

1

What is an ARCH process?

1.1 Introduction

Since the first decades of the twentieth century, asset returns have been assumed to form an independently and identically distributed (i.i.d.) random process with zero mean and constant variance. Bachelier (1900) was the first to contribute to the theory of random walk models for the analysis of speculative prices. If $\{P_t\}$ denotes the discrete time asset price process and $\{y_t\}$ the process of continuously compounded returns, defined by $y_t = 100 \log(P_t/P_{t-1})$, the early literature viewed the system that generates the asset price process as a fully unpredictable random walk process:

$$\begin{aligned} P_t &= P_{t-1} + \varepsilon_t \\ \varepsilon_t &\stackrel{i.i.d.}{\sim} N(0, \sigma^2), \end{aligned} \quad (1.1)$$

where ε_t is a zero-mean i.i.d. normal process. Figures 1.1 and 1.2 show simulated $\{P_t\}_{t=1}^T$ and $\{y_t\}_{t=1}^T$ processes for $T = 5000$, $P_1 = 1000$ and $\sigma^2 = 1$.

However, the assumptions of normality, independence and homoscedasticity do not always hold with real data.

Figures 1.3 and 1.4 show the daily closing prices of the London Financial Times Stock Exchange 100 (FTSE100) index and the Chicago Standard and Poor's 500 Composite (S&P500) index. The data cover the period from 4 April 1988 until 5 April 2005. At first glance, one might say that equation (1.1) could be regarded as the data-generating process of a stock index. The simulated process $\{P_t\}_{t=1}^T$ shares common characteristics with the FTSE100 and the S&P500 indices.¹ As they are clearly

¹ The aim of the visual comparison here is not to ascertain a model that is closest to the realization of the stochastic process (in fact another simulated realization of the process may result in a path quite different from that depicted in Figure 1.1). It is merely intended as a first step towards enhancing the reader's thinking about or conceiving of these notions by translating them into visual images. Higher-order quantities, such as the correlation, absolute correlation and so forth, are much more important tools in the analysis of stochastic process than their paths.

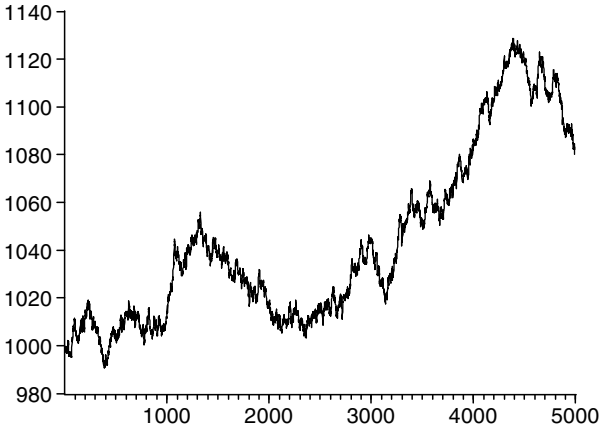


Figure 1.1 Simulated $\{P_t\}$ process, where $P_t = P_{t-1} + \varepsilon_t$, $P_1 = 1000$ and $\varepsilon_t \stackrel{i.i.d.}{\sim} N(0, 1)$.

non-stationary, the autocorrelations presented in Figure 1.5 are marginally less than unity in any lag order. Figure 1.6 plots the distributions of the daily FTSE100 and S&P500 indices as well as the distribution of the simulated process $\{P_t\}_{t=1}^T$. The density estimates are based on the normal kernel with bandwidths method calculated according to equation (3.31) of Silverman (1986). S&P500 closing prices and the simulated process $\{P_t\}_{t=1}^T$ have similar density functions.

However, this is not the case for the daily returns. Figures 1.7 and 1.8 depict the FTSE100 and S&P500 continuously compounded daily returns, $\{y_{FTSE100,t}\}_{t=1}^T$ and

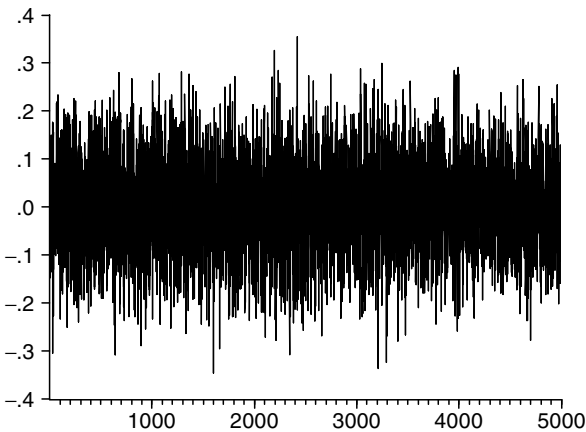


Figure 1.2 Simulated $\{y_t\}$ process, where $y_t = 100 \log(P_t/P_{t-1})$, $P_t = P_{t-1} + \varepsilon_t$, $P_1 = 1000$ and $\varepsilon_t \stackrel{i.i.d.}{\sim} N(0, 1)$.

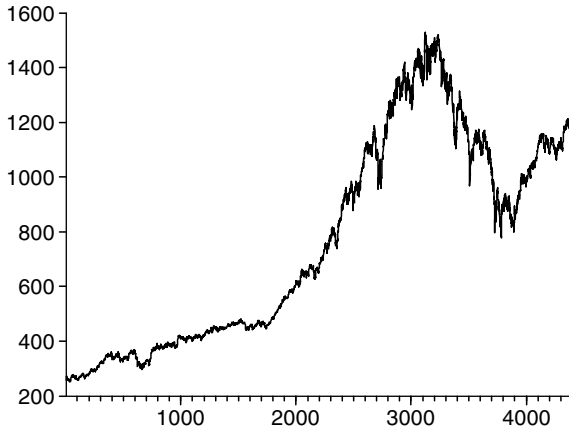


Figure 1.3 S&P500 equity index closing prices from 4 April 1988 to 5 April 2005.

$\{y_{SP500,t}\}_{t=1}^T$, while Figure 1.9 presents the autocorrelations of $\{y_t\}_{t=1}^T$, $\{y_{FTSE100,t}\}_{t=1}^T$ and $\{y_{SP500,t}\}_{t=1}^T$ for lags of order 1, \dots , 35. The 95% confidence interval for the estimated sample autocorrelation is given by $\pm 1.96/\sqrt{T}$, in the case of a process with independently and identically normally distributed components. The autocorrelations of the FTSE100 and the S&P500 daily returns differ from those of the simulated process. In both cases, more than 5% of the estimated autocorrelations are outside the above 95% confidence interval. Visual inspection of Figures 1.7 and 1.8 shows clearly that the mean is constant, but the variance keeps changing over time, so the return series does not appear to be a sequence of i.i.d. random variables. A characteristic of asset returns, which is noticeable from the figures, is the volatility clustering first noted by

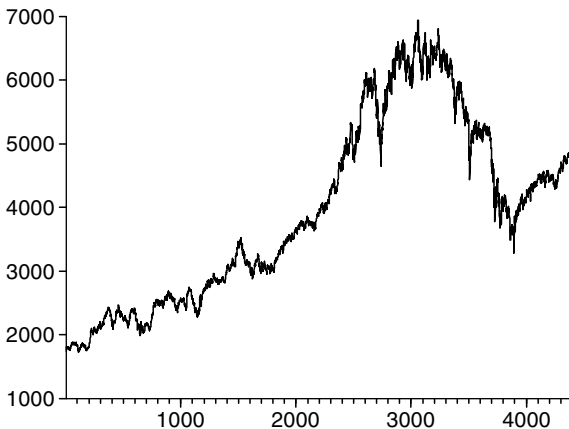


Figure 1.4 FTSE100 equity index closing prices from 4 April 1988 to 5 April 2005.

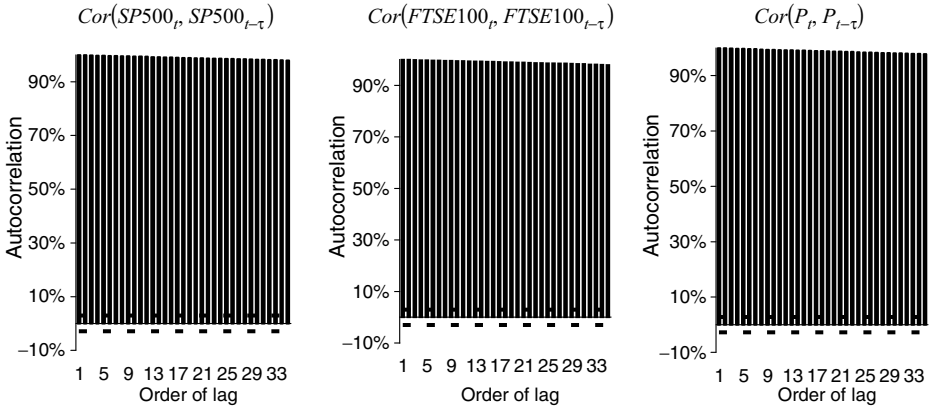


Figure 1.5 Autocorrelation of the S&P500 and the FTSE100 closing prices and of the simulated process $\{P_t\}_{t=1}^T$, for $\tau = 1(1)35$ lags. Dashed lines present the 95% confidence interval for the estimated sample autocorrelations given by $\pm 1.96/\sqrt{T}$.

Mandelbrot (1963): ‘Large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes’. Fama (1970) also observed the alternation between periods of high and low volatility: ‘Large price changes are followed by large price changes, but of unpredictable sign’.

Figure 1.10 presents the histograms of the stock market series. Asset returns are highly peaked (leptokurtic) and slightly asymmetric, a phenomenon observed by Mandelbrot (1963):

The empirical distributions of price changes are usually too peaked to be relative to samples from Gaussian populations . . . the histograms of price changes are indeed unimodal and their central bells [recall] the Gaussian ogive. But, there are typically so many outliers that ogives fitted to the mean square of price changes are much lower and flatter than the distribution of the data themselves.

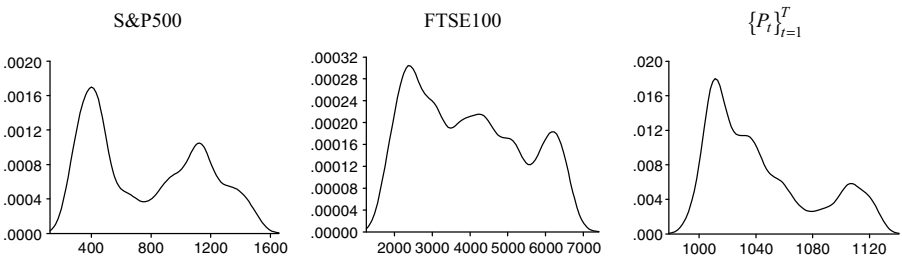


Figure 1.6 Density estimate of the S&P500 and FTSE100 closing prices, and of the simulated process $\{P_t\}_{t=1}^T$.

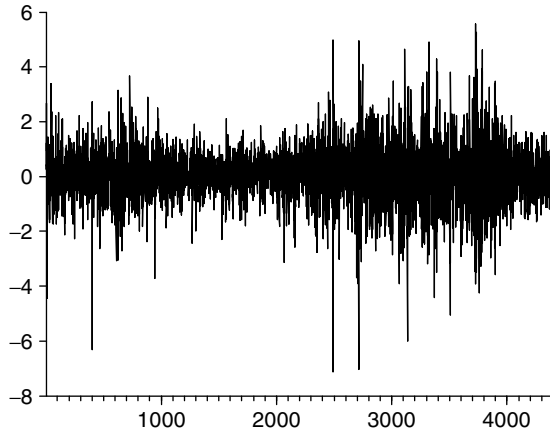


Figure 1.7 S&P500 equity index continuously compounded daily returns from 5 April 1988 to 5 April 2005.

According to Table 1.1, for estimated kurtosis² equal to 7.221 (or 6.241) and an estimated skewness³ equal to -0.162 (or -0.117), the distribution of returns is flat (platykurtic) and has a long left tail relative to the normal distribution. The Jarque and Bera (1980, 1987) test is usually used to test the null hypothesis that the series is normally distributed. The test statistic measures the size of the difference between the skewness, Sk , and kurtosis, Ku , of the series and those of the normal distribution. It is computed as $JB = T(Sk^2 + ((Ku-3)^2/4))/6$, where T is the number of observations. Under the null hypothesis of a normal distribution, the JB statistic is χ^2 distributed

² Kurtosis is a measure of the degree of peakedness of a distribution of values, defined in terms of a normalized form of its fourth central moment by μ_4/μ_2^2 (it is in fact the expected value of quartic standardized scores) and estimated by

$$Ku = T \frac{\sum_{t=1}^T (y_t - \bar{y})^4}{\left(\sum_{t=1}^T (y_t - \bar{y})^2 \right)^2},$$

where T is the number of observations and \bar{y} is the sample mean, $\bar{y} = \sum_{t=1}^T y_t / T$. The normal distribution has a kurtosis equal to 3 and is called *mesokurtic*. A distribution with a kurtosis greater than 3 has a higher peak and is called *leptokurtic*, while a distribution with a kurtosis less than 3 has a flatter peak and is called *platykurtic*. Some writers talk about *excess kurtosis*, whereby 3 is deducted from the kurtosis so that the normal distribution has an excess kurtosis of 0 (see Alexander, 2008, p. 82).

³ Skewness is a measure of the degree of asymmetry of a distribution, defined in terms of a normalized form of its third central moment of a distribution by $\mu_3/\mu_2^{3/2}$ (it is in fact the expected value of cubed standardized scores) and estimated by

$$Sk = \sqrt{T} \frac{\sum_{t=1}^T (y_t - \bar{y})^3}{\left(\sum_{t=1}^T (y_t - \bar{y})^2 \right)^{3/2}}.$$

The normal distribution has a skewness equal to 0. A distribution with a skewness greater than 0 has a longer right tail is described as *skewed to the right*, while a distribution with a skewness less than 0 has a longer left tail and is described as *skewed to the left*.

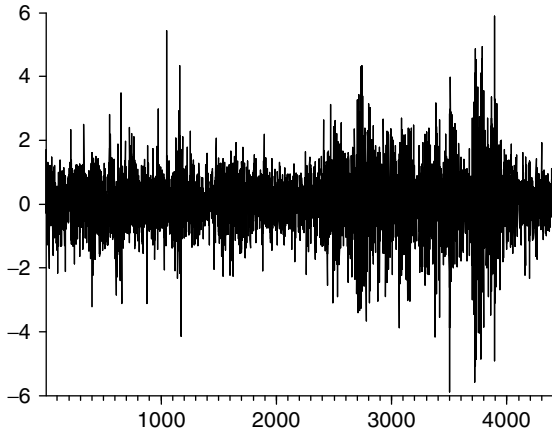


Figure 1.8 FTSE100 equity index continuously compounded daily returns from 5 April 1988 to 5 April 2005.

with 2 degrees of freedom. The two return series were tested for normality using *JB* resulting in a *p*-value that was practically zero, thus signaling non-validity of the hypothesis. Due to the fact that the *JB* statistic frequently rejects the hypothesis of normality, especially in the presence of serially correlated observations, a series of more powerful test statistics (e.g. the Anderson–Darling and the Cramér–von Mises statistics) were also computed with similar results. A detailed discussion of the computation of the aforementioned test statistics is given in Section 3.3.

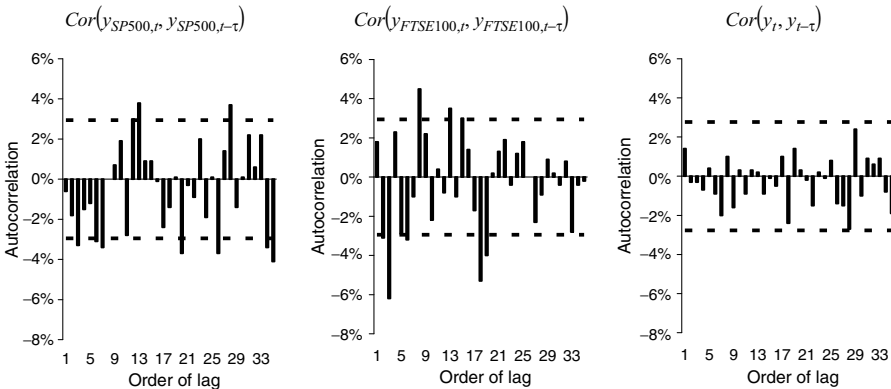


Figure 1.9 Autocorrelation of the S&P500 and FTSE100 continuously compounded daily returns and of the simulated process $\{y_t\}_{t=1}^T$, for $\tau = 1(1)35$ lags. Dashed lines present the 95% confidence interval for the estimated sample autocorrelations given by $\pm 1.96/\sqrt{T}$.

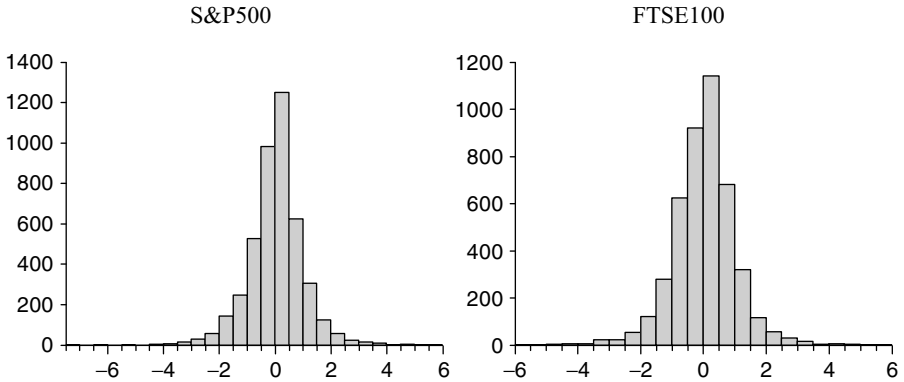


Figure 1.10 Histogram of the S&P500 and FTSE100 index log-returns.

In the 1960s and 1970s, the regularity of leptokurtosis led to a literature on modelling asset returns as i.i.d. random variables having some thick-tailed distribution (Blattberg and Gonedes, 1974; Clark, 1973; Hagerman, 1978; Mandelbrot, 1963, 1964; Officer, 1972; Praetz, 1972). These models, although able to capture the leptokurtosis, could not account for the existence of non-linear temporal dependence such as volatility clustering observed from the data. For example, applying an autoregressive model to remove the linear dependence from an asset returns series and testing the residuals for a higher-order dependence using the Brock–Dechert–Scheinkman (BDS) test (Brock et al., 1987, 1991, 1996), the null hypothesis of i.i.d. residuals was rejected.

Table 1.1 Descriptive statistics of the S&P500 and the FTSE100 equity index returns

	S&P500	FTSE100
Mean	0.034%	0.024%
Standard deviation	15.81%	15.94%
Skewness	-0.162	-0.117
Kurtosis	7.221	6.241
Jarque–Bera	3312.9	1945.6
[<i>p</i> -value]	[0.00]	[0.00]
Anderson–Darling	44.3	28.7
[<i>p</i> -value]	[0.00]	[0.00]
Cramér–von Mises	8.1	4.6
[<i>p</i> -value]	[0.00]	[0.00]

The annualized standard deviation is computed by multiplying the standard deviation of daily returns by $252^{1/2}$, the square root of the number of trading days per year. The Jarque–Bera, Anderson–Darling and Cramér–von Mises statistics test the null hypothesis that the daily returns are normally distributed.

1.2 The autoregressive conditionally heteroscedastic process

Autoregressive conditional heteroscedasticity (ARCH) models have been widely used in financial time series analysis and particularly in analysing the risk of holding an asset, evaluating the price of an option, forecasting time-varying confidence intervals and obtaining more efficient estimators under the existence of heteroscedasticity.

Before we proceed to the definition of the ARCH model, let us simulate a process able to capture the volatility clustering of asset returns. Assume that the true data-generating process of continuously compounded returns, y_t , has a fully unpredictable conditional mean and a time-varying conditional variance:

$$\begin{aligned} y_t &= \varepsilon_t, \\ \varepsilon_t &= z_t \sqrt{a_0 + a_1 \varepsilon_{t-1}^2}, \end{aligned} \tag{1.2}$$

where $z_t \stackrel{i.i.d.}{\sim} N(0, 1)$, z_t is independent of ε_t , $a_0 > 0$ and $0 < a_1 < 1$. The unconditional mean of y_t is $E(y_t) = E(z_t)E(\sqrt{a_0 + a_1 \varepsilon_{t-1}^2}) = 0$, as $E(z_t) = 0$ and z_t and ε_{t-1} are independent of each other. The conditional mean of y_t given the lag values of ε_t is $E(y_t | \varepsilon_{t-1}, \dots, \varepsilon_1) = 0$. The unconditional variance is $V(y_t) = E(\varepsilon_t^2) - E(\varepsilon_t)^2 = E(z_t^2(a_0 + a_1 \varepsilon_{t-1}^2)) = a_0 + a_1 E(\varepsilon_{t-1}^2)$. As $E(\varepsilon_t^2) = E(\varepsilon_{t-1}^2)$, $V(y_t) = a_0(1 - a_1)^{-1}$. The conditional variance is $V(y_t | \varepsilon_{t-1}, \dots, \varepsilon_1) = E(\varepsilon_t^2 | \varepsilon_{t-1}, \dots, \varepsilon_1) = a_0 + a_1 \varepsilon_{t-1}^2$. The kurtosis of the unconditional distribution equals $E(\varepsilon_t^4) / E(\varepsilon_t^2)^2 = 3(1 - a_1^2) / (1 - 3a_1^2)$. Note that the kurtosis exceeds 3 for $a_1 > 0$ and diverges if a_1 approaches $\sqrt{1/3}$. Figure 1.11 plots the unconditional kurtosis for $0 \leq a_1 < 1/\sqrt{3}$.

Both the unconditional and conditional means and the unconditional variance of asset returns remain constant, but the conditional variance has a time-varying character as it depends on the previous values of ε_t . Let us consider equation (1.2) as the true data-generating function and produce 5000 values of y_t . Figure 1.12

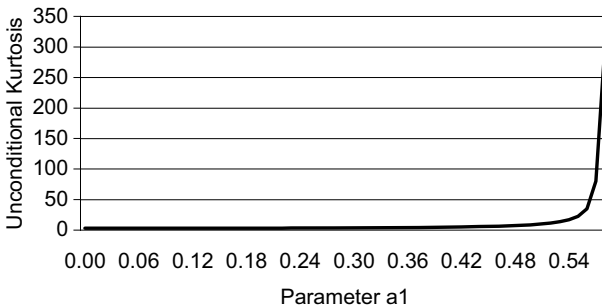


Figure 1.11 The unconditional kurtosis of $\varepsilon_t = z_t \sqrt{a_0 + a_1 \varepsilon_{t-1}^2}$ for $0 \leq a_1 < 1/\sqrt{3}$ and $a_0 > 0$.

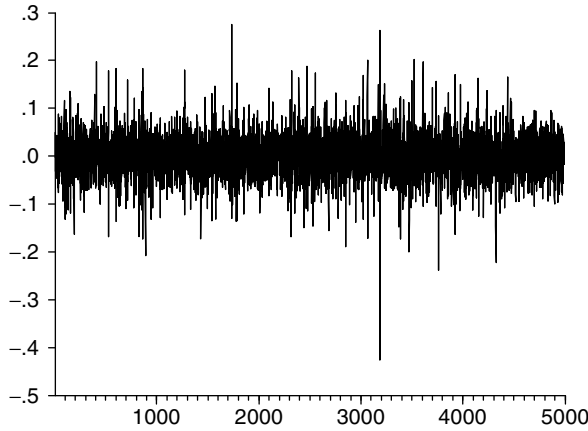


Figure 1.12 Simulated $\{y_t\}$ process, where $y_t = \varepsilon_t$, $\varepsilon_t = z_t \sqrt{a_0 + a_1 \varepsilon_{t-1}^2}$, $z_t \stackrel{i.i.d.}{\sim} N(0, 1)$, $\varepsilon_0 = 0$, $a_0 = 0.001$ and $a_1 = 0.55$.

presents the simulated series y_t for $\varepsilon_0 = 0$, $a_0 = 0.001$ and $a_1 = 0.55$. Equation (1.2) produces a series with a time-varying conditional variance $V(y_t | \varepsilon_{t-1}, \dots, \varepsilon_1) = 0.001 + 0.55 \varepsilon_{t-1}^2$. Because of the non-constant variance, the series is highly leptokurtic. Figure 1.13 presents both the histogram (the series is obviously highly leptokurtic) and the conditional variance (revealing volatility clustering) of the simulated process. On the other hand, the y_t series standardized by its conditional standard deviation $y_t / \sqrt{V(y_t | \varepsilon_{t-1}, \dots, \varepsilon_1)}$ is essentially the $z_t \stackrel{i.i.d.}{\sim} N(0, 1)$ series. Therefore, the series itself is leptokurtic and has a non-constant variance, but its standardized version has a kurtosis equal to 3 and constant variance. Table 1.2

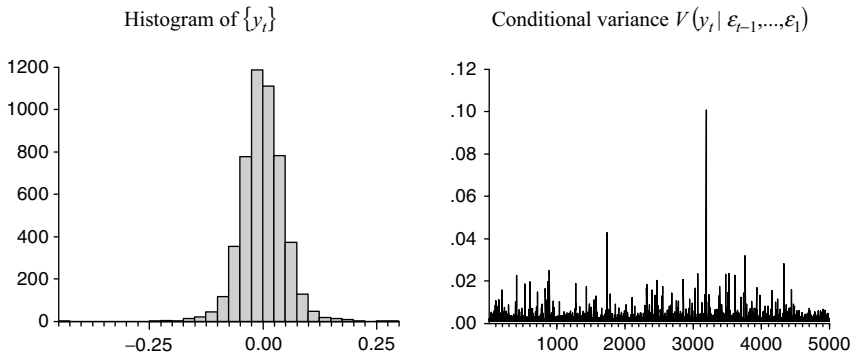


Figure 1.13 Histogram of the simulated $\{y_t\}$ process and conditional variance $V(y_t | \varepsilon_{t-1}, \dots, \varepsilon_1) = a_0 + a_1 \varepsilon_{t-1}^2$, where $y_t = \varepsilon_t$, $\varepsilon_t = z_t \sqrt{a_0 + a_1 \varepsilon_{t-1}^2}$, $z_t \stackrel{i.i.d.}{\sim} N(0, 1)$, $\varepsilon_0 = 0$, $a_0 = 0.001$ and $a_1 = 0.55$.

Table 1.2 Descriptive statistics of the simulated y_t and z_t series, where $y_t = \varepsilon_t$, $\varepsilon_t = z_t \sqrt{a_0 + a_1 \varepsilon_{t-1}^2}$, $z_t \sim N(0, 1)$, $\varepsilon_0 = 0$, $a_0 = 0.001$ and $a_1 = 0.55$

	y_t	z_t
Mean	0.0006	0.0109
Standard deviation	0.046	0.999
Skewness	-0.091	0.014
Kurtosis	6.387	2.973
Jarque–Bera	2395.7	0.3
[p -value]	[0.00]	[0.86]
Anderson–Darling	12.3	0.2
[p -value]	[0.00]	[0.87]
Cramér–von Mises	1.69	0.04
[p -value]	[0.00]	[0.77]

presents the descriptive statistics of the y_t and z_t series. The kurtosis of the simulated y_t series is 6.387.

In the sequel, we provide a formal definition of the ARCH process.

Let $\{y_t(\theta)\}$ refer to the univariate discrete time real-valued stochastic process to be predicted (e.g. the rate of return of a particular stock or market portfolio from time $t-1$ to t), where θ is a vector of unknown parameters and $E(y_t(\theta)|I_{t-1}) \equiv E_{t-1}(y_t(\theta)) \equiv \mu_t(\theta)$ denotes the conditional mean given the information set I_{t-1} available at time $t-1$. The innovation process for the conditional mean, $\{\varepsilon_t(\theta)\}$, is then given by $\varepsilon_t(\theta) = y_t(\theta) - \mu_t(\theta)$ with corresponding unconditional variance $V(\varepsilon_t(\theta)) = E(\varepsilon_t^2(\theta)) \equiv \sigma^2(\theta)$, zero unconditional mean and $E(\varepsilon_t(\theta)\varepsilon_s(\theta)) = 0$, $\forall t \neq s$. The conditional variance of the process given I_{t-1} is defined by $V(y_t(\theta)|I_{t-1}) \equiv V_{t-1}(y_t(\theta)) \equiv E_{t-1}(\varepsilon_t^2(\theta)) \equiv \sigma_t^2(\theta)$. Since investors would know the information set I_{t-1} when they make their investment decisions at time $t-1$, the relevant expected return to the investors and volatility are $\mu_t(\theta)$ and $\sigma_t^2(\theta)$, respectively. An ARCH process, $\{\varepsilon_t(\theta)\}$, can be represented as:

$$\begin{aligned}
 \varepsilon_t(\theta) &= z_t \sigma_t(\theta), \\
 z_t &\stackrel{i.i.d.}{\sim} f(w; 0, 1), \\
 \sigma_t^2(\theta) &= g(\sigma_{t-1}(\theta), \sigma_{t-2}(\theta), \dots; \varepsilon_{t-1}(\theta), \varepsilon_{t-2}(\theta), \dots; v_{t-1}, v_{t-2}, \dots),
 \end{aligned} \tag{1.3}$$

where $E(z_t) = 0$, $V(z_t) = 1$, $f(\cdot)$ is the density function of z_t , w is the vector of the parameters of f to be estimated, $\sigma_t(\theta)$ is a time-varying, positive and measurable function of the information set at time $t-1$, v_t is a vector of predetermined variables included in I_t , and $g(\cdot)$ is a linear or non-linear functional form of I_{t-1} . By definition, $\varepsilon_t(\theta)$ is serially uncorrelated with mean zero, but with a time-varying conditional variance equal to $\sigma_t^2(\theta)$. The conditional variance is a linear or non-linear function of lagged values of σ_t and ε_t , and predetermined variables (v_{t-1}, v_{t-2}, \dots) included in

I_{t-1} . For example, a simple form of the conditional variance could be

$$\sigma_t^2 = a_0 + a_1 \varepsilon_{t-1}^2, \quad (1.4)$$

which will be referred to as ARCH(1) in the next chapter. In the sequel, for notational convenience, no explicit indication of the dependence on the vector of parameters, θ , is given when obvious from the context.

Since very few financial time series have a constant conditional mean of zero, an ARCH model can be presented in a regression form by letting ε_t be the innovation process in a linear regression:

$$\begin{aligned} y_t &= x_t' \beta + \varepsilon_t, \\ \varepsilon_t | I_{t-1} &\sim f(w; 0, \sigma_t^2), \\ \sigma_t^2 &= g(\sigma_{t-1}(\theta), \sigma_{t-2}(\theta), \dots; \varepsilon_{t-1}(\theta), \varepsilon_{t-2}(\theta), \dots; v_{t-1}, v_{t-2}, \dots), \end{aligned} \quad (1.5)$$

where x_t is a $k \times 1$ vector of endogenous and exogenous explanatory variables included in the information set I_{t-1} and β is a $k \times 1$ vector of unknown parameters.

Of course, an ARCH process is not necessarily limited to an expression for the residuals of a regression model. The dependent variable, $\{y_t\}$, can be decomposed into two parts, the predictable component, μ_t , and the unpredictable component, ε_t . Hence, another general representation of an ARCH process is:

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \\ \mu_t &= \mu(\theta | I_{t-1}), \\ \varepsilon_t &= \sigma_t z_t, \\ \sigma_t &= g(\theta | I_{t-1}), \\ z_t &\stackrel{i.i.d.}{\sim} f(0, 1; w). \end{aligned} \quad (1.6)$$

Here $\mu(\theta | I_{t-1})$ and $g(\theta | I_{t-1})$ denote the functional forms of the conditional mean μ_t and the conditional standard deviation σ_t , respectively signifying that μ_t and σ_t are conditioned on the information available up to time I_{t-1} and depending on the parameter vector θ . The predictable component can be any linear or non-linear functional form of estimation. The conditional mean can also be expressed as a function of the conditional variance (see Section 1.6 for a detailed explanation). The form most frequently considered is an autoregressive moving average, or an ARMA(κ, l), process:

$$y_t = c_1 y_{t-1} + c_2 y_{t-2} + \dots + c_\kappa y_{t-\kappa} + \varepsilon_t + d_1 \varepsilon_{t-1} + d_2 \varepsilon_{t-2} + \dots + d_l \varepsilon_{t-l}, \quad (1.7)$$

or

$$\left(1 - \sum_{i=1}^{\kappa} c_i L^i\right) y_t = \left(1 + \sum_{i=1}^l d_i L^i\right) \varepsilon_t, \quad (1.8)$$

where L is the lag operator.⁴ Use of the AR(1) or MA(1) models is mainly imposed by the non-synchronous trading effect (see Section 1.5). The ARMA(κ, l) model can be

⁴That is, $L^2 y_t \equiv y_{t-2}$, $\sum_{i=1}^3 (L^i) y_t = y_{t-1} + y_{t-2} + y_{t-3}$.

expanded to take into account the explanatory power of a set of exogenous variables that belong to I_{t-1} . In this case we may refer to an ARMA model with exogenous variables, or ARMAX(κ, l):

$$C(L)(y_t - x_t' \beta) = D(L)\varepsilon_t, \quad (1.9)$$

where $C(L) = (1 - \sum_{i=1}^{\kappa} c_i L^i)$ and $D(L) = (1 + \sum_{i=1}^l d_i L^i)$. If the series under study, $\{y_t\}$, exhibits long-term dependence, it is best modelled by a fractionally integrated ARMAX, or ARFIMAX(κ, \tilde{d}, l) model,

$$C(L)(1-L)^{\tilde{d}}(y_t - x_t' \beta) = D(L)\varepsilon_t, \quad (1.10)$$

where $(1-L)^{\tilde{d}}$ is the fractional differencing operator and $\tilde{d} \in (-0.5, 0.5)$ is the fractional differencing parameter.⁵ The ARFIMAX(κ, \tilde{d}, l) specification was introduced by Granger (1980) and Granger and Joyeux (1980).

There is a plethora of formulations of the conditional mean in the literature. For example, Sarkar (2000) illustrated the ARCH model for a regression model in which the dependent variable is Box–Cox transformed. He referred to this as the Box–Cox transformed ARCH, or BCARCH, model:

$$\begin{aligned} y_t^{(BC)} &= \begin{cases} (y_t^\lambda - 1)\lambda^{-1}, & \lambda \neq 0, \\ \log(y_t), & \lambda = 0, \end{cases} \\ y_t^{(BC)} | I_{t-1} &\sim N(x_t' \beta, \sigma_t^2), \\ \varepsilon_t &= y_t^{(BC)} - x_t' \beta, \\ \sigma_t^2 &= g(\theta | I_{t-1}). \end{aligned} \quad (1.11)$$

Bickel and Doksum's (1981) extended form of the Box–Cox transformation allows negative values of y_t .

Feng et al. (2007) proposed an estimation method for semi-parametric fractional autoregressive (SEMIFAR) models with ARCH errors. A SEMIFAR ARCH model can be written as:

$$\begin{aligned} C(L)(1-L)^{\tilde{d}} \left((1-L)^{\tilde{d}} y_t - c(t/T) \right) &= D(L)\varepsilon_t, \\ \varepsilon_t &= z_t \sigma_t, \\ z_t &\sim N(0, 1), \\ \sigma_t^2 &= a_0 + a_1 \varepsilon_{t-1}^2, \end{aligned} \quad (1.12)$$

where t/T is the rescaled time, $c(\cdot)$ is a smooth function on $[0, 1]$, $\tilde{d} \in (-0.5, 0.5)$ is the fractional differencing parameter, and $\tilde{d} \in \{0, 1\}$ is the integer differencing operator. Beran (1995) first noted that the SEMIFAR class of models, is appropriate for simultaneous modelling of deterministic trends, difference stationarity and stationarity with short- and long-range dependence.

If one wishes to create a specification in which the conditional mean is described by an autoregressive model of order 1, the conditional variance is expressed by the simulated data-generating process in (1.4) and the standardized residuals are normally distributed, one may construct a model of the form:

⁵ More information about fractionally integrated specification can be found in Chapter 3.

$$\begin{aligned}
y_t &= c_1 y_{t-1} + \varepsilon_t, \\
\varepsilon_t &= z_t \sigma_t, \\
z_t &\sim N(0, 1), \\
\sigma_t^2 &= a_0 + a_1 \varepsilon_{t-1}^2.
\end{aligned}
\tag{1.13}$$

In this case, the vector of unknown parameters under estimation is $\theta' = (c_1, a_0, a_1)$. The elements of θ are now estimated simultaneously for the conditional mean and variance. However, the conditional mean can also be estimated separately. Artificial neural networks,⁶ chaotic dynamical systems,⁷ non-linear parametric and non-parametric models⁸ are some examples from the literature dealing with conditional mean predictions. One can apply any model in estimating the conditional mean and may subsequently proceed to the estimation of the conditional variance (see, for example, Feng et al., 2007;⁹ Pagan and Schwert, 1990¹⁰).

The rest of this chapter looks at the influence that various factors have on financial time series and in particular at the leverage effect, the non-trading period effect, the non-synchronous trading effect, the relationship between investors' expected return and risk, and the inverse relation between volatility and serial correlation.

1.3 The leverage effect

Black (1976) first noted that changes in stock returns often display a tendency to be negatively correlated with changes in returns volatility, i.e., volatility tends to rise in response to bad news and to fall in response to good news. This phenomenon is termed the *leverage effect* and can only be partially interpreted by fixed costs such as financial and operating leverage (see Black, 1976; Christie, 1982).

⁶ For an overview of the neural networks literature, see Poggio and Girosi (1990), Hertz et al. (1991), White (1992) and Hutchinson et al. (1994). Thomaidis and Dounias (2008) proposed a neural network parameterization for the mean and a linear ARCH parameterization for the variance. Plasmans et al. (1998) and Franses and Homelen (1998) investigated the ability of neural networks to forecast exchange rates. They supported the view that the non-linearity found in exchange rates is due to ARCH effects, so that no gain in forecasting accuracy is obtained by using a neural network. Saltoglu (2003) investigated the ability of neural networks to forecast interest rates and noted the importance of modelling both the first and second moments jointly. Jasic and Wood (2004) and Perez-Rodriguez et al. (2005) provided evidence that neural network models have a superior ability compared to other model frameworks in predicting stock indices.

⁷ Brock (1986), Holden (1986), Thompson and Stewart (1986) and Hsieh (1991) review applications of chaotic systems to financial markets. Adrangi and Chatrath (2003) found that the non-linearities in commodity prices are not consistent with chaos, but they are explained by an ARCH process. On the other hand, Barkoulas and Travlos (1998) mentioned that even after accounting for the ARCH effect, the chaotic structure of the Greek stock market is consistent with the evidence.

⁸ Priestley (1988), Tong (1990) and Teräsvirta et al. (1994) cover a wide variety of non-linear models. Applications of SETAR and ARFIMA models can be found in Peel and Speight (1996) and Barkoulas et al. (2000), respectively.

⁹ They proposed the estimation of the conditional mean and conditional variance parameters of the SEMIFAR ARCH model separately.

¹⁰ For example, a two-step estimator of conditional variance with ordinary least squares is a consistent but inefficient estimator.

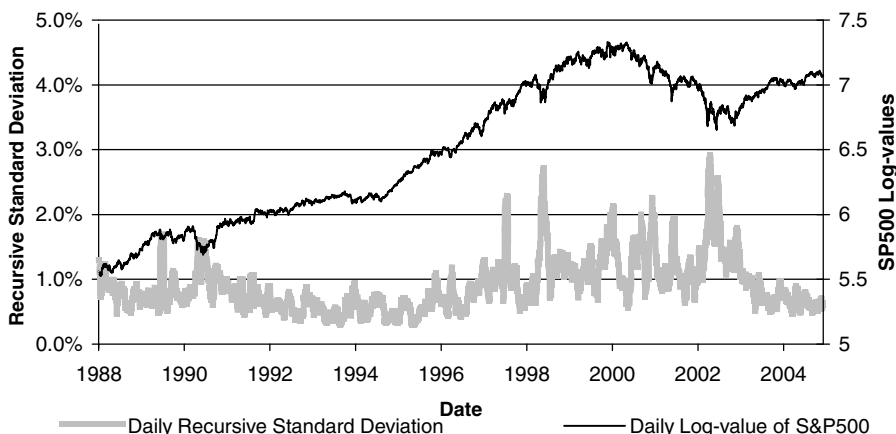


Figure 1.14 Daily log-values and recursive standard deviation of returns for the S&P500 equity index.

We can observe the phenomenon of leverage by plotting the market prices and their volatility. As a naive estimate of volatility at day t , the standard deviation of the 22 most recent trading days,

$$\sigma_t^{(22)} = \sqrt{\frac{\sum_{i=t-22}^t \left(y_i - \left(\sum_{i=t-22}^t y_i / 22 \right) \right)^2}{22}}, \tag{1.14}$$

is used. Figures 1.14 and 1.15 plot daily log-values of stock market indices and the relevant standard deviations of the continuously compounded returns. The periods

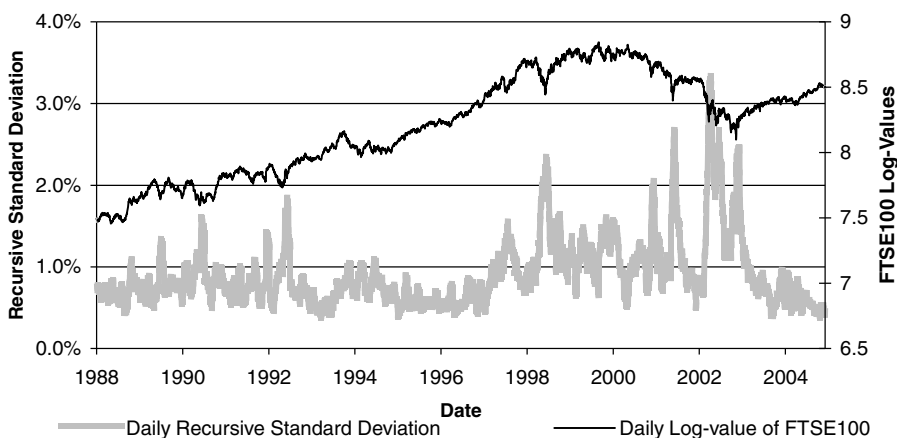


Figure 1.15 Daily log-values and recursive standard deviations of returns for the FTSE100 equity index.

Table 1.3 Mean and annualized standard deviation of the S&P500 and the FTSE100 equity index returns

	Monday	Tuesday	Wednesday	Thursday	Friday
S&P500					
Mean	0.061%	0.045%	0.045%	0.001%	0.021%
Std deviation	16.27%	16.13%	14.92%	15.59%	16.13%
Skewness	-0.895	0.315	0.450	0.153	-0.663
Kurtosis	10.494	6.039	6.332	5.767	7.017
FTSE100					
Mean	-0.014%	0.052%	-0.008%	0.033%	0.053%
Std deviation	16.60%	15.81%	14.94%	16.35%	15.93%
Skewness	-0.201	-0.200	-0.256	0.104	-0.069
Kurtosis	6.677	6.433	5.194	6.724	5.751

Annualized standard deviation is computed by multiplying the standard deviation of daily returns by $252^{1/2}$, the square root of the number of trading days per year.

of market drops are characterized by a large increase in volatility. The leverage effect is captured by a class of models that are explored in Section 2.4.

1.4 The non-trading period effect

Financial markets appear to be affected by the accumulation of information during non-trading periods, as reflected in the prices when the markets reopen following a close. As a result, the variance of returns displays a tendency to increase. This is known as the *non-trading period effect*. It is worth noting that the increase in the variance of returns is not nearly proportional to the market close duration, as would be anticipated if the information accumulation rate were constant over time. In fact, as Fama (1965) and French and Roll (1986) observed, information accumulates at a lower rate when markets are closed than when they are open. Also, as reflected by the findings of French and Roll (1986) and Baillie and Bollerslev (1989), the returns variance tends to be higher following weekends and holidays than on other days, but not by as much as it would be under a constant news arrival rate. Table 1.3 shows the annualized standard deviations of stock market returns for each day for the FTSE100 and S&P500 indices. In both cases, the standard deviation on Mondays is higher than on other days. The non-trading period effect is explored in the empirical example of Section 5.4.

1.5 The non-synchronous trading effect

The fact that the values of time series are often taken to have been recorded at time intervals of one length when in fact they were recorded at time intervals of another,

not necessarily regular, length is an important factor affecting the return series, an effect known as the *non-synchronous trading effect*. For details, see Campbell et al. (1997, p. 84) and Tsay (2002, p. 176). For example, the daily prices of securities usually analysed are the closing prices. The closing price of a security is the price at which the last transaction occurred. The last transaction of each security is not implemented at the same time each day. So, it is falsely assumed that the daily prices are equally spaced at 24-hour intervals. The importance of non-synchronous trading was first recognized by Fisher (1966) and further developed by many researchers such as Atchison et al. (1987), Cohen et al. (1978), Cohen et al. (1979, 1983), Dimson (1979), Lo and MacKinlay (1988, 1990a, 1990b) and Scholes and Williams (1977).

Non-synchronous trading in the stocks making up an index induces autocorrelation in the return series, primarily when high-frequency data are used. To control this, Scholes and Williams (1977) suggested a first-order moving average (MA(1)) form for index returns, while Lo and MacKinlay (1988) suggested a first order autoregressive (AR(1)) form. According to Nelson (1991), 'as a practical matter, there is little difference between an AR(1) and an MA(1) when the AR and MA coefficients are small and the autocorrelations at lag one are equal, since the higher-order autocorrelations die out very quickly in the AR model'.

1.6 The relationship between conditional variance and conditional mean

1.6.1 The ARCH in mean model

Financial theory suggests that an asset with a higher expected risk would pay a higher return on average. Let y_t denote the rate of return of a particular stock or market portfolio from time t to time $t-1$ and rf_t be the return on a riskless asset (e.g. treasury bills). Then, the excess return (asset return minus the return on a riskless asset) can be decomposed into a component anticipated by investors at time $t-1$, μ_t , and a component that was unanticipated, ε_t :

$$y_t - rf_t = \mu_t + \varepsilon_t. \quad (1.15)$$

The relationship between investors' expected return and risk was presented in an ARCH framework by Engle et al. (1987). They introduced the *ARCH in mean*, or ARCH-M, model where the conditional mean is an explicit function of the conditional variance of the process in framework (1.5). The estimated coefficient on the expected risk is a measure of the risk–return trade-off. Thus, the ARCH regression model in framework (1.5) can be written as

$$\begin{aligned} y_t &= x_t' \beta + \varphi(\sigma_t^2) + \varepsilon_t, \\ \varepsilon_t | I_{t-1} &\sim f(0, \sigma_t^2), \\ \sigma_t^2 &= g(\sigma_{t-1}, \sigma_{t-2}, \dots; \varepsilon_{t-1}, \varepsilon_{t-2}, \dots; v_{t-1}, v_{t-2}, \dots), \end{aligned} \quad (1.16)$$

where, as before, x_t is a $k \times 1$ vector of endogenous and exogenous explanatory variables included in the information set I_{t-1} and $\phi(\sigma_t^2)$ represents the risk premium, i.e., the increase in the expected rate of return due to an increase in the variance of the return.¹¹ Although earlier studies concentrated on detecting a constant risk premium, the ARCH-M model provided a new approach by which a time-varying risk premium could be estimated. The most commonly used specifications of the ARCH-M model are in the form

$$\phi(\sigma_t^2) = c_0 + c_1 \sigma_t^2 \quad (1.17)$$

(Nelson, 1991; Bollerslev et al., 1994),

$$\phi(\sigma_t^2) = c_0 + c_1 \sigma_t \quad (1.18)$$

(Domowitz and Hakkio, 1985; Bollerslev et al., 1988), or

$$\phi(\sigma_t^2) = c_0 + c_1 \log(\sigma_t^2) \quad (1.19)$$

(Engle et al., 1987). A positive as well as a negative risk–return trade-off could be consistent with the financial theory. A positive relationship is expected if we assume a rational risk-averse investor who requires a larger risk premium during the times when the payoff of the security is riskier. On the other hand, a negative relationship is expected under the assumption that during relatively riskier periods the investors may want to save more. In applied research work, there is evidence for both positive and negative relationships. The relationship between the investor’s expected return and risk is explored in the empirical example of Section 5.4. French et al. (1987) found a positive risk–return trade-off for the excess returns on the S&P500 composite portfolio, albeit not statistically significant in all the periods examined. Nelson (1991) observed a negative but insignificant relationship for the excess returns on the Center for Research in Security Prices (CRSP) value-weighted market index. Bollerslev et al. (1994) noted a positive, but not always statistically significant, relationship for the returns on the Dow Jones and S&P500 indices. Interesting studies employing the ARCH-M model were conducted by Devaney (2001) and Elyasiani and Mansur (1998). The former examined the trade-off between conditional variance and excess returns for commercial bank sector stocks, while the latter investigated the time-varying risk premium for real estate investment trusts. Theoretical results on the moment structure of the ARCH-M model were investigated by Arvanitis and Demos (2004) and Demos (2002).

¹¹ x_t is a vector of explanatory variables. The risk free rate of return, rf_t , could be incorporated as an explanatory variable, e.g. $x_t' = (rf_t)$ and $\beta = (1)$, or $rf_t + \mu_t = x_t' \beta + \phi(\sigma_t^2) = (rf_t, \dots)(1, \dots)'$.

1.6.2 Volatility and serial correlation

LeBaron (1992) noted a strong inverse relationship between volatility and serial correlation for the returns of the S&P500 index, the CRSP value-weighted market index, the Dow Jones and the IBM returns. He introduced the *exponential autoregressive GARCH*, or EXP-GARCH(p, q), model in which the conditional mean is a non-linear function of the conditional variance. Based on LeBaron (1992), the ARCH regression model, in framework (1.5), can be written as:

$$\begin{aligned} y_t &= x_t' \beta + (c_1 + c_2 \exp(-\sigma_t^2/c_3))y_{t-1} + \varepsilon_t, \\ \varepsilon_t | I_{t-1} &\sim f(0, \sigma_t^2), \\ \sigma_t^2 &= g(\sigma_{t-1}, \sigma_{t-2}, \dots; \varepsilon_{t-1}, \varepsilon_{t-2}, \dots; v_{t-1}, v_{t-2}, \dots). \end{aligned} \tag{1.20}$$

The model is a mixture of the GARCH model and the exponential AR model of Ozaki (1980). For the data set LeBaron used, c_2 is significantly negative and remarkably robust to the choice of sample period, market index, measurement interval and volatility measure. Generally, the first-order autocorrelations are larger during periods of lower volatility and smaller during periods of higher volatility. The accumulation of news (see Section 1.4) and the non-synchronous trading (see Section 1.5) were mentioned as possible reasons. The stocks do not trade close to the end of the day and information arriving during this period is reflected in the next day's trading, inducing serial correlation. As new information reaches the market very slowly, traders' optimal action is to do nothing until enough information is accumulated. Because of the non-trading, the trading volume, which is strongly positive related with volatility, falls. Thus, we have a market with low trade volume and high correlation. The inverse relationship between volatility and serial correlation is explored in the empirical example of Section 5.4.

Kim (1989), Sentana and Wadhvani (1991) and Oedegaard (1991) have also investigated the relationship between autocorrelation and volatility and found an inverse relationship between volatility and autocorrelation. Moreover, Oedegaard (1991) found that the evidence of autocorrelation, for the S&P500 daily index, decreased over time, possibly because of the introduction of financial derivatives (options and futures) on the index.