risklab germany | advanced asset management solutions

working paper

A Factor-Based Stochastic Implied Volatility Model

Dr. Reinhold Hafner
Dr. Bernd Schmid

March 2005

risklab germany GmbH
Nymphenburger Str. 112-116
80636 München
Tel. +49.89.1220 7750

wp 05-01

A Factor-Based Stochastic Implied Volatility Model

Dr. Reinhold Hafner*, Dr. Bernd Schmid†

Current Version: March 2005

Abstract

We present a factor-based model of the stochastic evolution of the implied volatility surface. The model allows for the integrated and consistent pricing and hedging, risk management, and trading of equity index derivatives as well as volatility derivatives. We develop a unifying theory for the analysis of contingent claims under both the real-world measure and the risk-neutral measure in an environment of stochastic implied volatility. On the basis of transaction data, we provide extensive statistical analyses on the dynamics of the implied volatility surface of German DAX options and propose a four-factor model to describe its evolution. The model is validated and tested on market data.

JEL classification: G10; G12; G13

Keywords: Implied Volatility; Smile; No-Arbitrage; Pricing; DAX option

*risklab germany GmbH, Managing Director, Nymphenburger Str. 112-116, 80636 München, Germany, e-mail: reinhold.hafner@risklab.de

†Fitch Risk, Senior Vice President, 101 Finsbury Pavement, London, EC2A 1RS, United Kingdom, e-mail: bernd.schmid@fitchrisk.com

1 Introduction

The increasing use and complexity of derivatives makes a framework necessary that enables for an *accurate* and *consistent* pricing and hedging, risk management, and trading of derivative products, including all kinds of exotic derivatives.

The first important attempt was the *Black-Scholes option pricing model*, developed by Black & Scholes (1973), formalized and extended in the same year by Merton (1973). This model provides a unique "fair price" for a (European) option that is traded on a frictionless market and whose underlying asset exhibits lognormally distributed prices. However, empirical investigation of the Black-Scholes model shows statistically significant and economically relevant deviations between market and model prices. A convenient way of illustrating these discrepancies is to express the option price in terms of its *implied volatility*, i.e. as the volatility parameter value that, when plugged into the Black-Scholes formula, results in a model price equal to the market price. If the Black-Scholes model was correct, then all options on the same underlying asset should provide the same implied volatility. However, Black-Scholes implied volatilities usually differ across exercise prices and times to maturity. The relationship between implied volatilities and exercise prices is commonly referred to as the "*volatility smile*" and the relationship between implied volatilities and times to maturity as the "volatility term structure". Volatility surfaces combine the volatility smile with the term structure of volatility.

The obvious shortcomings of the Black-Scholes model have led to the development of a considerable literature on alternative option pricing models, which attempt to identify and model the financial mechanisms that give rise to volatility surfaces, in particular to smiles. One large part of the literature generalizes the assumption of a constant volatility parameter to deterministically or stochastically varying volatility. Derman & Kani (1994b), Derman & Kani (1994a), Dupire (1994), and Rubinstein (1994) were the first ones to model volatility as a *deterministic* function of time and the stock price, known as local volatility. In their case, the unknown volatility function can be fitted to observed option prices to obtain an implied price process for the underlying asset. Unfortunately, in an empirical study Dumas et al. (1998) conclude that, as far as S&P 500 options are concerned, local volatility models are unreliable and not really useful for valuation and risk management. Therefore, the extension to a *stochastic* volatility approach was motivated by empirical studies on the time series behavior of (realized) volatilities. Specifications for a stochastic volatility process have been proposed by a number of authors, including Hull & White (1987), Wiggins (1987), Scott (1987), Stein & Stein (1991), and Heston (1993). A problem of stochastic volatility models is that unrealistically high parameters are required in order to generate volatility smiles that are consistent with those observed in option prices with short times to maturity.[1] Furthermore, these models are *incomplete*. Consequently, the requirement of no-arbitrage is no longer sufficient to determine a unique preference-free price of the contingent claim. In addition, quantities such as the local or stochastic volatility are not directly observable but have to be

[1] See, e.g., Andersen et al. (1999), p. 3, and Das & Sundaram (1999), p. 5.

filtered out either from pricing data on the underlying asset or "calibrated" to options data. In the first case, the quantity obtained is model-dependent and in the second case it is the solution to a non-trivial optimization problem.

A second strand of the literature identifies *market frictions* as another possible explanation for the smile pattern. Transaction costs, illiquidity, and other trading restrictions imply that a single arbitrage-free option price no longer exists. Longstaff (1995) and Figlewski (1989) examined the effects of transaction costs and found that they could be a major element in the divergences of implied volatilities across strike prices. Yet, Constantinides (1996) points out that transaction costs *cannot fully* explain the extent of the volatility smile.

The increasing liquidity in the market for standard options has had two major consequences:[2] First, there is no more need to theoretically price standard options. The market's liquidity ensures fair prices. Second, hedging of standard options becomes less important as positions can be unwound quickly. These developments and the inability of the models described above to accurately reflect the dynamic behavior of option prices or their implied volatilities have brought up a second modelling approach. In directly taking the implied volatility (surface) as *primitive*, this approach is usually referred to as a "*market-based*" approach.[3] Market-based models have the advantage to be automatically fitted to market option prices. In difference to fundamental quantities such as an (unobservable) instantaneous volatility, implied volatilities are continuously monitored by market participants. A market scenario described in terms of implied volatilities is therefore easy to understand for a practitioner.

Due to the noticeable standard deviation found in time series of implied volatilities, deterministic implied volatility models do not seem to be appropriate. Therefore, a straightforward generalization to stochastic implied volatility models is needed. In contrast to (traditional) stochastic volatility models, where the instantaneous volatility of the stock return is modelled, stochastic implied volatility models focus on the (stochastic) dynamics of either a single implied volatility (e.g., Lyons (1997)), the term structure of volatility (e.g., Schönbucher (1999)), or the whole volatility surface (e.g., Albanese et al. (1998) and Ledoit et al. (2002)). A major advantage of stochastic implied volatility models is their completeness. While the approaches of Schönbucher (1999) and Ledoit et al. (2002) dealt with the problem of stochastic implied volatility from a *theoretical perspective*, Rosenberg (2000), Cont & Fonseca (2002), Goncalves & Guidolin (2003), among others, focused on *empirical aspects* of the problem. For example, Cont & Fonseca (2002), using S&P 500 and FTSE 100 option data, suggest a factor-based stochastic implied volatility model where the *abstract* risk factors driving the volatility surface are obtained from a Karhunen Loève decomposition. A natural application of this model is the simulation of implied volatility surfaces under the real-world measure, for the purpose of risk management. However, the models are not intended to determine the consistent volatility drifts needed for risk-neutral pricing of exotic derivatives. "How best to

[2] See Schönbucher (1999).

[3] This approach is similar to the approach of Heath-Jarrow-Morton (HJM) in the field of interest rates. See Heath et al. (1992).

introduce the ideas from these models into a no-arbitrage theory remains an *open* question".[4] It has mainly been this question that motivated our work.

The *overall goal* of this work is to provide a stochastic implied volatility model that allows for *integrated* and *consistent* pricing and hedging, risk management, and trading of equity index derivatives as well as derivatives on index volatility. As we assume that the evolution of the volatility surface is driven by a small number of (fundamental or economic) risk factors, the model is termed "*factor-based*". Specifically, in the first, theoretical part of this work, we aim at developing a unifying *theory* for the analysis of contingent claims under both, the real-world measure *and* the risk-neutral measure, in a stochastic implied volatility environment. Based on the theory developed, the objective of the second, *empirical* part is to specify, estimate, and test a factor-based stochastic implied volatility model for DAX implied volatilities.[5]

The paper is organized as follows. Section 2 develops a general mathematical model of a financial market in continuous time where in addition to the usual underlying securities stock and risk-free bond, a collection of standard European options is traded. Section 3 investigates the properties of DAX implied volatilities, both in a cross-sectional ("structure") and a time-series ("dynamics") setting. In Section 4, we propose a four-factor model for the stochastic evolution of the DAX volatility surface. The paper concludes with a short summary.

2 The General Model

2.1 Definitions

In this section we repeat some basic definitions and results we use thereafter. We start with the Black-Scholes formula for European call options on non-dividend paying stocks: The arbitrage price C_t at time $t \in [0,T]$ of a European call option with strike price K and maturity date $T \leq T^*$ in the Black-Scholes market is given by the formula

$$C_t = C_{BS}\left(t, S_t\right), \qquad \forall t \in [0,T], \tag{1}$$

where $C_{BS} : [0,T] \times \mathbb{R}_+ \to \mathbb{R}$ is the Black-Scholes call option pricing formula

$$C_{BS}\left(t,s\right) = sN(d_1\left(t,s\right)) - Ke^{-r(T-t)}N(d_2\left(t,s\right)), \tag{2}$$

with

$$d_1(t,s) = \frac{\ln\left(\frac{s}{K}\right) + \left(r + \frac{1}{2}\upsilon^2\right)(T-t)}{\upsilon\sqrt{T-t}}, \qquad d_2(t,s) = d_1(t,s) - \upsilon\sqrt{T-t}, \tag{3}$$

and $N(\cdot)$ is the standard normal distribution function. The arbitrage price of a *standard European put option* can easily be obtained using the well known *put-call parity*. Denoting by $C_{BS} = C_{BS}(t,s,K,T,r,\upsilon)$ the Black-Scholes pricing function of a standard European

[4] See Lee (2002), p. 25.

[5] The DAX option is one of the most heavily traded equity index options in the world.

call, the most common Greeks are defined as

$$
\begin{aligned}
\delta_{BS} &= \frac{\partial C_{BS}}{\partial s} = N(d_1) > 0 \quad (\text{“Delta”}), \\
\Gamma_{BS} &= \frac{\partial^2 C_{BS}}{\partial s^2} = \frac{n(d_1)}{s\upsilon\sqrt{T-t}} > 0 \quad (\text{“Gamma”}), \\
\Lambda_{BS} &= \frac{\partial C_{BS}}{\partial \upsilon} = s\sqrt{T-t}\,n(d_1) > 0 \quad (\text{“Vega”}), \\
\Theta_{BS} &= \frac{\partial C_{BS}}{\partial t} = -\frac{s n(d_1)\upsilon}{2\sqrt{T-t}} - rKe^{-r(T-t)}N(d_2) < 0 \quad (\text{“Theta”}), \\
\frac{\partial C_{BS}}{\partial r} &= K(T-t)e^{-r(T-t)}N(d_2) < 0 \quad (\text{“Rho”}), \\
\mathcal{V}_{BS} &= \frac{\partial^2 C_{BS}}{\partial \upsilon^2} = s\sqrt{T-t}\,n(d_1)\frac{d_1 d_2}{\upsilon} \quad (\text{“DVegaDVol”}), \\
\Psi_{BS} &= \frac{\partial^2 C_{BS}}{\partial s \partial \upsilon} = -n(d_1) \cdot \frac{d_2}{\upsilon} \quad (\text{“DDeltaDVol”}),
\end{aligned}
\tag{4}
$$

where $d_1 = d_1(t,s,K,T,r,\upsilon) = \frac{\ln\left(\frac{s}{K}\right)+\left(r+\frac{1}{2}\upsilon^2\right)(T-t)}{\upsilon\sqrt{T-t}}$, $d_2 = d_2(t,s,K,T,r,\upsilon) = d_1 - \upsilon\sqrt{T-t}$ for $t \in [0,T]$.

If one may observe the market price of the option, then the volatility *implied* by the market price can be determined by inverting the option pricing formula. This volatility is known as the *implied volatility.*

Definition 1 (Implied Volatility) *Let $C_t(K,T)$ be the market price of a standard European call option with strike price $K > 0$ and maturity date T at time $t \in [0,T)$. The* (Black-Scholes) implied volatility *$\sigma_t(K,T)$ is then defined as the value of the volatility parameter which equates the market price of the option with the price given by the Black-Scholes formula (2)*:

$$C_t(K,T) = C_{BS}(t, S_t, K, T, r, \sigma_t(K,T)). \tag{5}$$

If the *Black-Scholes model* holds exactly, the volatility implied by an option's market price is widely regarded as the (subjective) *market's expectation of the future volatility υ* of the stock's continuously compounded returns.[6] If it *does not hold*, it is then just a *convenient and well-known mapping* from option prices to implied volatilities.

Volatility surfaces combine volatility smiles with the term structure of volatility to tabulate the implied volatilities appropriate for *market consistent pricing* of an option with any strike price and any maturity.

Definition 2 (Volatility Surface) *For any time $t \in [0,T^*]$, the function $\sigma_t : (0,\infty) \times (t,T^*] \to \mathbb{R}_+$, which assigns each strike price and maturity date tuple (K,T) its implied volatility $\sigma_t(K,T)$ is referred to as the* (implied) volatility surface (*or* volatility matrix).

Note, that for any fixed maturity date T, $T \le T^*$, the function $\sigma_t(K,\cdot)$ of implied volatility against strike price K, $K > 0$, is called the *(implied) volatility smile* or just *smile (*for maturity T*)* at date $t \in [0,T)$. In addition, for any fixed strike price K, $K > 0$, the function

6 See Christensen & Prabhala (1998), Figlewski (1997), and Mayhew (1995), among others.

$\sigma_t(\cdot, T)$ of implied volatility against maturity T, $T \leq T^*$, is called the *term structure of (implied) volatility* or *(implied) volatility term structure (*for strike K*)* at date $t \in [0, T)$.

Empirically it is often advantageous to reexpress the volatility surface in terms of moneyness and time to maturity, defined as $\tau = T - t$.

Definition 3 (Moneyness) *Let $m(t, s, K, T, r)$ be a function of time, underlying price, strike price, maturity date, and interest rate. Then the* moneyness M_t *at time $t \in [0, T^*]$ is generally defined as*

$$M_t = m(t, S_t, K, T, r). \tag{6}$$

The function m is referred to as the moneyness function. *It is required to be increasing in K.*

The choice of the adequate moneyness metric is mainly an empirical issue and depends strongly on the application under consideration. When expressed in terms of M and τ, we denote the volatility surface by $\widetilde{\sigma}_t(M, \tau)$, where the implied volatility function is defined as $\widetilde{\sigma}_t : \mathbb{R} \times (0, T^* - t] \to \mathbb{R}_+$. In contrast to the *absolute* volatility surface $\sigma_t(K, T)$, the function $\widetilde{\sigma}_t(M, \tau)$ is sometimes called the *relative* volatility surface. Note that there is a one-to-one correspondence between $\sigma_t(K, T)$ and $\widetilde{\sigma}_t(M, \tau)$ of the form

$$\widetilde{\sigma}_t(M, \tau) = \sigma_t(m^{-1}(M), t + \tau), \tag{7}$$

with $m^{-1}(M)$ being the inverse function of m with respect to M.

2.2 The Financial Market Model

MODEL SPECIFICATION

We consider a frictionless security market where investors are allowed to trade continuously up to some fixed finite planning horizon T^* (**Assumption 1**). The uncertainty in the financial market is characterized by the complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is the state space, $\mathcal{F}$ is the σ-algebra representing measurable events, and $\mathbb{P}$ is the objective probability measure. Information evolves over the trading interval $[0, T^*]$ according to the augmented, right continuous, complete filtration $\mathbb{F} = \{\mathcal{F}_t : t \in [0, T^*]\}$ generated by a $p+1$-dimensional standard Brownian motion $W = \{W_t : t \in [0, T^*]\}$, $W_t = (W_{0,t}, W_{1,t}, \ldots, W_{p,t})'$, initialized at zero. We assume that the σ-field $\mathcal{F}_0$ is trivial and contains all the $\mathbb{P}$-null sets of $\mathcal{F}$, and that $\mathcal{F}_{T^*} = \mathcal{F}$. The primary traded securities are a non-dividend paying stock[7], a money market account and a continuum of standard European call options on the stock. Under the objective probability measure $\mathbb{P}$, we make the following assumptions on the evolution of the money market account and the stock:

Assumption 2: *The price process of the money market account is given by the SDE*:

$$dB_t = rB_t dt, \qquad \forall t \in [0, T^*], \tag{8}$$

[7] All results can be easily extended to the case of a (even stochastic) dividend paying stock.

where $B_0 = 1$ and the interest rate r is supposed to be constant[8] *and nonnegative.*

Assumption 3: *The evolution of the stock price is governed by*

$$dS_t = S_t \mu_t dt + S_t \upsilon_t dW_{0,t}, \qquad \forall t \in [0, T^*], \tag{9}$$

with initial non-random stock price $S_0 > 0$. The drift process $\{\mu_t : t \in [0, T^*]\}$ is real-valued, progressively measurable and satisfies $\int_0^t |S_u \mu_u| \, du < \infty$ $\mathbb{P}$-a.s. for all $t \in [0, T^*]$. The volatility process $\{\upsilon_t : t \in [0, T^*]\}$ is supposed to be nonnegative, progressively measurable with $\int_0^t S_u^2 \upsilon_u^2 du < \infty$ $\mathbb{P}$-a.s. for all $t \in [0, T^*]$.

At time t the continuum of European call option prices $C_t(K, T)$ with strikes $K > 0$ and maturities $T > t$ can be represented by the volatility surface $\sigma_t(K, T)$. Because the volatility surface can be more easily parameterized and estimated as a function in moneyness M and time to maturity $\tau = T - t$, we rather consider the continuum of call prices $\widetilde{C}_t(M, \tau)$ which is represented by the *relative* volatility surface $\widetilde{\sigma}_t(M, \tau)$. There is a one-to-one correspondence between $\sigma_t(K, T)$ and $\widetilde{\sigma}_t(M, \tau)$ according to Equation 7. We demand the moneyness function and the corresponding moneyness to be *valid* according to the following definition:

Definition 4 (Valid Moneyness) *We call m a* valid moneyness function *and M defined as $M = M_t = m(t, S_t, K, T, r)$ a* valid moneyness *for our financial market model if m has the following properties:*

1. $m(t, s, K, T, r) \in C^2([0, T^*] \times \mathbb{R}_{++} \times \mathbb{R}_{++} \times (t, T^*] \times \mathbb{R}_+)$,
2. $\lim_{t \to T} m(t, S_t, K, T, r) < \infty$ $\mathbb{P}$-*a.s., and*
3. $\lim_{t \to T} \frac{\partial m}{\partial t}(t, S_t, K, T, r) < \infty$ $\mathbb{P}$-*a.s.*

Under the measure $\mathbb{P}$, we make the following assumption on the evolution of the relative volatility surface $\widetilde{\sigma}_t(M, \tau)$.

Assumption 4: *Let $g(t, M, \tau, y) \in C^2([0, T^*] \times \mathbb{R} \times (0, T^* - t] \times \mathbb{R}^p)$. The volatility surface $\widetilde{\sigma}_t(M, \tau)$ at time $t \in [0, T^*]$ is completely described by a p-dimensional vector of abstract risk factors $Y_t = (Y_{1,t}, Y_{2,t}, \ldots, Y_{p,t})$* such that:

$$\widetilde{\sigma}_t(M, \tau) = g(t, M, \tau, Y). \tag{10}$$

The dynamics of the i-th risk factor is modelled as

$$dY_{i,t} = \alpha_{i,t} dt + \sum_{j=0}^{p} \gamma_{i,j,t} dW_{j,t}, \qquad i = 1, \ldots, p, \ \forall t \in [0, T^*], \tag{11}$$

with initial value $Y_{i,0}$. For any $i = 1, \ldots, p$, $j = 0, \ldots, p$, the processes $\{\alpha_{i,t} : t \in [0, T^*]\}$ and $\{\gamma_{i,j,t} : t \in [0, T^*]\}$ are progressively measurable, and satisfy $\int_0^t |\alpha_{i,u}| \, du < \infty$ $\mathbb{P}$-a.s., $\int_0^t \gamma_{i,j,u}^2 du < \infty$ $\mathbb{P}$-a.s. for all $t \in [0, T^*]$.

[8] Our approach can be generalized to allow for stochastic interest rates by simply attaching a HJM-type model.

Movements of the Volatility Surface

The dynamics of an implied volatility of an option with moneyness M and time to maturity τ is determined by the dynamics of $Y = (Y_1, \ldots, Y_p)$. If we set $g(t, y) = g(t, M, \tau, y)$, then by application of Itô's rule to equation (10) we get

$$\begin{aligned} dg(t, Y_t) &= \frac{\partial g}{\partial t}(t, Y_t)\, dt + \sum_{i=1}^{p} \frac{\partial g}{\partial y_i}(t, Y_t)\, dY_{i,t} \\ &\quad + \frac{1}{2} \sum_{i=1}^{p} \sum_{k=1}^{p} \frac{\partial^2 g}{\partial y_i \partial y_k}(t, Y_t)\, d\langle Y_i, Y_k \rangle_t, \end{aligned} \tag{12}$$

with quadratic covariation $d\langle Y_i, Y_k \rangle_t = \sum_{j=0}^{p} \gamma_{i,j,t}\gamma_{k,j,t} dt$. Substituting (11) for $dY_{i,t}$ $(i = 1, \ldots, p)$ in (12) yields

$$\begin{aligned} dg(t, Y_t) &= \left\{ \frac{\partial g}{\partial t}(t, Y_t) + \sum_{i=1}^{p} \alpha_{i,t} \frac{\partial g}{\partial y_i}(t, Y_t) + \frac{1}{2} \sum_{i=1}^{p} \sum_{k=1}^{p} \sum_{j=0}^{p} \gamma_{i,j,t}\gamma_{k,j,t} \frac{\partial^2 g}{\partial y_i \partial y_k}(t, Y_t) \right\} dt \\ &\quad + \sum_{i=1}^{p} \sum_{j=0}^{p} \gamma_{i,j,t} \frac{\partial g}{\partial y_i}(t, Y_t)\, dW_{j,t}. \end{aligned} \tag{13}$$

Using (10), the dynamics of the implied volatilities of each fixed time to maturity τ and moneyness M can be expressed as

$$d\widetilde{\sigma}_t(M, \tau) = \widetilde{\eta}_t(M, \tau)dt + \widetilde{\vartheta}_t(M, \tau)dW_t, \tag{14}$$

where $\widetilde{\eta}_t = \widetilde{\eta}_t(M, \tau)$ is given by

$$\widetilde{\eta}_t = \frac{\partial g}{\partial t}(t, Y_t) + \sum_{i=1}^{p} \alpha_{i,t} \frac{\partial g}{\partial y_i}(t, Y_t) + \frac{1}{2} \sum_{i=1}^{p} \sum_{k=1}^{p} \sum_{j=0}^{p} \gamma_{i,j,t}\gamma_{k,j,t} \frac{\partial^2 g}{\partial y_i \partial y_k}(t, Y_t), \tag{15}$$

and $\widetilde{\vartheta}_t = \widetilde{\vartheta}_t(M, \tau)$ is the $(p+1)$-dimensional row vector

$$\widetilde{\vartheta}_t = \left(\widetilde{\vartheta}_{0,t}, \ldots, \widetilde{\vartheta}_{p,t} \right), \tag{16}$$

with

$$\widetilde{\vartheta}_{j,t} = \sum_{i=1}^{p} \gamma_{i,j,t} \frac{\partial g}{\partial y_i}(t, Y_t), \qquad j = 0, \ldots, p. \tag{17}$$

2.3 Risk-Neutral Implied Volatility Dynamics

Change of Measure and Drift Restriction

If we define $W_t^* = (W_{0,t}^*, \ldots, W_{p,t}^*)'$ for all $t \in [0, T^*]$, with dynamics given by

$$dW_t^* = dW_t + \psi_t dt, \tag{18}$$

where $\psi_t = (\psi_{0,t}, \ldots, \psi_{p,t})'$ is a $p+1$-dimensional progressively measurable process satisfying

$$\int_0^{T^*} \psi_t' \psi_t dt < \infty \qquad \mathbb{P}\text{-a.s.} \tag{19}$$

as well as Novikov's condition, i.e.

$$\mathbb{E}_{\mathbb{P}}\left[\exp\left\{\frac{1}{2}\int_0^{T^*}\psi_t^{'}\psi_t dt\right\}\right] < \infty, \tag{20}$$

then, by the virtue of Girsanov's theorem, under the measure $\mathbb{Q}$ with Radon-Nikodým derivative

$$\mathcal{E}_t = \left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{F}_t} = \exp\left(-\int_0^t \psi_s^{'} dW_s - \frac{1}{2}\int_0^t \psi_s^{'}\psi_s ds\right), \tag{21}$$

the process W_t^* is a multi-dimensional standard Brownian motion. The process ψ_i is interpreted as the *market price of risk* (process) associated with the random factor W_i ($i = 0, \ldots, p$).

Choosing

$$\psi_{0,t} = \frac{\mu_t - r}{\upsilon_t}, \qquad \forall t \in [0, T^*], \tag{22}$$

it is easy to prove that under the measure $\mathbb{Q}$ the discounted stock price $S_t^* = S_t/B_t$ is a martingale: $dS_t^* = S_t^* \upsilon_t dW_{0,t}^*$.

As in our model option contracts are primary traded assets as well, we have to show that under the measure $\mathbb{Q}$ the discounted option price processes are martingales, i.e. we need to look at the dynamics of the option price with fixed maturity date T and with fixed strike price K. Therefore, we have to determine the dynamics of the implied volatility $\sigma_t(K,T)$ of this option under the measure $\mathbb{P}$, given the dynamics of $\widetilde{\sigma}_t(M,\tau)$. Setting $g(t,M,\tau) = g(t,M,\tau,y)$, for convenience, we proceed in three steps:

1. For fixed T we make the second argument of the implied volatility process $\widetilde{\sigma}_t(M,\tau)$ change deterministically with time, $\tau(t) = T - t$. Then, by application of Itô's lemma,

$$\begin{aligned} d\widetilde{\sigma}_t(M, T-t) &= \left(\widetilde{\eta}_t\,(M, T-t) - \frac{\partial g}{\partial \tau}\,(t, M, T-t)\right) dt \\ &\quad + \widetilde{\vartheta}_t(M, T-t) dW_t. \end{aligned} \tag{23}$$

2. Using Itô's lemma, we get the dynamics of the moneyness with respect to the stock price as

$$\begin{aligned} dM_t(S_t) &= \frac{\partial m}{\partial t}(t, S_t)\,dt + \frac{\partial m}{\partial s}(t, S_t)\,dS_t + \frac{1}{2}\frac{\partial^2 m}{\partial s^2}(t, S_t)\,d\langle S, S\rangle_t \\ &= \left(\frac{\partial m}{\partial t}(t, S_t) + S_t\mu_t\frac{\partial m}{\partial s}(t, S_t) + \frac{1}{2}S_t^2\upsilon_t^2\frac{\partial^2 m}{\partial s^2}(t, S_t)\right) dt \\ &\quad + S_t\upsilon_t\frac{\partial m}{\partial s}(t, S_t)\,dW_{0,t}. \end{aligned} \tag{24}$$

3. Finally, we use a generalization of the Itô-Venttsel formula to the processes given by equations (23) and (24) to get

$$d\sigma_t(K,T) = d\widetilde{\sigma}_t(M_t, T-t) = \eta_t(K,T)\,dt + \vartheta_t(K,T)\,dW_t,$$

where

$$\eta_t(K,T) = \Bigg(\widetilde{\eta}_t(M_t, T-t) - \frac{\partial g}{\partial \tau}(t, M_t, T-t) \tag{26}$$

$$+\frac{1}{2}\left(S_t \upsilon_t \frac{\partial m}{\partial s}(t,S_t)\right)^2 \frac{\partial^2 g}{\partial^2 M}(t, M_t, T-t)$$

$$+\left(\frac{\partial m}{\partial t}(t,S_t) + S_t \mu_t \frac{\partial m}{\partial s}(t,S_t) + \frac{1}{2} S_t^2 \upsilon_t^2 \frac{\partial^2 m}{\partial s^2}(t,S_t)\right) \frac{\partial g}{\partial M}(t, M_t, T-t)$$

$$+S_t \upsilon_t \frac{\partial m}{\partial s}(t,S_t) \frac{\partial u}{\partial M}(t, M_t, T-t)\Bigg),$$

$$\vartheta_{0,t}(K,T) = \widetilde{\vartheta}_{0,t}(M_t, T-t) + S_t \upsilon_t \frac{\partial m}{\partial s}(t,S_t) \frac{\partial g}{\partial M}(t, M_t, T-t), \tag{27}$$

and

$$\vartheta_{i,t}(K,T) = \widetilde{\vartheta}_{i,t}(M_t, T-t), \qquad i = 1,\ldots,p, \tag{28}$$

and u is the deterministic function corresponding to $\widetilde{\vartheta}_{0,t}$.

Evaluating the Black-Scholes Greek functions at $\upsilon = \sigma_t(K,T)$ we define

$$\begin{aligned}
\delta_t &= \delta_{BS}(t, S_t, K, T, r, \sigma_t(K,T)), \ \Gamma_t = \Gamma_{BS}(t, S_t, K, T, r, \sigma_t(K,T)), \\
\Lambda_t &= \Lambda_{BS}(t, S_t, K, T, r, \sigma_t(K,T)), \ \Theta_t = \Theta_{BS}(t, S_t, K, T, r, \sigma_t(K,T)), \\
\mathcal{V}_t &= \mathcal{V}_{BS}(t, S_t, K, T, r, \sigma_t(K,T)), \ \Psi_t = \Psi_{BS}(t, S_t, K, T, r, \sigma_t(K,T)),
\end{aligned} \tag{29}$$

and by applying Itô's lemma to the Black-Scholes call option pricing formula (2), we find the dynamics of the call prices $C_t(K,T)$ under $\mathbb{P}$ to be

$$\begin{aligned}
dC_t(K,T) &= \Bigg(\Theta_t + \delta_t S_t \mu_t + \Lambda_t \eta_t(K,T) + \frac{1}{2}\Gamma_t S_t^2 \upsilon_t^2 \\
&\quad + \frac{1}{2}\mathcal{V}_t \vartheta_t(K,T)\, \vartheta_t^{'}(K,T) + \Psi_l S_t \upsilon_t \vartheta_{0,t}(K,T)\Bigg) dt \\
&\quad + \breve{\vartheta}_t(K,T)\, dW_t,
\end{aligned} \tag{30}$$

where

$$\breve{\vartheta}_{0,t}(K,T) = \delta_t S_t \upsilon_t + \Lambda_t \vartheta_{0,t}(K,T), \tag{31}$$

and

$$\breve{\vartheta}_{i,t}(K,T) = \Lambda_t \vartheta_{i,t}(K,T), \qquad i = 1,\ldots,p. \tag{32}$$

We choose $\psi_{i,t}$, $i = 1,\ldots,p$, such that the *drift* or *no-arbitrage condition*

$$\begin{aligned}
\vartheta_t(K,T)\psi_t &= \eta_t(K,T) + \frac{1}{2}\frac{d_1 d_2}{\sigma_t(K,T)} \vartheta_t(K,T)\, \vartheta_t^{'}(K,T) \\
&\quad + \frac{1}{2}\frac{1}{\sigma_t(K,T)\,(T-t)}\left(\upsilon_t^2 - \sigma_t^2(K,T)\right) \\
&\quad - \frac{d_2}{\sigma_t(K,T)\sqrt{T-t}} \upsilon_t \vartheta_{0,t}(K,T),
\end{aligned} \tag{33}$$

with[9]

$$d_1 = d_1(t, S_t, K, T, r, \sigma_t(K,T)), \ d_2 = d_2(t, S_t, K, T, r, \sigma_t(K,T)),$$

holds for all $t \in [0, T^*]$ and all (K,T).[10] Then, under this measure, the dynamics of $C_t(K,T)$

[9] For a definition of $d_1(\cdot)$ and $d_2(\cdot)$, see equation (3).

[10] A similar condition that restricts the drift coefficients of the instantaneous forward rates was derived by Heath et al. (1992).

are given by

$$\begin{aligned} dC_t(K,T) &= \Bigg(\Theta_t + \delta_t S_t \mu_t + \Lambda_t \eta_t(K,T) + \frac{1}{2}\Gamma_t S_t^2 \upsilon_t^2 \\ &\quad + \frac{1}{2}\mathcal{V}_t \vartheta_t(K,T)\, \vartheta_t^{'}(K,T) + \Psi_t S_t \upsilon_t \vartheta_{0,t}(K,T) - \breve{\vartheta}_t(K,T)\, \psi_t\Bigg)\, dt \\ &\quad + \breve{\vartheta}_t(K,T)\, dW_t^*. \end{aligned} \tag{34}$$

Discounting and using equation (33) shows that the relative call prices $C_t^*(K,T) = C_t(K,T)\,/B_t$ are indeed martingales:

$$dC_t^*(K,T) = -rC_t^*(K,T)\, dt + \frac{1}{B_t} dC_t(K,T) = \frac{1}{B_t}\breve{\vartheta}_t(K,T)\, dW_t^*. \tag{35}$$

Note that equation (35) holds because the Black-Scholes PDE

$$\frac{\Theta_t}{\Lambda_t} + \frac{1}{2}\frac{\Gamma_t}{\Lambda_t} S_t^2 \sigma_t^2(K,T) + \frac{\delta_t}{\Lambda_t} S_t r - \frac{r}{\Lambda_t} C_t(K,T) = 0. \tag{36}$$

is satisfied for $C_t(K,T)$ when the stock price volatility υ_t equals the option's implied volatility $\sigma_t(K,T)$. Finally, the dynamics of $\sigma_t(K,T)$ and $\widetilde{\sigma}_t(M,\tau)$ under $\mathbb{Q}$ are:

$$\begin{aligned} d\sigma_t(K,T) &= (\eta_t(K,T) - \vartheta_t(K,T)\, \psi_t)\, dt + \vartheta_t(K,T)\, dW_t^* \\ &= \eta_t^*(K,T)\, dt + \vartheta_t(K,T)\, dW_t^*, \end{aligned} \tag{37}$$

where

$$\begin{aligned} \eta_t^*(K,T) &= \Bigg(-\frac{1}{2}\frac{d_1 d_2}{\sigma_t(K,T)}\vartheta_t(K,T)\, \vartheta_t^{'}(K,T) \\ &\quad -\frac{1}{2}\frac{1}{\sigma_t(K,T)\,(T-t)}\left(\upsilon_t^2 - \sigma_t^2(K,T)\right) \\ &\quad + \frac{d_2}{\sigma_t(K,T)\,\sqrt{T-t}}\upsilon_t \vartheta_{0,t}(K,T)\Bigg), \end{aligned} \tag{38}$$

and

$$d\widetilde{\sigma}_t(M,\tau) = \left(\widetilde{\eta}_t(M,\tau) - \widetilde{\vartheta}_t(M,\tau)\, \psi_t\right) dt + \widetilde{\vartheta}_t(M,\tau)\, dW_t^*. \tag{39}$$

For an interpretation of each of the three terms in the risk-neutral drift $\eta_t^*(K,T)$ of $\sigma_t(K,T)$ given in (38) see, e.g., Albanese et al. (1998).

Existence and Uniqueness of the Risk-Neutral Measure

In Section 2.3, equation (33), we have implicitly assumed that we can always find a market price of risk process such that equation (33) holds for all traded options at all times. However, if there is a continuum of traded options, equation (33) can only be satisfied if we can define $\psi_{i,t}$, $i = 1, \ldots, p$, independently of K and T. In general, this might be impossible. If we consider exactly p traded options we can expect the implicit function to be uniquely defined by equation (33). Fix $(K_1,T_1), \ldots, (K_p,T_p)$ such that $(K_i,T_i) \neq (K_j,T_j)$ for all $i \neq j$, and $0 < T_1 < \ldots < T_p \leq T^*$. There exists an equivalent probability measure denoted by $\mathbb{Q}_{(K_1,T_1),\ldots,(K_p,T_p)}$ such that $C_t^*(K_1,T_1), \ldots, C_t^*(K_p,T_p)$ are martingales if and only if equation (33) holds for $(K_1, T_1), \ldots, (K_p, T_p)$ and for all $t \in [0, T_1)$. Depending on the specific choice of the processes ϑ, η, and υ the right-hand side of equation (33) might explode $\mathbb{P}$-a.s. and $\vartheta_t < \infty$ $\mathbb{P}$-a.s. if t tends to T_1.[11] In this case it is impossible to find a market price of risk process ψ such that $\int_0^t \psi_u' \psi_u du < \infty$ $\mathbb{P}$-a.s. and thus an equivalent martingale measure does not exist.[12] Nevertheless, if we can find an equivalent martingale measure for the call options $C_t(K_1, T_1), \ldots, C_t(K_p, T_p)$, the measure is unique if and only if

$$\begin{pmatrix} \vartheta_{1,t}(K_1, T_1) & \cdots & \vartheta_{p,t}(K_1, T_1) \\ \vdots & \cdots & \vdots \\ \vartheta_{1,t}(K_p, T_p) & \cdots & \vartheta_{p,t}(K_p, T_p) \end{pmatrix} \tag{40}$$

is nonsingular $\mathbb{P}$-a.s. Both the market prices of risk and the martingale measure, however, depend on the particular call options chosen. To guarantee that there exists a unique equivalent martingale measure $\mathbb{Q}$ simultaneously making all relative call option prices martingales, the market prices of risk must be independent of the vector of calls $(C_t(K_1, T_1), \ldots, C_t(K_p, T_p))$ chosen. Formally, the following conditions are equivalent:[13]

- $\mathbb{Q}$ defined by $\mathbb{Q} = \mathbb{Q}_{(K_1,T_1),\ldots,(K_p,T_p)}$ for any $(K_1, T_1), \ldots, (K_p, T_p) \in \mathbb{R}_{++} \times (0, T^*]$, $(K_i, T_i) \neq (K_j, T_j)$ $\forall i \neq j$, $0 < T_1 < \ldots < T_p \leq T^*$, is the unique equivalent probability measure such that $C_t^*(K, T) = C_t(K, T) / B_t$ is a martingale for all $T \in (0, T^*]$, $K \in \mathbb{R}_{++}$, $t \in [0, T_1)$.

- $\psi_{i,t}((K_1, T_1), \ldots, (K_p, T_p)) = \psi_{i,t}((K_{p+1}, T_{p+1}), \ldots, (K_{2p}, T_{2p}))$ for $i = 1, \ldots, p$ and for all $(K_1, T_1), \ldots, (K_p, T_p), (K_{p+1}, T_{p+1}), \ldots, (K_{2p}, T_{2p}) \in \mathbb{R}_{++} \times (0, T^*]$, $(K_i, T_i) \neq (K_j, T_j)$ $\forall$ $i \neq j$, $i, j \in \{1, \ldots, p\}$, $(K_i, T_i) \neq (K_j, T_j)$ $\forall$ $i \neq j$, $i, j \in \{p+1, \ldots, 2p\}$, $t \in [0, T^*]$ such that $0 \leq t < T_1 < \ldots < T_p \leq T^*$ and $0 \leq t < T_{p+1} < \ldots < T_{2p} \leq T^*$.

- For all $T \in (0, T^*]$ and $t \in [0, T)$

$$\begin{aligned} \vartheta_t(K, T)\, \psi_t &= \eta_t(K, T) + \frac{1}{2} \frac{d_1 d_2}{\sigma_t(K, T)} \vartheta_t(K, T)\, \vartheta_t'(K, T) \\ &\quad + \frac{1}{2} \frac{1}{\sigma_t(K, T)\,(T - t)} \left(\upsilon_t^2 - \sigma_t^2(K, T) \right) \\ &\quad - \frac{d_2}{\sigma_t(K, T) \sqrt{T - t}} \upsilon_t \vartheta_{0,t}(K, T), \end{aligned}$$

  where $\psi_{0,t} = \frac{\mu_t - r}{\upsilon_t}$ and $\psi_{i,t} = \psi_{i,t}((K_1, T_1), \ldots, (K_p, T_p))$ for any maturities $T_1, \ldots, T_p \in (t, T^*]$ and times $t \in [0, T_1)$.

In the following, we assume that there exists a unique equivalent martingale measure $\mathbb{Q}$.

[11] Similar problems occur in the HJM framework. See Filipovic (2000), for details.

[12] A similar problem also occurs if one specifies the implied volatility dynamics already under the risk neutral measure. See Schönbucher (1999).

[13] These conditions are very similar to those in the HJM model framework. See in particular, Proposition 3, p. 86. For this reason, we omit the proof.

2.4 Pricing and Hedging of Contingent Claims

This section demonstrates how to value and hedge contingent claims in the economy decribed above. For the valuation of a general stock price dependent claim the following result is central:[14]

Theorem 5 *The (Black-Scholes) ATM implied volatility converges to the stock price volatility when the time to maturity goes to zero, i.e. for* $K = S_t e^{r(T-t)}$ *the following is true:*

$$\upsilon_t = \lim_{T \to t} \widetilde{\sigma}_t \Big(m(t, S_t, S_t e^{r(T-t)}, T, r), T - t \Big) = \widetilde{\sigma}_t(m(t, S_t, S_t, t, r), 0)\,. \tag{41}$$

Proof. See appendix A. ■

Note that this result holds under both the objective measure $\mathbb{P}$ and the equivalent martingale measure $\mathbb{Q}$. Using (39), (41), and the general risk-neutral valuation formula, the arbitrage price process of a contingent claim in the (general) factor-based stochastic implied volatility model is given by the following theorem:

Theorem 6 *The arbitrage price* $\Pi_t(H)$ *at time* $t \in [0, T]$ *of any* $\mathbb{Q}$*-attainable claim* H *with maturity date* T *in the (general) factor-based stochastic implied volatility model is given by the formula*

$$\Pi_t(H) = \mathbb{E}_{\mathbb{Q}} \left[e^{-r(T-t)} H \Big| \mathcal{F}_t \right], \qquad \forall t \in [0, T]\,, \tag{42}$$

where the expectation is taken with respect to the joint diffusion of the stock price and its volatility under the measure $\mathbb{Q}$ *:*

$$dS_t = S_t r dt + S_t \upsilon_t dW^*_{0,t}, \qquad \forall t \in [0, T]\,, \tag{43}$$

and

$$\begin{aligned} d\upsilon_t \;\; = \;\; & \Big(\widetilde{\eta}_t(m(t, S_t, S_t, t, r), 0) - \widetilde{\vartheta}_t(m(t, S_t, S_t, t, r), 0)\, \psi_t \Big)\, dt \\ & + \widetilde{\vartheta}_t(m(t, S_t, S_t, t, r), 0)\, dW^*_t, \end{aligned} \tag{44}$$

for every $t \in [0, T]$ *. The processes* $\widetilde{\eta}_t(M, \tau)$ *and* $\widetilde{\vartheta}_t(M, \tau)$ *are defined by (15), (16) and (17).*

If an analytic solution is not available, standard Monte Carlo simulation techniques can be applied to approximate the expectation in (42). For *hedging* standard European call and put options we can use the Black-Scholes Greeks as defined in equation (29).

3 Properties of DAX Implied Volatilities

3.1 Data

Our database contains all reported transactions of options and futures on the DAX, traded on the DTB/Eurex over the sample period from January 1995 to December 2002.[15] The

[14] A similar result was derived by Ledoit et al. (2002).

[15] We are grateful to Eurex Deutschland for providing us with these data.

underlying of the option, the DAX index, comprises the 30 largest and most actively traded German companies. The DAX is a capital-weighted performance index, i.e. dividends are reinvested. DAX options are cash settled European-style options which expire on the third Friday of the contract month. At any point in time eight option maturities with lifetimes of up to two years are available: the three nearest calendar months, the three following months of the cycle March-June-September-December and the two following months of the cycle June-December. With more than 44 million traded contracts during 2002, the DAX option is the most liquid index option on Eurex and ranks among the top index options contracts in the world.[16] The futures contract on the DAX index is clearly associated with the option contract, nevertheless, one difference can be noted: the expiry months are only the three nearest months within the cycle March-June-September-December. Trading hours changed several times from 1995 to now, but DAX option and DAX future were traded at least from 9:30 a.m. to 4:00 p.m.

To compute the implied volatility for each options trade, we apply the standard Black-Scholes option pricing formula. Apart from the option price and the strike, three parameters are required to compute the implied volatilities: the time to expiration, the risk-free rate and the level of the underlying index. Let $n \in \{0, \ldots, N\}$, $N = 2009$, denote the trading day and t_n the corresponding trading date.[17] The option's expiration date is symbolized by T_O. The time to expiration $(T_O - t_n)$ is measured in calendar days.[18] Daily series of 1, 3, 6, and 12 months DM-LIBOR rates for the period from 1995 to 1998 and EURIBOR rates for the period from 1999-2002 serve as riskless interest rates r. The $(T_O - t_n)$-period interest rate is obtained by linear interpolation between the available rates enclosing $(T_O - t_n)$. The resulting value is then converted to a continously compounded rate.[19]

Let l $(l = 0, \ldots, L)$ be the trading minute of an options transaction.[20] The underlying index $S_{n,l}$ on day n at minute l is derived from the current price $F_{n,l}$ of the futures contract most actively traded on that day.[21] The maturity of this contract, which is normally the nearest available, is denoted by T_F. The value $F_{n,l}(T_F)$ corresponds to the *average* transaction price observed in the T_F-futures contract in minute l on day n. To obtain the corresponding index level we solve the theoretical futures pricing model

$$F_{n,l}\left(T_F\right) = S_{n,l} e^{r(T_F - t_n)} \tag{45}$$

[16] See FIA (2003).

[17] For our convenience, we call N the *sample size*, though N is in fact the sample size minus one.

[18] It is uncertain, whether volatility is related to trading or calendar days. The difference between calendar and trading days, expressed as a fraction of one year, is small except for very short-term options. These are not considered in this paper. See also Hull (2000). In the calculations the time to maturity is measured as a proportion of 365 days per year.

[19] The riskless rate r is dependent on day t_n and the investment horizon T_O. For ease of exposition, we suppress these indices.

[20] Since trading hours changed through time, L is time-dependent. To keep notation simple, we suppress the time index.

[21] We use $S_{n,l}$ as a shortcut for $S_{t_{n,l}}$. The same notational principle applies to forward prices, implied volatilities, etc., in discrete time.

for $S_{n,l}$.[22] If no future is traded at minute l, we exclude all options transactions that took place in this minute from our database. This procedure ensures simultaneous options and underlying prices, i.e. their respective time stamps diverge by not more than one minute.

Put-call-parity requires that the implied call volatilities do not systematically deviate from the implied put volatilities with the same degree of moneyness. However, on a number of trading days systematic violations of this relationship were observed. After a closer inspection, the problem could be traced back to a biased index level caused by dividend payments. The DAX index calculation rests on the assumption that cash dividends are reinvested after deduction of the corporate income tax for distributed gains from the gross dividend If the marginal investor's tax rate is smaller than the corporate income tax rate, he receives an extra dividend, here called difference dividend, which has the same effect as an ordinary dividend in the case of unprotected options and futures. To estimate the difference dividend, we use the method proposed by Hafner & Wallmeier (2001). This method relies on the assumption that put-call-parity holds. In fact, this a very weak assumption as put-call-parity does not depend on a particular stock price model, but only postulates the absence of arbitrage opportunities. More specifically, on each day n we select matching pairs of call and put options and solve the put-call parity for the difference dividend. This estimate is then used to recalculate the index levels and implied volatilities of all options transactions on that day. The procedure is repeated for each day in the sample. An inspection of all scatterplots reveals that this implied estimation of the relevant underlying index level solves the problem of the difference dividend. To correct for outliers, we finally eliminate all options that have implied volatilities higher than 150%.

At the DAX options market, liquidity is very much concentrated in short-term options and declines exponentially with increasing time to expiration. Of the total number of 6,171,700 options (transactions), 87.44% expire within the next 90 calendar days and 95.19% within the next 180 calendar days. The call trades distribution across strike prices is clearly skewed to the left whereas the put trades distribution is skewed to the right. This means that OTM options are traded far more frequently than ITM options. Since the estimation of the volatility surface requires a sufficient variety of strike prices, we include both calls and puts in our empirical study.

3.2 Structure of DAX Implied Volatilities

Estimation of the DAX Volatility Surface

For any *fixed time* t, we aim at finding a function $\widehat{g}(t, M, \tau, y_1, y_2, \ldots, y_p)$ and (abstract) risk factors $Y_1, Y_2, \ldots, Y_p$ such that the estimated implied volatilities are as close as possible to the observed implied volatilities. Because the relationship between the implied volatility

[22] Using the futures-based implied index level rather than the reported index level as the underlying price has also been suggested in a study for the S&P 500 options market by Jackwerth & Rubinstein (1996), p. 1616.

$\sigma_t(M,\tau)$, the moneyness M and the time to maturity $\tau = T - t$ is supposed to vary through time, we include only one trading day's data in any *cross-sectional analysis* and estimate the DAX volatility surface $\widetilde{\sigma}_n\,(M,\tau)$ separately for each day $n \in \{0,\ldots,N\}$.

Ait-Sahalia & Lo (1998), Cont & Fonseca (2002), and Fengler et al. (2000), among others, use *nonparametric methods* (e.g., kernel regressions) to estimate the daily volatility surface.[23] However, these methods tend to be very data intensive. Also, they tend to fill in missing data in a non-intuitive way. For instance, in a downward sloping implied volatility smile where a range of strike prices has no observations, we would find question-mark shaped segments. This could create artificial arbitrage opportunities.[24] For these difficulties, we employ a *parametric* approach and model the DAX *volatility surface* for an arbitrary but fixed day $n \in \{0,\ldots,N\}$ across degrees of moneyness M and times to maturity τ by the regression model

$$\widetilde{\sigma}_{n,j} = \widehat{g}(M_{n,j},\tau_{n,j},\beta_n) + \varepsilon_{n,j}, \qquad j = 1,\ldots,J_n \tag{46}$$

where $\widehat{g}(\cdot)$ is a regression function, β_n is a vector of regression coefficients, ε_n is a random disturbance, and j indexes the J_n sample observations on day n.[25] The *estimation procedure* involves three steps:

1. Specify the moneyness measure M.
2. Specify the regression model and choose an appropriate estimation method.
3. Test on violations of general no-arbitrage relations.

Subsequently, we discuss each step in some detail.[26]

In the literature, different *moneyness* measures have been proposed that explicitely account for time to maturity (*Step 1*). For example, Natenberg (1994) has suggested the measure $\ln\left(\frac{K}{F_t(T)}\right)/\sqrt{T-t}$. Other measures that fall into this category are the Tompkins (1994) moneyness and the option's Black-Scholes delta. Although the dependence on time to maturity is in general a desirable property of each moneyness measure, it can easily be shown that these measures are not *valid* in the sense of Definition 4. Considering further alternatives, one of the most common moneyness measures is the *log simple moneyness*[27], defined by

$$M = M_t = \ln\left(\frac{K}{S_t e^{r(T-t)}}\right) = \ln\left(\frac{K}{F_t(T)}\right), \qquad \forall t \in [0,T^*] \tag{47}$$

for all $S_t > 0$, $T \in (t,T^*]$, $r \geq 0$, and all $K > 0$.[28] As can easily be verified, the log simple moneyness (function) meets the requirements of a *valid* moneyness (function). When

[23] See Ait-Sahalia & Lo (1998), Cont & Fonseca (2002), and Fengler et al. (2000).

[24] See Jackwerth & Rubinstein (1996).

[25] Note that $\widetilde{\sigma}_{n,i} = \sigma_{n,i}$, by definition.

[26] Note, that step 1 and step 2 are not independent.

[27] Another very common measure is the simple moneyness, defined as the ratio of strike price to stock price or the inverse of it. It is used, e.g., by Cont & Fonseca (2002) and Ledoit et al. (2002).

[28] For a theoretical and empirical justification to define moneyness with respect to the forward price rather than the spot price, see Natenberg (1994), pp. 106-110 and Roth (1997). It should be noted, however, that under normal interest rate conditions, there is not much difference between the "spot moneyness" and the "forward moneyness", at least not for short-term options.

expressed in terms of log simple moneyness, an option is said to be ATM, if $M = 0$. A call (put) is said to be ITM (OTM) for $M < 0$ and OTM (ITM) for $M > 0$. This allows for an easy interpretation of the moneyness measure as the amount a put option is ITM or a call option is OTM as a proportion of the forward price.[29] In all further calculations, we use the log simple moneyness, often just called "the moneyness". Given our sample data, we assign each options trade its log simple moneyness. Specifically, for an option expiring in T_O, we compute

$$M_{n,l} = \ln\left(\frac{K}{F_{n,l}(T_O)}\right),$$

where $F_{n,l}\,(T_O)$ is the theoretical T_O-futures price computed on the basis of the adjusted index level.

The *second step* specifies the regression model. We choose the form:

$$\begin{aligned}\widetilde{\sigma}_{n,j} \quad &= \quad \beta_{1,n} + \beta_{2,n} M_{n,j} + \beta_{3,n} M_{n,j}^2 + \beta_{4,n} \ln(1+\tau_{n,j}) \\ &\quad + \beta_{5,n} M_{n,j} \ln(1+\tau_{n,j}) + \beta_{6,n} M_{n,j}^2 \ln(1+\tau_{n,j}) + \varepsilon_{n,j},\end{aligned} \tag{48}$$

where $n \in \{0, \ldots, N\}$ and $j = 1, \ldots, J_n$. This model allows the slope and the curvature of the volatility smile to vary with time to maturity. Hence, the model should be able to reproduce the "*flattening out*" effect that is commonly observed. This effect refers to the fact that the volatility smile becomes flatter when the time to maturity increases.[30] The volatility term structure in model (48) is represented by the function $\ln(1+\tau)$. This is necessary because the implied volatility processes (expressed in terms of K and T) do not exist if the volatility term structure is modelled by the more common square root function.

Although the model is expected to capture most of the cross-sectional variations in implied volatilities, it might yet not be well specified as the variables M and $M\ln(1+\tau)$, on the one hand, and M^2 and $M^2\ln(1+\tau)$, on the other hand, and hence also β_2 and β_5, and β_3 and β_6, are highly correlated. When we hypothesize that these relationships are stable over time, the model can be simplified by imposing the following *restrictions* on the regression coefficients:

$$\beta_{5,n} = \varrho_1 \beta_{2,n}, \qquad n = 0, \ldots, N, \tag{49}$$

and

$$\beta_{6,n} = \varrho_2 \beta_{3,n}, \qquad n = 0, \ldots, N, \tag{50}$$

where ϱ_1 and ϱ_2 are known constants. Substituting (49) and (50) into (48) leads to the following *unrestricted* regression model:

$$\begin{aligned}\widetilde{\sigma}_{n,j} \quad &= \quad \beta_{1,n} + \beta_{2,n} M_{n,j}\left(1 + \varrho_1 \ln\left(1+\tau_{n,j}\right)\right) \\ &\quad + \beta_{3,n} M_{n,j}^2 \left(1 + \varrho_2 \ln\left(1+\tau_{n,j}\right)\right) + \beta_{4,n} \ln\left(1+\tau_{n,j}\right) + \varepsilon_{n,j}.\end{aligned} \tag{51}$$

In practice, ϱ_1 and ϱ_2 are not known but have to be estimated. Consequently, it is not possible to apply the least squares method directly to (51). Instead, we propose a *two-step estimation*

[29] This interpretation results from a first order Taylor series expansion of the log moneyness: $\ln(K/F_t(T)) \approx K/F_t(T) - 1$.

[30] See also Das & Sundaram (1999).

procedure.[31] In the *first step*, we estimate the *original* regression model (48) for all days $n = 0, \ldots, N$. From the obtained time series of regression coefficients $\widehat{\beta}_{i,n}$ $(i = 2, 3, 5, 6)$, we then estimate ϱ_1 and ϱ_2 by fitting the *auxiliary* regression models

$$\widehat{\beta}_{5,n} = \varrho_1 \widehat{\beta}_{2,n} + \epsilon_n^{(1)}, \qquad n = 0, \ldots, N, \tag{52}$$

and

$$\widehat{\beta}_{6,n} = \varrho_2 \widehat{\beta}_{3,n} + \epsilon_n^{(2)}, \qquad n = 0, \ldots, N, \tag{53}$$

with $\epsilon^{(1)}$ and $\epsilon^{(2)}$ being random disturbances. The final estimates of the regression coefficients $\beta_1, \ldots, \beta_4$ are obtained in the *second step* by applying the least squares method to model (51) where the unknown constants ϱ_1 and ϱ_2 are replaced by their estimates $\widehat{\varrho}_1$ and $\widehat{\varrho}_2$, i.e. to:

$$\begin{aligned} \widetilde{\sigma}_{n,j} \;\; = \;\; & \beta_{1,n} + \beta_{2,n} M_{n,j} \left(1 + \widehat{\varrho}_1 \ln\left(1 + \tau_{n,j}\right)\right) \\ & + \beta_{3,n} M_{n,j}^2 (1 + \widehat{\varrho}_2 \ln\left(1 + \tau_{n,j}\right)) + \beta_{4,n} \ln\left(1 + \tau_{n,j}\right) + \varepsilon_{n,j}. \end{aligned} \tag{54}$$

The implied volatility of deep ITM calls and puts is very sensitive to changes in the index level. Since small errors in determining the appropriate index level are unavoidable, the disturbance variance of regression models (51) and (54) is supposed to increase as options go deeper ITM. Residual scatterplots support this presumption. To account for the heteroskedasticity of the disturbances we apply a *weighted least squares estimation* (WLS) assuming that the disturbance variance is proportional to the positive ratio of the option's delta and vega.[32] This ratio indicates how an increase in the index level by one (marginal) point affects the implied volatility of an option, if its price does not change.

In view of the large number of intraday transactions it is not astonishing that some extreme deviations occur representing "off-market" implied volatilities. To exclude such unusual events we discard all observations corresponding to large errors of more than four standard deviations of the regression residuals where the standard deviation is computed as the square root of the weighted average squared residuals. We then repeat the estimation on the basis of the reduced sample until no further observations are discarded. This procedure is known as applying the "*4-sigma-rule*" or "*trimmed regression*".[33] We examined the impact of this exclusion of outliers and found it to be negligible in all but very few cases.

A large percentage of all traded DAX options in the period from 1995 to 2002 features a degree of moneyness between -0.25 and 0.20 and a time to maturity below 180 days. We discard all observations outside this range in order to eliminate potential problems with extreme degrees of moneyness or time to maturity. As options with fewer than 5 days to maturity have relatively little or no time premium and hence the estimation of volatility is extremely

[31] This estimation procedure is similar to the two-step Cochrane-Orcutt method that is sometimes used to estimate regression models where the error terms are autocorrelated. For a detailed description of the Cochrane-Orcutt method, see Kmenta (1997), pp. 314-315.

[32] The delta and vega are computed using the implied volatility of the corresponding option. The delta of puts is multiplied by -1 to obtain a positive ratio.

[33] See Kmenta (1997), p. 219 and Sachs (1972), p. 265.

sensitive to measurement errors, we also exclude them. To always ensure a good fit to the data, we eliminate all days from the sample where the adjusted coefficient of determination is lower than 60%. In total, these are 29. Plotting the residuals did not reveal any remaining violations of the assumptions of the chosen regression model.

In the *last step* of the estimation procedure, we check for arbitrage opportunities. As can easily be shown, if the market for standard options at an arbitrary time t is free of arbitrage, then standard options satisfy four general arbitrage relations: the hedge relation, the bull spread relation, the butterfly spread relation, and the calendar spread relation.[34] To check these conditions, we consider for each day $n \in \{0, \ldots, N\}$ an equally spaced grid of 200 options exhibiting degrees of moneyness between -0.15 and 0.10 and times to maturity between 5 and 120 days. The call option prices are calculated on the basis of the estimated volatility surfaces. If any of the four no-arbitrage conditions is violated on a specific day, this day is excluded from the sample. In the overall sample, this happens on 42 days.[35] The remaining days are (re)numbered from 0 to N, with N now being equal to 1938.[36]

Empirical Results

For each day $n \in \{0, \ldots, N\}$, $N = 1938$, we estimate a regression of implied volatility on moneyness and time to maturity following the two-step procedure described above. The result of the first regression is a time series of daily coefficient estimates for the parameters $\beta_2, \beta_3, \beta_5$, and β_6. These are used to estimate the model constants ϱ_1 and ϱ_2. We get: $\widehat{\varrho}_1 = -1.6977$ and $\widehat{\varrho}_2 = -3.3768$. The corresponding R^2 values of 90.45% for model (52) and 95.61% for model (53) support the assumption of an almost deterministic, linear relationship between β_2 and β_5, on the one hand, and between β_5 and β_6, on the other hand. Repeating the estimation of ϱ_1 and ϱ_2, using different subsamples, the estimates turn out to be quite stable.

Based on the parameter estimates $\widehat{\varrho}_1 = -1.6977$ and $\widehat{\varrho}_2 = -3.3768$, we run regression (54). Across the 1939 days in the sample the average adjusted R^2 value is 92.44% and the median adjusted R^2 value amounts to 94.58%.[37] For comparison, the average adjusted R^2 value obtained under the original regression model (48), using the same sample, is 93.00%. The loss in accuracy of 0.70% seems acceptable, when contrasted with the increase in model parsimony. All in all, the high R^2 values suggest that our regression model captures most of the variation in implied volatilities attributable to moneyness and time to maturity. To further assess the quality of our model, the mean absolute error of the regression, i.e. the mean of the absolute deviations of the reported implied volatilities from the model's theoretical

[34] For details, see Brunner & Hafner (2003) and Carr (2001).

[35] Of the four no-arbitrage relations, the *butterfly spread relation*, stating that the value of a butterfly spread is always nonnegative, is violated most frequently.

[36] This results from $2010 - 29 - 42 - 1 = 1938$.

[37] It should be noted that in the case of a WLS regression model there exists no single generally accepted definition of R^2. The reported values are based on the non-weighted WLS regression residuals. The meaning of this R^2 is not exactly the same as in an ordinary least-squares regression (OLS) regression. For more details, see, e.g., Greene (1993), p. 399.

values, is computed each day. For almost all days in the sample, we find this measure to be well within the average bid-ask spread.[38]

Table 1 reports the mean and the standard deviation of the daily coefficient estimates for each parameter, as well as the t-statistic for the mean. Figure 1 gives a graphical representation of these results. It shows a plot of the average estimated volatility surface, i.e. the function $\widehat{g}(M, \tau, \overline{\widehat{\beta}})$, for different degrees of moneyness $M \in [M_L, M_U]$ and times to maturity $\tau \in [\tau_L, \tau_U]$.

$\overline{\widehat{\beta}}_1$ $s_{\widehat{\beta}_1}$ (t-value)	$\overline{\widehat{\beta}}_2$ $s_{\widehat{\beta}_2}$ (t-value)	$\overline{\widehat{\beta}}_3$ $s_{\widehat{\beta}_3}$ (t-value)	$\overline{\widehat{\beta}}_4$ $s_{\widehat{\beta}_4}$ (t-value)
0.2361 0.1031 (100.84)	−0.4966 0.1427 (-153.23)	1.4594 1.0369 (61.98)	0.0166 0.1397 (5.23)

Table 1: Mean, standard deviation and t-value of the daily parameter estimates over the period January 1995 to December 2002

For the interpretation of the regression results, it is convenient to recall the regression function:

$$\widehat{g}(M, \tau, \beta) = \beta_1 + \beta_2 M \left(1 + \varrho_1 \ln(1+\tau)\right) + \beta_3 M^2 (1 + \varrho_2 \ln(1+\tau)) + \beta_4 \ln(1+\tau). \quad (55)$$

The parameter β_1 is common to all implied volatilities constituting the volatility surface. It may therefore be interpreted as the general *level of volatility* in the market. It should be closely related to the volatility of the underlying index. During the sample period, the average estimated value of β_1, i.e. $\overline{\widehat{\beta}}_1$, amounts to 23.61%. The *shape of the volatility smile* is determined by the parameters β_2 and β_3 as the differentiation of the function $\widehat{g}(M, \tau, \beta)$ with respect to moneyness shows:

$$\begin{aligned} \frac{\partial \widehat{g}(M,\tau,\beta)}{\partial M} &= \beta_2 \left(1 + \varrho_1 \ln(1+\tau)\right) + 2\beta_3 M (1 + \varrho_2 \ln(1+\tau)), \quad (56) \\ \frac{\partial^2 \widehat{g}(M,\tau,\beta)}{\partial M^2} &= 2\beta_3 (1 + \varrho_2 \ln(1+\tau)). \end{aligned}$$

Given ϱ_1, the parameter β_2 reflects the common part of the *slope of the volatility smile*. As expected, its average estimated value of -0.4966 is negative. The *curvature of the smile* is represented by the parameter β_3. Since $(1 + \widehat{\varrho}_2 \ln(1+\tau))$ is positive for all $\tau \in [\tau_L, \tau_U]$ and the daily parameter estimates of β_3 are mostly positive, the volatility smile is typically convex. The degree of convexity is however often *small*. On average, the curvature of the smile amounts to 1.4594. The minimum of the smile is almost always located at degrees of moneyness clearly above zero. Given β_2 and β_3, the estimates $\widehat{\varrho}_1 = -1.6977$ and $\widehat{\varrho}_2 =$

[38] As our database does not contain information on bid and ask prices, we use the average bid-ask spread of all liquid option contracts quoted on December 31, 2002 as a proxy for the bid-ask spread in the whole sample period. We find this value to be roughly 0.3 volatility points. Since liquidity has very much increased over the years, this measure should provide a lower bound to the true bid-ask spread in the sample period.

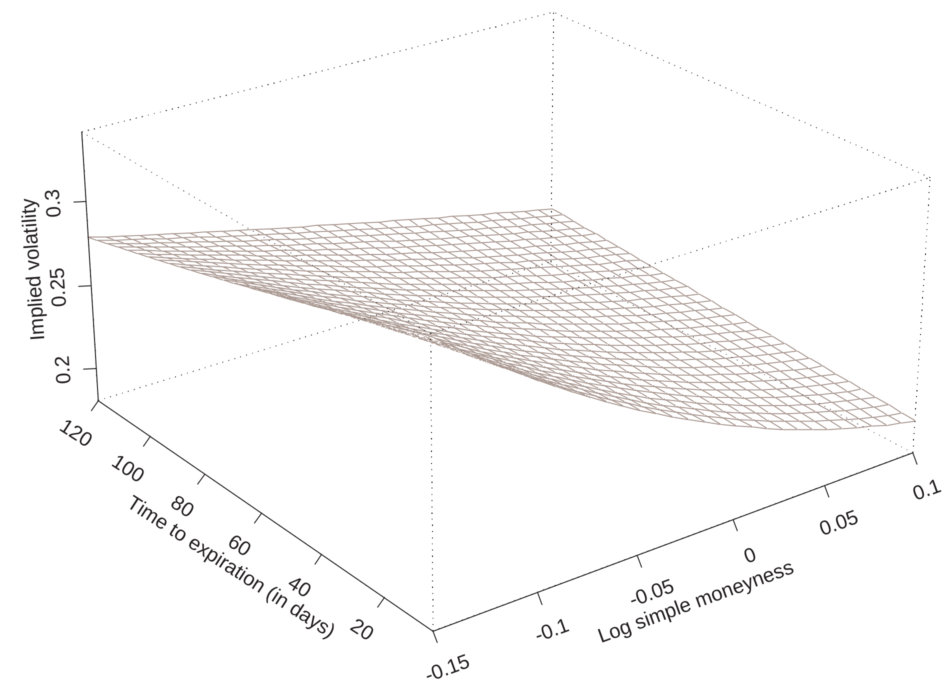

Figure 1: Average estimated volatility surface $\widehat{g}(M, \tau, \overline{\widehat{\beta}})$ for the sample period 1995-2002

-3.3768 suggest that, in general, the volatility smile is steeper and more convex for shorter-term options. These features of the average volatility smile are also apparent from Figure 2, which plots the function $\widehat{g}(M, \cdot, \overline{\widehat{\beta}})$ for three different times to maturity. With regard to the smile patterns, the skew pattern is the predominant pattern in our sample.[39] However, on some days, for instance on October 6, 1995, a nearly literal smile pattern can be observed (see Figure 3).

The *slope of the volatility term structure* of ATM options is represented by the parameter β_4. Its mean estimated value of 0.0166 indicates that, on average, the implied volatilities of shorter-term ATM options are lower than those of longer-term ATM options. This implies that the average volatility term structure features a "normal shape". A detailed investigation of the coefficient estimates $\widehat{\beta}_4$ reveals that the ATM volatility term structure exhibits on 1316 days a normal shape ($\widehat{\beta}_4 > 0$) and on 623 days ($\widehat{\beta}_4 < 0$) an inverse shape.

[39] Note, however, that a true skew, i.e. a linear function of implied volatility versus moneyness, hardly ever occurs.

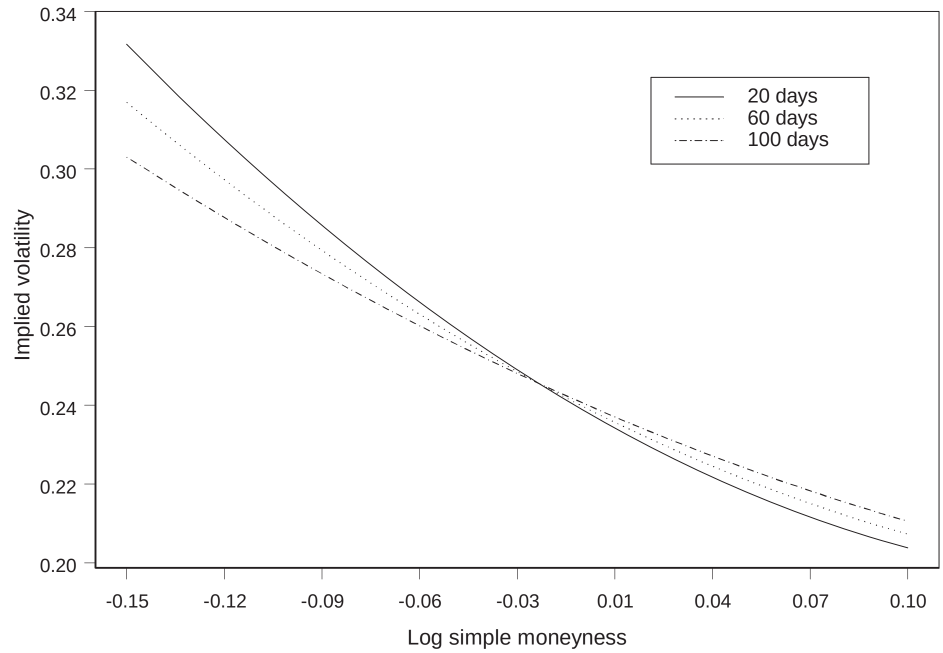

Figure 2: Average estimated volatility smiles for 20, 60, and 100 days to expiration. Sample period: 1995-2002

Identification and Selection of Volatility Risk Factors

The discussion so far has shown that the DAX volatility surface evolves randomly over time. However, as the volatility surface forms a highly correlated complex multivariate system, it is difficult to model. To reduce complexity, we search for a smaller set of *abstract* risk factors which represents, in the best possible way, the set of *original* risk factors, i.e. the implied volatilities for different degrees of moneyness and times to maturity. Given our regression model, the volatility surface on day n is completely described by the four regression coefficients $\beta_{i,n}$, $(i = 1, \ldots, 4)$ and the time-invariant parameters ϱ_1 and ϱ_2. Since the coefficient estimates have proven to be highly accurate, the regression coefficients $\beta_1, \ldots, \beta_4$ are the most natural candidates for being used as abstract risk factors.

To investigate the issue of model parsimony, we run a principal component analysis on the correlation matrix of the coefficient estimates. As a result, the first three principal components explain 92.90% of total variance. This value goes down to 85% when looked at different subsamples, implying that the fourth factor still explains a substantial part of total variance. Consequently, any further reduction of the number of risk factors would lead to a significant loss in accuracy. The last section has shown that the regression parameters are easy to interpret. The parameter β_1 represents the (overall) level of implied volatility, β_2 and β_3 stand respectively for the (overall) slope and the curvature of the implied volatility smile,

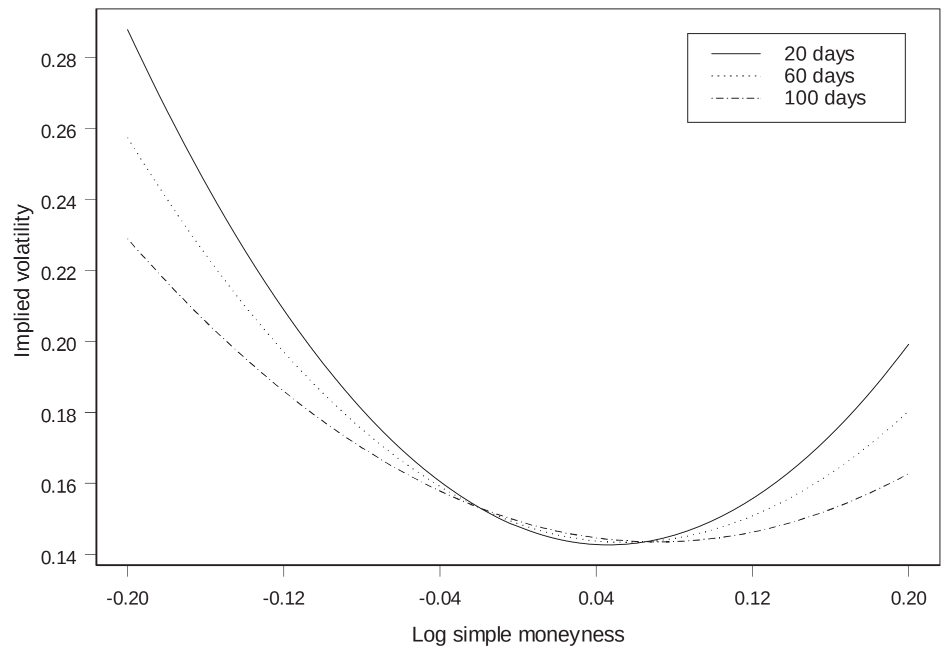

Figure 3: Estimated volatility smiles for 20, 60, and 100 days to expiration on October 6, 1995

and β_4 represents the slope of the (ATM) term structure of volatility. As these parameters can be thought of to capture systematic risks an option's investor is facing, and are therefore directly linked to economic activity, they are commonly called *fundamental* risk factors.

As opposed to fundamental risk factors, one could alternatively use methods such as factor analysis or principal component analysis to derive a set of *statistical risk factors* that characterize the dynamics of implied volatilities.[40] The main advantage of the statistical approach is the orthogonality of the obtained factors, i.e. the factors have a correlation of zero among each other. On the other hand, the factors are usually difficult to interpret. Moreover they are not unique, because a factor rotation can yield a different set of factors with the same degree of explanation.

Mainly due to their *better interpretability*, we decide for the regression coefficient estimates $\widehat{\beta}_1, \ldots, \widehat{\beta}_4$ or transformations of them to serve as our abstract implied volatility risk factors or just *volatility risk factors* $Y_1, \ldots, Y_4$. Concretely, we define

$$Y_{1,n} = \ln\left(\widehat{\beta}_{1,n}\right), \qquad Y_{i,n} = \widehat{\beta}_{i,n}, \quad i = 2, 3, 4, \tag{57}$$

for $n = 0, \ldots, N$. The variable $\widehat{\beta}_1$ was normalized by taking the natural logarithm, because

[40] For instance, Skiadopoulos et al. (1999), Cont (2001) and Cont & Fonseca (2002) analyze the dynamics of the S&P 500 volatility surface, using principal component analysis. Fengler et al. (2000) applies a similar technique to the DAX volatility surface.

it represents the level of volatility in the market and as such it has to be positive in any economically meaningful model. Note that the log transformation is not appropriate for the variable β_3, although β_3 mainly assumes positive values. The reason is that in order to be consistent with no-arbitrage the smile needs not to be convex in moneyness (and strike), but can also be concave.[41]

3.3 Dynamics of DAX Implied Volatilities

Statistical Analysis of the Evolution of the Volatility Risk Factors

Having identified the set of risk factors characterizing the volatility surface, this section is concerned with the question of finding what process is most appropriate for each factor.[42] For that purpose, we *individually* examine the historical time series of the volatility risk factors $\{Y_{i,n} : n = 0, \dots, N\}$, $i = 1, \dots, 4$, and determine their main statistical properties so that we can afterwards propose models that are suitable to capture most of the historical features.[43] To reduce complexity, we restrict our search for models to the class of *autoregressive integrated moving average* (ARIMA) models. The *correlation structure* defining the relationships between the volatility risk factors and the DAX index will be studied later on. The procedure for analyzing the data partly follows the *model identification stage* of the Box & Jenkings (1976) approach to time series analysis and involves the steps: 1) Graphical inspection of the data; 2) Identification of nonstationarity: testing for unit roots; 3) Analysis of the marginal distributions; 4) Determination of model order.

We start our analysis with a *graphical inspection* of the data. Figure 4 has four panels containing plots of the time series of the volatility factors $Y_1, \dots, Y_4$ for the sample period January 1995 to December 2002. During the sample period, the values of the estimated volatility level $\widehat{\beta}_1$ lie between 9.16% (August 10, 1996) and 63.13% (July 24, 2002). The values of Y_1, defined as the natural logarithm of β_1, therefore range between -2.39 and -0.46. As can be seen from the Y_1-graph, the volatility levels in the years 1995 and 1996 are distinctly lower than they are in the period from July 1997 to December 2002. Economically, the reason for this substantial increase in volatility in the first half of 1997 can be seen in the beginning of the Asian crisis. Reaching its peak in October 1997, the Asian crisis was followed by the Russian crisis just one year later in 1998 letting volatility levels – expressed in terms of $\widehat{\beta}_1$ – climb to more than 60% (see top left graph in Figure 4). Subsequently, such high levels of volatility have only been reached during the WTC terrorists attacks on September 11, 2001 and during the stock market's turmoil in fall 2002.

[41] See also Carr (2001).

[42] In the general factor-based stochastic implied volatility model, the price process of the underlying asset is assumed to follow a GBM with stochastic volatility, and the volatility process is implicitly defined in terms of implied volatility. Consequently, the price process of the underlying is *almost determined*, and therefore not considered here.

[43] A similar strategy for the specification of multivariate risk factor models is chosen by Algorithmics in their *Mark-to-Future* framework for scenario generation. See Reynolds (2001).

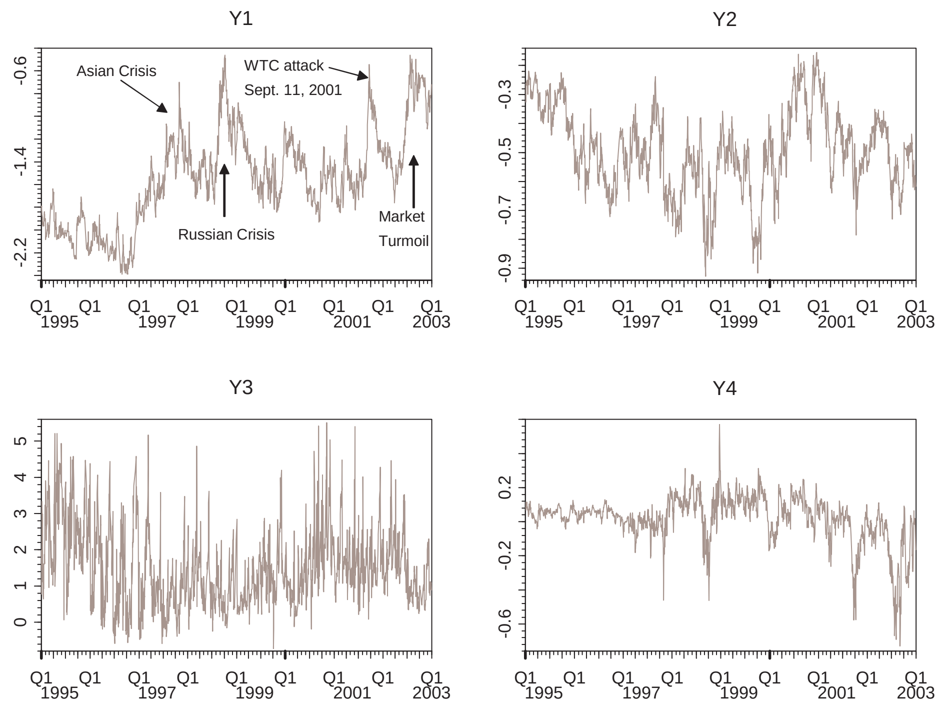

Figure 4: DAX volatility risk factors $Y_1, \ldots, Y_4$ on a daily basis over the period 1995-2002

As can be seen from Figure 4, the time series of the risk factors Y_2, Y_3, and Y_4 tend to be mean-reverting, i.e. they tend to fluctuate around a long-run mean. Consistent with our previous findings, the slope of the volatility smile Y_2 is always negative. On the other hand, the curvature of the volatility smile as represented by the factor Y_3 is mostly, but not always, positive. The Y_4 time series alternates between positive and negative values. Correspondingly the volatility term structure changes between normal and inverse shapes. The extreme points in the graphs of the Y_2 and Y_4 time series match quite well with the extreme points in the graph of the Y_1 series, although there seems to be a time lag in some cases. Exceptionally high levels of volatility Y_1 usually come along with a distinctly downward-sloping smile and a pronounced inverse term structure of volatility.

The second step of the procedure deals with the detection of potential nonstationarities. At a first glance, there is no evidence of a deterministic time trend, a seasonal component or a unit root in any of the four series. To further investigate the key issue of stationarity, we plot the *sample autocorrelation function* (sample ACF) for each of the series over the sample period 1995-2002 in Figure 5. It can serve as an indicator of whether a series is stationary. The slow decay of the sample ACFs of Y_1, Y_2 and Y_4 indicate either a large characteristic root, a unit root, or a trend stationary process.[44] For Y_3 a unit root is not suspected from the sample ACF. The high positive autocorrelations that are observed for Y_1 mean that high

[44] See Enders (1995), p. 211.

volatility levels follow high volatility levels and low volatility levels follow low volatility levels. This effect is well-documented for stock return volatilities and is commonly referred to as "*volatility clustering*".[45]

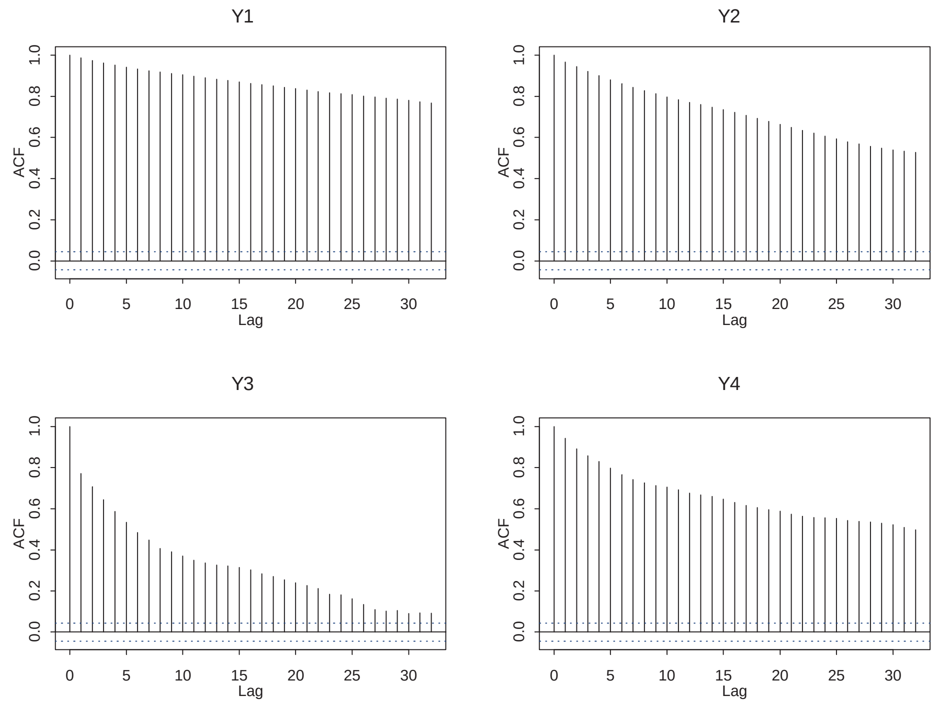

Figure 5: Sample ACF for each of the series $Y_1, \ldots, Y_4$ over the period 1995-2002

Although plots of the sample ACF are useful tools for detecting the possible existence of unit roots, the method is necessarily imprecise. For this reason, a number of formal unit root tests has been developed.[46] The most widely used tests are the *augmented Dickey-Fuller* (ADF) and the *Phillips-Perron* (PP) tests. Assuming that the data generation process is given by a *random walk with drift and deterministic time trend*, we apply these tests to $Y_1, \ldots, Y_4$. For all variables, except from Y_1, we can reject the null hypothesis of a unit root at the 1% significance level. For Y_1 the null of a unit root can be rejected at the 1% level when we employ the PP test, but it cannot be rejected according to the ADF test. Yet, as is well known, the available tests have low power to distinguish between unit root and near unit root processes.[47] Our failure to reject the null hypothesis according to the ADF test does not suffice to draw the conclusion that a unit root exists. As far as Y_1 is concerned, economic theory does not support a unit root since the variable should be subject to an upper arbitrage

[45] The volatility clustering effect dates back to Mandelbrot (1963) and Fama (1965).

[46] For a comprehensive treatment of unit root tests, see Hamilton (1994), Chapters 15-17.

[47] See Enders (1995), p. 261.

bound even if severe violations of the Black-Scholes assumptions occur.[48] Similar economic arguments hold for the other three variables. To conclude, we suppose that all volatility risk factors follow *stationary* processes, i.e. we assume that they are integrated of order 0.

	Y_1	Y_2	Y_3	Y_4
Minimum	-2.3896	-0.9276	-0.7275	-0.7291
25% Quantile	-1.7962	-0.5881	0.7127	-0.0235
Mean	-1.5297	-0.4966	1.4594	0.0166
Median	-1.5289	-0.4992	1.3057	0.0404
75% Quantile	-1.2872	-0.4012	2.0012	0.0954
Maximum	-0.4599	-0.1549	5.5113	0.5704
Standard deviation	0.4109	0.1427	1.0369	0.1397
Skewness	0.1912	-0.0893	0.8447	-1.6864
Excess Kurtosis	-0.3655	-0.2537	0.6381	4.7809

Table 2: Summary statistics for the volatility risk factor time series $Y_1,\ldots,Y_4$. Sample period: 1995-2002

Next, Table 2 presents descriptive statistics on the *marginal distribution* of the volatility risk factors $Y_1, \ldots, Y_4$. The distributions of Y_1 and Y_2 appear to be close to a normal distribution, as the skewness and excess kurtosis values are around zero and the median closely matches the mean in both cases. Performing a Jarque-Bera test and a Kolmogorov-Smirnov goodness-of-fit test, however, the null hypothesis that Y_1 follows a normal distribution is rejected at the 1% significance level in both tests.[49] In the case of Y_2, normality cannot be rejected at the 5% level using the Kolmogorov-Smirnov test, but is rejected at the 1% level, using the Jarque-Bera test. The skewness and excess kurtosis values for the factors Y_3 and Y_4 indicate a nonnormal distribution – both empirical distributions exhibit a skewness that is distinctly different from zero and excess kurtosis. Although the risk factors Y_3 and Y_4 are not normally distributed, in both cases the normal distribution still represents the "best" approximation to the empirical distribution from a broad spectrum of potential continuous distributions. More specifically, computing the Chi-Square and the Kolmogorov-Smirnov goodness-of-fit measures for the beta, the chi-square, the exponential, the gamma, the inverse Gaussian, the normal, the lognormal, the Student's t, and the Weibull distribution, we find the normal distribution (and for Y_3 also the lognormal distribution), to be the one where these measures are lowest.[50]

The focus of the last step of the procedure is on model order. The basic idea of the Box-Jenkins approach to ARIMA model identification is essentially to match the behavior of the

[48] See also Gatheral (2000). In contrast, Dartsch (1999), p. 141-146, argues that the VDAX follows a random walk as he cannot reject the null hypothesis of a unit root at conventional significance levels.

[49] For details on the Jarque-Bera test, see, e.g. Kmenta (1997), p. 265-266. More information on the Kolmogorov-Smirnov goodness-of-fit test can be found in Casella & Berger (2002).

[50] To fit a density, a first guess of parameters is made using maximum likelihood estimators. In a second step, the fit is optimized using the *Levenberg-Marquardt* method. See Press et al. (1992), p. 683 ff., for more information on this method.

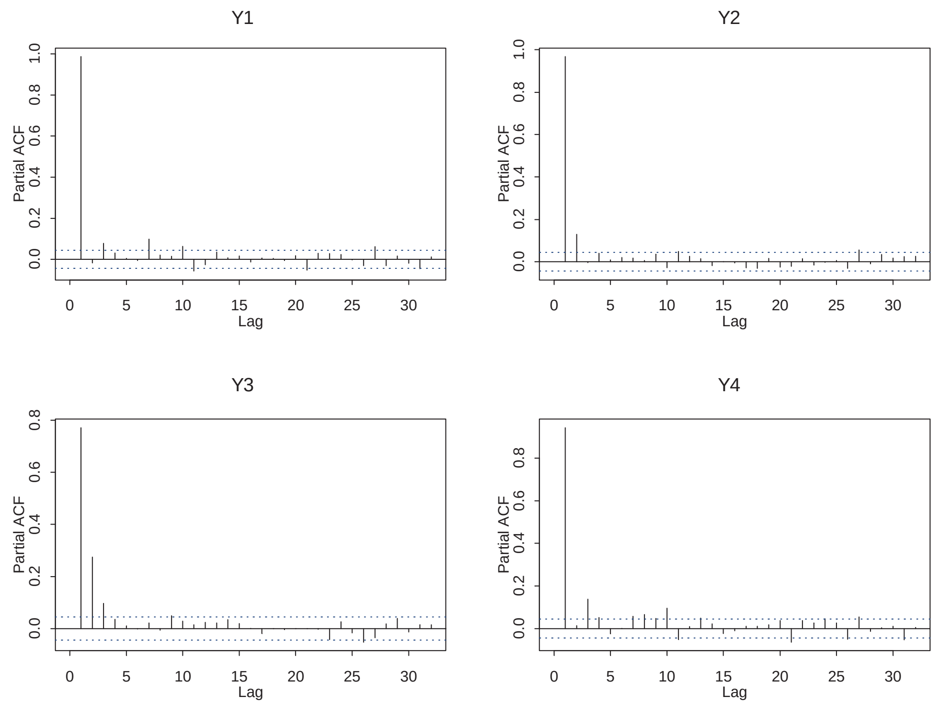

Figure 6: Sample PACF for each of the series $Y_1, \ldots, Y_4$ over the period 1995-2002

sample ACF and the *sample partial autocorrelation function* (sample PACF) with that of various theoretical ACFs and PACFs. For each of the series, the sample ACF is shown in Figure 5 and the sample PACF is plotted in Figure 6. The Y_1, Y_2 and Y_4 time series exhibit partial autocorrelations which are close to zero for lags greater than 1. Therefore, these variables might be reasonably well modelled by AR(1) processes. For Y_3 , representing the curvature of the volatility smile, an AR(1) process might not be the best choice, because the partial autocorrelations of lags 2 and 3 are significantly different from zero.[51] Instead, an AR(3) process should be more appropriate. Unfortunately, an AR(3) process is incompatible with our continuous-time model framework, as it only allows for processes of serial dependence of at most order 1.[52] For this reason, Y_3 is also presumed to follow an AR(1) process.

Relating Volatility Risk Factors to Index Returns

The high first-order autocorrelations suggest that most of the variation of the volatility risk factors is explained by the factors itself, or more precisely by their own lagged values. For

[51] For example, the sample partial autocorrelation for lag 2 is 0.2775. This is well above the 95% confidence level of 0.048.

[52] In theory, there exists a continuous-time version of a discrete AR(3) process. However, the process becomes very complex and the Itô theory is no longer applicable.

the unexplained part of the variance, a number of different market variables and fundamental factors may be responsible. With respect to our model, one factor, the *return of the underlying index*, is thereby of particular interest.

The relationship between index returns and both implied and realized volatility is usually negative: volatility increases when the index level falls. This well-established relationship is commonly referred to as the *leverage effect*.[53] In the following analysis we will examine the leverage effect for the German stock market, represented by the DAX index. The volatility level is represented by Y_1.[54] In addition to this analysis, we will also study the relation between the other three implied volatility risk factors and the DAX index level. Formally, we test on the relationship between the volatility risk factor Y_i $(i = 1, 2, 3, 4)$ and the DAX return R by running the *ordinary least squares* (OLS) regression

$$Y_{i,n} = b_{0,i} + b_{1,i} Y_{i,n-1} + b_{2,i} R_n + \varepsilon_n, \qquad n = 1, \ldots, N. \tag{58}$$

The variable R_n is defined as the continuously compounded DAX return on day n, i.e. $R_n = \ln(S_n/S_{n-1})$, where S_n denotes the last futures-implied DAX index level[55] on day n and ε denotes, as usual, a random disturbance. The regression coefficients for the predictors of the volatility risk factor i $(i = 1, \ldots, 4)$ are denoted by $b_{0,i}$, $b_{1,i}$, and $b_{2,i}$. The coefficient $b_{2,i}$ reflects the influence of the DAX return on the respective volatility factor. The lagged value $Y_{i,n-1}$ on the right-hand side of equation (58) is included to account for first-order autocorrelation in Y_i.

The OLS coefficient estimates and the t-statistics (in parentheses) from the regressions are displayed in Table 3. Due to the high autocorrelations, the adjusted coefficients of determination R^2_{adj} are also high. The parameter estimates for b_2 are all significant, at least at the 5% level. The negative coefficient estimate of -2.3043 for b_2 in the first regression supports the existence of a leverage effect: the volatility level Y_1 increases when the DAX index falls. The coefficient estimates $\widehat{b}_{2,i}$, $i = 2, 3, 4$, are interpreted as follows: when the DAX index falls, the slope of the volatility smile (Y_2) increases, the curvature of the volatility smile (Y_3) decreases, and the slope of the (ATM) volatility term structure (Y_4) decreases. In some further analyses, we find these relationsships to be stable over time.

[53] Among many examples in the literature, Schwert (1990) and Bollerslev et al. (1994) give an extensive analysis of the leverage effect.

[54] For a detailed study on the leverage effect for the US, see Figlewski & Wang (2000).

[55] S_n means in fact the *adjusted* futures-implied DAX index level $\breve{S}_n$.

Risk factor	$\widehat{b}_0$	$\widehat{b}_1$	$\widehat{b}_2$	R^2_{adj}
Y_1	-0.0173 (-3.93^{**})	0.9881 (353.62^{**})	-2.3043 (-34.06^{**})	98.52%
Y_2	-0.0164 (-5.62^{**})	0.9672 (171.82^{**})	-0.2769 (-5.73^{**})	93.87%
Y_3	0.3336 (12.90^{**})	0.7712 (53.40^{**})	2.1630 (2.40^{*})	59.64%
Y_4	0.0006 (0.66)	0.9430 (135.71^{**})	1.0598 (18.15^{**})	90.66%
* (**) significant at the 5% (1%) level				

Table 3: Relationship between volatility risk factors and DAX returns. Sample period: 1995-2002.

4 A Four-Factor Model for DAX Implied Volatilities

4.1 The Model under the Objective Measure

Model Specification

The empirical study in the previous chapter has shown that the evolution of the *DAX volatility risk factors* $Y_1, \ldots, Y_4$ may be well described by a system of correlated AR(1) processes with normal innovations. Since the continuous-time analogue of an AR(1) process is a *mean-reverting Ornstein-Uhlenbeck* (OU) *process*[56], we model the joint dynamics of the DAX volatility risk factors under the objective measure $\mathbb{P}$ by a system of correlated OU processes:[57]

$$dY_{i,t} = a_i(c_i - Y_{i,t})dt + \gamma_i d\widetilde{W}_{i,t}, \qquad i = 1, \ldots, 4, \forall t \in [0, T^*], \tag{59}$$

where $Y_{1,0}, Y_{2,0}, Y_{3,0}, Y_{4,0} \in \mathbb{R}$ are the initial values of the processes, $a_i \in \mathbb{R}_+, c_i \in \mathbb{R}, \gamma_i \in \mathbb{R}_+$ $(i = 1, \ldots, 4)$ are constant parameters, and T^* is, as usual, some finite planning horizon. The Brownian motions $\widetilde{W}_i$ and $\widetilde{W}_j$ are correlated with (instantaneous) correlation $\rho_{i,j}$, i.e.

$$d\widetilde{W}_{i,t} d\widetilde{W}_{j,t} = \rho_{i,j} dt,$$

for $i, j = 1, \ldots, 4$, and $i \neq j$. According to (59), the volatility risk factor Y_i is pulled to a level c_i at rate a_i. When Y_i is high, mean reversion tends to cause it to have a lower drift; when Y_i is low, mean reversion tends to cause it to have a higher drift. The parameter c_i is commonly referred to as the *mean-reversion level* or the long-run average and the parameter a_i is called the *speed of mean reversion* $(i = 1, \ldots, 4)$. The parameter γ_i may be interpreted as the "*volatility*" of the i-th volatility risk factor Y_i $(i = 1, \ldots, 4)$.

[56] For details on the (mean-reverting) OU process, see Øksendal (1998), p. 74 and Karatzas & Shreve (1988).

[57] For similar process specifications in the context of implied volatility modelling, see Cont & Fonseca (2002), Goncalves & Guidolin (2003), and Hafner & Wallmeier (2001).

For any $(t, M, \tau) \in [0, T^*] \times [M_L, M_U] \times [\tau_L, \tau_U]$, $\tau_U \leq T^* - t$, we describe the relative (DAX) *volatility surface* $\widetilde{\sigma}_t(M, \tau)$ by[58]

$$\widetilde{\sigma}_t\,(M, \tau) = g(t, M, \tau, Y_{1,t}, Y_{2,t}, Y_{3,t}, Y_{4,t}), \tag{60}$$

where $g(\cdot)$ is given by

$$\begin{aligned} g(t, M, \tau, y_1, y_2, y_3, y_4) \quad &= \quad e^{y_1} + y_2 M\,(1 + \varrho_1 \ln(1+\tau)) \\ &\quad\; + y_3 M^2 (1 + \varrho_2 \ln(1+\tau)) + y_4 \ln(1+\tau)\,, \end{aligned} \tag{61}$$

with constant parameters $\varrho_1, \varrho_2 \in \mathbb{R}$, and M is the log simple moneyness, i.e.:

$$\begin{aligned} M \quad &= \quad m(t, S_t, K, T, r) \\ &= \quad \ln\left(\frac{K}{F_t(T)}\right) = \ln\left(\frac{K}{S_t}\right) - r(T-t), \end{aligned} \tag{62}$$

Since $m(t, S_t, S_t, T, r) = 0$, the dynamics of the underlying's instantaneous volatility is obtained from Theorem 5 by

$$\upsilon_t = \widetilde{\sigma}_t\,(m(t, S_t, S_t, t, r), 0) = e^{Y_{1,t}}, \qquad \forall t \in [0, T^*]\,. \tag{63}$$

Equation (63) shows that the (stochastic) instantaneous volatility of the underlying asset is equal to the level of implied volatility (in absolute terms). The other implied volatility risk factors, i.e. the slope and the curvature of the volatility smile and the slope of the term structure of volatility, have no influence. The intuition behind this result is that these factors represent specific properties of the options market and as such they have no impact on the stock price process. Note that in contrast to traditional stochastic volatility models, the stochastic volatility component $\upsilon_t = \exp(Y_{1,t})$ is *observable*.[59]

Using relation (63) and assuming a constant drift rate $\mu \in \mathbb{R}$, the evolution of the (DAX) *index level* under $\mathbb{P}$ is given by

$$dS_t = S_t \mu dt + S_t e^{Y_{1,t}} d\widetilde{W}_{0,t} \qquad \forall t \in [0, T^*]\,, \tag{64}$$

where the initial stock price S_0 is supposed to be positive. To model the observed dependency between the DAX volatility surface and the DAX index, the Brownian motion that drives the DAX index $\widetilde{W}_0$ is correlated with the Brownian motions $\widetilde{W}_1, \ldots, \widetilde{W}_4$ that drive the DAX volatility surface. The instantaneous correlations $\rho_{0,j}$ $(j = 1, \ldots, 4)$, with $\rho_{0,j}$ defined by $d\widetilde{W}_{0,t} d\widetilde{W}_{j,t} = \rho_{0,j} dt$, are assumed to be constant.

For completeness, we also state the price process of the *money market account.* It is given by the ODE:

$$dB_t = rB_t dt, \qquad \forall t \in [0, T^*]\,, \tag{65}$$

where $B_0 = 1$ and the interest rate r is supposed to be constant and nonnegative. As can readily be shown, all processes fulfill the usual technical regularity conditions.

[58] For a motivation and empirical test of this representation of the DAX volatility surface, see Section 3.2.

[59] In the estimation of the DAX volatility surface we have only included options with a time to maturity of 5 calendar days and more. Thus, the implied volatility for $\tau = 0$ represents an extrapolated value. For the ATM option that is relevant here, we suppose this estimate to be very precise.

In the model specification, the Wiener process W driving the stock price and the volatility risk factors was assumed to be a *multi-dimensional standard Brownian motion.* This implies that each component of W is an independent standard Brownian motion in one dimension. In contrast, the formulation above involves a *vector Brownian motion* $\widetilde{W} = \left(\widetilde{W}_0, \ldots, \widetilde{W}_4\right)'$ with correlation matrix $\Xi = \left(\rho_{i,j}\right)_{i,j=0,\ldots,4}$.[60] To transform $\widetilde{W}$ into W we use a technique known as *Cholesky decomposition* or *Cholesky factorization*.[61] Defining

$$d\widetilde{W}_{i,t} = \sum_{j=0}^{4} e_{i,j} dW_{j,t}, \qquad i = 0, \ldots, 4, \tag{66}$$

or in vector notation

$$d\widetilde{W}_t = E \cdot dW_t,$$

where $E = (e_{i,j})_{i,j=0,\ldots,4}$ is a 5×5 lower triangular matrix with zeros in the upper right corners, i.e. E is of the form

$$E = \begin{pmatrix} e_{0,0} & 0 & 0 & 0 & 0 \\ e_{1,0} & e_{1,1} & 0 & 0 & 0 \\ e_{2,0} & e_{2,1} & e_{2,2} & 0 & 0 \\ e_{3,0} & e_{3,1} & e_{3,2} & e_{3,3} & 0 \\ e_{4,0} & e_{4,1} & e_{4,2} & e_{4,3} & e_{4,4} \end{pmatrix}. \tag{67}$$

the stock price process and the processes of the volatility risk factors can be (re)expressed in terms of 5-*dimensional standard Brownian motion* W.[62] We get

$$dS_t = S_t \mu dt + S_t e^{Y_{1,t}} dW_{0,t}, \tag{68}$$

and

$$dY_{i,t} = a_i(c_i - Y_{i,t})dt + \sum_{j=0}^{4} \gamma_{i,j} dW_{j,t}, \qquad i = 1, \ldots, 4, \tag{69}$$

for all $t \in [0, T^*]$, with $\gamma_{i,j}$ defined by

$$\gamma_{i,j} = \gamma_i e_{i,j}, \qquad i = 1, \ldots, 4, \; j = 0, \ldots, 4. \tag{70}$$

For convenience, the elements $\gamma_{i,j}$ are summarized in the matrix $\Gamma = (\gamma_{i,j})_{\substack{i=1,\ldots,4 \\ j=0,\ldots,4}}$. Depending on the specific application, we will henceforth use the correlated or the uncorrelated form of the model.

Model Estimation

Discretizing the explicit solution of the DAX price process (64) and applying the *Euler method*[63] to the DAX volatility risk factor processes (59), the discrete-time approximation

[60] Note that $\rho_{i,j} = 1$ for $i = j$, by definition.

[61] The Cholesky decomposition and other related techniques can only create uncorrelated, but not statistically independent, processes. However, in the case of normality, which is considered here, independency is equivalent to uncorrelatedness. See Hartung et al. (1999).

[62] For the matrix E to be rendered unique, we specify that all diagonal elements must be positive. See Dhrymes (2000), p. 80-82.

[63] See Kloeden et al. (2003), pp. 91-97.

of the continuous-time model (64) and (59) – the correlated form – is:[64]

$$\ln(S_n) = \left(\mu - \frac{1}{2}e^{2Y_{1,n-1}}\right)\Delta t + \ln(S_{n-1}) + e^{Y_{1,n-1}}\sqrt{\Delta t}\varepsilon_{0,n}, \tag{71}$$

$$Y_{i,n} = a_i c_i \Delta t + (1 - a_i \Delta t)\, Y_{i,n-1} + \gamma_i \sqrt{\Delta t}\varepsilon_{i,n}, \qquad i = 1, \ldots, 4, \tag{72}$$

for a given time discretization $0 = t_0 < t_1 < \ldots < t_n < \ldots < t_{N^*} = T^*$ with equal time increments $\Delta t = T^*/N^* = t_n - t_{n-1}$ $(n = 1, \ldots, N^*)$. The index level and the value of the i-th volatility risk factor at time t_n are denoted by $S_n = S_{t_n}$ and $Y_{i,n} = Y_{i,t_n}$, respectively. For each $i = 0, \ldots, 4$, ε_i follows a Gaussian white noise process with variance 1. This implies that $\varepsilon = (\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)'$ is a 5-dimensional white noise process – though not necessarily Gaussian. Since the diagonal elements of the covariance matrix $(Cov\,[\varepsilon_{i,n}, \varepsilon_{j,n}])_{i,j=0,\ldots,4}$ are equal to one, the covariance matrix equals the correlation matrix Ξ.

Given observations $\{(Y_{0,n}, Y_{1,n}, Y_{2,n}, Y_{3,n}, Y_{4,n}) : n = 0, \ldots, N\}$, $N < N^*$, where $Y_{0,n} = \ln(S_n)$ for $n = 0, \ldots, N$, we employ a *maximum likelihood* (ML) approach to estimate the model parameters.[65] To make maximum likelihood estimation feasible, the conditional distribution $\epsilon_n | Y_{n-1}$ is assumed to be multivariate normal with mean zero and covariance matrix $\Sigma_n | Y_{n-1}$. If the normality assumption is violated, the estimation technique detailed below is known as *quasi maximum likelihood estimation* (QMLE).

It is well known that the MLE of the mean stock return is a very noise estimate of the true parameter μ. Given that imprecision, we consider it more accurate simply to *impose* a value for the mean rather then trying to estimate the mean from the data.[66] To estimate the other model parameters we apply OLS equation-by-equation.[67] In fact, it can then be shown that for each i $(i = 1, \ldots, 4)$ the parameter estimates $\widehat{a}_i$ and $\widehat{c}_i$ are simple transformations of the estimated regression coefficients from an OLS regression of $Y_{i,n}$ on $Y_{i,n-1}$.

Estimation Results

In the sample period 1995-2002, the average number of trading days per year amounts to 251.[68] We thus set $\Delta t = 1/251$. The number of observations in the final sample is 1939; this implies $N = 1938$. The mean return of the DAX is supposed to be 8%, i.e. $\widehat{\mu} = 8\%$.[69]

Estimates of parameters a_i and c_i $(i = 1, \ldots, 4)$ are given in Table 4. Additionally, the coefficient of determination R^2 of each OLS regression is reported. To illustrate the strength

[64] See Taylor (1994) for a similar discrete-time specification.

[65] For simplicity, we use the same symbols to denote sample observations and process variables.

[66] See also Figlewski (1997), pp. 22-23.

[67] For a similar procedure, see Goncalves & Guidolin (2003), p. 10. Note that the OLS estimatior is not efficient here. To obtain efficiency in estimation, an iterative SUR estimation algorithm has to be applied (see, e.g., Hamilton (1994), pp. 315-317). In small samples, however, the OLS estimator often turns out to be more robust.

[68] The original sample consists of 2010 trading days, spread over 8 years.

[69] This assumption is based on the results of a study of Claus & Thomas (2001). Using an implicit estimation method, they report a mean DAX return of 8% for the year 1997 (see p. 1649). For 1995 and 1996, the estimates range between 8 and 9%.

of the pull back to mean, we compute the "*half-life*" of each series. The (estimated) half-life HL_i of process Y_i $(i = 1, \ldots, 4)$ is defined as the number of trading days h_i such that[70]

$$(1 - \widehat{a}_i \Delta t)^{h_i} = 0.5. \tag{73}$$

Hence:

$$HL_i = \frac{\ln(0.5)}{\ln(1 - \widehat{a}_i \Delta t)}. \tag{74}$$

Clearly, the higher $\widehat{a}$, the shorter the half-life and thus the stronger the pull back to mean. Apparently, the mean-reverting effect is most pronounced for Y_3 with a half-life of roughly 3 trading days and least pronounced for Y_1 with a half-life of 63 trading days. Furthermore, the values suggest that shocks to the structure of the volatility surface, represented by the variables Y_2, Y_3, and Y_4, are rather temporary, whereas shocks to the volatility level Y_1 tend to be much more persistent.

	$\widehat{a}$	$\widehat{c}$	HL	R^2
Y_1	2.7575	-1.4797	63	97.64%
Y_2	7.9951	-0.5013	21	93.77%
Y_3	57.3609	1.4601	3	59.52%
Y_4	14.0119	0.0144	12	89.07%

Table 4: Estimates of the parameters a_i and c_i $(i = 1, \ldots, 4)$ for the sample period 1995-2002. In addition: estimated half-lifes $HL_1, \ldots, HL_4$

Given the residuals of the OLS regressions, we calculate the "volatility" estimates $\gamma_1, \ldots, \gamma_4$ of the volatility risk factors $Y_1, \ldots, Y_4$. We get

$$\widehat{\gamma}_1 = 1.0006, \qquad \widehat{\gamma}_2 = 0.5646, \qquad \widehat{\gamma}_3 = 10.4561, \quad \text{and} \qquad \widehat{\gamma}_4 = 0.7317. \tag{75}$$

Finally, the estimate of the correlation matrix Ξ is obtained:

$$\widehat{\Xi} = \left(\widehat{\rho}_{i,j}\right)_{i,j=0,\ldots,4} \begin{pmatrix} 1.0000 & -0.6152 & -0.1787 & 0.0315 & 0.3446 \\ -0.6152 & 1.0000 & 0.0588 & -0.1668 & -0.8020 \\ -0.1787 & 0.0588 & 1.0000 & 0.2041 & 0.0276 \\ 0.0315 & -0.1668 & 0.2041 & 1.0000 & 0.0696 \\ 0.3446 & -0.8020 & 0.0276 & 0.0696 & 1.0000 \end{pmatrix}. \tag{76}$$

The negative correlation between $\widehat{\varepsilon}_0$ and $\widehat{\varepsilon}_1$ of -0.6152 is consistent with the leverage effect: when the DAX index (i.e. Y_0) falls, the overall level of DAX implied volatilities as represented by the factor Y_1 shifts upward. The other values in the first row of $\widehat{\Xi}$ can be interpreted as follows: when the DAX index falls, the slope of the DAX volatility smile (Y_2) increases, the curvature of the DAX volatility smile (Y_3) decreases, and the slope of the (ATM) DAX volatility term structure (Y_4) decreases.[71] The correlations among the volatility risk factors are in general quite low; however the estimate $\widehat{\rho}_{1,4} = -0.8020$ suggests that there exists a

[70] See, e.g., Taylor (2001), p. 58.

[71] These results are consistent with the findings in Section 3.3.

relatively strong negative relationship between the volatility level and the slope of the ATM term structure.

Applying the Cholesky decomposition to the matrix $\widehat{\Xi}$ yields

$$\widehat{E} = (\widehat{e}_{i,j})_{i,j=0,\dots,4} = \begin{pmatrix} 1.0000 & 0 & 0 & 0 & 0 \\ -0.6152 & 0.7884 & 0 & 0 & 0 \\ -0.1787 & -0.0648 & 0.9817 & 0 & 0 \\ 0.0315 & -0.1870 & 0.2013 & 0.9610 & 0 \\ 0.3446 & -0.7483 & 0.0415 & -0.0931 & 0.5574 \end{pmatrix}, \quad (77)$$

and with the help of equation (70) we get the following estimate of the matrix Γ:

$$\widehat{\Gamma} = (\widehat{\gamma}_{i,j})_{\substack{i=1,\dots,4 \\ j=0,\dots,4}} = \begin{pmatrix} -0.6155 & 0.7889 & 0 & 0 & 0 \\ -0.1009 & -0.0365 & 0.5543 & 0 & 0 \\ 0.3294 & -1.9553 & 2.1043 & 10.0484 & 0 \\ 0.2522 & -0.5476 & 0.0303 & -0.0682 & 0.40790 \end{pmatrix}. \quad (78)$$

Diagnostic Checking

After fitting the model to the data, the last step in the Box-Jenkings methodology to time series analysis involves *diagnostic checking*.[72] The standard practice is to plot the (standardized) residuals to look for outliers and evidence of periods in which the model does not fit the data well. If the fitted model is appropriate, the standardized residuals $\widehat{\varepsilon}_i$ $(i = 0, \dots, 4)$ should have properties similar to those of univariate Gaussian white noise processes with variance 1. Plotting $\widehat{\varepsilon}_0, \dots, \widehat{\varepsilon}_4$, the graphs gives no indication of a nonzero mean or nonconstant variance, though some very few outliers can be observed. So on this basis there is no evidence for a violation of the white noise assumption. The next step is to check that the sample ACFs of $\widehat{\varepsilon}_0, \dots, \widehat{\varepsilon}_4$ behave as they should under the assumption that the fitted model is appropriate. Under the white noise assumption, it is well known that for large sample size N the sample autocorrelations are approximately i.i.d with distribution $N(0, 1/N)$. We can therefore test whether or not the observed residuals are consistent with white noise by examining the sample ACF of the residuals and rejecting the white noise hypothesis if more than three out of 60 fall outside the confidence bound $\pm 1.96/\sqrt{N}$. As a result, only the sample ACFs of $\widehat{\varepsilon}_3$ and $\widehat{\varepsilon}_4$ indicate a violation of the white noise assumption, because more than 4 sample autocorrelations lie outside the critical bounds. Finally, the *Box-Pierce* and *Ljung-Box* Q-statistics can be used to test whether a *group of autocorrelations* is significantly different from zero. Under the null hypothesis of no significant autocorrelations, Q is approximately χ^2 distributed . Applying the Ljung-Box test to $\widehat{\varepsilon}_3$ and $\widehat{\varepsilon}_4$, the null hypothesis of no autocorrelation is rejected at the 1% level for almost all of the tested lag lengths. It is also rejected for $\widehat{\varepsilon}_1$, though only for some of the tested lag lengths.[73] A further examination reveals that for the residuals $\widehat{\varepsilon}_3$ and $\widehat{\varepsilon}_4$ to become white noise AR models of

[72] See Brockwell & Davis (1996), pp. 162-165 and Enders (1995), pp. 97-98.

[73] We used different lag lengths of up to 200 lags.

order 10 and more would be needed. Obviously, this would present a conflict with model parsimony and also with other models which mainly advocate AR(1) processes for implied volatilities.[74] Although the hypothesis of normally distributed residuals is formally rejected at the 5% level, large deviations from normality are in general not observed for any of the series.[75]

Out-of-Sample Forecasting

So far, we have only investigated the in-sample behaviour of the model. In this section, we complete the analysis by examining the model's out-of-sample forecasting ability.[76] Our forecasting procedure involves the following steps:

1. Divide the sample to an in-sample period and an out-of-sample period.
2. Estimate the model using in-sample data.
3. Obtain forecasts for the out-of-sample period.
4. Compare the forecasts to the actual observations in the out-of-sample period.

We choose the (initial) in-sample period to include the first 250 daily observations or approximately data from the year 1995 (step 1). Using these data, we apply the procedure described in the previous section to estimate the model parameters (step 2). In step 3, we employ the estimated model to produce forecasts of the risk factor vector $Y = (Y_1, Y_2, Y_3, Y_4)$ for three different forecasting horizons h: 1, 5, and 10 trading days. Due to specific structure of our VAR(1) model – the variables $Y_1, \ldots, Y_4$ do only depend on their own lagged values and $\epsilon = (\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4)$ is independent white noise so that ϵ_n and ϵ_m are independent for $n \neq m$ – it can readily be shown that the optimal h-days-ahead forecast $\widehat{Y}_{n+h}$, given the information at time n, equals the vector of the individually h-days-ahead optimal forecasts $\left(\widehat{Y}_{1,n+h}, \widehat{Y}_{2,n+h}, \widehat{Y}_{3,n+h}, \widehat{Y}_{4,n+h}\right)$.[77]

In evaluating the forecasts, we conduct the forecasting in a *rolling-window fashion,* i.e. every period the latest data is added, and the oldest data is removed. The size of the rolling window is thereby kept fix at 250 observations. For $h = 5$ or $h = 10$, we have to decide whether to use overlapping or nonoverlapping windows. In general, overlapping windows give us more forecasts, however, they have the drawback that the forecasts are autocorrelated, by

[74] Cont & Fonseca (2002), for instance, proposes a three-factor model for the S&P 500 implied volatility surface, where each factor follows an AR(1) process. For the DAX, Wallmeier (2003), pp. 194-195, suggests to model the ATM volatility and the difference between the implied volatility of a 5% OTM option and a 5% ITM option (called the "span") by an AR(1) process. These variables are similar to the risk factors Y_1 and Y_2 in our model.

[75] The violation of the normality assumption means that the model is not able to fully match the empirical distributions of the DAX index and the DAX volatility risk factors.

[76] Strictly speaking, the following tests are not truly out-of-sample as we have used all data from the sample to derive the functional form of the daily DAX volatility surface.

[77] See Lütkepohl (1993), pp. 28-30.

construction. Because our sample is sufficiently large and h is suffciently small, we decide for nonoverlapping windows.

To compare the forecast values with the realized or actual values (step 4), we perform two analyses. In the first analysis, we run for each factor i ($i = 1, \ldots, 4$) and for each forecast horizon $h \in \{1, 5, 10\}$ the following regression:

$$Y_{i,k}^{(h)} = b_{0,i}^{(h)} + b_{1,i}^{(h)} \widehat{Y}_{i,k}^{(h)} + u_{i,k}^{(h)}, \qquad k = 1, 2, \ldots, m_h. \tag{79}$$

where the k-th realization (forecast) of factor i for horizon h is denoted by $Y_{i,k}^{(h)}$ ($\widehat{Y}_{i,k}^{(h)}$), u_i^h is a random disturbance and the sample size is m_h.[78] At least three hypotheses can be tested using equation (79). First, if $\widehat{Y}_{i,k}^{(h)}$ contains *some* information about $Y_{i,k}^{(h)}$, the slope coefficient $b_{1,i}^{(h)}$ should be nonzero. Second, if $\widehat{Y}_{i,k}^{(h)}$ is an *unbiased* forecast of $Y_{i,k}^{(h)}$, we should find $b_{0,i}^{(h)} = 0$ and $b_{1,i}^{(h)} = 1$. Finally, if the forecasts are *efficient*, the residuals $\widehat{u}_{i,k}^{(h)}$ should be white noise. For judging the overall out-of-sample performance of the model, the coefficient of determination R^2 is used.

OLS estimates of (79) are shown in Table 5. Obviously, the slope estimate $\widehat{b}_1$ is statistically different from zero in all cases.[79] Hence, the forecast values contain at least some information about the future actual values. Since the estimates for b_1 are morevover close to unity and the estimates for b_0 are close to zero – none of them is significant at the 5% level – we would expect the forecasts to be unbiased. To formally check on this, we perform F-tests. In all cases, the joint hypothesis $H_0: b_0 = 0, b_1 = 1$ cannot be rejected; none of the $F(2, m-2)$-statistics is significant at the 5% level. The Durbin-Watson statistic *DW*, reported in the last column of Table 5, is only in three cases significantly different from two. Therefore we conclude that the residuals u are not autocorrelated and hence that the forecasts are *efficient*. The high R^2 values, especially for $h = 1$, suggest that the model forecasts explain most of the out-of-sample variations of the volatility risk factors. As expected, the out-of-sample R^2 value is negatively correlated with the forecast horizon. However, the loss in out-of-sample performance turns out to be quite different for the different risk factors. So is the decrease in R^2 relatively small for Y_1, Y_2 and Y_4 but quite large for Y_3. To conlude, our results suggest that the *forecasts of the volatility risk factors are unbiased and efficient*.

Besides the magnitude of the forecasts, the number of correct direction predicted by a model may also be of interest. In the second analysis, we therefore compute the proportion of correct direction (PCD) for each risk factor Y_i ($i = 1, \ldots, 4$) and for each forecast horizon $h \in \{1, 5, 10\}$, where the PCD measure is defined by

$$\text{PCD} = \frac{\text{Number of correct direction forecasts}}{m}. \tag{80}$$

The results are shown in Table 6. The first observation is that all PCD values are above 50%, i.e. the model produces more correct direction forecasts than wrong direction forecasts. To check on their statistical significance, we test the null hypothesis $H_0: \text{PCD} \leq 0.50$ against the alternative hypothesis $H_1: \text{PCD} > 0.50$. For all volatility risk factors but Y_2 and for all

[78] See Christensen & Prabhala (1998), p. 134, for a similar test.

[79] For convenience, we omit the risk factor index i and the forecast horizon index h.

		$\widehat{b}_0$	$\widehat{b}_1$	t-ratio ($b_1 = 0$)	F-ratio ($b_0 = 0, b_1 = 1$)	R^2	DW
$h = 1$	Y_1	0.0027	1.0008	244.02**	0.44	97.24%	1.87
	Y_2	-0.0023	0.9969	146.48**	0.39	92.71%	2.18
	Y_3	-0.0012	0.9867	46.54**	0.98	56.22%	2.23
	Y_4	-0.0004	0.9997	116.88**	0.41	89.01%	1.92
$h = 5$	Y_1	0.0027	0.9971	47.32**	0.41	86.99%	2.01
	Y_2	-0.0084	0.9890	29.26**	0.30	71.89%	1.97
	Y_3	0.0289	0.9307	9.02**	1.32	19.56%	1.37
	Y_4	-0.0034	1.0111	24.07**	0.24	63.38%	1.87
$h = 10$	Y_1	-0.0056	0.9855	23.95**	0.67	77.56%	1.80
	Y_2	-0.0197	0.9710	14.65**	0.34	56.40%	1.67
	Y_3	0.4429	0.6502	3.66**	2.12	7.50%	1.55
	Y_4	-0.0095	1.0568	13.15**	0.82	51.50%	1.62

* (**) significant at the 5% (1%) level

Table 5: Test on the out-of-sample performance of the model: regression results

forecast horizons h the null hypothesis of a PCD value lower than or equal to 0.50 (or 50%) can be rejected at conventional significance levels (1% and 5%).[80] The second observation is that the PCD values increase with the length of the forecast horizon. Probably this is due to the mean-reverting property of the series: the longer the forecast horizon, the stronger the pull back to mean. This explanation is also supported by a third observation: for a given forecasting horizon h the PCD value is highest for the series with the shortest half-life (here: Y_3), and lowest for the series with the longest half-life (here: Y_1 and Y_2).[81]

	$h = 1$	$h = 5$	$h = 10$
Y_1	52.81%*	58.16%**	60.11%**
Y_2	51.74%	54.30%	54.76%
Y_3	57.90%**	66.17%**	69.64%**
Y_4	55.06%**	60.23%**	68.45%**

* (**) significant at the 5% (1%) level

Table 6: Proportion of correct direction (PCD) values

The above analyses have shown that the model is able to forecast quite accurately the direction of the movements of the volatility risk factors as well as their future levels.

[80] The p-values for Y_2 are 7.63% ($h = 1$), 5.67% ($h = 5$), and 10.79% ($h = 10$).

[81] See Table 4.

4.2 The Model under the Risk-Neutral Measure

Risk-Neutral Stock Price and Volatility Dynamics

This section focuses on the pricing and hedging of contingent claims. For now, we assume that there exists a unique martingale measure $\mathbb{Q}$. Then, according to Theorem 6 in Section 2.4, the arbitrage price of any $\mathbb{Q}$-attainable claim H is given by the risk-neutral valuation formula where the expectation is taken with respect to the joint diffusion of the stock price and its volatility under the measure $\mathbb{Q}$:

$$dS_t = S_t r dt + S_t \upsilon_t dW^*_{0,t}, \tag{81}$$

$$\begin{aligned} d\upsilon_t &= \left(\widetilde{\eta}_t(m(t,S_t,S_t,t,r),0) - \widetilde{\vartheta}_t(m(t,S_t,S_t,t,r),0)\,\psi_t\right)dt \\ &\quad + \widetilde{\vartheta}_t(m(t,S_t,S_t,t,r),0)\,dW^*_t. \end{aligned} \tag{82}$$

Here, W^* is a $(p+1)$-dimensional $\mathbb{Q}$-standard Brownian motion, and $\psi = (\psi_0,\ldots,\psi_p)'$ is the $(p+1)$-dimensional market price of risk process. The drift process $\widetilde{\eta}_t = \widetilde{\eta}_t\,(M,\tau)$ and the volatility process $\widetilde{\vartheta}_t = \widetilde{\vartheta}_t\,(M,\tau)$ are given by equations (15) and (16). Considering the specific structure of our model, p is equal to 4 and $\gamma_{i,j,t} = \gamma_{i,j}$, $i = 1,\ldots,4$, $j = 0,\ldots,4$ for all $t \in [0,T^*]$. By using

$$\begin{aligned} g(t,M,\tau,y_1,y_2,y_3,y_4) &= e^{y_1} + y_2 M\,(1+\varrho_1 \ln(1+\tau)) \\ &\quad + y_3 M^2 (1+\varrho_2 \ln(1+\tau)) + y_4 \ln(1+\tau)\,, \end{aligned} \tag{83}$$

it follows that

$$\begin{aligned} \widetilde{\eta}_t(M,\tau) &= \alpha_{1,t} e^{Y_{1,t}} + \alpha_{2,t} M(1+\varrho_1 \ln(1+\tau)) \\ &\quad + \alpha_{3,t} M^2(1+\varrho_2 \ln(1+\tau)) + \alpha_{4,t} \ln(1+\tau) + \frac{1}{2}\sum_{j=0}^{4} \gamma_{1,j}^2 e^{Y_{1,t}}, \end{aligned} \tag{84}$$

and

$$\begin{aligned} \widetilde{\vartheta}_{j,t}(M,\tau) &= \gamma_{1,j} e^{Y_{1,t}} + \gamma_{2,j} M(1+\varrho_1 \ln(1+\tau)) \\ &\quad + \gamma_{3,j} M^2 (1+\varrho_2 \ln(1+\tau)) + \gamma_{4,j} \ln(1+\tau)\,, \end{aligned} \tag{85}$$

with $\alpha_{i,t} = a_i(c_i - Y_{i,t})$, $i = 1,\ldots,4$, for all $t \in [0,T^*]$ and for $j = 0,\ldots,4$. Since the moneyness function

$$m(t,s,K,T,r) = \ln\left(\frac{K}{s e^{r(T-t)}}\right) = \ln\left(\frac{K}{s}\right) - r(T-t), \tag{86}$$

is 0 if evaluated at $s = K = S_t$, equation (82) becomes:

$$d\upsilon_t = \left(a_1(c_1 - Y_{1,t})e^{Y_{1,t}} + \frac{1}{2}\sum_{j=0}^{4}\gamma_{1,j}^2 e^{Y_{1,t}} - \sum_{j=0}^{4}\gamma_{1,j} e^{Y_{1,t}}\psi_{j,t}\right)dt + \sum_{j=0}^{4}\gamma_{1,j} e^{Y_{1,t}}\psi_{j,t} dW^*_{j,t}.$$

Using the fact that $\gamma_{1,j} = 0$ for $j = 2,3,4$ and $\upsilon_t = e^{Y_{1,t}}$, $\forall t \in [0,T^*]$, the stock price volatility process under $\mathbb{Q}$ can finally be written as

$$\begin{aligned} \frac{d\upsilon_t}{\upsilon_t} &= \left(a_1\,(c_1 - \ln(\upsilon_t)) + \frac{1}{2}\left(\gamma_{1,0}^2 + \gamma_{1,1}^2\right) - \gamma_{1,0}\psi_{0,t} - \gamma_{1,1}\psi_{1,t}\right)dt \\ &\quad + \gamma_{1,0} dW^*_{0,t} - \gamma_{1,1} dW^*_{1,t}, \end{aligned} \tag{87}$$

where the market price of stock price risk is given by $\psi_{0,t} = \frac{\mu - r}{\upsilon_t}$. At a first glance, it seems that the dynamics of the stock price volatility are completely determined by the evolution of Y_1. However, this turns out not to be the case. The volatility risk factors Y_2, Y_3 and Y_4 enter equation (87) through ψ_1, as ψ_1 depends in general on the stock price and all volatility risk factors. Since the evolution of the volatility risk factors under the measure $\mathbb{Q}$ is given by

$$dY_{i,t} = \left[a_i(c_i - Y_{i,t}) - \sum_{j=0}^{4} \gamma_{i,j} \psi_{j,t} \right] dt + \sum_{j=0}^{4} \gamma_{i,j} dW^*_{j,t}, \qquad i = 1, \ldots, 4, \ \forall t \in [0, T^*], \quad (88)$$

ψ_1 also depends on ψ_0, ψ_2, ψ_3 and ψ_4.

The Market Price of Risk Process

Recall from Section 2.3, that equation (33) holds for all traded options at all times. To evaluate (33) for our specific case, we compute $\eta_t(K,T)$ and $\vartheta_t(K,T)$, generally defined in (26), (27) and (28):

$$\begin{aligned} \eta_t(K,T) &= \widetilde{\eta}_t(M,\tau) - \frac{Y_{2,t} M \varrho_1 + Y_{3,t} M^2 \varrho_2 + Y_{4,t}}{1+\tau} + \frac{1}{2} \upsilon_t^2 2 Y_{3,t} (1 + \varrho_2 \ln(1+\tau)) \quad (89) \\ &+ \left(r - \mu + \frac{1}{2} \upsilon_t^2 \right) \cdot \left[Y_{2,t}(1 + \varrho_1 \ln(1+\tau)) + 2 Y_{3,t} M (1 + \varrho_2 \ln(1+\tau)) \right] \\ &- \upsilon_t \left[\gamma_{2,0}(1 + \varrho_1 \ln(1+\tau)) + 2 \gamma_{3,0} M (1 + \varrho_2 \ln(1+\tau)) \right], \end{aligned}$$

and

$$\begin{aligned} \vartheta_{0,t}(K,T) &= \widetilde{\vartheta}_{0,t}(M,\tau) - \upsilon_t \left[Y_{2,t}(1 + \varrho_1 \ln(1+\tau)) + 2 Y_{3,t} M (1 + \varrho_2 \ln(1+\tau)) \right], \\ \vartheta_{j,t}(K,T) &= \widetilde{\vartheta}_{j,t}(M,\tau), \qquad j = 1, \ldots, 4, \end{aligned} \quad (90)$$

for all $t \in [0, T^*]$.

Because we assume that an equivalent martingale measure exists, equation (33) must hold especially for a specific option with strike price K_1 and maturity date T_1 and additionally $\lim_{t \to T_1} \psi_t < \infty$, $\mathbb{P}$-a.s. Since

$$\lim_{t \to T_1} \vartheta'_t(K_1, T_1) = \begin{pmatrix} \gamma_{1,0} e^{Y_{1,T_1}} - \gamma_{2,0} M + \gamma_{3,0} M^2 + \upsilon_{T_1} \left(Y_{2,T_1} + 2 Y_{3,T_1} M \right) \\ \gamma_{1,1} e^{Y_{1,T_1}} - \gamma_{2,1} M + \gamma_{3,1} M^2 \\ \gamma_{1,2} e^{Y_{1,T_1}} - \gamma_{2,2} M + \gamma_{3,2} M^2 \\ \gamma_{1,3} e^{Y_{1,T_1}} - \gamma_{2,3} M + \gamma_{3,3} M^2 \\ \gamma_{1,4} e^{Y_{1,T_1}} - \gamma_{2,4} M + \gamma_{3,4} M^2 \end{pmatrix} < \infty, \quad (91)$$

$\mathbb{P}$-a.s., the left-hand side of equation (33) is less than infinity $\mathbb{P}$-a.s. Let us consider the right-hand side now. As

$$\begin{aligned} \lim_{t \to T_1} \eta_t(K_1, T_1) &= \widetilde{\eta}_{T_1}(M, \tau_1) - Y_{2,T_1} M \varrho_1 + Y_{3,T_1} M^2 \varrho_2 + Y_{4,T_1} \quad (92) \\ &+ \left(r - \mu + \frac{1}{2} \upsilon_{T_1}^2 \right) \cdot \left[Y_{2,T_1} + 2 Y_{3,T_1} M \right] \\ &- \upsilon_{T_1} \left[\gamma_{2,0} + 2 \gamma_{3,0} M \right] < \infty \end{aligned}$$

$\mathbb{P}-$a.s., but

$$\frac{1}{2}\frac{d_1 d_2}{\sigma_t\left(K_1,T_1\right)}\vartheta_t(K_1,T_1)\vartheta_t^{'}(K_1,T_1)+\frac{1}{2}\frac{1}{\sigma_t(K_1,T_1)\left(T_1-t\right)}\left(\upsilon_t^2-\sigma_t^2(K_1,T_1)\right)$$
$$-\frac{d_2}{\sigma_t(K_1,T_1)\sqrt{T_1-t}}\upsilon_t\vartheta_{0,t}(K_1,T_1), \tag{93}$$

explodes as t approaches T_1, the right-hand side of equation (33) also explodes. This presents a contradiction to our initial assumption and thus gives rise to arbitrage opportunities in the model. One possibility to overcome this problem would be to modify the initially specified processes of the volatility risk factors such that the no-arbitrage condition (33) can be satisfied. But this would cause other problems, e.g., such processes are usually very complicated, and they often produce a worse fit to the observable implied volatility data. Therefore, we keep our initial real-world model specification. To be still able to use risk-neutral pricing, we try to find a new market price of risk process ψ and thus a measure $\mathbb{Q}$ such that the stock's instantaneous volatility process υ, defined by (87), exists, and the set of standard options is priced correctly within a specified range of degrees of moneyness and times to maturity.[82] Then we have found a measure $\mathbb{Q}$ that at least avoids arbitrage opportunities "*locally*". In the following we show how to specify ψ. This involves two basic steps:

1. Specify the functional dependence of ψ_i $(i=1,\ldots,4)$ on the state variables $Y_1,\ldots,Y_4$.
2. Calibrate the model to market data.

The remainder of this section is devoted to step 1. In order to specify the functional form of $\psi_{i,t}$ $(i=1,\ldots 4)$, we use a statistical approach that is based upon the observed market prices of risk. For this purpose, we select a set of options with degrees of moneyness

$$\mathbf{M}=\left\{M_i: M_i=M_L+\frac{M_U-M_L}{9}\cdot i,\ i=0,\ldots,9\right\}$$

and times to maturity $\boldsymbol{\tau}=\{10/365;40/365;70/365;100/365\}$.[83] The moneyness boundaries M_L and M_U are chosen to be -0.05 and 0.05. As for $(M_j,\tau_j)\in\mathbf{M}\times\boldsymbol{\tau}$, the right-hand side of (33) exists, (33) is well-defined, and we want to keep close to the original model, we deduce the functional form of $\psi_{i,t}$ $(i=1,\ldots,4)$ from (33). Therefore, for each day $n\in\{0,\ldots,N\}$ in the sample, we estimate the market prices of risk $\psi_{i,n}$, $i=1,\ldots 4$, from a cross-section of $k=40$ options specified by $(M_j,\tau_j)\in\mathbf{M}\times\boldsymbol{\tau}$. Estimation is done by running the regression model

$$\breve{\eta}_n\left(K_j,T_j\right)-\vartheta_{0,n}\left(K_j,T_j\right)\psi_{0,n}=\sum_{i=1}^{4}\vartheta_{i,n}\left(K_j,T_j\right)\psi_{i,n}+u_{n,j},\qquad j=1,\ldots,k, \tag{94}$$

where u_n is a random disturbance and $\breve{\eta}_n\left(K_j,T_j\right)$ is defined as the right-hand side of equation (33). The computation of $\breve{\eta}_n\left(K_j,T_j\right),\vartheta_{i,n}\left(K_j,T_j\right),\ i=0,\ldots,4$ and $\psi_{0,n}$ is based on the parameter estimates of Section 4.1 and the observation vector

[82] This implies that we now are in an *incomplete market*.

[83] The specification of the set of options in (M,τ) rather than (K,T) has practical reasons: If we fix at time $t=n\cdot\Delta t$ an option with maturity date T this option ceases to exist at time T and has to be replaced by another option. Yet, it is unclear by which one.

$(S_n, Y_{1,n}, Y_{2,n}, Y_{3,n}, Y_{4,n}, r_n)$, where r_n is the risk-free rate on day n. Across the 1939 days in the sample, the average adjusted R^2 is 80.93%.

Given the time series of coefficient estimates $\widehat{\psi}_{1,n}, \ldots, \widehat{\psi}_{4,n}$, we next examine their statistical properties and their relationships to the state variables $Y_1, \ldots, Y_4$ in more detail. Testing on the stationarity of $\widehat{\psi}_1, \ldots, \widehat{\psi}_4$, the null of a unit root can be rejected at the 1% level for all variables, using the ADF and PP-test. Performing various statistical analyses and balancing the issues complexity and accuracy, we suppose ψ_2, ψ_3 and ψ_4 to be constants. To find a suitable representation for ψ_1 in terms of $Y_1, \ldots, Y_4$, we use *stepwise regression*.[84] The set of explanatory variables consists of a constant, the variables $Y_1, \ldots, Y_4$, as well as all combinations of $Y_1, \ldots, Y_4$. The regression model that has the smallest value of Akaike's Information Criterion is considered the "best". Restricting the number of variables to five (including the constant), we find this to be:

$$\widehat{\psi}_{1,n} = \kappa_1 + \kappa_2 Y_{1,n} + \kappa_3 Y_{1,n}^2 + \kappa_4 Y_{1,n} Y_{4,n} + \kappa_5 Y_{2,n} Y_{3,n} + \zeta_n, \qquad n \in [0, N]. \tag{95}$$

where κ_i $(i = 1, \ldots, 5)$ are the regression coefficients and ζ is a random disturbance. The average adjusted R^2 value amounts to 97.55%. All parameter estimates turn out to be significant, and an inspection of the residuals shows no severe violation of the assumptions underlying the regression model. To test the robustness of the model, we compare the model's goodness-of-fit among four different subperiods of equal length. The average adjusted R^2 values of 96.04%, 98.90%, 99.12% and 98.19% suggest that the relationship between ψ_1 and the variables on the right-hand side of (95) is sufficiently stable to allow us to model

$$\psi_{1,t} = \kappa_1 + \kappa_2 Y_{1,t} + \kappa_3 Y_{1,t}^2 + \kappa_4 Y_{1,t} Y_{4,t} + \kappa_5 Y_{2,t} Y_{3,t}, \qquad t \in [0, T^*]. \tag{96}$$

Summing up the preceding results, we can state the following theorem:

Theorem 7 *Let $\mathbb{Q}$ be an equivalent martingale measure defined through*

$$\begin{aligned} \psi_{1,t} &= \kappa_1 + \kappa_2 Y_{1,t} + \kappa_3 Y_{1,t}^2 + \kappa_4 Y_{1,t} Y_{4,t} + \kappa_5 Y_{2,t} Y_{3,t}, \\ \psi_{2,t} &= \kappa_6, \quad \psi_{3,t} = \kappa_7, \quad \psi_{4,t} = \kappa_8, \end{aligned}$$

where $\kappa_1, \ldots, \kappa_8$ are real-valued constants and

$$dY_{i,t} = \left[a_i(c_i - Y_{i,t}) - \sum_{j=0}^{4} \gamma_{i,j} \psi_{j,t} \right] dt + \sum_{j=0}^{4} \gamma_{i,j} dW_{j,t}^*, \qquad i = 1, \ldots, 4.$$

The process ψ_1 is supposed to be well-defined. Then the arbitrage price $\Pi_t(H)$ at time $t \in [0, T]$ of any $\mathbb{Q}$-attainable claim H on the stock S with maturity date $T \leq T^*$ in the factor-based stochastic implied volatility model is given by

$$\Pi_t(H) = e^{-r(T-t)} \mathbb{E}_{\mathbb{Q}} \left[H | \mathcal{F}_t \right], \qquad \forall t \in [0, T], \tag{97}$$

where the joint evolution of the stock price and the stock price volatility under $\mathbb{Q}$ is

$$dS_t = S_t r dt + S_t \upsilon_t dW_{0,t}^*, \tag{98}$$

[84] For details on stepwise regression methods, see Miller (1990).

and

$$\frac{d\upsilon_t}{\upsilon_t} = \left(a_1\left(c_1 - \ln\left(\upsilon_t\right)\right) + \frac{1}{2}\left(\gamma_{1,0}^2 + \gamma_{1,1}^2\right) - \gamma_{1,0}\psi_{0,t} - \gamma_{1,1}\psi_{1,t}\right)dt + \gamma_{1,0}dW_{0,t}^* - \gamma_{1,1}dW_{1,t}^*. \quad (99)$$

In general, the expectation (97) cannot be solved in closed-form. Therefore, we apply monte Carlo simulation technniques. For the calculation of the Greeks within the Monte Carlo framework, we suggest the so called *explicit finite differencing approach with path recycling.*

Model Calibration

By observing today's stock price S_0, today's interest rate $r = r_0$ and today's volatility risk factors $Y_{1,0}, \ldots, Y_{4,0}$, the (estimated) *market price* of stock options is given. In addition, given the vector $\widehat{\theta}$ of "historical" parameter estimates, we can compute today's *theoretical prices* of option contracts. In order to calibrate the model to data, we choose the vector of parameters that determine the market price of risk processes such that the theoretical European call option prices are "as close as possible" to observed market prices, i.e. we minimize the *mean absolute percentage error* (MAPE).

In the following, we calibrate the model to option prices as of December 30, 2002. To keep the computational burden at a *reasonable* level, we select only two maturities $\tau_1 = T_1 = 20/365$ and $\tau_2 = T_2 = 50/365$, and for each maturity we consider only 5 degrees of moneyness: -0.05, -0.025, 0, 0.025 and 0.05. Solving the optimization problem, yields an optimal MAPE of 0.4498% or approximately 45 basis points.[85] This value indicates that the model is able to accurately price back the considered set of standard options. It is sufficiently low for most practical applications.

5 Summary and Conclusion

We develop a general mathematical model of a financial market in continuous time where in addition to the usual underlying securities stock and money market account, a continuum of standard European call options is traded. Given an initial volatility surface and a mechanism which describes how it fluctuates, we derive a condition that has to be imposed on the drift coefficients of the options' implied volatilities to ensure absence of arbitrage. Under the assumption that a unique equivalent martingale measure exists, the arbitrage price of a general stock price dependent claim can be obtained from the risk-neutral valuation formula. Because the at-the-money implied volatility converges to the stock price volatility when the time to maturity goes to zero, the stock's instantaneous volatility (process) is *implicitly* defined. In particular, this means that if implied volatilities are stochastic the stock price volatility is also required to be stochastic in order for the model to be consistent.

[85] Note that this value is effectively an estimate, since option prices are obtained by Monte Carlo simulation.

In the *empirical part* of this paper, we propose a two-stage approach to modelling the DAX volatility surface. In the first stage we model the daily DAX volatility surface along the dimensions degree of moneyness and time to maturity ("*structure*"). Using all call and put prices of each trading day in the sample period from 1995 to 2002, we estimate the DAX volatility surface via regression analysis. In the second stage we investigate and model the *dynamics* of the surface by applying time series techniques.

We propose a four-factor model for the stochastic evolution of the DAX implied volatility surface in continuous time. The DAX volatility risk factors are modelled by mean-reverting Ornstein-Uhlenbeck (OU) processes,[86] and the DAX index is given by a geometric Brownian motion with constant drift and stochastic volatility. In an examination of the model's out-of-sample performance, we find the forecasts of the volatility risk factors to be *unbiased* and *efficient*. Moreover, the model produces significantly *more correct direction forecasts* than wrong direction forecasts.

In contrast to classical stochastic volatility models, the instantaneous volatility in our model is *observable* and the same set of parameters can be used under the real-world measure *and* the risk-neutral measure. This allows for the *integrated* pricing and hedging, risk managing, and trading of index derivatives and derivatives on the index volatility.

[86] An OU process is the continuous-time analogue of an AR(1) process in discrete time.

A Proof of Theorem 5

Proof. Let us consider a standard European call option with strike price equal to the forward price, i.e. $K = S_t e^{r(T-t)}$. Then, the terms d_1 and d_2 from the Black-Scholes option pricing formula are given by

$$d_1(t, S_t, K) = \frac{1}{2}\widetilde{\sigma}_t\Big(m(t, S_t, S_t e^{r(T-t)}, T, r), T-t\Big)\sqrt{T-t},$$

and $d_2(t, S_t, K) = -d_1(t, S_t, K)$. Since for small $T-t$ the cumulative standard normal distribution function can be approximated by a first-order Taylor approximation

$$\begin{aligned} N\left(d_1\left(t, S_t, S_t e^{r(T-t)}\right)\right) &\approx \frac{1}{2} + \frac{\widetilde{\sigma}_t\big(m(t, S_t, S_t e^{r(T-t)}, T, r), T-t\big)\sqrt{T-t}}{2\sqrt{2\pi}}, \\ N\left(d_2\left(t, S_t, S_t e^{r(T-t)}\right)\right) &\approx \frac{1}{2} - \frac{\widetilde{\sigma}_t\big(m(t, S_t, S_t e^{r(T-t)}, T, r), T-t\big)\sqrt{T-t}}{2\sqrt{2\pi}}, \end{aligned}$$

the ATM forward call option price, i.e. $C_{BS}(t, S_t, K)$ for $K = S_t e^{r(T-t)}$, is given by

$$\begin{aligned} & C_{BS}\Big(t, S_t, S_t e^{r(T-t)}\Big) = S_t\left(N\left(d_1\Big(t, S_t, S_t e^{r(T-t)}\Big)\right) - N\left(d_2\Big(t, S_t, S_t e^{r(T-t)}\Big)\right)\right) \\ \approx\ & S_t\frac{\widetilde{\sigma}_t\big(m(t, S_t, S_t e^{r(T-t)}, T, r), T-t\big)\sqrt{T-t}}{\sqrt{2\pi}}. \end{aligned}$$

Therefore

$$\begin{aligned} & \widetilde{\sigma}_t(m(t, S_t, S_t, t, r), 0) \\ =\ & \lim_{T\to t}\widetilde{\sigma}_t\Big(m(t, S_t, S_t e^{r(T-t)}, T, r), T-t\Big) = \lim_{T\to t}\frac{\sqrt{2\pi}}{\sqrt{T-t}}\frac{C_{BS}\big(t, S_t, S_t e^{r(T-t)}\big)}{S_t}. \end{aligned}$$

Now it suffices to show that the last line of this equation is equal to υ_t. This can be done on the basis of the payoff function for the ATM forward call option using the distribution properties of the Brownian motion. The price of the call option can be described as the expected discounted payoff under the Black-Scholes martingale measure $\mathbb{Q}$:

$$C_{BS}\Big(t, S_t, S_t e^{r(T-t)}\Big) = \mathbb{E}_{\mathbb{Q}}\left[e^{-r(T-t)}\max\Big\{S_T - S_t e^{r(T-t)}; 0\Big\}\right],$$

Here, the option is considered close to maturity so that the stock price S_T can be approximated by

$$S_T \approx S_t e^{r(T-t)} + S_t\upsilon_t(W_T^* - W_t^*),$$

for small $T-t$; the last term is the Brownian increment which is normally distributed with mean zero and variance $T-t$. Consequently

$$\mathbb{E}_{\mathbb{Q}}\left[e^{-r(T-t)}\max\Big\{S_T - S_t e^{r(T-t)}, 0\Big\}\right] \approx \mathbb{E}_{\mathbb{Q}}\left[e^{-r(T-t)}\max\{S_t\upsilon_t(W_T^* - W_t^*); 0\}\right].$$

Taking the limit of the product of the right-hand side and $1/\sqrt{T-t}$ gives

$$\lim_{T\to t}\frac{1}{\sqrt{T-t}}\mathbb{E}_{\mathbb{Q}}\left[e^{-r(T-t)}\max\{S_t\upsilon_t(W_T^* - W_t^*); 0\}\right] = \lim_{T\to t}\frac{1}{\sqrt{2\pi}}S_t e^{-r(T-t)}\upsilon_t,$$

where the property $\mathbb{E}[\max\{z; 0\}] = \frac{\sqrt{\upsilon}}{\sqrt{2\pi}}$ for a normal random variable z with mean zero and variance υ is used. ■

References

Ait-Sahalia, Y. & Lo, A. W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices, *Journal of Finance* **53**: 499–547.

Albanese, C., Carr, P. & Madan, D. (1998). Stochastic implied volatility, *Working paper*, Bank of America Securities.

Andersen, T., Bollerslev, T., Diebold, F. & Labys, P. (1999). The distribution of exchange rate volatility, *Working paper*, Northwestern University.

Black, F. & Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy* **81**: 635–654.

Bollerslev, T., Engle, R. F. & Nelson, D. B. (1994). ARCH models, *in* R. Engle & D. McFadden (eds), *Handbook of Econometrics*, Vol. 4, Elsevier Science, North-Holland, Amsterdam, chapter 49, pp. 2959–3038.

Box, G. & Jenkings, G. (1976). *Time Series Analysis: Forecasting and Control*, Holden Day, San Francisco.

Brockwell, P. J. & Davis, R. A. (1996). *Introduction to Time Series and Forecasting*, Springer-Verlag, Berlin et al.

Brunner, B. & Hafner, R. (2003). Arbitrage-free estimation of the risk-neutral density from the implied volatility smile, *Journal of Computational Finance* **7**(1): 75–106.

Carr, P. (2001). Constraints on implied volatility, *Working paper*, Banc of America Securities.

Casella, G. & Berger, R. L. (2002). *Statistical Inference*, 2 edn, Duxbury, Pacific Grove.

Christensen, B. & Prabhala, N. (1998). The relation between implied and realized volatility, *Journal of Financial Economics* **50**: 125–150.

Claus, J. & Thomas, J. (2001). Equity premia as low as three percent ? Evidence from analysts' earnings forecasts for domestic and international stock markets, *Journal of Finance* **56**: 1629–1666.

Constantinides, G. (1996). Transactions costs and the implied volatility smile, *Working paper*, University of Chicago.

Cont, R. (2001). Dynamics of implied volatility surfaces, Ph.D. Course at the Aarhus Business School.

Cont, R. & Fonseca, J. D. (2002). Dynamics of implied volatility surfaces, *Quantitative Finance* **2**(1): 45–60.

Dartsch, A. (1999). *Implizite Volatilitäten am Aktien- und Optionsmarkt*, Deutscher Universitätsverlag, Wiesbaden.

Das, S. R. & Sundaram, R. K. (1999). Of smiles and smirks: A term structure perspective, *Journal of Financial and Quantitative Analysis* **34**(2): 211–239.

Derman, E. & Kani, I. (1994a). Riding on the smile, *RISK* **7**(2): 32–39.

Derman, E. & Kani, I. (1994b). The volatility smile and its implied tree, *Quantitative strategies research notes*, Goldman Sachs.

Dhrymes, P. J. (2000). *Mathematics for Econometrics*, 3 edn, Springer-Verlag, New York et al.

Dumas, B., Fleming, J. & Whaley, R. E. (1998). Implied volatility functions: empirical tests, *Journal of Finance* **53**: 2059–2106.

Dupire, B. (1994). Pricing with a smile, *RISK* **7**(1): 18–20.

Enders, W. (1995). *Applied Econometric Time Series*, John Wiley and Sons, New York et al.

Fama, E. (1965). The behavior of stock market prices, *Journal of Business* **38**: 34–105.

Fengler, M. R., Härdle, W. & Villa, C. (2000). The dynamics of implied volatilities: A common principal component approach, *Working paper*, Humboldt Universität Berlin.

FIA (2003). 2002 annual volume, FIA Press Release.

Figlewski, S. (1989). Options arbitrage in imperfect markets, *Journal of Finance* **34**: 1289–1311.

Figlewski, S. (1997). Forecasting volatility, *Financial Markets, Institutions and Instruments* **6**(1): 1–88.

Figlewski, S. & Wang, X. (2000). Is the "leverage effect" a leverage effect, *Working paper*, NYU Stern School of Business.

Filipovic, D. (2000). *Consistency Problems for HJM Interest Rate Models*, PhD thesis, ETH Zürich, Zürich.

Gatheral, J. (2000). Case studies in Financial Modeling, Course notes, Spring 2000, www.math.nyu.edu.

Goncalves, S. & Guidolin, M. (2003). Predictable dynamics in the S&P 500 index options implied volatility surface, *Working paper*, University de Montreal.

Greene, W. H. (1993). *Econometric Analysis*, 2 edn, Prentice-Hall, Englewood Cliffs.

Hafner, R. & Wallmeier, M. (2001). The dynamics of dax implied volatilities, *International Quartlerly Journal of Finance* **1**(1): 1–27.

Hamilton, J. D. (1994). *Time Series Analysis*, Princeton University Press, New Jersey.

Hartung, J., Elpelt, B. & Klösener, K. (1999). *Statistik: Lehr- und Handbuch der angewandten Statistik*, 12 edn, Oldenbourg Verlag, München.

Heath, D., Jarrow, R. & Morton, A. (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation, *Econometrica* **60**(1): 77–105.

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.

Hull, J. C. (2000). *Options, Futures, and Other Derivatives*, 4th edn, Prentice Hall, Englewood Cliffs, New Jersey.

Hull, J. & White, A. (1987). The pricing of options on assets with stochastic volatilties, *Journal of Finance* **42**: 281–300.

Jackwerth, J. C. & Rubinstein, M. (1996). Recovering probability distributions from option prices, *Journal of Finance* **51**: 1611–1631.

Karatzas, I. & Shreve, S. (1988). *Brownian Motion and Stochastic Calculus*, Springer-Verlag, Berlin et al.

Kloeden, P. E., Platen, E. & Schurz, H. (2003). *Numerical Solution of SDE Through Computer Experiments*, 3 edn, Springer-Verlag, Berlin et al.

Kmenta, J. (1997). *Elements of Econometrics*, 2 edn, The University of Michigan Press, Michigan.

Ledoit, O., Santa-Clara, P. & Yan, S. (2002). Relative pricing of options with stochastic volatility, *Working paper*, UCLA.

Lee, R. W. (2002). Implied volatility: statics, dynamics, and probabilistic interpretation, *Working paper*, Stanford University.

Longstaff, F. A. (1995). Option pricing and the martingale restriction, *Review of Financial Studies* **8**(4): 1091–1124.

Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis*, 2 edn, Springer-Verlag, Berlin et al.

Lyons, T. (1997). Derivatives as tradable assets, RISK 10th Anniversary Global Summit, Nov. 18-19, 1997.

Mandelbrot, B. (1963). The variation of certain speculative prices, *Journal of Business* **36**: 394–419.

Mayhew, S. (1995). Implied volatility, *Financial Analysts Journal* (4): 8–20.

Merton, R. C. (1973). Theory of rational option pricing, *Bell Journal of Economics and Management Science* **4**: 141–183.

Miller, A. J. (1990). *Subset Selection in Regression Models*, Chapman and Hall, New York.

Natenberg, S. (1994). *Option Volatility and Pricing*, Probus Publishing, Chicago, Cambridge.

Øksendal, B. (1998). *Stochastic Differential Equations*, 5 edn, Springer-Verlag, Berlin et al.

Press, W. H., Teucholsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical*

Recipes in C, 2 edn, Cambridge University Press, Cambridge.

Reynolds, D. (2001). A framework for scenario generation, *Algo Research Quarterly* **4**(3): 15–35.

Rosenberg, J. V. (2000). Implied volatility functions: a reprise, *Journal of Derivatives* **7**(3): 51–64.

Roth, R. (1997). Die Bestimmung des At-the-money-Punktes europäischer Optionen - Implikationen für die Einführung neuer Basispreise an der DTB, *Working paper*, Technische Universität Dresden.

Rubinstein, M. (1994). Implied binomial trees, *Journal of Finance* **69**: 771–818.

Sachs, L. (1972). *Statistische Auswertungsmethoden*, Vol. 3, Springer-Verlag, Berlin et al.

Schönbucher, P. J. (1999). A market model for stochastic implied volatility, *Philosophical Transactions of the Royal Society* **357**(1758): 2071–2092.

Schwert, W. (1990). Stock market volatility, *Financial Analysts Journal* **46**: 23–34.

Scott, L. O. (1987). Option pricing when the variance changes randomly: theory, estimation, and an application, *Journal of Financial and Quantitative Analysis* **22**(4): 419–438.

Skiadopoulos, G., Hodges, S. & Clewlow, L. (1999). The dynamics of the S&P 500 implied volatility surface, *Review of Derivatives Research* **3**: 263–282.

Stein, E. & Stein, J. (1991). Stock price distributions with stochastic volatility: An analytical approach, *Review of Financial Studies* **4**(4): 727–752.

Taylor, S. J. (1994). Modeling stochastic volatility: A review and comparative study, *Mathematical Finance* **4**(2): 183–204.

Taylor, S. J. (2001). Volatility, option information and empirical finance, PhD Course Aaarhus Business School.

Tompkins, R. G. (1994). *Options Analysis*, Probus Publishing, Chicago, Cambridge.

Wallmeier, M. (2003). *Der Informationsgehalt von Optionspreisen*, Physica-Verlag, Heidelberg.

Wiggins, J. B. (1987). Option values under stochastic volatility - theory and empirical estimates, *Journal of Financial Economics* **19**: 351–372.