CHAPTER ONE

Volatility Models

LUC BAUWENS, CHRISTIAN HAFNER, and SÉBASTIEN LAURENT

1.1 Introduction

This chapter presents an introductory review of volatility models and some applications. We link our review with other chapters that contain more detailed presentations. Section 1.2 deals with generalized autoregressive conditionally heteroskedastic models, Section 1.3 with stochastic volatility (SV) models, and Section 1.4 with realized volatility.

1.2 GARCH

1.2.1 UNIVARIATE GARCH

Univariate ARCH models appeared in the literature with the paper of Engle (1982a), soon followed by the generalization to GARCH of Bollerslev (1986). Although applied, in these pathbreaking papers, to account for the changing volatility of inflation series, the models and their later extensions were quickly found to be relevant for the conditional volatility of financial returns observed at a monthly and higher frequency (Bollerslev, 1987), and thus to the study of the intertemporal relation between risk and expected return (French et al., 1987; Engle et al., 1987). The reason is that time series of returns (even if

Handbook of Volatility Models and Their Applications, First Edition.
Edited by Luc Bauwens, Christian Hafner, and Sebastien Laurent.

adjusted for autocorrelation, typically through an ARMA model) have several features that are well fitted by GARCH models. The main stylized feature is volatility clustering: "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes" (Mandelbrot, 1963). This results in positive autocorrelation coefficients of squared returns, typically with a relatively slowly decreasing pattern starting from a first small value (say, <0.2). Said differently, volatility, measured by squared returns, is persistent, hence to some extent predictable even if it is noisy. Another stylized property of financial returns that was known long before ARCH models appeared is that their unconditional probability distributions are leptokurtic, that is, they have fatter tails and more mass around their center than the Gaussian distribution (Mandelbrot 1963). In this and later papers (e.g., Fama, 1963, 1965; Mandelbrot and Taylor, 1967), the returns are modeled as independently and identically distributed (i.i.d.) according to a stable Paretian distribution. But clearly, if squared returns are autocorrelated, they are not independent. A great advantage of GARCH models is that the returns are not assumed independent, and even if they are assumed Gaussian conditional to past returns, unconditionally they are not Gaussian, because volatility clustering generates leptokurtosis.

We illustrate the stylized facts with the percentage daily returns of the S&P 500 index, that is, the returns (y_t) are computed as $100(p_t - p_{t-1})$, where $p_t = \log P_t$ and P_t is the closing price index value adjusted for dividends and splits (available at http://finance.yahoo.com) and t is the time index referring to trading day t. The sample period starts on January 3, 1950 and ends on July 14, 2011 for a total of 15,482 returns. Table 1.1 contains descriptive statistics of the original and demeaned returns, the latter being the residuals of an AR(2) model fitted to the original returns. The descriptive statistics of the two series hardly differ and the large excess kurtosis coefficients confirm their leptokurtosis.

TABLE 1.1 Descriptive Statistics for S&P 500 Returns

	Returns	Demeaned Returns
Observations	15,482	15,480
Mean	0.02818	0
Standard deviation	0.97078	0.96897
Skewness	−1.0567	−1.0738
Kurtosis	32.035	31.623
Minimum	−22.900	−22.856
Maximum	10.957	10.571

Returns definition and source: (see text). Demeaned returns are residuals of an AR(2) model fitted to the returns by ordinary least squares (OLS). Skewness is the ratio of the third centered moment to the third power of the standard deviation. Kurtosis is the ratio of the fourth centered moment to the square of the variance.

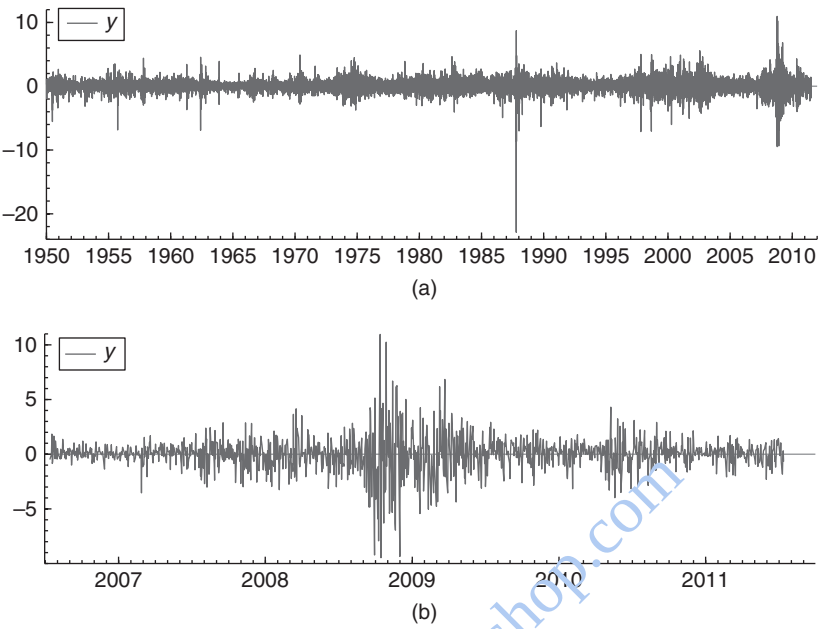

FIGURE 1.1 S&P 500 index returns (y).

Figure 1.1 displays the full sample series of returns (a) and the series for the last five years (2006/07/14–2011/07/14) (b). Figure 1.2 shows the full series of absolute demeaned returns (a), the sample ACF of the corresponding squared series until lag 100 (b), and the absolute demeaned returns or the last five years (c). The squared demeaned returns are positively autocorrelated: their ACF starts at 0.15, has a peak of 0.20 at lag 5, and decreases rather slowly. Volatility clustering is clearly visible on the top and bottom graphs of both figures. The leptokutrosis of the estimated density of the demeaned returns, shown over a truncated support—see maximum and minimum values in Table 1.1—is visible on Figure 1.3, where a Gaussian density with the same mean (0) and standard deviation (0.969) is drawn for comparison. The negative skewness coefficients reported in Table 1.1 illustrate that large negative returns are more probable than large positive ones. This is also a widespread feature, by no means universal, of financial return series, which we discuss below.

1.2.1.1 Structure of GARCH Models. We define a GARCH model for y_t (an asset return as defined above) by

$$y_t - \mu_t = \varepsilon_t = \sigma_t z_t, \tag{1.1}$$

where z_t is an unobservable random variable belonging to an i.i.d. process, with mean equal to 0 and variance equal to 1, $E(z_t) = 0$ and $\mathrm{Var}(z_t) = 1$. The symbols μ_t and σ_t denote measurable functions with respect to a σ-field $\mathcal{F}_{t-1}$ generated

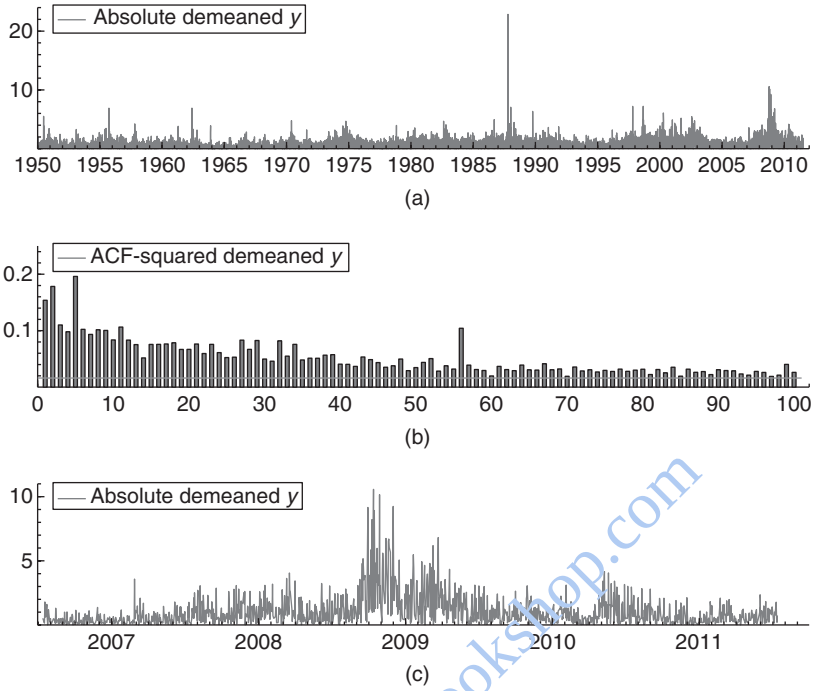

FIGURE 1.2 S&P 500 index demeaned absolute returns and ACF of their square.

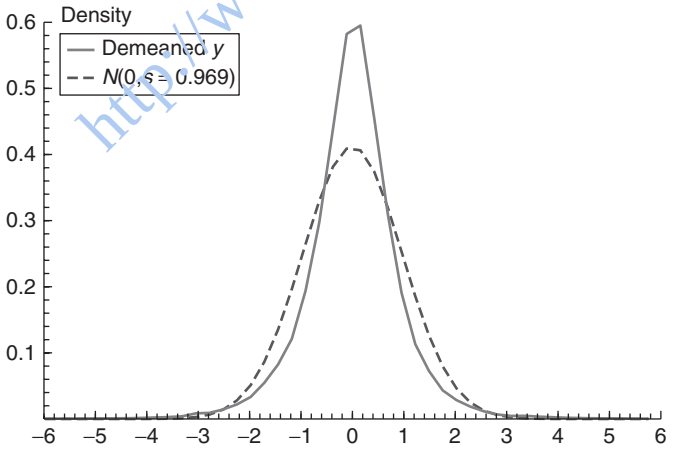

FIGURE 1.3 Density estimate of S&P 500 index demeaned returns and Gaussian density.

by y_{t-j} for $j \geq 1$ and possibly by other variables available at time $t-1$. It follows that μ_t and σ_t^2 are the conditional mean and variance of y_t, respectively, that is, $\mu_t = E(y_t|\mathcal{F}_{t-1}) = E_{t-1}(y_t)$ and $\sigma_t^2 = \text{Var}(y_t|\mathcal{F}_{t-1}) = \text{Var}_{t-1}(y_t)$, so that $E_{t-1}(\varepsilon_t) = 0$ and $\text{Var}_{t-1}(\varepsilon_t) = \sigma_t^2$. The i.i.d. hypothesis for the z_t process can be replaced by the assumption that the process is an m.d.s. (martingale difference sequence), such that $E_{t-1}(z_t) = 0$ and $\text{Var}_{t-1}(z_t) = 1$.

The model is fully parametric if μ_t, σ_t^2, and $f(z_t)$, the probability density function (pdf) of z_t (assumed to be time invariant), are indexed by a finite dimensional parameter vector denoted by $\theta \in \Theta$ (the parameter space). Otherwise, the model is nonparametric or semiparametric, see Su et al. (2012) in this Handbook for a review of this approach. In the parametric version, the conditional mean function is typically specified as an ARMA model, augmented by additional regressors according to the modeling objectives. We discuss briefly below the specification of the conditional variance as a function of the variables generating $\mathcal{F}_{t-1}$ and of the probability distribution of z_t.

1.2.1.2 Early GARCH Models. The GARCH(1,1) equation,

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\varepsilon_{t-1}^2, \tag{1.2}$$

where ω, β, and α are parameters, is the most widely used formulation. The positivity of σ_t^2 is ensured by the following sufficient restrictions: $\omega > 0$, $\alpha \geq 0$, and $\beta \geq 0$, but if $\alpha = 0$, β must be set to 0 as well, otherwise the sequence σ_t^2 tends to a constant and β is not identifiable. If q lags of ε_t^2 and p lags of σ_t^2 are included (instead of setting $p = q = 1$ as above), the model is named GARCH(p,q), as put forward by Bollerslev (1986). Tests of zero restrictions for the lag coefficients and model choice criteria result in choices of p and q equal to 1 in a vast diversity of data series and sample sizes, with p or q equal to two rarely selected and higher values almost never.

To understand why the GARCH(1,1) equation together with (Eq. 1.1) and the assumptions stated above is able to account for volatility clustering and leptokurtosis, let us note that the autocorrelation coefficients of ε_t^2, denoted by ρ_j, are equal to $\rho_1 = \alpha(1-\beta^2-\alpha\beta)/(1-\beta^2-2\alpha\beta)$, which is larger than α, and $\rho_j = (\alpha+\beta)\rho_{j-1}$ for $j \geq 2$, if $\alpha + \beta < 1$. The last inequality ensures that $\text{Var}(\varepsilon_t) = \omega/(1-\alpha-\beta)$ (denoted by σ^2) exists and that ε_t is covariance stationary. Moreover, the autocorrelations of ε_t^2 are positive and decaying at the rate $\alpha + \beta$. The sum $\alpha + \beta$ is referred to as the persistence of the conditional variance process. For financial return series, estimates of α and β are often in the ranges [0.02, 0.25] and [0.75, 0.98], respectively, with α often in the lower part of the interval and β in the upper part for daily series, such that the persistence is close to but rarely exceeding 1. Hence, ρ_1 is typically small, and the autocorrelations decay slowly, though still geometrically. Table 1.2 reports quasi-maximum likelihood (QML) estimates (see Section 1.2.1.3) for the demeaned S&P 500 returns over the full sample and 12 subsamples of 5 years of data (except for the first and last subsamples, which are a bit longer). The

TABLE 1.2 GARCH(1,1) QML Estimates for S&P 500 Demeaned Returns

Period	T	α	β	$\alpha+\beta$	σ^2	KC	ρ_1
1950-01-06	15,480	0.079	0.915	0.994	1.301	∞	0.405
2011-07-14					0.939	31.62	0.154
2006-01-03	1393	0.092	0.898	0.990	1.762	16.08	0.351
2011-07-14					2.306	11.53	0.209
2001-01-02	1256	0.073	0.920	0.993	1.178	18.34	0.344
2005-12-30					1.320	5.34	0.186
1996-01-02	1263	0.094	0.882	0.977	1.638	4.91	0.238
2000-12-29					1.348	6.71	0.209
1991-01-02	1264	0.026	0.963	0.989	0.389	3.20	0.056
1995-12-29					0.423	5.62	0.027
1986-01-02	1264	0.156	0.755	0.911	1.397	4.20	0.250
1990-12-31					1.632	87.53	0.118
1981-01-02	1264	0.033	0.956	0.989	0.720	3.31	0.076
1985-12-31					0.756	4.77	0.050
1976-01-02	1263	0.044	0.943	0.987	0.602	3.54	0.111
1980-12-31					0.595	4.34	0.128
1971-01-04	1262	0.072	0.923	0.995	1.006	35.83	0.362
1975-12-31					0.885	4.84	0.165
1966-01-03	1233	0.138	0.817	0.955	0.499	5.33	0.285
1970-12-31					0.493	5.99	0.233
1961-01-03	1258	0.237	0.733	0.970	0.550	∞	0.599
1965-12-31					0.417	20.40	0.427
1956-01-03	1260	0.088	0.868	0.956	0.472	3.64	0.158
1960-12-30					0.471	5.29	0.124
1950-01-03	1509	0.020	0.975	0.995	0.651	3.29	0.062
1955-12-30					0.537	12.81	0.104

T, number of observations; σ^2 is estimated as $\omega/(1-\alpha-\beta)$ using the (unreported) estimate of ω, KC using Equation 1.3 with $\lambda = 3$ (∞ means that the existence condition in Equation 1.3 is not satisfied), and ρ_1 as $\alpha(1-\beta^2-\alpha\beta)/(1-\beta^2-2\alpha\beta)$. The data in the second row (for each period) are the sample variance (σ^2 column), kurtosis coefficient (KC column), and first-order autocorrelation of squared returns. Results obtained with GARCH module of OxMetrics 6.20. demeaned returns are defined in Table 1.1.

kurtosis coefficient, defined as $E(\varepsilon_t^4)/\mathrm{Var}(\varepsilon_t)^2$ and denoted by KC, is equal to

$$\mathrm{KC} = \lambda\frac{(1-\alpha^2-\beta^2-\alpha\beta)}{(1-\lambda\alpha^2-\beta^2-2\alpha\beta)}, \quad \text{if } \lambda\alpha^2+\beta^2+2\alpha\beta<1, \tag{1.3}$$

where $\lambda = E(z_t^4)$ is the kurtosis coefficient of z_t, so that KC is larger than λ. In particular, if z_t is Gaussian, $\lambda = 3$ and ε_t is leptokurtic. However, it is not easy to obtain jointly a small value of ρ_1 and a high kurtosis with a small value of

α, a large value of β and a Gaussian distribution for z_t: for example, $\alpha = 0.05$, $\beta = 0.93$, yield $\rho_1 = 0.11$, while KC $= 3.43$ if $\lambda = 3$. If λ is set to 5, KC increases to 6.69. In Table 1.2, it can be seen that estimates of the parameters fit the unconditional variance much better than the first autocorrelation and especially the kurtosis coefficients. The extreme value of the kurtosis in the period 1986–1990 is due to the extreme negative return of 19 October, 1987.

The GARCH(1,1) equation (Eq. 1.2) is the universal reference with respect to which all its extensions and variants are compared. An (almost) exhaustive list of GARCH equations in 2008 is provided in the ARCH glossary of Bollerslev (2008). Several of them are presented by Teräsvirta (2009). The formulation in Equation 1.2 is linear in the parameters but others are not, and most of these are presented in Chapter 2 by Teräsvirta (2012) in this Handbook. A widely used extension introduces an additional term in Equation 1.2, as given in Glosten et al. (1993):

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\varepsilon_{t-1}^2 + \gamma\varepsilon_{t-1}^2 I(\varepsilon_{t-1} < 0). \tag{1.4}$$

With $\gamma = 0$, the conditional variance response to a past shock (ε_{t-1}) of given absolute value is the same whether the shock is positive or negative. The news impact curve, which traces σ_t^2 as a function of ε_{t-1} for given values of $\omega + \beta\sigma_{t-1}^2$ and α, is a parabola having its minimum at $\varepsilon_{t-1} = 0$. If γ is positive, the response is stronger for a past negative shock than for a positive one of the same absolute value and the news impact curve is asymmetric (steeper to the left of 0). This positive effect is found empirically for many (individual and index) stock return series and may be interpreted as the leverage effect uncovered by Black (1976). This effect for a particular firm says that a negative shock—a return below its expected value—implies that the firm is more leveraged, that is, has a higher ratio of debt to stock value, and is therefore more risky, so that the volatility should increase. The extended GARCH model (Eq. 1.4) is named *GJR-GARCH* or just GJR and referred to as an *asymmetric GARCH equation*. There exist several other GARCH equations that allow for an asymmetric news impact effect, in particular, the EGARCH model of Nelson (1991b) and the TGARCH model of Zakoian (1994). The positive asymmetric response of stock return volatility to past shocks is considered as a stylized fact, but there is no consensus that the finding of positive γ estimates corresponds actually to the financial leverage effect. Negative estimates of γ are found for commodity return series as documented in Carpantier (2010), who names it the *inverse leverage effect*. Engle (2011) also provides evidence of this effect for returns of a gold price series, volatility indexes, some exchange rates, and other series and interprets this as a hedge effect (the mentioned type of series are from typical hedge assets).

1.2.1.3 Probability Distributions for z_t. The Gaussian distribution was the first to be used for estimation by the method of maximum likelihood (ML). The likelihood function based on the Gaussian distribution has a QML interpretation, that is, it provides consistent and asymptotically Gaussian estimators of the conditional mean and GARCH equation parameters provided that the conditional

mean and variance are correctly specified. See Bollerslev et al. (1994, Section 2.3) for a short presentation of QML for GARCH and Francq and Zakoian (2010, Chapter 7) for more details and references. The (quasi-)log-likelihood function is based on the assumption of independence of the z_t innovations (even if the latter are only an m.d.s.). For a sample of T observations collected in the vector y, it is written

$$\ell_T(\theta; y) = \sum_{t=1}^{T} \ell(y_t; \theta), \tag{1.5}$$

where $\ell(y_t; \theta) = \log f(y_t|\theta)$, with $f(y_t|\theta)$ the density function of y_t obtained by the change of variable from z_t to y_t implied by Equation 1.1. Actually, $f(y_t|\theta)$ is conditional on $\mathcal{F}_{t-1}$ through μ_t and σ_t^2. For example, if $z_t \sim N(0,1)$ and $y_t \sim N(\mu_t, \sigma_t^2)$, and apart from a constant $\log f(y_t|\theta) = -0.5[\log \sigma_t^2 + (y_t - \mu_t)^2/\sigma_t^2]$. As mentioned above, the Gaussian assumption implies conditional mesokurtosis for y_t (i.e., a kurtosis coefficient equal to 3 for z_t) and unconditional leptokurtosis if a GARCH effect exists, but the degree of leptokurtosis may be too small to fit the kurtosis of the data. For this reason, Bollerslev (1987) proposed to use the t-distribution for z_t, since it implies conditional leptokurtosis and, therefore, stronger unconditional leptokurtosis. The functional expression of $\ell(y_t; \theta)$, if $f(z_t)$ is a t_ν density with ν degrees of freedom, is given by (apart from a constant) $\log[\Gamma((\nu+1)/2)/\Gamma(\nu/2)] - 0.5\{(\nu-2)\sigma_t^2 + (\nu+1)(y_t - \mu_t)^2/[(\nu-2)\sigma_t^2]\}$. Notice that θ includes ν in addition to the parameters indexing μ_t and σ_t^2, and the restriction that ν be larger than 2 is imposed to ensure the existence of the variance of y_t. When $\nu > 4$, the fourth moment exists and the conditional kurtosis coefficient, that is, the λ to be used in Equation 1.3, is equal to $3 + 6(\nu-4)^{-1}$. Another family of distributions for z_t, which is sometimes used in GARCH estimation is the generalized error (GE) distribution indexed by the positive parameter ν. It was proposed by Nelson (1991b). It implies conditional leptokurtosis, if $\nu > 2$; platykurtosis, if $\nu < 2$; and corresponds to the Gaussian distribution, if $\nu = 2$.

The Gaussian, t, and GE distributions are symmetric around 0. The symmetry of the conditional distribution does not necessarily imply the same property for the unconditional one. He et al. (2008) show that conditional symmetry combined with a constant conditional mean implies unconditional symmetry, whatever the GARCH equation is (thus, even if the news impact curve is itself asymmetric). They also show that a time-varying conditional mean is sufficient for creating unconditional asymmetry (even if the conditional density is symmetric), but the conditional mean dynamics has to be very strong to induce nonnegligible unconditional asymmetry. Empirically, the conditional mean dynamics is weak for return series as their autocorrelations are small. Since it is obvious that conditional asymmetry implies the same unconditionally, an easy way to account for the latter, which is not rare in financial return series as illustrated above, is to employ a conditionally asymmetric distribution. Probably the most used asymmetric (or skewed) distributions in GARCH modeling is the skewed-t of Hansen (1994). Bond (2001) surveys asymmetric conditional

TABLE 1.3 GJR-GARCH(1,1) ML Estimates for S&P 500 Demeaned Returns for the Period 2006-07-14 Until 2011-07-14 (1260 Observations)

Density	BIC	α	γ	β	ω	ν	ξ
Gaussian	3.132	−0.0225	0.181	0.915	0.0212	—	—
		(7.3)	(0.9)	(0.0)	(0.0)		
GE	3.082	−0.0209	0.182	0.914	0.0137	1.24	—
		(11)	(0.9)	(0.0)	(2.9)	(0.0)	
t	3.087	−0.0255	0.196	0.920	0.0106	5.05	—
		(6.7)	(0.0)	(0.0)	(7.5)	(0.0)	
Skewed-t	3.074	−0.0289	0.206	0.919	0.0140	5.90	−0.18
		(1.4)	(0.0)	(0.0)	(2.9)	(0.0)	(0.0)

Results obtained with G@RCH module of OxMetrics 6.20. Demeaned returns are defined as in Table 1.1. In parentheses: p-values in percentage.

densities for ARCH modeling. Another way to account for asymmetry and excess kurtosis is to estimate the conditional distribution nonparametrically, as given by Engle and Gonzalez-Rivera (1991)—see also Teräsvirta (2012) in this Handbook.

The use of an asymmetric conditional density often improves the fit of a model as illustrated in Table 1.3—the Bayesian information criterion (BIC) is minimized for the skewed-t choice—and may be useful in Value-at-Risk (VaR) forecasting (see below). The skewed-t-density is indexed by an asymmetry parameter ξ in addition to the degrees of freedom parameter ν also indexing the symmetric t-density used by Bollerslev (1987). A negative ξ corresponds to a left-skewed density, a positive ξ to right skewness, and for $\xi = 0$ the skewed-t reduces to the symmetric t. The estimation results in Table 1.3 show that the conditional skewed-t is skewed to the left, which generates unconditional left skewness, in agreement with the negative skewness coefficient of the data, equal to −0.23. Notice that ξ is not the skewness coefficient, that is, the values −0.18 and −0.23 are not directly comparable in magnitude. The data kurtosis coefficient is equal to 10.9, hence it is not surprising that the estimated degrees of freedom parameter is of the order of 6 for the skewed-t, 5 for the symmetric t, and that the estimated GE parameter value of 1.24 is well below 2. Notice that, perhaps with the exception of ω, the estimates of the GJR-GARCH equation parameters are not sensitive to the choice of the density used for the estimation. An unusual feature of the results are the negative estimates of α, but except in the skewed-t case, α is not significantly different from 0 at the level of 5%.

1.2.1.4 New GARCH Models. Although early GARCH models have been and are still widely used, a viewpoint slowly emerged, according to which these models may be too rigid for fitting return series, especially over a long span. This is related to the rather frequent empirical finding that the estimated persistence of conditional variances is high (i.e., close to 1), as illustrated by the results in Table 1.2. In the GARCH infancy epoch, Engle and Bollerslev (1986) suggested

that it might be relevant to impose the restriction that $\alpha + \beta$ be equal to 1 in the GARCH equation (Eq. 1.2), then named integrated GARCH (IGARCH) by analogy with the unit root literature. However, the IGARCH equation

$$\sigma_t^2 = \omega + \sigma_{t-1}^2 + \alpha(\varepsilon_{t-1}^2 - \sigma_{t-1}^2) \tag{1.6}$$

implies that the unconditional variance does not exist (since $\alpha + \beta < 1$ is necessary for this), and that the conditional expectation of the conditional variance at horizon s is equal to $\omega s + \sigma_{t+1}^2$.[1] Unless $\omega = 0$, there is a trend in $E_t(\sigma_{t+s}^2)$, which is not sensible for long-run forecasting.[2]

Diebold (1986), in his discussion of Engle and Bollerslev (1986), briefly mentions that the high persistence of conditional variances may be provoked by overlooking changes in the conditional variance intercept ω. The intuition for this is that changes in ω (or σ^2) induce nonstationarity, which is captured by high persistence. Lamoureux and Lastrapes (1990) document empirically this idea and show it to be plausible by Monte Carlo (MC) simulation, while Hillebrand (2005) provides a theoretical proof. Another possible type of change is in the persistence itself, as suggested by the results in Table 1.2 for some periods.

The GJR-GARCH equation (Eq. 1.4) has an undesirable drawback linked to the way it models the leverage effect for stocks ($\gamma > 0$). It implies that conditional variances persist more strongly after a large negative shock than after a large positive shock of the same magnitude ($\beta + \alpha + 0.5\gamma > \beta + \alpha$). This is somehow in disagreement with the view that after the October 87 crash, the volatility in US stock markets reverted swiftly to its precrash normal level. Evidence of this based on implied volatilities from option prices is provided by Schwert (1990) and Engle and Mustafa (1992).

All this has promoted the development of more flexible GARCH models, in particular, models allowing for changing parameters. There are many ways to do this, and somewhat arbitrarily, we present a selection of existing models into three classes.

1-Component and smooth transition models. Component models are based on the idea that there is a long-run component in volatilities, which changes smoothly, and a short-run one, changing more quickly and fluctuating around the long-run component. The components may be combined in an additive way or in a multiplicative way. The component model of Engle and Lee (1999) is additive and consists of the equations

$$\sigma_t^2 = q_t + \beta(\sigma_{t-1}^2 - q_{t-1}) + \alpha(\varepsilon_{t-1}^2 - q_{t-1}), \tag{1.7}$$

$$q_t = \sigma^2 + \rho q_{t-1} + \phi(\varepsilon_{t-1}^2 - \sigma_{t-1}^2), \tag{1.8}$$

[1] From the GARCH(1,1) equation, one gets that $E_t(\sigma_{t+s}^2) = \omega + (\alpha + \beta)E_t(\sigma_{t+s-1}^2)$, hence $E_t(\sigma_{t+s}^2) = \omega \sum_{i=1}^{s}(\alpha + \beta)^{i-1} + (\alpha + \beta)^s \sigma_{t+1}^2$. If $\alpha + \beta < 1$, this tends to σ^2 as s tends to ∞, but if $\alpha + \beta = 1$, this diverges because of the linear trend.

[2] The RiskMetrics model (J.P. Morgan, 1996) sets $\omega = 0$ in addition to $\alpha + \beta = 1$ and $\alpha = 0.94$ for daily returns. Thus, it avoids the trend but implies forecasts that stay at the level of date t.

where β, α, σ^2, ρ, and ϕ are parameters. If $\phi = \rho = 0$ and $\alpha + \beta < 1$, the equations above are equivalent to the GARCH(1,1) equation (Eq. 1.2), where $\omega = \sigma^2(1 - \alpha - \beta)$. If ϕ and ρ differ from 0, q_t is an AR(1) process with 0 mean error $\varepsilon_{t-1}^2 - \sigma_{t-1}^2$ (an m.d.s.). If $\rho = 1$, Equation 1.8 has the IGARCH format of Equation 1.6. The equation for σ_t^2 is a GARCH(1,1) allowing for volatility clustering around the component q_t that evolves more smoothly than the σ_t^2 component if $\rho > \alpha + \beta$, which justifies the interpretation of q_t as long-run component. If moreover $\rho < 1$, the forecasts of both q_t and σ_t^2 converge to $\sigma^2/(1 - \rho)$ as the forecast horizon tends to infinity. By combining the Equations 1.7 and 1.8, the model is found to be equivalent to a GARCH(2,2). In an application to the daily S&P 500 returns over the period 1971–1991, Engle and Lee (1999) do not reject the hypothesis that the q_t component is integrated ($\hat{\rho} = 0.9982$), and that shock effects are stronger on σ_t^2 than on q_t ($\hat{\alpha} = 0.089 > \hat{\phi} = 0.032$), while $\hat{\beta} = 0.80$, such that the persistence of the short-run component ($\hat{\alpha} + \hat{\beta} = 0.89$) is much lower than for the long-run one. However, the slowly moving component q_t reverts to a constant level (assuming $\rho < 1$), a feature that does not fit to the viewpoint that the level of unconditional volatility can itself evolve through time, as suggested by the different subsample estimates of σ^2 in Table 1.2. A related additive component model is put forward by Ding and Granger (1996), where the conditional variance is a convex linear combination of two components: $\sigma_t^2 = w\sigma_{1,t}^2 + (1 - w)\sigma_{2,t}^2$. One component is a GARCH(1,1)—$\sigma_{1,t}^2 = \omega_1 + \beta_1\sigma_{1,t-1}^2 + \alpha_1\varepsilon_{t-1}^2$—and the other is an IGARCH equation—$\sigma_{2,t}^2 = (1 - \alpha_2)\sigma_{2,t-1}^2 + \alpha_2\varepsilon_{t-1}^2$. The restriction to IGARCH form with 0 intercept is necessary for identifiability. Bauwens and Storti (2009) extend this model by letting the fixed weight w become time-varying and specifying w_t as a logistic transformation of σ_{t-1}^2. This allows to relax the restriction that one of the components must be integrated. That model is close to a smooth transition GARCH (STGARCH) model. In a STGARCH model, the parameters of the GARCH equation change more or less quickly through time. For example, to allow for a change of the intercept, ω in the GARCH(1,1) equation is replaced by $\omega_1 + \omega_2 G(\varepsilon_{t-1})$, where $G()$ is a "transition" function taking values in $[0, 1]$. For example, if $G(\varepsilon_{t-1}) = \{1 + \exp[-\gamma(\varepsilon_{t-1} - \kappa)]\}^{-1}$, the intercept is close to ω_1 if ε_{t-1} is very negative and to ω_2 if it is very positive. The parameter γ is restricted to be positive and represents the speed of the transition; if it is large, the transition function is close to a step function jumping at the value of κ. The parameter κ represents the location of the transition. Smooth transition models are presented in detail in Chapter 2 by Teräsvirta (2012) in this Handbook. Multiplicative component models are briefly discussed below and in more detail in Chapter 9 by Brownlees et al. (2012b) in this Handbook.

2-Mixture and Markov-switching models. The log-likelihood function of the component model of Ding and Granger (1996) is of the type of Equation 1.5, so that estimation is not complicated. A mixture model is also based on two (or more) variance components $\sigma_{i,t}^2 = \omega_i + \beta_i\sigma_{i,t-1}^2 + \alpha_i\varepsilon_{t-1}^2$ (for $i = 1, 2$) that appear in a mixture of two Gaussian distributions. It is assumed that $\varepsilon_t|\mathcal{F}_{t-1} \sim wN(\mu_1, \sigma_{1,t}^2) + (1 - w)N(\mu_2, \sigma_{2,t}^2)$. The means of the Gaussian

distributions are related by $w\mu_1 + (1-w)\mu_2 = 0$ to ensure that the mixture has a null expectation. This model is a particular "mixed normal GARCH" (MN-GARCH) model, see Haas et al. (2004a) for several extensions. One interpretation of it is that there are two possible regimes: for each t, a binary variable takes one of the values 1 and 2 with respective probabilities of w and $1-w$. Once the regime label is known, the model is a GARCH(1,1) with given mean. One regime could feature a low mean with high variance (bear market) and the other a high mean with low variance (bull), for example, if $\mu_1 < \mu_2$ and $\omega_1/(1-\beta_1-\alpha_1) > \omega_2/(1-\beta_2-\alpha_2)$. Haas et al. (2004a) derive the existence conditions for the fourth-order moments of MN-GARCH models. In the model described above, the unconditional variance exists if $w(1-\alpha_1-\beta_1)/(1-\beta_1) + (1-w)(1-\alpha_2-\beta_2)/(1-\beta_2) > 0$, so that it is not necessary that $\alpha_i + \beta_i < 1$ holds for $i = 1$ and $i = 2$. If $w = 1$, the model reduces to the GARCH(1,1) case and the previous condition to $\alpha_1 + \beta_1 < 1$. The model is useful to capture not only different levels of variance (according to the regimes) but also unconditional skewness and kurtosis, since a mixture of Gaussian densities can have such features. In an application to a series of NASDAQ daily returns over the period 1971–2001, for two components, the ML estimates are $\hat{w} = 0.82$, $\hat{\mu}_1 = 0.09$, $\hat{\alpha}_1 = 0.05$, $\hat{\beta}_1 = 0.92$, $\hat{\mu}_2 = -0.42$, $\hat{\alpha}_2 = 0.51$, and $\hat{\beta}_2 = 0.73$. These values are in agreement with the interpretation suggested above of bull and bear regimes. The second regime thus has $\hat{\alpha}_2 + \hat{\beta}_2 > 1$, yet the variance existence condition holds. The estimates imply a variance level equal to 0.53 in the first variance process and 1.74 in the second, thus on average 1.06. The single regime GARCH(1,1) Gaussian ML estimates are $\hat{\alpha} = 0.12$, $\hat{\beta} = 0.87$, and $\hat{\sigma}^2 = 0.99$. The likelihood ratio statistic is about 140, indicating a much better fit of the MN-GARCH model with two components.

The idea that the regime indicator variables that are implicit in the MN-GARCH model are independent through time does not seem realistic. Intuitively, if the market is bullish, it stays in that state for a large number of periods and likewise if it is bearish. Thus, some persistence is likely in each regime. Following the idea of Hamilton (1989), this is modeled by assuming that the regime indicator variables are dependent, in the form of a Markov process of order 1. Thus, once in a given regime, there is a high probability to stay in the same regime and a low to move to the other regime. This idea can be combined with the existence of two different means and conditional variance processes within each regime, as in the MN-GARCH model (the extension to more than two regimes is obvious). Haas et al. (2004b) develop this type of Markov-switching GARCH model. This model is much easier to estimate than a Markov-switching model featuring path dependence. Such a model is defined by assuming that the parameters of the GARCH equation change according to a Markov process. Let s_t denote a random variable taking the values 1 or 2 in the case of two regimes. Then, if $\varepsilon_t(s_t) = \sigma_t(s_t)z_t$ and $\sigma_t(s_t)^2 = \omega_{s_t} + \alpha_{s_t}\varepsilon_{t-1}(s_{t-1})^2 + \beta_{s_t}\sigma_{t-1}(s_{t-1})^2$, the model features path dependence. This means that to compute the value of the conditional variance at date t, one must know the realized values of all s_τ for $\tau \leq t$. Since the s_t process is latent, the realized values are not known and thus for

estimation by ML, these variables must be integrated out by summation over 2^t possible paths (K^t for K regimes). This renders ML estimation infeasible for the sample sizes typically used in volatility estimation. Notice that path dependence does not occur if $\beta_{s_t} = 0$ for all possible values of s_t, that is, in the ARCH case, see Cai (1994) and Hamilton and Susmel (1994). However, Bayesian estimation of a Markov-switching GARCH model using a MCMC algorithm is feasible, as shown by Bauwens et al. (2010). Chapter 3 by Haas and Paolella (2012) in this Handbook presents in detail the mixture and Markov-switching GARCH models and contains empirical illustrations.

3-Models with a changing level of the unconditional variance. The models in the previous classes (when stationary) have a constant level of unconditional variance even if they let the conditional variances fluctuate around a changing level. This moving level changes smoothly in the model of Engle and Lee (1999), and it changes abruptly in a Markov-switching GARCH model whenever there is a switch. In the third class discussed hereafter, the models are nonstationary since the unconditional variance is time-varying. The level of the unconditional variance is captured either by a smooth function or by a step function, independently of the short-run GARCH dynamics.

The models of Engle and Rangel (2008) and Amado and Teräsvirta (2012) let the unconditional variance change smoothly as a function of time.[3] In their models, Equation 1.1 is extended by including a factor τ_t multiplicatively, as follows:

$$\varepsilon_t = \tau_t \sigma_t z_t. \tag{1.9}$$

In the spline-GARCH model of Engle and Rangel (2008), the factor τ_t is an exponential quadratic spline function with k knots and is multiplied by a GARCH component:

$$\sigma_t^2 = (1 - \alpha - \beta) + \beta\sigma_{t-1}^2 + \alpha(\varepsilon_{t-1}/\tau_{t-1})^2, \tag{1.10}$$

$$\tau_t^2 = \omega \exp\left(\delta_0 t + \sum_{i=1}^{k} \delta_i [(t - t_{i-1})_+]^2\right), \tag{1.11}$$

where β, α, ω, and δ_i are parameters for $i = 0, 1, \ldots, k$, $x_+ = x$ if $x > 0$ and 0 otherwise, and $\{t_0 = 0, t_1, \ldots, t_{k-1}\}$ are time indices partitioning the time span into k equally spaced intervals. The specification of σ_t^2 may be chosen among other available GARCH equations[4] with an adapted identification constraint for the intercept (e.g., $1 - \alpha - \beta - \gamma/2$ for the GJR-GARCH(1,1) equation and a symmetric distribution for z_t). Given this type of constraint on the constant of the GARCH equation, it is obvious that $\mathrm{Var}(\varepsilon_t) = \tau_t^2$, so that the τ_t^2 component is interpretable as the smoothly changing unconditional variance, while σ_t^2 is the component of the conditional variance capturing the

[3]Another model with this feature is the STGARCH model where the variable triggering the transitions is the index of time, see Section 2.4.7 in Chapter 2 Teräsvirta (2012) in this Handbook.
[4]Notice that ε_{t-1} is divided by τ_{t-1} in Equation 1.10.

clustering effect. The model of Amado and Teräsvirta (2012) uses a transition function type of functional form for τ_t, see Section 8 of Teräsvirta (2012) in this Handbook for more details. Baillie and Morana (2009) put forward an additive model where the unconditional variance component τ_t^2 evolves smoothly via another type of function known as the *Fourier flexible form*, given by $\omega + \sum_{i=1}^{k}[\gamma_i \sin(2\pi it/T) + \delta_i \cos(2\pi it/T)]$. The GARCH component of their model is a fractionally integrated one (FIGARCH) that is useful to capture a long-memory aspect in squared returns, see Baillie et al., (1996). Table 1.4 shows the ML estimates of the spline-GARCH model with three knots for the period January 2006 to mid-July 2011, and Figure 1.4 displays the estimated spline component, which clearly reflects the volatility bump due to the financial

TABLE 1.4 Three Knot Spline-GARCH(1,1) ML Estimates for S&P 500 Demeaned Returns for the Period 2006-01-03 Until 2011-07-14 (1393 Observations)

Parameter	Estimate	*p*-Value, %
ω	0.727	0.04
δ_0	−4.432	39.7
δ_1	21.48	8.50
δ_2	−49.71	1.40
δ_3	49.56	0.62
α	0.077	0.00
β	0.892	0.00

Results obtained with OxMetrics 6.20. Demeaned returns are defined as in Table 1.1.

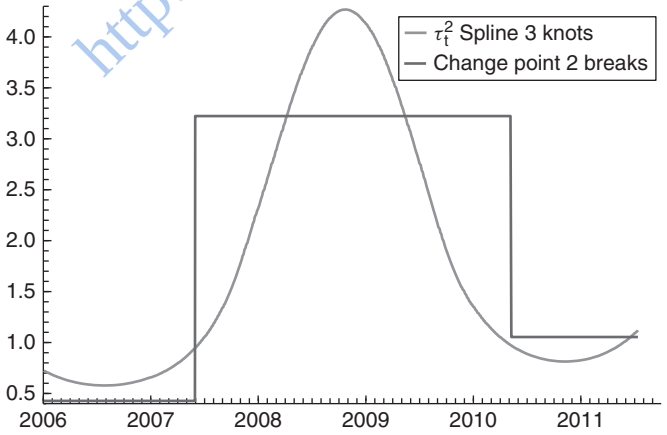

FIGURE 1.4 Three knot spline-GARCH component and variance of change-point model with two breaks of S&P 500 index demeaned returns, 2006-01-03/2011-07-14 (1393 observations).

crisis of 2008 and anticipates the increase of the summer of 2011. The persistence of the conditional variance component σ_t^2 is estimated to be 0.97 versus 0.99 in the simple GARCH(1,1) model (Table 1.2).

In Chapter 10 by Van Bellegem (2012) in this Handbook, more general multiplicative models that feature nonstationarity are reviewed. The spline-GARCH model is also briefly presented with an empirical illustration.

Models allowing sudden changes in the level of the unconditional variance are scarce. The model of Amado and Teräsvirta (2012) has this feature if the transition function becomes a step function (or a superposition of such functions). He and Maheu (2010) propose a change-point GARCH model based on the change-point modeling framework of Chib (1998). It is a Markov switching model that excludes recurrent states: once in a given state, the time series can only stay in it (with some probability) or move to the next state (with the complementary probability). He and Maheu (2010) use this approach for the univariate GARCH(1,1) model (with 0 mean and student errors), using a particle filter for implementing Bayesian estimation. Applying such a model (with Gaussian innovations) to the same data as for the spline-GARCH model above and assuming two breaks, the estimated unconditional variance increases from 0.43 to 3.22 on 2007-05-31 and decreases to 1.05 on 2010-12-10. This is shown graphically by the piecewise constant line in Figure 1.4. The estimates of α and β for the three successive regimes are $(0.034, 0.900)$, $(0.099, 0.890)$, and $(0.002, 0.753)$. For details on algorithms and model choice in this type of models, see Bauwens et al. (2011).

1.2.1.5 Explanation of Volatility Clustering. According to financial theory, the price of an asset should equal the expected present value of its future income flows. An asset price then changes because the expectations of investors about these future incomes change over time: as time passes, new information (news) about these is released, which modifies the expectations. This explains why prices and, hence, returns are random and therefore volatile. Volatility fluctuates over time because the contents and the arrival rate of news fluctuate. For example, crisis periods correspond to more news releases: in particular, bad news tend to happen in clusters. Volatility clustering is thus due to clusters of arrivals of different types of news. For a more extensive discussion, see Engle (2004). This fundamental explanation is difficult to test empirically. For the example of the S&P 500 index returns, there are many types of news that might be relevant in different importance: news affecting the constituent stocks (earnings announcements, profit warnings, etc.) and the industries to which they belong, news affecting the real activity of the US economy, news about the monetary policy... The contents of these news must be measured. The way they affect volatility is through expectations of many investors, raising an issue of aggregation. It is not known how these expectations are formed, and it is likely that there is a degree of heterogeneity in this process. Parke and Waters (2007) provide an evolutionary game theory model based on heterogeneous agents, who form different types of expectations and adjust these over time. The model is able to generate volatility clustering.

Thus, at best, a reduced form, partial, approach is feasible, that is, relating volatility to some macroeconomic variables and news measures. Relevant papers about the relation between macroeconomic variables and stock volatility include Schwert (1989a, b) and Engle and Rangel (2008). In the latter, the authors use the estimated unconditional variances (the τ^2 of Equation 1.11) of their spline-GARCH model to compute time series of annualized estimates of volatilities for different countries and relate them to macroeconomic variables through a panel data model. Other authors study the impact of news announcements on the intraday volatility of exchange rate returns using a GARCH model, by including variables representing news occurrences and measurements (Degennaro and Shrieves, 1997; Melvin and Yin, 2000; Bauwens et al., 2005).

1.2.1.6 Literature and Software. Extensive surveys of GARCH include Bollerslev et al. (1992), Bera and Higgins (1993), Bollerslev et al. (1994), Diebold and Lopez (1996), Pagan (1996), Palm (1996), Shephard (1996), Li et al. (2002), Giraitis et al. (2006), Teräsvirta (2009), and Hafner (2008). Introductory surveys include Engle and Patton (2001), Engle (2001, 2004), and Diebold (2004). Introductory econometric textbooks briefly mention or explain ARCH (see Stock and Watson (2007) and Wooldridge (2009)) intermediate and advanced books provide more details (Hamilton, 1994; Greene, 2011; Tsay, 2002; Verbeek, 2008). Specialized books are Gouriéroux (1997), Francq and Zakoian (2010), and Xekalaki and Degiannakis (2010). Andersen et al. (2009) contains nine chapters on GARCH modeling.

Several well-known software for econometrics and statistics (EVIEWS, OxMetrics, SAS, SPSS, STATA) contain menu-driven modules for GARCH modeling, avoiding the need to program inference tools for applying GARCH models. See Laurent (2009) for the OxMetrics documentation.

1.2.1.7 Applications of Univariate GARCH. Univariate GARCH models are useful for financial applications such as option pricing and risk measurement.

Option pricing. We take the example of a European call option. Such an option is an acquired right to buy a security (called the *underlying*) at a price (the *premium*) set in advance (the *exercise price*) and at a fixed date (the *maturity*). It is well known that the value of an option is a function of several parameters, among which is the volatility of the return on the underlying security until the maturity. According to the financial theory, see Cox and Ross (1976), "options are priced *as if* investors were risk-neutral and the underlying asset expected return were equal to the risk-free interest rate" (denoted by r below). This is called *"risk-neutral" pricing*. Let C_t^T denote the premium at t for maturity T, $T-t$ thus being the time to maturity. Let P_T be the random value of the underlying security at T and K the exercise price. Then,

$$C_t^T = \mathrm{e}^{-r(T-t)} E_Q[\max(P_T - K, 0)] \tag{1.12}$$

is the discounted expected cash flow of the option, where the expected value is computed using Q, the "risk-neutral" probability distribution. So, C_t^T is

a function of r, $T-t$, K, and the parameters of Q, which determine the variance of the return on the underlying. The risk-neutral density function must have as expected return (until maturity) the risk-free interest rate, and its variance must be the same as in the process generating the observed returns. Duan (1995) showed that risk-neutralization cannot be used with a GARCH model both for the unconditional variance and the conditional variance at all horizons. He uses a "locally" risk-neutral probability for GARCH processes, that is, for the one-step-ahead conditional variance. For the GARCH process defined by Equations 1.1 and 1.2, where $z_t \sim N(0,1)$, the locally risk-neutralized process is given by $y_t = r + v_t$, where $v_t = \mu_t - r + \varepsilon_t$ is $N(0, \sigma_t^2)$ and $\sigma_t^2 = \omega + \alpha(v_{t-1} - \mu_{t-1} + r)^2 + \beta\sigma_{t-1}^2$. The parameters of Q are denoted by θ and consist of the parameters indexing μ_t in addition to (ω, α, β). Thus, denoting $C_t^T = C_t^T(r, K, \theta)$, the premium can be computed by numerical simulation if θ is known or, in practice, replaced by an estimate. Given N simulated realizations $\{P_{T,i}\}_{i=1}^N$ of P_T using the risk-neutralized process,[5] $C_t^T(r, K, \theta)$ is estimated by $\widehat{C}_t^T(\theta) = \mathrm{e}^{-r(T-t)} \frac{1}{N} \sum_{i=1}^N \max(P_{T,i} - K, 0)$. Bauwens and Lubrano (2002) apply this procedure in a Bayesian setup, which makes it possible to compute a predictive distribution of the premium and not only a point estimate as is the case when θ is simply replaced by a point estimate. Such predictive distributions have a positive probability mass at 0, corresponding to the probability that the option will not be exercised, while the remaining probability is spread over the positive values through a continuous density function. Among many others, some references about option pricing in relation with GARCH models are Noh et al. (1994), Kallsen and Taqqu (1998), Hafner and Herwartz (2001), and Rombouts and Stentoft (2009).

Value-at-risk. The VaR of a financial position provides a quantitative measure of the risk of holding the position. It is an estimate of the loss that may be incurred over a given horizon, under normal market conditions, corresponding to a given statistical confidence level. For example, an investor holding a portfolio of stocks might say that the daily VaR of his trading portfolio is €5 million at the 99% confidence level. That means there is 1 chance in 100 that a loss >€5 million will occur the next day under normal market conditions. Indeed, the VaR is a quantile (the 1% quantile in the example above) of the probability distribution of the position. The distribution can be, for example, the conditional distribution implied by a GARCH model estimated at the date when the VaR must be computed. The model is estimated using historical data of past returns on the portfolio and provides a value of the 1% quantile of the next day return distribution. Multiplying this quantile by the portfolio value gives the VaR estimate.

Formally, assume that $y_t = \mu_t + \sigma_t z_t$, where σ_t is defined by a GARCH equation and $z_t \sim N(0,1)$. Let n_α be the left quantile at $\alpha\%$ of the $N(0,1)$ distribution, and $n_{1-\alpha}$ be the right quantile at $\alpha\%$ (e.g., $n_1 = -n_{99} = -2.326$). The one-step-ahead VaR (computed at date $t-1$) for a long position of

[5] Since $P_T = P_t \prod_{s=t+1}^T (1 + y_s)$, one must simulate sequentially the returns $y_{t+1}, y_{t+2}, \ldots, y_T$ from the risk-neutral GARCH process.

€1 is given by $\text{VaR}_t(\alpha) = \mu_t + n_\alpha \sigma_t$. For a short position, $\text{VaR}_t(1-\alpha) = \mu_t + n_{1-\alpha}\sigma_t$. In practice, the GARCH model is estimated with data until date $t-1$, and μ_t and σ_t are replaced by their one-step-ahead forecast in the VaR formula. If we assume another distribution for z_t, we use its quantiles. For example, for a t-density with ν degrees of freedom, we replace ν by its ML estimate and find the corresponding quantiles. Angelidis et al. (2004) evaluate GARCH models for VaR and illustrate that the use of a t-density instead of a Gaussian one improves VaR forecasts. Giot and Laurent (2003) show that the use of a skewed-t instead of a symmetric distribution may be beneficial. VaR forecasts are evaluated using statistical tests (Kupiec, 1995; Christoffersen 1998; Engle and Manganelli, 2004).

1.2.2 MULTIVARIATE GARCH

Multivariate ARCH models appeared almost at the same time as univariate models. Kraft and Engle (1982) was a first attempt, and Engle et al. (1984) put forward a bivariate ARCH model, applied to the forecast errors of two competing models of US inflation, so that their conditional covariance matrix adapts over time. The move to financial applications was done by Bollerslev et al. (1988) who also extended multivariate ARCH to GARCH. They used the capital asset pricing model (CAPM) in the framework of conditional moments rather than unconditional moments. The multivariate GARCH (MGARCH) model of that paper, known as the *VEC model*, has too many parameters to be useful for modeling more than two asset returns jointly. A natural research question was then to design models that can be estimated for larger dimensions. Important milestones are the BEKK model of Engle and Kroner (1995), the factor model of Engle et al. (1990), and the constant conditional correlation (CCC) model of Bollerslev (1990). The latter was followed 12 years later by the time-varying correlation (TVC) model of Tse and Tsui (2002) and the dynamic correlation model (DCC) of Engle (2002a).

In this section, we review briefly the conditional correlation models and factor models. Chapter 4 by Sheppard (2012) in this Handbook is partly complementary to what follows, since it contains more models and is oriented by their use in forecasting. Some surveys and books cited in Section 1.2.1.6 cover the topic of MGARCH models (Bollerslev et al., 1994; Hafner 2008; Francq and Zakoian, 2010). More detailed and extensive surveys of MGARCH models are those of Silvennoinen and Teräsvirta (2009) and Bauwens et al. (2006). The discussion paper version of the latter (Bauwens et al. (2003)) includes a review of applications of MGARCH models to asset pricing, volatility transmission, futures hedging, Value-at-Risk, and the impact of financial volatility on the level and volatility of macroeconomic variables. In Chapter 5 of this Handbook, Hashmi and Tay (2012) apply factor models that not only allow for volatility spillovers between different stock markets but also for time-varying skewness and spillovers in skewness effects. Multivariate models can be used also for pricing options that are written on more than a single underlying asset, so that their price depends on the correlations between the assets (Rombouts and Stentoft, 2011).

1.2.2.1 Structure of MGARCH Models. We denote by y_t a column vector of N asset returns, by μ_t the vector of conditional expectations of y_t, and by $\Sigma_t = (\sigma_{tij})$ the conditional variance–covariance matrix of y_t. The elements of μ_t and Σ_t must be measurable with respect to the σ-field $\mathcal{F}_{t-1}$ generated by y_{t-j} for $j \geq 1$ and possibly by other variables available at time $t-1$. An MGARCH model for y_t is then defined by

$$y_t - \mu_t = \varepsilon_t = \Sigma_t^{1/2} z_t, \tag{1.13}$$

where $\Sigma_t^{1/2}$ is any square matrix such that $\Sigma_t = \Sigma_t^{1/2}(\Sigma_t^{1/2})'$ and z_t is an unobservable random vector belonging to an i.i.d. process, with mean equal to 0 and variance–covariance equal to an identity matrix, $E(z_t) = 0$ and $\mathrm{Var}(z_t) = I_N$. It follows that $\Sigma_t = \mathrm{Var}(y_t|\mathcal{F}_{t-1}) = \mathrm{Var}_{t-1}(y_t)$, so that $\mathrm{Var}_{t-1}(\varepsilon_t) = \Sigma_t$ (note that $E_{t-1}(\varepsilon_t) = 0$). The model is parametric and the definition is complete when the pdf of z_t is defined and the functional form of μ_t and Σ_t is specified. These functions are altogether indexed by a parameter vector of finite dimension. In what follows, we assume that $\mu_t = 0$ and concentrate on the specification of the other elements.

Concerning the pdf of z_t, the reference is the multivariate Gaussian, that is, $z_t \sim N(0, I_N)$, since it provides the basis of QML estimation as in the univariate case. The quasi-log-likelihood function of a sample of T observed vectors y_t (altogether denoted by Y) for a model defined by Equation 1.13 and for known initial observation is

$$\ell_T(\theta; Y) = -\frac{1}{2}\sum_{t=1}^{T}(\log|\Sigma_t| + \varepsilon_t'\Sigma_t^{-1}\varepsilon_t), \tag{1.14}$$

where θ denotes the vector of parameters appearing in μ_t, Σ_t, and in the pdf of z_t (if any). Another choice of density for ε_t is the multivariate t. Multivariate skewed distribution, such as the skewed-t of Bauwens and Laurent (2005), can also be used. As in the univariate case, distributions with fat-tails and skewness are usually better fitting data than the Gaussian, see Giot and Laurent (2003) for an example in the context of Value-at-Risk evaluation.

1.2.2.2 Conditional Correlations. In conditional correlation models, what is specified is the conditional variances σ_{tii} (equivalently denoted by σ_{ti}^2) for $i = 1, 2, \ldots, N$, and the conditional correlations ρ_{tij} for $i < j$ and $j = 2, 3, \ldots, N$. The conditional covariance σ_{tij} is equal to $\rho_{tij}\sigma_{ti}\sigma_{tj}$. In matrix notations,

$$\Sigma_t = D_t R_t D_t, \tag{1.15}$$

where $D_t = \mathrm{diag}(\sigma_{t1}, \sigma_{t2}, \ldots, \sigma_{tN})$ is a diagonal matrix with σ_{ti} as ith diagonal element, and $R_t = (\rho_{tij})$ is the correlation matrix of order N (implying $\rho_{tii} = 1$ $\forall i$ and $\forall t$). The matrix Σ_t is positive-definite if σ_{ti}^2 is positive for all i and R_t is positive-definite.

With this approach, the specification of Σ_t is divided into two independent parts: a model choice for each conditional variance and a choice for the conditional correlation matrix.[6] Concerning the first part, an important simplification is obtained in QML estimation if each conditional variance is specified as a function of its own lags and the ith element of ε_t (denoted by ε_{ti}), for example, by a GARCH(1,1) equation written as

$$\sigma_{ti}^2 = \omega_i + \beta_i \sigma_{t-1,i}^2 + \alpha_i \varepsilon_{t-1,i}^2 \tag{1.16}$$

or any other univariate GARCH equation (Section 1.2.1). This type of model excludes transmission (or spillover) effects between different assets, that is, the presence of terms involving $\varepsilon_{t-1,j}$ or $\sigma_{t-1,j}^2$ for $j \neq i$ in the previous equation. To explain why the assumption of no spillovers simplifies the estimation of conditional correlation models, we substitute $D_t R_t D_t$ for Σ_t in Equation 1.14 to define "degarched" returns

$$\tilde{\varepsilon}_t = D_t^{-1} \varepsilon_t \tag{1.17}$$

and split the likelihood function into two parts:

$$\ell_T(\theta; Y) = -\frac{1}{2} \sum_{t=1}^{T} \left(2 \log |D_t| + \log |R_t| + \tilde{\varepsilon}_t' R_t^{-1} \tilde{\varepsilon}_t \right) \tag{1.18}$$

$$= -\frac{1}{2} \sum_{t=1}^{T} \left(2 \log |D_t| + \tilde{\varepsilon}_t' \tilde{\varepsilon}_t \right) \tag{1.19}$$

$$- \frac{1}{2} \sum_{t=1}^{T} \left(\log |R_t| + \tilde{\varepsilon}_t' R_t^{-1} \tilde{\varepsilon}_t - \tilde{\varepsilon}_t' \tilde{\varepsilon}_t \right). \tag{1.20}$$

It is clear that Equation 1.19 depends only on the parameters (denoted by θ_V) of the conditional variances that appear in D_t, while Equation 1.20 depends on the whole θ that includes, in addition to θ_V, the parameters (denoted by θ_C) of the conditional correlation matrix R_t. If there are no spillover terms in the conditional variance equations, maximizing Equation 1.19 with respect to θ_V provides a consistent and asymptotically normal estimator under usual regularity conditions. Moreover, it is easy to see that Equation 1.19 itself can be split into N functions that correspond to the quasi-log-likelihood functions of univariate

[6] A generalization of this model is the class of copula-MGARCH models. Such models are specified by univariate marginal GARCH models for each asset, and a copula function capturing the dependence between the different assets. If the margins are Gaussian and the copula is multivariate Gaussian, the dependence is captured by the correlation matrix. Other copula function can be used to model dependence in a more refined way, see Jondeau and Rockinger (2006) and Patton (2006a) for examples. Chapter 12 by Heinen and Valdesogo (2012) in this Handbook reviews copula-based volatility models.

GARCH models.[7] Once θ_V is estimated, its value can be injected in Equation 1.20 and the latter maximized with respect to θ_C. To do this, the term $\tilde{\varepsilon}_t'\tilde{\varepsilon}_t$ can be neglected in Equation 1.20, since it does not depend on θ_C.

The separate estimation of each conditional variance model and of the correlation model is the key to enable estimation of MGARCH models of conditional correlations when N is large, where large means more than, say, 5. The price to pay for this is the impossibility of including spillover terms in the conditional variance equations. If spillover effects are included, one can, in principle, maximize Equation 1.19 with respect to θ_V, and then Equation 1.20, where θ_V is replaced by the previous estimate, with respect to θ_C. The first step of maximization will be limited by the dimension of θ_V, which is of order N^2 if all spillover terms are included in each conditional variance equation.

Models for R_t. Several specifications are available from the literature. The challenge is to ensure that R_t be positive-definite and not depending on so many parameters by which the model is not estimable. Bollerslev (1990) solves the issue by setting $R_t = R\ \forall t$, that is, by assuming CCCs, where R is a correlation matrix. Notice that R has $N(N-1)/2$ parameters, but they can be estimated easily even if N is large. It follows from Equation 1.18 that if $R_t = R\ \forall t$, and if D_t is known, the ML estimator of R, given by

$$\widehat{R} = \frac{1}{T}\sum_{t=1}^{T}\tilde{\varepsilon}_t\tilde{\varepsilon}_t', \tag{1.21}$$

is consistent (under usual regularity conditions and if $T > N$) and remains so if D_t is replaced by a consistent estimator for all t (obtained by computing D_t using the consistent estimator of θ_V resulting from the maximization of Equation1.19 as explained above). In finite samples, the diagonal elements of $\widehat{R}$ are not exactly equal to 1, so that $\widehat{R}$ should be transformed to a correlation matrix. This is done by replacing the elements of $\widehat{R}_t$ by $\widehat{\rho}_{tij}/\sqrt{\widehat{\rho}_{tii}\widehat{\rho}_{tjj}}$. In matrix notation, the transformed matrix is

$$\widetilde{R} = (I_N \odot \widehat{R})^{-1/2}\widehat{R}(I_N \odot \widehat{R})^{-1/2}, \tag{1.22}$$

where the symbol $\odot$ is the element by element multiplication operator (Hadamard product).

The hypothesis of CCCs is not tenable except for specific cases and short periods. Several tests of the null hypothesis of constant correlations exist: see Longin and Solnik (1995), Tse (2000), Engle and Sheppard (2001), Bera and Kim (2002a), and Silvennoinen and Teräsvirta (2005). The tests differ because of the specification of the alternative hypothesis. Smooth transition-type CCC models are proposed by Silvennoinen and Teräsvirta (2005) and Silvennoinen and Teräsvirta (2007).

[7] If the GARCH equations are as in Equation 1.6, θ_V consists of the vectors $(\omega_i\ \beta_i\ \alpha_i)$, $i = 1, 2, \ldots, N$.

The CCC model has been generalized in different ways, so that the conditional correlations change over time. One dynamic model of conditional correlations is the TVC model of Tse and Tsui (2002). The dynamic process generating R_t is specified as

$$R_t = (1 - \alpha - \beta)R + \beta R_{t-1} + \alpha S_{t-1}, \tag{1.23}$$

where β and α are scalar parameters, R is a constant correlation matrix parameter, and $S_t = (s_{tij})$ is the correlation matrix computed from the past degarched returns $\tilde{\varepsilon}_{t-1}, \tilde{\varepsilon}_{t-2}, \ldots, \tilde{\varepsilon}_{t-M}$, with $M > N$ to ensure that S_t be positive-definite. Thus, if R_0 is a positive-definite correlation matrix, α and β are positive and satisfy $\alpha + \beta < 1$, R_t is a positive-definite correlation matrix for all t. By writing the above constant part of R_t as $(1 - \alpha - \beta)R$, R is interpretable as the expected value of R_t. Hence R is estimated consistently by $\widetilde{R}$ defined in Equation 1.22. This can be used to ease estimation of the model when N is large: instead of maximizing Equation 1.20 with respect to R, α, and β, we can replace R by $\widetilde{R}$ in Equation 1.20 and maximize it with respect to α and β, that is, only two parameters instead of $2 + N(N-1)/2$. This procedure is called *correlation targeting* (or *tracking*) and is unavoidable if N is large.

Another generalization of the CCC model is the DCC model of Engle (2002a), who specifies the dynamic process on the variance–covariance matrix of $\tilde{\varepsilon}_t$, denoted by Q_t, and transforms it to the correlation matrix R_t:

$$Q_t = (1 - \alpha - \beta)Q + \beta Q_{t-1} + \alpha \tilde{\varepsilon}_{t-1}\tilde{\varepsilon}'_{t-1}, \tag{1.24}$$

$$R_t = (I_N \odot Q_t)^{-1/2} Q_t (I_N \odot Q_t)^{-1/2}. \tag{1.25}$$

where β and α are scalar parameters and Q is a $N \times N$ symmetric and positive-definite matrix parameter. If Q_0 is symmetric and positive-definite and β and α satisfy the same restrictions as in the TVC model above, Q_t is symmetric and positive-definite and R_t is a correlation matrix for all t. The parameter matrix Q can be estimated by $\widehat{R}$ as defined in Equation 1.21 and inserted in Equation 1.20 to ease estimation as explained for the TVC model.[8] However, Aielli (2009) showed that the estimation of Q by $\widehat{R}$ is inconsistent since $E(\tilde{\varepsilon}_t\tilde{\varepsilon}'_t) = E\left(E\left(\tilde{\varepsilon}_t\tilde{\varepsilon}'_t|\mathcal{F}_{t-1}\right)\right) = E\left(R_t\right) \neq E\left(Q_t\right)$. He proposes a consistent specification of Q_t (cDCC, consistent DCC),

$$Q_t = (1 - \alpha - \beta)Q + \beta Q_{t-1} + \alpha P_t \tilde{\varepsilon}_{t-1}\tilde{\varepsilon}'_{t-1} P_t, \tag{1.26}$$

where $P_t = \operatorname{diag}\left(q_{t11}^{1/2}, q_{t22}^{1/2}, \ldots, q_{tNN}^{1/2}\right) = (I_N \odot Q_t)^{1/2}$, so that, by construction, Q is the unconditional variance–covariance matrix of $P_t\tilde{\varepsilon}_t$. The available empirical evidence suggests that the cDCC and DCC estimates are close to each other.

[8] Sheppard (2012) in this Handbook (Section 4.4.6) reviews alternative estimation methods to the maximization of Equation 1.20, which are especially useful when N is large.

Although estimation for a large dimension is in principle feasible, it raises problems. First, as one sees in Equation 1.18 directly, a large matrix (R_t) has to be inverted for each observation in the sample, which is time consuming for the sample sizes typically used in applications, and may raise numerical difficulties for large N. Even if estimation is done in two steps, based on Equation 1.20, with or without correlation targeting, the same issue arises. A model that circumvents this problem is the dynamic equicorrelation (DECO) model of Engle and Kelly (2008). Secondly, Engle et al. (2008) show by simulation that the QML estimates of α and β in the DCC model (estimated with correlation targeting) are subject to a bias problem (toward 0) that is more and more acute as the dimension N increases relative to the sample size T. The source of the problem seems to be that when N approaches T, the estimator (Eq. 1.21) is ill-conditioned, as it approaches a singularity. Better conditioned estimators can be built and are useful to reduce the bias problem, see Hafner and Reznikova (2010b).

Both the TVC and DCC models extend the CCC model by making the conditional correlations time-varying. Notice that a test of $\alpha = \beta = 0$ can be based on the Wald statistic to test the null hypothesis of CCCs. The fact that only two additional parameters suffice to render the correlations time-varying is very useful to deal with large dimensions, but of course, the price to pay is the constraint that all the correlations have the same dynamic pattern. This may be viewed as unrealistic. Several extensions of the DCC model have been proposed to relax this constraint, at the price of introducing more parameters in the process of Q_t and thus being applicable only for moderate values of N (say up to 5 or 10 depending on the number of additional parameters): see Engle (2002a), Billio and Caporin (2006), and Hafner and Franses (2009). Cappiello et al. (2006) introduce asymmetric ("leverage") effects in a DCC model. More extensions are on the agenda of several researchers.

The conditional correlation models described above share the same feature: that the conditional correlations revert to a constant level. Like for univariate GARCH models, this is considered too restrictive for long data series. In particular, the correlation level usually increases in periods of financial turbulence. Thus, models that allow for a smoothly changing level of the correlations are in development. The DCC-MIDAS model of Colacito et al. (2011) and the factor spline-GARCH model for high and low frequency correlations of Rangel and Engle (2009) are of this type. The latter is reviewed in Section 4.4.3 of this Handbook.

1.2.2.3 Factor Models. Factor MGARCH models rest on the idea that the volatilities of assets might be driven by a few common forces. This is related to factor models in finance where excess returns of financial assets are related to factors such as the market excess return in the CAPM or macroeconomic and financial factors, though it should be noted that these models were developed to explain the cross-section of returns rather than their time-series evolution. In the MGARCH literature, the factor structure is a convenient way to reduce the number of parameters with respect to the VEC and BEKK models. Basically, the factor structure says that the unexpected excess return vector $\varepsilon_t = y_t - \mu_t$

(of N elements) is a linear function of p factors (with $p < N$) collected in the vector f_t:

$$\varepsilon_t = Bf_t + \nu_t, \tag{1.27}$$

where B is a matrix of factor loadings, of dimension $N \times p$ and rank equal to p, and ν_t is a white noise vector, called the *indiosyncratic noise*. Assuming that $\text{Var}_{t-1}(\nu_t) = \text{Var}(\nu_t) = \Psi$ with Ψ of full rank, that $\text{Var}_{t-1}(f_t) = \Phi_t$, and that $\text{Cov}(f_t, \nu_t) = 0$, the conditional variance–covariance matrix of ε_t is given by

$$\Sigma_t = B\Phi_t B' + \Psi, \tag{1.28}$$

which is positive-definite.

The specification is completed by a choice of an MGARCH process for Φ_t. The most simple choice is to constrain Φ_t to be a diagonal matrix of univariate GARCH processes, $\Phi_t = \text{diag}(\phi_{t1}^2, \phi_{t2}^2, \ldots, \phi_{tp}^2)$, as in Engle et al. (1990).[9] Thus if $p = 1$, this yields $\Sigma_t = BB'(\omega_1 + \beta_1\phi_{t-1,1}^2 + \alpha_1 f_{t-1,1}^2) + \Psi$. If y_t is a vector of stock returns, the factor can be chosen as the market return. Another choice is to take the factor as a linear combination of ε_t denoted by $\lambda_1'\varepsilon_t$, where λ_1 is a vector of weights that can be estimated (after normalizing their sum to unity). This model implies that the conditional variances of the unexpected returns have the same dynamics. If λ_1 is known, the number of parameters to estimate is $N + 3 + N(N+1)/2$. The $N(N+1)/2$ elements of Ψ can be estimated by covariance targeting and injected in the (Gaussian) log-likelihood to estimate the remaining parameters. If the factor is observed directly (such as the market return), the parameters of its conditional variance can be estimated in a preliminary step, as an univariate GARCH model.

One can add more factors provided some identification restrictions are imposed, see Bauwens et al. (2006) for details. For two factors $f_{ti} = \lambda_i'\varepsilon_t$ $(i = 1, 2)$, these restrictions are that $B_1'\lambda_2 = B_2'\lambda_1 = 0$, where B_i is the ith column of B. It follows that the two factors have a constant conditional covariance. The fact that factors are conditionally correlated can be viewed as a drawback since they may catch similar features of the data. Several factor models avoid this feature; the orthogonal GARCH (O-GARCH) model of Alexander and Chibumba (1997), the generalized orthogonal GARCH model of van der Weide (2002), the full factor GARCH model of Vrontos et al. (2003), and the generalized orthogonal factor GARCH model of Lanne and Saikkonen (2007). See the surveys of Bauwens et al. (2006) and Silvennoinen and Teräsvirta (2009) for more information.

If the viewpoint is taken that the factors can be correlated, then a DCC model can be chosen for the factor vector f_t, for the idiosyncratic noise vector ν_t, or for the correlations between f_t and ν_t. Engle (2009b) and Rangel and Engle (2009) contain such extensions (in a single factor model) that enrich considerably the conditional correlation structure of the factor model.

[9] If Φ_t is not diagonal and the off-diagonal elements are constant, they can be absorbed in Ψ.

1.3 Stochastic Volatility

An alternative to GARCH-type models is the class of SV models, which postulate that volatility is driven by its own stochastic process. For example, the standard Gaussian autoregressive SV model in discrete time, as first introduced in this form by Taylor (1982), is given by

$$y_t - \mu_t = \varepsilon_t = \sigma_t z_t \quad z_t \sim N(0,1), \tag{1.29}$$

$$\log \sigma_{t+1}^2 = \omega + \beta \log \sigma_t^2 + \sigma_u u_t, \quad u_t \sim N(0,1), \tag{1.30}$$

and, in the standard case, the innovations z_t and u_t are independent. This discrete time model can be thought of as the Euler approximation of an underlying diffusion model,

$$\mathrm{d}p(t) = \sigma(t)\mathrm{d}W_1(t), \tag{1.31}$$

$$\mathrm{d}\log\sigma(t)^2 = \omega + \phi \log\sigma(t)^2 + \sigma_u \mathrm{d}W_2(t), \tag{1.32}$$

where $\mathrm{d}p(t)$ denotes the logarithmic price increment (i.e., $\mathrm{d}p(t) = \mathrm{d}\log P(t)$), and $W_1(t)$ and $W_2(t)$ are two independent Wiener processes.

The major difference to GARCH models is that, conditional on the information set $\mathcal{F}_{t-1}$, volatility σ_t^2 is not known but rather an unobserved random variable. As we will see, this renders estimation and inference of SV models more complicated than for GARCH models. On the other hand, SV models have some advantages compared with GARCH models. For example, SV models offer a natural economic interpretation of volatility, are easier to connect with continuous-time diffusion models with SV, and are often found to be more flexible in the modeling of financial returns.

The economic motivation is based on the so-called mixture-of-distributions hypothesis, which states that financial returns are driven by a convolution of two random variables as in Equation 1.29, one being an independent noise term, the other a stochastic process representing an information arrival process. For example, Clark (1973) uses trading volume as a proxy for the information arrival process, while Tauchen and Pitts (1983) study the joint distribution of returns and volume, which are driven by the latent information flow. A common feature of models motivated by the mixture-of-distributions hypothesis is that, conditional on the latent variable σ_t, returns follow a normal distribution: $\varepsilon_t|\sigma_t \sim N(0, \sigma_t^2)$. However, as σ_t is assumed to be a random variable, the unconditional distribution of ε_t is no longer Gaussian but, in particular, has fatter tails than the normal distribution, which corresponds to the empirical evidence for financial returns.

While Clark (1973) and Tauchen and Pitts (1983) did not specify any dynamics for the information flow process, Taylor (1982) was the first to propose the popular model in Equation 1.30, where the logarithm of volatility follows a first-order Gaussian autoregressive process. This allows, through a positive parameter β, to model volatility clustering as in GARCH models, that

is, alternating periods of high and low volatility. Moreover, because of the simplicity of the model, stochastic properties such as stationarity, distributions, or moments are straightforward to derive. For example, by Theorem 2.1 of Andersen (1994), the stochastic process $\{\varepsilon_t\}$ is strictly stationary (which in the SV case is equivalent to covariance stationary) and ergodic if $|\beta| < 1$. This contrasts with GARCH models, for which conditions for strict and covariance stationarity do not coincide, although they do coincide for the EGARCH model. Moreover, we know the unconditional distribution of log-volatility, given by $\log \sigma_t^2 \sim N(\omega/(1-\beta), \sigma_u^2/(1-\beta^2))$, which can, for example, be used to draw initial values for σ_1^2 when simulating the model. It also implies that volatility itself follows a log-normal distribution.

For the model in Equations 1.29 and 1.30, we can calculate the autocorrelation function of squared demeaned returns, ε_t^2, and the kurtosis of ε_t (Ghysels et al., 1996). They are given by, respectively,

$$\rho(\tau) = \frac{\exp(\sigma_u^2/(1-\beta^2)\beta^\tau) - 1}{\kappa - 1}, \tag{1.33}$$

$$\kappa = 3\exp(\sigma_u^2/(1-\beta^2)), \tag{1.34}$$

which shows that, unless the error term of volatility is degenerate (i.e., $\sigma_u = 0$), the kurtosis κ is strictly larger than 3, and returns have a leptokurtic or fat-tailed distribution. Furthermore, the ACF $\rho(\tau)$ decays exponentially with β. Both properties are shared with GARCH-type models. However, Carnero et al. (2004) show that the SV model in Equations 1.29 and 1.30 is more flexible than the standard GARCH(1,1) model with Gaussian innovations in fitting kurtosis and persistence of empirical autocorrelations of squared returns, although both models have the same number of parameters. They attribute the often used fat-tailed distributions for the innovations of a GARCH model to this lack-of-fit of standard GARCH models, which requires adding additional parameters such as the degrees of freedom parameter of a t distribution to better explain kurtosis and persistence of empirical data. In the SV model, however, it is typically not necessary to relax the normality assumption of innovations.

1.3.1 LEVERAGE EFFECT

The classical SV model in Equations 1.29 and 1.30 with independent error terms z_t and u_t cannot take into account the leverage effect mentioned above, that is, the effect that negative news tend to increase volatility stronger than positive news. It is, however, possible to incorporate this effect in the standard model by introducing a dependence between the two error terms. It turns out that there are basically two ways of doing this, which is discussed in the following sections, and the conclusion is that the second one should be preferred.

Jacquier et al. (2004) propose to let (z_t, u_{t-1}) follow a bivariate normal distribution with correlation ρ. They propose estimation and inference methods for this model in a Bayesian framework. A critique of this model, however, is the fact that returns are no longer martingale difference sequences in the sense that

$E[\varepsilon_t|\varepsilon_{t-1}, \sigma_{t-1}] \neq 0$, violating the efficient market hypothesis, see (Harvey and Shephard, 1996).

As an alternative to introduce the leverage effect, Harvey and Shephard (1996) propose to let (z_t, u_t) follow a bivariate normal distribution with correlation ρ. This makes a small but important difference: The two components of ε_t, that is, σ_t and z_t, remain independent and, hence, ε_t has the martingale difference property. Moreover, this model is the discrete time Euler approximation of the diffusion model in Equations 1.31 and 1.32, where $\mathrm{d}W_1(t)\mathrm{d}W_2(t) = \rho \mathrm{d}t$. Yu (2005) provides a comprehensive comparison of these two specifications of the leverage effect in the SV model and concludes that, both from a theoretical and empirical perspective, the model of Harvey and Shephard (1996) should be preferred to that of Jacquier et al. (2004).

1.3.2 ESTIMATION

The estimation problem is certainly the main reason why GARCH models have been more often used in empirical applications than SV models, although much progress has been made over the last 15 years. While in GARCH models the predictive density of returns depends on volatility, which is measurable with respect to the information set, estimation by ML is straightforward. Unfortunately, this is not the case for SV models, since the likelihood function for a sample of T observations can be written as

$$L(\theta; Y_T) \propto \int f(Y_T|H_T; \theta) f(H_T|\theta) \mathrm{d}H_T, \tag{1.35}$$

where Y_T is a vector containing all observed returns, $H_T = (\sigma_1^2, \ldots, \sigma_T^2)'$ is the vector containing all latent volatilities, and θ is the parameter vector, which in the classical model without leverage effect is $\theta = (\omega, \beta, \sigma_u)'$. The problem is the integral appearing in Equation 1.35, which is a multiple integral of dimension T. It cannot be solved analytically, and direct numerical methods are infeasible even for moderately large samples. Other techniques have to be employed, and this is what we discuss in the following paragraphs.

Chapter 6 by Bos (2012) in this handbook gives a broad overview of existing estimation methods of the standard univariate SV model, emphasizing the relationship between them and providing an empirical comparison. Estimation methods of SV models can be roughly categorized into moment methods and simulation methods, where the former are often simpler but inefficient, while the latter attempt to achieve a close approximation of the likelihood function through computationally expensive simulation methods.

Let us first discuss some estimators based on moment expressions. If only the estimation of the model parameter θ is of interest but not the filtration of the underlying volatility process, then simple moment-based estimators can be used based, for example, on the moments given in Equations 1.33 and 1.34. Even closed-form estimators for θ are available, see (Taylor, 1982; Dufour and Valéry, 2006; Hafner and Preminger, 2010), which however are quite inefficient.

Generalized method of moment (GMM) estimators have been proposed, for example, by Melino and Turnbull (1990) and Andersen and Sorensen (1996).

Harvey et al. (1994a) propose a QML estimator based on the linear state space representation of model (Eqs. 1.29 and 1.30), whose measurement equation is given by $\log \varepsilon_t^2 = \log \sigma_t^2 + \xi_t$ and transition equation for the volatility state variable given by Equation 1.30. If $u_t \sim N(0, \sigma_u^2)$, then the noise term $\xi_t = \log u_t^2$ has a highly skewed log chi-square distribution, which Harvey et al. (1994a) approximate by a Gaussian distribution with the same mean and variance. On the basis of this Gaussian linear state space model, they obtain a QML estimator for θ and filtered and smoothed estimates of volatility by using the Kalman filter. While this approach is simple and straightforward to implement, it is not fully efficient because of the skewness of ξ_t.

The influential paper by Kim et al. (1998) extends the approximation of the log chi-squared error term ξ_t to a Gaussian mixture with unobserved mixture weights. Since volatility depends on these latent state variables, the resulting state space model is no longer linear and the Kalman filter cannot be used directly as in Harvey et al. (1994a). Kim et al. (1998) propose to use a Bayesian Markov Chain MC algorithm with data augmentation. For given sampled mixture weights, the state space model is again linear and Kalman filtering can be employed in the estimation and inference procedure. Omori et al. (2007) extend the approach of Kim et al. (1998) to allow for the leverage effect.

Turning to the second category of estimation methods, those based on simulation to approximate as close as possible the likelihood function of the model, Bos (2012) emphasizes importance sampling methods in which much progress has been made recently. Early examples of estimation by simulated ML are Danielsson (1994), Durbin and Koopman (1997), and Sandmann and Koopman (1998). The basic idea of importance sampling, as first used in the SV context by Durbin and Koopman (1997), is to approximate the likelihood function given in Equation 1.35 by the simulation mean of $f(Y_T, H_T^{(i)}; \theta)/g(H_T^{(i)})$, where the ith sequence of volatilities, $H_T^{(i)}$, is drawn from the approximating importance density $g(H_T)$. Extensions of the basic importance sampling estimator of Durbin and Koopman (1997) have been proposed, for example, the efficient importance sampler (EIS) of Liesenfeld and Richard (2003) and Richard and Zhang (2007). The alternative methods differ in the way they construct the importance density $g(H_T)$, which depends on auxiliary parameters.

Bos (2012) discusses other estimation techniques such as simulated method of moments as in Gallant and Tauchen (1996) or the multimove sampler of Shephard and Pitt (1997). Methods based on MCMC and particle filtering are more extensively reviewed in Broto and Ruiz (2004) and Andersen (2009). In this case, estimation and inference is typically investigated in a Bayesian context, the earliest example being Jacquier et al. (1994).

1.3.3 MULTIVARIATE SV MODELS

As for multivariate GARCH models, the variety of multivariate stochastic volatility (MSV) models is remarkable, ranging from a rigid model with independent

volatilities and constant correlations to highly complex models incorporating dynamic correlations and leverage effects. A review of proposed MSV models is given by Asai et al. (2006), see also Yu and Meyer (2006).

The basic model due to Harvey et al. (1994a) can be written as

$$\varepsilon_t = H_t^{1/2} z_t, \quad z_t \sim N(0, \Sigma_z), \tag{1.36}$$

$$H_t = \text{diag}(\exp(h_{1t}), \ldots, \exp(h_{Nt}), \tag{1.37}$$

$$h_{t+1} = \omega + \beta \odot h_t + u_t, \quad u_t \sim N(0, \Sigma_u), \tag{1.38}$$

where $h_t = (h_{1t}, \ldots, h_{Nt})'$ is the vector of volatilities, ω and β are $(N \times 1)$ parameter vectors, Σ_z is a correlation matrix, and $\odot$ designates the Hadamard (elementwise) product operator. Note that, since Σ_z is constant, the model is similar to the CCC model of Bollerslev (1990) discussed above. As a straightforward extension of univariate SV models, Harvey et al. (1994a) propose to use QML with the Kalman filter to estimate this model, while Danielsson (1998) uses simulated ML methods. Obviously, the model may be too restrictive since correlations are restricted to constants, there is no Granger causality in volatilities and leverage effects are not present. However, it is a reasonable starting point and extensions are usually encompassing this basic model.

According to the empirical analysis of Yu and Meyer (2006), the two most successful models to explain volatilities and correlations of a bivariate exchange rate series were explicitly taking into account temporal variation of correlations. The first of these two is a model similar in spirit to the DCC-GARCH model of Engle (2002a) discussed above and can be written as

$$\varepsilon_t = H_t^{1/2} z_t, \quad z_t \sim N(0, \Sigma_{z,t}), \tag{1.39}$$

$$\Sigma_{z,t} = \text{diag}(Q_t^{-1/2}) Q_t \text{diag}(Q_t^{-1/2}), \tag{1.40}$$

$$Q_{t+1} = S \odot (\iota\iota' - A - B) + B \odot Q_t + A \odot v_t v_t', \tag{1.41}$$

$$v_t \sim N(0, I_N),$$

where $\iota = (1, \ldots, 1)'$. Restricting A, B, and $(\iota\iota' - A - B)$ to be positive-definite will ensure that Q_t is positive-definite and, hence, $\Sigma_{z,t}$ is a valid correlation matrix. The volatilities are given as in Equations 1.37 and 1.38.

The second possibility to allow for TVCs is a factor-type SV model of the form

$$\varepsilon_t = D f_t + z_t, \quad z_t \sim N(0, \Sigma_{z,t}), \tag{1.42}$$

$$f_t = \exp(h_t/2)\eta_t, \quad \eta_t \sim N(0, 1), \tag{1.43}$$

$$h_{t+1} = \omega + \beta h_t + u_t, \quad u_t \sim N(0, \sigma_u^2), \tag{1.44}$$

where $D = (\delta_1, \delta_2, \ldots, \delta_N)$ and z_t, η_t, u_t are mutually independent. For identification, one usually imposes $\delta_1 = 1$. The common factor f_t captures comovements in volatilities. In a bivariate framework, Yu and Meyer (2006) estimate this factor model using Bayesian MCMC, while Liesenfeld and Richard (2003) use the EIS in a frequentist approach. Note that the model

(Eqs. 1.42–1.44) has much less parameters than the DCC-SV model in Equations 1.39–1.41. However, it may again be quite restrictive. For example, in the bivariate case, it is straightforward to show that conditional on the factor volatility, the correlation coefficient is given by

$$\mathrm{Corr}(\varepsilon_{1t}, \varepsilon_{2t}|h_t) = \frac{\delta}{\sqrt{(1+\sigma_{z1}^2 \exp(-h_t))(\delta^2 + \sigma_{z2}^2 \exp(-h_t))}},$$

which depends on δ and h_t. Clearly, as h_t increases, correlations increase as well, which corresponds to empirical findings, but the functional form of the dependence between volatilities and correlations might be too rigid in many cases.

Omori and Ishihara (2013), in this handbook, propose very general forms of MSV models that allow, for example, for leverage effects, cross-leverage effects, and heavy-tailed innovation distributions. Their factor MSV model extends the model in Equations 1.42–1.44 to multiple factors, while estimation is performed using Bayesian MCMC based on the multimove sampler of Kim et al. (1998), see also Omori et al. (2007) and Omori and Watanabe (2008).

1.3.4 MODEL SELECTION

Caporin and McAleer (2012), in this handbook, give a survey of recent advances in model selection in the context of volatility models. GARCH and SV models are not nested, which renders the choice based on statistical criteria nontrivial. If the problem is to choose between the exponential GARCH model of Nelson (1991b) and the standard SV model, then an encompassing model could be specified such as

$$\log \sigma_{t+1}^2 = \omega + \alpha_1 z_t + \alpha_2 |z_t| + \beta \log \sigma_t^2 + \sigma_u u_t,$$

see Danielsson (1994) and Fridmann and Harris (1998). The SV model results if $\alpha_1 = \alpha_2 = 0$ and the EGARCH model if $\sigma_u = 0$. For the latter case, Kobayashi and Shi (2005) propose a Lagrange Multiplier test. Furthermore, one can test for the leverage effect by testing the hypothesis $\alpha_1 = 0$.

Choosing between GARCH and SV models can be more complicated since standard model selection criteria such as BIC or Bayesian posterior odds are inconsistent (Hong and Preston, 2005). Several approaches have been proposed to address this problem. Franses et al. (2008) suggest to augment the GARCH model by a contemporaneous stochastic error term, whose variance collapses to 0 if the true model is standard GARCH. Under the alternative of nonzero variance, the resulting model is a variant of an SV model but not equivalent to the standard SV model. Hafner and Preminger (2010) propose a set of simple, strongly consistent decision rules to choose GARCH or SV. Their selection procedure is based on a number of moment conditions that is allowed to increase with the sample size. This method leads to choosing the best and simplest model with probability 1 as the sample size increases.

Furthermore, statistics of standard tests such as likelihood ratio have non-standard distributions (Vuong, 1989). Kim et al. (1998) propose an algorithm based on simulations to obtain empirical p-values of testing one model against the other, which might be inconclusive when hypotheses are reversed. Caporin and McAleer (2012) then continue by giving an extensive review of out-of-sample comparisons, an area with a lot of new results.

1.3.5 EMPIRICAL EXAMPLE: S&P 500

We give a small illustration of estimation results for the daily returns of the S&P 500 index over the period 2006-07-14 until 2011-07-14 (1260 observations). This series was analyzed above using an asymmetric GARCH model, and the leverage effect was found to be highly significant. We therefore would like to allow for leverage effects, also in the SV model. To estimate the SV model, we choose a Bayesian framework and use the MCMC algorithm of Omori et al. (2007).[10] First, the algorithm also relaxes the assumption of normality of the innovation term z_t, which however in our case did not turn out to be necessary: The estimated posterior mean of the degrees of freedom parameter of a t-density is 22.8 with a large 95% confidence interval given by [15.71, 31.89], so that we decided to use the normal distribution for simplicity. We estimate a reparametrized version of model (Eqs. 1.29 and 1.30),

$$\varepsilon_t = \sigma_t z_t, \tag{1.45}$$

$$\log \sigma_{t+1}^2 = \mu + \beta(\log \sigma_t^2 - \mu) + \sigma_u u_t, \tag{1.46}$$

$$\begin{pmatrix} z_t \\ u_t \end{pmatrix} \sim N\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \tag{1.47}$$

where $\mu = \omega/(1-\beta)$. Prior distributions are chosen similar to Omori et al. (2007) as $(\beta+1)/2 \sim \beta(20, 1.5)$, $\sigma_u^{-2} \sim \Gamma(5/2, 0.025)$, $(\rho+1)/2 \sim \beta(1,1)$ and $\mu \sim N(-10, 1)$. For the MC sampler, 5500 draws of the posterior distribution are obtained, and the first 500 are discarded as in Omori et al. (2007).

Table 1.5 shows the estimation results for the posterior distributions of the parameters. It is quite common to find that the persistence parameter β is close to 1, which is again the case here. Remarkably, however, the correlation parameter ρ is strongly negative with a posterior mean of -0.7263, which is probably due to the financial crisis present in the sample. Previous precrisis studies such as Yu (2005) find that correlation is significantly negative but much smaller in absolute value, of the order -0.3 to -0.5. This may indicate that the leverage effect depends on time and, in particular, the state of the economy. The last column of Table 1.5 reports the inefficiency factor defined as $1 + 2\sum_{i=1}^{\infty} \rho_i$, where ρ_i is the sample autocorrelation of order i of the sampled parameter. The small values compared with Kim et al. (1998) indicate the efficiency of the employed sampler

[10] An OxMetrics program is available at http://jnakajima.web.googlepages.com.

TABLE 1.5 Posterior Statistics of SV Parameters Estimated for S&P500 Returns

Parameter	Mean	Standard Deviation	95%L	95%U	Inef.
β	0.9777	0.0050	0.9670	0.9868	4.23
σ_u	0.2030	0.0244	0.1593	0.2560	7.14
$\exp(\mu/2)$	0.0117	0.0011	0.0097	0.0140	1.15
ρ	−0.7263	0.0694	−0.8435	−0.5789	7.07

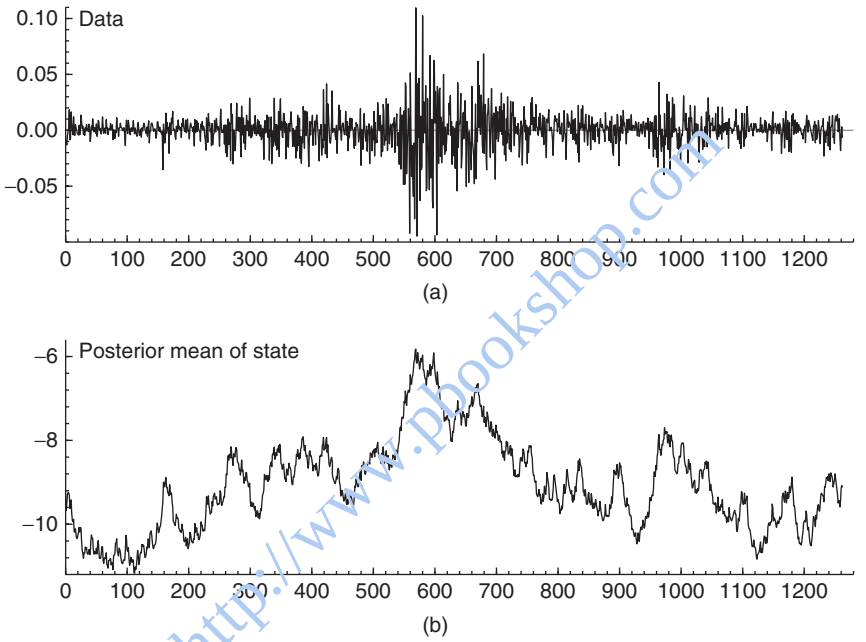

FIGURE 1.5 (a) S&P 500 index returns, July 14, 2006 to July 14, 2011. (b) Posterior mean of log-volatilities.

(Omori et al., 2007). Figure 1.5 shows the index returns and the posterior means of log-volatilities.

1.3.6 LITERATURE

As the research on the modeling, estimation, and inference of SV models, especially in the multivariate case, is huge and still growing, our account can only be partial. We refer to more extensive reviews of the subject: An early monograph that discusses in detail the SV model is given by Taylor (1986). Andersen (1994) is an early review of discrete and continuous time SV models and their applications in finance. Some computational aspects of estimation and inference in SV models are discussed in Bauwens and Rombouts (2004). Shephard (2005)

is a collection of important papers on SV, Broto and Ruiz (2004) provides a review of estimation methods in SV models, whereas Andersen (2009) gives a general review with particular focus on continuous-time SV models and their link with realized volatility measures, to be discussed in the next section.

1.4 Realized Volatility

The models described in the previous sections are essentially parametric and usually designed to estimate the daily, weekly, or monthly volatility using data sampled at the same frequency. Since French et al. (1987) and thanks to the widespread availability of databases providing the intradaily prices of financial assets (stocks, stock indices, bonds, currencies, etc.) econometricians have considered using data sampled at a very high frequency to compute ex-post measures of volatility at a lower frequency.

1.4.1 REALIZED VARIANCE

This method has been popularized by several authors, including Andersen, Barndorff-Nielsen, Bollerslev, Diebold, and Shephard, and is known as *realized volatility approach*.

It is clear that the trading and pricing of securities in many of today's liquid financial asset markets is evolving in a near continuous fashion throughout the trading day. It is thus natural to think of the price and return series of financial assets as arising through discrete observations from an underlying continuous-time process.

The intuition behind realized volatility is most readily conveyed within the popular continuous-time diffusion:

$$\mathrm{d}p(t) = \mu(t)\mathrm{d}t + \sigma(t)\mathrm{d}W(t), t \geq 0, \tag{1.48}$$

where $\mathrm{d}p(t)$ denotes the logarithmic price increment, where $\mu(t)$ is a continuous locally bounded variation process, $\sigma(t)$ is a strictly positive and càdlàg (right-continuous with left limits) SV process and $W(t)$ is a standard Brownian motion.

Assuming that the time length of 1 day is 1, what does model (Eq. 1.48) implies for the one-period daily return? It follows immediately that

$$r_t \equiv p(t) - p(t-1) = \int_{t-1}^{t} \mu(s)\mathrm{d}s + \int_{t-1}^{t} \sigma(s)\mathrm{d}W(s). \tag{1.49}$$

From Equation 1.49, we see that the volatility for the continuous-time process over $[t-1, t]$ is linked to the evolution of the spot volatility $\sigma(t)$. Furthermore, conditional on the sample path of the drift and the spot volatility processes,

$$r_t \sim N\left(\int_{t-1}^{t} \mu(s)\mathrm{d}s, \mathrm{IV}_t\right), \tag{1.50}$$

where IV_t denotes the so-called integrated variance (volatility), and is defined as follows:

$$IV_t \equiv \int_{t-1}^{t} \sigma^2(s)ds. \tag{1.51}$$

It is clear from the above equation that IV_t is latent because $\sigma^2(s)$ is not observable. GARCH and SV models typically infer IV_t from a model that links the daily volatility of day t to past realizations of the one-period daily returns, that is, $r_{t-1}, r_{t-2}, \ldots$ This approach raises some natural questions:

- Which model to choose?
- How good is our GARCH/SV estimate of IV_t?
- By conditioning on past daily returns, do we not lose a significant part of the available information by throwing away all the intraday returns (if intraday data are available of course)?

One of the most popular measures to check the forecasting performance of the volatility models is the Mincer-Zarnowitz regression, that is, ex-post volatility regression:

$$\breve{\sigma}_t^2 = a_0 + a_1\hat{\sigma}_t^2 + u_t, \tag{1.52}$$

where $\breve{\sigma}_t^2$ is the ex-post volatility, $\hat{\sigma}_t^2$ is the forecasted volatility, and a_0 and a_1 are parameters to be estimated. Recall that if the model for the conditional variance is correctly specified (and the parameters are known) and if $E(\breve{\sigma}_t^2) = \hat{\sigma}_t^2$, we have $a_0 = 0$ and $a_1 = 1$.

To judge the quality of the GARCH forecasts, econometricians first used daily squared returns to approximate the ex-post volatility, that is, $\breve{\sigma}_t^2 = r_t^2$. The R^2 of this regression is used to measure the degree of predictability of the volatility models. However, the R^2 of the above regression is typically lower than 5% for GARCH models and this could lead to the conclusion that GARCH models produce poor forecasts of the volatility (see, among others, Schwert (1990) or Jorion (1996)).

In their seminal paper, Andersen and Bollerslev (1998) have shown that if r_t follows a GARCH(1,1), for example, $r_t = \sigma_t z_t$ with $\sigma_t^2 = \omega + \alpha_1 r_{t-1}^2 + \beta_1\sigma_{t-1}^2$, the R^2 of this regression is nothing but $\frac{\text{var}(\breve{\sigma}_t^2)}{\text{var}(r_t^2)} = \frac{\alpha_1^2}{(1-\beta_1^2-2\alpha_1\beta_1)}$. If κ is the kurtosis of the innovations z_t, we have that $\kappa\alpha_1^2 + \beta_1^2 + 2\alpha_1\beta_1 < 1$ to ensure the existence of the unconditional kurtosis of r_t. It follows then that $\kappa\alpha_1^2 < 1 - \beta_1^2 - 2\alpha_1\beta_1$ and

$$R^2 \equiv \frac{\alpha_1^2}{(1 - \beta_1^2 - 2\alpha_1\beta_1)} < \frac{1}{\kappa}.$$

If z_t is i.i.d $N(0, 1)$, the R^2 is thus necessarily lower than 1/3 (and even smaller if z_t has fat-tails).

Let us now illustrate this result by means of a simple MC simulation. In model (Eq. 1.48), $\sigma(t)$ was deliberately let unspecified. The simulated model is designed to induce temporal dependencies consistent with the GARCH(1,1) model. We first consider the continuous-time GARCH diffusion of Nelson (1991a). It is formally defined by

$$\mathrm{d}p(t) = \sigma(t)\mathrm{d}W_p(t), \tag{1.53}$$

$$\mathrm{d}\sigma^2(t) = \theta[\omega - \sigma^2(t)]\mathrm{d}t + (2\lambda\theta)^{1/2}\sigma^2(t)\mathrm{d}W_d(t), \tag{1.54}$$

where $W_p(t)$ and $W_d(t)$ denote two independent Brownian motions.

We used a standard Euler discretization scheme to generate the continuous-time GARCH diffusion process, that is, $p(t+\Delta) = p(t) + \sigma(t)\sqrt{\Delta}Z_p(t)$ and $\sigma^2(t+\Delta) = \theta\omega\Delta + \sigma^2(t)\left[1 - \theta\Delta + \sqrt{2\lambda\theta\Delta}Z_d(t)\right]$, where $Z_p(t)$ and $Z_d(t)$ denote two independent standard normal variables.

We set $\theta = 0.054$, $\omega = 0.478$, and $\lambda = 0.480$ to replicate the behavior of the YEN-USD exchange rate during October 1987 to September 1992 like in Andersen and Bollerslev (1998). To simulate exchange rates, we choose $\Delta = 1/2880$, corresponding to 10 observations per 5-min interval. The number of simulated days is 510, but the first 10 days have been discarded, giving a total of 500 simulated days. Furthermore, we use the following initial values for the log-price and spot volatility: $p(0) = 1$ and $\sigma^2(0) = 0.1$. From the simulated log-prices, we computed 5-min log-prices (denoted $p_{t,i}$ for $i = 1, \ldots, M = 288$, and $t = 1, \ldots, T$) by selecting 1 price for every 10 observations. Five-minute returns $r_{t,i}$ are computed as the first difference of $p_{t,i}$. Finally, daily returns r_t are defined as $\sum_{i=1}^{M} r_{t,i}$.

Figure 1.6 graphs the simulated 5-min and daily returns for the above DGP.

Figure 1.7 plots four volatility measures computed on the simulated data. Let us concentrate on three of these for the moment.

1. Panel (a) displays the daily integrated volatility, that is, IV_t. Given the fact that $\mathrm{IV}_t \equiv \int_{t-1}^{t} \sigma^2(s)\mathrm{d}s$, the "daily" IV_t is computed as $\sum_{i=1}^{1/\Delta} \sigma^2(t - j/\Delta)\Delta$, where $1/\Delta = 2880$. Recall that in empirical applications this quantity is unknown.
2. Panel (c) displays the conditional variance obtained by estimating a GARCH(1,1) model by Gaussian QML on the daily returns r_t.
3. Finally, panel (d) plots the daily squared returns r_t^2.

Two comments are in order.

- Even though the daily squared return is known to be an unbiased measure of the daily volatility, this estimator is extremely noisy.
- Unlike the daily squared returns, the conditional variance of the GARCH(1,1) is much less noisy. Indeed, it generally tracks the level of the integrated volatility very well.

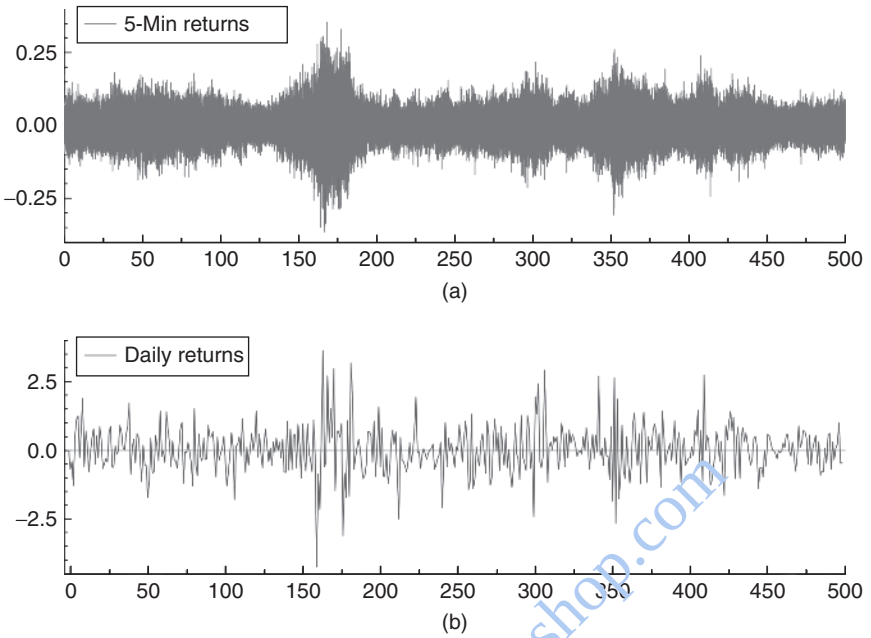

FIGURE 1.6 Simulated 5-min and daily returns from a continuous-time GARCH(1,1).

We also run the above Mincer-Zarnowitz regression to illustrate the findings of Andersen and Bollerslev (1998). When using the daily squared returns to measure the observed daily volatility, the R^2 of the regression is found to be extremely low, that is, 7% even though a_0 and a_1 are not significantly different from 0 and 1, respectively.

Naturally, this finding raises another question. Can we use the R^2 of this regression to discriminate between volatility models? Said differently, "is model 1 preferable to model 2 if the R^2 of this regression is higher for model 1?". It is not always true that using a conditionally unbiased proxy, such as r_t^2, will lead asymptotically to the same outcome that would be obtained if the true volatility was observed. When the evaluation is based on a target observed with error, such as r_t^2, the choice of the evaluation criterion becomes critical in order to avoid a distorted outcome. The problem of consistency, sometimes referred to as *robustness*, of the ordering between two or more volatility forecasts is discussed in Chapter 19 by Violante and Laurent (2012).

Note that if we consider the integrated volatility instead of the squared daily returns as an ex-post volatility measure, the R^2 now equals 53.3%, suggesting that the GARCH model explains more than 50% of the variability of the true volatility despite the fact that a large proportion of the data has been ignored. However, this regression in unfeasible because IV_t is not computable in practical applications.

Andersen and Bollerslev (1998) are the first to point out that a much more precise ex-post estimator than the daily squared return can be obtained by simply summing up intraday squared returns. They called this *estimator-realized*

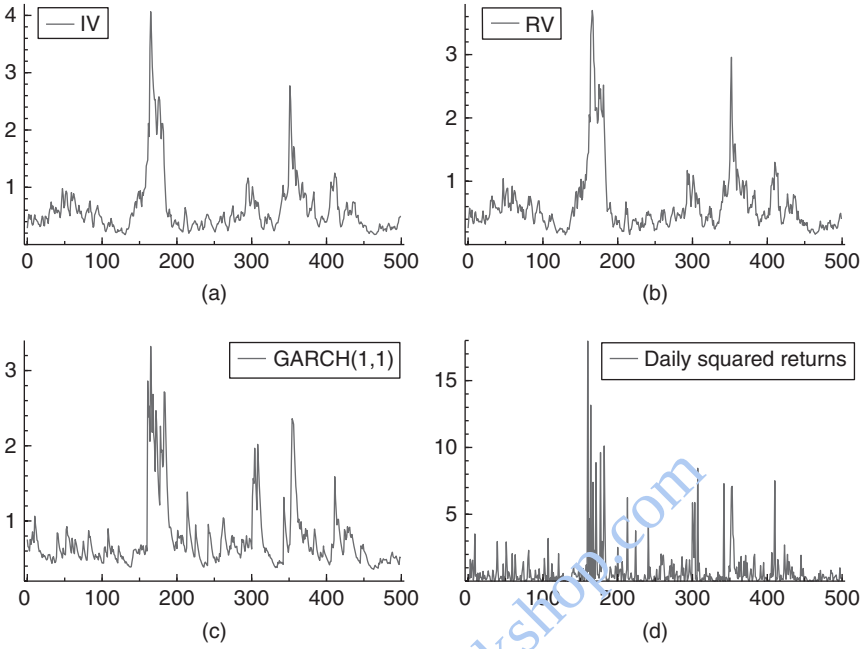

FIGURE 1.7 Four volatility measures: (a) integrated volatility; (b) realized volatility; (c) GARCH(1,1) on daily returns; and (d) daily squared returns.

volatility.[11] More formally, this estimator is defined as follows:

$$\mathrm{RV}_t = \sum_{i=1}^{M} r_{t,i}^2. \tag{1.55}$$

By summing high frequency squared returns, we may obtain an "error free" or "model free" measure of the daily volatility. This is illustrated in Figure 1.7. Panel (a) displays the daily realized volatility computed from the simulated 5-min returns, that is, $\mathrm{RV}_t = \sum_{i=1}^{288} r_{t,i}^2$. It is clear from this graph that realized volatility is indeed a very precise estimator of IV_t. The correlation between IV_t and RV_t equals 0.989.

We also computed the R^2 of the Mincer-Zarnowitz regression using the realized volatility as endogenous variable. Not surprisingly, the R^2 is very close to the value previously obtained for IV_t, that is, 52.7% versus 53.3%.

The properties of this estimator are presented in detail in Chapter 13 by Park and Linton (2012). The main findings of the literature are that under suitable conditions (such as the absence of serial correlation in the intraday returns) the

[11] The origin of realized volatility is not as recent as it would seem at first sight. Merton (1980) already mentioned that, provided data sampled at a high frequency are available, the sum of squared realizations can be used to estimate the variance of an i.i.d. random variable.

realized volatility is consistent for the integrated volatility in the sense that when $\Delta \to 0, \mathrm{RV}_t$ measures the latent integrated volatility IV_t perfectly. However, in practice, at very high frequencies, returns are polluted by microstructure noise (bid-ask bounce, unevenly spaced observations, discreteness, etc.). This "errors-in-variables" problem causes the high frequency returns to be autocorrelated. Recall that bid-ask bounce occurs in all high frequency transaction data as successive quotes tend to bounce between buys and sells, and sampling these as proxies for the mid-price gives an impression that markets are moving more than they actually are, adding an upward bias to the measured volatility. Note that Chapter 14 by Ait-Sahalia and Xiu (2012) show how maximum-likelihood estimators can be used to deal with the microstructure noise issue.

Empirical studies have shown that a continuous diffusion model as in Equation 1.48 fails to explain some characteristics of asset returns. Furthermore, standard GARCH models are not able to fully explain the excess kurtosis found in most financial time series. In a continuous-time framework, the inadequacy of the standard stochastic diffusion model has led to developments of alternative models. Jump diffusion and SV models have been proposed in the literature to overcome this inadequacy.

Suppose now that the log-price process belongs to the Brownian Semi-Martingale with Jumps (BSMJ) family of models. Under the BSMJ model, the diffusion component captures the smooth variation of the price process, while the jump component accounts for the rare, large discontinuities in the observed prices. Andersen et al. (2007) cite the work of several authors who found that this is a realistic model for the price series of many financial assets.

A BSMJ log-price diffusion admits the representation

$$\mathrm{d}p(t) = \mu(t)\mathrm{d}t + \sigma(t)\mathrm{d}W(t) + \kappa(t)\mathrm{d}q(t), t \geq 0, \tag{1.56}$$

where $\mathrm{d}q(t)$ is a counting process with $\mathrm{d}q(t) = 1$ corresponding to a jump at time t and $\mathrm{d}q(t) = 0$ otherwise. The (possibly time-varying) jump intensity is $l(t)$ and $\kappa(t)$ is the size of the corresponding jump.

Jumps in stock prices are often assumed to follow a probability law. For instance, the jumps may follow a Poisson process, which is a continuous-time discrete process.

Let us consider the following continuous-time GARCH diffusion process with jumps,

$$\mathrm{d}p(t) = \sigma(t)\mathrm{d}W_p(t) + \kappa(t)\mathrm{d}q(t), \tag{1.57}$$

$$\mathrm{d}\sigma^2(t) = \theta[\omega - \sigma^2(t)]\mathrm{d}t + (2\lambda\theta)^{1/2}\sigma^2(t)\mathrm{d}W_d(t), \tag{1.58}$$

$$\kappa(t) \sim \sigma(t)\sqrt{m}([-2,-1] \cup [1,2]) \tag{1.59}$$

$$\mathrm{d}q(t) \sim \mathrm{Poisson}(l). \tag{1.60}$$

The jump size $\kappa(t)$ is modeled as the product between $\sigma(t)$ and a uniformly distributed random variable on $\sqrt{m}([-2,-1] \cup [1,2])$. Note that in this DGP,

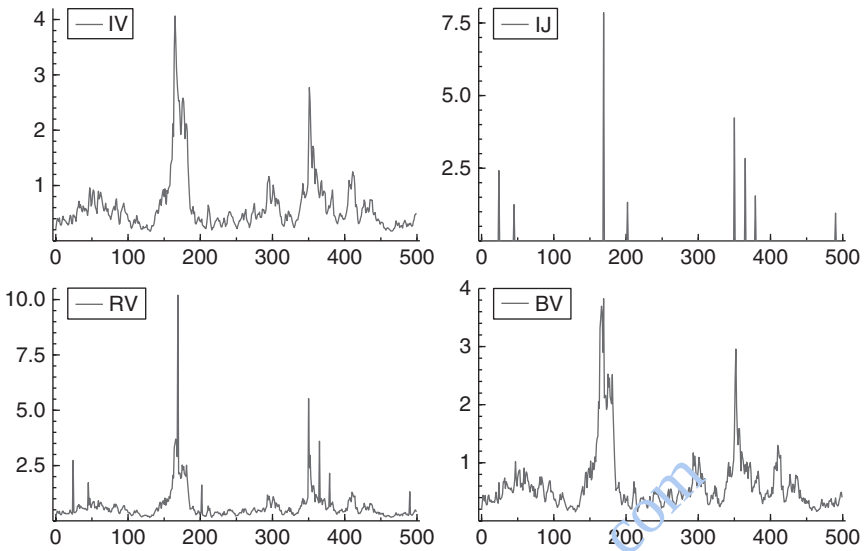

FIGURE 1.8 Four volatility measures in presence of jumps: (a) integrated volatility; (b) integrated jumps; (c) realized volatility, GARCH(1,1) on daily returns; and (d) bipower variation.

the intensity of the jumps (l) is assumed to be constant over time for simplicity. The parameter m determines the magnitude of the jumps.

Figure 1.8 plots four volatility estimates for 500 days of simulated intraday returns. The parameters of the continuous-time GARCH(1,1) are the same as for the previous simulation. For the jump component, l is chosen such that first jump is expected every 100 days (in this replication there are 8 days with at least one jump). About the magnitude of the jumps, we chose $m = 2$, which corresponds to a case of rare but very big jumps.

1. Panel (a) displays the daily integrated volatility, that is, IV_t;
2. Panel (b) displays the integrated jumps, defined as $IJ_t = \sum_{t-1<s\leq t} \kappa^2(s)$. Like IV_t, this quantity is latent and cannot be computed on real data;
3. Panel (c) displays the realized volatility computed from 5-min returns;
4. Finally, panel (d) plots the so-called bipower variation estimator BV_t (see below).

It is clearly visible that the realized volatility does not match the integrated volatility in presence of jumps. This result is not surprising since we know by the theory of quadratic variation that for $\Delta \to 0$, we have the following convergence in probability:

$$\text{RV}_t \to \int_{t-1}^{t} \sigma^2(s)\mathrm{d}s + \sum_{t-1<s\leq t} \kappa^2(s). \tag{1.61}$$

In other words, in the absence of jumps, the realized volatility is a consistent estimator of the integrated volatility but not in the presence of jumps.

Several robust to jumps estimators of IV_t are discussed in Chapters 17 and 18 by Mancini and Calvori (2012) and Boudt et al. (2012), respectively. The pioneers are Barndorff-Nielsen and Shephard (Barndorff-Nielsen and Shephard (2004b)), who showed that for a subclass of BSMJ price diffusions (i.e., BSM with Finite Activity Jumps), the normalized sum of products of the absolute value of contiguous returns (i.e., bipower variation) is a consistent estimator for IV_t. Mancini and Calvori (2012) also discuss the case of infinite activity jump processes (Levy jumps).

The bipower variation is defined as

$$\mathrm{BV}_t \equiv \mu_1^{-2} \frac{M}{M-1} \sum_{i=2}^{M} |r_{t,i}||r_{t,i-1}|, \tag{1.62}$$

where $\mu_1 \equiv \sqrt{2/\pi} \simeq 0.79788$.

Unlike the RV_t, BV_t is designed to be robust to jumps because its building block is the product between two consecutive returns instead of the squared return. If one of the returns corresponds to a jump and the next one follows the BSM diffusion process, then the product has a small impact on BV_t, being the sum of many of these building blocks. If the jump process has finite activity[12] then a.s. jumps cannot affect two contiguous returns for $\Delta \to 0$ (or equivalently $M \to \infty$) and the jump process has a negligible impact on the probability limit of BV_t, which coincides with the IVar. Under the BSM with finite activity jumps (BSMFAJ), one has

$$\mathrm{plim}_{\Delta \to 0} \mathrm{BV}_t = \int_{t-1}^{t} \sigma^2(s) \mathrm{d}s. \tag{1.63}$$

Looking at Figure 1.8d, we see that unlike RV_t, BV_t is indeed a robust estimate of the integrated volatility in presence of jumps.

1.4.1.1 Empirical Application. The series we consider is the Dow Jones index. We use a 5-min sampling frequency corresponding to seventy-eight 5-min intraday price observations for each trading day (from 9:30 EST until the market closes, i.e., at 16:00 EST). The data set covers the periods from 1995-01-03 to 2009-12-31. The Dow Jones index data is provided by the Tickdata.

Figure 1.9 plots the daily returns, realized volatility and bipower variation for the Dow Jones index series computed from 5-min returns.

Figure 1.10 plots the autocorrelation function (50 lags) of these three series. The displayed 95% confidence bands (doted lines) are computed with the generalized Bartlett's formula of Francq and Zakoian (2009). This figure clearly suggests the presence of long memory in the realized volatility and bipower

[12] A jump process is defined to be of finite activity if the number of jumps in *any* interval of time is finite.

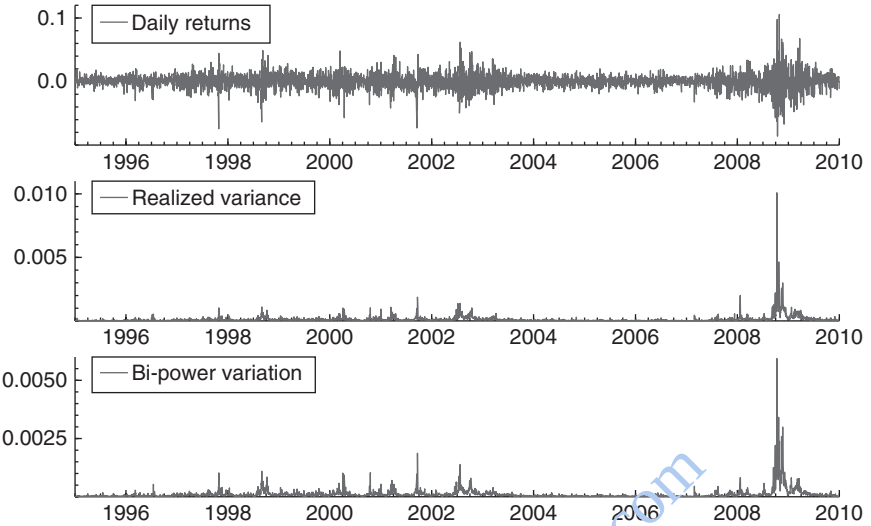

FIGURE 1.9 (a) Daily returns; (b) realized volatility; and (c) bipower variation for the Dow Jones index computed from 5-min returns.

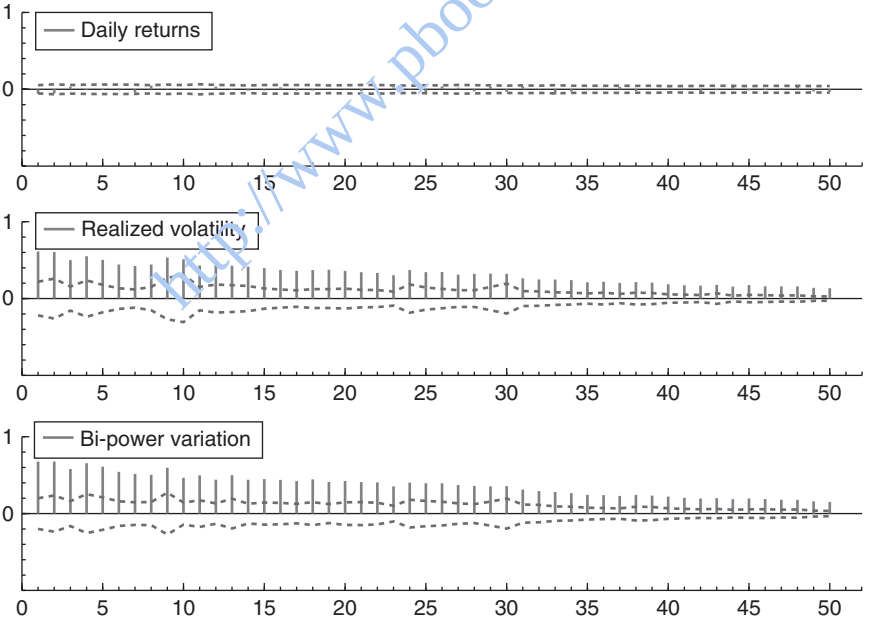

FIGURE 1.10 Autocorrelation function and robust 95% confidence interval on daily returns, realized volatility, and bipower variation for the Dow Jones index computed from 5-min returns.

variation but no serial correlation in the daily returns. The estimated long-memory parameters given by the log-periodogram regression method of Geweke and Porter-Hudak (1983) are equal to 0.28 and 0.33 for the realized volatility and bipower variation, respectively, suggesting that BV_t is slightly more persistent than RV_t (because of the presence of jumps in RV_t, see Andersen et al. (2007)).

The next step is naturally to formulate a model to forecast RV_t, and/or BV_t that takes into account their most important characteristics. ARFIMA models are usually estimated on these two series (or their log-transformation to ensure the positivity of the forecasts). Figure 1.11 plots $\log(BV_t)$ as well as the conditional mean and conditional variance of an ARFIMA$(1, d, 0)$-GARCH$(1, 1)$ estimated by ML with a skewed-t distribution (see Giot and Laurent (2003) and Bauwens and Laurent (2005) for more details on this distribution).

This figure suggests that this conditional mean captures the main features of the series. Furthermore, the conditional variance is not constant over time, suggesting that the variance of the variance is time-varying as well. Figure 1.11d plots a histogram of the standardized residuals of the estimated model, together with a kernel estimate (solid line) and the estimated (dotted line) of the unconditional density of the standardized residuals. This graph also suggest that the skewed-t density provides a good approximation of the true density (the estimated asymmetry coefficient is positive and highly significant and the degree of freedom is about 15).

In Chapter 15, Corsi et al. (2012) follow an alternative direction that generates very similar stylized facts for volatility series using a cascade of heterogeneous volatility components. This model leads to a simple AR-type

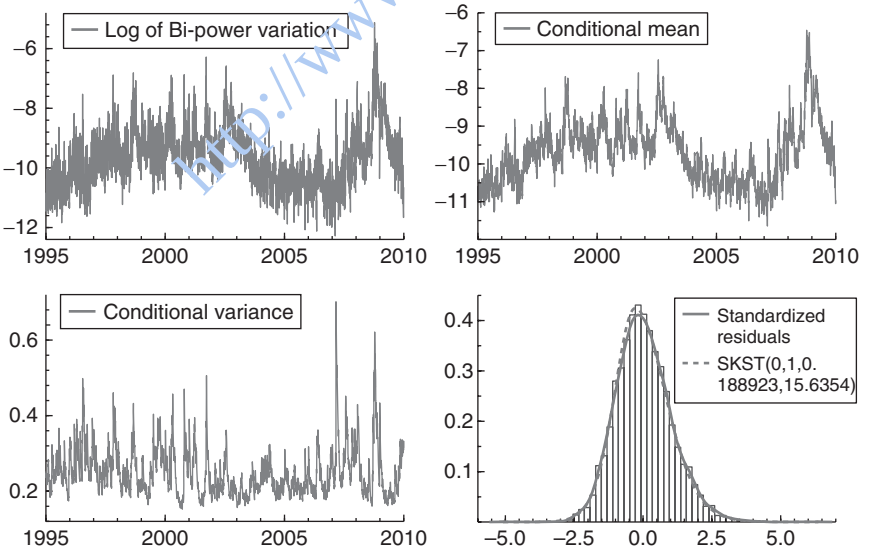

FIGURE 1.11 **Log of bipower variation, conditional mean, and conditional variance of an ARFIMA$(1, d, 0)$-GARCH$(1, 1)$ on $\log(BV_t)$ and density estimate of the innovations for the Dow Jones index.**

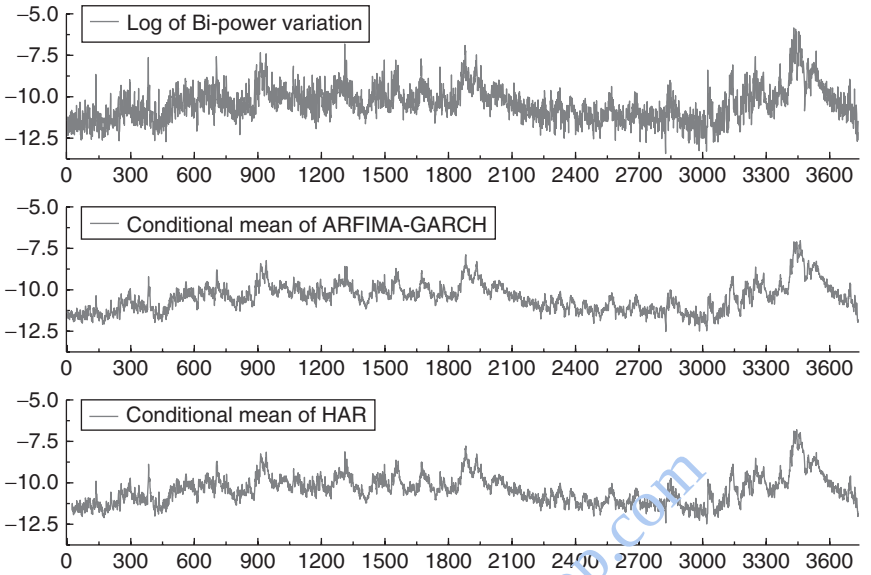

FIGURE 1.12 Log of bipower variation, conditional mean of the ARFIMA$(1, d, 0)$-GARCH$(1, 1)$, and HAR models on $\log(\mathrm{BV}_t)$ for the Dow Jones index.

model that considers volatilities realized over different time horizons and is thus called heterogeneous autoregressive (HAR). This framework turns out to be easier to handle than the above ARFIMA model, with a straightforward economic interpretation and an excellent fit to the data.

Figure 1.12 plots $\log(\mathrm{BV}_t)$ as well as the conditional mean of the above ARFIMA$(1, d, 0)$-GARCH$(1, 1)$ and the HAR model with a cascade of three volatility measures, that is, the log of the average of BV_t over the previous 1, 5, and 21 days. The difference between the fitted values of the ARFIMA and HAR models is hardly visible. Indeed, both models seem to track the dynamics of $\log(\mathrm{BV}_t)$ rather well. The in-sample average of the errors equal 0.018871 and $1.2188\mathrm{e}^{-14}$, respectively, for the ARFIMA and HAR models. The corresponding standard deviations are 0.56977 and 0.57120. Both models explain respectively 68.8879% and 68.5991% of the variability of $\log(\mathrm{BV}_t)$, while the correlation between the fitted values (conditional mean) of the two models is 98.358%. The two models are thus hardly distinguishable.

The HAR model and the ARFIMA model described above have something in common. Both the endogenous and explanatory variables are aggregated measures of volatility (e.g., realized volatility) or some transformation of these measures (e.g., square root, log). Chapter 16 by Ghysels and Valkanov (2012) reviews an alternative strategy called MIDAS (*mi*xed *da*ta *s*ampling). For example, when we forecast daily volatility, we want to preserve the information in the intradaily data without computing daily aggregates such as realized volatility. Likewise, when we focus on, say, weekly or monthly volatility forecasts, we want to use daily returns or daily realized volatility measures. They focus on the issues

pertaining to mixed frequencies—that arise typically because we would like to consider multistep volatility forecasts while maintaining information in high frequency data.

1.4.2 REALIZED COVARIANCE

In the case, where $y_{t,i}$ is an N-dimensional return vector generated by the multivariate counterpart of the BSMJ price diffusion model in Equation 1.56, the processes $p(s)$, $\mu(t)$, and $q(t)$ are all N-dimensional vector processes and $W(t)$ is a vector of N independent Brownian motions. Denoted by $\Omega(t)$, the $N \times N$ càdlàg process such that $\Sigma(t) = \Omega(t)\Omega'(t)$ is the spot covariance matrix process of the continuous component of the price diffusion. Let $K(t)$ be the $N \times N$ process controlling the magnitude and transmission of jumps such that $K(t)\mathrm{d}q(t)$ is the contribution of the jump process to the price diffusion. We then have that a N-dimensional log-price diffusion can be decomposed as follows:

$$\mathrm{d}p(t) = \mu(t)\mathrm{d}t + \Omega(t)\mathrm{d}w(t) + K(t)\mathrm{d}q(t) \tag{1.64}$$

The integrated covariance matrix (ICov) over $[t-1, t]$ is the matrix

$$\mathrm{ICov}_t = \int_{t-1}^{t} \Sigma(s)\mathrm{d}s. \tag{1.65}$$

Denoted by κ_j, the contribution of the jth jump in $[t-1, t]$ to the price diffusion.

1.4.2.1 Realized Quadratic Covariation. Andersen et al. (2003) have shown that the realized quadratic covariation (RCov)

$$\mathrm{RCov}_t \equiv \sum_{i=1}^{M} y_{t,i} y'_{t,i} \tag{1.66}$$

is a consistent estimator for the sum of the ICov and the realized jump variability

$$\mathrm{plim}_{\Delta \to 0} \mathrm{RCov}_t = \mathrm{ICov}_t + \sum_{j=1}^{j_t} \kappa_j \kappa'_j, \tag{1.67}$$

where $j_t = \int_{t-1}^{t} \mathrm{d}q^*(s)$, with $q^*(s)$ the univariate counting process derived from $q(s)$ such that $q^*(s)$ increases by 1 whenever $q(s)$ changes.

Compared to the univariate case, the additional issue of synchronicity arises, whereby trading for different assets occurs at different times. Park and Linton (2012) discuss two methods typically used to solve this problem, namely, the *fixed clock time* and the *refresh time*.

1.4.2.2 Realized Bipower Covariation. For disentangling the continuous and jump components in the RCov, we need an additional estimator for the

ICov that is robust to jumps. To this purpose, Barndorff-Nielsen and Shephard (2004b) introduce the Realized BiPower Covariation process (RBPCov) as the process whose value at time t is the N-dimensional square matrix with k-, lth element equal to

$$\frac{\pi}{8}\left(\sum_{i=2}^{M}\left|y_{(k)t,i}+y_{(l)t,i}\right|\ \left|y_{(k)t,i-1}+y_{(l)t,i-1}\right|\right.$$
$$\left.-\left|y_{(k)t,i}-y_{(l)t,i}\right|\ \left|y_{(k)t,i-1}-y_{(l)t,i-1}\right|\right), \tag{1.68}$$

where $y_{(k)t,i}$ is the kth component of the return vector $y_{t,i}$. The factor $\pi/8$ ensures that the RBPCov converges to the ICov under model (Eq. 1.64):

$$\text{plim}_{\Delta\to 0}\text{RBPCov}_t=\int_{t-1}^{t}\Sigma(s)\text{d}s. \tag{1.69}$$

Other estimators of the integrated covariance that are robust to the presence of jumps in assets returns are reviewed in the Chapters 17 and 13 by Mancini and Calvori (2012) and Park and Linton (2012), respectively, for example, the threshold realized covariation of Mancini and Gobbi (2012), and the outlyingness weighted quadratic covariation of Boudt et al. (2010).

A natural question is how these nonparametric covariance estimates can be used to model and forecast future values of the covariance, ensuring its symmetry and its positive semidefiniteness. Answers to this question are provided in this Handbook in the Chapters 4, 9, 13, and 15 by Sheppard (2012), Brownlees et al. (2012b), Park and Linton (2012), and Corsi et al. (2012), respectively. The existing models are the Wishart autoregressive (WAR) model of Gouriéroux et al. (2009), and standard univariate and multivariate models estimated on the elements of the Cholesky factorization (Chiriac and Voev, 2010) or the matrix log-transformation (Bauer and Vorkink, 2011) of the covariance estimates. New models based on the Wishart distribution are in development (Jin and Maheu, 2010; Golosnoy et al., 2010; Bauwens et al., 2012).

Acknowledgments

Research supported by the contract "Projet d'Actions de Recherche Concertées" 07/12-002 of the "Communauté francaise de Belgique," granted by the "Académie universitaire Louvain."