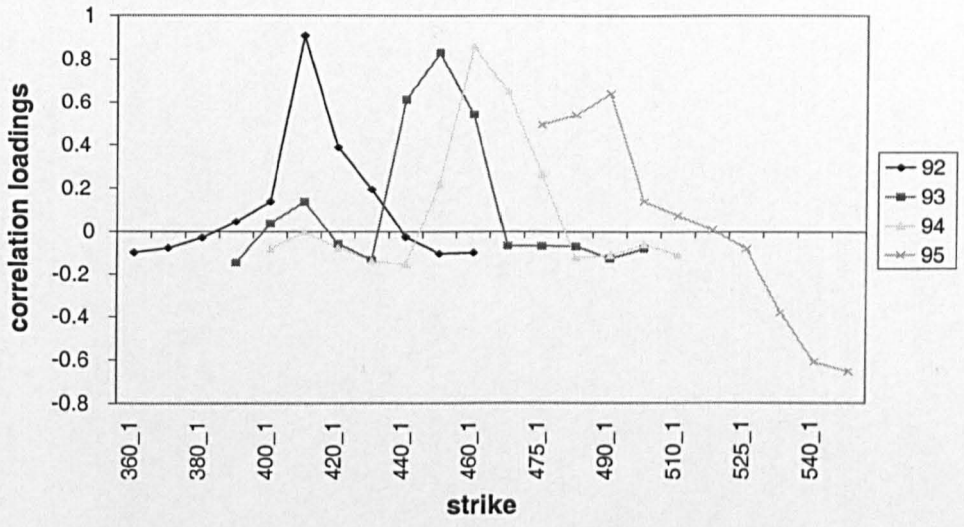


Figure 4.20: Smile Analysis on the Strike Metric: Interpretation of the Second PC for 240-150.

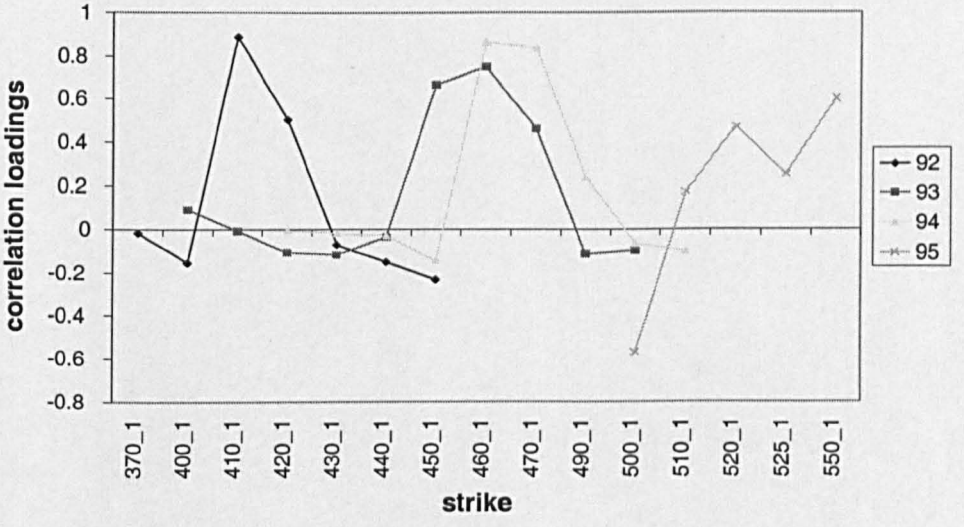


Figure 4.21: Smile Analysis on the Strike Metric: Interpretation of the Second PC for 360-240.

In Figures 4.22-4.27, we show the correlation loadings of the third PC, for each expiry over the years. The graphs reveal that for all the years and ranges, the third PC does not have a typical pattern. It is just noise, as already predicted by Velicer's criterion, and the explained communalities. Therefore, under the variety of criteria that we have applied, we can only identify two shocks driving the implied volatility smiles of the S&P 500 Futures Options in the Strike Metric.

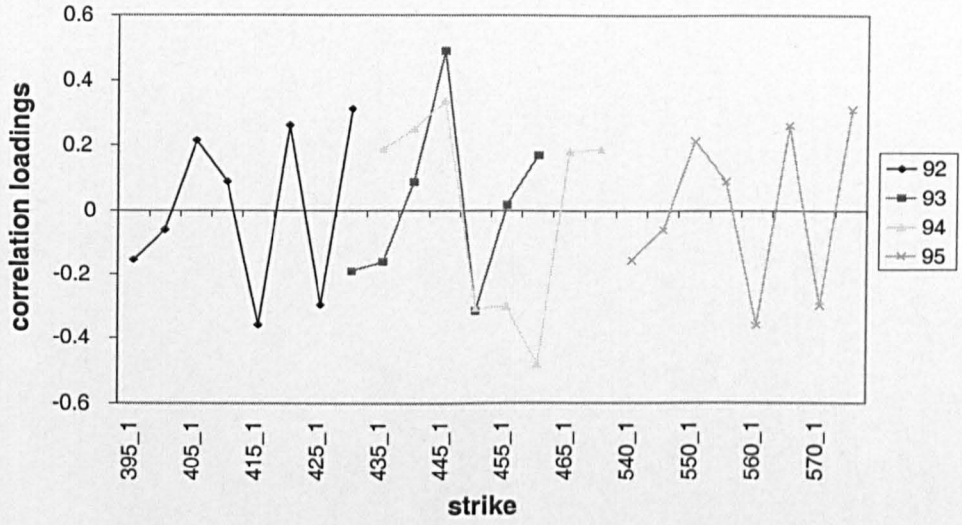


Figure 4.22: Smile Analysis on the Strike Metric: Interpretation of the Third PC for 30-10.

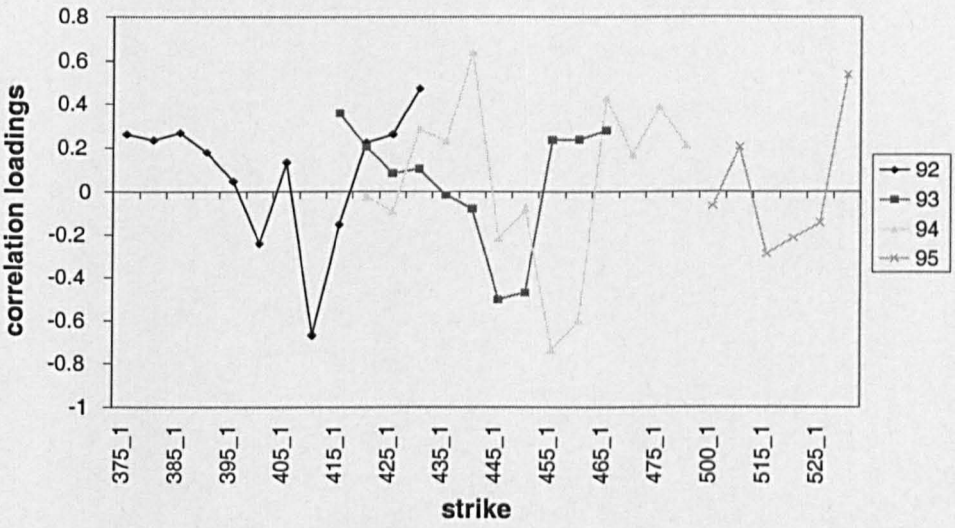


Figure 4.23: Smile Analysis on the Strike Metric: Interpretation of the Third PC for 60-30.

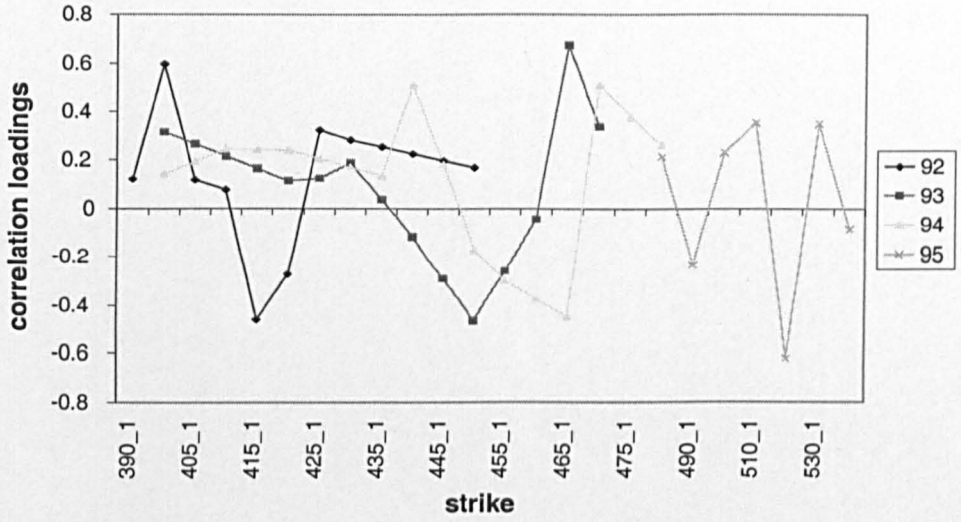


Figure 4.24: Smile Analysis on the Strike Metric: Interpretation of the Third PC for 90-60.

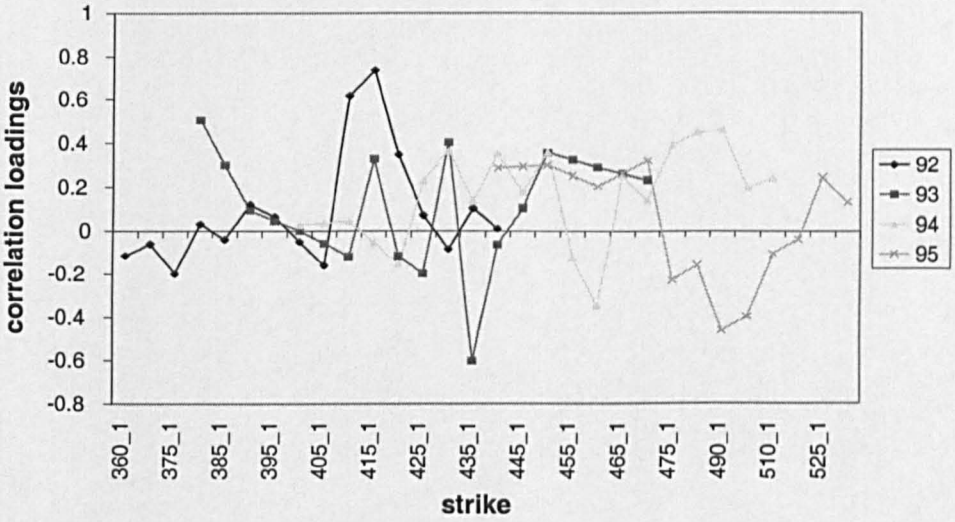


Figure 4.25: Smile Analysis on the Strike Metric: Interpretation of the Third PC for 150-90.

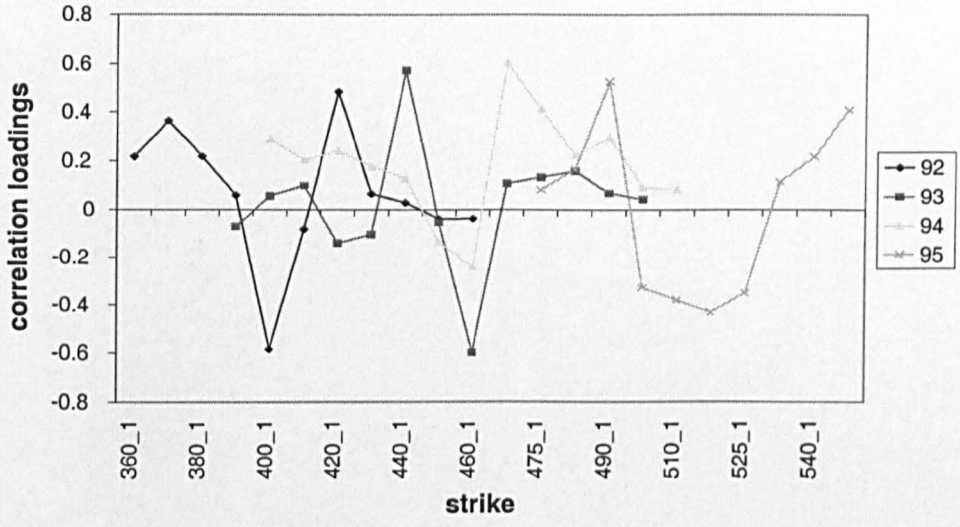


Figure 4.26: Smile Analysis on the Strike Metric: Interpretation of the Third PC for 240-150.

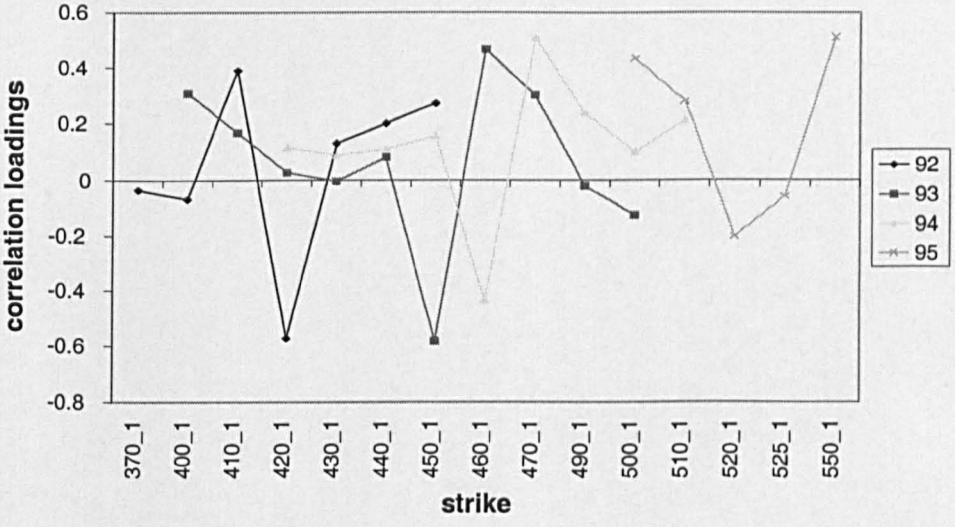


Figure 4.27: Smile Analysis on the Strike Metric: Interpretation of the Third PC for 360-240.

Although the first and second PC have a simple and consistent interpretation in some years for the ranges 30-10 and 60-30, and over all the expiries in 1995, we would like this to be the case for the other ranges and years, as well. In order to obtain a better interpretation, we will utilize a rotation method, explained in the next section.

4.3.4 The Rotation Method

The Idea of a Rotation

We use a rotation technique, so as to achieve a simple and consistent interpretation of the PCs. This is a common procedure followed by researchers in the context of PCA, when the interpretation of the PCs is not straightforward. In this section, first we describe the idea of a rotation. Then, we construct an appropriate for the purposes of our study rotation, and we apply it.

The way that the first two components are affecting the implied volatilities can be revealed by inverting equation (4.1), i.e.

$$\mathbf{X} = \mathbf{Z}\mathbf{P}' \tag{4.9}$$

We get $\mathbf{P}'$ by inverting equation (4.4), so that

$$\mathbf{P}' = \mathbf{L}^{-\frac{1}{2}}\mathbf{A}' \tag{4.10}$$

The interpretation of the first and second PCs is obtained then from the first and second row of $\mathbf{P}'$.

Rotating the r retained orthogonal PCs to a new orhtogonal position, is equivalent to multiplying the eigenvectors and the PCs, by an $r \times r$ orthogonal matrix $\mathbf{T}$ (which we will call from now on the transformation matrix) and its inverse $\mathbf{T}^{-1}$[26]. Then, equation (4.9) becomes

$$\mathbf{X} = \mathbf{Z}_{(r)}\mathbf{T}\mathbf{T}^{-1}\mathbf{P}'_{(r)} + \boldsymbol{\varepsilon} = \mathbf{V}_{(r)}\mathbf{G}'_{(r)} + \boldsymbol{\varepsilon} \tag{4.11}$$

where

$$\mathbf{V}_{(r)} = \mathbf{Z}_{(r)}\mathbf{T} \tag{4.12}$$

$$\mathbf{G}_{(r)} = \mathbf{P}_{(r)}\mathbf{T} \tag{4.13}$$

[26]For example, if we want to rotate two axes, i.e. we have kept two components, then $\mathbf{T} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$.

Equations (4.12), and (4.13) deliver the new rotated PCs and eigenvectors, respectively. The rotated components still explain the same total amount of variance, as the unrotated ones, but the total variance may have been re-distributed between the two. Moreover, as a property of the orthogonal rotation (rotation which maintains the axes at right angles), the rotated eigenvectors remain orthogonal, but the rotated principal components do not (see Basilevsky [9], Theorem 5.2, page 259).

In order to apply a rotation, there are three issues that need to be resolved. First, the rotated PCs are not invariant with respect to a change in their number. Hence, it is necessary to have decided first how many components we are going to retain. Second, rotation of the eigenvectors $\mathbf{P}$ (and corresponding unstandardized PCs), will not yield the same results as rotating the correlation loadings $\mathbf{A}'$ (and standardized components). For reasons that are explained in Appendix B, we are going to use the $\mathbf{P}$ eigenvectors (loadings) by extracting them from the already calculated correlation loadings by means of equation (4.10). Finally, the matrix $\mathbf{T}$ is not unique, since an infinite number of orthogonal rotations is possible. An additional criterion has to be introduced to fix the location of the axes, i.e. to fix the angle ϑ. *In order to determine ϑ we have to think what shape of components we would like to get.* We would like the first PC to be as close to a parallel shift, as possible (as in the 30-10 expiry), and the second to be a slope. This is also consistent with the intuition coming from the Taylor series expansion of order one[27].

The most popular criterion for determining ϑ, is the varimax criterion. This criterion minimizes the number of variables that have high loadings on a component. However, it is well suited for giving clusters, so it cannot give to us the interpretation that we are looking for. Figure 4.28 shows the result from applying varimax rotation to the loadings of the first two PCs in the 60-30 expiry bucket, in year 1992. We can see that even though the second PC has a slope character, the

[27] Any well-behaved function can be approximated by a Taylor series expansion of first order, where the zero order expansion is the level, and the first order expansion is the slope.

first one does not have the parallel shift interpretation. Other rotation methods, such as the quartimax, and the oblique, can not give us the desired interpretation either (for a description of these methods, see Jackson [93].) Consequently, we decide to use a "Procrustes" type rotation.

Figure 4.28: Smile Analysis on the Strike Metric: Rotated Correlation Loadings Obtained from Varimax Rotation, in the 60-30 Expiry, in Year 1992.

The Construction of the Procrustes Rotation

The idea behind "Procrustes" type rotation is the following: suppose that there are two $(p \times r)$ matrices, **A** and **B**. We need to find an $(r \times r)$ transformation matrix **T** which will best transform **A** into **B** (target matrix)[28]. For the purposes of our study, we choose **T,** so that the loadings on the first rotated PC are as flat, as possible. We accomplish this by using a regression to find the orthogonal rotation which minimizes the least squares distance between the loadings of the first PC, and a vector of constants. The construction of our Procrustes type rotation is described in Appendix B.

[28]For more details on the Procrustes rotation, see Jackson [93]. The name of the rotation has been inspired by the Greek mythology. When Thesus was cleaning up Greece's highways of criminals, one of those he terminated was Procrustes, the Stretcher. Procrustes had an iron bed on which he tied any traveller who fell into his hands. If the victim was shorter than the bed, he or she was stretched out to fit; if too long, Procrustes lopped off whatever was necessary.

Year	30-10	60-30	90-60	150-90	240-150	360-240
92	3.26	50.32	85.47	64.41	85.38	6.89
93	3.93	20.50	60.51	38.80	80.74	53.16
94	0.61	80.80	42.65	60.99	56.64	52.01
95	2.93	0.59	25.90	8.04	1.38	31.33

Table 4.5: Smile Analysis on the Strike Metric: Angle of the Procrustes Type of Rotation (Angles are Measured in Degrees).

Application of the Procrustes Rotation

We now discuss the results from applying our Procrustes rotation. Table 4.5 shows the angle ϑ that our rotation method has used, i.e. the extent to which we have moved the cartesian co-ordinates of the original components.

We see that in the range 30-10 there is virtually no rotation in all the years, so the shift shape is preserved. In all the other ranges we can see that we have rotated significantly. The results for the first rotated PC, for the different expiries, appear in Figures 4.29-4.34, while the results for the second rotated PC appear in Figures 4.35-4.40. These graphs have been produced by working with unstandardized variables[29].

The first rotated PC in the expiries 30-10 (as expected), 60-30, and 90-60, can be interpreted as a shift in all the years. In the expiries 150-90, and especially for 240-150, and 360-240 it has both positive and negative loadings in all the years, apart from 1995. The figures also show that the magnitude of the first PC is bigger for the longer expiries. The second rotated PC has a Z-shape, in all the expiries, and years[30]. Its magnitude, remains the same over the expiries.

[29] As we have already explained, the construction of our rotation method requires the rotation of the loadings and not of the correlation loadings. It does not require to unstandardize the variables.We do so, so as to see the rotated PCs, without imposing any standardization (the standardization of the variables does not affect the shape of the PCs, but only the scale).

[30] We can always flip over the Z-shape of the second PC. This can be done by taking the appropriate sign from equations (B.12),(B.13), and respecting equation (B.14).

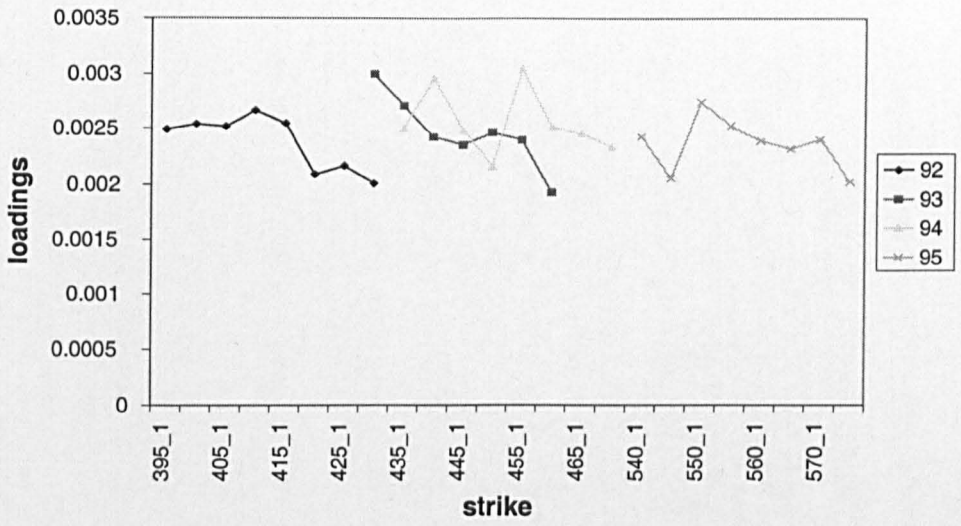


Figure 4.29: Smile Analysis on the Strike Metric: Interpretation of the First Rotated PC for 30-10.

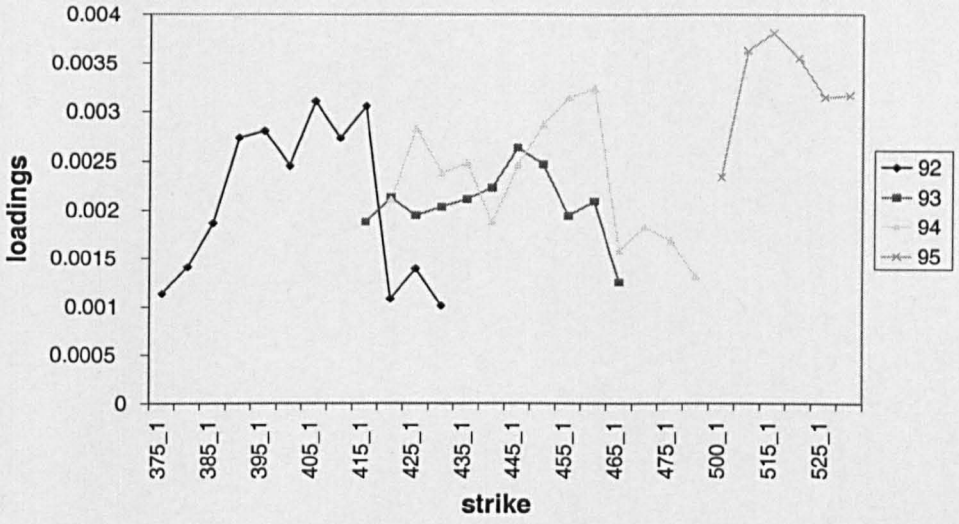


Figure 4.30: Smile Analysis on the Strike Metric: Interpretation of the First Rotated PC for 60-30.

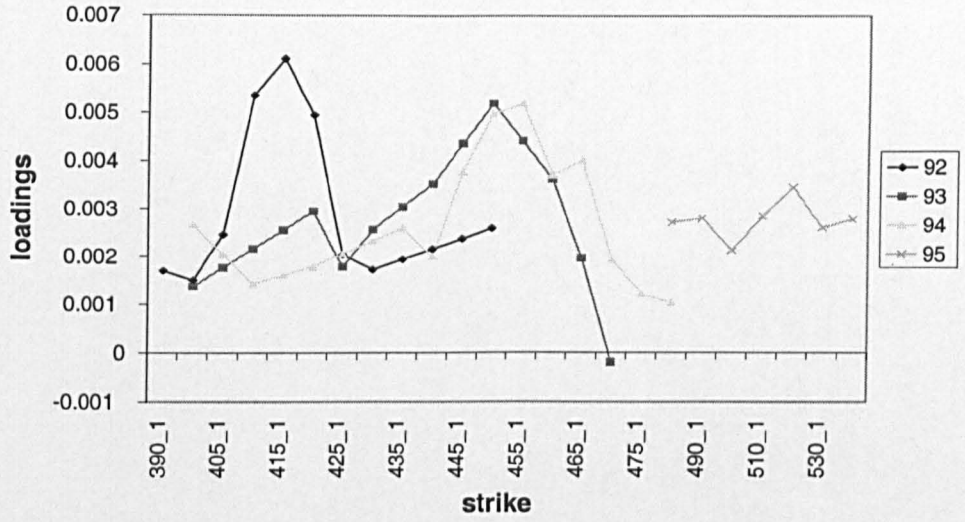


Figure 4.31: Smile Analysis on the Strike Metric: Interpretation of the First Rotated PC for 90-60.

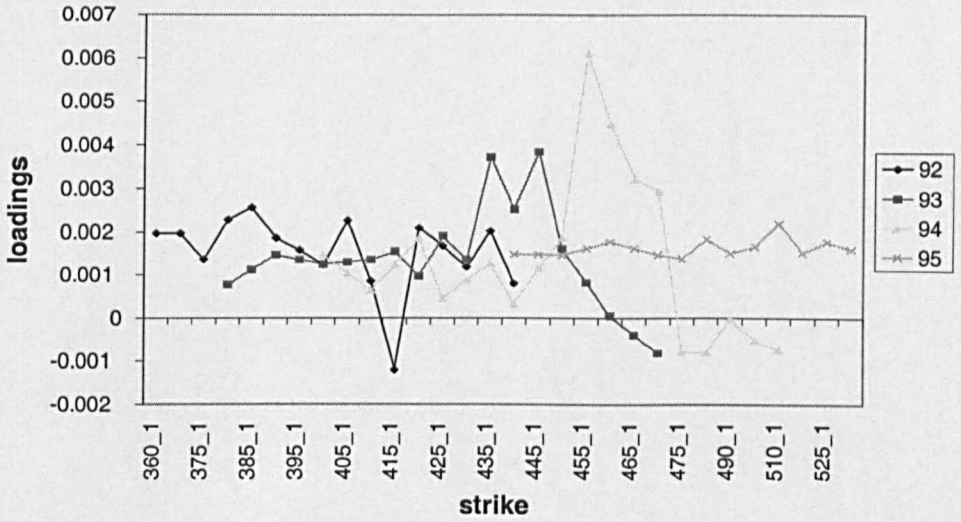


Figure 4.32: Smile Analysis on the Strike Metric: Interpretation of the First Rotated PC for 150-90.

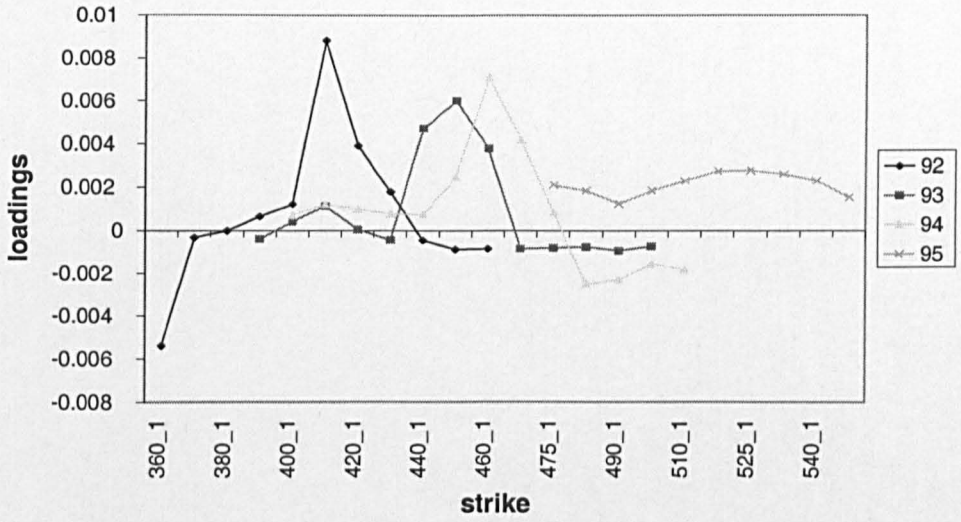


Figure 4.33: Smile Analysis on the Strike Metric: Interpretation of the First Rotated PC for 240-150.

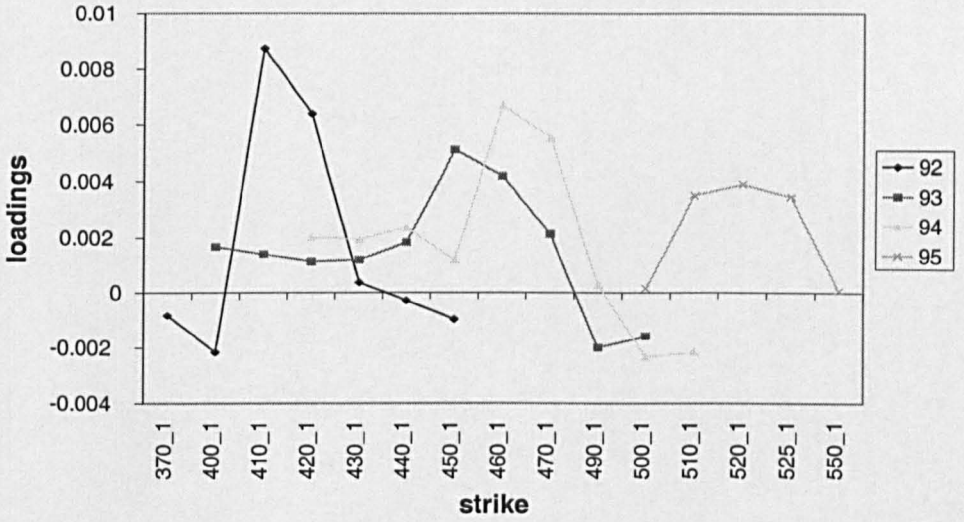


Figure 4.34: Smile Analysis on the Strike Metric: Interpretation of the First Rotated PC for 360-240.

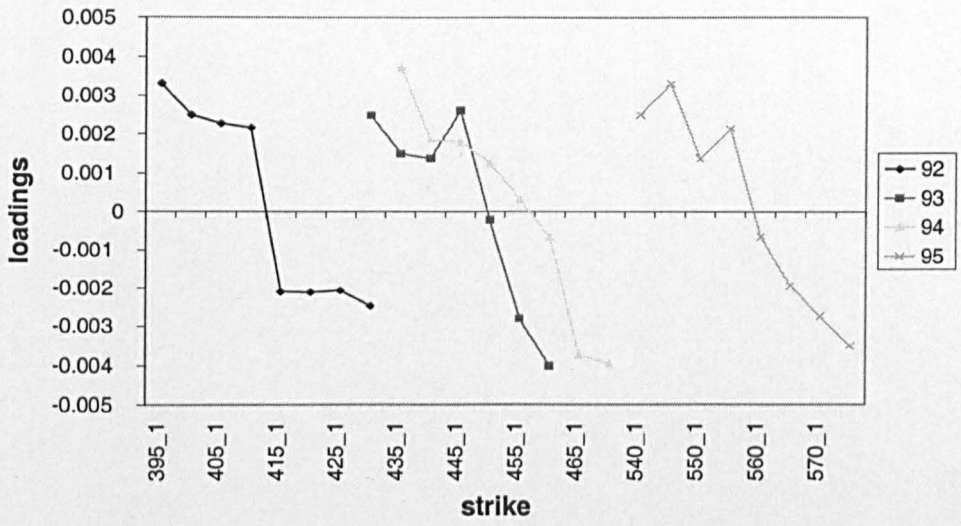


Figure 4.35: Smile Analysis on the Strike Metric: Interpretation of the Second Rotated PC for 30-10.

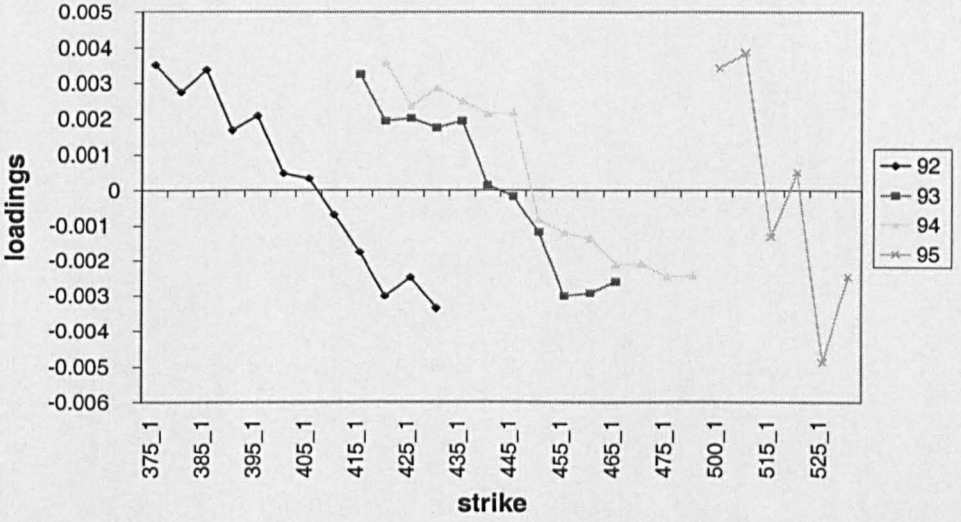


Figure 4.36: Smile Analysis on the Strike Metric: Interpretation of the Second Rotated PC for 60-30.

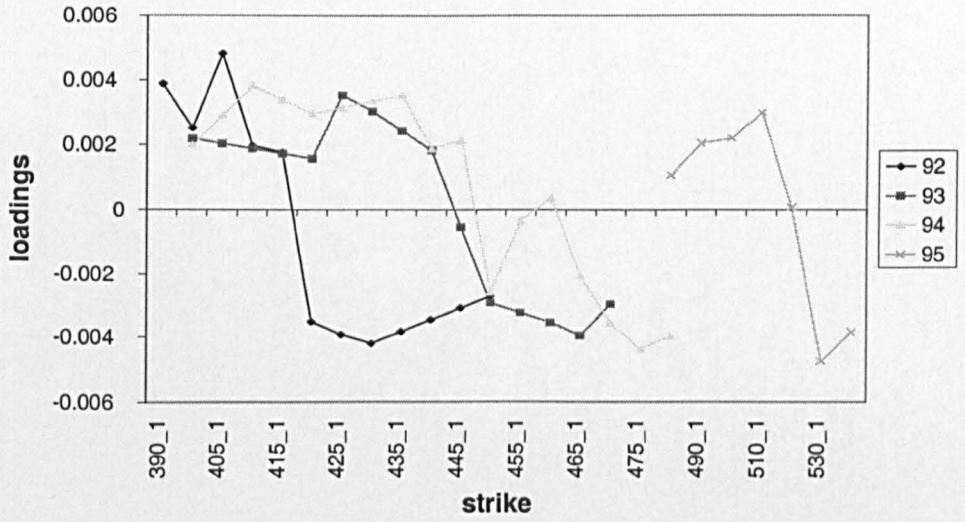


Figure 4.37: Smile Analysis on the Strike Metric: Interpretation of the Second Rotated PC for 90-60.

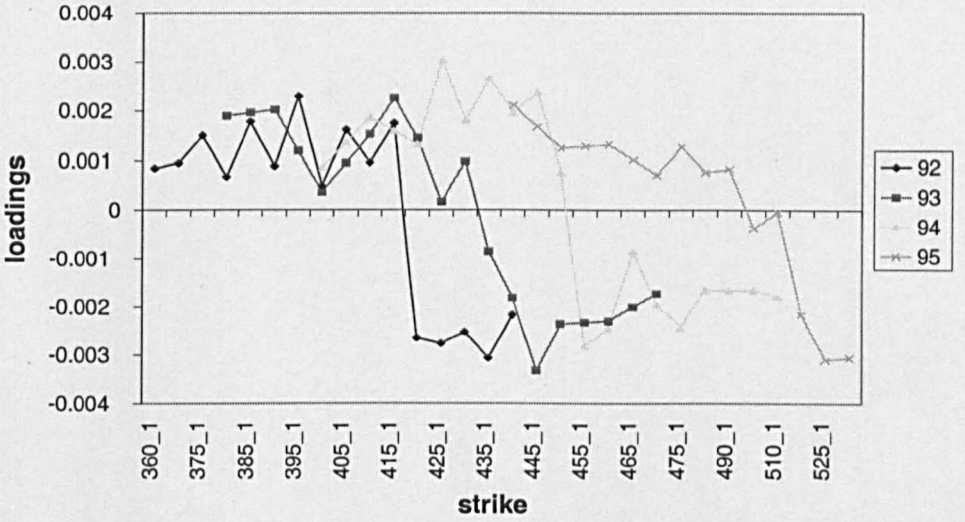


Figure 4.38: Smile Analysis on the Strike Metric: Interpretation of the Second Rotated PC for 150-90.

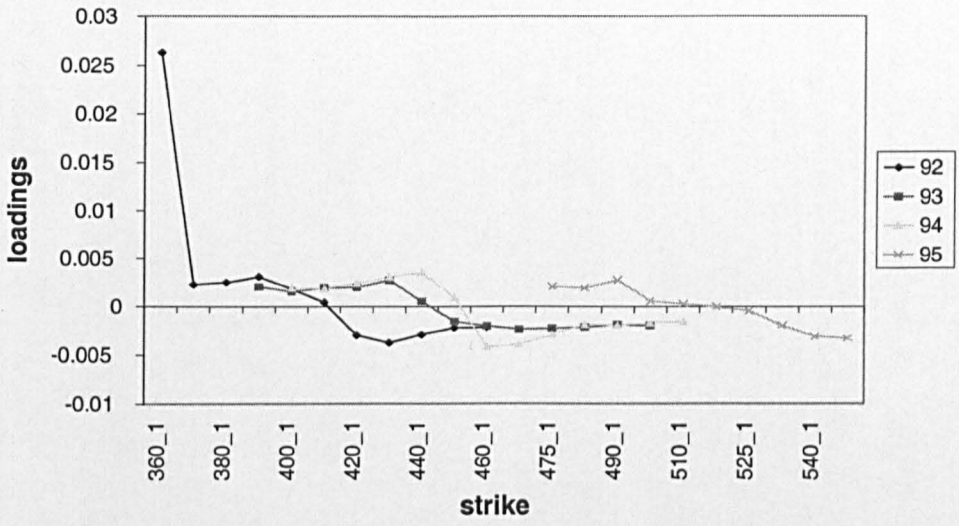


Figure 4.39: Smile Analysis on the Strike Metric: Interpretation of the Second Rotated PC for 240-150.

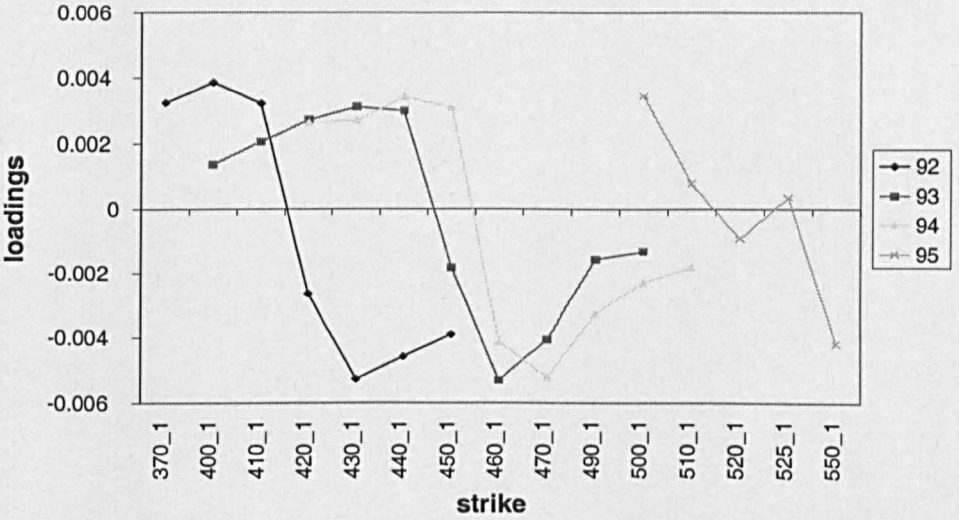


Figure 4.40: Smile Analysis on the Strike Metric: Interpretation of the Second Rotated PC for 360-240.

Table 4.6 shows the percentage of variance that the first and second rotated PCs explain (as well as the percentage for the original first component)[31]. We

[31]The calculation of the percentage of the variance that the rotated components explain, is done by using equations (B.9), (B.10), and the property that $Var(\underline{z_i}) = l_i$. Notice that a_1 and b_1 are the standardized coefficients.

Range	Year	Unrot. 1st PC	1st PC	2nd PC	Cumulative
30-10	92	57.60%	57.5%	16.9%	74.4%
	93	61.30%	61.1%	16.6%	77.7%
	94	55.90%	55.9%	14%	69.9%
	95	60.00%	59.9%	17.4%	77.3%
	Average	58.70%	58.6%	16.2%	74.8%
60-30	92	39.30%	26.8%	30.7%	57.4%
	93	34.90%	33.6%	25.4%	59%
	94	28.00%	23.5%	27.9%	51.4%
	95	64.40%	64.4%	11.4%	75.8%
	Average	41.65%	37.1%	23.9%	60.9%
90-60	92	46.40%	16%	46.2%	62.2%
	93	43.20%	24.7%	37.2%	61.9%
	94	36.90%	28.1%	26.5%	54.6%
	95	39.20%	35.7%	24.6%	60.3%
	Average	41.43%	26.1%	33.6%	59.8%
150-90	92	50.30%	16.4%	42.5%	58.9%
	93	41.20%	30.8%	25.1%	55.9%
	94	44.00%	43%	15%	58%
	95	46.60%	46%	16.1%	62.1%
	Average	45.53%	34.1%	24.7%	58.7%
240-150	92	61.80%	59.5%	12.1%	71.6%
	93	58.10%	13.3%	56.9%	70.2%
	94	56.70%	24.7%	42.9%	67.6%
	95	42.30%	42.2%	21.4%	63.7%
	Average	54.73%	34.9%	33.3%	68.3%
360-240	92	61.90%	61.3%	19.4%	80.7%
	93	64.90%	31.8%	46.3%	78.1%
	94	66.30%	36.6%	48.2%	84.8%
	95	56.30%	46.5%	29.9%	76.4%
	Average	62.35%	44.1%	36%	80%

Table 4.6: Smile Analysis on the Strike Metric: Percentage of Variance Explained by the Unrotated first PC and by the Rotated PCs.

Year	30-10, 60-30, 90-60, 150-90	240-150, 360-240
92	1.2%	2.4%
93	1.1%	2.2%
94	1.1%	2.2%
95	1.1%	2%

Table 4.7: Minimum Step-Size for the Moneyness Metric.

find that the parallel shock of the first component dominates in the expiries 30-10, 240-150, and 360-240. In the other ranges, it is the Z-shaped shock which explains more of the cumulative variance in many years.

4.4 PCA on the Moneyness Metric

4.4.1 Construction of the Moneyness Metric

To construct our variables, we first choose a grid of moneyness points for them. We measure implied volatilities at different s, fixed, $\frac{K_i - F_t}{F_t}\%$ levels for $i = 1, 2, \ldots s$. This involves interpolation across the implied volatilities for these fixed variables, since F_t is always changing[32].

In choosing this grid we need to be careful not to make it too fine. If we allowed two different grid points to fall between adjacent strikes, they would both be interpolated from the same two data points. This would produce spurious dependence which would distort our results. We have therefore chosen moneyness levels which are slightly coarser than the coarsest spacing of the strikes.

Table 4.7 shows the minimum distance that the moneyness levels should be apart (minimum step-size), for the different expiries over the years. This has been calculated as $\frac{\Delta K}{F_{\min}} \times 100$, where $\Delta K = K_{i+1} - K_i$, for $i = 1, 2, \ldots, s-1$, and $F_{\min}$ is the minimum futures price for each year (recall that ΔK is either 5, or 10). Ideally, we would like the spacing to be as coarse, as possible, so as to avoid any effect from noisy data, which may have remained in our data sets,

[32]We use linear interpolation instead of some other interpolation scheme, e.g. cubic spline, because we have screened the data for any noise. Hence, there are no abrupt changes in the data, and consequently cubic spline interpolation is not necessary (see Chapra and Canale [32]).

despite the screening. However, there is a trade-off between the spacing, and the number of variables to which we will apply the PCA. Hence, we have to check out how many variables we get for a given step-size.

Next, the range of moneyness variables (e.g. -18% up to 4%) that we apply the PCA to is determined, so as to get the maximum number of observations. For every day, we find the minimum and maximum moneyness, and we do distribution graphs for them. Subsequently, the range that we choose, is set by taking the right hand tail from the distribution of the minimum moneyness and by taking the left hand tail from the distribution of the maximum moneyness.

Table 4.8 shows the final choice of grid for each expiry range (chosen step size) between the moneyness levels. We also report the number of variables available on which to perform the PCA, the number of observations (after the listwise deletion), and the KMO measure of correlation between the variables.

We can see that (a) we have a sufficiently wide range of moneyness variables with a step-size which is greater than the coarsest spacing of the strikes; we set the spacing to 1.5%, for all the years in the ranges 30-10, 60-30, 90-60 and 150-90, so as to have consistency across the years in the same expiry. In the range 240-150 we set the spacing to 2.5%. In the range 360-240 we set the spacing to its lower bound, so as to maximize the number of variables, (b) in general, the number of observations does not fall below 100, and (c) the KMO coefficients are between 0.70 and 0.90. (a), (b), and (c) make the application of the PCA feasible[33].

4.4.2 Some Descriptives for the Chosen variables

Figures 4.41-4.44 show the average implied volatilities of the levels of the moneyness variables that we have chosen for the application of the PCA. Figures

[33]Since we have set fixed moneyness levels, the PCA could be applied to the whole sample from 1992-95, as well (rather than just to the separate years). However, we prefer to perform the technique on each year separately, so as to be consistent with the way that it was applied to the strike metric (where the analysis of the whole sample is not possible, since we have different strikes for each year). Moreover, the examination of the dynamics of implied volatilities for distinct years offers a more in-depth investigation of the issue.

Range	Step	Year	Number of Variables	Number of Obs.	KMO
30-10	1.5%	92	7	111	0.87241
		93	7	93	0.88451
		94	7	111	0.89955
		95	7	113	0.87486
60-30	1.5%	92	10	205	0.81014
		93	9	226	0.81702
		94	10	194	0.88138
		95	9	220	0.81545
90-60	1.5%	92	8	106	0.77715
		93	7	116	0.71837
		94	10	119	0.85393
		95	8	118	0.72414
150-90	1.5%	92	10	161	0.82847
		93	10	120	0.82835
		94	9	147	0.86238
		95	9	108	0.83825
240-150	2.5%	92	8	220	0.87038
		93	10	221	0.914
		94	8	224	0.87067
		95	9	192	0.8541
360-240	2.4%	92	7	99	0.84789
		93	5	104	0.80401
		94	6	102	0.83461
		95	5	57	0.79815
360-240	2.4%	92	7	99	0.84789
	2.2%	93	5	113	0.81593
	2.2%	94	7	101	0.88693
	2%	95	6	67	0.88729

Table 4.8: Smile Analysis on the Moneyness Metric: Number of Variables, Number of Observations, Chosen Step-Size, and the KMO measure.

4.45-4.48 show the standard deviations of the first differences of the chosen variables. The information is reported across the expiry buckets for every year. The average implied volatility in year 1992 ranges from 0.19 to 0.12. In 1993 it varies from 0.16 to 0.09, and in 1994 from 0.19 to 0.09. Finally, in 1995 it lies in the interval 0.18-0.09. Again, we can see that the standard deviations in the shorter expiries, are not more volatile than those in the longer expiries.

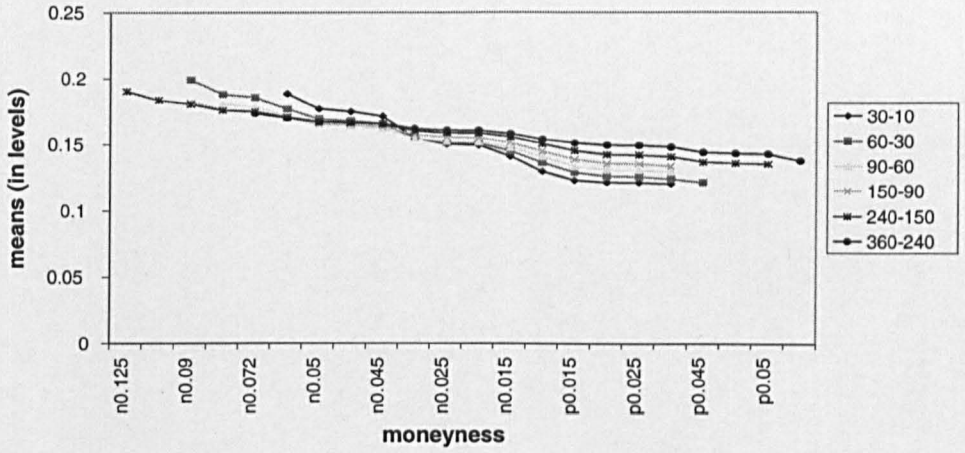


Figure 4.41: Smile Analysis on the Moneyness Metric: Average Implied Volatilities (levels) in 1992.

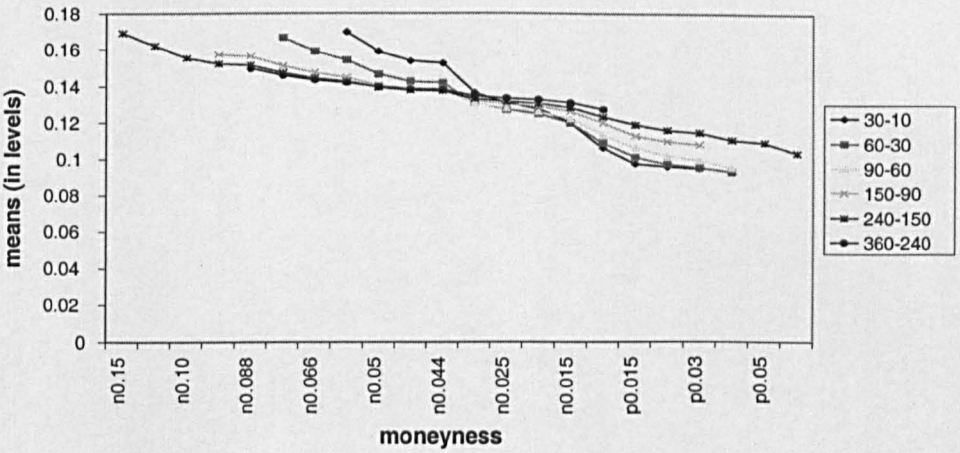


Figure 4.42: Smile Analysis on the Moneyness Metric: Average Implied Volatilities (levels) in 1993.

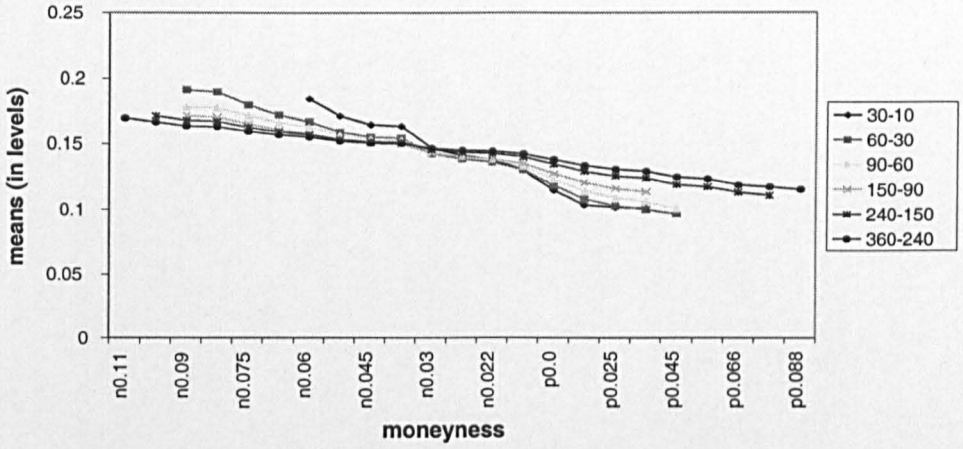


Figure 4.43: Smile Analysis on the Moneyness Metric: Average Implied Volatilities (levels) in 1994.

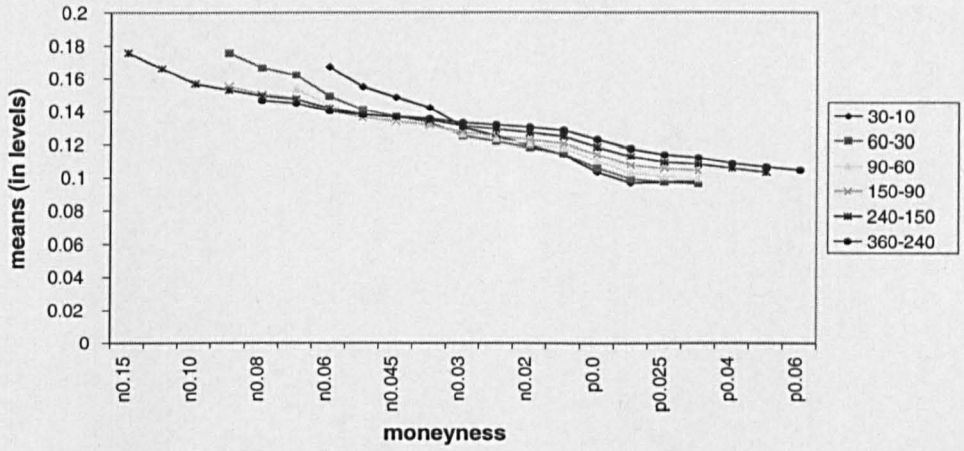


Figure 4.44: Smile Analysis on the Moneyness Metric: Average Implied Volatilities (levels) in 1995.

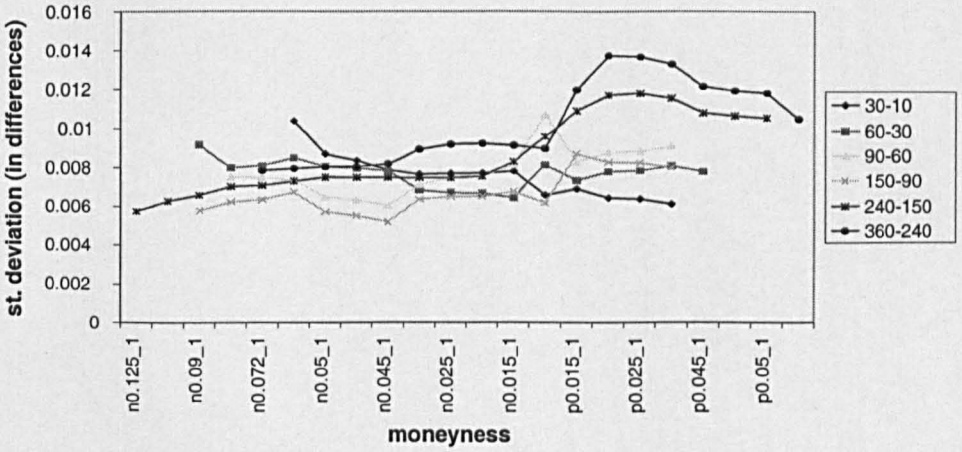


Figure 4.45: Smile Analysis on the Moneyness Metric: Standard Deviation of Implied Volatilities (in differences) in 1992.

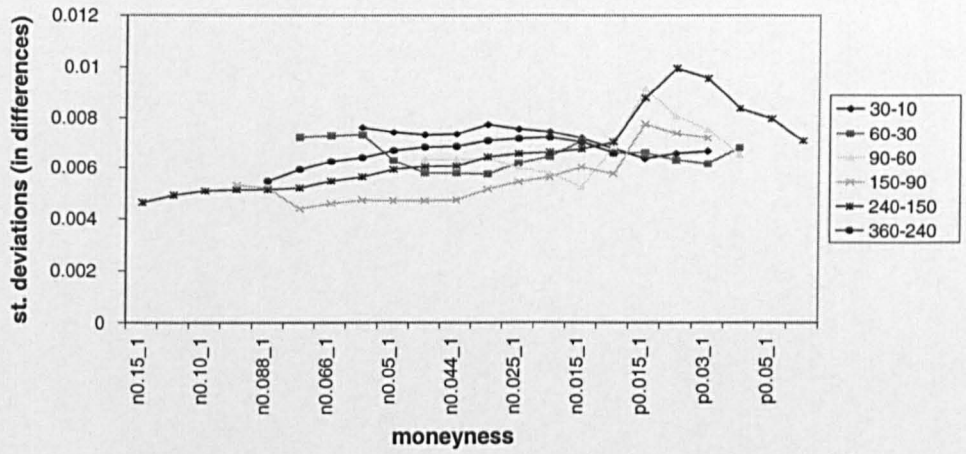


Figure 4.46: Smile Analysis on the Moneyness Metric: Standard Deviation of Implied Volatilities (in differences) in 1993.

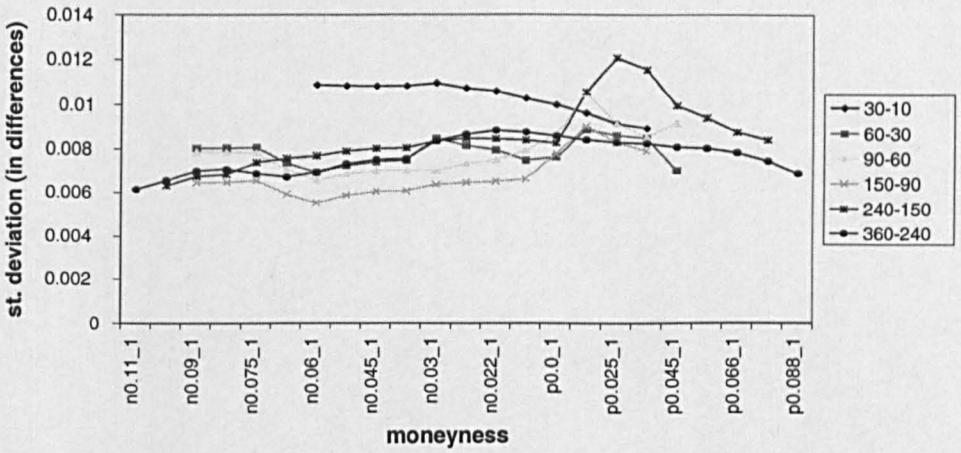


Figure 4.47: Smile Analysis on the Moneyness metric: Standard Deviation of Implied Volatilities (in differences) in 1994.

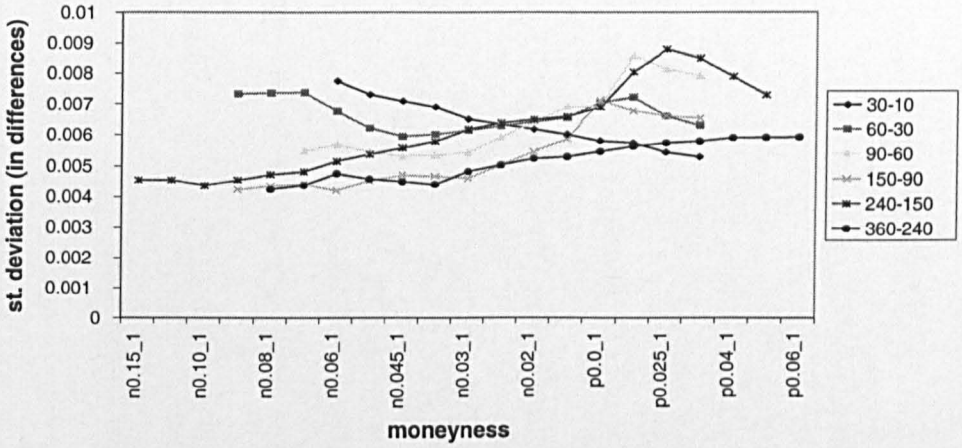


Figure 4.48: Smile Analysis on the Moneyness Metric: Standard Deviation of Implied Volatilities (in differences) in 1995.

4.4.3 Number of Retained Principal Components and a First Interpretation

In this section, we check first for multivariate normality of the moneyness variables. We conclude that we can not use any parametric tests for deciding on the number of components to retain. Then, we apply three criteria to decide how many components to retain, just as we did for the strike metric.

Testing for Normality

We test for multivariate normality, by applying the BJ test. Table 4.9 shows the results from the test. The null hypothesis of univariate normality is rejected in all the cases; the only exception is in the range 30-10 days to expiry, in 1995. Since acceptance of normality when using this test does not confirm it, we look at the Q-Q plot for this case. Figure 4.49 shows that the differences of implied volatilities for the variable -1.5% are not normally distributed.

Range	Year	Variable	n	Skewness	Ex. Kurtosis	Test Value	P-Value
30-10	92	P0.015_1	151	-0.014	1.579	15.6916	0.0004
	93	N0.015_1	146	-0.019	2.055	25.6989	0.0000
	94	P0.015_1	147	0.737	2.822	62.0852	0.0000
	95	N0.015_1	149	0.267	0.441	2.9777	0.2256
60-30	92	N0.015_1	238	-0.134	1.639	27.3516	0.0000
	93	N0.015_1	235	0.101	3.403	113.7910	0.0000
	94	P0.045_1	215	0.143	4.053	147.8896	0.0000
	95	P0.0_1	229	0	5.993	342.6990	0.0000
90-60	92	P0.0_1	159	0.473	3.894	106.3853	0.0000
	93	P0.015_1	183	-0.24	3.468	93.4629	0.0000
	94	P0.03_1	188	-0.85	7.072	414.4083	0.0000
	95	P0.0_1	170	0.527	2.204	42.2771	0.0000
150-90	92	N0.045_1	168	0.377	1.663	23.3386	0.0000
	93	N0.015_1	168	-0.012	2.047	29.3355	0.0000
	94	P0.015_1	167	0.164	10.979	839.4953	0.0000
	95	N0.06_1	146	0.038	4.956	149.4536	0.0000
240-150	92	P0.0_1	245	0.011	1.529	23.8704	0.0000
	93	P0.0_1	242	-0.343	1.831	38.5502	0.0000
	94	P0.0_1	240	0.302	2.15	49.8732	0.0000
	95	P0.0_1	203	-0.086	3.381	96.9388	0.0000
360-240	92	N0.048_1	184	0.133	2.935	66.5849	0.0000
	93	P0.0_1	164	0.001	3.534	85.3426	0.0000
	94	N0.11_1	176	0.819	3.534	111.2628	0.0000
	95	N0.08_1	166	0.797	2.153	49.6357	0.0000

Table 4.9: Smile Analysis on the Moneyness Metric: Bera-Jarque Test for Univariate Normality. P(N) X-1 denotes the differenced once implied volatilities corresponding to plus (minus) moneyness level X.

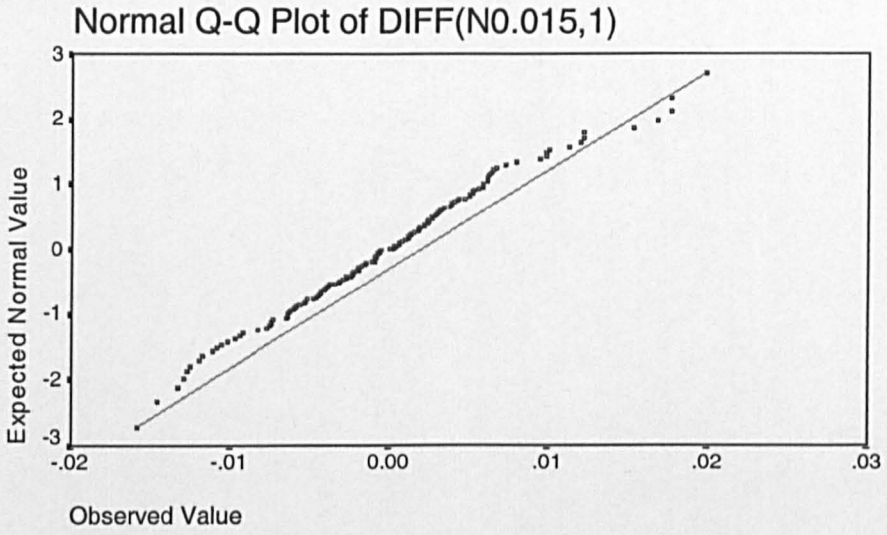


Figure 4.49: Smile Analysis on the Moneyness Metric: Q-Q Plot for the First Differences of Implied Volatilities of Moneyness -1.5%, in the range 30-10 for 1995.

Hence, in all the ranges and for all the years, the null hypothesis of univariate and therefore of multivariate normality is rejected. Consequently, just as in the strike metric, we can not use any parametric tests for determining the number of components to be retained. Therefore, we have to use again the non-parametric Velicer's criterion.

Velicer's Criterion

Table 4.10 shows the number of components retained under Velicer's criterion, and under the mean eigenvalue rule of thumb in the moneyness metric. It also shows the percentage of the variance explained by the first three components. These results are reported for the six separate expiry buckets and over the four one year periods. We can see that under Velicer's criterion we should retain either one or two PCs. This is more conservative than the mean eigenvalue rule of thumb which in some expiries retains one more PC than Velicer's. Note that $f_0 > f_1$ showing that at least one component can be extracted.

The first component explains between 40% and 86% of the variance, depending

on the year and the expiry range under consideration. The second component explains between 7% and 31%. The two components combined, explain between 65% and 92% of the variance. This is 5%-12% bigger than the amount of the variance explained by the two components in the strike metric. The highest proportions are explained for the shortest and longest expiry options, just as with the strike metric.

Communalities

We next look at the communalities that the retained, according to Velicer's criterion, PCs explain. In Appendix C, in Tables C.1 and C.2, we present the communalities (expressed in percentage terms) explained by retaining one and two PCs, respectively, for the year 1994.

In general, the explained by the two PCs communalities, are bigger than what they were in the strike metric. We can see that in the ranges 150-90 and 240-150, where Velicer's criterion keeps one PC, there are some cases where the explained by one PC communalities are rather low. Adding the second PC increases significantly the communalities (e.g. in the 240-150 range for the ATM implied volatility, there is an increase from 5.49% to 72.87%). The results are similar for the other years; in the ranges where Velicer's criterion keeps one PC we should add and the second. Hence, according to the explained communalities, we should retain two PCs on the moneyness metric.

Interpretation of the Components

We now look at the interpretation of the PCs. Figures 4.50-4.55 show the correlation loadings of the first PC on the moneyness variables that we apply the PCA to, for each expiry bucket over the different years. These figures show that in the range 30-10, the first PC is like a parallel shift with a slight attenuation at the edges. In the range 60-30, it has positive correlation loadings in all the years, but 1992. In the other ranges, it has a mixture of positive and negative

Range	Year	f_0	f_1	f_2	f_3	r*	l	1st PC	2nd PC	3rd PC
30-10	1992	0.4818	0.4792	0.4801	0.4807	1	1	73.30	11.00	5.50
	1993	0.5521	0.5491	0.5500	0.5508	1	1	77.80	8.10	6.00
	1994	0.6959	0.6936	0.6946	0.6949	1	1	85.70	6.40	2.30
	1995	0.4764	0.4735	0.4745	0.4752	1	1	73.00	11.10	5.90
60-30	1992	0.1816	0.1812	0.1810	0.1811	2	2	39.60	30.30	8.10
	1993	0.2065	0.2058	0.2058	0.2059	2	2	44.60	28.70	7.70
	1994	0.2761	0.2747	0.2751	0.2753	1	2	53.60	22.90	5.30
	1995	0.2189	0.2183	0.2183	0.2185	2	2	47.60	25.50	8.70
90-60	1992	0.2291	0.2276	0.2274	0.2278	2	2	48.80	27.30	9.00
	1993	0.2164	0.2133	0.2132	0.2137	2	2	46.80	30.60	9.70
	1994	0.2168	0.2153	0.2152	0.2154	2	2	44.20	29.60	7.40
	1995	0.1395	0.1383	0.1381	0.1384	2	2	39.70	25.50	9.50
150-90	1992	0.2326	0.2313	0.2316	0.2317	1	2	51.50	16.80	7.50
	1993	0.1989	0.1969	0.1972	0.1975	1	2	48.10	16.90	8.90
	1994	0.2578	0.2560	0.2563	0.2565	1	2	54.20	18.00	8.00
	1995	0.2008	0.1985	0.1988	0.1991	1	2	48.80	18.10	8.50
240-150	1992	0.4314	0.4302	0.4303	0.4307	1	1	69.20	11.90	7.60
	1993	0.4334	0.4322	0.4324	0.4327	1	2	68.70	10.40	6.50
	1994	0.3595	0.3585	0.3588	0.3590	1	2	63.40	16.90	7.40
	1995	0.3224	0.3211	0.3213	0.3215	1	2	60.70	12.80	7.30
360-240	1992	0.5668	0.5653	0.5657	0.5663	1	1	77.90	14.10	4.40
	1993	0.4933	0.4917	0.4922	0.4928	1	1	75.10	16.90	4.10
	1994	0.6852	0.6826	0.6839	0.6844	1	1	85.10	7.80	2.50
	1995	0.5932	0.5874	0.5894	0.5911	1	1	80.90	6.80	5.80

Table 4.10: Principal Components in the Smile Analysis on the Moneyness Metric: r* = number of components retained under Velicer's criterion (minimum f0,...f3), l = number of components retained under rule of thumb with percentage variance explained by components 1-3.

correlation loadings in all the years[34].

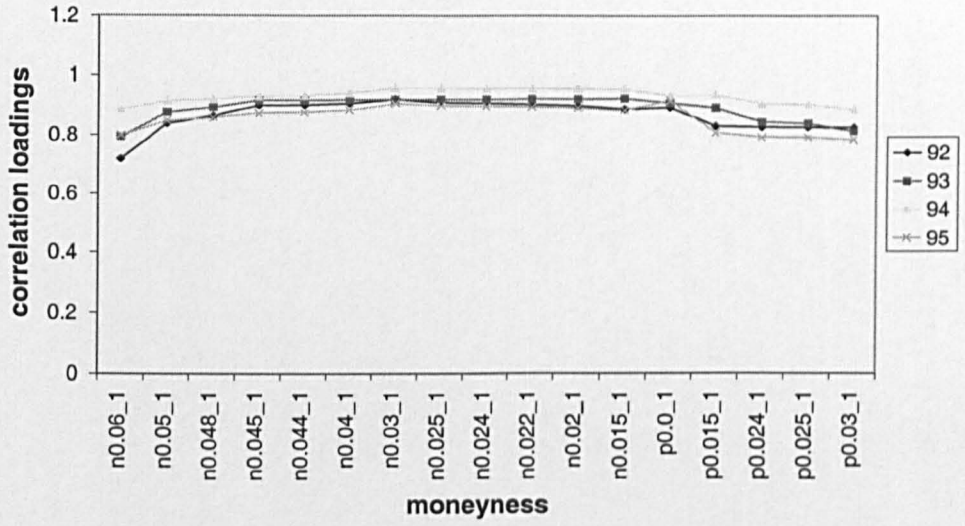


Figure 4.50: Smile Analysis on the Moneyness Metric: Interpretation of the First PC for 30-10.

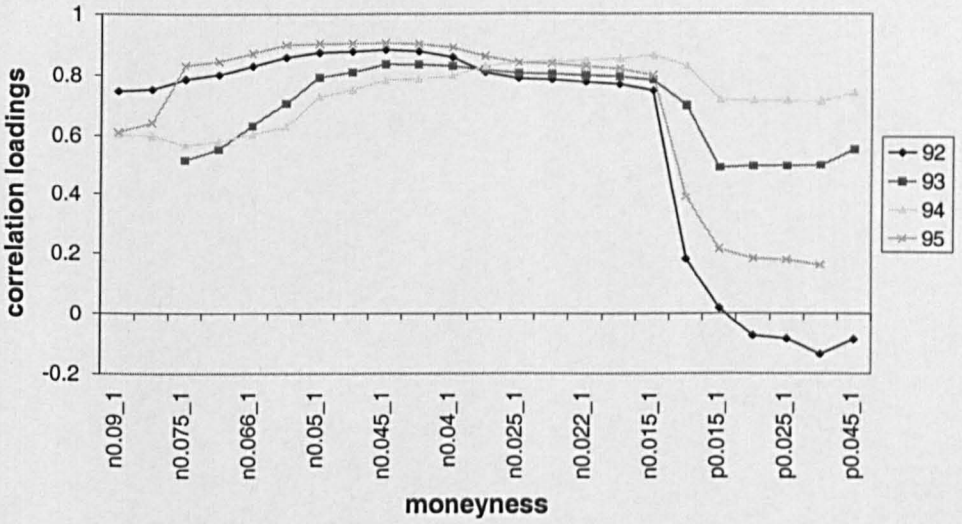


Figure 4.51: Smile Analysis on the Moneyness Metric: Interpretation of the First PC for 60-30.

[34]In some of the graphs, we have reversed the sign of the correlation loadings of the first PC, so as to have a consistent exhibition. This does not affect the scaling, nor the orthogonality property of the correlation loadings. Moreover, whether we reverse the sign or not, does not change the results from our Procrustes rotation.

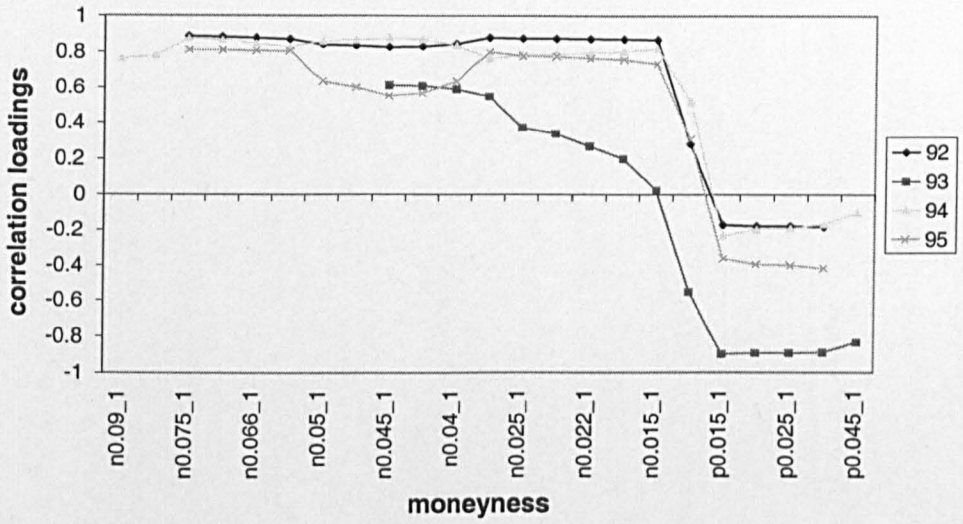


Figure 4.52: Smile Analysis on the Moneyness Metric: Interpretation of the First PC for 90-60.

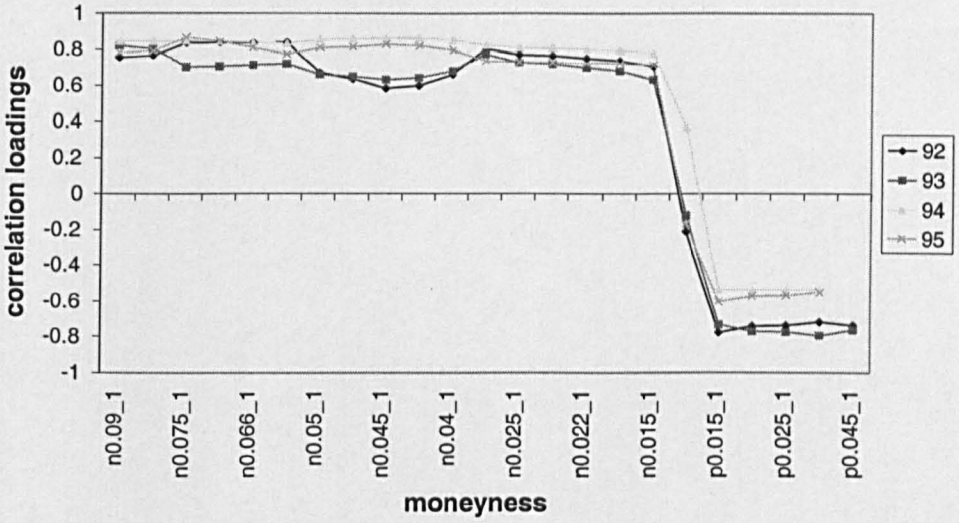


Figure 4.53: Smile Analysis on the Moneyness Metric: Interpretation of the First PC for 150-90.

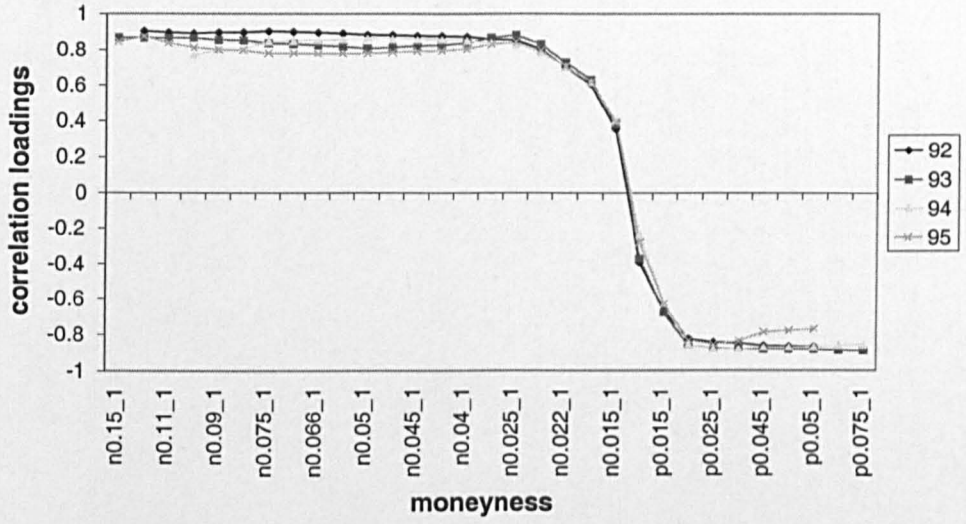


Figure 4.54: Smile Analysis on the Moneyness Metric: Interpretation of the First PC for 240-150.

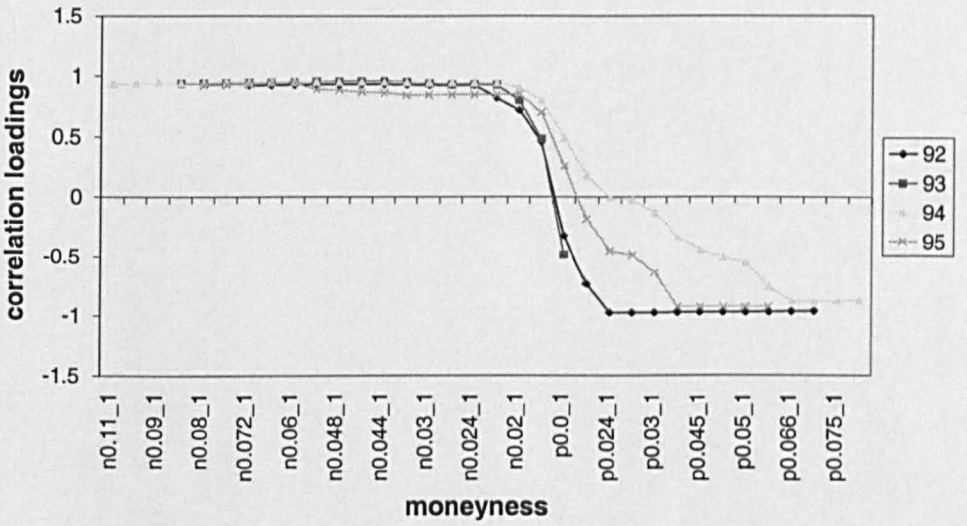


Figure 4.55: Smile Analysis on the Moneyness Metric: Interpretation of the First PC for 360-240.

Figures 4.56-4.61, show the correlation loadings of the second PC for each expiry over the different years. In the 30-10 and 60-30 expiry ranges, the second PC has a Z-shape for all years. However, in the range 90-60 it has positive correlation loadings in 1993, and a mixture of both positive and negative loadings in the other years. In the range 150-90, it has positive correlation loadings in

all the years, but 1994. In the range 240-150 the shape is triangular, with big correlation loadings for the ATM options. Finally, in the range 360-240 it has positive correlation loadings for all the years, but 1993.

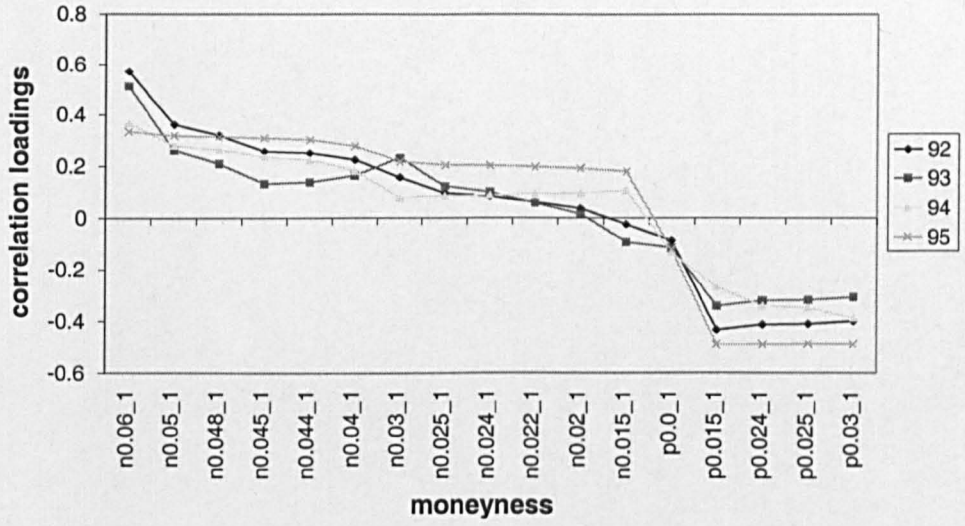


Figure 4.56: Smile Analysis on the Moneyness Metric: Interpretation of the Second PC for 30-10.

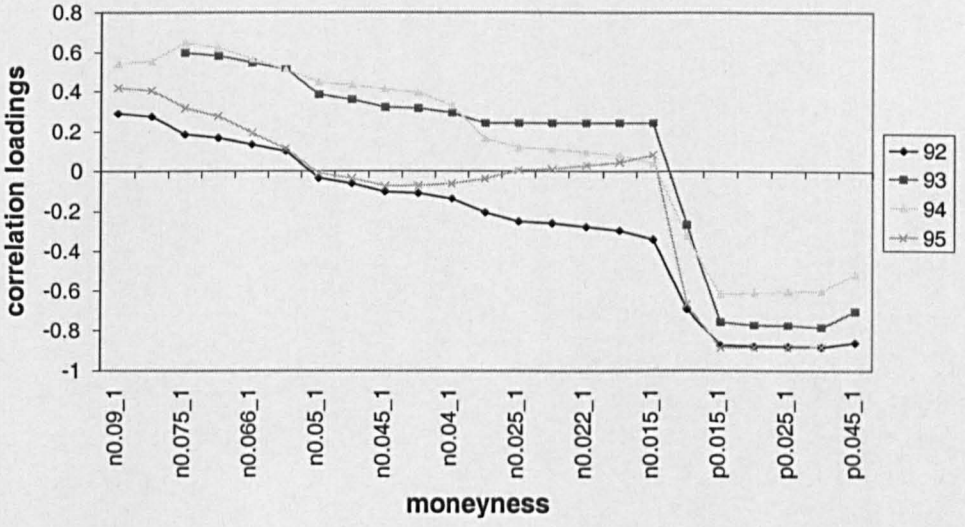


Figure 4.57: Smile Analysis on the Moneyness Metric: Interpretation of the Second PC for 60-30.

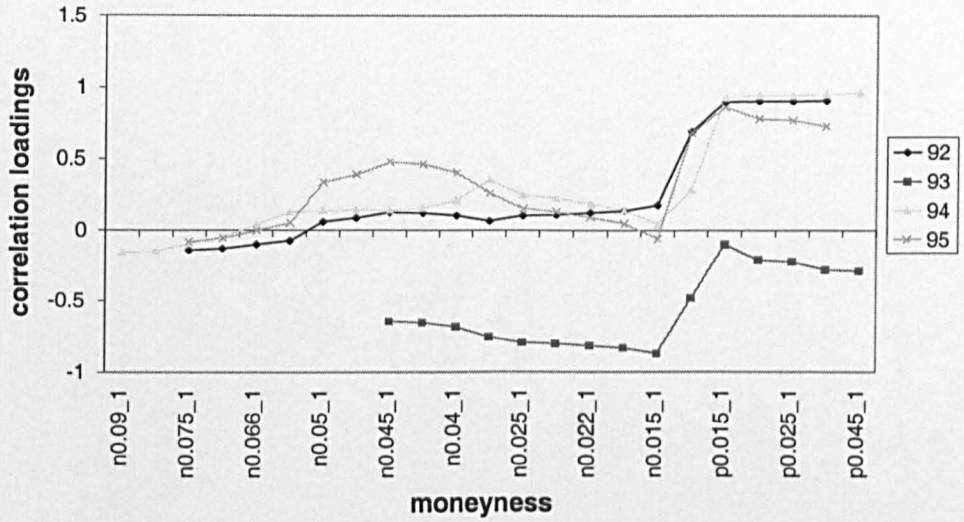


Figure 4.58: Smile Analysis on the Moneyness Metric: Interpretation of the Second PC for 90-60.

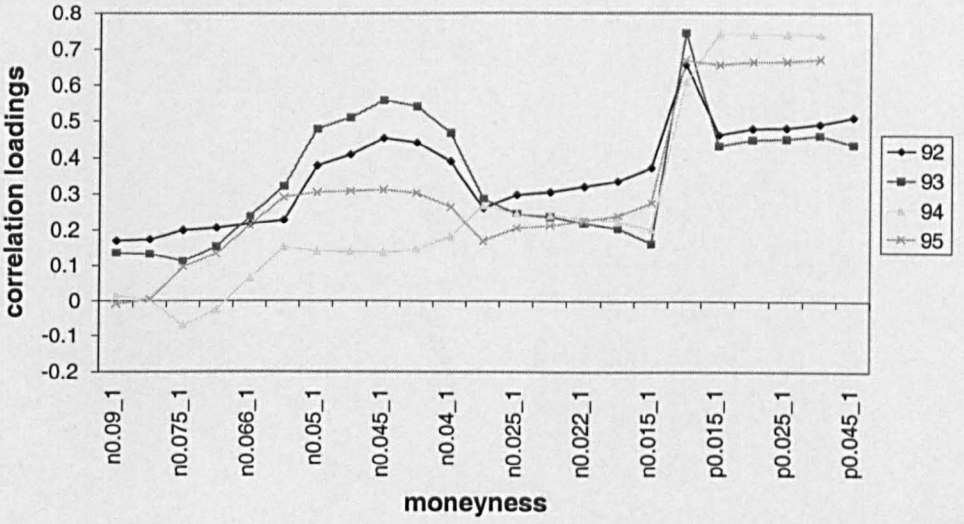


Figure 4.59: Smile Analysis on the Moneyness Metric: Interpretation of the Second PC for 150-90.

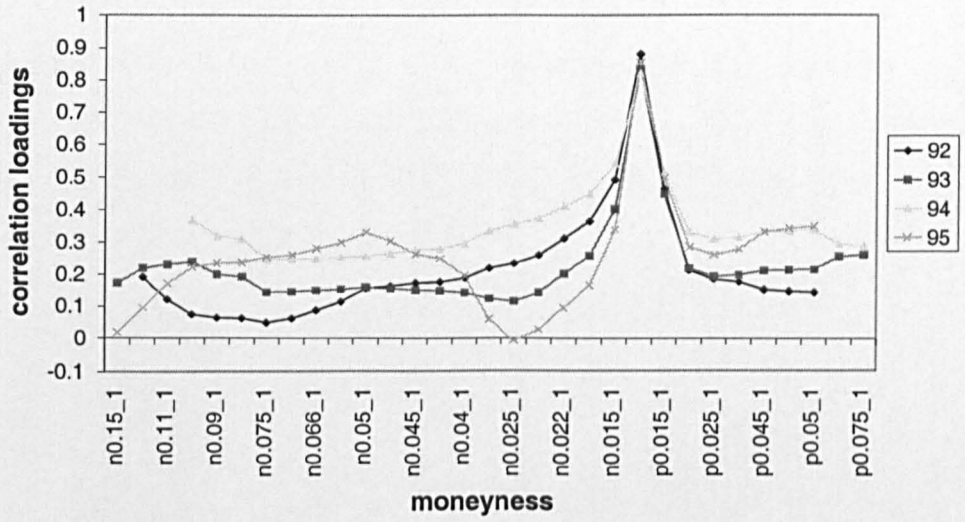


Figure 4.60: Smile Analysis on the Moneyness Metric: Interpretation of the Second PC for 240-150.

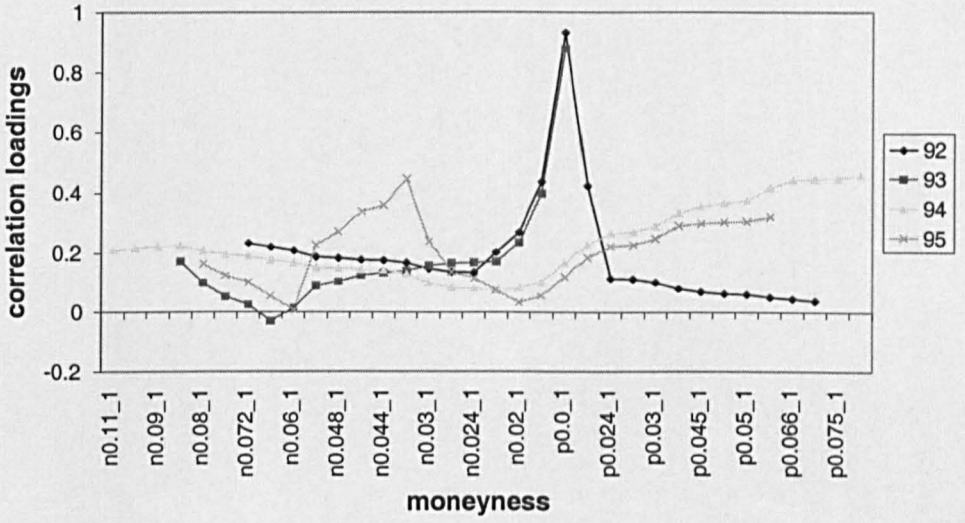


Figure 4.61: Smile Analysis on the Moneyness Metric: Interpretation of the Second PC for 360-240.

Figures 4.62-4.67 show the correlation loadings of the third PC for each expiry over the different years. The figures reveal that the third PC has a systematic pattern in some expiry ranges (e.g. 30-10 and 240-150), i.e. it increases the implied volatilities in the low and high strikes, and it decreases them in the middle strikes. On the other hand, in some other ranges (e.g. in the range 60-30 in year 1994, in the 90-60 in 1994 and 1995, or in the range 150-90) is just

noise. Given that Velicer's criterion retains at most two PCs, and the explained communalities by the first two PCs are satisfactory, we judge that the number of components to be retained in the moneyness metric is two. *Thus, the number of components that we retain in the moneyness metric, is the same as the number that we retain in the strike metric.*

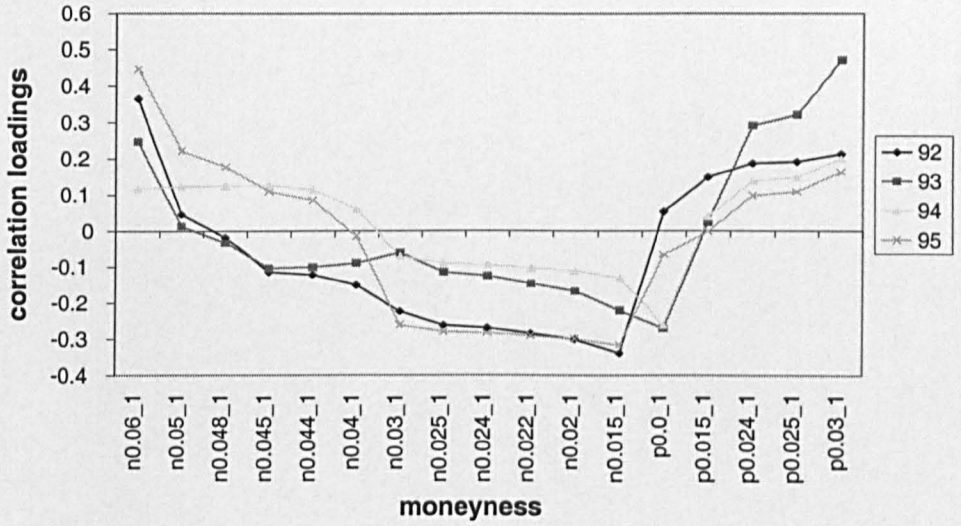


Figure 4.62: Smile Analysis on the Moneyness Metric: Interpretation of the Third PC for 30-10.

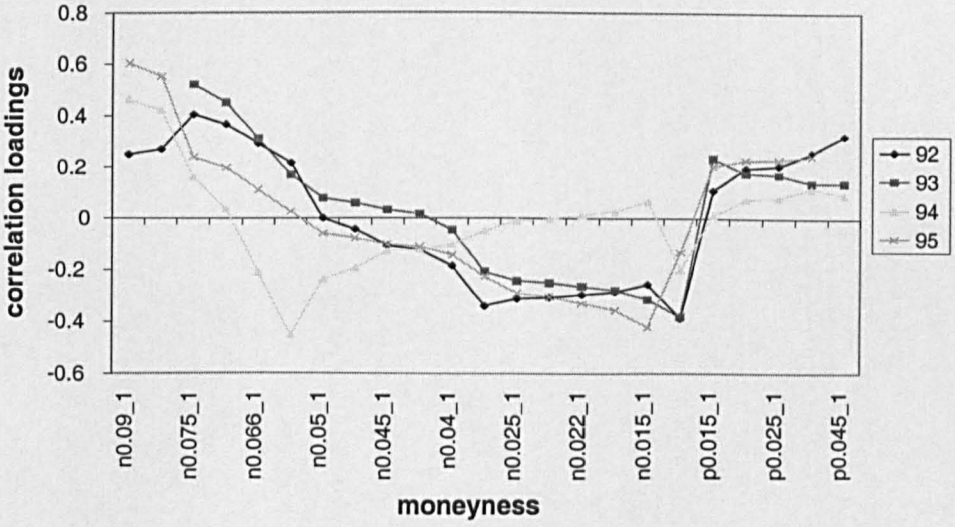


Figure 4.63: Smile Analysis on the Moneyness Metric: Interpretation of the Third PC for 60-30.

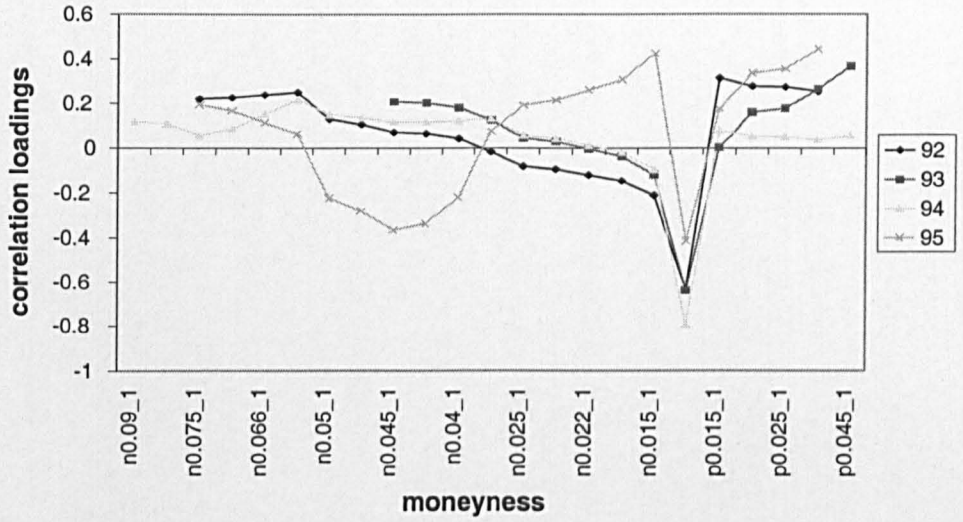


Figure 4.64: Smile Analysis on the Moneyness Metric: Interpretation of the Third PC for 90-60.

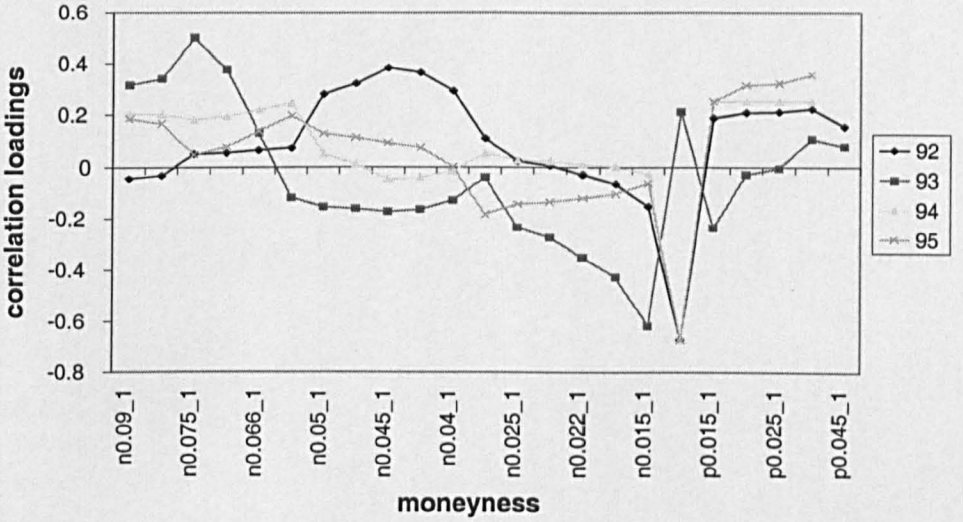


Figure 4.65: Smile Analysis on the Moneyness Metric: Interpretation of the Third PC for 150-90.

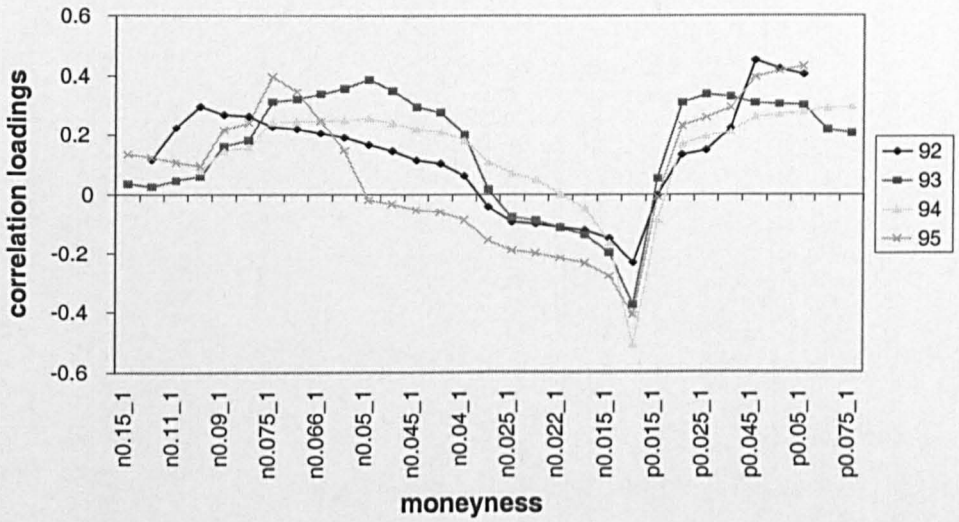


Figure 4.66: Smile Analysis on the Moneyness Metric: Interpretation of the Third PC for 240-150.

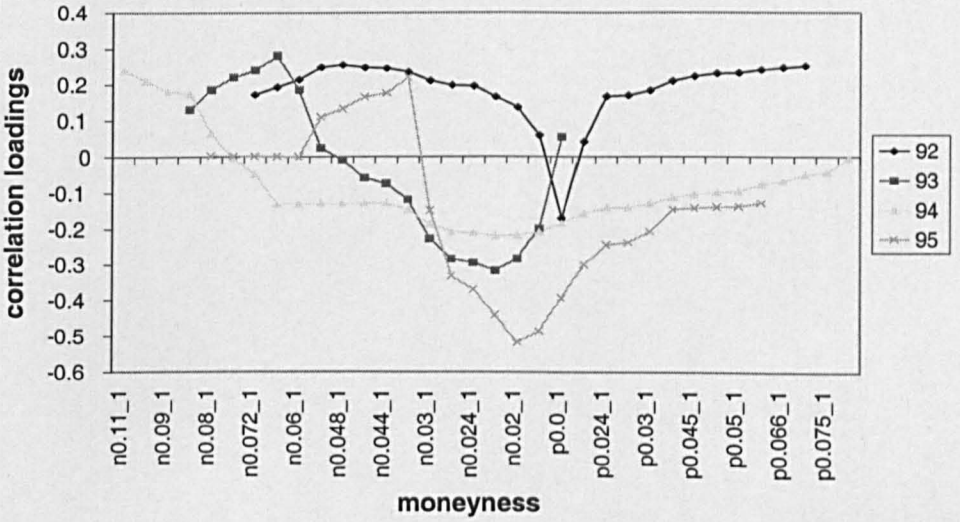


Figure 4.67: Smile Analysis on the Moneyness Metric: Interpretation of the Third PC for 360-240.

However, even if we were able to get the shift and slope interpretation for the ranges 30-10 and 60-30, we would like to get this neat interpretation for the other ranges, as well. Hence, we will perform again our "Procrustes" type rotation.

Year	30-10	60-30	90-60	150-90	240-150	360-240
92	1.12	38.54	39.72	72.57	64.98	83.70
93	0.7	6.95	64.73	73.00	72.20	40.34
94	0.14	3.33	41.13	48.45	85.40	62.34
95	1.53	21.08	48.41	60.44	59.57	68.46

Table 4.11: Smile Analysis on the Moneyness Metric: Angle of the Procrustes Type of Rotation (Angles are Measured in Degrees).

4.4.4 The Rotation Method

Table 4.11 shows the angle that the "Procrustes" type of rotation has achieved. We can see that in the ranges 30-10 and 60-30, we have not rotated much, in contrast to the other ranges. It is worth recalling that in the first two ranges, the unrotated first two PCs had a shift and a Z-shape interpretation, respectively. Hence, rotation in these ranges was not necessary.

Figures 4.68-4.73, and 4.74-4.79, show the results for the first and second PCs, respectively, for the different expiry buckets. The first rotated PC provides positive loadings for all the expiries and years with the single exception in the 240-150 range, in 1992. In the range 30-10, the parallel shift with the attenuation at the edges has been maintained. In the range 60-30 there are positive loadings in every year, with a much smoother shape than the unrotated ones. Again the shape indicates movement which is close to parallel. In the ranges 90-60, 150-90, and 360-240, there are positive loadings in every year. In the range 240-150 we see much greater attenuation for extreme moneyness, giving rise to a more triangular shape. This shape is also evident for one or two years in the adjacent expiry ranges. In addition, the figures show that the magnitude of the first PC attenuates with expiry. This is especially clear for the 30-10 and 60-30 expiries, but rather less so for the others, presumably because of less liquidity in these expiries.

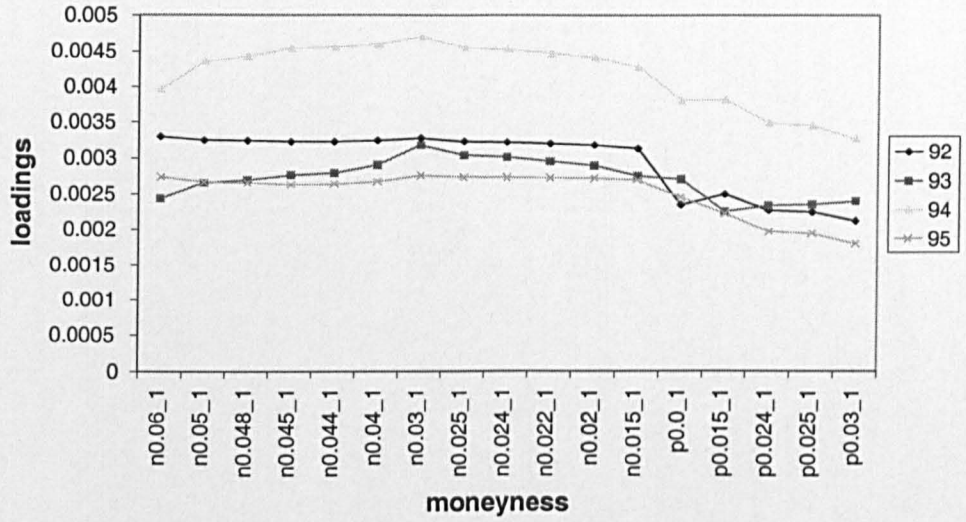


Figure 4.68: Smile Analysis on the Moneyness Metric: Interpretation of the First Rotated PC for 30-10.

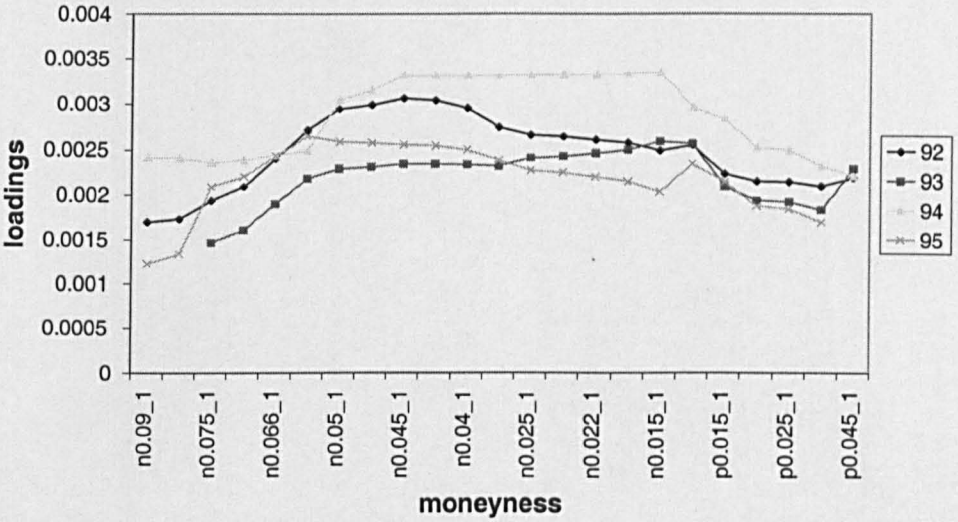


Figure 4.69: Smile Analysis on the Moneyness Metric: Interpretation of the First Rotated PC for 60-30.

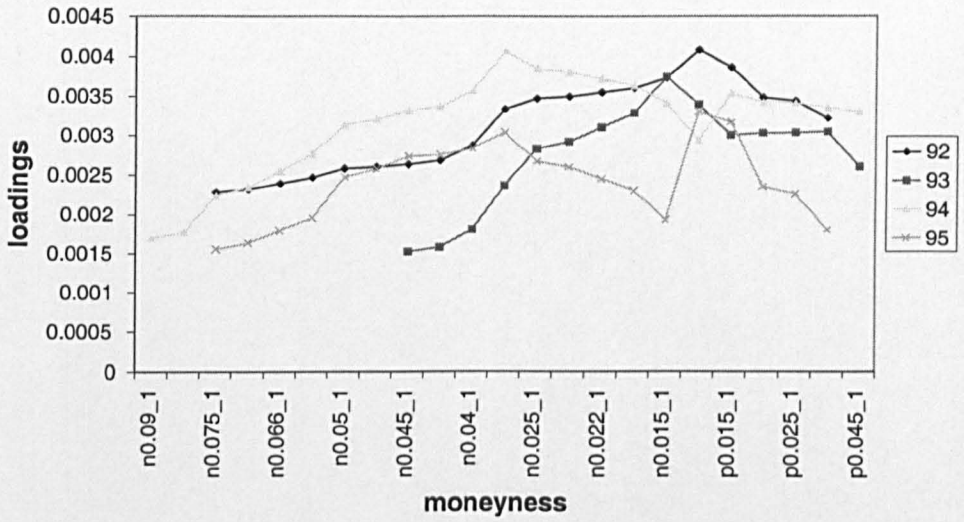


Figure 4.70: Smile Analysis on the Moneneyness Metric: Interpretation of the First Rotated PC for 90-60.

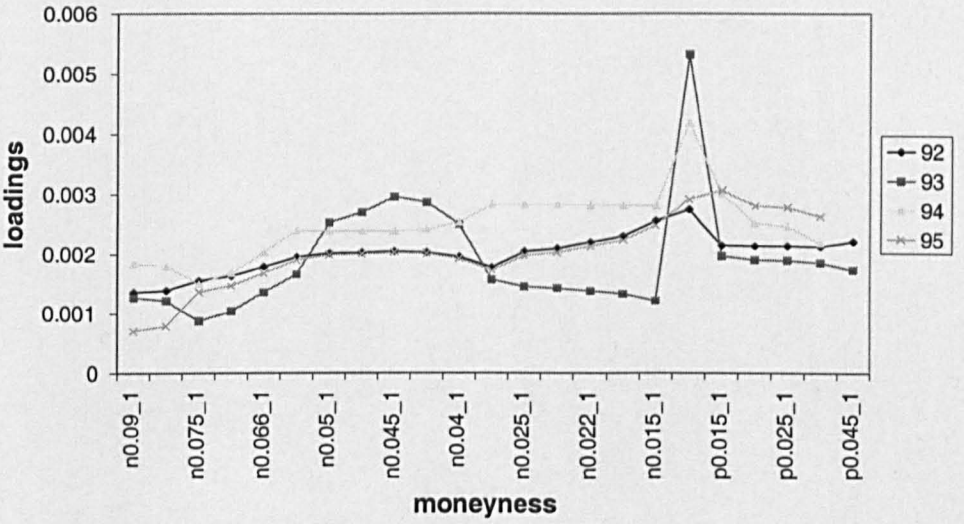


Figure 4.71: Smile Analysis on the Moneyness Metric: Interpretation of the First Rotated PC for 150-90.

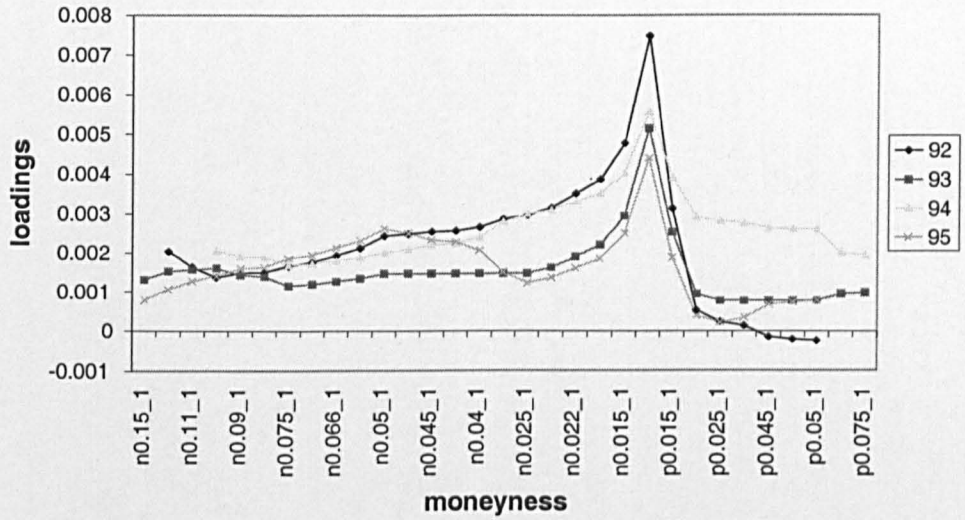


Figure 4.72: Smile Analysis on the Moneyness Metric: Interpretation of the First Rotated PC for 240-150.

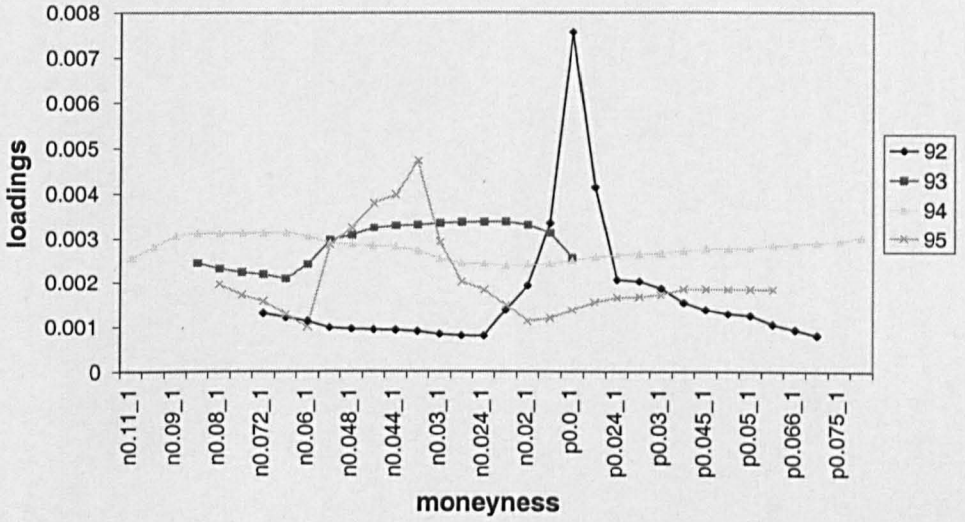


Figure 4.73: Smile Analysis on the Moneyness Metric: Interpretation of the First Rotated PC for 360-240.

The second rotated PC consistently produces a Z-shaped profile of loadings. This component does not show the clear attenuation with expiry that we find for the first component. Although its magnitude is greatest for the shortest expiry range, it changes rather little over the longer expiries.

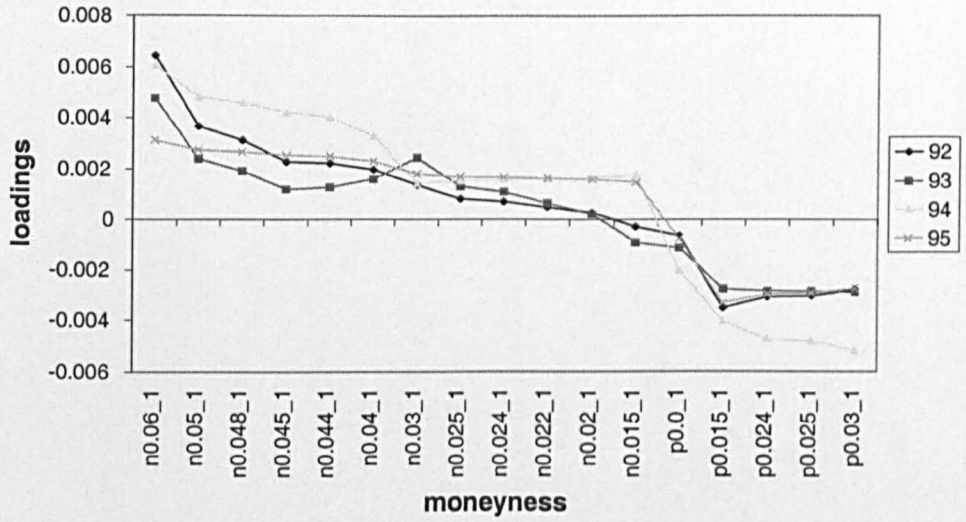


Figure 4.74: Smile Analysis on the Moneyness Metric: Interpretation of the Second Rotated PC for 30-10.

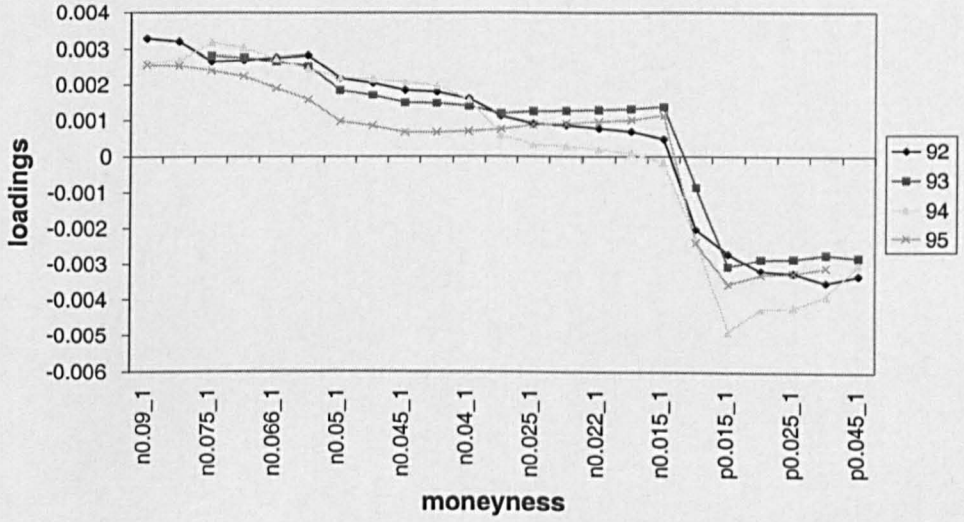


Figure 4.75: Smile Analysis on the Moneyness Metric: Interpretation of the Second Rotated PC for 60-30.

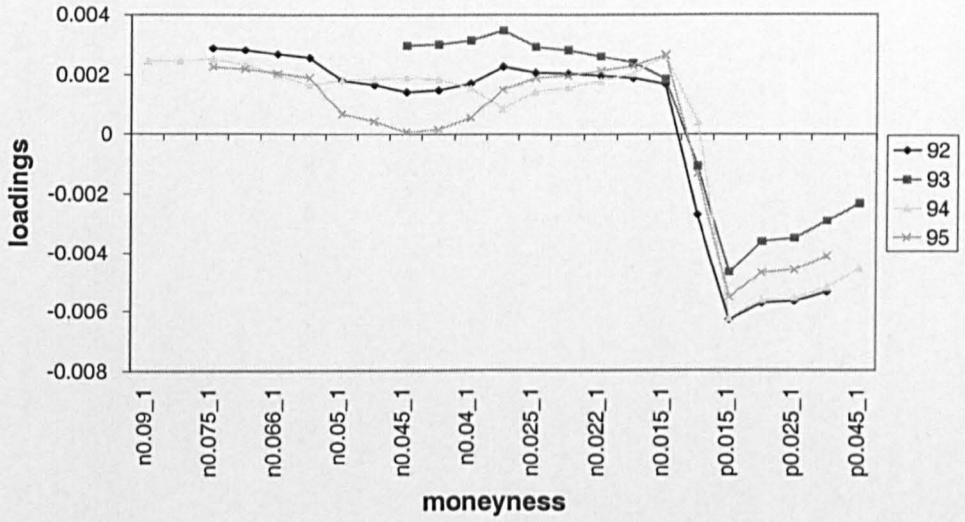


Figure 4.76: Smile Analysis on the Moneyness Metric: Interpretation of the Second Rotated PC for 90-60.

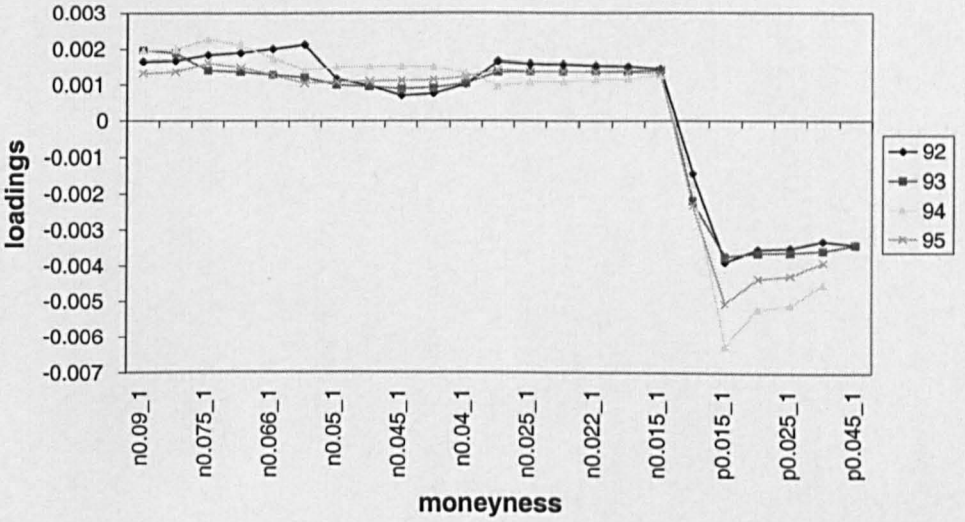


Figure 4.77: Smile Analysis on the Moneyness Metric: Interpretation of the Second Rotated PC for 150-90.

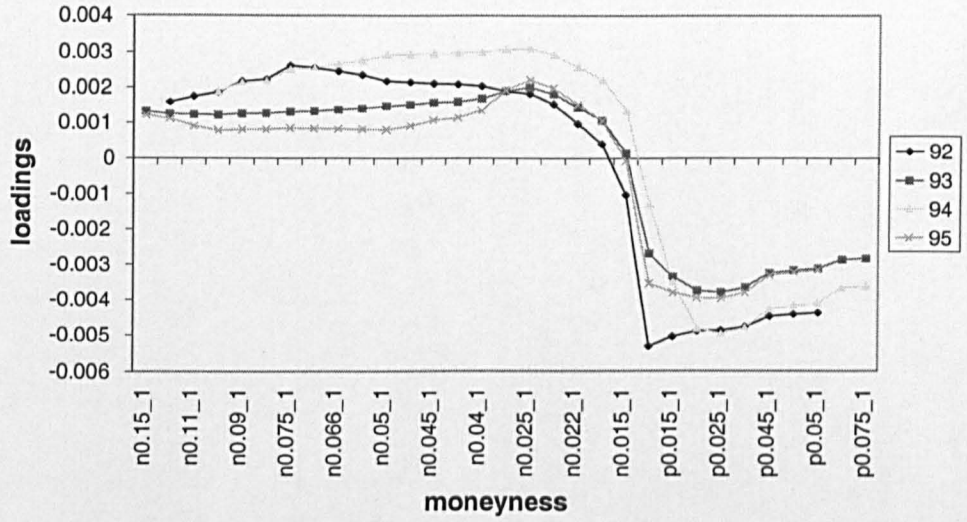


Figure 4.78: Smile Analysis on the Moneyness Metric: Interpretation of the Second Rotated PC for 240-150.

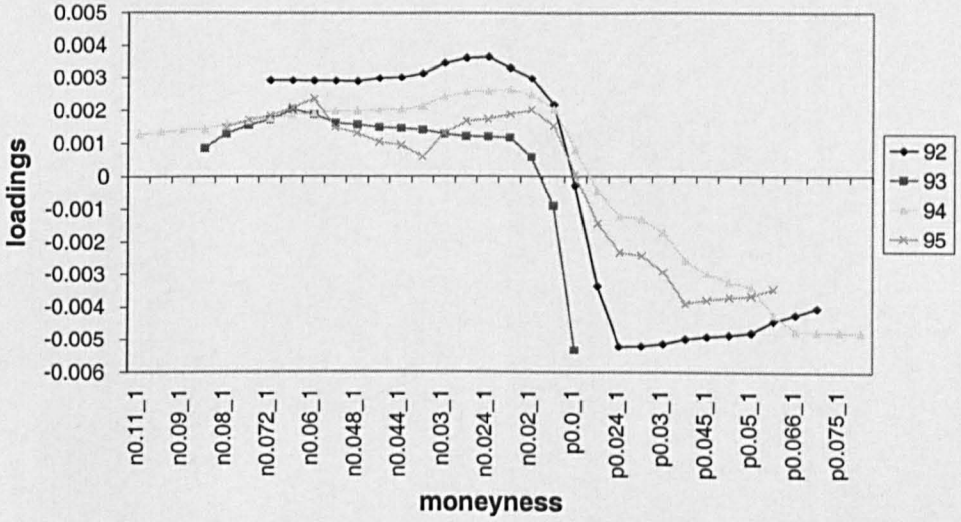


Figure 4.79: Smile Analysis on the Moneyness Metric: Interpretation of the Second Rotated PC for 360-240.

Table 4.12 shows the percentage of the variance that the first and second rotated PCs explain (as well as the percentage for the original first component) on the moneyness metric. We find that the parallel shock of the first component dominates in the shorter expiries of 30-10 and 60-30, while the Z-shaped shock explains more of the variance in the longer ones of 150-90, 240-150 and 360-240.

Range	Year	Unrot. 1st PC	1st PC	2nd PC	Cumulative
30-10	1992	73.30%	73.30%	11.10%	84.40%
	1993	77.80%	77.80%	8.10%	85.80%
	1994	85.70%	85.70%	6.40%	92.10%
	1995	73.00%	73.00%	11.10%	84.10%
	average	77.45%	77.50%	9.20%	86.60%
60-30	1992	39.60%	36.00%	33.90%	69.90%
	1993	44.60%	44.40%	28.90%	73.30%
	1994	53.60%	53.20%	23.30%	76.50%
	1995	47.60%	44.70%	28.30%	73.10%
	average	46.35%	44.60%	28.60%	73.20%
90-60	1992	48.80%	40.00%	36.10%	76.10%
	1993	46.80%	33.50%	43.90%	77.40%
	1994	44.20%	37.90%	35.90%	73.90%
	1995	39.70%	31.80%	33.50%	65.20%
	average	44.88%	35.80%	37.40%	73.20%
150-90	1992	51.50%	19.90%	48.40%	68.30%
	1993	48.10%	19.60%	45.40%	65.00%
	1994	54.20%	34.00%	38.30%	72.30%
	1995	48.80%	25.50%	41.30%	66.90%
	average	50.65%	24.80%	43.40%	68.10%
240-150	1992	69.20%	22.20%	59.00%	81.20%
	1993	68.70%	15.80%	63.20%	79.00%
	1994	63.40%	17.20%	63.10%	80.30%
	1995	60.70%	25.10%	48.40%	73.50%
	average	65.50%	20.10%	58.40%	78.50%
360-240	1992	77.90%	14.90%	77.10%	92.00%
	1993	75.10%	50.70%	41.30%	92.00%
	1994	85.10%	24.50%	68.50%	92.90%
	1995	80.90%	16.80%	70.90%	87.70%
	average	79.75%	26.70%	64.50%	91.20%

Table 4.12: Smile Analysis on the Moneyness Metric: Percentage of Variance Explained by the Unrotated first PC and by the Rotated PCs.

4.5 Serial Correlation and the use of the Differences

At this point we should examine the appropriateness of having chosen to work with first differences in implied volatilities. Performing PCA on the changes of implied volatilities was motivated by the idea that implied volatilities should be close to a random walk. The empirical literature on the serial correlation of the first differences of implied volatilities is slightly mixed. On the one hand, Brenner and Galai [26] find that the daily average at-the-money implied volatility is autocorrelated, but the first differences of it are not, as we would expect. On the other hand, negative serial correlation in the first differences of implied volatilities is likely to arise through non-synchronous data (as Harvey and Whaley [81] suggest) and from "bid-ask bounce" (see Roll [124]).

Even though we have been careful to minimize the sources of measurement error in our data, serial correlation effects could hamper the efficiency of the PCA. We investigated this by applying the following alternative method of filtering out the serial correlation:

Let $\hat{\sigma}_t$ be the observed implied volatility which is measured with noise, and σ_t be the true implied volatility. Then, $\hat{\sigma}_t = \sigma_t + u_t$, where u_t is the measurement error. Moreover, assume that $\sigma_t = \phi\sigma_{t-1} + \varepsilon_t$, where ε_t are the white noise disturbances. The above two equations are equivalent to an ARMA(1,1) process (see Harvey [83]), i.e. $\hat{\sigma}_t = \phi\hat{\sigma}_{t-1} + \eta_t$ -$\vartheta\eta_{t-1}$, where η_t are white noise disturbances. We estimated this model and then we applied PCA to η_t, for several subsets of our data. After correcting for the serial correlation in this way, we obtained results from the PCA which were very similar to those already reported. Hence, we can conclude that PCA is reasonably robust to the fairly minor serial correlation present in our data set. We have preferred in the main body of the study to work with the original first differences of implied volatilities.

4.6 Comparing the Results from the Two Metrics

In this section, we summarize how the results of the PCA differ between the two metrics. The number of variables that the PCA was applied to, the number of observations, and the correlations between the variables, are bigger for the strike metric. The assumption of multivariate normality was rejected for both metrics, something which necessitated the use of a non-parametric test for the number of PCs to be retained. This is Velicer's criterion, which retained either one or two PCs in both metrics, depending on the expiry that we looked at. Looking at another two criteria, we decided to keep two PCs in both metrics. Under the moneyness metric, the two retained PCs explained more of the total variance. They also had higher communalities than in the strike metric.

However, only in the shortest expiries the retained PCs had a simple interpretation. Hence, a rotation had to be done. We chose a "Procrustes" type of rotation, so as the loadings of the first PC to be as flat, as possible. In the strike metric, the rotation brought up the shift interpretation for the first PC in the ranges 30-10, 60-30, and 90-60. In the remaining buckets the first rotated PC had both positive and negative loadings. The second rotated PC had a Z-shape interpretation in all the expiries. In the moneyness metric, the rotation presented the first PC as being a parallel shift in the range 30-10. In the range 60-30, its shape was not far away from indicating a parallel shift, as well. In the remaining ranges, the rotated first PC had positive loadings (shift interpretation). On the other hand, the rotated second PC had a distinct Z-shape in all the ranges. The results about the interpretation of the unrotated and rotated PCs are presented in Tables 4.13 and 4.14, respectively.

In general, the rotated PCs more frequently have an intuitive interpretation in the moneyness metric than in the strike metric. The fact that smiles are driven by two shocks, suggests that implied volatility smiles have non-trivial dynamics that can only be captured by a complex stochastic volatility model. This is

Metric		**Interpretation of the Unrotated PCs**
Strike	1st PC	Shift in 30-10, and 60-30 (two years)
		Z-shape in the remaining ranges, and years
	2nd PC	Z-shape in 30-10, 60-30, and for 1995 in all the ranges
		Triangular shape in the remaining ranges and years
	3rd PC	Noise
Moneyness	1st PC	Shift in 30-10, and 60-30
		Z-shape in the remaining ranges
	2nd PC	Z-shape in 30-10, and 60-30
		Shift in the other ranges (Triangular shape in the last two)
	3rd PC	Mostly Noise

Table 4.13: Smile Analysis: Summarized Results for the Interpretation of the Unrotated PCs in both Metrics.

Metric		**Interpretation of the Rotated PCs**
Strike	1st PC	Shift in 30-10, 60-30, and 90-60
		Both positive and negative loadings in the other ranges
	2nd PC	Z-shape
Moneyness	1st PC	Shift in all the ranges and years
	2nd PC	Z-shape

Table 4.14: Smile Analysis: Summarized Results for the Interpretation of the Rotated PCs in both Metrics.

consistent with the results of Buraschi and Jackwerth [28], and Dumas, Fleming and Whaley [58].

4.7 Correlations between the Futures Price and the Principal Components

So far, we have found simple parameterizations of the dynamics of the implied volatility smiles. In this section, we investigate the correlations between the percentage changes in the futures price and the changes of the principal components in each expiry bucket. Knowledge of these is necessary in order to complete the specification of the process for the evolution of implied volatilities (see Hull [91]).

Table 4.15 presents the correlations between the proportional changes of the futures price $(\frac{F_{t+1}-F_t}{F_t})$ with the changes of each one of the first two rotated principal components, for both the strike and moneyness metric[35]. The correlations are measured by the Pearson correlation coefficient r_p. Pearson's coefficient is calculated by the following expression:

$$r_p = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2} \tag{4.14}$$

where $i = 1, 2, \ldots, n$, n is the number of observations, and $\overline{X}$, $\overline{Y}$ are the averages of the two variables X, Y respectively. The values of Pearson's coefficient range from -1 to $+1$, where the absolute value of the correlation coefficient indicates the strength of the relationship between the variables, with larger absolute values indicating stronger relationships. The sign of the coefficient indicates the direction of the relationship. To cope with the missing observations, r_p has been calculated by excluding cases pairwise, i.e. we exclude from the analysis cases with missing values for either, or both, of the pair of variables.

[35] Given that the rotated components deliver a clear interpretation, we use those in order to calculate the correlations. We calculate the rotated components from equation (4.12). Since SPSS uses standardized components $\mathbf{Z}^*$, we have to calculate first the unstandardized components $\mathbf{Z}$. This is done by using the relationship $\mathbf{Z}^* = \mathbf{Z}\mathbf{L}^{-\frac{1}{2}}$.

Since the correlations have been calculated from a specific sample, we test whether or not they are significantly different from zero. We use a two-tailed significance level; the null hypothesis is that the correlation coefficient is equal to zero, and the alternative is that the coefficient is different from zero. For the two-sided tests, acceptance of the alternative hypothesis occurs when $|r_p| \geq$critical value which corresponds to the significance level. One asterisk is displayed when the correlation is significant at 5% significance level, and two asterisks are displayed when the correlation is significant at 10% significance level[36].

The first rotated PC is positively correlated with the underlying in the strike metric; the only exceptions are in the range 150-90 in year 1994, and in the ranges 240-150, and 360-240 for the year 1992. On the other hand, the correlations are negative in the moneyness metric; the only exceptions occur in the range 360-240 for the years 1993, 1994, and 1995. The correlations for the rotated second PC in most of the cases, are positive in the strike and moneyness metric. The exceptions appear in the strike metric in the range 240-150 for the year 1995, and in the range 360-240 for the year 1992. We investigated the scatterplots for the anomalous correlations and found that although there were some outliers, removing them did not affect the sign of the correlation (though in some cases it became insignificantly different from zero).

The negative sign of correlation for the first rotated PC can be interpreted as a leverage effect (see Christie [33]), or in general it can be said that it is consistent with the well-known fact that implied volatilities are negatively correlated with the market index returns (see Franks and Schwartz [69], Rubinstein [126], Schmalensee and Trippi [130])[37]. The positive sign of correlation for the second

[36]Testing for the significance of the values of the Pearson coefficient, is based on the assumption that X and Y, are jointly normally distributed (see Clarke and Cooke [34]). We have also calculated the correlations by using the Spearman correlation coefficient. This is a nonparametric version of the Pearson coefficient, and captures non-linear association. However, the results for both coefficients were very similar, in terms of significance, size, and sign. Therefore, we report only the Pearson's results.

[37]There is a caveat, since this is a sign of the correlation of the asset's price with the first PC; not of the correlation coefficient between the asset and the volatility process. The correlation between the two processes is a function of the strike (moneyness variable) and of the time to

Range			1992	1993	1994	1995
30-10	Strike	$\Delta PC1$	0.04	-0.06	-0.18	-0.01
		$\Delta PC2$	0.38**	0.34**	0.48**	0.00
	Moneyness	$\Delta PC1$	-0.40**	-0.52**	-0.65**	-0.42**
		$\Delta PC2$	0.12	0.01	0.05	-0.06
60-30	Strike	$\Delta PC1$	0.23**	0.11	-0.10	0.19*
		$\Delta PC2$	0.31**	0.31**	0.36**	0.06
	Moneyness	$\Delta PC1$	-0.35**	-0.38**	-0.58**	-0.41**
		$\Delta PC2$	0.24**	0.22**	0.25**	0.11
90-60	Strike	$\Delta PC1$	-0.15	0.20*	0.10	0.29**
		$\Delta PC2$	0.32**	0.27**	0.39**	0.15
	Moneyness	$\Delta PC1$	-0.36**	-0.49**	-0.49**	-0.37**
		$\Delta PC2$	0.25*	0.17	0.33**	0.37**
150-90	Strike	$\Delta PC1$	0.30**	0.23**	-0.35**	0.28**
		$\Delta PC2$	0.36**	0.36*	0.24**	0.28**
	Moneyness	$\Delta PC1$	-0.28**	-0.31**	-0.19*	-0.32**
		$\Delta PC2$	0.33**	0.26**	0.37**	0.16
240-150	Strike	$\Delta PC1$	-0.31**	0.12	0.30**	0.34**
		$\Delta PC2$	0.15*	0.38**	0.38**	-0.17*
	Moneyness	$\Delta PC1$	0.08	0.04	-0.28**	0.08
		$\Delta PC2$	0.28**	0.36**	0.37**	0.38**
360-240	Strike	$\Delta PC1$	-0.38**	0.33**	0.33**	0.05
		$\Delta PC2$	0.04	0.40**	0.41**	0.27**
	Moneyness	$\Delta PC1$	-0.45**	0.33**	0.14	0.27*
		$\Delta PC2$	0.31**	0.35**	0.47**	0.52**

Table 4.15: Smile Analysis: Correlations between Percentage Changes of the Futures Price with Changes of the Rotated PCs on the Strike and Moneyness Metric.

rotated PC implies that when the asset price falls (rises), the twist in the volatility smile decreases (increases), i.e. the high strike implied volatilities increase less (more) than the low strikes ones.

The following proposition shows that the sign of the correlation changes as we change metrics, when some conditions are met. This confirms that the dynamics of implied volatilities depend on the metrics that we consider.

Proposition 1 *The correlation in the moneyness metric will be negative if*

$$\sigma'_{t+1} < -\frac{Cov_{strike}}{Var(\Delta F)}$$

and positive if the inequality is reversed. σ'_{t+1} *is the first derivative of the smile evaluated at a point* K*, and* Cov_{strike} *is the covariance in the strike metric.*

Proof. Since the correlation has the same sign as the covariance, the problem is reduced to looking at the covariance between $\Delta\sigma_t = \sigma_{t+1} - \sigma_t$ and ΔF under both metrics. The covariance in the strike metric is $Cov_{strike} = Cov(\Delta\sigma_t(K),\ \Delta F)$. Then, in the moneyness metric, for a given moneyness level we have $Cov_{mon} = Cov[\sigma_{t+1}(K + \Delta F) - \sigma_t(K),\ \Delta F]$. Expanding $\sigma_{t+1}(K + \Delta F)$ as a Taylor series of order one around a point K we get: $Cov_{mon} = Cov[\sigma_{t+1}(K) + \Delta F\sigma'_{t+1} - \sigma_t(K), \Delta F] = Cov_{strike} + \sigma'_{t+1}Var(\Delta F)$.

Therefore, whether the correlation sign is going to alter as we change metrics, depends on the slope of the skew, the point around which we do the expansion, and the variance of ΔF.

■

Finally, the magnitude of the correlations varies randomly over years and over expiries. This is similar to the findings of Clewlow and Xu [37]; It is not surprising because we expect the correlation to be a stochastic process, since the variance is a stochastic process, as well (see Engle and Mezrich [66]).

---

maturity and therefore its sign and value may, or may not be the same as, ρ_1 and ρ_2. However, the fact that the second rotated PC has both positive and negative loadings, implies that its impact to the *average* implied volatility is almost zero. Hence, the sign of the correlation for the first PC determines the sign of the overall correlation for the average implied volatility.

4.8 Conclusions

In this chapter we have investigated the dynamics of implied volatility smiles. We applied Principal Components Analysis (PCA) to the first differences of implied volatilities for six fixed ranges of days to expiry for the years 1992-1995. In particular, we answered three questions: what is the number of shocks that drive the smiles, what is their interpretation and what is their correlation with changes in the underlying asset. This is the first step in order to understand the dynamics of the more complex implied volatility surface.

We applied PCA under two metrics: the strike and the moneyness. This is because there are theoretical reasons which suggest that each metric may be relevant for our study. After considering three criteria (Velicer's criterion, communalities, interpretation), we found that two components could be reliably extracted. We constructed a Procrustes rotation, so as to obtain a clean interpretation of them. The first rotated component can be interpreted as a parallel shift, and the second as a Z-shape twist. These results were similar for both metrics. The two components explained between 56% and 85% of the variance in the strike metric, and between 65% to 92% in the moneyness metric.

Finally, we looked at the correlations between the changes of the futures price and each one of the first two rotated PCs in the strike and the moneyness metric. We found that the correlation for the first rotated PC is positive in the strike metric, and negative in the moneyness metric. The correlation for the second rotated PC is positive in both metrics.

The results from the smile analysis suggest that implied volatility smiles have non-trivial dynamics that can only be captured by a complex stochastic volatility model. This confirms the findings of other authors, e.g. Buraschi and Jackwerth [28] and Dumas, Fleming and Whaley [58], that deterministic volatility models are not capable of representing the dynamics of the volatility surface. Moreover, the smile analysis is a necessary step before examining the dynamics of the more complex volatility surface (Jackwerth and Rubinstein [95]). The results from the

analysis of separate smiles, can be used as a check of the robustness of the results that we will get from applying the PCA to the whole implied volatility surface. This will be done in the next chapter.

Chapter 5

The Dynamics of Implied Volatility Surfaces

5.1 Introduction

In this Chapter, we extend the analysis of the previous chapter in order to model the entire volatility surface rather than a single volatility smile (skew) at a time[1]. By doing so, we incorporate the effect of the term-structure on the dynamics of implied volatilities. It may be the case that short and long-term implieds are driven by different shocks (see Xu and Taylor [149]). This could result in a different number (and possibly shape) of shocks driving the whole implied volatility surface from those driving individual smiles.

Just as the previous Chapter, the investigation of the dynamics of the implied volatility surface is also motivated by the already described literature on smile-consistent stochastic volatility no-arbitrage models (see Chapter 2 and Section 4.1). However, the results from this Chapter allow us to hedge portfolios of options with different expiries which are subject to the same shock (rather than different shocks, as it was the case in Chapter 4).

[1]This also parallels work on implementing the HJM term structure model where Litterman and Scheinkman [107] and others, have applied Principal Components Analysis to analyze innovations in the yield curve.

We are able to identify two factors which explain about 60% of the variance. The first factor is interpreted as an essentially parallel shift and the second as a Z-shaped twist. Our results are remarkably consistent across years. To implement a "smile-consistent" no-arbitrage stochastic volatility model for the pricing and hedging of futures options we need three factors. One is required for the underlying asset, and the two more for the implied volatility. A related paper by Kamal and Derman [102] has also investigated the same issue. They have used a different data set and a somewhat different methodology; their results are markedly different from ours. We comment further on the differences between the two studies in the concluding section of the chapter.

The Chapter is structured in Six Sections. In the second, and third section we apply Principal Components Analysis to the strike and moneyness metric. We decide on how many components we should retain. Then, we interpret them by rotating them, so as to get a clean and consistent across years interpretation. In the fourth section, we compare the results from the two metrics. In the fifth section, we calculate the correlations between the changes of the principal components and the underlying asset price, under the two metrics. In the sixth section, we present concluding remarks and implications.

5.2 PCA on the Strike Metric

5.2.1 Determining the Expiry Buckets

In order to look into the dynamics of implied volatilities as a surface for a given year (surface analysis), we first choose the expiry buckets, and then the strike (moneyness) levels for each bucket. Finally, all the variables are brought together to perform the PCA. However, the original expiry buckets, used in the analysis of the dynamics of individual smiles, gave rise to a problem with this pooled analysis. Their spacing was too fine, so that we had insufficient dates with data points in every expiry range. Therefore, we redefine the new expiry buckets to

Year	Number of Variables	Number of Observations	KMO
92	26	182	0.89402
93	26	198	0.87585
94	32	187	0.89416
95	20	103	0.77463

Table 5.1: Surface Analysis on the Strike Metric: Number of Variables, Number of Observations and the KMO measure. The numbers have been calculated across both strikes and expiries.

Year	Variable	n	Skewness	Ex. Kurtosis	Test Value	P-value
92	4209.1_1	247	0.336	5.323	296.255	0.0000
93	4409.1_1	251	-0.898	7.067	556.0498	0.0000
94	4409.1_1	245	0.269	6.419	423.5744	0.0000
95	5101.9_1	235	0.458	3.925	159.0625	0.0000

Table 5.2: Surface Analysis on the Strike Metric: Bera-Jarque Test for Univariate Normality. XA.B-1 denotes the differenced once implied volatilities corresponding to strike level X, in the range B-A; B,A are the first digits of the three ranges that we examine.

90-10, 180-90 and 270-180 days to expiry. Table 5.1 shows that these intervals give us a satisfactory number of observations (not less than 100), they permit us to measure smiles across a wide range (not less than 20 variables), and they have a high KMO coefficient, once they are pooled together.

5.2.2 Number of Retained Principal Components and a First Interpretation

Testing for Normality

Next, we check the null hypothesis of univariate normality by using the BJ test. Table 5.2 shows that the null is rejected. Consequently, we can not use the tests which are based on the assumption of multivariate normality, in order to decide on the number of components to retain.

Year	f_0	f_1	f_2	f_3	r*	l	1st PC	2nd PC	3rd PC
92	0.18141	0.18070	0.18069	0.18080	2	6	42.5%	12.9%	6.8%
93	0.14154	0.14097	0.14091	0.14096	2	5	37.5%	12.7%	7.2%
94	0.05127	0.05112	0.05116	0.05117	1	6	40.8%	9.8%	7.2%
95	0.12350	0.12288	0.12255	0.12265	2	5	29.9%	23.7%	9.3%
average							37.6%	14.7%	7.6%

Table 5.3: Surface Analysis on the Strike Metric: r* = number of components retained under Velicer's criterion (minimum f0,...,f3), l = number of components retained under rule of thumb with percentage of variance explained by components 1-3.

Velicer's Criterion

We decide on the number of PCs to be retained, by applying the variety of criteria that we used in the smile analysis (Velicer's criterion, communalities and interpretation). In Table 5.3 we show the results from applying Velicer's criterion and the mean eigenvalue rule of thumb. We also show the percentage of the variance explained by each one of the first three PCs. We can see that Velicer's criterion keeps one PC in year 1994, and two PCs in the other years. The reduction in the dimensionality of the variables is legitimate, since $f_0 > f_1$. The mean eigenvalue rule of thumb retains too many PCs. This shows why we should not rely on such ad-hoc rules (see Jackson [93]). The average, across the years, explained by the two components variance is 52.5%.

Communalities

Table D.1, (Appendix D) shows the communalities explained by the first three (rather than two) PCs in year 1994. This is because there are certain cases that two retained PCs seem to perform poorly in terms of the explained communalities, e.g. in the range 90-10 for the strikes 435, 440, 465, and 470; therefore, one could argue that the inclusion of the third PC should be considered.

However, it is not clear that adding the third PC in these cases, improves substantially the explained communalities. In some cases the explained communality is only marginally improved, (e.g. for the variable 465 in the range 90-10),

and in some others there is a significant increase (e.g. for the variable 440 in 90-10). In general, retaining the third PC does not seem to be necessary for the vast majority of the variables. This is true for the other years, as well.

Interpretation of the Components

Figures 5.1-5.3 show the correlation loadings of the first PC. This has a Z-shape in all the ranges, apart from 90-10 in the years 1994, and 1995.

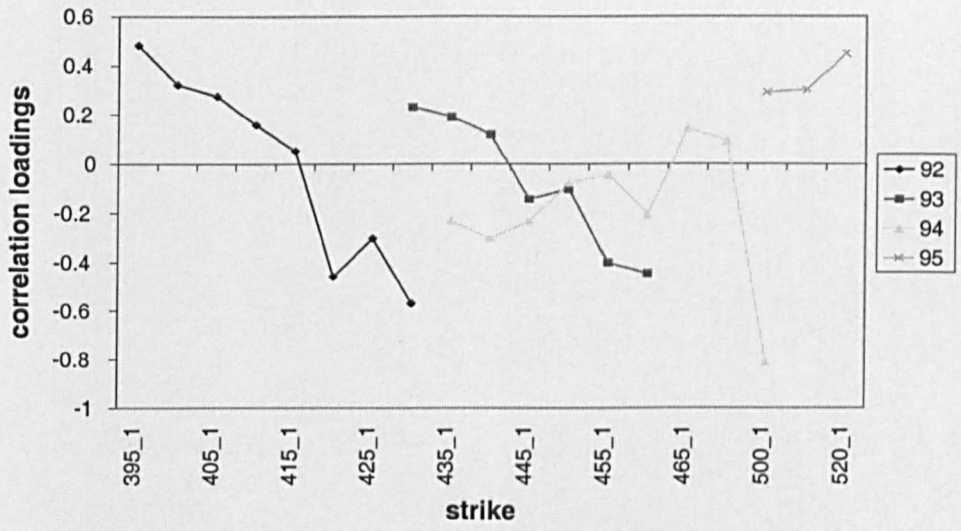


Figure 5.1: Surface Analysis on the Strike Metric: Interpretation of the First PC for 90-10.

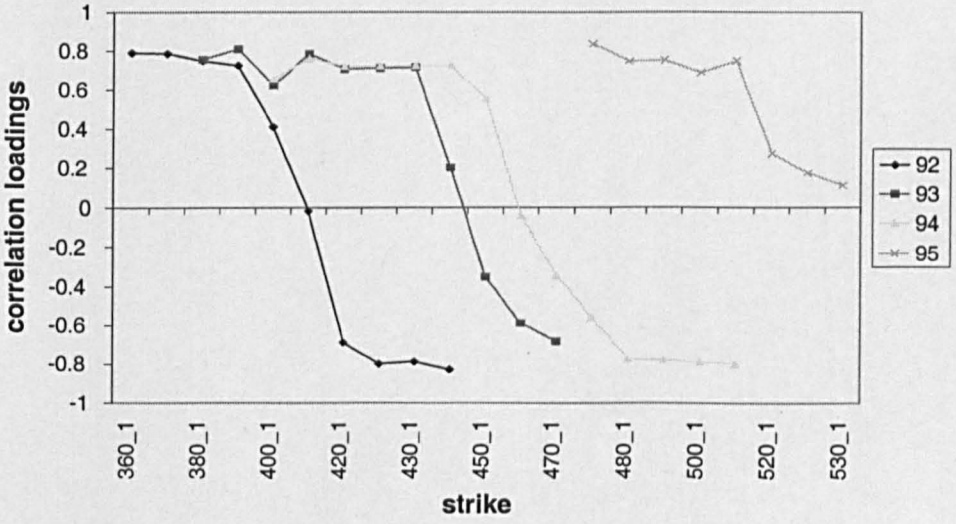


Figure 5.2: Surface Analysis on the Strike Metric: Interpretation of the First PC for 180-90.

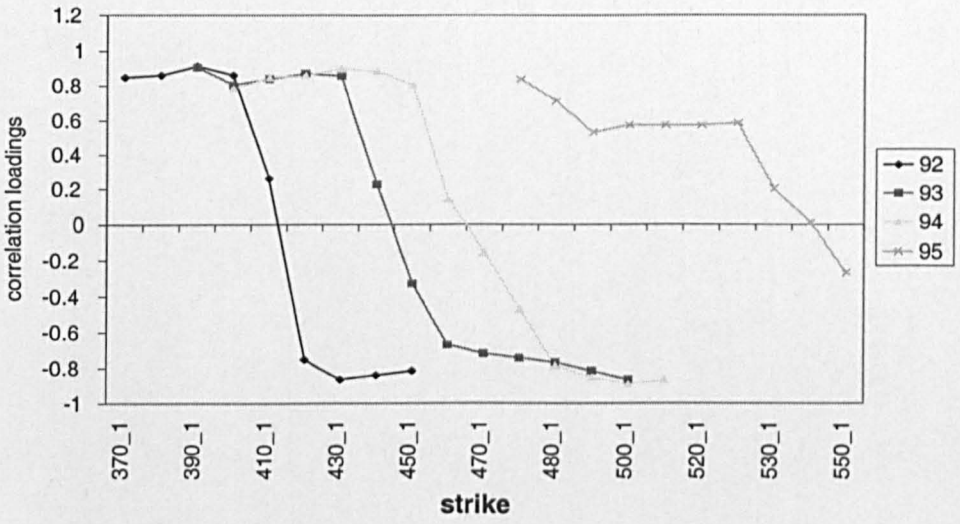


Figure 5.3: Surface Analysis on the Strike Metric: Interpretation of the First PC for 270-180.

Figures 5.4-5.6 show the interpretation of the second PC. In the range 90-10, it has a shift interpretation. In the longer expiry ranges, it has a triangular shape for the years 1992, 1993 and 1994, and a Z-shape in the year 1995. It is worth noting that this triangular shape was also apparent in the smile analysis in the strike metric for the expiries with more than 90 days. This similarity should not surprise us; we can think of the pooled expiries as some kind of average of the individual ones.

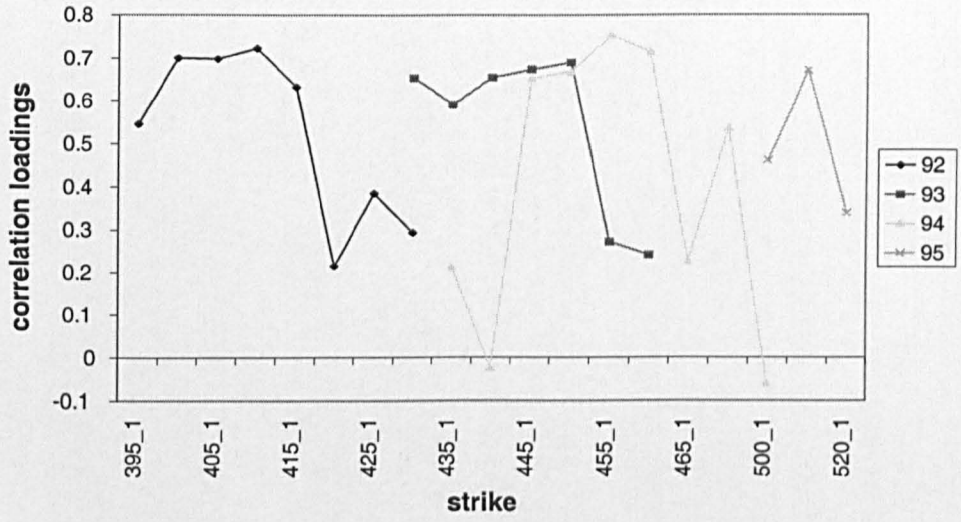


Figure 5.4: Surface Analysis on the Strike Metric: Interpretation of the Second PC for 90-10.

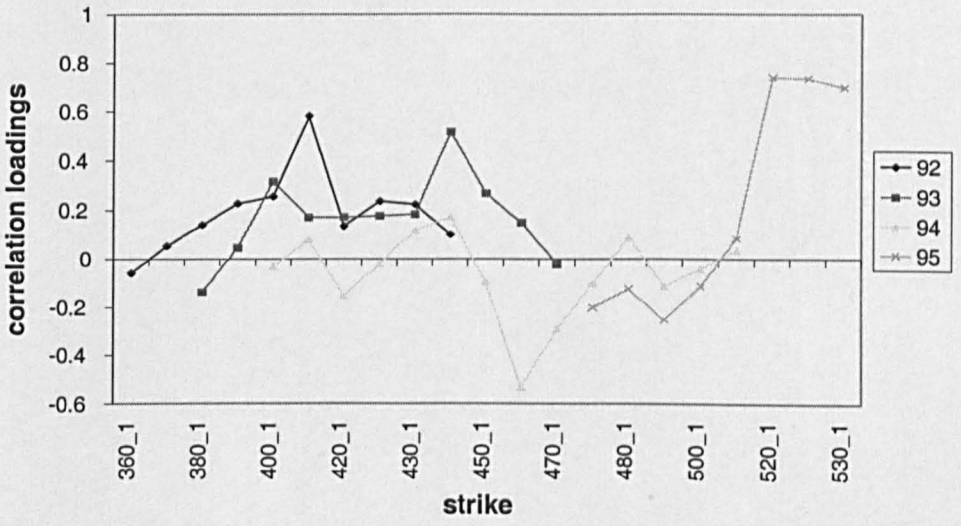


Figure 5.5: Surface Analysis on the Strike Metric: Interpretation of the Second PC for 180-90.

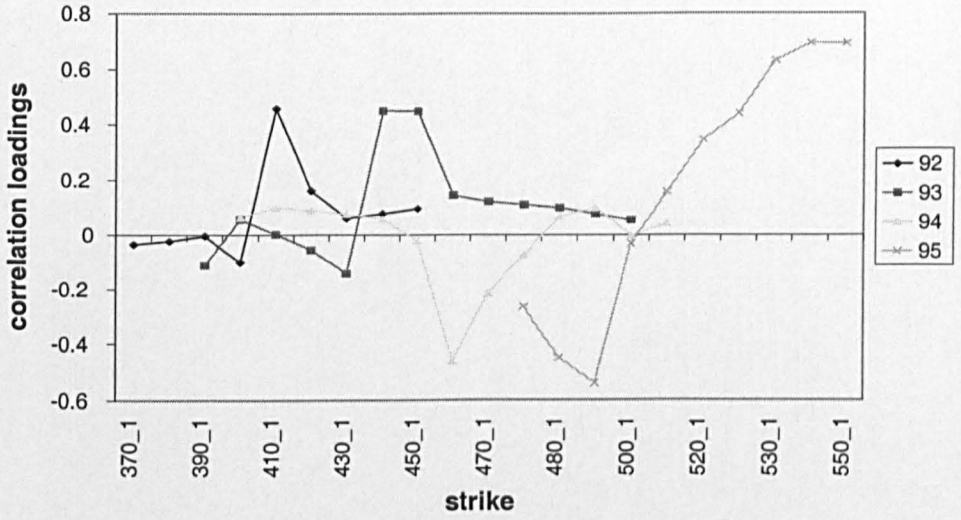


Figure 5.6: Surface Analysis on the Strike Metric: Interpretation of the Second PC for 270-180.

Figures 5.7-5.9 show that the third PC is just noise. Hence, all the criteria that we have applied suggest that two PCs are adequate in explaining the implied volatility surface dynamics, in the strike metric.

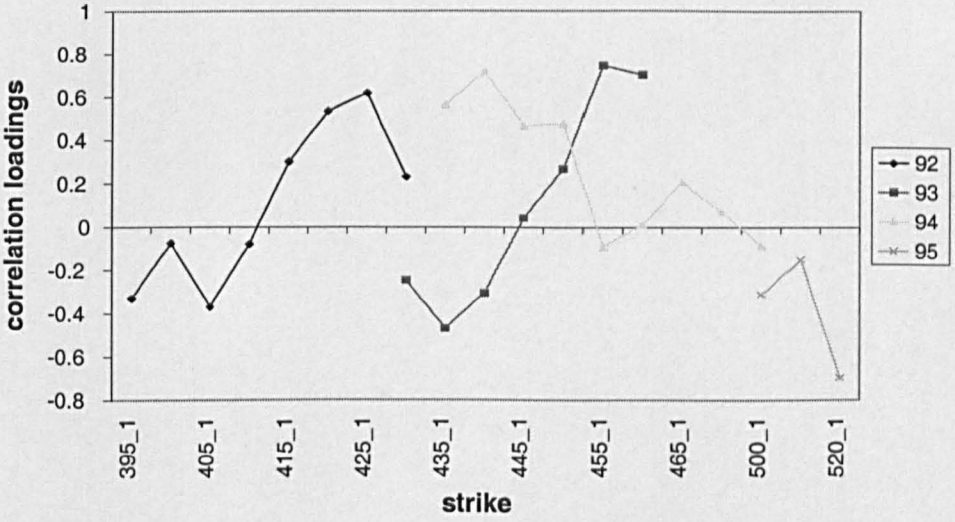


Figure 5.7: Surface Analysis on the Strike Metric: Interpretation of the Third PC for 90-10.

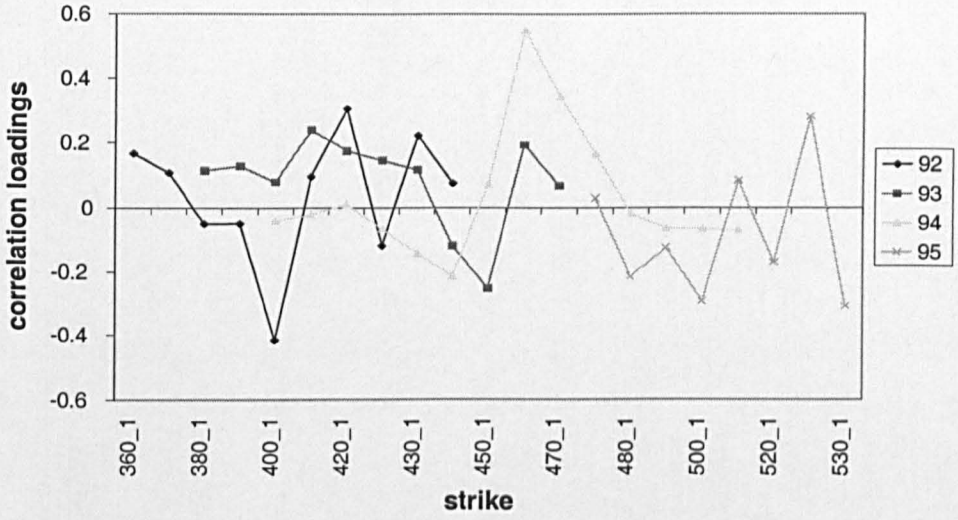


Figure 5.8: Surface Analysis on the Strike Metric: Interpretation of the Third PC for 180-90.

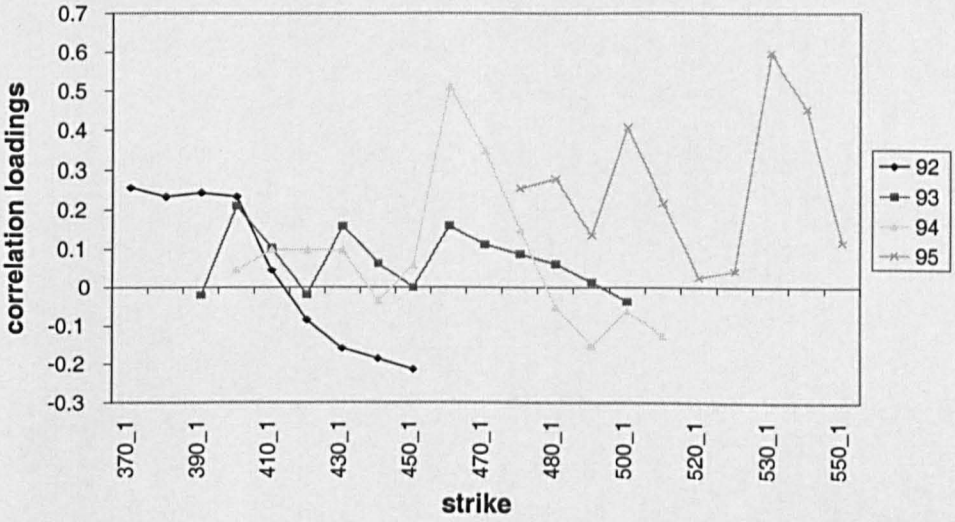


Figure 5.9: Surface Analysis on the Strike Metric: Interpretation of the Third PC for 270-180.

Next, we compare (roughly) the amount of variance explained by the retained PCs in each year, in the smile and the surface analysis[2]. In order to do this, we compute from Table 4.4 for each year, the average, across the separate expiries,

[2]This is not an exact comparison, since the ranges for the separate expiries in the Smile analysis, are not the same as the ranges that we use for the Surface analysis.

Year	1st PC	2nd PC	Cumulative
92	52.8%	14.61%	67.41%
93	50.6%	16.51%	67.11%
94	47.96%	16.4%	64.36%
95	51.46%	17.8%	69.26%
average	50.7%	16.33%	67.03%

Table 5.4: Constructed "Pooled" Variance explained by the retained PCs.

explained variance. In that way, we construct a "pooled" variance for each year (we will call this the constructed "pooled" variance in contrast to the actual "pooled" variance that we find from applying the PCA to the implied volatility surface).

The constructed "pooled" variance is shown in Table 5.4. Comparing Tables 5.3, and 5.4 we can see that for the first PC the constructed "pooled" variance is bigger than the actual. We would not expect to be the same, since the constructed "pooled" variance can not be measured just by a simple average; a weighted average, whose weights are unknown, has to be used, instead. For the second PC, the constructed and actual variances are quite similar. Hence, it seems that a simple average is a satisfactory measure for the constructed "pooled" variance for the second PC.

5.2.3 Interpretation of the Rotated Components

Given that the shapes of the retained two PCs are not consistent across years and expiries, we apply our "Procrustes" type rotation. Figures 5.10-5.12 show the shape of the first rotated PC across the individual expiries[3]. We can see that in the range 90-10, it has a shift interpretation in the years 1992, 1993 and 1995. In the year 1994 it has a shift interpretation for all, but the extreme strikes. In the other two ranges, the first PC has a triangular shape in every year (it has

[3]Alternatively, rather than showing the cross-sections of loadings for each expiry range, we could have plotted the estimated PCs across both strike levels and expiries for each year. The two ways of plotting the results reveal the same information. We prefer the cross-sectional way because we can not plot a proper surface of loadings with only three expiry ranges.

been flipped over for the year 1994)[4]. In general, the shape of the first rotated PC is very similar to that of the second unrotated PC. The size of its effect is bigger for the shorter expiries. Figures 5.13-5.15 show the shape of the second rotated PC. In the range 90-10, has a Z-shape in 1992, and it is a shift in the other years. In the other two ranges it has a Z-shape. The magnitude of the second PC is greatest for the longer expiries.

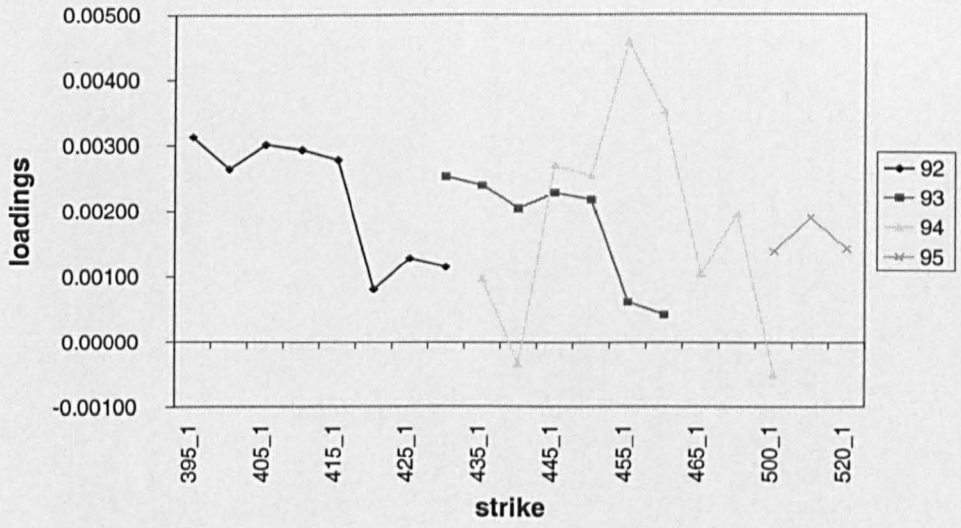
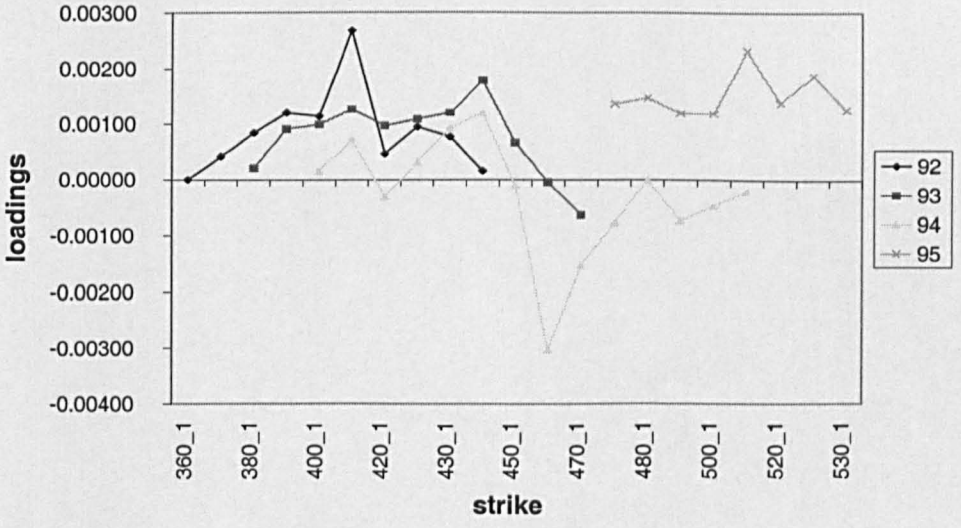


Figure 5.10: Surface Analysis on the Strike Metric: Interpretation of the First Rotated PC for 90-10.



[4]Notice that in contrast to the smile analysis, we can not reverse the sign of the first PC in a separate expiry bucket.

Figure 5.11: Surface Analysis on the Strike Metric: Interpretation of the First Rotated PC for 180-90.

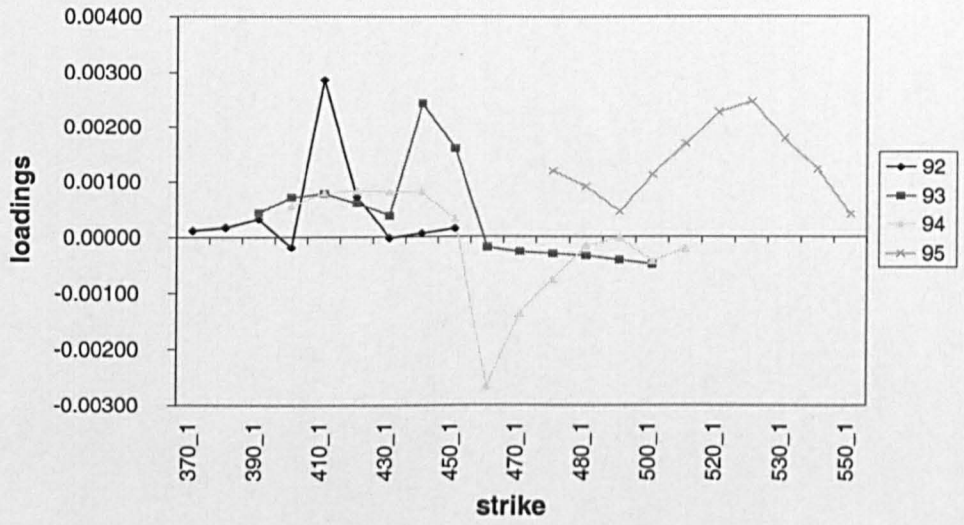


Figure 5.12: Surface Analysis on the Strike Metric: Interpretation of the First Rotated PC for 270-180.

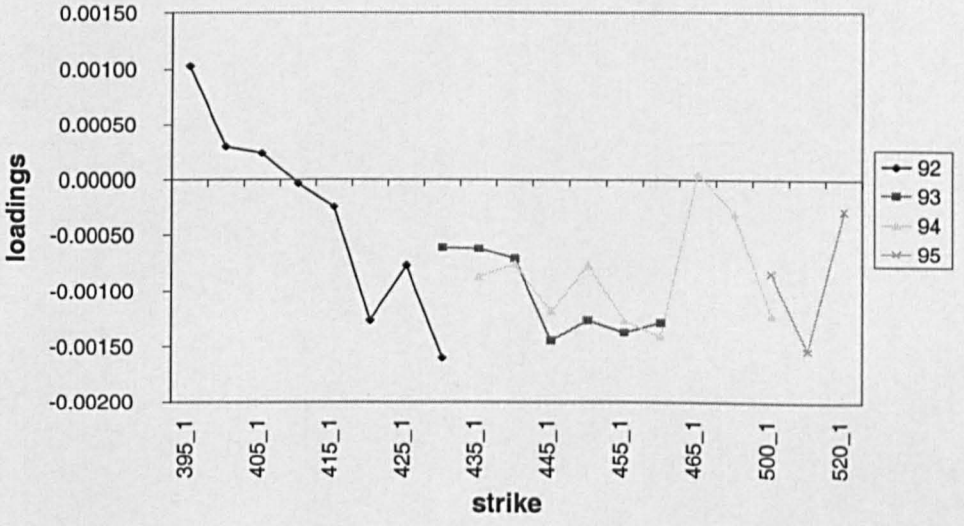


Figure 5.13: Surface Analysis on the Strike Metric: Interpretation of the Second Rotated PC for 90-10.

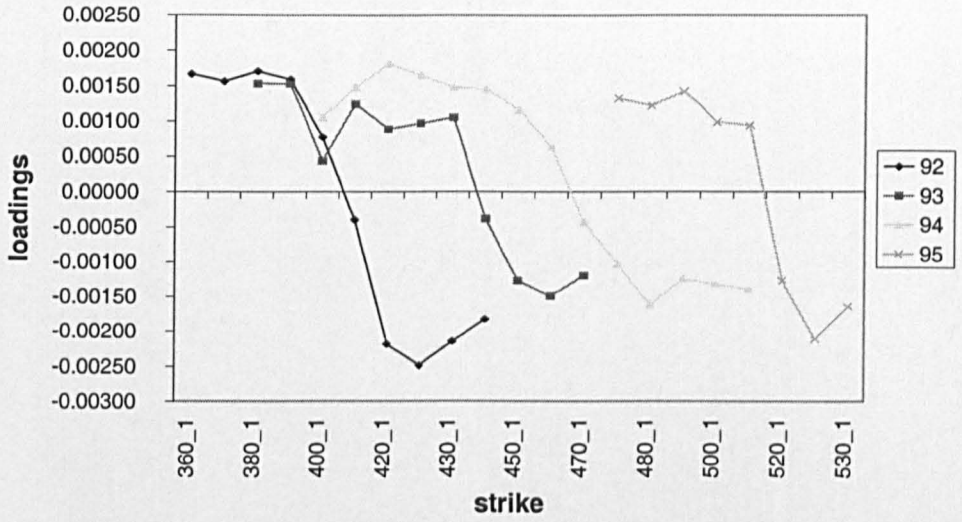


Figure 5.14: Surface Analysis on the Strike Metric: Interpretation of the Second Rotated PC for 180-90.

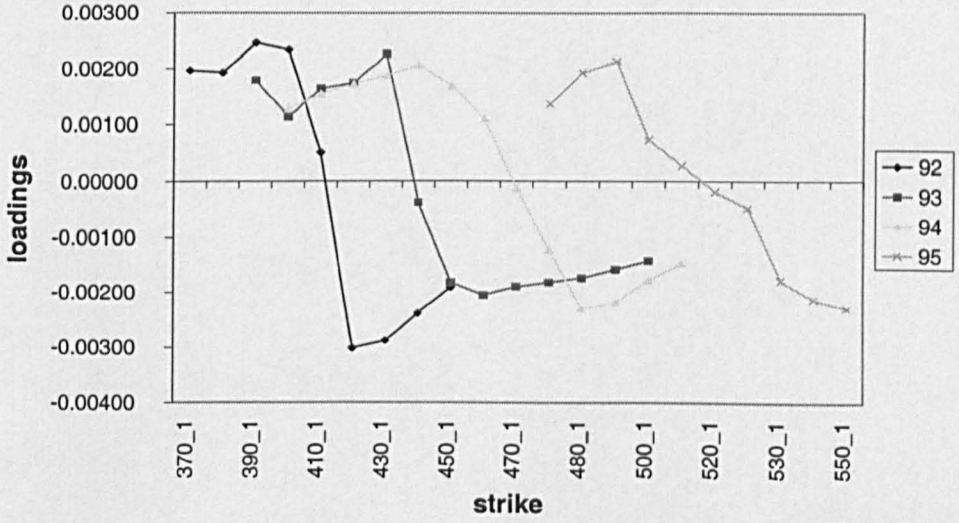


Figure 5.15: Surface Analysis on the Strike Metric: Interpretation of the Second Rotated PC for 270-180.

In Table 5.5 we show the percentage of the variance explained by each one of the two rotated PCs, and the cumulative explained variance (as well as the percentage for the original first component). We can see that the first rotated PC explains less amount of variance than the second rotated one.

Year	Unrot. 1st PC	1st PC	2nd PC	Cumulative
92	52.8%	13.4%	42%	55.4%
93	50.6%	17.3%	33%	50.3%
94	47.96%	11.5%	39%	50.5%
95	51.46%	28.4%	25.2%	53.6%
Average	50.7%	17.6%	34.8%	52.5%

Table 5.5: Surface Analysis on the Strike Metric: Percentage of Variance Explained by the Unrotated first PC and by the Rotated PCs.

Year		90-10	180-90	270-180
92	Minimum	1.18%	2.377%	2.377%
	Chosen	1.2%	2.4%	2.4%
93	Minimum	1.147%	2.295%	2.295%
	Chosen	1.2%	2.3%	2.3%
94	Minimum	1.073%	2.146%	2.146%
	Chosen	1.2%	2.2%	2.2%
95	Minimum	1.083%	2.16%	2.16%
	Chosen	1.2%	2.2%	2.2%

Table 5.6: Surface Analysis on the Moneyness Metric: Minimum and Chosen Step-Size for the Moneyness Metric.

5.3 PCA on the Moneyness Metric

5.3.1 Preliminary Descriptives

The moneyness metric in the surface analysis, is constructed just as in the smile analysis. The spacing between the moneyness levels has to be set so as to avoid inducing any spurious correlation between the interpolated implied volatilities (minimum step size). The final choice for the step-size (chosen step-size) was made after considering the number of variables on which we would perform the PCA (see Section 4.4.1). Table 5.6 shows the minimum and chosen step-size between the moneyness levels. The range of moneyness varies from -18% up to 6%, depending on the year that we look at.

Table 5.7 shows the number of variables, the number of observations, and the KMO for the surface analysis on the moneyness metric. These make us to expect

Year	Number of Variables	Number of Observations	KMO
92	24	125	0.85493
93	23	108	0.85693
94	25	129	0.86127
95	25	105	0.83056

Table 5.7: Surface Analysis on the Moneyness Metric: Number of Variables and Observations and the KMO measure. The numbers have been calculated across both strikes and expiries.

Year	Variable	n	Skewness	Ex. Kurtosis	Test Value	P-value
92	N4.8_1_1	207	0.237	3.247	92.8713	0.0000
93	P2.4_9_1	243	0.261	4.178	176.1297	0.0000
94	P2.2_1_1	237	-0.059	5.161	263.1672	0.0000
95	N15.4_2_1	195	0.523	3.424	104.1454	0.0000

Table 5.8: Surface Analysis on the Moneyness Metric: Bera-Jarque Test for Univariate Normality. P (N) X-A-1 denotes the differenced once implied volatilities corresponding to plus (minus) moneyness level X for the expiry range with upper limit A.

reliable results from the PCA.

5.3.2 Number of Retained Principal Components and a First Interpretation

Testing for Normality

Just as we did in the strike metric, we apply once more the BJ test, so that to see whether or not multivariate normality holds. Table 5.8 shows that the null hypothesis of univariate normality is rejected for all the years.

Velicer's Criterion

In Table 5.9 we show the results from applying Velicer's criterion and the mean eigenvalue rule of thumb. We also show the percentage of the variance explained by each one of the first three PCs. We can see that in every year, the criterion keeps two PCs, while the reduction in the dimensionality of the variables is legitimate, since $f_0 > f_1$. The mean eigenvalue rule of thumb retains again too

Year	f_0	f_1	f_2	f_3	r*	l	1st PC	2nd PC	3rd PC
92	0.18474	0.18384	0.18362	0.1838	2	4	38.6%	22.7%	10.7%
93	0.17218	0.17141	0.17109	0.17132	2	4	34.5%	26.7%	10.2%
94	0.1894	0.18842	0.1884	0.18851	2	6	40.6%	19.2%	10.2%
95	0.17256	0.17127	0.17116	0.17136	2	4	39.2%	18.3%	9.4%
average							38.2%	21.7%	10.1%

Table 5.9: Surface Analysis on the Moneyness Metric: r* = the number of components retained under Velicer's criterion (minimum of f0,...,f3), l = number of components retained under rule of thumb with percentage of variance explained by components 1-3.

many PCs. The two PCs explain on average, across the years, 60% of the total variance. Compared to the strike metric (see Table 5.3), the cumulative variance is bigger in the moneyness metric with the difference ranging from 4% up to 10%. This difference is stemming from the higher amount of variance explained by the second PC in the moneyness metric.

Communalities

Table E.1 (Appendix E) shows that the explained by the first two PCs communalities in year 1994, are satisfactory. The exceptions are for the at-the-money variables in the ranges 180-90 and 270-180. In these ranges, the explained by the two PCs communalities were low (only 4.71% and 1.23%, respectively). However, even if we add the third PC, the explained communalities would have risen only to 17.82% and 4.02%, respectively. There is a similar pattern for the other years, as well. Therefore, the communalities criterion suggests that we should keep two PCs in the surface analysis in the moneyness metric.

Interpretation of the Components

Figures 5.16-5.18, and 5.19-5.21 show the interpretation of the first and second PCs in the moneyness metric. We can see that the PCs do not have a consistent shape across years and expiries. The first PC has a Z-shape in the range 90-10 for the years 1992 and 1993. However, in the same range, it has a shift interpretation for the years 1994 and 1995. In the ranges 180-90 and 270-180, it has a Z-shape

in 1992 and 1993, but it has been flipped over in 1994 and 1995.

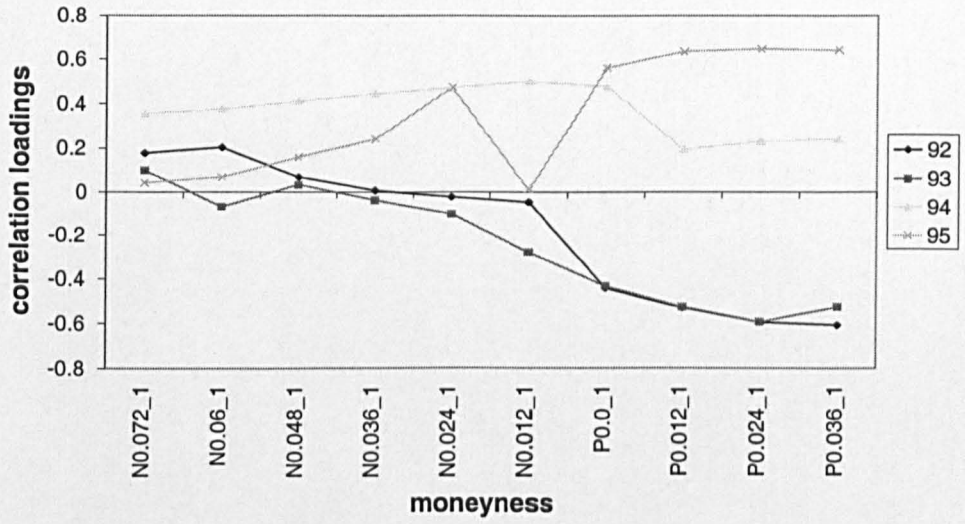


Figure 5.16: Surface Analysis on the Moneyness Metric: Interpretation of the First PC for 90-10.

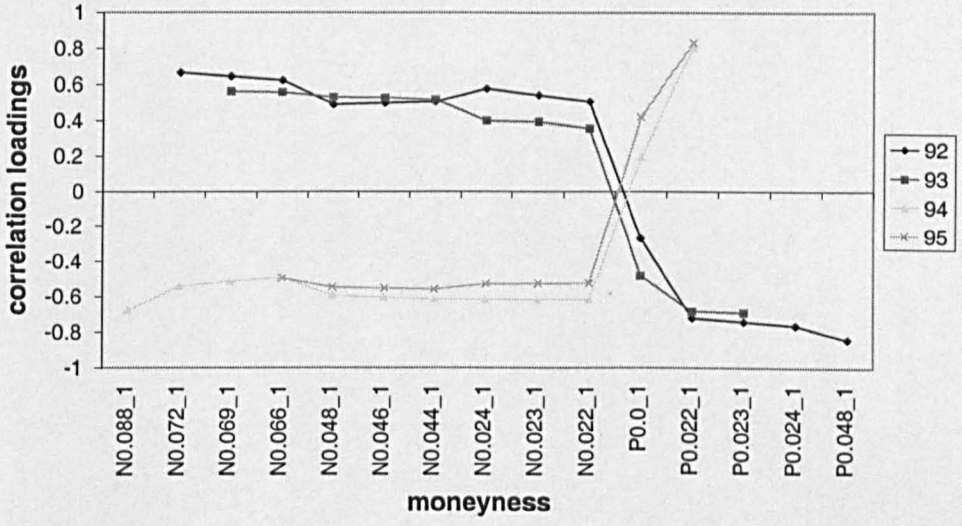


Figure 5.17: Surface Analysis on the Moneyness Metric: Interpretation of the First PC for 180-90.

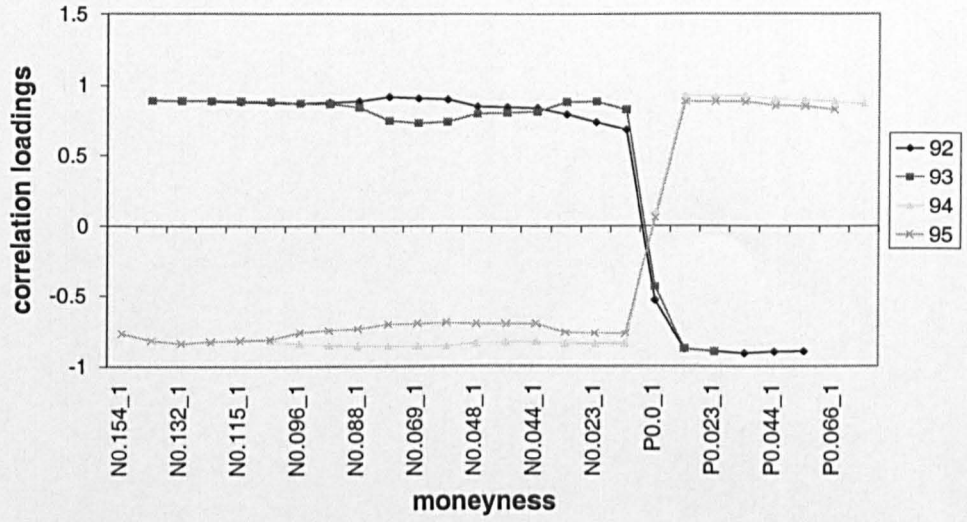


Figure 5.18: Surface Analysis on the Moneyness Metric: Interpretation of the First PC for 270-180.

The second PC, in general, has a shift interpretation across years and all the ranges. The exceptions occur in the range 90-10 for the year 1995, and in the range 270-180 for 1994.

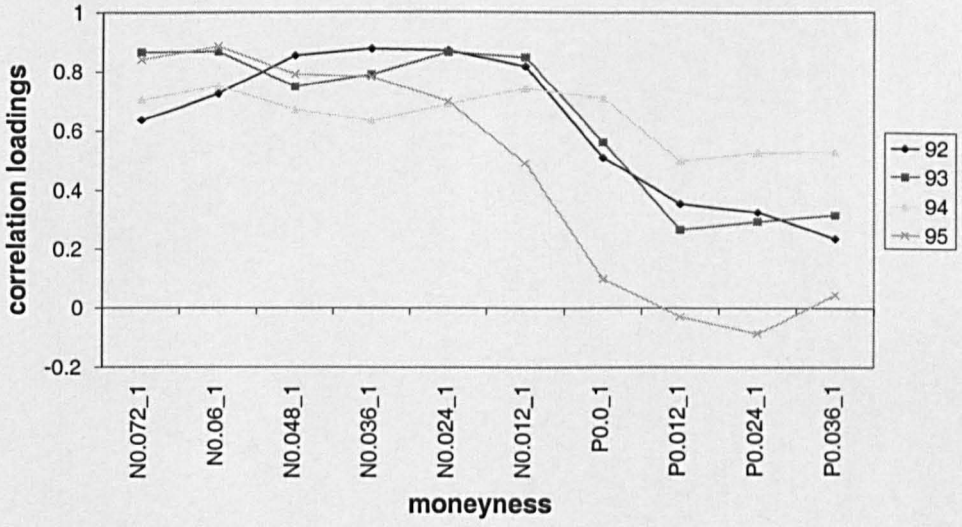


Figure 5.19: Surface Analysis on the Moneyness Metric: Interpretation of the Second PC for 90-10.

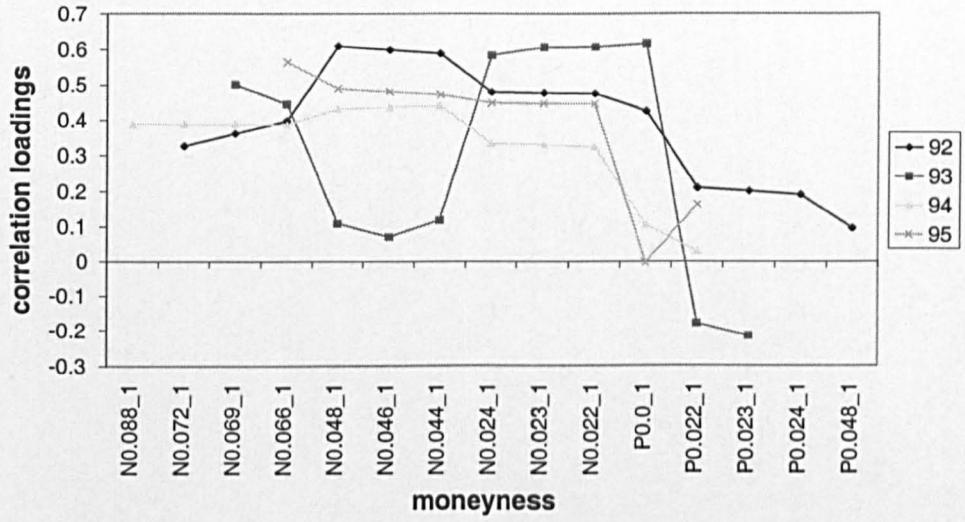


Figure 5.20: Surface Analysis on the Moneyness Metric: Interpretation of the Second PC for 180-90.

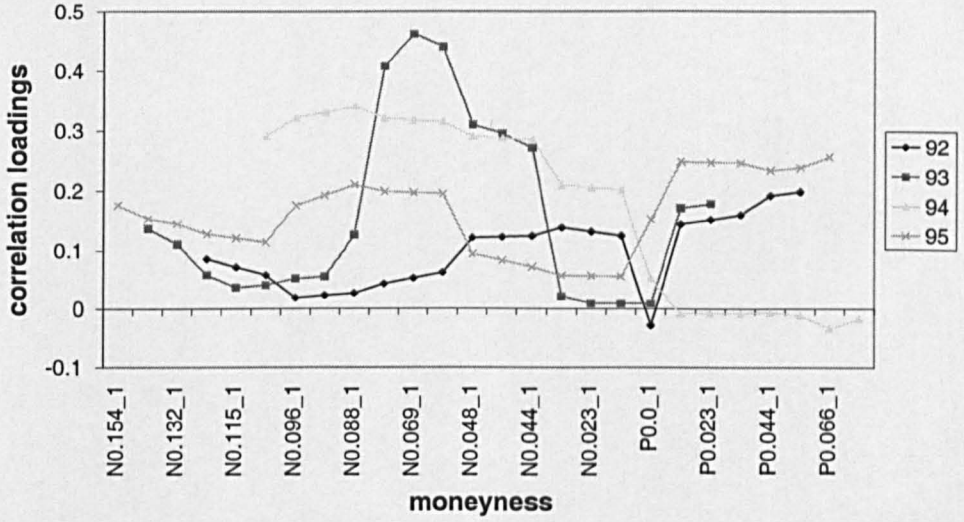


Figure 5.21: Surface Analysis on the Moneyness Metric: Interpretation of the Second PC for 270-180.

Figures 5.22-5.24 reveal that the shape of the third PC across the years and expiries is noise. Hence, all the criteria that we applied, suggest that two PCs are adequate in explaining the implied volatility surface dynamics in the moneyness metric, just as in the strike metric.

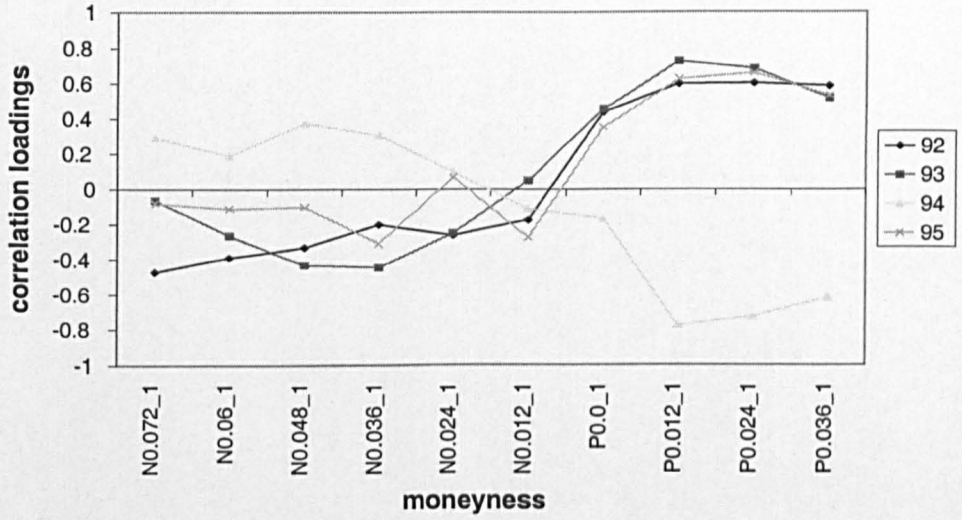


Figure 5.22: Surface Analysis on the Moneyness Metric: Interpretation of the Third PC for 90-10.

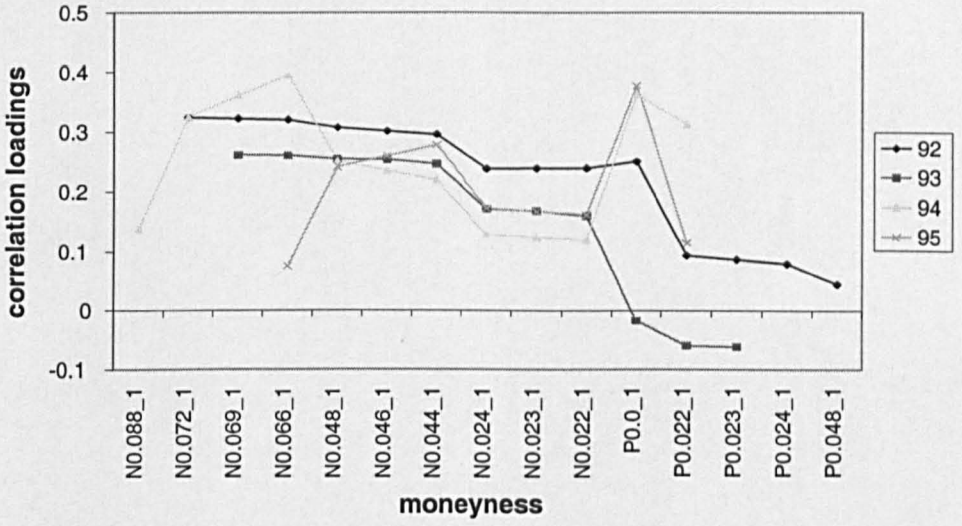


Figure 5.23: Surface Analysis on the Moneyness Metric: Interpretation of the Third PC for 180-90.

Year	1st PC	2nd PC	Cumulative
92	60.05%	18.56%	78.61%
93	60.18%	18.6%	78.78%
94	64.36%	16.93%	81.29%
95	58.3%	16.63%	74.93%
average	60.72%	17.68%	78.4%

Table 5.10: Constructed "Pooled" variance explained by the retained PCs.

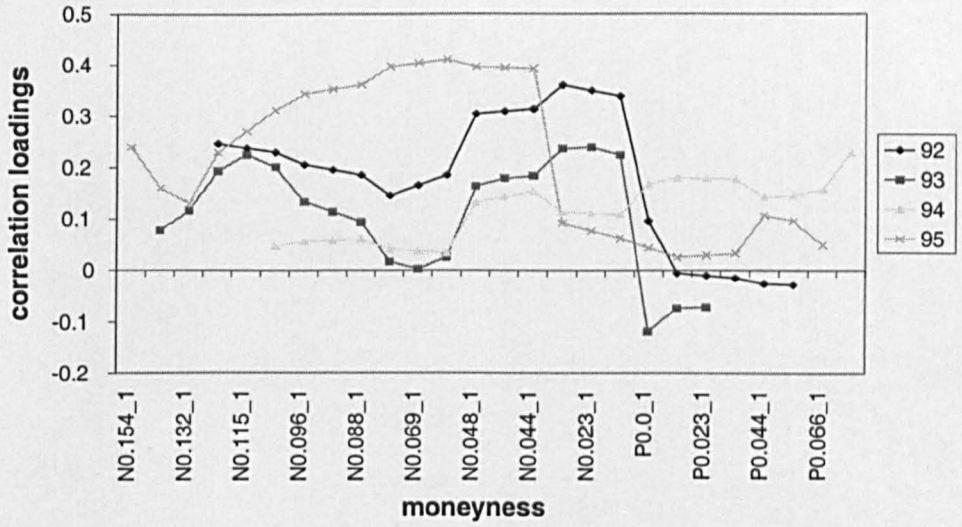


Figure 5.24: Surface Analysis on the Moneyness Metric: Interpretation of the Third PC for 270-180.

Table 5.10 shows the constructed "pooled" variance in the moneyness metric. We can see that, as it was the case in the strike metric, the actual "pooled" variance is less than the constructed "pooled" variance for the first PC (see Table 5.9). On the other hand, the actual and constructed "pooled" variances are similar for the second PC.

5.3.3 Interpretation of the Rotated Components

Given that the shape of the two retained PCs does not have a consistent interpretation, we will apply our rotation method. Figures 5.25-5.27 show the shape of the first rotated PC. The first PC moves the implied volatility surface consistently across ranges and across years, and its shape 'approaches' the shift interpreta-

tion. The figures also show that the effect of the shift component attenuates with expiry.

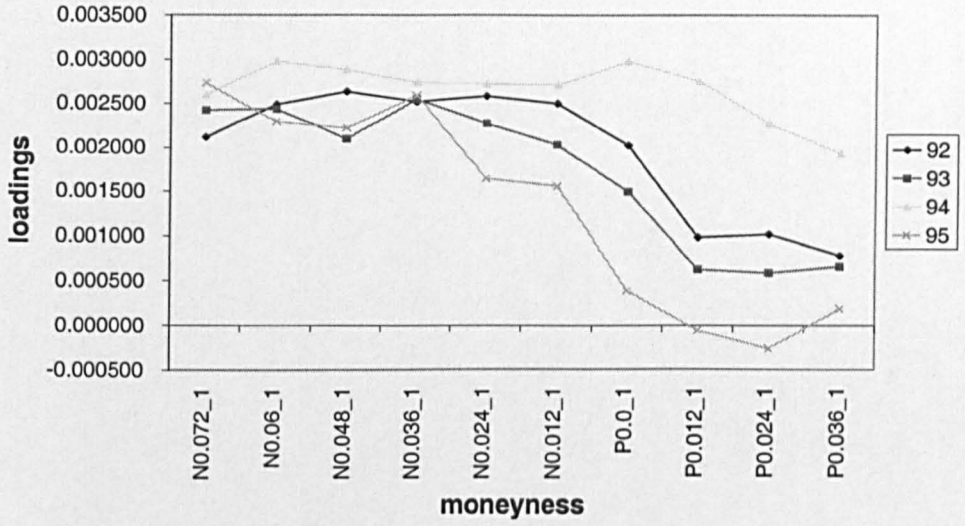


Figure 5.25: Surface Analysis on the Moneyness Metric: Interpretation of the First Rotated PC for 90-10.

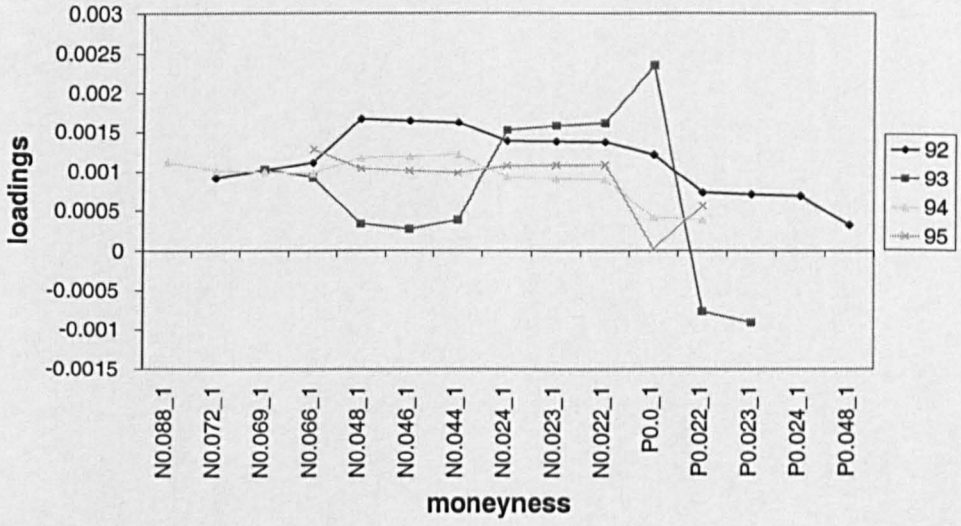


Figure 5.26: Surface Analysis on the Moneyness Metric: Interpretation of the First Rotated PC for 180-90.

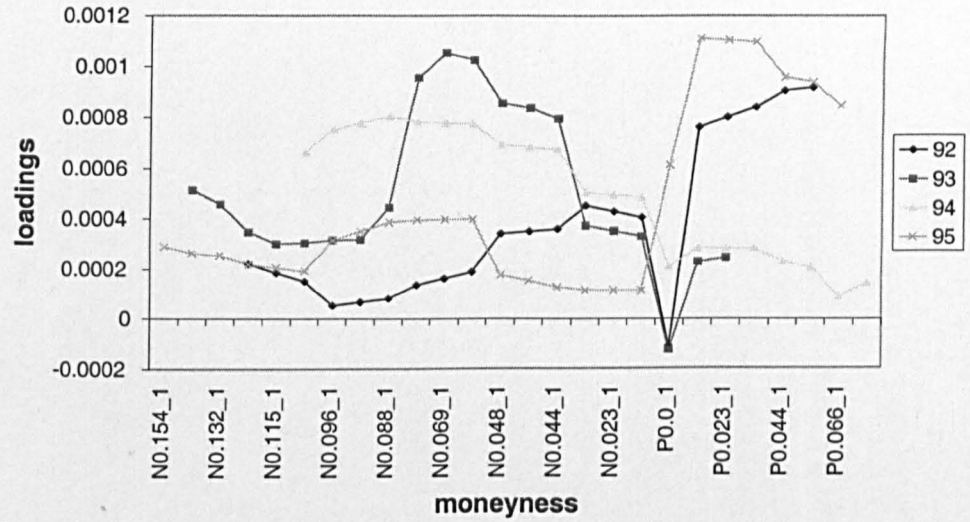


Figure 5.27: Surface Analysis on the Moneyness Metric: Interpretation of the First Rotated PC for 270-180.

Figures 5.28-5.30 reveal the shape of the rotated second PC. The second PC, in general, has a Z-shape across ranges and years (even though it has a shift interpretation in the range 90-10 in the years 1993, 1994 and 1995). There is not much sign of attenuation with expiry.

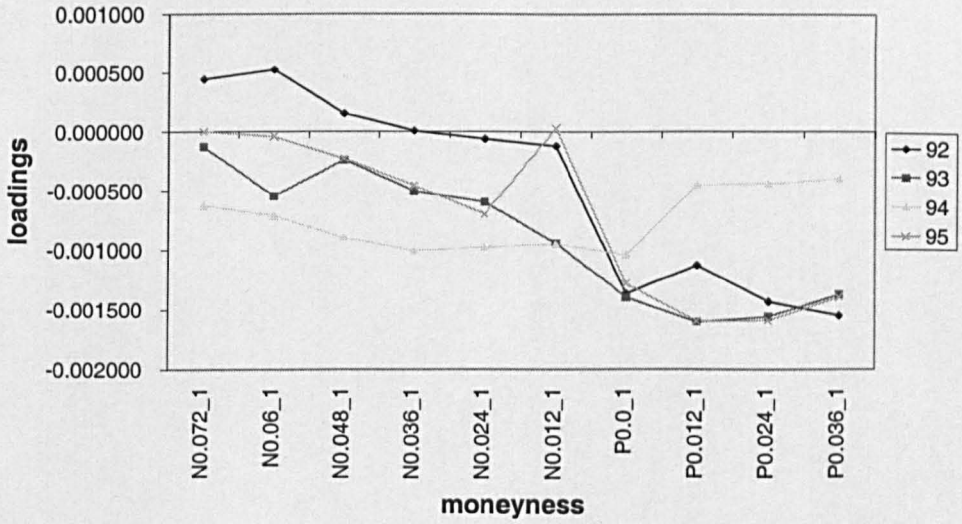


Figure 5.28: Surface Analysis on the Moneyness Metric: Interpretation of the Second Rotated PC for 90-10.

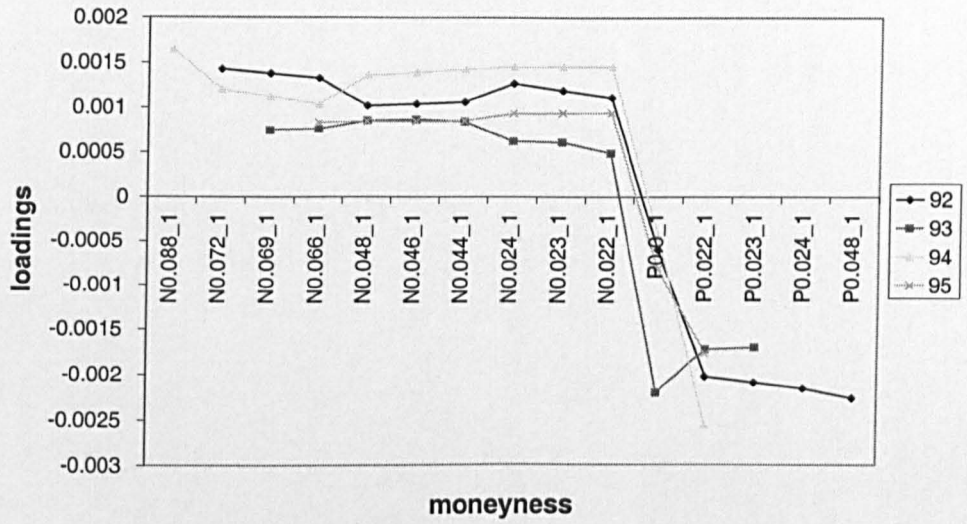


Figure 5.29: Surface Analysis on the Moneyness Metric: Interpretation of the Second Rotated PC for 180-90.

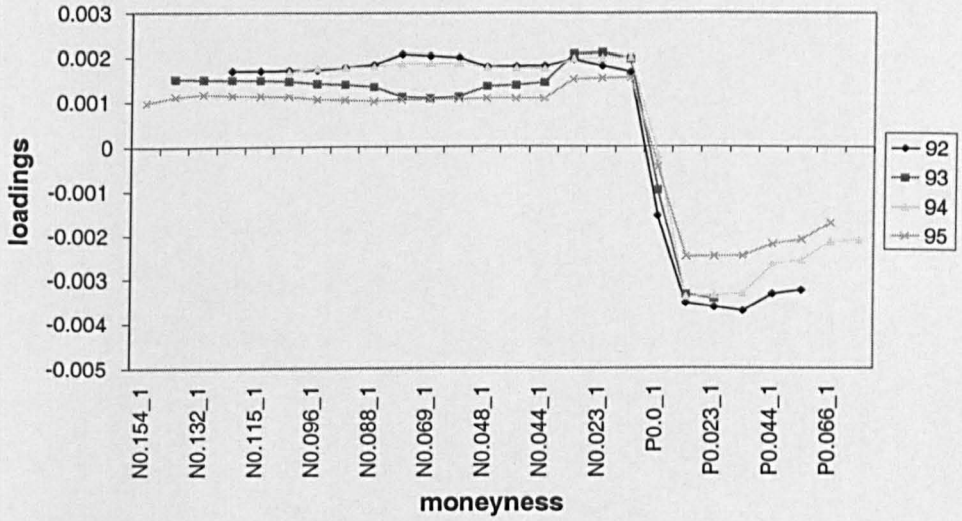


Figure 5.30: Surface Analysis on the Moneyness Metric: Interpretation of the Second Rotated PC for 270-180.

Our analysis shows that the implementation of a "smile-consistent" no-arbitrage stochastic volatility model suggests the need for three factors. One is required for the underlying asset, and two more for the implied volatility. The loadings of the two PCs may be useful for volatility risk management, even though we do not know whether they will be stable (in fact, they exhibit some variability over the

Year	Unrot. 1st PC	1st PC	2nd PC	Cumulative
92	38.6%	22.7%	38.6%	61.4%
93	34.5%	26.9%	34.3%	61.2%
94	40.6%	19.5%	40.4%	59.5%
95	39.2%	18.4%	39.2%	57.5%
Average	38.2%	21.8%	38.1%	60%

Table 5.11: Surface Analysis on the Moneyness Metric: Percentage of Variance Explained by the Unrotated First PC and by the Rotated PCs.

years). With this type of factor model, the definition of vega can be generalized to the sensitivities to each one of the volatility shocks.

Table 5.11 shows the percentage of the variance explained by the rotated first and second PCs. We find that after the rotation, it is not the shift which has the dominant effect on the implied volatility surface, but the second PC.

5.4 Comparing the Results from the Two Metrics

In this section, we compare the results of the PCA to the implied volatility surface between the two metrics. The number of variables that the PCA was applied to, the number of observations after the listwise deletion, and the KMO coefficients, were bigger for the strike metric, apart from year 1995. The assumption of multivariate normality was rejected for both metrics. This required the use of Velicer's criterion in order to decide on the number of PCs to retain. After considering three criteria, we decided that we should retain two PCs. In Table 5.12 we show the general interpretation of the unrotated PCs in both metrics; their interpretation was not consistent across expiries, and years.

The average, across the years, variance that the unrotated PCs explained in both metrics is shown in Table 5.13. We can see that the PCs in the moneyness metric explained more of the variance than in the strike metric. This is similar to what we found in the smile analysis.

Table 5.14 shows the general interpretation of the rotated PCs in both metrics.

Metric		
Strike	**1st PC**	Z-shape, Noisy in 1994 for 90-10
	2nd PC	Shift, Noisy in 1994
	3rd PC	Noise
Moneyness	**1st PC**	Z-shape for 1992, and 1993. Shift in 90-10 for 1994, 1995
		Z-shape for the other ranges
	2nd PC	Shift
	3rd PC	Noise

Table 5.12: Summarized Results for the Interpretation of the Unrotated PCs in the Surface Analysis in both Metrics.

Metric	AvVar 1stPC	AvVar 2ndPC	Cumulative
Strike	37.6%	14.7%	52.5%
Moneyness	38.2%	21.7%	60%

Table 5.13: Summary of the Average Variance that the Retained Unrotated PCs explain in the Surface Analysis in both Metrics.

The rotated first PC corresponded to a shift, and the second corresponded to a Z-shaped twist. It is worth recognizing that the interpretation of the first rotated PC was very close to the original second PC. Moreover, after the rotation it was not the first PC which explained most of the variation of the implied volatility surface, but the second one. Finally, considering the size of the effect of the rotated PCs across the expiries, the first PC affected more the shorter range implied volatilities in both metrics. On the other hand, the second PC affected more the longer ranges' implied volatilities, in both metrics.

Metric		
Strike	**1st PC**	Shift; For 180-90, 270-180 the shift has a triangular shape
	2nd PC	Z-shape
Moneyness	**1st PC**	Shift
	2nd PC	Z-shape, shift for 1993, 1994, 1995 in 90-10

Table 5.14: Summarized Results for the Interpretation of the Rotated PCs in the Surface Analysis in both Metrics.

5.5 Correlations of the Changes of the Futures Price with the Changes of the PCs under the Two Metrics

We calculate the correlations between the percentage changes of the futures price and the changes of the rotated PCs, by using the futures price from each expiry range and the common "pooled" PC. The correlations are calculated with the Pearson coefficient. One asterisk is displayed when the coefficient is significant at 5% significance level, and two asterisks are displayed when the coefficient is significant at 10% significance level.

In Table 5.15 we show the correlations between the percentage changes of the futures price and the changes of each one of the first two rotated PCs for both the strike and moneyness metric. In the strike metric, the correlation is positive for both the PCs (even though the correlation for the first PC is insignificant for year 1992). In the moneyness metric, the correlation for the first PC is negative (leverage effect), while for the second PC is positive, apart from year 1993. We investigated the scatterplots in order to explain the change of sign in the calculated correlations for year 1993. Although there were some outliers, removing them did not affect the sign of the correlation. The sign of the correlations in the moneyness metric, is the same for both the smile and the surface analysis (see Table 4.15).

Moreover, the magnitude of the correlations varies randomly over years. The instability of the correlations poses problems for the implementation of models such as Ledoit and Santa-Clara's [106].

5.6 Conclusions

In this Chapter, we have investigated the dynamics of implied volatilities surfaces of the S&P 500 Futures Options by applying Principal Components Analy-

Range			92	93	94	95
90-10	**Strike**	$\Delta PC1$	0.13	0.09	-0.11	0.26**
		$\Delta PC2$	0.40**	0.33**	0.06	0.24*
	Moneyness	$\Delta PC1$	-0.28**	-0.36**	-0.56**	-0.29**
		$\Delta PC2$	0.35**	-0.32**	0.16	0.26*
180-90	**Strike**	$\Delta PC1$	0.11	0.14	0.26**	0.26*
		$\Delta PC2$	0.41**	0.36**	0.38**	0.26*
	Moneyness	$\Delta PC1$	-0.29**	-0.27**	0.03	-0.29**
		$\Delta PC2$	0.35**	-0.43**	0.27**	0.26*
270-180	**Strike**	$\Delta PC1$	0.11	0.15*	0.26**	0.23*
		$\Delta PC2$	0.41**	0.37**	0.36**	0.26**
	Moneyness	$\Delta PC1$	-0.27**	-0.28**	0.06	-0.31**
		$\Delta PC2$	0.35**	-0.45**	0.27**	0.27**

Table 5.15: Surface Analysis: Correlations between Changes of the Futures Price with Changes of the Rotated PCs on the Strike and Moneyness Metric.

sis (PCA). This provides important insights into the behavior of traded option prices, and it is a prerequisite for the future development of models to realistically portray the dynamics of implied volatility surfaces.

The PCA was performed on the first differences of implied volatilities measured under the strike and the moneyness metric. After considering three criteria (Velicer's criterion, communalities, interpretation), we found that two components could be reliably extracted. We next applied a novel form of "Procrustes" rotation to the components, so as to obtain a clear interpretation for them. The first component can be interpreted as a parallel shift, and the second as a Z-shaped twist. These results were in most cases very consistent for the four separate years 1992, 1993, 1994 and 1995. On average, the two factors explained 53% and 60% of the surface variation in the strike and moneyness metric, respectively.

Our results have further implications. First, running PCA analysis both on individual volatility smiles and on the whole volatility surface gave consistent results on both the number of factors (two) and their interpretation. In this respect, the term structure of implied volatilities does not affect the dynamics of implieds. However, we have to bear in mind that the magnitude of shocks may

differ between the individual smiles and the surface. The fact that we had to use different expiry ranges for the smile and the surface analysis does not allow us to draw a definite conclusion on the effect of the maturity dimension on the dynamics of implied volatilities.

Second, the results again confirm the findings of other authors, e.g. Buraschi and Jackwerth [28], and Dumas, Fleming and Whaley [58], that deterministic volatility models are not capable of representing the dynamics of the volatility surface. Third, our results are in contrast with Kamal and Derman's [102]. They analyze the dynamics of implied volatilities of over the counter (OTC) S&P 500 and Nikkei 225 Index options. They find that three PCs explain about 95% of the variance of the volatility surface. Their interpretation is a level of volatilities for the first PC, a term structure of volatilities for the second PC and a skew for the third. Their results suggest that a four factor model for pricing and hedging options under a stochastic volatility "smile consistent no-arbitrage pricing" type model is an appropriate one. Our work, identifies a model with one less factor, but cautions that the factor structure has higher dimensionality since it explains only 60% of the surface variation[5]. Despite considerable care and effort to eliminate the main sources of measurement error in the traded market data and to minimize its effects, the market futures option data seem to exhibit a noisier volatility structure than that estimated from the quotations for OTC index options provided by Goldman and Sachs traders. Portfolios of options modelled as riskless under a three or four factor models, may in fact exhibit substantial market risk.

Finally, just as Fung and Hsieh [70] and Tompkins [143] have shown that the typical smile structure depends on the particular underlying asset, it seems certain that the dynamics of the volatility surface must also be specific to the

[5]It is worth noting that our model has similarities to that used recently in a time series analysis of index returns by Gallant, Chien and Tauchen [71]. Using daily data on close-to-close price movements and the high/low spread, they find that a model with two stochastic volatility shocks plus the underlying asset component, fits the data very well, and in particular it mimics the long-memory feature of volatility.

choice of asset[6]. The proper test of all such models lies in out-of-sample hedging analysis. This is well beyond the scope of the current study, but deserves to become a topic for future research.

[6]For example, short maturity options on indices display an implied volatility skew, while options on foreign currency exhibit an implied volatility smile.

Chapter 6

A New Method for Simulating the Evolution of the Implied Distribution

6.1 Introduction

There are two approaches that can be taken for the construction of an option pricing model. The first approach specifies the process for the underlying asset in advance, and then the model is developed. The simplest example of this approach is the Black-Scholes [20] model, which assumes a lognormal diffusion with constant volatility for the underlying. Richer processes which include either stochastic volatility (see among others Hull and White [90], Johnson and Shanno [101], Scott [131]), or jumps (see Bates [10] Merton [113]), or both (see Bates [11], [12], Scott [132]), can give rise to implied volatility patterns which are somewhat similar to the ones observed. However, none of these models fully explains the empirically observed implied volatilities (see Das and Sundaram [47], Taylor and Xu [141]).

The second approach reverses the option pricing problem. Rather than specifying the asset process exogenously, it starts from the observed European option

prices and it implies the process from them (implied process). There are two streams of literature falling within this approach. The first stream, extracts a deterministic volatility implied process (see Derman and Kani [49], Dupire [60], [62], Jackwerth [96], Rubinstein [126]). The second, derives a no-arbitrage stochastic volatility implied process (see Britten-Jones and Neuberger [27], Derman and Kani [54], Ledoit and Santa-Clara [106]).

The deterministic volatility implied process ensures an exact fit of the current observed smile, but is not able to account for its stochastic evolution (see Dumas, Fleming, and Whaley [58], and Buraschi and Jackwerth [28]). This drawback can be dealt with a stochastic volatility implied process; its evolution should be in a no-arbitrage fashion, so that option pricing to be feasible.

Derman and Kani combine an implied trinomial tree (see Derman, Kani, and Chriss [53] for the construction of such a tree) with Monte-Carlo simulation by starting from today's local volatilities. Ledoit and Santa-Clara, start from today's implied volatilities and they derive the risk-neutral implied volatility process; Britten-Jones and Neuberger construct a trinomial tree under stochastic volatility which is consistent with today's option prices. These models are very promising, but their implementation is subject to some theoretical and practical limitations. Derman and Kani's model is very computer intensive. Britten-Jones and Neuberger's is based on the assumption that the asset's transition probabilities can be decomposed, in a multiplicative way, into two other functions; in Ledoit's and Santa-Clara's, the implementation is very sensitive to the chosen interpolation method for the implied volatility surface.

In this chapter, we propose a new and general method for constructing smile-consistent stochastic volatility models. Rather than simulating the evolution of implied, or local volatilities, *we simulate the evolution of the implied risk-neutral distribution* through time. The simulation starts from today's implied distribution, and proceeds in an arbitrage free way. This type of approach provides a natural tool for risk assessment and for applications in pricing and hedging. However, even though the way to extract the risk-neutral distribution from Eu-

ropean option prices has been studied extensively (see for a survey Bahra [5], and Mayhew [111]), we are not aware of any research so far, on its simulation.

In principle, we can calculate the option price by integrating the payoff of the option over the risk-neutral distribution. Hence, the simulation of the implied distribution enables us to generate price paths for the options having the same expiry simultaneously. In addition, the technique can be used to examine the behavior of hedges. For example, a riskless hedge is constructed by using a deterministic volatility model. Then, its performance can be investigated under a stochastic volatility scenario generated by our model. The technique can also be used for economic policy purposes. Since the implied distributions reveal the expectations of risk-neutral agents, their evolution shows how these expectations change over time. Hence, provided that the subjective and the risk-neutral distributions do not differ a lot, issues like the credibility of economic policy announcements can be addressed (see Bahra [5]). On the other hand, the proposed algorithm can not be used for valuing simultaneously options with different expiries and American type products, as is the case with the conventional Monte Carlo simulation.

Our method uses the theory of mixture of distributions, and it can be implemented easily. Using mixture of distributions has the advantage that the simulated probabilities evolve as martingales, i.e. the model does not allow any arbitrage opportunities. Once the form of the mixtures has been specified, the only input that the algorithm requires is the current implied distribution.

The algorithm is developed in two stages. First, the implied distribution changes due to shocks to its mean. Second, the implied distribution changes because of shocks to the variance. The two shocks are assumed to be orthogonal, and therefore they can be studied separately. The simulation of the evolution of the implied distribution can be performed by applying them sequentially. Moreover, the algorithm can be extended to cases where more complicated assumptions for the mixture are considered.

The rest of the paper is structured as follows. In the second section, we explain why the simulation of the implied distribution is relevant for option pricing.

In the third section, we describe briefly the idea of mixture of distributions, and we present the main steps for the construction of the algorithm. In the fourth and fifth section, we present the algorithm for the simulation of the implied distribution when its mean and variance change, respectively. Numerical examples are presented. The sixth section concludes, and it addresses issues for further research.

6.2 Simulation of Implied Distributions and Option Pricing

There are two analogies between the conventional Monte Carlo simulation, and the simulation of the implied distribution. In the Monte Carlo simulation, for each simulation run we generate a path for the asset price S_t, until we reach the maturity T of the option. In the simulation of the implied distribution, a path for the evolution of the whole distribution is created, until we reach $t = T$, where the distribution degenerates to a single point S_T. The probability density function (PDF) enables us to value options for every strike, and its mean provides the value of S_t (and hence its path). In both cases, we require some thousands of simulation runs to obtain a credible estimate of the option price.

Denote by $\pi_t(S_T)$ the risk-neutral probability, as formed at current time t, that the asset price S_T will be reached at time T $(t < T)$. $\pi_t(S_T)$ can be extracted from the market call option prices C maturing at time T, which are observed at time t. The way to do this comes from the well-known relationship

$$\frac{\partial^2 C}{\partial K^2} = e^{-r(T-t)} f_T(K) \tag{6.1}$$

established by Breeden and Litzenberger [25], where K is the strike price, r is the interest rate, and $f_T(K)$ is the risk neutral PDF. Assuming that we observe option prices for a continuum of strikes, we can calculate the probabilities for reaching every S_T.

The simulation of the implied density, can reveal to us the path of the asset price. This is because the current asset price S_t is proportional to the expected value of S_T (martingale property of the implied distribution). The expectation is conditional on the information at time t, and it is formed with respect to the risk-neutral probability measure $\pi_t(S_T)$ (see Dothan [57]). Hence,

$$S_t = e^{-r(T-t)} E_t^*(S_T) \tag{6.2}$$

If we knew what the implied risk neutral probability $\pi_{t+1}(S_T)$ would be in the next time step $t+1$ for $\forall S_T$, then we would be able to calculate the asset price S_{t+1}, as equation (6.2) shows. In this way we would have in front of us a path of the asset price for all times n for $t < n < T$.

In fact, the implied density at $t+1$ is unobserved at time t; Today's known implied density is perturbed due to a number of shocks which affect its moments; as a result, the next time step implied density emerges. Hence, the simulation of the implied distribution is a multivariate problem since it should take into account these shocks. In the next sections, we perturb the initial implied density by assuming that orthogonal shocks affect its mean and variance. The orthogonality assumption is necessary, so as to convert the multivariate problem to a simpler univariate one, and to examine the effect of the shocks, separately. The shocks which are orthogonal, can be applied sequentially, and the multivariate problem becomes a sequence of multivariate ones. The shock on the mean is the shock on the asset price, and we expect to shift the PDF to the left, or to the right; the shock on the variance should shrink the PDF, as we approach the maturity T.

6.3 Partitions, Mixtures, and the Simulation of the Implied Distribution

In this section, we explain first briefly the concept of the mixture of distributions (for more details on the mixture of distributions, see Timmerington, Smith, and

Makov [142]). Then, we present, in a general context, the algorithm with which we can perturb the implied distribution, when one of its moments changes. A mixture of distributions is assumed for the construction of the algorithm.

6.3.1 Partitions and Mixtures

Let $f_0(S_T)$ be the PDF of the asset price S_T, as viewed from time 0 (i.e. conditional on the information at time 0), and $f_t(S_T)$ be the PDF of S_T, as viewed from a later time t. The evolution of the initial PDF $f_0(S_T)$ to $f_t(S_T)$ can be regarded as a partition of $f_0(S_T)$. $f_0(S_T)$ consists of infinite many different slices; one of them is the realized $f_t(S_T)$[1]. This is a result from the law of iterated expectations. More specifically, $\Pr[S_T \leq s \mid I_k] = E[1_{S_T \leq s} \mid I_k]$, and $\Pr[S_T \leq s \mid I_t] = E[1_{S_T \leq s} \mid I_t]$, where $k < t$, and the information set $I_k \subset I_t$. Then, from the law of iterated expectations (see Oksendal [117]) we get that

$$E[E[1_{S_T \leq s} \mid I_t] \mid I_k] = E[1_{S_T \leq s} \mid I_k]$$

Hence, the original PDF at time k, evolves to a distribution on date t, conditional on the information at the subsequent date t. The set of possible conditional distributions has to be constrained so that the original distribution is seen as a mixture over the possible subsequent conditional ones. Figure 6.1 shows how the original PDF is partitioned in two possible slices. One of them is going to be the subsequent conditional PDF (once its area is normalized to one).

[1]The area of each slice does not integrate up to one, since it is a subset of the original distribution. Therefore, it has to be scaled so that a proper PDF to be defined.

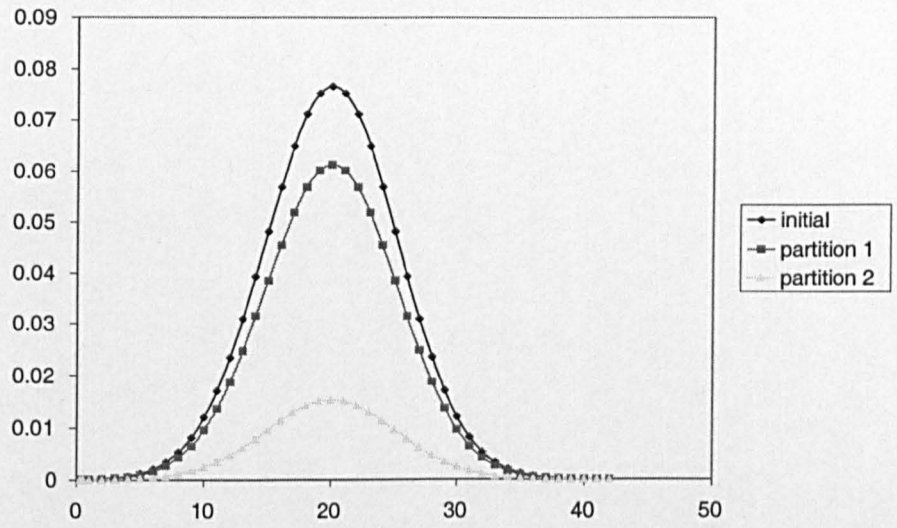


Figure 6.1: Partitioning of the Original Distribution in two possible realizations.

A mixed distribution is generated from two PDFs. One distribution, called the *structural distribution*, has a parameter which itself is distributed according to a second distribution, called the *mixing distribution*. The structural distribution of a variable X is assumed to have a density $g(X \mid \theta)$, where θ is a parameter which varies across subsets of X. In our context, θ will be one of the moments of the distribution. It is assumed that the distribution of the variable θ is specified by a second distribution, a mixing density $k(\theta)$. The resulting distribution for X is the mixture of the density functions $g(X \mid \theta)$ and $k(\theta)$. This observed, or mixed density is defined by

$$m(X) = \int g(X \mid \theta)k(\theta)d\theta \tag{6.3}$$

Equation (6.3) defines a continuous mixture. Similarly, a discrete mixture can be defined as

$$m(X) = \sum_{i=1}^{n} g(X \mid \theta_i)k(\theta_i) \tag{6.4}$$

Equations (6.3), and (6.4) are also satisfied by the cumulative density functions (CDF) $M(X)$, $G(. \mid \theta)$, and $K(\theta)$ (we denote with capital cases a CDF, and with lower cases a PDF, hereafter).

6.3.2 Simulating the Implied Distribution

The General Approach

We can relate now the concept of a mixture to the one of a partition by regarding $m(X)$ as a model for $f_0(S_T)$[2]. In order to apply this model so that we can simulate the evolution of $f_0(S_T)$ into $f_t(S_T)$, we need to establish a mapping between our model variable X, and the actual variable S_T. In the next section, we describe how we use the distributions of $m(X)$, and $f_0(S_T)$ to accomplish this.

The Details

We assume a mixture of distributions, so as to construct the algorithm which simulates the evolution of the implied distribution. The mixture is used as a model for today's implied PDF[3]. The simulation is performed by mixing over one of the moments of the PDF. To simplify the analysis, we simulate the implied distribution over only two time steps.

Let F_0^{-1} be the inverse function of the CDF F_0, as viewed from time 0, i.e.

$$F_0^{-1} : P \rightarrow S_T. \tag{6.5}$$

where P is the cumulative probability, i.e. $P \equiv \Pr[S_T \leq s \mid I_0] = \int_0^s f_0(S_T)dS_T$. Similarly, the inverse of the cumulative distribution as viewed from time t F_t^{-1},

[2] $k(\theta)$ is the area (less than one) of one of the possible slices of $m(X)$. It is chosen once the corresponding value of θ has been realized. The structural distribution $g(X \mid \theta)$ is $f_t(S_T)$; it is the slice given by $k(\theta)$, but its area has been normalized to one. The mixing density $k(\theta)$ provides the weights with which the $f_t(S_T)$ are averaged.

[3] Mixtures of distributions have long been used in the finance literature to explain the observed significant skewness, and kurtosis in the distribution of rates of returns (see Blattberg and Gonedes [21], Kon [104], and Praetz [120]). The motivation for using a mixture of distributions comes from the empirical evidence that daily returns deviate from normality more than monthly ones (see among others Blattberg and Gonedes [21], and Fama [67]). The mixture of distributions gives rise to a fat-tailed unconditional distribution with a finite variance, and finite higher moments. Since all moments are finite, the Central Limit theorem applies, and long-horizon returns will tend to be closer to the normal distribution than short-horizon returns (see Campbell, Lo, and Mackinlay [29]). Our work can be viewed as a generalization of this literature to enable us to work with an arbitrary initial pricing kernel.

is defined as

$$F_t^{-1} : Q \rightarrow S_T \tag{6.6}$$

where $Q \equiv \Pr[S_T \leq s \mid I_t] = \int_0^s f_t(S_T) dS_T$[4].

The aim is to obtain the next time step inverse function F_t^{-1} when one of the moments of today's PDF changes. We only know F_0^{-1}, as this is implied from the European option prices. However, we can derive F_t^{-1}, by establishing a mapping from Q to P ($Q \rightarrow P$). Then, since we know what F_0 looks like, we can calculate the corresponding asset value S_T from the cumulative probability P. Now we know $Q \rightarrow S_T$. The mapping ($Q \rightarrow S_T$) is established via four steps.

1. We assume a mixture distribution

$$M(X_T) = \Pr[X_T \leq x] = \int G(X_T \mid \theta) K(\theta) d\theta \tag{6.7}$$

with CDFs $G(. \mid \theta)$, and $K(\theta)$[5].

The mixing is done over θ which can be the mean, or the variance, or the skewness, etc. We use the mixture $M(X)$ as a model for today's implied distribution, i.e. $M(X) = P$, and $G(X_T \mid \theta)$ as a model for the new distribution. We have a great deal of flexibility in the choice of the CDFs $G(. \mid \theta)$, and $K(\theta)$ to provide the kind of stochastic evolution we seek. The main requirement for computational purposes is that $G(. \mid \theta)$ and $G^{-1}(. \mid \theta)$ can be calculated reasonably efficient.

2. For any given probability level $Q \in (0, 1)$ and a given θ, we calculate X_T by inverting the CDF $G(. \mid \theta)$ for every Q, i.e.

$$X_T = G^{-1}(Q \mid \theta) \tag{6.8}$$

3. From the calculated X_T, we calculate $M(X_T)$ from equation (6.7). Since

[4]We prefer using the inverse function of a CDF rather than that of a PDF. This enables us to work in the interval [0,1], rather than in the interval [-∞, +∞].

[5]Alternatively, the mixture can be discrete for the sake of tractability in the numerical implementation.

$M(X_T) = P$, the mapping $Q \rightarrow P$ is established.

4. Since $M(X_T)$ is only a model which may not represent accurately today's implied CDF, we make the pairs S and X equivalent through the relation $M(X_T) = P(S_T)$. So, we can scale S from X as $S = P^{-1}[M(X_T)]$ (or the less interesting case, X from S as $X = M^{-1}[P(S)]$). By doing this, we incorporate in our method the information from today's implied distribution. In order to invert $P(S_T)$ we need to approximate $P^{-1}(S_T)$ with a CDF whose inverse function has an analytic formula (see Section 6.3.3).

The whole procedure can be repeated, as described, for any probability values Q are needed. Figure 6.2 shows how the mapping $Q \rightarrow S_T$ is established via the four steps; it shows, for a given value of θ, the transition from Q to X_T, then from X_T to $M(X_T)$, and eventually from X_T to the actual level S_T.

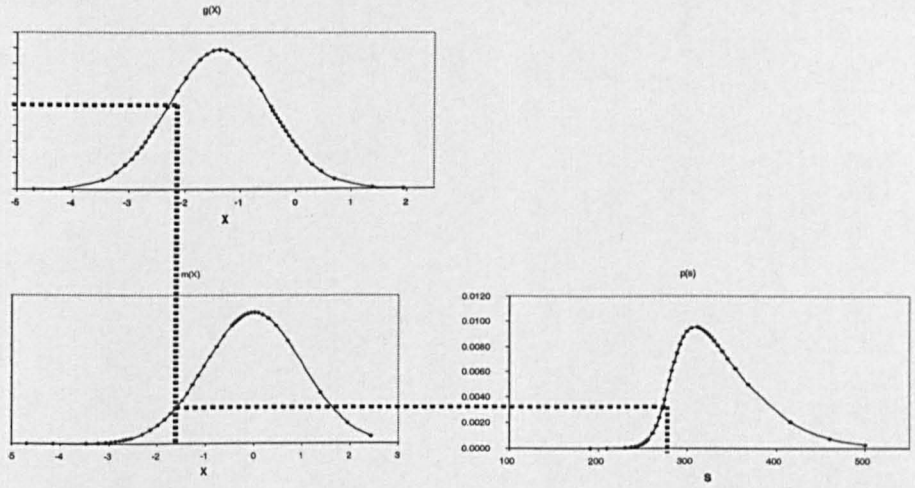


Figure 6.2: Establishing the Mapping from Q to S_T, using the mixture of distributions $M(X)$.

The process for establishing $Q \rightarrow P$ is a legitimate because it ensures that the date T risk-neutral distribution for S_T is a martingale, and hence so are all securities which are priced by it, e.g. options and the underlying asset. Figure 6.3 shows the martingale property of the implied distribution assuming that θ

can take only two values, i.e. $P(S_T) = \pi_1 Q(S_T \mid \theta_1) + \pi_2 Q(S_T \mid \theta_2)$, where S_T takes a discrete distribution.

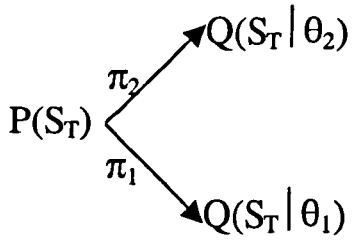


Figure 6.3: Martingale Property of the Implied Distribution.

The following theorem proves that the algorithm simulates the CDF so that it is a martingale.

Theorem 1 *The established mapping $Q \rightarrow S_T$ ensures that the simulated implied cumulative distribution function evolves as a martingale, i.e.*

$$\int Q(S_T \mid \theta) K(\theta) d\theta = P(S_T)$$

(martingale property).

Proof. $\int Q(S_T \mid \theta) K(\theta) d\theta = \int G(X_T \mid \theta) K(\theta) d\theta = M(X_T) = P(S_T)$. The last equality follows because X_T is chosen so as $M(X_T) = P(S_T)$.

■

The established mapping $Q \rightarrow S_T$, simulates the CDF. The following proposition shows how we can simulate the PDF, i.e. perturb $f_0(S_T)$ to $f_t(S_T)$, by using $Q \rightarrow P$.

Proposition 2 *The PDF for the underlying asset at the maturity of the option $f_t(S_T)$, as viewed from time t, is given by*[6]

$$f_t(S_T) = f_0(S_T) \times \frac{dQ}{dP} \tag{6.9}$$

[6] *The evolution of $f_0(S_T)$ to $f_t(S_T)$ can be regarded as a change of measure, and the ratio $\frac{dQ}{dP}$ can be interpreted as the Radon-Nikodym derivative.*

Proof. The mapping $Q \rightarrow P$, implies that $Q = Q(P)$. In addition, $P = P(S)$. Hence, $Q = Q(P(S))$. Differentiating Q with respect to S, we have

$$\frac{dQ}{dS} = \frac{dQ}{dP}\frac{dP}{dS} \tag{6.10}$$

which can be written as

$$f_t(S_T) = f_0(S_T)\frac{dQ}{dP}$$

■

Proposition 3 *The PDF $f_t(S_T)$ calculated from equation (6.9), integrates up to one, i.e.*

$$\int_{-\infty}^{+\infty} f_t(S_T)dS = 1$$

Proof. Using equations (6.9) and (6.10) we have that

$\int_{-\infty}^{+\infty} f_t(S_T)dS = \int_{-\infty}^{+\infty} f_0(S_T)\frac{dQ}{dP}dS = \int_{-\infty}^{+\infty} f_0(S_T)\frac{1}{f_0(S_T)}\frac{dQ}{dS}dS = \int_{-\infty}^{+\infty} dQ =$ 1.

■

It is worth noting that the mapping $Q \rightarrow P$ depends on the postulated mixture given by equation (6.7). For a different mixture, the mappings $Q \rightarrow X$, $X \rightarrow M(X)$, and $M(X) \rightarrow S_T$, described by Steps 1-4 would have been different. The analytic (or numerical calculation) of the derivative $\frac{dQ}{dP}$ depends on the assumed mixture of distributions, as well. In general, the analytic expression for the derivative is

$$\frac{dQ}{dP} = \frac{\frac{dQ}{dX}}{\frac{dP}{dX}} = \frac{g(X_T \mid \theta)}{m(X_T)} \tag{6.11}$$

In sections 6.4 and 6.5, we study the simulation of the evolution of the implied distribution when either the mean, or the variance changes. Imposing shocks only on the first two moments should make the simulated evolution of the implied distribution to be close to the empirical one. This is because the observed skewness may be explained by shifts in the mean, and the observed fat tails are consistent with shifts in the variance (see Kon [104]). However, the algorithm can be easily

generalized to other cases, by choosing more complicated assumptions for the mixture.

6.3.3 Inverting the Initial Implied Distribution

We approximate $P^{-1}(S_T)$ by using the Ramberg distribution [121]. This is because it is defined analytically in terms of the inverse of the PDF, given that we have pre-specified its first four moments. Moreover, it is a four parameter distribution, which includes a wide variety of curve shapes. Hence, it is useful for the representation of data when the underlying model is unknown. All these characteristics make it ideal for Monte Carlo simulation purposes[7]. Its percentile function is[8]

$$S = R(P) = \lambda_1 + [P^{\lambda_3} - (1-P)^{\lambda_4}]/\lambda_2 \tag{6.12}$$

where $0 \leq P \leq 1$, is the cumulative probability, λ_1 is a location parameter, λ_2 is a scale parameter and λ_3 and λ_4 are shape parameters. The values for the λ_1, and λ_2 parameters, are given in tables (see Ramberg et al. [121]) for a variable with a mean of zero and a variance of one. To calculate them for a variable with mean μ and variance σ, we use

$$\lambda_1(\mu, \sigma) = \lambda_1(0,1)\sigma + \mu$$

$$\lambda_2(\mu, \sigma) = \lambda_2(0,1)/\sigma$$

The values of λ_3, and λ_4 correspond to values of skewness and kurtosis, and they are given by tables, as well. The PDF corresponding to equation (6.12) is given by

$$f(x) = f(R(P)) = \lambda_2[\lambda_3 P^{\lambda_3 - 1} + \lambda_4(1-P)^{\lambda_4 - 1}]^{-1} \tag{6.13}$$

[7] Other rich classes of distributions defined in terms of the PDF (e.g. the Generalized Beta-2 distribution [23]) are less suitable for the purposes of our study.

[8] The percentile function is the inverse of the CDF, i.e. $R(P) : P \rightarrow S$.

In order to plot the density for given λ_1, λ_2, λ_3, λ_4, we need to evaluate equations (6.12) and (6.13) for values of P ranging from zero to one. Then, $f(R(P))$ is plotted on the y-axis versus $R(P)$ on the x-axis.

6.4 Simulating Changes in the Asset Price

In this section, we present the algorithm for the simulation of the PDF when its mean changes. This corresponds to changes in the asset price. We provide a numerical example, and we check the numerical accuracy of our procedure.

6.4.1 The Algorithm

In order to study how the PDF changes when the mean changes (keeping all the other moments fixed), we assume that the model variable X follows a standard Brownian motion W_t, i.e. $X_t = W_t$. This implies that X is distributed normally[9].

Imagine that we stand at time 0, and the terminal time is $T = 1$. From the properties of the standard Brownian motion process, the model's asset price as viewed from time 0, ${}_0X_1$, is normally distributed with mean zero, and variance one, i.e.

$$ {}_0X_1 \sim N(0,1) \tag{6.14} $$

The asset price as viewed from time t, is normally distributed with mean μ_t, and variance $1-t$, i.e.[10]

$$ {}_tX_1 \sim N(\mu_t, 1-t) \tag{6.15} $$

[9] In fact, the observed implied distributions do not differ a lot from the normal one. Corrado and Su [39] examine the skewness and kurtosis in the implied distribution of stock options. They find that the implied skewness and kurtosis are -0.55 and 4.92, respectively, on average across their samples. Gemmill and Saflekos [74] investigate the implied skewness and kurtosis of FTSE 100 options. They find an average, across years, skewness and kurtosis of -0.26 and 4.54. Similar results are also obtained by Jackwerth and Rubinstein [94].

[10] Equation (6.15) shows that the statement that we allow only for a change in the mean of the distribution, keeping all the other moments fixed, is not strictly true. There is a change in the variance, as well, due to the passage of time.

μ_t is normally distributed with zero mean and variance t, i.e.

$$\mu_t \sim N(0,t) \tag{6.16}$$

This is because under the equivalent martingale measure, the relative asset price at time t is[11]

$$X_t = E_t^*[X_1] = \mu_t \tag{6.17}$$

and μ_t is one of the possible realizations of the Brownian Motion as viewed from time 0 for the time interval $[0,t]$. Then, $Q \rightarrow S_T$ is defined through the four steps of the generic procedure that we have already described.

1. Equations (6.14), (6.15), and (6.16) make us to choose the mixture

$$P(S_T) = M(X_T) = \int_0^\infty N(X_T \mid \mu_t, 1-t) N(\mu_t \mid 0,t) d\mu \tag{6.18}$$

as a model for the implied distribution at time 0[12].

2. Since $\mu_t \sim N(0,t)$, we define the mapping

$$N^{-1}(R \mid 0,t) : R \rightarrow \mu_t \tag{6.19}$$

where $R \equiv \Pr[\mu_t \leq \mu]$, and $N^{-1}(R \mid 0,t)$ is the inverse of the normal CDF when a value for the probability mass R accumulated under μ, is drawn.

For a given drawing from the random variable R (i.e. for a given value of μ_t), we get by inverting the normal CDF in relationship (6.15) that

$$X_T = N^{-1}(Q \mid \mu_t, 1-t) \tag{6.20}$$

[11] We use the bank account as numeraire (see Geman, Karoui, and Rochet [72]).

[12] This is because if a variable $X \sim N(0,1)$, and $\mu \sim N(0,t)$, then

$$M(X) = \int N(X \mid \mu, 1-t) N(\mu \mid 0,t) d\mu$$

This follows from the properties of the convolution of the PDFs of two normally distributed variables.

Equation (6.20) is applied for every Q, with $Q \in [0,1]$.

3. We calculate $M(X_T)$ as

$$M(X_T) = N(X_T \mid 0,1) \tag{6.21}$$

Since $M(X_T) = P(S_T)$, this establishes the mapping $Q \rightarrow P$, i.e.

$$P = N[N^{-1}(Q \mid \mu_t, 1-t) \mid 0,1] \tag{6.22}$$

4. We scale S_T from X, as $S_T = P^{-1}[M(X_T)]$.

$f_t(S_T)$ is calculated from equation (6.9). The derivative $\frac{dQ}{dP}$ is calculated analytically by using equation (6.11), and our assumed mixture, i.e.[13]

$$\frac{dQ}{dP} = \frac{n(X_T \mid \mu_t, 1-t)}{n(X_T \mid 0,1)}$$

6.4.2 A Numerical Example

In this section, we provide a numerical example of the simulation of an implied distribution, when the mean changes. The simulation is performed by using the mapping Q to P obtained from our postulated mixture, given by equation (6.18). The mapping has been established for $t = 0.2$. First, we calculate $f_0(S_T)$ via the Ramberg PDF, by making an assumption about the values of its first four moments. Then, we perturb $f_0(S_T)$, by using equation (6.9), in order to see what it looks like 0.2 units of time later.

Figure 6.4 shows the mapping from Q to P, for random drawings R= 0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99, and 0.999, i.e. for several values of μ_t, when $t = 0.2$.

[13] We have also calculated the derivative numerically, by using several numerical schemes (see Chapra and Canale [32]). However, the simulated PDFs were not smooth enough, because of the numerical errors.

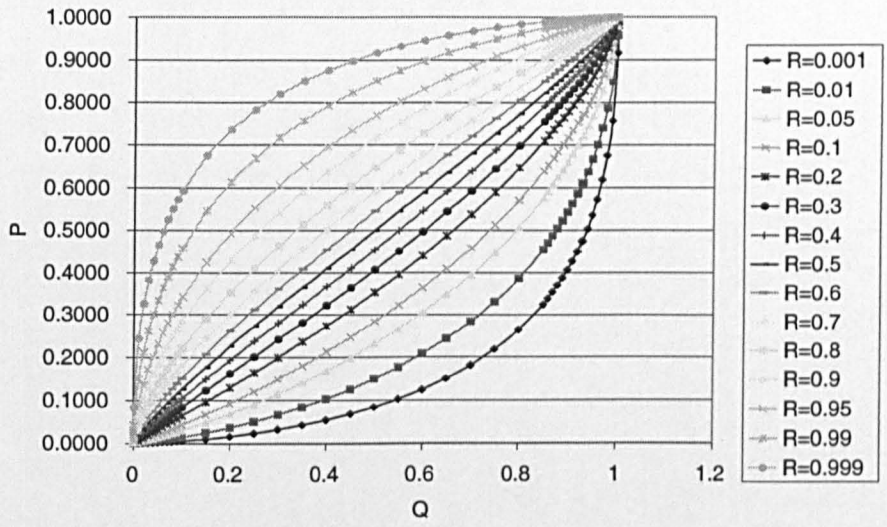


Figure 6.4: Mapping from Q to P when the Mean changes, for t=0.2.

The mapping shows that for the small values of R, $f_t(S_T)$ is shifted to the left of the initial $f_0(S_T)$, while for high values of R, $f_t(S_T)$ is shifted to the right. This is because for a given Q, R increases with P. An increase in R, increases μ_t. This shifts $F_t(S_T)$ to the right, since μ_t is its mean. Therefore, the values that S_T can take are higher, and consequently the probability mass P accumulated under them is larger, i.e.

$$R_1 < R_2 \Rightarrow \mu_{R_1} < \mu_{R_2} \Rightarrow S_{R_1} < S_{R_2} \Rightarrow P_{R_1} < P_{R_2}$$

Furthermore, the increase in P as R increases will be bigger, the bigger t is. This is because the variance of the distribution from which μ_t is drawn, increases when t increases. Hence, bigger shifts in the PDF are expected to happen. Therefore, for a given Q, an increase in R will produce a bigger increase in P, than with smaller t.

Another feature of the figures is that for a given R, P is increasing in Q. This is because increasing Q, increases S_T, and therefore P increases, as well.

For our numerical example, we assume that the mean of the initial implied PDF is 340, its variance is 2500, its skewness is 1, and its kurtosis is 4.2. From the tables of the Ramberg distribution, these values correspond to $\lambda_1 = -0.787$,

$\lambda_2 = 0.1142$, $\lambda_3 = 0.0212$, and $\lambda_4 = 0.1244$. Then, we adjust λ_1, and λ_2 for the mean and variance.

Figures 6.5, 6.6, and 6.7 show $f_0(S_T)$, and $f_t(S_T)$, for $R = 0.01$, 0.5, and 0.95, respectively. The figures show that as R increases, the new density shifts to the right. This reflects the property of the mapping that we have already explained, i.e. that P increases with R, for a given Q.

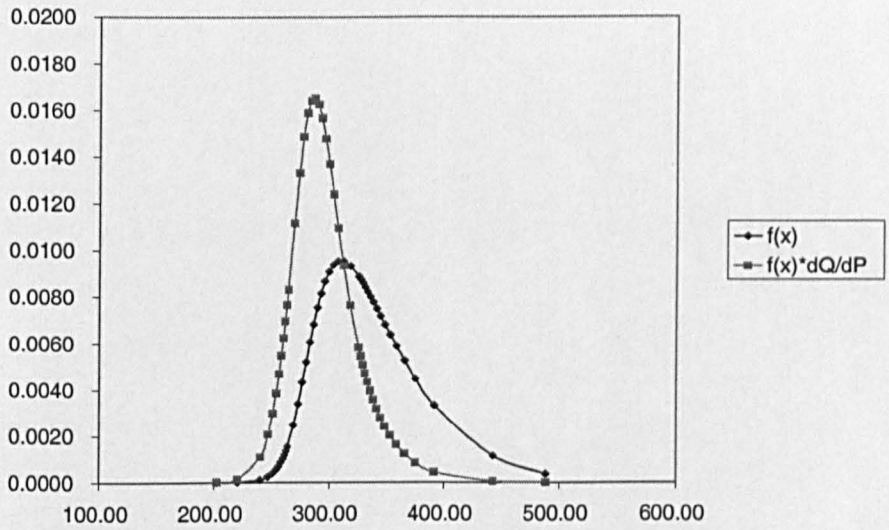


Figure 6.5: Initial and New PDF for R=0.01.

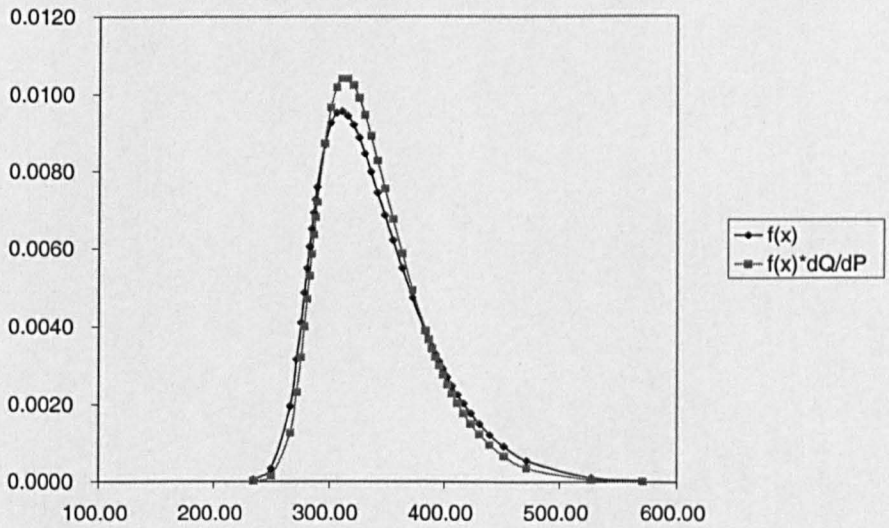


Figure 6.6: Initial and New PDF for R=0.5.

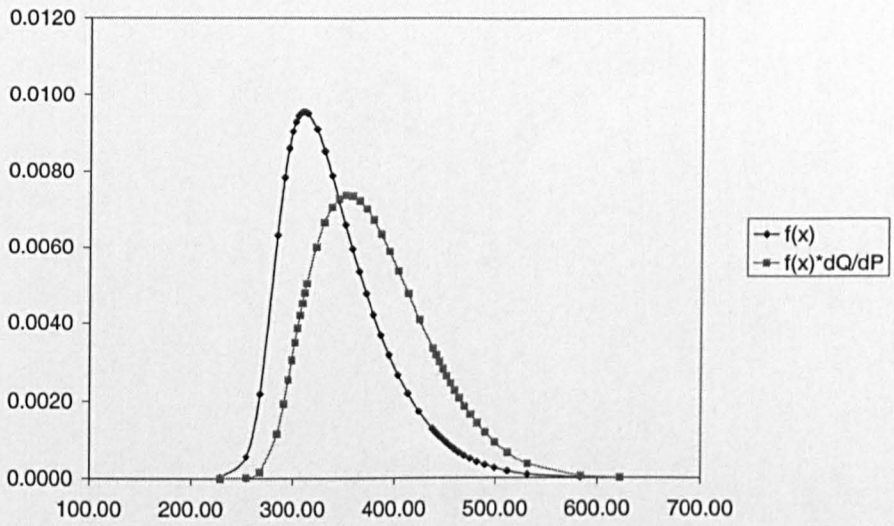


Figure 6.7: Initial and New PDF for R=0.95.

6.4.3 Checking the Accuracy of the Transformation

The mapping $Q \rightarrow P$, has been established for discrete values of Q and P, while the assumed mixture given by equation (6.18) is a continuous one. This may incur numerical errors in our simulation procedure. In this section we check for this by assuming that $S_T \sim N(0,1)$. Using $Q \rightarrow P$, we calculate $f_0(S_T)$ by means of the Ramberg distribution[14]. Then, we calculate $f_t(S_T)$ by means of equation (6.9). Finally, we compare $f_t(S_T)$ with the normal PDF when $S_T \sim N(\mu_t, 1-t)$. The comparison is shown in Figures 6.8, 6.9, and 6.10, for the cases that $R = 0.01$, 0.5, and 0.95, respectively.

[14]The values of the Ramberg parameters when the Ramberg PDF converges to the normal are $\lambda_1 = 0$, $\lambda_2 = 0.1975$, $\lambda_3 = 0.1349$, and $\lambda_4 = 0.1349$ (recall that the skewness and the kurtosis of the normal PDF are zero, and three, respectively).

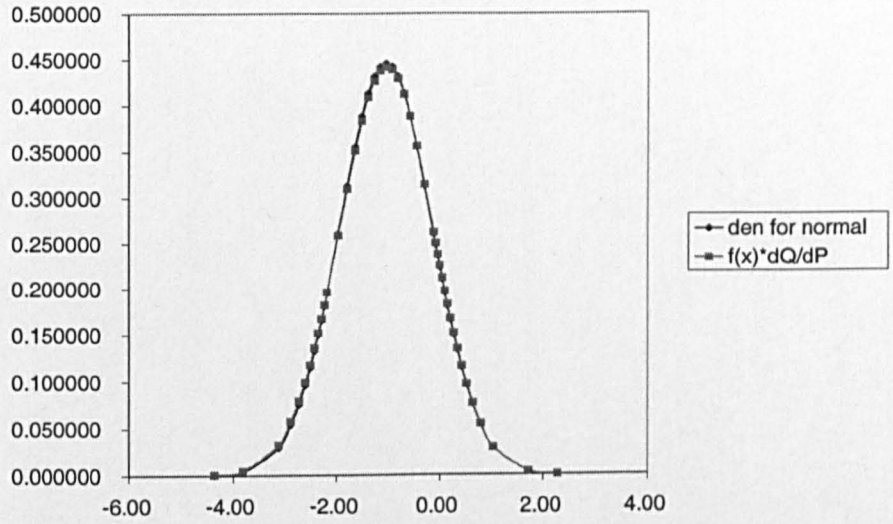


Figure 6.8: Comparison between the Ramberg-Normal and the Normal PDFs for R=0.01.

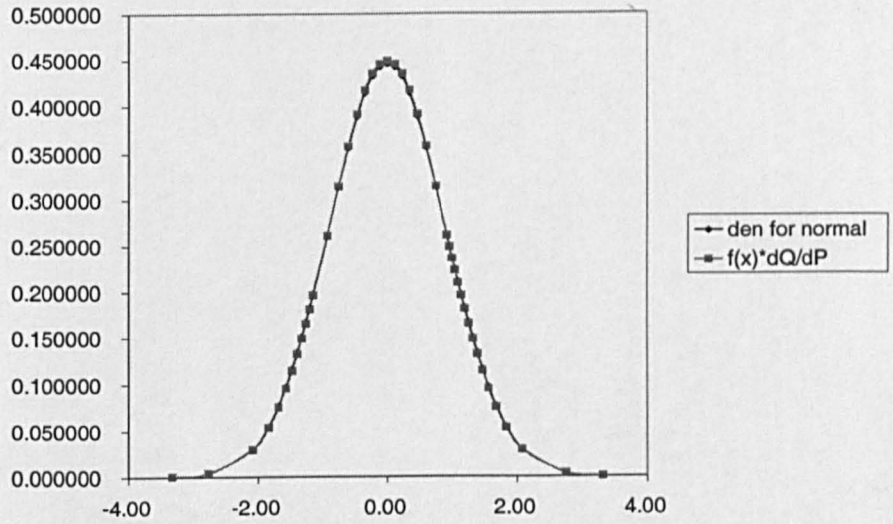


Figure 6.9: Comparison between the Ramberg-Normal and the Normal PDFs for R=0.5.

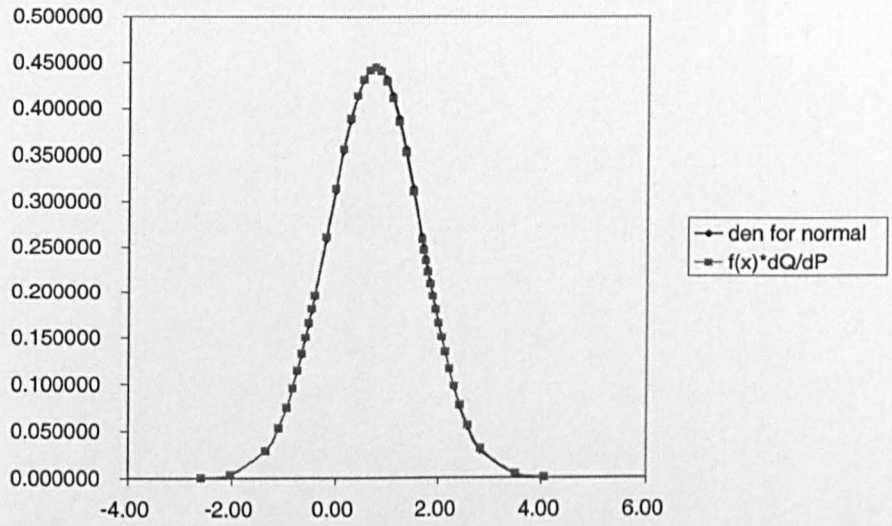


Figure 6.10: Comparison between the Ramberg-Normal and the Normal PDFs for R=0.95.

It is obvious that the Ramberg-Normal and the Normal PDFs coincide, showing that there is no numerical bias in our procedure.

6.5 Simulation of the Implied Distribution when the Volatility Changes

In this section, we develop the algorithm for the simulation of the PDF due to a change in its variance σ^2. We show how we determine the evolution of the variance, and we present some numerical examples.

6.5.1 The Algorithm

We need to establish the mapping $Q \rightarrow S_T$, in order to find F_t^{-1}, just as when the mean changes. However, now the mapping has to take into account the change in the variance; it is established through, the familiar by now, four steps.

1. We introduce the effect of the variance by assuming that the initial implied

CDF can be represented by a discrete mixture $M(X_T)$, i.e.

$$M(X_T) = \Pr[X_T \leq x] = \sum_{i=1}^{n} N(X_T \mid 0, \theta_i) K(\theta_i) \tag{6.23}$$

where θ_i is the *ith* realization for the variance variable, $K(\theta_i)$ is the CDF for the variance, and $N(X_T \mid 0, \theta_i)$ is the normal CDF of X_T conditional on θ_i occurring. In order to simplify the notation, we set $\pi_i = K(\theta_i)$.

2. We postulate a PDF for the variance for some discrete values of θ, so as to calculate π_i and θ_i. Then, a θ is drawn from the distribution $K(\theta)$, and we get

$$X_T = N^{-1}(Q \mid 0, \theta)$$

for every Q.

3. For the calculated from the second step X_T, we compute $M(X_T)$.

4. We calculate S_T, since $M(X_T) = P(S_T)$, as $S_T = P^{-1}[M(X_T)]$.

Having established the mapping $Q \rightarrow S_T$, $f_t(S_T)$ is calculated from equation (6.9), with the derivative being calculated from equations (6.11) and (6.23) as

$$\frac{dQ}{dP} = \frac{n(X_T \mid 0, \theta)}{\sum_{i=1}^{n} n(X_T \mid 0, \theta_i) K(\theta_i)}$$

6.5.2 Calculating the Mean and the Variance of the Variance

In a mixture of distributions with different variances, the kurtosis of the mixture depends directly on the variance of these variances. The choice of our mixture as a model for today's implied distribution, is constrained by the following equations (see Appendix F for the proof)

$$E(X) = 0 \tag{6.24}$$

$$Var(X) = E[\sigma^2] \tag{6.25}$$

$$Skewness(X) = 0 \tag{6.26}$$

$$K(X) = 3(1 + (\frac{S.D.(\sigma^2)}{E(\sigma^2)})^2) \tag{6.27}$$

The factor 3 in equation (6.27) is the kurtosis of the next time step model distribution, as represented by our conditional normal model. The abbreviation S.D. stands for the standard deviation[15].

Since, the chosen mixture is a model for the observed distribution, we choose $E(X^2) = E[\sigma^2] = \sigma_*^2$, where σ_*^2 is the variance of the observed implied distribution[16]. In addition, we choose the kurtosis of the mixture for X_T (model) to be less than the kurtosis of the observed implied distribution for S_T, K_*, i.e.

$$K_t[X] < K_{t,*} \tag{6.28}$$

Otherwise, we would be likely to generate conditional distributions for S_T which have negative excess kurtosis. This would contradict the empirical evidence which shows that implied volatility smiles are more pronounced for shorter maturity options (Taylor and Xu [141], Tompkins [143]), i.e. the kurtosis of the implied distribution increases as we approach the maturity of the option.

The inequality in equation (6.28) implies that

$$3(1 + (\frac{S.D.(\sigma^2)}{E(\sigma^2)})^2) < K_{t,*}$$

which can be written as[17]

$$S.D.(\sigma^2) < \sqrt{(\frac{K_{t,*}}{3} - 1)\sigma_*^2} \tag{6.29}$$

[15]Equation (6.27) shows that the chosen mixture has excess kurtosis. It also shows that the bigger the volatility of volatility, the higher is the kurtosis of the mixture.

[16]Since X is only a model for S, it would not have made a difference, if we had assumed that $E[X^2] = 1$. In that case, we would have constrained the simulated evolution of the implied distribution via the choise of the standard deviation of the variance.

[17]Hence, we express the standard deviation of the variance in terms of the kurtosis of the actual distribution. We have to restrict ourselves to implied distributions with positive kurtosis K_*, so as the square root to be defined.

6.5.3 Choosing a PDF for the Variance

We assume that the variance is distributed lognormally. This ensures that it can not take negative values. The variance is distributed with mean α, and variance β^2, i.e. $\sigma^2 \sim \Lambda(\alpha, \beta^2)$. Hence, $\ln \sigma^2 \sim N(\psi, v^2)$. Given that the lognormal distribution is the limit of the binomial distribution, as the time interval $\delta t \rightarrow 0$ (see Cox, Ross, Rubinstein [43]), we construct a binomial tree for the evolution of the variance. The tree has N time steps, which are $\delta t = \frac{T}{N}$ spaced apart. The terminal time step delivers to us the values of θ_i with their associated probabilities π_i.

In order to construct the tree, we choose the upward and downward movement factors for the evolution of σ^2, u, and d ($u > 1$, $d < 1$), in the same way as Cox, Ross and Rubinstein [43] do, i.e. $u = \frac{1}{d}$, with $u = e^{v\sqrt{\delta t}}$, and $d = e^{-v\sqrt{\delta t}}$. Appendix G shows that

$$v\sqrt{\delta t} = \sqrt{\frac{1}{N} \ln[1 + \frac{S.D.(\sigma^2)^2}{E(\sigma^2)^2}]} \tag{6.30}$$

Notice that the construction of the tree for the evolution of the variance, takes into account the information from the current implied PDF.

The transition probabilities p, are calculated by requiring that $E[\sigma^2] = \sigma_*^2$, i.e.

$$pu\sigma_*^2 + (1-p)d\sigma_*^2 = \sigma_*^2$$

which simplifies to

$$p = \frac{1-d}{u-d} \tag{6.31}$$

Once the tree has been constructed, we use the formula for the binomial distribution to calculate the probabilities π_i for reaching the values θ_i at the terminal level of the tree. This is

$$\pi_i = \Pr(\sigma^2 = \theta_i) = \frac{N!}{i!(N-i)!} p^i (1-p)^{N-i} \tag{6.32}$$

θ_i	π_i
9266.95	0.02
5487.04	0.1
3248.92	0.26
1923.71	0.34
1139.05	0.22
674.44	0.06

Table 6.1: Values for the Variance with their associated Probabilities.

for $i = 0, \ldots N$.

6.5.4 A Numerical Example

We give a numerical example for the construction of the binomial variance tree. We assume that the first four moments of the initial implied distribution are $E(S) = 340$, $Var(S) = 2500$, Skewness=1, and Kurtosis=8. Hence, $E[\sigma^2] = 2500$. Using equation (6.29), we find that $S.D.(Var) < 3227.49$. We choose $S.D.(Var) = 1600$. The tree is constructed with five time steps ($N = 5$), and spacing $\delta t = 0.2$. Hence, the time horizon is $T = 1$. From equation (6.30), we find $v\sqrt{\delta t} = 0.26$. Then, $u = 1.3$, $d = 0.769$, and $p = 0.435$ (equation (6.31)). Figure 6.11 shows the constructed tree for the evolution of the variance. Table 6.1 shows the values θ_i with their associated probabilities π_i.

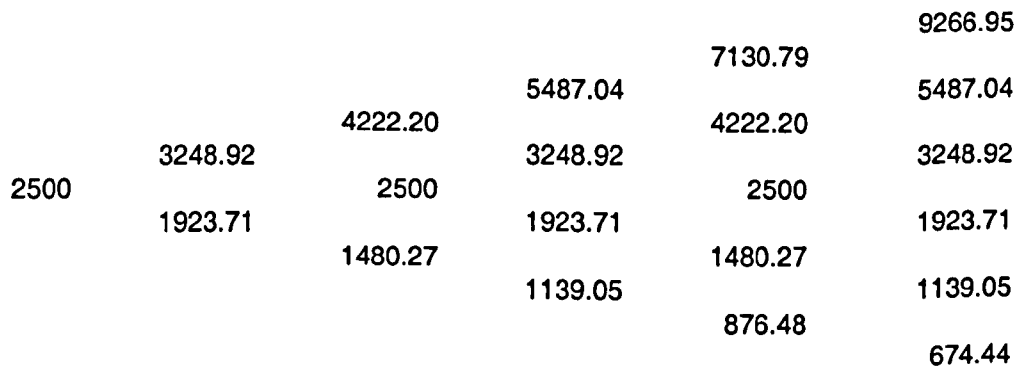


Figure 6.11: Constructed Tree for the Evolution of the Variance.

Figure 6.12 shows the mapping $Q \rightarrow P$ when the variance changes, for these

different values of θ_i. We can see that for very high θ_i the initial PDF accumulates less mass than the new simulated one, for the low values of the asset price, while it accumulates more mass for the high asset prices. On the other hand, for the low θ_i the reverse happens.

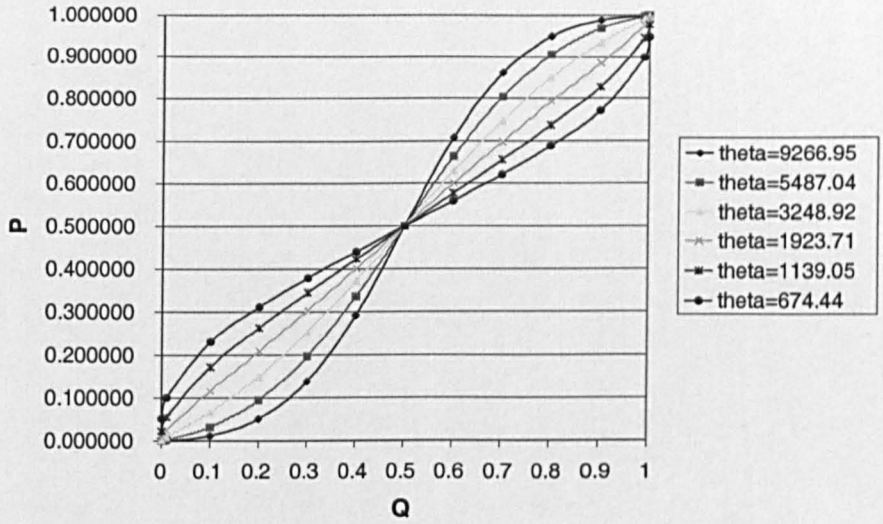


Figure 6.12: Mapping from Q to P for the different values of θ_i, when the Variance changes.

Figures 6.13-6.18 show the original and the simulated $f_t(S_T)$ across the θ_i which are shown in Table 6.1. The pictures are consistent with the above discussed implications of the mapping $Q \rightarrow P$.

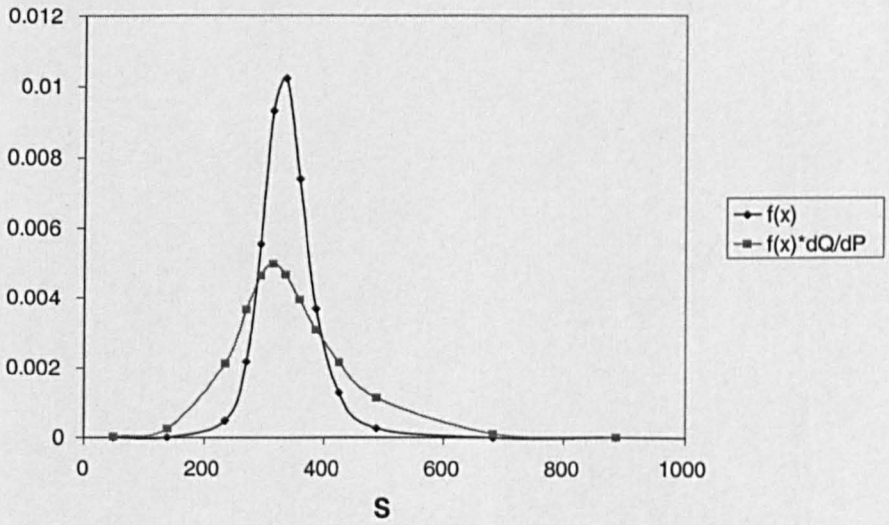


Figure 6.13: Initial and New PDF when the variance is 9266.95.

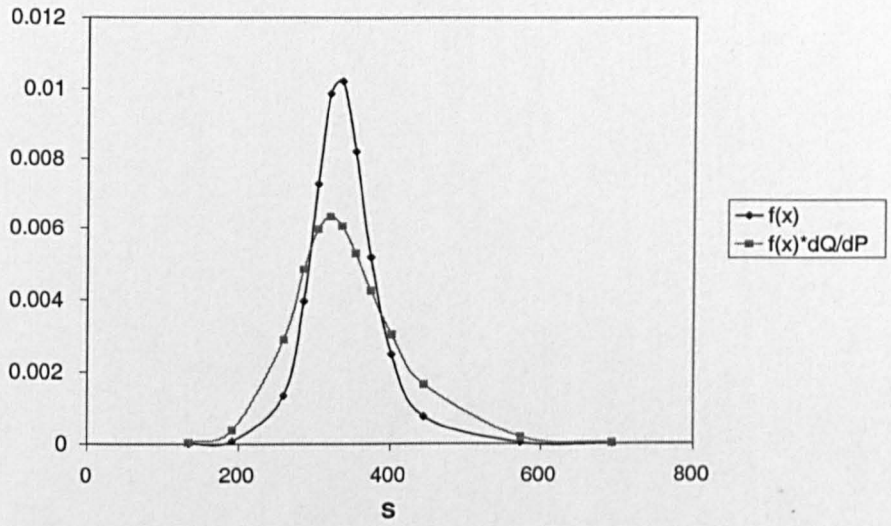


Figure 6.14: Initial and New PDF when the variance is 5487.04.

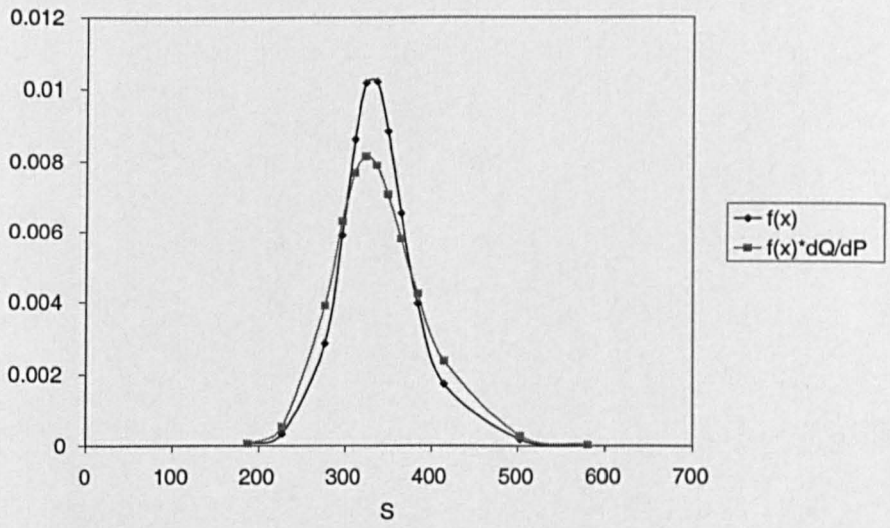


Figure 6.15: Initial and New PDF when the variance is 3248.92.

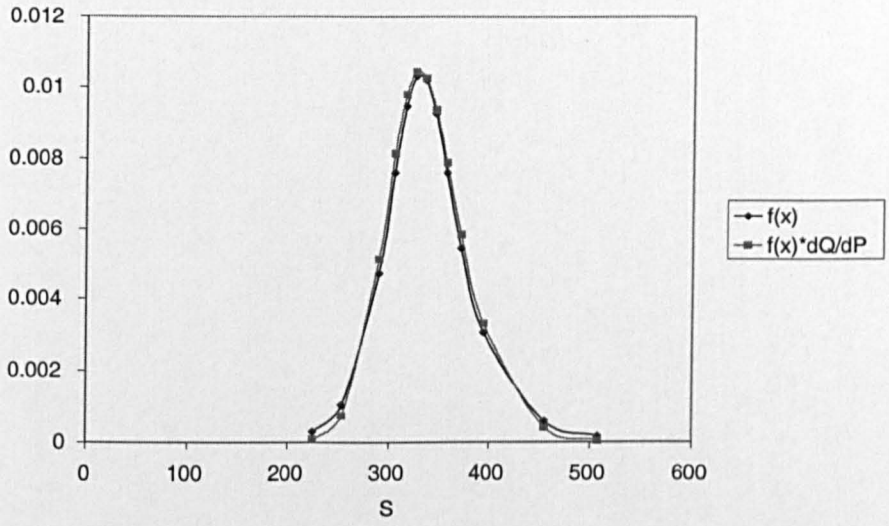


Figure 6.16: Initial and New PDF when the variance is 1923.71.

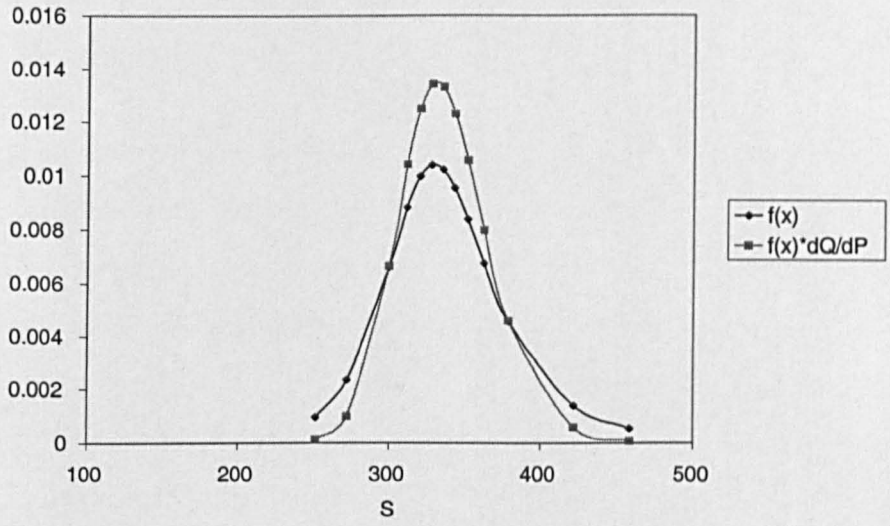


Figure 6.17: Initial and New PDF when the variance is 1139.05.

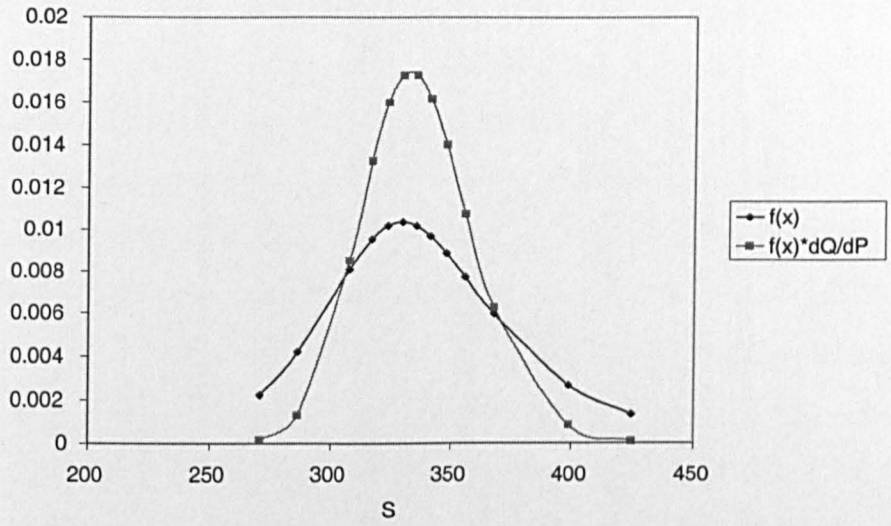


Figure 6.18: Initial and New PDF when the variance is 674.44

6.6 Conclusions and Issues for Further Research

We have presented a new method for smile consistent option pricing under stochastic volatility: the simulation of the implied risk-neutral probability density function (PDF). We have developed an algorithm for simulating the evolution of the implied PDF, by allowing for orthogonal shocks on its moments. The algorithm relies upon a postulated mixture of distributions. The mixture is used as a model for today's implied distribution. Then, a mapping between today's, and tomorrow's cumulative density function (CDF), for a given value of the asset price, is easily established.

We have dealt with the cases that the mean, or the variance changes, by assuming tractable, for the numerical implementation, mixtures. Some numerical examples are presented. Three are the main advantages of our algorithm. First, the mapping can be established through some very simple calculations. Second, all that the algorithm takes as input is today's implied PDF, which we can easily extract from the European option prices. Third, the algorithm can be extended to cases where more complicated specifications of the mixture are assumed.

These advantages make the simulation of the implied PDF a very promising

tool for option pricing. However, there is a number of issues that needs to be investigated by future research, in the context of our algorithm. First, in order to assess the performance of the algorithm, we need to test it empirically. Second, the performance of various specifications for the employed mixture should be examined, so that to develop the method into a fully operational tool. The dynamics of implied distributions should also be explored. This will facilitate the search for the mixture of distributions which will simulate the evolution of the distribution, so as to mimic the evolution of the empirical distribution. Future research should also generalize the method to allow for correlation between the shocks of the asset price and the volatility. Finally, the performance of hedges (constructed by a deterministic volatility model) should be investigated via our algorithm (see Section 7.2).

Chapter 7

Conclusions and Suggested Future Research

7.1 Conclusions

This study has contributed to the literature of "smile-consistent" no-arbitrage stochastic volatility models, in two ways. First, we have investigated the dynamics of implied volatilities of the S&P 500 Futures Options. This provides important insights into the evolution of traded option prices, and it is necessary for the implementation of such models. Second, we have presented a new method of simulating the evolution of the implied distribution.

We have investigated the dynamics of implied volatilities by answering three questions: (1) how many factors are needed to explain the dynamics of the implied volatility smiles and surfaces?, (2) what do these factors look like?, and (3) how are these factors correlated with the innovation in the underlying asset's process? The technique that we have used is Principal Components Analysis (PCA). It has been applied to time series of first differences of implied volatilities, when these are indexed in two different metrics: the strike, and the moneyness level. To avoid contamination from noisy data, the raw data were screened for data errors, and we excluded data likely to introduce errors into our volatility estimates.

In both metrics, two components are identified after examining a variety of

criteria. Next, we obtain a clear interpretation for them by constructing and applying to the retained components a Procrustes rotation. The first component can be interpreted as a parallel shift, and the second as a Z-shaped twist. The number of shocks and their interpretation, is the same for both the smile, and the surface analysis. In addition, the results were very consistent for the four separate years 1992, 1993, 1994, and 1995. Furthermore, for both the smile and the surface analysis, the first rotated PC is positively correlated with the underlying in the strike metric, and negatively correlated in the moneyness metric. The correlations for the rotated second PC are positive in both metrics. However, their magnitude varies randomly over the years.

Our research shows that three factors are required for the implementation of a "smile-consistent" no-arbitrage stochastic volatility model. One is needed for the underlying asset, and two more for the implied volatility. This conclusion may apply only to Futures Options on the S&P 500.

Finally, we have developed a new and general method for constructing a smile-consistent" no-arbitrage stochastic volatility model. This is the simulation of the implied risk-neutral distribution. The algorithm is based on an assumed mixture of distributions, and it is presented for the cases that the first two moments change over time. It can be implemented easily, and it can also be generalized to the cases that more complicated forms of the mixture are considered. Once the mixture has been specified, the algorithm only requires as input today's implied distribution. Our method can be used for pricing purposes, and it can assess the behavior of hedges. On the other hand, it can not be used for American type products.

7.2 Future Research

Both empirical and theoretical work of academic and practical interest can flow from the ideas introduced in this dissertation. Kamal and Derman [102] analyzed the dynamics of the implied volatility surface of over the counter (OTC) Index

Options by answering the same three questions by using PCA, as well. Their results suggest that a four factor stochastic volatility smile-consistent no-arbitrage model is an appropriate one. They find that three components explain about 95% of the variance of the volatility structure. Our work, identifies a model with one less factor which explains only about 53% (60%) of the surface variation in the strike (moneyness) metric.

These differences could be attributed to either one (or all) of the following: (a) the market futures option data have a noisier volatility structure than that estimated from the quotations for OTC index options, (b) Kamal and Derman have indexed their variables with the delta (delta metric), (c) the dynamics of the volatility surface may simply depend on the choice of the underlying asset. This is analogous to the documented fact that the magnitude of observed implied volatility smiles, depends on the underlying asset (Fung and Hsieh [70], Tompkins [143]). Future research should apply PCA to different data sets under different metrics, so as to solve these issues. We would expect that there should not be a difference in the results of the moneyness and the delta metric if the changes in the volatility are small and we control for the time to expiry (as we did by using expiry ranges).

The results from the application of the PCA under the strike and moneyness metrics, can be used to assess whether either the deterministic, or the stochastic volatility models (or both), are mis-specified. If a deterministic volatility model is the true model, then implieds should have a "small" variation in the strike metric, and a "large" variation in the moneyness metric. The reverse will happen if a stochastic volatility model is the correct description of the world. Simulating implied volatilities from different models, e.g. stochastic volatility, or jump diffusion models (or both), will provide a way to measure how small or large, the variation of the empirical implieds is in both metrics.

Another empirical question is whether portfolios of options modelled as riskless under a three, or four factor model, exhibit substantial market risk. This can be answered by looking at the out-of-sample hedging performance of smile-

consistent no-arbitrage stochastic volatility models. This brings up the next issue which is the extent to which this class of models really outperforms the "smile-consistent" no-arbitrage deterministic volatility models. The studies by Dumas, Fleming, and Whaley [58], and by Jackwerth and Rubinstein [95] are direct empirical tests of the out-of-sample effectiveness of deterministic volatility models. Dumas, Fleming and Whaley's results imply that the smile-consistent deterministic volatility models are mis-specified. Jackwerth and Rubinstein can not conclude whether stochastic volatility models provide a better description of the reality. However, these studies do not include the smile-consistent stochastic volatility models, so as to see whether the use of a more complex model is justified . Hence, a comparison of the out-of-sample pricing and hedging errors of the two types of models is necessary.

Next, the method for the simulation of the implied distribution should be tested with options data, and compared with alternative models (deterministic and stochastic volatility) for option pricing and hedging. Our method assumes that the shocks which affect the moments of the distribution are orthogonal, and therefore they can be studied sequentially. An issue which has to be investigated is whether the order that the shocks are applied, is important (we would expect that for instantaneous shocks, it is not). Moreover, the robustness of the method when we deviate from our postulated mixture of distributions, needs to be checked. Using a different mixture as a model for today's implied distribution is going to alter the simulated evolution of the probability density function. It would be interesting to investigate the cases where the conditional distribution employed in the mixture is not normal, but some kind of fat-tailed distribution. Also, the construction and implementation of the method when the shocks are correlated, should be explored.

A related topic to the implementation of our algorithm, is the study of the number of shocks which affect the implied distributions. This is both an empirical and theoretical issue. It will be interesting to see whether the number of shocks that drive implied distributions are different from those which affect im-

plied volatilities. Given that implied volatilities smiles (skews) reflect the shape of implied distributions, we would not expect to find significant differences. In addition, the empirical analysis of the dynamics of implied distributions will make apparent the kind of evolution that the simulated by our method distribution should mimic. This may facilitate the choice of the appropriate mixture of distributions.

Furthermore, the method for the simulation of the implied distribution should be extended to the case where we deal with options having different expiries, simultaneously. Then issues like the performance of static hedges which have been constructed according to a deterministic volatility tree, can be addressed. If a deterministic volatility model is the correct model, then the replication of a barrier option for example, will be exact along the barrier and at the terminal time T. However, if a stochastic volatility model is the correct model, then hedging errors will be incurred. We can examine the magnitude of these errors by simulating the asset price paths via a modification of our algorithm and calculating the value of the portfolio with which we want to hedge the barrier option.

A statistical problem in the application of the PCA, is the testing of the significance of the (correlation) loadings of the principal components. Kamal and Derman [102] find factors which have a different interpretation from ours. However, the interpretation of the factors could be an artifact; it should be done after considering the statistical significance of the loadings. The existing tests assume a multivariate normal distribution (Basilevsky [9]). Given, that this distributional assumption is too restrictive, non-parametric tests should be developed.

Since observed smiles can not be explained just by introducing a stochastic volatility process, but a jump process should be included, as well (Das and Sundaram [47]), researchers should look at how to create a "smile-consistent" no-arbitrage stochastic volatility-jump model. Pappalardo [118] provides a method to recover the volatility coefficient of a jump process from the prices of European options. His methodology is analogous to Dupire's [60], [62]. He assumes that the volatility is a deterministic function of the asset price and time, and he derives

the forward Kolmogorov equation for the jump process. He is able to reveal the implied volatility coefficient for certain distributions of the jump size. However, the problem of estimating the other parameters of the jump process, i.e. the probability that the jump occurs, and the parameters of the distribution for the jump size remains.

Future research should also investigate whether implied volatility smiles can be explained by transaction costs. Jackwerth and Rubinstein [94] point out that a potential trading-cost theory of the smile needs to explain why, given the extreme shift in the option smile, these costs were apparently of much less importance before the 1987 crash than after. We are aware of only Constandinides [38] transaction costs model which implies that transaction costs can not account for the volatility smile. However, the extent to which implied volatility patterns are affected by transaction costs, should be explored further.

A broader research topic which is promising is the measurement of the "degree" of market incompleteness (Bertsimas, Kogan, Lo [17]). If the market is incomplete, then the payoff of a complex security (e.g. a contingent claim) can not be replicated exactly by a self-financing dynamic portfolio strategy of simpler securities. In that case, the price of the complex security is not unique, but it varies within some bounds (see for example, Hodges [87]). The less incomplete the market is, the tighter the bounds will be. This will enable us to have an idea of whether the complex security is "cheap", or "dear".

Appendix A

Explained Communalities in the Smile Analysis in the Strike Metric

In this section, we show the communalities explained by one and two PCs in the smile analysis in the strike metric. The results are reported across the separate expiry buckets for the year 1994[1].

[1] X_1 denotes the differenced once implied volatilities corresponding to the strike level X.

Variables	30-10	60-30	90-60	150-90	240-150	360-240
400_1			64.83	46.98	55.82	
405_1						
410_1			52.93	60.38	69.84	
415_1						
420_1		54.55	52.77	51.98	70.97	91.46
425_1		34.11		65.64		
430_1		45.59		65.68	75.32	91.71
435_1	45.24	36.04	61.77	68.65		
440_1	61.93	26.05	36.38	56.43	75.72	90.37
445_1	57.42	36.01	59.12	67.98		
450_1	59.06	0.8	12.12	39.35	48.18	83.04
455_1	65.16	1.43	42.59	1.16		
460_1	57.41	3.14	42.58	0	0.71	3.21
465_1	53.48	24.81	8.02	8.57		
470_1	47.9	31.81	4.47	0.98	7.53	2.22
475_1		34.66	20.27	34.52		
480_1		35.10	21.82	54.41	59.68	
485_1						
490_1				46.99	70.07	
495_1						
500_1				63.48	73.77	86.37
505_1						
510_1				58.92	73.4	82.05

Table A.1: Smile Analysis on the Strike Metric: Communalities Explained by One Principal Component for 1994.

Variables	30-10	60-30	90-60	150-90	240-150	360-240
400_1			65.16	52.23	56.46	
405_1						
410_1			59.45	61.01	96.64	
415_1						
420_1		61.41	54.84	55.79	71.53	91.46
425_1		55.95		68.65		
430_1		59.01		65.74	77.12	91.74
435_1	69.25	51.79	62.86	68.73		
440_1	67.74	34.57	36.38	58.76	78.01	90.45
445_1	64.46	56.46	60.71	68.05		
450_1	63.74	42.35	56.69	47.55	53.04	85.15
455_1	65.3	28.68	65.35	68.09		
460_1	58.55	46.76	54.28	65.56	74.32	77.48
465_1	86.02	47.24	36.74	49.57		
470_1	84.44	70.24	40.27	49.85	49.83	72.52
475_1		60.86	58.18	34.93		
480_1		52.69	59.27	54.46	61.11	
485_1						
490_1				51.22	71.22	
495_1						
500_1				64.31	74.13	86.82
505_1						
510_1				59.18	74.59	83.08

Table A.2: Smile Analysis on the Strike Metric: Communalities Explained by Two Principal Components for 1994.

Appendix B

Construction of the Procrustes Rotation Method

Denote by $\mathbf{g}_1, \mathbf{g}_2$, the $(p \times 1)$ vectors of rotated loadings of the first and second retained PCs, respectively. Let the transformation matrix $\mathbf{T} = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix}$. Then, equation (4.13) delivers the rotated loadings as linear combinations of the unrotated ones, i.e.

$$\mathbf{g}_1 = a_1\mathbf{p}_1 + b_1\mathbf{p}_2 \tag{B.1}$$

$$\mathbf{g}_2 = a_2\mathbf{p}_1 + b_2\mathbf{p}_2 \tag{B.2}$$

where $\mathbf{p}_i$ is the $(p \times 1)$ vector of loadings of each component, for $i = 1, 2$. The inner product $< \mathbf{g}_1, \mathbf{g}_2 >$ is given by

$$< \mathbf{g}_1, \mathbf{g}_2 > = < a_1\mathbf{p}_1 + b_1\mathbf{p}_2, a_2\mathbf{p}_1 + b_2\mathbf{p}_2 > = \tag{B.3}$$

$$a_1a_2 < \mathbf{p}_1, \mathbf{p}_1. > + a_1b_2 < \mathbf{p}_1, \mathbf{p}_2 > + b_1a_2 < \mathbf{p}_2, \mathbf{p}_1 > + b_1b_2 < \mathbf{p}_2, \mathbf{p}_2 >$$

Given that in the construction of the PCA the normalization constraint $\mathbf{P}'\mathbf{P} = \mathbf{I}$ has been imposed, we have that

$$< \mathbf{p}_1, \mathbf{p}_1 >=< \mathbf{p}_2, \mathbf{p}_2 >= 1 \tag{B.4}$$

Also we have that

$$< \mathbf{p}_1, \mathbf{p}_2 >=< \mathbf{p}_2, \mathbf{p}_1 >= 0 \tag{B.5}$$

(Basilevsky [9], theorem 3.1, page 103). Hence, from equations (B.4) and (B.5), equation (B.3) becomes

$$< \mathbf{g}_1, \mathbf{g}_2 >= a_1 a_2 + b_1 b_2 \tag{B.6}$$

In the case of an orthogonal rotation, rotated eigenvectors remain orthogonal (Basilevsky [9])[1]. Thus, $< \mathbf{g}_1, \mathbf{g}_2 >= 0$, and from equation (B.6) we have that

$$a_1 a_2 + b_1 b_2 = 0 \tag{B.7}$$

The idea of a Procrustes type rotation is to choose the coefficients with which we form the linear combinations, so as to achieve a pre-defined interpretation. We choose them, so as to get the parallel shift for the first PC; our method does not guarantee a priori the shape of the second PC.

A natural way of getting the parallel shift character for the first PC, is to do an Ordinary Least Squares (OLS) regression of a vector of constants on $\mathbf{p}_1$,$\mathbf{p}_2$. This regression will deliver to us a_1 and b_1. In addition, we standardize the obtained a_1, and b_1, i.e. $a_1^* = \dfrac{a_1}{\sqrt{a_1^2 + b_1^2}}$ and $b_1^* = \dfrac{b_1}{\sqrt{a_1^2 + b_1^2}}$, so that $a_1^{*2} + b_1^{*2} = 1$. We will explain in a while the reason for this standardization.

In order to calculate a_2 and b_2 from equation (B.7), we need one more equa-

[1]Notice that in order to do an orthogonal rotation in such a way, so as to keep the resulting rotated eigenvectors orthogonal, *we must not* use the correlation loadings of the matrix **A**. This is because rotated correlation loadings are not orthogonal (see Basilevsky [9] theorem 5.3). Consequently, we have to work with the original loadings which come from unstandardized components and/or variables.

tion. We get this equation by imposing the constraint that

$$a_2^2 + b_2^2 = 1 \tag{B.8}$$

The constraint given by equation (B.8) has not been chosen arbitrarily, but it ensures, combined with the standardization of a_1 and b_1, that the cumulative variance explained by the unrotated PCs equals the explained variance by the rotated PCs.

Proposition 4 *A sufficient condition that the cumulative variance of the unrotated PCs is equal to this of the rotated is that* $a_1^{*2} + b_1^{*2} = 1$ *and* $a_2^2 + b_2^2 = 1$.

Proof.

Let $\mathbf{z}_i$, and $\mathbf{v}_i$ be the $(T \times 1)$ vectors of the unrotated and rotated PCs, respectively, for $i = 1, 2$. Then, from equation (4.12) we have that

$$\begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{z}_1 & \mathbf{z}_2 \end{pmatrix} \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} = \begin{pmatrix} \mathbf{z}_1 a_1 + \mathbf{z}_2 b_1 & \mathbf{z}_1 a_2 + \mathbf{z}_2 b_2 \end{pmatrix}$$

Since, by the construction of the PCA, the unrotated PCs are orthogonal, we get that the variance of $\mathbf{v}_i$ is given by

$$Var(\mathbf{v}_1) = Var(\mathbf{z}_1)a_1^2 + Var(\mathbf{z}_2)b_1^2 \tag{B.9}$$

$$Var(\mathbf{v}_2) = Var(\mathbf{z}_1)a_2^2 + Var(\mathbf{z}_2)b_2^2 \tag{B.10}$$

Since $Var(\mathbf{z}_i) = l_i$, the cumulative variance of the unrotated PCs will equal the cumulative variance of the rotated PCs, if

$$(a_1^2 + a_2^2)l_1 + (b_1^2 + b_2^2)l_2 = l_1 + l_2 \tag{B.11}$$

Equation (B.11) is true if $a_1^{*2} + b_1^{*2} = 1$ and $a_2^2 + b_2^2 = 1$.

This completes the proof.

■

Solving equations (B.7) and (B.8) we get

$$a_2 = \pm\sqrt{b_1^{*2}} \tag{B.12}$$

$$b_2 = \pm\sqrt{a_1^{*2}} \tag{B.13}$$

Note that a_2, and b_2 are chosen, so that the condition which results from equation (B.7), is respected, i.e.

$$b_2 = \frac{-a_1 a_2}{b_1} \tag{B.14}$$

Appendix C

Explained Communalities in the Smile Analysis in the Moneyness Metric

In this section, we show the communalities explained by one and two PCs in the smile analysis in the moneyness metric. The results are reported across the separate expiry buckets for the year 1994[1].

[1]P (or N) X_1 denotes the differenced once implied volatilities corresponding to the plus (minus) moneyness level X.

Variables	30-10	60-30	90-60	150-90	240-150	360-240
N0.15_1						
N0.125_1						
N0.11_1						87.48
N0.10_1					58.75	
N0.09_1		36.22	57.98	71.35		
N0.088_1						88.71
N0.075_1		32.04	77.06	72.00	71.79	
N0.072_1						
N0.066_1						89.73
N0.06_1	78.31	39.24	68.84	70.31		
N0.05_1					74.72	
N0.048_1						
N0.045_1	86.50	61.18	76.83	74.81		
N0.044_1						89.57
N0.03_1	91.82	68.52	57.97	68.02		
N0.025_1					69.05	
N0.024_1						
N0.022_1						88.29
N0.015_1	90.85	74.44	67.43	60.86		
P0.0_1	86.82	68.57	26.97	12.96	5.49	
P0.015_1	87.19	51.11	5.44	28.91		
P0.024_1						
P0.025_1					78.00	
P0.03_1	78.44	50.53	2.81	28.74		
P0.045_1		54.43	1.08			
P0.048_1						
P0.05_1					75.52	
P0.066_1						76.45
P0.072_1						
P0.075_1					73.70	
P0.088_1						75.57

Table C.1: Communalities Explained by One Principal Component on the Moneyness Metric for 1994.

Variables	30-10	60-30	90-60	150-90	240-150	360-240
N0.15_1						
N0.125_1						
N0.11_1						91.82
N0.10_1					72.27	
N0.09_1		65.81	60.49	71.37		
N0.088_1						93.67
N0.075_1		74.23	77.88	72.48	77.79	
N0.072_1						
N0.066_1						92.85
N0.06_1	91.76	65.77	70.39	72.64		
N0.05_1					81.32	
N0.048_1						
N0.045_1	91.96	78.26	78.88	76.70		
N0.044_1						91.62
N0.03_1	92.43	71.26	70.20	75.14		
N0.025_1					81.58	
N0.024_1						
N0.022_1						88.82
N0.015_1	91.97	74.59	67.60	64.86		
P0.0_1	88.63	78.98	34.93	49.80	72.87	
P0.015_1	94.38	89.22	91.41	83.88		
P0.024_1						
P0.025_1					87.57	
P0.03_1	93.36	86.36	93.67	83.49		
P0.045_1		80.63	93.08			
P0.048_1						
P0.05_1					86.93	
P0.066_1						95.58
P0.072_1						
P0.075_1					81.95	
P0.088_1						96.07

Table C.2: Communalities Explained by Two Principal Components on the Moneyness Metric for 1994.

Appendix D

Explained Communalities in the Surface Analysis in the Strike Metric

In this section, we show the communalities explained by the first three PCs in the surface analysis in the strike metric. The results are reported for the years 1994[1].

[1]XB.A_1 denotes the differenced once implied volatilities corresponding to the strike level X, in the range B-A; B, A are the first digits of the three ranges that we examine.

Variables for 94	1 PC	2 PCs	3 PCs
4001.9_1	43.28	43.38	43.56
4002.1_1	61.49	61.89	62.09
4101.9_1	58.35	58.96	59.01
4102.1_1	72.09	73.08	74.04
4201.9_1	52.36	54.76	54.77
4202.1_1	73.96	74.74	75.66
4301.9_1	52.90	54.25	56.24
4302.1_1	80.62	81.18	82.07
4359.10_1	5.36	9.9	41.46
4401.9_1	52.9	55.8	60.28
4402.1_1	78.32	78.66	78.8
4409.10_1	9.16	9.21	60.76
4459.10_1	5.4	48.1	70.05
4501.9_1	30.98	31.88	32.4
4502.1_1	64.56	64.61	64.95
4509.10_1	0.58	45.39	68.21
4559.10_1	0.19	56.92	57.68
4601.9_1	1.5	28.38	58.63
4602.1_1	2.08	23.45	49.52
4609.10_1	4.2	55.49	55.51
4659.10_1	2.05	7.3	11.47
4701.9_1	11.59	20.11	31.99
4702.1_1	2.34	6.8	18.88
4709.10_1	0.8	29.79	30.27
4801.9_1	59.53	60.35	60.38
4802.1_1	62.19	62.63	62.9
4901.9_1	59.92	61.18	61.6
4902.1_1	72.97	73.94	76.24
5002.1_1	79.64	79.64	80.05
5009.10_1	65.13	65.48	66.2
51101.9_1	63.77	63.9	64.39
5102.1_1	75.47	75.65	77.26

Table D.1: Surface Analysis on the Strike Metric: Communalities Explained by Three Principal Components for 1994.

Appendix E

Explained Communalities in the Surface Analysis in the Moneyness Metric

In this section, we show the communalities explained by the first two PCs in the surface analysis in the moneyness metric. The results are reported for the years 1994[1].

[1]P (or N) X.Y_A_1 denotes the differenced once implied volatilities corresponding to the plus (or minus) moneyness level X.Y%, for the expiry range with upper limit A.

Variables for 94	1 PC	2 PCs
N7.2_9_1	12.64	62.59
N6.0_9_1	14.14	70.65
N4.8_9_1	16.84	61.96
N3.6_9_1	19.83	60.13
N1.2_9_1	24.99	80.17
P0.0_9_1	22.71	73.09
P1.2_9_1	3.66	28.39
P2.4_9_1	5.2	32.8
P3.6_9_1	5.7	33.63
N8.8_1_1	46.18	61.25
N6.6_1_1	24.11	39.05
N4.4_1_1	38.04	57.48
N2.2_1_1	38.37	48.81
P0.0_1_1	3.62	4.71
P2.2_1_1	64.63	64.71
N11_2_1	67.76	76.2
N8.8_2_1	74.35	85.9
N6.6_2_1	73.73	83.6
N4.4_2_1	68.78	76.88
N2.2_2_1	70.44	74.47
P0.0_2_1	0.954	1.225
P2.2_2_1	0.8583	85.83
P4.4_2_1	0.80494	80.49
P6.6_2_1	0.77376	77.49
P8.8_2_1	0.75193	75.22

Table E.1: Surface Analysis on the Moneyness Metric: Communalities Explained by One and Two Principal Components for 1994.

Appendix F

The Moments of a Mixture of (Normal) Distributions, Mixed across the Variance

We calculate the first four moments of the asset's price distribution for a mixture of a general zero mean distribution, mixed across the variance, i.e.

$$m(X) = \sum_{i=1}^{n} \pi_i f(X \mid 0, \theta_i)$$

where $\sigma_i^2 = \theta_i$. The mean is

$$E(X) = \int Sm(X)dS = \int X \sum_{i=1}^{n} \pi_i f(X \mid 0, \theta_i)dX =$$

$$\sum_{i=1}^{n} \pi_i \int Xf(X \mid 0, \theta_i)dS = \sum_{i=1}^{n} \pi_i E^C(X) = 0$$

where the superscript C, denotes the expectation of the conditional distribution. The variance is

$$Var(X) = E(X^2) = \int X^2 m(X)dX =$$

$$\int X^2 \sum_{i=1}^{n} \pi_i f(X \mid 0, \theta_i)dX = \sum_{i=1}^{n} \pi_i \int X^2 f(X \mid 0, \theta_i)dX$$

$$= \sum_{i=1}^{n} \pi_i E^C(X^2) = \sum_{i=1}^{n} \pi_i \theta_i = E[\sigma^2] \qquad \text{(F.1)}$$

The skewness is

$$Skewness(X) \equiv \frac{E(X^3)}{[Var(X)]^{\frac{3}{2}}} = \frac{\int X^3 m(X) dX}{[Var(X)]^{\frac{3}{2}}} = \frac{\int X^3 \sum_{i=1}^{n} \pi_i f(X \mid 0, \theta_i) dX}{[Var(X)]^{\frac{3}{2}}} =$$

$$\frac{\sum_{i=1}^{n} \pi_i \int X^3 f(X \mid 0, \theta_i) dX}{[Var(X)]^{\frac{3}{2}}} = \frac{\sum_{i=1}^{n} \pi_i E^C(X^3)}{[Var(X)]^{\frac{3}{2}}} = 0 \qquad \text{(F.2)}$$

The last step follows if $E^C(X^3) = 0$ (which is the case for a normal distribution).

Using the definition of the Kurtosis and equation (F.1) we find that

$$Kurtosis(X) = K(X) = \frac{E(X^4)}{[Var(X)]^2} = \frac{\int X^4 m(X) dX}{(E[\sigma^2])^2} =$$

$$\frac{\int X^4 \sum_{i=1}^{n} \pi_i f(X \mid 0, \theta_i) dX}{(E[\sigma^2])^2} = \frac{\sum_{i=1}^{n} \pi_i \int X^4 f(X \mid 0, \theta_i) dX}{(E[\sigma^2])^2} = \frac{\sum_{i=1}^{n} \pi_i E^C(X^4)}{(E[\sigma^2])^2}$$

Let the kurtosis of the conditional distribution be

$$K^C(X) = K = \frac{E^C(X^4)}{[Var^C(X)]^2}$$

and rearranging gives

$$E^C(X^4) = K\sigma^4$$

We also have that

$$Var(\sigma^2) = E[\sigma^4] - [E(\sigma^2)]^2$$

Hence,

$$K^M(X) = \frac{K \sum \pi_i \sigma_i^4}{(E[\sigma^2])^2} = K \frac{E[\sigma^4]}{(E[\sigma^2])^2} = K \frac{Var(\sigma^2) + [E(\sigma^2)]^2}{(E[\sigma^2])^2} = K(1 + [\frac{\sqrt{Var(\sigma^2)}}{E[\sigma^2]}]^2)$$

If the conditional distribution is normal, then $K = 3$.

Appendix G

Constructing the Binomial Tree for the Evolution of the Variance

We calculate $v\sqrt{\delta t}$ by using the properties of the lognormal distribution . The mean and the variance of the lognormal distribution are given by (see Aitchison and Brown [1])

$$\alpha = e^{\psi + \frac{1}{2}v^2} \tag{G.1}$$

$$\beta^2 = e^{2\psi + v^2}(e^{v^2} - 1) \tag{G.2}$$

Rearranging expressions (G.1), and (G.2) we can calculate v^2, i.e.

$$v^2 = \log(\frac{\beta^2}{\alpha^2} + 1) \tag{G.3}$$

Since σ^2 is distributed lognormally, $\mathrm{Var}(\ln \sigma_T^2 \mid \sigma_0^2) = v^2 T$. Therefore,

$$v^2 T = \ln[1 + \frac{S.D.(\sigma^2)^2}{E(\sigma^2)^2}] \tag{G.4}$$

The standard deviation of $\ln \sigma_T^2$ per step is

$$v\sqrt{\delta t} = \sqrt{\frac{v^2 T}{N}} \tag{G.5}$$

From equations (G.4), and (G.5) we get that

$$v\sqrt{\delta t} = \sqrt{\frac{1}{N}\ln[1 + \frac{S.D.(\sigma^2)^2}{E(\sigma^2)^2}]} \tag{G.6}$$

Bibliography

[1] Aitchison, J., and Brown, J. (1981): *The Lognormal Distribution with special reference to its uses in economics*, Cambridge University Press.

[2] Andersen, L. (1996): "The Equity Option Volatility Smile: an Implicit Finite Difference Approach", Five Essays on the Pricing of Contingent Claims, Unpublished Ph.D. Dissertation, The Aarhus School of Business.

[3] Anderson, A., Basilevsky, A., and Hum A. (1983): "Missing Data: A Review of the Literature". In Rossi, P., Wright, J., and Anderson, A., *Handbook of Survey Research*, Academic Press.

[4] Andreasen, J. (1996): "Implied Modelling: Stable Implementation, Hedging and Duality", Working Paper, The Aarhus School of Business.

[5] Bahra, B. (1997):"Implied Risk-Neutral Probability Density Functions From Option Prices: Theory and Application", Bank of England Working Paper Series.

[6] Bakshi, G., Cao, C., and Chen, Z. (1997): "Empirical Performance of Alternative Option Pricing Models", *Journal of Finance* 52, pp. 2003-2049.

[7] Barle, S. and Cakici, N. (1995): "Growing a Smiling Tree", *Risk* 8, pp. 76-81.

[8] Barone-Adesi, G., and R. Whaley (1987): "Efficient Analytic Approximation of American Option Values", *Journal of Finance* 42, pp. 301-320.

[9] Basilevsky, A. (1994): *Statistical Factor Analysis and Related Methods, Theory and Applications*, Wiley Series in Probability and Mathematical Statistics.

[10] Bates, D. (1988): "Pricing Options under Jump-Diffusion Processes", Working Paper, Rodney L. White Center.

[11] Bates, D. (1996): "Dollar Jump Fears, 1984-1992: Distributional Abnormalities Implicit in Currency Futures Options", *Journal of International Money and Finance* 15, pp. 65-93.

[12] Bates, D. (1996): "Testing Option Pricing Models", In Maddala, G and C. Rao (eds.), *Handbook of Statistics*, Vol. 14, Elsevier Science B.V.

[13] Bates, D. (1996). "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options", *Review of Financial Studies* 9, pp. 69-107.

[14] Bates, D. (1997): "Post-87 Crash Fears in the S&P 500 Futures Option Market", *Journal of Econometrics*, Forthcoming.

[15] Bates, D. (1998): "Hedging the Smirk", Working Paper, University of Iowa.

[16] Bates, P., and Clewlow, L. (1992): "Testing for Overreaction in Short Sterling Options". In Hodges, S., *Options, Recent Advances in Theory and Practice* 2, Manchester University Press.

[17] Bertsimas, D., Kogan, L., and Lo, A. (1997): "Pricing and Hedging Derivative Securities in Incomplete Markets: An ϵ-Arbitrage Approach", Working Paper, Massachusetts Institute of Technology.

[18] Black, F. (1975): "Fact and Fantacy in the Use of Options", *Financial Analysts Journal* 31, pp. 36-72.

[19] Black, F. (1976): "The Pricing of Commodity Contracts", *Journal of Financial Economics* 3, pp. 167-179.

[20] Black, F., and Scholes, M. (1973): "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy* 81, pp. 637-654.

[21] Blattberg, R., and Gonedes, N. (1974): "A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices", *Journal of Business* 47, pp. 274-80.

[22] Bick, A., and Reisman, H. (1993): "Generalized Implied Volatility", Working Paper, Simon Fraser University, Canada.

[23] Bookstaber, R., and McDonald, J. (1987): "A General Distribution for Describing Security Price Returns", *Journal of Business* 60, pp. 401-424.

[24] Boyle, P. (1977): "Options: A Monte Carlo Approach", *Journal of Financial Economics* 4, pp. 323-338.

[25] Breeden, D., and Litzenberger, R. (1978): "Prices of State-Contingent Claims Implicit in Option prices", *Journal of Business* 51, pp. 621-651.

[26] Brenner, M., and Galai, D. (1987): "On the Prediction of the Implied Standard Deviation", *Advances in Futures and Options Research* 2, pp. 167-177.

[27] Britten-Jones, M., and Neuberger, A. (1998): "Option Prices, Implied Price Processes, and Stochastic Volatility", Working Paper, London Business School.

[28] Buraschi, A., and Jackwerth, J. (1998): "Explaining Option Prices: Deterministic vs. Stochastic Models", Working Paper, London Business School.

[29] Cambell, J., Lo, A., and Mackinlay, A. (1997): *The Econometrics of Financial Markets*, Princeton University Press.

[30] Canina, L., and Figlewski, S. (1993): "The Informational Content of Implied Volatility", *Review of Financial Studies* 6, pp. 659-681.

[31] Carr, P., Ellis, K., and Gupta, V. (1998): "Static Hedging of Exotic Options", *Journal of Finance* 53, pp. 1165-1190.

[32] Chapra, S., and Canale, A. (1998): *Numerical Methods for Engineers: With Programming and Software Applications*, Mc Graw Hill, Third Edition.

[33] Christie, A. (1982): "The Stochastic Behavior of Common Stock Variances, Value, Leverage and Interest Rate Effects", *Journal of Financial Economics* 10, pp. 407-432.

[34] Clarke, G., and Cooke, D. (1983): *A Basic Course in Statistics*, Edward Arnold, 2nd Edition.

[35] Clewlow, L., and Strickland, C. (1998): *Implementing Derivatives Models*, John Wiley and Sons.

[36] Clewlow, L., and Xu, X. (1992): "A Review of Option Pricing with Stochastic Volatility", Working Paper, Financial Options Research Centre, University of Warwick.

[37] Clewlow, L., and Xu, X. (1994): "The Dynamics of Stochastic Volatility", Working Paper, Financial Options Research Centre, University of Warwick.

[38] Constandinides, G. (1998): "Transaction Costs and the Volatility Implied by Option Prices", Working Paper, University of Chicago.

[39] Corrado, C., and Su, T. (1997): "Implied Volatility Skews and Stock Return Skewness and Kurtosis implied by Stock Option Prices", *European Journal of Finance* 3, 73-85.

[40] Cox, D., and Miller, H. (1965): *Theory of Stochastic Processes*, Chapman and Hall.

[41] Cox, J. (1996): "The Constant Elasticity of Variance Option Pricing Model", *The Journal of Portfolio Management* 22 (Special Issue), pp. 15-17.

[42] Cox, J., and Ross, S. (1976): "The Valuation of Options for Alternative Stochastic Processes", *Journal of Financial Economics* 3, pp. 145-166.

[43] Cox, J., Ross, S., and Rubinstein, M. (1979): "Option Pricing: A Simplified Approach", *Journal of Financial Economics* 7, pp. 229-263.

[44] Cox, J, and Rubinstein, M. (1985): *Options Markets*, Prentice-Hall, Englewood Cliffs, N.J.

[45] Cox, J., Ingersoll, J., and Ross, S. (1985): "A Theory of the Term Structure of Interest Rates", *Econometrica* 53, pp. 385-407.

[46] Day, T., and Lewis, C. (1988): "The Behavior of Volatility Implicit in the Prices of Stock Index Options", *Journal of Financial Economics* 22, 103-122.

[47] Das, S, and Sundaram, R. (1998): "Of Smiles and Smirks: A Term-Structure Perspective", *Journal of Financial and Quantitative Analysis*, forthcoming.

[48] Davydov, D., and Linetsky, V. (1999): "Pricing Options on One-Dimensional Diffusions: A Unified Approach", Working Paper, University of Michigan.

[49] Derman, E., and Kani, I (1994): "Riding on a Smile", *Risk* 7, pp. 32-39.

[50] Derman, E., Ergener, D., and Kani, I. (1995): "Static Options Replication", *Journal of Derivatives* 2, pp. 78-95.

[51] Derman, E., Kani, I., Zou, J. (1996): "The Local Volatility Surface: Unlocking the Information in Index Option Prices", *Financial Analysts Journal* 52, pp. 25-36.

[52] Derman, E., Kamal, M., Kani, I., McClure, J., Pirasteh, C., Zou, J. (1996): "Investing in Volatility", Goldman Sachs Quantitative Strategies Research Notes.

[53] Derman, E., Kani, I. and Chriss, N. (1996): "Implied Trinomial Trees of the Volatility Smile", *The Journal of Derivatives* 3, pp. 7-22.

[54] Derman, E., M., Kani, I. (1997): "Stochastic Implied Trees: Arbitrage Pricing With Stochastic Term and Strike Structure of Volatility", *International Journal of Theoretical and Applied Finance* 1, pp. 61-110.

[55] Diz, F., and Finucane, T. (1991): "The Time Series Properties of Implied Volatility of S&P 100 Index Options", Working Paper, School of Management, Syracuse University.

[56] Diz, F., and Finucane, T. (1993): "Do the Options Markets Really Overreact?", *Journal of Futures Markets* 13, pp. 299-312.

[57] Dothan, M. (1990): *Prices in Financial Markets*, Oxford University Press.

[58] Dumas, B., Fleming, J. and Whaley, R. (1998): "Implied Volatility functions: Empirical Tests", *Journal of Finance* 53, pp. 2059-2106.

[59] Dupire, B. (1992): "Arbitrage Pricing with Stochastic Volatility", *Proceedings of French Finance Association Meetings, Paris 1992.*

[60] Dupire, B. (1993): "Pricing and Hedging with Smiles", *Proceedings of French Finance Association Meetings, La Baule 1993.*

[61] Dupire, B. (1993): "Model Art", *Risk* 6, pp. 118-121.

[62] Dupire, B. (1994): "Pricing with a Smile", *Risk* 7, pp. 18-20.

[63] Dupire, B. (1994): "A Unified Theory of Volatility", Working Paper, Paribas Capital Markets.

[64] Eastment, H., and Krzanowski, W. (1982): "Cross Validatory Choice of the Number of Components from a Principal Component Analysis", *Technometrics* 24, pp. 73-77.

[65] Emanuel, D., and MacBeth J. (1982): "Further Results on the Constant Elasticity of Variance Option Pricing Model", *Journal of Financial and Quantitative Analysis* 17, pp. 533-554.

[66] Engle, R., and Mezrich, J. (1996): "How to Model Correlation; GARCH for Groups", Working Paper, Salomon Brothers.

[67] Fama, E. (1965): "The Behaviour of Stock Market Prices", *Journal of Business* 38, pp. 34-105.

[68] Frachot, A., Jansi, D., and Lacoste, V. (1992): "Factor Analysis of the Term Structure: A Probabilistic Approach", NER, Bank of France.

[69] Franks, J., and Schwartz, E. (1991): "The Stochastic Behaviour of Market Variance Implied in the Prices of Index Options", *Economic Journal* 101, pp. 1460-1475.

[70] Fung, K., and Hsieh, D. (1991): "Empirical Analysis of Implied Volatility: Stocks, Bonds and Currencies", Working Paper, Fuqua School of Business, Duke University.

[71] Gallant, A., Chien, H., and Tauchen, G. (1998): "Calibrating Volatility Diffusions and Extracting Integrated Volatility", Working Paper, Duke University.

[72] Geman, H., Karoui, N., and Rochet, J. (1995): "Changes of Numeraire, Changes of Probability Measure and Option Pricing", *Journal of Applied Probability* 32, pp. 443-458.

[73] Gemmill, G. (1996): "Did Option Traders Anticipate the Crash? Evidence from Volatility Smiles in the U.K., with U.S. Comparisons", *Journal of Futures Markets* 16, pp. 881-897.

[74] Gemmill, G., and Saflekos, A. (1999): "How Useful are Implied Distributions? Evidence from Stock-Index Options", Working Paper, City University Business School.

[75] Geske, R., and Shastri, K. (1985): "Valuation of Approximation: a Comparison of Alternative Approaches", *Journal of Financial and Quantitative Analysis* 20, pp. 45-72.

[76] Greene, W. (1990): *Econometric Analysis*, Second Edition, Prentice Hall.

[77] Grossman, S., and Zhou, Z. (1996): "Equilibrium Analysis of Portfolio Insurance", *Journal of Finance* 51, pp. 1379-1403.

[78] Gruenbichler, A., Longstaff, F. (1996): "Valuing Futures and Options on Volatility", *Journal of Banking and Finance* 20, pp. 985-1001.

[79] Hamilton, J (1994): *Time Series Analysis*, Princeton University Press.

[80] Harvey, C., and Whaley, R. (1989): "Conditional Volatility Estimation and the Pricing of S&P 100 Options", Working Paper, Fuqua School of Business, Duke University.

[81] Harvey, C., and Whaley, R. (1991): "S&P 100 Index Option Volatility", *Journal of Finance* 46, pp. 1551-1561.

[82] Harvey, C., and Whaley, R. (1992): "Market Volatility Prediction and the Efficiency of the S&P 100 Index Option Market", *Journal of Financial Economics* 31, pp. 43-73.

[83] Harvey, A. (1993): *Time Series Models*, Harvester Wheatsheaf, Second Edition.

[84] Heath, D., Jarrow, R., Morton, A. (1992): "Bond Pricing and the Term Structure of Interest Rates : A New Methodology For Contingent Claims Valuation", *Econometrica* 60, pp. 77-105.

[85] Heynen, R. (1994): "An Empirical Investigation of Observed Smile Patterns", *Review of Futures Markets* 13, pp. 317-354.

[86] Ho, L., and Lee, S. (1986): "Term Structure Movements and Pricing Interest Rate Contingent Claims", *Journal of Finance* 41, pp. 1011-1029.

[87] Hodges, S. (1998): "A Generalization of the Sharpe Ratio and its Applications to Valuation Bounds and Risk Measures", Working Paper, Financial Options Research Centre, University of Warwick.

[88] Hsieh, D. (1993): "Estimating the Dynamics of Volatility", *Proceedings of the Conference on Financial Innovations: 20 Years of Black/Scholes and Merton*, Duke University.

[89] Hsieh, D. (1995): "Nonlinear Dynamics in Financial Markets: Evidence and Implications", *Financial Analysts Journal* 51, pp. 55-62.

[90] Hull, J., and White, A. (1987): "The Pricing of Options on Assets with Stochastic Volatilities", *Journal of Finance* 3, pp. 281-300.

[91] Hull, J. (1991): *Options, Futures, and Other Derivative Securities*, Prentice Hall, Second Edition.

[92] Hull, J., and White, A. (1992): "One-Factor Interest-Rate Models and the Valuation of Interest-Rate Contingent Claims", *Journal of Financial and Quantitative Analysis* 28, pp. 235-254.

[93] Jackson, E. (1991): *A User's Guide to Principal Components*, Wiley Series in Probability and Mathematical Statistics.

[94] Jackwerth, J. and Rubinstein, M. (1996): "Recovering Probability Distributions from Option Prices", *Journal of Finance* 51, pp. 1611-1631.

[95] Jackwerth, J. and Rubinstein, M. (1996): "Recovering Stochastic Processes from Options Prices", Working Paper, University of California, Berkeley.

[96] Jackwerth, J. (1997): "Generalized Binomial Trees", *Journal of Derivatives* 5, pp. 7-17.

[97] Jamshidian, F. (1991): "Forward Induction and Construction of Yield Curve Diffusion Models", *Journal of Fixed Income* 1, pp. 62-74.

[98] Jamshidian, F. (1993): "Option and Futures Evaluation with Deterministic Volatilities", *Mathematical Finance* 3, pp. 149-159.

[99] Jarrow, R. and Rudd, A. (1982): "Approximate Option Valuation for Arbitrary Stochastic Processes", *Journal of Financial Economics* 10, pp. 347-369.

[100] Johnson, N., and Kotz, S. (1970): *Continuous Univariate Distributions*, Vol 1 and 2. New York: John Wiley & Sons.

[101] Johnson, H., and Shanno, D. (1987): "Option Pricing when the Variance is Changing", *Journal of Financial and Quantitative Analysis* 22, pp. 143-151.

[102] Kamal, M., and Derman, E. (1997): "The Patterns of Change in Implied Index Volatilities", Goldman Sachs, Quantitative Strategies Research Notes.

[103] Kani, I., Derman, E., Kamal, M. (1997): "Trading and Hedging Local Volatility", *Journal of Financial Engineering* 6, pp. 1233-1268.

[104] Kon, S. (1984): "Models of Stock Returns: A Comparison", *Journal of Finance* 39, pp. 147-166.

[105] Krzanowski, W. (1983): "Cross-Validatory Choice in Principal Component Analysis; Some Sampling Results", *Journal of Statistical Computational Simulation* 18, pp. 299-314.

[106] Ledoit, O., and Santa-Clara, P. (1998): "Relative Pricing of Options with Stochastic Volatility", Working Paper, University of California, Los Angeles.

[107] Litterman, R., and Scheinkman J. (1988): "Common Factors Affecting Bond Returns", Goldman Sachs, Financial Strategies Group.

[108] MacBeth, J., and Merville, L. (1979): "An Empirical Estimation of the Black-Scholes Call Option Pricing Model", *Journal of Finance* 34, pp. 1173-1186.

[109] Macbeth, J., and Merville, L. (1980): "Tests of the Black-Scholes and Cox Call Option Valuation Models", *Journal of Finance* 35, pp. 285-301.

[110] Mayhew, S. (1995a): "Implied Volatility", *Financial Analysts Journal* 51, pp. 8-13.

[111] Mayhew, S. (1995b): "On Estimating The Risk-Neutral Probability Distribution Implied by Option Prices", Working Paper, University of California, Berkeley.

[112] Melick, W., and Thomas, C. (1997): "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, pp. 91-115.

[113] Merton, R. (1976): "Option Pricing when Underlying Stock Returns are Discontinuous", *Journal of Financial Economics* 3, pp. 125-144.

[114] Merville, L., and Pieptea, D. (1989): "Stock Price Volatility, Mean-Reverting Difussion and Noise", *Journal of Financial Economics* 21, pp. 193-214.

[115] Nelson, D., and Ramaswamy, K. (1990): "Simple Binomial Processes as Diffusion Approximations in Financial Models", *Review of Financial Studies* 3, pp. 393-430.

[116] Neuberger, A. (1994): "The Log Contract: A New Instrument to Hedge Volatility", *Journal of Portfolio Management* 75, pp. 74-80.

[117] Oksendal, B. (1992): *Stochastic Differential Equations: An Introduction with Applications*, Springer-Verlag.

[118] Pappalardo, L. (1996): "Option Pricing and Smile Effect when Underlying Stock Prices are Driven by a Jump Process", Working Paper, Financial Options Research Centre, University of Warwick.

[119] Poterba, J., and Summers, L. (1986): "The Persistence of Volatility and Stock Market Fluctuations", *American Economic Review* 76, pp. 1142-1151.

[120] Praetz, P. (1972): "The Distribution of Share Price Changes", *Journal of Business* 45, pp. 49-55.

[121] Ramberg, J., Dudewicz, E., Tadikamalla, P., and Mykytka, E. (1979): "A Probability Distribution and its Uses in Fitting Data", *Technometrics* 21, pp. 201-209.

[122] Rebonato, R. (1996): *Interest-Rate Option Models: Understanding, Analysing and Using Models for Exotic Interest-Rate Options*, John Wiley and Sons.

[123] Reddon, J. (1985): "Monte Carlo type I Error Rates for Velicer's Partial Correlation test for the Number of principal Components", *Criminometrica* 1, pp. 13-23.

[124] Roll, R. (1984): "A Simple Implicit Measure of the Effective Bid-Ask Spread", *Journal of Finance* 39, pp. 1127-1139.

[125] Rubinstein, M. (1985). "Non-Parametric Tests of Alternative Option Pricing Models using all Reported Trades and Quotes on the 30 most Active CBOE Option Classes from August 23, 1976 through August 31, 1978", *Journal of Finance* 40, pp. 455-480.

[126] Rubinstein, M. (1994): "Implied Binomial Trees", *Journal of Finance* 69, pp. 771-818.

[127] Rubinstein, M. (1995): "As Simple as One, Two, Three", *Risk* 8, pp. 44-47.

[128] Rubinstein, M. (1998): "Edgeworth Binomial Trees", *Journal of Derivatives*, pp. 20-27.

[129] Platen, E. and Schweitzer, M. (1995): "On Smile and Skewness", Working Paper, Australian National University.

[130] Schmalensee, R., and Trippi, R. (1978): "Common Stock Volatility Expectations Implied by Option Premia", *Journal of Finance* 33, pp. 129-147.

[131] Scott, L. (1987): "Option Pricing when the Variance changes Randomly: Theory, Estimation and an Application", *Journal of Financial and Quantitative Analysis* 22, pp. 419-438.

[132] Scott, L. (1997): "Pricing Stock Options in a Jump-Diffusion Model with Stochastic Volatility and Interest Rates: Applications of Fourier Inversion Methods", *Mathematical Finance* 7, pp. 413-426.

[133] Shastri, K., and Wethyavivorn, K. (1987): "The valuation of currency options for alternate stochastic processes", *Journal of Financial Research* 10, pp. 283-293.

[134] Sheikh, A. (1991): "Transaction Data Tests of S&P 100 Call Option Pricing", *Journal of Financial and Quantitative Analysis* 26, pp.459-475.

[135] Shimko, D. (1993): "Bounds of Probability", *Risk* 6, pp. 33-37.

[136] *SPSS for Windows Professional Statistics* Release 6.0, Marija Norusis/SPSS INc.

[137] Stein, J. (1989): "Overreactions in the Options Market", *Journal of Finance,* XLIV, pp. 1011-1023.

[138] Strickland, C. (1996): "A Comparison of Diffusion Models of the Term Structure", *The European Journal of Finance* 2, pp. 103-123.

[139] Sundaram, R. (1997): "Equivalent Martingale Measures and Risk-Neutral Pricing: An Expository Note", *Journal of Derivatives* 5, pp. 85-98.

[140] Taylor, S., and Xu, X. (1994): "The Magnitude of Implied Volatility Smiles: Theory and Empirical Evidence for Exchange Rates", *Review of Futures Markets* 13, pp. 355-380.

[141] Taylor, S., and Xu, X. (1994): "Implied Volatility Shapes when Price and Volatility Shocks are Correlated", Working Paper, University of Lancaster.

[142] Timmerington, D., Smith, A., and Makov, U. (1985): *Statistical Analysis of Finite Mixture Distributions*, John Wiley & Sons.

[143] Tompkins, R. (1998): "Implied Volatility Surfaces: Uncovering Regularities for Options on Financial Futures", Working Paper, Financial Options Research Centre, University of Warwick.

[144] Vasicek, O. (1977): "An Equilibrium Characterization of the Term Structure", *Journal of Financial Economics* 5, pp. 177-88.

[145] Velicer, W. (1976): "Determining the Number of Components from the Matrix of Partial Correlations", *Psychometrica* 41, pp.321-327.

[146] Whaley, R. (1986): "Valuation of American Futures Options: Theory and Empirical Tests", *Journal of Finance* 41, pp. 127-150.

[147] Wiggins, J. (1987): "Option Values under Stochastic Volatility", *Journal of Financial Economics* 19, pp. 351-372.

[148] Wilson, T. (1994): "Debunking the Myths", *Risk* 7, pp. 67-73.

[149] Xu, X., and Taylor, S. (1994): "The Term Structure of Volatility Implied by Foreign Exchange Options", *Journal of Financial and Quantitative Analysis* 29, pp. 57-74.

[150] Zou, J., and Derman, E. (1997): "Monte Carlo Valuation of Path-Dependent Options on Indexes with a Volatility Smile", *Journal of Financial Engineering* 6, pp. 149-168.