

Classical Time Series Models and Financial Series

The standard time series analysis rests on important concepts such as stationarity, autocorrelation, white noise, innovation, and on a central family of models, the autoregressive moving average (ARMA) models. We start by recalling their main properties and how they can be used. As we shall see, these concepts are insufficient for the analysis of financial time series. In particular, we shall introduce the concept of volatility, which is of crucial importance in finance.

In this chapter, we also present the main stylized facts (unpredictability of returns, volatility clustering and hence predictability of squared returns, leptokurticity of the marginal distributions, asymmetries, etc.) concerning financial series.

1.1 Stationary Processes

Stationarity plays a central part in time series analysis, because it replaces in a natural way the hypothesis of independent and identically distributed (iid) observations in standard statistics.

Consider a sequence of real random variables $(X_t)_{t \in \mathbb{Z}}$, defined on the same probability space. Such a sequence is called a time series, and is an example of a discrete-time stochastic process.

We begin by introducing two standard notions of stationarity.

Definition 1.1 (Strict stationarity) *The process (X_t) is said to be strictly stationary if the vectors $(X_1, \dots, X_k)'$ and $(X_{1+h}, \dots, X_{k+h})'$ have the same joint distribution, for any $k \in \mathbb{N}$ and any $h \in \mathbb{Z}$.*

The following notion may seem less demanding, because it only constrains the first two moments of the variables X_t , but contrary to strict stationarity, it requires the existence of such moments.

Definition 1.2 (Second-order stationarity) The process (X_t) is said to be second-order stationary if:

- (i) $EX_t^2 < \infty, \quad \forall t \in \mathbb{Z};$
- (ii) $EX_t = m, \quad \forall t \in \mathbb{Z};$
- (iii) $\text{Cov}(X_t, X_{t+h}) = \gamma_X(h), \quad \forall t, h \in \mathbb{Z}.$

The function $\gamma_X(\cdot)$ ($\rho_X(\cdot) := \gamma_X(\cdot)/\gamma_X(0)$) is called the autocovariance function (autocorrelation function) of (X_t) .

The simplest example of a second-order stationary process is white noise. This process is particularly important because it allows more complex stationary processes to be constructed.

Definition 1.3 (Weak white noise) The process (ϵ_t) is called weak white noise if, for some positive constant σ^2 :

- (i) $E\epsilon_t = 0, \quad \forall t \in \mathbb{Z};$
- (ii) $E\epsilon_t^2 = \sigma^2, \quad \forall t \in \mathbb{Z};$
- (iii) $\text{Cov}(\epsilon_t, \epsilon_{t+h}) = 0, \quad \forall t, h \in \mathbb{Z}, h \neq 0.$

Remark 1.1 (Strong white noise) It should be noted that no independence assumption is made in the definition of weak white noise. The variables at different dates are only uncorrelated and the distinction is particularly crucial for financial time series. It is sometimes necessary to replace hypothesis (iii) by the stronger hypothesis

(iii') the variables ϵ_t and ϵ_{t+h} are independent and identically distributed.

The process (ϵ_t) is then said to be strong white noise.

Estimating Autocovariances

The classical time series analysis is centered on the second-order structure of the processes. Gaussian stationary processes are completely characterized by their mean and their autocovariance function. For non-Gaussian processes, the mean and autocovariance give a first idea of the temporal dependence structure. In practice, these moments are unknown and are estimated from a realization of size n of the series, denoted $X_1, \dots, X_n$. This step is preliminary to any construction of an appropriate model. To estimate $\gamma(h)$, we generally use the sample autocovariance defined, for $0 \leq h < n$, by

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{j=1}^{n-h} (X_j - \bar{X})(X_{j+h} - \bar{X}) := \hat{\gamma}(-h),$$

where $\bar{X} = (1/n) \sum_{j=1}^n X_j$ denotes the sample mean. We similarly define the sample autocorrelation function by $\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0)$ for $|h| < n$.

The previous estimators have finite-sample bias but are asymptotically unbiased. There are other similar estimators of the autocovariance function with the same asymptotic properties (for instance, obtained by replacing $1/n$ by $1/(n-h)$). However, the proposed estimator is to be preferred over others because the matrix $(\hat{\gamma}(i-j))$ is positive semi-definite (see Brockwell and Davis, 1991, p. 221).

It is of course not recommended to use the sample autocovariances when h is close to n , because too few pairs (X_j, X_{j+h}) are available. Box, Jenkins and Reinsel (1994, p. 32) suggest that useful estimates of the autocorrelations can only be made if, approximately, $n > 50$ and $h \leq n/4$.

It is often of interest to know – for instance, in order to select an appropriate model – if some or all the sample autocovariances are significantly different from 0. It is then necessary to estimate the covariance structure of those sample autocovariances. We have the following result (see Brockwell and Davis, 1991, pp. 222, 226).

Theorem 1.1 (Bartlett's formulas for a strong linear process) *Let (X_t) be a linear process satisfying*

$$X_t = \sum_{j=-\infty}^{\infty} \phi_j \epsilon_{t-j}, \quad \sum_{j=-\infty}^{\infty} |\phi_j| < \infty,$$

where (ϵ_t) is a sequence of iid variables such that

$$E(\epsilon_t) = 0, \quad E(\epsilon_t^2) = \sigma^2, \quad E(\epsilon_t^4) = \kappa_\epsilon \sigma^4 < \infty.$$

Appropriately normalized, the sample autocovariances and autocorrelations are asymptotically normal, with asymptotic variances given by the Bartlett formulas:

$$\begin{aligned} \lim_{n \rightarrow \infty} n \text{Cov}\{\hat{\gamma}(h), \hat{\gamma}(k)\} &= \sum_{i=-\infty}^{\infty} \gamma(i) \gamma(i+k-h) + \gamma(i+k) \gamma(i-h) \\ &\quad + (\kappa_\epsilon - 3) \gamma(h) \gamma(k) \end{aligned} \quad (1.1)$$

and

$$\begin{aligned} \lim_{n \rightarrow \infty} n \text{Cov}\{\hat{\rho}(h), \hat{\rho}(k)\} &= \sum_{i=-\infty}^{\infty} \rho(i) [2\rho(h)\rho(k)\rho(i) - 2\rho(h)\rho(i+k) - 2\rho(k)\rho(i+h) \\ &\quad + \rho(i+k-h) + \rho(i-k-h)]. \end{aligned} \quad (1.2)$$

Formula (1.2) still holds under the assumptions

$$E\epsilon_t^2 < \infty, \quad \sum_{j=-\infty}^{\infty} |j| \phi_j^2 < \infty.$$

In particular, if $X_t = \epsilon_t$ and $E\epsilon_t^2 < \infty$, we have

$$\sqrt{n} \begin{pmatrix} \hat{\rho}(1) \\ \vdots \\ \hat{\rho}(h) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_h).$$

The assumptions of this theorem are demanding, because they require a strong white noise (ϵ_t) . An extension allowing the strong linearity assumption to be relaxed is proposed in Appendix B.2. For many nonlinear processes, in particular the ARCH processes studied in this book, the asymptotic covariance of the sample autocovariances can be very different from (1.1) (Exercises 1.6 and 1.8). Using the standard Bartlett formula can lead to specification errors (see Chapter 5).

1.2 ARMA and ARIMA Models

The aim of time series analysis is to construct a model for the underlying stochastic process. This model is then used for analyzing the causal structure of the process or to obtain optimal predictions.

The class of ARMA models is the most widely used for the prediction of second-order stationary processes. These models can be viewed as a natural consequence of a fundamental result due to Wold (1938), which can be stated as follows: *any centered, second-order stationary, and ‘purely nondeterministic’ process¹ admits an infinite moving-average representation of the form*

$$X_t = \epsilon_t + \sum_{i=1}^{\infty} c_i \epsilon_{t-i}, \quad (1.3)$$

where (ϵ_t) is the linear innovation process of (X_t) , that is

$$\epsilon_t = X_t - E(X_t | \mathcal{H}_X(t-1)), \quad (1.4)$$

where $\mathcal{H}_X(t-1)$ denotes the Hilbert space generated by the random variables $X_{t-1}, X_{t-2}, \dots$ ² and $E(X_t | \mathcal{H}_X(t-1))$ denotes the orthogonal projection of X_t onto $\mathcal{H}_X(t-1)$. The sequence of coefficients (c_i) is such that $\sum_i c_i^2 < \infty$. Note that (ϵ_t) is a weak white noise.

Truncating the infinite sum in (1.3), we obtain the process

$$X_t(q) = \epsilon_t + \sum_{i=1}^q c_i \epsilon_{t-i},$$

called a moving average process of order q , or $MA(q)$. We have

$$\|X_t(q) - X_t\|_2^2 = E\epsilon_t^2 \sum_{i>q} c_i^2 \rightarrow 0, \quad \text{as } q \rightarrow \infty.$$

It follows that the set of all finite-order moving averages is dense in the set of second-order stationary and purely nondeterministic processes. The class of ARMA models is often preferred to the MA models for parsimony reasons, because they generally require fewer parameters.

Definition 1.4 (ARMA(p, q) process) A second-order stationary process (X_t) is called ARMA(p, q), where p and q are integers, if there exist real coefficients $c, a_1, \dots, a_p, b_1, \dots, b_q$ such that,

$$\forall t \in \mathbb{Z}, \quad X_t + \sum_{i=1}^p a_i X_{t-i} = c + \epsilon_t + \sum_{j=1}^q b_j \epsilon_{t-j}, \quad (1.5)$$

where (ϵ_t) is the linear innovation process of (X_t) .

This definition entails constraints on the zeros of the autoregressive and moving average polynomials, $a(z) = 1 + \sum_{i=1}^p a_i z^i$ and $b(z) = 1 + \sum_{i=1}^q b_i z^i$ (Exercise 1.9). The main attraction of this model, and the representations obtained by successively inverting the polynomials $a(\cdot)$ and $b(\cdot)$, is that it provides a framework for deriving the *optimal linear predictions* of the process, in much simpler way than by only assuming the second-order stationarity.

Many economic series display trends, making the stationarity assumption unrealistic. Such trends often vanish when the series is differentiated, once or several times. Let $\Delta X_t = X_t - X_{t-1}$ denote the first-difference series, and let $\Delta^d X_t = \Delta(\Delta^{d-1} X_t)$ (with $\Delta^0 X_t = X_t$) denote the differences of order d .

¹A stationary process (X_t) is said to be purely nondeterministic if and only if $\bigcap_{n=-\infty}^{\infty} \mathcal{H}_X(n) = \{0\}$, where $\mathcal{H}_X(n)$ denotes, in the Hilbert space of the real, centered, and square integrable variables, the subspace constituted by the limits of the linear combinations of the variables X_{n-i} , $i \geq 0$. Thus, for a purely nondeterministic (or regular) process, the linear past, sufficiently far away in the past, is of no use in predicting future values. See Brockwell and Davis (1991, pp. 187–189) or Azencott and Dacunha-Castelle (1984) for more details.

²In this representation, the equivalence class $E(X_t | \mathcal{H}_X(t-1))$ is identified with a random variable.

Definition 1.5 (ARIMA(p, d, q) process) Let d be a positive integer. The process (X_t) is said to be an ARIMA(p, d, q) process if, for $k = 0, \dots, d - 1$, the processes $(\Delta^k X_t)$ are not second-order stationary, and $(\Delta^d X_t)$ is an ARMA(p, q) process.

The simplest ARIMA process is the ARIMA(0, 1, 0), also called the random walk, satisfying

$$X_t = \epsilon_t + \epsilon_{t-1} + \dots + \epsilon_1 + X_0, \quad t \geq 1,$$

where ϵ_t is a weak white noise.

For statistical convenience, ARMA (and ARIMA) models are generally used under stronger assumptions on the noise than that of weak white noise. *Strong* ARMA refers to the ARMA model of Definition 1.4 when ϵ_t is assumed to be a strong white noise. This additional assumption allows us to use convenient statistical tools developed in this framework, but considerably reduces the generality of the ARMA class. Indeed, assuming a strong ARMA is tantamount to assuming that (i) the *optimal predictions* of the process are linear ((ϵ_t) being the strong innovation of (X_t)) and (ii) the amplitudes of the prediction intervals depend on the horizon but not on the observations. We shall see in the next section how restrictive this assumption can be, in particular for financial time series modeling.

The orders (p, q) of an ARMA process are fully characterized through its autocorrelation function (see Brockwell and Davis, 1991, pp. 89–90, for a proof).

Theorem 1.2 (Characterization of an ARMA process) Let (X_t) denote a second-order stationary process. We have

$$\rho(h) + \sum_{i=1}^p a_i \rho(h - i) = 0, \quad \text{for all } |h| > q,$$

if and only if (X_t) is an ARMA(p, q) process.

To close this section, we summarize the method for time series analysis proposed in the famous book by Box and Jenkins (1970). To simplify presentation, we do not consider seasonal series, for which SARIMA models can be considered.

Box–Jenkins Methodology

The aim of this methodology is to find the most appropriate ARIMA(p, d, q) model and to use it for forecasting. It uses an iterative six-stage scheme:

- (i) *a priori* identification of the differentiation order d (or choice of another transformation);
- (ii) *a priori* identification of the orders p and q ;
- (iii) estimation of the parameters $(a_1, \dots, a_p, b_1, \dots, b_q)$ and $\sigma^2 = \text{Var } \epsilon_t$;
- (iv) validation;
- (v) choice of a model;
- (vi) prediction.

Although many unit root tests have been introduced in the last 30 years, step (i) is still essentially based on examining the graph of the series. If the data exhibit apparent deviations from stationarity, it will not be appropriate to choose $d = 0$. For instance, if the amplitude of the variations tends

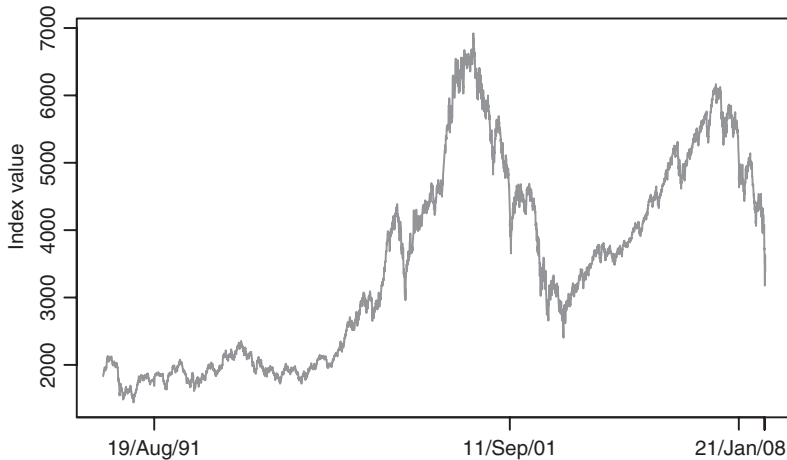


Figure 1.1 CAC 40 index for the period from March 1, 1990 to October 15, 2008 (4702 observations).

to increase, the assumption of constant variance can be questioned. This may be an indication that the underlying process is heteroscedastic.³ If a regular linear trend is observed, positive or negative, it can be assumed that the underlying process is such that $EX_t = at + b$ with $a \neq 0$. If this assumption is correct, the first-difference series $\Delta X_t = X_t - X_{t-1}$ should not show any trend ($E\Delta X_t = a$) and could be stationary. If no other sign of nonstationarity can be detected (such as heteroscedasticity), the choice $d = 1$ seems suitable. The random walk (whose sample paths may resemble the graph of Figure 1.1), is another example where $d = 1$ is required, although this process does not have any deterministic trend.

Step (ii) is more problematic. The primary tool is the sample autocorrelation function. If, for instance, we observe that $\hat{\rho}(1)$ is far away from 0 but that for any $h > 1$, $\hat{\rho}(h)$ is close to 0,⁴ then, from Theorem 1.1, it is plausible that $\rho(1) \neq 0$ and $\rho(h) = 0$ for all $h > 1$. In this case, Theorem 1.2 entails that X_t is an MA(1) process. To identify AR processes, the partial autocorrelation function (see Appendix B.1) plays an analogous role. For mixed models (that is, ARMA(p, q) with $pq \neq 0$), more sophisticated statistics can be used, as will be seen in Chapter 5. Step (ii) often results in the selection of several candidates $(p_1, q_1), \dots, (p_k, q_k)$ for the ARMA orders. These k models are estimated in step (iii), using, for instance, the least-squares method. The aim of step (iv) is to gauge if the estimated models are reasonably compatible with the data. An important part of the procedure is to examine the residuals which, if the model is satisfactory, should have the appearance of white noise. The correlograms are examined and portmanteau tests are used to decide if the residuals are sufficiently close to white noise. These tools will be described in detail in Chapter 5. When the tests on the residuals fail to reject the model, the significance of the estimated coefficients is studied. Testing the nullity of coefficients sometimes allows the model to be simplified. This step may lead to rejection of all the estimated models, or to consideration of other models, in which case we are brought back to step (i) or (ii). If several models pass the validation step (iv), selection criteria can be used, the most popular being the Akaike (AIC) and Bayesian (BIC) information criteria. Complementing these criteria, the predictive properties of the models can be considered: different models can lead to almost equivalent predictive formulas. The parsimony principle would thus lead us to choose the simplest model, the one with the fewest parameters. Other considerations can also come into play: for instance, models frequently involve a lagged variable at the order 12 for monthly data, but this would seem less natural for weekly data.

³ In contrast, a process such that $\text{Var}X_t$ is constant is called (marginally) homoscedastic.

⁴ More precisely, for $h > 1$, $\sqrt{n}|\hat{\rho}(h)|/\sqrt{1 + 2\hat{\rho}^2(1)}$ is a plausible realization of the $|\mathcal{N}(0, 1)|$ distribution.

If the model is appropriate, step (vi) allows us to easily compute the best linear predictions $\hat{X}_t(h)$ at horizon $h = 1, 2, \dots$. Recall that these linear predictions do not necessarily lead to minimal quadratic errors. Nonlinear models, or nonparametric methods, sometimes produce more accurate predictions. Finally, the interval predictions obtained in step (vi) of the Box–Jenkins methodology are based on Gaussian assumptions. Their magnitude does not depend on the data, which for financial series is not appropriate, as we shall see.

1.3 Financial Series

Modeling financial time series is a complex problem. This complexity is not only due to the variety of the series in use (stocks, exchange rates, interest rates, etc.), to the importance of the frequency of observation (second, minute, hour, day, etc) or to the availability of very large data sets. It is mainly due to the existence of statistical regularities (*stylized facts*) which are common to a large number of financial series and are difficult to reproduce artificially using stochastic models.

Most of these stylized facts were put forward in a paper by Mandelbrot (1963). Since then, they have been documented, and completed, by many empirical studies. They can be observed more or less clearly depending on the nature of the series and its frequency. The properties that we now present are mainly concerned with daily stock prices.

Let p_t denote the price of an asset at time t and let $\epsilon_t = \log(p_t/p_{t-1})$ be the continuously compounded or log return (also simply called the return). The series (ϵ_t) is often close to the series of relative price variations $r_t = (p_t - p_{t-1})/p_{t-1}$, since $\epsilon_t = \log(1 + r_t)$. In contrast to the prices, the returns or relative prices do not depend on monetary units which facilitates comparisons between assets. The following properties have been amply commented upon in the financial literature.

- (i) *Nonstationarity of price series*. Samples paths of prices are generally close to a random walk without intercept (see the CAC index series⁵ displayed in Figure 1.1). On the other hand, sample paths of returns are generally compatible with the second-order stationarity assumption. For instance, Figures 1.2 and 1.3 show that the returns of the CAC index

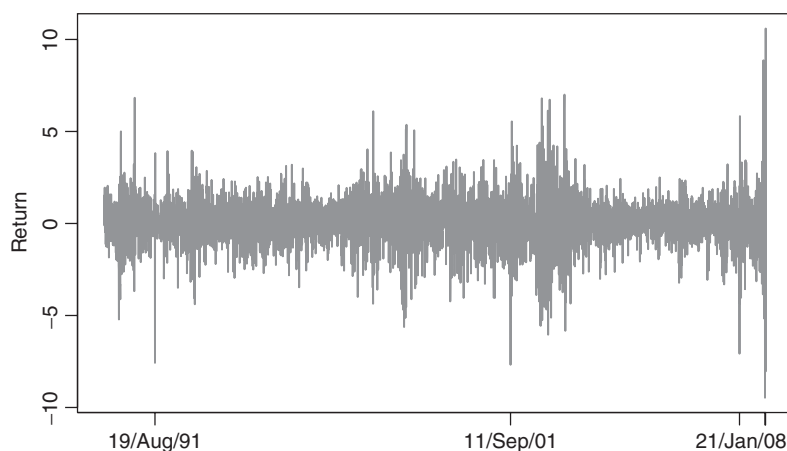


Figure 1.2 CAC 40 returns (March 2, 1990 to October 15, 2008). August 19, 1991, Soviet Putsch attempt; September 11, 2001, fall of the Twin Towers; January 21, 2008, effect of the subprime mortgage crisis; October 6, 2008, effect of the financial crisis.

⁵ The CAC 40 index is a linear combination of a selection of 40 shares on the Paris Stock Exchange (CAC stands for ‘Cotations Assistées en Continu’).

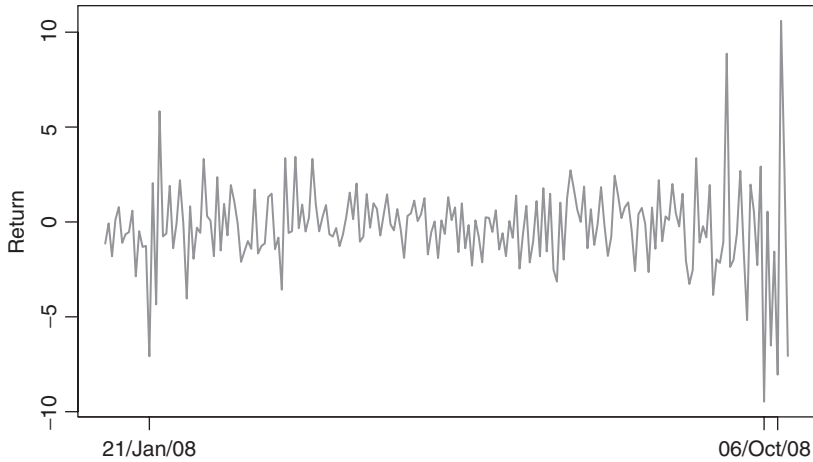


Figure 1.3 Returns of the CAC 40 (January 2, 2008 to October 15, 2008).

oscillate around zero. The oscillations vary a great deal in magnitude, but are almost constant in average over long subperiods. The recent extreme volatility of prices, induced by the financial crisis of 2008, is worth noting.

- (ii) *Absence of autocorrelation for the price variations.* The series of price variations generally displays small autocorrelations, making it close to a white noise. This is illustrated for the

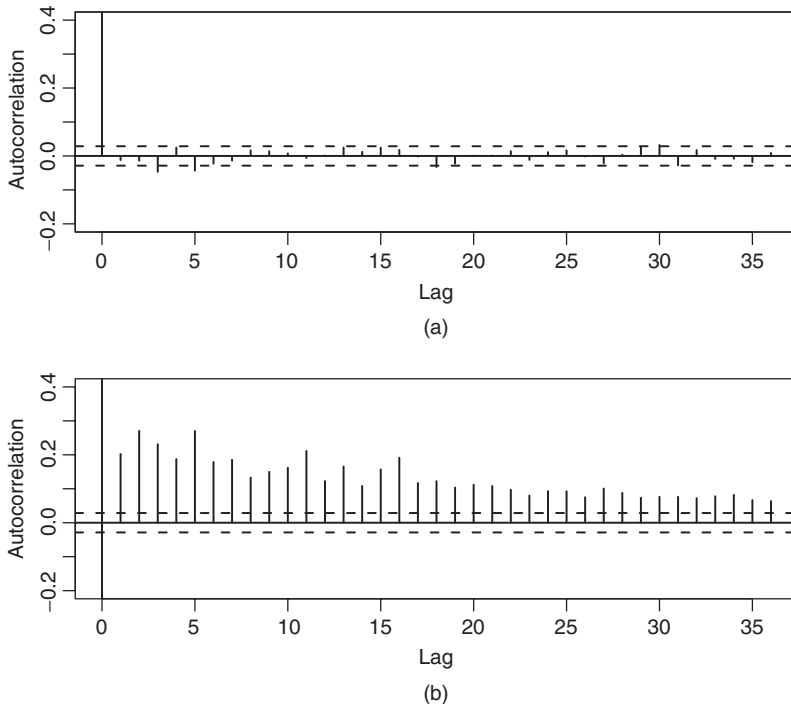


Figure 1.4 Sample autocorrelations of (a) returns and (b) squared returns of the CAC 40 (January 2, 2008 to October 15, 2008).

CAC in Figure 1.4(a). The classical significance bands are used here, as an approximation, but we shall see in Chapter 5 that they must be corrected when the noise is not independent. Note that for intraday series, with very small time intervals between observations (measured in minutes or seconds) significant autocorrelations can be observed due to the so-called *microstructure* effects.

- (iii) *Autocorrelations of the squared price returns.* Squared returns (ϵ_t^2) or absolute returns ($|\epsilon_t|$) are generally strongly autocorrelated (see Figure 1.4(b)). This property is not incompatible with the white noise assumption for the returns, but shows that the white noise is not strong.
- (iv) *Volatility clustering.* Large absolute returns $|\epsilon_t|$ tend to appear in clusters. This property is generally visible on the sample paths (as in Figure 1.3). Turbulent (high-volatility) subperiods are followed by quiet (low-volatility) periods. These subperiods are recurrent but do not appear in a periodic way (which might contradict the stationarity assumption). In other words, volatility clustering is not incompatible with a homoscedastic (i.e. with a constant variance) marginal distribution for the returns.
- (v) *Fat-tailed distributions.* When the empirical distribution of daily returns is drawn, one can generally observe that it does not resemble a Gaussian distribution. Classical tests typically lead to rejection of the normality assumption at any reasonable level. More precisely, the densities have fat tails (decreasing to zero more slowly than $\exp(-x^2/2)$) and are sharply peaked at zero: they are called leptokurtic. A measure of the leptokurticity is the kurtosis coefficient, defined as the ratio of the sample fourth-order moment to the squared sample variance. Asymptotically equal to 3 for Gaussian iid observations, this coefficient is much greater than 3 for returns series. When the time interval over which the returns are computed increases, leptokurticity tends to vanish and the empirical distributions get closer to a Gaussian. Monthly returns, for instance, defined as the sum of daily returns over the month, have a distribution that is much closer to the normal than daily returns. Figure 1.5 compares

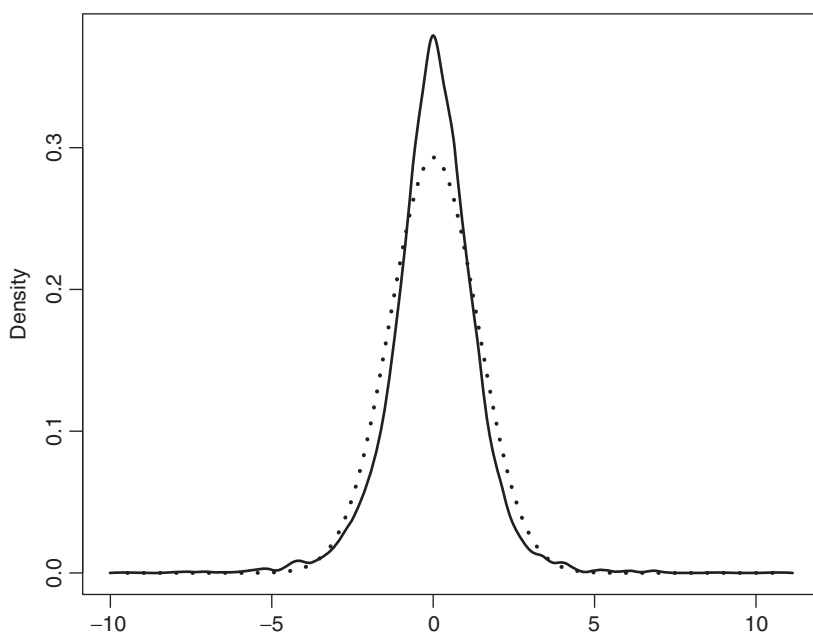


Figure 1.5 Kernel estimator of the CAC 40 returns density (solid line) and density of a Gaussian with mean and variance equal to the sample mean and variance of the returns (dotted line).

Table 1.1 Sample autocorrelations of returns ϵ_t (CAC 40 index, January 2, 2008 to October 15, 2008), of absolute returns $|\epsilon_t|$, sample correlations between ϵ_{t-h}^+ and $|\epsilon_t|$, and between $-\epsilon_{t-h}^-$ and $|\epsilon_t|$.

h	1	2	3	4	5	6	7
$\hat{\rho}_\epsilon(h)$	-0.012	-0.014	-0.047	0.025	-0.043	-0.023	-0.014
$\hat{\rho}_{ \epsilon }(h)$	0.175	0.229	0.235	0.200	0.218	0.212	0.203
$\hat{\rho}(\epsilon_{t-h}^+, \epsilon_t)$	0.038	0.059	0.051	0.055	0.059	0.109	0.061
$\hat{\rho}(-\epsilon_{t-h}^-, \epsilon_t)$	0.160	0.200	0.215	0.173	0.190	0.136	0.173

We use here the notation $\epsilon_t^+ = \max(\epsilon_t, 0)$ and $\epsilon_t^- = \min(\epsilon_t, 0)$.

a kernel estimator of the density of the CAC returns with a Gaussian density. The peak around zero appears clearly, but the thickness of the tails is more difficult to visualize.

- (vi) *Leverage effects*. The so-called leverage effect was noted by Black (1976), and involves an asymmetry of the impact of past positive and negative values on the current volatility. Negative returns (corresponding to price decreases) tend to increase volatility by a larger amount than positive returns (price increases) of the same magnitude. Empirically, a positive correlation is often detected between $\epsilon_t^+ = \max(\epsilon_t, 0)$ and $|\epsilon_{t+h}|$ (a price increase should entail future volatility increases), but, as shown in Table 1.1, this correlation is generally less than between $-\epsilon_t^- = \max(-\epsilon_t, 0)$ and $|\epsilon_{t+h}|$.
- (vii) *Seasonality*. Calendar effects are also worth mentioning. The day of the week, the proximity of holidays, among other seasonalities, may have significant effects on returns. Following a period of market closure, volatility tends to increase, reflecting the information cumulated during this break. However, it can be observed that the increase is less than if the information had cumulated at constant speed. Let us also mention that the seasonal effect is also very present for intraday series.

1.4 Random Variance Models

The previous properties illustrate the difficulty of financial series modeling. Any satisfactory statistical model for daily returns must be able to capture the main stylized facts described in the previous section. Of particular importance are the leptokurticity, the unpredictability of returns, and the existence of positive autocorrelations in the squared and absolute returns. Classical formulations (such as ARMA models) centered on the second-order structure are inappropriate. Indeed, the second-order structure of most financial time series is close to that of white noise.

The fact that large absolute returns tend to be followed by large absolute returns (whatever the sign of the price variations) is hardly compatible with the assumption of constant *conditional* variance. This phenomenon is called *conditional heteroscedasticity*:

$$\text{Var}(\epsilon_t \mid \epsilon_{t-1}, \epsilon_{t-2}, \dots) \neq \text{const.}$$

Conditional heteroscedasticity is perfectly compatible with stationarity (in the strict and second-order senses), just as the existence of a nonconstant conditional mean is compatible with stationarity. The GARCH processes studied in this book will amply illustrate this point.

The models introduced in the econometric literature to account for the very specific nature of financial series (price variations or log-returns, interest rates, etc.) are generally written in the multiplicative form

$$\epsilon_t = \sigma_t \eta_t \quad (1.6)$$

where (η_t) and (σ_t) are real processes such that:

- (i) σ_t is measurable with respect to a σ -field, denoted $\mathcal{F}_{t-1}$;
- (ii) (η_t) is an iid centered process with unit variance, η_t being independent of $\mathcal{F}_{t-1}$ and $\sigma(\epsilon_u; u < t)$;
- (iii) $\sigma_t > 0$.

This formulation implies that the sign of the current price variation (that is, the sign of ϵ_t) is that of η_t , and is independent of past price variations. Moreover, if the first two conditional moments of ϵ_t exist, they are given by

$$E(\epsilon_t | \mathcal{F}_{t-1}) = 0, \quad E(\epsilon_t^2 | \mathcal{F}_{t-1}) = \sigma_t^2.$$

The random variable σ_t is called the *volatility*⁶ of ϵ_t .

It may also be noted that (under existence assumptions)

$$E(\epsilon_t) = E(\sigma_t)E(\eta_t) = 0$$

and

$$\text{Cov}(\epsilon_t, \epsilon_{t-h}) = E(\eta_t)E(\sigma_t \epsilon_{t-h}) = 0, \quad \forall h > 0,$$

which makes (ϵ_t) a weak white noise. The series of squares, on the other hand, generally have nonzero autocovariances: (ϵ_t) is thus not a strong white noise.

The kurtosis coefficient of ϵ_t , if it exists, is related to that of η_t , denoted κ_η , by

$$\frac{E(\epsilon_t^4)}{\{E(\epsilon_t^2)\}^2} = \kappa_\eta \left[1 + \frac{\text{Var}(\sigma_t^2)}{\{E(\sigma_t^2)\}^2} \right]. \quad (1.7)$$

This formula shows that the leptokurticity of financial time series can be taken into account in two different ways: either by using a leptokurtic distribution for the iid sequence (η_t) , or by specifying a process (σ_t^2) with a great variability.

Different classes of models can be distinguished depending on the specification adopted for σ_t :

- (i) *Conditionally heteroscedastic* (or GARCH-type) processes for which $\mathcal{F}_{t-1} = \sigma(\epsilon_s; s < t)$ is the σ -field generated by the past of ϵ_t . The volatility is here a deterministic function of the past of ϵ_t . Processes of this class differ by the choice of a specification for this function. The standard GARCH models are characterized by a volatility specified as a linear function of the past values of ϵ_t^2 . They will be studied in detail in Chapter 2.
- (ii) *Stochastic volatility* processes⁷ for which $\mathcal{F}_{t-1}$ is the σ -field generated by $\{v_t, v_{t-1}, \dots\}$, where (v_t) is a strong white noise and is independent of (η_t) . In these models, volatility is a latent process. The most popular model in this class assumes that the process $\log \sigma_t$ follows an AR(1) of the form

$$\log \sigma_t = \omega + \phi \log \sigma_{t-1} + v_t,$$

where the noises (v_t) and (η_t) are independent.

- (iii) *Switching-regime models* for which $\sigma_t = \sigma(\Delta_t, \mathcal{F}_{t-1})$, where (Δ_t) is a latent (unobservable) integer-valued process, independent of (η_t) . The state of the variable Δ_t is here interpreted as a regime and, conditionally on this state, the volatility of ϵ_t has a GARCH specification. The process (Δ_t) is generally supposed to be a finite-state Markov chain. The models are thus called Markov-switching models.

⁶ There is no general agreement concerning the definition of this concept in the literature. Volatility sometimes refers to a conditional standard deviation, and sometimes to a conditional variance.

⁷ Note, however, that the volatility is also a random variable in GARCH-type processes.

1.5 Bibliographical Notes

The time series concepts presented in this chapter are the subject of numerous books. Two classical references are Brockwell and Davis (1991) and Gouriéroux and Monfort (1995, 1996).

The assumption of iid Gaussian price variations has long been predominant in the finance literature and goes back to the dissertation by Bachelier (1900), where a precursor of Brownian motion can be found. This thesis, ignored for a long time until its rediscovery by Kolmogorov in 1931 (see Kahane, 1998), constitutes the historical source of the link between Brownian motion and mathematical finance. Nonetheless, it relies on only a rough description of the behavior of financial series. The stylized facts concerning these series can be attributed to Mandelbrot (1963) and Fama (1965). Based on the analysis of many stock returns series, their studies showed the leptokurticity, hence the non-Gaussianity, of marginal distributions, some temporal dependencies and nonconstant volatilities. Since then, many empirical studies have confirmed these findings. See, for instance, Taylor (2007) for a recent presentation of the stylized facts of financial times series. In particular, the calendar effects are discussed in detail.

As noted by Shephard (2005), a precursor article on ARCH models is that of Rosenberg (1972). This article shows that the decomposition (1.6) allows the leptokurticity of financial series to be reproduced. It also proposes some volatility specifications which anticipate both the GARCH and stochastic volatility models. However, the GARCH models to be studied in the next chapters are not discussed in this article. The decomposition of the kurtosis coefficient in (1.7) can be found in Clark (1973).

A number of surveys have been devoted to GARCH models. See, among others, Bollerslev, Chou and Kroner (1992), Bollerslev, Engle and Nelson (1994), Pagan (1996), Palm (1996), Shephard (1996), Kim, Shephard, and Chib (1998), Engle (2001, 2002b, 2004), Engle and Patton (2001), Diebold (2004), Bauwens, Laurent and Rombouts (2006) and Giraitis et al. (2006). Moreover, the books by Gouriéroux (1997) and Xekalaki and Degiannakis (2009) are devoted to GARCH and several books devote a chapter to GARCH: Mills (1993), Hamilton (1994), Franses and van Dijk (2000), Gouriéroux and Jasiak (2001), Tsay (2002), Franke, Härdle and Hafner (2004), McNeil, Frey and Embrechts (2005), Taylor (2007) and Andersen et al. (2009). See also Mikosch (2001).

Although the focus of this book is on financial applications, it is worth mentioning that GARCH models have been used in other areas. Time series exhibiting GARCH-type behavior have also appeared, for example, in speech signals (Cohen, 2004; Cohen, 2006; Abramson and Cohen, 2008), daily and monthly temperature measurements (Tol, 1996; Campbell and Diebold, 2005; Romilly, 2005; Huang, Shiu, and Lin, 2008), wind speeds (Ewing, Kruse, and Schroeder, 2006), and atmospheric CO₂ concentrations (Hoti, McAleer, and Chan, 2005; McAleer and Chan, 2006).

Most econometric software (for instance, GAUSS, R, RATS, SAS and SPSS) incorporates routines that permit the estimation of GARCH models. Readers interested in the implementation with Ox may refer to Laurent (2009).

Stochastic volatility models are not treated in this book. One may refer to the book by Taylor (2007), and to the references therein. For switching regimes models, two recent references are the monographs by Cappé, Moulines and Rydén (2005), and by Frühwirth-Schnatter (2006).

1.6 Exercises

1.1 (*Stationarity, ARMA models, white noises*)

Let (η_t) denote an iid centered sequence with unit variance (and if necessary with a finite fourth-order moment).

1. Do the following models admit a stationary solution? If yes, derive the expectation and the autocorrelation function of this solution.

- (a) $X_t = 1 + 0.5X_{t-1} + \eta_t$;
- (b) $X_t = 1 + 2X_{t-1} + \eta_t$;
- (c) $X_t = 1 + 0.5X_{t-1} + \eta_t - 0.4\eta_{t-1}$.

2. Identify the ARMA models compatible with the following recursive relations, where $\rho(\cdot)$ denotes the autocorrelation function of some stationary process:

- (a) $\rho(h) = 0.4\rho(h-1)$, for all $h > 2$;
- (b) $\rho(h) = 0$, for all $h > 3$;
- (c) $\rho(h) = 0.2\rho(h-2)$, for all $h > 1$.

3. Verify that the following processes are white noises and decide if they are weak or strong.

- (a) $\epsilon_t = \eta_t^2 - 1$;
- (b) $\epsilon_t = \eta_t \eta_{t-1}$;

1.2 (A property of the sum of the sample autocorrelations)

Let

$$\hat{\gamma}(h) = \hat{\gamma}(-h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_t - \bar{X}_n)(X_{t+h} - \bar{X}_n), \quad h = 0, \dots, n-1,$$

denote the sample autocovariances of real observations $X_1, \dots, X_n$. Set $\hat{\rho}(h) = \hat{\rho}(-h) = \hat{\gamma}(h)/\hat{\gamma}(0)$ for $h = 0, \dots, n-1$. Show that

$$\sum_{h=1}^{n-1} \hat{\rho}(h) = -\frac{1}{2}.$$

1.3 (It is impossible to decide whether a process is stationary from a path)

Show that the sequence $\{(-1)^t\}_{t=0,1,\dots}$ can be a realization of a nonstationary process. Show that it can also be a realization of a stationary process. Comment on the consequences of this result.

1.4 (Stationarity and ergodicity from a path)

Can the sequence $0, 1, 0, 1, \dots$ be a realization of a stationary process or of a stationary and ergodic process? The definition of ergodicity can be found in Appendix A.1.

1.5 (A weak white noise which is not semi-strong)

Let (η_t) denote an iid $\mathcal{N}(0, 1)$ sequence and let k be a positive integer. Set $\epsilon_t = \eta_t \eta_{t-1} \dots \eta_{t-k}$. Show that (ϵ_t) is a weak white noise, but is not a strong white noise.

1.6 (Asymptotic variance of sample autocorrelations of a weak white noise)

Consider the white noise ϵ_t of Exercise 1.5. Compute $\lim_{n \rightarrow \infty} n \text{Var } \hat{\rho}(h)$ where $h \neq 0$ and $\hat{\rho}(\cdot)$ denotes the sample autocorrelation function of $\epsilon_1, \dots, \epsilon_n$. Compare this asymptotic variance with that obtained from the usual Bartlett formula.

1.7 (ARMA representation of the square of a weak white noise)

Consider the white noise ϵ_t of Exercise 1.5. Show that ϵ_t^2 follows an ARMA process. Make the ARMA representation explicit when $k = 1$.

1.8 (Asymptotic variance of sample autocorrelations of a weak white noise)

Repeat Exercise 1.6 for the weak white noise $\epsilon_t = \eta_t/\eta_{t-k}$, where (η_t) is an iid sequence such that $E\eta_t^4 < \infty$ and $E\eta_t^{-2} < \infty$, and k is a positive integer.

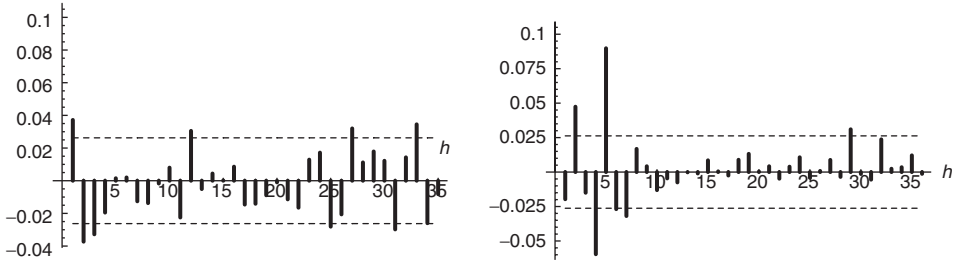


Figure 1.6 Sample autocorrelations $\hat{\rho}(h)$ ($h = 1, \dots, 36$) of (a) the S&P 500 index from January 3, 1979 to December 30, 2001, and (b) the squared index. The interval between the dashed lines ($\pm 1.96/\sqrt{n}$, where $n = 5804$ is the sample length) should contain approximately 95% of a strong white noise.

1.9 (Stationary solutions of an AR(1))

Let $(\eta_t)_{t \in \mathbb{Z}}$ be an iid centered sequence with variance $\sigma^2 > 0$, and let $a \neq 0$. Consider the AR(1) equation

$$X_t - aX_{t-1} = \eta_t, \quad t \in \mathbb{Z}. \quad (1.8)$$

1. Show that for $|a| < 1$, the infinite sum

$$X_t = \sum_{k=0}^{\infty} a^k \eta_{t-k}$$

converges in quadratic mean and almost surely, and that it is the unique stationary solution of (1.8).

2. For $|a| = 1$, show that no stationary solution exists.
3. For $|a| > 1$, show that

$$X_t = - \sum_{k=1}^{\infty} \frac{1}{a^k} \eta_{t+k}$$

is the unique stationary solution of (1.8).

4. For $|a| > 1$, show that the causal representation

$$X_t - \frac{1}{a} X_{t-1} = \epsilon_t, \quad t \in \mathbb{Z}, \quad (1.9)$$

holds, where $(\epsilon_t)_{t \in \mathbb{Z}}$ is a white noise.

1.10 (Is the S&P 500 a white noise?)

Figure 1.6 displays the correlogram of the S&P 500 returns from January 3, 1979 to December 30, 2001, as well as the correlogram of the squared returns. Is it reasonable to think that this index is a strong white noise or a weak white noise?

1.11 (Asymptotic covariance of sample autocovariances)

Justify the equivalence between (B.18) and (B.14) in the proof of the generalized Bartlett formula of Appendix B.2.

1.12 (*Asymptotic independence between the $\hat{\rho}(h)$ for a noise*)

Simplify the generalized Bartlett formulas (B.14) and (B.15) when $X = \epsilon$ is a pure white noise.

In an autocorrelogram, consider the random number M of sample autocorrelations falling outside the significance region (at the level 95%, say), among the first m autocorrelations. How can the previous result be used to evaluate the variance of this number when the observed process is a white noise (satisfying the assumptions allowing (B.15) to be used)?

1.13 (*An incorrect interpretation of autocorrelograms*)

Some practitioners tend to be satisfied with an estimated model only if all sample autocorrelations fall within the 95% significance bands. Show, using Exercise 1.12, that based on 20 autocorrelations, say, this approach leads to wrongly rejecting a white noise with a very high probability.

1.14 (*Computation of partial autocorrelations*)

Use the algorithm in (B.7) – (B.9) to compute $r_X(1)$, $r_X(2)$ and $r_X(3)$ as a function of $\rho_X(1)$, $\rho_X(2)$ and $\rho_X(3)$.

1.15 (*Empirical application*)

Download from <http://fr.biz.yahoo.com/bourse/accueil.html> for instance, a stock index such as the CAC 40. Draw the series of closing prices, the series of returns, the autocorrelation function of the returns, and that of the squared returns. Comment on these graphs.

