The time-variation of volatility and the evolution of expectations

van der Weide, R.

Chapter 4

Volatility: Expectations and realizations

"Not to be absolutely certain is, I think, one of the essential things in rationality" (Bertrand Russell, 1947).

4.1 Introduction

At the heart of financial markets are market expectations. Embedded in option prices are market expectations regarding future volatility of the asset on which the option contract is written. While the assumption of rational expectations has been a popular (as well as convenient) paradigm, it is difficult to ignore the subjective nature of expectations as evidenced in recent empirical studies. Han (2008) and Constantinides *et al.* (2009) denote two such examples for the options market. "A growing literature shows that S&P 500 options are mispriced or not efficiently priced relative to a large class of rational option pricing models ... These lead to calls for research outside traditional rational option pricing" (Han, 2008; page 387). The proposed way forward is to extend the option pricing framework "to incorporate imperfect market and/or imperfect rationality" (Han, 2008; page 410), which will ultimately fit the empirical data better.[1]

[1] Han (2008) finds that investor sentiment, for example, plays an important role in determining option prices; "there is no guarantee that sentiment-induced mispricing will get corrected over a given horizon because of unpredictable investor sentiment in the future" (Han, 2008; page 390). As is to be expected, "the impact of sentiment is stronger when there are more limits to arbitrage" (Han, 2008; page 408).

The objective of this chapter is to make market expectations visible as they evolve over time, and to price options in line with prevailing expectations, be they rational or non-rational. Like financial volatility, market expectations are hidden. If not rational, market expectations can take on many different forms. Capturing expectations accurately is key for pricing options accurately. The expectation formation process denotes an integral part of the option pricing model.

Consider a European Call option contract that gives the owner the right to buy a specified quantity of an asset on which the contract is written (the underlying) at a specified future time (the expiration date) for a specified price (the strike price). Given the current price of the underlying asset, and the risk-free interest rate prevailing until expiration, the value of the option contract will be determined by the anticipated future volatility of the asset price. Agents trading option contracts are speculating on future volatility, in effect making the option market a market for financial volatility. "Because option value depends critically on expected future volatility, the volatility expectation of market participants can be recovered by inverting the option valuation formula" (Dumas, Fleming and Whaley, 1998).

Traditionally, implied volatility expectations are obtained by inverting the Black-Scholes and Merton (BSM) option pricing model that is built on the assumption that volatility fluctuates deterministically over time. The BSM model can be inverted for each strike price, but it predicts that all implied volatilities are the same. The well-documented smile and smirk pattern, that emerges when implied volatility is plotted against the strike price, contradicts the assumption of deterministic volatility.

We now know that volatility is best described by a stochastic process. Assuming that the market acknowledges the stochastic nature of volatility, implied volatility expectations will take on the form of a probability density function (pdf). Unlike traditional implied volatility, the implied pdf cannot be obtained from data on a single strike price. Instead, the option data for all strike prices jointly determine the implied pdf of volatility. So there can be no disagreement between different strike prices, there is only one implied pdf between them. The latter collapses to the traditional measure of implied volatility only in the event that the market believes volatility to be deterministic.

The 1987 stock market crash provides a great example where market expectations can be seen to change over time, arguably due to learning behaviour. Constantinides *et al.* (2009) note that "before the crash, option traders were using average historical volatility to price options and were not actively forecasting volatility changes". Although justifiable at the time perhaps, it does not sit well with rational expectations; "option traders were extensively using the BSM pricing model and the dictates of this model were imposed on the option prices even though these dictates were not necessarily consistent with the time series behavior of index prices" (Constantinides *et al.*, 2009; page 1271). Market participants like most humans are creatures of habit, until they are shaken out of their comfort zone. In their studies of this transition, Jackwerth and Rubinstein (1996) and Bates (2000), among others, compare the subjective distribution of asset prices obtained from option prices to the objective distribution. They find that through the eyes of the market, the volatility process before and after the crash of 1987 is fundamentally different. As the objective distributions show no sign of such change in the volatility process, this suggests that the market has changed instead.[2] Market expectations are more aligned with objective data after the crisis. Yet, recent findings by Constantinides *et al.* (2009) still "cast doubts on the hypothesis that the option market is becoming more rational over time, particularly after the crash" (Constantinides *et al.*, 2009; page 1271).

We put forward an analytically convenient option pricing framework that accommodates both stochastic volatility and asymmetric volatility. The latter, also known as the Fischer Black effect, captures an important stylized fact of asset prices, namely that volatility rises in times of negative price changes and drops in times of positive price changes. Option pricing models that do not accommodate this feature are found to be seriously misspecified (see e.g. Andersen *et al.*, 2002). The way we incorporate asymmetric volatility may be viewed as the simplest possible extension of the Hull and White (1987) stochastic volatility model. It offers an alternative to the Heston (1993) model. Both our model and Heston (1993) are nested as a special case in the generalized Black-Scholes model of Garcia *et al.* (2003a, 2010) (see also Garcia and Renault, 2001, and Garcia *et al.*, 2003b).

Daily estimates of the implied pdf of volatility are obtained by estimating the option pricing model one day at a time.[3] This means that we choose not to capture the dynamics of the volatility process, as perceived by the market, into a parametric model and then estimate the model parameters using the time series of option prices (from which one could re-construct the implied pdfs of volatility). Instead we keep the volatility model as simple as possible. Our sole interest is the implied distribution of volatility which

[2]Note that over time, agents may change their beliefs as well as their preferences. They may learn about the volatility process, and they may adjust their risk-aversion. If either of these changes happened abruptly, and were large enough, both might in principle be able to explain the findings linked to the crash of 1987. In this chapter, our interest is to inspect the subjective expectations held by the market as they evolve over time during every-day-trading-days, in a post-crash period, and assess market rationality.

[3]Estimation is based on the cross-section of option prices for different strike prices.

can be estimated directly, without having to first identify the model that underlies the distribution of volatility. Each new day's implied distribution of volatility is treated as a new (independent) parameter, whose estimate is not tied to estimates for other days in the sample. We worry not whether our simplistic model for volatility can rationalize the obtained time-series of implied pdfs. Market expectations are driven by many factors, including sentiment, therefore it is unlikely that any one model will be able to capture its time-variation. We explicitly do not want to impose too much structure a priori on how the market updates its expectations over time, but allow market expectations to take their course, and permit them to deviate freely from rational expectations.

Note that volatility here refers to the average variance over the remaining lifetime of the option. Even though the implied pdf of volatility is the natural measure to consider for volatility expectations in a world where volatility is stochastic, studies of the implied pdf of volatility and its time-variation are still uncommon. Upon replacing tradition implied volatility with the implied pdf of volatility, a popular empirical question that is worthwhile revisiting is whether volatility expectations contained in option prices yield better predictions of future volatility than predictors based on past (realized) volatility (see e.g. Busch *et al.* (2011), and the references therein).

Under the standard assumption that volatility risk is not priced, comparing the implied pdf to estimates of the objective pdf of average volatility opens the door for testing of market rationality. The objective pdf, however, is not easily estimated. For any given period in time we observe at best the level of average volatility that has been realized, but not the distribution from which it has been drawn. Yet, under suitable assumptions, daily obervations of realized volatility will lend itself for estimating moments of the objective pdf that can then be compared to moments of the subjective pdf. An example of such a moment is the degree of persistence in volatility, which determines how fast the volatility of average volatility declines when increasing the time to expiration. Similarly, estimates of realized volatility may be compared to the first moment of the implied pdf (although the former denotes a noisy estimate).

Our first empirical findings include: (i) the first moment of the implied pdf closely follows estimates of average realized variance, (ii) estimates of implied persistence in variance suggest that the market is fully aware of the fact that variance exhibits long-range dependence, and (iii) market expectations about future average variance appear to exhibit a degree of foresight.

In addition to fitting our option pricing model to the data, we include an analytic study of the implied variance function. The analytic expressions obtained show how the implied variance function is shaped by the model parameters. These results help with gauging the ability of the model to fit the empirical data, and may also be used in the estimation of the model parameters.

The remainder of the chapter is organized as follows. Section 4.2 presents our data. The theoretical framework is presented in Section 4.3 which includes the derivation of the option pricing model. Subsequently we study the implied variance function in Section 4.4. In Section 4.5 we discuss the method of estimation and provide a brief evaluation of model performance. Section 4.6 presents the empirical results. Finally, Section 4.7 concludes.

4.2 Data

4.2.1 Underlying value

For the underlying value we consider the German DAX stock index. The data used in our empirical illustration covers a period of one year: December 15, 2005 until December 15, 2006. We derive the price data from the most actively traded future on the DAX, which is the one with the shortest time to expiration. For the future's price we take the average between the bid and ask price recorded daily at 17:30 (the option prices too are recorded at 17:30).

The DAX stock index is corrected for dividends on the underlying stocks, which means that dividends are automatically reinvested. We view this as a convenient property. Alternatively, we would first need to derive the (expected) dividends until expiration ourselves, and correct the underlying prices accordingly.

Since we need not worry about dividends, the price of the underlying at time t denoted by S_t can be obtained by:

$$S_t = F_{t,T} e^{-r(T-t)}, \tag{4.1}$$

where $F_{t,T}$ denotes the future price at time t with expiration $T > t$, and where r denotes the instanteneous interest rate (which for the sake of simplicity is assumed constant). Our interest rate data consists of quotes on the 1-12 month Euribor rate. The time t anual interest rate for any $0 < t < T$ is obtained by means of linear interpolation.

4.2.2 Realized volatility

We construct a measure of realized variance for the underlying value by using tick-by-tick data on the DAX future(s) for the period December 15, 2005 until December 15, 2006. While the DAX future is traded from 9:00 until 20:00, the future is not traded very actively any more after the closing of stock trading at 17:30. For this reason we will focus on intraday future prices between 9:00 and 17:30.

Let $F_{t_{n,j}}$ denote the jth observation of the future price at day n corresponding to time t_n.[4] The length of the interval between subsequent obervations measures two minutes, which yields a total of 255 observations per day. The sum of the intraday squared returns will define our measure of quadratic variation rv_{t_n} at day n:

$$rv_{t_n} = \left(\log F_{t_{n,1}} - \log F_{t_{n-1,255}}\right)^2 + \sum_{j=2}^{255} \left(\log F_{t_{n,j}} - \log F_{t_{n,j-1}}\right)^2 .$$

The DAX future is among the most liquid futures on the European market which helps curb the effects of market microstructure, even at two minute intervals.

We use the daily realized variances to construct measures of average (anual) variance over periods with varying times to expiration $T-t$ (as t approaches T). A confirmation that realized variance provides accurate estimates of the stochastic variance process, such that standardized returns are indeed standard normal, can be found in e.g. Peters and de Vilder (2006). For a more elaborate discussion of realized variance we refer to Barndorff-Nielsen and Shephard (2002), and the references therein.

4.2.3 Option prices

Our option data consists of both puts and calls on the German DAX stock index, which are of the European type, with expiration date December 15, 2006 (the third Friday of the month). For each trading day between December 15, 2005 and December 15, 2006 (one year to expiration), we have both bid and ask prices recorded at 17:30 for a large series of strike prices. The number of strikes available varies by day. There tends to be no trade in options when the difference between their strike price and the value of the DAX has become too large. Also new strikes are introduced as the underlying value reaches new price levels over time.

[4]For notational convenience we supressed the indicator for expiration time.

Let us briefly summarize the construction of the option price data we use in the empirical illustration. For each strike price we have at most four daily observations:

$$(C_b, C_a, P_b, P_a) \tag{4.2}$$

where C_b, P_b (C_a, P_a) denote the bid (ask) price of the put and call option, respectively. For notational convenience we supressed the indicator for t. As there are no quoting obligations at the German option market, it is possible that the vector above is only partially observable. Provided there exists at least one bid price, and one ask price, the vector (C_b, C_a, P_b, P_a) can in principle be fully reconstructed using the put-call parity:

$$S + P^a - C^b = Ke^{-r(T-t)}, \tag{4.3}$$

where K denotes the strike price, and where r denotes the interest rate.

We estimate r by exploiting the put-call parity for strike prices for which we have both the call and the put price.[5] This yields an implied interest rate for each strike price. We take the average as our estimate for r. It follows that the estimated r closely matches the trend of the (interpolated) Euribor rate. We do observe a small bias. For much of the period, our estimate for r is slightly lower (2.7 versus 3 percent at the start of the period in December, 2005; 3 versus 3.5 half a year later). The two converge however toward expiration. Our estimate for r is used in the estimation of the option pricing model.

As puts and calls are directly related, we focus on one of them, which will be the call price. We either use the actual quote on the call option, or the price derived from the corresponding put using the put-call parity. If both are available, the one with the smaller bid-ask spread is used. We decided to drop option prices if they fall below 50 cents. Once the appropriate adjustments have been made, the price of the call we will work with is the mid-price:

$$C(t, T, K) = \frac{C_b(t, T, K) + C_a(t, T, K)}{2}. \tag{4.4}$$

Let N_t denote the number of option prices available at time t, and let $K_i(t, T)$ with $i = 1, \ldots, N_t$ denote the corresponding strike prices. As we will be estimating implied pdf's (of average volatility) for each day t separately, it is important that N_t is sufficiently large as it denotes the number of observations available to construct these daily estimates. Let m_t (M_t) measure the minimum (maximum) strike price as a fraction of the price of

[5]We exclude options whose price is less than 50 cents and options that are far out of the money (with a log moneyness below -0.10).

the underlying at day t:

$$m_t = \min\{K_1(t,T), \ldots, K_{N_{t,T}}(t,T)\}/S_t \quad (4.5)$$

$$M_t = \max\{K_1(t,T), \ldots, K_{N_{t,T}}(t,T)\}/S_t. \quad (4.6)$$

Expiry	# Days	$\bar{N}$	$\bar{m}$	$\bar{M}$
2006-12	257	26.2	0.852	1.184

Table 4.1: *Summary statistics of option price data set*

Table 4.1 presents the summary statistics for our option price data set; it shows the total number of trading days, the average number of observations per trading day, and the average of m_t and M_t. Note that averages are taken over the total number of trading days.

4.3 Option pricing framework

4.3.1 Model for underlying value

This section introduces our model for the stock price process. We formulate our model in discrete time, which provides an analytically convenient framework that is accommodating to asymmetric volatility.[6] The model is built on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $0 = t_0 < \ldots < t_N = T$ denote the timepoints at which we observe price quotes. For notational convenience we assume they are equidistant[7], $\Delta t = t_n - t_{n-1}$. At each timepoint t_n the information available to the market is denoted by $\mathcal{F}_n$, $n = 0, \ldots, N$. Formally, $\mathcal{F}_n, n = 0, \ldots, N$ is an increasing sequence of σ-algebras with $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_N = \mathcal{F}$.

Conditional on $\mathcal{F}_{n-1}$ the price of a stock S_n is determined by:

$$S_n = S_{n-1} \exp\left(\mu_n \Delta t - \frac{V_n}{2}\Delta t + (\beta V_n - \gamma_n)\Delta t + V_n^{\frac{1}{2}} \Delta t^{\frac{1}{2}} U_n\right) \quad (4.7)$$

$$V_n = h_n(W_n), \quad (4.8)$$

where $V_0, \ldots, V_N$ denotes the variance process, which is assumed bounded in L^2. The random variables U_n and W_n, $n = 1, \ldots, N$, are measurable with respect to $\mathcal{F}_n$. Conditional

[6]For arguments in favour of discrete time, other than analytical convenience, see e.g. Brennan (1979) and Rubinstein (1976a).

[7]It is straightforward to generalize this to allow for non-equidistant timepoints.

on $\mathcal{F}_{n-1}$ the random variable U_n is normally distributed with mean zero and unit variance. The variable W_n is independent of $\mathcal{F}_{n-1}$. Moreover, the sequences $(U_n)_{n=1}^N$ and $(W_n)_{n=1}^N$ are independent.

Assumption 4.1 *The processes* $(\gamma_n)_{n=0}^N$ *is deterministic and satisfies:*

$$\gamma_n = \Delta t^{-1} \log(E[\exp(\beta V_n \Delta t)|\mathcal{F}_{n-1}]), \tag{4.9}$$

which yields:

$$\mu_n = \Delta t^{-1} \log E[(S_n/S_{n-1})|\mathcal{F}_{n-1}]. \tag{4.10}$$

With Assumption 4.1 we rule out the possibility of arbitrage opportunities where persistently low (high) values of V_n would yield predictably higher (lower) returns. While Assumption 4.1 places strong restrictions on the stochastic volatility model, it is found to be effective in the sense that the framework accommodates asymmetric (stochastic) volatility, fits the empirical data well (as we will show later), and preserves analytical convenience. For a number of popular stochastic volatility models, for which γ_n would be stochastic, the option pricing formula may be viewed as an accurate approximation where it is no longer exact.[8]

β governs the degree of asymmetric volatility.[9] For non-zero β we have that stock returns are correlated with the variance process. Negative values are consistent with the Fischer Black effect, a key stylzed fact of the stock price process: Volatility is high in times of negative price changes, and volatility is low in times of positive price changes.

V_n measures the variance of the log asset price return when V_n is known. In what follows we will derive quadratic variation rv_{t_n} and treat this as realizations of V_n. Note that at any time prior to time t_n both V_n and rv_{t_n} are random variables. While we observe rv_{t_n} *ex post*, we do not observe realizations of V_n. Instead we estimate the distribution of V_n conditional on the information that is available at times prior to t_n, and then consider the first moment of this distribution as an estimate for V_n (and compare this to rv_{t_n}). Conditional on $\mathcal{F}_{n-1}$

[8]Examples of such stochastic volatility models are the autoregressive models and simple regime switching models for parameter values that correspond to a strong persistence in volatility.

[9]Asymmetric volatility can be captured in different ways. In our model the volatility shocks that correlate with the asset return are modeled as deviations from the 'local mean'. This choice is largely motivated by analytic convenience; it follows that the latent stochastic volatility process in this case only impacts option prices via its average over time to expiration. The stochastic volatility model of Heston (1993) denotes a well-known alternative where volatity shocks are modeled as incremental changes in volatility. While this imposes fewer restrictions on the volatility process, the pricing of options now require integration over the joint distribution of average volatility and the volatility at expiration.

we have that $var[\log(S_n/S_{n-1})|\mathcal{F}_{n-1}] = E[V_n|\mathcal{F}_{n-1}]\Delta t + (\beta - \frac{1}{2})^2 var[V_n|\mathcal{F}_{n-1}]\Delta t^2$. The second term however can be made arbitrarily small relative to the first term by choosing Δt arbitrarily small, so that $var[\log(S_n/S_{n-1})|\mathcal{F}_{n-1}] \approx E[V_n|\mathcal{F}_{n-1}]\Delta t$.

Average variance V is given by:

$$V = \sum_{n=1}^{N} V_n/N. \tag{4.11}$$

Let φ denote the probability density function of V. The first two moments of φ will be denoted by $\lambda = E[V]$ and $\nu = var[V]$. Let us define $\gamma = T^{-1}\log(E[\exp(\beta TV)])$. If the variance of V is positive then the distribution of stock price returns is leptokurtic. It follows that skewness largely depends on the parameter β.

Without loss of generality, we pick $t_0 = 0$ as our reference point at which options with expiration time T will be priced.

4.3.2 Pricing of options

When "trading takes place only at discrete intervals, it is in general not possible to construct a portfolio containing the contingent claim and the underlying asset in such proportions that the resulting portfolio return is non-stochastic" (Brennan, 1979).[10] But even in continuous time, when volatility is stochastic and the corresponding risks cannot be hedged, markets are incomplete, such that we cannot price derivatives via risk-free replicating portfolios, as can be done in e.g. Black and Scholes (1973) and Cox, Ross and Rubinstein (1979). In the more realistic setups where volatility is stochastic and trading is not continuous, it is however still possible to derive option pricing formulas. It follows that preferences and consumption will then play a role.

The Stochastic Discount Factor (SDF) provides a general framework for asset pricing. The fundamental equation states that the price P_n of an asset at time t_n with payoff G_{n+1} solves:

$$P_n = E\left[m_{n+1}G_{n+1}|\mathcal{F}_n\right], \tag{4.12}$$

where the (positive) random variable m_{n+1} denotes the SDF. Expectations are taken over the objective probability distribution. The SDF m_{n+1} is used to price all assets (including derivatives), with the bond price $B_{n,n+1}$ ($E[m_{n+1}] = B_{n,n+1}$) and the stock price S_n

[10]For conditions under which the Black-Scholes formula can be obtained within a discrete-time model see e.g. Merton (1973) and Rubinstein (1976b). They assume, however, that volatility is non-stochastic.

($E[m_{n+1}S_{n+1}] = S_n$) nested as obvious special cases. For a recent overview on option pricing see Garcia *et al.* (2003a, 2010), who adopt the SDF framework to derive a generalized Black-Scholes model.

In a multi-period setting the SDF is given by the concatenation of the one-period SDFs:

$$m_0^N = \prod_{n=1}^{N} m_n. \tag{4.13}$$

Accordingly, the price of a contingent claim P_0 at t_0 with payoff $G(S_N)$ is given by:

$$P_0 = E\left[m_0^N G(S_N)|\mathcal{F}_0\right]. \tag{4.14}$$

Consider the following two assumptions on the SDF which we borrow from Garcia *et al.* (2003a):

Assumption 4.2 *The distribution of* $(\log(m_n), \log(S_n/S_{n-1}))$ *conditionally on* $\mathcal{F}_{n-1}$ *and* W_n *is bivariate normal for* $n = 1, \dots, N$.

Assumption 4.3 *The pair* $(\log(m_n), \log(S_n/S_{n-1}))$ *is independent of* $\mathcal{F}_{n-1}$ *given* $W_1, ...W_{n-1}$ *for* $n = 1, \dots, N$.

It follows that when Assumptions 4.2 and 4.3 are satisfied, and $U_1, \dots, U_N$ are i.i.d. normal, contingent claims can be priced via risk neutral valuation relationships (RNVR), which permits the following convenient representation (see Garcia *et al.*, 2003a):

$$P_0 = e^{-rT} E_{\mathbb{Q}}\left[G(S_N)|\mathcal{F}_0\right]. \tag{4.15}$$

Here $\mathbb{Q}$ denotes the risk-neutral probability measure. Under this new measure the drift $\exp(\sum_{n=0}^{N} \mu_n \Delta t)$ of the stock price process is shifted to $\exp(rT)$.

Albeit not explicitly, Assumptions 4.2 and 4.3 make assumptions about investors preferences. By definition the SDF is determined by:

$$m_{n+1} = \beta \frac{u'(C_{n+1})}{u'(C_n)}, \tag{4.16}$$

where C_n denotes aggregate consumption at time t_n, and u the investor's utility function. Equation (4.16) can be derived from optimizing utility. Note that Assumption 4.2 is satisfied when we assume a power utility function, and that returns on aggregate wealth $\log(C_n/C_{n-1})$ equal returns on the stock $\log(S_n/S_{n-1})$. Brennan (1979) was the first to show that these are necessary and sufficient for RNVR in a one-period setting[11]. Assumption 4.3 is included to extend these results to the multi-period case. Note that we implicitly make the standard assumption that volatility risk is noncompensated.

The following theorem derives the price of an European call option, given our model for the underlying value, as a function of β and φ, and the Black-Scholes option pricing formula.

Theorem 4.1 *Given our model for the underlying value, and Assumptions 4.1, 4.2 and 4.3, we have that the price of a call* $C_\varphi(K,S,r,T) = e^{-rT}E_{\mathbb{Q}}[(S_T-K)^+]$ *at time* $t_0 = 0$ *with expiration date* T *and strike price* K *solves:*

$$\begin{aligned} C_\varphi(K,S,r,T) &= E_V[e^{(\beta V-\gamma)T}C_{BS}(K,S,r(V),T,V)] && (4.17)\\ &= \int_0^\infty e^{(\beta v-\gamma)T}C_{BS}(K,S,r(v),T,v)\varphi(v)dv, && (4.18)\end{aligned}$$

with $r(v) = r + (\beta v - \gamma)$*, and where* $C_{BS}(K,S,r(v),T,v)$ *denotes the Black-Scholes option pricing formula with volatility* $\sqrt{v}$*:*

$$\begin{aligned} C_{BS} &= S\Phi(d_+) - Ke^{-r(v)T}\Phi(d_-) && (4.19)\\ d_\pm &= (\ln(S/K) + (r(v) \pm \frac{v}{2})T)/\sqrt{vT}, && (4.20)\end{aligned}$$

with Φ *the standard normal distribution function.*

Proof. The proof follows directly from Garcia *et al.* (2003a, 2010). It can be verified that our model is nested as a special case of their generalized Black-Scholes model (GBS); equations (4.7) and (4.8), and Assumptions 4.1 to 4.3, are seen to satisfy Assumptions 2.2 to 2.4 of Garcia *et al.* (2003a).[12] The GBS solves (see eq. (2.16) in section 2.6.2 of Garcia

[11] Assumption 4.2 thus implicitly assumes that preferences can be described by a power utility function, as Brennan (1979) does. Yet, Brennan (1979) assumes these preferences only to establish conditional joint lognormality of (m_{n+1}, S_{n+1}), which allows him to derive RNVR. For this reason Garcia *et al.* (2003a, 2010) directly assume Assumption 4.2 instead. Note that Assumption 4.3 is also borrowed from Garcia *et al.* (2003a, 2010).

[12] In Garcia *et al.* (2010), the same assumptions are implied by the model given in eq. (2.27) and (2.28) in section 2.6.

et al., 2003a):[13]

$$C_0 = E_0[\xi_{0,T} S_0 \Phi(d_+) - Ke^{-rT}\Phi(d_-)], \tag{4.21}$$

where $\xi_{0,T}$ solves:

$$\begin{aligned} \xi_{0,T} &= E\left[m_0^N \frac{S_N}{S_0} | W_1, ..., W_N\right] && (4.22)\\ &= e^{-rT} E_{\mathbb{Q}}\left[\frac{S_N}{S_0} | W_1, ..., W_N\right] && (4.23)\\ &= e^{(\beta V - \gamma)T}, && (4.24) \end{aligned}$$

with $E[\xi_{0,T}] = 1$. The first $E[.]$ takes expectations over the joint distribution of the stochastic discount factor m_0^N and the log-return S_N/S_0. In the second step we make use of the RNVR that holds due to Assumptions 4.2 and 4.3 (see eq. (2.8) and (2.9) on page 12 of Garcia *et al.*, 2003a). Substituting this into equation (2.16) from Garcia *et al.* (2003a) then yields:

$$C_\varphi(K, S, r, T) = SE_V\left[e^{(\beta V - \gamma)T}\Phi(d_+) - Ke^{-rT}\Phi(d_-)\right], \tag{4.25}$$

where expectations are taken over stochastic (average) variance V. After rearranging terms we obtain equation (4.17). ■

It can be verified that our model is nested as a special case of the generalized Black-Scholes model (see e.g. Garcia *et al.*, 2010). Note that for $\beta = 0$ we obtain the Hull and White (1987) option pricing model.[14]

4.4 Implied variance function

Let us define money-ness by $x = \log(\frac{K}{\exp(rT)S_0})$. It will also be convenient to define:

$$\begin{aligned} c_{BS}(x, T, v) &= C_{BS}(K, S, r, T, v)/S\\ &= \Phi(d_+(x, T, v)) - e^x\Phi(d_-(x, T, v)), \end{aligned}$$

[13]The original equation (2.16) features the ratio $B^*(0,T)/B(0,T)$ where $B^*(0,T)$ denotes the bond price conditional on the latent state variable W. For our model $B^*(0,T) = B(0,T)$.

[14]For an early overview on the empirical performance of alternative option pricing models, which include the popular Hull and White (1987) and Heston (1993) models, see e.g. Bakshi *et al.* (1997).

where $d_{\pm}(x,T,v) = -\frac{x}{\sqrt{vT}} \pm \frac{1}{2}\sqrt{vT}$. Similarly, we define c_φ by:

$$\begin{aligned}
c_\varphi(x,T) &= C_\varphi(K,S,r,T)/S \\
&= E_V[e^{(\beta V-\gamma)T} C_{BS}(K,S,r(V),T,V)/S] \\
&= E_V[e^{(\beta V-\gamma)T} c_{BS}(x_V,T,V)] \\
&= E_V[e^{(\beta V-\gamma)T} \left(\Phi(d_+(x_V,T,V)) - e^{x_V}\Phi(d_-(x_V,T,V))\right)] \\
&= E_V[e^{(\beta V-\gamma)T}\Phi(d_+(x_V,T,V)) - e^{x}\Phi(d_-(x_V,T,V))],
\end{aligned}$$

where $x_V = x - (\beta V - \gamma)T$. (Note that in obtaining the last equation we used that: $e^{(\beta V-\gamma)T}e^{x_V} = e^x$.)

Definition 4.1 *The implied variance function $I(x)$ for the option pricing model from Theorem 4.1 is defined as the solution to the following integral equation:*

$$\begin{aligned}
c_{BS}(x,T,I(x)) &= \int_0^\infty e^{(\beta v-\gamma)T} c_{BS}(x-(\beta v-\gamma)T,T,v)\varphi(v)dv \\
&= c_\varphi(x,T).
\end{aligned}$$

The function $I(x)$ is single-valued for each x, which follows from the invertibility of c_{BS} as a function of variance v.

Next we will derive selected properties of $I(x)$. As symmetric volatility ($\beta = 0$) versus asymmetric volatility ($\beta \neq 0$) denotes an important distinction, both analytically and empirically, we will discuss them in separate subsections.

4.4.1 Symmetric stochastic volatility

In this subsection we will assume $\beta = 0$.

Theorem 4.2 *The implied variance function $I(x)$ is quasi-convex:*

$$\begin{aligned}
\partial I(x)/\partial x < 0 &\quad if \quad x < x^* \\
\partial I(x)/\partial x > 0 &\quad if \quad x > x^*,
\end{aligned}$$

where x^* denotes the unique minimum that is attained at $x^* = 0$.

Proof. To proof follows from Theorem 4.2 of Renault and Touzi (1996), in which it is assumed that volatility risk is noncompensated (Assumption 2.3), as it is in our framework. ∎

Theorem 4.3 *The implied variance function $I(x)$ is symmetric:*

$$I(x) = I(-x) \quad \forall x. \tag{4.26}$$

Proof. The proof follows directly from Proposition 3.1 of Renault and Touzi (1996). ■

Theorem 4.2 and 4.3 show that $I(x)$ is both quasi-convex and symmetric for $\beta = 0$. Note that $I(x)$ need not be a convex function, as is illustrated by the following example.

Example 4.1 *Suppose that the pdf of V is given by:*

$$\varphi(v) = \frac{1}{v_1 - v_0} 1_{[v_1, v_0]}(v), \qquad v_1 > v_0 > 0. \tag{4.27}$$

Then $v_0 < I(x) < v_1$, and hence the function $I(x)$ cannot be convex.

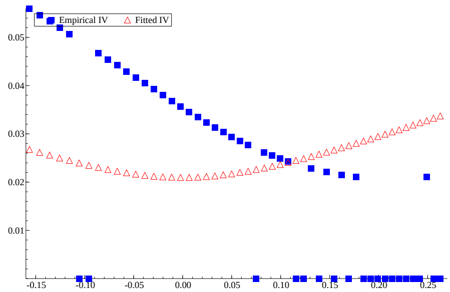

Figure 4.1: *Mismatch when imposing symmetric volatility (June 23, 2006)*

It follows that the symmetry and the location of the minimum of the implied variance function do not depend on φ. Provided that φ has positive dispersion (the market believes in stochastic volatility), we will observe a smile where the minimum is attained at $x^* = 0$. While the empirical implied variance function exhibits a smile pattern, the theoretical implied variance function in case of symmetric volatility ($\beta = 0$) typically does not fit the empirical data well. Figure 4.1 shows a typical mismatch. Two observations are apparant. First, and most notably, the horizontal alligment of the theoretical miminum ($x^* = 0$) does not match with the empirical minimum of the smile. Second, while the theoretical smile is always symmetric, the empirical smile is not.

4.4.2 Asymmetric stochastic volatility

Asymmetric volatility (obtained for $\beta \neq 0$) will introduce both asymmetry in the implied variance function and a shift in where it attains its minimum.

Conjecture 4.1 *The minimum of $I(x)$ satisfies:*

$$x^* \approx -2\beta T I(x^*). \tag{4.28}$$

If the relationship from Conjecture 4.1 is not exact, our simulation results suggest that it holds as a close approximation. Assuming that Conjecture 4.1 holds true in general, it confirms that for $\beta < 0$ the minimum of the implied variance function is attained at $x^* > 0$, which is consistent with the empirical data. Given that β and $I(x^*)$ are finite, the conjecture also predicts that the location of the minimum tends to zero ($x^* \to 0$) when time to expiration tends to zero ($T \to 0$).

Corollary 4.1 *Assuming Conjecture 4.1 holds true and $|\beta| < \infty$, we have:*

$$\lim_{T \to 0} x^*(T) = 0. \tag{4.29}$$

For the empirical variance function we indeed see that the minimum moves closer to $x^* = 0$ as time to expiration T becomes smaller. Where the empirical minimum does not exactly respect $x^* \to 0$ for $T \to 0$, it holds approximately. Finally, note that $\beta = 0$ implies $x^* = 0$ for any time to expiration T, such that the conjecture is also consistent with what we know for $\beta = 0$.

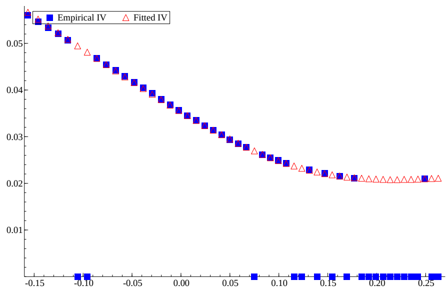

Figure 4.2: *Fit for asymmetric empirical smile (June 23, 2006)*

Figure 4.2 plots the empirical implied volatility function together with the fitted theoretical values where we allowed for $\beta < 0$. We used the same empirical data as in Figure 4.1. Now that asymmetric volatility is accounted for, we observe a nearly perfect match. The minimum of the implied variance function perfectly coincides with the empirical minimum (compare Figure 4.1 with Figure 4.2), and also the skewness introduced for $\beta < 0$ is in line with the empirical smile.

Conjecture 4.1, assuming it holds true, also suggests that we can estimate β as a function of the minimum $(x^*, I(x^*))$ of the empirical implied variance function: $\hat{\beta} = -\frac{1}{2}\frac{x^*}{TI(x^*)}$.

Next we will derive the asymptotic left and right slopes of the theoretical implied variance function. If both the left and right slope coefficients can be observed for the empirical implied variance function, they too can be used to construct an estimate for β.

Let α_R and α_L measure the slopes of linear "asymptotes" to implied variance:

$$\alpha_R(T) := \lim_{x\to\infty} \sup \frac{I(x,T)}{|x|/T} \tag{4.30}$$

$$\alpha_L(T) := \lim_{x\to-\infty} \sup \frac{I(x,T)}{|x|/T}. \tag{4.31}$$

Using this notation, the tail slopes of $I(x)$ (in absolute value) will be α_R/T and α_L/T. For any model for the underlying value S_T, the coefficients α_R and α_L belong to the interval $[0, 2]$ (see Lee, 2004), which confirms that $I(x)$ becomes flat when T tends to infinity. Given a model, the exact values are entirely determined by the following two moments of the distribution of S_T:

$$p := \sup\{p : E[S_T^{1+p}|F_0] < \infty\} \tag{4.32}$$

$$q := \sup\{q : E[S_T^{-q}|F_0] < \infty\}. \tag{4.33}$$

The relationship between (p, q) and (α_R, α_L) is given by (see Lee, 2004):

$$\alpha_R = 2 - 4\left(\sqrt{p^2+p} - p\right) \tag{4.34}$$

$$\alpha_L = 2 - 4\left(\sqrt{q^2+q} - q\right). \tag{4.35}$$

In the next theorem p and q are derived for our model.

Theorem 4.4 *Let $E[V] = \lambda$ and $var[V] = \nu$. In case average variance V is Gamma or Inverse Gaussian distributed, the moments p and q, as defined in eq. (4.32) and (4.33),* *are given by:*

$$p = (\sqrt{\xi} - \frac{1}{2}) - \beta \quad (4.36)$$
$$q = (\sqrt{\xi} - \frac{1}{2}) + \beta, \quad (4.37)$$

where

$$\xi = (\beta - \frac{1}{2})^2 + \frac{\lambda}{\nu T} c_\varphi. \quad (4.38)$$

with $c_\varphi = 1$ *in case of Inverse Gaussian, and* $c_\varphi = 2$ *in case of Gamma.*

Proof. The proof is given in the Appendix. ■

For $\beta < 0$ the result implies $\alpha_R > \alpha_L$, which is consistent with empirical data. Note that only for $\beta = 0$ we have $p = q$. In other words, $\alpha_R = \alpha_L$ if and only if $\beta = 0$, which denotes the symmetric case.

Corollary 4.2 *Let* φ *denote the Gamma or Inverse Gaussian pdf. Then, the degree of asymmetry* β *can be derived from:*

$$\beta = \frac{1}{2}(q - p), \quad (4.39)$$

where p *and* q *are uniquely determined from* α_R *and* α_L*. Subsequently,* λ/ν *can be solved from the expression for* q *(or* p*) from Theorem 4.4:*

$$\frac{\lambda}{\nu} = q^2 T c_\varphi^{-1} - 2q(\beta - \frac{1}{2}) T c_\varphi^{-1}. \quad (4.40)$$

Proof. The proof is straightforward, and therefore omitted. ■

It thus follows that the degree of asymmetry β can easily be derived from the empirical values of α_R and α_L, independent of the values of the other model parameters.

The next theorem provides an analytical approximation of $I(x)$ as a function of the moments of φ without making further assumptions about the functional form of φ.

Theorem 4.5 *Let* $E[V] = \lambda$*,* $var[V] = \nu$*, and* $\gamma = T^{-1}\log(E[\exp(\beta T V)])$*. Without further assumptions about the pdf* φ*, the implied variance function* $I(x)$ *can be approximated* *by:*

$$I(x) \approx \lambda + \frac{\kappa}{4v_0}\left(\frac{x^2}{v_0T} - 1 - \frac{v_0T}{4}\right) + \beta\kappa T\left(\beta + \frac{1}{2} + \frac{x}{v_0T} + (v_0T)^{\frac{1}{2}}(\beta - \frac{2\psi}{\kappa T})a(x)\right),$$

where

$$a(x) = \frac{\Phi(d_-(x))}{\phi(d_-(x))},$$

with $d_-(x) = \frac{1}{2}(v_0T)^{\frac{1}{2}} - x(v_0T)^{-\frac{1}{2}}$. *The functions* Φ *and* ϕ *denote the standard normal cumulative distribution and density function. Variance level* v_0 *is defined by:*[15]

$$v_0 = \begin{cases} \gamma/\beta & if \quad \beta \neq 0 \\ \lambda & if \quad \beta = 0 \end{cases}$$

The parameters (ψ, κ) *are given by:* $\psi = v_0 - \lambda$, *and* $\kappa = \nu + (v_0 - \lambda)^2$.

Proof. The proof is given in the Appendix. ■

In the symmetric volatility case ($\beta = 0$), the approximation of the implied variance function is indeed symmetric around $x = 0$ where it attains its minimum:[16]

$$I(x|\beta = 0) \approx \lambda + \frac{\nu}{4\lambda^2}\left(\frac{x^2}{T} - \lambda - \frac{\lambda^2}{4}T\right). \tag{4.41}$$

It follows that the curvature of the 'smile' is determined by $\nu/T\lambda^2$. The implied variance function becomes flat when $\nu/T \to 0$, while the smile will be most prominent close to expiration ($T \to 0$), or when ν is large. For $\nu = 0$, regardless of β, we know that $I(x) = \lambda$, which also holds for our approximation. ($\nu \to 0$ implies $v_0 \to \lambda$ and $\kappa \to 0$.)

The approximation for $I(x)$ is seen to share the important features of empirical implied variance. For $\beta \neq 0$ asymmetry is introduced by the asymmetric function $a(x)$. For $\beta < 0$ the minimum indeed shifts to the right. As may be expected, accuracy of the approximation is found to be highest for at- and around the money options, while divergence can be observed for far out of the money and far in the money options.

The analytical properties derived in this subsection show that the location and shape of the implied variance function is sensitive to the choice of φ. If we keep the first two moments of φ fixed but vary a third moment, the implied variance function will generally

[15] Note that $\beta \to 0$ implies $\gamma \to \beta\lambda \to 0$.
[16] This approximation for the symmetric case can also be found in Ball and Roma (1994).

alter its shape. That is not to say that the time-variation in φ can a priori not accurately be described by two degrees of freedom. Higher moments, which largely determine the type of distribution (e.g. Gamma, Inverse Gamma and Inverse Gaussian), may or may not be constant over time.

4.5 Estimation

This first subsection describes the estimation procedure adopted. The second subsection provides a brief evaluation of the goodness-of-fit of the option pricing model.

4.5.1 Loss-function for parametric estimator

For this chapter we consider a parametric estimator for the implied pdf φ. We assume a Generalized Inverse Gaussian (GIG) distribution for φ with three degrees of freedom, which has Gamma, Inverse Gamma and Inverse Gaussian nested as special cases.[17] β will be estimated jointly with φ.

Common loss functions measure either the errors in option prices or the errors in implied volatility.[18] Our experience is that minimizing the errors in option prices is numerically more attractive. Even so, we are keen to obtain an accurate fit of the implied volatility function. As our objective is to estimate the model parameters for each day separately, the number of observations will typically range between 20 and 40.

We choose to minimize the following loss function l:

$$l(\theta) = \sum_i \left(\log\left(\tau + C_{\varphi,i}(\theta)\right) - \log\left(\tau + C_i\right)\right)^2, \tag{4.42}$$

where θ denotes the parameter vector, $C_{\varphi,i}(\theta)$ denotes the model prediction, and where C_i denotes the observed option price at strike price K_i. $\tau > 0$ is included to curb the divergence of $\log(\tau + C)$ for small option prices. We have set $\tau = 5$ in our empirical example, but results are robust to the choice of τ provided it is not too close to zero. (To check whether the numerical optimization procedure converged to a global minimum, and not a local minimum, we repeat the procedure with different initial values.)

[17] Note that the GIG distribution is closed under aggregation, which denotes a convenient property. By averaging variance we will not leave the class of probability density functions.

[18] In view of consistency one may want to employ the same loss function for both estimation and evaluation (see e.g. Christoffersen and Jacobs, 2004).

Each of the model parameters is found to have a distinctive effect on the location and shape of the implied volatility function; the first moment of φ largely controls the vertical allignment. Curvature increases with the dispersion of φ, while β (the degree of asymmetry in volatility) largely governs the location of the minimum and the skewness of the implied volatility function. This should help with the identification of the parameters. It also helps that while the number of observations is limited, they are often found to provide an accurate description of the implied volatility function with little noise.

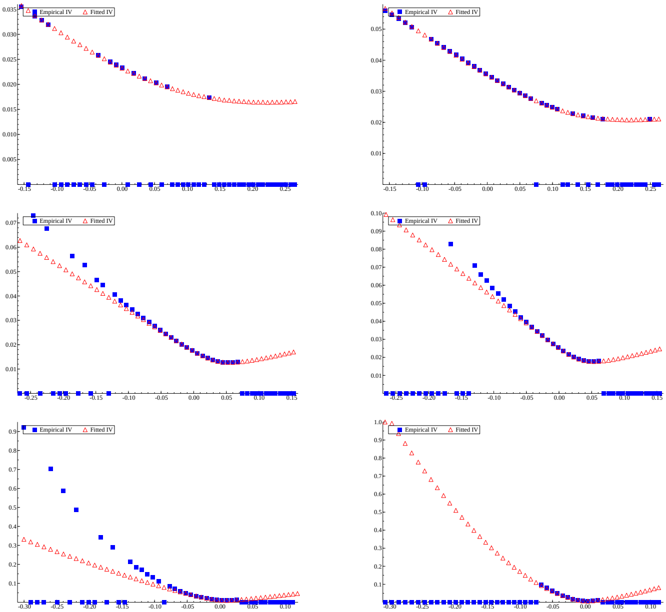

Figure 4.3: *Fitted empirical implied variance for: (a) 1 year (top-left); (b) half a year (top-right); (c) 6 weeks (middle-left); (d) 2 weeks (middle-right); (e) 3 days (bottom-left); and (f) 2 days (bottom-right) to expiration*

4.5.2 Model performance

This subsection provides a first evaluation of model performance. Figure 4.3 shows six fits for different, yet typical, implied variance functions. The time to expiration varies from two days in the top left figure to a year in the bottom right figure. It is stimulating to observe that our model, given its simplicity and analytical convenience, fits the data surprisingly well for all strike prices with longer maturities. For shorter maturities the best fits are observed for options that are not too far out of the money.[19]

When discrepancies between empirical and predicted values are observed, they are mostly at the far left or right end of the figure. By the same token the option price is not very sensitive to volatility for far in and out of the money options, such that the fit of the option price is still very good (which is not shown in the figure). Moreover, the right tail of the implied volatility function is often poorly represented in the empirical data. It refers to high values of the strike price where call option prices are small. Due to the lack of data it will generally be harder to fit this side of the implied volatility function. (Measurement error may also be more of an issue at this end since option prices are not quoted with infinite precision, which matters when prices are very small.) Overall, the model appears well equiped to handle the empirical smiles observed in practice.

4.6 Empirical results

This section shows by means of a modest empirical example how estimates of the implied pdf of volatility φ evolves over time. We assume a GIG distribution for φ with three degrees of freedom. The parameters are estimated jointly with β for each day of the sample. Our focus will be on the first two moments of the implied pdf φ: $E[V]$ and $sd[V]$. The estimates will comprise a time-series of market expectations concerning the distribution of future average variance. These will be compared with the time-series of realized variance.

The time-variation in (φ, β), as we move foward t_0, may reflect both the conditioning on new information over time and the variation in the model h_n (see eq. 4.8) that is adopted by the market. While these are different types of variation, they may well co-exist over time. We will not try to disentangle them, but focus on their joint outcome.

The first panel of Figure 4.4 shows the first moment ($E[V]$) of φ overlayed with realized variance (qv) as well as realized variance averaged over the time left to expiration ($avg[qv]$) for a period of one year with expiration at the end of the time-series. The second panel shows the dispersion of φ measured by the standard deviation ($sd[V]$).

Observation 4.1 *The first moment of the implied pdf of volatility closely matches average realized volatility.*

[19] We arguably need more degrees of freedom for the implied pdf of volatility to obtain a better fit of the implied variance function for shorter maturities.

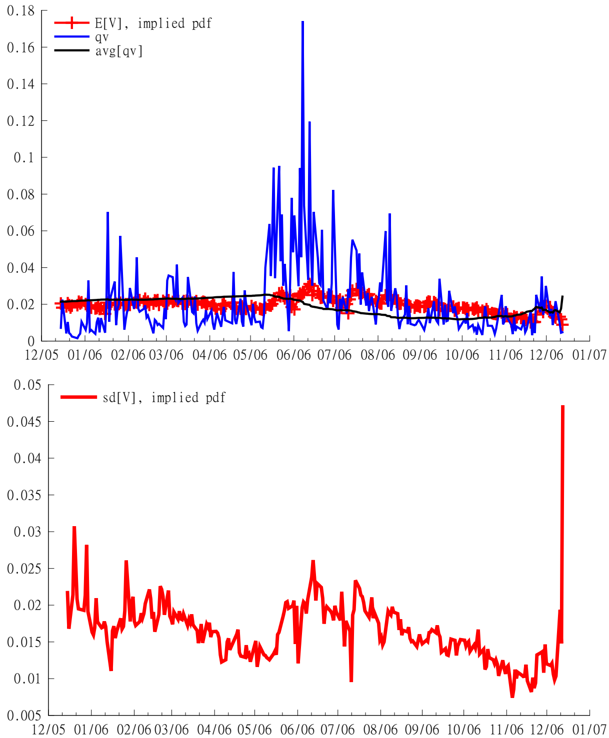

Figure 4.4: *Time-series of: (a) first moment of implied pdf E[V], realized variance qv, average realized variance avg[qv] (top panel), and (b) second moment of implied pdf sd[V] (bottom panel)*

It can be seen how $E[V]$ is particularly close to $avg[qv]$ during the first months of the year. Between six and seven months prior to expiration, realized variance jumps to a higher level and exhibits frequent peaks. $E[V]$ climbs during this period, and remains high for an extended period of time, noticably higher than $avg[qv]$. Expectations are nevetheless steadily reduced as realized variance continues to decline following the turbulent summer of 2006 (June to August). The fact that expectations do not immediately close the gap with average realized variance is consistent with a market contemplating the possibility that the remaining months leading up to expiration may bring a new surge in volatility. In November 2006, with 6 weeks left to expiration, expectations and realizations were fully aligned. Note that a modest rise in volatility emerged in December 2006 that was immediately picked up by the market.

Observation 4.2 *The implied volatility of average volatility (second moment of the implied pdf) does not tend to zero when the time to expiration is increased.*

This observation suggests that the market is fully aware of the fact the volatility exhibits strong persistence (long-range dependence). The variance of average variance is indeed expected not to show a rapid decline when we increase the time to expiration. Whether the subjective degree of volatility persistence matches objective estimates of persistence is an empirical question left for future research. In the bottom panel of Figure 4.4 it can be seen how $sd[V]$ moves up and down over the course of a year. While it increases during the weeks/days prior to expiration (as one would expect), it does not fade away as we move further away from expiration (where the period over which variance is averaged becomes larger).

Observation 4.3 *Market expectations about future average volatility appear to exhibit a degree of foresight.*

Figure 4.5 shows the one year period divided into three parts of about 4 months each. What is interesting is that expectations remain close to future average realized variance even in periods where present day realized variance is noticably lower (best seen in top panel) or noticably higher (best seen in middle panel). This may be an indication that the market is not fooled by current events as it correctly anticipates future levels of average variance.

Figure 4.6 zooms in on the last six weeks. We can see here how expectations respond faster to movements in realized variance closer to expiration. In the first three weeks of this period realized variance is seen to be on the decline, which is closely followed by market expectations, even though future average variance remains steady. Then, three weeks to expiration, realized variance rises to a higher level. This is again immediately picked up by the market which stays on top of movements in volatility until expiration. The bottom panel shows once more how the perceived volatility of average variance increases towards expiration, as one would expect, but that the trend is visibly not exponential.

Observation 4.4 *Estimates of β appear to diverge in the limit where time to expiration tends to zero.*

See Figure 4.7. Note that this occurs when the empirical minimum of the implied variance function does not tend to zero money-ness (i.e. β diverges if we do not have $x^* \to 0$ for

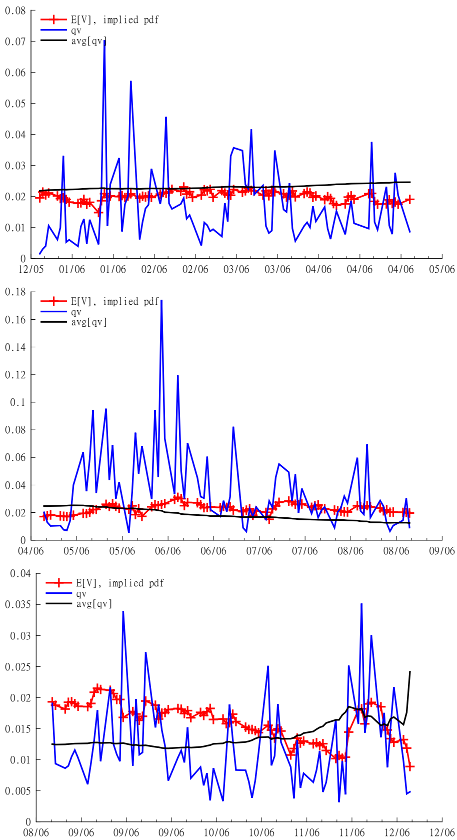

Figure 4.5: *Time-series of first moment of implied pdf E[V], realized variance qv, and average realized variance avg[qv] for: (a) end of Dec 2005 - beginning of May 2006 (top panel), (b) beginning of May 2006 - 3rd week of Aug 2006 (middle panel), and (c) 3rd week of Aug 2006 - 3rd week of Dec 2006 (bottom panel)*

$T \to 0$), see the section on the implied variance function. Whether this is particular for options on the DAX futures, or whether the minimum of the implied variance function does not tend to zero at expiration more generally remains an empirical question. In this case, our option pricing model appears better equipped to fit options with medium to long time to expiration. While our model still provides reasonable fits for options with short maturities, estimates of β in that case become unreasonably large.

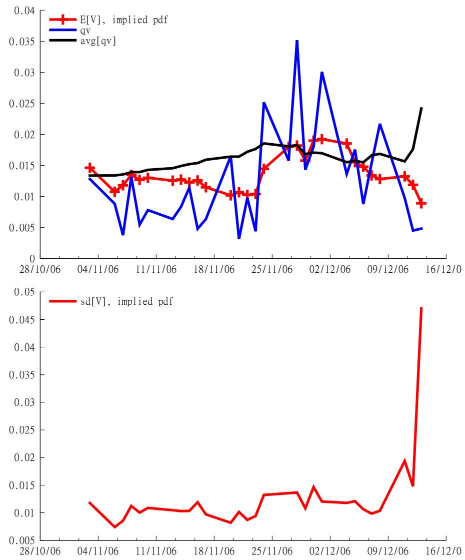

Figure 4.6: *Time-series of: (a) first moment of implied pdf E[V], realized variance qv, average realized variance avg[qv] (top panel), and (b) second moment of implied pdf sd[V] (bottom panel)*

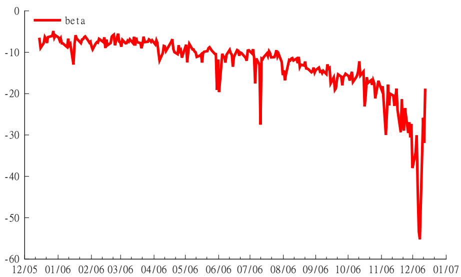

Figure 4.7: *Time-series of β*

4.7 Concluding remarks

The value of European options depends crucially on market expectations about the distribution of future average volatility. The options market may therefore be viewed as a market for volatility.

The objectives of this chapter are three-fold. First, we put forward an option pricing model that accommodates both stochastic and asymmetric volatility, that fits the data well, and preserves analytical convenience. Second, we advocate for the use of the implied pdf of volatility as an alternative to the traditional measure of implied volatility, which was introduced when volatility was still considered a deterministic process. The former is more meaningful in a world where volatility is stochastic. Third, we adopt the option pricing model to estimate the implied pdf of volatility for each day in the sample. The time-series of the implied pdf describes how market expectations about future volatility evolve over time. Data on the German stock index (DAX) is used for the empirical application.

When we compare the first moment of the implied pdf of volatility to estimates of realized volatility, it appears that the market is able to gauge the level of future average volatility quite well. This can perhaps best be seen during periods of low volatility when future volatility is high, and vice versa. Estimates of the second moment of the implied pdf suggest that the market adopts a model that predicts long-range dependence in stochastic volatility. This too is consistent with the stylized facts of realized volatility.

Let us conclude with suggestions for future research. First on the agenda is to test the option pricing model for misspecification using a variety of option price data. Second, given estimates of market expectations over time, a logical next step would be to infer possible learning behaviour. How is the subjective probability distribution of future average volatility updated over time as new observations become available? A better understanding of the interaction between beliefs and observables would allow for more accurate estimation of the implied pdf's, and hence for more accurate option pricing, and hedging.

A third direction for future research is to explore the role of heterogeneity. "A further potential advantage of the discrete time approach is that it permits the introduction of heterogeneous probability assessments across investors and even individual uncertainty as to the parameters of the underlying probability distributions, thus removing the most restrictive feature of the continuous time model which is the assumption that the parameters of the underlying stochastic processes are known with certainty and agreed upon by all investors" (Brennan, 1979). Note, however, that introducing heterogeneity may call for agent based modelling where different demands for option contracts and the underlying stock, due to different expectations, are cleared by some market maker mechanism. For early studies both model uncertainty and heterogeneity of beliefs within an option pricing model see e.g. Buraschi and Jiltsov (2006) and Brock at al. (2009).

4.8 Appendix

Theorem 4.6 *Let $E[V] = \lambda$ and $var[V] = \nu$. In case average variance V is Gamma or Inverse Gaussian distributed, the moments p and q, as defined in eq. (4.32) and (4.33), are given by:*

$$p = (\sqrt{\xi} - \frac{1}{2}) - \beta \tag{4.43}$$
$$q = (\sqrt{\xi} - \frac{1}{2}) + \beta, \tag{4.44}$$

where

$$\xi = (\beta - \frac{1}{2})^2 + \frac{\lambda}{\nu T} c_\varphi. \tag{4.45}$$

with $c_\varphi = 1$ in case of Inverse Gaussian, and $c_\varphi = 2$ in case of Gamma.

Proof. It follows that $z^+ = \sup\{z : E[S_T^z|F_0] < \infty | z > 1\}$ and $z^- = \sup\{z : E[S_T^z|F_0] < \infty | z \leq 0\}$. We have $q = -z^-$ (for $z \leq 0$), and $p = z^+ - 1$ (for $z > 1$). Observe that S_T^z condtional on V (and on F_0) is log-normal distributed, with:

$$E[S_T^z|V] = c \exp\left((\beta - \frac{1}{2})VTz + \frac{1}{2}VTz^2\right) \tag{4.46}$$
$$= c \exp(\tau(z)V), \tag{4.47}$$

where $\tau(z) = (\beta - \frac{1}{2})Tz + \frac{1}{2}Tz^2$. The constant c depends on S_0, and on the model parameters, but not on z. $E[S_T^z] = E_\varphi[E[S_T^z|V]]$ may thus be derived as the expected value of the RHS of eq. (4.47), with variance V as the random variable with pdf φ. Given the functional form of $E[S_T^z|V]$, we have that $E[S_T^z]$ follows directly from the moment-generating-function of φ, which we shall denote $g(t) = E_\varphi[\exp(tV)]$ (with slight abuse of notation, as t is also used to indicate time).

For the Gamma distribution with two parameters θ and k, such that $E[V] = k\theta$ and $var[V] = k\theta^2$, it follows that $g(t) = (1 - \theta t)^{-k}$ for $t < \theta^{-1}$. This implies:

$$E[S_T^z] = c\left(1 - \theta((\beta - \frac{1}{2})Tz + \frac{1}{2}Tz^2)\right)^{-k}. \tag{4.48}$$

Given that $k > 0$, this will tend to infinity when $(\beta - \frac{1}{2})Tz + \frac{1}{2}Tz^2 \to 1/\theta$. Solving this for z yields:

$$z^{+,-} = (\frac{1}{2} - \beta) \pm \sqrt{(\beta - \frac{1}{2})^2 + 2\frac{\lambda}{\nu T}}, \tag{4.49}$$

where $1/\theta = \lambda/\nu$, which implies the followig values for p and q:

$$
\begin{aligned}
p &= -1 + z^{+} && (4.50)\\
&= (\sqrt{\xi_2} - \frac{1}{2}) - \beta && (4.51)\\
q &= -z^{-} && (4.52)\\
&= (\sqrt{\xi_2} - \frac{1}{2}) + \beta, && (4.53)
\end{aligned}
$$

where $\xi_2 = (\beta - \frac{1}{2})^2 + 2\frac{\lambda}{\nu T}$.

For the Inverse Gaussian distribution with two parameters μ and l, such that $E[V] = \mu$ and $var[V] = \mu^3/l$, the moment-generating-function is given by:

$$g(t) = \exp\left(\frac{l}{\mu}\left(1 - \sqrt{1 - 2\mu^2 t/l}\right)\right). \tag{4.54}$$

This implies that $E[S_T^z]$ will exist only when $1 - 2\mu^2\tau(z)/l \geq 0$. The values of z for which we border non-existance thus satisfy:

$$\tau(z) = \frac{1}{2}\frac{l}{\mu^2} = \frac{1}{2}\frac{\lambda}{\nu}. \tag{4.55}$$

Solving for z yields:

$$z^{+,-} = (\frac{1}{2} - \beta) \pm \sqrt{(\beta - \frac{1}{2})^2 + \frac{\lambda}{\nu T}}. \tag{4.56}$$

If we define $\xi_1 = (\beta - \frac{1}{2})^2 + \frac{\lambda}{\nu T}$, and derive p and q given the solution for z, we will obtain the expressions stated in the proposition. ■

Theorem 4.7 *Let $E[V] = \lambda$, $var[V] = \nu$, and $\gamma = T^{-1}\log(E[\exp(\beta TV)])$. Without further assumptions about the pdf φ, the implied variance function $I(x)$ can be approximated by:*

$$I(x) \approx \lambda + \frac{\kappa}{4v_0}\left(\frac{x^2}{v_0 T} - 1 - \frac{v_0 T}{4}\right) + \beta\kappa T\left(\beta + \frac{1}{2} + \frac{x}{v_0 T} + (v_0 T)^{\frac{1}{2}}(\beta - \frac{2\psi}{\kappa T})a(x)\right),$$

where

$$a(x) = \frac{\Phi(d_-(x))}{\phi(d_-(x))},$$

with $d_-(x) = \frac{1}{2}(v_0 T)^{\frac{1}{2}} - x(v_0 T)^{-\frac{1}{2}}$*. The functions* Φ *and* ϕ *denote the standard normal cumulative distribution and density function. Variance level* v_0 *is defined by:*[20]

$$v_0 = \begin{cases} \gamma/\beta & if \quad \beta \neq 0 \\ \lambda & if \quad \beta = 0 \end{cases}$$

The parameters (ψ, κ) *are given by:* $\psi = v_0 - \lambda$*, and* $\kappa = \nu + (v_0 - \lambda)^2$*.*

Proof. Let us begin with restating the expression for the option price:

$$C_\varphi = \int_0^\infty e^{(\beta v - \gamma)T} C_{BS}(r + (\beta v - \gamma), v)\varphi(v)dv, \tag{4.57}$$

where φ denotes the probability density function for average variance V. Expressed as an expected value, we have:

$$\begin{aligned} C_\varphi &= E[e^{(\beta V - \gamma)T} C_{BS}(r + (\beta V - \gamma), V)] \qquad (4.58) \\ &= E[f(V)]. \qquad (4.59) \end{aligned}$$

Consider the following second-order Taylor expansion for f at $V = v_0$:

$$\begin{aligned} f(V) &\approx f(v_0) + f^{'}(v_0)(V - v_0) + \frac{1}{2}f^{''}(v_0)(V - v_0)^2 \\ &= f(v_0) + f^{'}(v_0)((V - \lambda) - (v_0 - \lambda)) + \frac{1}{2}f^{''}(v_0)((V - \lambda) - (v_0 - \lambda))^2. \end{aligned}$$

If we take expectations over V, with $E[V] = \lambda$ and $var[V] = \nu$, we obtain:

$$E[f(V)] \approx f(v_0) - f^{'}(v_0) + \frac{1}{2}f^{''}(v_0)\left(\nu + (v_0 - \lambda)^2\right), \tag{4.60}$$

which provides a second-order approximation for our option pricing formula C_φ from eq. (4.59).

Let us take $v_0 = \gamma/\beta$ in case $\beta \neq 0$, and $v_0 = \lambda$ otherwise. Note that our price equation features the expectation of the Black-Scholes price $C_{BS}(r(V), V)$ where the interest rate depends on the variance: $r(V) = r + (\beta V - \gamma)$. For our choice of v_0, this reduces to the risk-free interest rate: $r(v_0) = r$. As a result, $f(v_0) = C_{BS}(r, v_0)$.

[20] Note that $\beta \to 0$ implies $\gamma \to \beta\lambda \to 0$.

A first order Taylor approximation of the Black-Scholes price $C_{BS}(I)$ yields:

$$C_\varphi = C_{BS}(I) \approx C_{BS}(r, v_0) + (I - v_0)\frac{\partial C_{BS}}{\partial V}(r, v_0). \tag{4.61}$$

For ease of notation we will supress the indication that the derivatives are evaluated at $V = v_0$ with $r(v_0) = r$. Note that the partial derivatives will be a function of x.

By combining the approximations from eq. (4.60) and eq. (4.61), we obtain:

$$I(x) \approx v_0 - (v_0 - \lambda)f'(v_0)\left(\frac{\partial C_{BS}}{\partial V}\right)^{-1} + \frac{1}{2}\left(\nu + (v_0 - \lambda)^2\right) f''(v_0)\left(\frac{\partial C_{BS}}{\partial V}\right)^{-1}.$$

Let us restate the definition for f: $f(V) = e^{(\beta V - \mu)T} C_{BS}(r + (\beta V - \mu), V)$. Differentiating this with respect to V yields:

$$\frac{\partial f}{\partial V} = e^{(\beta V - \mu)T}\left(\beta T C_{BS} + \frac{\partial C_{BS}}{\partial V} + \frac{\partial C_{BS}}{\partial r}\frac{\partial r}{\partial V}\right). \tag{4.62}$$

It follows that $\frac{\partial r}{\partial V} = \beta$. If we evaluate the derivative at $V = v_0$, we find: $f'(v_0) = \beta T C_{BS} + \frac{\partial C_{BS}}{\partial V} + \beta\frac{\partial C_{BS}}{\partial r}$. By differentiating once more with respect to V, and evaluating the derivative at $V = v_0$, we obtain:

$$\frac{\partial^2 f}{\partial V^2} = \beta^2 T^2 C_{BS} + 2\beta T\frac{\partial C_{BS}}{\partial V} + 2\beta^2 T\frac{\partial C_{BS}}{\partial r} + \frac{\partial^2 C_{BS}}{\partial V^2} + 2\beta\frac{\partial^2 C_{BS}}{\partial r \partial V} + \beta^2\frac{\partial^2 C_{BS}}{\partial r^2}.$$

This would be a good moment to review the 'Greeks', the partial derivatives of the Black-Scholes formula $C_{BS} = S_0\Phi(d_1(x)) - Ke^{-rT}\Phi(d_2(x))$:

$$\begin{aligned}
\frac{\partial C_{BS}}{\partial V} &= \frac{1}{2}T^{\frac{1}{2}}v_0^{-\frac{1}{2}}S_0\phi(d_1(x)) \\
\frac{\partial^2 C_{BS}}{\partial V^2} &= \frac{1}{4}T^{\frac{1}{2}}v_0^{-\frac{3}{2}}S_0\left[\left((v_0T)^{-\frac{1}{2}}x + \frac{1}{2}(v_0T)^{\frac{1}{2}}\right)\phi'(d_1(x)) - \phi(d_1(x))\right] \\
\frac{\partial C_{BS}}{\partial r} &= KTe^{-rT}\Phi(d_2(x)) \\
\frac{\partial^2 C_{BS}}{\partial r^2} &= Ke^{-rT}T^{\frac{1}{2}}v_0^{-\frac{1}{2}}\left(T\phi(d_2(x)) - T^{\frac{3}{2}}v_0^{\frac{1}{2}}\Phi(d_2(x))\right) \\
\frac{\partial^2 C_{BS}}{\partial r \partial V} &= \frac{1}{2}T^{\frac{1}{2}}v_0^{-\frac{1}{2}}\left(\frac{x}{v_0T} - \frac{1}{2}\right)KTe^{-rT}\phi(d_2(x)),
\end{aligned}$$

where $d_2(x) = d_1(x) - (v_0T)^{\frac{1}{2}}$, with $d_1(x) = \frac{1}{2}(v_0T)^{\frac{1}{2}} - x(v_0T)^{-\frac{1}{2}}$. Dividing these derivatives, as well as C_{BS} itself, by $\frac{\partial C_{BS}}{\partial V}$, gives us:

$$C_{BS}\left(\frac{\partial C_{BS}}{\partial V}\right)^{-1} = 2T^{-\frac{1}{2}}v_0^{\frac{1}{2}}\left(\frac{\Phi(d_1(x))}{\phi(d_1(x))}-\frac{\Phi(d_2(x))}{\phi(d_2(x))}\right) \tag{4.63}$$

$$\frac{\partial^2 C_{BS}}{\partial V^2}\left(\frac{\partial C_{BS}}{\partial V}\right)^{-1} = \frac{1}{2}v_0^{-2}\left(\frac{x^2}{T}-v_0-\frac{1}{4}Tv_0^2\right) \tag{4.64}$$

$$\frac{\partial C_{BS}}{\partial r}\left(\frac{\partial C_{BS}}{\partial V}\right)^{-1} = 2(v_0T)^{\frac{1}{2}}\frac{\Phi(d_2(x))}{\phi(d_2(x))} \tag{4.65}$$

$$\frac{\partial^2 C_{BS}}{\partial r^2}\left(\frac{\partial C_{BS}}{\partial V}\right)^{-1} = 2T\left(1-(v_0T)^{\frac{1}{2}}\frac{\Phi(d_2(x))}{\phi(d_2(x))}\right) \tag{4.66}$$

$$\frac{\partial^2 C_{BS}}{\partial r\partial V}\left(\frac{\partial C_{BS}}{\partial V}\right)^{-1} = \frac{x}{v_0}-\frac{1}{2}T. \tag{4.67}$$

In these derivations we used that $\frac{\phi(d_1(x))}{\phi(d_2(x))}=e^x$, and $\frac{\phi'(d_1(x))}{\phi(d_1(x))}=-d_1(x)$.

Now define: $\psi=(v_0-\lambda)$, $\kappa=\nu+(v_0-\lambda)^2$, and $a_i(x)=\frac{\Phi(d_i(x))}{\phi(d_i(x))}$ for $i=1,2$. If we substitute these results into our expression for the approximation of $I(x)$, we find:

$$\begin{aligned} I(x) &\approx \lambda+\beta(v_0T)^{\frac{1}{2}}(\beta\kappa T-2\psi)(a_1(x)-a_2(x))+2\beta(v_0T)^{\frac{1}{2}}(\beta\kappa T-\psi)a_2(x) \\ &+ \left(\beta\kappa T+\beta\kappa(\frac{x}{v_0}-\frac{1}{2}T)+\beta^2\kappa T\right)-\beta^2\kappa T(v_0T)^{\frac{1}{2}}a_2(x) \\ &+ \frac{1}{4}\kappa v_0^{-1}\left(\frac{x^2}{v_0T}-1-\frac{1}{4}v_0T\right). \end{aligned}$$

It can be verified that all terms with $a_2(x)$ cancel each other out. Finaly, the result stated in the proposition can be obtained by rearranging the expression that remains. ■