MPRA

Forecasting and testing a non-constant volatility

Abramov, Vyacheslav and Klebaner, Fima

6 June 2006

FORECASTING AND TESTING A NON-CONSTANT VOLATILITY

VYACHESLAV M. ABRAMOV AND FIMA C. KLEBANER

ABSTRACT. In this paper we study volatility functions. Our main assumption is that the volatility is deterministic or stochastic but driven by a Brownian motion independent of the stock. We propose a forecasting method and check the consistency with option pricing theory. To estimate the unknown volatility function we use the approach of [12] based on filters for estimation of an unknown function from its noisy observations. One of the main assumptions is that the volatility is a continuous function, with derivative satisfying some smoothness conditions. The two forecasting methods correspond to the the first and second order filters, the first order filter tracks the unknown function and the second order tracks the function and it derivative. Therefore the quality of forecasting 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. Further, in deterministic volatility models the price of options is given by the Black-Scholes formula with averaged future volatility [16], [29]. This enables us to compare the implied volatility with the averaged estimated historical volatility. This comparison is done for five companies and shows that the implied volatility and the historical volatilities are not statistically related.

1. INTRODUCTION

The aim of this paper is to propose a method of forecasting a volatility function, and then check whether the models agree with option pricing theory. The concept of volatility is associated with fluctuations of a time series. More specifically, in finance volatility $\sqrt{v_t}$ is the function appearing in the the Black-Scholes model for the stock price S_t

$$(1.1) \qquad \mathrm{d}S_t = rS_t\mathrm{d}t + \sqrt{v_t}S_t\mathrm{d}W_t,$$

where W_t is a standard Wiener process. The function $\sqrt{v_t}$ is referred to as the spot volatility process. In the standard Black-Scholes model [6], [23], the spot volatility is assumed to be constant σ, i.e. $\sqrt{v_t} \equiv \sigma$. Recently (e.g. [11], [12], [13], [20] [21], [28] and many others), there has been an increasing attention to non-constant volatility models. We assume that in the above model the stock price is the only observable, and only at discrete times t_1, t_2, ..., 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 [12], which rests on nonparametric approach due to [18] and

[19]. The method is essentially prediction by solving of a control problem. This method might be especially useful in the context of volatility forecasting due to the extra information about volatility in the future that is derived from options, e.g. values of the implied volatility.

Volatility estimation and forecasting is discussed in a large number of papers (see [1] and the references therein, which also gives a review of this field). Specifically, [1] discusses a wide circle of problems of volatility forecasting based on GARCH, stochastic volatility and realized volatility; and includes different forecast evaluation methods for univariate and multivariate cases. Another recent paper [2] also discusses univariate and multivariate forecasting models for realized volatility in Australian stocks.

Our method of forecasting is based on the development of recent results of tracking historical volatility [12]. The approach of [12] represents a wide class of approximations and is based on adaptive algorithm for tracking historical volatility using ideas of non-parametrical statistics. We assume that unknown spot volatility function belongs to the Ibragimov-Khasminskii class [17], [18]. [1] If the spot volatility function is continuous and satisfies the Lipschitz condition with constant L, i.e.

$$|v(t) - v(s)| \leq L|t - s|$$

then it belongs to the class $\Sigma(1, L)$. Note that the integrated spot volatility is a differentiable function and therefore automatically belongs to the class $\Sigma(1, L)$. In this situation the first order filter is used and given by the formula (see Section 3 of [12])

$$(1.2) \qquad \widehat{v}_n = \left(1 - \frac{a_1}{N}\right)\widehat{v}_{n-1} + \frac{a_1\kappa}{N} + \frac{\vartheta}{N^{2/3}}\left(X_n - \widehat{v}_{n-1}\right),$$

where N is the number of observations, $n = 1, 2, \ldots, N$, and a_1 and κ are specific parameters of this filter, ϑ is the (unique) parameter chosen to minimize the innovation difference

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

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

and $\Delta = t_{n+1} - t_n = \frac{T}{N}$.

It is worth noting that the first order filter in [12] is derived under the assumption

$$(1.3) \qquad v_n = \left(1 - \frac{a_1}{N}\right)v_{n-1} + \frac{a_1\kappa}{N} + \text{white noise}.$$

---

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

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

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. If f belongs to the class $\Sigma(\beta, L)$, where β is a positive real but not integer number, then it is assumed that f has k derivatives, $k = \lfloor\beta\rfloor$, where $\lfloor\beta\rfloor$ denotes the integer part of β, and the kth derivative of f satisfies the Hölder condition with parameter $\alpha = \beta - k$.

More specifically, in this equation the standard deviation parameter of white noise is small having order $O\left(\dfrac{1}{N^{2/3}}\right)$. It is derived exactly in Chow, Khasminskii and Liptser [8] to be consistent with mean square error of kernel estimates obtained earlier in Ibragimov and Khasminskii [17], Theorem 2.1.

Therefore the best linear predictor $\widehat{v}_n$ is defined by the representation similar to (1.3), using the formula with the same coefficients and replacing the last term by zero when future observations are not available, i.e.

$$(1.4) \qquad \widehat{v}_n = \left(1 - \frac{a_1}{N}\right)\widehat{v}_{n-1} + \frac{a_1\kappa}{N}.$$

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 Section 3 of [12])

$$(1.5) \qquad \begin{aligned} \widehat{v}_n &= \widehat{v}_{n-1} + \frac{1}{N}\widehat{v}_{n-1}^{(1)} + \frac{\sqrt{2\vartheta}}{N^{4/5}}(X_n - \widehat{v}_{n-1}), \\ \widehat{v}_n^{(1)} &= \left(1 - \frac{a_1}{N}\right)\widehat{v}_{n-1}^{(1)} - \frac{a_2}{N}\widehat{v}_{n-1} + \frac{a_2\kappa}{N} + \frac{\vartheta}{N^{3/5}}(X_n - \widehat{v}_{n-1}), \end{aligned}$$

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 derived by assumption:

$$(1.6) \qquad \begin{aligned} v_n &= v_{n-1} + \frac{1}{N}v_{n-1}^{(1)} + \text{white noise}, \\ v_n^{(1)} &= \left(1 - \frac{a_1}{N}\right)v_{n-1}^{(1)} - \frac{a_2}{N}v_{n-1} + \text{white noise}, \end{aligned}$$

where the white noises in the first and second equations of (1.6) are independent. Similarly to the case of the first order filter, the standard deviation parameters of these noises are small, and correspondingly having orders $O\left(\dfrac{1}{N^{4/5}}\right)$ and $O\left(\dfrac{1}{N^{3/5}}\right)$. They are derived exactly in [8] to be consistent with mean square error of kernel estimates obtained in [17], Theorem 2.1.

Then the last terms of the first and second equations of (1.5) are replaced by zero for the projection and therefore we have

$$(1.7) \qquad \begin{aligned} \widehat{v}_n &= \widehat{v}_{n-1} + \frac{1}{N}\widehat{v}_{n-1}^{(1)}, \\ \widehat{v}_n^{(1)} &= \left(1 - \frac{a_1}{N}\right)\widehat{v}_{n-1}^{(1)} - \frac{a_2}{N}\widehat{v}_{n-1}. \end{aligned}$$

The above parameters a_1, a_2, N, κ of these two filters can be found by tuning procedure.

The accuracy of volatility approximation in [12] depends on the class of that volatility. If the class of volatility is higher, then there is more information on volatility function is used and accuracy of approximation is higher. However the numerical experiments of [12] show that the difference between approximations given by the first and second order filters is so small, that in most cases it is not actually visible on the graph. For this reason the paper discusses volatility classes of the first two orders and studies corresponding volatility forecasting by the first and second order filters only.

Along with the method of approximation volatility suggested in [12], there is a number of different methods of approximation of volatility in the literature which are discussed in the next section.

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 according to open periodical information about option prices, i.e. the value of $\frac{1}{T-t} \int_t^T v(s) \mathrm{d}s$ is assumed to be known, where T is the last point of time interval and t is an initial (current) time moment. According to our notation, where daily information is considered in discrete time scale, the above integral is approximately written as 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 $\overline{V}$.

The standard problem of interpolation is formulated as follows. 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. The detailed description of this minimization problem for the first and second order filters is given in next section.

The paper is structured as follows. In Section 2 we review briefly other methods of approximation and forecasting volatility known in the literature making a comparison when possible. In Section 3 we approach forecasting by the control method for the first and second order filters respectively. Section 4 discusses numerical examples of forecasting volatility. Section 5 checks volatility functions on options. Specifically, it checks whether the volatility obtained by the method of [12] agrees with the observed implied volatility.

2. Other methods of approximation and forecasting volatility

Along with the approach of [12] there is a number of different approaches to approximation of volatility in the literature (e.g. [2], [3], [9], [10], [14], [15], [22]). In this section we briefly describe these methods and compare with that of [12].

One of the simplest methods of approximation is discussed in [3] (see also [4] and [5]). The method is based on calculation of realized variance (realized volatility) for fixed intervals of given length h, containing a large number M of observations. The estimator is consistent (as $M \to \infty$). An approximated realized volatility is piecewise constant taking for intervals of length h some specific value. Accuracy of calculation essentially depends on chosen value M: It was shown in [4] that the above convergence is at rate $\frac{1}{\sqrt{M}}$ to asymptotically normal distribution of estimator. Some examples for exchange of US dollar/DM is given in [5] for different values of $M = 1, 8$ and 48.

To obtain satisfactory volatility approximation it is required a large volume of information. In our experiments with stock data it is taken $M = 100$ from daily information of a number of companies. Compared with the rate of convergence $\frac{1}{\sqrt{M}}$, this volume of information is small and does not give satisfactory accuracy. Only in a small number of cases the comparison results of two methods seem to be relatively
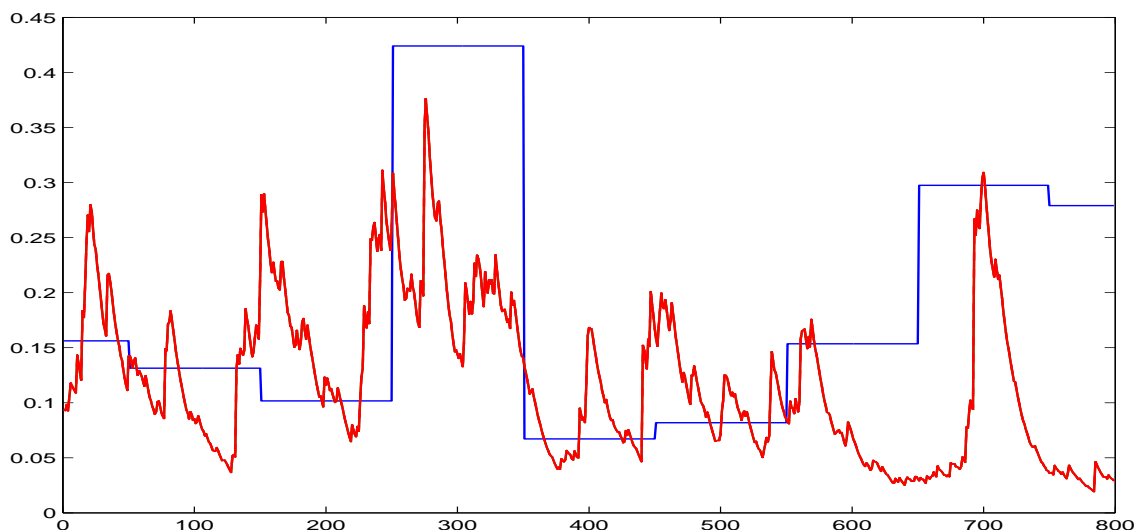
FIGURE 1. Comparison results for calculation of volatility by two methods for CECO Environmental Corporation: red line - the first order approximation by the method of [12]; blue line - piecewise constant approximation by the approach suggested by [3].

close to each other as in figure 1. Relatively close results by two approaches [3] and [12] can be expected in cases when price variation is not too large.

Anderson and Vahid [2] also study realized volatility in Australian stocks, and their approach is closely related to the approach of [5]. They use multi-factor models and show that application of the methods of factor analysis can improve forecasting of volatility. In the case when the cross-sectional factor dimension is not large, the estimation procedures for approximate factor models is robust to jumps. However similarly to the approach of Barndorff-Nielsen and Shephard [3], [4], [5], the approach of [2] requires a large volume of information to obtain piecewise constant approximation for volatility functions.

Another approach is suggested by Mercurio and Spokoiny [22]. They suppose that volatility can be locally approximated by a constant, that is for every time moment t there exists *interval of time homogeneity* $[t-m, t]$ where the volatility v_t varies very slowly. An algorithm for estimation of these intervals of time homogeneity has been proposed, and the estimate of volatility is obtained by local averaging over that interval. The local averaging adaptive estimate has been constructed in order to perform this local averaging and create the volatility function. However, the proposed algorithm of [22] seems to be hard. Another more simple way of estimating and forecasting volatility is based on the so-called *adaptive weight smoothing* (AWS) introduced by Polsehl and Spokoiny [24]. The AWS procedure is a method of non-parametric estimation which is based on locally constant smoothing with adaptive choice of weights for every pair of data points. The AWS procedure was then developed in [25], [26] and applied to extended GARCH models with varying coefficients in [27]. The suggested there adaptive procedure could estimate the GARCH coefficients as a function of time and was applied to short term forecasting in GARCH(1,1) models.

Hillebrand [14], [15] studied short-term forecasts for GARCH models as well as generalizations of GARCH models allowing several time scales. Specifically presence of two time scales, short and long, and their influence to the GARCH parameters has been investigated. Multi-scale stochastic volatility processes have been also studied in [9] and [10]. Specifically Fouque et al [9], [10] studied stochastic volatility asymptotics proposing to use a combination of regular and singular perturbations to analyze parabolic partial differential equations arising in context of prising options when a stochastic volatility varies in several time scales. They showed efficiency of asymptotic methods in presence of separation time scales between the main observed process and stochastic volatility.

3. Approach to forecasting by the control method

3.1. The first order filter.
Denote $a = 1 - \dfrac{a_1}{N}$ and $b = \dfrac{a_1 \kappa}{N}$ in (1.4), and rewrite (1.4) in the form:

$$(3.1) \qquad \widehat{v}_n = a\widehat{v}_{n-1} + b.$$

Since $a < 1$, recursion in (3.1) converges to the fixed value $\widehat{v}_\infty$, and

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

Hence

$$(3.2) \qquad \widehat{v}_\infty = \frac{b}{1-a} = \kappa.$$

It is clear that following recursion (3.1) we most likely won't end up at the specified at time N point $\overline{v}$. Therefore introduce a control sequence u_n ($n = n_0, \ldots, N$), to be determined by minimization of the mean square error of approximation, and the following recurrence relation

$$(3.3) \qquad \widetilde{v}_n = a\widetilde{v}_{n-1} + b + u_n,$$

(which in the case $u_n \equiv 0$ gives us approximation (3.1))

Let N be a large number (of steps). The values $\widetilde{v}_n$ are assumed to be known for all $n = 1, 2, \ldots, n_0$. For $n = N$, $\widetilde{v}_N = \overline{v}$, where $\overline{v}$ can be deduced from the known value $\overline{V}$ (it turns out $\overline{v} = \overline{V} - \sum_{i=n_0}^{N-1} \widehat{v}_i$.)

Thus, the value $\overline{v}$ is known, and the sequence u_n should be chosen such that

$$(3.4) \qquad \begin{cases} \widetilde{v}_N = \overline{v}, \\ \sum_{n=n_0+1}^{N} u_n^2 \quad \text{is minimal.} \end{cases}$$

The control sequence u_n is found as follows. For the point $\overline{v} = \widetilde{v}_N$ we have

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

On the other hand, according to (3.1)

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

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,

$$(3.5) \qquad \left(\sum_{n=n_0+1}^{N} a^{N-n} u_n \right)^2 \leq \sum_{n=n_0+1}^{N} a^{2(N-n)} \cdot \sum_{n=n_0+1}^{N} u_n^2.$$

The equality in (3.5) is achieved if and only if $a^{N-n} = cu_n$ for some constant c, and since the equality in (3.5) is associated with the minimum of the left-hand side of (3.5), 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:

$$(3.6) \qquad c^* = \frac{\widetilde{v}_N - \widehat{v}_N}{\sum_{i=n_0+1}^{N} a^{2(N-i)}}.$$

Thus, the sequence u_n satisfying (3.4) is

$$(3.7) \qquad u_n = \frac{\overline{v} - \widehat{v}_N}{\sum_{i=n_0+1}^{N} a^{2(N-i)}} \cdot a^{N-n},$$

and its substitution for (3.3) yields

$$(3.8) \qquad \widetilde{v}_n = a\widetilde{v}_{n-1} + b + \frac{\overline{v} - \widehat{v}_N}{\sum_{i=n_0+1}^{N} a^{2(N-i)}} \cdot a^{N-n}.$$

The aforementioned explicit values $a = 1 - \dfrac{a_1}{N}$ and $b = \dfrac{a_1 \kappa}{N}$ are finally substituted for (3.1) and (3.8) in order to obtain the desired values of $\widetilde{v}_n$, $n = n_0+1, \ldots, N-1$.

3.2. The second order filter.

The equations for the second order filter can be written in the form of a two dimensional analogue of the equations for the first order filter. Specifically, the analogue of (3.3) is written as

$$(3.9) \qquad \widetilde{\mathbf{v}}_n = A\widetilde{\mathbf{v}}_{n-1} + \mathbf{b} + \mathbf{u}_n,$$

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. To be specific note, that the control vector $\mathbf{u}_n$ in (3.9) 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

$$(3.10) \qquad \widehat{\mathbf{v}}_n = A\widehat{\mathbf{v}}_{n-1} + \mathbf{b},$$

where $\widehat{\mathbf{v}}_n = \begin{pmatrix} \widehat{v}_n \\ \widehat{v}_n^{(1)} \end{pmatrix}$, and the components of this vector $\widehat{v}_n$ and $\widehat{v}_n^{(1)}$ are defined by

(1.5). Note also that explicit form of A and $\mathbf{b}$ are $A = \begin{pmatrix} 1 & \dfrac{1}{N} \\ -\dfrac{a_2}{N} & 1 - \dfrac{a_1}{N} \end{pmatrix}$,

$\mathbf{b} = \begin{pmatrix} 0 \\ \dfrac{a_2 \kappa}{N} \end{pmatrix}$. Similarly to the above one-dimensional case, the control sequence u_n should be chosen such that

$$(3.11) \qquad \begin{cases} \widetilde{v}_N = \overline{v}, \\ \sum_{n=n_0+1}^{N} u_n^2 \quad \text{is minimal}, \end{cases}$$

where $\widetilde{v}_N$ is the first component of the vector $\widetilde{\mathbf{v}}_N$ which is determined by (3.9).

Next, 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:

$$(3.12) \qquad (\widetilde{\mathbf{v}}_N - \widehat{\mathbf{v}}_N)^\top (\widetilde{\mathbf{v}}_N - \widehat{\mathbf{v}}_N) = \left(\sum_{n=n_0+1}^{N} A^{N-n} \mathbf{u}_n \right)^\top \left(\sum_{n=n_0+1}^{N} A^{N-n} \mathbf{u}_n \right),$$

where $\top$ is the notation for the matrix (vector) transpose operation.

Now, let $(a_{i,j})_n$ denote the element of the matrix A^n taken in an intersection of the ith row and jth column. Then, taking into account that the second component of all vectors $\mathbf{u}_n$ is equal to zero, the right-hand side of (3.12) reduces to

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

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

Therefore, applying the Cauchy-Schwartz inequality, we obtain

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

$$(3.13) \qquad \leq \sum_{n=n_0+1}^{N} \begin{pmatrix} (a_{1,1})_{N-n} \\ (a_{2,1})_{N-n} \end{pmatrix}^\top \begin{pmatrix} (a_{1,1})_{N-n} \\ (a_{2,1})_{N-n} \end{pmatrix} \cdot \sum_{n=n_0+1}^{N} u_n^2$$

$$= \sum_{n=n_0+1}^{N} [(a_{1,1})_{N-n}^2 + (a_{2,1})_{N-n}^2] \cdot \sum_{n=n_0+1}^{N} u_n^2.$$

Here in (3.13) we use the standard notation for the square of $(a_{i,j})_n$: $(a_{i,j})_n^2 = (a_{i,j})_n \cdot (a_{i,j})_n$.

The equality for the left-hand side of (3.13) is achieved if and only if

$$\sqrt{(a_{1,1})_{N-n}^2 + (a_{2,1})_{N-n}^2} = c u_n$$

for some constant c, and since the equality is associated with the minimum of the left-hand side of (3.13), the problem reduces to find an appropriate value $c = c^*$ such that

$$(3.14) \qquad u_n = c^* \sqrt{(a_{1,1})_{N-n}^2 + (a_{2,1})_{N-n}^2}.$$

Therefore,

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

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:

$$(3.15) \qquad c^* = \frac{\widetilde{v}_N - \widehat{v}_N}{\sum_{i=n_0+1}^{N}[(a_{1,1})_{N-i}^2 + (a_{2,1})_{N-i}^2]}.$$

Substituting (3.15) for (3.14) and taking into account (3.11) we finally obtain

$$(3.16) \qquad u_n = \frac{\overline{v} - \widehat{v}_N}{\sum_{i=n_0+1}^{N}[(a_{1,1})_{N-i}^2 + (a_{2,1})_{N-i}^2]} \sqrt{(a_{1,1})_{N-n}^2 + (a_{2,1})_{N-n}^2}.$$

Recall that in the case of the second order filter, $A = \begin{pmatrix} 1 & \dfrac{1}{N} \\ -\dfrac{a_2}{N} & 1 - \dfrac{a_1}{N} \end{pmatrix}$,

$\mathbf{b} = \begin{pmatrix} 0 \\ \dfrac{a_2 \kappa}{N} \end{pmatrix}$, and these matrix and vector are used for calculations in (3.16) and (3.9).

4. NUMERICAL EXAMPLES OF FORECASTING VOLATILITY

The numerical examples are based on the real data on financial market.

In figure 2 the volatility dynamics for the IBM corporation stock is presented. The figure consists of two graphs of real volatility dynamics compared with its approximations with the first and the second order filters respectively. At the end the graphs are split into two colors: the real dynamics of volatility is marked by red while the approximated by one or other filter is marked by blue. It is seen from the figure that the second order approximation in the given case has a visible advantage over the first order approximation.

This advantage is slightly less visible in figure 3, where dynamics of the exchange volatility of the US dollar vis the Australian dollar is given. Then in figure 4 the dynamics of the exchange volatility of US dollar vis Russian ruble is given. The difference between the first and the second order approximation seems not to be visible at all in the given scaling. Which approximation is more appropriate in this
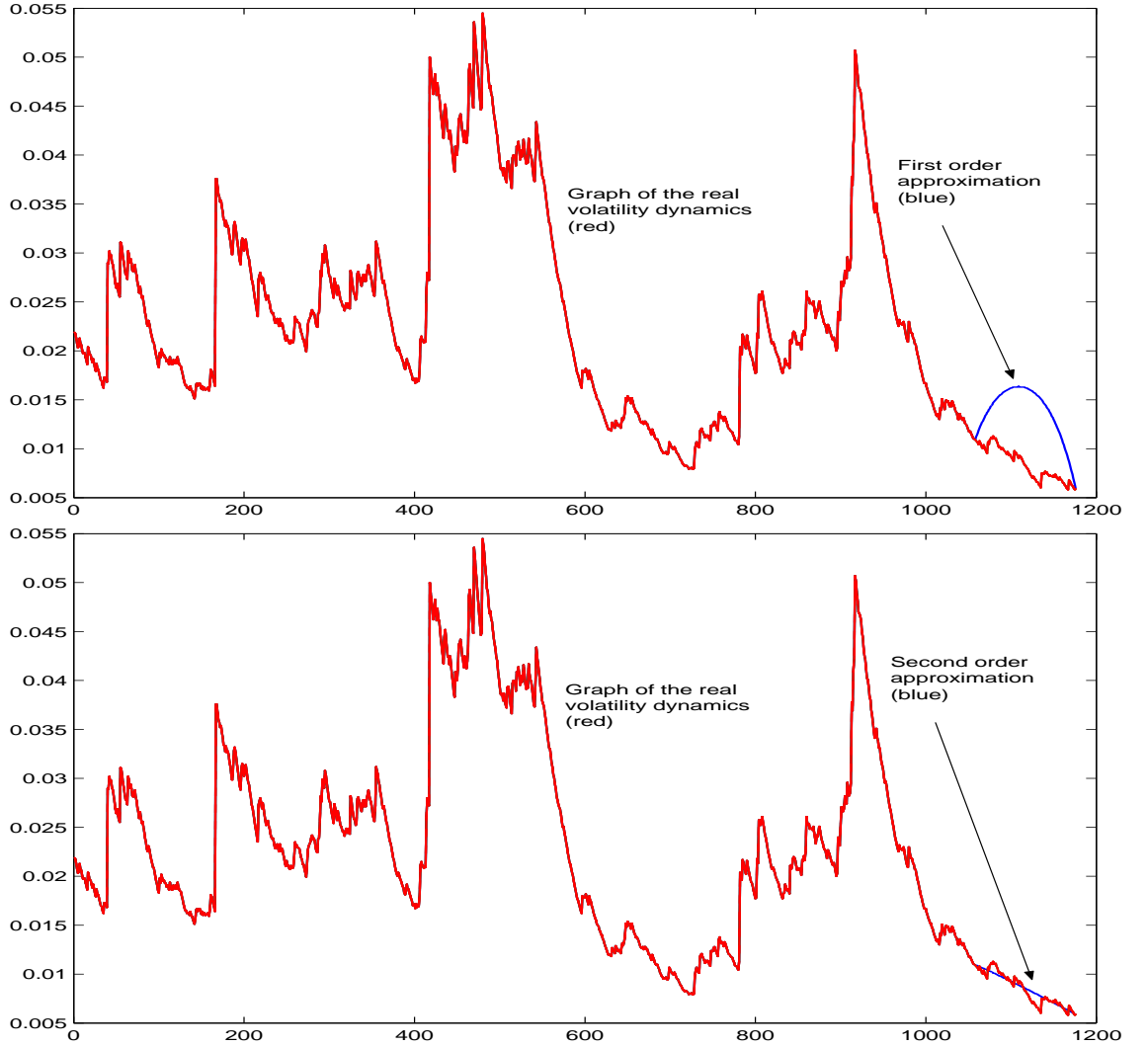
FIGURE 2. Dynamics of the stock volatility of IBM Corporation:
(a) First order approximation; (b) Second order approximation

case? We made the following elementary calculations. For the observed volatility
dynamics v_n, $n = n_0, n_0 + 1, \ldots, N$ its average is

$$\mathcal{V}(n_0, N) = \frac{1}{N - n_0} \sum_{i=n_0+1}^{N} v_i.$$

Similar averaging was done for the data corresponding to the first and second order
approximations:

$$\mathcal{V}^{(i)}(n_0, N) = \frac{1}{N - n_0} \sum_{i=n_0+1}^{N} v_i^{(i)},$$
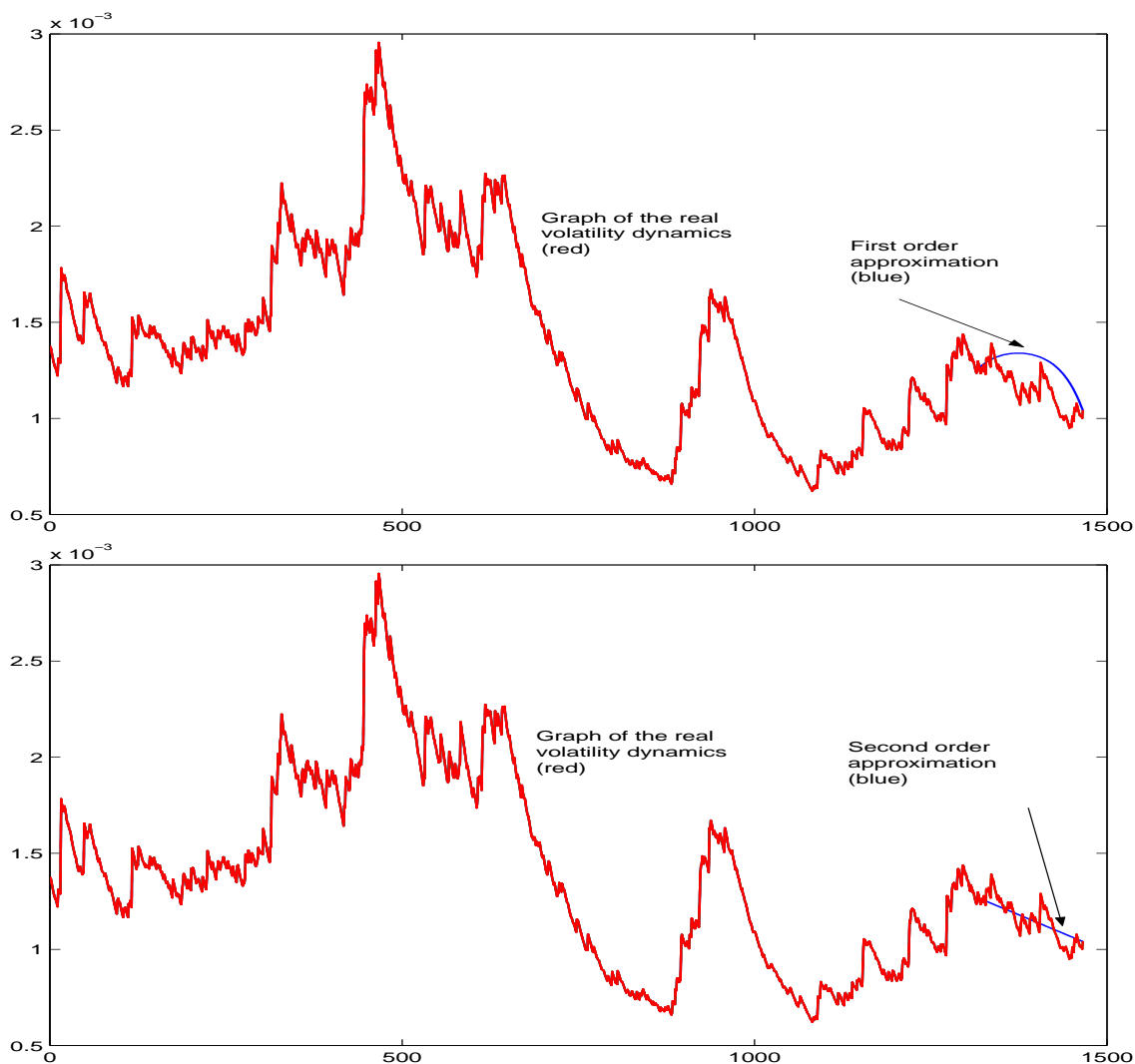
FIGURE 3. Dynamics of the exchange volatility of the US dollar
vis the Australian dollar: (a) First order approximation; (b) Second
order approximation

where the superscript $^{(i)}$, $i = 1, 2$, characterizes first or second order approximation.
We observed the inequality

$$(4.1) \qquad \left| \mathcal{V}^{(1)}(n_0, N) - \mathcal{V}(n_0, N) \right| \leq \left| \mathcal{V}^{(2)}(n_0, N) - \mathcal{V}(n_0, N) \right|,$$

the left-hand side of which was 0.000132 while the right-hand side 0.000697. The
results obtained by the simple calculation above justify a possible advantage of the
first filter. But as seen this advantage is negligible. The advantage of the first order
filter in this case can be explained by analytic properties of the forecasting curves
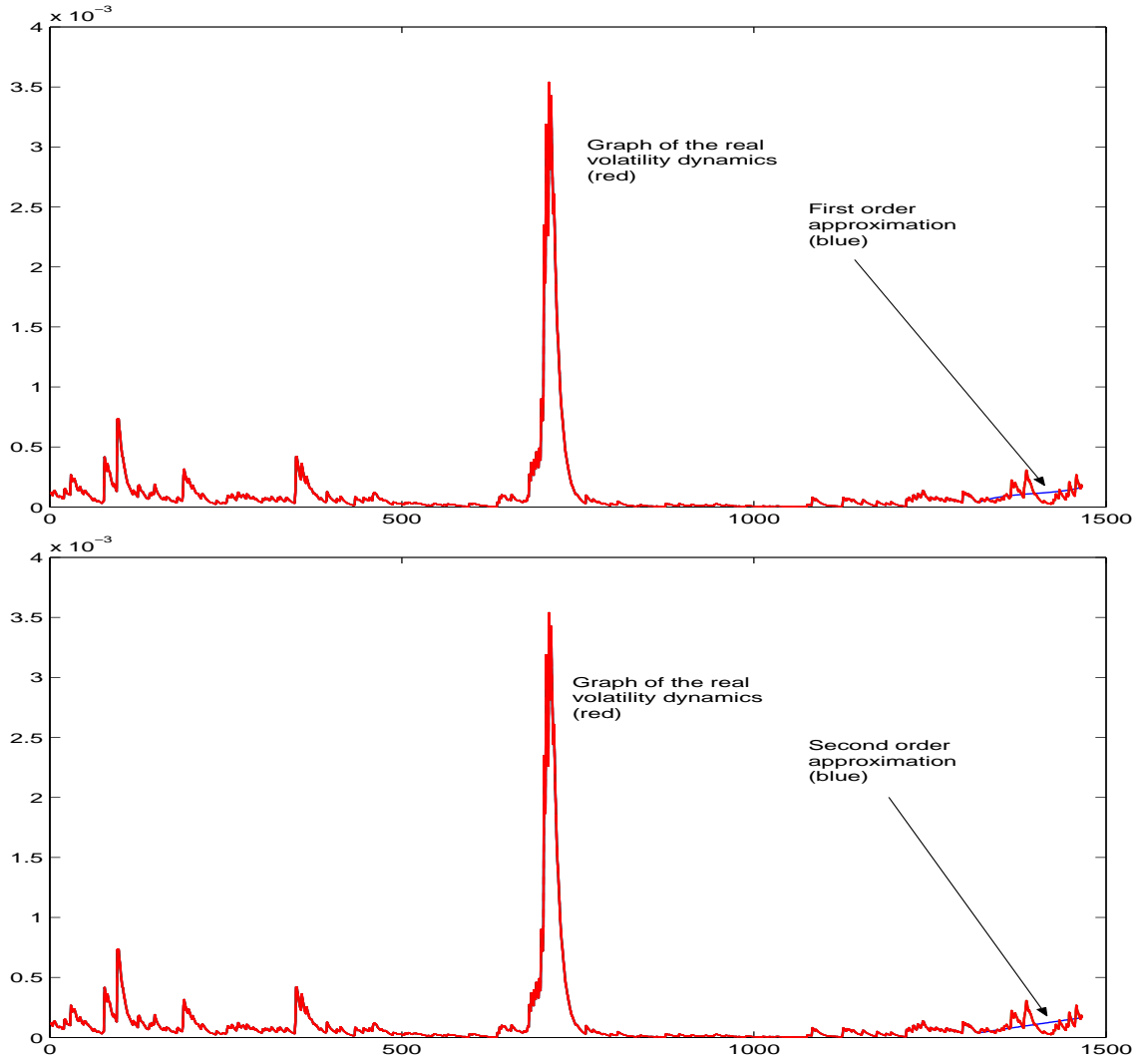of these two filters.

FIGURE 4. Dynamics of the exchange volatility of the US dollar
vis the Russian ruble: (a) First order approximation; (b) Second
order approximation

The main property of the first order filter is based on (3.1) converging geomet-
rically fast to the limit κ, hence and the corresponding forecasting curve tends
sharply to κ and then smoothly changes towards the point $\overline{v}$ (see Figures 2 (a) and
3 (a)).

The behaviour of the second order forecasting curve also depends on parameter κ,
but this dependence is much weaker. The curvature of the second order forecasting
curve is small, its trajectory is close to the segment of straight line connecting two
points (see Figures 2 (b) and 3 (b)). This is observed experimentally and can be
proven using formulæ (3.1) and (3.10) above. Both filters "remember" historical

information (κ the overall average volatility) but the first filter has better memory than the second.

First order filter will be better than the second one in the situations when volatility oscillates symmetrically about its mean, but in all other cases the second order filter is superior.

5. TESTING VOLATILITY ON OPTIONS

The aim of this section is to analyze the spot volatility functions of a number of companies and to check whether the obtained volatility as the function of time t can be considered as a function deterministic or stochastic process but driven by Brownian motion independent of stock. Such type of analysis goes back to the classical results of Hull and White [16] and Stein and Stein [29].

Assume that the model of stock is described by

$$(5.1) \qquad \mathrm{d}S_t = \mu_t S_t \mathrm{d}t + \sqrt{v_t} S_t \mathrm{d}W_t,$$

where W_t is a standard Wiener process, and v_t is a function of time. Then the price of the option is given by the Black-Scholes formula with the *averaged future volatility*, see [29].

General expression for the call option that expires at T with exercise price κ is

$$C(T, K) = \mathrm{e}^{r(T-t)} \mathbf{E}_Q (S_T - K)^+,$$

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 (5.1) reduces to (1.1).

Therefore, according to Itô's formula,

$$S_T = S_0 \exp\left[\int_0^T \left(r - \frac{v_t}{2} \right) \mathrm{d}t + \int_0^T \sqrt{v_t} \mathrm{d}W_t \right].$$

$\int_0^T \sqrt{v_t} \mathrm{d}W_t$ has normal distribution, mean zero and variance $\int_0^T v_t \mathrm{d}t$.

Then S_T is a lognormal random variable, and therefore $\mathbf{E}(S_T - K)^+$ is given by the Black-Scholes formula. So that when the spot volatility is a deterministic function then the price of options is given by

$$C(T, K) = E(S_T - K)^+ = B\left(\frac{1}{T} \int_0^T v_t \mathrm{d}t \right),$$

with the notation $B(\sigma^2) = S_0 \Phi(h) - K \mathrm{e}^{-rT} \Phi\left(h - \sigma\sqrt{T} \right)$,

$h = \dfrac{\log \dfrac{S_0}{K} + \left(r + \dfrac{\sigma^2}{2} \right) T}{\sigma\sqrt{T}}$, and $\Phi(u) = \dfrac{1}{\sqrt{2\pi}} \int_{-\infty}^u \mathrm{e}^{-x^2/2} \mathrm{d}x$ is the standard normal distribution.

Options prices from the following companies were sampled, where applicable, between 27th of December 1995 to 14th of May 1997.

- Australia & New Zealand Banking Group Ltd. (ANZ)
- BHP Billiton Ltd. (BHP)
- National Australia Bank Ltd. (NAB)
- News Corporation Ltd. (NCP)
- Thomas Nation-Wide Transportation (TNT)

Regression	ANZ		BHP		NCP	
and correlation	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

TABLE 1. Regression equations and correlations for the call and put options of companies ANZ, BHP and NCP

In figures 5-7 I denotes integrated volatility

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

and V implied volatility, i.e. that value of σ in the Black-Scholes formula that gives the observed market price of an option.

The computations below related to three companies ANZ, BHP and NCP show that there is visible difference between I and V. Specifically, we have the following estimations. Conclusion that I and V are distinct can be made on the base of the available statistical information. The regression equations in the form $I = aV + b$ and correlation coefficients $r_{I,V}$ are provided in Table 4 for call and put options of these companies. It is seen from Table 4 that the correlations $r_{I,V}$ vary in the bounds 0.14 - 0.17. These bounds enable us to conclude that there is no correlation between I and V, and our computational experiments support this hypothesis at the level of probability 0.95 (we used software ITSM-2000 from the book of Brockwell and Davis [7] for this hypothesis). Then absence of correlation helps to support the conclusion that I and V are distinct for all these three companies by using the standard statistical tests.

REFERENCES

[1] ANDERSEN, T.G., BOLLERSLEV, T., CHRISTOFFERSEN, P.F. AND DIEBOLD, F.X. Volatility forecasting. *NBER working paper 11188* (March, 2005). Accessible at: http://www.nber.org/papers/w11188

[2] ANDERSON, H.M. AND VAHID, F. Forecasting the volatility of Australian stock returns: Do common factor help? *Working paper 451* (2005). Australian National University, Canberra.

[3] BARNDORFF-NIELSEN, O.E. AND SHEPHARD, N. Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 17 (2002), 457-477.

[4] BARNDORFF-NIELSEN, O.E. AND SHEPHARD, N. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society*, Ser. B, 64 (2002), 253-280.

[5] BARNDORFF-NIELSEN, O.E. AND SHEPHARD, N. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2 (2004), 1-37.

[6] BLACK, F. AND SCHOLES, M. The pricing of options and corporate liabilities. *Journal of Political Economics*, 81 (1973), 637-659.

[7] BROCKWELL, P.J. AND DAVIS, R.A. *Introduction to Time Series Analysis and Forecasting*, 2nd edn. Springer, New York, 2002.
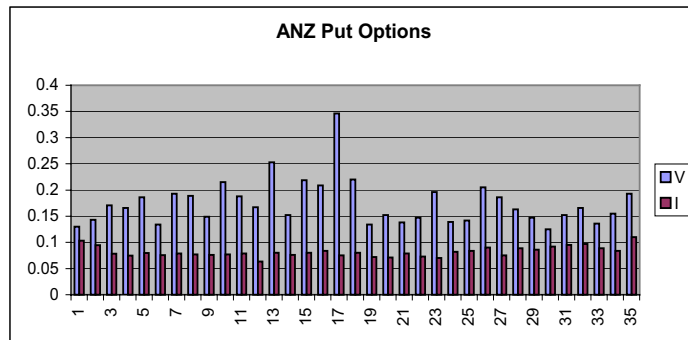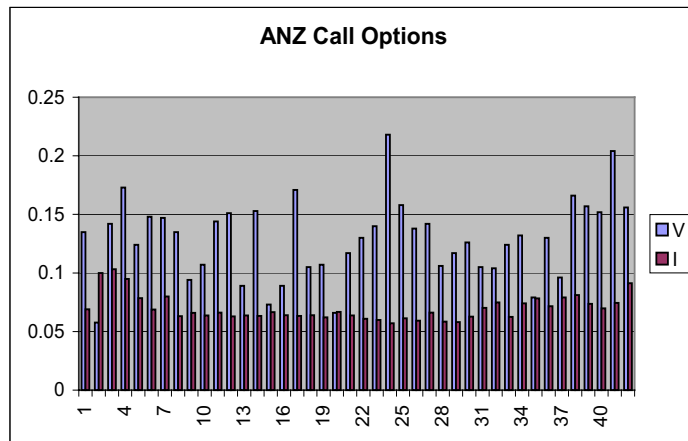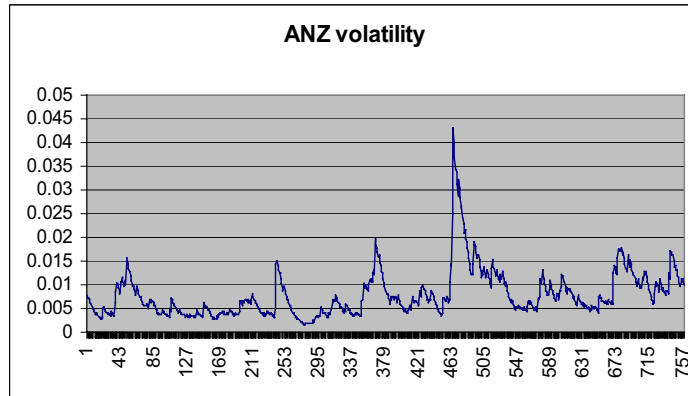
FIGURE 5. Volatility function for ANZ and comparison of I and V for ANZ call and put options
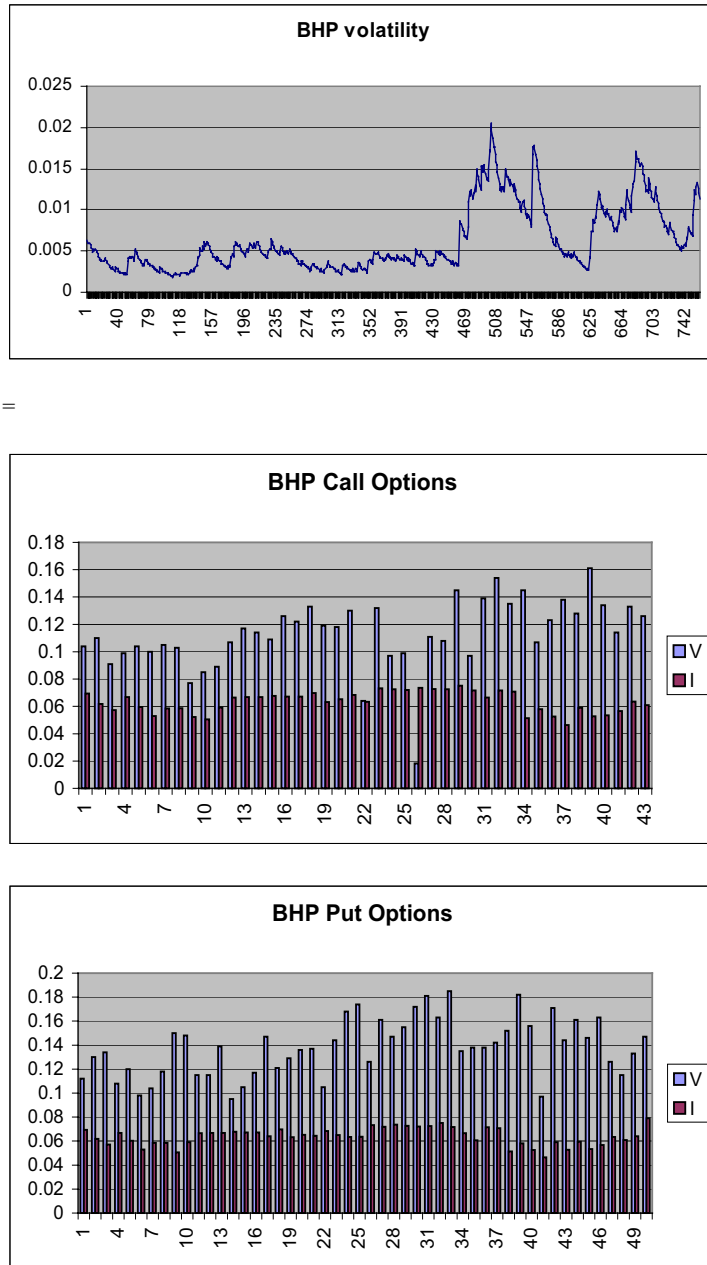
=

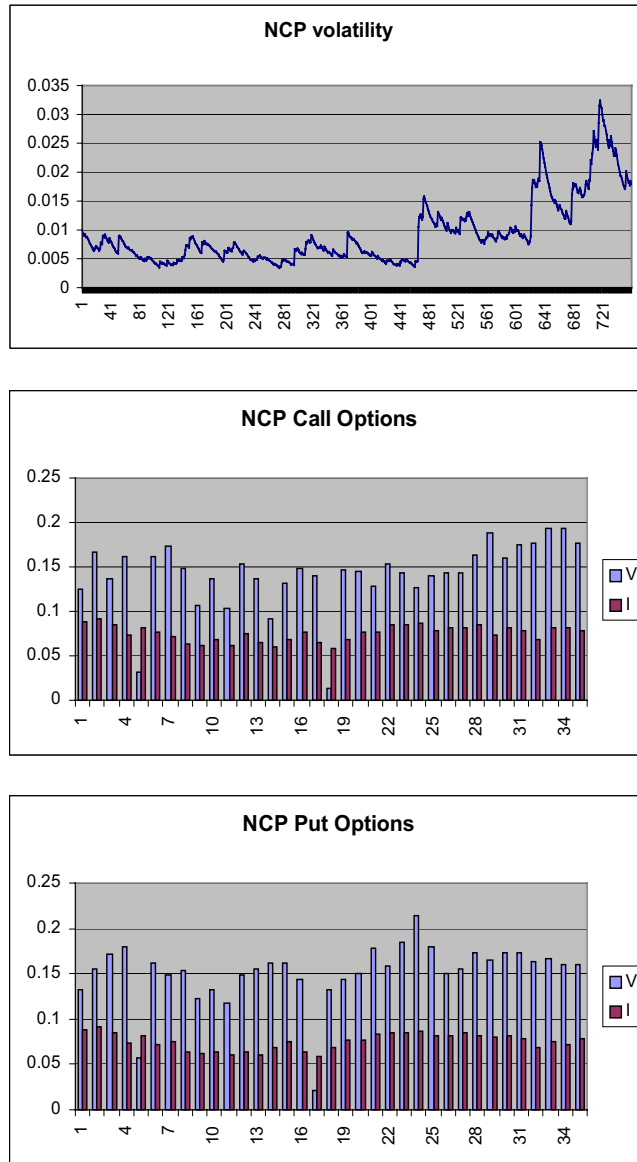FIGURE 6. Volatility function for BHP and comparison of I and V for BHP call and put options

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

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

[9] Fouque, J.-P., Papanicolaou, G., Sircar, R. and Solna, K. Singular perturbation in option prices. *SIAM Journal on Applied Mathematics*, 63 (2003), 1648-1665.

[10] Fouque, J.-P., Papanicolaou, G., Sircar, R. and Solna, K. Multiscale stochastic volatility asymptotics. *SIAM Journal for Multiscale Modelling and Simulation*, 2 (2003), 22-42.

[11] Ghysels, E., Harvey, A.C. and Renault, E. Stochastic volatility. In: *Statistical Methods in Finance* C.R.Rao and G.S.Maddala (eds) Amsterdam, North-Holland, 1996, pp. 119-191.

[12] Goldentayer, L., Klebaner, F. and Liptser, R. Tracking volatility, *Problems of Information Transmission*, 41 (2005), 212-229.

[13] Harvey, A.C., Ruiz, E. and Shephard, N. Multivariate stochastic variance models, *Review of Economic Studies*, 61 (1994), 247-264.

[14] Hillebrand, E. Overlaying time scales and persistence estimation in GARCH(1,1) models. Econometrics 0301003 EconWPA (2003).

[15] Hillebrand, E. Overlaying time scales in financial volatility. Econometrics 0501015 EconWPA (2005).

[16] Hull, J.C. and White, A. The pricing of options on asset with stochastic volatilities. *Journal of Finance*, 42 (1987), 281-300.

[17] Ibragimov, I.A. and Khasminskii, R.Z. On nonparametric estimation of regression. *Soviet Mathematical Doklads* 21 (1980), 810-814.

[18] Ibragimov, I.A. and Khasminskii, R.Z. *Statistical Estimation. Asymptotic Theory.* Springer, New York, 1981.

[19] Khasminskii, R and Liptser, R. On-line estimation of a smooth regression function, *Theory of Probability and its Applications*, 47 (2002), 541-550.

[20] Kim, S., Shephard, N. and Chib, S. Stochastic volatility: likelihood inference and comparison with ARCH models, *Review of Economic Studies*, 65 (1998), 363-393.

[21] Klebaner, F.C., Le, T. and Liptser, R. Estimating and predicting volatility by using implied and historical volatilities, submitted.

[22] Mercurio, D. and Spokoiny, V. Statistical inference for time-inhomogeneous volatility models. *The Annals of Statistics*, 32 (2004), 577-602.

[23] Merton, R.C. Theory of rational option pricing. *Bell Journal of Economics and Management Sciences*, 4 (1973), 141-183.

[24] Polsehl, J. and Spokoiny, V. Adaptive weights smoothing with applications to image restoration. *Journal of the Royal Statistical Society* 62, Ser. B, (2000), 335-354.

[25] Polsehl, J. and Spokoiny, V. Local likelihood modelling in adaptive weights smoothing. Preprint 787, WIAS, 2002.

[26] Polsehl, J. and Spokoiny, V. Varying coefficients regression modelling by adaptive weights smoothing. Preprint 818, WIAS, 2003.

[27] Polsehl, J. and Spokoiny, V. Varying coefficient GARCH versus local constant volatility modelling. Comparison of the predictive power. Preprint 977, WIAS, 2004.

[28] Shephard, N. Statistical aspects of ARCH and stochastic volatility. In: *Time Series Models in Econometrics, Finance and Other Fields* D.R.Cox, D.V.Hinkley and O.E.Barndorff-Nielsen (eds) London, Chapman and Hall, 1996, pp. 1-67.

[29] Stein, E. and Stein, J.C. Stock price distributions with stochastic volatility: An analytic approach. *Review of Financial Studies*, 4 (1991) 727-752.

School of Mathematical Sciences, Monash University, Building 28M, Clayton campus, Clayton, VIC 3800, Australia

  E-mail address: vyacheslav.abramov@sci.monash.edu.au

  E-mail address: fima.klebaner@sci.monash.edu.au