Semi-parametric implied volatility surface models and forecasts based on a regression tree-boosting algorithm

Dominik Colangelo

Submitted for the degree of Ph.D. in Economics at

Swiss Finance Institute
Faculty of Economics
Università della Svizzera italiana, USI
Lugano, Switzerland

Thesis Committee:
Prof. F. Audrino, advisor, Universität St. Gallen
Prof. F. Trojani, Università della Svizzera italiana
Prof. W. Härdle, Humboldt-Universität zu Berlin

November 2009

Acknowledgments

I would like to thank my supervisor Prof. Francesco Audrino for his guidance throughout my doctoral studies. He taught me a lot by pushing me to my limits and beyond. This thesis is an offspring of my collaboration in his research project '*Multivariate FGD techniques for implied volatility surfaces estimation and term structure forecasting*' that was funded by the Foundation for Research and Development of USI.

During the time spent in Lugano working at the Institute of Finance, I had the chance to make a lot of great friends, discuss my research with colleagues on numerous occasions and follow a top PhD program provided by the Swiss Finance Institute. I also discovered my own italianità and found the love of my life.

Special thanks and gratitude go to my wife, my parents, my sister and my extended family for their love, support and hospitality. A substantial part of the thesis has been written in the land down under.

Gabriela and Kathy take credit for proofreading, remaining typos are my sole responsibility.

Abstract

A new methodology for semi-parametric modelling of implied volatility surfaces is presented. This methodology is dependent upon the development of a feasible estimating strategy in a statistical learning framework. Given a reasonable starting model, a boosting algorithm based on regression trees sequentially minimizes generalized residuals computed as differences between observed and estimated implied volatilities. To overcome the poor predicting power of existing models, a grid is included in the region of interest and a cross-validation strategy is implemented to find an optimal stopping value for the boosting procedure. Back testing the out-of-sample performance on a large data set of implied volatilities from S&P 500 options provides empirical evidence of the strong predictive power of the model. Accurate IVS forecasts also for single equity options assist in obtaining reliable trading signals for very profitable pure option trading strategies.

Contents

Acknowledgments **iii**

Abstract **v**

1 Introduction **1**
1.1 Goals 3
1.2 Outline 4

2 The Black-Scholes model revisited **5**
2.1 GBM as a stock price process 6
2.2 Pricing European plain vanilla options 8
2.2.1 Ingredients of the BS framework 9
2.2.2 The BS formula 10
2.2.3 Comments and clarifications 12
2.3 The Greeks 14
2.4 No-arbitrage conditions and option bounds 17
2.5 Criticism of BS framework 20

3 The Implied Volatility Surface **21**
3.1 Popularity of Black-Scholes IV 25
3.1.1 Sublimation of model ambiguity into IV 25
3.1.2 IV as an instrument for empirical option pricing 25
3.1.3 IV as a risk-neutral expectation of future volatility 27
3.2 Explaining the smile 28
3.2.1 Deterministic instantaneous volatility 28
3.2.2 Local volatility in terms of implied volatility 29
3.2.3 Stochastic volatility 30
3.3 Modelling IVS directly 31

3.3.1 IVS as a link to other volatility concepts 31
3.3.2 Predictor space 32
3.3.3 Challenges 33
3.3.4 Models 34

4 Supervised learning 39
4.1 Classification and regression trees 41
4.2 Functional gradient descent 42

5 Model and estimation procedure 47
5.1 Desired properties 47
5.1.1 Improve upon a starting model 48
5.1.2 Keep extremal IV in the sample 51
5.1.3 Local focus 53
5.1.4 OS prediction 54
5.2 Inspiration 55
5.3 The model 57
5.4 Estimation 58
5.4.1 Empirical local criterion 59
5.4.2 A feasible algorithm 60

6 OS analysis of the S&P 500 IVS 65
6.1 Settings 66
6.1.1 Data 66
6.1.2 Special days of interest 71
6.1.3 Model specification 71
6.1.4 Filtered historical simulation of exogenous factors 78
6.1.5 OS performance measures 78
6.2 Empirical results 79
6.2.1 OS forecasts of predictor variables 79
6.2.2 Cross-validation 82
6.2.3 Relative importance of predictor variables 86
6.2.4 Comparison of different models 86
6.2.5 Dispersion trading 90
6.2.6 Robustness check 94
6.3 Summary 95

7 Trading strategy 103
7.1 Settings 104
7.1.1 Data 104
7.1.2 Calculating option returns 105

7.1.3 Inspiration for an option trading strategy 108
7.2 Method 110
7.2.1 Predicting IV changes 111
7.2.2 Predicting option price changes 113
7.2.3 Predicting option returns 114
7.2.4 Portfolio formation 115
7.3 Empirical results 117
7.3.1 Portfolio return time series 117
7.3.2 Sensitivity analysis 123
7.3.3 Risk measures 128

8 Conclusions **131**

Appendices **135**

A History of options **137**

B Asset pricing and contingent claims **139**

C Volatility **141**
C.1 Instantaneous volatility 142
C.2 Stochastic volatility 145
C.3 Local volatility 148
C.4 Implied volatility 150

Bibliography **155**

List of Figures **167**

List of Tables **169**

Abbreviations **171**

Chapter 1

Introduction

The liquidity of option markets has steadily grown since the seminal work of Black and Scholes (1973) and Merton (1973). They showed that the price of an option is the initial cost of a self financing replicating strategy and derived the well known analytical Black-Scholes (BS) formula for European options. At the current time t, the expiry date T, the underlying stock price S_t as well as the constant risk-free interest rate r are directly observable. However, the instantaneous volatility of the underlying stock return process is unknown. Using the market price of an option, it is possible to numerically solve the BS formula for the unknown volatility parameter. The resulting number is called implied volatility (IV). It is a well known empirical fact that the IV is not constant as actually assumed for deriving the BS formula. Instead, it varies over time, strike and expiry date. The concept of implied volatility surface (IVS) specifies IV as a function of moneyness m and time to maturity τ, where the former quantifies the degree of intrinsic value in the option price and the latter the time value. m is an increasing function in the strike K, in general eventually also depending on t, T, S_t and r.

IV is regarded as a state variable that reflects current market situations and expections about future states. Hence it makes sense to model the IVS directly although the degenerated structure of option data makes this task difficult. Only options with a few distinct maturities, but various different strikes are traded. Certain regions of the IVS exhibit a strong dynamic that is hard to capture. A thoughtfully constructed estimation strategy needs to be considered to avoid all sorts of pitfalls (smoothness, no-arbitrage conditions, computational feasibility, overfitting, etc.).

Recently, a great deal of effort has been put into modelling the IVS directly. Gonçalves and Guidolin (2006) combined a cross-sectional approach similar to that of Dumas, Fleming, and Whaley (1998) with vector autoregressive models. They tried, and partially succeeded depending on transaction costs, to exploit single- and multi-step ahead volatility predictions produced by their model to form profitable volatility-based trading strategies. Semi- and nonparametric smoothing methods as well as dimension-reduction techniques have also been introduced. Skiadopoulos, Hodges, and Clewlow (2000) popularized principal components analysis (PCA) in the IVS literature. They applied PCA on a multivariate time series of IV differences for a given moneyness level and within a certain expiry range. For a surface analysis, they only used three ‘expiry buckets’ with 10 to 90, 90 to 180 and 180 to 270 days to expiry.

Cont and da Fonseca (2002) presented a functional data analysis approach based on the Karhunen-Loève decomposition, an extension of the PCA method for random surfaces. Fengler, Härdle, and Villa (2003) argued that IVs of different maturity groups have a common eigenstructure and defined a common principal component (CPC) framework. Fengler, Härdle, and Mammen (2007) combined methods from functional PCA and backfitting techniques for additive models in their dynamic semiparametric factor model (dsfm). By taking the degenerated option data structure explicitly into account, they overcame some of the difficulties that the models based on PCA had encountered. They fitted their functional model directly on the aggregated data, without the need to estimate IV with a nonparametric smoothing estimator on a fixed grid or to sort IV into moneyness/time to expiry buckets in order to obtain a high dimensional time series of IV classes as an approximation of the IVS. In a comparison of the one-day out-of-sample prediction error, the dsfm performes only 10% better on DAX option data than a simple sticky-moneyness model, where IV is taken to be constant over time at a fixed moneyness.

1.1 Goals

The first goal of this thesis is to set up a statistical learning framework that improves any given starting model for the IVS with an extended predictor space. The classical predictor space consisting of only m and τ is enhanced to higher dimensions by including a call/put dummy variable, exogenous factors and time-lagged as well as forecasted time-leading versions of themselves. Supervised learning is achieved by iteratively applying a tree-boosting algorithm.

Tree-boosting is a simple version of an optimization technique in function space called functional gradient descent (FGD), using regression trees (Breiman, Friedman, Stone, and Olshen, 1984) as base learners and a quadratic loss function. Audrino and Bühlmann (2003) developed this machine learning technique for financial time series. FGD has shown its power in improving volatility forecasts in high-dimensional GARCH models for risk management purposes (Audrino and Barone-Adesi, 2005), modelling interest rates (Audrino, Barone-Adesi, and Mira, 2005) and expected bond returns (Audrino and Barone-Adesi, 2006). It also helps to improve the filtered historical simulation method, for example to compute reliable out-of-sample yield curve scenarios and confidence intervals (Audrino and Trojani, 2007).

The second goal is to focus on out-of-sample predictions of the IVS. For certain regions in the (m, τ) domain, the prediction errors shall be controlled such that the peformance of any reasonable starting model in forecasting IV is also improved under possible structural breaks in the time series.

The third goal is to investigate the practical use of the proposed IVS methodology. Only a few studies link option trading with IV analysis (Ahoniemi, 2006; Goyal and Saretto, 2009). This thesis defines option trading strategies and analyzes their performances, also in the context of dispersion trades (Driessen, Maenhout, and Vilkov, 2009).

1.2 Outline

After thoroughly revisiting the Black-Scholes framework in Chapter 2, the related concept of implied volatility is compared in Chapter 3 to other volatility concepts that emerge from generalizing the dynamics of the underlying security. It is possible to analyze the shape of the IVS for any local volatility or stochastic volatility model[1], but the opposite direction is more promising as IV provides an exact link to them. Modelling the IVS directly (as a random field) raises a lot of questions about possible predictors of IVS.

Chapter 4 introduces supervised learning methods that perform automatic variable selection. Chapter 5, based on a forthcoming article in Statistics and Computing[2], defines the new methodology for modelling the IVS in a statistical learning framework.

The two following chapters are empirical. In Chapter 6, the out-of-sample (OS) performance of IVS predictions for the S&P 500 index is analyzed, also for a possible application with dispersion trading. In Chapter 7, single equity option returns (the constituents of the S&P 100 index) are forecasted 10 days OS. A pure option trading strategy is defined based on that signal, relying on stability of the moneyness state during the last 20 calendar days until maturity. Conclusions are presented in Chapter 8.

---

[1] All such models generate an IVS with similar shape (Gatheral, 2006, Chapter 7).

[2] Audrino, F. and D. Colangelo (2009). Semi-parametric forecasts of the implied volatility surface using regression trees. Forthcoming in *Statistics and Computing*. DOI: `10.1007/s11222-009-9134-y`.

Chapter 2

The Black-Scholes model revisited

The model of Black and Scholes (1973) is set in a continuous-time financial market. Assume there are two securities in a frictionless market[3], a risky asset S_t and a risk-free security B_t that acts as a numéraire, i.e. a saving account paying a risk-free interest rate r, here assumed to be constant and equal for borrowing and lending. The dynamics of the two securities are given by

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{2.1}$$

$$dB_t = rB_t dt \tag{2.2}$$

where W_t is a $\mathbb{F}$-adapted standard Wiener process (a.k.a. Brownian motion) defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The filtration $\mathbb{F}$ is an increasing sequence of σ-algebras on $(\Omega, \mathcal{F})$, consisting of $\mathcal{F}_t = \sigma(W_s : s \leq t)$, the smallest σ-algebra such that all $\{W_s, s \leq t\}$ are $\mathcal{F}_t$-measurable, for $t \in [0, T]$. Furthermore, all $\mathbb{P}$-nullsets are included in $\mathcal{F}_0$. In other words, the investors know the history of S from time 0 up to present time t, but they have no information about later values.

---

[3] Assets are perfectly (infinitesimally) divisible, there are no short sale restrictions and no transaction costs occur either for buying or selling.

2.1 Geometric Brownian motion as a process for stock prices

The solution of the ordinary differential equation for the numéraire (2.2) with boundary condition $B_0 = 1$ is straightforward, given by $B_t = e^{rt}$. Dividing both sides of Eq. (2.1) by $S_t > 0$ reveals that μ is the instantaneous drift and σ the instantaneous volatility of dS_t/S_t, the percentage change process of S_t over an infinitesimally small period dt. Both μ and σ are assumed to be constant in the Black-Scholes (BS) framework. A process following such a stochastic differential equation (SDE) is called geometric Brownian motion (GBM). The solution to Eq. (2.1) is analytically given by

$$S_t = S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_t\right) \tag{2.3}$$

for any initial value $S_0 > 0$. This can be checked with the help of Itô's lemma. For an Itô process of the form

$$X_t = X_0 + \int_0^t a_s ds + \int_0^t b_s dW_s \tag{2.4}$$

with a predictable and Lebesgue integrable, b a predictable W-integrable process, Itô's lemma states that a twice continuously differentiable function f on X_t is itself an Itô process with dynamics given by

$$df(X_t) = f'(X_t)dX_t + \frac{1}{2} f''(X_t) d\left\langle X \right\rangle_t, \tag{2.5}$$

adding half of the second derivative of f times the differential of the quadratic variation process to the standard chain rule part[4]. For a partition of the interval $[0, t]$, $0 = t_0 < t_1 < ... < t_n = t$, the quadratic variation $\sum_{k=1}^{n} (X_{t_k} - X_{t_{k-1}})^2$ converges in probability to $\left\langle X \right\rangle_t = \int_0^t b_s^2 ds$ as the mesh of the partition tends to 0. Therefore, in differential notation, we have $d\left\langle X \right\rangle_t = b_t^2 dt$.

Applying Itô's lemma (2.5) to $f(S_t) = \log S_t$ helps find the solution of the SDE for a geometric Brownian motion.

---

[4] More generally, if $f(t, X_t)$ is continuously differentiable in t and twice continuously differentiable in X_t, then $df(t, X_t) = \left(\frac{\partial f(t,X_t)}{\partial t} dt + \frac{\partial f(t,X_t)}{\partial X_t} dX_t\right) + \frac{1}{2} \frac{\partial^2 f(t,X_t)}{\partial X_t^2} d\left\langle X \right\rangle_t$.

$$\begin{aligned} d\log S_t &= \frac{1}{S_t}dS_t + \frac{1}{2}\left(-\frac{1}{S_t^2}\right)\sigma^2 S_t^2 dt \\ &= \frac{1}{S_t}(\mu S_t dt + \sigma S_t dW_t) - \frac{1}{2}\sigma^2 dt \\ &= (\mu - \frac{1}{2}\sigma^2)dt + \sigma dW_t, \end{aligned} \tag{2.6}$$

the right-hand side being independent of S_t. It follows that $\log S_t = \log S_0 + (\mu - \frac{1}{2}\sigma^2)t + \sigma W_t$, and solving for S_t leads to expression (2.3). The defining properties of a standard Wiener process[5] together with the derived results imply that the log return process of S_t has a normal distribution,

$$\log\left(\frac{S_t}{S_s}\right) \overset{d}{\sim} \mathcal{N}\left((\mu - \frac{1}{2}\sigma^2)(t-s), \sigma^2(t-s)\right), \;\; 0 \le s < t \le T. \tag{2.7}$$

Hence, $S_t|S_s$ is log-normally distributed with probability density function (PDF)

$$\begin{aligned} p_{s,t}(x) &= \frac{1}{xb\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\log x - a}{b}\right)^2\right\} \qquad (2.8) \\ a &:= a(s,t,\mu,\sigma,S_s) = \left(\mu - \frac{1}{2}\sigma^2\right)(t-s) + \log S_s \\ b &:= b(s,t,\sigma) = \sigma\sqrt{t-s} \end{aligned}$$

and cumulative distribution function (CDF)

$$\begin{aligned} \mathbb{P}[S_t \le x|S_s] = \mathbb{P}[S_t \le x|\mathcal{F}_s] &= \int_0^x p_{s,t}(y)dy \qquad (2.9) \\ &= \int_{-\infty}^{\frac{\log x - a}{b}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\} dz \qquad (2.10) \\ &= \int_{-\infty}^{\frac{\log x - a}{b}} \varphi(z)dz = \Phi\left(\frac{\log x - a}{b}\right) \quad . \qquad (2.11) \end{aligned}$$

A change of variable takes place in (2.10), $z := \frac{\log y - a}{b}$. $\varphi(\cdot)$ and $\Phi(\cdot)$ denote the PDF and CDF of a standard normal random variable.

---

[5] A standard Wiener process W_t on $[0,T]$ is defined by the following properties: $W_0 = 0$, W_t is almost surely continuous, has independent increments and $W_t - W_s \overset{d}{\sim} \mathcal{N}(0, t-s)$ for $0 \le s < t \le T$.

The conditional expectation and variance of $S_t|S_s$ under the phyiscal probability measure $\mathbb{P}$ are

$$\begin{aligned} \mathrm{E}_{\mathbb{P}}[S_t|S_s] &= e^{a+\frac{1}{2}b^2} = e^{\mu(t-s)}S_s && (2.12) \\ \mathrm{Var}_{\mathbb{P}}(S_t|S_s) &= e^{2a+b^2}\left(e^{b^2}-1\right) \\ &= e^{2\mu(t-s)}S_s^2\left\{e^{\sigma^2(t-s)}-1\right\}. && (2.13) \end{aligned}$$

Remark 2.1 Note that the instantaneous drift μ is the expected percentage change in the stock price per infinitesimally small period dt, $\mathrm{E}_{\mathbb{P}}[dS_t/S_t]/dt = \mu$, but the expected continuously compounded return over the period $[0,T]$ is $\mathrm{E}_{\mathbb{P}}\left[\frac{1}{T}\log\left(\frac{S_T}{S_0}\right)\right] = \mu - \frac{1}{2}\sigma^2$.

2.2 Pricing European plain vanilla options

The term “plain vanilla option” describes the standard version of an option that does not have any special component. This is unlike an exotic option which is more complex and non-standard.

Definition 2.2 *A stock option is a contract between a buyer (holder) and a seller (writer) that guarantees the buyer the right, but not the obligation, to buy (call option) or sell (put option) a share of the underlying stock at a fixed strike price K in the future at (European-style) or up to (American-style) a fixed maturity date T (a.k.a. expiry date). In financial jargon, the holder is said to be long and the writer short an option. If the option is exercised, the writer is obliged to fulfill the terms of the contract.*

The frictionless BS financial market consisting of a risk-free security $B_t = e^{rt}$ with constant r and a (non-dividend paying) risky stock S_t that follows a geometric Brownian motion with constant μ and σ is complete and does not allow for arbitrage opportunities (Hafner, 2004, p. 24). A complete market is one in which any contingent claim is attainable, i.e. for any contingent claim, there exists a self-financing strategy investing in the given securities such that it replicates the final value of that contingent claim. Therefore, by the fundamental theorem of asset pricing, a unique risk-neutral measure to price contingent claims exists (Schachermayer, 2009). The principles of contingent claim pricing are explained in Appendix B.

2.2.1 Ingredients of the BS framework

Equations (2.1), (2.3), (2.7) specify the stock price process and Eq. (2.8) its PDF; the pricing kernel is given by the following change of measure

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left\{-\int_0^t \left(\frac{\mu - r}{\sigma}\right) dW_s - \frac{1}{2}\int_0^t \left(\frac{\mu - r}{\sigma}\right)^2 ds\right\} \tag{2.14}$$

and Girsanov's theorem states that

$$\tilde{W}_t = \left(\frac{\mu - r}{\sigma}\right) t + W_t \tag{2.15}$$

is a standard Brownian motion under the new measure $\mathbb{Q}$, which together with Eq. (2.1) implies that the stock price process satisfies

$$dS_t = rS_t dt + \sigma S_t d\tilde{W}_t. \tag{2.16}$$

The density of $\mathbb{Q}$ is called risk-neutral PDF or state-price density (SPD),

$$\begin{aligned} q_{s,t}(x) &= d\mathbb{Q}[S_t \leq x | S_s] \qquad (2.17) \\ &= \frac{1}{x\sigma\sqrt{2\pi(t-s)}} \exp\left\{-\frac{1}{2}\left(\frac{\log(\frac{x}{S_s}) - \left(r - \frac{1}{2}\sigma^2\right)(t-s)}{\sigma\sqrt{t-s}}\right)^2\right\}. \end{aligned}$$

The risk-neutral PDF is Log-normal distributed like the physical PDF in Eq. (2.8), but with r instead of μ.

The discounted stock price process $\tilde{S}_t = e^{-rt}S_t$ is a martingale under $\mathbb{Q}$; to prove this, we have $d\tilde{S}_t = \tilde{S}_t \sigma d\tilde{W}_t$ by virtue of Itô's lemma, and an Itô integral is a martingale (Elliott and Kopp, 2005, Theorem 6.3.3)[6]. Alternatively, the martingale property is directly checked by

---

[6] The martingale representation theorem proves the converse statement. Any almost sure continuous martingale can be expressed as an Itô integral with unique integrand process w.r.t. a standard Brownian motion (Elliott and Kopp, 2005, Theorem 7.3.9).

$$\mathrm{E}_{\mathbb{Q}}[\tilde{S}_t|\mathcal{F}_s] = S_0\mathrm{E}_{\mathbb{Q}}\left[e^{\sigma\tilde{W}_t - \frac{\sigma^2}{2}t} \middle| \mathcal{F}_s \right] \overset{(*)}{=} S_0 e^{\sigma\tilde{W}_s - \frac{\sigma^2}{2}s} = \tilde{S}_s \tag{2.18}$$

for all $0 \leq s < t \leq T$. $(*)$ follows from the defining properties of a Wiener process (Elliott and Kopp, 2005, Theorem 6.2.5).

According to Ait-Sahalia and Lo, *"SPDs are 'sufficient statistics' in an economic sense – they summarize all relevant information about preferences and business conditions for purposes of pricing financial securities"* (1998, p. 503). Detlefsen, Härdle, and Moro (2007) show how to recover the market utility function $U(s)$ implicit in the BS framework by equating the stochastic discount factor $M_{t,T} = \beta U'(S_T)/U'(S_t)$ obtained in a preference-based equilibrium model, where β is a fixed discount factor, with the state price density per unit probability $e^{-r(T-t)}q_{t,T}(S_T)/p_{t,T}(S_T)$ that appears in the context of risk-neutral pricing. The implicit utility is a power utility of the form

$$U(S_T) = \left(1 - \frac{\mu - r}{\sigma^2}\right)^{-1} S_T^{\left(1 - \frac{\mu - r}{\sigma^2}\right)} . \tag{2.19}$$

The contract specifications of a European plain vanilla stock option determine its payoff function; $\psi_T(S_T) = \max(S_T - K, 0)$ for a call and $\psi_T(S_T) = \max(K - S_T, 0)$ for a put. All relevant quantities to price these contingent claims (Appendix B) have now been defined. Option prices can be obtained by calculating $\pi_t(\psi_T) = \mathrm{E}_{\mathbb{P}}\left[\psi_T M_{t,T}|\mathcal{F}_t\right] = \mathrm{E}_{\mathbb{Q}}\left[\psi_T e^{-r(T-t)}\middle| \mathcal{F}_t\right]$.

2.2.2 The BS formula

Black and Scholes derive their famous option pricing formula by showing that *"it is possible to create a hedged position, consisting of a long position in the stock and a short position in the* [call] *option* [on the same stock], *whose value will not depend on the price of the stock"* (1973, p. 641). Since such a hedge portfolio is risk-free, its rate of return must equal r by the assumption of no-arbitrage.

More generally, this method of arbitrage-free pricing leads to a partial differential equation (PDE) for the price $H(t, S_t)$ of a European contingent claim[7]. Merton (1973) derives the BS model from weaker assumptions than in the original paper and also includes dividends. If the stock provides a dividend yield at constant rate q, then the BS PDE turns out to be

$$\frac{\partial H}{\partial t} + (r-q)S\frac{\partial H}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 H}{\partial S^2} - rH = 0 \tag{2.20}$$

with boundary condition $H(T, S_T) = \psi_T(S_T)$. For European plain vanilla stock options, the solution of the PDE can be analytically calculated and is known as BS formula,

$$C_t^{\text{BS}} = S_t e^{-q(T-t)}\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2) \qquad \text{(call)} \tag{2.21}$$

$$P_t^{\text{BS}} = Ke^{-r(T-t)}\Phi(-d_2) - S_t e^{-q(T-t)}\Phi(-d_1) \qquad \text{(put)} \tag{2.22}$$

where

$$\begin{aligned} \Phi(u) &= \int_{-\infty}^{u} \varphi(z)dz & \qquad d_1 &= \frac{\log(S_t/K) + (r - q + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \\ \varphi(z) &= \frac{1}{\sqrt{2\pi}} e^{-z^2/2} & \qquad d_2 &= d_1 - \sigma\sqrt{T-t} \end{aligned}$$

Definition 2.3 *The cp flag denotes a binary variable that equals 1 for a call and 0 for a put option.*

The BS formula can then be written as

$$\text{BS}_t(S_t, \sigma, cp \text{ flag}, K, T, r, q) = \begin{cases} C_t^{\text{BS}} & \text{if } cp \text{ flag} = 1 \\ P_t^{\text{BS}} & \text{if } cp \text{ flag} = 0 \end{cases}. \tag{2.23}$$

---

[7] Its payoff function $\psi_t = \psi_t(S_t)$ must be path-independent and a non negative random variable that is $\mathcal{F}_t$-measurable. An integrability condition for ψ_t can be found in Fengler (2005, Section 2).

2.2.3 Comments and clarifications

The solution of the BS PDE (2.20) with boundary condition $H(T, S_T) = \psi_T(S_T)$ is equivalent to the 'linear pricing rule' result that is inherent in the state price approach, $H(t, S_t) \equiv \pi_t(\psi_T(S_T))$. For example, the price of a European call option on a non-dividend paying stock is

$$\begin{aligned} C_t(S_t, K, T) &= \pi_t(\max(S_T - K, 0)) \\ &= e^{-r(T-t)} \mathrm{E}_{\mathbb{Q}}[\max(S_T - K, 0)|\mathcal{F}_t] \\ &= e^{-r(T-t)} \int_0^\infty \max(S_T - K, 0) \ d\mathbb{Q}(S_T|S_t) \\ &= e^{-r(T-t)} \int_K^\infty (S_T - K) q_{t,T}(S_T) \ dS_T. \end{aligned} \tag{2.24}$$

The first part of the integral in (2.24) is

$$\begin{aligned} \int_K^\infty S_T q_{t,T}(S_T) \ dS_T &= \mathrm{E}_{\mathbb{Q}}[S_T|S_t] - \int_0^K S_T q_{t,T}(S_T) \ dS_T \\ &= e^{r(T-t)} S_t - \int_0^K S_T q_{t,T}(S_T) \ dS_T \end{aligned} \tag{2.25}$$

and the second part

$$\begin{aligned} \int_K^\infty K q_{t,T}(S_T) \ dS_T &= K\mathbb{Q}[S_T > K|S_t] \\ &= K(1 - \mathbb{Q}[S_T \leq K|S_t]) \\ &= K - K \int_0^K q_{t,T}(S_T) \ dS_T. \end{aligned} \tag{2.26}$$

Indeed, it can be shown that

$$C_t(S_t, K, T) \equiv \mathrm{BS}_t(S_t, \sigma, cp \text{ flag} = 1, K, T, r, q = 0).$$

Remark 2.4 Breeden and Litzenberger (1978) prove that $q_{t,T}(x) = d\mathbb{Q}[S_T \leq x|S_t]$ is the second derivative of the price of a call option with strike x at maturity T w.r.t. the strike of the price when *"the relation between the future cash flow* *and the underlying portfolio may be of any type – not necessarily linear or jointly normal"* (p. 649),

$$q_{t,T}(x) = e^{r(T-t)} \left. \frac{\partial^2 C_t(S_t, K, T)}{\partial K^2} \right|_{K=x} . \tag{2.27}$$

Note 2.5 This result is only based on the specific form of the call payoff function $\psi_T = \max(S_T - K, 0)$, as we shortly verify for Eq. (2.24) with help of Equations (2.25) and (2.26):

$$\frac{\partial C_t(S_t, K, T)}{\partial K} = e^{-r(T-t)} \left[\int_0^K q_{t,T}(S_T) \ dS_T - 1 \right] \tag{2.28}$$

$$\frac{\partial^2 C_t(S_t, K, T)}{\partial K^2} = e^{-r(T-t)} q_{t,T}(K). \tag{2.29}$$

Remark 2.6 The BS PDE (2.20) and therefore also the BS formula (2.23) do not depend on μ. No individual investor preferences or agreements on expectations amongst investors are assumed in the BS framework.

> *It is quite reasonable to expect that investors may have quite different estimates for current (and future) expected returns due to different levels of information, techniques of analysis, etc. However, most analysts calculate estimates of variances and covariances in the same way: namely, by using previous price data. Since all have access to the same price history, it is also reasonable to assume that their variance-covariance estimates may be the same* (Merton, 1973, p. 163).

This seems to be a contradiction to the found implicit market utility (2.19). Using Eq. (2.27), Breeden and Litzenberger clarify this issue by showing that *"a necessary and sufficient condition for the Black-Scholes option-pricing formula to correctly price options on aggregate consumption is that individuals' preferences aggregate to a utility function displaying constant relative risk aversion"* (1978, Theorem 3).

2.3 The Greeks

The Greeks of a European contingent claim represent the sensitivities of the value process $H(t, S_t)$ to a small change in underlying parameters of the financial model. Usually, they are denoted by Greek letters. Table 2.1 defines the Greeks as partial derivatives of $H(t, S_t)$. The most common ones are delta, gamma, vega, theta and rho;

$$\Delta := \frac{\partial H}{\partial S}, \quad \Gamma := \frac{\partial^2 H}{\partial S^2}, \quad \nu = \frac{\partial H}{\partial \sigma}, \quad \theta := \frac{\partial H}{\partial t}, \quad \rho = \frac{\partial H}{\partial r} \;.$$

The Greeks can be analytically calculated in the case of European plain vanilla stock options with given price

$$\mathrm{BS}_t(S_t, \sigma, cp\ flag, K, T, r, q) = C_t^{\mathrm{BS}} \mathbb{I}_{cp\ flag=1} + P_t^{\mathrm{BS}} \mathbb{I}_{cp\ flag=0},$$

where $\mathbb{I}_{\text{expression}}$ is a dummy variable that equals 1 if the expression is true and 0 otherwise.

$$\begin{aligned}
\Delta_t^{\mathrm{BS}} = \frac{\partial \mathrm{BS}_t}{\partial S_t} &= \left\{ e^{-q\tau}\Phi(d_1) \right\} \mathbb{I}_{cp\ flag=1} \\
&\quad + \left\{ -e^{-q\tau}\Phi(-d_1) \right\} \mathbb{I}_{cp\ flag=0} && (2.30)\\
\Gamma_t^{\mathrm{BS}} = \frac{\partial \mathrm{BS}_t}{\partial S_t^2} &= \frac{e^{-q\tau}\varphi(d_1)}{S_t \sigma \sqrt{\tau}} && (2.31)\\
\nu_t^{\mathrm{BS}} = \frac{\partial \mathrm{BS}_t}{\partial \sigma} &= e^{-q\tau} S_t \sqrt{\tau}\varphi(d_1) && (2.32)\\
\theta_t^{\mathrm{BS}} = \frac{\partial \mathrm{BS}_t}{\partial t} &= \left\{ -\frac{e^{-q\tau} S_t \sigma \varphi(d_1)}{2\sqrt{\tau}} + q e^{-q\tau} S_t \Phi(d_1) \right. \\
&\quad \left. - r e^{-r\tau} K \Phi(d_2) \right\} \mathbb{I}_{cp\ flag=1} \\
&\quad + \left\{ -\frac{e^{-q\tau} S_t \sigma \varphi(d_1)}{2\sqrt{\tau}} - q e^{-q\tau} S_t \Phi(-d_1) \right. && (2.33)\\
&\quad \left. + r e^{-r\tau} K \Phi(-d_2) \right\} \mathbb{I}_{cp\ flag=0} \\
\rho_t^{\mathrm{BS}} = \frac{\partial \mathrm{BS}_t}{\partial r} &= \left\{ \tau e^{-r\tau} K \Phi(d_2) \right\} \mathbb{I}_{cp\ flag=1} \\
&\quad + \left\{ -\tau e^{-r\tau} K \Phi(-d_2) \right\} \mathbb{I}_{cp\ flag=0} && (2.34)
\end{aligned}$$

Definition of the Greeks

	Spot price S	Volatility σ	Time t	Time to expiry $\tau := T - t$	Risk-free rate r
Value H	delta $\Delta := \frac{\partial H}{\partial S}$	vega $\nu := \frac{\partial H}{\partial \sigma}$	theta $\theta := \frac{\partial H}{\partial t}$	$[\theta = -\frac{\partial H}{\partial \tau}]$	rho $\rho := \frac{\partial H}{\partial r}$
Delta Δ	gamma $\Gamma := \frac{\partial \Delta}{\partial S} = \frac{\partial^2 H}{\partial S^2}$	vanna $\frac{\partial \Delta}{\partial \sigma} = \frac{\partial \nu}{\partial S} = \frac{\partial^2 H}{\partial S \partial \sigma}$		charm $\frac{\partial \Delta}{\partial \tau} = -\frac{\partial \theta}{\partial S} = \frac{\partial^2 H}{\partial S \partial \tau}$	
Gamma Γ	speed $\frac{\partial \Gamma}{\partial S} = \frac{\partial^3 H}{\partial S^3}$	zomma $\frac{\partial \Gamma}{\partial \sigma} = \frac{\partial^3 H}{\partial S^2 \partial \sigma}$		color $\frac{\partial \Gamma}{\partial \tau} = \frac{\partial^3 H}{\partial S^2 \partial \tau}$	
Vega ν	vanna $\frac{\partial \Delta}{\partial \sigma} = \frac{\partial \nu}{\partial S} = \frac{\partial^2 H}{\partial S \partial \sigma}$	vomma $\frac{\partial \nu}{\partial \sigma} = \frac{\partial^2 H}{\partial \sigma^2}$		DvegaDtime $\frac{\partial \nu}{\partial \tau} = \frac{\partial^2 H}{\partial \sigma \partial \tau}$	
Vomma		ultima $\frac{\partial \text{vomma}}{\partial \sigma} = \frac{\partial^3 H}{\partial \sigma^3}$			

Table 2.1: *“The table shows the relationship of the more common sensitivities to the four primary inputs into the Black-Scholes model (spot price of the underlying security, time remaining until option expiration, volatility and the rate of return of a risk-free investment) and to the option’s value, delta, gamma, vega and vomma. Greeks which are a first-order derivative are in* [**blue**]*, second-order derivatives are in* [**green**]*, and third-order derivatives are in* [**orange**]*. Note that vanna is used, intentionally, in two places as these two sensitivities are mathematically equivalent”* (Wikipedia contributors, 2009).

Remark 2.7 First-order linear approximations of the loss distribution play an important role in risk management, for example when estimating the value at risk (VaR) of a stock portfolio. If the risk-factor changes have a multivariate normal distribution, then a linear combination of them is also normally distributed and it is not difficult to find the mean μ_p and variance σ_p of the portfolio. VaR is the α-quantile of the loss distribution over a specified period. In this case, the calculations simplify to $\mathrm{VaR}_\alpha = \mu_p + \sigma_p \Phi^{-1}(\alpha)$ because the normal distribution belongs to a location-scale family. Hence it is clear why this procedure is called variance-covariance method or delta-normal approach in the literature (see e.g. McNeil, Frey, and Embrechts, 2005, Section 2.3.1).

> *For spot or forward positions in the underlying, the delta approach is fully accurate, because the associated price function . . . is linear in the underlying. The delta approximation . . . is the foundation of delta hedging: A position in the underlying asset whose size is minus the delta of the derivative is a hedge of changes in price of the derivative, if continually re-set as delta changes, and if the underlying price does not jump* (Duffie and Pan, 2001, Section 3.1).

Remark 2.8 Applying the delta method to an option portfolio results in a poor approximation of the true change in value because an option price is a highly nonlinear function of (t, S_t, σ, r, q). A better solution is given by a second-order Taylor extension. For a general portfolio value process $V(t, X_t)$ that depends on a d-dimensional risk factor X_t, the delta-gamma method

$$\delta V_t \approx \theta \delta t + \Delta' \delta X_t + \frac{1}{2}\ \delta X_t' \Gamma \delta X_t \tag{2.35}$$

approximates the change in portfolio value $\delta V_t = V_{t+\delta t} - V_t$ over a short fixed time δt as a function of risk-factor changes $\delta X_t = X_{t+\delta t} - X_t$. The symbol $'$ stands for the transpose sign. The Greeks of the portfolio are $\theta = \frac{\partial V_t}{\partial t}$, $\Delta = [\frac{\partial V_t}{\partial X_{t,1}}, \dots \frac{\partial V_t}{\partial X_{t,d}}]'$ (gradient) and $\Gamma = [\Gamma_{ij}]$, a $d \times d$ matrix (Hessian) with $\Gamma_{ij} = \frac{\partial^2 V_t}{\partial X_{t,i} \partial X_{t,j}}$. Duffie and Pan (2001, Section 4) show how to calculate the portfolio VaR.

Remark 2.9 In his PhD thesis, Studer studied the delta-gamma method and noted that it *"captures a part of the non-linearity of option portfolios. Nevertheless heavy-tailedness is not included and we have the problem of estimating a* *covariance matrix* [of X_t]. *Finally for the last step* [finding the distribution of δV_t for risk management purposes] *we have to rely on numerical methods* (2001, p. 11). Assuming a BS framework, Studer refined the delta-gamma method in Proposition 4.9 by using stochastic Taylor expansions to approximate the *"distribution of the change in value of a portfolio … of positions in assets and derivatives in the market"* (2001, p. 68).

2.4 No-arbitrage conditions and option bounds

The value of a contingent claim at expiry date T equals its payoff, $\pi_T(\psi_T(S_T)) = \psi_T(S_T)$ and hence it is obvious that $C_T(S_T, K, T) - P_T(S_T, K, T) = \max(S_T - K, 0) - \max(K - S_T, 0) = S_T - K$. A simple no-arbitrage argument shows that this equality must also hold for $t < T$ when K is discounted appropriately, the options are of European style and the stock does not pay dividends,

$$C_t - P_t = S_t - e^{-r(T-t)}K. \tag{2.36}$$

Eq. (2.36) is called **put-call parity** and is model-free, i.e. only based on the specific form of European option payoff functions similar to the Breeden-Litzenberger result in Remark 2.4. The put-call parity also holds for the BS formula,

$$\begin{aligned} C_t^{\text{BS}} &= S_t\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2) \\ P_t^{\text{BS}} &= Ke^{-r(T-t)}\Phi(-d_2) - S_t\Phi(-d_1) \\ \hline \Rightarrow C_t^{\text{BS}} - P_t^{\text{BS}} &= S_t - Ke^{-r(T-t)} \end{aligned}$$

since $\Phi(d_i) + \Phi(-d_i) = \Phi(d_i) + (1 - \Phi(d_i)) = 1$ for $i = 1, 2$. If the stock pays dividends, the present value of the dividends that will be paid out before the option's expiry date T needs to be subtracted from S_t in Eq. (2.36). If we assume a dividend yield at constant rate q, the **put-call parity** becomes

$$C_t - P_t = e^{-q(T-t)}S_t - e^{-r(T-t)}K. \tag{2.37}$$

$C_t, P_t \geq 0$, hence the following **lower and upper bounds** for European option prices are implied by Eq. (2.37):

$$\max(e^{-q(T-t)}S_t - e^{-r(T-t)}K, 0) \leq C_t \leq S_t \tag{2.38}$$

$$\max(e^{-r(T-t)}K - e^{-q(T-t)}S_t, 0) \leq P_t \leq e^{-r(T-t)}K. \tag{2.39}$$

From Eq. (2.28) follows that $\frac{\partial C_t}{\partial K} < 0$ since $0 \leq \int_0^K q_{t,T}(S_T)\ dS_T < 1$ for $0 \leq K < \infty$ and by combining Eq. (2.28) with Eq. (2.37), we conclude that

$$\begin{aligned}\frac{\partial P_t}{\partial K} &= \frac{\partial\left(C_t - e^{-q(T-t)}S_t + e^{-r(T-t)}K\right)}{\partial K} \\ &= e^{-r(T-t)}\int_0^K q_{t,T}(S_T)\ dS_T > 0.\end{aligned} \tag{2.40}$$

In the BS framework, a similar change of variable as in Eq. (2.11) leads to an explicit expression for $\mathbb{Q}[S_T \leq K|S_t]$ since

$$\begin{aligned}\int_0^K q_{t,T}(S_T)\ dS_T &= \Phi\left(\frac{\log K - \left(r - \frac{1}{2}\sigma^2\right)(T-t) - \log S_t}{\sigma\sqrt{T-t}}\right) \\ &= \Phi(-d_2) = 1 - \Phi(d_2).\end{aligned} \tag{2.41}$$

The partial derivations w.r.t. K are then given by

$$\frac{\partial C_t^{\mathrm{BS}}}{\partial K} = -e^{-r(T-t)}\Phi(d_2) < 0 \tag{2.42}$$

$$\frac{\partial P_t^{\mathrm{BS}}}{\partial K} = e^{-r(T-t)}\{1 - \Phi(d_2)\} > 0. \tag{2.43}$$

As a consequence of $\frac{\partial C_t}{\partial K} < 0$ and $\frac{\partial P_t}{\partial K} > 0$, the **strike monotonicity** of European options is a no-arbitrage condition: for $K_1 < K_2$ and all other variables fixed,

$$C_t(K_1) > C_t(K_2) \tag{2.44}$$

$$P_t(K_1) < P_t(K_2). \tag{2.45}$$

When dealing with European options on non-dividend paying stocks ($q = 0$), there exists a **maturity monotonicity** due to $\mathbb{Q}(S_{T_1} > K|S_t) < \mathbb{Q}(S_{T_2} > K|S_t)$

for $T_1 < T_2$ under usual economical conditions ($r > 0$). In such a case, $\frac{\partial C_t}{\partial \tau} = -\frac{\partial C_t}{\partial t} > 0$: for $\tau_1 < \tau_2$ and all other variables fixed,

$$C_t(\tau_1) \leq C_t(\tau_2) \tag{2.46}$$

$$P_t(\tau_1) - e^{-r\tau_1}K \leq P_t(\tau_2) - e^{-r\tau_2}K. \tag{2.47}$$

A long (short) **butterfly spread** with strikes $K_1 < K_2 < K_3$ is an option strategy that is neutral in the underlying at level K_2 and bearish (bullish) in volatility, i.e. a trader taking such a position does not assume anything about the direction in which S_t moves relative to K_2 as $t \to T$ ($S_T \lesseqgtr K_2$?), but she believes in decreasing (increasing) volatility such that S_T is close to (far away from) K_2. A long butterfly can be created by going long 1 call with strike K_1, short 2 calls with strike K_2 and long 1 call with strike K_3, all with the same maturity T, or by doing the same with puts. The payoff function of a long butterfly is shaped like an upside-down V. There is non-zero probability that the payoff Ψ_T is positive, hence it must have a non-negative price by no-arbitrage:

$$C_t(K_1) - 2C_t(K_2) + C_t(K_3) \geq 0 \tag{2.48}$$

$$P_t(K_1) - 2P_t(K_2) + P_t(K_3) \geq 0. \tag{2.49}$$

Cassese and Guidolin (2004) also discuss additional no-arbitrage conditions: **reverse strike monotonicity** (for $K_1 < K_2$ and all other variables fixed)

$$(K_1 - K_2)e^{-r\tau} \leq C_t(K_2) - C_t(K_1) \tag{2.50}$$

$$P_t(K_2) - P_t(K_1) \leq (K_2 - K_1)e^{-r\tau}, \tag{2.51}$$

box spreads

$$[P_t(K_2) - C_t(K_2)] - [P_t(K_1) - C_t(K_1)] = (K_2 - K_1)e^{-r\tau} \tag{2.52}$$

and **maturity spreads** (for $\tau_1 < \tau_2$ and all other variables fixed)

$$[P_t(\tau_2) - C_t(\tau_2)] - [P_t(\tau_1) - C_t(\tau_1)] = K(e^{-r\tau_2} - e^{-r\tau_1}). \tag{2.53}$$

Testing these no-arbitrage conditions and option price bounds empirically is rather tricky. Synchrony of option and equity prices is absolutely essential, but not necessarily ensured when using end-of-day settlement data. It is also important to mind the persistence of detected arbitrage opportunities. Market microstructure (Corsi, 2005; Bandi and Russell, 2008), transaction costs and dividends need to be taken into consideration. Put-call parity (2.37) and all derived results only hold for European options. For an overview of the classical empirical literature on testing no-arbitrage conditions in option prices see Hull (2002, Section 8.8).

2.5 Criticism of BS framework

Undoubtedly, the assumptions made in the BS framework are unrealistic. The fact that financial markets are not frictionless lies at the bottom of the market microstructure theory. A continously rebalanced hedge with or without transaction costs of option positions can not be realized in practice.

> *"The many improvements on Black-Scholes are rarely improvements, the best that can be said for many of them is that they are just better at hiding their faults. Black-Scholes also has its faults, but at least you can see them"* (Wilmott, 2008).

The main flaw of the BS framework is the assumed asset price dynamics with constant volatility, only driven by independent Gaussian increments. This has led to extensive research in option pricing theory. More realistic continuous-time models and different concepts of volatility will be introduced in the next chapter.

Chapter 3

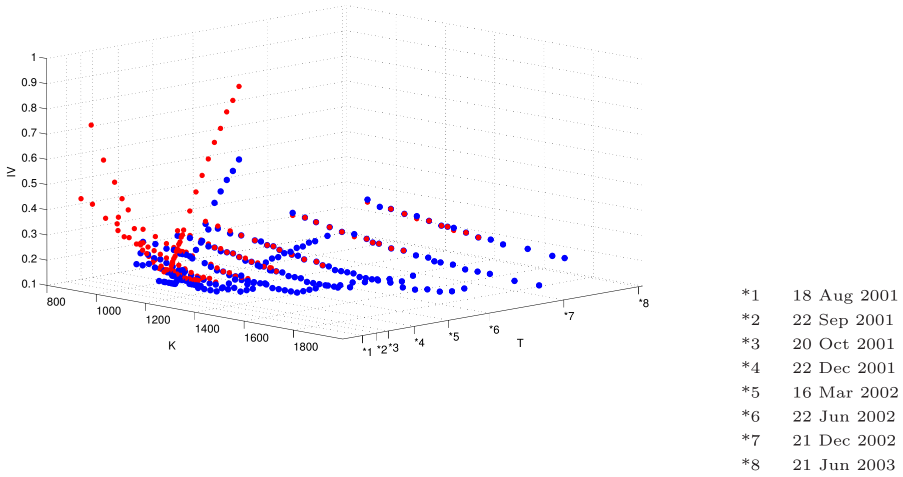
The Implied Volatility Surface

The only unobservable variable in the BS framework is the most crucial one, the volatility σ. By equating the observed market price (C_t, P_t) of an option with the BS price and implicitly solving for

$$\tilde{\sigma}^{\mathrm{IV}}: \quad \mathrm{BS}_t(S_t, \tilde{\sigma}^{\mathrm{IV}}, cp\ flag, K, T, r, q) \overset{!}{=} C_t \mathbb{1}_{cp\ flag=1} + P_t \mathbb{1}_{cp\ flag=0}, \tag{3.1}$$

an implied volatility (IV) can be numerically found. $\tilde{\sigma}^{\mathrm{IV}}$ is unique, due to the monotonicity of the BS price in σ, see Eq. (2.30). According to the BS assumptions, this implicitly calculated volatility should be constant. Cassese and Guidolin remark that *"since Rubinstein (1985), it is well known that option markets are characterized by systematic deviations from the constant volatility benchmark of Black and Scholes (1973), a fact that has become even more evident after the world market crash of October 1987"* (2006, p. 146).

To visualize how far BS assumptions and reality are apart, IVs for options on the S&P 500 index with different strikes K and expiry dates T are calculated on t = 10 August 2001 and plotted in Figure 3.1. IV is not constant as actually assumed for deriving the BS formula. Instead, 'smiles' and 'smirks' across the K-axis as well as a term structure along the T-axis can be seen.



Implied volatilities of S&P 500 index options, t = 10 August 2001

Figure 3.1: Scatter-plot of IVs for 214 calls and 147 puts with different strikes K and expiry dates T. The underlying S&P 500 index closed at 1,190.16 points on 10 August 2001.

Definition 3.1 (IVS in absolute coordinates) *The mapping*

$$\tilde{\sigma}_t^{IV}: \quad (K,T) \longmapsto \tilde{\sigma}_t^{IV}(K,T) \tag{3.2}$$

is called the implied volatility surface (IVS) in absolute coordinates.

Plugging S_t, K, r, T, and $\tilde{\sigma}_t^{\mathrm{IV}}(K,T)$ back in the BS formula leads (by definition of IV) to the observed market price. As it is usually done in the IV literature, we describe the IVS in relative coordinates.

Definition 3.2 (Relative coordinates) *Moneyness is an increasing function in the strike K, in general eventually also depending on time t, expiry date T, the spot price of the underlying security S_t and risk-free interest rate r. If not stated otherwise, moneyness is defined as $m = K/S_t$ throughout this thesis. $\tau = T - t$ is called time to maturity.*

Strike and expiry date are fixed in the contract specification of each option.

	In-the-money (ITM)	At-the-money (ATM)	Out-of-the-money (OTM)
Call	$m < 1$	$m = 1$	$m > 1$
Put	$m > 1$	$m = 1$	$m < 1$

Table 3.1: Moneyness categories for options when moneyness is defined as $m = K/S_t$.

They can be easily derived from relative coordinates, $(K, T) = (m \cdot S_t, \tau + t)$. At any point in time during its lifetime, an option is either in-the-money (ITM), at-the-money (ATM) or out-of-the-money (OTM).

Definition 3.3 (Intrinsic value, time value) *The ITM part of the option value is called intrinsic value at time t,*

$$\max(S_t - K, 0)\mathbb{1}_{cp\ flag=1} + \max(K - S_t, 0)\mathbb{1}_{cp\ flag=0}. \tag{3.3}$$

The time value is the difference between option value and intrinsic value.

The former reflects the (hypothetical) value of immediate exercise of the option and the latter the value of holding on to the option. The time value is usually decreasing as time t approaches the expiry date T. As Hull points out, *"an exception to this could be an in-the-money European put option on a non-dividend-paying stock or an in-the-money European call option on a currency with a very high interest rate"* (2002, p. 310). This can also be seen in the BS framework. From Eq. (2.33) follows that $\theta_t^{\text{BS}} = \frac{\partial \text{BS}_t}{\partial t}$ is not always < 0 and might be positive in the aforementioned cases. A rational option holder exercises a European option at T only if its intrinsic value is positive, which implies that the option is ITM. A monyeness classification is given in Table 3.1.

Definition 3.4 (IVS in relative coordinates)

$$\sigma_t^{IV} : (m, \tau) \longmapsto \sigma_t^{IV}(m, \tau) = \tilde{\sigma}_t^{IV}(m \cdot S_t, t + \tau) \tag{3.4}$$

is called the IVS in relative coordinates.

Figure 3.2 shows the IV plot of the earlier mentioned S&P 500 index options in relative coordinates. ‘Smiles’ and ‘smirks’ appear to form a string; although there are only a few expiry dates, a higher number of options with different strikes per string exist. This is due to institutional and practical conventions. Terms and conditions of exchange-traded options are standardized. The difference between successive expiry dates for the range of small time to maturities ($\tau \leq 3$ months) is usually one month, for large time to maturities three months. It is clear that the scatter-plot looks different on another day as IV levels, string shapes and (m, τ) location are all functions of time.

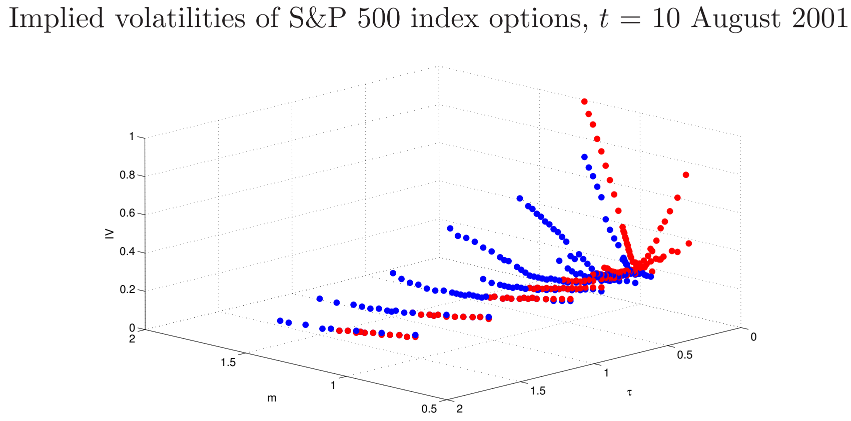

Implied volatilities of S&P 500 index options, $t = 10$ August 2001

Figure 3.2: IVs of S&P 500 index options on $t = 10$ August 2001 are plotted against **relative coordinates** $(m, \tau) = (K/S_t, T - t)$. IVs of 214 calls are blue and IVs of 147 puts are red.

Definition 3.5 (Degenerated option data structure) *The fact that there is only a discrete set of strikes with a small number of maturities available at each moment in time is called the degenerated structure of option data. The data is sparse and unequally distributed over the* (m, τ) *plane and arranged in IV strings that are moving deterministically along the* τ*-axis towards zero and randomly along the* m*-axis according to time.*

3.1 Popularity of Black-Scholes implied volatility

The implicit function theorem for Eq. (3.1) states that a locally unique continuously differentiable function $\tilde{\sigma}^{\mathrm{IV}}$ exists that maps observed market option prices to IV, even if it is not possible to write down the function explicitly. The sufficient condition $\frac{\partial \mathrm{BS}_t}{\partial \sigma} \neq 0$ for its existence is met because, from Eq. (2.32), it follows that the BS vega is strictly positive for call and put options, $\frac{\partial C_t^{\mathrm{BS}}}{\partial \sigma} = \frac{\partial P_t^{\mathrm{BS}}}{\partial \sigma} > 0$. IV can be easily calculated with an iterative numerical procedure like the Newton-Raphson method as the BS vega is explicitely known.

3.1.1 Sublimation of model ambiguity into IV

IV as a mapping of observed option prices to volatility parameters clearly has practical advantages. *"Specifying the implied volatility surface at a given date is therefore synonymous with specifying prices of all (vanilla) calls and puts at that date"* (Cont and da Fonseca, 2002, p. 45). Although IV does not say anything about the 'real' latent instantaneous volatility per se, traders can compare options with different characteristics using one single observable quantity. All model ambiguity is sublimated into σ_t^{IV}. An option with higher IV is priced relatively higher compared to another one with lower IV. Theoretically, even the option type does not play a role because put-call parity (2.36) implies equality of IV for European call and put options on the same underlying with identical strike price and maturity date.

3.1.2 IV as an instrument for empirical option pricing

Option prices can move very quickly. IV as a standardization of option prices tends to be more smooth (Shimko, 1993; Rosenberg, 2000) and less subject to noise, at least when the expiry date is $\tau > 10$ days (see Section 5.1.2). IV needs to be less frequently updated than option prices themselves. The smile curve in Figure 3.2 flattens out with longer time to maturity. Daglish, Hull, and Suo (2007) show that simple rules of thumb (one of them is the sticky moneyness, see Section 3.3.4) for the IVS have a reasonable one day OS prediction performance, even in cases where they are inconsistent with no-arbitrage conditions. Constraints on option prices (see Section 2.4) translate into bounds on IV. Lee (2005) describes the behaviour of IV at extreme strikes. Although the no-arbitrage conditions are quite complex, Kahalé (2004) and Wang, Yin, and Qi (2004) are able to outline interpolation strategies. Fengler (2009) deals with the smoothing of the IVS under no-arbitrage conditions.

Benko (2006) provides a step-by-step derivation for an expression of the state-price density (2.17) in terms of IV by combining Eq. (2.23) and (2.27),

$$\begin{aligned} q_{t,T}(K) = e^{r\tau} S_t \varphi(d_1)\sqrt{\tau} \left\{ \frac{1}{K^2 \sigma_t^{\mathrm{IV}} \tau} + \frac{\partial \sigma_t^{\mathrm{IV}}}{\partial K} \frac{2d_1}{\sqrt{\tau} K \sigma_t^{\mathrm{IV}}} \right. \\ \left. + \frac{\partial^2 \sigma_t^{\mathrm{IV}}}{\partial K^2} + \left(\frac{\partial \sigma_t^{\mathrm{IV}}}{\partial K} \right)^2 \left[\frac{d_1 d_2}{\sigma_t^{\mathrm{IV}}} \right] \right\}. \end{aligned} \tag{3.5}$$

Bliss and Panigirtzoglou (2002); Brunner and Hafner (2003) find that an SPD estimation inferred from the IVS can be superior to classical approaches directly based on option prices (Breeden and Litzenberger, 1978; Aït-Sahalia and Lo, 1998; Britten-Jones and Neuberger, 2000). Latest approaches focus directly on the pricing kernel $M_{t,T} = q_{t,T}/p_{t,T}$, taking into account both option prices and the time series of S_t (Detlefsen et al., 2007; Barone-Adesi, Engle, and Mancini, 2008). Gagliardini, Gouriéroux, and Renault (2008) obtain an estimator of $M_{t,T}$ by an extended method of moments that combines no-arbitrage restrictions with IVS interpolation.

Delta hedging reduces the risk of price movements in the underlying asset. For example, a long equity call option is delta hedged by shorting $\Delta = \frac{\partial C_t}{\partial S_t}$ stocks. The portfolio value process $H_t = C_t - \Delta S_t$ has a zero delta in this way, $\frac{\partial H_t}{\partial S_t} = \frac{\partial C_t}{\partial S_t} - \frac{\partial C_t}{\partial S_t} = 0$. Practitioners often use the BS delta calibrated at the IV as hedge ratio, $\Delta = \Delta_t^{\mathrm{BS}} = \frac{\partial \mathrm{BS}_t}{\partial S_t}(S_t, \sigma_t^{\mathrm{IV}}, \mathit{cp\ flag}, K, T, r, q)$. Barone-Adesi and Elliott (2007) explain the good performance of this so called 'IV compensated BS hedge' with local first degree homogeneity[8] of C_t w.r.t. S_t and K. Fengler (2005, Section 2.9.1) shows how a 'corrected hedge ratio' can be obtained from Δ_t^{BS} when an explicit dependence of IV on S_t is assumed, $\Delta = \Delta_t^{\mathrm{BS}} + \nu_t^{\mathrm{BS}} \frac{\partial \sigma_t^{\mathrm{IV}}}{\partial S_t}$. IVS is a function of (m, τ) and not of S_t, but Lee (2005) proves in a stochastic volatility framework that $\frac{\partial \sigma_t^{\mathrm{IV}}}{\partial S_t} = -\frac{K}{S_t} \frac{\partial \sigma_t^{\mathrm{IV}}}{\partial K}$. The previous examples show how different volatility concepts are connected with each other and consolidates the importance of BS IV as a building block in risk management (see Remarks 2.7, 2.8 and 2.9).

---

[8] This property also holds for BS option prices. For $a > 0$, $\mathrm{BS}_t(aS_t, \sigma, \mathit{cp\ flag}, aK, T, r, q) = a \cdot \mathrm{BS}_t(S_t, \sigma, \mathit{cp\ flag}, K, T, r, q)$, as d_1 and d_2 in Eq. (2.21), (2.22) only depend on S_t/K.

3.1.3 IV as a risk-neutral expectation of future volatility

IV can be interpreted as the market's expectation of average volatility through the lifetime of the option. In a BS similar framework where volatility is allowed to be a deterministic function of time only and under mild conditions (Karatzas and Shreve, 1991, Theorem 5.2.9), the complete market setting is preserved and it can be shown that the fair price of European plain vanilla stock options is given by

$$\mathrm{BS}_t\left(S_t, \sqrt{\bar{\sigma}^2}, cp\ flag, K, T, r, q\right), \tag{3.6}$$

the common BS formula (2.23) with

$$\bar{\sigma}^2 := \frac{1}{T-t} \int_t^T \sigma^2(s) ds. \tag{3.7}$$

In a general stochastic volatility setup (see Appendix C.2), where $\sigma(t, Y_t) = f(Y_t)$ is a function of time and an Itô process Y_t that drives the volatility dynamics, one can show that

$$\mathrm{BS}_t\left(S_t, \mathrm{E}_{\tilde{\mathbb{Q}}}\left[\sqrt{\bar{\sigma}^2}\middle| \mathcal{F}_t\right], cp\ flag, K, T, r, q\right) \tag{3.8}$$

with

$$\bar{\sigma}^2 := \frac{1}{T-t} \int_t^T f(Y_s)^2 ds \tag{3.9}$$

and $\tilde{\mathbb{Q}}$ a risk-neutral measure[9] is only an approximation to the fair price of ATM options if additionally zero correlation between the Brownian motions in the SDE for S_t and the SDE for Y_t is assumed. References to both extensions of the constant volatility case are given by Fengler (2005, Section 2.8).

---

[9] As the market in this setting is incomplete, there is no unique risk-neutral measure.

Eq. (3.7) and (3.9) are time averages of the integrated volatility over the remaining option lifetime. Assuming that the market prices options either with Eq. (3.6) or (3.8), IV can thus be understood as an option based estimator of average future volatility. IV at time t is consequently a forward-looking measure, depending on the expected path of S_t and possibly other factors for $t \in [t, T]$.

3.2 Explaining the smile

We have seen that there are good reasons to stick to the volatility implied by the BS model although the assumption of constant instantaneous volatility is demonstrably false. In order to explain the observed non-flat IVS, the restrictions in the BS model have been relaxed over the last three decades by allowing more flexible asset price dynamics. It is difficult to keep track of the different volatility concepts that have evolved over time. A short introduction to these, including references to the literature, is presented in Appendix C.

3.2.1 Deterministic instantaneous volatility

Assume that instantaneous drift and volatility in the SDE of an asset price dynamics are deterministic functions of the asset price and time only[10],

$$\frac{dS_t}{S_t} = \mu(t, S_t)dt + \sigma(t, S_t)dW_t. \tag{3.10}$$

The complete market setting of the BS model is maintained. The generalized version of the BS PDE (2.20) with $\sigma^2(t, S_t)$ replacing the constant σ^2 is valid for the pricing of any European contingent claim, although analytical solutions do not exist in general, only for particular specifications of instantaneous volatility like $\sigma(t, S_t) := \sigma(t)$ [see Eq. (3.6)]. Under the risk-neutral measure $\mathbb{Q}$, the SDE becomes

$$\frac{dS_t}{S_t} = (r_t - q_t)dt + \sigma(t, S_t)d\tilde{W}_t. \tag{3.11}$$

---

[10] A global Lipschitz and a linear growth condition (Karatzas and Shreve, 1991, Theorem 5.2.9) assure existence and uniqueness of an Itô process S_t that is a solution of the SDE (3.10).

Dupire's formula (C.25) characterizes instantaneous volatility $\sigma(t, S_t)$ expressed in terms of observed market prices such that SDE (3.11) is the unique stochastic process that is consistent with the risk-neutral density obtained directly from observed market prices of European vanilla options (Dupire, 1994; Derman and Kani, 1994).

3.2.2 Local volatility in terms of implied volatility

In broader terms, local volatility $\mathrm{LV}_{T,K}(t, S_t)$ can be defined as an expectation of an instantaneous volatility that depends on additional state variables in a more general setup (see Appendix C.3). It is presumed that *"the instantaneous volatility will evolve entirely along today's market expectations sublimated in the local volatility function"* (Fengler, 2005, p. 50). Under these assumptions, Eq. (3.10) together with

$$\sigma(t, S_t, \cdot) = \mathrm{LV}_{t,S_t}(t, S_t) \equiv \mathrm{LV}_{T,K}(t, S_t)|_{T=t,K=S_t}$$

preserves the complete market setting. The generalized BS PDE (2.20) still prices any European contingent claim and Dupire's formula (C.25) also holds. The local volatility surface $(T, K) \longmapsto \mathrm{LV}_{T,K}(t, S_t)$ can be recovered from observed market prices C_t via binomial or trinomial trees[11]. Parametric approaches are given by the constant elasticity of variance model (Cox and Ross, 1976), polynomial LV functions (Dumas et al., 1998; McIntyre, 2001) or the LV mixture diffusion model (Brigo and Mercurio, 2001).

Calibrating LV models to observed market prices is an ill posed inverse problem (Fengler, 2005, Section 3.10.3). The obtained local volatility surface can be very rough and spiky, contradicting any intuition. LV models have recently been criticized for several other reasons (see Remark C.17) and practitioners seem to prefer stochastic volatility models.

If one is still willing to use LV models after all criticism and if a 'reasonable' dynamic IVS model is available that allows arbitrage-free interpolation in the (m, τ)

---

[11] See for example Rubinstein (1994), Derman, Kani, and Chriss (1996), Jackwerth (1997), Derman and Kani (1998) and Britten-Jones and Neuberger (2000).

domain, then a link from IV to LV can be exploited. $C_t = C_t^{\mathrm{BS}}(S_t, \sigma_t^{\mathrm{IV}}, K, T, r, q)$ is used to replace C_t, $\frac{\partial C_t}{\partial T}$, $\frac{\partial C_t}{\partial K}$ and $\frac{\partial^2 C_t}{\partial K^2}$ in Dupire's formula (C.25). Fengler (2005) derives an expression in terms of the IVS and its derivatives,

$$\mathrm{LV}_{T,K}(t, S_t) = \sqrt{\frac{\frac{\sigma_t^{\mathrm{IV}}}{\tau} + 2\frac{\partial \sigma_t^{\mathrm{IV}}}{\partial T} + 2K(r-q)\frac{\partial \sigma_t^{\mathrm{IV}}}{\partial K}}{K^2 \left\{ \frac{1}{K^2 \sigma_t^{\mathrm{IV}} \tau} + 2\frac{d_1}{K \sigma_t^{\mathrm{IV}} \sqrt{\tau}} \frac{\partial \sigma_t^{\mathrm{IV}}}{\partial K} + \frac{d_1 d_2}{\sigma_t^{\mathrm{IV}}} \left(\frac{\partial \sigma_t^{\mathrm{IV}}}{\partial K} \right)^2 + \frac{\partial^2 \sigma_t^{\mathrm{IV}}}{\partial K^2} \right\}}}. \tag{3.12}$$

Gatheral (2006, p. 13) does the same in terms of the BS implied total variance $w := \sigma_t^{\mathrm{IV}} \tau$ and the log-strike $y := \log(K/F_{t,T})$, where $F_{t,T} = S_t \exp\left\{\int_t^T (r_t - q_t)dt\right\}$ denotes the forward price of S_t.

3.2.3 Stochastic volatility

Stochastic volatility (SV) models allow the volatility itself to be a stochastic process (see Appendix C.2). The financial market is thus incomplete and option pricing is no longer preference free (see Note C.9). An additional assumption about market price of volatility risk is needed[12] to identify the risk-neutral pricing measure $\mathbb{Q}$. Bakshi and Kapadia (2003) summarize evidence for systematic market volatility risk. Empirically observed equity index options are found to be non-redundant securities and the volatility risk premium is negative. The latter is no longer fully supported when correlation risk is incorporated within an asset pricing framework (see Remark C.10).

Modern SV models (Remark C.11) capture a great deal of empirically observed features of the asset price dynamics: SV accounts for longer dated smiles on the IVS, jumps for shorter dated smiles and default risk. The parameters of a SV model can be very difficult to estimate, but they have a clear financial interpretation, as opposed to the parameters of LV models. Affine jump diffusions

---

[12] In the early literature, the market price of volatility risk is often assumed to be zero (Hull and White, 1987; Scott, 1987) or constant (Stein and Stein, 1991). The SDE under the risk-neutral measure of modern SV models is assumed to be of the same type as under $\mathbb{P}$, which implicitly requires a risk-adjustment of the coefficients and determines the form of the market price of volatility risk (see e.g. Jones, 2003).

(Duffie, Pan, and Singleton, 2000) are a very popular class of SV models because of their analytical solutions for option pricing (see Note C.12).

3.3 Modelling IVS directly

The presented volatility concepts have their deficiencies, but the relationships amongst them are well understood and could be usefully exploited for typical financial purposes (hedging, pricing, trading strategy) when starting from a proper dynamic IVS model that is well defined over the whole (m, τ) domain.

3.3.1 IVS as a link to other volatility concepts

Deterministic volatility models have been generalized to LV models since instantaneous volatility might also depend on stochastic variables other than S_t. Recovering the LV surface from a few observed option prices is very sensitive to changes in the data and LV is better recovered from observed IV. Derman and Kani (1998) and Britten-Jones and Neuberger (2000) apply stochastic perturbation techniques to merge trinomial trees and stochastic volatility. The stochastic nature of IV can also be modelled directly with a SDE for fixed K and T (Schönbucher, 1999; Ledoit, Santa-Clara, and Yan, 2002). Calibrated to a single contingent claim, stochastic IV fails to accurately reprice options with different strike and time to maturity. Durrleman (2004) shows that the spot volatility dynamics can be expressed in terms of the IVS dynamics. Durrleman (2008) extends the BS robustness formula of El Karoui, Jeanblanc-Picqué, and Shreve (1998) to the case of jumps with finite variation and proves for $m_{\text{ATM}} = 1$ that

$$\lim_{\tau \downarrow 0} \sigma_t^{\text{IV}}(m_{\text{ATM}}, \tau) = \sigma(t, S_t) \ \ \mathbb{P}\text{-a.s.} \tag{3.13}$$

with the help of a central limit theorem for martingales.

All of these different volatility concepts have the aim to explain the IVS smile in common, but it is apparent that the concepts are either more efficiently estimated or even appointed to some kind of structural form by IV. An analogy to this is a snake that bites its own tail, which is not saying that these volatility concepts are useless just because the body of a snake performing such an act forms a circle. This illustrative thought has at least once advanced natural sciences, namely when Kekulé discovered the structure of benzene in organic chemistry (Benfey, 1958).

3.3.2 Predictor space

In order to directly model the IVS in a general statistical framework, let us introduce the following predictor space

$$\mathbf{x}^{\text{pred}} := (m, \tau, \textit{cp flag}, \textit{factors}), \tag{3.14}$$

where *cp flag* allows the IVS to depend on the option type. While $m = K/S_t$ and $\tau = T - t$ are time dependent, *cp flag* is a categorical variable that takes on either 1 (call) or 0 (put). According to Noh, Engle, and Kane (1994), there are advantages in separately modelling the IVS for call and put options. Ahoniemi and Lanne (2009) also distinguish between calls and puts in their bivariate mixture multiplicative error model to fit Nikkei 225 index options. They argue that

> *"With no market imperfections such as transaction costs or other frictions present, option prices should always be determined by no-arbitrage conditions, making the implied volatilities of identical call and put options the same. However, in real-world markets the presence of imperfections may allow option prices to depart from no-arbitrage bounds if there is, for example, an imbalance between supply and demand in the market"* (2009, p. 239)[13].

Figure 3.2 shows the violation of the put-call parity for small τs.

An arbitrary number of exogenous[14] factors, directly or indirectly time dependent, is represented by *factors* in the predictor space. As already indicated in Section 3.2.2, instantaneous volatility might depend on more than just the underlying stock price dynamics. It is not appropriate to use the three month US Treasury-bill rate, fixed at the day when an option is issued, as the constant r in the BS formula to price an option with one year until expiry. Allowing for stochastic interest rates or including proxies for the term structure of interest rates is common in modern option pricing. Other possible exogenous factors would be implied asset prices (Garcia, Luger, and Renault, 2003), the bid-ask spread, net buying pressure (Bollen and Whaley, 2004), trading volume, other stock or index returns.

---

[13] In Section 2, Ahoniemi and Lanne support their reasoning by various references to the option literature.

[14] An exogenous factor is uncorrelated with the error term in a classical linear regression framework. Here, the term 'exogenous' refers only to explanatory variables 'from outside the system'. It is only assumed that the sigma algebra generated by the underlying stock price and all factors at time t is $\subseteq \mathcal{F}_t$, see Remark C.6, similar to a dependent stochastic regression setting. Classical exogeneity is a sufficient, but not necessary, condition for this assumption.

3.3.3 Challenges

At least two challenges arise from defining IV as a function of $\mathbf{x}^{\text{pred}}$ in general,

$$\sigma_t^{\text{IV}}(\mathbf{x}^{\text{pred}}) = \sigma_t^{\text{IV}}(m, \tau, \textit{cp flag}, \textit{factors}). \tag{3.15}$$

1. The set of *factors* could eventually contain all observed option prices, e.g. when a smoothing method like a kernel regression procedure is used to model the IVS. It is not clear how to obtain out-of-sample (OS) predictions from such a model that is only dynamic because the observed data and the number of observations change from day to day in-sample (IS). However, it is not able to incorporate the dynamics of the IVS as a whole. What kinds of nonparametric or semi-parametric IVS models can be calibrated to observed data and evaluated at any (m, τ) location?

2. IV can be seen as predictor of future volatility, so from a supervised learning perspective time-lagged as well as (forecasts of) time-leading factors could be included in the predictor space. Consider the latter as conditional expectations under $\mathbb{P}$ on the information availabe today. How is the IVS predicted OS? What further assumptions on the (multivariate) time series of exogenous factors are required?

Example 3.6 Suppose daily ATM IV and end-of-day settlement prices for the underlying asset over the n most recent days (= IS period) have been recorded. Time is denoted by $t \in \mathbb{N}$, today is $\bar{t} = N$ and IS $= \{1, 2, ..., N\}$. ATM IV shall be modelled as ATM $\mathrm{IV}_t = f(\mathit{factors}_t)$ with f a not otherwise specified statistical learning function and $\mathit{factors}_t = \{\mathrm{E}_{\mathbb{P}}[S_{t-1}|\mathcal{F}_{\bar{t}}], \mathrm{E}_{\mathbb{P}}[S_t|\mathcal{F}_{\bar{t}}], \mathrm{E}_{\mathbb{P}}[S_{t+1}|\mathcal{F}_{\bar{t}}]\}$. Then the subset $\mathrm{IS}\backslash\{1, N\}$ can be used without further assumptions on the dynamics of S_t for supervised learning since ATM $\mathrm{IV}_t = f(S_{t-1}, S_t, S_{t+1})$ for $t \in \{2, 3, \ldots, N-1\}$.

3.3.4 Models

Statistical modelling is a methodical procedure. At its core lies the specification of a model (dependent/independent variables, forms of interaction and relationships amongst them, degrees of freedom). When modelling the IVS directly, it is self-evident that we want IV to be explained depending upon m and τ. The model is eventually also based on a number of parameters or other predictors and is restricted to a certain form. A model selection criteria that balances goodness of fit with complexity (number of free parameters) can be consulted to determine the best model among a set of possible models. Model fitting deals with finding the best values for the free parameters of a specified model. An estimate is obtained through optimization of a fit criterion (e.g. the likelihood). A model is calibrated to observed data when predictions of the fitted model correspond very closely with observed data.

The five possible IVS models that are presented here act as a starting point to a new methodology for building semi-parametric models that take up the challenges posed in the previous section. In this context, the models only need to provide a first rough approximation of the true IVS model (to leave room for improvement in a statistical learning framework). Default values are imposed for certain parameters to reduce the model complexity.

Regression tree (regtree) We fit a regression tree with 10 leaves to all observed IS call options and another regression tree with 10 leaves to all observed IS put options. Thus, the model depends on three location parameters (m, τ, *cp flag*) and 36 regression tree parameters (9 split variables and 9 cut values per regression tree). Positivity of the IVS is guaranteed since the model depends on the aggregated observed positive IVs. See also Section 4.1 for further information.

Ad hoc BS model (adhocbs) Dumas et al. (1998) performed a goodness-of-fit test for several functions of quadratic form in a deterministic volatility framework. They found that the best parametrization was given by

$$\sigma(K,\tau) = \max(0.01, a_0 + a_1 K + a_2 K^2 + a_3\tau + a_4 K\tau).$$

Since relative coordinates $m = K/S_t$ and $\tau = T - t$ are used in this thesis, the model

$$\sigma^{\mathrm{IV}}_{ti} = a_{t0} + a_{t1} m_{ti} S_t + a_{t2}(m_{ti}S_t)^2 + a_{t3}\tau_{ti} + a_{t4} m_{ti} S_t \tau_{ti} + \epsilon_{ti} \tag{3.16}$$

is fitted by least square, using observations on day t. In case of negatively estimated IV, values are also set to 0.01. The adhocbs model depends on (m,τ), *factors* $= \{S_t\}$ and time-varying parameters $\mathbf{a}_t = (a_{t0}, a_{t1}, a_{t2}, a_{t3}, a_{t4})$. The last IS day $\bar{t}$ is used as the reference day. Set $\mathbf{a}_{\tilde{t}} = \mathbf{a}_{\bar{t}}$ to evaluate the IVS model on a future date $\tilde{t} > \bar{t}$.

Sticky moneyness model (stickym) The term 'sticky moneyness' denotes a broad class of 'naïve trader models' (Cont and da Fonseca, 2002; Daglish et al., 2007). These models assume time invariance of the IVS for fixed moneyness. Such an assumption is only realistic for a short period and has to be understood in a 'relative coordinate IVS random walk' sense for OS predictions: the best guess for a point on the IVS of tomorrow at a fixed m location is a point on the IVS of today with the same m location. Further assumptions on the τ location are required to fully identify the point.

The stickym model is defined here in the following way: IS evaluation at any (m,τ) location is provided by data gridding. The focus of the used interpolation method[15] is set to the geometrical aspect of the observed data $\{(m_{ti}, \tau_{ti}, \sigma^{\mathrm{IV}}_{ti}) | i =$

---

[15] First, Delaunay triangulation is applied. The algorithm forms special triangles out of any given set of scattered data points in the (m,τ) plane such that the minimum angle of all triangles is maximized. Next, an estimate of the IVS is obtained via cubic interpolation over these triangles. Delaunay triangulation is important for computer graphics and finite element methods to numerically solve PDEs. Watson (1992) is a reference for Delaunay triangulation-based applications in spatial data analysis.

$1, ..., L_t\}$ on day t; the IVS is smooth, the monotonicity and the shape of the data are preserved. No extrapolation is conducted, and out-of-range values are set to the average IV on day t. The term structure of the IVS at the last IS day $\bar{t}$ is used to interpolate the IV on a future date $\tilde{t} > \bar{t}$.

Bayesian vector autoregression (bvar) Doan, Litterman, and Sims (1984) introduced a spatial econometric model that uses Bayesian prior information to overcome problems with high correlations in the data.

We implement the model as follows: First, a linearly spaced 10 × 10 grid with values from $m = 0.2$ to 2 and from $\tau = 1/365$ to 3 is laid in the (m, τ) domain. For each IS day, IV is estimated on this grid using a Nadaraya-Watson estimator with a normal product kernel and stepwidth set according to the normal reference rule[16].

Next, a Bayesian vector autoregression model of order 2 is fitted to this 100 dimensional time series of IV estimates on the fixed (m, τ) grid. A normal distributed prior with mean 1 for coefficients associated with the lagged dependent variable in each equation of the vector autoregression and mean 0 for all other coefficients is imposed. In equation i of the vector autoregression, the standard deviation of the prior imposed on the dependent variable j at lag k is

$$\mathrm{sd}_{ijk} = 0.2 \frac{w(i,j)}{k} \frac{\widehat{\mathrm{sd}}_{uj}}{\widehat{\mathrm{sd}}_{ui}}$$

where $\widehat{\mathrm{sd}}_{ui}$ is the estimated standard error from a univariate autoregression involving variable i and $W = [w(i,j)]_{i=1,...,100}^{j=1,...,100}$ is a matrix containing the values 1 for $i = j$. If the grid points corresponding to time series components i and j are neighbours, $w(i,j) = 0.8$ is set. All other entries of W are set to 0.1. An internal grid point has at most eight neighbours. As a result of all these arrangements, only ca. 2% of the parameters are estimated significantly different from zero. LeSage (1999, p. 126) explains Bayesian vector autoregression models in the manual of his Econometrics Toolbox and provides functions to estimate, evaluate and forecast them. Between the fixed 100 grid points, we apply bicubic interpolation to evalutate the IVS at any (m, τ) location.

---

[16] Scott (1992) shows how to minimize the asymptotic integrated squared bias and asymptotic integrated variance for the multivariate normal product kernel. In the bivariate case, it is given by the sample standard deviation times number of observations to the power of $-1/6$.

Dynamic semiparametric factor model (dsfm) Fengler, Härdle, and Mammen (2005; 2007) describe dsfm as a type of functional coefficient model. *"Surface estimation and dimension reduction is achieved in one single step.* [The dsfm] *can be seen as a combination of functional principal component analysis, nonparametric curve estimation and backfitting for additive models"* (2005, p. 6).

The dsfm model consists of smooth basis functions g_k that are multiplied by time-varying latent factor loadings $\beta_{t,k}$,

$$\underbrace{\sigma^{\mathrm{IV}}(m_{ti}, \tau_{ti})}_{=\sigma^{\mathrm{IV}}_{ti}} = g_0(m_{ti}, \tau_{ti})) + \sum_{k=1}^{K} \beta_{t,k} g_k(m_{ti}, \tau_{ti}) + \varepsilon_{m_{ti}, \tau_{ti}}. \tag{3.17}$$

The model is fitted to aggregated observed data $t \in \mathrm{IS} = \{1, \ldots, N\}$, $(m_{ti}, \tau_{ti}, \sigma^{\mathrm{IV}}_{ti})$, $i \in \{1, ..., L_t\}$ by minimizing a localized least square criterion. The estimated $\hat{g}_k$ and $\hat{\beta}_{t,k}$ are not uniquely defined. The $\hat{g}_k$ are iteratively orthogonalized such that each $\sum_{t=1}^{N} \hat{\beta}^2_{t,k}$ is maximized.

$K = 4$ smooth basis functions are chosen, each a linear combination of cubic B-splines on a uniformly spaced knot sequence of length 6 between the minimum and maximum of time-aggregated observed (m, τ) locations. The algorithm directly runs on IV rather than on log IV. This improves OS prediction of the approximated latent factor loadings.

Remark 3.7 A comparison of a 'complexity reduced' model (where default values are used for parameters) with the 'improved' version (where that 'complexity reduced' model is used as a starting model in a statistical learning framework) is obviously in favour of the latter. To be fair, the 'improved' model should be compared to the 'best fitted' model (first model selection, then fitting the best model to observed data).

Example 3.8 Let us compare the OS performance of the regtree model with an improved version of regtree.

- First, the 'complexity reduced' regtree model is calibrated to observed data and IV predictions at some future date are obtained. These forecasts are referred to as A.
- Next, a statistical learning algorithm is applied to obtain the improved version of regtree and OS predictions of the IV. These forecasts are referred to as B.
- Last, the set of regtree models with $\{10, 20, \ldots, 100\}$ leaves (separately fitted to calls and puts) is analyzed and the best model is selected, for example the regtree model with 60 leaves for calls and 40 leaves for puts because this combination minimizes the expected square prediction error approximated by the same cross-validation scheme that was adopted in the statistical learning algorithm. The OS predictions of the 'best fitted' regtree model are referred to as C.

Forecasts A are worse than forecasts B, provided the statistical learning algorithm does not overfit the data. Forecasts B are expected to be better than forecasts C, otherwise the purpose of statistical learning is defeated. Model selection for the best fitted model can become computationally infeasible when increasing the size of the set of possible models.

Chapter 4

Supervised learning

The pedagogical paradigm of supervised learning is closely related to what is known as cognitivism in social sciences. Learning occurs not only within the human brain, but mainly through or as a consequence of social interactions with other individuals. Learning follows the actio et reactio principle: the learner puts effort into solving a task (willingness to learn), the teacher provides feedback in the form of a correct answer or by showing a way to solve the problem. The learner increases her knowledge by adapting her behaviour according to the received feedback. Supervised and unsupervised learning are distinguished by the availability of feedback.

The meaning of supervised learning in the theory of statistical learning is similar. It denotes a technique of function approximation in a prediction framework, where given inputs X are related to outputs Y. The available feedback is represented by the training sample $\mathcal{T} = \{(x_i, y_i) | i \in \{1, 2, \ldots, N\}\}$. The goal of supervised learning is to find a useful approximation $\hat{f}(x)$ to the function $f(x)$ relating x to y for $(x, y) \in \mathcal{U} \supset \mathcal{T}$ in general.

A statistical model for the joint distribution of X and Y addresses uncertainty between input and output variables. For example, the additive error model $Y = f(X) + \varepsilon$ assumes that ε has $\mathrm{E}[\varepsilon] = 0$ and is independent of X. Hastie, Tibshirani, and Friedman (2009) introduce supervised learning in their book in the context of such an additive error model.

> *"Here the data pairs (x_i, y_i) are viewed as points in a $(p+1)$-dimensional Euclidean space. The function $f(x)$ has domain equal to the p-dimensional input subspace, and is related to the data via a model such as $y_i = f(x_i) + \varepsilon_i$. For convenience … we will assume the domain is $\mathbb{R}^p$, a p-dimensional Euclidean space, although in general the inputs can be of mixed type. The goal* [of supervised learning] *is to obtain a useful approximation to $f(x)$ for all x in some region of $\mathbb{R}^p$, given the representations in $\mathcal{T}$. Although somewhat less glamorous than the learning paradigm, treating supervised learning as a problem in function approximation encourages the geometrical concepts of Euclidean spaces and mathematical concepts of probabilistic inference to be applied to the problem. This is the approach taken in this book"* (2009, p. 29).

Supervised learning methods can be split into two main categories according to the structural assumptions of the models. The first category of local methods contains unstable models (high variance, low bias) that suffer from the curse of dimensionality. The nearest neighbors estimator is a type of instance-based learner that belongs to this first category. The second category consists of more structured regression models that are more stable (low variance, high bias). A variety of different classes of restricted estimators are contained in this category. Complexity restrictions that are imposed by most learning methods guarantee model identification. A complexity or smoothing parameter controls the variance-bias tradeoff.

Statistical decision theory requires a loss function $\lambda(Y, f(X))$ for penalizing prediction errors. A quadratic loss $\lambda(Y, f(X)) = (Y - f(X))^2$ leads to an optimal choice of

$$\hat{f}(x) = \mathrm{E}[Y|X = x] = \arg\min_{f} \mathrm{E}[\{Y - f(X)\}^2] = \int (y - f(x))^2 \ \mathbb{P}(dx, dy), \quad (4.1)$$

when the goodness of fit is measured by average squared error. The nearest neighbors estimator is based on this loss function and directly estimates $\mathrm{E}[Y|X = x]$ in a neighborhood of x.

In the following sections, two structured regression techniques that belong to the family of linear expansions in the large class of methods that depend on basis functions are discussed,

$$f_\theta(x) = \sum_{j=1}^{M} \theta_j h_j(x), \tag{4.2}$$

where $f_\theta(\cdot)$ is linear in the parameter $\theta \in \mathbb{R}^M$ and $h_j : \ \mathbb{R}^p \mapsto \mathbb{R}$, $j = 1, 2, \ldots, M$ are basis functions.

4.1 Classification and regression trees

Breiman et al. (1984) introduce the classification and regression trees (CART),

$$f_\Theta(x) = \sum_{j=1}^{M} c_j \mathbb{I}_{x \in S_j}, \tag{4.3}$$

with $\Theta = \{S_j, c_j\}_{j=1}^{M}$. In a regression framework, where both the response variable Y and the predictor variable X are continuous, S_j is of the form $S_j(u,v) = \{X \in \mathbb{R}^p | X_u \leq v\}$. X_u is called the split variable, v the cut value.

Each component of the predictor variable is checked for a best cut value, such that the resulting two groups are homogeneous with respect to the response variable. The split that yields the smallest variance within a group is chosen, and the procedure is repeated, leading to a sequence of binary splits that forms a maximal regression tree. The tree is then pruned by a cross-validation scheme to prevent it from over-fitting.

In each step of the iteration, $\hat{S}_j(u,v)$ is determined such that

$$\min_{u,v} \left[\min_{c_j} \sum_{i:\ x_i \in S_j} (y_i - c_j)^2 + \min_{c_{j+1}} \sum_{i:\ x_i \notin S_j} (y_i - c_{j+1})^2 \right] \tag{4.4}$$

and

$$\hat{c}_j = \frac{1}{\sum_{i=1}^{N} \mathbb{I}_{x_i \in S_j}} \sum_{i=1}^{N} y_i \mathbb{I}_{x_i \in S_j}. \tag{4.5}$$

Example 4.1 In the case of IVS modelling, assuming *factors* = $\{X\}$, a very simple example of a regression tree with three end-nodes (or leaves) may be the following:

$$\hat{f}_{\hat{\Theta}}(\mathbf{x}^{\text{pred}}) = \begin{cases} \hat{c}_1 & \text{if } m \leq \hat{v}_1 \\ \hat{c}_2 \ , & \text{if } m > \hat{v}_1 \text{ and } X \leq \hat{v}_2, \\ \hat{c}_3 \ , & \text{if } m > \hat{v}_1 \text{ and } X > \hat{v}_2. \end{cases}$$

Example 4.2 Figure 4.1 displays separately fitted regression trees for calls and puts that only depend on predictor variables m and τ (as described for the regtree model in Section 3.3.4).

4.2 Functional gradient descent

Gradient descent methods are an iterative way for finding a minimum of a function f of several real-valued variables. The negative gradient $g_j = -\nabla f(P_j)$ is the direction of the steepest descent at the point P_j. In the line search step, we find $\lambda_j \in \mathbb{R}$, such that $P_{j+1} = P_j + \lambda_j g_j$ is the lowest point along this path. Iterating those two steps leads to a sequence of points which converges to the minimum of f. The drawback of this method is that it converges slowly for functions which have a long, narrow valley. A better choice for the direction would be in this case the conjugate gradient.

Applying the steepest descent method in a function space $\mathcal{F} = \{f \mid f : \mathbb{R}^p \longrightarrow \mathbb{R}\}$ leads, as the name indicates, to the functional gradient descent (FGD) technique. Based on data $\{(x_i, y_i) | i \in \{1, 2, \ldots, N\}\}$, an estimation of a function $F \in \mathcal{F}$ is developed which minimizes an expected loss function $\mathrm{E}[\lambda(Y, F(X))]$, where $\lambda : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}^+$. The FGD estimate of $F(\cdot)$ is found by minimizing Λ, the empirical risk, defined as:

$$\Lambda(F)(x_1, ..., x_N, y_1, ..., y_N) = \frac{1}{N} \sum_{i=1}^{N} \lambda(y_i, F(x_i)). \tag{4.6}$$

Starting from an initial function $\hat{F}$, the steepest descent direction would be given by the negative functional derivative $-d\Lambda(\hat{F})$. Due to smoothness and regularization constraints on the minimizer of $\Lambda(\hat{F})$, we must restrict the search to finding

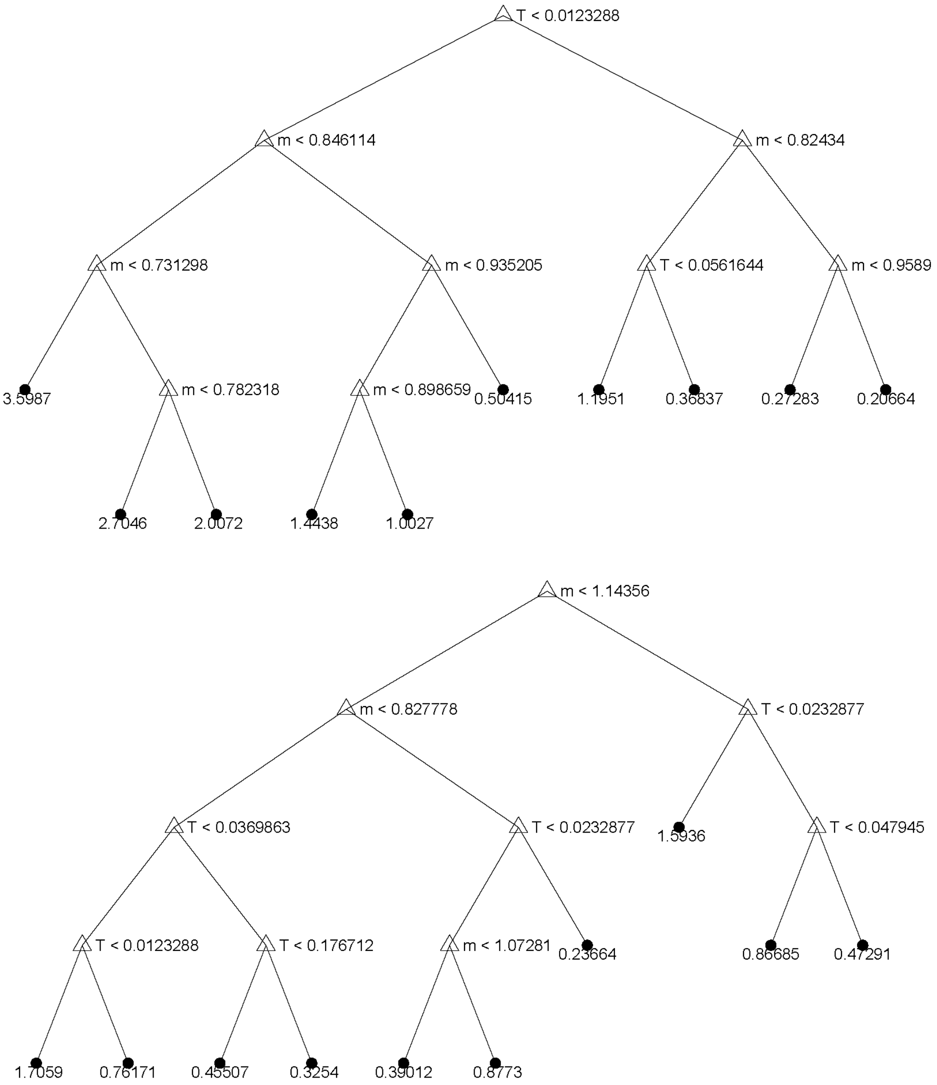

T < 0.0123288
m < 0.846114
m < 0.82434
m < 0.731298
m < 0.935205
T < 0.0561644
m < 0.9589
3.5987
m < 0.782318
m < 0.898659
0.50415
1.1951
0.36837
0.27283
0.20664
2.7046
2.0072
1.4438
1.0027
m < 1.14356
m < 0.827778
T < 0.0232877
T < 0.0369863
T < 0.0232877
1.5936
T < 0.047945
T < 0.0123288
T < 0.176712
m < 1.07281
0.23664
0.86685
0.47291
1.7059
0.76171
0.45507
0.3254
0.39012
0.8773

Figure 4.1: Starting model $F_0(m, \tau, \mathit{cp\ flag})$, a three location regression tree (see Section 3.3.4), is fitted on subsample 3 of S&P 500 index options (see Table 6.3). The *upper* panel displays the regression tree fitted on call options, the *lower* panel the regression tree fitted on put options. If a condition at a split point is met, one proceeds to the left in the graphical representation of the regression tree. Hence, the IV predicted with $\hat{F}_0(\cdot)$ for $m = 0.9$, $\tau = 0.1644$, $\mathit{cp\ flag} = 0$ is 0.23664.

a function $\hat{f}$ which is in the linear span of a class of simple base learners $\mathcal{S}$ and close to $-d\Lambda(\hat{F})$ in the sense of a functional metric. This is equivalent to fitting the base learner $h(x,\theta) \in \mathcal{S}$ to the negative gradient vectors:

$$U_i = - \left. \frac{\partial \lambda(Y_i, Z)}{\partial Z} \right|_{Z=\hat{F}(X_i)}, \quad i = 1, ..., N \tag{4.7}$$

The minimal function $F \in \mathcal{F}$ is approximated in an additive way with simple functions $\hat{f}_j(\cdot) = h(\cdot, \hat{\theta}_{U,X}) \in \mathcal{S}$:

$$\hat{F}_M(\cdot) = \sum_{j=0}^{M} \hat{w}_j \hat{f}_j(\cdot), \tag{4.8}$$

where the $\hat{w}_j$s are obtained in a line search step as in the previous procedure.

Note 4.3 FGD is the statistical view of boosting (Friedman, Hastie, and Tibshirani, 2000; Friedman, 2001). Bühlmann and Hothorn (2007, Section 2) explain in a historical context how boosting was understood to have a represention as steepest descent algorithm in function space. Properties of L_2 Boosting, i.e. FGD for regression problems with squared error loss which amounts to repeated fitting of ordinary residuals, are presented in (2007, Section 5). It is worth pointing out that there is a connection between L_2 Boosting and the Lasso (Tibshirani, 1996) and that regularizations (early stopping, shrinkage etc.) play an important role in boosting. Both mentioned properties give insight to asymptotic convergence results (2007, Section 9.2).

Remark 4.4 AdaBoost (Freund and Schapire, 1997) is the most well-known boosting algorithm for binary classification. Overfitting is a major concern from the statistical view of boosting as an optimization in function space; overfitting deters the algorithm from being consistent[17]. The misclassification error of AdaBoost for an increasing number of iterations $M \to \infty$ does not approach the Bayes error arbitrarily close for N big enough in the case of a $p = 1$ dimensional predictor space, unless regularization is employed (Jiang, 2004). The proof of this proposition relies on the equivalence between AdaBoost and the nearest neighbor classifier. Mannor, Meir, and Mendelson (2001); Mannor, Meir, and Zhang (2003) provide sufficient conditions for the consistency of boosting algorithms in the case where $p > 1$ and for deterministic data. Mease and Wyner (2008) construct counterexamples of noisy data on which boosting with larger classification trees overfits the data less than with conventionally weak learners. It remains an open question why AdaBoost is quite immune against overfitting in practice. Mease and Wyner empirically find a self-averaging performance improvement effect and other evidence against the statistical view of boosting.

---

[17] The classification rule converges in probability to the optimal classification rule (Bayes rule).

Remark 4.5 Statistical learning algorithms are becoming popular in finance. Audrino and Bühlmann (2003) successfully applied FGD to estimate volatility in high-dimensional GARCH models. Gavrishchaka (2006) provided an overview of boosting-based frameworks in finance. Audrino and Trojani (2007) used it to model interest rates. Rossi and Timmermann (2009) modelled the shape of the risk-return relation with a boosting algorithm.

Chapter 5

Model and estimation procedure

In this chapter, a new methodology to capture the salient features for modelling the IVS is proposed. Semi-parametric models can be easily estimated using a classical boosting algorithm based on an additive expansion of simple fitted regression trees that does not resort to variance reduction techniques like factor analysis or PCA to forecast the dynamics of the IVS. These new models take into account the degenerated structure of option data and no information is discarded through preliminary filtering.

5.1 Desired properties

In the statistical theory of point estimation, unbiasedness, consistency and efficiency are goodness properties of estimators and play an important role. To thoroughly analyse and predict the time-varying dynamics of the IVS, the aspects of handling observed data and goodness of IVS models also need to be addressed. The new methodology is based on the following catalogue of desired properties.

5.1.1 Improve upon a starting model

Let F_0 denote a generic function that models the IVS with predictor space defined as in Eq. (3.14),

$$\sigma_t^{\mathrm{IV}}(m,\tau) = F_0(m,\tau,\mathit{cp\ flag},\mathit{factors}) + \epsilon_t(m,\tau). \tag{5.1}$$

Eventually, the IVS model needs to be fitted to historical data during the in-sample (IS) period. The fitted model is denoted $\hat{F}_0$ and the residuals

$$\hat{\epsilon}_t(m,\tau) = \sigma_t^{\mathrm{IV}}(m,\tau) - \hat{F}_0(m,\tau,\mathit{cp\ flag},\mathit{factors}) \tag{5.2}$$

are typically different from zero for any option at any time $t \in \mathrm{IS}$.

Example 5.1 End-of-day settlement data for Microsoft stock options during 10 December 1996 – 2 March 1999 (IS period of 500 sample days) are obtained from OptionMetrics Ivy database (99,445 observations). The five IVS models of Section 3.3.4 (regtree, adhocbs, stickym, bvar and dsfm) are fitted to the IS data. Every day, the IV of the call and put option closest to $m \approx 1$ and $\tau >\approx 10$ days is estimated for each model $\hat{F}_0$. The ex-post performance of each model is depicted in Figure 5.1 and 5.2 by plotting the estimated residuals.

They seem to lie most of the time in the band delineated by the horizontal dotted lines. The stickym model has very small residuals (besides a few exceptions). There is enough room for improvements for the relatively simple regtree model, which assigns a constant ex-post IV of either 0.3542 or 0.4361 for calls and 0.4152 for puts over the whole IS period. The other models (adhocbs, bvar and dsfm) are of higher complexity but their residuals show similar behaviour. This might be due to the chosen (m,τ) locations (near to expiry ATM options).

Comparison of estimated (**call** option) residuals for different IVS models

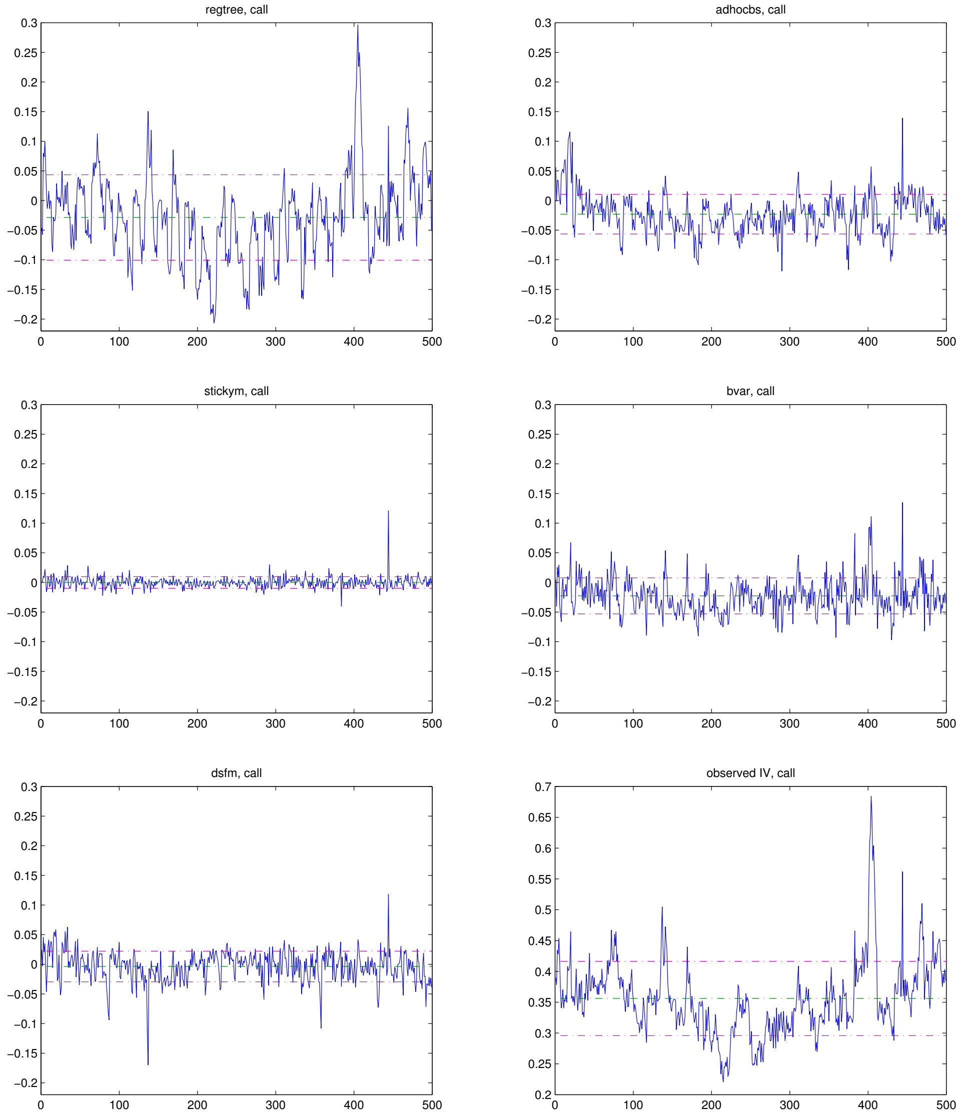


Figure 5.1: In the *lower right* panel, daily observed IVs for Microsoft **call** options closest to $m \approx 1$ and $\tau >\approx 10$ days are plotted against time (IS period of 500 sample days between 10 December 1996 – 2 March 1999). The *other panels* contain time series of estimated residuals $\sigma_t^{\mathrm{IV}}(m,\tau) - \hat{F}_0(m,\tau,\cdot)$ for these call options obtained with different IVS models, $F_0 \in \{$regtree, adhocbs, stickym, bvar and dsfm$\}$. $\hat{F}_0$ is fitted to the IS data. The dotted horizontal lines denote the time series sample mean plus/minus one sample standard deviation.

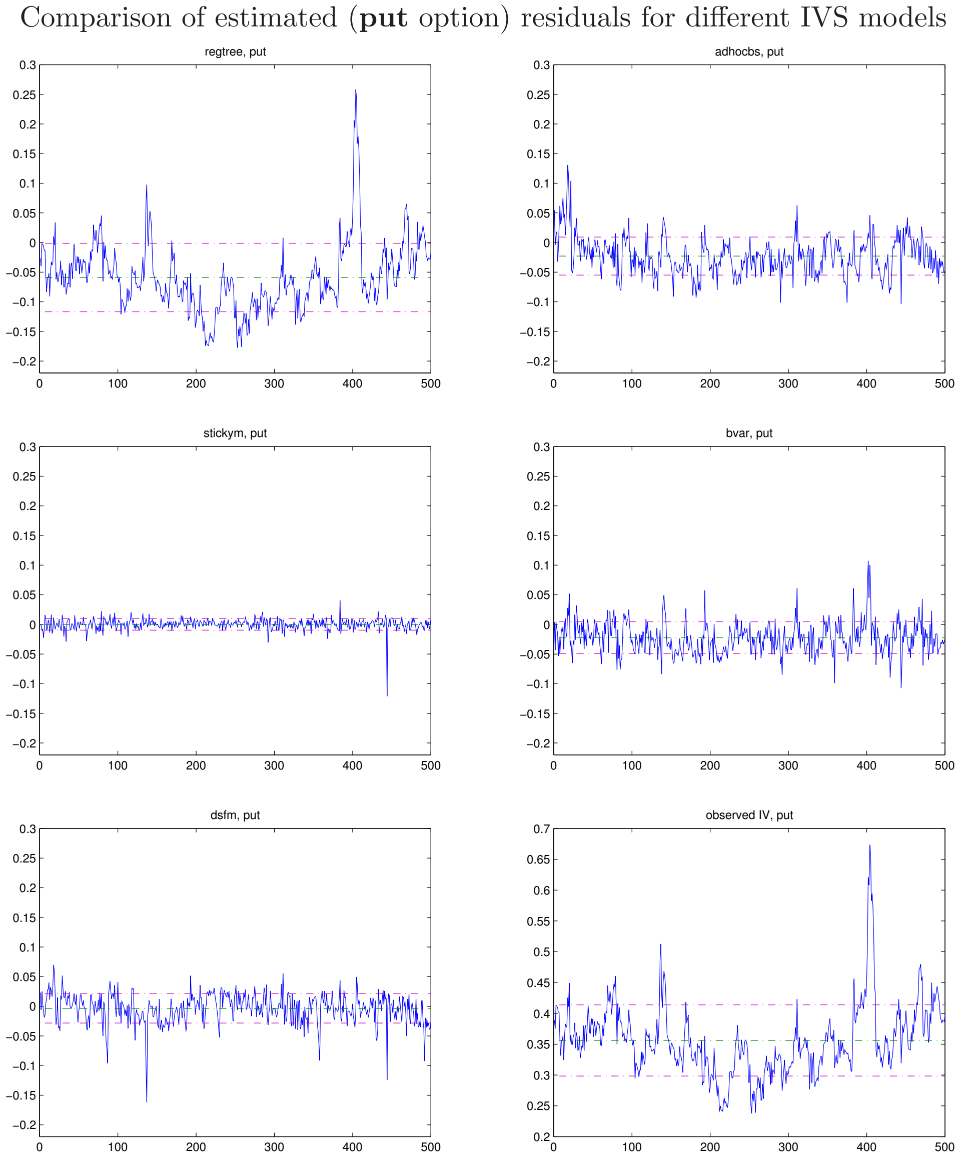


Figure 5.2: In the *lower right* panel, daily observed IVs for Microsoft **put** options closest to $m \approx 1$ and $\tau >\approx 10$ days are plotted against time (IS period of 500 sample days between 10 December 1996 – 2 March 1999). The *other panels* contain time series of estimated residuals $\sigma_t^{\mathrm{IV}}(m,\tau) - \hat{F}_0(m,\tau,\cdot)$ for these put options obtained with different IVS models, $F_0 \in \{$regtree, adhocbs, stickym, bvar and dsfm$\}$. $\hat{F}_0$ is fitted to the IS data. The dotted horizontal lines denote the time series sample mean plus/minus one sample standard deviation.

An obvious way to include the option type besides the two location parameters (m, τ) is indicated by the regtree model; a three location parameter model is obtained by combining a separate model for calls and puts each with an indicator function,

$$F_0(\mathbf{x}^{\text{pred}}) = F_0^c(m, \tau, \mathit{factors}) \mathbb{1}_{cp\ \mathit{flag}=\text{call}} + F_0^p(m, \tau, \mathit{factors}) \mathbb{1}_{cp\ \mathit{flag}=\text{put}} \ . \qquad (5.3)$$

Only adhocbs depends on a set of 'exogenous'[18] $\mathit{factors}_t = \{S_t\}$. As pointed out in Section 3.3, including other time-lagged as well as *forecasted* time-leading factors would extend the predictor space in a very appealing way to supervised learning strategies. We could try to minimize the magnitude of estimated residuals by adding a nonparametric expansion to F_0. If this procedure is iterated, a series of $\{F_k\}_{k \geq 0}$ is obtained and we can hope that $F_k \underset{k \to \infty}{\longrightarrow} \sigma_t^{\text{IV}}$ in the sense of an (unspecified) functional norm[19].

5.1.2 Keep extremal IV in the sample

ITM options are often excluded in the IVS literature. They contain a liquidity premium: ITM options have an intrinsic value, therefore they cost more and there is less leverage for speculation. The costs in portfolio hedging are higher with those options, hence they are traded less frequently. Cont and da Fonseca (2002) claim that OTM options contain the most information about the IVS.

Gonçalves and Guidolin (2006) apply five exclusionary criteria to filter their IVS data. They exclude thinly traded options, options that violate at least one basic no-arbitrage condition (see Section 2.4), options with fewer than six trading days to maturity or more than one year, options with moneyness smaller than 0.9 and larger than 1.1, and, finally, contracts with prices lower than three-eighths of a dollar. Cassese and Guidolin (2006) investigate the pricing efficiency in a bid-ask spread and transaction cost framework. They find a frictionless data set by dropping 51% of the original observations. Skiadopoulos et al. (2000) also screen the raw data. They eliminate data where the option price is less or equal to its intrinsic value, where prices are less than 10 cents and with $\tau < 10$ days. They construct smiles using OTM puts for low strikes and OTM calls for high strikes only, relying on the put-call parity. They also set a vega cutoff: options with vega less than eight are dropped from the sample. In this way, only 40% of the observations for calls and 70% for puts are retained in the sample.

---

[18] An estimate of the linear correlation in this concrete case yields $\text{corr}(S_t, \hat{\epsilon}_t) = -0.1585$, significantly different from zero. From what is already noted in Section 3.3.2, it follows that the right expression for S_t is here 'state variable' instead of 'exogenous factor'.

[19] Reducing such errors is the aim of functional gradient descent methods, see Note 4.3.

It is known that ITM calls and OTM puts are traded at higher prices compared to corresponding ATM options in general. This is especially true when the expiry date nears as observed prices and IV react violently, see Hentschel (2003). Options with expiry further in the future have more vega and less gamma than shorter expiring ones (see Section 7.3.2). Low IV precision close to expiration is inherent to options. These options have very little vega, so inverting the pricing formula gives a bigger change of volatility for a tiny price change. This is usually amplified by the wider bid-ask spreads for ITM options close to maturity. The usual trick is to focus on ATM and OTM options. Excluding the strangely behaving options from the sample helps any model to perform better, but this neglects the reality of having higher IV values at all. Regardless of what causes very high IV, removing them leads to a loss of information that may be important for prediction.

Example 5.2 The IVS dynamics of S&P 500 index options will be analyzed in Chapter 6. The highest IV value found in the whole sample belongs to a call option issued on 20 April 1998 with a strike of 700 points and expiry date 20 March 1999. On the day of option issue, the S&P 500 index closes at 1,123.70 points ($m = 0.6229$), on the maturity date at 1,299.30 ($m = 0.5388$) and the index level never drops below 957.30 ($m = 0.7312$) during the entire lifetime of the option. With three days left to maturity, the mid option price of \$598.625 translates into an IV of 4.9899, which is extremely high compared to the mean IV of 0.5209 during the option's lifetime. Figure 5.3 plots the time series of IVs for this option.

The proposed methodology needs to make sure that all kind of options can

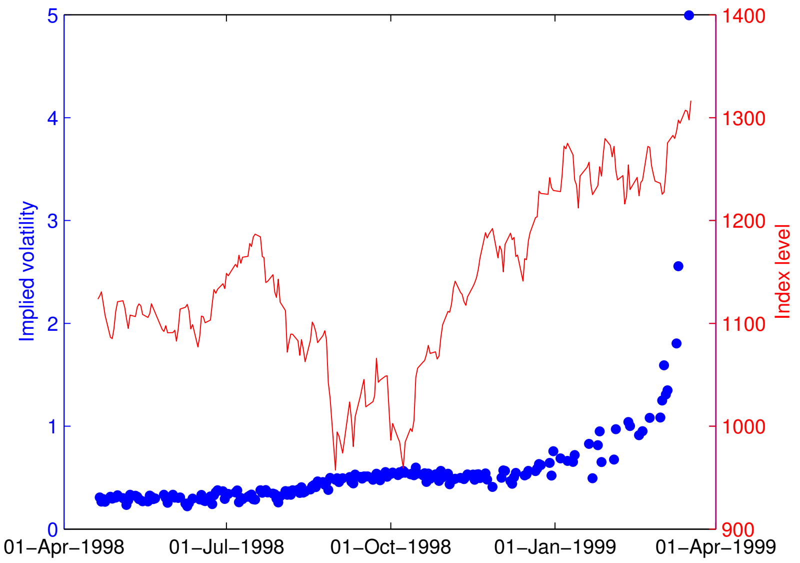


Figure 5.3: Call option on the S&P 500 index. Issued 20 April 1998, expired 20 March 1999, strike $K = 700$ points. IV is plotted in blue, the S&P 500 index level in red. IV time series average over the lifetime of the option is 0.5209, the maximum IV value of 4.9899 is reached on the 17 March 1999. The Greeks on that day are delta = 0.9937, gamma = 0.000052, vega = 1.2065, theta = -1118.905.

be handled. A weight function guarantees that extremal IV can be kept in the sample and controls their influence.

5.1.3 Local focus

One of the aims of the new methodology is to control prediction errors for certain regions in the (m, τ) domain, IS and especially also OS. The approach relies on a fixed grid in the (m, τ) domain, which should be laid over the region where forecasts of the IVS are to be calculated. Using the following indexing

$$[1] := (m_{(1)}, \tau_{(1)}), \;\; [2] := (m_{(2)}, \tau_{(1)}), \;\; \ldots, \;\; [N_m] := (m_{(N_m)}, \tau_{(1)}),$$

$$
\begin{gathered}
[N_m+1] := (m_{(1)}, \tau_{(2)}),\ \ [N_m+2] := (m_{(2)}, \tau_{(2)}),\ \ ...,\ \ [2 \cdot N_m] := (m_{(N_m)}, \tau_{(2)}), \\
\vdots \\
[(y-1) \cdot N_m + x] := (m_{(x)}, \tau_{(y)}),\ \ x \in \{1, ..., N_m\},\ \ y \in \{1, ..., N_\tau\} \\
(m_{(1)} < m_{(2)} < ... < m_{(N_m)},\ \ \tau_{(1)} < \tau_{(2)} < ... < \tau_{(N_\tau)}),
\end{gathered}
\tag{5.4}
$$

a grid with grid points GP $= \{[1], [2], ..., [N_m \cdot N_\tau]\}$ is obtained, on which the weight function from the previous section are also dependent. The grid is not used for smoothing the IVS, but for calibrating the series of $\{F_k\}_{k \geq 0}$. In this way, the estimation focus is set to the region of the grid.

5.1.4 OS prediction

OS predictions are of particular interest. IVS models may all provide a reasonable IS fit at any (m, τ) location, but ex-post analysis is of limited importance. When it comes to OS predictions, questionable assumptions like constant IV at fixed moneyness (sticky moneyness) or time invariance of model parameters (ad hoc BS model) need to be made to evaluate the model even just one day into the future. OS prediction is not possible for smoothing techniques, because kernel functions explicitly depend on observed data.

Example 5.3 (Cont. of Example 5.1) The 30 sample days from 3 March 1999 – 19 April 1999 are subsequent to the last IS day and represent the OS period during which we want to test the OS forecasting abilities of the five IVS models. Actually observed IVs of Microsoft (call and put) options closest to $m \approx 1$ and $\tau >\approx 10$ days are compared to the estimated OS IV forecasts obtained from the five IVS models. The differences are plotted in Figure 5.4.

Surprisingly, the upper two panels show that regtree can be as accurate with OS forecasts of the IV as the technically more complex dsfm model. Note that we are not tracking one specific option contract like in Example 5.2. Every day, another option fulfils the criteria ‘closest to ATM Microsoft stock option with nearest maturity of at least 10 days' such that $m_t \approx 1$ and $\tau_t >\approx 10$ days for all $t \in \text{IS}$. Thus, it makes sense to evaluate $\hat{F}_0$ at constant $\bar{m} = 1$ and $\bar{\tau} = 10/365$ in order to obtain an OS forecast of the IV. If the set of $factors_t$ contained another element than S_t, then we would have to consider how to predict $\widehat{factors}_{\tilde{t}}$. Since the adhocbs is the only one of the five models that directly depends on $factors_t = \{S_t\}$ and since its model coefficients have been estimated on the reference day $\bar{t} = 500$ (last IS day), it is therefore justified to use constant $\overline{factors} = \{S_{500}\}$ for all $\tilde{t} \in$ OS.

The exact future value of $\mathbf{x}_{\tilde{t}}^{\text{pred}}$ for $\tilde{t} \in \text{OS}$ is of course unknown at the end of the last IS day when IV needs to be forecasted. Nevertheless, the supervised learning method aims to improve OS prediction such that

$$(\sigma_{\tilde{t}}^{\text{IV}}(m,\tau) - \hat{F}_k(\mathbf{x}_{\tilde{t}}^{\text{pred}}))^2 < (\sigma_{\tilde{t}}^{\text{IV}}(m,\tau) - \hat{F}_0(\mathbf{x}_{\tilde{t}}^{\text{pred}}))^2 \tag{5.5}$$

on average for any $\tilde{t} \in \text{OS}$ and an integer $k > 0$. If the predicted $\widehat{\mathbf{x}_{\tilde{t}}^{\text{pred}}}$ is close enough to $\mathbf{x}_{\tilde{t}}^{\text{pred}}$ and if $\hat{F}_k$ is robust[20], then we would expect that $\hat{F}_k(\widehat{\mathbf{x}_{\tilde{t}}^{\text{pred}}}) \approx \hat{F}_k(\mathbf{x}_{\tilde{t}}^{\text{pred}})$. Eq. (5.5) typically does not hold for $k \to \infty$, at some stage $\hat{F}_k$ overfits the data.

5.2 Inspiration

Gouriéroux, Monfort, and Tenreiro (1995) and Aït-Sahalia and Lo (1998) introduce nonparametric kernel smoothing estimators in the option pricing literature. The least-square kernel (LSK) smoothing estimator of Gouriéroux et al. (1995) is defined by

$$\hat{\sigma}^{\text{IV}}(m,\tau) = \arg\min_{\tilde{\sigma}} \sum_{i=1}^{n} \left(c_{t_i} - c^{\text{BS}}(\cdot,\tilde{\sigma})\right)^2 \omega(m_{t_i}) K_1\left(\frac{m - m_{t_i}}{h_1}\right) K_2\left(\frac{\tau - \tau_{t_i}}{h_2}\right).$$

The observed call prices are normalized by the price of the underlying stock, $c_t = C_t/S_t$, and c^{BS} is the BS formula in terms of moneyness and time to maturity (see Appendix C.4). The estimate for a particular point on the IVS is given by the

---

[20] In mathematical statistics, a robust estimator is distorted only slightly by small departures from model assumptions.

Comparison of OS forecasting abilities of different IVS models,

$F_0 \in \{\text{regtree, adhocbs, stickym, bvar and dsfm}\}$

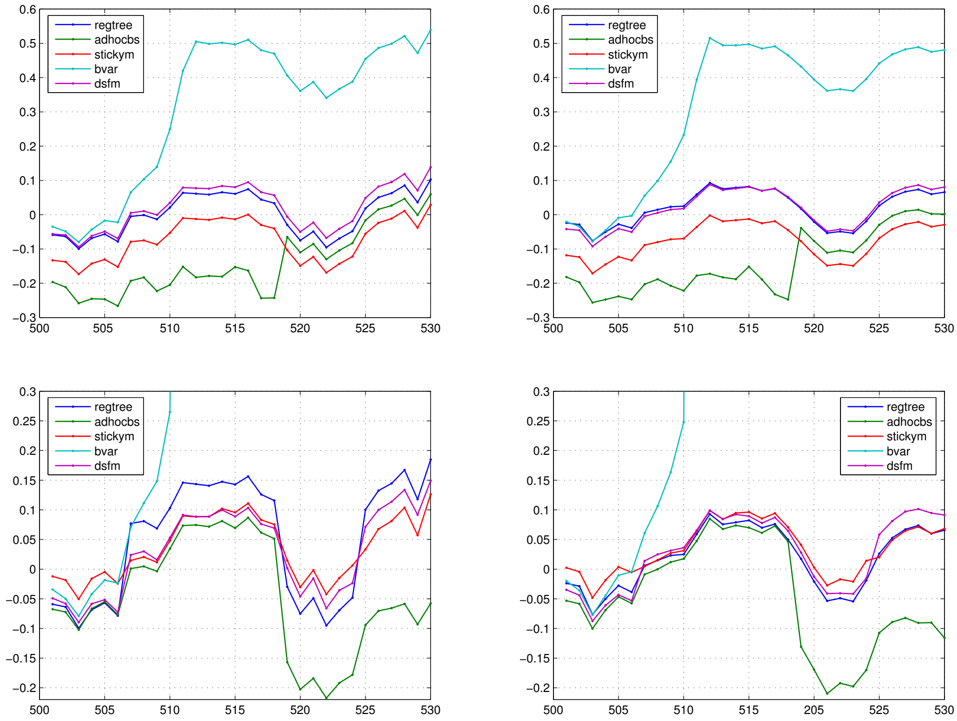


Figure 5.4: IVS models $F_0(m, \tau, cp\ flag, factors)$ have been fitted to Microsoft option data from 10 December 1996 – 2 March 1999 (IS period of 500 days). $\sigma^{\mathrm{IV}}_{\tilde{t}}(m, \tau)$ denotes actually observed IVs for $\tilde{t} \in \mathrm{OS} = \{501, \ldots, 530\}$, representing 30 sample days between 3 March 1999 – 19 April 1999. The forecasts in the *upper* panels are obtained by evaluating the fitted $\hat{F}_0$ at constant $\overline{\mathbf{x}^{\mathrm{pred}}} = (1, 10/365, cp\ flag, S_{500})$ during the whole OS period. The *upper left* panel plots $\sigma^{\mathrm{IV}}_{\tilde{t}}(m, \tau) - \hat{F}_0(\overline{\mathbf{x}^{\mathrm{pred}}})$ against $\tilde{t} \in \mathrm{OS}$ for calls ($cp\ flag = 1$), the *upper right* panel for puts ($cp\ flag = 0$). In the *lower* panels, $\hat{F}_0$ is evaluated exactly at $\mathbf{x}^{\mathrm{pred}}_{\tilde{t}} = (m_{\tilde{t}}, \tau_{\tilde{t}}, cp\ flag, S_{\tilde{t}}) \notin \mathcal{F}_{500}$ since $\tilde{t} \in \mathrm{OS}$. The *lower left* panel plots $\sigma^{\mathrm{IV}}_{\tilde{t}}(m, \tau) - \hat{F}_0(\mathbf{x}^{\mathrm{pred}}_{\tilde{t}})$ against $\tilde{t} \in \mathrm{OS}$ for calls ($cp\ flag = 1$), the *lower right* panel for puts ($cp\ flag = 0$).

minimum of the weighted sum of least squares. K_1 and K_2 are univariate kernel functions with bandwidths of h_1 and h_2, respectively. $\omega(m)$ denotes a uniformly continuous and bounded weight function, depending on m. Gouriéroux, Monfort, and Tenreiro (1994) prove under some weak conditions that the LSK estimator is consistent. They further show that it belongs to the class of kernel M estimators. Thus, the LSK estimator is also asymptotically normal distributed.

To be able to obtain accurate OS forecasts, it is desirable to modify the LSK estimator along several directions that will be discussed in the next section.

5.3 The model

In a general nonparametric model, IV is regressed on a vector of predictors through unspecified functions $f_{m,\tau}$ such that

$$\sigma^{\mathrm{IV}}_{m,\tau} = f_{m,\tau}(\mathbf{x}^{\mathrm{pred}}) + \varepsilon_{m,\tau} \tag{5.6}$$

with $E[\varepsilon_{m,\tau}] = 0$ and $E[\varepsilon^2_{m,\tau}] < \infty$ for each m, $\tau > 0$. The regression functions $f_{m,\tau}(\cdot)$ are implicitly defined in such a way that the expectation of a given loss function λ (which is known as *risk* in supervised learning),

$$\mathrm{E}\left[\lambda(\sigma^{\mathrm{IV}}_{m,\tau}, f_{m,\tau}(\mathbf{x}^{\mathrm{pred}}))\right] \tag{5.7}$$

is minimized for each m, $\tau > 0$.

The proposed methodology is based on semi-parametric models. A given (parametric or nonparametric) starting model $F_0(\mathbf{x}^{\mathrm{pred}})$ might fit the IVS quite well in certain (m, τ) areas, but not necessarily everywhere. To be able to apply classical boosting algorithms, for each m, $\tau > 0$ we restrict the regression function $f_{m,\tau}$ to a linear additive expansion of the form

$$f_{m,\tau}(\mathbf{x}^{\mathrm{pred}}) = F_0(\mathbf{x}^{\mathrm{pred}}) + \sum_{j=1}^{M} B_j(\mathbf{x}^{\mathrm{pred}}) \tag{5.8}$$

where each B_j denotes a base learner (see Section 4.2).

Regression trees are chosen as base learners for several reasons:

- Previous studies (see Section 4) already illustrated that accurate OS predictions can be obtained using regression trees as base learners, in particular when the number of end-nodes (or leaves) L is kept small, i.e. $L \leq 5$.
- A robust base learner is required when including time-lagged and *forecasted* time-leading factors in the predictor space. Let $\bar{t}$ denote the last IS day. Suppose that $\mathrm{E}_{\mathbb{P}}[X_{\tilde{t}}|\mathcal{F}_{\bar{t}}] \in$ *factors* and $\bar{t} < \tilde{t} \in$ OS. The regression tree B_j partitions the predictor space into $L+1$ different cells. Such a base learner is robust since $B_j(\mathrm{E}_{\mathbb{P}}[X_{\tilde{t}}|\mathcal{F}_{\bar{t}}], \cdot) = B_j(X_{\tilde{t}}, \cdot)$, provided that $\mathrm{E}_{\mathbb{P}}[X_{\tilde{t}}|\mathcal{F}_{\bar{t}}]$ lies in the same cell as the true future $X_{\tilde{t}}$.
- The lack of smoothness of the prediction surface obtained using regression trees is not a disadvantage: often m or τ are chosen as split variables when fitting the jth regression tree, and plotting the contribution of B_j to $f_{m,\tau}$ mainly shows that IV residuals for small τs are improved. This is in line with results from the stochastic volatility literature, where the shape of the IVS for small τs is better fitted when introducing jumps in the dynamics of the underlying (Remark C.11).
- Regression trees can handle the degenerated option data structure and deal with an extended predictor space. From a huge number of predictors, only the most relevant ones are automatically chosen as split variables.

5.4 Estimation

Since only a finite sample of observed IV is available, an estimate of $f_{m,\tau}(\cdot)$ is constructed with the help of the functional gradient descent technique (see Section 4.2) from a constrained minimization of the average observed loss function (*empirical risk*), the empirical analogon of Eq. (5.7). The constraints (5.8) require that $\hat{f}_{m,\tau}(\cdot)$ is an additive expansion of base learner functions. Boosting based on regression trees is a simple version of FGD, using regression trees as base learners and a quadratic loss function.

5.4.1 Empirical local criterion

Let $(m_{ti}, \tau_{ti}, \sigma_{ti}^{\mathrm{IV}})$, $i \in \{1, ..., L_t\}$ denote the observations of moneyness, time to maturity and IV at day t. The daily number of observations L_t is varying over time $t \in \{1, \dots, N\}$. The degenerated structure of option data demands aggregation over time. It is necessary to obtain a region where observed location parameters form quasi a continuum. The time to expiry needs to be controlled since long dated options can appear daily in the aggregated sample, whereas short dated ones soon expire and are replaced by others. An empirical local criterion is proposed to make sure that $\hat{f}_{m,\tau}(\cdot)$ lives up to all desired properties of Section 5.1.

The proposed approach relies on a fixed grid in the (m, τ) domain, as defined in Section 5.1.3, and on a quadratic loss function which depends directly on implied volatilities,

$$\lambda(\sigma_t^{\mathrm{IV}}, \hat{\sigma}_t^{\mathrm{IV}}) = \left(\sigma^{\mathrm{IV}} - \hat{\sigma}^{\mathrm{IV}}\right)^2 .$$

Definition 5.4 *The empirical local criterion to minimize over the set of grid points $GP = \{[1], [2], ..., [N_m \cdot N_\tau]\}$ is defined as*

$$\Lambda_{grid} := \sum_{t=1}^{N} \sum_{i=1}^{L_t} \sum_{[g] \in GP} \left(\sigma_{ti}^{IV} - \hat{\sigma}_{ti}^{IV}\right)^2 w_t(i, [g]), \tag{5.9}$$

with weights specified by

$$w_t(i, [g]) = \omega_1(m_{ti}) \cdot \omega_2(\tau_{ti}) \cdot K\left(\frac{m_{(x)} - m_{ti}}{h_1}, \frac{\tau_{(y)} - \tau_{ti}}{h_2}\right). \tag{5.10}$$

In the above equation, the different quantities are defined as

$$\begin{aligned}
[g] &= (m_{(x)}, \tau_{(y)}) \in GP, \;\; x \in \{1, ..., N_m\}, \;\; y \in \{1, ..., N_\tau\} \\
K(u, v) &= \frac{1}{2\pi} \cdot e^{-\frac{1}{2}(u^2+v^2)} \\
\omega_1(m_{ti}) &= \begin{cases} 1/\pi \cdot \arctan(\alpha_1(m_{ti} - 1)) + 1/2, & \textit{if option } i \textit{ is a call} \\ 1/\pi \cdot \arctan(\alpha_1(1 - m_{ti})) + 1/2, & \textit{if option } i \textit{ is a put} \end{cases} \\
\omega_2(\tau_{ti}) &= 1/\pi \cdot \arctan(\alpha_2(1 - \tau_{ti})) + 1/2.
\end{aligned}$$

Remark 5.5 The weight function (5.10) consists of three factors. The first is ω_1, taken from Fengler et al. (2007), with slight corrections such that OTM options have more influence than ITM options. The second one is ω_2. It depends on the time to maturity and was chosen to reduce the influence of options which expire far in the future, and to increase the importance of options that are soon due. The third one is a bivariate normal product kernel that sets the local focus to the grid points. From a numerical point of view, it is convenient to normalize the weight function in such a way that

$$\sum_{i=1}^{L_t} w_t(i, [g]) = 100$$

for every $[g] \in \mathrm{GP}$ and t, because the product of three small factors can become very small. Figure 5.5 shows a plot of ω_1 and ω_2.

Note 5.6 The weight function (5.10) depends on the chosen grid, kernel function, bandwidths h_1, h_2 and α_1, α_2. Of utmost importance is the choice of the grid, which is provided by the user. Finding optimal coefficients in the weight function is more difficult. To avoid complex adjustment procedures, we set $h_1 = h_2 = 0.5$, $\alpha_1 = 5$ and $\alpha_2 = 0.5$ if not mentioned otherwise. These fixed settings specify the local focus w.r.t. the grid. The kernel function is not used to smooth the IVS directly, it provides a measure of spatial distance between daily observed IVs and the grid points.

5.4.2 A feasible algorithm

A cross-validation scheme is needed to prevent the model from over-fitting. The first 70% of the days in the training data are considered to be a learning sample, and the remaining 30% are used as the validation sample. The model is fitted on the aggregated IV observations in the learning sample only. The more additive components in the expansion there are, the smaller the error in the learning sample becomes. It tends to zero as the number of iterations increases, but this is generally accompanied by worsening predictive power.

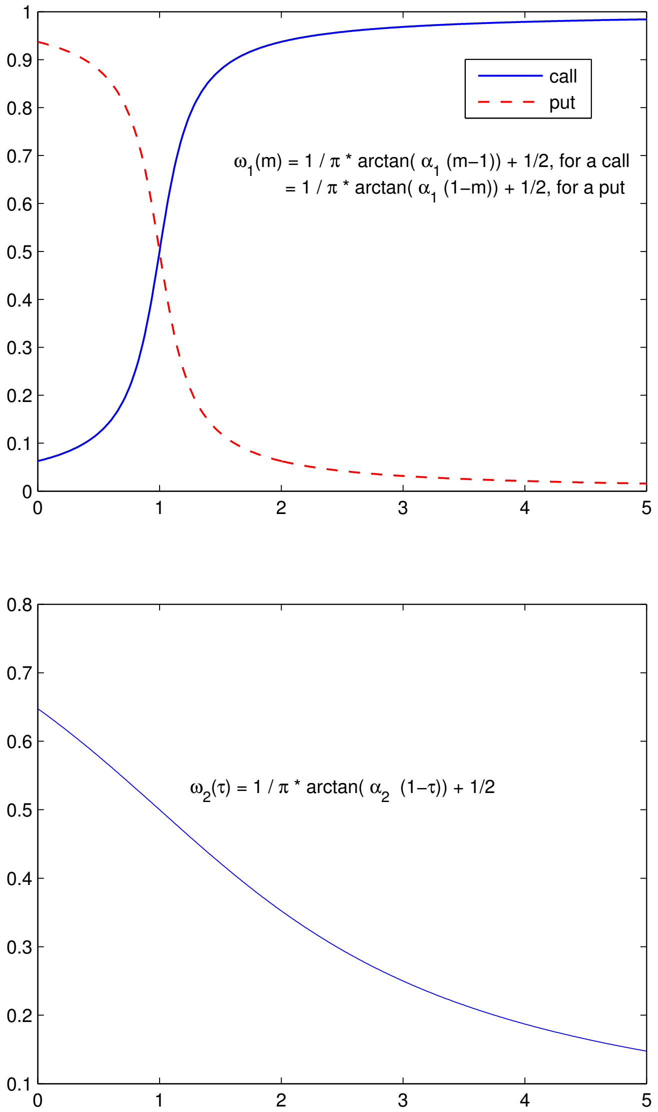


Figure 5.5: Plot of the function $\omega_1(m)$ with $\alpha_1 = 5$ on the *top* and $\omega_2(\tau)$ with $\alpha_2 = 0.5$ on the *bottom*. With these coefficients, the IV of an ITM (OTM) call with $m = 0.9$ ($m = 1.1$) has 30% less (more) weight ω_1 than the IV of an ATM call. The extremely high IV value of Example 5.2 is observed when $m = 0.5394$, hence its weight ω_1 is 74% less than for an ATM call. The IV of an option with $\tau = 10$ days has 29% more weight ω_2 than an option with $\tau = 1$ year. For an option with half a year to maturity, $\omega_2(0.5) \approx 1.15 \cdot \omega_2(1)$.

The empirical local criterion (5.9) is tailored to highlight importance of prediction errors in the grid region. The addition of expansions ceases when the empirical criterion takes its minimum on the validation sample. Slow convergence on the learning sample helps find the optimal number of iterations M. Indeed, using

$$\sum_{[g]\in GP} 2(\sigma^{\mathrm{IV}}_{ti} - \hat{\sigma}^{\mathrm{IV}}_{ti}) w_t(i,[g])$$

as a negative gradient to fit the base learner of each FGD step requires numerous iterations ($M > 500$). To make computation feasible, we resort to un-weighted residuals in growing the additive expansion. The learning rate can be further controlled by introducing a shrinkage factor $0 < \eta \leq 1$.

Taking all the above considerations into account, the following algorithm is proposed to estimate the IVS.

Algorithm: Tree-boosting for Implied Volatility Surfaces (treefgd)

Let $\mathbf{x}^{\mathrm{pred}}_{ti} = (m_{ti}, \tau_{ti}, \mathit{cp\ flag}_{ti}, \mathit{factors}_t)$ denote the observed predictor variables for option $i \in \{1,\dots,L_t\}$ on day $t \in \mathrm{IS} = \{1,\dots,N\}$. Split the in-sample period into a learning sample period $\mathrm{LS} = \{1,\dots,\lfloor 0.7\cdot N\rfloor\}$ and a validation sample period $\mathrm{VS} = \{\lfloor 0.7\cdot N\rfloor + 1,\dots,N\}$.

1. Fit the starting model $F_0(\mathbf{x}^{\mathrm{pred}})$ to the aggregated LS data. Evaluate the fitted model $\hat{F}_0$ at all $\mathbf{x}^{\mathrm{pred}}_{ti}$ to obtain estimated IVs for all options in the learning sample,
$$\hat{\sigma}^{\mathrm{IV},0}_{ti} = \hat{F}_0(\mathbf{x}^{\mathrm{pred}}_{ti}).$$

2. For $j = 1,\dots,M$:

   (a) Compute the residuals for all options in the learning sample,

$$\begin{aligned}\mathrm{residual}_{ti} &= \sigma^{\mathrm{IV}}_{ti} - \hat{\sigma}^{\mathrm{IV},j-1}_{ti}\\ &= \sigma^{\mathrm{IV}}(m_{ti}, \tau_{ti}, \mathit{cp\ flag}_{ti}) - \hat{F}_{j-1}(\mathbf{x}^{\mathrm{pred}}_{ti}).\end{aligned}$$

(b) Fit regression tree^c with L leaves to $\{\text{residual}_{ti} | \textit{cp flag}_{ti} = 1\}$ for all call options in the learning sample $\rightsquigarrow \widehat{\text{tree}^c}(m, \tau, \textit{factors})$.

(c) Fit regression tree^p with L leaves to $\{\text{residual}_{ti} | \textit{cp flag}_{ti} = 0\}$ for all put options in the learning sample $\rightsquigarrow \widehat{\text{tree}^p}(m, \tau, \textit{factors})$.

(d) Base learner j is then given by

$$\hat{B}_j(\mathbf{x}^{\text{pred}}) = \widehat{\text{tree}^c}(m, \tau, \textit{factors})\mathbb{1}_{\textit{cp flag}=1} + \widehat{\text{tree}^p}(m, \tau, \textit{factors})\mathbb{1}_{\textit{cp flag}=0}.$$

(e) Update

$$\hat{F}_j(\mathbf{x}^{\text{pred}}) = \hat{F}_{j-1}(\mathbf{x}^{\text{pred}}) + \eta \cdot \hat{B}_j(\mathbf{x}^{\text{pred}}),$$

with small shrinkage factor $0 < \eta \leq 1$.

3. Choose j such that Λ_{grid} is minimal on the validation sample. The optimal value is denoted $\hat{M}$ and satisfies

$$\hat{M} = \arg\min_j \; \Lambda_{\text{grid}} \left\{ \sigma_{ti}^{\text{IV}} \in \text{VS}, \;\; \hat{\sigma}_{ti}^{\text{IV}} = \hat{F}_j(\mathbf{x}^{\text{pred}}) \right\} \tag{5.11}$$

4. Repeat Steps 1 and 2 for the aggregated in-sample data (instead of only the learning sample) and for the number of $\hat{M}$ FGD iterations. An estimate for the unspecified function $f_{m,\tau}$ in the general nonparametric model of Eq. (5.6) is then given by

$$\hat{f}_{m,\tau}(\mathbf{x}^{\text{pred}}) = \hat{F}_{\hat{M}}(\mathbf{x}^{\text{pred}}) = \hat{F}_0(\mathbf{x}^{\text{pred}}) + \eta \sum_{j=1}^{\hat{M}} \hat{B}_j(\mathbf{x}^{\text{pred}}) \tag{5.12}$$

Remark 5.7 (Default values) The distributed computing capability of today's standard software makes it possible to optimize parameters like η, L and the number of time-lagged or leading factors by running the algorithm for different parameter values and choosing the best combination of them. Regression trees as base learners are in such a way flexible that various parameters can be set to (reasonable) default values. $\eta = 0.5$, $L = 5$, a number of 5 time-lagged or leading factors and the coefficients mentioned in Note 5.6 ($h_1 = h_2 = 0.5$, $\alpha_1 = 5$, $\alpha_2 = 0.5$) proved satisfactory in fitting the IVS.

Note 5.8 Step 2 of the algorithm does not depend on the chosen grid points because we decided to resort to un-weighted residuals to make the computation feasible. If we include only one exogenous factor X, we end up with a 14 dimensional $\mathbf{x}^{\text{pred}}$ when using the default values from the remark above. IVs of call options at time t are regressed on m_t, τ_t, X_{t-5}, X_{t-4}, $\ldots$, X_t, X_{t+1}, $\ldots$, X_{t+5}, but only four split variables and cut values per regression tree are automatically chosen to obtain a binary partition of the predictor space into five cells. The same is separately done for put options. The two regression trees are combined to B_j with the help of *cp flag*. Thus, we are counting three location parameters (m_t, τ_t and *cp flag*) and eleven versions of the exogenous factor X. If we include 25 exogenous factors, then the extended predictor space becomes 278 dimensional.

Remark 5.9 $M = 250$ iterations of Step 2 on a standard PC are calculated within 10 minutes for a 14 dimensional $\mathbf{x}^{\text{pred}}$ and a learning sample of 175 days, containing about 70,000 observed IVs. The same calculations for a 278 dimensional $\mathbf{x}^{\text{pred}}$ will take half a day. The amount of available RAM is the biggest bottleneck. Running the programs on a workstation with 64 bit operating system and 16 GB RAM is way faster (7 minutes, 25 minutes)[21]. Further gains in efficiency are achievable through parallelization (Panda, Herbach, Basu, and Bayardo, 2009, Section 8).

Remark 5.10 The cross-validation in Step 3 requires us to evaluate $\hat{F}_j(\mathbf{x}_{ti}^{\text{pred}})$ for all options in the validation sample to calculate $\Lambda_{\text{grid}}(\hat{F}_j)$. Computation time depends on the size of the chosen grid and should not be underestimated. Once the optimal stopping value $\hat{M}$ has been found, $F_{\hat{M}}$ needs to be estimated again, now using the whole sample and not only the first 70% of the data.

---

[21] The most recent implementation of the CART algorithm in MATLAB (`classregtree`) uses compiled MEX code and is significantly faster than the older `treefit` M-file.

Chapter 6

OS analysis of the S&P 500 IVS

After having introduced a new methodology for modelling the IVS in the previous chapter, it is time to empirically verify how much a given starting model can be improved by tree-boosting (treefgd). For a start, the analysis is limited to index options. The Standard & Poor's 500 (S&P 500) suggests itself as the best candidate for this kind of study[22]. It is a capitalization-weighted (free flow-weighted) index of 500 U.S. stocks and serves as performance benchmark for the U.S. equity markets[23]. Many derivatives are based on the S&P 500. The Chicago Board Options Exchange (CBOE) is offering European-style options on the S&P 500 since 1983. The VIX is a popular measure of the IV of S&P 500 index options. CBOE introduced this volatility index in 1993 to represent a measure of the market's expectation of volatility over the next 30 days. Futures and options on the VIX started trading in 2004 and 2006, respectively. Standard & Poor's Depositary Receipts (SPDR) are shares of an exchange-traded fund (ETF) that tracks the S&P 500 index. American-style SPDR options were launched in 2005.

---

[22] Noh et al. (1994) review the use of S&P 500 options in the literature.

[23] The S&P 500 is not strictly rule based. An investment committee selects the constituents w.r.t. liquidity, industry representation and place of business amongst other criteria.

6.1 Settings

In this section, the data, special days of interest and models are presented in detail. In addition to that, measures are defined to assess the OS performance of the fitted IVS.

6.1.1 Data

IVs of call and put options with different strikes and maturities on the S&P 500 index from 4 January 1996 to 29 August 2003 have been downloaded from OptionMetrics Ivy database. The whole sample consists of 777,887 observations on 1,928 days, approximately 400 observations of IV per day on average. The sample also features the degenerated structure of option data defined in Definition 3.5. Table 6.1 shows summarized statistics of the option data under investigation.

Time series plots of the S&P 500 index level and daily log returns are given in Figure 6.1. To visualize the time-varying volatility $\sigma_t \equiv \sqrt{h_t}$, an ARMA(1,1) model is combined with an asymmetric GARCH(1,1) model of Glosten, Jagannathan and Runkle (GJR, 1993) with scaled t-distributed innovations[24]

$$\begin{aligned} R_t &= \log\left(\frac{S_t}{S_{t-1}}\right) \\ R_t &= c + \phi R_{t-1} + \theta z_{t-1} + z_t \\ z_t &= \sqrt{h_t}\varepsilon_t, \quad \varepsilon_t \overset{d}{\sim} \text{iid scaled } \mathrm{t}_{\mathrm{df}} \\ h_t &= \kappa + \alpha z_{t-1}^2 + \beta h_{t-1} + \gamma \mathbb{I}_{z_{t-1}<0} \cdot z_{t-1}^2. \end{aligned} \tag{6.1}$$

The model is then fitted by the maximum likelihood method to the S&P 500 log returns over the whole period. Table 6.2 reports the estimated parameters of the model and compares them to the parameters of a standard GARCH(1,1) with normal distributed innovations. Figure 6.2 plots the time series of annualized conditional volatilities obtained with the GJR GARCH model.

---

[24] A t-distributed random variable X can be scaled to unit variance by setting $\tilde{X} = \frac{\mathrm{df}-2}{\mathrm{df}} X$, where df denotes the degrees of freedom, df > 2.

Descriptive statistics of σ^{IV} (in %)

		Maturity in days					
		Less than 90		90 to 240		More than 240	
Moneyness		Call	Put	Call	Put	Call	Put
$m \leq 0.8$	mean	76.75	71.96	35.33	24.49	29.92	19.73
	std	61.99	62.21	9.34	7.77	6.57	3.13
	# obs	15,176	4,667	16,302	2,510	25,227	4,694
$0.8 < m \leq 0.94$	mean	35.50	27.53	25.56	20.12	24.79	21.02
	std	21.75	18.87	5.08	3.53	4.59	3.33
	# obs	41,347	20,673	22,489	12,782	30,309	22,011
$0.94 < m \leq 1.04$	mean	22.10	22.15	21.06	21.17	22.14	22.36
	std	7.09	6.34	4.33	4.36	4.33	3.97
	# obs	61,768	61,707	19,771	19,910	25,368	25,443
$1.04 < m \leq 1.2$	mean	21.83	32.06	18.64	25.64	19.86	25.13
	std	10.27	10.97	3.50	4.96	3.65	4.33
	# obs	44,562	49,061	21,285	23,037	31,920	30,544
$1.2 < m$	mean	43.07	49.25	20.65	34.95	18.38	30.25
	std	34.72	19.41	5.67	7.33	2.90	5.65
	# obs	19,642	24,037	19,521	22,724	29,906	29,494

Table 6.1: Descriptive statistics of implied volatilities of S&P 500 index options from 4 January 1996 to 29 August 2003, 777,887 observations on 1,928 days. Sample average (mean) and standard deviation (std) of IVs are reported in percentage. Moneyness m is defined as strike price divided by the closing price of the underlying asset. Maturity is measured in calendar days. The intervals in the moneyness column represent the following moneyness categories for call options from top to bottom: deep in-the-money (DITM), in-the-money (ITM), at-the-money (ATM), out-of-the-money (OTM) and deep out-of-the-money (DOTM). For put options, the reverse order has to be considered.

Estimated model parameters for S&P 500 log returns

Glosten, Jagannathan and Runkle (GJR) GARCH model

Parameter	Estimated value	Standard error	T statistic
c	0.00061495	0.00045581	1.3491
ϕ	-0.90656	0.11378	-7.9674
θ	0.89092	0.12363	7.2063
κ	$3.211 \cdot 10^{-6}$	$8.7856 \cdot 10^{-7}$	3.6548
α	0	0.015349	0.0000
β	0.90335	0.016851	53.6072
γ	0.15195	0.022684	6.6983
df	11.328	2.1929	5.1658

Standard GARCH(1,1) model

Parameter	Estimated value	Standard error	T statistic
c	0.00061057	0.00026214	2.3291
κ	$3.8946 \cdot 10^{-6}$	$9.0714 \cdot 10^{-7}$	4.2933
α	0.094194	0.010285	9.1580
β	0.88399	0.013365	66.1418

Table 6.2: The models are fitted on 1,927 S&P 500 log returns from 4 January 1996 until 29 August 2003. The log likelihood for the GJR model from Eq. (6.1) is 5,883.3, for the standard GARCH(1,1) model with iid $\mathcal{N}(0,1)$ distributed ε_t is 5,820.5. The T statistic measures the number of standard deviations that the parameter estimate is away from zero. T statistic ≥ 2 in magnitude supports the case of a true non-zero parameter.

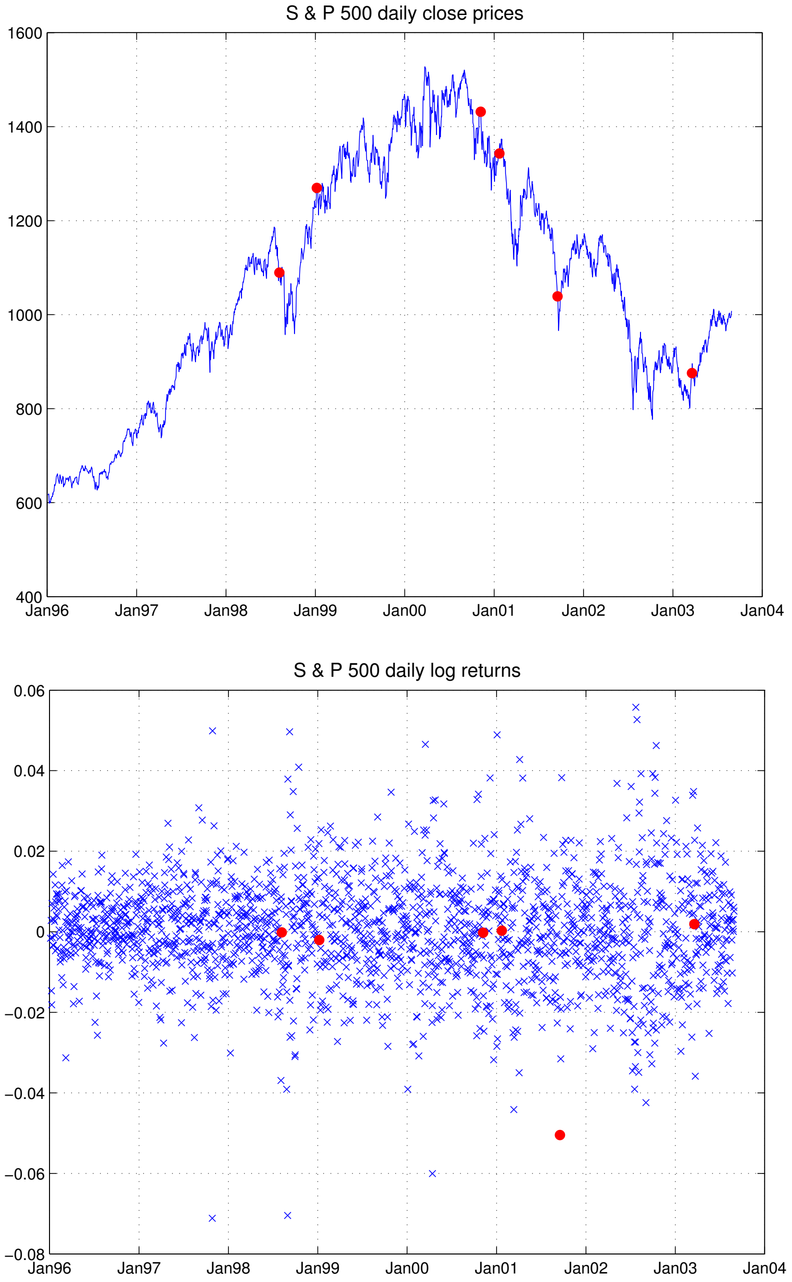


Figure 6.1: End-of-day level and daily log returns of the S&P 500 index. The red circles mark the following dates: 7 August 1998, 7 January 1999, 7 November 2000, 22 January 2001, 17 September 2001, 20 March 2003.

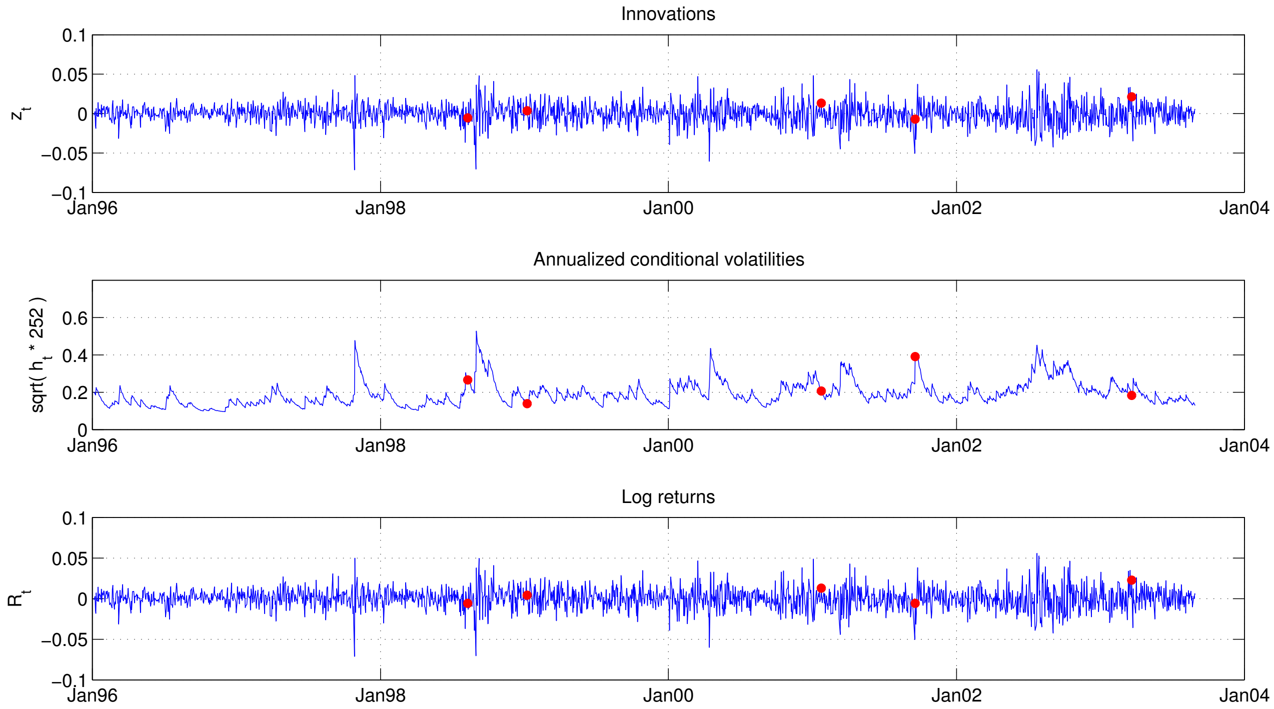


Figure 6.2: The **GJR GARCH model** from Eq. (6.1) is fitted to S&P 500 log returns from 4 January 1996 until 29 August 2003. The time series of estimated innovations z_t, annualized conditional volatilities $\sqrt{h_t \cdot 252}$ and the log returns R_t are plotted. The red circles mark the following dates: 7 August 1998, 7 January 1999, 7 November 2000, 22 January 2001, 17 September 2001, 20 March 2003.

6.1.2 Special days of interest

The IVS models will only be fitted to five different subsamples of length 250 days. They are chosen to occur before special days of interest, where – from today's perspective – a more or less heavy structural break is expected to happen. On 7 August 1998 two bomb attacks on US embassies in Africa occurred. The impeachment trial of President Clinton was opened in the senate on 7 January 1999. The first date for which we have IVS observations in our sample after President Bush's oath of office and the disruption caused by the unclear outcome in the presidential election 2000 is 22 January 2001. 17 September 2001 is six days after the 11 September terrorist attacks, and 20 March 2003 marks the official beginning of the military campaign against Iraq. In Figures 6.1 and 6.2, these dates are marked by red circles. Table 6.3 summarizes the chosen subsamples. They end four to 25 days before the special days of interest. The goal is to attain 60 days OS predictions of the IVS such that we can compare the observed IVs with their predicted values, before and after a supposed structural break. The accuracy is measured by evaluating a fitted model at the exactly same (m, τ) locations as observed.

6.1.3 Model specification

The five IVS models presented in Section 3.3.4 – the three-location regression tree model (regtree), the ad hoc BS model (adhocbs), the data gridding sticky moneyness model (stickym), the high-dimensional Bayesian vector autoregression (bvar) and the dynamic semiparametric factor model (dsfm) – are fitted to each subsample. The choice of tuning parameters (for example number of split variables or autoregressive lags) is driven by a trade-off between computational feasibility and optimization. Given some restrictions on the maximum number of parameters, the optimal tuning parameters are derived by minimizing the same cross-validation scheme adopted in Step 3 of the tree-boosting algorithm in Section 5.4.2. Such a 'best fitted' model is denoted $\hat{G}(\cdot)$.

Subsamples description

Subsample	Training data In-sample period	Forecasting Out-of-sample period	Special day of interest	
1	21 Jul 1997 - 16 Jul 1998	17 Jul 1998 - 09 Oct 1998	07 Aug 1998	16th OS day
2	06 Jan 1998 - 31 Dec 1998	04 Jan 1999 - 30 Mar 1999	07 Jan 1999	4th OS day
3	20 Dec 1999 - 13 Dec 2000	14 Dec 2000 - 13 Mar 2001	22 Jan 2001	25th OS day
4	15 Aug 2000 - 10 Aug 2001	13 Aug 2001 - 09 Nov 2001	17 Sep 2001	21st OS day
5	18 Mar 2002 - 13 Mar 2003	14 Mar 2003 - 09 Jun 2003	20 Mar 2003	5th OS day

Table 6.3: The table specifies the special days of interest in the OS periods for the five different subsamples under investigation. In-sample (IS) periods are of length 250 days and out-of-sample (OS) periods of length 60 days.

Note 6.1 The IVS models of Section 3.3.4 with preset default tuning parameters[25] are used as starting model $F_0(\cdot)$ in the tree-boosting algorithm, where the parameters for boosting are also preset to default values (see Remark 5.7). The performance of the fitted $\hat{F}_{\hat{M}}(\cdot)$ is compared to $\hat{G}(\cdot)$ (see Remark 3.7). To distinguish the different models, the notation 'adhocbs-treefgd' for example stands for the IVS model $F_M(\cdot)$ that consists of the starting model $F_0(\cdot) \equiv$ adhocbs enhanced with M additive expansions coming from the tree-boosting (treefgd) algorithm. Other word combinations have the same implied meaning.

The empirical local criterion Λ_{grid} of Eq. (5.9) is designed w.r.t. to a grid in the (m, τ) domain. Three sets of grid points are needed for the analysis in this chapter,

$$\begin{aligned} \text{GP}_1 = \{ & \text{linearly spaced } 15 \times 15 \text{ grid with values} \\ & \text{from } m = 0.2 \text{ to } 2 \text{ and from } \tau = \frac{1}{365} \text{ to } 3\} \end{aligned} \tag{6.2}$$

$$\text{GP}_2 = \{m = 1, \tau = 0\} \tag{6.3}$$

$$\text{GP}_3 = \{m \in \{0, 0.1, 0.2, \ldots, 3\},\ \ \tau = \Delta_{\text{ns30d}}\} \tag{6.4}$$

with Δ_{ns30d} = the time (in years) to the next option exchange settlement date (the third Friday of each month) at least 30 days in the future.

We also limit ourselves to three different sets of predictor variables (pv set) in this study. They are defined such that

$$\text{pv set } 1 \subset \text{pv set } 2 \subset \text{pv set } 3.$$

Five time-lagged and five forecasted time-leading versions of each exogenous factor are included in the extended predictor space.

---

[25] It is a nested setting: The discrete set of tuning parameters also contains the default tuning parameters of $F_0(\cdot)$. In the model selection step, the optimal tuning parameters for $G(\cdot)$ are determined. Hence it is possible that $F_0(\cdot)$ and $G(\cdot)$ coincide.

pv set 1 consists of the closing prices of the underlying.

pv set 2 additionally includes the {3,6}-month and {1, 3, 5, 10}-year Treasury Constant Maturity Rates as representatives of the interest rates term structure. The time series are available on the St. Louis Fed Homepage in the FRED database and their series IDs are labeled {DGS3MO, DGS6MO, DGS1, DGS3, DGS5, DGS10}.

pv set 3 comprises another 18 factors that are actually prices of options with different fixed characteristics obtained with the Heston-Nandi-GARCH (HNG) model. Table 6.4 specifies the characteristics for the HNG factors.

Specifications of the Heston-Nandi-GARCH (HNG) factors

	HNG01	HNG02	HNG03	HNG04	HNG05	HNG06
m	0.8	1	1.2	0.8	1	1.2
τ	30 days	30 days	30 days	60 days	60 days	60 days
cp flag	call	call	call	call	call	call
	HNG07	HNG08	HNG09	HNG10	HNG11	HNG12
m	0.8	1	1.2	0.8	1	1.2
τ	90 days	90 days	90 days	30 days	30 days	30 days
cp flag	call	call	call	put	put	put
	HNG13	HNG14	HNG15	HNG16	HNG17	HNG18
m	0.8	1	1.2	0.8	1	1.2
τ	60 days	60 days	60 days	90 days	90 days	90 days
cp flag	put	put	put	put	put	put

Table 6.4: Each HNG factor is the price of an option with fixed (time invariant) characteristic. A time series of option prices for each constant option specification in the table can be calculated via closed form solution (6.6).

The model of Heston and Nandi (2000) is based on an asymmetric GARCH process for the spot asset price under the assumption that an option with one period to expiration obeys the BS formula. The risk-neutral dynamics in this model are

$$\begin{aligned} R_t &= r - \frac{1}{2}h_t + \sqrt{h_t}\varepsilon_t, \qquad \varepsilon_t \overset{d}{\sim} \text{iid } \mathcal{N}(0,1) \\ h_t &= \kappa + \alpha\left(\varepsilon_{t-1} - \gamma\sqrt{h_{t-1}}\right)^2 + \beta h_{t-1} \end{aligned} \tag{6.5}$$

and the price of a European call on a non-dividend stock is given by

$$C_t(S_t, h_t, K, T, r) = S_t P_1 - K e^{-(T-t)} P_2 \tag{6.6}$$

$$P_1 = \frac{1}{2} + \frac{1}{\pi S_t e^{r(T-t)}} \int_0^\infty \mathrm{Re}\left[\frac{K^{-i\phi} f(i\phi + 1)}{i\phi}\right] d\phi \tag{6.7}$$

$$P_2 = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \mathrm{Re}\left[\frac{K^{-i\phi} f(i\phi)}{i\phi}\right] d\phi \ . \tag{6.8}$$

The option price depends on h_t through the conditional characteristic function of $\log S_T$, $f(i\phi) = \mathrm{E}_{\mathbb{Q}}\left[e^{i\phi \log S_T} | \mathcal{F}_t\right]$. It can be calculated recursively from terminal conditions (see Heston and Nandi, 2000, Proposition 2). The HNG model belongs to the class of affine models. In the continuous-time limit, it contains the stochastic model of Heston (1993), which is still very popular in practice (see Note C.12).

All parameters needed for $f(i\phi)$ in the valuation formula (6.6) are estimated directly from the history of underlying asset prices. The maximum likelihood estimator (MLE) is obtained on a rolling window of 500 days. The model could also be calibrated to observed option prices by nonlinear least squares estimation techniques. This is not necessary, the rolling window MLE approach works well as can be seen in Figure 6.3. The time series of annualized conditional volatilities looks similar to the one obtained with the GJR GARCH model in Figure 6.2.

Closest ATM, 30 days to maturity call and put options are chosen every day over the whole sample period. Figure 6.4 compares observed and calculated IVs obtained from the HNG model for these options. The two time series look very similar, differences between them are small[26].

---

[26] Including option prices obtained by other competing methods as exogenous factors aligns $\hat{F}_{\hat{M}}(\cdot)$ if necessary (supervised learning).

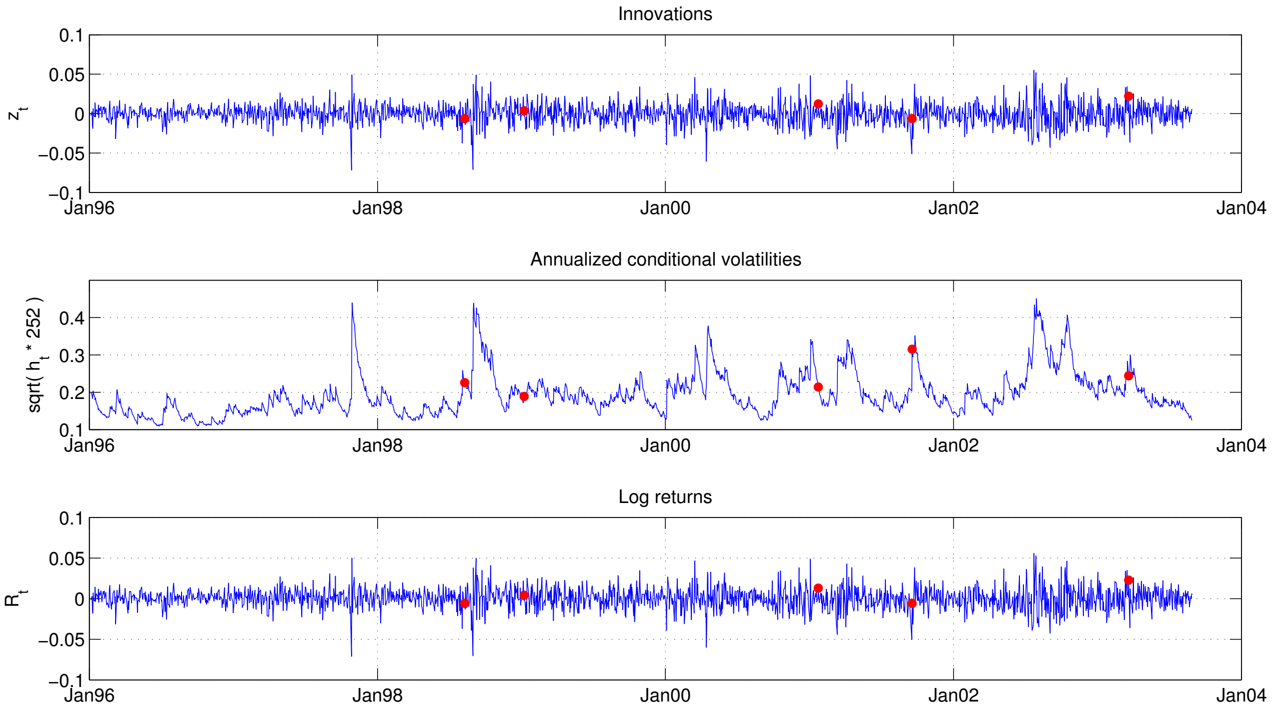


Figure 6.3: The **HNG model** from Eq. (6.5) is fitted to S&P 500 log returns from 4 January 1996 until 29 August 2003. The time series of estimated innovations z_t, annualized conditional volatilities $\sqrt{h_t \cdot 252}$ and the log returns R_t are plotted. The red circles mark the following dates: 7 August 1998, 7 January 1999, 7 November 2000, 22 January 2001, 17 September 2001, 20 March 2003.

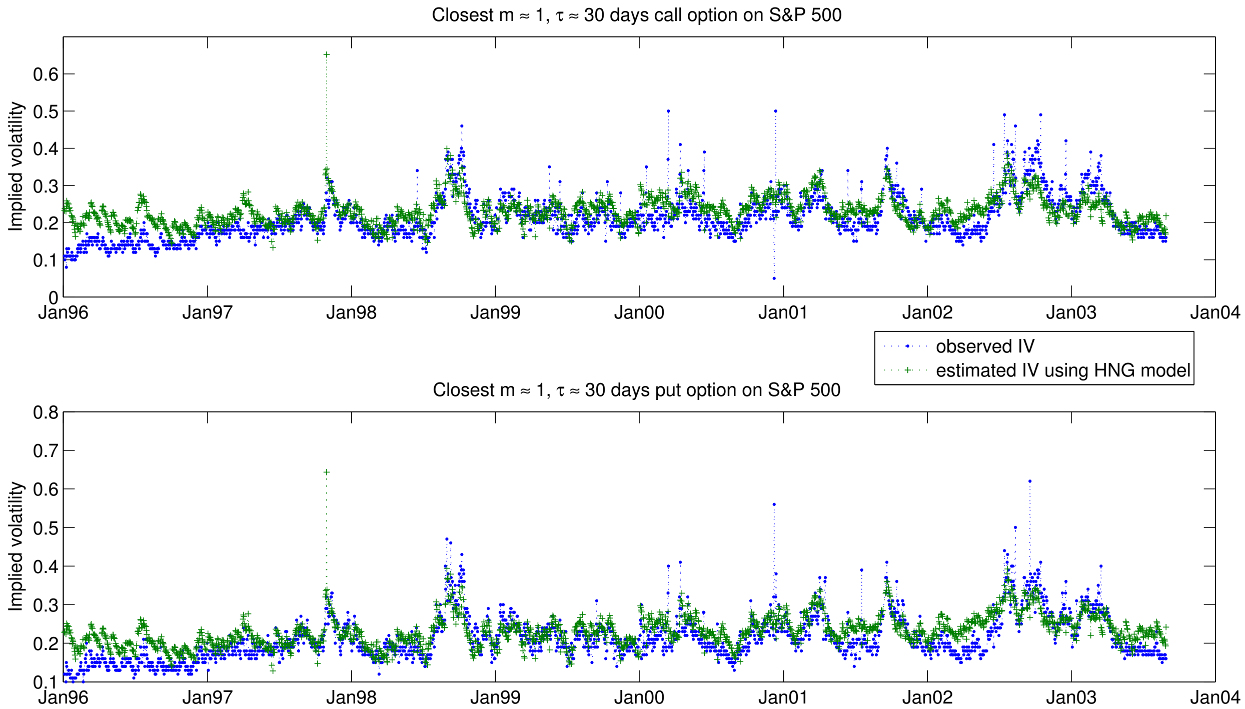


Figure 6.4: From 4 January 1996 until 29 August 2003, daily call and put options with m closest to ATM and τ nearest to $30/365$ are chosen. Daily HNG option prices are computed and mapped to volatility values by solving Eq. (3.1). Time series for calls are plotted in the *upper* graphic, for puts in the *lower* graphic.

6.1.4 Filtered historical simulation of exogenous factors

Let $\bar{t} = 250$ denote the last IS day of a subsample. To obtain a 60 days OS prediction of the IVS, forecasts of exogenous factors up to day $\bar{t}+65$ are needed since five forecasted time-leading factors are included in the predictor space. For this reason, we model the log returns in the case of an asset price, the first differences in the case of interest rates, as a univariate ARMA(1,1)-GJR GARCH(1,1) process (6.1) and apply standard filtered historical simulation (FHS) to obtain $\widehat{factors}_{\bar{t}+k}$ for $k \in \{1, 2, \ldots, 65\}$.

Filtered historical simulation[27] is a particular technique based on the bootstrap of the estimated residuals to reduce the forecast errors. Parameters of the ARMA(1,1)-GJR GARCH(1,1) process are estimated on a rolling time-window of past 500 observations. Sixty-five of the fitted $\{\hat{\varepsilon}_t | 1 \leq t \leq 500\}$ are randomly chosen and labeled $\{\hat{\varepsilon}^{(i)}_{\bar{t}+1}, \hat{\varepsilon}^{(i)}_{\bar{t}+2}, \ldots \hat{\varepsilon}^{(i)}_{\bar{t}+65}\}$. This sample is treated as if it were a future realization of ε over the next 65 days. With the help of Eq. (6.1), it is possible to construct iteratively $\hat{h}^{(i)}_{t+k}$, $\hat{z}^{(i)}_{\bar{t}+k}$,$\hat{R}^{(i)}_{\bar{t}+k}$ and finally $\hat{S}^{(i)}_{\bar{t}+k}$. The procedure is repeated 10,000 times. $\left\{ \hat{S}^{(i)}_{t+k} \middle| 1 \leq i \leq 10,000 \right\}$ is the bootstrapped distribution of S_{t+k} and its sample mean is taken as OS forecast of the exogenous factor,

$$\widehat{factors}_{\bar{t}+k} = \frac{1}{10,000} \sum_{i=1}^{10,000} \hat{S}^{(i)}_{\bar{t}+k}. \tag{6.9}$$

6.1.5 OS performance measures

The OS performances of $\hat{F}_{\hat{M}}(\cdot)$ and $\hat{G}(\cdot)$ are compared to observed IVs by evaluating the IVS models at exactly the same $(m, \tau, cp\ flag)$ locations as the ones of recorded entries in our database. Let $\hat{\sigma}^{\mathrm{IV}}_{ti}$ denote this predicted value for $t \in \mathrm{OS}$, $i \in \{1, \ldots, L_t\}$. The goodness-of-fit of the different competitors is measured w.r.t.

---

[27] Barone-Adesi, Bourgoin, and Giannopoulos (1998) and Barone-Adesi, Giannopoulos, and Vosper (1999) give a detailed description of filtered historical simulation.

the daily and overall averaged mean square forecast error:

$$\text{daily averaged SSR}_t = \frac{1}{L_t}\sum_{i=1}^{L_t}(\sigma_{ti}^{\text{IV}} - \hat{\sigma}_{ti}^{\text{IV}})^2 \tag{6.10}$$

$$\text{overall averaged SSR} = \frac{1}{N}\sum_{t=1}^{N}\text{daily SSR}_t. \tag{6.11}$$

As additional performance measures, the daily and the overall averaged empirical criterion, daily EC_t and overall EC, are considered:

$$\text{daily averaged EC}_t = \frac{1}{L_t}\sum_{i=1}^{L_t}\sum_{[g]\in\text{GP}}(\sigma_{ti}^{\text{IV}} - \hat{\sigma}_{ti}^{\text{IV}})^2\, w_t(i,[g]) \tag{6.12}$$

$$\text{overall averaged EC} = \frac{1}{N}\sum_{t=1}^{N}\text{daily EC}_t. \tag{6.13}$$

6.2 Empirical results

This section summarizes the results when applying the new methodology for building semi-parametric models on the five S&P 500 option data subsamples.

6.2.1 OS forecasts of predictor variables

During subsample 1, the S&P 500 index rises from 912.94 points at the beginning of the period to 1,184 points at the last day $\bar{t}$ of IS $= \{1,2,\ldots,250\}$. The FHS prediction (6.9) for the S&P 500 on the 60th OS day is 1,273 points. Compared to the effectively observed index level of 984.4, this prediction overestimates the true future S&P 500 level by 29.32%. The forecasts for the same quantity differ by $\{1.46\%, 19.78\%, 11.34\%, -11.73\%\}$ for subsamples 2 to 5, respectively. Figure 6.5 plots all OS realizations of the FHS estimators for a selected group of exogenous factors for subsample 1. Only the HNG option prices seem to be accurately forecasted.

Our database contains in total 25,020 IVs during the 60 OS days following directly after subsample 1. Table 6.5 reports optimal stopping values $\hat{M}$ (5.11)

Robustness of regression trees as base learners on subsample 1

	Starting model $F_0()$				
same cell ratio	regtree	adhocbs	stickym	bvar	dsfm
pv set 1	0.7411	0.7334	0.25	0.9661	0.7261
pv set 2	0.6419	0.5601	0.2394	0.9661	0.7164
pv set 3	0.5764	0.5903	0.1773	0.8512	0.7253

$\hat{M}$ w.r.t. GP_1	regtree	adhocbs	stickym	bvar	dsfm
pv set 1	108	8	1	3	57
pv set 2	234	69	1	3	2
pv set 3	114	19	1	3	1

Table 6.5: Each base learner partitions the predictor space into mutually exclusive hypercube cells. The 'same cell ratio' reports the percentage of cases for which the OS forecasts and the actually observed multivariate predictor variables fall into the same cells. Example: for the 60 day OS period after subsample 1 and all base learners in the regtree-treefgd model, $\widehat{\mathbf{x}^{\text{pred}}}$ and $\mathbf{x}^{\text{pred}}$ fall into the same partition cell in 2,002,571 of the 108 · 25,020 = 2,702,160 cases (74.11%).

w.r.t. GP_1 for tree-boosted models with different starting models and the ratio of observations for which $\widehat{\mathbf{x}^{\text{pred}}}_{\bar{t}+k}$ and $\mathbf{x}^{\text{pred}}_{\bar{t}+k}$ fall into the same hypercube partition cell of the predictor space generated by base learner B_j for $j \in \{1, 2, \ldots, \hat{M}\}$ and $k \in \{1, 2, \ldots, 60\}$. The obtained high percentages indicate the robustness of regression trees as base learners. This property is relevant in practical cases to preserve the benefits of supervised learning. Due to the time series nature of the data, future predictor variables can only be estimated. Univariate FHS forecasts (6.9) can be very inaccurate, maybe a multivariate FHS with dynamic conditional correlation could improve the quality of $\widehat{\mathbf{x}^{\text{pred}}}_{\bar{t}+k}$.

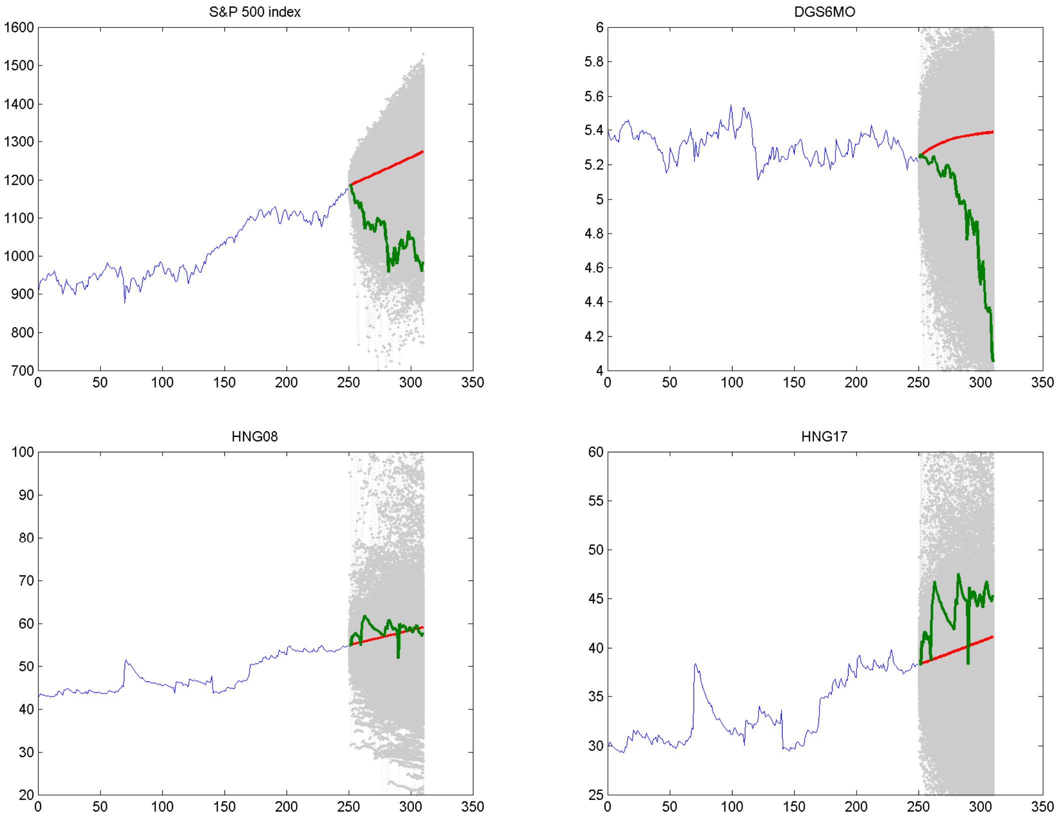


Figure 6.5: Time series plot of exogenous factors (S&P 500 index level, DGS6MO, HNG08, HNG17; see Section 6.1 for detailed specifications) during subsample 1 (blue line in the time interval [1,250]). For each exogenous factor, 10,000 filtered historical simulation scenarios for a 60 days OS period are drawn (dark gray scatter-plot in the time interval [251,310]). The red line marks the expected OS path, the green line the actually observed evolution of each exogenous factor.

6.2.2 Cross-validation

The optimal stopping values $\hat{M}$ that controls the number of additive expansions to the starting model is obtained by cross-validation. The top left graphic in Figure 6.6 plots $\Lambda_{\text{grid}}(\hat{F}_j)$ w.r.t. GP_1 on the learning sample LS=$\{1, 2 \ldots, 175\}$ against $j \in \{1, 2, \ldots, 250\}$ for the dsfm-treefgd model on subsample 1. As expected, this graph is decreasing. The top right graphic is more interesting; the local empirical criterion is evaluated on the validation sample VS=$\{176, 177, \ldots, 250\}$ instead. This corresponds to Step 3 of the tree-boosting algorithm, where the cross-validated $\hat{M}$ satisfies Eq. (5.11). The plot of the sum of squared residuals (SSR) against $j \in \{1, 2, \ldots, 250\}$ in the bottom right graphic behaves similarly, but the minimum is reached later than for the local empirical criterion. Cross-validation w.r.t. to SSR would lead to over-fitting.

Based on subsample 1, the lower panel of Table 6.5 shows how $\hat{M}$ varies with the choice of predictor variable set and starting model. The latter has more influence on $\hat{M}$. The more complex a starting model, the less chance for base learners to improve upon the model. For subsample 3, Tables 6.6 and 6.7 report the cross-validated $\hat{M}$, for all combinations of $\{(\text{pv set}_i, \text{GP}_j) | i, j \in \{1, 2, 3\}\}$ and starting models, together with overall averaged SSR (6.11) and empirical criterion (6.13) on the validation sample.

	Optimal stopping value $\hat{M}$			Overall averaged SSR			Overall averaged EC		
regtree	pv set 1	pv set 2	pv set 3	pv set 1	pv set 2	pv set 3	pv set 1	pv set 2	pv set 3
GP 1	231	246	250	0.0062	0.0058	0.0057	0.1062	0.0928	0.0814
GP 2	250	246	250	0.0061	0.0058	0.0057	0.0013	0.0013	0.0012
GP 3	250	246	250	0.0061	0.0058	0.0057	0.0466	0.0438	0.0414

	Optimal stopping value $\hat{M}$			Overall averaged SSR			Overall averaged EC		
adhocbs	pv set 1	pv set 2	pv set 3	pv set 1	pv set 2	pv set 3	pv set 1	pv set 2	pv set 3
GP 1	66	10	90	0.0171	0.0199	0.0175	0.8401	0.9378	0.8681
GP 2	228	16	131	0.0166	0.0193	0.0174	0.0034	0.0041	0.0035
GP 3	219	16	237	0.0166	0.0193	0.0180	0.1861	0.2071	0.1926

	Optimal stopping value $\hat{M}$			Overall averaged SSR			Overall averaged EC		
stickym	pv set 1	pv set 2	pv set 3	pv set 1	pv set 2	pv set 3	pv set 1	pv set 2	pv set 3
GP 1	6	10	4	0.0284	0.0249	0.0303	3.5360	3.5139	3.5490
GP 2	250	181	132	0.0259	0.0263	0.0267	0.0017	0.0018	0.0018
GP 3	250	184	132	0.0259	0.0263	0.0267	0.0639	0.0641	0.0661

Table 6.6: Optimal stopping value $\hat{M}$, obtained by cross-validation on the validation sample VS of **subsample 3**, overall averaged SSR (6.11) and empirical criterion (6.13) on VS for all combinations of $\{(\text{pv set}_i, \text{GP}_j) | i, j \in \{1, 2, 3\}\}$ and starting model $\in$ {regtree, adhocbs, stickym}.

	Optimal stopping value $\hat{M}$			Overall averaged SSR			Overall averaged EC		
bvar	pv set 1	pv set 2	pv set 3	pv set 1	pv set 2	pv set 3	pv set 1	pv set 2	pv set 3
GP 1	190	213	243	0.0040	0.0105	0.0044	0.0806	0.1434	0.0897
GP 2	235	224	243	0.0039	0.0104	0.0044	0.0011	0.0020	0.0013
GP 3	235	227	243	0.0039	0.0104	0.0044	0.0377	0.0777	0.0417

	Optimal stopping value $\hat{M}$			Overall averaged SSR			Overall averaged EC		
dsfm	pv set 1	pv set 2	pv set 3	pv set 1	pv set 2	pv set 3	pv set 1	pv set 2	pv set 3
GP 1	163	2	2	0.0025	0.0028	0.0028	0.0614	0.0624	0.0624
GP 2	163	2	2	0.0025	0.0028	0.0028	0.0008	0.0008	0.0008
GP 3	163	2	2	0.0025	0.0028	0.0028	0.0243	0.0269	0.0269

Table 6.7: Optimal stopping value $\hat{M}$, obtained by cross-validation on the validation sample VS of **subsample 3**, overall averaged SSR (6.11) and empirical criterion (6.13) on VS for all combinations of $\{(\text{pv set}_i, \text{GP}_j)|i, j \in \{1, 2, 3\}\}$ and starting model $\in$ {bvar, dsfm}.

Detailed cross-validation process w.r.t. GP_1 for dsfm-treefgd on subsample 1

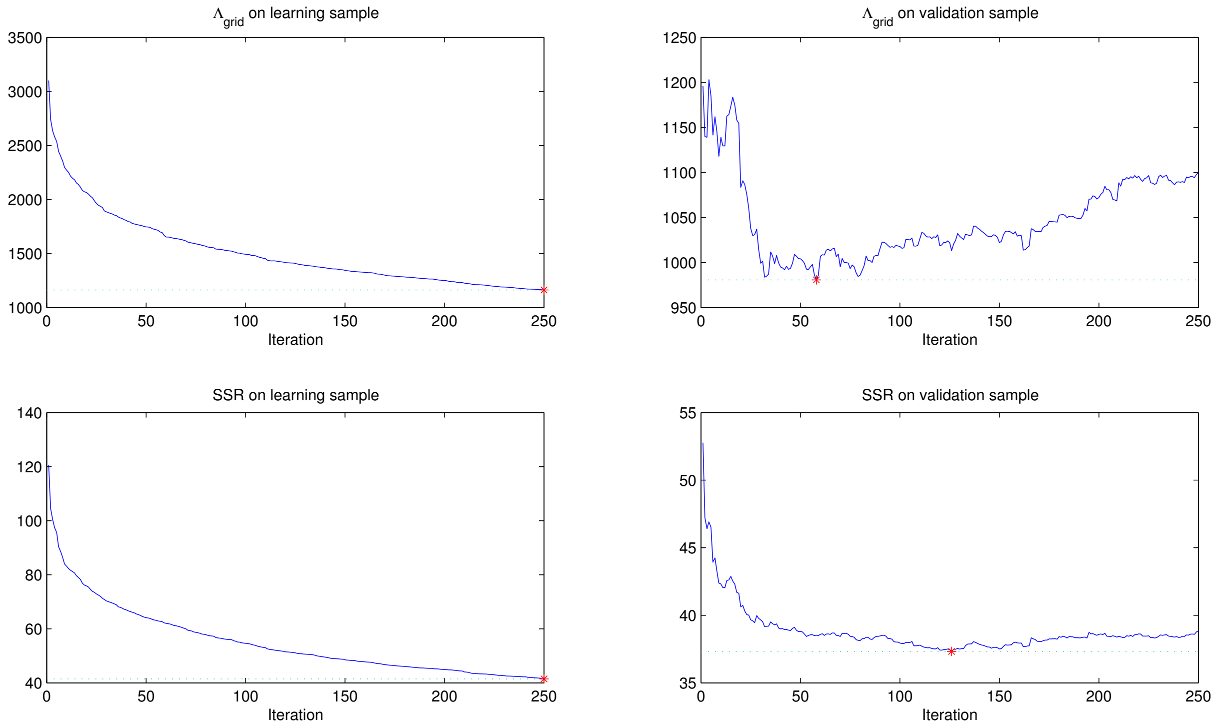


Figure 6.6: The in-sample period IS $= \{1, \ldots, 250\}$ of subsample 1 is split into a learning sample LS $= \{1, \ldots, 175\}$ and a validation sample VS $= \{176, \ldots, 250\}$. Plot of the empirical local criterion Λ_{grid} (5.9) w.r.t. GP_1 on LS (*top left*) and on VS (*top right*) in dependence of $\hat{F}_j \equiv$ dsfm-treefgd model with j additive expansions, $j \in \{1, 2, \ldots, 250\}$. The graphics on the *bottom* contain the same plots for the sum of squared residuals. The red dots and the horizontal dotted line mark the minimal value for each criterion.

6.2.3 Relative importance of predictor variables

When interpreting boosting algorithms, the relevance of the different predictors needs to be addressed. Performing cross-validation w.r.t. GP_1 on all five subsamples leads to a total number of $\hat{M}_{\text{total}} = \sum_{i=1}^{5} \hat{M}_i$ base learners. Each additive expansion consists of two regression trees with $L = 5$ leaves, one for the calls and one for the puts. Hence, the number of split variables sums up to a total of $\hat{M}_{\text{total}} \cdot (5 - 1) \cdot 2$. A regression tree as base learner automatically selects split variables and cut values when it is fitted to data. Table 6.8 summarizes how often each group of predictor variables is chosen in the boosting procedure.

Regardless which starting model or predictor variable set is used, the location parameters m and τ are chosen about 70% of times. The cut values for m lie uniformly distributed over the interval $[0.4, 1.5]$. Only 5% of the cut values are greater than 1.5, the mean is 0.9067, and the maximum 2.4007. The distribution of cut values for split variable τ is concentrated around small values. 25% are smaller than 0.0164 (6 days), 50% are smaller than 0.0466 (17 days), the average is 0.2112 (77 days).

In only 30% of times, regression trees select time-lagged or time-leading factors as split variables. Including forecast of time-leading factors in the predictor variable sets turns out to be as important as including time-lagged factors: both are chosen about the same number of times.

6.2.4 Comparison of different models

Besides minimizing the overall averaged EC (6.13) on the validation sample, the tree-boosting algorithm also leads to a reduction of the SSR. The upper panel of Figure 6.7 contains boxplots of the daily averged SSR_t (6.10) over the whole IS period for different IVS $G(\cdot)$ models, in 'best fitted' form as discussed in Section 6.1.3, and their tree-boosted versions $F_M(\cdot)$, based on 'complexity reduced' models $F_0(\cdot)$, for subsample 4, pv set 3 and GP_1.

The variances of daily averaged SSR_t for all $F_M(\cdot)$ models are several times smaller than for the $G(\cdot)$ models. Tree-boosting also moves all quartiles as well

Split variables

		$\hat{M}_{\text{total}}$	# splits	m	τ	close	DGS	HNG	past	contemp	future
pv set 1	regtree	1,066	8,528	39.8%	29.6%	30.6%			1,146	269	1,191
	adhocbs	228	1,824	46.4%	22.8%	30.8%			262	49	251
	stickym	10	80	36.3%	53.7%	10.0%			2	4	2
	bvar	430	3,440	35.6%	29.5%	34.9%			496	138	568
	dsfm	282	2,256	20.8%	35.6%	43.6%			459	87	437
pv set 2	regtree	1,185	9,480	42.0%	30.7%	9.2%	18.1%		1,173	289	1,124
	adhocbs	152	1,216	43.9%	22.3%	7.9%	25.9%		199	46	166
	stickym	21	168	29.8%	37.4%	1.8%	31.0%		7	3	45
	bvar	446	3,568	36.2%	28.4%	12.2%	23.2%		614	105	544
	dsfm	11	88	36.4%	43.2%	3.4%	17.0%		10	1	7
pv set 3	regtree	890	7,120	42.3%	31.1%	0.8%	6.6%	19.1%	841	243	806
	adhocbs	182	1,456	43.7%	19.3%	0.8%	9.3%	26.9%	235	53	250
	stickym	13	104	35.6%	37.5%	1.0%	1.9%	24.0%	2	20	6
	bvar	497	3,976	38.8%	25.9%	1.0%	9.6%	24.7%	658	127	619
	dsfm	11	88	31.8%	47.7%	1.2%	10.2%	9.1%	10	2	6
	TOTAL	5,424	43,392	39.4%	29.3%	15.9%	9.1%	6.4%	6,114	1,436	6,022

Table 6.8: Summary of automatically chosen split variables. Each row shows the composition of selected predictor variables when applying the tree-boosting algorithm on the five subsamples for a variety of starting models $F_0(\cdot)$ and predictor variable sets. m and τ are location parameters, close is the option's underlying closing price. DGS are treasury constant maturity rates with different maturities. HNG stands for option prices calculated according to the Heston Nandi GARCH model. Time-lagged and leading versions of close, DGS and HNG are included as predictor variables; past $= \{t-5, ..., t-1\}$, contemporaneous $= t$ and future $= \{t+1, ..., t+5\}$.

as upper and lower whiskers towards zero. Outliers are not shown in the boxplot, but the range

$$\max\{\text{daily averaged } \mathrm{SSR}_t, t \in \mathrm{IS}\} - \min\{\text{daily averaged } \mathrm{SSR}_t, t \in \mathrm{IS}\}$$

also shrinks in all cases. The numbers are

$$\{0.1121, 0.5167, 799.9573, 0.1920, 0.0493\}$$

for regtree, adhocbs, stickym, bvar and dsfm, respectively. Cross-validated optimal stopping values $\hat{M}$ w.r.t. GP_1 are {40, 64, 2, 7, 1}. After applying treefgd, the ranges of daily averaged SSR_t for the improved $F_M(\cdot)$ models shrink to

$$\{0.0192, 0.1891, 776.7169, 0.0558, 0.0438\}.$$

The high IS variations of daily averaged SSR_t for adhocbs and stickym are irritating at first, but this is a result of the model selection process for $G(\cdot)$ that chooses the best tuning parameters such that Eq. (5.11) holds. The average reductions of daily averaged SSR_t for all combinations of subsamples and predictor variable sets obtained by treefgd are $\{89\%, 75\%, 11\%, 81\%, 28\%\}$ compared to regtree, adhocbs, stickym, bvar and dsfm, respectively.

The lower panel of Figure 6.7 shows a boxplot of daily averaged SSR_t over a 60 days OS period for subsample 4, pv set 3 and GP_1. Results for bvar are not obtained due to instable forecasts of this model. The distribution of daily averaged SSR_t for stickym seems stable, but this is not the case in general. If we had chosen for example subsample 5, then the OS prediction error would have been several thousand times higher. This is actually interesting because the special day of interest for subsample 4 is the first business day after 9/11.

This event definitively marks a structural break. The 20 OS days before 9/11 have an overall averaged SSR of

$$\{0.0030, 0.0191, 0.0402, 0.0065\}$$

for regtree-treefgd, adhocbs-treefgd, stickym-treefgd and dsfm-treefgd. The average over the whole 60 OS days triples at least. Tree-boosting does not bridge over the structual break implied by 9/11, but it reduces at least the variation of daily averaged SSR_t in some cases.

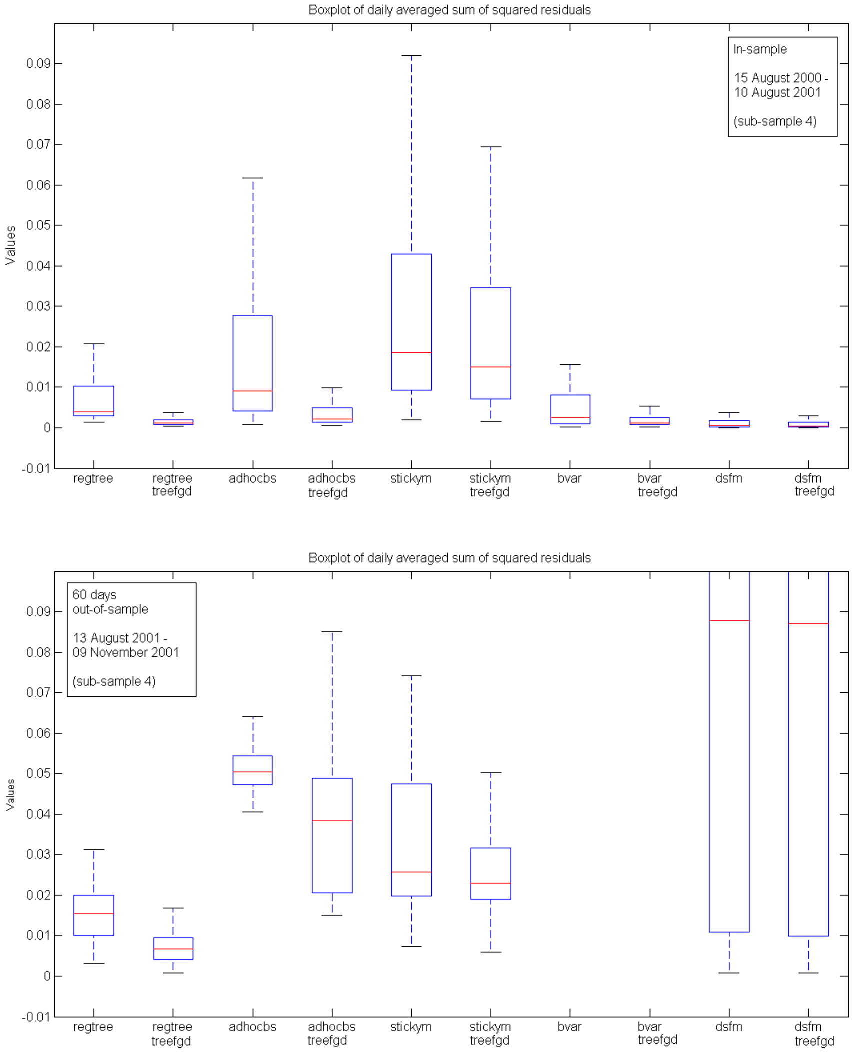


Figure 6.7: Boxplot of $Y :=$ daily averaged SSR_t (6.10) over the whole IS period (*top*) and over a 60 days OS period (*bottom*) of subsample 4. A boxplot visualizes the distribution of observed Ys and consists of two short horizontal lines, the lower whisker $= \arg\max(Y \leq 1.5 \cdot \mathrm{IQR})$ and the upper whisker $= \arg\min(Y \geq 1.5 \cdot \mathrm{IQR})$, the blue box (with lower edge at Q1 $= 25\%$ quantile and upper edge at Q3 $= 75\%$ quantile, representing the interquartile range IQR $=$ Q3 $-$ Q1) and the red horizontal line in the blue box at Q2 $= 50\%$ quantile (median). Observed Ys that lie outside the area between lower and upper whisker are considered as outliers and are not plotted. OS prediction for bvar is not possible due to instable forecasts of this model.

Table 6.9 compares the overall averaged SSR performances over 60 OS days for all subsamples. Two models are not listed, stickym and bvar. The former because the use of the term structure of the IVS on the last IS day to interpolate IV 60 days in the future leads to a very large OS error[28]. The latter because the Bayesian vector autoregression model predicts unexpected high IV values after maximum 10 OS days. Tree-boosting reduces the overall averaged SSR over 60 OS days by $\{58\%, 31\%, 3\%\}$ on average for all combinations of subsamples and predictor variable sets compared to regtree, adhocbs and dsfm. Subsample 4 raises most problems for starting models, especially for dsfm. The daily averaged mean square forecast errors (6.10) remain very high for a period of 10 days after the special day of interest and then coincidentally approach normal levels. A single tree-boosting iteration can not change much in such a situation.

6.2.5 Dispersion trading

Dispersion trades bet on the degree to which constituent stocks of an index disperse, i.e. how the components evolve relatively to each other. In the early 1990s, hedge funds started selling index options and simultaneously buying options on the constituents. When appropriately hedged, such a strategy is in fact short of correlation risk (Driessen et al., 2009).

> *The basic idea of a dispersion trade is trading index volatility against component's volatility, thereby taking exposure in the average correlation of the index. Traditionally, a dispersion trade was set up with vanilla options since it was the only way of trading volatility. Then, variance swaps, which allow for trading volatility directly, have been developed. Today, these are quite liquid and are most common for trading dispersion* (Vonhoff, 2006, p. 44).

As pointed out by Vonhoff, variance swaps offer a more comfortable way for dispersion trading. The payoff ψ_T of a variance swap at expiry T is given by

---

[28] The IVS predicted by the stickym model in Figure 6.12 on the special day of interest of subsample 4 (21 days OS) has with 0.2597 actually one of the lowest daily averaged SSR_t values during the 60 OS period. They are considerably higher otherwise.

Overall averaged SSR performance over 60 OS days

		Only starting model			Improved with treefgd		
	Subsample	regtree	adhocbs	dsfm	regtree	adhocbs	dsfm
pv set 1	1	0.0188	0.0223	0.0280	0.0166	0.0267	0.0260
	2	0.0118	0.0661	0.4993	0.0032	0.0191	0.4979
	3	0.0112	0.0264	0.1325	0.0022	0.0115	0.1304
	4	0.0225	0.0755	15.4498	0.0061	0.0252	15.4485
	5	0.0117	0.1362	0.0267	0.0066	0.0910	0.0240
pv set 2	1	0.0188	0.0223	0.0280	0.0149	0.0234	0.0271
	2	0.0120	0.0661	0.4993	0.0027	0.0173	0.4987
	3	0.0112	0.0264	0.1325	0.0026	0.0378	0.1314
	4	0.0225	0.0755	15.4498	0.0066	0.0259	15.4492
	5	0.0117	0.1362	0.0267	0.0055	0.0968	0.0243
pv set 3	1	0.0188	0.0223	0.0280	0.0149	0.0279	0.0275
	2	0.0120	0.0661	0.4993	0.0028	0.0260	0.4987
	3	0.0112	0.0264	0.1325	0.0028	0.0213	0.1314
	4	0.0225	0.0755	15.4498	0.0101	0.0376	15.4492
	5	0.0117	0.1362	0.0267	0.0045	0.0968	0.0244

Table 6.9: 60 days out-of-sample performance in terms of overall averaged sum of squared residuals (6.11) for different predictor variable sets and subsamples (see Section 6.1 for detailed specifications).

a notional amount paid per variance point multiplied by the difference of future realized variance $\frac{1}{T-t}\int_t^T \sigma(\tilde{t}, S_{\tilde{t}})^2 d\tilde{t}$ and the strike variance K^2_{var}, which is fixed at time t such that the fair value of the swap $\pi_t(\psi_T)$ is zero,

$$K^2_{\text{var}} = \frac{1}{T-t}\mathrm{E}_{\mathbb{Q}}\left[\int_t^T \sigma(\tilde{t}, S_{\tilde{t}})^2 d\tilde{t} \,\middle|\, \mathcal{F}_t\right]. \tag{6.14}$$

Demeterfi, Derman, Kamal, and Zou (1999) derived a formula for K^2_{var} by constructing a hedge portfolio of OTM put and call options,

$$K^2_{\text{var}} = \frac{2}{T-t}e^{r(T-t)}\left(\int_0^{F_{t,T}} \frac{P_t(S_t, K, T)}{K^2}dK + \int_{F_{t,T}}^{\infty} \frac{C_t(S_t, K, T)}{K^2}dK\right) \tag{6.15}$$

with $F_{t,T} = S_t e^{r(T-t)}$. Carr and Wu (2009) showed that the return on a variance swap from t to T is

$$\frac{\int_t^T \sigma(\tilde{t}, S_{\tilde{t}})^2 d\tilde{t}}{\mathrm{MFIV}^2_T(t)} - 1, \tag{6.16}$$

with $\mathrm{MFIV}_T(t)$ the model-free IV (see Definition C.19). Jiang and Tian (2005) describe a procedure based on $\frac{1}{T-t}\mathrm{MFIV}^2_T(t)$ that delivers an estimator of (6.14).

The goal is to obtain an estimate of Eq. (6.16) for a variance swap that expires on the next option exchange settlement date at least 30 days in the future. Subsample 3 ends on $t = 13$ December 2000, one day before the third Friday of that month. Hence the chosen variance swap expires on the settlement day of the following month, $T = 19$ January 2001 (24th OS day in the database), $\Delta_{\text{ns30d}} = \frac{T-t}{365} = 37/365 \approx 0.1014$.

Estimation of $\mathbf{K^2_{var}}$ Cross-validation w.r.t. to GP_3 for a regtree-treefgd IVS model with predictor variables pv set 3 identifies an optimal stopping value $\hat{M} = 250$. The fitted model is evaluated on a fine grid with $m \in [0, 3]$ and $\tau = \Delta_{\text{ns30d}}$,

$$\hat{\sigma}^{\text{IV}}_t(m, \Delta_{\text{ns30d}}) = \hat{F}_{\hat{M}}(m, \Delta_{\text{ns30d}}, \mathbb{I}_{m>\exp(r\Delta_{\text{ns30d}})}, \mathit{factors}_t).$$

The estimated IVs are mapped to option prices

$$f_t(m) := \mathrm{BS}_t(S_t, \hat{\sigma}^{\text{IV}}_t(m, \Delta_{\text{ns30d}}), \mathbb{I}_{m>\exp(r\Delta_{\text{ns30d}})}, mS_t, T, r, q)$$

and K_{var}^2 is approximated by numerical integration,

$$\widehat{K_{\text{var}}}^2 = \frac{2}{T-t} e^{r(T-t)} \int_0^3 \frac{f_t(m)}{(mS_t)^2}\, dm. \tag{6.17}$$

For $r = 0.0676$, the interpolated T-bill rate at t with maturity at T from OptionMetrics, and $q = 0$ for simplicity, $\widehat{K_{\text{var}}}^2$ is approximatively $0.0575 = 0.2398^2$.

Estimation of RV Based on the limit theorem of Durrleman (2008) (see Section 3.3.1), $\sigma_{\tilde{t}}^{\text{IV}}(1,0)$ can be interpreted as spot volatility $\sigma(\tilde{t}, S_{\tilde{t}})$ of the underlying S&P 500 index. Cross-validation w.r.t. GP_2 yields $\hat{M} = 250$, hence regtree-treefgd does not need to be refitted. $\hat{F}_{\hat{M}}(1, 0, \mathit{cp\ flag}, \mathit{factors}_{\tilde{t}})$ is evaluated separately for calls and puts for all $\tilde{t} \in \text{IS}$. The average of the two time series is considered to be the estimated time series of spot volatilities $\hat{\sigma}_{\tilde{t}}^{\text{IV}}(1,0)$ and is plotted in Figure 6.8. The obtained values are in the range of $[0.23, 0.47]$, higher than estimated annualized conditional volatilities in Figures 6.2 and 6.3 throughout the year 2000.

Predictions of the spot volatility until the 24th OS day of subsample 3 are calculated and numerical integration under the graph of $\left\{\hat{\sigma}_{\tilde{t}}^{\text{IV}}(1,0)\right\}^2$ for $\tilde{t} \in [t, T]$ yields

$$\frac{1}{T-t} \int_t^T \left\{\hat{\sigma}_{\tilde{t}}^{\text{IV}}(1,0)\right\}^2 d\tilde{t} = 0.1158 = 0.3402^2 \tag{6.18}$$

as an approximation of actually observed average future realized variance

$$\frac{1}{T-t} \sum_{i=1}^{24} (\log S_{t_i} - \log S_{t_{i-1}})^2 = 0.0708 = 0.2660^2, \tag{6.19}$$

with $t_0 = t$ and $\{t_i | i \in \{1, 2, \ldots, 24\}\} \subset \text{OS}$. The estimate in Eq. (6.18) is quite accurate compared to the naïve estimate

$$\frac{1}{T-t} \sum_{i=1}^{24} (\log \hat{S}_{t_i} - \log \hat{S}_{t_{i-1}})^2 = 1.8747 \cdot 10^{-4} = 0.0137^2 \tag{6.20}$$

induced by a filtered historical simulation of the underlying S&P 500 index level. Estimates obtained from other IVS models are $0.4009 = 0.6332^2$ (adhocbs-treefgd), $0.6051 = 0.7779^2$ (stickym-treefgd), $0.0016 = 0.0404^2$ (dsfm-treefgd).

Further analysis of estimators (6.17) and (6.18) are problems left to future research.

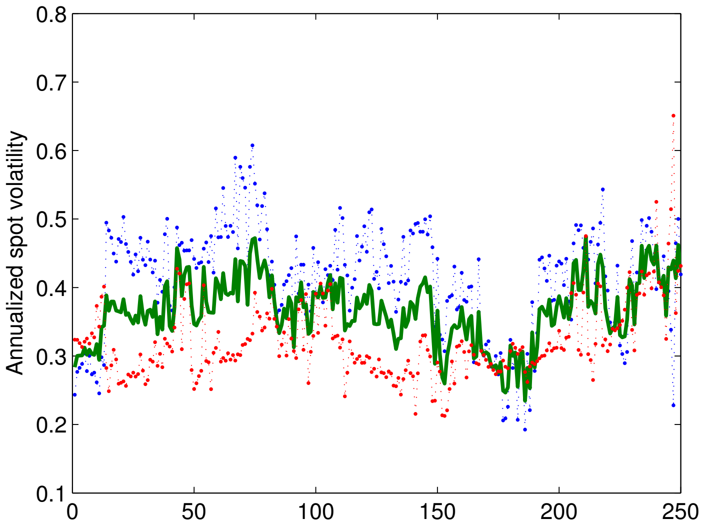

Recovering the spot volatility $\sigma(t, S_t)$ from implied volatility $\sigma_t^{\text{IV}}(1,0)$

Figure 6.8: ATM IV converges to the spot volatility $\sigma(t, S_t)$ as $\tau \to 0$ (see Section 3.3.1). The *blue dotted* line represents the time series of $\hat{F}_{\hat{M}}(\mathbf{x}^{\text{pred}}_{t,\text{call}})$ and the *red dotted* lines the time series of $\hat{F}_{\hat{M}}(\mathbf{x}^{\text{pred}}_{t,\text{put}})$ for the regtree-treefgd model $\hat{F}_{\hat{M}}(\cdot)$ fitted on subsample 3, optimal stopping value $\hat{M} = 250$ w.r.t. GP$_2$, $t \in \text{IS} = \{1, 2, \ldots, 250\}$, $\mathbf{x}^{\text{pred}}_{t,\text{call}} = (1, 0, 1, \mathit{factors}_t)$ and $\mathbf{x}^{\text{pred}}_{t,\text{put}} = (1, 0, 0, \mathit{factors}_t)$. The *solid green* line represents the times series of estimated spot volatilities $\hat{\sigma}^{\text{IV}}_{\tilde{t}}(1,0) = 0.5 \cdot \left(\hat{F}_{\hat{M}}(\mathbf{x}^{\text{pred}}_{t,\text{call}}) + \hat{F}_{\hat{M}}(\mathbf{x}^{\text{pred}}_{t,\text{put}})\right)$.

6.2.6 Robustness check

Recently, Battalio and Schultz (2006) discussed problems related to the use of the OptionMetrics Ivy database for academic studies when arbitrage violations must be taken into account. The problems are mainly due to the non-synchronicity of the prices stored in the database: in many cases, time stamps of the options differ from time stamps of the underlying[29]. To verify that the forecasting results discussed in this chapter are not a consequence of this non-synchronicity issue, a small robustness check is performed on subsample 1. A new data set of option prices is constructed. By minimizing the sum of the squared differences between calculated BS option prices and daily reported option prices with the nonlinear least-square method, implied estimates of index levels and dividend yields are calculated each day, similar to the procedure proposed by Manaster and Rendleman (1982). Using these values, moneyness and IVs of all options are recalculated. This new data set of option IVs is not depending on underlying prices that are asynchronous w.r.t. the reported closing option prices.

Table 6.10 shows no qualitative difference in the results. In particular, regtree-treefgd is still outperforming all other competitors and bvar still produces unreasonable forecasts after 10 OS days. Only small changes are observed when comparing the overall averaged SSR of the OS predictions based on the newly constructed database with the results obtained from the forecasts using the OptionMetrics Ivy database. The averaged overall EC values change because the weight function (5.10) depends on the location parameter m. The results for the new option data are qualitatively identical to the original data.

6.3 Summary

A regression tree (regtree) as starting model $F_0(\cdot)$ in combination with the tree-boosting algorithm (treefgd) beats all other models that have been considered in this chapter. $F_0(\cdot)$ partitions the $(m, \tau, cp\ flag)$ domain into regions with 10 different IV levels for calls and puts each. Two piecewise constant IVS are obtained, one for calls and one for puts. Each of them captures to a certain extent an average IVS over time. The additive expansions B_j in the boosting algorithm are regression trees as well, but with larger predictor spaces, including various time

---

[29] OptionMetrics collects closing prices for the underlying stock that occur no later than 4:00 pm, while option prices consist of quotes posted at 4:02 pm. If there is no closing price for the underlying at 4:00 pm, the price of the last transaction (could be hours before the close) is taken.

Robustness check 60 OS days performance

	New sample, overall averaged		Original sample, overall averaged	
	EC	SSR	EC	SSR
stickym	1.3198	0.0502	0.1418	0.0608
adhocbs	0.3647	0.0170	0.0360	0.0223
dsfm	0.8703	0.0205	0.0688	0.0280
bvar (10 OS days)	0.0513	0.0021	0.0069	0.0024
regtree-treefgd	0.3799	0.0101	0.0383	0.0149

Table 6.10: 60 days OS performance in terms of overall averaged SSR (6.11) and overall averaged EC (6.13) for different IVS models fitted on an alternative data set of option prices on subsample 1. A cross-validation w.r.t. GP_1 yields an optimal stopping value $\hat{M} = 105$ for regtree-treefgd on predictor variable set 3.

series of exogenous factors. The suggested cross-validation strategy for determining the optimal number of B_js in $F_M(\cdot) = F_0(\cdot) + \eta \sum_{j=1}^{M} B_j(\cdot)$ works very well. The fitted $\hat{F}_{\hat{M}}(\cdot)$ is a precise dynamical IVS model that does not overfit the data.

Example 6.2 As an illustrative example, let us examine the best OS predictions obtained from regtree-treefgd on t = 23 February 1999, the 35th OS day of subsample 2. There are 235 calls and 206 puts in the database for that day. Figure 6.9 illustrates how close the IV predictions to the effectively observed IV values are. The daily averaged SSR_t equals 0.0010, the daily averaged $\text{EC}_t = 0.0002$. The relative error (estimated IV − observed IV) / observed IV lies between -18% (regtree-treefgd underestimates true IV) and 50% (regtree-treefgd overestimates true IV) although the shape of each IV string is almost perfectly retained.

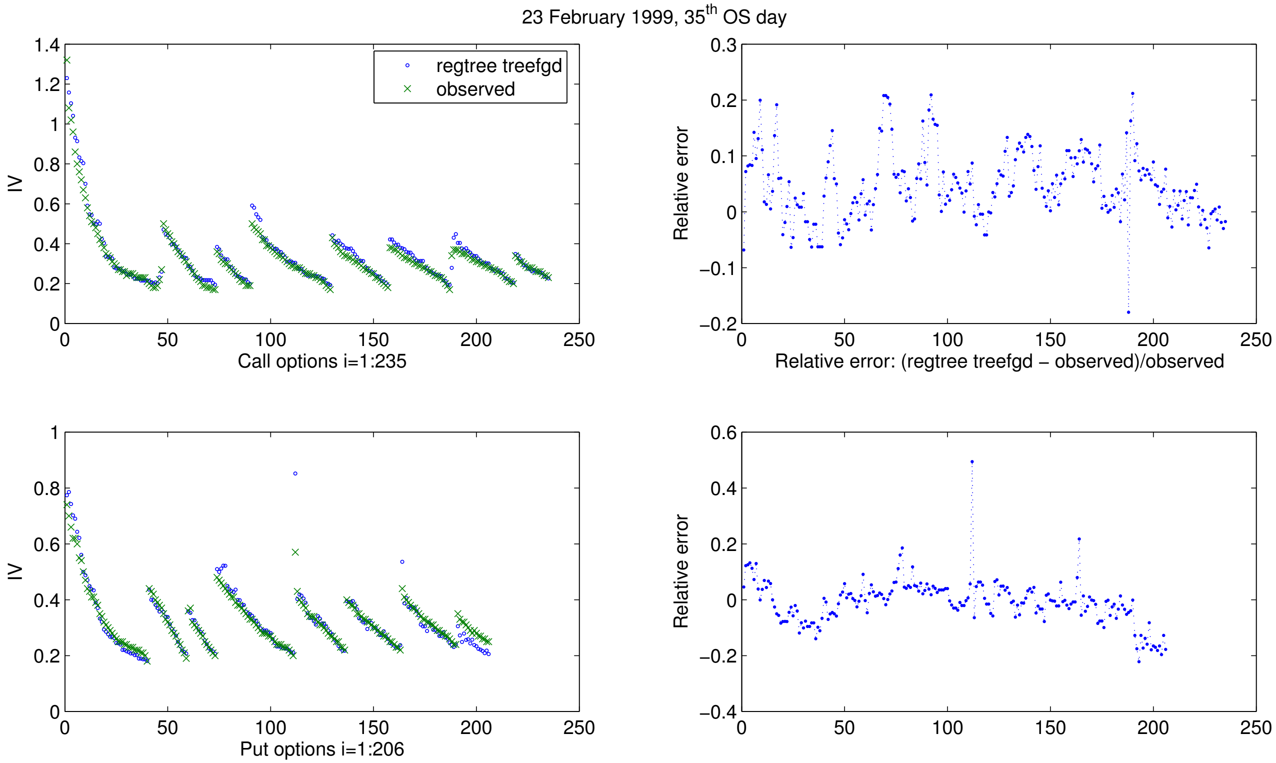


Figure 6.9: Comparison of estimated IVs $\hat{F}_{\hat{M}}(\mathbf{x}_{ti}^{\text{pred}})$ and actually observed IVs $\sigma_t^{\text{IV}}(m_{ti}, \tau_{ti}, cp\ flag_{ti})$ with $F_M(\cdot) \equiv$ regtree-treefgd ($\hat{M} = 242$, GP$_1$, pv set 3), $\mathbf{x}_{ti}^{\text{pred}} = (m_{ti}, \tau_{ti}, cp\ flag_{ti}, factors_t)$ for $t =$ 23 February 1999 (35th OS day of subsample 2) and $i \in \{1, 2, \ldots, 441\}$. The plotted IVs of calls (*upper left*) and puts (*lower left*) are sorted first by τ, then by m in ascending order. Relative errors $\hat{F}_{\hat{M}}(\mathbf{x}_{ti}^{\text{pred}}) \Big/ \sigma_t^{\text{IV}}(m_{ti}, \tau_{ti}, cp\ flag_{ti}) - 1$ are separately plotted for calls (*upper right*) and puts (*lower right*).

The ad hoc BS model (adhocbs) works best when its parameters are estimated on a filtered sample[30]. The sticky moneyness model (stickym) is a simple interpolation scheme that produces good IS results, but it is inappropriate for longer dated OS predictions. The Bayesian vector autoregressive model (bvar) catches the IVS dynamics reasonably well in-sample. Unfortunately, it does not produce stable OS predictions. OS prediction errors increase exorbitantly after a maximum of 10 OS days. The tree-boosting algorithm is not able to stabilize the function. The dynamic semi-parametric factor model (dsfm) provides the best dynamical IS fit of the IVS, problems only occur at the edges of observed (m, τ) locations. OS predictions are prone to errors because the latent factor loadings $\hat{\beta}_{t,k}$ are instable for $t \in$ OS.

Example 6.3 For Figures 6.10 and 6.11, fitted IVS models of the S&P 500 index are evaluated on a fine grid in the (m, τ) domain at t = 10 August 2001, the last IS day of subsample 4. This is the same day as in Figures 3.1 and 3.2. The improvements of the starting models thanks to the tree-boosting algorithm are apparent in the right column. Figures 6.12 and 6.13 plot forecasts of the IVS 21 days OS on the special day of interest (17 September 2001). bvar and dsfm have problems with IV predictions at the edge of the grid in the (m, τ) domain.

---

[30] A filtration in the spirit of Gonçalves and Guidolin (2006) removes up to 70% in a subsample. OLS estimation of the time-varying parameters $\mathbf{a}_t$ is not very robust.

$\hat{F}_0(\mathbf{x}_t^{\text{pred}})$ versus $\hat{F}_{\hat{M}}(\mathbf{x}_t^{\text{pred}})$, $t \in \text{IS}$
for {regtree, adhocbs, stickym}

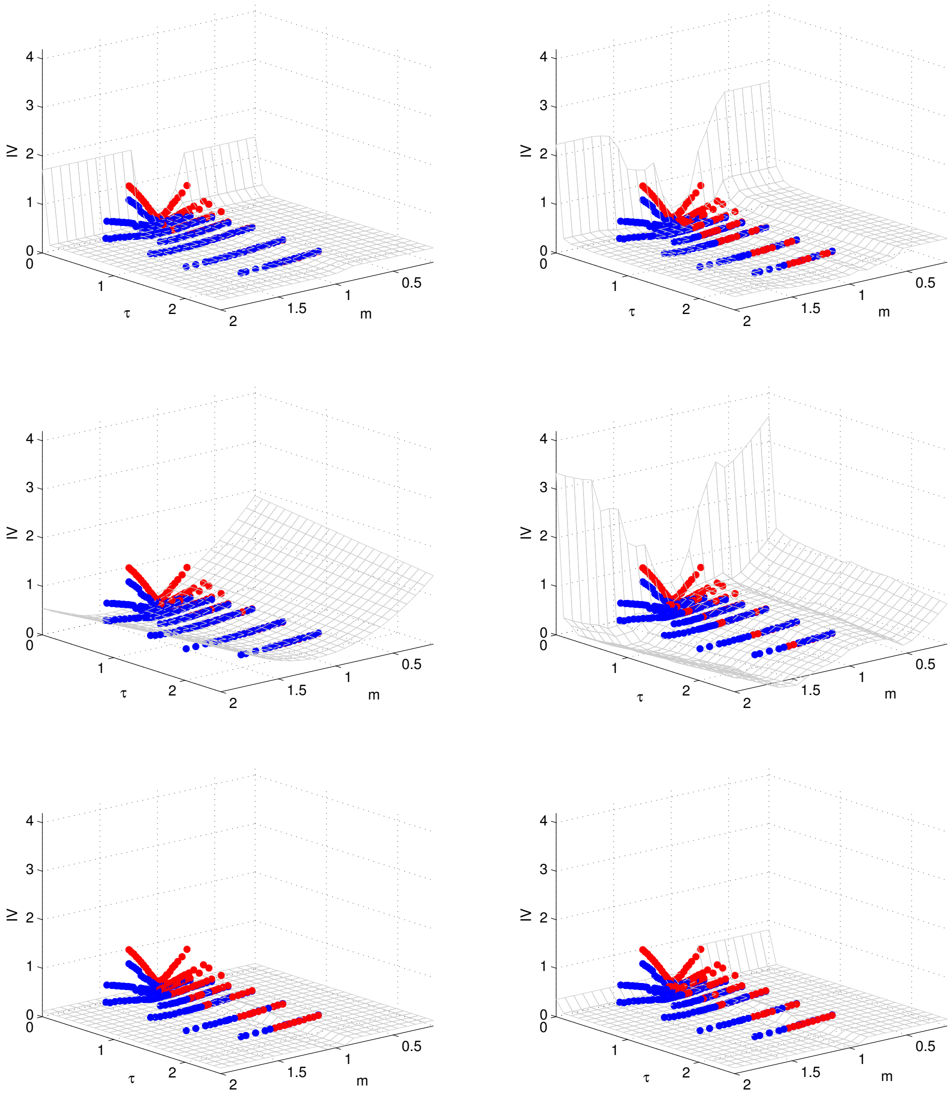


Figure 6.10: IS plot of the IVS using $\hat{F}_0(\mathbf{x}_t^{\text{pred}})$ (*left* column) and $\hat{F}_{\hat{M}}(\mathbf{x}_t^{\text{pred}})$ (*right* column), fitted w.r.t. GP_1 and pv set 3 for t = 10 August 2001 (last IS day of subsample 4). The *top* row is regtree, the *middle* row is adhocbs and the *bottom* row is stickym. *Blue* dots mark observed IVs of calls, *red* dots observed IVs of puts.

$\hat{F}_0(\mathbf{x}_t^{\text{pred}})$ versus $\hat{F}_{\hat{M}}(\mathbf{x}_t^{\text{pred}})$, $t \in \text{IS}$
for $\{\text{bvar, dsfm}\}$

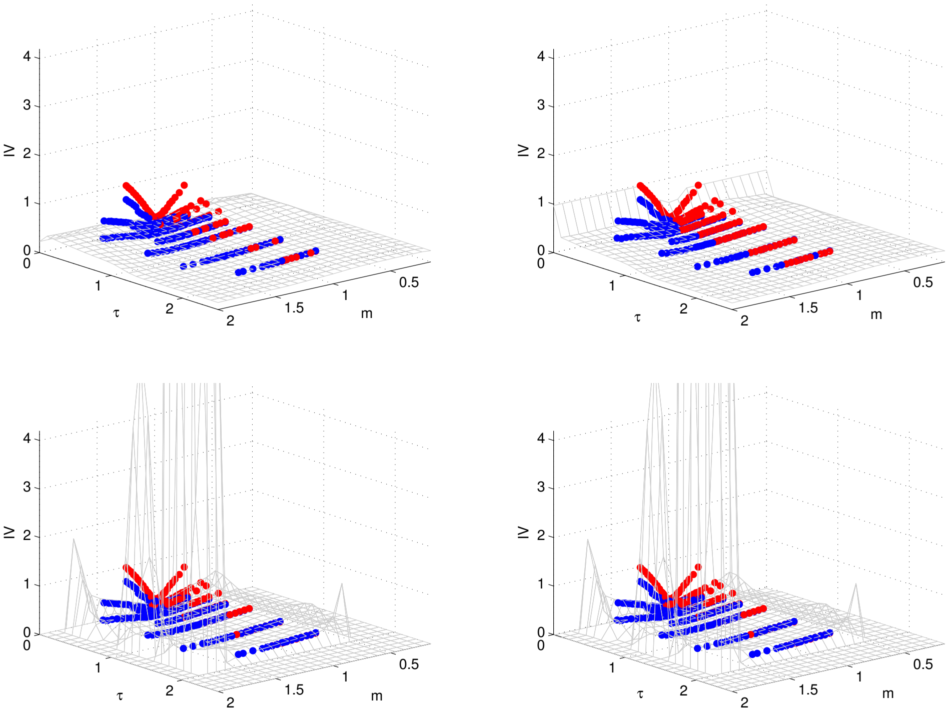
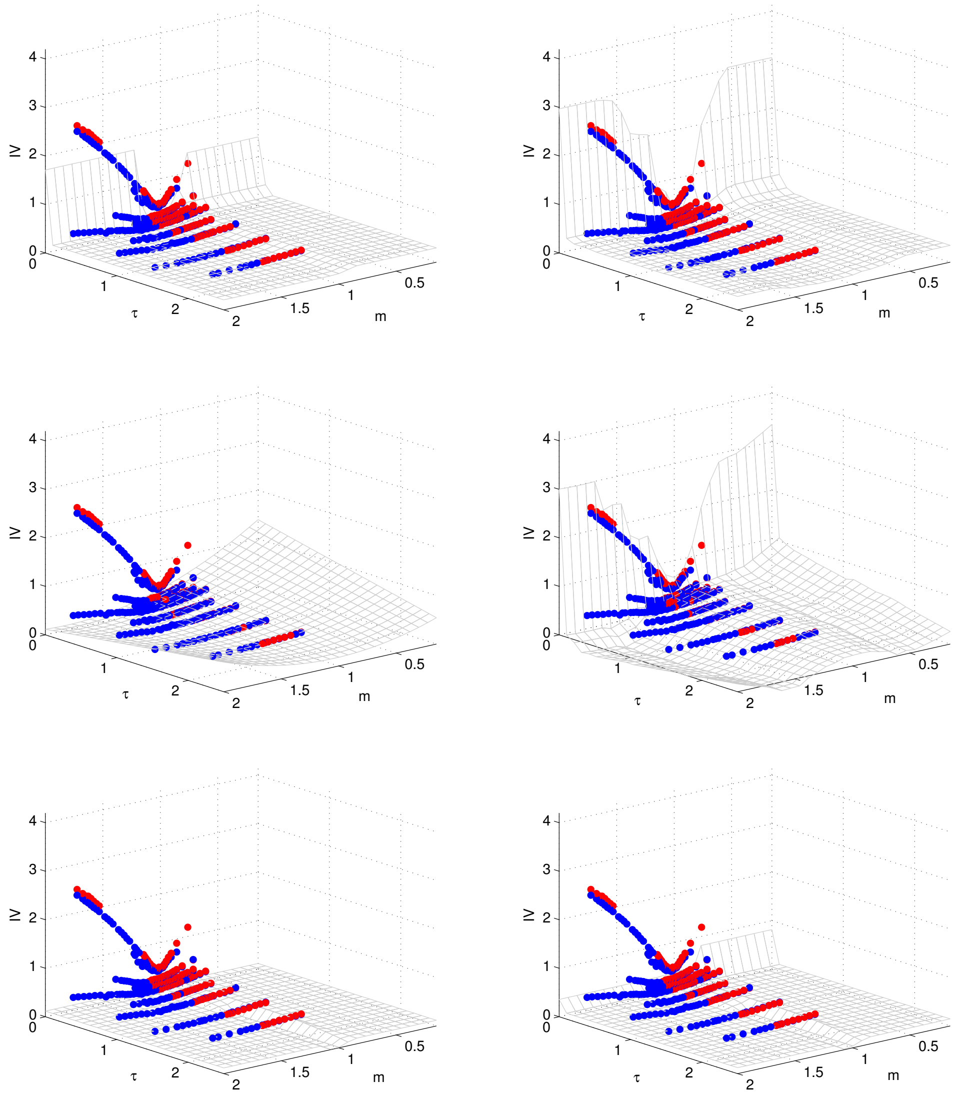


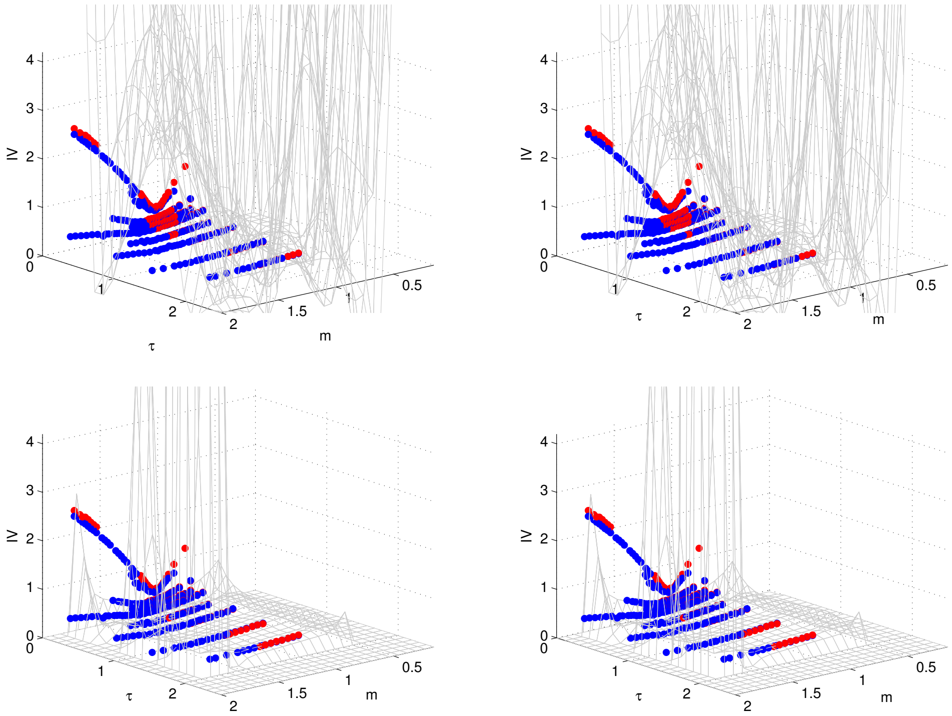
Figure 6.11: IS plot of the IVS using $\hat{F}_0(\mathbf{x}_t^{\text{pred}})$ (*left* column) and $\hat{F}_{\hat{M}}(\mathbf{x}_t^{\text{pred}})$ (*right* column), fitted w.r.t. GP_1 and pv set 3 for t = 10 August 2001 (last IS day of subsample 4). The *top* row is bvar and the *bottom* row is dsfm. *Blue* dots mark observed IVs of calls, *red* dots observed IVs of puts.



$\hat{F}_0(\mathbf{x}_t^{\text{pred}})$ versus $\hat{F}_{\hat{M}}(\mathbf{x}_t^{\text{pred}})$, $t \in \text{OS}$

for {regtree, adhocbs, stickym}

Figure 6.12: OS plot of the IVS using $\hat{F}_0(\mathbf{x}_t^{\text{pred}})$ (*left* column) and $\hat{F}_{\hat{M}}(\mathbf{x}_t^{\text{pred}})$ (*right* column), fitted w.r.t. GP_1 and pv set 3 for t = 17 September 2001 (21st OS day of subsample 4, special day of interest). The *top* row is regtree, the *middle* row is adhocbs and the *bottom* row is stickym. *Blue* dots mark observed IVs of calls, *red* dots observed IVs of puts.



$\hat{F}_0(\mathbf{x}_t^{\text{pred}})$ versus $\hat{F}_{\hat{M}}(\mathbf{x}_t^{\text{pred}})$, $t \in \text{OS}$
for {bvar, dsfm}

Figure 6.13: OS plot of the IVS using $\hat{F}_0(\mathbf{x}_t^{\text{pred}})$ (*left* column) and $\hat{F}_{\hat{M}}(\mathbf{x}_t^{\text{pred}})$ (*right* column), fitted w.r.t. GP_1 and pv set 3 for t = 17 September 2001 (21st OS day of subsample 4, special day of interest). The *top* row is bvar and the *bottom* row is dsfm. *Blue* dots mark observed IVs of calls, *red* dots observed IVs of puts.

Chapter 7

Trading strategy

The aim of this final chapter is to find option-only trading strategies for a limited set of investment instruments, consisting of 100 call and 100 put options on the constituents of the S&P 100 index. Investment decisions need to be made at the end of an option exchange settlement day t, i.e. usually the third Friday of a month. The available investment instruments that come into consideration at that time are restricted to approximatively one month to expiry ATM options, hence $m_t = \frac{K}{S_t} \approx 1$ and $\tau_t = T - t \approx 30/365 \approx 0.0822$. An option strategy is allowed to choose at most k long and k short option positions, and the portfolio needs to be held until the options expire. Both long and short positions have to be equally weighted in such a way that zero costs occur to form the portfolio, i.e. the cash flows CF_t when entering long and short option positions at time t must sum up to zero. Single option returns as well as portfolio returns are calculated in terms of hold-to-expiration returns, in other words net profits divided by (gross) exposure.

The assigned task is very particular and the given constraints limit the scope of possible investment strategies. The options are not hedged (‘naked’). The only way to limit possible loss is choosing the right options at time t. The data are extracted from OptionMetrics Ivy database, actually a subset of the sample used in Goyal and Saretto (2009) (henceforth called GS). According to these authors, volatility is mispriced because forming long-short option portfolios based on the log difference between historical volatility (HV) of the underlying and the option's IV earns high positive average returns.

Inspired by the sorting method proposed by GS, profitable long-short option trading strategies are defined that benefit from movements in the underlying stocks and exposure to volatility.

7.1 Settings

In the following sections, the used option data are presented and it is shown how to calculate returns for a single investment instrument as well as for a portfolio of options. Applying the sorting method of GS on two years worth of our data, a promising long-short option trading strategy is identified.

7.1.1 Data

The S&P 100 index consists of 100 large cap, blue chip U.S. companies across diverse industries and is dynamically reconstituted according to a set of published guidelines and policies. The primary criterion for index inclusion is the availability of individual stock options. The constituent list of 30 November 2006 is used as a basis and all option data are collected from OptionMetrics Ivy database for this fixed composition over the period 1 January 2002 until 31 December 2006. Index reconstitutions are ignored, only options on these 100 fixed underlyings are considered. This approach results in approximately 15 million data records. The sample is split into two complementary subsets, the first consisting of the 2002 to 2003 data mainly for checking the method of GS and the second consisting of the 2004 to 2006 data used for backtesting option strategies [out-of-sample (OS) period].

Options on the S&P 100 index and on its constituents have an American-style exercise feature. They are subject to PM settlement, i.e. the closing price of the underlying on the last trading day. The third Friday of the expiration month is the basis for settlements of exercises and assignments. The American-style nature of the available options has an influence on how implied volatilities are obtained. The options reported by OptionMetrics are calculated with the help of a proprietary algorithm based on the Cox-Ross-Rubinstein (CRR) binomial tree model, adapted to securities that pay dividends. Theoretically, there should be 100 ATM call and 100 ATM put options for all underlying stocks with one month expiry in each sample month. In practice, only options with $0.95 \leq m_t \leq 1.05$ closest to 1 and expiry date on the next option exchange settlement day T are chosen. Due to missing IV values and the degenerated option data structure (see Definition 3.5), there are only between 150 and 196 options available each month, with an average of 176. In total, our sample contains 10,358 such options.

7.1.2 Calculating option returns

Available investment instruments are analyzed in terms of hold-to-expiration return, which is calculated as the sum of cash flows at times t and T (net profit) divided by exposure at t, i.e. for $t < T$

$$\begin{array}{lll} \text{long call} & r_{t,T} = \frac{-C_{t,\text{ask}}+\max(S_T-K,0)}{|-C_{t,\text{ask}}|} \\ \text{long put} & r_{t,T} = \frac{-P_{t,\text{ask}}+\max(K-S_T,0)}{|-P_{t,\text{ask}}|} \\ \text{short call} & r_{t,T} = \frac{C_{t,\text{bid}}-\max(S_T-K,0)}{|C_{t,\text{bid}}|} \\ \text{short put} & r_{t,T} = \frac{P_{t,\text{bid}}-\max(K-S_T,0)}{|P_{t,\text{bid}}|} & . \end{array} \tag{7.1}$$

Remark 7.1 Portfolio returns are calculated as total net profit divided by gross exposure. For a long-short portfolio, the gross exposure $E_{\text{gross}} = E_{\text{long}} + E_{\text{short}}$, the sum of (absolute) exposure in long (E_{long}) and short positions (E_{short}), represents the absolute level of investment bets. Option transaction costs are taken into account in (7.1) by using bid and ask prices. Therefore, buying a call option and simultaneously shorting the same contract does not yield a 0% return over a one month holding period to expiration because the total net profit $\text{CF}_T - \text{CF}_t = C_{t,\text{bid}} - C_{t,\text{ask}} < 0$.

Note 7.2 The constraint of forming only zero-cost, equally weighted long-short portfolios implies that $E_{\text{long}} = E_{\text{short}} = E_{\text{gross}}/2$ and determines the number of contracts that need to be bought at ask prices or sold at bid prices. The return of a long only (short only) portfolio is the average of the single option returns. The return of a zero-cost, equally weighted long-short portfolio is given by the average of the long only and the short only equally weighted portfolios.

Example 7.3 Let us consider a portfolio long 1 call, 5 puts and short 6 puts, each option position on a different underlying. Let the ask prices at time t be \$1.50 (long call) and \$0.30 (long put), the bid price \$0.50 (short put) such that the portfolio is equally weighted and zero costs occur to set it up. Assume that the payoffs of the three single options at expiration T are \$1.80, \$0.57 and \$0.00. Therefore, the returns are 20%, 90% and 100%, respectively. The total net profit is $1 \cdot (\$\text{-}1.50 + \$1.80) + 5 \cdot (\$\text{-}0.30 + \$0.57) + 6 \cdot (\$0.50 - \$0.00) = \$4.65$, the gross exposure of the portfolio \$6.00. The portfolio return is $\$4.65/\$6 = 77.5\%$, which is the same as $0.5 \cdot [0.5 \cdot (20\% + 90\%) + 100\%]$.

The left half of Table 7.1 reports descriptive statistics of single option returns from the unfiltered aggregated dataset. The right half contains the same statistics for the dataset filtered by the method of GS:

> *"We apply a series of data filters to minimize the impact of recording errors. First we eliminate prices that violate arbitrage bounds. Second we eliminate all observations for which the ask price is lower than the bid price, the bid price is equal to zero, or the bid-ask spread is lower than the minimum tick size (equal to \$0.05 for option trading below \$3 and \$0.10 in any other cases). Finally, following Driessen et al. (2009), we remove all observations for which the option open interest is equal to zero, in order to eliminate options with no liquidity"* (2009, p. 3).

Using the filters of GS excludes 26% of the data in the sample. In either case, long returns are right skewed and short returns strongly left skewed. Put returns have heavier tails than call returns. Only short put options have a positive average return.

	Unfiltered Data				Filtered Data			
	Long Call	Short Call	Long Put	Short Put	Long Call	Short Call	Long Put	Short Put
max	12.6667	1.0000	39.6667	1.0000	11.8000	1.0000	27.6739	1.0000
Q3	0.5732	1.0000	0.0955	1.0000	0.6000	1.0000	0.2340	1.0000
Q2	-0.6957	0.6672	-1.0000	1.0000	-0.6250	0.5900	-0.9825	0.9821
Q1	-1.0000	-0.7000	-1.0000	-0.1926	-1.0000	-0.7484	-1.0000	-0.3375
min	-1.0000	-19.5000	-1.0000	-39.6667	-1.0000	-18.2000	-1.0000	-29.6744
IQR	1.5732	1.7000	1.0955	1.1926	1.6000	1.7484	1.2340	1.3375
uw	2.9227	1.0000	1.7333	1.0000	2.9800	1.0000	2.0833	1.0000
lw	-1.0000	-3.2308	-1.0000	-1.9792	-1.0000	-3.3500	-1.0000	-2.3429
out > uw	225	0	400	0	162	0	203	0
out < lw	0	236	0	408	0	165	0	206
returns >= 0	1865	3231	1371	3747	1561	2592	1018	2395
returns < 0	3314	1947	3808	1425	2660	1629	2437	1060
mean	-0.0181	-0.0886	-0.2034	0.1204	-0.0158	-0.0956	-0.1583	0.0797
std	1.4204	1.6411	1.8629	2.0697	1.3572	1.5783	1.7443	1.9258
var	2.0174	2.6932	3.4704	4.2838	1.8420	2.4909	3.0427	3.7086
skewness	2.2941	-2.8999	7.9937	-7.6436	2.0104	-2.6573	6.9271	-6.8615
kurtosis	11.2176	18.6415	111.6780	98.9235	9.0371	16.5645	82.1616	78.6428

Table 7.1: Descriptive statistics of aggregated option returns in terms of hold-to-expiration (7.1). The table reports quartiles (min = Q0, Q1, Q2, Q3, max =Q4), interquartile range (IQR = Q3 − Q1), upper and lower whisker (uw $= \max\{x \in \text{aggregated data} | x \leq \text{Q3} + 1.5 \cdot \text{IQR}\}$, lw $= \min\{x \in \text{aggregated data} | x \geq \text{Q1} - 1.5 \cdot \text{IQR}\}$), number of outliers (out) above/below upper/lower whisker, number of positive and negative returns, mean, standard deviation (std), variance (var), skewness and kurtosis. The unfiltered data set consists of all returns that can be calculated from the 10,358 options in our sample (see Section 7.1.1). Filtration is done as in Goyal and Saretto (2009).

7.1.3 Inspiration for an option trading strategy

GS form option portfolios by sorting them in deciles[31] based on the log difference between historical volatility (HV) and implied volatility (IV) to *“neutralize the impact of movements in the underlying stocks as much as possible . . . by forming straddle portfolios and delta-hedged call portfolios”* (2009, p. 5). GS find highest (lowest) returns for decile portfolio 10 (1) and conclude that it is underpriced (overpriced). Their strategy exploits this mispricing and goes long in decile 10 and short in decile one. This essentially represents a bet on IV mean-reversion during the options’ remaining lifetime of one month. The zero-cost, long-short option spread strategy of GS is highly profitable in the long run when transaction costs are not taken into account.

GS use close to ATM options with one month expiry from the entire U.S. equity option market from 1996 to 2006. In total, there are 75,627 pairs of monthly call and put observations after filtration, on average 53 option pairs per month and decile portfolio. Their long-short strategy is costly since it involves buying and selling hundreds of options each month. Our sample is only a fraction of theirs, limited in the time series dimension and also cross-sectionally.

The portfolio performance (considering transaction costs) based on the sorting method of GS is checked on our dataset for the years 2002 and 2003. IV is calculated as an average of call and put option IV in order to limit measurement errors. HV is computed as the annualized standard deviation of realized daily log returns of the underlying stock over the most recent twelve months. Decile portfolios are formed following exactly the sorting procedure of GS. Table 7.2 summarizes descriptive statistics of monthly {long call, short call, long put, short put} portfolio return time series.

The average portfolio return when going long put options on stocks in decile 10 during 2002 and 2003 is -3.99%. Going short the same options, one would expect at least a positive return on average, but Panel D reports -4.45% instead. This is due to an outlier: without a loss of -585.60% for the short put portfolio chosen at $t = 21$

---

[31] The first decile contains 10% of the data with the lowest sorting criterion values, the last decile 10% with the highest ones.

Decile	1	2	3	4	5	6	7	8	9	10
Panel A: Long call decile portfolio returns										
mean	-0.0913	-0.0574	0.0149	-0.1015	-0.1772	-0.0780	-0.1049	0.0931	-0.1035	0.1278
std	0.7661	1.0664	0.9563	0.7784	0.8203	0.9076	0.8855	1.2756	1.1866	1.2341
min	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000
max	2.1474	3.1917	2.2205	1.7465	1.5348	1.9995	1.9742	3.9083	4.4106	3.2690
SR	-0.1192	-0.0539	0.0156	-0.1304	-0.2160	-0.0860	-0.1184	0.0730	-0.0873	0.1036
Panel B: Short call decile portfolio returns										
mean	0.0071	-0.0300	-0.1143	0.0091	0.0922	-0.0376	0.0074	-0.2500	-0.0243	-0.3890
std	0.8281	1.1834	1.0489	0.8521	0.9173	1.0793	1.0039	1.5423	1.3539	1.6512
min	-2.3318	-3.8099	-2.5061	-2.0438	-1.9826	-2.9498	-2.2319	-5.1605	-5.2202	-5.1194
max	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SR	0.0086	-0.0253	-0.1089	0.0107	0.1005	-0.0349	0.0074	-0.1621	-0.0179	-0.2356
Panel C: Long put decile portfolio returns										
mean	0.0661	-0.1798	-0.3178	-0.3442	-0.0922	-0.3484	-0.1404	-0.1342	-0.0319	-0.0399
std	1.4495	0.8584	0.9455	0.6977	1.1341	0.8080	1.3727	0.9749	1.0393	1.3502
min	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000
max	5.0014	2.6226	3.4850	1.1987	3.6864	2.7098	5.2405	2.1913	2.7537	5.1062
SR	0.0456	-0.2095	-0.3361	-0.4933	-0.0813	-0.4311	-0.1023	-0.1377	-0.0307	-0.0295
Panel D: Short put decile portfolio returns										
mean	-0.1490	0.1051	0.2581	0.2836	0.0176	0.2813	0.0344	0.0724	-0.0551	-0.0445
std	1.5605	0.9258	1.0397	0.7478	1.2231	0.9247	1.6219	1.0436	1.1392	1.5025
min	-5.4296	-2.9274	-3.9510	-1.3410	-3.9539	-3.3045	-6.5057	-2.4327	-3.1294	-5.8560
max	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9576	1.0000
SR	-0.0955	0.1135	0.2482	0.3792	0.0144	0.3043	0.0212	0.0693	-0.0484	-0.0296

Table 7.2: Decile portfolio returns sorted on log difference between HV and IV. The sorting method (inclusive filtration) is as in Goyal and Saretto (2009). The sample contains the year 2002 to 2003 option data (see Section 7.1.1). Equally weighted portfolios consist either of long call, short call, long put or short put option positions. Single option returns as well as portfolio returns are calculated in terms of hold-to-expiration returns (7.1). The table reports mean, standard deviation, minimum, maximum and Sharpe ratio (SR) of monthly decile portfolio return time series.

June 2002, the average portfolio return would be 21.96%. Big losses are uniformly distributed over the deciles. Up to {61%, 23%, 54%, 22%} of the {long call, short call, long put, short put} option returns in a decile portfolio are $\leq -100\%$. There is no clear order or monotonicity observable for decile portfolio performances. The sorting criterion seems to be able to assign enough winner options to decile portfolio 10, but long decile 10 and short decile 1 does not appear to be superior to any other long/short decile combination when the number of available investment instruments is small and the length of the sample period is short. A strategy that goes long calls and short puts on the stocks in decile 10 might work well on a larger cross-section of options over a longer period. These findings are used to construct a GS inspired trading strategy in Section 7.2.4.

7.2 Method

In this section, it is suggested to form portfolios based on predicted option returns. If the option's underlying stock price at expiry, S_T, were known, option returns $r_{t,T}$ in Eq. (7.1) could be easily calculated. With the help of filtered historical simulation (see Section 6.1.4), it is possible to generate future paths of the underlying stock price from the empirical distribution of the fitted standardized innovations, to obtain a distribution of possible S_T and to derive an estimator $\hat{S}_T$ of the future stock price. Even if a very large number of simulations were conducted, the predicted option return obtained by this approach would most likely be inaccurate simply because of the large standard error of $\hat{S}_T$ for a 30 days OS prediction (see Figure 6.5).

In order to limit the influence of this kind of error, option price changes are predicted over a shorter period $\delta t < T - t$. Having ATM options at time t, it can be assumed that options are either ITM or OTM at time $t + \delta t$. If it is believed that an option will be deep ITM at time $t + \delta t$, for example more than $2 \cdot \hat{\sigma}^{\text{IV}}_{t+\delta t}\sqrt{\tau_{t+\delta t}}$ away from strike level K, and $\hat{r}_{t,t+\delta t} = 87\%$ is predicted over such a short period, then this option should be bought. It is likely that the option's hold-to-expiration return is significantly positive, provided the underlying stock price does not behave too erratically over the remaining time to maturity, $\tau_{t+\delta t} = T - (t + \delta t)$.

An optimal value for the tuning parameter δt is empirically determined by checking the ratio of correctly predicted directions of IV change during in-sample periods. Clearly, δt should be chosen to be large enough, otherwise the desired moneyness state stability of the form

$$\operatorname{sign}(m_{t+\delta t} - 1) = \operatorname{sign}(m_T - 1) \tag{7.2}$$

is not guaranteed. On the other hand, IV and option return predictions are more reliable for smaller δt. Setting $\delta t = 10$ trading days, condition (7.2) holds for 75% of the available investment instruments in the sample. In other words, the observed long option returns over the period from t until $t + \delta t$, $r_{t,t+\delta t}$, and hold-to-expiration returns $r_{t,T}$ have 75% of the time the same sign. Hence, it should be possible to form profitable portfolios based on predicted option returns $\hat{r}_{t,t+\delta t}$.

The following sections explain how to predict IV changes $\widehat{\delta\sigma_t^{\mathrm{IV}}} = \hat{\sigma}_{t+\delta t}^{\mathrm{IV}} - \sigma_t^{\mathrm{IV}}$, option price changes $\widehat{\delta OP_t} = OP_{t+\delta t} - OP_t$, and finally option returns $\hat{r}_{t,t+\delta t} = \widehat{\delta OP_t}/OP_t$.

7.2.1 Predicting IV changes

The methodology for modelling the IVS introduced in Chapter 5 is used to track the evolution of an option's IV over a short OS period δt. The IVS of an underlying stock is fitted with a $F_M(\cdot) \equiv$ regtree-treefgd model for each subsample according to the procedure of Section 5.4.2 and the following specifications:

- $F_M(\cdot)$ is fitted on the last 100 trading days. The observations of the first 70 days form the learning sample, the remaining 30 days the validation sample.
- The cross-validation procedure in Step 2 of the tree-boosting algorithm is repeated 150 times.
- Since we are interested in OS forecasting short expiring ATM options,

  $$\mathrm{GP} = \{(m, \tau) | m \in \{0.9, 1, 1.1\}, \tau \in \{5/365, 20/365, 35/365\}\}$$

  is chosen as grid for determining the optimal stopping value $\hat{M}$.

- All tuning parameters in the tree-boosting algorithm and for the weight function (5.10) are set to default values (see Remark 5.7).
- The set of exogenous factors includes the closing price of the underlying stock, S&P 100 and S&P 500, the 13-week US Treasury-bill rate, CBOE volatility index (VIX), HNG02 and HNG11 option prices (see Table 6.4), IV values of fixed 30 days ATM call and put option on the underlying stock computed by the ad hoc BS model and the sticky moneyness model (see Section 3.3.4 and Footnote [26]).
- The extended predictor space $\mathbf{x}_t^{\text{pred}} = (m_t, \tau_t, \mathit{cp\ flag}, \mathit{factors}_t)$ contains five time-lagged and forecasted time-leading versions of each exogenous factor. Thus, $\mathbf{x}_t^{\text{pred}}$ is 124 dimensional.
- In order to obtain the required forecasts of exogenous factors, the filtered historical simulation procedure of Section 6.1.4 is implemented.

In this way, IV predictions

$$\hat{\sigma}^{\text{IV}}_{t+\delta t} = \hat{F}_{\hat{M}}(\hat{m}_{t+\delta t}, \tau_{t+\delta t}, \mathit{cp\ flag}, \widehat{\mathit{factors}}_{t+\delta t}) \tag{7.3}$$

are obtained for any underlying stock in our database and $t + \delta t \in \text{OS}$.

Definition 7.4 (Option tracking) *Tracking an option means following the evolution of IV over the remaining lifetime for a specific option with fixed contract characteristics (type, strike, expiry date).*

OS option tracking is difficult because one needs to evaluate $\hat{F}_{\hat{M}}(\cdot)$ at location $(m_{t+\delta t}, \tau_{t+\delta t})$ that corresponds exactly to the tracked option. Time to maturity $\tau_{t+\delta t}$ is deterministic since T is known. Moneyness $m_{t+\delta t}$ is a random variable and needs to be estimated by $\hat{m}_{t+\delta t} = K/\hat{S}_{t+\delta t}$. Prediction errors in $\widehat{\mathit{factors}}_{t+\delta t}$, $\hat{S}_{t+\delta t}$ and $\hat{m}_{t+\delta t}$ can amplify to a big prediction error in $\hat{\sigma}^{\text{IV}}_{t+\delta t}$.

Remark 7.5 In the option trading literature[32], this problem is usually circumvented by fitting a univariate time series model to all observed IVs of a tracked option, $\{\sigma_t^{\text{IV}}(m_t, \tau_t)|t \in \text{IS}\}$. Then the minimum mean squared error (MSE) forecasts for the desired number of periods into the future are used as $\hat{\sigma}_{t+\delta t}^{\text{IV}}$. Although $\hat{F}_{\hat{M}}(\cdot)$ explicitly depends on $\hat{S}_{t+\delta t}$, it is superior to forecasts of a parametrical time series models. The regtree-treefgd model improves upon four exogenous factors that are IV time series of fixed specification options with $m_0 = 1$ and $\tau_0 = 30/365$ obtained with adhocbs and stickym. The minimum MSE forecasts as well as the filtered historical forecasts for any of them have a lower ratio of correctly predicted IV changes than regtree-treefgd.

---

[32] See for example Harvey and Whaley (1992), Noh et al. (1994), Brooks and Oozeer (2002), Ahoniemi (2006).

Note 7.6 The proposed trading strategies will rely on an accurate IV forecast as close as possible to the expiry date T; otherwise moneyness does not remain stable until expiry, and expected hold-to-expiration returns $r_{t,T}$ are not in accordance with $r_{t,t+\delta t}$. Tracking accuracy tests show that IV predictions for $\delta t = 10$ trading days offer a good tradeoff between MSE and ratio of correctly predicted direction of IV changes

$$\widehat{\delta\sigma}_t^{\text{IV}} = \hat{\sigma}_{t+\delta t}^{\text{IV}} - \sigma_t^{\text{IV}} = \hat{F}_{\hat{M}}(\hat{m}_{t+\delta t}, \tau_{t+\delta t}, cp\ flag, \widehat{factors}_{t+\delta t}) - \sigma_t^{\text{IV}}, \tag{7.4}$$

which is of great importance for short-dated options (see Section 7.3.2).

7.2.2 Predicting option price changes

Using the BS formula (2.23) as mapping from option prices to IVs is a promising approach to predicting option price changes. The only problem is that σ_t^{IV} and the Greeks provided by OptionMetrics are *not* calculated with the BS formula; they are actually derived from a CRR binomial tree algorithm because the sample consists of American-type options only (see Section 7.1.1). BS Greeks (see Section 2.3) are functions of $(S_t, \sigma, cp\ flag, K, T, r, q)$. The delta reported by OptionMetrics, Δ_t, is usually different from $\Delta_t^{\text{BS}}(S_t, \sigma_t^{\text{IV}}, cp\ flag, K, T, r, q)$. Let

$$OP_t = \begin{cases} \frac{1}{2} \cdot (C_{t,\text{bid}} + C_{t,\text{ask}}) & \text{if } cp\ flag = 1 \\ \frac{1}{2} \cdot (P_{t,\text{bid}} + P_{t,\text{ask}}) & \text{if } cp\ flag = 0 \end{cases} \tag{7.5}$$

denote the mid price of an option. It is possible to find an implied risk-free interest rate $\hat{r}$ and dividend yield $\hat{q}$ by minimizing the least square errors,

$$(\hat{r}, \hat{q}) = \arg\min_{r,q} \left\| \begin{pmatrix} \mathrm{BS}_t(S_t, \sigma_t^{\mathrm{IV}}, cp\ \mathit{flag}, K, T, r, q) - OP_t \\ \Delta_t^{\mathrm{BS}}(S_t, \sigma_t^{\mathrm{IV}}, cp\ \mathit{flag}, K, T, r, q) - \Delta_t \\ \Gamma_t^{\mathrm{BS}}(S_t, \sigma_t^{\mathrm{IV}}, cp\ \mathit{flag}, K, T, r, q) - \Gamma_t \\ \nu_t^{\mathrm{BS}}(S_t, \sigma_t^{\mathrm{IV}}, cp\ \mathit{flag}, K, T, r, q) - \nu_t \\ \theta_t^{\mathrm{BS}}(S_t, \sigma_t^{\mathrm{IV}}, cp\ \mathit{flag}, K, T, r, q) - \theta_t \end{pmatrix} \right\|_2 \tag{7.6}$$

and to consider them as constant values between t and $t + \delta t$. This enables us to predict the option price changes over a period of length δt either using the direct BS mapping

$$\widehat{\delta OP_t} = \mathrm{BS}_{t+\delta t}(\hat{S}_{t+\delta t}, \hat{\sigma}_{t+\delta t}^{\mathrm{IV}}, cp\ \mathit{flag}, K, T, \hat{r}, \hat{q}) - OP_t \tag{7.7}$$

or the indirect option price sensitivity approach based on Equations (2.5) and (2.35)

$$\widehat{\delta OP_t} = \Delta_t \widehat{\delta S_t} + \frac{1}{2}\Gamma_t \left(\widehat{\delta S_t}\right)^2 + \nu_t \widehat{\delta \sigma_t^{\mathrm{IV}}} + \theta_t \delta t, \tag{7.8}$$

with $\widehat{\delta S_t} = \hat{S}_{t+\delta t} - S_t$ and $\widehat{\delta \sigma_t^{\mathrm{IV}}} = \hat{\sigma}_{t+\delta t}^{\mathrm{IV}} - \sigma_t^{\mathrm{IV}}$.

7.2.3 Predicting option returns

The predicted return for a long option position from t to $t+\delta t$ is denoted $\hat{r}_{t,t+\delta t} = \widehat{\delta OP_t}/OP_t$. It can be computed with the help of Eq. (7.7),

$$\widehat{POR1} := \frac{\mathrm{BS}_{t+\delta t}(\hat{S}_{t+\delta t}, \hat{\sigma}_{t+\delta t}^{\mathrm{IV}}, cp\ \mathit{flag}, K, T, \hat{r}, \hat{q}) - OP_t}{OP_t}. \tag{7.9}$$

Another way to forecast $\hat{r}_{t,t+\delta t}$ is derived from Eq. (7.8). A daily update of the Greeks can be incorporated to relax the strong assumptions of constant parameters $(\Delta_t, \Gamma_t, \nu_t, \theta_t, \hat{r}, \hat{q})$ over a relatively long period of $\delta t = 10$ trading days,

$$\widehat{POR2} := \left(\sum_{k=0}^{\delta t - 1} \hat{\Delta}_{t_k}^{\mathrm{BS}} \widehat{\delta S_{t_k}} + \frac{1}{2}\hat{\Gamma}_{t_k}^{\mathrm{BS}} \left(\widehat{\delta S_{t_k}}\right)^2 + \hat{\nu}_{t_k}^{\mathrm{BS}} \widehat{\delta \sigma_{t_k}^{\mathrm{IV}}} + \hat{\theta}_{t_k}^{\mathrm{BS}} \delta t_k \right) \Big/ OP_t \tag{7.10}$$

with

$$\begin{aligned}
t_k &= t + k \text{ trading days}\\
\delta t_k &= t_{k+1} - t_k = 1 \text{ trading day}\\
\widehat{\delta S_{t_k}} &= \hat{S}_{t_{k+1}} - \hat{S}_{t_k}\\
\widehat{\delta\sigma^{\mathrm{IV}}_{t_k}} &= \hat{\sigma}^{\mathrm{IV}}_{t_{k+1}} - \hat{\sigma}^{\mathrm{IV}}_{t_k}\\
\hat{\Delta}^{\mathrm{BS}}_{t_k} &= \Delta^{\mathrm{BS}}_{t_k}(\hat{S}_{t_k}, \hat{\sigma}^{\mathrm{IV}}_{t_k}, \mathit{cp\ flag}, K, T, \hat{r}, \hat{q})\\
\hat{\Gamma}^{\mathrm{BS}}_{t_k} &= \Gamma^{\mathrm{BS}}_{t_k}(\hat{S}_{t_k}, \hat{\sigma}^{\mathrm{IV}}_{t_k}, \mathit{cp\ flag}, K, T, \hat{r}, \hat{q})\\
\hat{\nu}^{\mathrm{BS}}_{t_k} &= \nu^{\mathrm{BS}}_{t_k}(\hat{S}_{t_k}, \hat{\sigma}^{\mathrm{IV}}_{t_k}, \mathit{cp\ flag}, K, T, \hat{r}, \hat{q})\\
\hat{\theta}^{\mathrm{BS}}_{t_k} &= \theta^{\mathrm{BS}}_{t_k}(\hat{S}_{t_k}, \hat{\sigma}^{\mathrm{IV}}_{t_k}, \mathit{cp\ flag}, K, T, \hat{r}, \hat{q}).
\end{aligned}$$

A third possibility is given by the filtered historical simulation forecasts of the underlying stock price,

$$\widehat{POR3} = \begin{cases} \frac{\max(\hat{S}_{t+\delta t} - K, 0)}{OP_t} - 1 & \text{if } \mathit{cp\ flag} = 1 \\ \frac{\max(K - \hat{S}_{t+\delta t}, 0)}{OP_t} - 1 & \text{if } \mathit{cp\ flag} = 0 \end{cases} \tag{7.11}$$

It is possible that $\widehat{POR1}$ and $\widehat{POR2}$ produce values that are in the range of -300% and less, although a long option return theoretically has a lower bound of -100%. Such negative $\widehat{\delta OP_t}/OP_t$ values only emerge when $\hat{S}_T < K \approx S_t$ and OP_t is small, hence indicating profitable short investments.

7.2.4 Portfolio formation

Zero-cost, equally weighted long-short option portfolios are formed by sorting the available investment instruments based on the predicted option return for a long option position from t to $t + \delta t$, $\hat{r}_{t,t+\delta t} = \widehat{\delta OP_t}/OP_t$. Possible sorting criteria are given by Equations (7.9), (7.10), and (7.11). The following strategies are defined, choosing at most k long and k short option positions:

Bullish strategy Call options with highest positive sorting criteria form the long positions and put options with smallest negative sorting criteria the short positions of the portfolio.

Bearish strategy Put options with highest positive sorting criteria form the long positions and call options with smallest negative sorting criteria the short positions of the portfolio.

Both strategies depend on the number of allowed long and short positions k as well as on the sorting criterion. Bull(5, $\widehat{POR1}$) denotes a bullish strategy with $k = 5$ sorted on $\widehat{POR1}$ and in the same way Bear(10, $\widehat{POR2}$) stands for a bearish strategy with $k = 10$ and $\widehat{POR2}$ as a sorting criterion.

Linking A single bad investment can have a great impact on portfolio returns. In the worst case, a long option position generates a loss of 100%, but the downside of a short option position is unlimited. A portfolio following the bullish (bearish) strategy is defined to be **short linked** if its short positions are put (call) options on the same underlyings as the ones of the chosen long call (put) option positions. **Long linked** if its long positions are call (put) options on the same underlyings as the ones of the chosen short put (call) option positions. The advantage of linking is that only one tail of the cross-sectional option return distribution needs to be correctly predicted. For example, when forming a Bear($k, \widehat{POR2}$, long linked) portfolio, if a short call turns out to be profitable, then the corresponding long put will also be profitable.

Goyal Saretto inspired strategy As suggested in Section 7.1.3, we chose the k long call and k short put options with the highest positive log difference between historical HV and IV from the filtered dataset. The strategy is denoted GS(k) and is actually a bullish short linked strategy based on a $\mathcal{F}_t$-predictable sorting criterion, i.e.

$$\mathrm{GS}(k) \equiv \mathrm{Bull}(k, \log(\mathrm{HV}/\mathrm{IV}), \text{short linked}). \tag{7.12}$$

Ranking strategy All available investment instruments from the unfiltered dataset (calls and puts) are sorted based on $\widehat{POR1}$, $\widehat{POR2}$ and $\widehat{POR3}$. Linearly spaced values between 0 (least favorable) and 1 (most favorable) are assigned to them. In the same manner, ranking values are assigned to calls (puts) of the filtered dataset when sorting them based on log(HV/IV) in order to form the long (short) portfolio. Using the weights $\mathbf{w} = (w_1, w_2, w_3, w_4)$, the options with the k highest weighted average ranking points form the long and the short portfolio, respectively.

This last strategy is denoted Ranking($k, \mathbf{w}$) and can also be short or long linked. $\tilde{\mathbf{w}} = (1, 1, 1, 4)$ is set as standard weights. The last ranking has slightly more weight than the three predicted option return sorting criteria together. They are supposed to choose options from the unfiltered dataset to enhance the filtered log(HV/IV) sorting. For $w_4 \to \infty$, Ranking($k, \mathbf{w}$) converges to GS(k).

7.3 Empirical results

In this section, the results of the empirical studies for the OS period between 2004 and 2006 is presented. The first settlement date is 16 January 2004, and the last one is 15 December 2006, for a total of 36 monthly subsamples.

7.3.1 Portfolio return time series

Tables 7.3, 7.4 and 7.5 list descriptive statistics of portfolio return time series for different strategies with $k = 5$, $k = 10$, and $k = 20$, respectively.

In general, the bullish strategy performs better than the bearish strategy. This is consistent with the direction in which the S&P 100 is moving, from initially 540.26 points on 19 December 2003, when the first portfolio formation in our sample took place, to 652.60 points on the last settlement date. The average annualized S&P 100 return during this period is 7.09%, as a result of 22 positive and 14 negative monthly returns. Short linking improves the bullish strategy most of the time. The average portfolio return of the short linked bearish strategy is negative for different k and sorting criteria, but at least somewhat better than without short linking. The performance of the long linked bullish strategy relative to the one without linking improves slightly, whereas the opposite is true for the bearish strategy. For increasing k, the mean portfolio return decays much more

	mean	std	SR	min	Q1	Q2	Q3	max
Bull(5, $\widehat{POR1}$)	-0.0401	1.1525	-0.0348	-4.8769	-0.2579	0.2710	0.7111	1.1455
Bull(5, $\widehat{POR1}$, short linked)	0.0791	0.8413	0.0941	-1.9316	-0.4234	0.4092	0.7019	1.1356
Bull(5, $\widehat{POR1}$, long linked)	0.0458	1.1978	0.0382	-4.9582	-0.0382	0.2201	0.6989	1.6497
Bear(5, $\widehat{POR1}$)	0.2098	0.9369	0.2240	-1.6218	-0.3599	0.1312	0.6975	2.7097
Bear(5, $\widehat{POR1}$, short linked)	0.1678	1.0235	0.1639	-1.6705	-0.4757	0.1524	0.8100	2.7180
Bear(5, $\widehat{POR1}$, long linked)	0.1309	0.7943	0.1648	-1.5735	-0.2045	0.0869	0.5722	2.7097
Bull(5, $\widehat{POR2}$)	-0.0802	1.2813	-0.0626	-6.0689	-0.4370	0.2278	0.6651	1.1629
Bull(5, $\widehat{POR2}$, short linked)	0.1148	0.8064	0.1423	-2.1101	-0.4839	0.3584	0.6651	1.0961
Bull(5, $\widehat{POR2}$, long linked)	-0.0561	1.3217	-0.0424	-6.3114	-0.4274	0.2259	0.6421	1.2063
Bear(5, $\widehat{POR2}$)	-0.1360	0.7423	-0.1833	-1.9540	-0.6204	-0.0389	0.4397	0.8692
Bear(5, $\widehat{POR2}$, short linked)	-0.0333	0.6460	-0.0515	-1.6120	-0.4309	0.0993	0.4312	0.8312
Bear(5, $\widehat{POR2}$, long linked)	-0.1326	0.7862	-0.1686	-1.9540	-0.6362	-0.1147	0.5678	0.9776
Bull(5, $\widehat{POR3}$)	0.2868	0.5401	0.5310	-0.7888	-0.0817	0.3051	0.6252	1.2814
Bull(5, $\widehat{POR3}$, short linked)	0.1331	0.8729	0.1525	-2.8320	-0.1894	0.2905	0.7498	1.2814
Bull(5, $\widehat{POR3}$, long linked)	0.2663	0.5112	0.5209	-0.8138	-0.0376	0.2777	0.5827	1.7044
Bear(5, $\widehat{POR3}$)	0.1289	1.0095	0.1277	-3.2515	-0.3338	0.0319	0.6149	2.5232
Bear(5, $\widehat{POR3}$, short linked)	0.1120	1.1285	0.0993	-2.2828	-0.5095	0.1506	0.6774	2.8119
Bear(5, $\widehat{POR3}$, long linked)	-0.1515	0.7697	-0.1969	-3.0677	-0.4712	-0.0612	0.2387	0.9683
GS(5)	0.1872	0.7952	0.2354	-1.8245	-0.1961	0.2034	0.6636	1.6935
Ranking(5, $\tilde{\mathbf{w}}$)	0.2155	0.8391	0.2568	-2.4823	-0.0758	0.3275	0.6836	1.5310
Ranking(5, $\tilde{\mathbf{w}}$, short linked)	0.2120	0.8739	0.2427	-3.0016	-0.0769	0.3140	0.8343	1.5120
Ranking(k=5, $\tilde{\mathbf{w}}$, long linked)	0.2001	0.9084	0.2203	-2.4969	-0.3804	0.4359	0.6328	1.9264

Table 7.3: Option portfolio returns for different strategies with at most $\mathbf{k} = \mathbf{5}$ long and $\mathbf{k} = \mathbf{5}$ short option positions. Monthly portfolio formations take place over a period of 36 months from 19 December 2003 until 17 November 2006. The table reports mean, standard deviation, Sharpe ratio (SR) and quartiles of the monthly zero-cost, equally weighted portfolio return time series.

	mean	std	SR	min	Q1	Q2	Q3	max
Bull(10, $\widehat{POR1}$)	-0.0706	1.1577	-0.0610	-3.4590	-0.2103	0.2565	0.5906	1.2480
Bull(10, $\widehat{POR1}$, short linked)	0.1723	0.7454	0.2312	-2.3958	-0.1849	0.2541	0.7684	1.1787
Bull(10, $\widehat{POR1}$, long linked)	-0.0781	1.1865	-0.0658	-3.5052	-0.2072	0.3034	0.6192	1.1398
Bear(10, $\widehat{POR1}$)	0.0344	0.7352	0.0468	-1.4652	-0.4208	0.0819	0.5133	1.3174
Bear(10, $\widehat{POR1}$, short linked)	0.0472	0.7749	0.0609	-1.8513	-0.5330	0.1144	0.6234	1.1539
Bear(10, $\widehat{POR1}$, long linked)	-0.0223	0.6769	-0.0330	-1.4814	-0.5059	-0.0814	0.5371	1.5362
Bull(10, $\widehat{POR2}$)	0.0330	0.8881	0.0372	-3.3178	-0.3708	0.1952	0.5733	1.0243
Bull(10, $\widehat{POR2}$, short linked)	0.0952	0.7832	0.1215	-2.5931	-0.5263	0.2838	0.6678	1.1235
Bull(10, $\widehat{POR2}$, long linked)	0.0464	0.8865	0.0523	-3.3573	-0.2883	0.2479	0.6346	1.1503
Bear(10, $\widehat{POR2}$)	-0.1486	0.7046	-0.2110	-1.9741	-0.5279	-0.0325	0.3505	1.1218
Bear(10, $\widehat{POR2}$, short linked)	-0.1264	0.7041	-0.1796	-1.8179	-0.4727	-0.1188	0.4533	1.1218
Bear(10, $\widehat{POR2}$, long linked)	-0.1425	0.7207	-0.1978	-1.9758	-0.5291	0.0173	0.3806	1.0874
Bull(10, $\widehat{POR3}$)	0.1639	0.5390	0.3041	-1.2584	-0.2815	0.3146	0.5182	0.9975
Bull(10, $\widehat{POR3}$, short linked)	0.1239	0.7263	0.1706	-2.2075	-0.3304	0.2980	0.6923	1.0124
Bull(10, $\widehat{POR3}$, long linked)	0.1828	0.5632	0.3245	-1.3284	-0.2108	0.2492	0.5067	1.0601
Bear(10, $\widehat{POR3}$)	-0.0347	0.7442	-0.0467	-2.2581	-0.4828	0.0076	0.3806	1.2310
Bear(10, $\widehat{POR3}$, short linked)	0.0608	0.7982	0.0761	-1.9784	-0.4328	0.1435	0.7014	1.5401
Bear(10, $\widehat{POR3}$, long linked)	-0.1644	0.6986	-0.2353	-2.2073	-0.6371	-0.0850	0.3179	1.2505
GS(10)	0.1399	0.7493	0.1867	-1.5836	-0.3116	0.1813	0.5027	1.7669
Ranking(10, $\tilde{\mathbf{w}}$)	0.1287	0.6824	0.1886	-1.7348	-0.1977	0.1787	0.6465	1.2083
Ranking(10, $\tilde{\mathbf{w}}$, short linked)	0.1431	0.6679	0.2142	-1.2206	-0.3330	0.1821	0.7049	1.2236
Ranking(10, $\tilde{\mathbf{w}}$, long linked)	0.1222	0.7402	0.1651	-1.8107	-0.1966	0.1239	0.6115	1.8249

Table 7.4: Option portfolio returns for different strategies with at most $\mathbf{k} = \mathbf{10}$ long and $\mathbf{k} = \mathbf{10}$ short option positions. Monthly portfolio formations take place over a period of 36 months from 19 December 2003 until 17 November 2006. The table reports mean, standard deviation, Sharpe ratio (SR) and quartiles of the monthly zero-cost, equally weighted portfolio return time series.

	mean	std	SR	min	Q1	Q2	Q3	max
Bull(20, $\widehat{POR1}$)	-0.0004	0.7933	-0.0005	-1.8986	-0.3902	0.2761	0.4882	0.9841
Bull(20, $\widehat{POR1}$, short linked)	0.0911	0.6405	0.1422	-1.5303	-0.3648	0.2719	0.5803	0.9805
Bull(20, $\widehat{POR1}$, long linked)	0.0408	0.8127	0.0501	-1.8255	-0.3887	0.3136	0.5854	1.0781
Bear(20, $\widehat{POR1}$)	-0.0469	0.6133	-0.0764	-1.1500	-0.5345	-0.1090	0.5050	1.0374
Bear(20, $\widehat{POR1}$, short linked)	-0.0569	0.6643	-0.0857	-1.2655	-0.5035	-0.0286	0.4570	1.2061
Bear(20, $\widehat{POR1}$, long linked)	-0.0649	0.6088	-0.1067	-1.1620	-0.5351	-0.1936	0.4972	0.9175
Bull(20, $\widehat{POR2}$)	0.0499	0.7723	0.0647	-1.9288	-0.3307	0.3365	0.5511	0.9614
Bull(20, $\widehat{POR2}$, short linked)	0.0782	0.6580	0.1189	-1.5361	-0.1929	0.2413	0.5351	0.9554
Bull(20, $\widehat{POR2}$, long linked)	0.0948	0.7741	0.1225	-1.9330	-0.2791	0.3731	0.6461	0.9553
Bear(20, $\widehat{POR2}$)	-0.0985	0.5939	-0.1658	-1.1647	-0.4910	-0.1134	0.4682	1.2043
Bear(20, $\widehat{POR2}$, short linked)	-0.1260	0.6187	-0.2036	-1.2818	-0.4164	-0.1297	0.4056	1.2093
Bear(20, $\widehat{POR2}$, long linked)	-0.1150	0.6233	2.0000	-1.2457	-0.5309	-0.1321	0.3931	1.0454
Bull(20, $\widehat{POR3}$)	0.1252	0.5303	0.2361	-1.2194	-0.2524	0.2005	0.4924	1.0325
Bull(20, $\widehat{POR3}$, short linked)	0.1472	0.6162	0.2389	-1.4799	-0.2240	0.2858	0.5868	0.9769
Bull(20, $\widehat{POR3}$, long linked)	0.1280	0.5082	0.2518	-1.2283	-0.1601	0.1562	0.4894	0.9194
Bear(20, $\widehat{POR3}$)	-0.1025	0.6463	-0.1587	-1.7524	-0.5367	-0.0862	0.4520	0.9147
Bear(20, $\widehat{POR3}$, short linked)	-0.0381	0.6415	-0.0593	-1.3861	-0.3688	-0.0228	0.5416	0.8837
Bear(20, $\widehat{POR3}$, long linked)	-0.1225	0.6547	-0.1871	-1.6884	-0.5960	-0.0943	0.4356	1.2084
GS(20)	0.0850	0.6307	0.1348	-1.4979	-0.3667	0.1686	0.4928	1.6114
Ranking(20, $\tilde{\mathbf{w}}$)	0.0583	0.6855	0.0851	-1.6359	-0.3137	0.0740	0.6103	1.2022
Ranking(20, $\tilde{\mathbf{w}}$, short linked)	0.1168	0.5834	0.2003	-0.8186	-0.3489	0.0854	0.6253	1.2144
Ranking(20, $\tilde{\mathbf{w}}$, long linked)	0.0722	0.7258	0.0994	-1.7061	-0.3426	0.0696	0.5860	1.5014

Table 7.5: Option portfolio returns for different strategies with at most $\mathbf{k} = \mathbf{20}$ long and $\mathbf{k} = \mathbf{20}$ short option positions. Monthly portfolio formations take place over a period of 36 months from 19 December 2003 until 17 November 2006. The table reports mean, standard deviation, Sharpe ratio (SR) and quartiles of the monthly zero-cost, equally weighted portfolio return time series.

slowly for the bullish than for the bearish strategy. All this indicates that the upper tail of the cross-sectional option return distribution is predicted more accurately than the lower tail.

A χ^2 test for goodness of fit rejects the null hypotheses that each of $\{\widehat{POR1}, \widehat{POR2}, \widehat{POR3}\}$ and $\delta OP_t/OP_t$ have the same distribution: the p-value is in all cases 0. Even the null hypotheses of zero median for $\widehat{POR1} - (\delta OP_t/OP_t)$ and $\widehat{POR3} - (\delta OP_t/OP_t)$ are rejected by a two-sided Wilcoxon signed-rank test at the 1% significance level. It can not be rejected only for $\widehat{POR2} - (\delta OP_t/OP_t)$. Truncating $\widehat{POR1}$ and $\widehat{POR2}$ values less than -100% does not solve the issue. Indeed, it would decrease mean portfolio returns for unlinked and long linked bullish and bearish strategies, but not for short linked strategies.

It seems that an unbounded support helps identify profitable short investment opportunities. Even though $\{\widehat{POR1}, \widehat{POR2}, \widehat{POR3}\}$ fail to mimic the distribution of $\delta OP_t/OP_t$, they are valid trading signals. The ratio of $\widehat{POR1}$ for which $\text{sign}(\widehat{POR1}) = \text{sign}(\delta OP_t/OP_t)$ holds is 50.23%, for $\widehat{POR2}$ 57.75%, and for $\widehat{POR3}$ even 64.25%. Table 7.6 summarizes different measures of concordance between observed and estimated option returns.

Our GS inspired strategy performs well: no extreme portfolio losses like those of the 2002 to 2003 period are observed. Ranking(k, $\tilde{\mathbf{w}}$, short linked) improves GS(k) for all k in terms of average monthly portfolio return. Ranking(5, $\tilde{\mathbf{w}}$, short linked) and GS(5) have 138 out of total 360 option positions in common (36 subsamples, 5 long and 5 short positions). They have 290 out of 720 in common for $k = 10$ and 684 out of 1,440 for $k = 20$. Table 7.7 shows that the chosen weights have an influence on the performance of Ranking(5, $\mathbf{w}$, short linked).

The average portfolio return benefits from $w_4 > 1$ because the last sorting criterion is applied in a pure bullish way, i.e. only calls (puts) are ranked in order to find long (short) option positions. The three other criteria also try to include long put or short call positions if predicted option returns indicate an eligible opportunity. Too many positions of the bullish portfolio are replaced when $w_4 = 1$. Figure 7.1 plots average monthly portfolio returns of Ranking(5, $\mathbf{w}$, short linked) for varying $\mathbf{w}$. Standard weights $\tilde{\mathbf{w}} = (1, 1, 1, 4)$ seem to do a

	$r_{t,T}$	$r_{t,t+\delta t}$	$\delta OP_t/OP_t$	$\widehat{POR1}$	$\widehat{POR2}$	$\widehat{POR3}$
$r_{t,T}$	**1.0000**	0.4373	0.4270	-0.0009	0.0512	0.1516
$r_{t,t+\delta t}$	0.7489	**1.0000**	0.9281	0.0275	0.0749	0.1465
$\delta OP_t/OP_t$	0.7408	0.9403	**1.0000**	0.0433	0.0710	0.1147
$\widehat{POR1}$	0.4983	0.4918	0.5023	**1.0000**	0.6558	0.2725
$\widehat{POR2}$	0.5893	0.5925	0.5775	0.6767	**1.0000**	0.4744
$\widehat{POR3}$	0.6723	0.6745	0.6425	0.5050	0.7858	**1.0000**

Table 7.6: Measures of concordance between observed and estimated returns of long option positions, aggregated from all available investment instruments in the 36 sub-samples used in Tables 7.3, 7.4, and 7.5. The percentage of option returns (as specified by row and column headers) with corresponding signs is displayed in the *lower triangular part* of the matrix. The *diagonal* represents the percentage of appearance in the sample, where 100% means that there are no missing entries. The *upper triangular part* of the matrix contains the sample correlation in terms of Kendall's tau between option returns as specified by row and column headers. t denotes the point in time when portfolios are formed, T when the options expiry and δt equals 10 trading days. Hold-to-expiration option returns over the period from t until T, $r_{t,T}$, are defined in Equation (7.1). Similarly, $r_{t,t+\delta t}$ is calculated using time $t + \delta t$ bid prices and time t ask prices. Mid option prices (7.5) are used for obtaining $\delta OP_t/OP_t = (OP_{t+\delta t} - OP_t)/OP_t$. Predicted option returns $\widehat{POR1}, \widehat{POR2}, \widehat{POR3}$ are defined by Equations (7.9), (7.10), and (7.11).

$\mathbf{w} =$	(1,1,1,1)	(1,1,1,2)	(1,0,0,4)	(0,1,0,4)	(0,0,1,4)	(1,0,1,4)
avg return	0.0586	0.2025	0.2180	0.1856	0.2292	0.2359

Table 7.7: Performance of the Ranking(5, $\mathbf{w}$, short linked) strategy in dependence of $\mathbf{w}$ over a period of 36 months from 19 December 2003 until 17 November 2006. Performance is measured as time-series average of portfolio returns generated by the strategy. Weights $\mathbf{w} = (w_1, w_2, w_3, w_4)$ are assigned to the four corresponding sortings based on $\widehat{POR1}$, $\widehat{POR2}$, $\widehat{POR3}$ and log(HV/IV). Five long option positions are chosen with the help of a weighted average ranking procedure. The short positions are automatically chosen to be options of opposite type on the same underlyings as for the long positions (short linking).

reasonable job, although using $\mathbf{w} = (1, 1, 1, 5)$ or $(1, 0, 0, 3)$ would result in higher average portfolio returns. This ex-post analysis is of course highly dependent on the chosen sample period. Given the high ratio of correctly predicted directions of option price changes of $\widehat{POR3}$ (see Table 7.6), it might be beneficial to increase the component w_3 in the Ranking strategy. Figure 7.2 reveals that this is only the case for $1 < w_3 < 4$ or $w_3 \geq 77$.

7.3.2 Sensitivity analysis

The more accurate the forecasts of $\hat{S}_{t+\delta t}$, the better the predictions of $\widehat{POR1}$, $\widehat{POR2}$, $\widehat{POR3}$ and the bigger the average portfolio returns. Table 7.8 reports the average portfolio returns when we predict option returns under perfect foresight of the underlying stock prices 10 days OS such that $\hat{S}_{t+\delta t} = S_{t+\delta t}$. This consequently simplifies the OS option tracking problem as

$$\begin{aligned}\left|\hat{F}_{\hat{M}}(m_{t+\delta t}, \tau_{t+\delta t}, cp\ flag,\ \widehat{factors}_{t+\delta t}) - \sigma^{\mathrm{IV}}_{t+\delta t}\right|& \\ < \left|\hat{F}_{\hat{M}}(\tilde{m}, \tau_{t+\delta t}, cp\ flag, \widehat{factors}_{t+\delta t}) - \sigma^{\mathrm{IV}}_{t+\delta t}\right|&\end{aligned}$$

for $\tilde{m} \neq K/S_{t+\delta t}$. Notice that the closing price of the underlying stock is used as one of the exogenous factors. Hence, whenever one of the time-lagged/leading

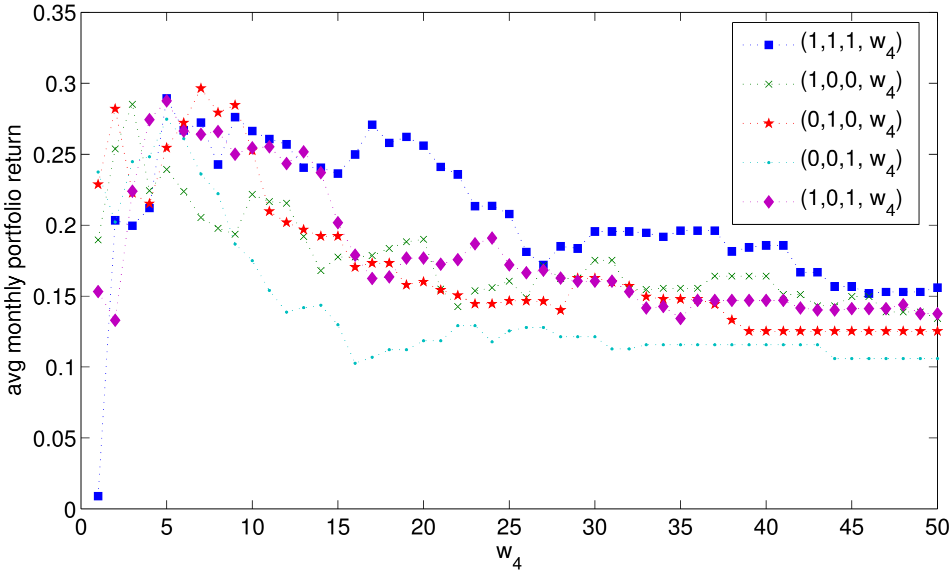


Figure 7.1: Performance of the Ranking(5, **w**, short linked) strategy in dependence of **w** over a period of 36 months from 19 December 2003 until 17 November 2006. Performance is measured as time-series average of portfolio returns generated by the strategy. The performance of the strategy is plotted for varying **w**.

$S_{t\pm i}$, $i = 1, \dots, 5$ is a split variable in the regtree-treefgd model $\hat{F}_{\hat{M}}(\cdot)$, its OS forecast is exactly $S_{t+\delta t-i}$ for a time-lagged version of $\hat{S}_{t+\delta t}$ but different from $S_{t+\delta t+i}$ for a time-leading version. The settlement price S_T is besides unknown at time t. Such a comparison is given in Table 7.8; it allows us to identify the potential of the proposed strategies.

Both the bullish and the bearish strategy have huge potential returns of up to 147% per month under perfect foresight of the underlying stock price up to time $t + \delta t$. Only a small extra performance gain would be realized if $\hat{\sigma}^{\mathrm{IV}}_{t+\delta t} = \sigma^{\mathrm{IV}}_{t+\delta t}$ could also be perfectly predicted because short-dated options have relatively more gamma than vega compared to long-dated options. The average recorded

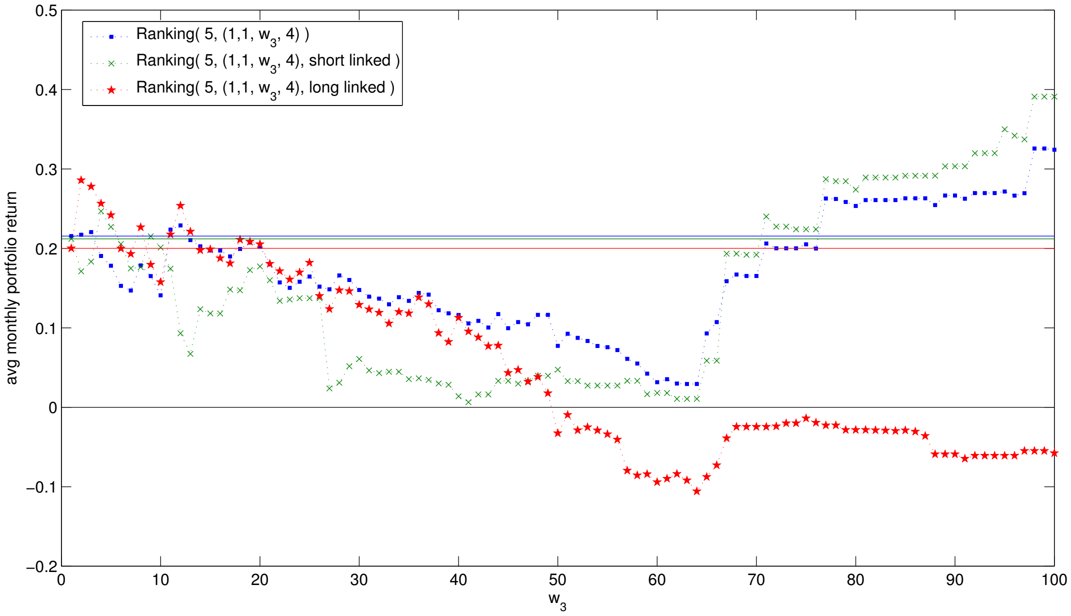


Figure 7.2: Performance of the Ranking strategy with $\mathbf{w} = (1, 1, w_3, 4)$ in dependence of w_3 over a period of 36 months from 19 December 2003 until 17 November 2006. Performance is measured as time-series average of portfolio returns generated by the strategy. The performance of the strategy is plotted for w_3 varying from 1 to 100. The solid {blue, green, red} vertical line represent the performance of the {not linked, short linked, and long linked} Ranking strategy with $k = 5$ as in Table 7.3, where $\tilde{\mathbf{w}} = (1, 1, 1, 4)$.

	$k=5$		$k=10$		$k=20$	
Bull(k, $\widehat{POR1}$)	-0.0401	1.2364	-0.0706	0.9339	-0.0004	0.7768
Bull(k, $\widehat{POR1}$, short linked)	0.0791	1.3803	0.1723	1.0219	0.0911	0.8400
Bull(k, $\widehat{POR1}$, long linked)	0.0458	0.9787	-0.0781	0.8636	0.0408	0.7407
Bear(k, $\widehat{POR1}$)	0.2098	1.2574	0.0344	0.9068	-0.0469	0.6360
Bear(k, $\widehat{POR1}$, short linked)	0.1678	1.1905	0.0472	0.8610	-0.0569	0.5934
Bear(k, $\widehat{POR1}$, long linked)	0.1309	1.2008	-0.0223	0.9735	-0.0649	0.6558
Bull(k, $\widehat{POR2}$)	-0.0802	1.0887	0.0330	0.9567	0.0499	0.7762
Bull(k, $\widehat{POR2}$, short linked)	0.1148	0.8898	0.0952	0.8361	0.0782	0.7665
Bull(k, $\widehat{POR2}$, long linked)	-0.0561	0.9890	0.0464	0.9249	0.0948	0.7618
Bear(k, $\widehat{POR2}$)	-0.1360	1.1823	-0.1486	0.8861	-0.0985	0.6020
Bear(k, $\widehat{POR2}$, short linked)	-0.0333	1.1252	-0.1264	0.8354	-0.1260	0.5459
Bear(k, $\widehat{POR2}$, long linked)	-0.1326	0.9091	-0.1425	0.7727	-0.1150	0.5318
Bull(k, $\widehat{POR3}$)	0.2868	1.4692	0.1639	1.1562	0.1252	0.8357
Bull(k, $\widehat{POR3}$, short linked)	0.1331	1.4637	0.1239	1.2207	0.1472	0.9082
Bull(k, $\widehat{POR3}$, long linked)	0.2663	0.6490	0.1828	0.6177	0.1280	0.5145
Bear(k, $\widehat{POR3}$)	0.1289	1.2899	-0.0347	0.9791	-0.1025	0.6599
Bear(k, $\widehat{POR3}$, short linked)	0.1120	1.4790	0.0608	1.0934	-0.0381	0.7049
Bear(k, $\widehat{POR3}$, long linked)	-0.1515	0.3610	-0.1644	0.4182	-0.1225	0.4617
GS(k)	0.1872	0.1872	0.1399	0.1399	0.0850	0.0850
Ranking(k, $\tilde{\mathbf{w}}$)	0.2155	0.8503	0.1287	0.7829	0.0583	0.5958
Ranking(k, $\tilde{\mathbf{w}}$, short linked)	0.2120	0.8918	0.1431	0.8393	0.1168	0.6534
Ranking(k, $\tilde{\mathbf{w}}$, long linked)	0.2001	0.8313	0.1222	0.7207	0.0722	0.5468

Table 7.8: Option portfolio returns under the **assumption of perfect OS foresight of the underlying stock prices** for $\delta t = 10$ trading days. Portfolio returns over a period of 36 months from 19 December 2003 until 17 November 2006 are calculated. The table reports the time series averages of zero-cost, equally weighted portfolio returns twice for each strategy: when no knowledge of future information is used at all (*left value* in a column) and under perfect foresight of $\hat{S}_{t+\delta t} = S_{t+\delta t}$ (*right value*).

$(\Delta_t, \Gamma_t, \nu_t, \Theta_t)$ for calls in the 36 subsamples are

$$(0.5252, 0.1501, 5.0669, -8.6026)$$

and for puts

$$(-0.4831, 0.1477, 5.0818, -7.4427).$$

The average BS Greeks $(\Delta_t^{\mathrm{BS}}, \Gamma_t^{\mathrm{BS}}, \nu_t^{\mathrm{BS}}, \Theta_t^{\mathrm{BS}})$ calculated for long-dated calls with $(S_t, \sigma_t^{\mathrm{IV}}, \textit{cp flag}, K, T + 365 \text{ days}, \hat{r}, \hat{q})$ are

$$(0.5790, 0.0424, 17.6851, -2.5862)$$

and for puts

$$(-0.3696, 0.0418, 17.7615, -1.1831).$$

The average relative contributions of the Greeks to option price changes δOP_t in terms of mid option prices OP_t are

$$\left(\Delta_t \delta S_t, \frac{1}{2}\Gamma_t \delta S_t^2, \nu_t \delta\sigma_t^{\mathrm{IV}}, \Theta_t \delta t\right) \Big/ OP_t = (13.09\%, 35.63\%, 2.17\%, -29.45\%)$$

for short-dated calls and

$$(-12.89\%, 29.25\%, 3.59\%, -25.74\%)$$

for short-dated puts versus

$$(17.24\%, 6.65\%, 6.48\%, -9.08\%)$$

for long-dated calls and

$$(-11.45\%, 7.49\%, 14.68\%, -4.93\%)$$

for long-dated puts. Therefore, improving the accuracy of $\hat{S}_{t+\delta t}$ would definitely be more worthwhile than minimizing

$$\mathrm{MSE}(\widehat{\delta\sigma_t^{\mathrm{IV}}}, \delta\sigma_t^{\mathrm{IV}}) = \mathrm{MSE}(\hat{\sigma}_{t+\delta t}^{\mathrm{IV}}, \sigma_{t+\delta t}^{\mathrm{IV}}).$$

The ratio of correctly predicted direction of IV changes, for which $\mathrm{sign}(\widehat{\delta\sigma_t^{\mathrm{IV}}}) = \mathrm{sign}(\delta\sigma_t^{\mathrm{IV}})$, is relevant for sortings based on predicted option returns.

7.3.3 Risk measures

The proposed strategies have an average monthly return of up to 28.68% over the 2004 to 2006 period, expressed in terms of portfolio gross exposure. Theoretically, no costs are incurred to set up long-short option portfolios, but an initial margin deposit is required. The maintenance requirement must be very high because standard deviations of the monthly portfolio return time series soar up to 132.17% for the different strategies. Given the performances shown before, a closer look is only taken at the risks involved in the Bull(5, $\widehat{POR3}$), GS(5) and Ranking(5, $\tilde{\mathbf{w}}$, short linked) strategies over an *extended* period of 1,610 days from 19 July 2002 until 15 December 2006. The first half of 2002 is used for the initial fit of the regtree-treefgd model. Long-short option portfolios for the additional 17 monthly subsamples are formed in the same way as described in Section 7.2.

Assume that an investor has $V_0 = \$100,000$ on a bank account that pays 1% p.a. risk-free interest. At each of the 53 trading dates, a zero-cost, equally weighted long-short portfolio is formed. The portfolio's gross exposure is constrained to 20% of the bank account balance at each trading date. That is also the amount of money that the broker demands as initial margin. This means that 80% of total wealth V_t remain on the bank account at the beginning of each month and that losses of up to 500% of the risky gross exposure can be covered with the initial margin and the remaining money on the bank account at the end of a month. Figure 7.3 shows how the total wealth process V_t evolves over time.

V_t grows from \$100,000 initially to \$325,535.81 (Bull), \$616,582.55 (GS) and \$729,114.60 (Ranking), respectively. Table 7.9 reports performance and risk measures for the returns of the total wealth process V_t. The results are a good illustration of the superior profitability of the GS(5) and Ranking(5, $\tilde{\mathbf{w}}$, short linked) strategies over the simpler Bull(5, $\widehat{POR3}$) strategy that has difficulties to recover from an early loss of 283.79% (17 October 2003). This loss is cushioned by the 20% risky option / 80% risk-free bank account investment plan and "only" results in a monthly loss of 56.69%, but the recovery from the maximum drawdown requires 15 months.

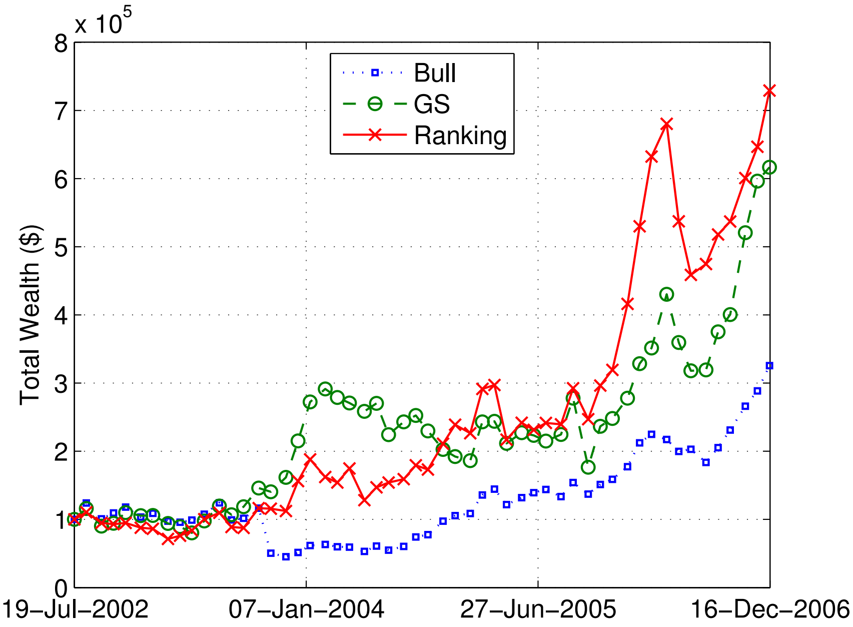


Figure 7.3: Evolution of the total wealth process V_t. Plot of total wealth against time when investing \$100,000 according to Bull(5, $\widehat{POR3}$), GS(5) and Ranking(5, $\tilde{\mathbf{w}}$, short linked) under the condition that the portfolio's gross exposure at each monthly trading date is limited to 20% of the bank account balance.

Table 7.9: Performance and risk measures for the returns of the total wealth process V_t over a period of 1,610 days from 19 July 2002 until 15 December 2006. An investor starts with $V_0 = \$100,000$. At each of the 53 monthly trading dates, she invests 20% of total wealth V_t according to the option strategy and keeps 80% for maintenance requirement on the bank account, which pays 1% p.a. risk-free interest. The table reports the number of monthly gains and losses, biggest gains and losses (in absolute terms), ex-post value-at-risk and expected shortfall (one month, 95%), maximum drawdown over the entire investment period and number of months to recover from it, cumulated return, annualized return and standard deviation and Sharpe ratio.

	Bull	GS	Ranking
# of monthly gains	36	32	35
# of monthly losses	17	21	18
Biggest gain	25.70%	33.94%	50.40%
Biggest loss	56.69%	36.42%	30.92%
VaR(0.95, 1 month)	18.68%	16.98%	22.90%
ES(0.95 1 month)	31.86%	25.23%	27.15%
Max drawdown	63.52%	39.37%	34.75%
# of recovery periods	15	4	4
Cumulated return	225.54%	516.58%	667.26%
Annualized return	30.68%	51.04%	58.71%
Annualized std	55.77%	52.96%	52.98%
Sharpe ratio	0.5502	0.9638	1.1082

Monthly gain $= (V_{t+1} - V_t)/V_t$
Monthly loss $= -(V_{t+1} - V_t)/V_t$
VaR(0.95, 1 month) 95% quantile of monthly losses
ES(0.95 1 month) average of monthly losses above VaR(0.95, 1 month)
drawdown_i $= \max\left(0, 1 - \frac{V_{\text{start}+i}}{\max_{j=1,\ldots,i}(V_{\text{start}+j})}\right)$
Max drawdown $= \max_i(\text{drawdown}_i)$
Cumulated return $= (V_{\text{end}} - V_{\text{start}})/V_{\text{start}}$
Annualized return $= r$, solves $V_{\text{start}}(1+r)^{t_y} = V_{\text{end}}$ with $t_y = (t_{\text{end}} - t_{\text{start}})$ in years
Annualized std sample standard deviation scaled by the square root of time rule
Sharpe ratio (SR) annualized return / annualized std

Chapter 8

Conclusions

A new approach to model and forecast the implied volatility surfaces has been proposed in this thesis. The methodology is based on a starting model that is improved by semi-parametric additive expansions of regression trees. A modified version of classical boosting procedures can handle very high dimensional predictor variable sets. Consequently, there is no need for variance reduction or other excluding data techniques to fit the model to real data, avoiding the possibility of a dangerous information loss. Focussing on out-of-sample predictions of the IVS, the statistical learning framework substantially reduces the sum of squared residuals, i.e. the squared difference between observed and estimated IVs, for a variety of possible starting models including the (rule of thumb) sticky moneyness model (Derman and Kani, 1998; Daglish et al., 2007), the ad-hoc BS model of Dumas et al. (1998) with deterministic volatility function, a high-dimensional Bayesian vector autoregression model and the dynamic semiparametric factor model of Fengler et al. (2007).

The predictive potential was tested on a huge data set of S&P 500 options collecting strong empirical evidence that the proposed methodology improves the performance of any reasonable starting model in forecasting short- and middle-term future implied volatilities (up to 60 days), and also under possible structural breaks in the time series. Similar results were also obtained when fitting the models to more volatile stock option data. The regtree-treefgd model is completely based on regression trees, i.e. regression tree as starting model and regression trees as base learners. It turned out to be the best performing model and a powerful tool in forecasting IVS dynamics.

In the final application, several trading strategies were proposed for 1 month ATM options based on predicted option returns over $\delta t = 10$ trading days. A predicted increase in option price is assumed to coincide with an increase in intrinsic value as the time value close to expiry converges to zero, and a positive correlation between predicted option returns $\hat{r}_{t,t+\delta t}$ and observed hold-to-expiration returns $r_{t,T}$ was indeed found. The option trading strategies generated high positive average monthly returns, unfortunately at the cost of high volatility. The distribution of $\hat{r}_{t,t+\delta t}$ poorly fitted that of $r_{t,t+\delta t}$ mainly in the lower tail. Short linking, i.e. shorting options of opposite type on the same underlyings as the long positions, circumvented this problem as the upper tail was reasonably fitted. In particular, bullish strategies with long call and short put option positions have profited from this because all positions had positive delta and the long calls also had positive gamma, which adjusted the delta in the right way for up or down moves in the underlying stock price.

Predicted option returns have turned out to be valuable trading signals. The influence of better forecasts of the underlying stock price $\hat{S}_{t+\delta t}$ on the average option portfolio return was analyzed for all strategies. The information contained in historical stock prices up to time t had limited influence on predicting $\hat{r}_{t,t+\delta t}$; even a filtered historical simulation generated OS forecasts that were prone to errors. A different approach to improve the quality of $\hat{r}_{t,t+\delta t}$ as a trading signal was used. First, the regtree-treefgd model managed to increase the ratio of correctly predicted directions of implied volatility changes by squeezing as much information as possible from the whole implied volatility surface and a set of exogenous factors that included the underlying stock price as well as alternative IV models. Second, three ways to estimate $\hat{r}_{t,t+\delta t}$ were defined, two of them allowing returns $< -100\%$ for long option positions. That feature turned out to be beneficial for the ranking strategy, as it replaced a few option positions that were originally assigned by the GS inspired strategy with better alternatives.

Nevertheless, predicted option returns only based on $\hat{S}_{t+\delta t}$ ($\widehat{POR3}$) seemed to outperform more sophisticated predicted option return models based on IVS forecasts ($\widehat{POR1}$, $\widehat{POR2}$). Up to 60% of the option positions of a simple option trading strategy were replaced by the more complex models. Further empirical analysis is needed to prove the claimed robustness of these methods with respect to the chosen sample period. Although the relative Greek contribution induced by the IV change to the change of option prices over a short period of $\delta t = 10$ trading days was approximatively four times smaller than the one of the underlying stock price, it is more likely that the accuracy of $\widehat{\delta\sigma_t^{\mathrm{IV}}} = \hat{\sigma}_{t+\delta t}^{\mathrm{IV}} - \sigma_t^{\mathrm{IV}}$ and not of $\widehat{\delta S_t} = \hat{S}_{t+\delta t} - S_t$ can be improved in the future. Advanced option tracking strategies with less sensitivity of $\hat{\sigma}_{t+\delta t}^{\mathrm{IV}}$ with respect to $\hat{S}_{t+\delta t}$ are currently being developed.

Finally, a possible implementation of the proposed option trading strategies was shown from an investor's point of view. A monthly loss of more than 100% would have put the investor out of business if no additional funds had been available. Hence, the gross exposure of the long-short option portfolio was limited to 20% of the invested capital, which left 80% of the capital for maintenance requirements. Backtesting three strategies from 19 July 2002 through 15 December 2006, average annualized returns up to 58.72% were obtained with an annualized volatility of at most 55.77%.

Appendices

Appendix A

History of options

Options are the main objects in this thesis. Hull describes them in his popular book *Options, Futures and Other Derivatives* as follows:

> *A call option is the right to buy an asset for a certain price; a put option is the right to sell an asset for a certain price. A European option can be exercised only at the end of its life; an American option can be exercised at any time during its life. There are four types of option positions: a long position in a call, a long position in a put, a short position in a call, and a short position in a put.... Options are fundamentally different from the forward, futures, and swap contracts discussed in the last few chapters. An option gives the holder of the option the right to do something. The holder does not have to exercise this right. By contrast, in a forward, futures, or swap contract, the two parties have committed themselves to some action. It costs a trader nothing (except for the margin requirements) to enter into a forward or futures contract, whereas the purchase of an option requires an up-front payment* (2002, p. 151).

The general concept of an option, having the right but not the obligation to do something in the future at a predetermined price, has been around for a long time.

> *"In book 1, Chapter 11 of Politics, Aristotle tells the story of Thales of Miletus (624-547 BC), one of the seven sages of the ancient world. People had been telling Thales that his philosophy was useless, since it had left him a poor man. 'But he, deducing from his knowledge of*

> *stars that there would be a good crop of olives, while it was still winter raised a little capital and used it to pay deposits on all the oil-presses in Miletus and Chios, thus securing an option on their hire. This cost him only a small sum as there were no other bidders. Then the time of the harvest came and as there was a sudden and simultaneous demand for oil-presses, he hired them out at any price he liked to ask. He made a lot of money and so demonstrated that it is easy for philosophers to be rich, if they want to; but that is not their object in life. Such is the story of Thales, how he gave proof of his cleverness but, as we have said, the principle can be generally applied; the way to make money in business is to get, if you can, a monopoly for yourself. Hence we find governments also on certain occasions employing this method when they are short of money. They secure a sales monopoly for themselves'."* (Makropoulou and Markellos, 2005).

Option contracts were originally sold 'over the counter', i.e. not standardized in terms or conditions, tailored to specialized people or institutions. Although Murphy (2009) shows that options were actively traded in late 17th century London, the birth of modern financial options market took place during 1973 in the United States of America. The Chicago Board of Trade (CBOT) opened the Chicago Board Options Exchange (CBOE) and started trading contracts standardized in terms and conditions. Quoted prices of 'exchange traded option contracts' were published and a market maker system was set up such that options could be traded on a secondary or resell market. The development of exchange-traded option markets over the last few decades is stunning. On the first trading day, 26 April 1973, 911 contracts were traded[33]. Nowadays, CBOE is the largest U.S. options exchange. By the end of 2008, it had an annual trading volume of about 1.2 billion contracts, corresponding to a traded amount of USD 970 billion[34].

---

[33] See `http://www.cboe.com/aboutcboe/History.aspx`

[34] See `http://www.cboe.com/data/marketstats-2008.pdf`

Appendix B

Asset pricing and contingent claims

The three basic forces in asset pricing theory are arbitrage, optimality and equilibrium.

> *The most important unifying principle is that any of these three conditions* [absence of arbitrage, single-agent optimality, market equilibrium] *implies that there are 'state prices', meaning positive discount factors, one for each state and date, such that the price of any security is merely the state-price weighted sum of its future payoffs"* (Duffie, 2001, p. xiii).

Dybvig and Ross provide a good introduction to single-period portfolio choice problems in complete markets and depict the equivalence of different pricing approaches in their Pricing Rule Representation Theorem, *"which asserts that a positive linear pricing rule can be represented as using state prices, risk-neutral expectations, or a state-price density"* (2003, p. 607).

As explained by Pliska, a contingent claim is a random variable that represents the time T payoff from a seller to a buyer (1997, p. 112). The payoff ψ of a contingent claim depends on the unknown future state of an underlying asset price process S, therefore $\psi_T = \psi_T(S_T)$. To obtain its fair price, a pricing measure that values all possible future payoffs as defined in the contract specification is needed.

A pricing kernel $M_{t,T}$ (a.k.a. stochastic discount factor a.k.a. state price density per unit probability) combines the probability distribution of future states of the underlying asset with assumed investor preferences for payoffs in these states. Furthermore, it provides the link between the physical measure $\mathbb{P}$ and the risk-neutral measure $\mathbb{Q}$. The former describes the distribution of S_t as originally defined by the assumed probability space, the latter relates to a hypothetical market where investors are risk-neutral. Based on the results of Harrison and Kreps (1979) and assuming a constant continuously compounded risk-free rate r, the price π_t at time $t < T$ of a contingent claim with payoff ψ_T at time T is given by

$$\begin{aligned}\pi_t(\psi_T) &= \mathrm{E}_{\mathbb{P}}\left[\psi_T M_{t,T} | \mathcal{F}_t\right] = \int_0^\infty \psi_T(S_T) M_{t,T}(S_T) p_{t,T}(S_T)\ dS_T \\ &= \mathrm{E}_{\mathbb{Q}}\left[\psi_T e^{-r(T-t)} \middle| \mathcal{F}_t\right] = e^{-r(T-t)} \int_0^\infty \psi_T(S_T) q_{t,T}(S_T)\ dS_T\end{aligned}$$

where $\mathcal{F}_t$ is the filtration on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ representing the set of available information generated by the stochastic process S up to time t, $p_{t,T}$ is the probability density function (PDF) under $\mathbb{P}$ and $q_{t,T}$ the PDF under $\mathbb{Q}$.

Appendix C

Volatility

Volatility measures the degree of unpredictable change over time of continuously compounded returns of a financial instrument. A discrete set of prices S_t is observable over a period $t \in [0, T]$, but volatility is not directly observable. It refers to the standard deviation of $R_i := \frac{1}{t_i - t_{i-1}} \log\left(\frac{S_{t_i}}{S_{t_{i-1}}}\right)$, $i \in \{1, 2, \ldots, n\}$, $0 \leq t_0 < t_1 < \ldots < t_n \leq T$. The standard deviation is a dispersion measure of the probability distribution of the R_is. Whenever volatility or an estimator of it is defined, some distributional properties have to be assumed.

Example C.1 Let us assume that $t_i - t_{i-1} = 1$ unit of time (day, week) for all $i \in \{1, 2, \ldots, n\}$ and that R_i are independent and identically distributed (iid) with finite first and second moment. In this case, the sample variance

$$s^2 := \frac{1}{n-1} \sum_{i=1}^{n} (R_i - \bar{R})^2 \tag{C.1}$$

$$\bar{R} := \frac{1}{n} \sum_{i=1}^{n} R_i \tag{C.2}$$

is a an unbiased and consistent estimator of $\sigma^2 = \mathrm{Var}(R_i) < \infty$. However, s is in general *not* an unbiased estimator of the volatility since $\mathrm{E}[s] = \mathrm{E}[\sqrt{s^2}] \leq \sqrt{\mathrm{E}[s^2]} = \sigma$ by Jensen's inequality. If $R_i \overset{d}{\sim} \mathcal{N}(\mu, \sigma^2)$, then

$$R_{[t_0,t_n]} := \frac{1}{n} \log\left(\frac{S_{t_n}}{S_{t_0}}\right) \equiv \frac{1}{\sum_{i=1}^{n}(t_i - t_{i-1})} \sum_{i=1}^{n} \underbrace{(t_i - t_{i-1})}_{=1} R_i = \bar{R} \tag{C.3}$$

is $\overset{d}{\sim} \mathcal{N}(\mu, \frac{\sigma^2}{n})$ and $(n-1)\frac{s^2}{\sigma^2} \overset{d}{\sim} \chi^2_{n-1}$ by Cochran's theorem, which would allow us to construct an unbiased estimator of σ.

Note C.2 Due to its importance in finance, the volatility literature is tremendous. For a good start, Abken and Nandi (1996) provide an overview of volatility concepts and models used in option pricing. Bates (1996a) discusses the commonly used methods for testing option pricing models, Dumas et al. (1998) empirically test local volatility models. Fengler (2005, Section 3.13) summarizes the relations amongst the concepts of instantaneous, local and implied volatility. Figlewski (1997); Poon and Granger (2003) review volatility forecasts in financial markets.

C.1 Instantaneous volatility

Let us assume that the stock price S_t is modelled in continuous time over the interval $[0, T]$ as a continous Itô process on a filtered probability space, i.e. S_t is an adapted stochastic process which can be expressed as the sum of an integral w.r.t. time and a stochastic integral wrt Brownian motion W_t,

$$S_t = S_0 + \int_0^t a(\tilde{t}, S_{\tilde{t}})d\tilde{t} + \int_0^t b(\tilde{t}, S_{\tilde{t}})dW_{\tilde{t}}, \tag{C.4}$$

such that

$$\int_0^t |a(\tilde{t}, S_{\tilde{t}})| + b(\tilde{t}, S_{\tilde{t}})^2 d\tilde{t} < \infty \tag{C.5}$$

for each $t \in [0, T]$.

Note C.3 A diffusion is mathematically correctly defined as in Eq. (C.4). Its representation as a stochastic differential equation (SDE)

$$dS_t = a(t, S_t)dt + b(t, S_t)dW_t \tag{C.6}$$

should always be interpreted as short form of the stochastic integral

$$S_t - S_0 = \int_0^t dS_{\tilde{t}} \equiv \int_0^t a(\tilde{t}, S_{\tilde{t}})d\tilde{t} + \int_0^t b(\tilde{t}, S_{\tilde{t}})dW_{\tilde{t}}. \tag{C.7}$$

Definition C.4 (Instantaneous volatility) *The percentage change in the stock price over an infinitesimally small period* dt *is*

$$\frac{dS_t}{S_t} = \frac{\int_0^{t+dt} dS_{\tilde{t}} - \int_0^t dS_{\tilde{t}}}{S_t} = \frac{a(t,S_t)dt + b(t,S_t)dW_t}{S_t} \tag{C.8}$$

and the square root of its conditional variance at time t *per infinitesimally small period* dt

$$\sqrt{Var\left(\frac{dS_t}{S_t}\bigg|\mathcal{F}_t\right)\Bigg/ dt} \tag{C.9}$$

is called instantaneous or spot volatility.

Note C.5 Suppose $S_t = h(t, W_t)$ is an explicit solution of Eq. (C.4). Itô's lemma yields

$$dS_t = dh(t,W_t) = \underbrace{\left(\frac{\partial h}{\partial t} + \frac{1}{2}\frac{\partial^2 h}{\partial W_t^2}\right)}_{\equiv a(t,S_t)} dt + \underbrace{\frac{\partial h}{\partial W_t}}_{\equiv b(t,S_t)} dW_t.$$

W_t and S_t are $\mathcal{F}_t$-measurable, hence $a(t,S_t) = a$ and $b(t,S_t) = b$ are known given all information up to time t and for small $\delta t > 0$ we have

$$\frac{S_{t+\delta t} - S_t}{S_t}\bigg|\mathcal{F}_t = \frac{h(t+\delta t, W_{t+\delta t}) - h(t,W_t)}{h(t,W_t)}\bigg|\mathcal{F}_t \overset{d}{\sim} \mathcal{N}((a/S_t)\delta t, (b/S_t)^2 \delta t).$$

It follows for $\delta t \to 0$ that the spot volatility is equal to $b(t,S_t)/S_t$.

Remark C.6 Assume

$$a(t,S_t,Z_t) = \tilde{a}(t,S_t,Z_t)S_t \tag{C.10}$$

$$b(t,S_t,Z_t) = \tilde{b}(t,S_t,Z_t)S_t \tag{C.11}$$

where Z_t is a collection of other $\mathcal{F}_t$-adapted state variables such that $\mathbb{P}[S_{t_2} \leq x|\mathcal{F}_{t_1}] = \mathbb{P}[S_{t_2} \leq x|S_{t_1}, Z_{t_1}]$ for $0 \leq t_1 < t_2 \leq T$. If the integrability condition (C.5) holds, it can be shown that $\tilde{a}(t,S_t,Z_t)$ is the instantaneous drift $\mathrm{E}_{\mathbb{P}}[dS_t/S_t|\mathcal{F}_t]/dt$ and $\tilde{b}(t,S_t,Z_t)$ the instantaneous volatility.

Remark C.7 A Taylor expansion of $\log(x)$ in a neighbourhood of $x_0 = 1$ shows that $\log(x) \approx (x-1)$ for $|x - x_0|$ small. Hence for $\delta t \to 0$

$$\log\left(\frac{S_{t+\delta t}}{S_t}\right) \approx \frac{S_{t+\delta t}}{S_t} - 1 = \underbrace{\frac{S_{t+\delta t} - S_t}{S_t}}_{\xrightarrow[\delta t \to 0]{} \frac{dS_t}{S_t}}.$$

Remark C.8 (Realized variance) The availability of high-frequency intraday returns provides a deeper insight into daily return variability. Assuming the stock price process S_t to be an Itô process as in Eq. (C.4) with $b(t, S_t) = \sigma(t, S_t)S_t$, the quadratic variation[35] of $X_t := \log S_t$ is equal to the integrated variance, $\langle X \rangle_t = \int_0^t \sigma(\tilde{t}, S_{\tilde{t}})^2 d\tilde{t}$.

Realized variance (RV) is an ex-post observable proxy for integrated variance. Having M equally spaced intra-day returns over a time interval of one day, Andersen, Bollerslev, Diebold, and Labys (2003); Corsi (2005) define RV as

$$\mathrm{RV}_t^{(d)} := \sum_{j=0}^{M-1} r_{t-j\cdot\Delta}^2 \tag{C.12}$$

with $\Delta = 1$ day$/M$, $r_{t-j\cdot\Delta} = \log(S_{t-j\cdot\Delta}) - \log(S_{t-(j+1)\cdot\Delta})$. It follows that

$$\mathrm{RV}_t^{(d)} \xrightarrow[M\to\infty]{\mathbb{P}} \int_{t-1\text{ day}}^{t} \sigma(\tilde{t}, S_{\tilde{t}})^2 d\tilde{t}. \tag{C.13}$$

Morgenson and Harvey (2002) explain realized volatility $= \sqrt{\mathrm{RV}}$ in the following way:

> *"Sometimes referred to as the historical volatility, this term is usually used in the context of derivatives. While the implied volatility refers to the market's assessment of future volatility, the realized volatility measures what actually happened in the past. The measurement of the volatility depends on the particular situation. For example, one could*

---

[35] Let $\mathcal{P} = \{0 = t_0 < t_1 < ... < t_n = t\}$ be a partition of the interval $[0, t]$. The quadratic variation of a stochastic process X_t is defined as

$$\langle X \rangle_t := \lim_{\substack{||\mathcal{P}|| \to 0 \\ n\to\infty}} \sum_{k=1}^{n} (X_{t_k} - X_{t_{k-1}})^2.$$

> *calculate the realized volatility for the equity market in March of 2003 by taking the standard deviation of the daily returns within that month. One could look at the realized volatility between 10:00AM and 11:00AM on June 23, 2003 by calculating the standard deviation of one minute returns"* (2002).

C.2 Stochastic volatility

A stochastic volatility (SV) model in continuous time typically assumes an asset price dynamcis of the following form (Fengler, 2005, Section 2.8.2):

$$\frac{dS_t}{S_t} = \mu_t \; dt + \sigma(t, Y_t) \; dW_t^{(0)} \tag{C.14}$$

$$\sigma(t, Y_t) = f(Y_t) \tag{C.15}$$

$$dY_t = \alpha(t, Y_t) \; dt + \beta(t, Y_t) \; dW_t^{(1)} \tag{C.16}$$

$$\left\langle dW_t^{(0)}, dW_t^{(1)} \right\rangle = \rho \; dt. \tag{C.17}$$

The two Brownian motions are defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the filtration $\mathbb{F} = \{\mathcal{F}_t, t \in [0, T]\}$ is generated by both of them. Realistically, $W_t^{(0)}$ and $W_t^{(1)}$ would be negatively correlated (Black, 1976), but independence is often assumed for simplicity. *"The function $f(y)$* [is] *chosen for positivity and analytical tractability"* (Fengler, 2005, p. 37). Typical examples of stochastic volatility models are

1. Hull and White (1987): $f(y) = \sqrt{y}$, $dY_t/Y_t = mdt + \xi dW_t^{(1)}$, $\rho = 0$
2. Wiggins (1987): Hull and White model with $\rho \neq 0$
3. Scott (1987): $f(y) = \exp\{y\}$, $dY_t = k(\theta - Y_t)dt + \xi dW_t^{(1)}$, $\rho \neq 0$
4. Stein and Stein (1991): $f(y) = |y|$, $dY_t = k(\theta - Y_t)dt + \xi dW_t^{(1)}$, $\rho = 0$
5. Heston (1993): $f(y) = \sqrt{y}$, $dY_t = k(\theta - Y_t)dt + \xi f(Y_t) dW_t^{(1)}$, $\rho \neq 0$

with m, k, θ, ξ constants. SV models 3 and 4 follow a mean-reverting Ornstein-Uhlenbeck process, model 5 a mean-reverting square root process. An overview of SV models is provided by Ghysels, Harvey, and Renault (1996), estimation techniques are discussed in Chernov and Ghysels (2000).

Note C.9 The financial market is incomplete for SV models. There is no unique risk-neutral measure $\mathbb{Q}$ – there exists a set of equivalent martingale measures $\mathcal{Q}$, characterized by the Radon-Nikodym derivative and obtained by Girsanov's theorem such that

$$\begin{aligned}\frac{d\tilde{\mathbb{Q}}}{d\mathbb{P}} = \exp\Bigg\{ - \int_0^t \lambda_s dW_s^{(0)} - \frac{1}{2}\int_0^t \lambda_s^2 ds \\ - \int_0^t \kappa_s dW_s^{(1)} - \frac{1}{2}\int_0^t \kappa_s^2 ds \Bigg\}. \end{aligned} \tag{C.18}$$

The Brownian motions on $(\Omega, \mathcal{F}, \tilde{\mathbb{Q}})$ are

$$\tilde{W}_t^{(0)} = W_t^{(0)} + \int_0^t \lambda_s ds \tag{C.19}$$

$$\tilde{W}_t^{(1)} = W_t^{(1)} + \int_0^t \kappa_s ds. \tag{C.20}$$

The discounted 'reinvestment stock portfolio'[36] $\tilde{S}_t = S_t \exp\left\{\int_0^t q_s - r_s ds\right\}$ is a $\tilde{\mathbb{Q}}$-martingale if and only if

$$\lambda_t := \frac{\mu_t + q_t - r_t}{\sigma(t, Y_t)}. \tag{C.21}$$

$(\kappa_t)_{t \geq 0}$ is an arbitrary $\mathbb{F}$-adapted process on $(\Omega, \mathcal{F}, \mathbb{P})$ such that the Novikov condition

$$\mathrm{E}\left\{\exp\left(\frac{1}{2}\int_0^T \lambda_s^2 + \kappa_s^2 \; ds\right)\right\} < \infty$$

holds. In that case, Itô's lemma and Eq. (C.19) imply $d\tilde{S}_t = \tilde{S}_t \sigma(t, Y_t) d\tilde{W}_t^{(0)}$ and contingent claims can be priced as described in Appendix B. κ_t is called 'market price of volatility risk'[37] and enters the option pricing PDE ('Merton-Garman equation').

---

[36] Invest an amount of S_0 in the stock at $t = 0$ and continuously reinvest the dividend pay out in the stock. The time t value of this strategy is $Z_t := S_t \exp\left\{\int_0^t q_s ds\right\}$. The dynamics $dZ_t = (\mu_t + q_t) Z_t dt + \sigma(t, Y_t) Z_t dW_t^{(0)}$ under $\mathbb{P}$ are inferred from Itô's lemma. The ex-dividend stock process S_t does not represent the value of a tradable asset (as required by the fundamental theorem of asset pricing to prove no-arbitrage), but Z_t does.

Remark C.10 An economic basis for a negative volatility risk premium is provided by van der Ploeg (2006, Section II.7) with the help of consumption-based asset pricing theory and the permanent-income hypothesis from macroeconomics. Numerous papers of the recent literature[38] find a negative volatility risk premium although inference is based on different definitions of this expression[39].

Driessen and Maenhout (2007); Carr and Wu (2009) find indication of a negative volatility risk premium by looking at investment strategies that are exposed to volatility risk only. Driessen et al. (2009) contribute the significant positive average excess return of dispersion trades to priced correlation risk. Comparing average model-free implied variances (Remark C.18) to average realized variances, they find insignificant differences, hence no volatility risk premium. *"In light of the general decomposition of index variance risk . . . , this analysis provides indirect evidence on the importance of priced correlation risk"* (2009, p. 1385). The authors develop and implement a simple option-based trading strategy that exploits priced correlation risk to prove their claim. Buraschi, Porchia, and Trojani (2009) analyze correlation risk in a new multivariate framework for intertemporal portfolio choice with stochastic second moments of asset returns.

Remark C.11 Lévy processes provide the basis for a broad generalization of the SV models. Applebaum (2004) reviews basic properties of Lévy processes and their applications in finance. The models of Merton (1976) and Bates (1996b) add jumps in the stock price dynamics only. Bakshi, Cao, and Chen (1997) derive a 'very general' option model with stochastic volatility, stochastic interest rate and with jumps (SVSI-J) and compare its performance to the BS model, models with stochastic volatility only (SV) and models with stochastic volatility and jumps (SVJ). According to them, most improvement over the benchmark BS model is achieved by SV models. Adding stochastic interest rate and jumps reduces option pricing errors only marginally. Gatheral lobbies for jumps as *"the impact on the shape of the volatility surface is all at the short-expiration end"* (2006, Chapter 5). He sees the SVJ model as the clear winner in a comparison with the Heston model (1993) and a stochastic volatility with simultaneous jumps in stock prices and volatility on S&P 500 index option data.

---

[37] Given a contingent claim H_t, $\kappa_t dt$ is the extra return per unit of volatility risk $dW_t^{(1)}$ for the delta-hedged (but not vega-hedged) portfolio $V(t, S_t, Y_t) = H_t - \frac{\partial H_t}{\partial S_t} S_t$. See Gatheral (2006, p. 6) for an example.

[38] E.g. Chernov and Ghysels (2000), Shumway and Coval (2001), Buraschi and Jackwerth (2001), Pan (2002), Jones (2003) and van der Ploeg (2006).

[39] Chernov and Ghysels (2000) characterize the risk premium for the continuous time affine class of SV models, van der Ploeg (2006) for multifactor extension of this class.

Note C.12 The Heston model (1993) is still popular nowadays because it has a closed-form solution for option prices and hedge ratios like delta and vega. This is a consequence of the assumed volatility risk premium (direct proportional to the level of volatility) and the parametrization. The instantaneous variance process in the Heston model follows a (mean-reverting) square root process (Cox, Ingersoll, and Ross, 1985), a special case of an affine jump-diffusion process (Duffie et al., 2000). Drifts, covariances and jump intensities are affine functions of the state variables. This class of process features analytical solutions for option prices by an extended transform technique.

C.3 Local volatility

Suppose asset prices follow the SDE

$$\frac{dS_t}{S_t} = (r_t - q_t)dt + \sigma(t, S_t, \cdot)d\tilde{W}_t \tag{C.22}$$

under an equivalent martingale measure $\mathbb{Q}$. It is not necessarily unique because $\sigma(t, S_t, \cdot)$ shall indicate the fact that instantaneous volatility might depend on other state variables that are either driven by the same Brownian motion (as in Remark C.6), by additional correlated ones or possibly by another random factor. The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and filtration $\mathbb{F} = \{\mathcal{F}_t, t \in [0, T]\}$ of the σ-algebra $\mathcal{F}$ need to be defined accordingly.

Definition C.13 (Local volatility) *The square root of the risk-neutral expectation of future instantaneous variance conditional on the market level at the same future date is called local volatility,*

$$\mathrm{LV}_{T,K}(t,S_t) := \sqrt{E_{\mathbb{Q}}\left[\sigma^2(T,S_T,\cdot)|S_T=K,\mathcal{F}_t\right]}. \tag{C.23}$$

Note C.14 For a one-factor diffusion process, i.e. when instantaneous volatility is a deterministic function of t and $S_t = h(t, W_t)$, then

$$\begin{aligned}\mathrm{LV}_{T,K}(t,S_t) &= \sqrt{\mathrm{E}_{\mathbb{Q}}\left[\sigma^2(T,S_T)|S_T=K,\mathcal{F}_t\right]}\\ &= \sqrt{\mathrm{E}_{\mathbb{Q}}\left[\sigma^2(T,h(T,W_T))|S_T=K,\mathcal{F}_t\right]}\\ &= \sqrt{\sigma^2(T,K)} = \sigma(T,K),\end{aligned}$$

hence instantaneous and local volatility coincide. *"In this case, instantaneous volatility evolves along the static local volatility function, since the right-hand side is independent of S and t"* (Fengler, 2005, p. 51). More trivially, the following equality always holds by definition of local volatility,

$$\sigma(t,S_t,\cdot) = \mathrm{LV}_{t,S_t}(t,S_t) \equiv \mathrm{LV}_{T,K}(t,S_t)|_{T=t,K=S_t}. \tag{C.24}$$

Remark C.15 Dupire's formula (1994) derives local volatility as a function of observed market prices of European plain vanilla call options,

$$\mathrm{LV}_{T,K}(t,S_t) = \sqrt{2\frac{\frac{\partial C_t(S_t,K,T)}{\partial T} + qC_t(S_t,K,T) + (r-q)K\frac{\partial C_t(S_t,K,T)}{\partial K}}{K^2\frac{\partial^2 C_t(S_t,K,T)}{\partial K^2}}}. \tag{C.25}$$

Historically, this formula has been derived in the context of a one-factor diffusion setting and both Dupire (1996) and Derman and Kani (1998) have independently expressed local variance as a conditional expectation of instantaneous variance.

Remark C.16 Rodrigo and Mamon (2008) use an ansatz approach to find a semi-explicit solution $\tilde{C}_t(S_t,T,K)$ to Dupire's forward equation (C.25), from which they are able to derive an explicit formula for the local volatility given by

$$\mathrm{LV}_{T,K}(t, S_t) = \frac{2}{K}\left[\frac{\partial z_2(K,\tau)/\partial\tau}{\partial z_2(K,\tau)/\partial K} - \frac{\partial z_1(K,\tau)/\partial\tau}{\partial z_1(K,\tau)/\partial K}\right] \tag{C.26}$$

$$z_1(K,\tau) = \Phi^{-1}(\beta_1(K,\tau)) \tag{C.27}$$

$$z_2(K,\tau) = \Phi^{-1}(\beta_2(K,\tau)) \tag{C.28}$$

$$\beta_1(K,\tau) = \frac{\exp\left\{\int_0^\tau q(s)ds\right\}}{S_t}\left[C_t(S_t,K,T) - K\frac{\partial C_t(S_t,K,T)}{\partial K}\right] \tag{C.29}$$

$$\beta_2(K,\tau) = -\exp\left\{\int_0^\tau r(s)ds\right\}\frac{\partial C_t(S_t,K,T)}{\partial K}. \tag{C.30}$$

Remark C.17 LV models are conceptionally questioned. Ayache et al. criticize that *"Dupire has not discovered a smile model. His great discovery was the forward PDE for pricing vanilla options of different strikes and different maturities in one solve"* (2004, p. 79). LV just simplifies a more complex stochastic instantaneous volatility process by integrating away all stochastic state variables (with the exception of S_t). Because the local volatility surface (like the IVS) flattens out for longer time horizons, not all exotic options can be priced with it, only short dated ones.

> *"The observed prices of vanilla options do not contain any information about the smile dynamics as they are just the snapshot of the present smile. In other words, from the prices of vanilla options (even a continuum thereof, in strike and maturity) we can only infer the probability distribution of the underlying price at the maturity dates of the options, as seen from today and from the spot price. Only in a . . . local volatility model does this impose the conditional, or forward, probability distributions. . . .* [In other models, they] *. . . are underdetermined"* (Ayache, 2007, p. 25).

C.4 Implied volatility

Implied volatility (IV) of an option contract is always linked to an option pricing model. It is the volatility that needs to be plugged in the model such that the model-based theoretical value for the option equals the observed market price of that option. The implied volatility surface (IVS) introduced in Chapter 3 is based on the model of Black and Scholes (1973). IV is throughout this thesis depending on the Black Scholes model if not stated otherwise.

Gouriéroux et al. (1995) rewrite the BS formula (2.23) in terms of moneyness $m = K/S_t$ and time to maturity $\tau = T - t$,

$$\begin{aligned} \mathrm{BS}_t(S_t, \sigma, \mathit{cp\ flag}, K, T, r, q) = S_t \{ & c^{\mathrm{BS}}(m, \tau, \sigma, r, q) \mathbb{I}_{\mathit{cp\ flag}=1} \\ & + p^{\mathrm{BS}}(m, \tau, \sigma, r, q) \mathbb{I}_{\mathit{cp\ flag}=0} \} \end{aligned} \tag{C.31}$$

$$c^{\mathrm{BS}}(m, \tau, \sigma, r, q) = e^{-q\tau}\Phi(d_1) - me^{-r\tau}\Phi(d_2) \tag{C.32}$$

$$p^{\mathrm{BS}}(m, \tau, \sigma, r, q) = me^{-r\tau}\Phi(-d_2) - e^{-q\tau}\Phi(-d_1) \tag{C.33}$$

where

$$d_1 = \frac{-\log m + (r - q + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}} \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{\tau}. \tag{C.34}$$

In the general statistical framework of Section 3.3.2, the predictor space

$$\mathbf{x}^{\mathrm{pred}} = (m, \tau, \mathit{cp\ flag}, \mathit{factors})$$

is introduced to identify the dependencies of IV. The set of *factors* contains exogenous factors or state variables. From Eq. (C.31) follows that $\mathit{factors} = \{S_t, r, q\}$ in the BS framework. By definition of implied volatility,

$$c^{\mathrm{BS}}(m, \tau, \sigma_t^{\mathrm{IV}}(\mathbf{x}^{\mathrm{pred}}), r, q) = C_t/S_t \tag{C.35}$$

$$p^{\mathrm{BS}}(m, \tau, \sigma_t^{\mathrm{IV}}(\mathbf{x}^{\mathrm{pred}}), r, q) = P_t/S_t. \tag{C.36}$$

The IVS is defined as the mapping

$$\begin{aligned} \sigma_t^{\mathrm{IV}} : \ & \mathbb{R} \times \mathbb{R}_+ \longrightarrow \mathbb{R}_+ \\ & (m, \tau) \longmapsto \sigma_t^{\mathrm{IV}}(\mathbf{x}^{\mathrm{pred}}) \ . \end{aligned} \tag{C.37}$$

Remark C.18 Britten-Jones and Neuberger (2000) derive a model-free implied volatility from the observed market prices of European options. Without loss of generality, they assume that $r = q = 0$. *"In the presence of nonzero interest rates and dividends, the option and underlying asset prices are viewed as forward* *prices"* (2000, footnote 5). With the help of a *"simple condition characterizing the set of all continuous price processes that are consistent with a given set of option prices"* (2000, p. 857), they find

$$\begin{aligned} \mathrm{E}_{\mathbb{Q}}[\sigma^2(S_T,T,\cdot)|\mathcal{F}_t] &= \int_0^\infty \overbrace{\mathrm{E}_{\mathbb{Q}}[\sigma^2(S_T,T,\cdot)|S_T=K,\mathcal{F}_t]}^{\rightsquigarrow\text{Eq. (C.25)}}\,\overbrace{q(S_T=K|\mathcal{F}_t)}^{\rightsquigarrow\text{Eq. (2.27)}}\,dK \\ &= \int_0^\infty \left(2\frac{\frac{\partial C_t(S_t,K,T)}{\partial T}}{K^2\frac{\partial^2 C_t(S_t,K,T)}{\partial K^2}}\right)\cdot\left(\frac{\partial^2 C_t(S_t,K,T)}{\partial K^2}\right) \\ &= 2\int_0^\infty \frac{1}{K^2}\frac{\partial C_t(S_t,K,T)}{\partial T}dK \end{aligned} \tag{C.38}$$

and hence for $T_1 < T_2$, the risk-neutral expected integrated variance is

$$\mathrm{E}_{\mathbb{Q}}\left[\int_{T_1}^{T_2}\sigma^2(S_T,T,\cdot)dT\,\middle|\,\mathcal{F}_t\right] = 2\int_0^\infty \frac{C_t(S_t,K,T_2)-C_t(S_t,K,T_1)}{K^2}dK. \tag{C.39}$$

"Jiang and Tian (2005) show that [this] *method also yields an accurate measure of the (total) risk-neutral expected integrated variance in a jump-diffusion setting"* (Driessen et al., 2009, p. 1383).

Definition C.19 (Model-free implied volatility) *The spot volatility $b(t,S_t)$ that makes $dS_t/S_t = b(t,S_t)dS_t$ a continuous price processes that is consistent with the set of T-expiry option prices is called model-free implied volatility,*

$$\mathrm{MFIV}_T(t) = \sqrt{2\int_0^\infty \frac{C_t(S_t,K,T)-\max(S_t-K,0)}{K^2}dK}. \tag{C.40}$$

Note C.20 Instantaneous volatility, LV and IV are all equal to the same constant σ^{BS} if the BS framework matched reality. From Note C.14 follows that $\mathrm{LV}_{t,S_t}(t,S_t) = \sigma(t,S_t) = \sigma^{\text{BS}} = \sigma(T,K) = \mathrm{LV}_{T,K}(t,S_t)$. Rodrigo and Mamon's LV representation in Remark C.16 reduces to

$$\mathrm{LV}_{T,K}(t,S_t) = (z_1(K,\tau)-z_2(K,\tau))/\sqrt{\tau},$$

with $z_i(K,\tau) = \Phi^{-1}(\beta_i(K,\tau)) \equiv d_i$, hence $d_2 = d_1 - \mathrm{LV}_{T,K}(t,S_t)\sqrt{\tau}$ as in the BS formula (2.23). Thus, $\mathrm{LV}_{T,K}(t,S_t) = \mathrm{IV}$.

Remark C.21 Durrleman (2004) provides a link between the implied volatility dynamics and the instantaneous volatility dynamics of the underlying stock. He notes that the latter can be recovered from the behavior of close ATM option prices near the expiry date. By observing implied volatilities dynamics, he concludes that the general spot volatility dynamics follows a mean reverting square root process, however with random coefficients.

Bibliography

Abken, P. A. and S. Nandi (1996, December). Options and volatility. *Economic Review 81*, 21–35.

Ahoniemi, K. (2006). Modeling and forecasting implied volatility - an econometric analysis of the VIX index. Discussion Paper No. 129, Helsinki Center of Economic Research.

Ahoniemi, K. and M. Lanne (2009). Joint modeling of call and put implied volatility. *International Journal of Forecasting 25*, 239–258.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71*, 529–626.

Applebaum, D. (2004). Lévy processes - from probability to finance and quantum groups. *Notices of the American Mathematical Society 51*(11), 1336–1347.

Aït-Sahalia, Y. and A. W. Lo (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance 53*, 499–548.

Audrino, F. and G. Barone-Adesi (2005). Functional gradient descent for financial time series with an application to the measurement of market risk. *Journal of Banking and Finance 29*(4), 959–977.

Audrino, F. and G. Barone-Adesi (2006). A dynamic model of expected bond returns: a functional gradient descent approach. *Computational Statistics and Data Analysis 51*(4), 2267–2277.

Audrino, F., G. Barone-Adesi, and A. Mira (2005). The stability of factor models of interest rates. *Journal of Financial Econometrics 3*(3), 422–441.

Audrino, F. and P. Bühlmann (2003). Volatility estimation with functional gradient descent for very high-dimensional financial time series. *Journal of Computational Finance 6*(3), 1–26.

Audrino, F. and F. Trojani (2007). Accurate short-term yield curve forecasting using functional gradient descent. *Journal of Financial Econometrics 5*(4), 591–623.

Ayache, E. (2007, January). Dial 33 for your local cleaner. *Wilmott magazine*, 24–33.

Ayache, E., P. Henrotte, S. Nassar, and X. Wang (2004, January). Can anyone solve the smile problem? *Wilmott magazine*, 78–96.

Bakshi, G., C. Cao, and Z. Chen (1997). Empirical performance of alternative option pricing models. *Journal of Finance 52*(5), 2003–2049.

Bakshi, G. and Kapadia (2003). Delta-hedged gains and the negative market volatility risk premium. *Review of Financial Studies 16*(2), 527–566.

Bandi, F. M. and J. R. Russell (2008, April). Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies 75*(2), 339–369.

Barone-Adesi, G., F. Bourgoin, and K. Giannopoulos (1998). Don't look back. *Risk 11*, 100–104.

Barone-Adesi, G. and R. J. Elliott (2007). Cutting the hedge. *Computational Economics 29*(2), 151–158.

Barone-Adesi, G., R. F. Engle, and L. Mancini (2008, May). A GARCH option pricing model with filtered historical simulation. *Review of Financial Studies 21*, 1223–1258.

Barone-Adesi, G., K. Giannopoulos, and L. Vosper (1999). VaR without correlations for portfolios of derivative securities. *Journal of Futures Markets 19*, 583–602.

Bates, D. M. (1996a). *Handbook of Statistics*, Volume 14, Chapter "Testing option pricing models", pp. 567–646. Elsevier Science.

Bates, D. S. (1996b). Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche Mark options. *Review of Financial Studies 9*(1), 69–107.

Battalio, R. and P. Schultz (2006). Options and the bubble. *Journal of Finance 61*(5), 2071–2102.

Benfey, O. T. (1958). August Kekulé and the birth of the structural theory of organic chemistry in 1858. *Journal of Chemical Education 35*, 21–23.

Benko, M. (2006). *Functional Data Analysis with Applications in Finance.* Ph. D. thesis, Wirtschaftwissenschaftlichen Fakultät, Humboldt-Universität zu Berlin.

Bühlmann, P. and T. Hothorn (2007). Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science 22*(4), 477–505.

Black, F. (1976). Studies of stock price volatility changes. In *Proceedings of the 1976 Meetings of the American Statistical Association*, pp. 177–181.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy 81*(3), 637–654.

Bliss, R. R. and N. Panigirtzoglou (2002). Testing the stability of implied probability density functions. *Journal of Banking and Finance 26*(2), 381–422.

Bollen, N. P. B. and R. E. Whaley (2004). Does net buying pressure affect the shape of implied volatility functions? *Journal of Finance 59*(2), 711–753.

Breeden, D. T. and R. H. Litzenberger (1978, October). Prices of state-contingent claims implicit in option prices. *The Journal of Business 51*(4), 621–651.

Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen (1984). *Classification and Regression Trees.* Chapman & Hall/CRC.

Brigo, D. and F. Mercurio (2001). Displaced and mixture diffusions for analytically-tractable smile models. In H. Geman, D. Madan, S. R. Pliska, and T. Vorst (Eds.), *Mathematical Finance Bachelier Congress 2000.*

Britten-Jones, M. and A. Neuberger (2000). Option prices, implied price processes, and stochastic volatility. *Journal of Finance 55*(2), 839–866.

Brooks, C. and M. C. Oozeer (2002). Modelling the implied volatility of options on long gilt futures. *Journal of Business Finance & Accounting 29*, 111–137.

Brunner, B. and R. Hafner (2003). Arbitrage-free estimation of the risk-neutral density from the implied volatility smile. *Journal of Computational Finance 7*(1), 75–106.

Buraschi, A. and J. C. Jackwerth (2001). The price of a smile: hedging and spanning in option markets. *Review of Financial Studies 14*(2), 495–527.

Buraschi, A., P. Porchia, and F. Trojani (2009). Correlation risk and optimal portfolio choice. *Journal of Finance*, forthcoming.

Carr, P. and L. Wu (2009). Variance risk premiums. *Review of Financial Studies 22*(3), 1311–1341.

Cassese, G. and M. Guidolin (2004). Pricing and informational efficiency of the MIB30 index options market. An analysis with high frequency data. *Economic Notes 33*, 275–321.

Cassese, G. and M. Guidolin (2006). Modelling the implied volatility surface: Does market efficiency matter?: An application to MIB30 index options. *International Review of Financial Analysis 15*(2), 145–178.

Chernov, M. and E. Ghysels (2000). *Computational Finance 1999*, Chapter "Estimation of Stochastic Volatility Models for the Purpose of Option Pricing", pp. 567–582. MIT Press.

Cont, R. and J. da Fonseca (2002, February). Dynamics of implied volatility surfaces. *Quantitative Finance 2*(1), 45–60.

Corsi, F. (2005). *Measuring and Modelling Realized Volatility: from Tick-by-tick to Long Memory.* Ph. D. thesis, Università della Svizzera italiana.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica 53*, 385–407.

Cox, J. C. and S. A. Ross (1976). The valuation of options for alternative stochastic processes. *Journal of Financial Economics 76*, 145–166.

Daglish, T., J. C. Hull, and W. Suo (2007). Volatility surfaces: Theory, rules of thumb, and empirical evidence. *Quantitative Finance 7*(5), 507–524.

Demeterfi, K., E. Derman, M. Kamal, and J. Zou (1999). A guide to volatility and variance swaps. *Journal of Derivatives 6*(4), 9–32.

Derman, E. and I. Kani (1994). Riding on a smile. *Risk 7*, 32–39.

Derman, E. and I. Kani (1998, January). Stochastic implied trees: Arbitrage pricing with stochastic term and strike structure of volatility. *International Journal of Theoretical and Applied Finance 1*(1), 61–110.

Derman, E., I. Kani, and N. Chriss (1996). Implied trinomial trees of the volatility smile. *Journal of Derivatives 3*(4), 7–22.

Detlefsen, K., W. Härdle, and R. A. Moro (2007). Empirical pricing kernels and investor preferences. Working paper, Humboldt-Universität zu Berlin.

Doan, T., R. Litterman, and C. Sims (1984). Forecasting and conditional projections using realistic prior distributions. *Econometric Reviews 3*(1), 1–100.

Driessen, J. and P. J. Maenhout (2007). An empirical portfolio perspective on option pricing anomalies. *Review of Finance 11*, 561–603.

Driessen, J., P. J. Maenhout, and G. Vilkov (2009, June). The price of correlation risk: Evidence from equity options. *Journal of Finance 64*(3), 1377–1406.

Duffie, D. (2001). *Dynamic Asset Pricing Theory.* Princeton University Press.

Duffie, D. and J. Pan (2001). *Options Markets*, Volume 6 of *The International Library of Critical Writings in Financial Economics*, Chapter "An Overview of Value at Risk". Edward Elgar.

Duffie, D., J. Pan, and K. Singleton (2000). Transform analysis and asset pricing for affine jump diffusions. *Econometrica 68*(6), 1343–1376.

Dumas, B., J. Fleming, and R. E. Whaley (1998). Implied volatility functions: Empirical tests. *Journal of Finance 53*(6), 2059–2106.

Dupire, B. (1994). Pricing with a smile. *Risk 7*(1), 18–20.

Dupire, B. (1996). A unified theory of volatility. Discussion paper Paribas Capital Markets, reprinted in *Derivatives Pricing: The Classic Collection*, edited by Peter Carr, 2004 (Risk Books, London).

Durrleman, V. (2004). *From Implied to Spot Volatilities*. Ph. D. thesis, Department of Operations Research and Financial Engineering, Princeton University.

Durrleman, V. (2008). Convergence of at-the-money implied volatilities to the spot volatility. *Journal of Applied Probability 45*(2), 542–550.

Dybvig, P. H. and S. A. Ross (2003). *Handbook of the Economics of Finance: Financial Markets and Asset Pricing*, Volume 1B of *Handbooks in Economics*, Chapter "Arbitrage, State Prices and Portfolio Theory", pp. 605–634. North Holland.

El Karoui, N., M. Jeanblanc-Picqué, and S. E. Shreve (1998, April). Robustness of the Black and Scholes formula. *Mathematical Finance 8*(2), 93–126.

Elliott, R. J. and P. E. Kopp (2005). *Mathematics of financial markets* (2nd ed.). Springer.

Fengler, M. R. (2005). *Semiparametric Modeling of Implied Volatility.* Springer.

Fengler, M. R. (2009, June). Arbitrage-free smoothing of the implied volatility surface. *Quantitative Finance 9*(4), 417–428.

Fengler, M. R., W. Härdle, and E. Mammen (2005). A dynamic semiparametric factor model for implied volatility string dynamics. Working paper, SFB 649 Discussion Paper 2005-020, Humboldt-Universität zu Berlin.

Fengler, M. R., W. Härdle, and E. Mammen (2007). A semiparametric factor model for implied volatility surface dynamics. *Journal of Financial Econometrics 5*(2), 189–218.

Fengler, M. R., W. Härdle, and C. Villa (2003). The dynamics of implied volatilities: A common principal components approach. *Review of Derivatives Research 6*, 179–202.

Figlewski, S. (1997, February). Forecasting volatility. *Financial Markets, Institutions & Instruments 6*(1), 1–88.

Freund, Y. and R. E. Schapire (1997, August). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences 55*(1), 119–139.

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics 29*(5), 1189–1232.

Friedman, J., T. Hastie, and R. Tibshirani (2000). Additive logistic regression: A statistical view of boosting. *Annals of Statistics 28*(2), 337–407.

Gagliardini, P., C. Gouriéroux, and E. Renault (2008, October). Efficient derivative pricing by the extended method of moments. Working paper.

Garcia, R., R. Luger, and E. Renault (2003). Pricing and hedging options with implied asset prices and volatilities. Working Paper, CIRANO, CIREQ and Université de Montréal.

Gatheral, J. (2006). *The Volatility Surface: A Practitioner's Guide.* Wiley & Sons.

Gavrishchaka, V. V. (2006). *Econometric Analysis of Financial and Economic Time Series*, Volume 20 Part 2 of *Advances in Econometrics*, Chapter "Boosting-Based Frameworks in Financial Modeling: Application to Symbolic Volatility Forecasting", pp. 123–151. Elsevier.

Ghysels, E., A. Harvey, and E. Renault (1996). *Handbook of Statistics*, Volume 14, Chapter "Stochastic volatility", pp. 119–192. Elsevier Science.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance 48*(5), 1779–1801.

Gonçalves, S. and M. Guidolin (2006, May). Predictable dynamics in the S&P 500 index options implied volatility surface. *Journal of Business 79*(3), 1591–1635.

Gouriéroux, C., A. Monfort, and C. Tenreiro (1994). Nonparametric diagnostics for structural models. Document de travail 9405, CREST, Paris.

Gouriéroux, C., A. Monfort, and C. Tenreiro (1995). Kernel M-estimators and functional residual plots. Document de travail 9546, CREST, Paris.

Goyal, A. and A. Saretto (2009, November). Cross-section of option returns and volatility. *Journal of Financial Economics 94*(2), 310–326.

Hafner, R. (2004). *Stochastic Implied Volatility: A Factor-Based Model*, Volume 545 of *Lecture Notes in Economics and Mathematical Systems*. Springer.

Harrison, J. M. and D. M. Kreps (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory 20*(3), 381–408.

Harvey, C. R. and R. E. Whaley (1992). Market volatility prediction and the efficiency of the S&P 100 index option market. *Journal of Financial Economics 31*(1), 43–73.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.

Hentschel, L. (2003). Errors in implied volatility estimation. *Journal of Financial and Quantitative Analysis 38*(4), 779–810.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies 6*(2), 327–343.

Heston, S. L. and S. Nandi (2000). A closed-form GARCH option valuation model. *Review of Financial Studies 13*(3), 585–625.

Hull, J. (2002). *Options, Futures, and Other Derivatives* (5 ed.). Pearson Education.

Hull, J. and A. White (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance 42*(2), 281–300.

Jackwerth, J. C. (1997). Generalized binomial trees. *Journal of Derivatives 5*, 7–17.

Jiang, G. J. and Y. S. Tian (2005). The model-free implied volatility and its information content. *Review of Financial Studies 18*(4), 1305–1342.

Jiang, W. (2004). Process consistency for AdaBoost (with discussion). *Annals of Statistics 32*(1), 13–29.

Jones, C. S. (2003). The dynamics of stochastic volatility: evidence from underlying and options markets. *Journal of Econometrics 116*, 181–224.

Kahalé, N. (2004, May). An arbitrage-free interpolation of volatilities. *Risk 17*(5), 102–106.

Karatzas, I. and S. E. Shreve (1991). *Brownian Motion and Stochastic Calculus* (2nd ed.), Volume 113 of *Graduate Texts in Mathematics*. Springer.

Ledoit, O., P. Santa-Clara, and S. Yan (2002). Relative pricing of options with stochastic volatility. Working Paper, Anderson Graduate School of Management, Los Angeles.

Lee, R. W. (2005). *Recent Advances in Applied Probability*, Chapter "Implied Volatility: Statics, Dynamics, and Probabilistic Interpretation", pp. 241–268. Springer.

LeSage, J. P. (1999, October). Applied econometrics using MATLAB. Manual to Econometrics Toolbox for MATLAB, `http://www.spatial-econometrics.com/html/mbook.pdf` (accessed 28 October 2009).

Makropoulou, V. and R. N. Markellos (2005). What is the fair rent Thales should have paid? *7th Hellenic-European Conference on Computer Mathematics and its Applications* at Athens University of Economics and Business.

Manaster, S. and R. J. Rendleman (1982). Option prices as predictors of equilibrium stock prices. *Journal of Finance 37*(4), 1043–1057.

Mannor, S., R. Meir, and S. Mendelson (2001). On the consistency of boosting algorithms. Unpublished manuscript.

Mannor, S., R. Meir, and T. Zhang (2003). Greedy algorithms for classification - consistency, convergence rates, and adaptivity. *Journal of Machine Learning Research 4*, 713–742.

McIntyre, M. L. (2001). Performance of dupire's implied diffusion approach under sparse and incomplete data. *Journal of Computational Finance 4*(4), 33–84.

McNeil, A. J., R. Frey, and P. Embrechts (2005). *Quantitative Risk Management: Concepts, Techniques and Tools.* Princeton Series in Finance. Princeton University Press.

Mease, D. and A. Wyner (2008). Evidence contrary to the statistical view of boosting. *Journal of Machine Learning Research 9*, 131–156.

Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science 4*(1), 141–183.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics 3*, 125–144.

Morgenson, G. and C. R. Harvey (2002). *The New York Times Dictionary of Money and Investing: The Essential A-to-Z Guide to the Language of the New Market.* Times Books.

Murphy, A. L. (2009, August). Trading options before Black-Scholes: a study of the market in late seventeenth-century London. *The Economic History Review 62*, 8–30.

Noh, J., R. F. Engle, and A. Kane (1994). Forecasting volatility and option prices of the S&P 500 index. *Journal of Derivatives 2*, 17–30.

Pan, J. (2002). The jump-risk premia implicit in options: evidence from an integrated time-series study. *Journal of Financial Economics 63*, 3–50.

Panda, B., J. S. Herbach, S. Basu, and R. J. Bayardo (2009). PLANET: massively parallel learning of tree ensembles with MapReduce. *Proceedings of the VLDB Endowment 2*(2), 1426–1437.

Pliska, S. R. (1997). *Introduction to Mathematical Finance: Discrete Time Models.* Blackwell.

Poon, S.-H. and C. W. J. Granger (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature 41*(2), 478–539.

Rodrigo, M. R. and R. S. Mamon (2008). A new representation of the local volatility surface. *International Journal of Theoretical and Applied Finance 11*(7), 691–703.

Rosenberg, J. (2000). Implied volatility functions: A reprise. *Journal of Derivatives 7*, 51–64.

Rossi, A. and A. G. Timmermann (2009, July). What is the shape of the risk-return relation? Working paper, available at SSRN: `http://ssrn.com/abstract=1364750`.

Rubinstein, M. (1985). Nonparametric 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*, 455–480.

Rubinstein, M. (1994). Implied binomial trees. *Journal of Finance 49*, 771–818.

Schachermayer, W. (2009). *Encyclopedia of Quantitative Finance*, Volume 4, Chapter "The fundamental theorem of asset pricing". Wiley & Sons.

Schönbucher, P. J. (1999). A market model for stochastic implied volatility. *Philosophical Transactions of the Royal Society 357*(1758), 2071–2092.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization.* Wiley & Sons.

Scott, L. O. (1987, December). Option pricing when the variance changes randomly: Theory, estimation, and an application. *The Journal of Financial and Quantitative Analysis 22*(4), 419–438.

Shimko, D. (1993). Bounds of probability. *Risk 6*(4), 33–37.

Shumway, T. and J. D. Coval (2001). Expected option returns. *Journal of Finance 56*(3), 983–1009.

Skiadopoulos, G. S., S. D. Hodges, and L. Clewlow (2000). The dynamics of the S&P 500 implied volatility surface. *Review of Derivatives Research 3*(3), 263–282.

Stein, E. M. and J. C. Stein (1991). Stock price distributions with stochastic volatility: An analytic approach. *Review of Financial Studies 4*(4), 727–752.

Studer, M. (2001). *Stochastic Taylor Expansions and Saddlepoint Approximations for Risk Management.* Ph. D. thesis, ETH Zürich, Diss. ETHNo. 14242.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological) 58*(1), 267–288.

van der Ploeg, A. (2006). *Stochastic Volatility and the Pricing of Financial Derivatives.* Ph. D. thesis, Tinbergen Institute.

Vonhoff, V. (2006). Dispersion trading. Master's thesis, Universität Konstanz.

Wang, Y., H. Yin, and L. Qi (2004, March). No-arbitrage interpolation of the option price function and its reformulation. *Journal of Optimization Theory and Applications 120*(3), 627–649.

Watson, D. F. (1992). *Contouring: A Guide to the Analysis and Display of Spatial Data*. Pergamon.

Wiggins, J. B. (1987). Option values under stochastic volatility: Theory and empirical estimates. *Journal of Financial Economics 19*(2), 351–372.

Wikipedia contributors (2009). "Greeks (finance)", *Wikipedia, The Free Encyclopedia*. `http://en.wikipedia.org/w/index.php?title=Greeks_(finance)&oldid=312362269` (accessed 7 September 2009).

Wilmott, P. (2008). Science in Finance IX: In defence of Black, Scholes and Merton. Blog entry, `http://www.wilmott.com/blogs/paul/index.cfm/2008/4/29/Science-in-Finance-IX-In-defence-of-Black-Scholes-and-Merton` (accessed 28 October 2009).

List of Figures

3.1 IV scatter-plot of S&P 500 options on $t = 10$ August 2001 in absolute coordinates (K, T) . . . 22
3.2 IV scatter-plot of S&P 500 options on $t = 10$ August 2001 in relative coordinates (m, τ) . . . 24

4.1 A three location regression tree as a starting model $F_0(\cdot)$ for the S&P 500 IVS . . . 43

5.1 Estimated IS residuals $\sigma_t^{\mathrm{IV}} - \hat{F}_0$ for Microsoft call options closest to $m \approx 1$ and $\tau >\approx 10$ days . . . 49
5.2 Estimated IS residuals $\sigma_t^{\mathrm{IV}} - \hat{F}_0$ for Microsoft put options closest to $m \approx 1$ and $\tau >\approx 10$ days . . . 50
5.3 Time series plot of σ_t^{IV} for the S&P 500 option with highest IV in the sample . . . 53
5.4 Estimated OS residuals $\sigma_t^{\mathrm{IV}} - \hat{F}_0$ for Microsoft options closest to $m \approx 1$ and $\tau >\approx 10$ days . . . 56
5.5 Plot of weight functions $\omega_1(m)$ and $\omega_2(\tau)$ . . . 61

6.1 Closing prices and log returns of the S&P 500 index . . . 69
6.2 Estimated residuals and conditional volatilities of S&P 500 log returns fitted with a Glosten-Jagannathan-Runkle GARCH model . . . 70
6.3 Estimated residuals and conditional volatilities of S&P 500 log returns fitted with a Heston-Nandi GARCH model . . . 76
6.4 Time series plot of observed σ_t^{IV} and estimated $\hat{\sigma}_t^{\mathrm{IV}}$ obtained from a Heston-Nandi GARCH model for S&P 500 options closest to $m \approx 1$ and $\tau \approx 30$ days . . . 77
6.5 Filtered historical simulation of exogenous factors for a 60 days OS period . . . 81
6.6 Cross-validation plot of the empirical local criterion Λ_{grid} and SSR 85
6.7 Boxplot of daily averaged SSR_t . . . 89

6.8 Time series plot of estimated spot volatilities obtained from IVs . . 94
6.9 Comparison of the 35 days OS forecasts with the actually observed IVs of all S&P 500 options on that day 97
6.10 IS plot of the S&P 500 IVS on $t = 10$ August 2001 for {regtree, adhocbs, stickym} and their tree-boosted versions. 99
6.11 IS plot of the S&P 500 IVS on $t = 10$ August 2001 for {bvar, dsfm} and their tree-boosted versions. 100
6.12 OS plot of the S&P 500 IVS on $t = 17$ September 2001 for {regtree, adhocbs, stickym} and their tree-boosted versions. 101
6.13 OS plot of the S&P 500 IVS on $t = 17$ September 2001 for {bvar, dsfm} and their tree-boosted versions. 102

7.1 The performance of the strategy Ranking(5, $\mathbf{w}$, short linked) for varying $\mathbf{w}$ 124
7.2 Plot of Ranking(5, $\mathbf{w}$, short linked) performance against $\mathbf{w}$ 125
7.3 Evolution of the total wealth process V_t when investing 20% in the option strategy and 80% in the risk-free asset 129

List of Tables

2.1 Definition of the Greeks 15

3.1 Moneyness categories for options when moneyness is defined as $m = K/S_t$. 23

6.1 Descriptive statistics of S&P 500 IV data 67
6.2 Estimated parameters for a Glosten-Jagannathan-Runkle GARCH and a standard GARCH model fitted on S&P 500 log returns . . . 68
6.3 Description of subsamples and special days of interest for the analysis of the S&P 500 IVS 72
6.4 Specifications of the Heston-Nandi-GARCH (HNG) factors 74
6.5 Robustness of regression trees as base learners 80
6.6 Cross-validated optimal stopping value $\hat{M}$ for tree-boosted versions of starting models {regtree, adhocbs, stickym} in dependence of chosen grid points and predictor variable set 83
6.7 Cross-validated optimal stopping value $\hat{M}$ for tree-boosted versions of starting models {bvar, dsfm} in dependence of chosen grid points and predictor variable set 84
6.8 Relative importance of predictor variables 87
6.9 Overall averaged SSR performance over 60 OS days of S&P 500 IV predictions 91
6.10 OS robustness check of IV predictions based on an alternative data set of option prices 96

7.1 Descriptive statistics of aggregated call and put option returns . . 107
7.2 Decile portfolio returns sorted on log difference between HV and IV 109
7.3 Option portfolio returns for $k = 5$ 118
7.4 Option portfolio returns for $k = 10$ 119
7.5 Option portfolio returns for $k = 20$ 120
7.6 Measures of concordance 122

7.7 Performance of Ranking(5, $\mathbf{w}$, short linked) in dependence of $\mathbf{w}$ . . 123
7.8 Option portfolio returns under perfect foresight of $\hat{S}_{t+\delta t} = S_{t+\delta t}$ . . 126
7.9 Performance and risk measures of the total wealth process V_t when investing 20% in the option strategy and 80% in the risk-free asset 130

Abbreviations

adhocbs	ad hoc Black-Scholes model
a.k.a.	also known as
ARMA	auto regressive moving average
ATM	at-the-money
BS	Black-Scholes
bvar	bayesian vector autoregression
CART	classification and regression tree
CBOE	Chicago Board Options Exchange
CBOT	Chicago Board of Trade
CDF	cumulative distribution function
CF	cash flow
CPC	common principal component
CRR	Cox-Ross-Rubinstein
DITM	deep in-the-money
DOTM	deep out-of-the-money
dsfm	dynamic semiparametric factor model
EC	empirical criterion
e.g.	exempli gratia
Eq.	equation
ES	expected shortfall

etc.	et cetera
ETF	exchange-traded fund
FGD	functional gradient descent
FHS	filtered historical simulation
GARCH	generalized autoregressive conditional heteroskedasticity
GBM	geometric brownian motion
GJR	Glosten-Jagannathan-Runkle
GP	grid points
GS	Goyal-Saretto
HNG	Heston-Nandi-GARCH
HV	historical volatility
i.e.	id est
IS	in-sample
ITM	in-the-money
IV	implied volatility
IVS	implied volatility surface
LSK	least-square kernel
LV	local volatility
MFIV	model-free implied volatility
MSE	mean squared error
OS	out-of-sample
OTM	out-of-the-money
PCA	principal components analysis
PDE	partial differential equation
PDF	probability density function
regtree	regression tree
RV	realized variance

SDE	stochastic differential equation
SPD	state-price density
SPDR	Standard & Poor's Depositary Receipts
SR	Sharpe ratio
SSR	sum of squared residuals
std	standard deviation
stickym	sticky moneyness
SV	stochastic volatility
treefgd	tree-boosting algorithm
VaR	value at risk
VIX	CBOE volatility index
w.r.t.	with respect to