Asia-Pacific Finan Markets (2007) 14:1–23
DOI 10.1007/s10690-007-9044-y

Estimation and Prediction of a Non-Constant Volatility

Vyacheslav M. Abramov · Fima C. Klebaner

Published online: 17 August 2007

Abstract In this paper we study volatility functions. Our main assumption is that the volatility is a function of time and is either deterministic, or stochastic but driven by a Brownian motion independent of the stock. Our approach is based on estimation of an unknown function when it is observed in the presence of additive noise. The set up is that the prices are observed over a time interval $[0, t]$, with no observations over (t, T), however there is a value for volatility at T. This value is may be inferred from options, or provided by an expert opinion. We propose a forecasting/interpolating method for such a situation. One of the main technical assumptions is that the volatility is a continuous function, with derivative satisfying some smoothness conditions. Depending on the degree of smoothness there are two estimates, called filters, the first one tracks the unknown volatility function and the second one tracks the volatility function and its derivative. Further, in the proposed model the price of option is given by the Black–Scholes formula with the *averaged future* volatility. This enables us to compare the implied volatility with the averaged estimated historical volatility. This comparison is done for three companies and has shown that the two estimates of volatility have a weak statistical relation.

Keywords Non-constant volatility · Approximating and forecasting volatility · Black–Scholes formula · Best linear predictor

V. M. Abramov (✉) · F. C. Klebaner
School of Mathematical Sciences, Monash University, Building 28M, Clayton Campus, Clayton, VIC 3800, Australia
e-mail: vyacheslav.abramov@sci.monash.edu.au

F. C. Klebaner
e-mail: fima.klebaner@sci.monash.edu.au

1 Introduction

The concept of volatility is associated with fluctuations of a time series. Historical observations on an asset provide information that can be used to estimate the volatility of this asset. However, in finance volatility has an additional importance, because it also enters prices of options on stocks. In the classical model of Black and Scholes the same parameter that describes historical volatility enters the option price. Therefore we have potentially two sources for estimation of volatility: the primary source, consisting of historical prices, and the secondary source, consisting of prices of financial derivatives. The estimates obtained from historical prices are called historical volatilities, and from options-implied. It has been known as a "well-known empirical fact" that historical volatility and implied volatility are distinct, e.g. Hull (2002). This conclusion is based on the analysis of prices using the classical Black–Scholes model with constant volatility. This paper studies a simple non-constant volatility model, in which volatility is a function of time. Volatility is allowed to be stochastic with the Brownian motion in volatility independent of the Brownian motion in stock, assumption needed for comparison with options theory. When volatility is not constant the standard statistical estimate of the variance ($s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$) does not apply, and in turn volatility studies based on this formula. In the proposed model the price of option is given by the Black–Scholes formula with the *averaged future* volatility, (Hull and White 1987; Stein and Stein 1991). The situation with two sources for estimation of unknown function (volatility) is very peculiar to financial markets, because in these markets both the prices of stocks and options can be used. It is also specific for the chosen model, because this model (and only this model, Hamza and Klebaner (submitted)) gives option prices by the Black–Scholes formula. Having an analytical expression for the volatility from the observed options prices is like having a look into the future: the knowledge of averaged *future* volatility over the interval $(t, T]$. In other models, such as Heston's stochastic volatility, the implied volatility is some complicated functional of future volatility, not known analytically.

The first objective of the paper is to propose a method of estimating, interpolating and forecasting a volatility function. This is important because the volatility is a widespread measure of risk and average integrated volatility can be used as a risk measure of the asset over the specified future time interval. Generally, volatility is itself traded, as well as quotes are often formulated in terms of volatilities. It will also provide fund managers with useful information about future volatility of stocks. Volatility estimates are needed not only for options pricing and hedging but also for efficient portfolio management, where variance is needed in Markovitz theory of portfolio optimization. We also compare estimates of historical and implied volatilities, which is the second aim of the paper. To our knowledge no such study has been done before, although recently there has been an increasing attention to non-constant volatility models (e.g. Ghysels et al. (1996); Goldentayer et al. (2005); Harvey et al. (1994); Heston's (1993); Kim et al. (1998); Klebaner et al. (2006); Shephard (1996) and many others). Our approach allows to predict volatility function over time intervals which have no observed prices. One can argue that interpolation can be done by 'practical' methods by connecting points by straight lines or splines. These methods may or may not 'perform well' in practice, however they lack theoretical justification. The methods for estimation and interpolation proposed here have the advantage of coherent mathematical foundation.

The paper is structured as follows. General methods of estimation and prediction of volatility are reviewed in Sect. 2. In Sect. 3 we give a precise description of the model and the method. Sections. 4.2 and 4.3 give new prediction formulae in Theorems 4.1 and 4.2. In Sect. 4.4 we compare quadratic variation method with the method of Goldentayer et al. (2005) numerically. Section. 5 discusses numerical examples of forecasting volatility. Section 6 checks whether the volatility obtained by the functional estimation method of Goldentayer et al. (2005) agrees with the observed implied volatility.

2 Methods of Estimation of Volatility

Volatility estimation and forecasting is discussed in a large number of papers (see Andersen et al. (2005); Poon and Granger (2003) and the references therein, which also give a review of this field).

Essentially, there are two statistical approaches for estimating and predicting volatility: a model-based approach and a model-free approach. A model-based approach assumes a model for the volatility process, while the other method does not assume any particular model, although certain regularity assumptions on the behaviour of volatility function are made. The most famous representative example of the model-based approach is the GARCH model, and the most famous example of the model-free approach is the (QV) Quadratic variation approach.

2.1 The GARCH Model for Volatility

GARCH-based modelling and forecasting volatility has been initiated by Engle's seminal ARCH paper Engle (1982), and studied by Bollerslev (1986) and Taylor (1986). The model for volatility of the returns has form

$$\sigma_t^2 = a_0 + a_1\epsilon_{t-1}^2 + \cdots + a_q\epsilon_{t-q}^2 + \beta_1\sigma_{t-1}^2 + \cdots + \beta_p\sigma_{t-p}^2, \tag{2.1}$$

where conditional on the past up to time $t-1, \epsilon_t$ has a Gaussian distribution with mean 0 and variance σ_t^2. Therefore the model for log returns they consider is of the form of regression with stochastic variance as in (2.1)

$$\log \frac{S_t}{S_{t-1}} = \mu(\mathbf{d}) + \epsilon_t,$$

where μ is the conditional mean function with vector argument $\mathbf{d}$, e.g. a linear regression $\mathbf{x}_t^\top \mathbf{d}$, where $\mathbf{x}_t$ denote the vector of independent explanatory variables.

The GARCH methods have been studied in papers (Hillebrand 2005, 2006; Hillebrand and Madeiros 2006). Hillebrand (2006) gives short-term forecasts as well as generalizations to several time scales. Hillebrand and Madeiros (2006) studies estimation and forecasting of volatility in the presence of structural breaks and regime switches.

2.2 Quadratic Variation Method

The QV method is based on the following facts. A continuous semimartingale X_t has a decomposition into a finite variation part A_t and a martingale part $M_t, X_t = X_0 + A_t + M_t$. The quadratic variation of X_t can be calculated from the observed paths as

$$[X,X]_t = \lim \sum_{i=1}^{n-1} (X_{t_{i+1}} - X_{t_i})^2 = [M,M]_t$$

when the $\max_i(t_{i+1} - t_i)$ of partitions of the interval $[0,t]$ tends to 0. In diffusion models for stock prices integrated volatility is in fact the quadratic variation of returns, $[X,X]_t = \int_0^t \sigma^2(s)\mathrm{d}s$. This approach is used in Anderson and Vahid (2005); Barndorff-Nielsen and Shephard (2002a,b, 2004), and others. The estimate obtained by this method is called realized variance. An important observation is that the quadratic variation methods do not allow extrapolation to the future where prices are not yet observed, unless some specific model for volatility is assumed. In Sect. 4.4 we give a comparison of QV approach used in Barndorff-Nielsen and Shephard (2002a,b) with the method used in this paper.

2.3 Local Approximation Methods

These methods are based on the assumption that the volatility can be locally approximated by constant, i.e. for any time moment t there exists interval of time homogeneity $[t-m,t]$ over which the volatility is a constant, Mercurio and Spokoiny (2004). They give an algorithm for estimation of the intervals of time homogeneity as well as the estimate of volatility, obtained by averaging over that interval. The so-called *adaptive weight smoothing* (AWS) introduced in Polsehl and Spokoiny (2000) is a non-parametric method of estimation based on locally constant smoothing with adaptive choice of weights for every pair of data points, Polsehl and Spokoiny (2002, 2003). It is applied to extended GARCH models with varying coefficients in Polsehl and Spokoiny (2004).

2.4 The Malliavin Calculus Methods

The methods based on Malliavin calculus (e.g. Bichteler et al. (1987); Kusuoka and Yoshida (2000); Malliavin and Thalmaier (2006); Yoshida (1997)) that have been applied to stochastic volatility models are also model-based. Malliavin calculus allows to obtain approximations of functionals of diffusion processes, and more general semimartingales, in terms of functionals of simpler processes, Gaussian for example, akin to the Edgeworth expansions of arbitrary probability distribution functions around the Normal distribution with specified error bounds. This was done for Barndorff-Nielsen and Shephard's stochastic volatility model (2001) by Masuda and Yoshida (2005). Barndorff-Nielsen and Shephard's model is a generalization of Heston's (1993) stochastic volatility model with a jump term, in which volatility is an ergodic non-Gaussian Orstein-Uhlenbeck type process. It would be interesting to do a numerical study based on the analytical results of Masuda and Yoshida (2005), but it is too involved to be done here.

2.5 Other Methods

Del Moral et al. (2001) proposes the Monte-Carlo method for filtering of discrete observations. The likelihood methods of estimation for parameters of diffusion processes with jumps with combination of the methods of Malliavin calculus have been used by Shimizu and Yoshida (2006). Hayashi, T. (2004) proposes a two-step procedure consisting of the method of principal components and EM (Expectation-Maximization) algorithm to investigate the parameters of the discrete-time model of high frequency stock returns considered in that paper. Other statistical methods are given in Hayashi and Mykland (2005), Lee et al. (2006).

3 Functional Estimation Method of the Present Paper

In this section we define the model precisely and describe the forecasting method. The volatility $\sqrt{v_t}$ is the function appearing in the Black–Scholes model for the stock price S_t (also known as the spot volatility)

$$\mathrm{d}S_t = rS_t\mathrm{d}t + \sqrt{v_t}S_t\mathrm{d}W_t, \tag{3.1}$$

where W_t is a standard Wiener process. In the classical Black–Scholes model Black and Scholes (1973), Merton (1973), the spot volatility is assumed to be constant σ, i.e. $\sqrt{v_t} \equiv \sigma$.

We assume that in the above model the stock price is the only observable, and only at discrete times $t_1, t_2, \ldots, t_N$ so that the challenge is: firstly to extract information about the volatility function from past stock prices, and secondly to predict this function into the future where no stock prices are yet observed.

We propose a new method of volatility forecasting based on the technique of functional estimation in the presence of noise developed in the context of volatility by Goldentayer et al. (2005) which is based on nonparametric approach of functional estimation due to Ibragimov and Khasminskii (1981) and Khasminskii and Liptser (2002). The method is essentially given by solving a control problem. It is especially useful in the context of volatility forecasting due to the extra information about volatility derived from options.

3.1 Nonparametric Functional Estimation of Volatility

The approach of Goldentayer et al. (2005) gives on adaptive algorithm for tracking historical volatility. It is assumed that the spot volatility belongs to the Ibragimov–Khasminskii class $\Sigma(\beta, L)$.[1] Only integrated variance $\int_{t_{n-1}}^{t_n} v(s)ds$ over intervals $[t_{n-1}, t_n]$, can be estimated from the observation of stock prices at the times $t_n, n = 1, \ldots, N$. We denote $v_{n-1} = \frac{1}{\Delta}\int_{t_{n-1}}^{t_n} v(s)ds$, $\Delta = t_n - t_{n-1} = T/N$, and its estimates by $\hat{v}_{n-1}$.

Note that the integrated spot volatility is a differentiable function and therefore automatically belongs to class $\Sigma(1, L)$. If no further assumption is made on the smoothness of $v(t)$ then the first order filter is used and given by the formula (see Sect. 3 of Goldentayer et al. (2005))

$$\widehat{v}_n = \left(1 - \frac{a_1}{N}\right)\widehat{v}_{n-1} + \frac{a_1\kappa}{N} + \frac{\vartheta}{N^{2/3}}\left(X_n - \widehat{v}_{n-1}\right), \tag{3.2}$$

where X_n is the squared log-return

$$X_n = \frac{1}{\Delta}\left[\log\left(\frac{S_{t_n}}{S_{t_{n-1}}}\right)\right]^2,$$

a_1 and κ are specific parameters, ϑ is the (unique) parameter chosen to minimize the innovation difference

$$S_N(\vartheta) = \frac{1}{N}\sum_{n=1}^{N}(X_n - \widehat{v}_{n-1})^2.$$

[1] Recall that the Ibragimov–Khasminskii class of functions Σ has the properties

$$\Sigma(\beta, L) = \left\{ f : \begin{bmatrix} f \text{ has } k \text{ derivatives with } k\text{-th derivative satisfying} \\ |f^{(k)}(t_2) - f^{(k)}(t_1)| \le L|t_2 - t_1|^{\alpha},\ \forall\, t_1, t_2 \text{ and } \alpha \in (0, 1]; \\ \beta = k + \alpha. \end{bmatrix} \right\}.$$

Thus if the function f belongs to the class $\Sigma(\ell, L)$, where ℓ is a positive integer, then it is assumed that the function f has the $\ell - 1$st derivative satisfying the Lipschitz condition. For example, $\Sigma(1, L)$ is the class of the Lipschitz-continuous functions, $|v(t) - v(s)| \le L|t - s|$.

It is worth noting that the first order filter in Goldentayer et al. (2005) is derived by using representation for the integrated variance

$$v_n = \left(1 - \frac{a_1}{N}\right) v_{n-1} + \frac{a_1\kappa}{N} + \text{white noise}, \tag{3.3}$$

where the standard deviation of white noise has the optimal order $O\left(\frac{1}{N^{2/3}}\right)$, see Chow et al. (1997), same as the mean squared error of kernel estimates Ibragimov and Khasminskii (1980), Theorem 2.1.

Therefore the best linear predictor $\widehat{v}_n$ is defined by the representation similar to (3.3), where the last term is replaced by zero when future observations are not available,

$$\widehat{v}_n = \left(1 - \frac{a_1}{N}\right) \widehat{v}_{n-1} + \frac{a_1\kappa}{N}. \tag{3.4}$$

If the volatility function has a derivative that satisfies the Lipschitz condition (and therefore belongs to $\Sigma(2, L)$), then the second order filter is used, and is given by the system (see Sect. 3 of Goldentayer et al. (2005))

$$\begin{aligned} \widehat{v}_n &= \widehat{v}_{n-1} + \frac{1}{N}\widehat{v}^{(1)}_{n-1} + \frac{\sqrt{2\vartheta}}{N^{4/5}}(X_n - \widehat{v}_{n-1}), \\ \widehat{v}^{(1)}_n &= \left(1 - \frac{a_1}{N}\right)\widehat{v}^{(1)}_{n-1} - \frac{a_2}{N}\widehat{v}_{n-1} + \frac{a_2\kappa}{N} + \frac{\vartheta}{N^{3/5}}(X_n - \widehat{v}_{n-1}), \end{aligned} \tag{3.5}$$

where the superscript$^{(1)}$ stands for the derivative of volatility function, and a_1, a_2 and κ are the specific parameters of this filter.

Similarly to the first order filter, the second order filter is based on representation:

$$\begin{aligned} v_n &= v_{n-1} + \frac{1}{N} v^{(1)}_{n-1} + \text{white noise}, \\ v^{(1)}_n &= \left(1 - \frac{a_1}{N}\right) v^{(1)}_{n-1} - \frac{a_2}{N} v_{n-1} + \text{white noise}, \end{aligned} \tag{3.6}$$

where the white noises are independent. Similarly to the case of the first order filter, the standard deviation parameters of these noises are optimal having orders $O\left(\frac{1}{N^{4/5}}\right)$ and $O\left(\frac{1}{N^{3/5}}\right)$, Chow et al. (1997), Ibragimov and Khasminskii (1980). The parameters a_1, a_2, N, κ of these two filters are found by a tuning procedure.

For prediction into the future, the last terms of the first and second equations of (3.5) are replaced by zero

$$\begin{aligned} \widehat{v}_n &= \widehat{v}_{n-1} + \frac{1}{N}\widehat{v}^{(1)}_{n-1}, \\ \widehat{v}^{(1)}_n &= \left(1 - \frac{a_1}{N}\right)\widehat{v}^{(1)}_{n-1} - \frac{a_2}{N}\widehat{v}_{n-1}. \end{aligned} \tag{3.7}$$

4 Approach to Forecasting by the Control Method

4.1 Extension of the Method to Include Future Information

The main idea in the proposed method for forecasting is to attain a given point in the future. That is, having a historical volatility dynamics in the first n_0 points and assuming its value at the last point N to be known ($N > n_0$), we interpolate volatility dynamics in all intermediate points between n_0 and N. The volatility at point N is assumed to be known. It can be taken as volatility implied by option prices, i.e. the value of $\left(\frac{1}{T-t}\int_t^T v(s)\mathrm{d}s\right)^{1/2}$, but not necessarily. It can also be an expert opinion, or a value from any other source. Alternatively, this value can be taken as a parameter, and the obtained predictions studied as a function of this parameter, for example for their sensitivity. When daily information is considered in discrete time scale, the above integral is approximated by the sum: $\frac{1}{N-n_0+1}\sum_{n=n_0}^{N} v_n$, and this value is just assumed to be known. Denote the known value of $\sum_{n=n_0}^{N} v_n$ by $\overline{V}$.

In this setup the problem of predicting volatility can be seen as a standard problem of interpolation. Assume that $\sum_{n=n_0}^{N} v_n = \overline{V}$. Then the problem is to find a control sequence $u_n, n = n_0, \ldots, N$ minimizing the mean squared error of approximation. We solve this problem and gives simple formulae for both types of filters described in the previous section.

4.2 Prediction/Interpolation with the First Order Filter

Here we assume that we have an adaptive estimating procedure given by (3.4) for time points $t_i, i = 1, \ldots, n_0$, as well as some value $\overline{v}$ at point $t_N, N > n_0$. We want to extend estimators $\widehat{v}_n$ to the time region $n_0 \le n \le N$ with the last one being $\overline{v}$. In application to implied volatility $\overline{v}$ is deduced from the known value $\overline{V}$, and it turns out to be $\overline{v} = \overline{V} - \sum_{i=n_0}^{N-1}\widehat{v}_i$.

Denote the predicted values by $\widetilde{v}_n, n_0 \le n \le N$. Put $a = 1 - \frac{a_1}{N}$ and $b = \frac{a_1\kappa}{N}$ in (3.4), and rewrite it as

$$\widehat{v}_n = a\widehat{v}_{n-1} + b. \tag{4.1}$$

Since $a < 1$, recursion in (4.1) converges to the fixed value $\widehat{v}_\infty$, and

$$\widehat{v}_\infty = a\widehat{v}_\infty + b.$$

Hence

$$\widehat{v}_\infty = \frac{b}{1-a} = \kappa. \tag{4.2}$$

It is clear that the recursion (4.1) most likely will not end up at the specified value $\overline{v}$ at time N. Introduce a control sequence u_n ($n = n_0, \ldots, N$), to be determined later and the following recurrence relation

$$\widetilde{v}_n = a\widetilde{v}_{n-1} + b + u_n, \tag{4.3}$$

which in the case $u_n \equiv 0$ gives (4.1). Take $\widetilde{v}_n = \widehat{v}_n$ for $n = 1, 2, \ldots, n_0$, and $\widetilde{v}_N = \overline{v}$. The sequence u_n is chosen so that

$$\begin{cases} \widetilde{v}_N = \overline{v}, \\ \sum_{n=n_0+1}^{N} u_n^2 \quad \text{is minimal.} \end{cases} \tag{4.4}$$

Theorem 4.1 *The optimal predictor satisfying $\widetilde{v}_N = \overline{v}$ is given by*

$$\widetilde{v}_n = a\widetilde{v}_{n-1} + b + \frac{\overline{v} - \widehat{v}_N}{\sum_{i=n_0+1}^{N} a^{2(N-i)}} \cdot a^{N-n}, \tag{4.5}$$

with $a = 1 - \dfrac{a_1}{N}$ and $b = \dfrac{a_1\kappa}{N}$.

Proof For the point $\overline{v} = \widetilde{v}_N$ we have from (4.2)

$$\overline{v} = \widetilde{v}_N = \widetilde{v}_{n_0} a^{N-n_0} + \sum_{n=n_0+1}^{N} a^{N-n}(b + u_n).$$

On the other hand, according to (4.1)

$$\widehat{v}_N = \widehat{v}_{n_0} a^{N-n_0} + b \sum_{n=n_0+1}^{N} a^{N-n}.$$

Therefore

$$\widetilde{v}_N = \widehat{v}_N + \sum_{n=n_0+1}^{N} a^{N-n} u_n.$$

This enables us to write

$$(\widetilde{v}_N - \widehat{v}_N)^2 = \left(\sum_{n=n_0+1}^{N} a^{N-n} u_n\right)^2.$$

By the Cauchy-Schwartz inequality,

$$\left(\sum_{n=n_0+1}^{N} a^{N-n} u_n\right)^2 \leq \sum_{n=n_0+1}^{N} a^{2(N-n)} \cdot \sum_{n=n_0+1}^{N} u_n^2. \tag{4.6}$$

The equality in (4.6) is achieved if and only if $a^{N-n} = cu_n$ for some constant c, and since the equality in (4.6) is associated with the minimum of the left-hand side of (4.6), the problem reduces to find an appropriate value $c = c^*$ such that

$$u_n = c^* a^{N-n}.$$

Therefore,

$$\widetilde{v}_N = \widehat{v}_N + c^* \sum_{i=n_0+1}^{N} a^{2(N-i)},$$

and then finally for c^* we have:

$$c^* = \frac{\widetilde{v}_N - \widehat{v}_N}{\sum_{i=n_0+1}^{N} a^{2(N-i)}}. \tag{4.7}$$

Thus, the sequence u_n satisfying (4.4) is

$$u_n = \frac{\overline{v} - \widehat{v}_N}{\sum_{i=n_0+1}^{N} a^{2(N-i)}} \cdot a^{N-n}, \tag{4.8}$$

and its substitution to (4.3) yields the result. □

4.3 Prediction/Interpolation with the Second Order Filter

The equations for the second order filter (3.7) can be written in the form of a two dimensional analogue of the equations for the first order filter. The analogue of (4.3) is

$$\widetilde{\mathbf{v}}_n = A\widetilde{\mathbf{v}}_{n-1} + \mathbf{b} + \mathbf{u}_n, \tag{4.9}$$

where $\widetilde{\mathbf{v}}_n, \mathbf{b}, \mathbf{u}_n$ are two-dimensional vectors corresponding to $\widetilde{v}_n, b$ and u_n in the one-dimensional case, and A is a 2×2 order matrix corresponding to the constant a in the one-dimensional case. The control vector $\mathbf{u}_n$ in (4.9) is of the form

$$\mathbf{u}_n = \begin{pmatrix} u_n \\ 0 \end{pmatrix},$$

u_n is a sequence chosen to minimize the error.

In the case where $u_n \equiv 0$ we obtain the following equation

$$\widehat{\mathbf{v}}_n = A\widehat{\mathbf{v}}_{n-1} + \mathbf{b}, \tag{4.10}$$

where $\widehat{\mathbf{v}}_n = \begin{pmatrix} \widehat{v}_n \\ \widehat{v}_n^{(1)} \end{pmatrix}$, where the components $\widehat{v}_n$ and $\widehat{v}_n^{(1)}$ are defined by (3.5) and $A = \begin{pmatrix} 1 & \dfrac{1}{N} \\ -\dfrac{a_2}{N} & 1 - \dfrac{a_1}{N} \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 0 \\ \dfrac{a_2\kappa}{N} \end{pmatrix}$. Similarly to the above one-dimensional case, the control sequence u_n should be chosen such that

$$\begin{cases} \widetilde{v}_N = \overline{v}, \\ \sum_{n=n_0+1}^{N} u_n^2 \quad \text{is minimal}, \end{cases} \tag{4.11}$$

where $\widetilde{v}_N$ is the first component of $\widetilde{\mathbf{v}}_N$ given by (4.9).

Denote by $a_{i,j}^{(n)}$ the elements of the matrix $A^n, i,j = 1,2$, and $\widetilde{v}_n$ is the first component of the vector $\widetilde{\mathbf{v}}_n, n = n_0 + 1, \ldots, N$.

Theorem 4.2 *The optimal predictor satisfying $\widetilde{v}_N = \overline{v}$, is given by $\widetilde{v}_n = \widehat{v}_n + u_n$, where $\widehat{v}_n$ is the first component of the vector $\widehat{\mathbf{v}}_n$ determined by* (4.10) *and*

$$u_n = \frac{\overline{v} - \widehat{v}_N}{\sum_{i=n_0+1}^{N} \left[\left(a_{1,1}^{(N-n)}\right)^2 + \left(a_{2,1}^{(N-n)}\right)^2 \right]} \sqrt{\left(a_{1,1}^{(N-n)}\right)^2 + \left(a_{2,1}^{(N-n)}\right)^2}. \tag{4.12}$$

Proof We have the following:

$$\widetilde{\mathbf{v}}_N = A^{N-n_0}\widetilde{\mathbf{v}}_{n_0} + \sum_{n=n_0+1}^{N} A^{N-n}(\mathbf{b} + \mathbf{u}_n).$$

On the other hand,

$$\widehat{\mathbf{v}}_N = A^{N-n_0}\widehat{\mathbf{v}}_{n_0} + \sum_{n=n_0+1}^{N} A^{N-n}\mathbf{b}.$$

This enables us to write:

$$(\widetilde{\mathbf{v}}_N-\widehat{\mathbf{v}}_N)^\top(\widetilde{\mathbf{v}}_N-\widehat{\mathbf{v}}_N)=\left(\sum_{n=n_0+1}^{N}A^{N-n}\mathbf{u}_n\right)^\top\left(\sum_{n=n_0+1}^{N}A^{N-n}\mathbf{u}_n\right), \quad (4.13)$$

where $\top$ denotes the transpose.

Taking into account that the second component of all vectors $\mathbf{u}_n$ is equal to zero, the right-hand side of (4.13) reduces to

$$\begin{aligned}&\left[\sum_{n=n_0+1}^{N}\begin{pmatrix}a_{1,1}^{(N-n)}u_n\\ a_{2,1}^{(N-n)}u_n\end{pmatrix}\right]^\top\left[\sum_{n=n_0+1}^{N}\begin{pmatrix}a_{1,1}^{(N-n)}u_n\\ a_{2,1}^{(N-n)}u_n\end{pmatrix}\right]\\ &=\left[\sum_{n=n_0+1}^{N}u_n\begin{pmatrix}a_{1,1}^{(N-n)}u_n\\ a_{2,1}^{(N-n)}u_n\end{pmatrix}\right]^\top\left[\sum_{n=n_0+1}^{N}u_n\begin{pmatrix}a_{1,1}^{(N-n)}\\ a_{2,1}^{(N-n)}\end{pmatrix}\right]\end{aligned}$$

Therefore, applying the Cauchy-Schwartz inequality, we obtain

$$\begin{aligned}&\left[\sum_{n=n_0+1}^{N}u_n\begin{pmatrix}a_{1,1}^{(N-n)}\\ a_{2,1}^{(N-n)}\end{pmatrix}\right]^\top\left[\sum_{n=n_0+1}^{N}u_n\begin{pmatrix}a_{1,1}^{(N-n)}\\ a_{2,1}^{(N-n)}\end{pmatrix}\right]\\ &\leq\sum_{n=n_0+1}^{N}\begin{pmatrix}a_{1,1}^{(N-n)}\\ a_{2,1}^{(N-n)}\end{pmatrix}^\top\begin{pmatrix}a_{1,1}^{(N-n)}\\ a_{2,1}^{(N-n)}\end{pmatrix}\cdot\sum_{n=n_0+1}^{N}u_n^2\\ &=\sum_{n=n_0+1}^{N}\left[\left(a_{1,1}^{(N-n)}\right)^2+\left(a_{2,1}^{(N-n)}\right)^2\right]\cdot\sum_{n=n_0+1}^{N}u_n^2. \quad (4.14)\end{aligned}$$

The equality for the left-hand side of (4.14) is achieved if and only if

$$\sqrt{\left(a_{1,1}^{(N-n)}\right)^2+\left(a_{2,1}^{(N-n)}\right)^2}=cu_n$$

for some constant c, and since the equality is associated with the minimum of the left-hand side of (4.14), the problem reduces to find an appropriate value $c=c^*$ such that

$$u_n=c^*\sqrt{\left(a_{1,1}^{(N-n)}\right)^2+\left(a_{2,1}^{(N-n)}\right)^2}. \quad (4.15)$$

Therefore,

$$\widetilde{v}_N=\widehat{v}_N+c^*\sum_{n=n_0+1}^{N}\left[\left(a_{1,1}^{(N-n)}\right)^2+\left(a_{2,1}^{(N-n)}\right)^2\right],$$

where $\widetilde{v}_N$ and $\widehat{v}_N$ are the first components of the vectors $\widetilde{\mathbf{v}}_N$ and $\widehat{\mathbf{v}}_N$ respectively. Then for c^* we have:

$$c^* = \frac{\widetilde{v}_N - \widehat{v}_N}{\sum_{i=n_0+1}^{N} \left[\left(a_{1,1}^{(N-n)} \right)^2 + \left(a_{2,1}^{(N-n)} \right)^2 \right]}. \tag{4.16}$$

Substituting (4.16) into (4.15) and taking into account (4.11) we finally obtain (4.12). □

Remark The quality of forecast depends on the type of the volatility function: if oscillations of volatility around its average are frequent, then the first order filter seems to be appropriate, otherwise the second order filter is better.

4.4 Comparison with Quadratic Variation Method

In this section we compare the quadratic variation method with the method of Goldentayer et al. (2005) numerically.

The quadratic variation approach in Barndorff-Nielsen and Shephard (2002a,b, 2004) is based on realized volatility for fixed intervals of given length h, containing a large number M of observations. For a fixed interval of length $h > 0$ the returns over the i-th interval are defined as

$$x_i = x^*(ih) - x^*((i-1)h), \ i = 1, 2, \ldots,$$

where $x^*(t)$ denotes the log-prices of stocks. During any i-th interval, we also can compute M intra h-returns:

$$x_{j,i} = x^* \left((i-1)h + \frac{hj}{M} \right) - x^* \left((i-1)h + \frac{h(j-1)}{M} \right), \quad j = 1, 2, \ldots, M.$$

The realized variance is defined as

$$[x_M^*] = \sum_{j=1}^{M} x_{j,i}^2,$$

and as $M \to \infty$, $[x_M^*]$ converge in probability to the integral $\int_{(i-1)h}^{ih} \sigma_t^2 dt$. This method requires a large volume of data. The Goldentayer–Klebaner–Liptser method of Goldentayer et al. (2005) computes instantaneous (spot) volatility by using observations on prices at discrete points $\{t_i = ih\}$. This method estimates v_i' s, $v_i = \frac{1}{h} \int_{(i-1)h}^{ih} \sigma_t^2 dt$, the average integrated volatility over intervals $[(i-1)h, ih)$. Therefore these two methods are comparable after the adjustment by the length of the interval h.

Figure 4.4 shows comparison of this method with the approach of the present paper. In our experiments with stock data the value of parameter M is taken

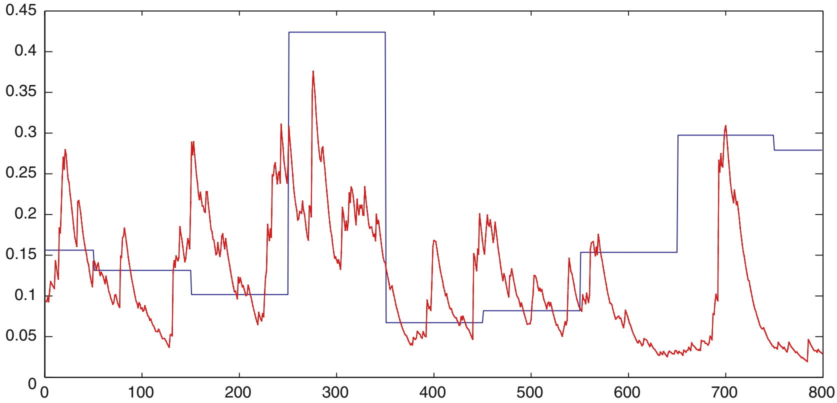

Fig. 1 CECO environmental corporation volatility estimates: red line — first order by Goldentayer et al. (2005); blue line—piecewise constant by Barndorff-Nielsen and Shephard (2002a)

$M = 100$ from daily information of a number of companies (see Barndorff-Nielsen and Shephard (2002a) for details of approximation). Compared with the rate of convergence $O\left(\frac{1}{\sqrt{M}}\right)$, the number of observations is not large enough to give satisfactory accuracy. Only in a small number of cases the comparison results of two methods seem to be relatively close to each other as in Fig. 1.

5 Case Studies of Prediction/Interpolation of Volatility

The numerical examples for interpolating/predicting volatility were done for IBM and FX rates AUD/USD, Ruble/USD. Figure 2 represents the volatility IBM corporation during 1,175 days since December 25, 1999. Volatility is estimated and then interpolated/ predicted by the two methods. The interpolation period is 117 days, and the past history period is 1,058 days. The estimated volatility is in red, and interpolation is in blue. Our results show that the second order filter has a visible advantage over that by the first order filter. Figure 3 gives the volatility of the exchange rates of the US dollar vis Australian dollar during 1,480 days since December 25, 1999. The interpolation period is 148 days, and the past history is 1,338 days. The advantage of the second order filter is less visible than in Fig. 2. Figure 4 shows the volatility dynamic of the US dollar vis Russian ruble. There is no difference between the first and the second order filters on the chosen scale. However, more detailed analysis shows that the first order filter is closer on average to the estimated volatility than the second.

The main property of the first order filter derived from (4.1) is geometric convergence to the limit κ. Hence the corresponding forecasting curve tends sharply to κ and then smoothly changes towards the final point $\overline{v}$ (see Figs. 2(a) and 3(a). κ is approximated by the average of all the points on the graph). The behaviour of the second order forecasting curve also depends on parameter κ, but this dependence is much weaker (see Figs. 2(b) and 3(b)). Both filters

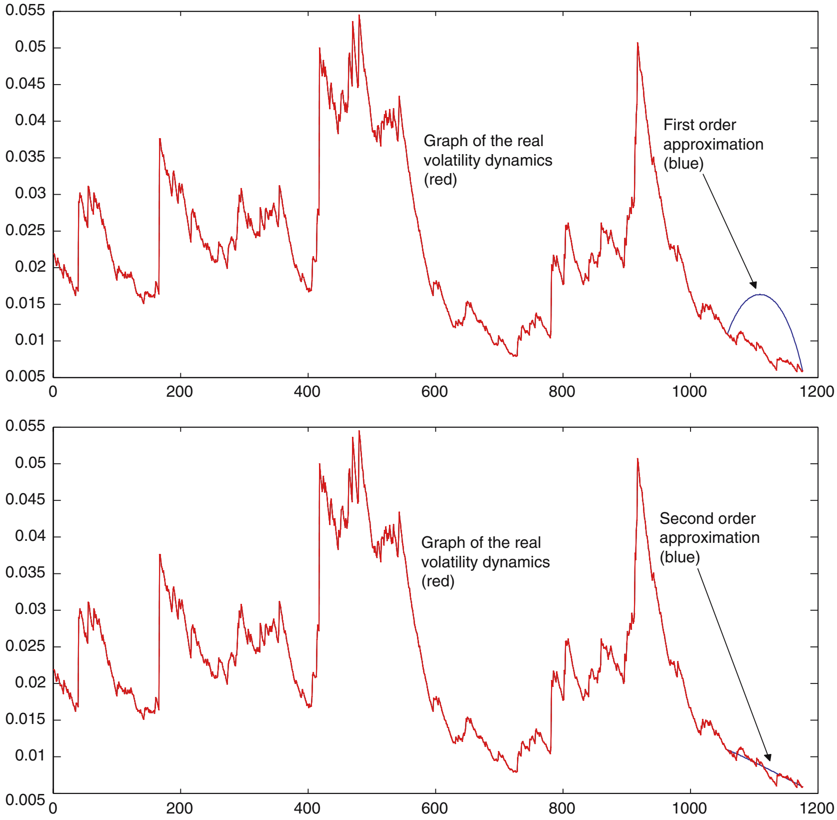

Fig. 2 Volatility of IBM: (**a**) First order; (**b**) Second order

"remember" historical information (κ the overall average volatility) but the first filter has better memory than the second one. The first order filter will be more appropriate than the second one in the situations when a volatility oscillates symmetrically about its mean, but in all other cases the second order filter is expected to be superior. Using Goldentayer et al. (2005) method and software we built volatility functions for about 500 world companies. In most cases, our analysis showed that the interpolation method should be provided by the second order filter. Yet, in 5% of cases, the first order filter was better (closer to the estimated curve).

6 Comparison of Historical and Implied Volatilities

The aim of this section is firstly to estimate the volatility functions of a number of companies under the model that volatility is a function of time, using the method of Goldentayer et al. (2005). Then to obtain estimates of integrated volatility and then to check the agreement with the estimate obtained from

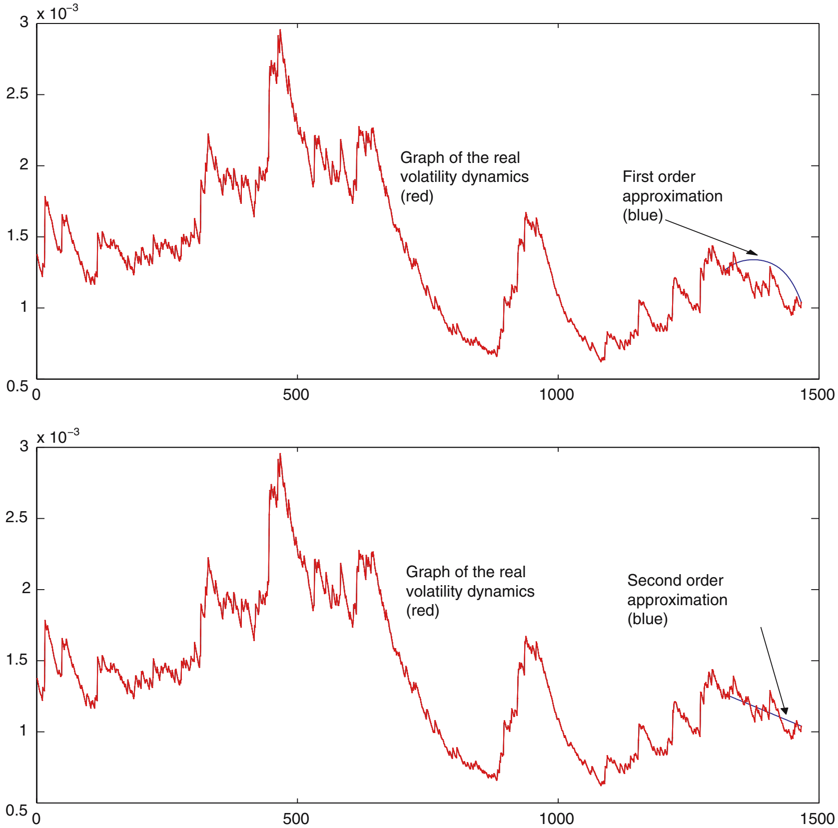

Fig. 3 Dynamics of the exchange volatility of the US dollar vis the Australian dollar: **(a)** The first order approximation; **(b)** The second order approximation

option prices by inverting the Black–Scholes formula. The model for prices where volatility is a deterministic function of time t or a stochastic process driven by Brownian motion independent that of Brownian motion of the stock goes back to Hull and White (1987) and Stein and Stein (1991). Hull and White (1987) and Stein and Stein (1991) show that the price of an option is given by the Black–Scholes formula where the constant volatility parameter is replaced by the averaged future volatility over the life of the option. Their derivations use PDE methods. Here we give a short derivation of the price of the option in a slightly more general model where μ_t mean return on stock is also a function of time, using the probability approach of arbitrage-free pricing. Assume the following model for stock

$$\mathrm{d}S_t = \mu_t S_t \mathrm{d}t + \sqrt{v_t} S_t \mathrm{d}W_t, \tag{6.1}$$

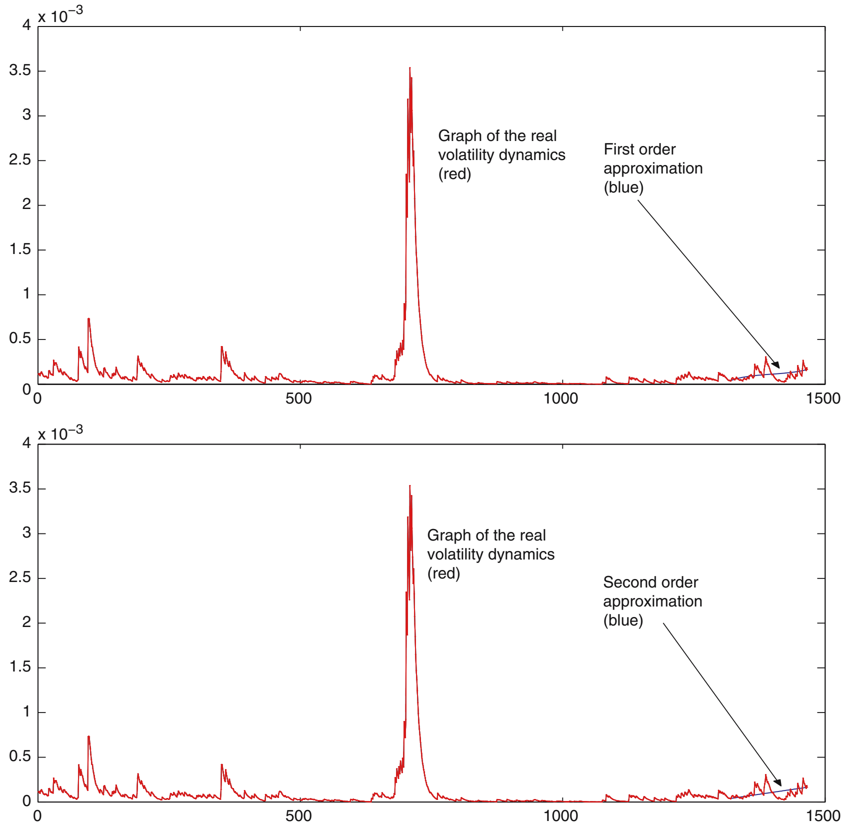

Fig. 4 Dynamics of the exchange volatility of the US dollar vis the Russian ruble: **(a)** The first order approximation; **(b)** The second order approximation

where W_t is a standard Wiener process, and v_t is a function of time. A general expression for the call option that expires at T with exercise price K is

$$C_t(T,K) = \mathrm{e}^{r(T-t)}\mathbf{E}_Q\left((S_T - K)^+ \middle| \mathcal{F}_t\right),$$

where Q is the so-called equivalent martingale measure, i.e. under Q the process $S_t\mathrm{e}^{-rt}$ is a martingale. The net effect of this is that the drift parameter μ_t does not enter the option formula, and (6.1) reduces to (3.1), where μ_t is replaced by r and a different Brownian motion is used, although we kept same notation.

By Itô's formula

$$S_T = S_t \exp\left[\int_t^T (r - v_s/2)\,\mathrm{d}t + \int_t^T \sqrt{v_s}\mathrm{d}W_s\right].$$

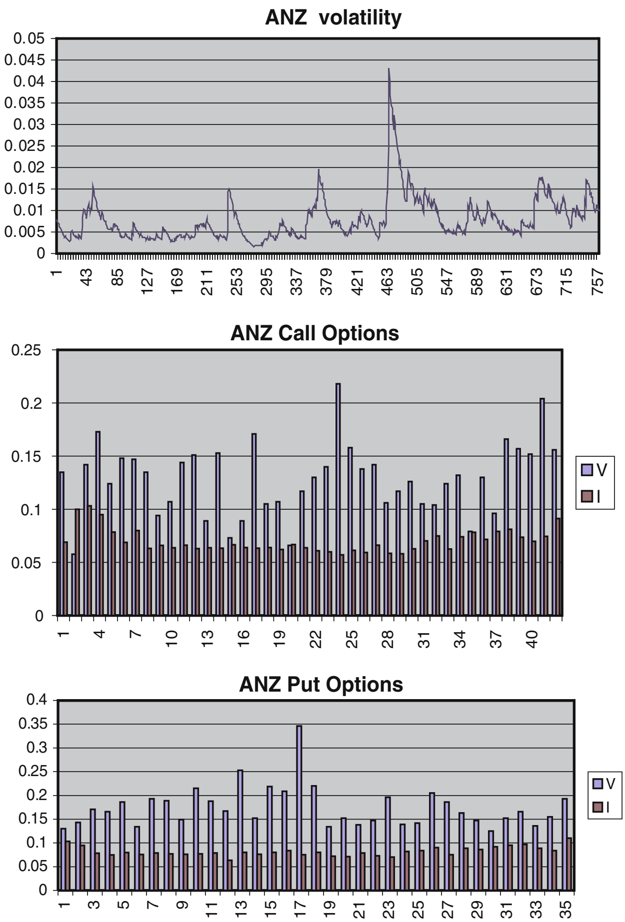

Fig. 5 Volatility function for ANZ and comparison of I and V for ANZ call and put options

The Itô integral with a deterministic integrand is Gaussian, hence $\int_t^T \sqrt{v_s}\mathrm{d}W_s$ has normal distribution, mean zero and variance $\int_t^T v_s \mathrm{d}s$. Therefore conditional distribution of S_T given S_t is lognormal and conditional expectation is given by the Black–Scholes formula,

$$C_t(T,K) = BS\Big(\frac{1}{T-t}\int_t^T v_s \mathrm{d}s\Big),$$

with the notation $BS(\sigma^2) = S_t\Phi(h) - K\mathrm{e}^{-r(T-t)}\Phi\left(h - \sigma\sqrt{T-t}\right)$, $\Phi(u)$ is the standard normal distribution function and $h = \dfrac{\log S_t/K + (r+\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}$.

I denotes the estimate of integrated volatility

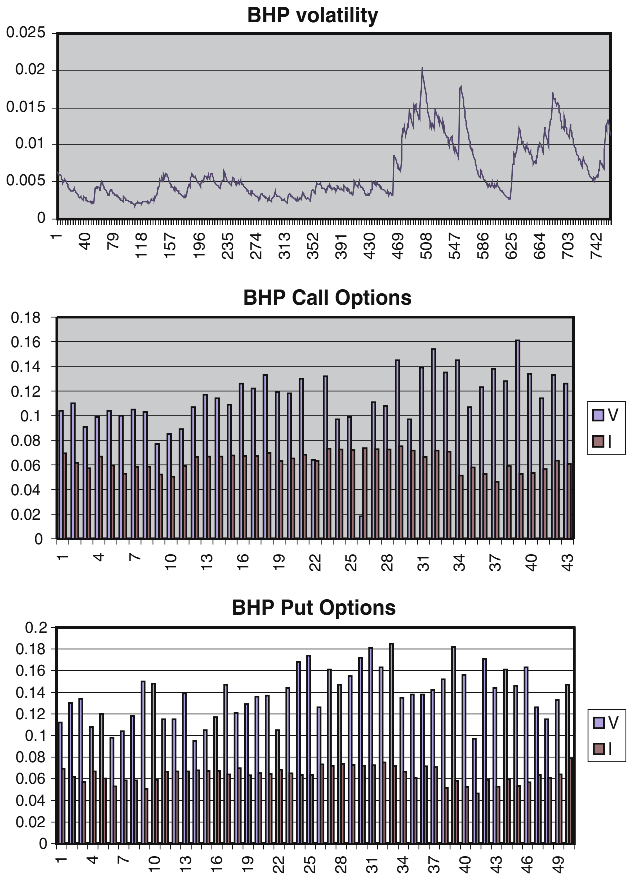

Fig. 6 Volatility function for BHP and comparison of I and V for BHP call and put options

$$I = \sqrt{\frac{1}{T-t}\int_t^T \widehat{v}(s)ds},$$

where $\widehat{v}(s)$ is obtained by using the first order filter (both filters are very close). V denotes the implied volatility, i.e. that value of σ in the Black–Scholes formula that gives the observed market price of an option.

We provide estimated volatility curves and comparisons for three companies between 27th of December 1995 to 14th of May 1997: Australia & New Zealand Banking Group (ANZ), BHP Billiton (BHP), News Corporation (NCP). Implied volatility is calculated separately for call and put options. The strikes K's for at the money options are given in Table 1, and maturities are chosen so that direct comparisons between the two estimates are possible. Namely, from the data available on options implied volatilities are calculated, and then for the same maturities integrated volatility estimator is computed.

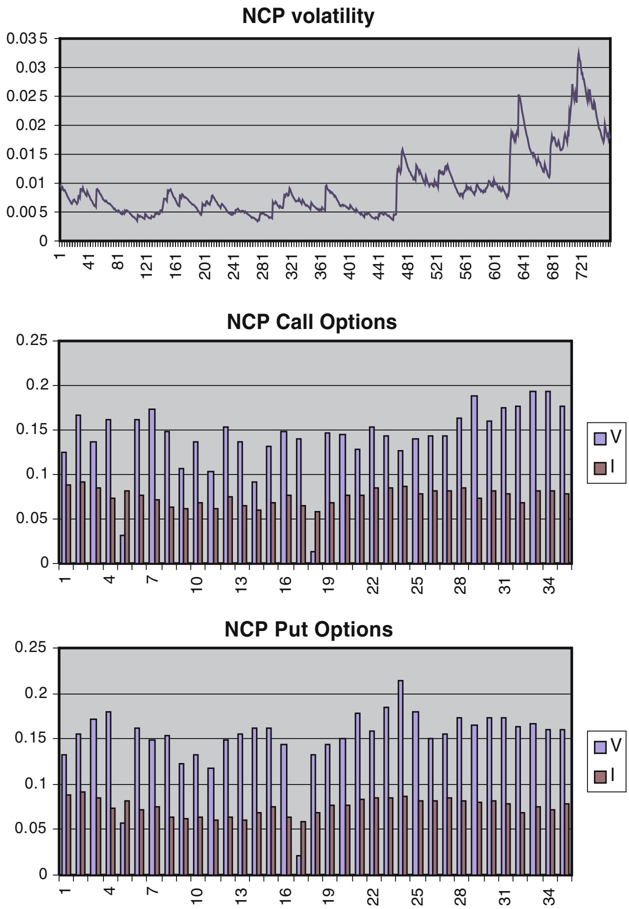

Fig. 7 Volatility function for NCP and comparison of I and V for NCP call and put options

The equation for integrated volatility was approximated numerically by averaging. The values of $T-t$ are plotted in Figs. 5–7 in the x-axis. They show substantial difference between integrated and implied volatilities. The upper panels in

Table 1 Options for ANZ, BHP and NCP at maturity

Option	Strike
ANZ call	6.75
ANZ put	6.75
BHP call	17.5
BHP put	17.5
NCP call	7.0
NCP put	7.0

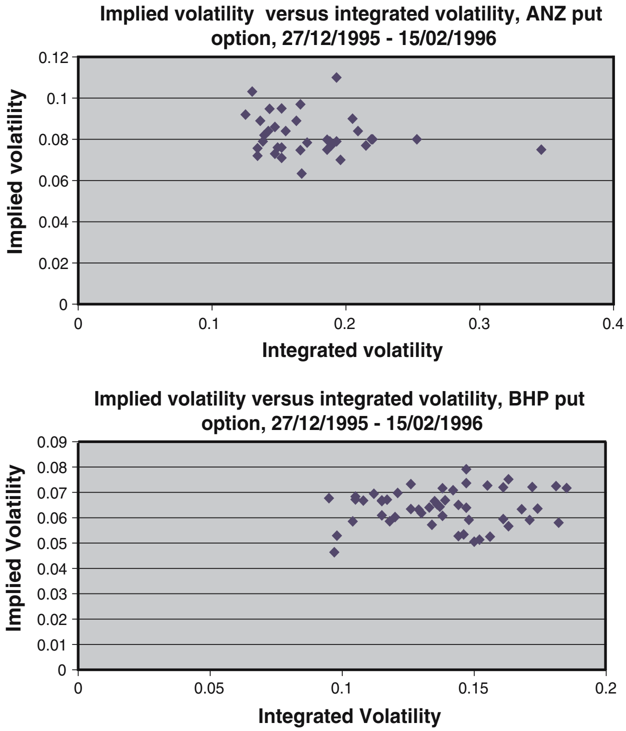

Fig. 8 Implied volatility versus integrated volatility

Table 2 Regression equations and correlations for the call and put options for ANZ, BHP and NCP

Regression and correlation	ANZ		BHP		NCP	
	Call	Put	Call	Put	Call	Put
$I = aV + b$						
a	−0.00585	−0.03937	0.020830	0.035903	0.081272	0.112515
b	0.07070	0.08890	0.065760	0.058934	0.063890	0.058198
$r_{I,V}$	0.154246	0.168609	0.152362	0.141581	0.169936	0.169961

Figs. 5– 7 are historical volatility estimates obtained by the Goldentayer–Klebaner–Liptser method of Goldentayer et al. (2005).

Figure 8. below shows the relationship between implied and integrated volatilities for ANZ and BHP put options.

Our conclusion that I and V are not related is made on the basis of the available statistical information. The regression equations in the form $I = aV + b$ and correlation coefficients $r_{I,V}$ are provided in Table 2 for call and put options for these companies. Table 2 shows weak correlations 0.14 – 0.17.

Acknowledgements The authors thank R. Liptser for useful discussions, L. Goldentayer for software and C. Lill for help with numerical calculations. The authors also thank the referees for their comments.

References

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., & Diebold, F. X. (2005). Volatility forecasting. NBER working paper 11188 (March, 2005). Accessible at: http://www.nber.org/papers/w11188

Anderson, H. M., & Vahid, F. (2005). Forecasting the volatility of Australian stock returns: Do common factor help? Working paper 451 (2005). Canberra: Australian National University.

Barndorff-Nielsen, O. E., & Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics (with discussion). *Journal of the Royal Statistical Society, Series B*, 63, 167–241.

Barndorff-Nielsen, O. E., & Shephard, N. (2002a). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics, 17*, 457–477.

Barndorff-Nielsen, O. E., & Shephard, N. (2002b). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B*, 64, 253–280.

Barndorff-Nielsen, O. E., & Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics, 2*, 1–37.

Bichteler, K., Gravereaux, J., & Jacod, J. (1987). *Malliavin calculus for processes with jumps*. Stochastic Monographs (vol. 2). New York: Gordon and Breach Science Publishers.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economics, 81*, 637–659.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*, 307–327.

Chow, P.-L., Khasminskii, R., & Liptser, R. (1997). Tracking of signal and its derivatives in Gaussian white noise. *Stochastic Processes and Their Application, 69*, 259–273.

Del Moral, P., Jacod, J., & Protter, P. (2001). The Monte-Carlo method for filtering with descrete observations. *Probability Theory and Related Fields, 120*, 346–368.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica, 50*, 987–1008.

Ghysels, E., Harvey, A. C., & Renault, E. (1996). Stochastic volatility. In C.R. Rao & G.S. Maddala (Eds.), *Statistical methods in finance* (pp. 119–191). Amsterdam, North-Holland.

Goldentayer, L., Klebaner, F., & Liptser, R. (2005). Tracking volatility. *Problems of Information Transmission, 41*, 212–229.

Hamza, K., & Klebaner, F.C. On the implicit Black–Scholes (submitted).

Harvey, A. C., Ruiz, E., & Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies, 61*, 247–264.

Hayashi, T. (2004). A discrete time model of high-frequency stock returns. *Quantitative Finance, 4*, 140–150.

Hayashi, T., & Mykland, P.A. (2005). Evaluating hedging errors: an asymptotic approach. *Mathematical Finance, 15*, 309–343.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with application to bond and currency options. *Review of Financial Studies, 6*, 327–343.

Hillebrand, E. (2005). Neglecting parameters changes in GARCH models. *Journal of Econometrics, 129*, 121–138.

Hillebrand, E. (2006). Overlaying time scales in financial volatility data. In T.B. Fomby & D. Terell (Eds.), *Advances in Econometrics*, (vol. 20) *Econometric analysis of financial and economic time series. Part B* (pp. 153–178). Elsevier.

Hillebrand, E., & Madeiros, M. C. (2006). Estimating and forecasting GARCH models in the presence of structural breaks and regime switches. In *Fronters of economics and globalization*. Elsevier (forthcoming).

Hull, J. C. (2002). *Options, futures and other derivative securities.* Prentice Hall.

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

Ibragimov, I. A., & Khasminskii, R. Z. (1980). On nonparametric estimation of regression. *Soviet Mathematical Doklads, 21*, 810–814.
Ibragimov, I. A., & Khasminskii, R. Z. (1981). *Statistical estimation. Asymptotic theory*. New York: Springer.
Khasminskii, R., & Liptser, R. (2002). On-line estimation of a smooth regression function. *Theory of Probability and its Applications, 47*, 541–550.
Kim, S., Shephard, N., & Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies, 65*, 363–393.
Klebaner, F. C., Le, T., & Liptser, R. (2006). On estimation of volatility surface and prediction future spot volatility. *Applied Mathematical Finance, 13*, 13–25.
Kusuoka, S., & Yoshida, N. (2000). Malliavin calculus, geometric mixing, and expansion of diffusion functionals. *Probability Theory and Related Fields, 116*, 457–484.
Lee, S., Nishiyama, Y., & Yoshida, N. (2006). Test for parameter change in diffusion processes by cusum statistics based on one-step estimators. *Annals of Institute of Statistical Mathematics, 58*, 211–222.
Malliavin, P., & Thalmaier, A. (2006). *Stochastic calculus of variation and mathematical finance*, Heidelberg: Springer-Verlag.
Masuda, H., & Yoshida, N. (2005). Asymptotic expansion for Barndorff-Nielsen and Shephard's stochastic volatility model. *Stochastic Processes and Applications, 115*, 1167–1186.
Mercurio, D., & Spokoiny, V. (2004). Statistical inference for time-inhomogeneous volatility models. *The Annals of Statistics, 32*, 577–602.
Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Sciences, 4*, 141–183.
Polsehl, J., & Spokoiny, V. (2000). Adaptive weights smoothing with applications to image restoration. *Journal of the Royal Statistical Society, 62*, Series B, 335–354.
Polsehl, J., & Spokoiny, V. (2002). Local likelihood modelling in adaptive weights smoothing. preprint 787, WIAS.
Polsehl, J., & Spokoiny, V. (2003). Varying coefficients regression modelling by adaptive weights smoothing. Preprint 818, WIAS.
Polsehl, J., & Spokoiny, V. (2004). Varying coefficient GARCH versus local constant volatility modelling. Comparison of the predictive power. Preprint 977 WIAS.
Poon, S.-H., & Granger, C. W. J. (2003). Forecasting volatility in financial market: A review. *Journal of Economic Literature, 41*, 478–539.
Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In D. R. Cox, D. V. Hinkley O. E. Barndorff-Nielsen (Eds.), *Time series models in econometrics, finance and other fields.* (pp. 1–67). Chapman and Hall: London.
Shimizu, Y., & Yoshida, N. (2006). Estimation of parameters for diffusion processes from discrete observations. *Stochastic Interference of Stochastic Processes, 9*, 227–277.
Stein, E., & Stein, J. C. (1991). Stock price distributions with stochastic volatility: An analytic approach. *Review of Financial Studies, 4*, 727–752.
Taylor, S. J. (1986). *Modelling financial time series*. Chichester: John Wiley.
Yoshida, N. (1997). Malliavin calculus and asymptotic expansion for martingales. *Probability Theory and Related Fields, 109*, 301–342.