CHAPTER
TWENTY

Time Series Analysis

The historian is a prophet in reverse.
Friedrich von Schlegel (1772–1829)

A sequence of observations indexed by the time of each observation is called a **time series**. Time series analysis is the art of specifying the most likely stochastic model that could have generated the observed series. The aim is to understand the relations among the variables in order to make better predictions and decisions. One particularly important application of time series analysis in financial econometrics is the study of how volatilities change over time because financial assets demand returns commensurate with their volatility levels.

20.1 Introduction

A time series of prices is called a financial time series: The prices can be those of stocks, bonds, currencies, futures, commodities, and countless others. Other time series of financial interest include those of prepayment speeds of MBSs and various economic indicators.

Models for the time series can be conjectured from studying the data or can be suggested by economic theory. Most models specify a stochastic process. Models should be consistent with past data and amenable to testing for specification error. They should also be simple, containing as few parameters as possible. Model parameters are to be estimated from the time series. However, because an observed series is merely a sample path of the proposed stochastic process, this is possible only if the process possesses a property called **ergodicity**. Ergodicity roughly means that sample moments converge to the population moments as the sample path lengthens.

The basic steps are illustrated with the maximum likelihood (ML) estimation of stock price volatility. Suppose that after the historical time series of prices $S_1, S_2, \ldots, S_{n+1}$, observed at Δt apart, is studied, the geometric Brownian motion process

$$\frac{dS}{S} = \mu \, dt + \sigma \, dW \tag{20.1}$$

https://doi.org/10.1017/CBO9780511546839.021 Published online by Cambridge University Press

is proposed. Transform the data to simplify the analysis. As the return process $r \equiv \ln S$ follows $dr = \alpha\, dt + \sigma\, dW$, where $\alpha \equiv \mu - \sigma^2/2$, we take the logarithmic transformation of the series and perform the difference transformation:

$$R_i \equiv \ln S_{i+1} - \ln S_i = \ln(S_{i+1}/S_i).$$

Clearly $R_1, R_2, \ldots, R_n$ are independent, identically distributed, normal random variables distributed according to $N(\alpha\,\Delta t, \sigma^2 \Delta t)$. The log-likelihood function is

$$-\frac{n}{2}\ln(2\pi\sigma^2\Delta t) - \frac{1}{2\sigma^2\Delta t}\sum_{i=1}^{n}(R_i - \alpha\,\Delta t)^2.$$

Differentiate it with respect to α and σ^2 to obtain the ML estimators:

$$\widehat{\alpha} \equiv \frac{\sum_{i=1}^{n} R_i}{n\Delta t} = \frac{\ln(S_{n+1}/S_1)}{n\Delta t}, \tag{20.2}$$

$$\widehat{\sigma^2} \equiv \frac{\sum_{i=1}^{n}(R_i - \widehat{\alpha}\Delta t)^2}{n\Delta t}. \tag{20.3}$$

We note that the simple rate of return, $(S_{i+1} - S_i)/S_i$, and the continuously compounded rate of return, $\ln(S_{i+1}/S_i)$, should lead to similar conclusions because

$$\ln(S_{i+1}/S_i) = \ln(1 + (S_{i+1} - S_i)/S_i) \approx (S_{i+1} - S_i)/S_i.$$

EXAMPLE 20.1.1 Consider a time series generated by

$$S_{i+1} = S_i \times \exp[\,(\mu - \sigma^2/2)\Delta t + \sigma\sqrt{\Delta t}\,\xi\,], \quad \xi \sim N(0, 1),$$

with $S_1 = 1.0$, $\Delta t = 0.01$, $\mu = 0.15$, and $\sigma = 0.30$. Note that $\alpha = \mu - \sigma^2/2 = 0.105$. For the sample time series with $n = 5999$ in Fig. 20.1, the ML estimates of the parameters α and σ^2 are $\widehat{\alpha} = 0.118348$ and $\widehat{\sigma^2} = (0.299906)^2$. Because the variance of $\widehat{\sigma^2}$ is asymptotically $2\sigma^4/n$, increasing n definitely helps. However, because the variance for the estimator of α (hence μ) is asymptotically $\sigma^2/(n\Delta t)$, increasing n by sampling ever more frequently over the *same* time interval does not narrow

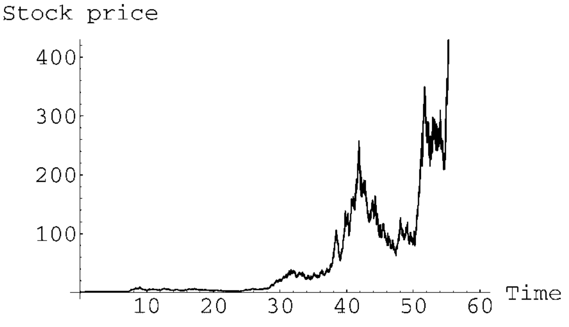

Figure 20.1: A simulated time series. The process is $dS = 0.15 \times S\,dt + 0.3 \times S\,dW$, starting at $S(0) = 1$.

the variance because the length of the sampling period, $n\Delta t$, remains unchanged [147, 611]. The variance can be reduced by sampling over a longer sampling period. For example, with $\Delta t = 2$ and $n = 999$ (the realized time series is not shown), $\widehat{\alpha}$ becomes 0.108039, a substantial improvement over the previous estimate even though the process is sampled *less* frequently (with a larger Δt).

As another illustration, consider the Ogden model for the short rate r,

$$dr = \beta(\mu - r)\, dt + \sigma r\, dW, \tag{20.4}$$

where $\beta > 0$, μ, and σ are the parameters [702]. Because approximately

$$\Delta r - \beta(\mu - r)\, \Delta t \sim N(0, \sigma^2 r^2 \Delta t),$$

conditional on r_1 the likelihood function for the n observations $\Delta r_1, \Delta r_2, \ldots,$ Δr_n is

$$\prod_{i=1}^{n} \left(2\pi\sigma^2 r_i^2 \Delta t\right)^{-1/2} \exp\left[-\frac{\{\, \Delta r_i - \beta(\mu - r_i)\, \Delta t\,\}^2}{2\sigma^2 r_i^2 \Delta t} \right],$$

where $\Delta r_i \equiv r_{i+1} - r_i$. The log-likelihood function after the removal of the constant terms and simplification is

$$-n \ln \sigma - \frac{1}{2\sigma^2 \Delta t} \sum_{i=1}^{n} [\, \Delta r_i - \beta(\mu - r_i)\, \Delta t\,]^2 r_i^{-2}.$$

Differentiating the log-likelihood function with respect to β, μ, and σ and equating them to zero gives rise to three equations in three unknowns:

$$0 = \sum_i [\, \Delta r_i - \beta(\mu - r_i)\, \Delta t\,](\mu - r_i)\, r_i^{-2}, \tag{20.5}$$

$$0 = \sum_i [\, \Delta r_i - \beta(\mu - r_i)\, \Delta t\,]\, r_i^{-2}, \tag{20.6}$$

$$\sigma^2 = \frac{1}{n\Delta t} \sum_i [\, \Delta r_i - \beta(\mu - r_i)\, \Delta t\,]^2 r_i^{-2}. \tag{20.7}$$

The ML estimators are not hard to obtain (see Exercise 20.1.3).

EXAMPLE 20.1.2 Consider the time series generated by $r_{i+1} = r_i + \beta(\mu - r_i)\, \Delta t + \sigma r_i \sqrt{\Delta t}\, \xi$, $\xi \sim N(0, 1)$, with $r_1 = 0.08$, $\Delta t = 0.1$, $\beta = 1.85$, $\sigma = 0.30$, $\mu = 0.08$, and $n = 999$. For the time series in Fig. 20.2, the ML estimates are $\widehat{\mu} = 0.0799202$, $\widehat{\beta} = 1.86253$, and $\widehat{\sigma} = 0.300989$.

➤ **Exercise 20.1.1** Derive the ML estimators for μ and σ^2 based on simple rates of returns, $\Delta S_i \equiv S_{i+1} - S_i$, $i = 1, 2, \ldots, n$.

➤ **Exercise 20.1.2** Assume that the stock price follows Eq. (20.1). The simple rate of return is defined as $[\, S(t) - S(0)\,]/S(0)$. Suppose that the volatility of the stock is that of simple rates of return, σ_s, instead of the instantaneous rates of return, σ. Express σ in terms of σ_s, the horizon t, and the expected simple rate of return at the horizon, μ_s.

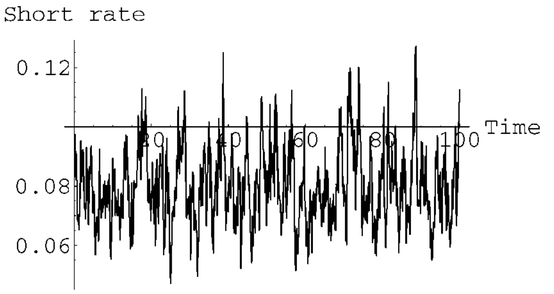

Figure 20.2: A simulated time series of the Ogden model. The process is $dr = 1.85 \times (0.08 - r)\, dt + 0.3 \times r\, dW$, starting at $r(0) = 0.08$.

➤ **Exercise 20.1.3** Derive the formulas for β and μ.

➤ **Exercise 20.1.4** Use the process for $\ln r$ to obtain the ML estimators for the Ogden model.

➤ **Exercise 20.1.5** The **constant elasticity variance** (**CEV**) **process** follows $dS/S = \mu\, dt + \lambda S^{\theta}\, dW$, where $\lambda > 0$. Derive the ML estimators with $\mu = 0$.

➢ **Programming Assignment 20.1.6** Write the simulator in Fig. 18.5 to generate stock prices. Then experiment with the ML estimators for goodness of fit.

20.1.1 Basic Definitions and Models

Although processes in this chapter are discrete-time by default, most of the definitions have continuous-time counterparts in Chap. 13. To simplify the presentation, we rely on context instead of notation to distinguish between random variables and their realizations.

Consider a discrete-time stochastic process $X_1, X_2, \ldots, X_n$. The **autocovariance** of X_t at lag τ is defined as $\text{Cov}[\, X_t, X_{t-\tau}\,]$, and the **autocorrelation** of X_t at lag τ is defined as $\text{Cov}[\, X_t, X_{t-\tau}\,]/\sqrt{\text{Var}[\, X_t\,] \times \text{Var}[\, X_{t-\tau}\,]}$, which clearly lies between -1 and 1. In general, autocovariances and autocorrelations depend on the time t as well as on the lag τ. The process is stationary if it has an identical mean μ and the autocovariances depend on only the lag; in particular, the variance is a constant. Because stationary processes are easier to analyze, transformation should be applied to the series to ensure stationarity whenever possible. The process is said to be **strictly stationary** if $\{\, X_s, X_{s+1}, \ldots, X_{s+\tau-1}\,\}$ and $\{\, X_t, X_{t+1}, \ldots, X_{t+\tau-1}\,\}$ have the same distribution for any s, t, and $\tau > 0$. This implies that the X_ts have identical distributions. A strictly stationary process is automatically stationary. A general process is said to be (**serially**) **uncorrelated** if all the autocovariances with a nonzero lag are zero; otherwise, it is **correlated**.

A stationary, uncorrelated process is called **white noise**. A strictly stationary process $\{\, X_t\,\}$, where the X_t are independent, is called **strict white noise**. A stationary process is called **Gaussian** if the joint distribution of $X_{t+1}, X_{t+2}, \ldots, X_{t+k}$ is multivariate normal for every possible integer k. A Gaussian process is automatically strictly stationary; in particular, Gaussian white noise is strict white noise.

Denote the autocovariance at lag τ of a stationary process by λ_τ. The autocovariance λ_0 is then the same variance shared by all X_t. The autocorrelations at lag τ of a stationary process are denoted by $\rho_\tau \equiv \lambda_\tau/\lambda_0$.

➤ **Exercise 20.1.7** Prove that $\lambda_\tau = \lambda_{-\tau}$ for stationary processes.

➤ **Exercise 20.1.8** Show that, for a stationary process with known mean, the optimal predictor in the mean-square-error sense is the mean μ.

➤ **Exercise 20.1.9** Consider a stationary process $\{ X_t \}$ with known mean and autocovariances. Derive the optimal linear prediction $a_0 + a_1 X_t + a_2 X_{t-1} + \cdots + a_t X_1$ in the mean-square-error sense for X_{t+1}. (Hint: Exercise 6.4.1.)

➤ **Exercise 20.1.10** Show that if price changes are uncorrelated, then the variance of prices must increase with time.

➤ **Exercise 20.1.11** Verify that if $\{ X_t \}$ is strict white noise, then so are $\{ |X_t| \}$ and $\{ X_t^2 \}$.

20.1.2 The Efficient Markets Hypothesis

The **random-walk hypothesis** posits that price changes are random and prices therefore behave unpredictably. For example, Bachelier in his 1900 thesis assumed that price changes have independent and identical normal distributions. The more modern version asserts that the return process has constant mean and is uncorrelated. One consequence of the hypothesis is that expected returns cannot be improved by use of past prices.

The random walk hypothesis is a purely statistical statement. By the late 1960s, however, the hypothesis could no longer withstand the mounting evidence against it. It was the work of Fama and others that shifted the focus from the time series of returns to that of cost- and risk-adjusted returns: Excess returns should account for transactions costs, risks, and information available to the trading strategy. Thereafter, the efficient-markets debate became a matter of economics instead of a matter of pure statistics [666].

The economic theory to explain the randomness of security prices is called the **efficient markets hypothesis**. It holds that the market may be viewed as a great information processor and prices come to reflect available information immediately. A market is efficient with respect to a particular information set if it is impossible to make abnormal profits by using this set of information in making trading decisions [799]. If the set of information refers to past prices of securities, we have the weak form of market efficiency; if the set of information refers to all *publicly available* information, we have the **semistrong** form; finally, if the set of information refers to all public and *private* information, we have the **strong** form. These terms are due to Roberts in 1967 [147]. The evidence suggests that major U.S. markets are at least weak-form efficient [767].

➤ **Exercise 20.1.12** Why are near-zero autocorrelations important for returns to be unpredictable?

20.1.3 Three Classic Models

Let ϵ_t denote a zero-mean white-noise process with constant variance σ^2 throughout the section. The first model is the **autoregressive (AR) process**, whose simplest version is

$$X_t - b = a(X_{t-1} - b) + \epsilon_t.$$

This process is stationary if $|a| < 1$, in which case $E[X_t] = b$, $\lambda_0 = \sigma^2/(1-a^2)$ and $\lambda_\tau = a^\tau \lambda_0$. The autocorrelations decay exponentially to zero because $\lambda_\tau/\lambda_0 = a^\tau$.[1] The general AR($p$) process follows

$$X_t - b = \epsilon_t + \sum_{i=1}^{p} a_i (X_{t-i} - b).$$

The next model is the **moving average (MA) process**,

$$X_t - b = \epsilon_t + c\epsilon_{t-1},$$

which has mean b and variance $\lambda_0 = (1+c^2)\sigma^2$. The autocovariance λ_τ equals 0 if $\tau > 1$ and $c\sigma^2$ if $\tau = 1$. The process is clearly stationary. Note that the autocorrelation function drops to zero beyond $\tau = 1$. The general MA(q) process is defined by

$$X_t - b = \epsilon_t + \sum_{j=1}^{q} c_j \epsilon_{t-j}.$$

It is not hard to show that observations more than q periods apart are uncorrelated. Repeated substitutions for the MA(1) process with $|c| < 1$ yield

$$X_t - b = -\sum_{i=1}^{\infty} (-c)^i (X_{t-i} - b) + \epsilon_t.$$

This can be seen as an AR(∞) process in which the effect of past observations decrease with age. An MA process is said to be **invertible** if it can be represented as an AR(∞) process.

The third model, the **autoregressive moving average (ARMA) process**, combines the AR(1) and MA(1) processes; thus

$$X_t - b = a(X_{t-1} - b) + \epsilon_t + c\epsilon_{t-1}.$$

Assume that $|a| < 1$ so that the ARMA process is stationary. Then the mean is b and

$$\begin{aligned}
\lambda_0 &= \sigma^2 \frac{1 + 2ac + c^2}{1 - a^2}, \\
\lambda_1 &= \sigma^2 \frac{(1+ac)(a+c)}{1 - a^2}, \\
\lambda_\tau &= a\lambda_{\tau-1} \quad \text{for } \tau \geq 2.
\end{aligned}$$

We define the more general ARMA(p, q) process by combining AR(p) and MA(q):

$$X_t - b = \sum_{i=1}^{p} a_i (X_{t-i} - b) + \sum_{j=0}^{q} c_j \epsilon_{t-j},$$

where $c_0 = 1$, $a_p \neq 0$, and $c_q \neq 0$.

Repeated substitutions for the stationary AR(1) process, with $|a| < 1$, yield

$$X_t - b = \sum_{j=0}^{\infty} a^j \epsilon_{t-j},$$

an MA(∞) process. This is a special case of **Wold's decomposition**, which says that any stationary process $\{ X_t \}$, after the linearly deterministic component has been removed, can be represented as an MA(∞) process:

$$X_t - b = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}, \tag{20.8}$$

where $c_0 = 1$ and $\sum_{j=1}^{\infty} c_j^2 < \infty$. Both stationary AR and ARMA processes admit such a representation.

Stationary processes have nice asymptotic properties. Suppose that a stationary process $\{ X_t \}$ is represented as Eq. (20.8), where $\sum_{j=1}^{\infty} |c_j| < \infty$ and ϵ_t are zero-mean, independent, identically distributed random variables with $E[\epsilon_t^2] < \infty$. A useful central limit theorem says that the sample mean is asymptotically normal in the sense that

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} X_i - b \right) \to N \left(0, \sum_{j=-\infty}^{\infty} \lambda_j \right) \quad \text{as } n \to \infty.$$

We note that $\sum_{j=1}^{\infty} |c_j| < \infty$ implies that $\sum_{j=1}^{\infty} c_j^2 < \infty$ but not vice versa, and it guarantees ergodicity for MA(∞) processes [415].

➤ **Exercise 20.1.13** Let $\{ X_t \}$ be a sequence of independent, identically distributed random variables with zero mean and unit variance. Prove that the process $\{ Y_t \equiv \sum_{k=0}^{l} a_k X_{t-k} \}$ with constant a_k is stationary.

➤ **Exercise 20.1.14** Given Wold's decomposition (20.8), show that λ_τ, the autocovariance at lag τ, equals $\sigma^2 \sum_{j=0}^{\infty} c_j c_{j+\tau}$.

Conditional Estimation of Gaussian AR Processes

For Gaussian AR processes, the ML estimation reduces to OLS problems. Write the AR(p) process as

$$X_t = c + \sum_{i=1}^{p} a_i X_{t-i} + \epsilon_t,$$

where ϵ_t is a zero-mean Gaussian white noise with constant variance σ^2. The parameters to be estimated from the observations $X_1, X_2, \dots, X_n$ are $a_1, a_2, \dots, a_p, c$, and σ^2. Conditional on the first p observations, the log-likelihood function can be easily seen to be

$$-\frac{n-p}{2} \ln(2\pi) - \frac{n-p}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=p+1}^{n} \left(X_t - c - \sum_{i=1}^{p} a_i X_{t-i} \right)^2.$$

The values of $a_1, a_2, \ldots, a_p$, and c that maximize the preceding function must minimize

$$\sum_{t=p+1}^{n} \left(X_t - c - \sum_{i=1}^{p} a_i X_{t-i} \right)^2, \tag{20.9}$$

an LS problem (see Exercise 20.1.15). This methodology is called the **conditional ML estimation**. The estimate will be consistent for any stationary ergodic AR process even if it is not Gaussian. The estimator of σ^2 is

$$\frac{1}{n-p} \sum_{t=p+1}^{n} \left(X_t - \widehat{c} - \sum_{i=1}^{p} \widehat{a}_i X_{t-i} \right)^2,$$

which can be found by differentiation of the log-likelihood function with respect to σ^2.

➤ **Exercise 20.1.15** Write the equivalent LS problem for function (20.9).

20.2 Conditional Variance Models for Price Volatility

Although a stationary model has constant variance, its *conditional* variance may vary. Take for example a stationary AR(1) process $X_t = a X_{t-1} + \epsilon_t$. Its conditional variance,

$$\mathrm{Var}[\, X_t \mid X_{t-1}, X_{t-2}, \ldots \,],$$

equals σ^2, which is smaller than the *unconditional* variance $\mathrm{Var}[\, X_t \,] = \sigma^2/(1-a^2)$. Note that the conditional variance is independent of past information; this property holds for ARMA processes in general. Past information thus has no effect on the variance of prediction. To address this drawback, consider models for returns X_t consistent with a changing conditional variance in the form of

$$X_t - \mu = V_t U_t.$$

It is assumed that (1) U_t has zero mean and unit variance for all t, (2) $E[\, X_t \,] = \mu$ for all t, and (3) $\mathrm{Var}[\, X_t \mid V_t = v_t \,] = v_t^2$. The process $\{ V_t^2 \}$ thus models the conditional variance.

Suppose that $\{ U_t \}$ and $\{ V_t \}$ are independent of each other, which means that $\{ U_1, U_2, \ldots, U_n \}$ and $\{ V_1, V_2, \ldots, V_n \}$ are independent for all n. Then $\{ X_t \}$ is uncorrelated because

$$\mathrm{Cov}[\, X_t, X_{t+\tau} \,] = E[\, V_t U_t V_{t+\tau} U_{t+\tau} \,] = E[\, V_t U_t V_{t+\tau} \,]\, E[\, U_{t+\tau} \,] = 0 \tag{20.10}$$

for $\tau > 0$. Furthermore, if $\{ V_t \}$ is stationary, then $\{ X_t \}$ has constant variance because

$$E[\, (X_t - \mu)^2 \,] = E\left[\, V_t^2 U_t^2 \,\right] = E\left[\, V_t^2 \,\right] E\left[\, U_t^2 \,\right] = E\left[\, V_t^2 \,\right], \tag{20.11}$$

making $\{ X_t \}$ stationary.

EXAMPLE 20.2.1 Here is a lognormal model. Let the processes $\{ U_t \}$ and $\{ V_t \}$ be independent of each other, $\{ U_t \}$ is Gaussian white noise, and $\ln V_t \sim N(a, b^2)$. One simple way to achieve this as well as to make both $\{ |X_t - \mu| \}$'s and $\{ (X_t - \mu)^2 \}$'s autocorrelations positive is to posit the following AR(1) model for $\{\ln V_t\}$:

$$\ln(V_t) - \alpha = \theta(\ln(V_{t-1}) - \alpha) + \xi_t, \quad \theta > 0.$$

In the preceding equation, $\{\xi_t\}$ is zero-mean, Gaussian white noise independent of $\{U_t\}$. To ensure the above-mentioned variance for $\ln V_t$, let $\text{Var}[\xi_t] = b^2(1-\theta^2)$. The four parameters in this model – a, b, θ, and α – can be estimated by the method of moments [839].

➤ **Exercise 20.2.1** Assume that the processes $\{V_t\}$ and $\{U_t\}$ are stationary and independent of each other. Show that the kurtosis of X_t exceeds that of U_t provided that both are finite.

➤ **Exercise 20.2.2** For the lognormal model, show that (1) the kurtosis of X_t is $3e^{4b^2}$, (2) $\text{Var}[V_t] = e^{2a+b^2}(e^{b^2}-1)$, (3) $\text{Var}[|X_t-\mu|] = e^{2a+b^2}(e^{b^2}-2/\pi)$, (4) $\text{Var}[V_t^2] = e^{4a+4b^2}(e^{4b^2}-1)$, and (5) $\text{Var}[(X_t-\mu)^2] = e^{4a+4b^2}(3e^{4b^2}-1)$.

20.2.1 ARCH and GARCH Models

One trouble with the lognormal model is that the conditional variance evolves independently of past returns. Suppose we assume that conditional variances are deterministic functions of past returns: $V_t = f(X_{t-1}, X_{t-2}, \ldots)$ for some function f. Then V_t can be computed given $I_{t-1} \equiv \{X_{t-1}, X_{t-2}, \ldots\}$, the information set of past returns. An influential model in this direction is the **autoregressive conditional heteroskedastic (ARCH) model**.

Assume that U_t is independent of $V_t, U_{t-1}, V_{t-1}, U_{t-2}, \ldots$, for all t. Consequently $\{X_t\}$ is uncorrelated by Eq. (20.10). Assume furthermore that $\{U_t\}$ is a Gaussian white-noise process. Hence $X_t \mid I_{t-1} \sim N(\mu, V_t^2)$. The ARCH($p$) process is defined by

$$X_t - \mu = \left[a_0 + \sum_{i=1}^{p} a_i (X_{t-i} - \mu)^2 \right]^{1/2} U_t,$$

where $a_0, a_1, \ldots, a_p$ are nonnegative and $a_0 > 0$ is usually assumed. The variance V_t^2 thus equals $a_0 + \sum_{i=1}^{p} a_i (X_{t-i} - \mu)^2$. This suggests that V_t depends on $U_{t-1}, U_{t-2}, \ldots, U_{t-p}$, which is indeed the case; $\{U_t\}$ and $\{V_t\}$ are hence not independent of each other. In practical terms, the model says that the volatility at time t as estimated at time $t-1$ depends on the p most recent observations on squared returns.[2]

The ARCH(1) process $X_t - \mu = [a_0 + a_1(X_{t-1} - \mu)^2]^{1/2} U_t$ is the simplest for which

$$\text{Var}[X_t \mid X_{t-1} = x_{t-1}] = a_0 + a_1(x_{t-1} - \mu)^2. \tag{20.12}$$

The process $\{X_t\}$ is stationary with finite variance if and only if $a_1 < 1$, in which case $\text{Var}[X_t] = a_0/(1-a_1)$. The kurtosis is a finite $3(1-a_1^2)/(1-3a_2^2)$ when $3a_1^2 < 1$ and exceeds three when $a_1 > 0$. Let $S_t \equiv (X_t - \mu)^2$. Then

$$E[S_t \mid S_{t-1}] = a_0 + a_1 S_{t-1}$$

by Eq. (20.12). This resembles an AR process. Indeed, $\{S_t\}$ has autocorrelations a_1^τ when the variance of S_t exists, i.e., $3a_1^2 < 1$. Because $X_t \mid I_{t-1} \sim N(\mu, a_0 + a_1 S_{t-1})$, the log-likelihood function equals

$$-\frac{n-1}{2}\ln(2\pi)-\frac{1}{2}\sum_{i=2}^{n}\ln(a_0+a_1(X_{i-1}-\mu)^2)-\frac{1}{2}\sum_{i=2}^{n}\frac{(X_i-\mu)^2}{a_0+a_1(X_{i-1}-\mu)^2}.$$

We can estimate the parameters by maximizing the above function. The results for the more general ARCH(p) model are similar.

A popular extension of the ARCH model is the **generalized autoregressive conditional heteroskedastic** (**GARCH**) **process**. The simplest GARCH(1, 1) process adds $a_2 V_{t-1}^2$ to the ARCH(1) model:

$$V_t^2 = a_0 + a_1(X_{t-1}-\mu)^2 + a_2 V_{t-1}^2.$$

The volatility at time t as estimated at time $t-1$ thus depends on the squared return and the estimated volatility at time $t-1$. By repeated substitutions, the estimate of volatility can be seen to average past squared returns by giving heavier weights to recent squared returns (see Exercise 20.2.3, part (1)). For technical reasons, it is usually assumed that $a_1 + a_2 < 1$ and $a_0 > 0$, in which case the unconditional long-run variance is given by $a_0/(1-a_1-a_2)$. The model also exhibits mean reversion (see Exercise 20.2.3, part (2)).

➤ **Exercise 20.2.3** Assume the GARCH(1, 1) model. Show that (1) $V_t^2 = a_0\sum_{i=0}^{k-1} a_2^i + \sum_{i=1}^{k} a_1 a_2^{i-1}(X_{t-i}-\mu)^2 + a_2^k V_{t-k}^2$, where $k > 0$ and (2) $\mathrm{Var}[\, X_{t+k} \mid V_t\,] = V + (a_1+a_2)^k(V_t^2 - V)$, where $V \equiv a_0/(1-a_1-a_2)$.

Additional Reading

Many books cover time-series analysis and the important subject of model testing skipped here [22, 415, 422, 667, 839]. A version of the efficient markets hypothesis that is due to Samuelson in 1965 says asset returns are martingales (see Exercise 13.2.2) [594]. However, constant expected returns that a martingale entails have been rejected by the empirical evidence if markets are efficient [333]. See [148, 333, 424, 587, 594, 767] for more information on the efficient markets hypothesis. Wold's decomposition is due to Wold in 1938 [881]. The instability of the variance of returns has consequences for long-term investors [69]. Stochastic volatility has been extensively studied [49, 293, 440, 444, 450, 471, 520, 579, 790, 823, 839, 840]. Besides the Ito process approach to stochastic volatility in Section 15.5, jump processes have been proposed for the volatility process [177]. Consult [255] for problems with such approaches to volatility. One more approach to the smile problem is the implied binomial tree [502, 503]. The ARCH model is proposed by Engle [319]. The GARCH model is proposed by Bollerslev [99] and Taylor [839]. It has found widespread empirical support [464, 552, 578]. See [286, 287, 288, 442] for option pricing models based on the GARCH process and [749] for algorithms. Consult [517] for a survey on the estimation of Ito processes of the form $dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t$. As illustrated in Example 20.1.1, direct estimation of the drift μ is difficult in general from discretely observed data over a short time interval, however frequently sampled; the diffusion σ can be precisely estimated.

The **generalized method of moments** (**GMM**) is an estimation method that extends the method of moments in Subsection 6.4.3. Like the method of moments, the GMM formulates the moment conditions in which the parameters are implicitly defined. However, instead of solving equations, the GMM finds the parameters that jointly minimize the weighted "distance" between the sample and the population moments. The typical conditions for the GMM estimate to be consistent include stationarity, ergodicity, and the existence of relevance expectations. The GMM method of moments is due to Hansen [418] and is widely used in the analysis of time series [173, 248, 384, 526, 754, 819].

NOTES

1. The stationarity condition rules out the random walk with drift in Example 13.1.2 as a stationary AR process. Brownian motion, the limit of such a random walk, is also not stationary (review Subsection 13.3.3).
2. In practice μ is often assumed to be zero. This is reasonable when Δt is small, say 1 day, because the expected return is then insignificant compared with the standard deviation of returns.

> Someone who tried to use modern observations from London and Paris to judge mortality rates of the Fathers before the flood would enormously deviate from the truth.
>
> Gottfried Wilhelm von Leibniz (1646–1716)

https://doi.org/10.1017/CBO9780511546839.021 Published online by Cambridge University Press