Chapter 3

Testing the diffusion model for the S&P500

3.1 Introduction

It is well known that the distributions of daily log-returns of stock indices display heavy tails and are asymmetric. These characteristics are in contradiction with the assumption that the underlying model is a geometric Brownian motion with constant volatility and motivated several authors, e.g. Rosenberg (1972), to propose that volatility itself is random.

In the general random volatility model the price process is described by the diffusion

\[ \frac{dS(t)}{S(t)} = dA(t) + dX(t). \] (3.1)

Here \( A \) is the (smooth) drift and \( X \) is a continuous martingale. Or equivalently, one could also assume that \( S \) is a semi-martingale with continuous sample paths.

The diffusion model (3.1) has received a lot of attention in the literature. The popularity of this model is explained mainly by the compatibility with the no-arbitrage condition. However, no convincing empirical evidence has been reported in the literature, to our knowledge, that warrants the use of this model. Finding this evidence is the aim of this chapter. To this purpose, we shall use the well-known result that every continuous local martingale \( X \) starting at the origin has the form

\[ X(t) = B(Q(t)). \] (3.2)

Here \( B \) is standard Brownian motion and \( Q \) is the quadratic variation of \( X \).

The question treated in this chapter can thus be rephrased as: “is the log price process, corrected for drift, a time-changed Brownian motion?” To an-
answer this question one has to perform three tasks: estimate the deterministic drift $A$, compute the realization of the quadratic variation $Q$ and test for a Brownian motion. We shall do this for a data-set of over 10 million intra-day observations of the S&P 500 stock index future.

Call the value of $Q$ at time $t$ the financial time at physical time $t$. We shall show that on this new time scale the process is a standard Brownian motion. Under the realistic economic assumption that $Q$ does not contain information on the future behaviour of $B$ the time-changed Brownian motion is a local-martingale and model (3.1) holds. Most interestingly one does not require independence between the processes $B$ and $Q$. Indeed, as pointed out by Andersen, Bollerslev, Diebold and Labys (2003) standardized returns need not be i.i.d. standard normal if there is dependence between $B$ and $Q$. The only assumption we shall make in this chapter is that the price process is continuous. There are many continuous processes. The diffusion is only one of them.

Let us now briefly discuss the model and the literature. The diffusion model has been around for more than a century. Bachelier's papers on one-dimensional diffusions were based on his experience at the Paris stock exchange, see Bachelier (1900). However, volatility is not constant. That is why one needs a diffusion with time-dependent coefficients. Such a model is able to explain the heavy tails of daily log-returns and the asymmetry of its distribution.

Clark (1973) and Mandelbrot (1963) introduced financial time to account for the fat tails. As in more recent work on stochastic volatility, see for example Barndorff-Nielsen and Shephard (2003), it is assumed that $X = B \circ T$, where $T$ is a random time-change independent of $B$, usually an increasing Lévy process. In this setting daily log-returns are mixtures of centered normal distributions, which may have thick tails but are symmetric. For other applications of the Lévy process to financial processes see for example Barndorff-Nielsen and Shephard (2000), Carr, Geman, Madan and Yor (2002), and Bingham and Kiesel (2001).

Ané and Geman (2000) fitted a model of Gaussian returns to time-changed log returns for certain well-traded stocks like IBM and Cisco. Due to the bid-ask spread the tick data for these stocks show a characteristic square saw-tooth effect and can not be interpreted as sample functions of a diffusion, see figure 3.1. They handle this difficulty by estimating the time-change from the number of transactions, and show that this is a better measure than market volume as used by Clark (1973). See also the paper by Ané and Geman (1996), where high frequency data of one future contract, for a period of six months, of the S&P500 index is investigated.

It is only more recently that a number of papers have surfaced based
3.2 Financial time

This section treats financial time. In the first subsection we shall go into the relation between quadratic variation and financial time. We discuss the dif-

Figure 3.1. The intra-day tick graphs of respectively IBM (left) and the S&P 500 (right) from 10 a.m. until 11 a.m. of February 12, 1996.

on quadratic variation. We mention the papers by Andersen et. al. (2000, 2001) on realized volatility of foreign exchange rates. There is an important difference between their approach and ours. As pointed out above, we do not require independence between the time-change and the price process. In fact, if we standardize the daily returns by the realized daily volatilities we have to reject i.i.d. standard normality as will be shown in section 3.4. This apparently is not the case for the foreign exchange data studied by Andersen et. al. (2000) and the data studied by Ané and Geman (2000). In the paper by Peters and de Vilder (2002b) both qualitative and quantitative properties of the realized volatility of the main Dutch Stock exchange index, using 15 second-interval data, are analyzed.

The chapter is organized as follows: Section 3.2 treats financial time. It consists of three parts. Subsection 3.2.1 discusses quadratic variation. Subsection 3.2.2 is concerned with increments of the log price process. Subsection 3.2.3 gives a detailed description of the data-set and discusses the empirical time-change. In section 3.3 we perform a series of tests on the log price process to test the hypothesis that the time-changed log prices of the S&P 500 corrected for drift describe a sample function of Brownian motion. Section 3.4 analyzes the dependence between the price process and the time-change. Section 3.5 contains some concluding remarks.

3.2 Financial time

This section treats financial time. In the first subsection we shall go into the relation between quadratic variation and financial time. We discuss the dif-
ference between estimating the drift of a diffusion and its quadratic variation. In subsection 3.2.2 we analyze some of the statistical problems encountered in estimating financial time and the effects of using an estimate of financial time in normalizing the data. Subsection 3.2.3 is devoted to a prescription of the data-set that we use in the analysis and to the construction of financial time for this data set.

3.2.1 Quadratic variation

The cornerstone result we use in this chapter is the Time-Change for Martingales Theorem (Dambis (1965), Dubins-Schwartz (1965)), which states the following. If \( X \) is a continuous local martingale started at the origin then \( X = B \circ Q \), where \( Q \) is the quadratic variation of \( X \) and \( B \) is a standard Brownian motion. The converse implication holds if the random variables \( Q(t) \) are stopping times for \( B \). See Monroe (1978). So under the assumption that the quadratic variation does not encode future events the tests of this chapter validate the proposition that the S&P 500 future index is a semi-martingale and that model (3.1) holds.

To see how subtle this result is, we focus on two extreme cases. The first case is the classical problem of estimating a continuously differentiable drift function \( f \) given the sample function \( \psi = f + \varphi \) over the interval \([0, c]\), with \( \varphi \) a standard Brownian sample function. For simplicity assume the drift is linear, \( f(t) = at \). The minimum variance unbiased estimate of the drift is \( \hat{a} = \psi(c)/c \) then. This estimator has a normal \( N(a, 1/c) \) distribution. Note in particular that the estimate is based on one observation only.

Now suppose we are given the function \( x = \varphi \circ q \) where \( q \) is an unknown continuous, strictly increasing function starting at the origin. Here the situation is completely different. Almost every sample path \( \varphi \) of Brownian motion has the property that given the distorted path \( x = \varphi \circ q \) it is possible to reconstruct \( \varphi \) and \( q \) for any time-change \( q \).

So it is possible to recover the exact form of \( q \) in contrast to the drift function \( f \). Indeed, this observation relates to Merton (1980) who showed that errors in the estimators of variances decrease with increasing sampling frequency while this does not hold for the mean.

The quadratic variation \( Q = \langle X, X \rangle \) of a continuous martingale \( X : [0, c] \to \mathbb{R} \) may be approximated by step-processes \( Q_n \) starting at 0 at time \( s_0 = 0 \) and with jumps \( (X(s_i) - X(s_{i-1}))^2 \) at the times \( 0 = s_0 < s_1 < \cdots < s_n = c \). Fisk’s Theorem, see for example Kallenberg (1997), gives simple conditions for convergence of the step-processes \( Q_n \) to \( Q \). This result is also valid for a semi-martingale \( Y = A + X \), with \( A \) a continuous process of bounded variation and \( \langle Y, Y \rangle = \langle X, X \rangle \). In Barndorff-Nielsen and Shephard
(2002) the distribution of $Q_n$ is derived for stochastic volatility models. In that paper it is also shown that the error between $Q_n$ and $Q$ decreases with the square root of the number of observations $n$.

3.2.2 Increments of the process $X$

There are essentially two types of increments that have our interest: standardized increments in physical time and increments in financial time. Most interest in the financial literature has been given to the first type of increments $U_i$ which are defined as

$$U_i = \frac{X(s_i) - X(s_{i-1})}{\sqrt{Q(s_i) - Q(s_{i-1})}}$$  \hspace{1cm} (3.3)

where $s_0 < \cdots < s_n$ are equidistant time points in physical time, e.g. market closing times. The increments in equation (3.3) are i.i.d. standard normal if $B$ and $Q$ are independent. If $B$ and $Q$ are dependent then this need not hold. In fact, if $X$ is a continuous martingale and the variables $U_i$ are not i.i.d. standard normal then $B$ and $Q$ are dependent. In that case the distribution of $U_i$ is not known.

On the other hand, dependence between $B$ and $Q$ does not influence the law of the scaled increments of $X$ in financial time. The latter increments are constructed as follows. Define financial time $\tau$ as the value of the time-change $Q$ at physical time $t$. Let $\tau_0 < \cdots < \tau_m$ be equidistant time points in financial time. If $X = B(Q)$ is a time-changed Brownian motion then the $m$ increments

$$\tilde{B}_i = \frac{X(Q^{-1}(\tau_i)) - X(Q^{-1}(\tau_{i-1}))}{\sqrt{\Delta \tau}} = \frac{B(\tau_i) - B(\tau_{i-1})}{\sqrt{\tau_i - \tau_{i-1}}}$$  \hspace{1cm} (3.4)

are i.i.d. $N(0, 1)$ random variables regardless the dependence structure of $B$ and $Q$. Here the inverse process $Q^{-1}(\tau)$ is defined as

$$Q^{-1}(\tau) = \inf_{\tau > 0} \{ Q(\tau) > \tau \},$$

and $\Delta \tau = \tau_i - \tau_{i-1}, i = 1, \ldots, m$.

If we replace the quadratic variation $Q$ by an approximating sum of squares, then the increments in (3.4) will be independent and symmetric, but the kurtosis may be smaller than that of the standard normal distribution. This is supported by the next proposition.
Proposition 3.1  Fix \( i, i = 1, \ldots, m \), and assume that the time-change \( Q \) is linear on the time interval \( \left[ Q^{-1}(\tau_{i-1}), Q^{-1}(\tau_{i}) \right] \). Let \( t^{(i)}_j, j = 0, \ldots, n_i \), be equidistant time points in this interval with \( Q^{-1}(\tau_{i-1}) = t^{(i)}_0 < \cdots < t^{(i)}_{n_i} = Q^{-1}(\tau_{i}) \). Define \( B_i \) as \( \tilde{B}_i \) in formula (3.4), where \( \Delta \tau \) is replaced by the approximating sum \( Q_i = \sum_{j=1}^{n_i} \left( X(t^{(i)}_j) - X(t^{(i)}_{j-1}) \right)^2 \). The density of \( B_i \) is given then by

\[
 f_{B_i}(b) = \frac{\Gamma(n_i/2)}{\sqrt{\pi n_i^{1/2}}((n_i - 1)/2)} \left( 1 - \frac{b^2}{n_i} \right)^{n_i - 3/2} 1_{(-\sqrt{n_i}, \sqrt{n_i})}(b), \quad b \in \mathbb{R},
\]  

(3.5)

with \( \Gamma(\cdot) \) the gamma function.

Proof. The increments of \( X \) on the interval \( \left[ Q^{-1}(\tau_{i-1}), Q^{-1}(\tau_{i}) \right] \) are independent normal variables with zero mean. Together with the linearity of \( Q \) and the equidistance of \( t^{(i)}_0, \ldots, t^{(i)}_{n_i} \) this implies that \( B_i \) has the same distribution as \( \sum_{j=1}^{n_i} Z_j \{ \sum_{k=1}^{n_i} Z_k \}^{-1/2} \), with \( Z_1, \ldots, Z_{n_i} \) i.i.d standard normal. Consequently, the random variable \( T_i = g(B_i) = \sqrt{n_i - 1} B_i / \sqrt{n_i - B_i^2} \) is \( \text{Student}(n_i - 1) \) distributed with density \( f_{T_i}(t) \) and the density \( f_{B_i}(b) \) of \( B_i \) is given by \( f_{T_i}(g(y)) | g'(y) | \). Some computation yields the density in (3.5). \( \square \)

It is straightforward to compute the second and fourth moment of \( B_i \) via

\[
 E(B_i^k) = \frac{\Gamma(n_i/2)}{\sqrt{\pi n_i^{1/2}}((n_i - 1)/2)} \int_{-\sqrt{n_i}}^{\sqrt{n_i}} b^{2k} \left( 1 - \left( \frac{b^2}{n_i} \right)^2 \right)^{n_i - 3/2} db
\]

\[
 = \frac{n_i k \Gamma(n_i/2)}{\sqrt{\pi} \Gamma((n_i - 1)/2)} \int_0^1 w^{k+\frac{1}{2}-1} (1 - w)^{n_i - 1 - 1} dw
\]

(3.6)

In particular \( EB_i^2 = 1 \) and \( EB_i^4 = 3n_i/(n_i + 2) \). For \( n_i \to \infty \) one can proof that the random variable \( B_i \) converges to a standard normal random variable. In the next section we are confronted with a finite number of observations. For certain time-intervals, where the volatility is extremely high, financial time runs so fast that there are only a few observation points available in the interval \( \left[ Q^{-1}(\tau_{i-1}), Q^{-1}(\tau_{i}) \right] \) in physical time, i.e. \( n_i \) is small. Proposition 3.1 implies that in these cases the observed distribution will have thin tails, as compared to the tails of the normal distribution. We shall avoid this problem by choosing sufficiently large financial time-intervals.
3.2 Financial time

3.2.3 The data and some notation

Our data-set is the U.S. Standard & Poor 500 stock index future \((S)\), traded on the Chicago Mercantile Exchange (CME), for the period 1st of January, 1988 until September 1st, 2001. The data-set was supplied by tickdata.com. For reasons pointed out in the previous subsection (too few observations are available during periods of high volatility) we left out October 1987 from the analysis. In total we then have 3430 trading days \(d = 1, \ldots, 3430\), where a day starts at 9:30 and ends at 16:00 CET. We always use the future contract with the shortest time horizon, being the most actively-traded contract. There are four expiration months: March, June, September, and December. So we start with the future contract that expired on the third Friday of March 1988. We next use the future contract that expired on the third Friday of June 1988, and so on. The last contract we use is the one that expired on the third Friday of September 2001. Hence the contracts we use always have a time horizon of at most 3 months.

We use future data rather than the S&P500 cash index to avoid non-synchronous trading effects which causes positive autocorrelation between successive observations, see Dacorogna, Gençay, Müller, Olsen and Pictetna (2001). As in the cash index there are bid-ask effects in the future prices which induce negative autocorrelation between successive observations, cf. figure 3.1. One can deal with this by taking larger time-intervals. For two-minute time intervals the bid-ask effects are no longer visible. Hence, we restrict attention to two-minute time-intervals.

We assume that at each time-point the S&P 500 future 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 in the future contract. Transactions update the value. One could argue that trade intensity, volume, liquidity and volatility all are determined by the financial weather.

Throughout this chapter we shall use capital letters for random variables and lower case letters for the corresponding realizations. The sample function of the process \(X\) is a construct. Each day yields \(j_d + 1 = 196\) two-minute observations \(s(d,0), \ldots, s(d,j_d)\), and thus \(j_d\) intervals. The increment \(z(d,j)\) of the process \(X\) over the \(j\)th interval of day \(d\) is defined as

\[
z(d,j) = \log(s(d,j)/s(d,j-1)) - \bar{\mu}/j_d,
\]

where \(\bar{\mu}\) is the average of the 3430 daily log-increments, see below. Now define

\[
z_d = z(d,1) + \cdots + z(d,j_d), \quad \sigma^2[d] = z^2(d,1) + \cdots + z^2(d,j_d) \tag{3.7}
\]

as the daily increase of the process \(X\) and the estimate of the daily increase of financial time, respectively. So the sample function \(x\) of the process \(X\)
and the estimate $\hat{q}$ of the time-change $Q$ associated with the sample function $x$ on time-intervals $t \in [d - 1 + j/j_d, d - 1 + (j + 1)/j_d)$ are given by the (step-functions) sums

$$
x(t) = z_1 + \cdots + z_{d-1} + z(d, 1) + \cdots + z(d, j),
$$
\begin{align*}
\hat{q}(t) &= \sigma^2[1] + \cdots + \sigma[d - 1]^2 + z^2(d, 1) + \cdots + z^2(d, j).
\end{align*}

(3.8)

Notice that we focus on intra-day prices. This is due to the absence of information on the quadratic variation during nights, weekends, and holidays. Hence the process $X$ is constructed in terms of a subset of increments of the 'true' price process. This procedure does not affect the main point of this chapter since a martingale is defined in terms of its increments. The sample path of the resulting process $X$ may be found in figure 3.2. The empirical function $\hat{q}$ may be found in figure 3.3. The flatter parts in the quadratic variation function correspond to phases where the financial clock runs slowly while in the steeper parts the financial clock runs faster. Observe from this figure that after 3430 trading days the total quadratic variation is $c \approx 0.285$.

The main results of this chapter are also valid for time intervals larger than two minutes. From the results on realized volatility of Barndorff-Nielsen and Shephard (2002) it is known that the variance of statistical error already doubles for four minute time intervals. On the other hand, for time intervals smaller than two minutes the bid-ask effect is dominant, leading to a positive bias in the quadratic variation. This explains the choice of two minute time-intervals in this chapter.
3.2 Financial time

Figure 3.2. Top: the log of the S&P500 corrected for drift in physical time. Bottom: the log of the S&P500 corrected for drift in financial time.
The slope of the drift $\tilde{\mu} \approx 7.80 \cdot 10^{-5}$ is the average of $\log(s(d, j_d)/s(d, 0))$ over the 3430 days of our data-set. To see the influence of the drift on the total quadratic variation we have computed the difference between the quadratic variation of $x(t)$ and the quadratic variation of $x(t) + \tilde{\mu}t$, which is equal to

$$3430 \cdot 195(\tilde{\mu}/195)^2 \approx 1.07 \cdot 10^{-7}.$$ 

Hence, smooth drift functions with a slope of the order $\tilde{\mu}$ have a negligible influence on the quadratic variation.

The time-points $\tau_i = \hat{q}(t_i)$ where the process $X$ is evaluated, are chosen according to a special rule. Since $\hat{q}$ is a step-function we can not choose the time-points $t_i$ so that $\Delta \tau_i = \tau_i - \tau_{i-1}$ are equal. Instead we first fix $\Delta \tau$ and then choose the time-points $t_1, t_2, \ldots$ successively so that $\Delta \tau_i \geq \Delta \tau$ and $\Delta \tau_i$ is minimal. So $\hat{q}(t_i) - \hat{q}(t_{i-1}) \geq \Delta \tau$ and $\hat{q}(t'_i) - \hat{q}(t_{i-1}) < \Delta \tau$ for any time-point $t'_i < t_i$. In periods of high volatility $\Delta \tau_i - \Delta \tau$ may be large.

We choose $\Delta \tau = 0.2ft$ where $ft = 0.001$ units of financial time. This gives us 1372 financial time points. We emphasize that the results presented in this chapter apply equally well for different values of $\Delta \tau$.

The scaled increments $b_i$ are given by

$$b_i = \frac{x(t_i) - x(t_{i-1})}{\Delta \tau_i}.$$ 

Figure 3.2 depicts also the process $X$ of the S&P 500 future index in financial time. Observe that, by construction, the sample path in this figure
is a sample path of a Brownian bridge instead of a Brownian motion. See also figure 3.4 where the increments $b_i$ are presented. Observe from figure 3.2a-b that patterns of the process in physical time are conserved, possibly in stretched (compressed) form, in financial time. It is only the horizontal axis which is deformed. The exact way in which the patterns are transformed is determined by the time-change function depicted in figure 3.3.

![Figure 3.4](image)

**Figure 3.4.** The time series of the increments $b_i$.

Table 3.1 displays some of the basic statistical properties of the increments $b_i$. Also included in this table are some statistical properties of the non-standardized intra-day returns $r_d = z_d + \bar{\mu}$. Observe that the skewness of $b_i$ is close to zero. On the other hand, the kurtosis of the series is less than 3. Along the lines set out at the end of the previous subsection we have also included the expected average kurtosis

$$k^* = m^{-1} \sum_{i=1}^{m} E b_i^4 = m^{-1} \sum_{i=1}^{m} 3n_i/(n_i + 2)$$

in table 1, where $n_i$ denotes the number of observations needed to construct $b_i$. Notice that $k^*$ is only slightly smaller than the theoretical value for the standard normal distribution.
Table 3.1. Statistical properties of the returns
Statistical properties of intra-day returns $r_d$ and the standardized returns in financial time $b_t$. The expected average kurtosis, based on proposition 3.1, is displayed in the column $k^*$.  

<table>
<thead>
<tr>
<th>Series</th>
<th>Mean</th>
<th>St.dev.</th>
<th>$\hat{\alpha}$</th>
<th>$k$</th>
<th>$k^*$</th>
<th>size</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_d$</td>
<td>$7.80 \cdot 10^{-5}$</td>
<td>$9.35 \cdot 10^{-3}$</td>
<td>$-0.453$</td>
<td>$13.4$</td>
<td>-</td>
<td>$3430$</td>
</tr>
<tr>
<td>$b_t$</td>
<td>$2.68 \cdot 10^{-3}$</td>
<td>$0.989$</td>
<td>$-0.0120$</td>
<td>$2.75$</td>
<td>$2.95$</td>
<td>$1372$</td>
</tr>
</tbody>
</table>

3.3 Testing the hypothesis

This section tests the validity of the diffusion model (3.1) for the S&P 500. We do this by checking whether the process $X$ is a Brownian motion in financial time. For a continuous function $\varphi : [0, a] \rightarrow \mathbb{R}$ it is not hard to test whether it is a Brownian sample path. Choose a large integer $m$. Now observe that for a standard Brownian motion $B$ the scaled increments

$$u_i = (\varphi(a(i)/m) - \varphi(a(i - 1)/m))/\sqrt{a/m}, \quad i = 1, \ldots, m,$$

are independent standard normal random variables. So we shall test whether the sequence $b_i$ may be regarded as a sample from a standard normal distribution.

3.3.1 The Tests

We apply five tests: two for normality, one for the tails and two for independence. We also present some graphical evidence.

(i) The Kolmogorov-Smirnov test.

The first test we use is the Kolmogorov-Smirnov test. See Shorack and Wellner (1986) for a detailed discussion. We test normality with fixed mean ($=0$) and fixed variance ($=1$). The outcomes are presented in table 3.2. We also apply the test to the intra-day returns $r_d$. As can be observed from table 3.2, standard normality is not rejected for the series $b_t$. It is rejected for the unscaled return series $r_d$.

(ii) The Jarque-Bera test.

The next we use the Jarque-Bera (1980) test to test for normality. The results of this test may also be found in table 3.2. As can be observed from this table normality is not rejected for the series $b_t$. If we take into account the expected value of the kurtosis $k = k^* < 3$, then the Jarque-Bera statistic
3.3 Testing the hypothesis

Table 3.2. Testing for normality

The test statistics of the Kolmogorov-Smirnov (KS), Jarque-Bera (JB) and the adjusted Jarque-Bera (JB*) tests applied to the intra-day log-returns $r_d$ and the series $b_i$. For the series $r_d$ we used a KS test with unknown mean and variance for the series $b_i$ with fixed mean (=0) and variance (=1). The values of $jb$ and $ks$ were obtained from Eviews and S-Plus respectively. The adjusted Jarque-Bera test (JB*) is corrected for the expected average $k^*$ given in table 3.1. We have also included the p-values for the different statistics.

<table>
<thead>
<tr>
<th>Series</th>
<th>$ks$</th>
<th>$P(KS &gt; ks)$</th>
<th>$jb$</th>
<th>$P(JB &gt; jb)$</th>
<th>$jb^*$</th>
<th>$P(JB^* &gt; jb^*)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_d$</td>
<td>0.0631</td>
<td>0.000</td>
<td>1.56·$10^4$</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$b_i$</td>
<td>0.0138</td>
<td>0.9572</td>
<td>3.54</td>
<td>0.170</td>
<td>2.42</td>
<td>0.457</td>
</tr>
</tbody>
</table>

$\frac{\tilde{s}^2}{\hat{s}^2} + \frac{1}{4} (\hat{k} - 3)^2$ is replaced by $(2n/3)(\tilde{s}^2 + (\hat{k} - k^*)^2)/4$. Asymptotically this adjusted Jarque-Bera statistic is $\chi^2$ distributed with two degrees of freedom. The new values for the Jarque-Bera statistic may be found in the column $jb^*$ in table 3.2. Observe that normality is again not rejected for $b_i$ and that the p-value has tripled for the adjusted Jarque-Bera test. As expected, standard normality is rejected for the unscaled series $r_d$.

(iii) The binomial test.

For financial data extreme values are important. We perform three simple binomial tests to check whether the series $b_i$ has heavy tails by counting the number of observations exceeding the values of 2, 3, and 3.5 standard deviations respectively. Let $p_c = P\{|U| > c\}$, where $U$ is standard normal. For $n$ independent observations of $|U|$, the number $N_c$ of observations exceeding the level $c$ has a binomial-$\binom{n}{p_c}$ distribution with mean $\mu_c = np_c$ and variance $\sigma_c^2 = np_c(1 - p_c)$. The observed number of observations exceeding $c$ is $n_c$. The results are presented in table 3.3. As can be seen from the table the hypothesis that the tails behave like those of a standard normal distribution is not rejected at a level of 10% for the series $b_i$.

(iv) 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 (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.
Table 3.3. Observations in the tails

The number of observations of $|b_i|$ exceeding 2, 3 and 3.5 standard deviations together with the mean, the standard deviation and the corresponding probabilities $P(B_c > n_c)$, where $B_c$ is Binomial(1370, $p_c$) distributed.

<table>
<thead>
<tr>
<th>Series</th>
<th>$c$</th>
<th>$p_c$</th>
<th>$n_c$</th>
<th>$\mu_c$</th>
<th>$\sigma_c$</th>
<th>$P(B_c &gt; n_c)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>b_i</td>
<td>$</td>
<td>2.0</td>
<td>0.0456</td>
<td>58</td>
<td>62.6</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>0.0027</td>
<td>2</td>
<td>3.70</td>
<td>1.92</td>
<td>0.715</td>
</tr>
<tr>
<td></td>
<td>3.5</td>
<td>4.6·10^{-4}</td>
<td>0</td>
<td>0.631</td>
<td>0.794</td>
<td>0.468</td>
</tr>
</tbody>
</table>

sequence. Let $x_1, \ldots, x_N$ be a time series and let $y_j$ be a reconstruction vector defined by $y_j := (x_{j-m+1}, \ldots, x_j)'$, where $m \in \mathbb{N}^+$ is the embedding dimension. Let $y_j$ and $y_{j'}$ be two arbitrary vectors and $x_k$ and $x_{k'}$ two arbitrary observations. If $(x_j)_{j=1}^N$ is an i.i.d. sequence then

$$P(|Y_j - Y_{j'}| < \epsilon) = (P(|X_k - X_{k'}| < \epsilon))^m,$$

where $|\cdot|$ denotes the sup-norm, $(|Y| = \sup_{i=1\ldots m} |Y_i|) \in \mathbb{R}^m$ and $\epsilon > 0$. The BDS test utilizes the difference of estimates of these two probabilities, $\hat{P}(|Y_j - Y_{j'}| < \epsilon) - (\hat{P}(|X_k - X_{k'}| < \epsilon))^m$. The estimates $\hat{P}(|Y_j - Y_{j'}| < \epsilon)$ and $\hat{P}(|X_k - X_{k'}| < \epsilon)$ are obtained by counting the number of vectors $y_j$ of observations and the number of observations $x_j$ that lie within an $\epsilon$ distance. Under the null hypothesis, the difference between the probabilities is approximately normally distributed.

We calculate the BDS statistic for different embedding dimensions $m$ and distances $\epsilon$ for $b_i$, see table 3.4. The p-values in table 3.4 for each $m$ and $\epsilon$ were obtained by reshuffling the sequence $b_i$ 2000 times, calculating for each reshuffled sequence the BDS statistic and then counting the number of absolute values that exceeded the absolute value of the corresponding statistic. Observe from this table that we can not reject i.i.d. on any reasonable level.

(v) (Partial) autocorrelation function.

As a final check for i.i.d. we investigate the (partial) autocorrelation function for the sequences $b_i$. If the series are indeed i.i.d. then we expect the coefficients of these functions to be close to zero. In figure 3.5 the autocorrelation functions for $b_i$ for lags of $1, \ldots, 10$ are displayed. Observe from this figure that for both functions only lag number eight lies outside the 95% confidence level. Moreover, the coefficients are on average negative. This is partly due to the fact that the process $X$, by construction, forms a Brownian bridge. It is well known that increments of such processes are negatively correlated. The theoretical autocorrelation coefficient for $b_i$
3.3 Testing the hypothesis

Table 3.4. The BDS test
The BDS statistics and the corresponding p-values for the series $b_t$. 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 $b_t$ 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 $b_t$.

<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>$-2.76 \cdot 10^{-4}$</td>
<td>$-5.39 \cdot 10^{-5}$</td>
<td>$-1.70 \cdot 10^{-5}$</td>
<td>$6.17 \cdot 10^{-6}$</td>
<td>$3.55 \cdot 10^{-6}$</td>
</tr>
<tr>
<td>p</td>
<td>0.581</td>
<td>0.687</td>
<td>0.599</td>
<td>0.554</td>
<td>0.311</td>
</tr>
<tr>
<td>bds 0.5</td>
<td>$-6.07 \cdot 10^{-4}$</td>
<td>$-3.50 \cdot 10^{-4}$</td>
<td>$-1.03 \cdot 10^{-4}$</td>
<td>$2.11 \cdot 10^{-6}$</td>
<td>$-1.26 \cdot 10^{-5}$</td>
</tr>
<tr>
<td>p</td>
<td>0.747</td>
<td>0.669</td>
<td>0.748</td>
<td>0.988</td>
<td>0.776</td>
</tr>
<tr>
<td>bds 0.75</td>
<td>$-1.45 \cdot 10^{-3}$</td>
<td>$-1.50 \cdot 10^{-3}$</td>
<td>$-9.65 \cdot 10^{-4}$</td>
<td>$-5.13 \cdot 10^{-4}$</td>
<td>$-3.09 \cdot 10^{-4}$</td>
</tr>
<tr>
<td>p</td>
<td>0.701</td>
<td>0.529</td>
<td>0.465</td>
<td>0.454</td>
<td>0.396</td>
</tr>
<tr>
<td>bds 1</td>
<td>$-1.99 \cdot 10^{-3}$</td>
<td>$-2.65 \cdot 10^{-3}$</td>
<td>$-2.35 \cdot 10^{-3}$</td>
<td>$-1.91 \cdot 10^{-3}$</td>
<td>$-1.41 \cdot 10^{-3}$</td>
</tr>
<tr>
<td>p</td>
<td>0.738</td>
<td>0.556</td>
<td>0.483</td>
<td>0.422</td>
<td>0.343</td>
</tr>
</tbody>
</table>

is approximately $-0.001$. So actually we have to shift the confidence level downwards by this number. Since this adjustment will not have a great impact on the results we omit this exercise. We next test whether the coefficients differ significantly from zero by means of the Ljung-Box $Q$-statistic. Under the null hypotheses the $Q$-statistic is asymptotically $\chi^2$ distributed with one degree of freedom. In table 3.5 we present the results. Note that we can not reject that the $b_t$ are realizations of an i.i.d. sequence, at a 10% level.

Table 3.5. Ljung-Box $Q$-statistics of $b_t$
$Q$-statistics (q) for lags 1 to 10 for the series $b_t$. The p-values are denoted by p.

<table>
<thead>
<tr>
<th>Lags</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>q</td>
<td>0.0551</td>
<td>1.12</td>
<td>2.24</td>
<td>2.95</td>
<td>2.96</td>
</tr>
<tr>
<td>p</td>
<td>0.814</td>
<td>0.572</td>
<td>0.523</td>
<td>0.567</td>
<td>0.706</td>
</tr>
<tr>
<td>Lags</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>q</td>
<td>5.47</td>
<td>5.75</td>
<td>9.83</td>
<td>12.3</td>
<td>12.4</td>
</tr>
<tr>
<td>p</td>
<td>0.485</td>
<td>0.570</td>
<td>0.277</td>
<td>0.198</td>
<td>0.260</td>
</tr>
</tbody>
</table>

We performed the same analysis for the squared sequence $b_t^2$. Figure 3.5 displays also the (partial) autocorrelation function for lags 1, \ldots, 10 and ta-
Figure 3.5. Above: Left the autocorrelation function of the increments $b_i$ and right the partial autocorrelation function. Below: the (partial) autocorrelation function of the squared increments $b_i^2$.

Table 3.6 presents the corresponding Ljung-Box $Q$-statistics. Note that we cannot reject independence for the squared sequence at a 10% level.

In addition to the tests described above we performed a number of qualitative graphical tests. Figure 3.6 displays the Gaussian kernel density estimate of $b_i$. The figure also includes the Gaussian kernel density estimate for a sequence of 1372 simulated i.i.d. standard normal variables. Observe that the two kernel density estimates have a large degree of resemblance.

Figure 3.7 presents the Quantile-Quantile plot. As can be seen from this plot most observations lie on the diagonal which is an indication of the presence of normality. The points that deviate from the diagonal are accounted for by the slightly thinner tails of the empirical distribution of $b_i$ as expressed by the empirical kurtosis $k^*$, see table 3.1, and the binomial test.

Figure 3.7 also depicts the scatterplot of $(\Phi(b_i), \Phi(b_{i+1}))$ for $i=1,...,1372$. Here $\Phi$ denotes the standard normal distribution function. Observe that the
3.3 Testing the hypothesis

Table 3.6. Ljung-Box $Q$-statistics of $b_i^2$

$Q$-statistics ($q$) for lags 1 to 10 for the series $b_i^2$. The p-values are denoted by $p$.

<table>
<thead>
<tr>
<th>Lags</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q$</td>
<td>0.0635</td>
<td>1.48</td>
<td>3.05</td>
<td>3.07</td>
<td>3.16</td>
</tr>
<tr>
<td>$p$</td>
<td>0.801</td>
<td>0.476</td>
<td>0.384</td>
<td>0.547</td>
<td>0.675</td>
</tr>
<tr>
<td>Lags</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>$q$</td>
<td>3.82</td>
<td>3.82</td>
<td>3.82</td>
<td>4.36</td>
<td>9.30</td>
</tr>
<tr>
<td>$p$</td>
<td>0.702</td>
<td>0.800</td>
<td>0.873</td>
<td>0.886</td>
<td>0.504</td>
</tr>
</tbody>
</table>

Figure 3.6. The Gaussian kernel density estimate of the increments $b_i$. The dashed line corresponds with the kernel density estimate of 1372 simulated i.i.d. standard normal variables.

transformed increments are uniformly distributed over the unit square. This finding is consistent with i.i.d. standard normality of the variables.

In summary, the statistical tests we have performed in this section on the process $X$ in financial time show that it constitutes a standard Brownian motion. An alternative way to state this result is that in physical time the process $X$ associated with $S$ is a time-changed Brownian motion. Under the standing assumption on the time-change $Q$ we may conclude that $X$ is a martingale and the diffusion model (3.1) is valid. Hence, $S$ is a semi-martingale.
Testing the diffusion model for the S&P500

3.4 Dependence between financial time and the price process

In this section the dependence structure between $Q$ and $B$ is investigated. Since in the previous section we have shown that we can not reject the hypothesis that in financial time the process $X$ is a Brownian motion, it is natural to define volatility over the interval $[s, t]$ by

$$\sigma[s, t] = \sqrt{\frac{Q(t) - Q(s)}{t - s}}.$$  \hspace{1cm} (3.9)

So in our setting the squared volatility measures the speed of financial time with respect to physical time. Volatility is high when the financial clock runs fast and low when the financial clock runs slowly.

Many studies present evidence that there is an asymmetric relation between volatility and the price process, see for example Engle and Ng (1993). An appealing theoretical foundation of this relation has been provided by Black (1976) and Christie (1982), who refer to it as the financial leverage effect. They reason that a decline in the stock price increases the company’s financial leverage which makes the stock riskier and leads to a higher volatility level. A different but possibly complementary explanation for this asymmetric relation is the so-called volatility feedback effect proposed by Pindyck (1984) and French, Schwert and Stambaugh (1987). These authors argue that anticipated changes in volatility affect the required return on the investment which implies an immediate adjustment of the stock price. Although the previously mentioned studies were concerned primarily with the
behaviour of individual stocks, it seems reasonable to assume that similar arguments apply for weighted sums of individual stocks. For an extensive study and discussion of both effects see Wu (2001).

As indicated in section 3.2.2 a clear indication that \( B \) and \( Q \) are dependent would be that the standardized daily observations

\[
u[d] = z[d] / \sigma[d]
\]

are not realizations of an i.i.d. \( N(0,1) \) sequence. Here \( \sigma[d] \) stands for the daily volatility, see equations (3.7) and (3.9). Indeed, the mean (=0.053) and the skewness (=0.14) of the series \( u[d] \) differ significantly from zero and standard normality has to be rejected. Hence, we conclude that \( B \) and \( Q \) are dependent. Similar results hold if we would take into account simple linear relations between the conditional mean of \( z[d] \) and \( \sigma[d] \) or \( \log(\sigma[d]) \). These findings seem to be in contradiction with those reported by Anderson et al. (2000) and Ané and Geman (2000). They present evidence that foreign exchange rates and two individual stock returns respectively, normalized by the daily volatilities, are Gaussian. Concerning the foreign exchange rates there is a priori no reason to assume that standardized returns and volatility are dependent. This is, as pointed out above, not obvious for asset prices.

So, if the process \( X \) is a continuous martingale then in physical time the standardized returns of an asset need not be i.i.d. standard normal, while scaled returns in financial time will be i.i.d. normal. This is a clear demonstration of the advantage of the approach propagated in this chapter.

Let us now investigate, to some extent, the dependence structure that causes the rejection of i.i.d. standard normality for daily standardized returns. We focus on the relation between the increments \( b_i \) and the volatility \( \sigma_i = \sigma[t_i, t_{i-1}] \) on the intervals \([t_i, t_{i-1}] \). The time points \( t_i \) were derived in section 3.2.3 where \( \tau_i = \tilde{q}(t_i) \) and \( \tau_i, i = 0, ..., 1371 \), fixed financial time points.

We investigate the relation by considering the regressions of \( \log \sigma_i \) and \( \log(\sigma_i / \sigma_{i-1}) \) on the orthogonal covariates \( |b_{i-k}| \) and \( \text{sign}(b_{i-k}) \) for \( k = 0, 1 \):

\[
\begin{align*}
\log \sigma_i &= \alpha_1 + \beta_{0,1} |b_i| + \gamma_{0.1} \text{sign}(b_i) + \beta_{1,1} |b_{i-1}| + \gamma_{1.1} \text{sign}(b_{i-1}), \\
\log \frac{\sigma_i}{\sigma_{i-1}} &= \alpha_2 + \beta_{0.2} |b_i| + \gamma_{0.2} \text{sign}(b_i) + \beta_{1.2} |b_{i-1}| + \gamma_{1.2} \text{sign}(b_{i-1}).
\end{align*}
\]  

The estimates of the coefficients may be found in table 3.7. Observe that both regressions have an identical form and that all \( \gamma' \)'s have negative estimates. These findings show that on average the speed of financial time is higher for negative increments \( b_i \) and/or lower for positive increments. Thus, financial time speeds up when increments are negative. So there is an asymmetric relation between the price process and financial time. Similar conclusions hold for the lagged covariates.
Table 3.7. Dependence relation
The least squares estimates of the parameters in the regression relations (3.11). Between brackets we display the standard errors of the estimates. The LSE's are obtained from Eviews.

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\alpha}_k$</th>
<th>$\hat{\beta}_0^k$</th>
<th>$\hat{\beta}_{1,k}$</th>
<th>$\hat{\gamma}_{0,k}$</th>
<th>$\hat{\gamma}_{1,k}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-19.3</td>
<td>0.118</td>
<td>0.241</td>
<td>-0.0983</td>
<td>-0.163</td>
</tr>
<tr>
<td></td>
<td>(0.0693)</td>
<td>(0.0498)</td>
<td>(0.0498)</td>
<td>(0.0289)</td>
<td>(0.0289)</td>
</tr>
<tr>
<td>2</td>
<td>-0.217</td>
<td>0.136</td>
<td>0.135</td>
<td>-0.118</td>
<td>-0.0696</td>
</tr>
<tr>
<td></td>
<td>(0.0447)</td>
<td>(0.0346)</td>
<td>(0.0346)</td>
<td>(0.0201)</td>
<td>(0.0201)</td>
</tr>
</tbody>
</table>

3.5 Conclusions

The aim of this chapter was to show that the diffusion model for asset prices is valid over the S&P500 stock index during the period 1988-2001. In the diffusion model the process $X$ associated with the stock index is a continuous martingale. The time-change for martingales theorem states that any continuous local martingale is a time-changed Brownian motion. The time change is given by the quadratic variation of the process $X$. Using this time-change we could not reject the hypothesis that in financial time the process $X$ is a standard Brownian motion. More precisely, we could not reject the hypothesis that the scaled $0.2ft$ increments of the process $X$ are i.i.d. standard normal. On the other hand, we rejected the hypothesis that daily standardized increments are i.i.d. standard normal. So there is significant dependence between the time-change and the Brownian motion. As a final remark we mention that preliminary analysis indicates that similar conclusions, as reported in this chapter, also hold for other indices.