4

Modeling Volatility

GARCH models have been developed to account for empirical features in the volatility of financial returns. As described in Chapter 2, return data appear stationary and barely autocorrelated. However, squared returns appear to be serially correlated. More specifically, the volatility of returns appears to cluster, so that large variations of price (positive or negative) are expected after a large variation of price (of either sign). Theoretical models justifying this volatility clustering have been discussed in Chapter 3. This serial correlation in squared returns has been initially modeled by Engle (1982) with the ARCH (AutoRegressive Conditional Heteroskedasticity) model and by Bollerslev (1986) with the Generalized ARCH (GARCH) model. Several extensions have been built on these early models to capture additional empirical features of asset returns. Section 4.2 presents the class of ARCH models, and Section 4.3 is devoted to GARCH models. In Section 4.4, we address the issues of introducing asymmetry in GARCH models. Finally, Section 4.5 describes GARCH models with jumps. Some aspects of these models are more particularly described: (1) how to forecast volatility using these models; (2) how to estimate (G)ARCH model efficiently; (3) how to test the presence of GARCH effects. Aggregation issues are discussed in Section 4.6. The multivariate extension is described in Chapter 6.

Another set of models, the so-called stochastic volatility models, has been introduced by Taylor (1982, 1986). These models, discussed in Section 4.7, introduce a second source of randomness in the price dynamics. Finally, a recent approach considers the use of intraday data to measure daily volatility. The construction and the use of this so-called realized volatility are described in Section 4.8.

4.1 Volatility at lower frequencies

In Chapter 3, we investigated the consequences of the functioning of financial markets on the intraday evolution of volatility. Often, such data is not avail-

able. Only data at some lower, say daily, frequency is available. For such a situation, so-called ARCH and GARCH models are useful in describing time-variation in conditional variance, which in turn explains, at least partially, the fat-tail phenomenon present in returns. Also, returns tend to be negatively correlated with changes in volatility, a feature that can be explained by the leverage effect (Black, 1976). It is easy to adapt GARCH models to accommodate this type of feature. This has given birth to so-called asymmetric GARCH models. As mentioned, GARCH models have been an extremely successful way to model several features of asset prices. A huge literature has emerged, and several surveys of GARCH models have been published. We may mention for instance Bollerslev, Chou, and Kroner (1992), Bera and Higgins (1993), Bollerslev, Engle, and Nelson (1994), Palm (1996), or more recently Li, Ling, and McAleer (2002). We will cover, in this chapter, the main features related to GARCH models, because these features will be building blocks for more complex models discussed in subsequent chapters.

The structure of a volatility model can be described as

$$x_t = \mu_t(\theta) + \varepsilon_t, \tag{4.1}$$

$$\varepsilon_t = \sigma_t(\theta)\, z_t, \tag{4.2}$$

where

$$\mu_t(\theta) = E\left[x_t \,|\mathcal{F}_{t-1}\right],$$
$$\sigma_t^2(\theta) = E\left[\left(x_t - \mu_t(\theta)\right)^2 |\mathcal{F}_{t-1}\right].$$

In (4.1), we decompose the return x_t into a conditional mean $\mu_t(\theta)$ and a residual term ε_t. The dynamics of the conditional mean $\mu_t(\theta)$ may be an ARMA(p,q) process or could consist of seasonality features. $\mathcal{F}_t$ is the information set available at time t. It may include current and past returns, current and past residuals, or any other variable known at time t. According to (4.2), the residual term ε_t has a volatility conditional on the information available at time $t-1$, denoted σ_t, which may vary over time. θ is the vector of unknown parameters. The variable z_t will be assumed to follow some distribution with mean 0 and variance 1. Even though normality will be assumed in this chapter, this need not be. Other distributions such as the Student t distribution will be discussed in Chapter 5.

A volatility model is a model that describes the evolution of $\sigma_t^2(\theta)$. There are essentially two types of models for describing the dynamics of volatility:

1. In the first category, volatility is described as an exact function of a given set of variables. This category includes (G)ARCH models.
2. In the second category, volatility is described as a stochastic function. It includes Stochastic Volatility models.

The first sections of the chapter are devoted to the main aspects of the ARCH and GARCH methodology. We then turn to the stochastic volatility models, and eventually we describe the realized volatility approach.

4.2 ARCH model

The ARCH(p) model, originally introduced by Engle (1982), assumes that the conditional variance is a linear function of the past p squared innovations:

$$\sigma_t^2(\theta) = \omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2. \tag{4.3}$$

According to (4.3), the conditional volatility is assumed to be a moving average of squared innovations. For this model to be well defined and the conditional variance to be positive, the parameters must satisfy the following constraints: $\omega > 0$, and $\alpha_i \geq 0$, $i = 1, \cdots, p$.

The unconditional variance of innovation, denoted σ^2, is the unconditional expectation of σ_t^2: $\sigma^2 = E\left[\varepsilon_t^2\right] = E\left[E_{t-1}\left[\varepsilon_t^2\right]\right] = E\left[\sigma_t^2\right]$. In the case of an ARCH(p) process, it is easy to compute that $\sigma^2 = \omega / \left(1 - \sum_{i=1}^{p} \alpha_i\right)$. This shows that the process ε_t is covariance stationary if and only if the sum of the autoregressive parameters is less than one, i.e., $\sum_{i=1}^{p} \alpha_i < 1$. (See Section 4.3 below for more details on stationarity conditions.)

Although the innovation ε_t is serially uncorrelated, they are not time independent, because, defining $v_t = \varepsilon_t^2 - \sigma_t^2$, we can rewrite (4.3) as

$$\varepsilon_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + v_t,$$

with $E_{t-1}[v_t] = 0$. Therefore, the plain ARCH(p) model can be viewed as an AR(p) model for the squared innovation ε_t^2.

Notice that we do not describe in this section how to estimate an ARCH(p) model. Estimation is detailed in Section 4.3 devoted to GARCH models.

4.2.1 Forecasting

Forecasts of the ARCH model are obtained recursively. In the following, parameters with a hat correspond to estimates. Let t be the starting date for forecasting. Then, the 1-step ahead forecast for σ_{t+1}^2 is

$$\sigma_t^2(1) = \hat{\omega} + \hat{\alpha}_1 \hat{\varepsilon}_t^2 + \cdots + \hat{\alpha}_p \hat{\varepsilon}_{t+1-p}^2 = \hat{\omega} + \sum_{i=1}^{p} \hat{\alpha}_i \hat{\varepsilon}_{t+1-i}^2,$$

where $\hat{\varepsilon}_t$ is the estimated residual. For the 2-step ahead forecast for σ_{t+2}^2, we need a forecast of ε_{t+1}^2. It is given by $\sigma_t^2(1)$. We therefore obtain:

$$\sigma_t^2(2) = \hat{\omega} + \hat{\alpha}_1 \sigma_t^2(1) + \hat{\alpha}_2 \hat{\varepsilon}_t^2 + \cdots + \hat{\alpha}_p \hat{\varepsilon}_{t+2-p}^2.$$

The κ-step ahead forecast for $\sigma_{t+\kappa}^2$ is

$$\sigma_t^2(\kappa) = \hat{\omega} + \hat{\alpha}_1 \sigma_t^2(\kappa - 1) + \cdots + \hat{\alpha}_p \sigma_t^2(\kappa - p) = \hat{\omega} + \sum_{i=1}^{p} \hat{\alpha}_i \sigma_t^2(\kappa - i).$$

with $\sigma_t^2(\kappa - i) = \hat{\varepsilon}_{t+\kappa-i}^2$ if $\kappa - i \leq 0$.

4.2.2 Kurtosis of an ARCH model

ARCH models are able to generate excess kurtosis. Indeed, even if the conditional distribution for the standardized innovations $z_t = (x_t - \mu_t(\theta))/\sigma_t(\theta)$ is assumed to be normal, the unconditional distribution for ε_t has fatter tails than the normal distribution. For instance, consider an ARCH(1) model

$$\sigma_t^2(\theta) = \omega + \alpha_1 \varepsilon_{t-1}^2,$$

where $\varepsilon_t = \sigma_t z_t$. The unconditional variance is given by $\sigma^2 = \omega/(1-\alpha_1)$ so that we should have $\alpha_1 < 1$ for σ^2 to be finite. Now, under normality, the conditional fourth moment is given by

$$E\left[\varepsilon_t^4 | \mathcal{F}_{t-1}\right] = E\left[\sigma_t^4 z_t^4 | \mathcal{F}_{t-1}\right] = 3E\left[\sigma_t^4 | \mathcal{F}_{t-1}\right] = 3\left(\omega + \alpha_1 \varepsilon_{t-1}^2\right)^2,$$

whereas the unconditional fourth moment is

$$\begin{aligned} m_4 = E\left[\varepsilon_t^4\right] &= E\left[E\left[\sigma_t^4 z_t^4 | \mathcal{F}_{t-1}\right]\right] = 3E\left[\left(\omega + \alpha_1 \varepsilon_{t-1}^2\right)^2\right] \\ &= 3\left(\omega^2 + 2\frac{\omega^2 \alpha_1}{1-\alpha_1} + \alpha_1^2 m_4\right) = \frac{3\omega^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}. \end{aligned}$$

Since the fourth moment of ε_t is positive, we must have that $\alpha_1^2 < 1/3$. The unconditional kurtosis is

$$\kappa = \frac{m_4}{\sigma^4} = \frac{3\omega^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)} \times \frac{(1-\alpha_1)^2}{\omega^2} = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2},$$

which exceeds 3 for $\alpha_1 > 0$ and $3\alpha_1^2 < 1$.[1] So, the excess kurtosis is always positive and the tails of the distribution of ε_t are fatter than that of a normal distribution, even if the conditional distribution is normal.[2]

This result helps understanding the success of ARCH modeling, because it is able to capture both volatility clustering as well as fat-tailedness of the distribution. However, we will see in the following chapter that it is not able to capture all the fat-tailedness found in asset returns.

4.2.3 Testing for ARCH effects

A widely-used test for ARCH effects is the Lagrange-Multiplier (LM) test proposed by Engle (1982). A definite advantage of this test is that it is very easy to perform. Under the null hypothesis, the error term ε_t is assumed to be a normal white noise process $\varepsilon_t | \mathcal{F}_{t-1} \sim \mathcal{N}\left(0, \sigma^2\right)$. The alternative hypothesis is that the error term is driven by an ARCH(p) model, so that

[1] For $3\alpha_1^2 \geq 1$, the kurtosis becomes infinite.

[2] For a more general formulation of the ARCH model, it becomes more and more complicated to obtain the unconditional kurtosis, but the excess kurtosis is still positive. See Section 5.1.

$$\varepsilon_t = \sigma_t z_t,$$
$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2.$$

The test for ARCH(p) effects is based on the null hypothesis $H_0 : \alpha_1 = \cdots = \alpha_p = 0$ against the alternative $H_a : \alpha_1 \geq 0, \cdots, \alpha_p \geq 0$ with at least one strict inequality.

The LM test statistic of this hypothesis is shown to be asymptotically equivalent to the $T \times R^2$ test statistic, where T is the sample size and R^2 is computed from the regression (Engle, 1982)

$$\hat{\varepsilon}_t^2 = a_0 + a_1 \hat{\varepsilon}_{t-1}^2 + \cdots + a_p \hat{\varepsilon}_{t-p}^2 + v_t.$$

Under the null of no ARCH effect, the LM and $T \times R^2$ test statistics are asymptotically distributed as a $\chi^2(p)$.

As an alternative form of the LM test, we may use the asymptotically equivalent portmanteau tests, such as the Ljung and Box (1978) statistic, for ε_t^2.

It should be noticed that because the parameters of the ARCH model must be positive, the ARCH test should be formulated as a one-sided test. Such a test has been proposed by Demos and Sentana (1998).

4.2.4 ARCH-in-mean model

The ARCH-in-Mean model, proposed by Engle, Lilien, and Robbins (1987), was designed to capture the effect of conditional volatility on the conditional mean of the process:

$$x_t = \delta \varphi(\sigma_t) + \sigma_t \varepsilon_t, \tag{4.4}$$

where the conditional mean depends on conditional volatility σ_t. Equation (4.4) appears as a natural way to describe the trade-off between risk and expected return, which is consistent with many theories in finance. Several specifications have been proposed to incorporate volatility in the conditional equation. In particular, $\varphi(\sigma_t) = \sigma_t$ has been associated with the CAPM, whereas $\varphi(\sigma_t) = \sigma_t^2$ has been associated with the model of asset demand. Finally, $\varphi(\sigma_t) = \log(\sigma_t)$ has been estimated for instance by Engle, Lilien, and Robbins (1987).

Because of the dependence of the conditional mean on the conditional variance, several problems arise in the estimation and testing of ARCH-in-Mean models. The reason is that the second-order derivatives are not block-diagonal anymore, i.e.,

$$E\left[\frac{\partial^2 \ell_t(\theta, \delta)}{\partial \theta \partial \delta'}\right] \neq 0.$$

Consequently, the estimation of the conditional mean and the conditional variance parts of the model has to be performed jointly.

4.2.5 Illustration

As an illustration of some properties of ARCH models, we report the estimation of an ARCH(10) for the SP500, DAX, FT-SE, and Nikkei daily returns between January 1980 and August 2004. The complete model is given by

$$\begin{aligned} x_t &= \mu + \varepsilon_t, \\ \varepsilon_t &= \sigma_t z_t, \\ \sigma_t^2 &= \omega + \sum_{i=1}^{10} \alpha_i \varepsilon_{t-i}^2, \end{aligned}$$

so that the expected return is assumed to be constant over time.

Table 4.1 reports the parameter estimates and the associated t-stat for the volatility equation. We first notice that several lags of the innovation process are necessary to explain the dynamics of conditional volatility. For most markets, all lags up to the 10th are significant. This dynamics appears to be rather persistent, because the sum of parameters α_i is larger than 0.7. In the case of the Nikkei, we may even conclude that the volatility process is non-stationary, because the sum of ARCH parameters reaches its boundary of 1.

Figures 4.1 to 4.4 present the estimated dynamics of volatility for the different indices at hand. We notice that a large range of volatility patterns exists in the markets, the Nikkei being particularly agitated. Notice also that although the sum of α_i parameters is equal to 1, the Nikkei volatility does not seem to explode. These figures illustrate the volatility clustering phenomenon already mentioned. We finally observe that the October 1987 crash has a pronounced effect on the volatility of the SP500 and FT-SE daily returns. For the DAX and Nikkei daily returns, the level of volatility reached in October 1987 is similar to the level reached during other episodes.

4.3 GARCH model

Due to the large persistence in volatility, the ARCH model often requires a large p to fit the data. In such cases, it is more parsimonious to use the GARCH (Generalized ARCH) model proposed by Bollerslev (1986). The conditional variance of a GARCH(p, q) is:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2. \tag{4.5}$$

For this model to be well defined and the conditional variance to be positive, the parameters must satisfy the following constraints: $\omega > 0$, $\alpha_i \geq 0$, $i = 1, \cdots, p$, $\beta_j \geq 0$, for $j = 1, \cdots, q$.

Table 4.1. *Estimation of an ARCH(10) model for daily log-returns*

	SP500	DAX	FT-SE	Nikkei
μ	0.0370	0.0330	0.0360	0.0090
	(0.0134)	(0.0173)	(0.0118)	(0.0175)
ω	0.2770	0.3010	0.2350	0.1910
	(0.0192)	(0.0243)	(0.0161)	(0.0166)
α_1	0.0970	0.0520	0.1150	0.2240
	(0.0144)	(0.0136)	(0.0161)	(0.0194)
α_2	0.1000	0.1090	0.1280	0.1630
	(0.0155)	(0.0166)	(0.0178)	(0.0184)
α_3	0.0750	0.1950	0.0710	0.0880
	(0.0148)	(0.0197)	(0.0160)	(0.0158)
α_4	0.0830	0.0950	0.1080	0.0890
	(0.0154)	(0.0172)	(0.0178)	(0.0152)
α_5	0.0750	0.1030	0.0400	0.0780
	(0.0151)	(0.0166)	(0.0143)	(0.0146)
α_6	0.0580	0.0780	0.0730	0.0650
	(0.0137)	(0.0158)	(0.0161)	(0.0138)
α_7	0.0650	0.0740	0.0590	0.1020
	(0.0147)	(0.0152)	(0.0156)	(0.0162)
α_8	0.1040	0.0980	0.0690	0.0590
	(0.0165)	(0.0175)	(0.0155)	(0.0140)
α_9	0.0660	0.0700	0.0400	0.0560
	(0.0151)	(0.0160)	(0.0131)	(0.0135)
α_{10}	0.0570	0.0160	0.0370	0.0760
	(0.0136)	(0.0121)	(0.0143)	(0.0147)
$\sum_{i=1}^{10} \alpha_i$	0.7800	0.8910	0.7380	1.0000
log-lik.	−8447.04	−9619.05	−7710.99	−9243.7

The unconditional variance is given by

$$\sigma^2 = \omega / \left(1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j \right).$$

Therefore, the process ε_t is covariance stationary if and only if $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$. Notice that this is a sufficient but not necessary condition for ε_t to be strictly stationary (Bollerslev, 1986, Nelson, 1990, Bougerol and Picard, 1992). To see why, assume the following GARCH$(1,1)$ model

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

This model can be rewritten by successive iterations

$$\sigma_t^2 = \omega + \sigma_{t-1}^2 \left(\alpha_1 z_{t-1}^2 + \beta_1 \right) = \omega \left(1 + \sum_{\tau=1}^{\infty} \prod_{k=1}^{\tau} \left(\alpha_1 z_{t-k}^2 + \beta_1 \right) \right).$$

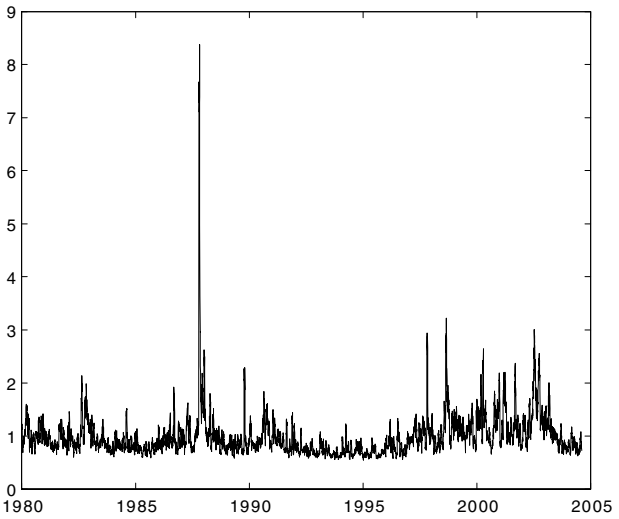

Fig. 4.1. *SP500 volatility estimated using an* $ARCH(10)$ *model.*

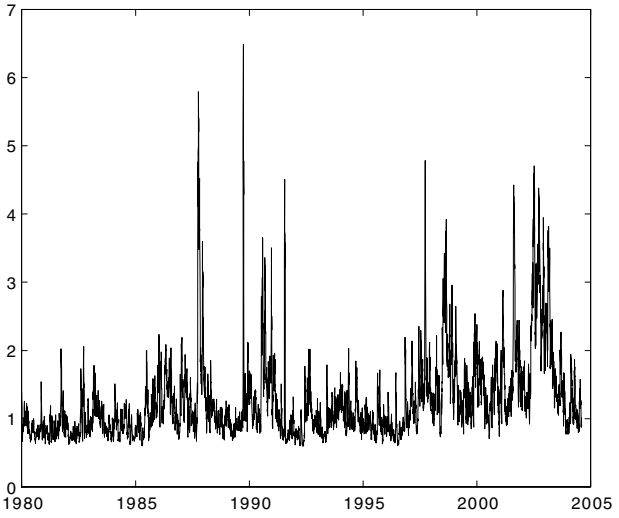

Fig. 4.2. *DAX volatility estimated using an* $ARCH(10)$ *model.*

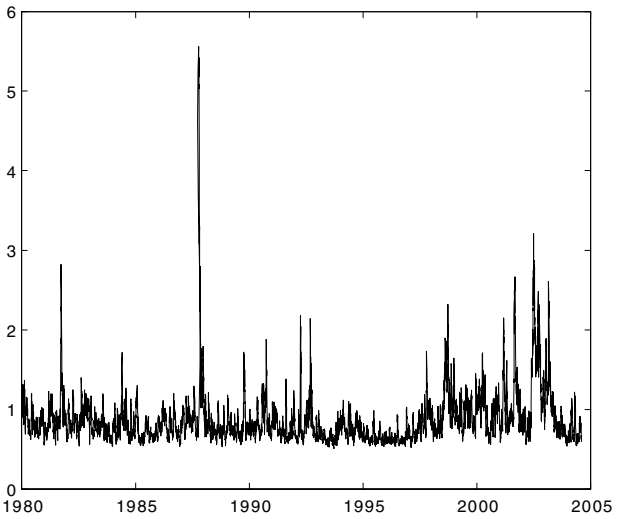

Fig. 4.3. *FT-SE volatility estimated using an* $ARCH(10)$ *model.*

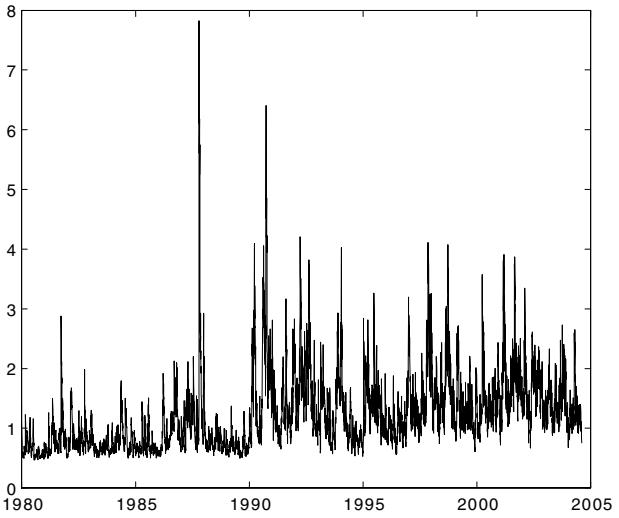

Fig. 4.4. *Nikkei volatility estimated using an* $ARCH(10)$ *model.*

Then, for this expression to converge toward a finite limit, so that the process ε_t is strictly stationary, we need

$$E\left[\log\left(\beta_1 + \alpha_1 z_{t-1}^2\right)\right] < 0.$$

We see that, due to the Jensen's inequality,

$$E\left[\log\left(\beta_1 + \alpha_1 z_{t-1}^2\right)\right] \leq \log\left(E\left[\beta_1 + \alpha_1 z_{t-1}^2\right]\right) = \log\left(\alpha_1 + \beta_1\right),$$

so that, even when $\alpha_1 + \beta_1 = 1$, the process is still strictly stationary.[3]

4.3.1 Forecasting

Forecasts of the GARCH model are obtained recursively in a similar way as for the ARCH model. Let t be the starting date for forecasting. Then, the 1-step ahead forecast for σ_{t+1}^2 is

$$\sigma_t^2(1) = \hat{\omega} + \hat{\alpha}_1 \hat{\varepsilon}_t^2 + \hat{\beta}_1 \hat{\sigma}_t^2.$$

Since $\varepsilon_t^2 = \sigma_t^2 z_t^2$, the GARCH(1, 1) model can be rewritten as

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 = \omega + \left(\alpha_1 + \beta_1\right)\sigma_{t-1}^2 + \alpha_1 \sigma_{t-1}^2\left(z_{t-1}^2 - 1\right),$$

so that, at time $t + 2$, we have

$$\sigma_{t+2}^2 = \omega + \left(\alpha_1 + \beta_1\right)\sigma_{t+1}^2 + \alpha_1 \sigma_{t+1}^2\left(z_{t+1}^2 - 1\right),$$

with $E\left[\left(z_{t+1}^2 - 1\right)|\mathcal{F}_t\right] = 0$. We deduce the following 2-step ahead forecast for σ_{t+2}^2:

$$\sigma_t^2(2) = \hat{\omega} + (\hat{\alpha}_1 + \hat{\beta}_1)\sigma_t^2(1).$$

Generally speaking, the κ-step ahead forecast for $\sigma_{t+\kappa}^2$ is

$$\sigma_t^2(\kappa) = \hat{\omega} + (\hat{\alpha}_1 + \hat{\beta}_1)\sigma_t^2(\kappa - 1) \qquad \text{for } \kappa > 1.$$

Alternatively, this last equation can be expressed as

$$\sigma_t^2(\kappa) = \hat{\sigma}^2 + (\hat{\alpha}_1 + \hat{\beta}_1)^{\kappa-1}\left(\sigma_t^2(1) - \hat{\sigma}^2\right) \qquad \text{for } \kappa > 1,$$

where $\hat{\sigma}^2 = \hat{\omega}/(1 - \hat{\alpha}_1 - \hat{\beta}_1)$. This expression shows that $\sigma_t^2(\kappa) \to \hat{\sigma}^2$ as $\kappa \to \infty$.

In empirical estimations, the parameters α and β tend to sum to a value that is very close to 1. It makes therefore sense to study the limit case, called an integrated GARCH model, where the sum of these parameters equals precisely one.

[3] For the ARCH(1) model, the condition for strict stationary becomes

$$E\left[\log(\alpha_1 z_{t-1}^2)\right] < 0.$$

Consequently, even when $\alpha_1 = 1$, we have $E\left[\log(z_{t-1}^2)\right] \leq \log(1)$, so that the process ε_t is strictly, but not covariance, stationary.

4.3.2 Integrated GARCH model

When $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1$, the unconditional variance σ^2 is not definite anymore. The process x_t is not covariance stationary anymore. However, it remains strictly stationary, because the unconditional density of ε_t does not change over time. Such a case corresponds to the Integrated GARCH (or IGARCH) described by Engle and Bollerslev (1986).

The IGARCH$(1,1)$ model can be written as

$$\sigma_t^2 = \omega + (1-\beta_1)\,\varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

with $0 < \beta_1 \leq 1$. We then see that

$$\sigma_t^2 = \omega + \sigma_{t-1}^2 + (1-\beta_1)\left(\varepsilon_{t-1}^2 - \sigma_{t-1}^2\right).$$

So, the κ-step ahead forecast for $\sigma_{t+\kappa}^2$ is

$$\sigma_t^2(\kappa) = (\kappa - 1)\,\hat{\omega} + \sigma_t^2(1), \qquad \text{for } \kappa \geq 1.$$

As it has been shown by Lumsdaine (1996), the statistical properties of the estimator of $(\omega, \beta_1)'$ are asymptotically normal.[4]

4.3.3 Estimation

Since this is the most widely adopted estimation, we only consider the Maximum Likelihood, ML, estimation of a GARCH(p,q) model. Assume that we have a time series of returns $\{x_1, \cdots, x_T\}$, and denote $m = \max(p,q)$ the number of observations lost for initializing the process. Under the normality assumption, the likelihood function of a GARCH(p,q) model is

$$\begin{aligned} f(\varepsilon_1, \cdots, \varepsilon_T|\theta) &= f(\varepsilon_T|\mathcal{F}_{T-1}) \times f(\varepsilon_{T-1}|\mathcal{F}_{T-2}) \times \cdots \times f(\varepsilon_{m+1}|\mathcal{F}_m) \\ &\quad \times f(\varepsilon_1, \cdots, \varepsilon_m|\theta) \\ &= \prod_{t=m+1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right) \times f(\varepsilon_1, \cdots, \varepsilon_m|\theta), \end{aligned}$$

with $\theta = \left(\omega, \alpha_1, \cdots, \alpha_p, \beta_1, \cdots, \beta_q\right)'$ the vector of unknown parameters and $f(\varepsilon_1, \cdots, \varepsilon_m|\theta)$ is the joint *pdf* of $(\varepsilon_1, \cdots, \varepsilon_m)$. In general, this term is simply dropped, to avoid specifying the joint distribution of the first observations. It is possible to justify this by observing that this term represents only one term in an usually long sequence of numbers. Omitting this term only introduces a negligible error in the likelihood. Hence, we focus on the *conditional likelihood function*

[4] Given that regression parameters involving integrated time series have non-normal distributions, which tend to be difficult to characterize explicitly, this is rather welcome news.

$$f\left(\varepsilon_{m+1},\cdots,\varepsilon_T|\theta,\varepsilon_1,\cdots,\varepsilon_m\right)=\prod_{t=m+1}^{T}\frac{1}{\sqrt{2\pi\sigma_t^2}}\exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right).$$

The ML estimator is obtained by maximizing this expression or, equivalently, the log-likelihood function

$$L_T\left(\theta|x_t,t=1,\cdots,T\right)=\sum_{t=m+1}^{T}\ell_t\left(\theta\right),$$

where

$$\ell_t\left(\theta\right)=-\frac{1}{2}\log\left(2\pi\right)-\frac{1}{2}\log\left(\sigma_t^2\right)-\frac{\varepsilon_t^2}{2\sigma_t^2}$$

is the log-likelihood of the observation t, with $\varepsilon_t=x_t-\mu_t\left(\theta\right)$.

Derivating the log-likelihood function yields the *gradient*

$$\frac{\partial\ell_t\left(\theta\right)}{\partial\theta}=-\frac{1}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}+\frac{\varepsilon_t^2}{2\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}=\frac{1}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\left(\frac{\varepsilon_t^2}{\sigma_t^2}-1\right),$$

with

$$\begin{aligned}\frac{\partial\sigma_t^2}{\partial\omega}&=1,\\ \frac{\partial\sigma_t^2}{\partial\alpha_i}&=\varepsilon_{t-i}^2,\qquad\text{for } i=1,\cdots,p,\\ \frac{\partial\sigma_t^2}{\partial\beta_j}&=\sigma_{t-j,}^2\qquad\text{for } j=1,\cdots,q.\end{aligned}$$

The second-order derivatives yield a matrix called *Hessian*, consisting of terms such as

$$\frac{\partial^2\ell_t\left(\theta\right)}{\partial\theta\partial\theta'}=-\frac{1}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'}\frac{\varepsilon_t^2}{\sigma_t^2}+\left(\frac{\varepsilon_t^2}{\sigma_t^2}-1\right)\frac{\partial}{\partial\theta'}\left(\frac{1}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\right),$$

with

$$\begin{aligned}E\left[\frac{\partial^2\ell_t\left(\theta\right)}{\partial\theta\partial\theta'}\right]&=-E\left[\frac{1}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'}\frac{\varepsilon_t^2}{\sigma_t^2}+\left(\frac{\varepsilon_t^2}{\sigma_t^2}-1\right)\frac{\partial}{\partial\theta'}\left(\frac{1}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\right)\right]\\&=-E\left[\frac{1}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'}\right].\end{aligned}$$

If we define $\xi_t=\left(1,\varepsilon_{t-1}^2,\cdots,\varepsilon_{t-p}^2,\sigma_{t-1}^2,\cdots,\sigma_{t-q}^2\right)'$, we may write $\sigma_t^2=\xi_t'\theta$, so that

$$E\left[\frac{\partial^2\ell_t\left(\theta\right)}{\partial\theta\partial\theta'}\right]=-E\left[\frac{\xi_t\xi_t'}{2\sigma_t^2}\right].$$

The covariance matrix of the ML estimator is given by the inverse of the information matrix, which is defined as minus the expected Hessian

$$I(\theta) = -E\left[\frac{\partial^2 \ell_t(\theta)}{\partial\theta\partial\theta'}\right] = E\left[\frac{\xi_t \xi_t'}{2\sigma_t^2}\right].$$

This expression can be estimated by

$$\hat{I}(\theta) = \frac{1}{2T}\sum_{t=1}^{T}\frac{\hat{\xi}_t'\hat{\xi}_t}{\hat{\sigma}_t^2}.$$

As shown by Weiss (1986), provided standardized innovations have finite fourth moments, the ML estimator is consistent and has the following asymptotic distribution

$$\sqrt{T}\left(\hat{\theta} - \theta\right) \Rightarrow \mathcal{N}\left(0, I(\theta)^{-1}\right).$$

For a regression model of the type

$$\begin{aligned} x_t &= X_t\delta + \varepsilon_t, \\ \varepsilon_t &= \sigma_t z_t, \\ \sigma_t^2 &= \omega + \sum_{i=1}^{p}\alpha_i\varepsilon_{t-i}^2 + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^2, \end{aligned}$$

where X_t is a set of exogenous explanatory variables and δ a set of unknown parameters pertaining to the conditional mean equation, we should estimate parameters δ and θ simultaneously. Fortunately, it turns out that the second derivatives of the regression model log-likelihood are such that (Engle, 1982)

$$E\left[\frac{\partial^2 \ell_t(\theta,\delta)}{\partial\theta\partial\delta'}\right] = 0.$$

Consequently, the information matrix is block-diagonal and the two sets of parameters can be estimated separately.

In practice, the estimation of a GARCH model should be performed as follows:

1. Estimate the mean equation $x_t = \mu_t + \varepsilon_t$. Deduce $\hat{\varepsilon}_t = x_t - \hat{\mu}_t$ and $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2$.
2. Select initial values for θ, say $\theta^0 = \left(\omega^0, \alpha_1^0, \cdots, \alpha_p^0, \beta_1^0, \cdots, \beta_q^0\right)$ and set $\hat{\varepsilon}_1^2 = \cdots = \hat{\varepsilon}_m^2 = \hat{\sigma}_1^2 = \cdots = \hat{\sigma}_m^2 = \hat{\sigma}^2$, with $m = \max(p, q)$.
3. Compute the conditional variance $\hat{\sigma}_t^2 = \hat{\omega}^0 + \sum_{i=1}^{p}\alpha_i^0\hat{\varepsilon}_{t-i}^2 + \sum_{j=1}^{q}\beta_j^0\hat{\sigma}_{t-j}^2$ for $t = m+1, \cdots, T$.
4. Compute the log-likelihood $L_T\left(\theta^0\right) = \sum_{t=m+1}^{T}\ell_t\left(\theta^0\right)$, where

$$\ell_t\left(\theta^0\right) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\hat{\sigma}_t^2\right) - \frac{\hat{\varepsilon}_t^2}{2\hat{\sigma}_t^2}.$$

   Change the value of parameters, say θ^1, so that the log-likelihood increases $L_T\left(\theta^1\right) > L_T\left(\theta^0\right)$.
5. Iterate steps 3 and 4 until convergence of the log-likelihood to a fixed value.

4.3.4 Testing for GARCH effects

Assume we want to test the null hypothesis that the process is homoskedastic against the alternative that the variance follows a GARCH$(1,1)$ process. Then, the null is $H_0 : \alpha_1 = \beta_1 = 0$, against $H_a : \alpha_1 \geq 0, \beta_1 \geq 0$ with at least one strict inequality.

A difficulty in constructing a statistic for this test is that, under the null hypothesis, the model is not identified. Indeed, for a GARCH$(1,1)$ process $\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$, we have when $\alpha_1 = 0$

$$\sigma_t^2 = \omega + \beta_1 \sigma_{t-1}^2 = \frac{\omega}{1-\beta_1} + \beta_1 \left(\sigma_{t-1}^2 - \frac{\omega}{1-\beta_1} \right) = \vartheta + \beta_1 \left(\sigma_{t-1}^2 - \vartheta \right).$$

Therefore, if we set, at date $t = 0$, $\sigma_0^2 = \vartheta$, then $\sigma_t^2 = \vartheta$ for all t. Then, β_1 is unidentified under the null. Under the null, the GARCH$(1,1)$ effect and the ARCH(1) effect are locally equivalent (see Lee, 1991). So, the test of $H_0 : \alpha_1 = \beta_1 = 0$, against $H_a : \alpha_1 \geq 0, \beta_1 \geq 0$ with at least one strict inequality, is equivalent to the test of no ARCH(1).

4.3.5 Software to estimate ARCH and GARCH models

Many packages exist that estimate the parameters of ARCH and GARCH models. Without pretending to be complete, we wish to dress a list for the reader who is interested in the actual estimation of the models discussed in this book. First, there exist software programs that are specific to econometric time series analysis and that allow the estimation of a wide range of ARCH and GARCH models. We may mention in this list the programs Eviews and Rats. There is also a list of statistics programs that either have modules or packages that estimate such models. In this class, we have STATA, S-Plus (see also Zivot and Wang, 2003), R. This latter is an open source program based on the S language.

Then there is a list of programming languages that use matrix language, which is convenient if own development is required. We may mention in this family Ox, which is a free program. Ox being a matrix language, it is possible to extend this language by additional ad hoc, user-written, modules. A module that estimates GARCH has been written by Laurent and Peters (2005). In this family of matrix language programs we have also MATLAB and Gauss. For the former program, GARCH code may be found in the UCSD Garch toolbox, written by Kevin Sheppard.[5] For the latter, we may consult Ronald Schoenberg's web page.[6]

[5] See http://www.kevinsheppard.com/research/ucsd_garch/ucsd_garch.aspx. Some GARCH codes may also be found under http://www.hec.unil.ch/mrockinger.

[6] See http://faculty.washington.edu/rons/.

For some really complex situations, we may also use Fortran code and sophisticated optimization code.[7] Using such a relatively low-level programming language allows one to obtain a most efficient optimized compiled code. The longevity of Fortran guarantees the existence of repeatedly tested modules.

4.3.6 Illustration

Table 4.2 reports the estimation of a GARCH$(1,1)$ model for the daily returns of our four market indices. The main interest of the GARCH model as compared with the initial ARCH model is that often the GARCH$(1,1)$ specification is enough to capture most dynamic in volatility. Consequently, the number of unknown parameters is reduced considerably.

We notice that the parameters α_1 and β_1 are found to be rather close for all indices at hand. In most cases, estimates of the GARCH$(1,1)$ model on returns yield $\alpha_1 + \beta_1 \approx 1$, so that the conditional variance is nearly integrated (Integrated GARCH model). We conclude that the volatility process may be unbounded although the return process is still strictly stationary.

The IGARCH process may reflect other dynamics for volatility. For instance, if the true model is a regime-switching model for volatility, estimating a GARCH model will generally result in a nearly integrated volatility process, a feature discussed in Gouriéroux and Jasiak (2001b). Finally, we observe that the log-likelihood significantly improves for all indices but the DAX, while the number of unknown parameters dramatically reduces. This suggests that the GARCH model is able to produce dynamics for the volatility that are very close to the ARCH model in a more parsimonious way.[8]

Table 4.2. *Estimation of a GARCH*(1,1) *model for daily log-returns*

	SP500	DAX	FT-SE	Nikkei
ω	0.012	0.037	0.023	0.018
	(0.0014)	(0.0029)	(0.0028)	(0.0019)
α_1	0.072	0.115	0.102	0.136
	(0.0016)	(0.0041)	(0.0064)	(0.0031)
β_1	0.919	0.868	0.872	0.864
	(0.0031)	(0.0063)	(0.0084)	(0.0040)
$\alpha_1 + \beta_1$	0.991	0.983	0.974	1.000
log-lik.	−8398.130	−9630.780	−7686.280	−9198.150

[7] Such as SNOPT or NPSOL, developed by Gill, Murray, and Saunders (1997, 1999). See also Jondeau and Rockinger (2003a, 2003b).

[8] The figures for the GARCH volatility are not reproduced here, because they look extremely similar to those for the ARCH volatility.

4.4 Asymmetric GARCH models

GARCH models are able to capture volatility clustering as well as some amount of fat-tailedness. However, in these models, positive and negative past values have a symmetric effect on the conditional variance. Yet, an abundant literature has documented that negative returns "bad news" tend to be followed by larger increases in volatility than equally large positive returns "good news" (Black, 1976, Christie, 1982, or French, Schwert, and Stambaugh, 1987).

Pagan and Schwert (1990) and Engle and Ng (1993) have defined the concept of *news impact curve*, which relates past return shocks (news) to current volatility. This curve measures how new information is incorporated into volatility estimates. In the GARCH model, this curve is a quadratic function centered on $\varepsilon_{t-1} = 0$. For asymmetric models, the curve is designed to increase differently in the two directions.

Several parameterizations have been proposed to capture such asymmetry in the response of volatility to shocks. To mention some of them, we have the Exponential GARCH (EGARCH) model of Nelson (1991), the Threshold GARCH (TGARCH) of Zakoïan (1994), the GJR model of Glosten, Jagannathan, and Runkle (1993), and the Absolute GARCH (AGARCH) model of Hentschel (1995). Some general expressions have been designed to incorporate the most well-known asymmetric GARCH models (Higgins and Bera, 1992, or Hentschel, 1995).

4.4.1 EGARCH model

The Exponential GARCH (EGARCH) model has been introduced by Nelson (1991) to improve two annoying aspects of the standard GARCH model: (1) the parameters α and β have to be constrained during the course of the estimation to ensure positivity of the variance process; (2) as already mentioned, empirical evidence suggests an asymmetric response of volatility to shocks. In the EGARCH model, σ_t^2 depends on both the size and the sign of lagged shocks:

$$\log\left(\sigma_t^2\right) = \omega + \sum_{i=1}^{p} \alpha_i g\left(z_{t-i}\right) + \sum_{j=1}^{q} \beta_j \log\left(\sigma_{t-j}^2\right), \tag{4.6}$$

with $\alpha_1 = 1$ and $g\left(z_t\right) = \left[\gamma\left(\left|z_t\right| - E\left[\left|z_t\right|\right]\right) + \psi z_t\right]$. In this specification, parameters are not restricted to be non-negative, because the conditional volatility is always positive. The process is strictly and covariance stationary if and only if $\sum_{j=1}^{q} \beta_j < 1$. Notice that $E\left[\left|z_t\right|\right] = \sqrt{2/\pi}$ under normality.[9]

This specification introduces the desired asymmetry in the following way. Over the range $0 < z_t < \infty$, $g\left(z_t\right)$ is linear in z_t with slope $\gamma + \psi$, whereas over the range $-\infty < z_t \leq 0$, $g\left(z_t\right)$ is linear in z_t with slope $\gamma - \psi$.

[9] For z_t distributed as a standardized Student t distribution with ν degrees of freedom, we have

$$E\left[\left|z_t\right|\right] = \frac{2\sqrt{\nu-2}\Gamma\left(\left(\nu+1\right)/2\right)}{\left(\nu-1\right)\Gamma\left(\nu/2\right)\sqrt{\pi}}.$$

4.4.2 TGARCH model

The Threshold GARCH (TGARCH) model (Zakoïan, 1994) is defined by:

$$\sigma_t = \omega + \sum_{i=1}^{p} \left[\alpha_i \left|\varepsilon_{t-i}\right| + \gamma_i \Pi_{t-i}^{-} \left|\varepsilon_{t-i}\right|\right] + \sum_{j=1}^{q} \beta_j \sigma_{t-j}, \tag{4.7}$$

with Π_t^- equal to 1 if $\varepsilon_t < 0$, and 0 otherwise.

The conditional volatility is positive when parameters satisfy $\omega > 0$, $\alpha_i \geq 0$, $\alpha_i + \gamma_i \geq 0$ and $\beta_j \geq 0$, for $i = 1, \cdots, p$ and $j = 1, \cdots, q$. Notice that ensuring covariance stationarity is not trivial, because it involves a rather non-linear constraint on parameters. In addition, this model does not nest the plain vanilla GARCH model.

4.4.3 GJR model

The *GJR* model (Glosten, Jagannathan, and Runkle, 1993) is closely related to the TGARCH model, because it is defined by:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \left[\alpha_i \varepsilon_{t-i}^2 + \gamma_i \Pi_{t-i}^{-} \varepsilon_{t-i}^2\right] + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2. \tag{4.8}$$

The conditional volatility is positive when parameters satisfy $\omega > 0$, $\alpha_i \geq 0$, $\alpha_i + \gamma_i \geq 0$ and $\beta_j \geq 0$, for $i = 1, \cdots, p$ and $j = 1, \cdots, q$. The process is covariance stationary if and only if $\sum_{i=1}^{p} (\alpha_i + \gamma_i/2) + \sum_{j=1}^{q} \beta_j < 1$ (Hentschel, 1995).

4.4.4 Cox-Box transform

Hentschel (1995) has put forward that many members of the GARCH family can be embedded in a Cox-Box transform of the form (in the case $p = q = 1$)

$$\frac{\sigma_t^{\gamma} - 1}{\gamma} = \omega + \alpha_1 \sigma_{t-1}^{\gamma} f^{\nu}(z_{t-1}) + \beta_1 \frac{\sigma_{t-1}^{\gamma} - 1}{\gamma}, \tag{4.9}$$

where $f(z_t) = |z_t - b| - c(z_t - b)$ is the news impact curve introduced by Pagan and Schwert (1990) and Engle and Ng (1993). The GARCH model is obtained from $\gamma = \nu = 2$, $b = c = 0$. The EGARCH model arises as a limit case when $\gamma = 0$, $\nu = 1$ and $b = 0$. The TGARCH model corresponds to $\gamma = \nu = 1$, $b = 0$, and $|c| \leq 1$. The GJR model is obtained from $\gamma = \nu = 2$, and $b = 0$.

For a generalized error distribution (GED) with fat-tailedness parameter ν, analyzed by Nelson (1991), we have

$$E\left[|z_t|\right] = \lambda 2^{1/\nu} \Gamma(2/\nu) / \Gamma(1/\nu),$$

where $\lambda = \left[2^{-2/\nu} \Gamma(1/\nu) / \Gamma(3/\nu)\right]^{1/2}$.

4.4.5 News impact curve

A comparison of the GARCH and GJR models suggests an interesting metric by which to analyze the effect of news on conditional volatility. Holding constant the information dated $t-2$ and earlier, we can examine the relation between ε_{t-1} and σ_t. This curve is called *news impact curve*. It relates past return shocks (news) to current volatility. It measures how new information is incorporated into volatility estimates.

For the GARCH$(1,1)$ model, the curve is a quadratic function centered on $\varepsilon_{t-1}=0$,

$$\sigma_t^2 = A + \alpha_1 \varepsilon_{t-1}^2,$$

with $A=\omega+\beta_1\sigma^2$ and σ^2 the unconditional variance.

For the GJR$(1,1)$ model, it has a minimum at $\varepsilon_{t-1}=0$ with

$$\sigma_t^2 = \begin{cases} A + \alpha_1 \varepsilon_{t-1}^2 & \text{for } \varepsilon_{t-1} > 0, \\ A + (\alpha_1 + \gamma_1)\, \varepsilon_{t-1}^2 & \text{for } \varepsilon_{t-1} \leq 0, \end{cases}$$

with $A=\omega+\beta_1\sigma^2$.

For the EGARCH$(1,1)$ model, it has a minimum at $\varepsilon_{t-1}=0$ with

$$\sigma_t^2 = \begin{cases} A \exp\left(\frac{\psi+\gamma}{\sigma}\varepsilon_{t-1}\right) & \text{for } \varepsilon_{t-1} > 0, \\ A \exp\left(\frac{\psi-\gamma}{\sigma}\varepsilon_{t-1}\right) & \text{for } \varepsilon_{t-1} \leq 0, \end{cases}$$

with $A=\sigma^{2\beta_1}\exp\left(\omega-\gamma\sqrt{2/\pi}\right)$.

Figure 4.5 illustrates the difference in the news impact curve between a symmetric GARCH model, represented by a continuous line, and an asymmetric (GJR) model (with $\alpha_1=0.05$ and $\gamma_1=0.1$), represented by a dotted and dashed line. For the GARCH model, the effect on volatility of a shock is the same, be it positive or negative. For the GJR model, by contrast, we notice a much stronger effect on the volatility when the lagged shock is negative.

4.4.6 Partially non-parametric estimation

Engle and Ng (1993) have proposed an alternative approach to estimating the news impact curve based on a partially non-parametric procedure, inspired by Gouriéroux and Monfort (1992). The range of ε_t is divided into m^- intervals when ε_{t-1} is negative and m^+ intervals when ε_{t-1} is positive, with $2m = m^- + m^+$. We define the $2m+1$ break points τ_i, $i=-m^-,\cdots,m^+$. Assuming $\tau_0=0$ allows one to test asymmetry. Finally, we define

$$P_{i,t} = 1, \text{ if } \varepsilon_t > \tau_i \text{ and 0 otherwise and}$$
$$N_{i,t} = 1, \text{ if } \varepsilon_t \leq \tau_{-i} \text{ and 0 otherwise.}$$

Then, the piecewise linear specification of the heteroskedasticity function is

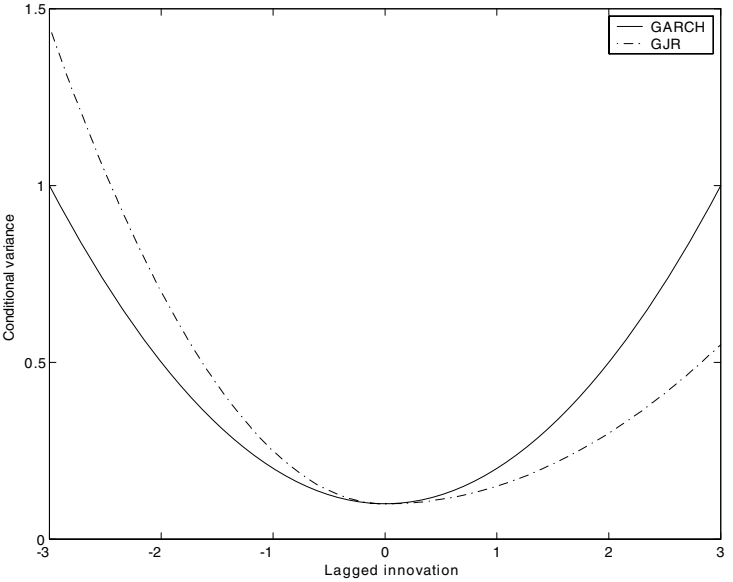

Fig. 4.5. *News impact curve.*

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \sum_{i=0}^{m^+} \theta_i P_{i,t-1} \times (\varepsilon_{t-1} - \tau_i) + \sum_{i=0}^{m^-} \delta_i N_{i,t-1} \times (\varepsilon_{t-1} - \tau_{-i})\,.$$

The functional form is a linear spline with knots at the τ_is. Between 0 and τ_1 the slope is θ_0, between τ_1 and τ_2 it is $\theta_0 + \theta_1$, and so forth. Between 0 and τ_{-1}, the slope is δ_0, between τ_{-1} and τ_{-2} it is $\delta_0 + \delta_1$, and so forth. A simple approach to select the τ_is is to use equally spaced bins with break points at $i \times \sigma$ for $i = 0, \pm 1, \pm 2, \cdots$, where σ is the unconditional standard deviation of ε_t. In such case, we have

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \sum_{i=0}^{m} \theta_i P_{i,t-1} \times (\varepsilon_{t-1} - i\sigma) + \sum_{i=0}^{m} \delta_i N_{i,t-1} \times (\varepsilon_{t-1} + i\sigma)\,.$$

4.4.7 Testing for asymmetric effects

Tests for asymmetry in the conditional variance have been proposed by Engle and Ng (1993). These tests examine whether we can predict the squared standardized residuals by some past variables that are not included in the hypothetized volatility model and may reflect an asymmetric response of volatility to shocks.

Assume that the hypothetized volatility model is a standard GARCH$(1,1)$ model. The volatility dynamic thus depends on $\xi_{0,t} = \left(1, \varepsilon_{t-1}^2, \sigma_{t-1}^2\right)'$, so that we have $\sigma_{0,t}^2 = \xi_{0,t}'\theta_0$, with $\theta_0 = (\omega, \alpha_1, \beta_1)'$. Consider a set of

missing explanatory variables $\xi_{a,t}$ and their pertaining parameters θ_a, with $\log\left(\sigma^2_{a,t}\right) = \xi'_{a,t}\theta_a$.[10] The nesting model can then be defined as

$$\log\left(\sigma_t^2\right) = \log\left(\xi^{'}_{0,t}\theta_0\right) + \xi'_{a,t}\theta_a.$$

Now, define $z_t = \varepsilon_t/\sigma_{0,t}$ the standardized residual obtained under the hypothetized (GARCH) model, and consider the auxiliary regression

$$z_t^2 = \xi^{*^{'}}_{0,t}\theta_0 + \xi^{*^{'}}_{a,t}\theta_a + e_t,$$

with

$$\xi^*_{0,t} = \frac{1}{\sigma^2_{0,t}}\frac{\partial\sigma_t^2}{\partial\theta_0},$$

$$\xi^*_{a,t} = \frac{1}{\sigma^2_{0,t}}\frac{\partial\sigma_t^2}{\partial\theta_a} = \frac{\sigma_t^2\xi_{a,t}}{\sigma^2_{0,t}},$$

and e_t the error term.

Then, we construct a Lagrange-Multiplier test statistic for the null hypothesis $H_0 : \theta_a = 0$ in the auxiliary regression above. Therefore, all variables in $\xi^*_{0,t}$ and $\xi^*_{a,t}$ are evaluated at ML estimator of θ_0 and $\theta_a = 0$, so that we finally have

$$\xi^*_{0,t} = \frac{\xi_{0,t}}{\sigma^2_{0,t}},$$

$$\xi^*_{a,t} = \xi_{a,t},$$

In addition, under the null, the left-hand side variable z_t^2 is theoretically orthogonal to the two right-hand side sets of variables. Therefore, the LM test statistic can be computed as the $T \times R^2$ associated with the regression. Under the null, the test statistic is asymptotically distributed as a χ^2 with a degree of freedom equal to the number of restrictions.

In practice, variables in $\xi_{a,t}$ have to be chosen in order to test the presence of asymmetric effects. Engle and Ng (1993) propose a series of complementary tests based on the following regressions:

$$z_t^2 = a + b\Pi^-_{t-1} + \xi^{*^{'}}_{0,t}\theta_0 + e_t, \tag{4.10}$$

$$z_t^2 = a + b\Pi^-_{t-1}\varepsilon_{t-1} + \xi^{*^{'}}_{0,t}\theta_0 + e_t, \tag{4.11}$$

$$z_t^2 = a + b\Pi^+_{t-1}\varepsilon_{t-1} + \xi^{*^{'}}_{0,t}\theta_0 + e_t, \tag{4.12}$$

$$z_t^2 = a + b_1\Pi^-_{t-1} + b_2\Pi^-_{t-1}\varepsilon_{t-1} + b_3\Pi^+_{t-1}\varepsilon_{t-1} + \xi^{*^{'}}_{0,t}\theta_0 + e_t, \tag{4.13}$$

[10] For instance, the missing variables may have the form of an EGARCH(1, 1) model. The variables involved in this model would then be

$$\xi_{a,t} = \left(\frac{\varepsilon_{t-1}}{\sigma_{t-1}}, \left(\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| - \sqrt{2/\pi}\right), \log\left(\sigma^2_{t-1}\right)\right)',$$

and the parameter set would be $\theta_a = \left(\tilde{\psi}, \tilde{\gamma}, \tilde{\beta}_1\right)'$, with $\log\left(\sigma^2_{a,t}\right) = \xi'_{a,t}\theta_a$.

where $\Pi_{t-1}^{-} = 1_{(\varepsilon_{t-1} \leq 0)}$ and $\Pi_{t-1}^{+} = 1 - \Pi_{t-1}^{-}$. The first test, the *Sign bias test*, examines the impact of positive and negative shocks on the conditional variance not predicted by the model under consideration. This is done by testing whether the variable Π_{t-1}^{-} has a predictive power on squared standardized residuals z_t^2 (equation (4.10)). The test statistic is the t-ratio of parameter b. The *Negative size bias test* examines whether the given model can explain the different effects that large and small negative shocks have on the conditional variance. The variable used for this test is $\Pi_{t-1}^{-}\varepsilon_{t-1}$ (equation (4.11)). The *Positive size bias test* examines whether the linear model can explain the different effects that large and small positive shocks have on the conditional variance. The variable used for this test is $\Pi_{t-1}^{+}\varepsilon_{t-1}$ (equation (4.12)). The last test considers the three previous hypotheses simultaneously. Under the null hypothesis $b_1 = b_2 = b_3 = 0$ and $\theta_0 = 0$, the test statistic is the $T \times R^2$ associated with (4.13). It is distributed as a $\chi^2(3)$. Notice that, if $\xi_{0,t}^*$ is not included in the regression (4.13), the test will be conservative.

4.4.8 Illustration

In the following of our previous estimates, we now consider the estimation of asymmetric GARCH models. Table 4.3 reports the estimation of three asymmetric GARCH$(1,1)$ models for daily returns under study: the EGARCH, GJR, and TGARCH models. Loosely speaking, the three models provide rather similar evidence: in almost all cases, the estimates suggest, as expected, that "bad news" has a stronger effect on volatility than "good news".[11]

If we turn to the log-likelihood estimates, we observe that no model systematically dominates the others. For instance, the EGARCH model performs very well for the Nikkei but very badly for the FT-SE. In contrast, the GJR model captures the dynamic of the DAX volatility best.

4.5 GARCH model with jumps

As seen in Chapter 3, the release of news is likely to generate volatility. So far, in this chapter, we have addressed the issue of modeling volatility under the assumption that news arrives, in a way that may be modeled with a GARCH model. An alternative is to assume that there is a sudden information release, followed by a period where this news gets gradually incorporated into the market. This suggests that in addition to the GARCH phenomenon, it is possible that returns suddenly jump. Such a model may be reasonable because the residuals of a GARCH model are in general not distributed according to a Gaussian distribution (see Chapter 5). This observation suggests that

[11] Notice that a negative asymmetry translates into a negative ψ parameter for the EGARCH model but a positive γ_1 for the GJR and TGARCH models. The only exception is therefore the estimate of the EGARCH model for the FT-SE.

Table 4.3. *Parameter estimates of asymmetric GARCH models*

	SP500	DAX	FT-SE	Nikkei
GARCH(1, 1)				
ω	0.012	0.037	0.023	0.018
	(0.0014)	(0.0029)	(0.0028)	(0.0019)
α_1	0.072	0.115	0.102	0.136
	(0.0016)	(0.0041)	(0.0064)	(0.0031)
β_1	0.919	0.868	0.872	0.864
	(0.0031)	(0.0063)	(0.0084)	(0.0040)
log-lik.	−8398.13	−9630.78	−7686.28	−9198.15
EGARCH(1, 1)				
ω	0.003	0.013	−0.036	0.016
	(0.0011)	(0.0010)	(0.0069)	(0.0014)
γ	0.133	0.174	0.441	0.244
	(0.0064)	(0.0061)	(0.0223)	(0.0061)
ψ	−0.077	−0.062	0.036	−0.111
	−(0.0042)	−(0.0043)	(0.0107)	−(0.0046)
β_1	0.981	0.974	0.715	0.974
	(0.0016)	(0.0023)	(0.0156)	(0.0020)
log-lik.	−8325.7	−9610.37	−7905.72	−9087.97
GJR(1, 1)				
ω	0.018	0.041	0.023	0.02
	(0.0014)	(0.0031)	(0.0026)	(0.0019)
α_1	0.025	0.053	0.055	0.053
	(0.0047)	(0.0074)	(0.0079)	(0.0051)
γ_1	0.095	0.1	0.069	0.168
	(0.0052)	(0.0069)	(0.0067)	(0.0064)
β_1	0.911	0.874	0.882	0.863
	(0.0038)	(0.0069)	(0.0084)	(0.0042)
log-lik.	−8343.49	−9595.55	−7664.5	−9103.32
TGARCH(1, 1)				
ω	0.019	0.034	0.027	0.034
	(0.0029)	(0.0044)	(0.0039)	(0.0041)
α_1	0.029	0.061	0.07	0.059
	(0.0057)	(0.0078)	(0.0086)	(0.0068)
γ_1	0.088	0.074	0.063	0.129
	(0.0079)	(0.0082)	(0.0079)	(0.0098)
β_1	0.926	0.897	0.89	0.876
	(0.0063)	(0.0085)	(0.0096)	(0.0082)
log-lik.	−8325.7	−9604.9	−7661.7	−9105.1

additional phenomena take place that are not captured by the simple GARCH model. In this section, we discuss some contributions that combine discrete GARCH models with jump models that may capture this sudden release of information.

An early contribution in this field is Press (1967) who considers the possibility that returns are generated by a superposition of a Gaussian distribution and a jump component. For estimating this model, Press (1967) proposes using the method of moments.[12] The next milestone is Jorion (1989) who assumes that returns are generated by an ARCH model that incorporates jumps. He finds that exchange rates are jumpier than stock market returns. This model has been extended by Vlaar and Palm, (1993) to GARCH models. As a more recent contribution, we may mention Chan and Maheu (2002) and Maheu and McCurdy (2004). Since their contribution nests the earlier contributions that appeared in the literature, we focus here on their model.

4.5.1 A model with time-varying jump intensity

Consider a time series of returns x_t, $t = 1, \cdots, T$. The following model, named by Maheu and McCurdy (2004) GARJI, follows the spirit of the GARCH literature and superposes a compound Poisson process to the GARCH with a time-varying jump intensity. The model is

$$x_t = \mu + \varepsilon_t, \tag{4.14}$$

$$\varepsilon_t = \varepsilon_{1,t} + \varepsilon_{2,t}, \tag{4.15}$$

$$\varepsilon_{1,t} = \sigma_t z_t, \tag{4.16}$$

$$z_t \sim \mathcal{N}(0,1), \tag{4.17}$$

$$\sigma_t^2 = \omega + g(\Lambda, \mathcal{F}_{t-1})\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \tag{4.18}$$

$$\begin{aligned} g(\Lambda, \Phi_{t-1}) = \exp\big(&\alpha + \alpha_J E[n_{t-1}|\mathcal{F}_{t-1}] \\ &+1_{\{\varepsilon_{t-1}<0\}}\left[\alpha_N + \alpha_{N,J} E[n_{t-1}|\mathcal{F}_{t-1}]\right]\big). \end{aligned} \tag{4.19}$$

Equation (4.14) indicates that returns have a mean μ and an innovation, ε_t. This innovation consists, as (4.15) indicates, of two components. The first component follows a GARCH model. This is described by the (4.16)–(4.19). The source of uncertainty of the GARCH, z_t, is assumed to be normally distributed as indicated in (4.17). The second component involves a Poisson process n_t whose specification is discussed below. The volatility dynamics in (4.19) depends on the innovation ε_t, that is both the part coming from the GARCH as well as from the jumps. This is a reasonable assumption, in that we would expect that a sudden burst of information, reflected in an increase of ε_t, affects subsequent volatility.

[12] These estimates, obtained for a simple model without the GARCH effect, may be used as starting values for more complicated models.

Equation (4.19) allows for a general news impact curve nesting the GJR and Threshold GARCH models. As in general for conditional models, the specifications depend on some information set $\mathcal{F}_{t-1}$. This can be past return variables or other observable variables. The only requirement, in order for some variable to belong to that information set, is that the variable is known at time t. In (4.19) figures an indicator variable $1_{\{\varepsilon<0\}}$, taking the value of 1 if $\varepsilon < 0$, and the value 0, otherwise. The chosen specification allows for the introduction of a differential impact if past news are deemed good or bad. If past news are "business as usual", in the sense that no jump occurred, and are positive, then the impact on current volatility will be $\exp(\alpha)\varepsilon_{t-1}^2$. If no jump took place but news is bad, the volatility impact becomes $\exp(\alpha+\alpha_N)\varepsilon_{t-1}^2$. If a jump took place, with good news, the impact is $\exp(\alpha+\alpha_J)\varepsilon_{t-1}^2$, and last, if there is a jump with bad news, then the impact becomes $\exp(\alpha+\alpha_J+\alpha_N+\alpha_{N,J})$.

The arrival rate of the jumps is assumed to follow a Poisson distribution. Leaving the derivation of the Poisson distribution to Chapter 16, we wish to mention here that the probability that $n_t = j$ jumps occurred on a given day t is given by

$$\Pr[n_t = j|\mathcal{F}_{t-1}] = e^{-\lambda_t}\frac{\lambda_t^j}{j!}. \tag{4.20}$$

The parameter λ_t measures the intensity of jumps. It measures the expected value of jump arrivals over unit time intervals. Here, it is assumed that the jump frequency is time dependent. A Poisson process with time-varying parameter is also called a non-homogenous Poisson process in the statistics literature. The jump component, $\varepsilon_{2,t}$, can be written as

$$\varepsilon_{2,t} = \sum_{k=1}^{n_t} Y_{t,k} - \lambda_t\theta, \tag{4.21}$$

$$Y_{t,k} \sim \mathcal{N}(\theta, \delta^2). \tag{4.22}$$

The first equation, (4.21), specifies that the jump component is the sum of a certain amount of shocks, assumed here to be Gaussian, with the parameters indicated in (4.22). Jumps have expected size θ and dispersion δ^2. The term $\lambda_t\theta$ in (4.21) has been introduced so that, for identification purposes, $E[\varepsilon_{2,t}] = 0$.[13]

In order to estimate the model, it is necessary to provide an expression for the jump intensity. Maheu and McCurdy (2004) assume that it evolves through time as an autoregressive process given by

$$\lambda_t = \lambda_0 + \rho\lambda_{t-1} + \gamma\xi_{t-1}. \tag{4.23}$$

[13] Without this condition, the expected return would be $E[x_t] = \mu + \theta\lambda_t$ and it would not be possible to distinguish the μ from the θ. By subtracting $\lambda_t\theta$, we obtain a so-called compensated Poisson process. Its expected value is zero.

The source of innovation to the jump component is assumed to be conditional on time $t-1$ rather than being generated at time t. It is this key assumption that allows Maheu and McCurdy (2004) to estimate the entire model. Using the definition of an innovation as being the difference between the expected value and the actual value, they obtain

$$\xi_{t-1} = E[n_{t-1}|\mathcal{F}_{t-1}] - \lambda_{t-1} = \sum_{j=0}^{\infty} j \Pr[n_{t-1} = j|\mathcal{F}_{t-1}] - \lambda_{t-1}. \tag{4.24}$$

This expression could be estimated if $\Pr[n_{t-1} = j|\mathcal{F}_{t-1}]$ were known. As shown by Maheu and McCurdy (2004), this expression may be inferred using Bayes' formula. We have

$$\begin{aligned}
\Pr\left[n_t = j|\mathcal{F}_t\right] &= \Pr[n_t = j|x_t, \mathcal{F}_{t-1}] = \frac{\Pr[n_t = j, x_t, \mathcal{F}_{t-1}]}{\Pr[x_t, \mathcal{F}_{t-1}]} && (4.25)\\
&= \frac{\Pr[x_t|n_t = j, \mathcal{F}_{t-1}] \times \Pr[n_t = j|\mathcal{F}_{t-1}] \Pr[\mathcal{F}_{t-1}]}{\Pr[x_t|\mathcal{F}_{t-1}] \times \Pr[\mathcal{F}_{t-1}]} \\
&= \frac{\Pr[x_t|n_t = j, \mathcal{F}_{t-1}] \times \Pr[n_t = j|\mathcal{F}_{t-1}]}{\Pr[x_t|\mathcal{F}_{t-1}]} && (4.26)\\
&= \frac{f(x_t|n_t = j, \mathcal{F}_{t-1}) \times \Pr[n_t = j|\mathcal{F}_{t-1}]}{f(x_t|\mathcal{F}_{t-1})}. && (4.27)
\end{aligned}$$

The first equation (4.25) indicates that the fundamental information available at time t is $\mathcal{F}_t$, which contains the new information on x_t and the one of the previous returns captured in $\mathcal{F}_{t-1}$. The next two equations use the formula of conditional probabilities. The step from (4.26) to (4.27) is just notation. From there on, it is possible to evaluate (4.24), since the various components of (4.27) are known. Indeed, conditional on knowing λ_t, σ_t, and the number of jumps that took place over a time interval, $n_t = j$, the density of x_t is

$$f\left(x_t|\mathcal{F}_{t-1}\right) = \sum_{j=0}^{\infty} f(x_t|n_t = j, \mathcal{F}_{t-1}) \times \Pr[n_t = j|\mathcal{F}_{t-1}], \tag{4.28}$$

with

$$f\left(x_t|n_t = j, \mathcal{F}_{t-1}\right) = \frac{1}{\sqrt{2\pi(\sigma_t^2 + j\delta^2)}} \exp\left(-\frac{(x_t - m_t)^2}{2(\sigma_t^2 + j\delta^2)}\right), \tag{4.29}$$

and $m_t = \mu - \theta\lambda_t + \theta j$. Equation (4.29) states that, if the number of jumps is known, the overall return will be normal. This follows from the fact that returns are the sum of the fundamental Gaussian part augmented by the Gaussian jumps. The second equation specifies the density of a return unconditionally on the number of jumps. The term $\Pr[n_t = j|\mathcal{F}_{t-1}]$ is given by (4.20). Hence, at this stage, all the components are computable.

Before proceeding, we need to provide conditions ensuring that all the dynamics are well defined. In an empirical illustration, we use a GARCH dynamics such as discussed earlier (equation (4.5)) for which positivity and stationarity conditions are easy to impose. Concerning (4.23), Chan and Maheu (2002), and Maheu and McCurdy (2004) notice that this equation may be rewritten as

$$\lambda_t = \lambda_0 + (\rho - \gamma)\lambda_{t-1} + \gamma E[n_{t-1}|\mathcal{F}_{t-1}].$$

This equation shows that a sufficient condition ensuring positivity of λ_t is that $\lambda_0 > 0$, $\rho \geq \gamma$, and $\gamma \geq 0$. The condition for stationarity is that $|\rho| < 1$. The expected level of jumps is then given by

$$E[\lambda_t] = \frac{\lambda_0}{1-\rho}.$$

We may now discuss some of the restrictions to the model. The most important restriction is to impose constancy to the jump intensity parameter, $\lambda_t = \lambda$, for all t. In that case, it is possible to estimate the parameters with a direct Maximum Likelihood estimation. The algorithm for the estimation is similar to the one outlined for the ARCH model, see Section 4.3.3, with the difference being the density of the returns. Here the likelihood of one observation is given by (4.28). We notice that the log-likelihood will not take the usual simple expression but will consist of a sum. In theory, the sum involved in computing (4.28) is infinite. In practice, the probability of multiple jumps quickly becomes small. Hence, the infinite sum can be truncated by a small number of terms.[14] This restriction corresponds to the model of Vlaar and Palm (1993). By assuming that the GARCH component is only an ARCH model, we obtain the model of Jorion (1989). Last, if we assume that the ARCH model reduces to a constant volatility, we obtain the model proposed by Press (1967).

Useful moments computed by Press (1967), see also Das and Sundaram (1997), adapted to a conditional setting, are

[14] In practice, truncation to the first 10 or 20 terms of the sum appears to be largely sufficient to obtain stable estimates. Jorion (1989), using exchange rate data and asset returns, truncates after 10 terms. Maheu and McCurdy (2004) propose, for stock market data, truncation after 25 terms. In practice, it is a good idea to perform the estimation for a set of increasing terms 10, 20, 30, and to verify the stability of the estimations. It is also possible to inspect the probabilites of multiple jumps, intervening in the likelihood. This probability will be found to be very small in practical implementations.

$$
\begin{aligned}
E[x_t|\mathcal{F}_{t-1}] &= \mu, \\
V[x_t|\mathcal{F}_{t-1}] &= \sigma_t^2 + (\theta^2 + \delta^2)\lambda_t, \\
Sk[x_t|\mathcal{F}_{t-1}] &= \frac{\lambda_t(\theta^3 + 3\theta\delta^2)}{(\sigma_t^2 + \lambda_t\delta^2 + \lambda_t\theta^2)^{3/2}}, \\
Ku[x_t|\mathcal{F}_{t-1}] &= 3 + \frac{\lambda_t(\theta^4 + 6\theta^2\delta^2 + 3\delta^4)}{(\sigma_t^2 + \lambda_t\delta^2 + \lambda_t\theta^2)^2},
\end{aligned}
$$

where $Sk[x_t|\mathcal{F}_{t-1}]$ and $Ku[x_t|\mathcal{F}_{t-1}]$ denote the conditional skewness and kurtosis, respectively. We notice that the sign of the jumps, θ, will determine the sign of the skewness.

A further useful result concerns the conditional probability of observing several jumps. This probability may be computed as

$$
\Pr[n_t \geq 1|\mathcal{F}_t] = 1 - \Pr[n_t = 0|\mathcal{F}_t]. \tag{4.30}
$$

This latter term is already computed during the estimation of the model.

4.5.2 An empirical illustration

To illustrate the working of the model with jumps, we consider the SP500 as well as the DAX daily returns. We estimate model (4.14)–(4.17) with a simple GARCH volatility given by

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

and the dynamics for the intensity parameter given by (4.23). We also estimate the parameters of the model when the volatility dynamics and the jump intensity constant over time, $\sigma_t = \sigma$, $\forall t$, and $\lambda_t = \lambda$, $\forall t$.

Table 4.4 displays the parameter estimates for the restricted and the general model. The jump size, θ, is always negative corroborating the stylized fact that returns are negatively skewed. We notice that θ is larger when the GARCH effect is introduced. For instance, for the SP500, the jump size evolves from $\theta = -0.024$, obtained for the model without GARCH nor time variation in the jump intensity, to $\theta = -0.503$ for the general model. Our interpretation is that without the GARCH dynamics, the jump component is forced to capture many of the variations due to the time-varying volatility and therefore does not necessary correspond to actual jumps. Clearly, as the jump size becomes larger, the intensity of the jumps diminishes. We notice for instance for the SP500 that $\lambda = 0.755$ for the simple model and $E[\lambda_t] = 0.21$ for the general model. This means that, for the general model, a jump takes place about every fifth day.

When we focus on the dynamics of the jump intensity, we notice by inspecting the parameter ρ that there is important persistence in the λ parameter implying a rather smooth evolution of this parameter through time. If we consider the log-likelihoods of the restricted and general models, we notice

Table 4.4. *Parameters for a GARCH model combined with a jump process with time-varying intensity and a restricted model without time-varying volatility nor time-varying jump intensity*

	SP500	DAX	SP500	DAX
μ	0.035	0.035	0.036	0.032
	(0.010)	(0.012)	(0.013)	(0.016)
θ	−0.503	−0.646	−0.024	−0.226
	(0.113)	(0.093)	(0.023)	(0.069)
δ	1.099	1.191	1.009	1.981
	(0.111)	(0.105)	(0.051)	(0.116)
λ	–	–	0.755	0.252
			(0.098)	(0.035)
λ_0	0.015	0.016	–	–
	(0.005)	(0.005)		
ρ	0.930	0.944	–	–
	(0.034)	(0.014)		
γ	0.508	0.791	–	–
	(0.111)	(0.160)		
σ	–	–	0.480	0.855
			(0.032)	(0.024)
w	0.007	0.009	–	–
	(0.003)	(0.007)		
α	0.028	0.029	–	–
	(0.007)	(0.018)		
β	0.950	0.950	–	–
	(0.011)	(0.029		
log-lik.	−8363.0	−9603.9	−8761.7	−10351.1
$E[\lambda_t]$	0.2130	0.2920	–	–

a statistically significant change as we introduce jumps to a simple GARCH model and a further statistically significant change as we introduce time variation in the jump intensity.

It is also possible to estimate the temporal evolution of the probability that a jump took place over a given day, see (4.30). The evolution of the jump probability for the SP500 over the period starting with the beginning 1999 and ending with the year 2002 is provided in Figure 4.6.[15] As we could expect from the relatively high parameter ρ and the ensuing smoothness of λ_t, the jump probability evolves relatively smoothly over time. Moreover, the period in the middle of 2002 related to the war in Iraq also corresponds to a period of increased jump probabilities. Last, we can consider the temporal evolution implied by this model for skewness and kurtosis. This is given in Figure 4.7.

[15] We could have chosen a much larger window, however, the picture becomes very erratic if we consider too many observations.

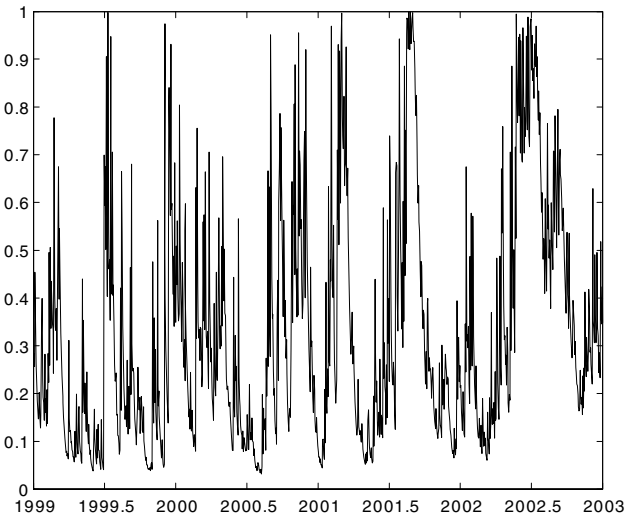

Fig. 4.6. *Estimation of the jump probability* $\Pr[n_t \geq 1|I_t]$ *for the SP500.*

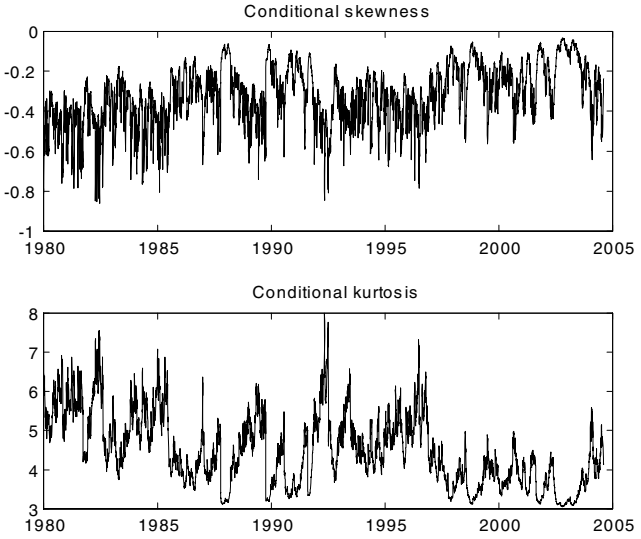

Fig. 4.7. *Evolution of the conditional skewness and kurtosis for the SP500.*

As these figures show, the higher moments of the data are found to evolve over time. For the October 1987 crash, we notice some limitations of this model. Indeed, the skewness parameter becomes very small, close to 0, and kurtosis approaches 3, the value a Gaussian distribution would take. This finding comes from the fact that the 1987 crash induced a huge upward jump in volatility. This increase in volatility, captured by the GARCH model, absorbs all the return variations that might otherwise be judged as a jump. This is related to an old debate of the empirical finance literature namely if we should include or not the 1987 crash in the model. Some pundits claim that any satisfactory model should be able to describe this outlier. Others are willing to discard those few observations surrounding the 1987 crash on the ground that the variation of that day cannot be encompassed in a general framework. With the model at hand, the λ_t parameter takes the value of 6, suggesting that on that day, 6 crashes took place. If this is a possibility remains an open question. We also notice that the range of variation of these two moments is rather large. The conditional skewness tends to be less negative over the recent period, while kurtosis tends to be closer to 3, suggesting that the distribution of the SP500 has been recently closer to the normal distribution. The time variation of such moments is important for portfolio allocation and will also be considered in Chapter 9.

4.6 Aggregation of GARCH processes

So far, we have seen that the description of the temporal evolution of return series, and in particular its time-varying volatility, may be relatively well captured by GARCH processes. At this stage, new questions arise. The first one concerns the evolution of the parameters of GARCH processes, as we use different time frequencies for the estimation. For instance, we may be tempted to use daily data to estimate a process that should, however, be scaled to some other frequency, say a monthly one. The answer is that the GARCH process, studied so far, does not aggregate, a result shown in an important paper by Drost and Nijman (1993). These authors introduce a new class of processes, that they name weak GARCH model, that aggregate, however.

The second question concerns the cross-section aggregation, i.e., what will be the resulting process for a portfolio of assets, if each asset has GARCH type returns? The answer is again disappointing. As shown by Nijman and Sentana (1996), a linear combination of assets, each with a GARCH dynamics, will no longer be distributed as a GARCH model in the usual sense.

In this section, we discuss the main results, relevant for the temporal as well as the cross-sectional aggregation of GARCH processes.

4.6.1 Temporal aggregation

Theoretical aspects

Drost and Nijman (1993) prove that the GARCH process, of the type considered so far, may not aggregate in a simple way.[16] For this reason, they introduce different categories of GARCH processes. The GARCH models studied so far will belong to the class of strong GARCH processes. At the other extreme, Drost and Nijman (1993), aware of the temporal aggregation results of Nijman and Palm (1990), realized that projections of squared returns on ARMA processes would lead to aggregation results. For this reason, they introduced so-called weak GARCH processes that are linear projections. The following definitions give a more precise sense.

Definition 4.1. *(Strong GARCH, Drost and Nijman, 1993) The series* ε_t*, with* $t \in \mathbb{Z}$*, and* $\varepsilon_t = \sigma_t z_t$ *where* $\sigma_t^2 = w + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$ *is a strong GARCH process if for some parameters* ω*,* α*, and* β*, the innovation* z_t *is iid with zero mean and unit variance.*

Definition 4.2. *(Semi-strong GARCH) The series* ε_t*, with* $t \in \mathbb{Z}$*, is a semi-strong GARCH process if* ω*,* α*, and* β *may be chosen so that*

$$E[\varepsilon_t|\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots] = 0, \text{ and}$$
$$E[\varepsilon_t^2|\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots] = \sigma_t^2,$$

with σ_t^2 *defined as for the strong GARCH process.*

Define now the linear projection operator $\mathrm{P}[\cdot]$ as the best linear predictor of some random variable given the information structure. Formally, we define $\mathrm{P}[x_t|\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots]$ the best linear predictor of x_t in terms of $(1, \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \cdots)'$. Then the following definition may be introduced.

Definition 4.3. *(Weak GARCH) The sequence* ε_t*, with* $t \in \mathbb{Z}$*, is a weak GARCH process if* ω*,* α*, and* β *may be chosen so that*

$$P[\varepsilon_t|\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots] = 0,$$
$$P[\varepsilon_t^2|\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots] = \sigma_t^2,$$

with σ_t^2 *defined as for the strong GARCH process.*

At this stage, Drost and Nijman (1993) assume that they observe a sample $\{\varepsilon_t\}_{t=1}^T$ of some asset returns. If P_t denotes the value of a given asset, then the continuously compounded return at time t is defined as $\varepsilon_t = \log(P_t/P_{t-1})$. This return is assumed to be sampled at, say, the daily frequency. The question is what parameters may be obtained if we had sampled at a different

[16] Alternatively, we may say that the class of GARCH processes is not closed under aggregation where the notion of closedness is taken in the mathematical sense.

frequency such as weekly. For instance, we may consider returns over a lower frequency consisting of the instants $m, 2m, \cdots, T$, where m is an integer and T is assumed to be a multiple of m. Drost and Nijman (1993) introduce the notation

$$\overline{\varepsilon}_{(m)t} \equiv \sum_{i=0}^{m-1} \varepsilon_{t-i} = \log\left(\frac{P_t}{P_{t-m}}\right), \qquad \text{for } t = m, 2m, \cdots, T.$$

Thus, $\overline{\varepsilon}_{(m)t}$ represents the return over a longer period than the initial returns series.[17] The main result of their paper is given by the following proposition.

Proposition 4.4. *(Drost and Nijman, 1993) If the sequence ε_t, $t \in \mathbb{Z}$, is a weak GARCH$(1,1)$ process with symmetric marginal distributions, $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$, and unconditional kurtosis κ_ε, then $\overline{\varepsilon}_{(m)tm}$, $t \in \mathbb{Z}$, is a symmetric weak GARCH$(1,1)$ process with*

$$\overline{\sigma}^2_{(m)tm} = \overline{w}_{(m)} + \overline{\alpha}_{(m)}\overline{\varepsilon}^2_{(m)tm-m} + \overline{\beta}_{(m)}\overline{\sigma}^2_{(m)tm-m},$$

and kurtosis $\overline{\kappa}_{(m)\varepsilon}$ where

$$\begin{aligned}
\overline{w}_{(m)} &= m\omega\frac{1-(\alpha+\beta)^m}{1-(\alpha+\beta)},\\
\overline{\alpha}_{(m)} &= (\alpha+\beta)^m - \overline{\beta}_{(m)},\\
\overline{\kappa}_{(m)\varepsilon} &= 3 + (\kappa_\varepsilon - 3)/m + 6(\kappa_\varepsilon - 1)\\
&\quad \times \frac{[m-1-m(\alpha+\beta)+(\alpha+\beta)^m]\,[\alpha-\alpha\beta(\alpha+\beta)]}{m^2(1-\beta-\alpha)^2(1-\beta^2-2\alpha\beta)},
\end{aligned}$$

and where $|\overline{\beta}_{(m)}| < 1$ is the solution to the quadratic equation

$$\frac{\overline{\beta}_{(m)}}{1+\overline{\beta}^2_{(m)}} = \frac{A(\alpha+\beta)^m - B}{A\{1+(\alpha+\beta)^{2m}\} - 2B},$$

where

$$\begin{aligned}
A &= m(1-\beta^2) + 2m(m-1)\frac{(1-\alpha-\beta)^2(1-\beta^2-2\alpha\beta)}{(\kappa_\varepsilon-1)\,[1-(\alpha+\beta)^2]}\\
&\quad + 4\frac{[m-1-m(\alpha+\beta)+(\alpha+\beta)^m]\,[\alpha-\alpha\beta(\alpha+\beta)]}{1-(\alpha+\beta)^2},
\end{aligned}$$

and

$$B = \alpha - \alpha\beta(\alpha+\beta)\frac{1-(\alpha+\beta)^{2m}}{1-(\alpha+\beta)^2}.$$

[17] Drost and Nijman (1993) introduce a distinction between flow and stock variables. For finance applications, involving asset returns, the aggregation of flows is of relevance.

Given that this proposition requires the kurtosis of the original error, κ_ε, for empirical applications the following reminder by Drost and Nijman (1993) may be useful

$$\kappa_\varepsilon = \kappa_z \frac{1-(\alpha+\beta)^2}{1-(\alpha+\beta)^2-(\kappa_z-1)\alpha^2},$$

where z_t denotes the standardized innovation. Thus, if we estimate the kurtosis of the innovations z_t, after having fitted the GARCH model, we may infer κ_ε once α and β have been estimated. Inspection of the formula for kurtosis reveals that as the aggregation step m increases, $\overline{\kappa}_{(m)\varepsilon}$ converges to 3. Thus, temporal aggregation will render the data more normal. We may intuitively view this result as resulting from the smoothing of observations as we sum the innovations. Heteroskedasticity will get smoothed. This result has been noticed already by Diebold (1988) and has been verified in Chapter 2.

An empirical illustration

To illustrate how temporal aggregation changes the parameters, we consider a situation where $(\alpha, \beta) = (0.1, 0.85)$ and $(\alpha, \beta) = (0.05, 0.94)$ as well as the cases where the kurtosis equals $\kappa_z = 3$, corresponding to the Gaussian case, and $\kappa_z = 8.75$, corresponding to the case where innovations are distributed as a Student t with 5 degrees of freedom. As in Drost and Nijman (1993), we consider various values of $m = 1, 2, 4, 5, 16, 32, 64$. The corresponding values of $(\overline{\beta}_{(m)}, \overline{\alpha}_{(m)})$ will be represented in Figure 4.8 by symbols, joined for graphical convenience by lines. The case $(\alpha, \beta) = (0.1, 0.85)$ corresponds to the upper curves represented by continuous lines. The case $(\alpha, \beta) = (0.05, 0.94)$ is represented by the dashed lower set of curves.

Among each curve, squares correspond to those pairs of the α and β parameters generated under a Gaussian scenario. Bullets are obtained under the assumption that the innovations follow a Student t with 5 degrees of freedom. We, thus, observe that the distribution and fat-tailedness of the innovations affects the properties of aggregation. Depending on the parameters α and β as well as on the kurtosis, various patterns how the parameters evolve as m increases are possible. Interestingly, therefore, the kurtosis of the innovation z_t affects the temporal aggregation properties of the parameters. In other words, in practical applications, if the kurtosis is not known, it is not possible to correctly infer how the parameters will aggregate.

We may also ask how aggregated parameters compare with those obtained for a GARCH model estimated on aggregated data. To answer this question, we estimate a symmetric GARCH$(1, 1)$ model to the SP500 returns sampled at the daily and weekly frequencies assuming normality of innovations. The sampling at weekly frequency corresponds to $m = 5$. The results are represented in Table 4.5. The table summarizes the results obtained. We notice that the kurtosis of innovations is very high in both cases. As we switch frequency, the parameters evolve in the right direction.

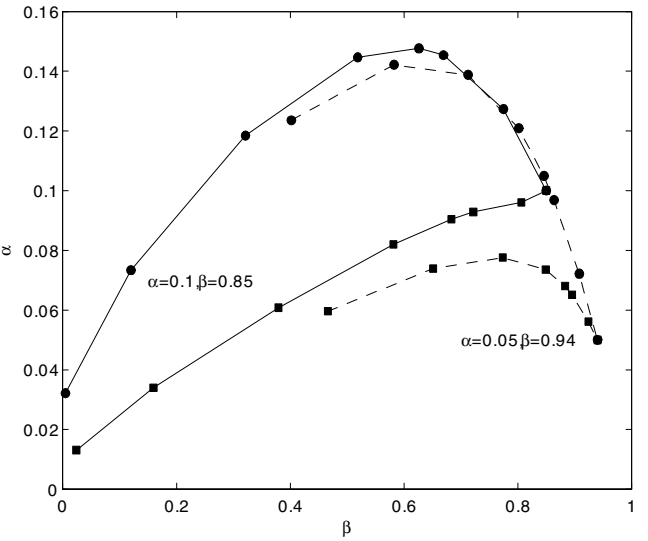

Fig. 4.8. *Evolution of the parameters of a weak GARCH model under temporal aggregation. The symbols correspond (from right to left) to sampling every 1, 2, 4, 5, 8, 16, 32, 64 observations. Curves with squares correspond to Gaussian innovations and bullets to Student t distributed innovations*

Table 4.5. *GARCH parameters for the SP500 at both weekly and daily frequencies*

Parameter	Estimate	Std error
	Weekly frequency	
ω	0.2834	(0.0945)
α	0.1826	(0.0271)
β	0.7862	(0.0341)
Kurtosis, κ_z	8.21	–
	Daily frequency	
w	0.0197	(0.0045)
α	0.0743	(0.0069)
β	0.9057	(0.0109)
Kurtosis, κ_z	8.71	–

To investigate how correct this switch is, we trace in Figure 4.9 the parameters $\overline{\alpha}_{(m)}$ and $\overline{\beta}_{(m)}$ obtained under temporal aggregation when we start with higher frequency data (here daily) $(\alpha, \beta) = (0.0743, 0.9057)$. The curve with the stars represents a curve corresponding to aggregated parameters for the same sequence of m as above. The parameters obtained for the GARCH model, estimated at weekly frequency, are represented by the single large diamond in the upper part of the figure. As we can see, there is a significant

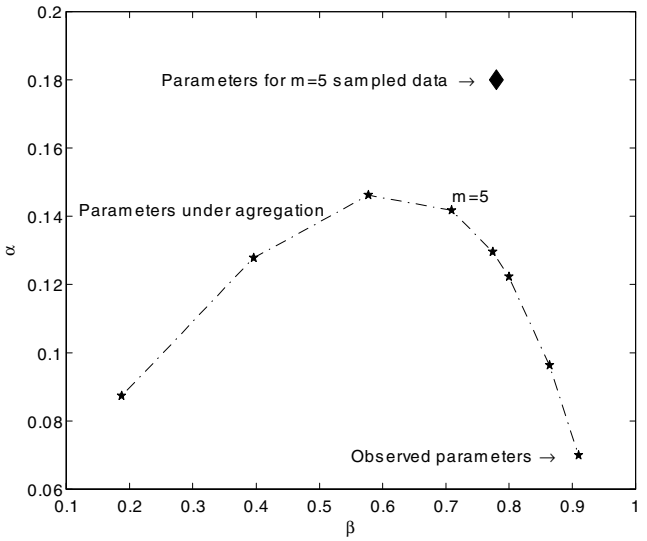

Fig. 4.9. *Evolution of the parameters of a weak GARCH model under temporal aggregation.*

distance between the actual parameters and those parameters that should hold under aggregation. This brings about the question of the correctness of the model.

In an interesting contribution, Drost and Werker (1996) consider a continuous time model with jumps that they discretize so as to obtain a weak GARCH representation. Estimation of this model allows these authors to provide some inference about the frequency of jumps in the data. If we contemplate the model used so far, we realize that several features that are present in the data have not been treated so far. We may mention the necessity that the fourth moment be finite (and as we will see in Chapter 7, this is not necessarily the case). The next issue is that the results are obtained for so-called weak GARCH models. These models are, however, not consistently estimated as demonstrated in a simulation exercise by Meddahi and Renault (2004). Also, these models do not allow for leverage effect. Meddahi and Renault (2004) circumvent these difficulties by building on a model in the spirit of stochastic volatility.

4.6.2 Cross-sectional aggregation

In this section, we consider the important issue of how GARCH processes aggregate across assets, an issue initially addressed by Nijman and Sentana

(1996). These authors show that the contemporaneous aggregation of independent univariate strong GARCH processes yields a weak GARCH process as defined in the previous section. This result may be seen as follows: Assume that $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are both generated by strong GARCH models such as

$$\varepsilon_{i,t} = \sigma_{i,t} z_{i,t}, \quad i = 1, 2,$$

where the $z_{i,t}$ are *iid* with mean 0, variance 1, and kurtosis κ_i, and that volatility is generated by the symmetric GARCH$(1, 1)$ model

$$\sigma^2_{i,t} = \omega_i + \alpha_i \varepsilon^2_{i,t-1} + \beta_i \sigma^2_{i,t-1}.$$

Then Nijman and Sentana (1996) follow Bollerslev (1988) and introduce the forecast volatility error $\eta_{i,t} = \varepsilon^2_{i,t} - \sigma^2_{i,t} = \varepsilon^2_{i,t} - E[\varepsilon^2_{i,t}|\mathcal{F}_{t-1}]$, where $\mathcal{F}_{t-1}$ represents the past information, to obtain an ARMA$(1, 1)$ representation for $\varepsilon^2_{i,t}$, such as

$$\varepsilon^2_{i,t} = \sigma^2_i + [1 - (\alpha_i + \beta_i)L]^{-1}[1 - \beta_i L]\eta_{i,t},$$

where L is the lag operator, defined by $L\eta_{i,t} = \eta_{i,t-1}$, and where $\sigma^2_i = E[\varepsilon^2_{i,t}] = \omega_i/(1 - \alpha_i - \beta_i)$. Taking the sum of the two processes yields

$$\begin{aligned}(\varepsilon_{1,t} + \varepsilon_{2,t})^2 = \sigma^2_1 + \sigma^2_2 &+ [1 - (\alpha_1 + \beta_1)L]^{-1}[1 - \beta_1 L]\eta_{1,t} \\ &+ [1 - (\alpha_2 + \beta_2)L]^{-1}[1 - \beta_2 L]\eta_{2,t} + 2\varepsilon_{1,t}\varepsilon_{2,t}.\end{aligned}$$

Given that $\eta_{1,t}$, $\eta_{2,t}$ and the product $\varepsilon_{1,t}\varepsilon_{2,t}$ are uncorrelated, and all series are non autocorrelated, Nijman and Sentana (1996) conclude that the sum of two strong GARCH$(1, 1)$ processes yields a weak GARCH$(2, 2)$ process using the results from time series analysis according to which a sum of two ARMA$(1, 1)$ processes yields an ARMA$(2, 2)$ process. A remaining difficulty is of course to obtain a parsimonious independent representation of this process. The reason for this difficulty comes from the cross-product term $\varepsilon_{1,t}\varepsilon_{2,t}$. Furthermore, Nijman and Sentana (1996) show that, if $\alpha_1 + \beta_1 = \alpha_2 + \beta_2 = \gamma$, then the resulting process will be a weak GARCH$(1, 1)$ process.

4.6.3 Estimation of the weak GARCH process

Clearly, the parametric specifications of a strong and a weak GARCH process are not the same. This raises as first question what would be the error if we estimated a strong GARCH process and we directly used the resulting parameters in a weak GARCH process. Drost and Nijman (1993) report that they performed simulations where the true data generating process was a strong GARCH process and where they estimate the parameters using a QML estimation. For parameters similar to the dynamic of exchange rates, they report that the results obtained in the QML estimation are close to the true parameters.

On the other hand, Nijman and Sentana (1996) construct weak GARCH processes as the sum of strong GARCH processes and by performing QML

estimation. They report a small, yet systematic bias in one of the parameters of the GARCH model.

These observations have led Francq and Zakoïan (2000) to investigate the properties of QML estimation of weak GARCH processes. For the setting of Drost and Nijman (1993), they confirm that the errors are small. For other settings and specifications, they show that the bias may be large. This in turn requires an estimation procedure for weak GARCH processes, because QML estimation appears to be troubled. The reason for this failure is that $\eta_{i,t}$ may not satisfy the properties of a martingale difference sequence, on which the QML theory in a time series context builds. Francq and Zakoïan (2000) further propose a two-stage least-squares estimation building on the observations, ε_t, and the weak GARCH representation yielding to an estimation such as

$$\varepsilon_t^2 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 = \omega + u_t + \sum_{i=1}^{q} \beta_i u_{t-i},$$

where u_t is now some white noise, with a constant variance.

To conclude this section, we need to mention that Meddahi and Renault (2004) propose an estimation procedure based on a state-space representation to avoid biasedness of the estimates.

4.7 Stochastic volatility

4.7.1 From GARCH models to stochastic volatility models

So far we have discussed many models in the spirit of GARCH models. Typically, for a simple model in this class, the centred returns may be written

$$\varepsilon_t = \sigma_t z_t, \tag{4.31}$$

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2. \tag{4.32}$$

Inspection of these equations reveals that there is only one source of uncertainty, z_t, with $E[z_t] = 0$, and $V[z_t] = 1$. This way of modeling returns is therefore very simple, which may explain, at least partially, its tremendous success. The drawback is that this model does not allow for a specific error in the dynamics of volatility. For given parameters, if we imagine a large return, in absolute value, then the volatility equation (4.32) will imply a subsequent large volatility. In general, this is certainly a desirable feature. However, the impact of a given large return will not be the same all the time. This implies that the σ_t that this model predicts, in comparison with the true unknown volatility, will sometimes be too large and other times too small.

This observation has led Taylor (1982, 1986) to consider the first discrete time stochastic volatility (SV) model in which volatility has a specific source of randomness. This model may be expressed as follows

$$\varepsilon_t = \sigma_t z_t = z_t \exp\left(\frac{1}{2} h_t\right), \tag{4.33}$$

$$h_t = \omega + \beta h_{t-1} + v_t, \tag{4.34}$$

where v_t is an *iid* $\mathcal{N}\left(0, \sigma_v^2\right)$ process. The two processes z_t and v_t are usually assumed to be independent. Adding an error term v_t in the dynamics of volatility introduces another source of randomness in the model that may improve the description of the actual volatility. Obviously, the true volatility is never observed and, thus, h_t is a latent process that needs to get estimated. The probable improvement in terms of volatility comes at the price of higher complexity at the estimation level.

Since (4.34) represents an AR(1) process, stationarity conditions and moments are well-known. Indeed, stationarity of the h_t process imposes $|\beta| < 1$. Then, we have

$$E\left[h_t\right] = \mu_h = \frac{\omega}{1-\beta}, \qquad \text{and} \qquad V\left[h_t\right] = \sigma_h^2 = \frac{\sigma_v^2}{1-\beta^2}.$$

Therefore, since z_t is assumed to be *iid* $\mathcal{N}(0,1)$ and uncorrelated with ν_t, we can compute the moments of $\varepsilon_t = z_t \exp\left(\frac{1}{2} h_t\right)$ (see also Taylor 1986, Chapter 3),[18]

$$\begin{aligned}
E\left[\varepsilon_t\right] &= E\left[z_t\right] E\left[\exp\left(\frac{1}{2} h_t\right)\right] = 0, \\
E\left[\varepsilon_t^2\right] &= E\left[z_t^2\right] E\left[\exp\left(h_t\right)\right] = \exp\left(\mu_h + \frac{\sigma_h^2}{2}\right), \\
E\left[\varepsilon_t^3\right] &= 0, \\
E\left[\varepsilon_t^4\right] &= E\left[z_t^4\right] E\left[\exp\left(2h_t\right)\right] = 3\exp\left(2\mu_h + \sigma_h^2\right).
\end{aligned}$$

It follows that all odd moments are zero, which results from the assumption of independence between z_t and ν_t and the symmetry of z_t. Kurtosis is equal to

$$\frac{E\left[\varepsilon_t^4\right]}{\left(E\left[\varepsilon_t^2\right]\right)^2} = \frac{3\exp\left(2\mu_h + \sigma_h^2\right)}{\left(\exp\left(\mu_h + \frac{\sigma_h^2}{2}\right)\right)^2} = 3\exp\left(\sigma_h^2\right) > 3.$$

This expression indicates that the kurtosis can become arbitrarily large.

In addition, taking the square and the logarithm of (4.33) and using (4.34) yields

$$\log\left(\varepsilon_t^2\right) = \log\left(\sigma_t^2\right) + \log\left(z_t^2\right) = h_t + \log\left(z_t^2\right).$$

Since the first component, h_t is an AR(1) process and $\log(z_t^2)$ is a white noise, the dynamics of the log of ε_t^2 will be an ARMA$(1,1)$ process. As Harvey, Ruiz, and Shephard (1994) indicate, extensions of the dynamics of h_t to an

[18] We remind that if a random variable X is distributed as a $\mathcal{N}(\mu, \sigma^2)$, then we have the rth moment $E[(e^X)^r] = E[e^{rX}] = e^{(r\mu + r\sigma^2/2)}$.

ARIMA dynamics are straightforward. Some of the implications of changing the dynamics of h_t are discussed in that paper.

The SV model is much more flexible than the GARCH models. It has been found to fit asset returns better and have residuals closer to standard normal. In the SV (and GARCH) models, the distribution of returns has immediately fat tails. Persistence in volatility is captured by autoregressive terms β_j. Correlation between z_t and v_t produces volatility asymmetry. The SV model is therefore able to capture most statistical features of return volatility.

4.7.2 Estimation of the discrete time SV model

However, although the error term v_t makes the SV model much more flexible, it also implies that the SV model cannot be estimated directly by ML. Indeed, the process h_t is a latent variable. The likelihood function of the SV model is

$$f(\varepsilon_1, \cdots, \varepsilon_T|\theta) = f(\varepsilon_T|\mathcal{F}_{T-1}) \times f(\varepsilon_{T-1}|\mathcal{F}_{T-2}) \times \cdots \times f(\varepsilon_1|\mathcal{F}_0),$$

where $\theta = \left(\omega, \beta, \sigma_v^2\right)'$ denotes the vector of unknown parameters. The difficulty is that the expression $f(\varepsilon_t|\mathcal{F}_{t-1})$ depends on an unobservable variable σ_t. To go further, we introduce the notation $\varepsilon_{\underline{t}} = \{\varepsilon_1, \cdots, \varepsilon_t\}$. This allows us to express the joint likelihood of $\varepsilon_{\underline{T}}$ and $\sigma_{\underline{T}}$. Since the distribution of ε_t is known to be normal with mean 0 and variance σ_t^2, conditionally on a given path of volatility, the joint density can be written as

$$f\left(\varepsilon_{\underline{T}}, \sigma_{\underline{T}}\right) = f\left(\varepsilon_{\underline{T}}|\sigma_{\underline{T}}\right) f\left(\sigma_{\underline{T}}\right),$$

where $f\left(\varepsilon_{\underline{T}}|\sigma_{\underline{T}}\right) = \prod_{t=1}^{T} f(\varepsilon_t|\sigma_t)$ and $f\left(\sigma_{\underline{T}}\right)$ is some multivariate distribution. Therefore, to compute the likelihood $f\left(\varepsilon_{\underline{T}}|\theta\right)$, we may use the expression above, because the marginal is obtained from the joint density by integrating out the variable we want to get rid of, so that,

$$L_T\left(\theta|\varepsilon_{\underline{T}}\right) = f\left(\varepsilon_{\underline{T}}|\theta\right) = \int\int\cdots\int f\left(\varepsilon_{\underline{T}}|\sigma_{\underline{T}}\right) f\left(\sigma_{\underline{T}}|\theta\right) d\sigma_1 d\sigma_2 \cdots d\sigma_T.$$

Since this expression does not require that volatility is observable anymore, it would be possible in principle to evaluate and therefore maximize the likelihood function L_T. However, evaluating L_T using this expression would require the numerical integration of a T-dimensional integral, which is impossible in practice. Several alternative estimation procedures have therefore been proposed to simplify this problem.

A quasi-maximum likelihood estimation (QMLE) has been proposed by Harvey, Ruiz, and Shephard (1994). First, from Abramowitz and Stegun (1970, p. 943), it is known that $E[\log\left(z_t^2\right)] = -1.27$ and $V[\log\left(z_t^2\right)] = 4.93$. Introducing the notation $\xi_t \equiv \log(z_t^2) + 1.27$, which yields a mean zero innovation, and squaring the returns, we can rewrite the system (4.33)–(4.34) as

$$\log\left(\varepsilon_t^2\right) = -1.27 + h_t + \xi_t,$$
$$h_t = \omega + \beta h_{t-1} + \nu_t.$$

It is possible to apply the Kalman filter to this system. Given that the Kalman filter assumes Gaussian innovations, which is clearly not the case for ξ_t, it is not possible to obtain an exact likelihood this way. Ruiz (1994) performs experiments that show that the estimates obtained via the Kalman filter have to be preferred over method-of-moments estimates.

A numerical difficulty is to deal with zero returns, i.e., when $\varepsilon_t = 0$ or cases where returns are close to zero. Breidt and Carriquiry (1996) suggest the use of the transformation

$$\log\left(\varepsilon_t^2\right) \approx \log(\varepsilon_t^2 + c \cdot \hat{\sigma}^2) - c \cdot \hat{\sigma}^2/(\varepsilon_t^2 + c \cdot \hat{\sigma}^2),$$

where $\hat{\sigma}^2$ stands for the sample variance of the ε_t and c is a small number such as 0.02. The impact of those observations equal or close to zero will therefore be diminished. Ghysels, Harvey, and Renault (1996) suggest the use of some conditional volatility estimate instead of the constant $\hat{\sigma}^2$.

Alternative estimation procedures, based on simulation methods have been proposed. The first one is *indirect inference* (Gouriéroux, Monfort, and Renault, 1996). The second one is the *Markov Chain Monte Carlo* (Jacquier, Polson, and Rossi, 1994, see also Kim, Shephard, and Chib, 1998, or Eraker, Johannes, and Polson, 2003). Both methods are very computationally intensive even in a univariate setting. Other estimation methods are based on *simulated method of moments* (Duffie and Singleton, 1993), *efficient method of moments* (Gallant and Tauchen, 1996), or *simulated maximum likelihood* (Brandt and Santa-Clara, 2002, Durham and Gallant, 2002). Recently, there has been an interest in *empirical characteristic function methods* (Singleton, 2001; Chacko and Viceira, 2003, Jiang and Knight, 2002, Rockinger and Semenova, 2005). None of these methods is truly simple to implement. Andersen, Benzoni, and Lund (2002) compare some of these estimation techniques. Multivariate extensions appear to be very difficult to estimate. The stochastic volatility model is therefore a very important building block for empirical finance. However, because of the difficulty of estimation, it may not have attracted as much attention as it deserves.

4.8 Realized volatility

Over the recent years, it has become possible to use data at a tick-by-tick level. Such data is also called *high-frequency data*. In parallel with the availability of this data, a new research strand emerged. This strand builds on results of stochastic integration.[19] The following is mainly based on Barndorff-Nielsen and Shephard (2004b, thereafter BN/S).

[19] A set of worthwile readings in the stochastic integration area is Dellacherie and Meyer (1982), Jacod and Shiryaev (2002), Protter (2004), Rogers and Williams

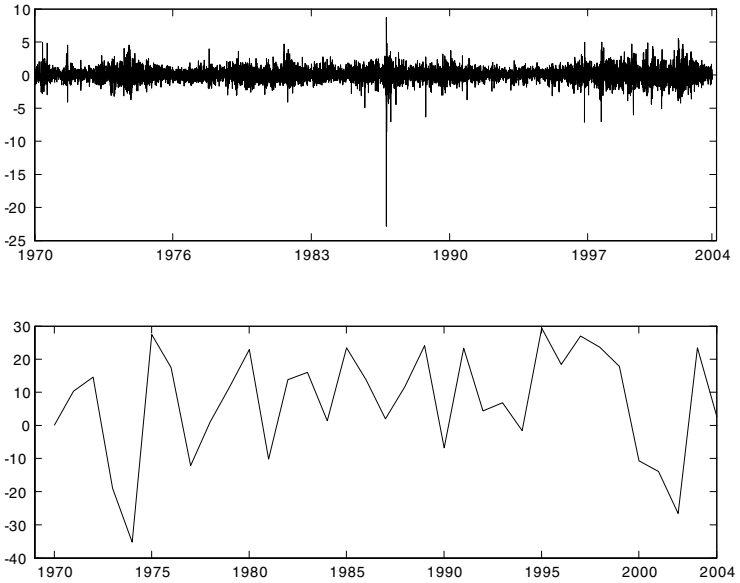

Fig. 4.10. *Evolution of SP500 returns at daily and annual frequencies.*

4.8.1 The difficulty to disentangle jumps

So far, probability theory tells us that in the continuous time setting, a reasonable model for asset prices is the semi-martingale. In actual empirical work, we observe, however, only discretely sampled data. This issue brings up the problem of distinguishing between the stochastic, yet continuous, part and the jumps. A recent contribution by Aït-Sahalia (2004) shows that, intuitively, data sampled at a higher and higher frequency should indeed provide information on the actual components. There are several difficulties with discrete sampling that render the detection of jumps difficult.

Figure 4.10 displays realizations of the SP500 at different frequencies.[20] As this figure illustrates, the shift from the daily to a yearly frequency renders the detection of the jump impossible. The crash of 1987 disappears totally at the yearly frequency. To the opposite, years without a crash, such as 2001 and 2002, saw a much larger drop in returns than 1987.

We may, therefore, reasonably ask from what timescale on does it become possible to visualize jumps in the data. Again, Aït-Sahalia (2004) provides some elements. He considers a simple model without temporal dependency of volatility in continuous time composed by a continuous part, provided by a

(2000), or the excellent recent book by Cont and Tankov (2004). Some of the following definitions may also be found in Andersen et al. (2003).

[20] Aït-Sahalia (2004) presents a similar plot with the Dow Jones Industrial Index for which the 1987 crash is even more pronounced.

Brownian motion, and discrete finite large jumps. Letting P_t denote the price at time t, the model that he considers assumes that the return over a given time interval $[t, t+\Delta]$ behaves as follows

$$r_{t,t+\Delta} = 100 \times \log(P_{t+\Delta}/P_t) = \mu\Delta + \sigma\sqrt{\Delta}\varepsilon_t + \sum_{s=t}^{t+\Delta} J_s N_s,$$

where ε_t is a normal $\mathcal{N}(0,1)$ realization, N_t is a Poisson process, increasing by 1 whenever a jump takes place. When a jump takes place, it does so over a time interval of length Δ with an intensity of $\lambda\Delta$. The distribution of the jump size J is $\mathcal{N}(\beta, \eta^2)$. All sources of uncertainty, ε_t, N_t, and J_t, are assumed independent. The probability of having n jumps between t and $t+\Delta$ is

$$\Pr[N_{t+\Delta} - N_t = n] = \exp(-\lambda\Delta)\frac{(\lambda\Delta)^n}{n!}.$$

In this model, returns have, therefore, a simple continuous part, provided by the discretized Brownian motion, $\sigma\sqrt{\Delta}\varepsilon_t$, and the jump part. The question is then what the probability of observing a jump is, given that the return $r_{t,t+\Delta}$ exceeds a certain threshold, say r. Using Bayes' rule, we obtain that the probability that there is exactly one jump during the considered time interval is

$$\Pr[N_{t+\Delta} - N_t = 1 | r_{t,t+\Delta} > r] = \frac{e^{-\lambda\Delta}\lambda\Delta\left(1 - \Phi\left(\frac{r-\mu\Delta-\beta}{2(\eta^2+\Delta\sigma^2)^{1/2}}\right)\right)}{\Pr\left[r_{t,t+\Delta} > r\right]},$$

where

$$\Pr\left[r_{t,t+\Delta} > r\right] = \sum_{n=0}^{\infty} e^{-\lambda\Delta}\frac{(\lambda\Delta)^n}{n!}\left(1 - \Phi\left(\frac{r-\mu\Delta-\beta}{2(n\eta^2+\Delta\sigma^2)^{1/2}}\right)\right).$$

The probability of seeing more than one jump is

$$\Pr[N_{t+\Delta} - N_t \geq 1 | r_{t,t+\Delta} > r] = \frac{\sum_{n=1}^{\infty} e^{-\lambda\Delta}\frac{(\lambda\Delta)^n}{n!}\left(1 - \Phi\left(\frac{r-\mu\Delta-\beta}{2(n\eta^2+\Delta\sigma^2)^{1/2}}\right)\right)}{\Pr\left[r_{t,t+\Delta} > r\right]}.$$

and obviously, the probability that no jump occurs is the complement of this probability. Using similar computations, we may show that the probability that two jumps occur in the time interval $[t, t+\Delta]$ is

$$\Pr[N_{t+\Delta} - N_t = 2 | r_{t,t+\Delta} > r] = \frac{e^{-\lambda\Delta}\frac{(\lambda\Delta)^2}{2!}\left(1 - \Phi\left(\frac{r-\mu\Delta-\beta}{2(2\eta^2+\Delta\sigma^2)^{1/2}}\right)\right)}{\Pr\left[r_{t,t+\Delta} > r\right]}.$$

Aït-Sahalia calibrates the model with the values $\mu = \beta = 0$, $\sigma = 0.3$, $\lambda = 0.2$, $\eta = 0.6, \Delta = 1/12$. These parameters correspond to data that is calibrated annually, but for which we consider one single month as a time frame. In

order to compare returns, it is useful to standardize them. Indeed, a return that appears as a crash for a series with a small volatility may appear to be a completely normal return for a series with higher volatility. To render returns comparable, it is useful to standardize with respect to the unconditional volatility of the series. To do so, we consider u such that the threshold is given by $r = u(\Delta(\sigma^2 + (\beta^2 + \eta^2)\lambda))^{1/2}$. This parameter u will then have the units of standard deviation. Thus a return with $u = 1$ will have a size that corresponds to one unconditional standard deviation of the series.

With this choice of parameter values and scaling of the possible threshold values, Aït-Sahalia obtains the probabilities of jumps as represented in Figure 4.11, where the continuous line represents the probability of no jump, the dotted line the probability of exactly one jump, and eventually the line with dots and dashes measures the probability of having two jumps in a given month provided that we observe one return exceeding the threshold r, this corresponding to a certain standard deviation level presented along the horizontal axis.

As this figure displays, if we have a threshold of about one standard deviation, the observation of a return higher than this threshold is not very much informative about it being a jump. This probability is only about $1 - 0.98 = 2\%$. It takes relatively high values before we are able to state that a jump really occurs. Still, if we observe a return exceeding a three standard deviation threshold, the probability of this being explained by the continuous part of the process is 80%. Only at a level of about 3.5, it becomes impossible to distinguish if the return is due to a jump or to the continuous part. We may therefore conclude that the detection of jumps is a difficult issue because returns with more than 3.5 standard deviations are not very likely to take place.

The theoretical value of using data sampled at higher frequencies is demonstrated by Aït-Sahalia in Figure 4.12. This figure traces the probability that a jump occurred if we observe a return of 10% over a given time horizon. We notice that the probability of detecting a jump is very low when considering a weekly frequency. In other words, if we believe that the parameter values correspond to actual market activity, it will be nearly impossible to state if a crash happened just by eyeballing weekly returns. When considering shorter and shorter time horizons, the probability to detect jumps gets higher and higher as intuition suggests. At the daily frequency, we would think that, with 30% probability, a jump took place. It is, however, only when considering data sampled at the hourly frequency that we can actually detect a jump.

With these caveats that justify the use of high-frequency data, we now turn to tools that allow for a measurement of jumps as the data gets sampled at higher and higher frequencies. The discussion of the deep nature of jumps is not trivial. The literature on Lévy process shows that jumps can also come in the variety of *infinite activity*. This class of jumps does not appear to be detectable with the tools that follow. It appears that further research needs to be done to uncover techniques to measure such processes.

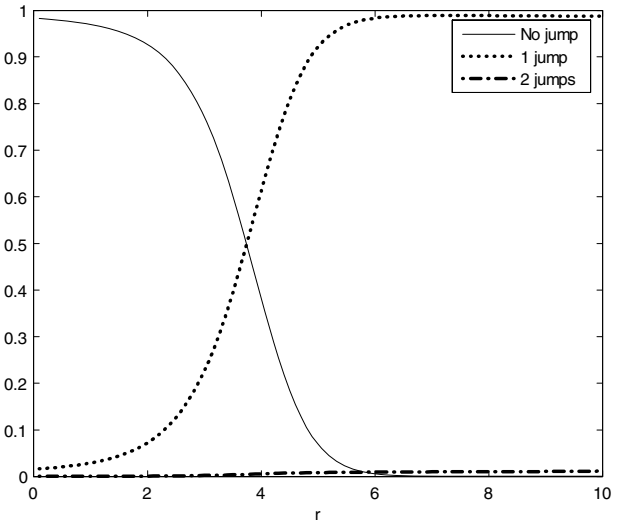

Fig. 4.11. *Probability of jump, depending on the threshold r.*

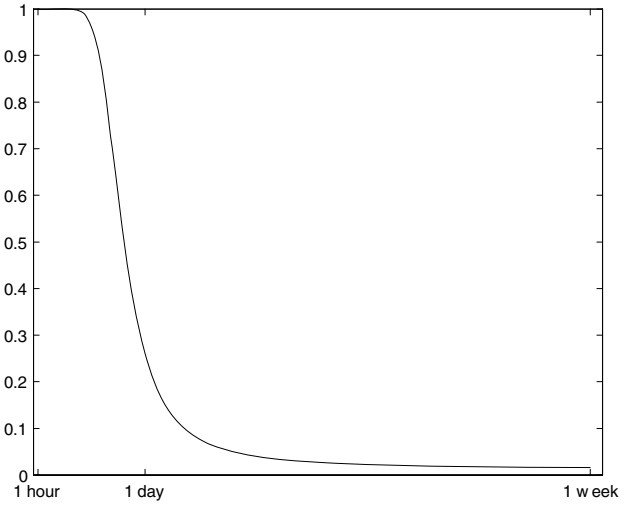

Fig. 4.12. *Probability that a jump occurred if we observe a return of 10%, depending on the time horizon.*

4.8.2 Quadratic variation

Consider a time horizon $[0, t]$. For this section, we will consider a partition of this interval with constant time increment, i.e., $\Delta = t/N$.[21] Many of the results reported below have been generalized to a non-constant-step partition. In practice, with our interpretation of constant time-step, it means that data should be sampled, say, every 5, 10, or 30 minutes. Those results obtained for a non-constant time increment would hold for the raw data, that is with their original time stamp.

Let p_t denote the log-price of a given asset. The concept of quadratic variation is specified in the following definition.

Definition 4.5. *Quadratic variation of a process p_t is given as*

$$[p]_t^{[2]} \equiv \underset{\Delta \to 0}{plim} \sum_{j=1}^{\lfloor t/\Delta \rfloor} \left(p_{j\Delta} - p_{(j-1)\Delta} \right)^2,$$

where $\lfloor x \rfloor$ denotes the function returning the largest integer less than or equal to x.

The resulting process is sometimes called the *bracket process*. This expression defines the quadratic variation process as the limit case of a cumulative sum of squared returns when the sampling sequence becomes infinitely fine. The question that emerges is, what this process actually represents, to which entity does this quantity converge? A key result of probability theory states that the quadratic variation captures the quadratic variation of the local continuous martingale, as well as the measure of squared jumps that are of finite size. We may therefore write

$$[p]_t^{[2]} = [p^C]_t^{[2]} + \sum_{s \leq t} (\Delta p_s)^2,$$

where by definition $\Delta p_s = \lim_{u \to s, u < s} p_s - p_u$, and where $[p^C]_t^{[2]}$ captures the quadratic variation of the continuous part of the process. Thus, the notation above $\sum_{s \leq t} (\Delta p_s)^2$ designates the squared sum of jumps that occur in the price process. Clearly, even if the quadratic variation was known, i.e., by estimating it from the actual data, it still would not allow for disentangling the continuous and the jump parts.[22]

[21] Actually, in most of the following, we will consider limit cases where Δ will converge to zero. This means that the interval $[0, t]$ will be divided into some N intervals, with N becoming very large. We assume that the difference between $N \cdot \Delta$ and t is negligible.

[22] If we use explicit models such as GARCH models combined with jump components, then it is possible to disentangle the continuous and jump parts. The issue when doing this is that strong assumptions on the stochastic volatility model and the type of jumps are made.

4.8.3 Power variation

At this stage, we have determined that the quadratic variation of the process contains the quadratic variation of the continuous part plus the measure of the jumps in the price process. A similar result can be obtained by using the concept of power variation, to which we turn now.

Definition 4.6. *The rth power variation is defined as*

$$[p]_t^{[r]} \equiv \underset{\Delta \to 0}{plim} \Delta^{1-r/2} \sum_{j=1}^{\lfloor t/\Delta \rfloor} |p_{j\Delta} - p_{(j-1)\Delta}|^r.$$

BN/S remark that the scaling factor Δ is crucial to get results that are comparable across series. The parameter r is called the power with which the variation is measured. For $r = 2$, power variation is the same as quadratic variation. Obviously, in that case, the scaling factor disappears. For $r \in (0, 2)$ the scaling factor converges to 0. For $r > 2$ it explodes as Δ converges to 0.

There remains the question what the power variation captures. To obtain further results, BN/S focus in their study of the power variation on semi-martingales with stochastic volatility. Before describing this class, let us define what semi-martingales are.

In the following, we consider, in general, time as belonging to the set $[0, \infty]$. It is assumed that there exists a filtration, $\{\mathcal{F}_t, t \geq 0\}$, describing the information available through time.

Definition 4.7. *A process M_t is a local martingale if M_0 is $\mathcal{F}_0$ measurable and if there exists an increasing sequence of stopping times τ_n with $\lim_{n\to\infty} \tau_n = \infty$ such that each process*

$$\{M_{\tau_i \wedge t} - M_0 = t \geq 0\}$$

is a martingale.[23]

This definition states that a process is a local martingale if there exists a sequence of moments, such that for each moment, the best forecast of the future values of the process is the current value. For instance, a Brownian motion is a martingale and a local martingale. To see this, it is sufficient to take $\tau_i = i$, for $i \in \mathbb{N}$. If the time horizon is finite, $[0, T]$, then it is relatively safe to say that most local martingales encountered in finance will also be martingales.[24]

[23] The symbols $\vee$ and $\wedge$ stand for the supremum and the infimum operators.

[24] Steve Shreve kindly communicated to us a discrete-time example of a process that is a local martingale but not a martingale. Inspection of this example shows that counter-examples are not easy to construct.

Definition 4.8. *Let X_t, $t \geq 0$ be a stochastic process. X_t is a semi-martingale if it can be written as*

$$X_t = X_0 + M_t + A_t,$$

where X_0 is known at time $t = 0$, where M_t is a local martingale, and where A_t is a process of finite variation, meaning that for any interval of time $[0, t]$, and for whatever partition $t_0 = 0 < t_1 < \cdots < t_N = t$, whatever N,

$$\sup \sum_{j=0}^{N} \left| A_{t_j} - A_{t_{j-1}} \right| < \infty.$$

By definition, we set $M_0 = A_0 = 0$.

As shown in Andersen et al. (2003), there is a no-arbitrage condition that the A_t and M_t components must satisfy. This condition states that if there exists a predictable jump in the asset price, there must be an offsetting variation in the martingale part. In our setting, the previous decomposition is the most useful to uncover, using recent convergence results, the continuous part and the jump part of a semi-martingale. A martingale is a stochastic volatility process if it can be written as

$$m_t = \int_0^t \sigma_u dW_u, \tag{4.35}$$

where σ_t is called the instantaneous volatility process and where W_t is a Brownian motion. It is assumed that σ_t is a càd-làg process, bounded away from zero. Since we can write, over a short interval of time, that $dm_t = \sigma_t dW_t$, the instantaneous variability is $(dm_t)^2 = \sigma_t^2 dt$. Over a longer time period, the variation will be defined as an integral

$$\sigma_t^{2*} = \int_0^t \sigma_u^2 du,$$

assumed to be finite, and which will be called the *integrated variance process*. Clearly, the integrated variance process measures the cumulative activity of a process where for the standard Brownian motion $\sigma_t^{2*} = t$. If the process σ_t is known, and this would be the case in a simulation exercise, then the integrated variance process may be computed with a numerical integration. A stochastic volatility process combined with a predictable component, a_t, will yield the $\mathcal{SVSM}$ (for Stochastic Volatility Semi-Martingale) class. Processes in the class with continuous paths will be denoted by $\mathcal{SVSM}^C$.

BN/S prove the following property: If p belongs to the class of $\mathcal{SVSM}^C$ and (σ, a) are independent of W, then

$$[p]_t^{[r]} = \mu_r \int_0^t \sigma_s^r ds,$$

where, in this expression, $\mu_r = 2^{r/2}\Gamma(\frac{1}{2}(r+1))/\Gamma(\frac{1}{2})$. Notice that μ_r corresponds to the rth moment of a zero mean, unit variance normal process, i.e., $\mu_r = E[|u|^r]$, if $u \sim \mathcal{N}(0,1)$. The following result demonstrates the importance of power variation. It shows that jumps will vanish if we compute power variations.

Proposition 4.9. *If $p = p^1 + p^2$ where $p^1 \in \mathcal{SVSM}^C$, i.e., $p^1_t = a_t + m_t$, with a_t the predictible part and m_t the martingale part as in (4.35), and p^2 is a process with a finite number of jumps over finite time increments, where the processes (a_t, σ_t) are independent of W_t, as well as p^1 and p^2 are independent, then*

$$[p]_t^{[r]} = \mu_r \int_0^t \sigma_s^r ds,$$

as long as $r \in (0,2)$.

This result means that if a process has a $\mathcal{SVSM}^C$ component and discrete jump components, the power variation will eliminate the jump parts and only capture the continuous infinite variation variance. Intuitively, this result stems from the fact that the increment $|p_{j\Delta} - p_{(j-1)\Delta}|^r$ measures increments of the type $|p^1_{j\Delta} - p^1_{(j-1)\Delta}|^r$ and of the type $|J|^r$ where J stands for jump. The scaling by $\Delta^{1-r/2}$ yields for the p^1 part, convergence to a well behaved integral, but for the jump part, convergence to 0.

If we consider, from here on, a small time increment $[t, t+dt]$ then, over this small time interval, for $r < 2$, it must hold that

$$[p]_{t+dt}^{[r]} - [p]_t^{[r]} = \mu_r \sigma_t^r dt \quad \Longrightarrow \quad \sigma_t = \left[\frac{[p]_{t+dt}^{[r]} - [p]_t^{[r]}}{\mu_r dt} \right]^{1/r}.$$

These computations show that in theory at least the instantaneous volatility may be easily estimated. For practical applications, where this expression would have to be estimated from data sampled over discrete time intervals and subject to microstructure noise, the usefulness of such a computation requires further investigations.

4.8.4 Bipower variation

BN/S extend this notion of power variation by focusing on the covariance properties of adjacent increments. This yields to the notion of bipower processes. For further extensions to multipower variation, we let the reader refer to their papers. An early investigation, from a probability theoretical point of view, was led by Lepingle (1976).

Definition 4.10. *The bipower variation is defined by*

$$\{p\}_t^{[r,s]} \equiv \underset{\Delta\to 0}{plim}\Delta^{1-\frac{r+s}{2}} \sum_{j=1}^{\lfloor t/\Delta\rfloor-1} |p_{j\Delta} - p_{(j-1)\Delta}|^r \cdot |p_{(j+1)\Delta} - p_{j\Delta}|^s, \quad r,s \geq 0.$$

We may trivially notice that, if $r = 0$ or $s = 0$, the bipower variation reduces to power variation. The notion of bipower variation, therefore, measures the behavior of powers of adjacent returns. If returns are in general driven by some continuous martingale component, given that this component will figure in both returns, the product will certainly capture it. In addition, following the logic of semi-martingales, there may be an occasional jump. Since the probability to have jumps in two adjacent returns tends to zero, this implies that the contribution of jumps to the bipower variation will be down-weighted. We might, therefore, intuitively expect that the bipower variation converges to an object where the jump component has been eliminated. The following key property shown by BN/S corroborates this intuition.

Proposition 4.11. *If $p = p^1 + p^2$ where $p^1 \in \mathcal{SVSM}^C$, i.e., $p_t^1 = a_t + m_t$, with a_t the predictible part and m_t the martingale part as in (4.35), and p^2 is a process with finite jumps over finite time increments, where the processes (a_t, σ_t) are independent of W_t, as well as p^1 and p^2 are independent then, for $r + s \leq 2$, the following holds*

$$\{p\}_t^{[r,s]} = \mu_r\mu_s \int_0^t \sigma_u^{r+s} du,$$

with μ_r defined as indicated above. For $r + s > 2$, the quantity diverges.

This result is interesting in that by setting $s = r-2$, the expression reduces to

$$\{p\}_t^{[r,2-r]} = \mu_r\mu_{2-r} \int_0^t \sigma_u^2 du.$$

This result shows that the bipower variation allows recovery of the integrated variance process, furthermore, it shows that by using the bipower variation, it is possible to get rid of the jumps. Combining these various results, we notice that the difference between power variation and bipower variation may be used to capture jumps

$$[p]_t^{[2]} - \frac{\{p\}_t^{[r,2-r]}}{\mu_r\mu_{2-r}} = \sum_{s\leq t} (\Delta p_s)^2.$$

So far, the results hold for a time period $[0, t]$. For empirical purposes, it may be interesting to consider these measures for finite time intervals. For instance, in empirical research, we may wish to compute the importance of jumps and of the continuous part, for a given day. Adapting the various concepts, introduced above, to finite time intervals is relatively easy. Rather than starting at 0, we may start at discrete points of time. Over finite time intervals, we may talk of

increments of quadratic variation, increments of bipower variation, etc. Also, rather than taking the theoretical limit cases of continuous time, by taking a discrete number of observations, it becomes possible to estimate the various ingredients by using actual data. Considering daily increments of the various variation measures, it is customary to call the estimate of the integrated daily quadratic variation *realized variance* and its square root *realized volatility*.

4.8.5 Estimation over finite time intervals

Some actual data

In this section, we indicate how to translate these various concepts into estimable entities. We also illustrate the type of results we may expect. To do so, we use a database consisting of high-frequency data of the French stock market. We focus on three companies, AXA, an insurance company, LVMH, a luxury fashion producer, as well as Société Générale, a large bank. The three companies under consideration stem from various sectors. For each of these companies, we extract data covering the second half of 2003 and the entire year 2004. The raw data is relatively large, for instance, on a typical day, there are about 12,000 transactions for AXA. From there on, we interpolated a price for predetermined 5-minute intervals.[25] We did this after dropping the opening price as well as those trades occurring after the official daily close, 8 hours and 30 minutes later. In our investigation, we do not work with returns from close to opening. Table 4.6 provides elementary statistics for these companies and for various frequencies of the data.

We find that the maximum return tends to be larger than the absolute value of the minimum return. This translates into the fact that during that period, the market was rather bullish. Contemplating the values of the standard deviation after converting them to a daily format, we obtain for AXA, $0.20 \times \sqrt{12 \times 8.5} = 2.01$ as a daily standard deviation using 5-minute returns, $0.66 \times \sqrt{8.5} = 1.92$ as daily standard deviation starting with hourly data, 1.73 using half-daily data, and eventually 1.67 using daily returns. These figures suggest that if we use high-frequency returns to compute daily standard deviations, we are led to an over estimation. This observation may be explained by possible autocorrelation in the data or by the presence of jumps in the data. The data appears in general positively skewed. The measures of kurtosis are larger than 3, suggesting that the data is generated by a process that is not distributed as a normal.[26]

To examine the temporal dependency in the data, we consider first-order autocorrelation of returns, taking care not to use returns from a previous

[25] The companies under study contained such a huge number of trades that if we had used the raw price rather than the interpolated price, the difference would have been negligible.

[26] Diebold et al. (1998) notice similar problems to obtain low-frequency volatility estimates, such as monthly ones, from daily estimates of volatility.

Table 4.6. *Various statistics for high-frequency data*

	AXA	LVMH	Soc. Gén.
Minimum			
5 minutes	−2.63	−1.41	−1.70
1 hour	−4.39	−2.87	−3.28
4 hours	−7.50	−5.65	−4.37
1 day	−5.21	−4.60	−5.47
Maximum			
5 minutes	3.23	1.58	1.78
1 hour	11.60	6.35	5.33
4 hours	12.55	6.48	6.45
1 day	13.96	7.52	8.82
Standard deviation			
5 minutes	0.20	0.16	0.17
1 hour	0.66	0.50	0.53
4 hours	1.23	0.97	0.97
1 day	1.67	1.31	1.35
Skewness			
5 minutes	0.12	−0.04	0.03
1 hour	1.33	0.78	0.37
4 hours	0.80	0.43	0.32
1 day	1.20	0.80	0.58
Kurtosis			
5 minutes	13.59	8.16	9.08
1 hour	33.10	13.35	9.82
4 hours	16.74	8.14	6.26
1 day	13.16	6.82	7.38
Autocorrelation			
5 minutes	−0.095	−0.129	−0.109
1 hour	0.013	0.062	0.057
4 hours	−0.018	0.080	0.059
1 day	−0.097	0.058	−0.085

day. From the table, we notice that, at the 5-minute frequency, the first-order autocorrelation is negative. This negative sign of the autocorrelation may be explained by the so-called bid-ask bounce. As trades occur, sometimes they take place at the bid and sometimes at the ask. Even if the value of a company does not change, given that the prices at which the trades occur oscillate, this translates into a negative first-order autocorrelation. As the time frequency changes, we find changing signs of the autocorrelations. Already at the hourly level, all the correlations are positive. This observation raises the issue of the optimal sampling frequency, which we will discuss in the last section of this chapter.

Realized measures

In order to implement estimation of the concepts seen above, such as the realized variance, we need to provide some notations. Assume that each day has been cut into M intervals that we suppose to be of constant step size, so that the instants where prices are measured are given by $j\delta$ where δ is the step size and $j = 0, \cdots, M$. For instance, δ could correspond to 5 minutes and $M = 102$ for a day with 8 hours and 30 minutes. We index by i the ith day.

The following expression defines the day i realized variance

$$\widehat{[p]}^{[2]}_{i|M} = \sum_{j=1}^{M} \left(p_{(i-1)M\delta+\delta j} - p_{(i-1)M\delta+\delta(j-1)}\right)^2. \tag{4.36}$$

Notice that the day figures as an index to the brackets and so does the sampling frequency. The realized volatility of day i is obtained as the square root of the realized variance:

$$\sqrt{\widehat{[p]}^{[2]}_{i|M}}. \tag{4.37}$$

The discrete measures for the realized (rth) power variation and the realized bipower variation are respectively

$$\widehat{[p]}^{[r]}_{i|M} = \delta^{1-\frac{r}{2}} \sum_{j=1}^{M} |p_{(i-1)M\delta+\delta j} - p_{(i-1)M\delta+\delta(j-1)}|^r, \tag{4.38}$$

$$\begin{aligned}\widehat{\{p\}}^{[r,s]}_{i|M} = \delta^{1-\frac{r+s}{2}} &\sum_{j=1}^{M-1} |p_{(i-1)M\delta+\delta j} - p_{(i-1)M\delta+\delta(j-1)}|^r \\ &\times |p_{(i-1)M\delta+(\delta+1)j} - p_{(i-1)M\delta+\delta j}|^s.\end{aligned} \tag{4.39}$$

Clearly, in terms of convergence, associating a global time index t to the various days, we obtain

$$\widehat{[p]}^{[2]}_{i|M} \underset{M\to\infty}{\to} [p]^{[2]}_t - [p]^{[2]}_{t-1} = \int_{t-1}^{t} \sigma_u^2 du + \sum_{t-1\le s<t} (\Delta p_s)^2,$$

where $t \in (iM\delta, iM\delta + \delta)$ and $t - 1 \in ((i-1)M\delta, (i-1)M\delta + \delta)$, and where $\sum_{t-1\le s<t}(\Delta p_s)^2$ stands for the jumps that occurred between $t-1$ and t. Also, for $r + s < 2$, we have

$$\widehat{[p]}^{[r]}_{i|M} \underset{M\to\infty}{\to} [p]^{[r]}_t - [p]^{[r]}_{t-1} = \mu_r \int_{t-1}^{t} \sigma_u^r du,$$

$$\widehat{\{p\}}^{[r,s]}_{i|M} \underset{M\to\infty}{\to} \{p\}^{[r,s]}_t - \{p\}^{[r,s]}_{t-1} = \mu_r \mu_s \int_{t-1}^{t} \sigma_u^{r+s} du.$$

As formulae (4.36)–(4.39) show, it is relatively easy to compute the various expressions of interest. These expressions indicate that it is also possible to focus on a time increment such as a day rather than on an ongoing process.

To illustrate these various concepts, we consider LVMH over three days around April 1, 2004. This is a day with a relatively large negative return. Figure 4.13 traces the daily returns. As we notice, the returns are relatively large on the negative side for the three consecutive days. To provide a further insight on how the returns of these days got generated, we consider higher frequencies, namely returns at an hourly and 5-minute frequencies.

Figure 4.14 focuses on the hourly frequency. We have added bars to this figure. The longer vertical bar represents the last return of a given day. The shorter lines delimit the various hours. We included the remaining 30 minutes into the computation of the last hour. As this picture reveals, on the first day, there are many relatively small negative returns that cumulate into something larger at the daily level. On the second day, the second hour finishes with a large negative return of about -3%.

To investigate whether this is a jump or just some continuous variation, we further focus on data where the sampling takes place at a 5-minute frequency. This data is displayed in Figure 4.15. These returns reveal that the first few hours of the second day are quite agitated. To investigate if the relatively large drop of -3% stems from a single event or a succession of events, we consider also the raw price process displayed in the next figure.

Figure 4.16 represents price increments at a 5-minute interval. We notice that the prices dropped in a rather steady manner during the second hour suggesting that the -3% return is likely due to some continuous event rather than to a jump. To investigate this conjecture more formally, we turn to

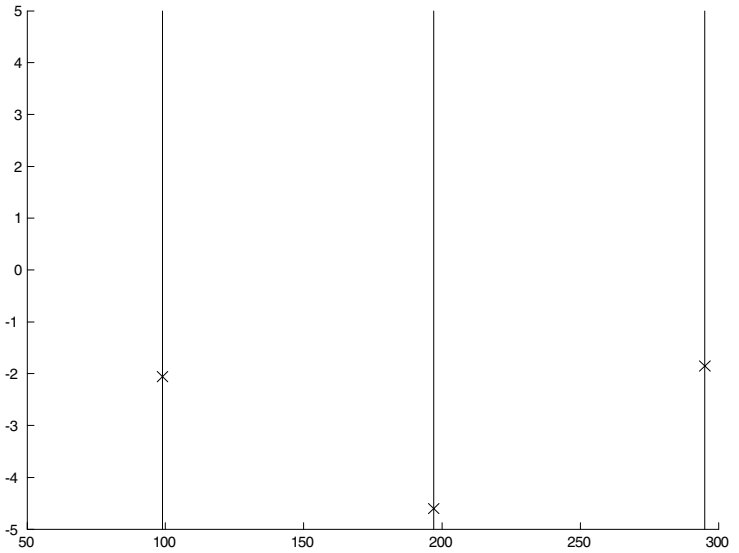

Fig. 4.13. *Returns of LVMH sampled once per day.*

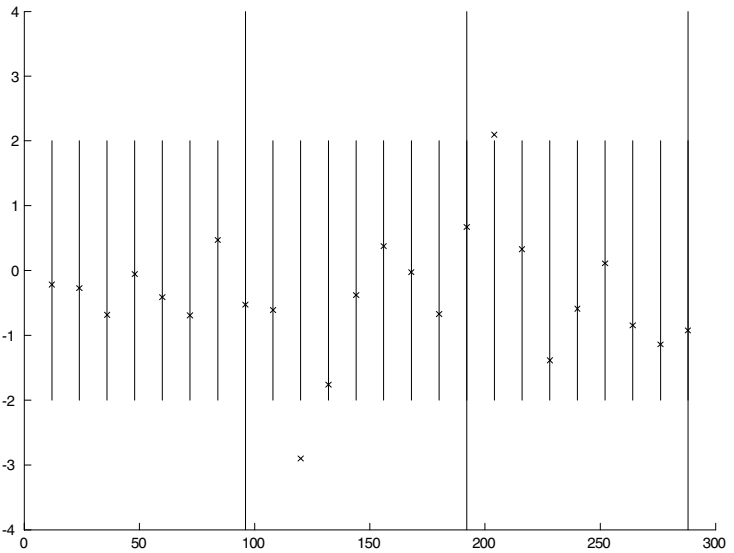

Fig. 4.14. *Returns of LVMH sampled at hourly frequency.*

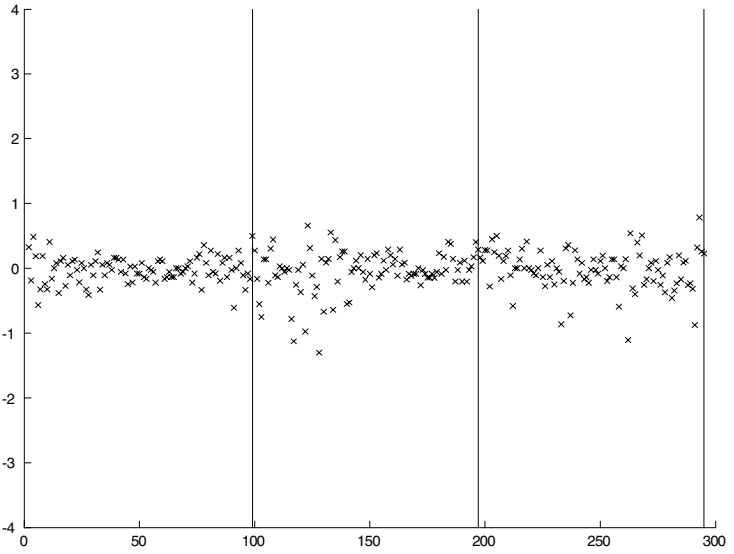

Fig. 4.15. *Returns of LVMH sampled every 5 minutes.*

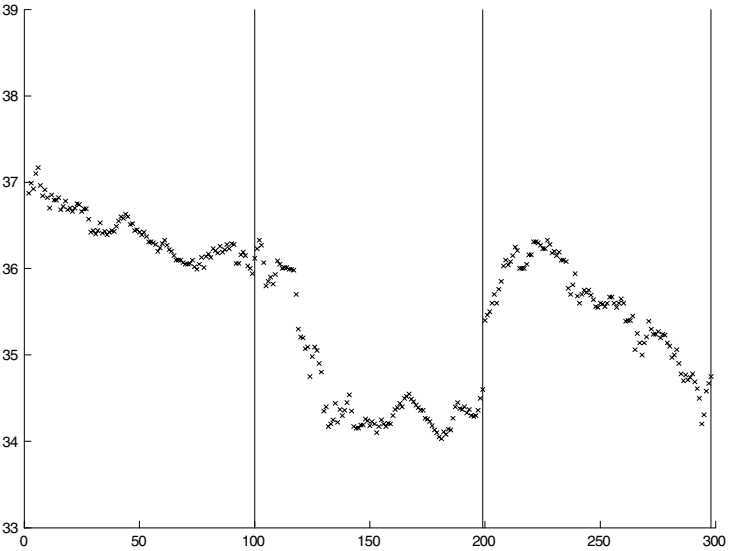

Fig. 4.16. *Raw prices of LVMH sampled every 5 minutes.*

the concepts of realized variance and bipower variation that should allow the detection of jumps.

Figure 4.17 represents the results of the computation of various statistics. For comparison purposes, we slightly increased the number of days. The largest price variation (on April 1, 2004) is now on day 4. The upper panel of the figure represents the daily estimate of quadratic variation, which is realized variance. This figure also contains the power variation obtained with $r = 1$ over the various days. As this figure shows, the two measures are relatively close. Obviously, given that quadratic variation captures the integrated variance and power variance, for the given power, captures the rth power of the absolute value of standard deviation, there is no reason for these two statistics to take the same value. We just notice that the two statistics evolve in the same direction.

In the lower panel of Figure 4.17, we display the difference between realized variance and realized bipower variation. As indicated by Barndorff-Nielsen and Shephard (2004a), in small samples, the estimate of the jump may yield to a negative estimate. Since the difference of these two statistics should theoretically capture the sum of squared jumps, the negativity of the measure is counterfactual. We may improve this computation by setting a lower bound at zero.

In terms of interpretation, as the lower panel of the figure indicates, there are several other days with jump measures as large as for the given day. Hence, the process for that day should be considered to be generated by a process

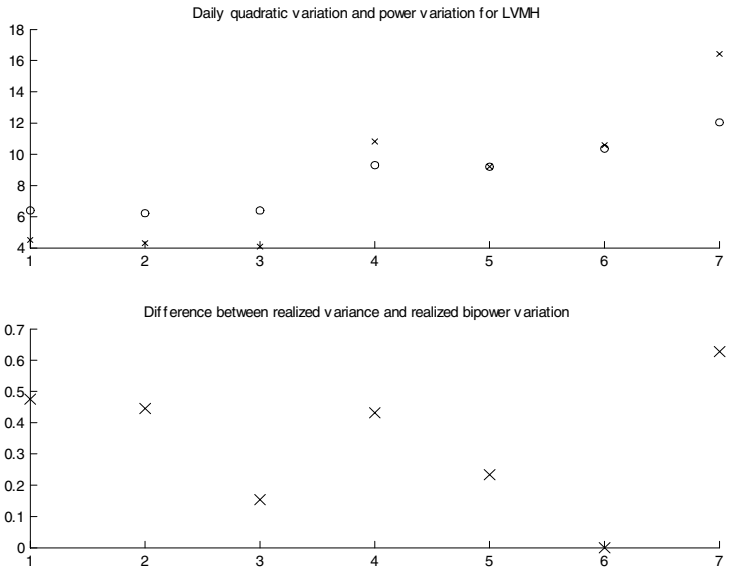

Fig. 4.17. *Daily quadratic variation and bipower variation for LVMH.*

without discrete jumps. Alternatively, one could imagine that the process jumps very frequently. This, however, appears in contradiction with the fact that jumps are supposed to take place only infrequently.

We may also ask how good the fit is. To do so, we compute for data, covering one year and a half, the various daily returns and standardize them by the square root of daily realized variance. Formally, for day i, we take $r_i / \sqrt{\widehat{\{p\}}^{[2]}_{i|M}}$, with M equal to 102 intervals. We could expect that the resulting measure is distributed as a normal if quadratic variation captures all activity. To perform this test, we trace a so-called quantile plot in Figure 4.18. As the left tail indicates, there remains some information in the returns that does not appear to be captured by realized variance, computed over a daily frequency.[27] An additional question that we can ask is the relation between volume and realized variance. Indeed, according to the mixture of distributions hypothesis, volume could represent the daily information content. It was already shown by Clark (1973) that even though these two measures appear correlated, this relation is not a good one. We may verify this result here by contemplating in Figure 4.19 the scatterplot between daily volumes on the horizontal axis (measured by the number of shares traded), and the realized variance along the vertical axis.

[27] For stock returns and exchange rates, Andersen et al. (2001) and Andersen et al. (2003) provide empirical evidence that returns scaled by realized standard deviations are approximately Gaussian.

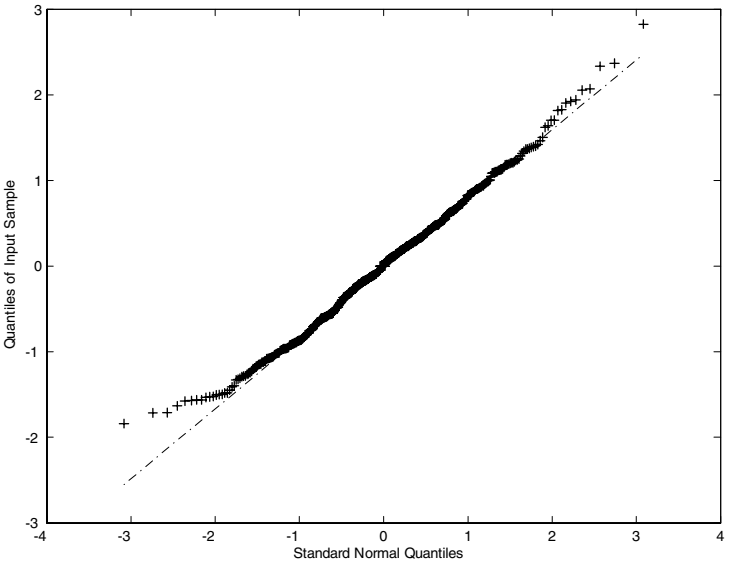

Fig. 4.18. *Daily devolatilized returns.*

We notice that the largest daily realized variances are associated with high volume but not the largest ones. Thus, even though there certainly exists a relation between volume and daily realized variances, this relation is far from being simple.

So far, our focus has been on univariate data. We turn now to the behavior of asset prices within a multivariate setting. In the following section, we start by providing theoretical concepts before turning to some empirical illustrations.

4.8.6 Realized covariance

Theory

It is possible to obtain a construction for high-frequency measures of covariance and correlation. Whereas Andersen et al. (2003) indicate how to compute the covariance from high-frequency data and its properties, BN/S provide a distribution theory.

Definition 4.12. *Consider two processes, written* p^1 *and* p^2 *defined over some time interval* $[0, t]$*. The realized covariance is given by the expression*

$$[p^1, p^2]_t \equiv \underset{\Delta \to 0}{plim} \sum_{j=1}^{\lfloor t/\Delta \rfloor} (p^1_{t_j} - p^1_{t_{j-1}})(p^2_{t_j} - p^2_{t_{j-1}}).$$

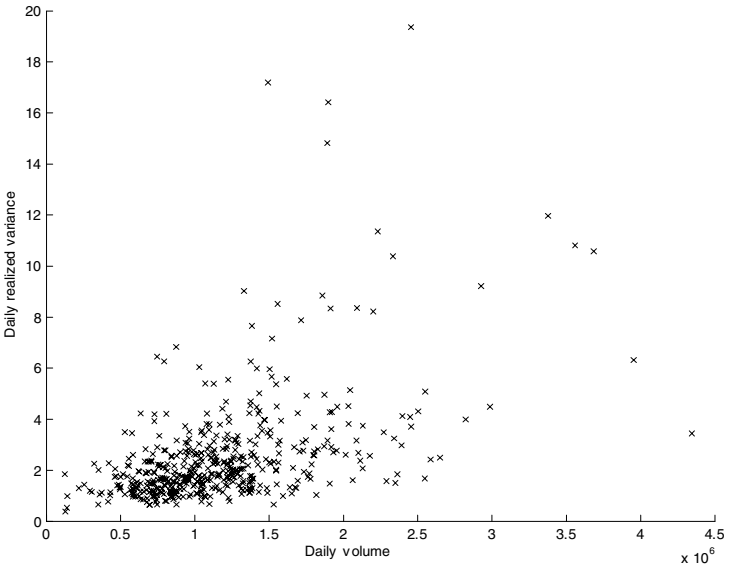

Fig. 4.19. *Daily realized variance and volume for LVMH.*

If both p^1 and p^2 belong to the $\mathcal{SVSM}^C$ class, then we can always, by definition, write that

$$\begin{pmatrix} p_t^1 \\ p_t^2 \end{pmatrix} = a_t + m_t, \qquad \text{and} \qquad m_t = \int_0^t \theta_u dW_u,$$

where a_t is a $(2,1)$ vector of drift terms and m_t is a two-dimensional martingale. The θ_u is a $(2,2)$ matrix, and W_u represents a two-dimensional Brownian motion. From there on, we may write

$$\Sigma \equiv \theta_u \theta_u' = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}.$$

Clearly, σ_{12} plays the role of a covariance. It may be shown that the realized covariance converges to an integrated covariance

$$[p^1, p^2]_t = \int_0^t \sigma_{12,u} du.$$

Interestingly, if we have such estimates of a covariance matrix, it becomes possible to also discuss portfolio allocations from a high-frequency point of view. As previously, it is possible to specialize realized covariance to an interval of time. We consider a given day i that is decomposed into M time intervals, each of size δ. Then we have

$$\widehat{[p^1,p^2]}_{i|M} = \sum_{j=1}^{M} \left(p^1_{(i-1)M\delta+\delta j} - p^1_{(i-1)M\delta+\delta(j-1)}\right) \times \left(p^2_{(i-1)M\delta+\delta j} - p^2_{(i-1)M\delta+\delta(j-1)}\right).$$

This expression is, thus, an estimate of the integrated covariance and will be called realized covariance. In the limit as M converges to infinity, realized covariance will converge to the increment of integrated covariance.

Using the realized covariance as well as realized variances, it is possible to compute a measure of realized correlation defined as

$$\frac{\widehat{[p^1,p^2]}_{i|M}}{\sqrt{\widehat{[p^1]}_{i|M}\widehat{[p^2]}_{i|M}}}.$$

We notice that this expression can be easily computed.

An empirical measure of realized correlation

Given the concepts of realized covariance and realized variance, it is also possible to construct a realized correlation. We just measure a covariance for data sampled at some frequency over some time intervals, for instance daily, and standardize it by the corresponding realized volatility. Turning to the three companies already considered, it is possible to investigate the correlation between the assets. We take all the data over the one year and a half, extract returns at various frequencies and then compute realized covariances, realized volatilities, and the resulting realized correlation. Table 4.7 presents the resulting measures. We notice that the correlation between AXA and LVMH is about the same as between LVMH and Société Générale, if we focus on 5-minute returns and that it takes a value of about 0.3. The corresponding correlation between AXA and Société Générale is 0.41. Investigation of the table shows that as the sampling frequency increases, the correlations become larger even though the overall ranking of the correlations remains the same.

It is also possible to investigate the intertemporal evolution of correlation. In Figure 4.20, we represent daily correlations computed for the three companies under investigation. The correlations are obtained by taking 5-minute returns for each day. To ease the detection of patterns, the correlations have been smoothed with a 10-day gliding average. The thick line represents the correlations between AXA and LVMH. The thin continuous line and the dotted line correspond to the correlations between AXA and Société Générale, respectively LVMH and Société Générale.

As this figure suggests, the correlations between the various companies are systematically anchored at different levels. We also notice that the correlations vary through time. To investigate the stability of the correlations through time, we take again AXA–LVMH and consider daily correlations computed

Table 4.7. *Realized correlations for high-frequency data*

	AXA	LVMH	Soc. Gén.
5-minute returns			
AXA	1.00	0.30	0.41
LVMH	0.30	1.00	0.31
Soc. Gén.	0.41	0.31	1.00
Hourly returns			
AXA	1.00	0.53	0.63
LVMH	0.53	1.00	0.51
Soc. Gén.	0.63	0.51	1.00
Returns over 4 hours			
AXA	1.00	0.60	0.65
LVMH	0.60	1.00	0.53
Soc. Gén.	0.65	0.53	1.00
Daily returns			
AXA	1.00	0.63	0.69
LVMH	0.63	1.00	0.55
Soc. Gén.	0.69	0.55	1.00

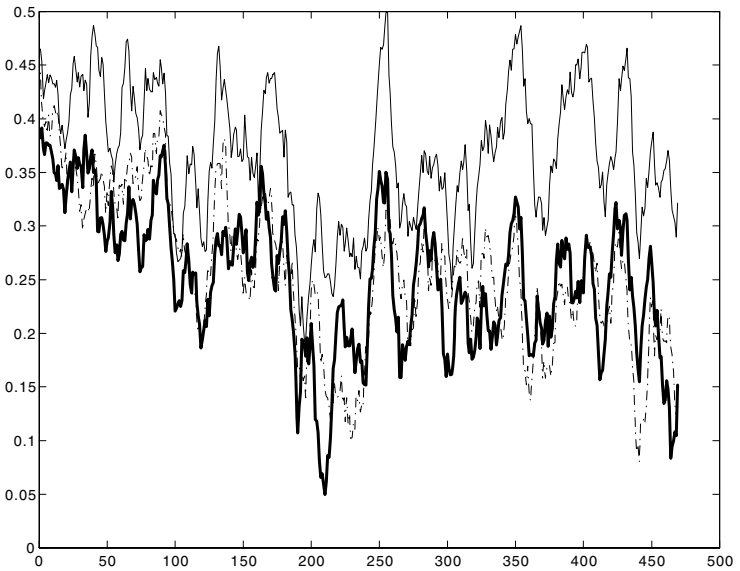

Fig. 4.20. *Daily realized correlations computed with 5-minute returns.*

with data sampled at the 5-minute frequency and the hourly frequency. Again, we smooth the data with a 10-day gliding average. As Figure 4.21 represents, there is quite a significantly different pattern in the correlations. This raises the question of the frequency at which the data should be sampled for this type of measure.

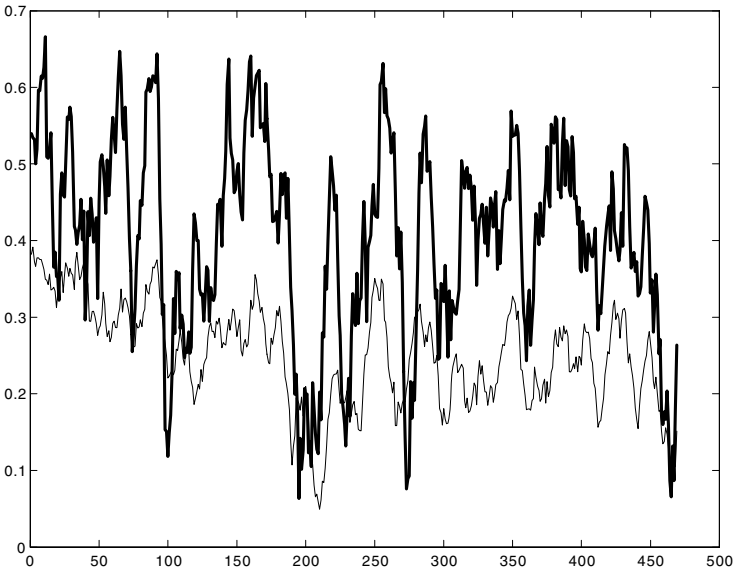

Fig. 4.21. *Daily realized correlations computed at 5-minute and 1-hour frequency.*

The time variability of correlation is an issue that is also addressed at the beginning of Chapter 6 with multivariate GARCH models. The link between measures of correlation, based on high-frequency data, and GARCH correlations appears to be an open issue.

Testing formally for jumps

So far, we have indicated how to decompose a price process into a continuous part and a discrete part. The following elements, drawn from Barndorff-Nielsen and Shephard (2004a), may be useful to formally test for the existence of jumps. See also BN/S.

As before, denote by $\mu_r = E[|u|^r]$, the rth moment of u. When a formula contains u, u', and u'', we assume that the associated distributions are all normal $\mathcal{N}(0,1)$ and independent. If p belongs to the class of $\mathcal{SVSM}^C$, then as Δ converges toward 0, for any $r > 0$,

$$\frac{\Delta^{-\frac{1}{2}}\left(\{p\}_t^{[r,r]} - \mu_r^2 \int_0^t \sigma_s^{2r} ds\right)}{\sqrt{\mu_{4r}^{-1}\left\{V\left[|u|^r |u'|^r\right] + 2Cov\left[|u|^r |u'|^r, |u'|^r |u''|^r\right]\right\}[p]_t^{[4r]}}} \Rightarrow \mathcal{N}(0,1),$$

where $\Rightarrow$ corresponds to convergence in distribution. Even though this result has been derived for the general case with r taking a general value, this result may prove most useful for the case where $r = 1$. Indeed, this is the case

corresponding to quadratic variation. In this case, the formula allows for a construction of a confidence interval to which the true quadratic variation belongs. If the expression of $\int_0^t \sigma_s^{2r} ds$ was known, we could even consider testing for jumps. Indeed, if the statistics fall outside the confidence interval, this may be interpreted as a failure of the null hypothesis. This could then indicate the presence of a jump in the data. Since, in practice, the integral is not known, we need further results.

Fortunately, BN/S also provide results for this case, where we work with quadratic variation and bipower variation. To test if there are jumps, we need indeed a statistics concerning the difference between a quadratic variation and a bipower variation. This is where the following result is useful. This result holds for the case when the price process is again in the $\mathcal{SVSM}^C$ class. We have

$$\frac{\Delta^{-\frac{1}{2}}}{\sqrt{\int_0^t \sigma_s^4 ds}} \begin{pmatrix} \sum_{j=1}^{\lfloor t/\Delta \rfloor} (p_{j\Delta} - p_{(j-1)\Delta})^2 - \int_0^t \sigma_s^2 ds \\ \mu_1^{-2} \sum_{j=1}^{\lfloor t/\Delta \rfloor - 1} |p_{j\Delta} - p_{(j-1)\Delta}||p_{(j+1)\Delta} - p_{j\Delta}| - \int_0^t \sigma_s^2 ds \end{pmatrix} \Rightarrow \mathcal{N}(0, \Sigma)$$

where

$$\Sigma = \begin{pmatrix} V\left[u^2\right] & 2\mu_1^{-2} Cov\left[u^2, |u|\,|u'|\right] \\ 2\mu_1^{-2} Cov\left[u^2, |u|\,|u'|\right] & \mu_1^{-4}\{V\left[|u|\,|u'|\right] + 2Cov\left[|u|\,|u', |u'|\,|u''|\right]\} \end{pmatrix}.$$

This covariance matrix may be computed once and for all numerically because none of the terms depends on the actual data. BN/S obtain

$$\Sigma = \begin{pmatrix} 2 & 2 \\ 2 & 2.60907 \end{pmatrix}.$$

These computations show that the efficiency of quadratic variation is somewhat better than of realized bipower variation. Indeed, once the variance is 2, the other time it is about 2.6. The correlation between the two statistics is equal to 0.87. This implies that the information content of both statistics will be very similar. The previous results are clearly obtained under the null hypothesis of no jumps. If we wish to test whether there are jumps, we may use the following expression:

$$\frac{\Delta^{-\frac{1}{2}}}{\sqrt{\int_0^t \sigma_s^4 ds}} \left[\sum_{j=1}^{\lfloor t/\Delta \rfloor} (p_{j\Delta} - p_{(j-1)\Delta})^2 - \mu_1^{-2} \sum_{j=1}^{\lfloor t/\Delta \rfloor - 1} |p_{j\Delta} - p_{(j-1)\Delta}|\ |p_{(j+1)\Delta} - p_{j\Delta}| \right] \Rightarrow \mathcal{N}(0, 0.6091).$$

Hence, as mentioned, this statistic can be used to test for the presence of jumps. To render this formula operational, it is necessary to evaluate the

integrated fourth power of volatility, also called *integrated quarticity,* that appears under the square-root. This can be done by using the fact that,

$$\operatorname*{plim}_{\Delta \longrightarrow 0} \Delta^{-1} \sum_{j=1}^{\lfloor t/\Delta \rfloor - 1} \left(p_{j\Delta} - p_{(j-1)\Delta}\right)^2 \left(p_{(j+1)\Delta} - p_{j\Delta}\right)^2 = \int_0^t \sigma_s^4 ds,$$

provides a feasible estimate of *realized quarticity*, an empirical measure of integrated quarticity. It is, thus, sufficient for practical purposes to approximate the integral by a finite sum obtained for a selected Δ.

4.8.7 Further related results

So far, we have discussed some recent theoretical developments. We now present some of the empirical results that have been found in the literature. First, we may mention a literature that deals with comparing volatility forecasts to actual volatility realizations. One such contribution is the one by Andersen et al. (2003). It focuses, like much of this literature, on exchange rate data. These contributions show that realized volatility is best forecast by using a vector autoregression involving multivariate realized volatility and where some long memory is allowed for by using a fractional difference, rather than by a model belonging to the GARCH family. As we have seen in the above statistical analysis of French stock market data, the sampling frequency plays an important role. Sampling data at a 5-minute frequency will not yield to the same pattern of autocorrelation than if hourly data is used. The reason for this is that at an increasing frequency, so-called *microstructure noise* comes into play. This noise may be explained by the bid-ask spread, by the fact that a given asset price may be coming from different markets, or due to the fact that prices are traded at discrete values. The contribution by Andersen, Bollerslev, and Meddahi (2004) shows that, in the presence of microstructure noise, the realized volatility forecast is even better than originally thought.

As just mentioned, the role of microstructure noise may hamper the estimation of realized volatility or of realized covariances. The question then is of course what time frequency we should use. As seen at the beginning of this section, the data should be sampled at a high enough frequency to allow for a disentangling of jumps. On the other hand, a finer and finer sampling, say at the level of minutes or seconds, will introduce an increased level of microstructure noise. Several contributions, such as Aït-Sahalia, Mykland, and Zhang (2005a and 2005b) investigate on how to improve the measure of realized volatility in the presence of microstructure noise. Their starting point is a simplified model where observed log-prices, $\tilde{p}_t$, are the sum of fundamental, unobservable log-prices, p_t, and some noise, ε_t. Beyond the already mentioned sources of noise, such as price discreteness, the noise may represent elements such as price impact due to trade size, inventory components of the bid-ask spread, etc. These authors adopt the following model

$$\tilde{p}_t = p_t + \varepsilon_t,$$
$$dp_t = \mu_t dt + \sigma_t dW_t,$$

where ε_t can be *iid* or dependent. They show that a good estimate, named Two Scales Realized Volatility (TSRV), is based on two components. First, one computes at a given slow timescale, such as at the 5-minute frequency, different daily realized volatilities where each measure starts at a different point of time. For instance, the first such realized volatility would start at 8:00:00 and then pursue at the 5-minute frequency, the second one would start at 8:00:30 and then pursue at 5-minute frequency, and so on. These realized volatilities would then be averaged. In other words, the first measure would be obtained by subsampling. As shown by Barndorff-Nielsen and Shephard (2004a), and as has been used already by Andersen, Bollerslev, and Meddahi (2004), this estimate will not be a good estimate of integrated variance because of some discretization bias. By combining this first averaged realized volatility with a realized volatility computed at a different timescale, for instance a fast timescale such as every 10 seconds, a more precise estimate may be obtained. The proposed estimator of realized volatility is

$$\widehat{[p]}_i^{[TSRV]} = \frac{1}{K}\sum_{k=1}^{K}\widehat{[p]}_{k;i|M}^{[2]} - \frac{M}{M'}\widehat{[p]}_{i|M'}^{[2]}$$

where

$$\widehat{[p]}_{k;i|M}^{[2]} = \sum_{j=1}^{M}\left(p_{(i-1)M\delta+t_k+\delta j} - p_{(i-1)M\delta+t_k+\delta(j-1)}\right)^2,$$

and where t_k, for $k = 1, \cdots, K$ are starting moments during the first daily interval $[0, \delta]$, M and M' are the number of intervals taken for the slow and the fast timescale. The suggestion to use estimates based on different timescales may also be found in Bandi and Russel (2004).

A better measurement of volatility is likely to be useful for option pricing. A decomposition of prices into continuous and discrete components may improve hedging. An analysis of volatility may also shed light on the behavior of investors. One paper in this direction is the one by Andersen, Bollerslev, and Diebold (2003). The authors consider realized volatilities measured over timescales and show, in a vector autoregression framework, that realized volatility contains complementary information at each of these timescales. This suggests that there are heterogeneous investors whose time horizons are not the same. Such an observation suggests that a better understanding of investor behavior from an economic point of view may be useful.