Financial Time and Volatility
Peters, R.T.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 2

The Realized Volatility of the Main Dutch (AEX) Stock Index

2.1 Introduction

Volatility, the "Holy Grail of financial mathematics", Goodhart and O'Hara (1997), is wrapped up as a free gift in sample functions (i.e. realizations) of financial processes. High frequency data yield bivariate sample functions, describing the behaviour both of the price and the realized volatility.

To obtain the realized (squared) volatility process one computes the quadratic variation of the sample function. This is theoretically justified by assuming that the underlying price process is a semi-martingale. Along these lines Andersen, Bollerslev, Diebold and Labys (2000) report estimates of the "realized volatility" of high frequency exchange rate returns and present evidence that the standardized returns are normally distributed. See also the work by Andersen, Bollerslev, Diebold and Labys (2001), Andersen, Bollerslev, Diebold and Ebens (2001) and Andersen, Bollerslev, Diebold and Labys (2003) where different aspects of the realized volatility process are investigated. All these papers assume that volatility and standardized returns are independent. This implies that the distribution of the returns can not be skewed. However, it is generally believed that the distribution of stock returns is skewed and so for these processes the independence assumption is too restrictive.

In this chapter the hypothesis that the standardized daily (scaled) log returns of the Dutch main stock index (AEX) are i.i.d. standard normal, is tested. It can not be rejected. We obtain this result by normalizing each daily
The realized volatility of the Main Dutch ...

return of the AEX, for the period May 1996 until September 2000, by means of the square root of the quadratic variation of that day. The latter quantity is computed by using, for each day, equally spaced (15 second interval) intraday observations of the AEX index. In view of our remark with respect to continuous martingales our assumption has an alternative formulation: the financial process associated with the Dutch AEX main stock index is a time-changed Brownian motion. In this interpretation the underlying argument is simple and attractive: the clock for financial time runs fast when trading is brisk and slowly when trading is light, see Mandelbrot (1963) and Clark (1973). In these two papers volatility is regarded as a non-visible or latent process. This is also the case in the paper by Ané and Geman (2000) where high frequency data of two technology stocks are used to show that the assumption of a time-changed Brownian motion can not be rejected. There is an important difference in our approach. Contrary to the above cited papers we use the Time-Change for Martingales Theorem. Here the volatility process is not regarded as latent and the assumption of independence between the time-change and the scaled return process is not required.

Essentially five problems will be discussed in this chapter. The first problem is: Are the daily centered log returns normalized by the daily volatilities, independent standard normal? This question will take up the first half of the chapter. In the second half we take a closer look at the volatility itself. We discuss the daily pattern for volatility and the effect of news on volatility. In the classical Black & Scholes theory volatility is assumed to be constant. In bivariate SDE models the volatility may vary by a factor two over a ten day period. We therefore ask the question: how variable is the volatility? Now that we have the realized volatility, we compare it to the volatility estimates from GARCH and EGARCH models based on the daily return series. Finally, we briefly touch on the fundamental question of the relation between the Brownian motion and the quadratic variation.

The chapter is organized as follows. Section 2.2 contains a leisurely exposition of the ideas underlying our investigation. Section 2.3 gives a detailed description of our data set. In section 2.4 we test the hypothesis of i.i.d. normality. Section 2.5 treats the daily volatility cycle. In section 2.6 we shall investigate the variability of the volatility process. In section 2.7 we test the validity of some of the commonly used models for financial processes and section 2.8 tests for dependence between volatility and standardized returns by making use of the visibility of the volatility process. We close by making some concluding remarks which are inspired by our analysis.
2.2 Time-change and volatility

This section contains an exposition of the basic ideas underlying our investigation of volatility and introduces some notation and terminology. Volatility is a concept which has become basic for pricing options and other derivative products. It is related to the parameter $\sigma$ in the Black & Scholes theory, see Black and Scholes (1973). It is well known that this parameter is not constant in time. If the volatility becomes higher, then the variability of the stock price at the expiration date of the option increases. The effect is as if the expiration date had been shifted further forward in time. Indeed by self-similarity of Brownian motion an increase in the scale parameter $\sigma$ may be interpreted as an increase in speed. Let $B$ denote standard Brownian motion. Then

$$(\sigma B(t), t \geq 0) \overset{d}{=} (B(\sigma^2 t), t \geq 0).$$

So in this model volatility is closely related to time-change: doubling the volatility $\sigma$ will speed up financial time by a factor four.

The term volatility has different meanings. It is used both in the sense of instantaneous volatility, and in the sense of expected future volatility over some time period. In the latter case, which is the common interpretation given to the term by traders, volatility is computed from the price of an option by applying the Black & Scholes formula and solving for $\sigma$. In the present chapter we shall mainly focus on realized volatility which we denote by $\sigma[s,t]$, where $[s,t]$ stands for the relevant time period. Ultimately one is interested in modeling and forecasting volatility. This however is not the aim of the present analysis.

There is wide spread belief that volatility is not visible and that even given the complete sample function of the financial process one needs more information, at least more than is contained in the sample function $x$ of the financial process, in order to obtain the volatility. Sources of such extra information may be trading volume, or the spacing of transaction epochs. We present a different point of view here. We assume that the financial process is a continuous local martingale. The sample function $x$ of the financial process $X$ may then be written in the form $x = \varphi \circ q$ where $q$ is a sample function of the time-change $Q$ and $\varphi$ a sample function of a Brownian motion $B$. From almost any sample function $x$ it is possible to reconstruct both the unknown $\varphi$ and the unknown $q$. Brownian sample functions are so wriggly that it is possible, in contrast to smooth functions, to define an internal clock which is based on counting wriggles and which therefore is not affected by stretching or contracting the horizontal axis. This result holds by the Time-Change for Martingales Theorem, which formally states
Theorem 2.1 (Dambis(1965), Dubins-Schwartz(1965))
If $X$ is a $(\mathcal{F}_t, \mathbb{P})$-continuous local martingale vanishing at 0 and such that $\langle X, X \rangle_\infty = \infty$ and if we set
\[ R(t) = \inf \{ s : \langle X, X \rangle_s > t \}, \]
then, $B(t) = X(R(t))$ is a $(\mathcal{F}_{R(t)})$-Brownian motion and $X(t) = B(\langle X, X \rangle_t)$.

Proof. See for example Revuz and Yor (1994, p. 173).

In our terminology $X$ is the financial process; the time change $Q(t) = \langle X, X \rangle_t$ is the quadratic variation of $X$. Recall, Kallenberg (1997, Proposition 15.18), that the quadratic variation of a continuous semi-martingale $Y$ starting at the origin may be constructed as follows: for $n = 1, 2, \ldots$ let $0 = t_{n0} < \cdots < t_{nn} = c$ be partitions of the interval $[0, c]$ and define step-functions
\[ Q_n = Z_{n1}^2 1_{[t_{n1}, c]} + \cdots + Z_{nn}^2 1_{[t_{nn}, c]} \quad Z_{ni} = Y(t_{ni}) - Y(t_{ni-1}) \]
Then $Q_n(t) \to \langle Y, Y \rangle_t$ for each $t \in [0, c]$, as $n \to \infty$ and $\max_i(t_{ni} - t_{ni-1}) \to 0$. Now write $Y = X + A$ where $X$ is a local martingale and $A$ is continuous, adapted, and of bounded variation. Then $\langle Y, Y \rangle = \langle X, X \rangle$. So the presence of a drift $A$ does not affect the quadratic variation.

Let us now briefly go into some questions surrounding volatility. For the moment assume the time-change $Q = q$ is non-random. If volatility is approximately constant, $\approx \sigma_0$, during the time interval $[s, t]$ then the increase in financial time over $[s, t]$ is approximately linear and
\[ \sigma_0^2 \approx (q(t) - q(s))/(t - s). \]
The daily volatility is the scaling factor needed to transform the daily increments of the financial process, $z(t)$, into standard normal observations. Given a variable volatility function $\sigma(t)$ one may express the increase of financial time $q$ over the interval $[s, t]$ as
\[ q(t) - q(s) = \int_s^t \sigma^2(r)dr. \]

We have mentioned above that for traders volatility is generally defined for time intervals. In that case $\sigma[s, t]$ denotes the 'average' volatility over the time interval $[s, t]$. It is defined by the difference quotient
\[ \sigma^2[s, t] = (q(t) - q(s))/(t - s) = \int_s^t \sigma^2(r)dr/(t - s). \]
Note that it is the squared volatility which is averaged.

For standard Brownian motion it is well known that \( B(t) - B(s) \) is normal with variance \( \sigma^2 = t - s \). So in this framework the standardized observations for day \( d \) are

\[
    u_d = \frac{z_d}{\sigma_d}
\]

and one would like to know whether these observations are realizations of an i.i.d. \( N(0,1) \) sequence. Here \( z_d = x(d) - x(d - 1) - \mu, \ d = 1, 2, \ldots \) stands for the daily centered increments of the financial process \( x \) with drift \( \mu \) and \( \sigma_d \) for the daily volatility. If \( Q \) is random then normality holds if \( Q \) and \( B \) are independent. It also holds if there is a time lag in the influence of the price on the volatility, for instance as in the GARCH model where the volatility during day \( d + 1 \) is allowed to depend on the behaviour of the financial process during the previous \( d \) days. However, in general if there is dependence between \( B \) and \( Q \) then normality of the standardized daily returns need not hold.

As a null hypothesis in this chapter we have:

\[ H_0: \text{The standardized daily increments } u_d \text{ in (2.1) of the log price process of the AEX index corrected for drift are independent observations from a standard normal distribution.} \]

### 2.3 Description of the data

Here we give a description of the data set that was used in the analysis presented in this chapter. As data set we use the (Dutch) Amsterdam exchanges (AEX) index for the period 1st of May, 1996 until the 29th of September, 2000. During the period May 1996 until July 1999 the trading hours were 9.30 until 16.30, during the period August 1999 until June 2000 the trading hours were 9.00 until 16.30 and from 9.00 until 17.00 during the period thereafter, where all times are GMT+1 hour.

The AEX cash index is a composite of the 25 most traded stocks in the previous year (a new year starts on the first trading day in March) that are listed on the Amsterdam stock exchange. The weight of a stock in the index is in turn determined by its market capitalization, with a maximum weight of 10%, at the outset of the new year. On the AEX index European options and future contracts are traded. At each third (non-holiday) Friday of the month a series of options expires at 16.00. If the third Friday is a holiday then the series expires on the last trading day before the holiday. To prevent price manipulation the value at 16.00 is determined by taking a weighted
average of the index over the previous 30 minutes. The expiration days are given a special treatment, see below.

The data set was obtained by reading the value of the AEX index each 15-th second. In general, transactions only concern one of the underlying stocks of the index, so effectively there is only a partial update. If none of the underlying stocks has been traded in a 15 second interval then the AEX does not return a value. So, a priori, the number of 15 second intervals gives some information on the trading intensity of a particular day. On average we have a value of the AEX index each 15-th second. In total we have 1117 trading days at our disposal. As an illustration of our data set the sample paths of the AEX index of four successive days are displayed in figure 2.1.

![Sample paths of the AEX index on four successive days. Top left: 2/2/1998; top right: 3/2/1998; bottom left: 4/2/1998; and bottom right: 5/2/1998.](image)

Figure 2.1. The sample paths of the AEX index on four successive days. Top left: 2/2/1998; top right: 3/2/1998; bottom left: 4/2/1998; and bottom right: 5/2/1998.

Denote the daily log returns during this period by \( r_d = \log(aex_d/aex_{d-1}) \), where \( d = 1, \ldots, 1117 \) and \( aex_d \) is the closing price at date \( d \). One may write \( r_d = r^*_d + r^0_d \), where \( r^*_d \) stands for the intra-day returns (opening to closing, see figure 2.2 for the time series and its histogram) and \( r^0_d \) for the overnight returns (closing to opening).

The total real return of the AEX index, for the period we consider, equals
163%. On a daily basis this amounts to an average daily log return ($\bar{r}$) of 0.0864% which may be decomposed into an average intra-day return ($\bar{r}^*$) of 0.0119% and an average overnight return ($\bar{r}^o$) of 0.0745%. The standard deviations of these returns are $\sigma_r = 1.38\%$, $\sigma_r^* = 1.19\%$ and $\sigma_r^o = 0.712\%$ respectively.

The data are filtered according to the following rules: as an opening price for day $d$ we use the second observation, as a closing price the second last tick. If $d$ is an expiration day we use as a closing price the value of the AEX index at 15.25. Moreover, if during a trading day the trading was halted (i.e. due to computer failure etc.) we disregard the returns between the last tick before the trading was stopped and the first tick after the trading has resumed. We correct the intra-day returns and volatilities of these days by the following procedure: denote the total time of a trading day by $T$ and the total time of missing values of a trading day by $T_m$. Both the returns and volatility of such a trading day are multiplied by $\sqrt{T/(T - T_m)}$. Finally, we omit intra-day observations for four days of the whole data set since the recorded data 'look' suspicious. The dates of these days are: 1/6/97, 23/6/97, 3/8/97 and 12/01/98. The first day is skipped because the time points were incorrectly recorded. The remaining three are skipped because the AEX was only measured over small time periods of the trading day.

2.4 Testing on i.i.d. normality

In this section we shall test whether the standardized daily return series in (2.1) is a sequence of realizations of i.i.d. standard normal random variables.
We apply six tests. Two for normality, two for the tails, and two for independence.

Let us start by discussing the use of 15 second interval data. We take an idealist point of view. At each time point the AEX cash index has a value. The value is a real number. It changes continuously in time. It may be observed to a certain degree of accuracy by making a transaction. Transactions update the value. In general, transactions only concern one of the underlying stocks of the index, so effectively there is only a partial update. Several studies, see e.g. Bai, Russel and Tiao (2001), have shown that the choice of the sampling frequency is not trivial. In fact, a high sampling frequency like 15 seconds may introduce bias into the estimate of the quadratic variation.

Essentially two sources of bias have to be distinguished. The first source is related to the bid-ask quotes of the underlying stocks. This effect is referred to as the bid-ask effect and induces a 'spurious' negative autocorrelation into the high frequency return process. This translates into positive bias of the estimator of the quadratic variation. The other source is caused by non-synchronous trading of the underlying stocks in the index. As pointed out stocks are not updated simultaneously but sequentially. This causes a 'spurious' positive autocorrelation into the high frequency return process. This results in a negative bias in the estimator of the quadratic variation. For more details on these two sources of bias, see Dacorogna, Gençay, Müller, Olsen and Pictetna (2001).

Rather than applying some cleaning operations to our data set we have opted for treating the raw data in order to keep close to the original hypothesis that the log increments of the AEX over the period 1996-2000, corrected for a linear drift, are martingale increments. For the 15 second AEX data the graph of the quadratic variation $Q$ hardly changes if one restricts the observations to 30 second, 1 minute or 2 minute intervals. All the results presented in this section also hold for these lower frequencies. Note that for 1 minute intervals the statistical error already increases by a factor two.

The two, above discussed, sources of bias in high frequency data possibly cancel each other out for the AEX. In table 2.1 we present some descriptive statistics of the non-standardized 15 second return series. As can be observed from this table there is still a small but statistically significant (at a level $\alpha = 0.01$) negative autocorrelation between successive observations. This may indicate that the bid-ask effect dominates the effect due to non-synchronous trading. The negative autocorrelation yields a positive bias in the estimate of the quadratic variation. We return to this point later. Autocorrelations on a lower frequency, as observed by LeBaron (1992) for the indices S&P 500 and DJIA, were not detected in the daily AEX data.
Table 2.1. 15 second return statistics

<table>
<thead>
<tr>
<th>mean</th>
<th>st.dev.</th>
<th>skew</th>
<th>kurt</th>
<th>corr(r(d,j),r(d,j-1))</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.23 \times 10^{-8}</td>
<td>2.61 \times 10^{-4}</td>
<td>-0.392</td>
<td>34.5</td>
<td>-0.0808</td>
</tr>
</tbody>
</table>

We next formally introduce the notation that we will use throughout the applied part (remainder) of this chapter. We use capital letters for random variables and lower case letters for the corresponding realizations. The financial process $X$ is a construct. We have deleted the first and last AEX observations of each day in order to avoid boundary effects. This leaves us with $j_d + 1 \approx 1700$ observations $aex(d,0), \ldots, aex(d,j_d)$ on the $d$th day, and hence with $j_d$ intervals. The 15-second increment $z(d,j)$ of the financial process $X$ over the $j$th interval of day $d$ is defined as

$$z(d,j) = \log(aex(d,j)/aex(d,j-1)) - \bar{\mu}/j_d$$

where $\bar{\mu}$ is the average daily log increment, see below. Now define

$$z_d = z(d,1) + \cdots + z(d,j_d) \quad \sigma_d^2 = z^2(d,1) + \cdots + z^2(d,j_d)$$

as the daily increase of the financial process and the estimate of the daily increment of financial time. So the sample function $x$ of the financial process $X$ and the estimate $\hat{q}$ of the time-change $Q$ associated with the sample function $x$ at time $t = d - 1 + j/j_d$ are given by the sums

$$x(t) = z_1 + \cdots + z_{d-1} + z(d,1) + \cdots + z(d,j),$$

$$\hat{q}(t) = \sigma_1^2 + \cdots + \sigma_{d-1}^2 + z^2(d,1) + \cdots + z^2(d,j).$$

The price at time $t$ of the AEX index is

$$AEX(t) = e^{x(t)-Q(t)/2}e^{Z^0(t)}e^\alpha(t) \quad t \geq 0$$

where $Z^0$ denotes the partial sum of log increments of the AEX during non trading hours, $e^{\alpha(t)}$ is the true drift term of the index during trading hours and $e^{X-Q/2} = e^{R_0-Q/2}$ is a geometric Brownian motion at the financial time $Q$. So we may write

$$X(t) = \log AEX(t) - Z^0(t) + Q(t)/2 - \alpha(t) = \log AEX(t) - Z^0(t) - \mu(t).$$
Since \( \alpha(t) \) and \( \mu(t) \) are not known we replace \( \mu(t) \) by \( t\bar{\mu} \), where \( \bar{\mu} = 0.000206 \) is the average of \( \log(aex(d,j_d)/aex(d,0)) \) over the 1117 days \( d \) of our data set, see section 2.3. In this chapter we are only interested in increments and quadratic increments over periods of a day or less. Hence the precise form of the drift function is of less importance. The average daily volatility is \( \bar{\sigma} = 0.0119 \). So \( \bar{\mu} << \bar{\sigma} \) and the influence of \( \mu \) on our results is negligible.

In order to avoid pedantic notation we use the same symbol \( X \) for the drift corrected process \( \log AEX(t) - \mu(t) \) (in the theoretical part) and the linearly corrected process \( \log AEX(t) - t\bar{\mu} \) (in the more applied part).

The realizations \( u_d \) in (2.1) for the days \( d = 1, \ldots, 1117 \) are given by

\[
u_d = \frac{x(d) - x(d-1)}{\sqrt{q[d]}} = \frac{z_d}{\sigma_d}.
\]

Here \( q[d] \) is the increase of financial time over day \( d, d = 1, \ldots, 1117 \).

Before discussing the six tests in detail we present some graphical characteristics of the return series \( u_d, d = 1, \ldots, 1117 \). Figure 2.3 displays the standardized return series \( u_d \) and the corresponding histogram.

Figure 2.3. Left the time series of the normalized daily returns \( u_d \) and right its histogram. The dotted line is the corresponding normal density function.

The series look similar to a simulated i.i.d. standard normal sequence. Figure 2.4 depicts the 1116 points \( (\Phi(U_d), \Phi(U_{d+1})) \). Here \( \Phi \) denotes the standard normal distribution function. Note that the random variable \( \Phi(U) \) is uniformly distributed on the interval \([0,1]\) if \( U \) has a standard normal distribution. If the variables \( U_d \) are also independent then the observations \( (\Phi(U_d), \Phi(U_{d+1})) \) will be uniformly distributed over the unit square.

If we look at the standard deviation of the normalized increments \( u_d \) we find that it equals 0.983 instead on 1. Although, the difference (0.017) is
too small to be significant, it might be due to the dominant bid-ask effect as pointed out at the outset of this section.

We shall now discuss the six tests for the null hypothesis $H_0$, which was formulated at the end of Section 2.2.

(i) The Kolmogorov-Smirnov test.
The first test we use is the Kolmogorov-Smirnov test. See Shorack and Wellner (1986) for a detailed discussion. This test utilizes the difference between the observed empirical distribution function $F_n$ and the distribution under the null hypothesis $F$. Under the null hypothesis this difference (blown up by $\sqrt{n}$) tends to a Brownian bridge. The test is based on the supremum of the absolute values of this Brownian bridge: under the null hypothesis the supremum should be smaller than some critical value. We test on normality both with unknown mean and variance and with fixed mean ($=0$) and fixed variance ($=1$). The outcomes are presented in table 2.2. For comparison the same test has been applied to the intra-day returns $r_d^*$ of the AEX. As can be observed from table 2.2 (standard) normality is not rejected for the standardized return series. It is rejected for the unscaled AEX return series $r_d^*$.

(ii) The Jarque-Bera test.
The Jarque-Bera (1980) test uses the skewness, $s$, and the kurtosis, $k$, from
the empirical sequence and checks whether they can be distinguished from the skewness 0 and the kurtosis 3 of a standard normal distribution. This test is particularly useful for financial time series where the empirical distributions may be skewed and fat tailed. Under the null hypothesis the statistic \( \frac{3}{2}(s^2 + \frac{1}{4}(k-3)^2) \) is asymptotically \( \chi^2 \) distributed with two degrees of freedom as the sample size tends to infinity. As can be observed from table 2.2 normality is not rejected at a significance level of 10%.

### Table 2.2. Testing for normality

The test statistics of the Kolmogorov-Smirnov (KS) and Jarque-Bera (JB) tests applied to the intra-day log returns \( r^*_d \) and the standardized returns \( u_d \) respectively. The p-values are presented in parenthesis. The p-values of the Kolmogorov-Smirnov statistic for unknown mean and variance were obtained from Lilliefors (1967). Under the hypothesis of normality the Jarque-Bera statistic is asymptotically \( \chi^2 \) distributed. The values of JB and KS were obtained from Eviews and S-Plus respectively.

<table>
<thead>
<tr>
<th>Series</th>
<th>KS:N(0,1)</th>
<th>KS:N(( \mu, \sigma^2 ))</th>
<th>JB</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r^*_d )</td>
<td>-</td>
<td>0.0588</td>
<td>406</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>( u_d )</td>
<td>0.0357</td>
<td>0.0129</td>
<td>0.527</td>
</tr>
<tr>
<td></td>
<td>(0.116)</td>
<td>(0.996)</td>
<td>(0.768)</td>
</tr>
</tbody>
</table>

(iii) The binomial test.

For financial data extreme values are important. We perform three simple binomial tests to check whether the distribution of \( U_d \) has heavy tails by counting the number of data points exceeding the values of 2, 3, and 3.5 standard deviations. Let \( p_c = P\{|U| > c\} \) where \( U \) has a standard normal distribution. For \( n \) independent observations \( \{ U \} \) the number \( N_c \) of observations exceeding the level \( c \) has a binomial\((n, p_c)\) distribution with mean \( \mu_c = np_c \) and variance \( \sigma^2_c = np_c(1 - p_c) \). The observed number is \( n_c \). In our situation \( n = 1117 \). The results are presented in table 2.3. The probabilities are neither close to 0 nor close to 1. The hypothesis that the tails behave like those of a standard normal distribution is not rejected.

(iv) A test for the maximum.

The normal distribution \( \Phi \) lies in the domain of attraction of the double exponential extreme value distribution \( \Lambda(x) = \exp(-e^{-x}) \) for maxima, and so does \( \Psi(x) = (2\Phi(x) - 1) \lor 0 \), the distribution of \( |U| \) where \( U \) is standard normal. The maximum \( M_n \) of \( |U_1|, \ldots, |U_n| \) for independent standard normal variables \( U_i \) satisfies the asymptotic relation \( M_n = b_n + a_n Z_n \) where \( Z_n \)
2.4 Testing on i.i.d. normality

Table 2.3. Observations in the tails
The number of observations |u_i| exceeding 2.3 and 3.5 standard deviations together with the mean, the standard deviation and the corresponding probabilities.

<table>
<thead>
<tr>
<th>c</th>
<th>p_c</th>
<th>n_c</th>
<th>μ_c</th>
<th>σ_c</th>
<th>P{N_c ≤ n_c}</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>0.0456</td>
<td>47</td>
<td>50.88</td>
<td>6.97</td>
<td>0.320</td>
</tr>
<tr>
<td>3.0</td>
<td>0.00270</td>
<td>2</td>
<td>3.02</td>
<td>1.73</td>
<td>0.419</td>
</tr>
<tr>
<td>3.5</td>
<td>4.6 · 10⁻⁴</td>
<td>0</td>
<td>0.51</td>
<td>0.72</td>
<td>0.598</td>
</tr>
</tbody>
</table>

converges in law to a variable Z with distribution function Λ. Since \( M_n \) has distribution function \( \Psi^n \) we may and do compute \( a_n \) and \( b_n \) so that \( \Psi^n(b_n + a_n u) = \Lambda(u) \) for \( u = 0, 1 \), rather than using the asymptotic expression in Embrechts, Klüppelberg and Mikosch (1997). The 90% confidence interval for \( Z \) will yield an asymptotic 90% confidence interval \([\bar{c}_0, \bar{c}_1]\) for \( M_n \). An exact 90% confidence interval \([c_0, c_1]\) may be constructed by using \( \Psi^n(x) \) rather than \( \Lambda((x - b)/a) \). Set \( \psi(x) = -n \log \Psi(x) \). Then one obtains \( b_n, a_n, \bar{c}_0, \bar{c}_1 \) and \( c_0, c_1 \) by solving \( \psi(b_n) = 1, \psi(b_n + a_n) = 1/e, \Lambda((\bar{c}_0 - b_n)/a_n) = 0.05, \Lambda((\bar{c}_1 - b_n)/a_n) = 0.95 \) and \( \psi(c_0) = \log(20), \psi(c_1) = \log(20/19) \). Taking \( n = 1117 \) we find

\[
a = 0.2694, b = 3.322, \bar{c}_0 = 3.026, \bar{c}_1 = 4.122, c_0 = 3.002, c_1 = 4.075.
\]

The observed value 3.168 lies in the interval \([c_0, c_1]\). The null hypothesis is not rejected at the significance level of 10%.

(v) The BDS test.
Next, we check whether the standardized return series are i.i.d. There are several ways to do this. We have chosen for the BDS test, introduced in the financial literature by Brock, Dechert and Scheinkman (BDS) (1987). This test has often been (mis)used to investigate the presence of deterministic (chaotic) structures in empirical time series. See Takens (1993) for an enlightening discussion on this matter. The BDS statistic has power against a wide variety of departures from i.i.d. processes.

The BDS, which was originally derived from the correlation integral, is based on the following property of an i.i.d. sequence. Let \( Y_d \) be a reconstruction vector given by \( Y_d := (U_{d-m+1}, \ldots, U_d)' \), where \( m \in \mathbb{N}^+ \) is called the embedding dimension. Let \( Y_d \) and \( Y_{d'} \) be two arbitrary vectors and \( U_s \) and \( U_{s'} \) two arbitrary observations. If \( (Y_d)_{d=1}^{1117} \) is an i.i.d. sequence then

\[
P(|Y_d - Y_{d'}| < \epsilon) = (P(|U_s - U_{s'}| < \epsilon))^m,
\]

where \( |\cdot| \) denotes the supnorm, \( (|y| = \sup_{i=1}^{m} |y_i|) \) in \( \mathbb{R}^m \) and \( \epsilon > 0 \). The BDS test utilizes the difference of estimates of these two probabilities. The
The realized volatility of the Main Dutch ...

Table 2.4. The BDS test

The BDS statistics and the corresponding p-values for the series \( u_d \). For embedding dimension \( m = 2, \ldots, 6 \) each column contains the BDS statistic (bds) and the p-value (p) for different values of \( \epsilon \). The p-values are obtained by reshuffling \( u_d \) 2000 times, computing the BDS statistic for these new series and then counting the number of absolute values exceeding the absolute value of bds of \( u_d \).

<table>
<thead>
<tr>
<th>( \epsilon )</th>
<th>m=2</th>
<th>m=3</th>
<th>m=4</th>
<th>m=5</th>
<th>m=6</th>
</tr>
</thead>
<tbody>
<tr>
<td>bds 0.25</td>
<td>-4.13 \cdot 10^{-5}</td>
<td>3.10 \cdot 10^{-5}</td>
<td>1.92 \cdot 10^{-5}</td>
<td>1.01 \cdot 10^{-5}</td>
<td>4.75 \cdot 10^{-6}</td>
</tr>
<tr>
<td>p</td>
<td>0.976</td>
<td>0.912</td>
<td>0.721</td>
<td>0.461</td>
<td>0.353</td>
</tr>
<tr>
<td>bds 0.5</td>
<td>-7.31 \cdot 10^{-5}</td>
<td>3.20 \cdot 10^{-4}</td>
<td>2.11 \cdot 10^{-4}</td>
<td>6.80 \cdot 10^{-5}</td>
<td>9.60 \cdot 10^{-6}</td>
</tr>
<tr>
<td>p</td>
<td>0.987</td>
<td>0.835</td>
<td>0.709</td>
<td>0.732</td>
<td>0.91</td>
</tr>
<tr>
<td>bds 0.75</td>
<td>2.92 \cdot 10^{-4}</td>
<td>1.01 \cdot 10^{-3}</td>
<td>1.01 \cdot 10^{-3}</td>
<td>8.47 \cdot 10^{-4}</td>
<td>6.06 \cdot 10^{-4}</td>
</tr>
<tr>
<td>p</td>
<td>0.972</td>
<td>0.834</td>
<td>0.688</td>
<td>0.507</td>
<td>0.351</td>
</tr>
<tr>
<td>bds 1</td>
<td>1.87 \cdot 10^{-4}</td>
<td>1.94 \cdot 10^{-3}</td>
<td>2.46 \cdot 10^{-3}</td>
<td>2.46 \cdot 10^{-3}</td>
<td>2.31 \cdot 10^{-3}</td>
</tr>
<tr>
<td>p</td>
<td>0.986</td>
<td>0.834</td>
<td>0.707</td>
<td>0.531</td>
<td>0.374</td>
</tr>
</tbody>
</table>

estimates \( \hat{P}(|Y_d - Y_d'| < \epsilon) \) and \( \hat{P}(|U_d - U_d'| < \epsilon) \) are computed by counting the number of vectors \( y_d \) and the number of observations \( u_d \) that lie within an \( \epsilon \) distance of each other. Under the null hypothesis, this difference is approximately normally distributed. We have calculated the BDS statistic for different embedding dimensions \( m \) and distances \( \epsilon \), see table 2.4. To determine the p-values we performed the following (bootstrap) power enhancing procedure. For each \( m \) and \( \epsilon \) we have reshuffled the sequence \( (u_d)_{d=1}^{17} \) 2000 times and calculated for each reshuffled sequence the BDS statistic. Then counting the number of absolute values exceeding the absolute value of bds of \( u_d \) results in the corresponding p-value. The outcome may be found in table 2.4. Clearly the test does not reject the hypothesis of i.i.d. increments.

(vi) (Partial) autocorrelation functions.

As a final check for i.i.d. we present the (partial) autocorrelation function of the standardized returns \( u_d \). We expect that the coefficients of these functions are close to 0. In figure 2.5 we display these functions for lags of 1, \ldots, 50. Note that four of the fifty autocorrelation coefficients are just outside the 95% confidence interval as do four of the fifty partial autocorrelation coefficients. This is an indication that no ARMA\((p,q)\) model fits the data. Moreover, the Akaike information criterion (see for example Brockwell and Davis (1991)) gives the orders \( p = 0 \) and \( q = 0 \) as the optimal order for an ARMA\((p,q)\) model.

To see if the volatility persistence has been removed from the original
2.5 The daily volatility cycle

Figure 2.5. Left the autocorrelation function of the standardized returns $u_d$ for lags 1...50, and right the partial autocorrelation function. In both figures the dashed lines represent the limits of the 95% confidence interval. The estimates of these functions were obtained from S-Plus.

return series we have computed both the (partial) autocorrelation function of the squared centered returns $z_d^2$ and the squared standardized returns $u_d^2$, see figure 2.6. As can be seen from this figure almost all autocorrelation present in the series $z_d^2$ has been removed.

Based on the statistical analysis described above we conclude that we can not reject the null hypothesis that the standardized return series $U_d$ are i.i.d. standard normal.

2.5 The daily volatility cycle

In the previous section we have shown that the quadratic variation (and the drift) of the log price process accounts for all the structure in the daily returns of the AEX index. Scaling the daily log returns by the estimated volatility for the day leads to a standard normal distribution. One is of course interested in the properties of the volatility associated with the log price process. The availability of the high frequency data gives us the opportunity to analyze volatility even on a time scale less than one trading day. In this section we address the question whether the volatility process displays any regularity on time intervals of one day. This question has been motivated by the work of Andersen and Bollerslev (1997,1998) who analyzed high frequency (5 minutes) data of the DM-$\$ exchange rate. In Andersen and Bollerslev (1997) long-run volatility dynamics is analyzed by annihilating short term intra-day patterns in volatility. From this analysis the authors find indications that macro economic announcements influence the absolute return values of the
The realized volatility of the Main Dutch...

![Figure 2.6](image)

**Figure 2.6.** Top the (partial) autocorrelation function of the squared centered returns $z_d^2$ for lags 1...50, and bottom the (partial) autocorrelation of the squared standardized returns $u_d^2$ for lags 1...50, and in (d) the partial autocorrelation function. In these figures the dashed lines represent the limits of the 95% confidence interval. The estimates of these functions were obtained from S-Plus.

exchange rate. In Andersen and Bollerslev (1998) this intra-day pattern is investigated in more detail. From this detailed analysis the authors conclude that calendar effects as well as different macro-economic indicators do indeed influence the absolute returns of DM-$ exchange rate.

Along the lines of Andersen and Bollerslev (1998, 1997) we partitioned each day into 420 one minute intervals and computed for each interval the average volatility (square root of the average squared volatility) based on the time-change $q$ for the period May 1996 until 1st July 1999, see figure 2.7. We have chosen this period of 818 trading days because opening and closing hours during this period were constant, see section 2.3 for details.

As may be observed from figure 2.7 the average volatility starts the day high then decreases and stays low until 14.30. At this time point something unexpected occurs. The one minute average volatility spikes. The spike is asymmetric. It achieves its maximum in less than a minute and it then takes an hour to decrease to its original level. In fact, a careful look at
2.5 The daily volatility cycle

![Graphs showing daily volatility cycle](image)

**Figure 2.7.** Left the average of the one minute volatility for the period May, 1996 until July, 1999 from 9.30 until 16.30 and in right the average one minute volatility of this period from 14.00 until 16.00.

Figure 2.7 (right) shows that the decrease of volatility after the impact of macro-economic news may be decomposed into two periods. The first period starts at the point of the big spike and ends five minutes later. The second period starts at 14.35 and ends at 15.30. In this 55 minutes time interval volatility is decreasing slowly. For both periods the increase in financial time due to the news is roughly the same. The time point 14:30 coincides with the time point of the release of important macro-economic data, such as GNP and inflation rate data, in the U.S. As described above the influence of news seems to have a short but clear impact on the volatility, cf. Anderson and Bollerslev (1998). That we can single out the effect of important U.S. macro-economic news on the volatility process is a special property of our data set. At 14:30 the U.S. markets are still closed. Also observe from figure 2.7 that at 15:30 volatility jumps to a higher level. This jump coincides with the opening of US markets. From this time on the one minute average volatility increases until the close, at 16.30, of the Dutch stock market. A careful look reveals a second smaller spike at 16.00. This spike may also be associated to the release of economic news, such as the National Association of Purchasing Managers index (NAPM) and forecasts of the University of Michigan.

The graph of the minute by minute averaged volatility in figure 2.7 is more wriggly than one would expect for an average of $n_0 = 818$ observations. Partly this is due to statistical error: ultimately our results are based on a finite number of observations. The time-change $\hat{q}$, even with daily 15 second observations, contains a certain error which is magnified by taking difference quotients in calculating the volatilities. To see if the volatility process displays a daily returning pattern we analyze the average daily volatility cycle.
Such a pattern could be of interest to many practitioners. Although we have clearly seen the effects of news on the process we shall find no evidence for the presence of a regular daily cycle. We shall test the hypothesis that the daily volatility follows the average pattern of figure 2.7 except for a random scale factor which may differ from day to day.

\[ H_0: \text{The daily volatility follows an average daily cycle except for a random scale factor.} \]

To test this hypothesis we divide the seven hour trading day into two equal parts, the morning from 9.30 till 13.00 and the afternoon from 13.00 till 16.30. We compute the increase of financial time over these two parts for each of the \( n_0 = 818 \) days and denote the increments by \( q[m_d] \) and \( q[a_d] \) respectively for \( d = 1, \ldots, 818 \). The square brackets indicate that we are working with increments. Under the null hypothesis the quotient \( q[m_d]/q[a_d] \) is constant. The average of the quotients (over the \( n_0 \) days) is a good approximation of this constant. Define the variables

\[ Y_d = \log(\hat{Q}[m_d]/\hat{Q}[a_d]) \]

and the statistic

\[ sd_Y^2 = \sum (Y_d - \bar{Y})^2/(n_0 - 1). \]

Under the null hypothesis the variance of \( Y_d \) is equal to the variance of the statistical error of \( Y_d \). Since we presume that the quadratic variation over a period of one day has a fixed density, given in figure 2.7, we have a statistical model and we can compute the coefficient of variation (cv)

\[ \text{cv}(\hat{Q}[I]) = \sqrt{\text{var}(\hat{Q}[I]/Q[I])} \quad (2.3) \]

of the estimator \( \hat{Q}[I] \) of the increase \( Q[I] \) of the quadratic variation over the interval \( I \). The increments of \( \hat{Q} \) are independent squared normal random variables whose variance is determined by the density. Under the null hypothesis the cv for \( \hat{Q}[m_d] \) and \( \hat{Q}[a_d] \) has the form \( cv = \sqrt{2c/n} \) where \( 2 = \text{var}(U^2/(E(U^2))) \) for a centered Gaussian variable \( U \), \( n = 1680/2 \) the number of successive intra-day observations, and \( c = c_m \) or \( c_a \) has the form

\[ c = n \sum w_i^2/(\sum w_i)^2 \]

where \( w_i \) is the average of the squared volatility during the \( i \)th minute, which may be read off from figure 2.7. The factor \( c \) equals \( c_m = 2.633 \) for the morning and \( c_a = 1.175 \) for the afternoon. Let \( se_Y \) denote the statistical error in \( Y_d \). Under the null hypothesis

\[ se_Y^2 = 2(c_m + c_a)/840 = 0.00906. \]
2.6 Variability of volatility

The observed standard deviation $sd_Y$ of the $n_0$ observations equals 0.183 (a factor 40!) and we reject the hypothesis that the observed variation is caused by the statistical error in the observation of the time-change. Hence, we have to reject the null hypothesis that figure 2.7 describes the volatility during the day up to a scaling constant.

To give an impression of the changes in volatility in the course of a day we display, in figure 2.8, the intra-day volatility measured over the past 30 minutes on four successive days: 2/2/1998 until 5/2/1998. The corresponding sample paths of the AEX index on these four days may be found in figure 2.1.

![Graphs](image)

**Figure 2.8.** The half-hour volatility, for the thirty minute time interval preceding time $t$, $10.00 \leq t \leq 16.30$, on four successive days: 2/2/1998; 3/2/1998; 4/2/1998 and 5/2/1998.

2.6 Variability of volatility

In the previous section we have presented one application to our method of making the volatility process visible: changes in the one minute volatility at different time points during a day can unambiguously be attributed to concrete events. In the present section we are interested in obtaining some
knowledge about the basic properties of the time-change. The main objective of this section is to investigate the variability of the volatility. Our original assumption was that volatility is a continuous process. In the bivariate SDE the volatility component functions as a parameter which has to monitor the variability of the sample function of the financial process. So one expects it to vary less than the financial process itself.

We begin with two remarks on the behaviour of the sample function \( \hat{q} \) plotted in figure 2.9.

**Figure 2.9.** Left the time-change calculated by using all the observations and right a piecewise linear approximation to the graph of the time-change.

**Figure 2.10.** Left the average daily log volatility and right the histogram of the changes in the log volatility.

After 1117 days the time-change \( \hat{q} \) sums up to 0.146, see figure 2.9. So roughly speaking, one unit of financial time corresponds to thirty years in calendar time. One should not take this unit of thirty years too seriously.
2.6 Variability of volatility

however. The AEX index is an average of twenty-five stocks on the Amsterdam stock exchange and averaging decreases volatility. Preliminary analysis of high frequency data of individual stocks shows that the financial clock of individual stocks runs faster than that of the composite index. This can also be understood in terms of the modern portfolio theory: adding stocks of companies that operate in different sectors of the economy reduces the volatility of the portfolio.

The second observation is that a visual inspection of the left graph of the time-change in figure 2.9 shows that different regimes may be distinguished. There are regimes when financial time runs fast (steep parts of the function) and regimes when financial time runs slowly (flatter parts of the function). As we may see from figure 2.10 volatility is obviously not constant over these regimes. However, it may fluctuate around a central value which changes (abruptly) from time to time. The idea of abruptly changing parameters was already investigated in the paper by Tyssedal and Tjøstheim (1988). See also Kokoszka and Leipus (2000). We distinguish nineteen regimes in the four and a half year period of our observations. Figure 2.9 (right) shows the piecewise linear approximation to $\hat{q}$ based on this subdivision into nineteen intervals.

We now return to the question how variable the volatility process is. As mentioned earlier in section 2.2 there is a simple relation between the daily volatilities and the daily increments of financial time. Let $\sigma_d$ denote the volatility over the $d$th day. Then $\sigma_d^2$ is the increase of financial time over the $d$th trading day. (In finance the day is the unit of physical time). The average daily volatility is approximately 0.01. We are not so much interested in absolute volatilities but rather in the relative change in the volatility on successive days. Hence we introduce the logarithmic increments

$$\delta_d = \log(\sigma_d/\sigma_{d-1}) = \log(\hat{q}[d]/\hat{q}[d-1])/2, \quad d = 2, \ldots, 1117,$$

where $\hat{q}[d]$ denotes the increase in the (estimated) financial time over day $d$. (Note that the quantity $\delta_d$ does not depend on the unit of one day for physical time.) Figure 2.10 shows the daily logarithmic volatilities $\log(\sigma_d)$ and the histogram of the logarithmic increments $\delta_d$. The histogram looks Gaussian, but it has extensive tails. The standard deviation of $\delta_d$ over the 1117 observations is $sd(\delta) = 0.22$. A difference of 0.22 in the logarithm scale is large. It corresponds to an increase by 25% or a decrease by 20% in the daily volatility.

In order to understand how large the variability in the volatility is, we compare the volatility process to the price process of the AEX. For the standard deviation of the logarithmic increments $z$ we find $sd(z) = 0.012$. This
quantity is much smaller than the standard deviation \( sd(\delta) = 0.22 \) of the logarithmic increments of the volatility. We conclude that the logarithm of the volatility process is approximately twenty times more variable than the logarithm of the AEX index price process. The idea of volatility as a parameter to monitor the financial process \( X \) will have to be reevaluated.

In an attempt to establish continuity of the volatility we have deleted all observations outside the four hour time interval 10.30-14.30 from May 1996 until July 1999. We have selected this time interval and period for the following reasons: (i) absence of opening effects, (ii) there are no U.S. macro-economic news releases in this time interval, (iii) during this period the Dutch stock exchange had 7 trading hours instead of 7.5 and 8 trading hours respectively in the periods thereafter, see section 2.3 for details. So this leaves us with a homogeneous data set, see figure 2.7. We have divided this four hour period into disjoint periods of two, 4/3, one, 2/3 and half hours. For \( m = 2, 3, 4, 6, 8 \) we define \( \sigma_m(d,j) \) as the volatility over the \( j \)th period of \( 4/m \) hours on day \( d \). We are again interested in the change of the logarithmic volatility between two successive periods. So we define

\[
\delta_m(d,j) = \log(\sigma_m(d,j)/\sigma_m(d,j - 1)) = \log(\hat{q}_d[j/m]/\hat{q}_d[(j - 1)/m])/2
\]

for \( j = 2, \ldots, m \), where \( \hat{q}_d[k/m] \) is the increment of financial time over the \( k \)th \( 4/m \)-hour period after 10.30 of the \( d \)th day. Similarly we define \( z_m(d,j) \) as the increment of the financial sample function \( x \) over the \( j \)th period of length \( 4/m \) hours of the \( d \)th day after 10.30 for \( j = 1, \ldots, m \). The sample standard deviation of \( \delta_m(d,j) \) and \( z_m(d,j) \) will be denoted by \( sd_m(\delta) \) and \( sd_m(z) \) respectively.

The squared standard deviations \( sd_m^2(\delta) \) and \( sd_m^2(z) \) are set out in figure 2.11. For the relative changes in the volatility \( \delta_m \) we have also computed the coefficient of variation \( cv_m(\delta) \) cf. equation (2.3) with

\[
(cv_m^2(\delta))^2 = 1/4 \left( cv^2(\hat{Q}[j/m]) + cv^2(\hat{Q}[(j - 1)/m]) \right).
\]

Under the assumption that the volatility is random but constant on these short time intervals we find, as in the previous section, \( cv_m^2(\delta) = 2 \cdot (1/4) \cdot 2/(960/m) = (m/960) \) since the four hour period contains 960 15-second intervals.

Figure 2.11 (right) shows that the sample variance \( sd_m^2(z) \) of the increments \( z_m(d,j) \) of the financial process \( X \) decreases substantially as the increments are taken over shorter time intervals. This is what one expects if \( X \) is a time-changed Brownian motion. For the volatility process the sample variance \( sd_m^2(\delta) \) increases as the increments are taken over shorter time intervals, even if we subtract the squared statistical error \( se_m^2(\delta) \). On an hourly
2.7 Confrontation with some standard models

In this section we confront standard volatility models with the empirically observed volatility $\sigma_d$. We focus on the GARCH and EGARCH models which were introduced by Bollerslev (1986) and Nelson (1991) respectively. In these models daily return data are used to estimate the parameters, and to visualize the volatility series. The estimates of the parameters are obtained by maximizing the log likelihood function.

The GARCH(1,1) model has the form:

\[ Z_d = \mu + h_d V_d \quad V_d \sim N(0,1), \]
\[ h_d^2 = \alpha_0 + \alpha_1 (Z_{d-1})^2 + \beta_1 h_{d-1}^2. \]  

(2.4)

In table 2.5 we present the maximum likelihood estimates of the parameters of equation (2.4). Table 2.6 contains the Ljung-Box Q-statistics of the

Figure 2.11. On the vertical axis: left the sample variance $s_\delta^2_m(t,j)$ associated with the increments $\delta_m(t,j)$ and beneath this line the sample variance corrected for the statistical error $s_\delta^2_m(\delta) - cv^2_m(\delta)$, and right the sample variance $s_d^2_m(z)$ of the increments $z_m(t,j)$ of the financial process $X$. The length of the interval, $240/m$ minutes, may be read off from the horizontal axis.

basis, $m = 4$, the sample variance for the increments $\delta$ is five thousand times larger than the sample variance for the increments $z$. The unstable behaviour of the volatility over short time intervals suggest that the use of a continuous process (e.g. an Ito process) is inappropriate for modelling the volatility process.
squared standardized residual series of the GARCH(1,1) model. The autocorrelations in that table do not significantly differ from zero. Moreover, from the Lagrange multiplier test it follows that the null hypothesis of "no further structures in the squared residuals" can not be rejected. So this GARCH model gives a good description of the data.

Table 2.5. GARCH and EGARCH estimates
Maximum likelihood estimates (MLE) of the parameters of the GARCH(1,1) model (see equation (2.4)) and of the EGARCH(1,1) model (see equation (2.5)) based on the daily increments \(z_d\) of the financial process and least squares estimates (LSE) of these parameters based on the bivariate data \((z_d, \sigma_d)\). The estimates were obtained using Eviews. The standard errors of the estimates are displayed between brackets.

<table>
<thead>
<tr>
<th></th>
<th>(\hat{\alpha}_0)</th>
<th>(\hat{\alpha}_1)</th>
<th>(\hat{\beta}_1)</th>
<th>(\gamma_1)</th>
<th>(R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>LSE</td>
<td>5.61 (\cdot) 10(-5)</td>
<td>0.065</td>
<td>0.50</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(4.4 (\cdot) 10(-6))</td>
<td>(0.011)</td>
<td>(0.097)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MLE</td>
<td>2.5 (\cdot) 10(-6)</td>
<td>0.094</td>
<td>0.89</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(7.9 (\cdot) 10(-7))</td>
<td>(0.016)</td>
<td>(0.018)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGARCH</td>
<td>LSE</td>
<td>-2.0</td>
<td>0.079</td>
<td>0.79</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.021)</td>
<td>(0.018)</td>
<td>(0.013)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MLE</td>
<td>-0.42</td>
<td>0.21</td>
<td>0.97</td>
<td>-0.029</td>
</tr>
<tr>
<td></td>
<td>(0.089)</td>
<td>(0.032)</td>
<td>(0.0084)</td>
<td>(0.013)</td>
<td></td>
</tr>
</tbody>
</table>

Similarly to the GARCH model we found that the EGARCH(1,1) model was optimal within its class. The EGARCH(1,1) model is given by

\[
Z_d = \mu_d + h_d V_d \quad V_d \sim N(0,1), \\
\log(h_d^2) = \alpha_0 + \alpha_1 \frac{Z_{d-1}}{h_{d-1}} + \gamma_1 \frac{Z_{d-1}}{h_{d-1}} + \beta_1 \log(h_{d-1}^2). \tag{2.5}
\]

The maximum likelihood estimates of the parameters in equation (2.5) may be found in table 2.5. Table 2.6 contains the Ljung-Box Q-statistics for the squared standardized residuals.

Because we now have the realized volatility \(\sigma_d\), the parameters in the GARCH and EGARCH models may also be estimated by means of least squares. Recall that the least squares method also gives information on the goodness of fit of the model in terms of the coefficient \(R^2\). Here a word of warning is in place. In our setting the variable \(\sigma_d^2\) is itself an estimate. This provides a possible additional source of noise which may result in biased least squares estimators. We assume, because of the large number of observations per day, that this bias is small and may be neglected.
2.7 Confrontation with some standard models

Table 2.6. Q-statistics of the squared (E)GARCH residuals
Ljung-Box Q-statistic of the squared standardized residual series of the (E)GARCH(1,1) models. The corresponding p-values are in parentheses.

<table>
<thead>
<tr>
<th>Lag</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>0.339</td>
<td>0.818</td>
<td>2.34</td>
<td>2.34</td>
<td>2.76</td>
</tr>
<tr>
<td></td>
<td>(0.560)</td>
<td>(0.664)</td>
<td>(0.504)</td>
<td>(0.673)</td>
<td>(0.737)</td>
</tr>
<tr>
<td>EGARCH</td>
<td>0.460</td>
<td>1.06</td>
<td>3.70</td>
<td>3.71</td>
<td>4.21</td>
</tr>
<tr>
<td></td>
<td>(0.498)</td>
<td>(0.588)</td>
<td>(0.296)</td>
<td>(0.447)</td>
<td>(0.519)</td>
</tr>
</tbody>
</table>

The least squares estimates (LSE), obtained from the bivariate data \((z_d, \sigma_d)\) of the daily log returns series and the empirical volatilities, of the parameters in equation (2.4) and (2.5) are also presented in table 2.5. If one uses \(R^2\) as a criterion then the values in the table indicate that the EGARCH(1,1) model gives a much better fit to the data than the GARCH(1,1) model.

Empirically, there is evidence of long term memory or persistence in the volatility process. This is evident from the left graph of financial time in figure 2.9, which shows a constant slope over considerable periods of time. Bollerslev and Engle (1993) observed that the high persistence is captured in the GARCH(1,1) and EGARCH(1,1) models through the estimates \(\hat{\alpha}_1 + \hat{\beta}_1 = 0.98\) of equation and \(\hat{\beta}_1 = 0.97\). Our values, obtained from the bivariate series, are 0.6 and 0.8 respectively. For these parameter values the effect of the present value of the volatility disappears in a few days time. The two graphs in figure 2.9 also reveal abrupt changes in the volatility (slope), rather than a continuous change. This points to an underlying change-point mechanism rather than persistence in the volatility.

The right graph in figure 2.12 displays the volatility series generated by the EGARCH(1,1) model. If one compares the two graphs in figure 2.12 one can only marvel at the statistical ingenuity of the EGARCH model which manages to capture the intricacy of the volatility process in such detail using less than one tenth of one percent of the data we have. A closer look reveals differences. The extreme variability has been filtered out. More disturbing is the tendency to eliminate single high values and to widen peaks. Clustered high observations are transformed into continuous sharp peaks. Note though, that for clusters this is done without substantially diminishing the heights as one would expect under a smoothing operation. In fact, one might prefer to work with the volatility as depicted on the right in figure 2.12 since it gives the main features and leaves out the unpredictable details. Unfortunately this graph does not show the real thing.
2.8 The relation between $B$ and $Q$

We shall now discuss the relation between the time-change $Q$ and the Brownian motion $B$. A basic question is whether these processes are independent. As mentioned, Mandelbrot (1963) and Clark (1973) proposed a subordinated Brownian motion as a model. In this model the time-change $Q$ (subordinator) and the Brownian motion $B$ are independent processes. Traditionally the subordinator is assumed to be an increasing Lévy process, see Kallenberg (1997, p. 239). Mandelbrot assumes that the subordinator is a one-sided stable process, with index $\alpha \in (0, 1)$ and Clark suggests a log-normal process. Ané and Geman (2000, p. 2262) drop the assumption of independent stationary increments for the time-change. In our approach we do not make any assumptions on the time-change $Q$ except continuity. Below we consider the hypothesis that the sequence $(\sigma_d)$ of daily volatilities and the sequence $(u_d)$ of normalized daily increments of the financial process $X$ are independent.

It is believed (and empirically observed in the skewness of the distribution) that stock prices tend to move downwards faster than they move upwards. This has led to the development of asymmetric GARCH-type models such as EGARCH and to bivariate diffusion models in which the driving Brownian motions have non-zero correlation, see for example Heston (1993). Engle and Ng (1993) present additional evidence of this phenomenon by showing the asymmetric response of volatility to news.

If there is dependence between the subordinated process $B$ and the subordinator $Q$ of the kind described above, one would expect the increments of the Brownian motion to be negatively correlated with increments of the time-change. To check this we compute the correlation coefficient between $u_d$ and $\sigma_d$.
2.8 The relation between $B$ and $Q$

log($\sigma_d$), and the correlation coefficient between $u_d$ and $\delta_d = \log(\sigma_d/\sigma_{d-1})$. The results confirm those reported by Engle and Ng (1993): we find negative values for these coefficients, $-0.159$ and $-0.225$, respectively.

We are also interested in the question whether there is dependence between log volatility and increments of the Brownian motion and its past. In order to measure dependence between $B$ and $Q$ we estimate the six coefficients in the following equation relating the log increments of volatility to the absolute value and sign of the standardized log daily increments of prices on successive days:

$$
\delta_d = \alpha_0 + \alpha_1 \delta_{d-1} + \beta_0 |u_d| + \beta_1 |u_{d-1}| + \gamma_0 \text{sign}(u_d) + \gamma_1 \text{sign}(u_{d-1}) + \epsilon_d, \quad (2.6)
$$

where $\text{sign}(u) = 1$ for $u \geq 0$ and $\text{sign}(u) = -1$ for $u < 0$ and $\epsilon_d$ the error term.

In order to get some information on the robustness of the parameters $\beta_i, \gamma_i$, $i = 0, 1$, we also run regressions on:

$$
\delta_d = \alpha_0 + \beta_0 |u_d| + \gamma_0 \text{sign}(u_d) + \epsilon_d, \quad (2.7)
$$

$$
\delta_d = \alpha_0 + \beta_1 |u_{d-1}| + \gamma_1 \text{sign}(u_{d-1}) + \epsilon_d. \quad (2.8)
$$

**Table 2.7. Dependence between $\sigma_d$ and $u_d$**

Least squares estimates of the parameters of equations (2.6,2.7,2.8). Between brackets we display the standard errors of the estimates. The LSE were obtained from Eviews.

<table>
<thead>
<tr>
<th>eqn.</th>
<th>$\hat{\alpha}_0$</th>
<th>$\hat{\alpha}_1$</th>
<th>$\hat{\beta}_0$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\gamma}_0$</th>
<th>$\hat{\gamma}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2.6)</td>
<td>-0.0471</td>
<td>-0.362</td>
<td>0.0263</td>
<td>0.0357</td>
<td>-0.0391</td>
<td>0.00678</td>
</tr>
<tr>
<td></td>
<td>(0.0124)</td>
<td>(0.0277)</td>
<td>(0.00999)</td>
<td>(0.00990)</td>
<td>(0.00681)</td>
<td>(0.00698)</td>
</tr>
<tr>
<td>(2.7)</td>
<td>-0.0201</td>
<td>0.0291</td>
<td>-0.0402</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0104)</td>
<td>(0.0107)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2.8)</td>
<td>-0.0214</td>
<td>0.0267</td>
<td>0.0210</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0106)</td>
<td>(0.0108)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The estimates of the coefficients of (2.6,2.7,2.8) are given in table 2.7. The entries in this table show that the time-change is not independent of the Brownian path. The negative value of $\hat{\gamma}_0$ indicates that the volatility increases if standardized returns are positive. Note the strong dependence between $\delta_d$ and $|u_d|$, and between $\delta_d$ and $\text{sign}(u_d)$, as expressed in the estimates of the coefficients $\beta_0$ and $\gamma_0$ in table 2.7. This destroys the idea that volatility is predictable.
These results suggest that the subordinator model, where $B$ and $Q$ are independent, does not fit the data from the AEX index.

2.9 Conclusions

In this chapter we have tested the hypothesis that the standardized log returns of the AEX index are independent standard normal random variables. This hypothesis can not be rejected. The results are based on the relation between time-change and volatility of continuous local martingales. For the AEX index our data are sufficiently detailed to compute, to a high precision, daily volatility. By averaging the intra-day volatility over a period of three years one may distinguish events on a time scale of minutes. It is possible to discern news announcements on the U.S. economic indicators at 14.30 local time as a sharp peak and the opening of the NYSE one hour later as a clear jump to a higher level of volatility on the Amsterdam stock exchange.

The sequence of daily volatilities has been used to test GARCH and EGARCH models. The estimates of the coefficients of these models obtained from the observed volatility of the AEX lie far outside the range of the estimates obtained in the traditional approach, although the graphs of the volatility processes show a striking degree of resemblance.

The behaviour of the sample function of financial time for the AEX index suggests that volatility exists but may not be continuous. High peaks in the sequence of daily volatilities may be interpreted as evidence of a singular component in the time-change. The time change $Q$ or its inverse might be a more robust parameter to use in the statistical analysis of financial processes than the volatility itself.

We have uncovered two interesting paradoxes: (i) Volatility which is used as a parameter to monitor the price process is many times more variable than the price process itself. (ii) The variability in the volatility is high for time periods of a day; on the other hand the time-change exhibits long periods (up to several months) during which the slope is practically constant. Future research will focus on these paradoxes.