8

Risk Management and VaR

Until quite recently, the variance was a widely accepted measure of risk. It is very easy to understand and to compute. A shortcoming of this measure is that it is a symmetric one, in the sense that large gains and losses are equally penalized. Financial institutions however are much more concerned by large losses than by large gains.

Financial and regulatory institutions realized that there are many different sources of risk. These sources of risk have been progressively more precisely described and understood. For instance, three major types of risk are nowadays acknowledged; market risk, credit risk, and operational risk. Step by step, the Basel Committee on Banking Supervision (BCBS), at the Bank of International Settlements, imposes new capital requirements to financial institutions to cover these different sources of risk. The first step was the implementation of new standards for computing the exposition to risk and for measuring margin requirements. While the initial *1988 Basel Accord* only covered credit risk, in the *1996 Amendment to the Capital Accord to Incorporate Market Risks*, the BCBS incorporates market risk and explicitly introduces the Value-at-Risk as the main quantitative tool for financial institutions to calculate their capital requirement. VaR then became one of the widely used measures of market risk in the risk-management and fund-management industries.

Let θ denote a small percentage. Then the VaR of a portfolio can be defined as the minimum potential loss that the portfolio can suffer in the $\theta\%$ worst cases, over a given time horizon. Therefore, from a statistical point of view, the VaR is a quantile on the lower tail of the distribution of portfolio returns. Since its naming by the BCBS, the VaR is an actively researched topic, because it raises several interesting theoretical issues and has important implications for financial institutions. Early work on VaR has been done by J.P. Morgan (1996), Jorion (1997), Duffie and Pan (1997). J.P. Morgan, through its RiskMetrics methodology, has played an important role in the increasing popularity of VaR as a risk measure. Then, a huge literature emerged on the practical estimation of the VaR. Several approaches have been proposed to provide a precise evaluation of this measure. These approaches have been based, for instance, on the use of univariate or multivariate GARCH models, on the modeling of the tails of the distribution, and on the modeling of non-normality. Interestingly, most of the elements needed for such computation have been analyzed in the previous chapters.

More recently, the VaR has been shown to be a potentially misleading measure of risk. A first criticism is that VaR is not a coherent measure because it does not satisfy the sub-additivity property, so that diversification does not necessarily result in a reduction of risk, as measured by VaR (see below for more formal definitions). Perhaps more importantly, it has been argued (BIS Committee on the Global Financial System, 2000) that VaR (as well as variance) misregards the risk of extreme loss. As has been highlighted by Basak and Shapiro (2001), because VaR disregards risk of extreme losses behind the confidence level, it may induce large losses. Consequently, this measure of risk may induce a larger risk exposure than the variance in case of falling markets. To cope with these shortcomings, some authors (Artzner et al., 1999, Basak and Shapiro, 2001) have proposed the use of the so-called Expected Shortfall (ES) as an alternative measure of risk. ES is the expected value of the loss of the portfolio in the $\theta\%$ worst cases over a given time horizon.

This chapter is organized as follows. Section 8.1 defines the two main notions we are going to analyze in this chapter, namely the VaR and the ES of a portfolio. The following sections are devoted to the practical evaluation of VaR as well as ES. There are actually four broad categories of VaR models: historical simulation (Section 8.2); semi-parametric models (Section 8.3), parametric models (Section 8.4); and finally non-linear techniques, which are designed to compute the VaR in presence of derivatives (Section 8.5). In the last section, we provide some tools to compare the performances of the various techniques developed in the previous sections.

8.1 Definitions and measures

8.1.1 Definitions

Let us now formally define the notions of VaR and ES. The VaR at probability $\theta \in (0,1)$ of a portfolio is defined as the minimum potential loss that the portfolio may suffer in the $\theta\%$ worst cases, over a given time horizon. We define $P_{i,t}$ the price at date t of asset i, so that the return of asset i between date $t-1$ and date t is $r_{i,t} = (P_{i,t} - P_{i,t-1})/P_{i,t-1}$. Then the value of the portfolio at date t, for a vector N_t that contains the number of shares in asset i, is simply given by $W_t = \sum_{i=1}^{n} N_{i,t}P_{i,t} = N_t'P_t$. If we assume that portfolio composition is held constant from t to $t+1$, the change in the market value of the portfolio is given by $W_{t+1} - W_t = N_t'(P_{t+1} - P_t)$. We deduce that $(W_{t+1} - W_t)/W_t = \alpha_t' r_{t+1} = r_{p,t+1}$, where $\alpha_{i,t} = N_{i,t}P_{i,t}/\left(\sum_{i=1}^{n} N_{i,t}P_{i,t}\right)$. We denote the wealth of the investor for a portfolio weight vector α_t as $W_t(\alpha_t)$.

Value at Risk

If we denote $\Delta W_{t+1}(\alpha_t) = W_{t+1}(\alpha_t) - W_t(\alpha_t)$ the VaR of a portfolio is defined by the relation

$$\theta = \Pr\left[\Delta W_{t+1}(\alpha_t) \leq -\overline{VaR}_{\theta,t} | \mathcal{F}_t\right],$$

where $\mathcal{F}_t$ denotes the information set at date t. Alternatively, we have[1]

$$\theta = \Pr\left[\frac{\Delta W_{t+1}(\alpha_t)}{W_t(\alpha_t)} \leq -\frac{\overline{VaR}_{\theta,t}}{W_t(\alpha_t)} | \mathcal{F}_t\right] = \Pr\left[r_{p,t+1} \leq -VaR_{\theta,t} | \mathcal{F}_t\right],$$

where $VaR_{\theta,t}(r_{p,t+1}) = \overline{VaR}_{\theta,t} / W_t(\alpha_t)$ denotes the VaR for probability θ for \$1 invested. In the following, it will be our definition of VaR. If we also define the conditional *cdf* $F_{p,t}(x) = \Pr[r_{p,t} \leq x | \mathcal{F}_{t-1}]$, with $F_{p,t}^{-1}$ the inverse of the conditional *cdf*, we observe that

$$VaR_{\theta,t} = -F_{p,t}^{-1}(\theta).$$

This expression indicates that the VaR of the portfolio at time t for the next period is (minus) the θ-quantile of the conditional *cdf* of the portfolio return. Evidently, if we assume that returns are *iid*, the VaR is constant over time and is simply given by the inverse of the unconditional *cdf*: $VaR_\theta = -F_p^{-1}(\theta)$, where $F_p(x) = \Pr[r_{p,t} \leq x]$ is the (assumed to be continuous) *cdf* of the portfolio return. Computing the VaR therefore "reduces" to estimating a quantile of the conditional distribution of the portfolio return. Notice that this *cdf* is very likely to be time varying. One reason is that the volatility of returns varies over time, see also Chapter 2.

The distribution of the portfolio return can also be described as depending on the joint distribution of asset returns. This brings about two fashions to consider the VaR and ES. A first one, that we will call portfolio approach will consider the distribution of an aggregate return. The second one will consider the induividual assets in an approach that we call the asset approach. In fact, the use of the asset-level approach is worthy to be considered, because it may be used for more in-depth analysis. This point has been outlined by Gouriéroux, Laurent, and Scaillet (2000). For instance, it allows computing the sensitivity of the portfolio VaR to changes in the weights of the portfolio. Ultimately, it can be used to directly optimize the weights of the portfolio under VaR constraints (see for instance, Huisman, Koedijk, and Pownall, 1999, or Krokhmal, Palmquist, and Uryasev, 2002). This issue is also addressed in Section 9.2. Notice that a reduction of the dimensionality can be attained by

[1] Rigorously, the VaR is defined as

$$VaR_{\theta,t} = -\sup\{x | \Pr[r_{p,t+1} \leq x] \leq \theta\},$$

because the *cdf* may be constant over some interval. We use simpler notations for ease of exposition.

defining some factors supposed to capture the main sources of risk affecting the portfolio return.

To estimate the VaR of a large portfolio over time, we can adopt several approaches. The first approach is based on the computation of the VaR of the portfolio. An obvious advantage of such an approach is that it avoids modeling the joint dynamics of asset returns. A drawback is that it will probably miss some important links between asset returns such as the time-varying correlation. The second set of approaches is based on the modeling of the joint distribution of asset returns.

In general, the VaR is computed over a time horizon k (10 day, for instance). In such a case, we have, assuming the position is held constant over the horizon, the following definition for the multi-period VaR

$$\theta = \Pr\left[\frac{\Delta W_{t+k}\left(\alpha_t\right)}{W_t\left(\alpha_t\right)} \leq -\frac{\overline{VaR}_{\theta,t:t+k}}{W_t\left(\alpha_t\right)}\right] = \Pr\left[r_{p,t}\left[k\right] \leq -VaR_{\theta,t:t+k}\right],$$

where $r_t[k]$ is the cumulative return between t and $t+k$ and we defined the VaR for 1\$ invested as $VaR_{\theta,t:t+k} = \overline{VaR}_{\theta,t:t+k}/W_t\left(\alpha_t\right)$. To obtain the multi-period VaR, we therefore need an estimate of the conditional distribution of the multi-period portfolio return $r_{p,t}\left[k\right]$.

Coherent measure of risk

Artzner et al. (1997, 1999) proposed a set of conditions that a coherent measure of risk should satisfy:

Definition 8.1. *Let V be a set of real-valued random variables (typically, the net final wealth). The function $\rho : V \to \mathbb{R}$ is a coherent risk measure if it safisties:*

1. *Translation invariance: $X \in V$, $\alpha \in \mathbb{R}$, then $\rho\left(X+\alpha\right) = \rho\left(X\right) - \alpha$.*
2. *Sub-additivity: $X, Y \in V$, then $\rho\left(X+Y\right) \leq \rho\left(X\right) + \rho\left(Y\right)$.*
3. *Positive homogeneity: $X \in V$, $\lambda \geq 0$, then $\rho\left(\lambda X\right) = \lambda\rho\left(X\right)$.*
4. *Monotonicity: $X, Y \in V$, with $X \leq Y$, then $\rho\left(X\right) \geq \rho\left(Y\right)$.*

Translation invariance means that if we add a sure amount α to the position, it will decrease the risk measure by α. Sub-additivity implies that the risk of a portfolio constituted of two sub-portfolios is smaller than the sum of the risk of the two sub-portfolios. Positive homogeneity means that if we increase the size of the portfolio by a factor λ with the same weights, we increase the risk measure by the same factor λ. Monotonicity means that the risk is greater for more negative random outcomes.

They then show that the VaR fails to be a coherent measure because it does not satisfy the sub-additivity property, so that diversification does not necessarily result in a reduction of risk as measured by VaR. Perhaps more importantly, it has been argued (BIS Committee on the Global Financial

System, 2000) that VaR (as well as variance) misregards the risk of extreme loss. As has been highlighted by Basak and Shapiro (2001), because VaR disregards risk of extreme losses behind the confidence level, it may induce large losses. Consequently, this measure may induce a larger risk exposure than the variance in case of down markets.

Expected Shortfall

To cope with these shortcomings, Artzner et al. (1999) and Basak and Shapiro (2001) propose the use of the so-called Expected Shortfall (ES) as an alternative measure of risk. Notice that the terminology is still not clearly established, because Conditional Expected Loss, Conditional VaR, or Tail Conditional Expectations are very closely related notions. ES is the expected value of the loss of the portfolio in the $\theta\%$ worst cases over a given time horizon

$$\overline{ES}_{\theta,t} = -E\left[\Delta W_{t+1}(\alpha_t)\,|\Delta W_{t+1}(\alpha_t) \leq -\overline{VaR}_{\theta,t}\right].$$

Alternatively, we have[2]

$$\begin{aligned} ES_{\theta,t} &= \frac{\overline{ES}_{\theta,t}}{W_t(\alpha_t)} = -E\left[\frac{\Delta W_{t+1}(\alpha_t)}{W_t(\alpha_t)}\,\middle|\,\frac{\Delta W_{t+1}(\alpha_t)}{W_t(\alpha_t)} \leq -VaR_{\theta,t}\right] \\ &= -E\left[r_{p,t}|r_{p,t} \leq -VaR_{\theta,t}\right], \end{aligned}$$

where $ES_{\theta,t}$ denotes the ES for probability θ for 1\$ invested.

The main advantages of the ES for asset allocation are the following: (i) ES is a coherent measure of risk, because it satisfies the sub-additivity property and consequently can be reduced by diversification; (ii) ES directly controls the risk in the left tail of the distribution, so that extreme losses are explicitly taken into account in the allocation process.

An additional advantage of ES over VaR from an asset allocation point of view is that portfolio optimization is easier to implement with ES objectives than with VaR objectives. The reason is that ES is convex, so that the problem can be solved by linear programming techniques, once the *cdf* has been approximated by its empirical counterpart (see Rockafellar and Uryasev, 2000, Krokhmal, Palmquist, and Uryasev, 2002, Rockafellar and Uryasev, 2002). Fermanian and Scaillet (2005) discuss how the VaR and the ES of a portfolio vary when the weights of the portfolio are slightly altered. See Section 9.2 for additional details on the asset allocation under expected shortfall constraints.

---

[2] Rigorously, the ES is defined as (see Acerbi and Tasche, 2001)

$$ES_{\theta,t} = -\frac{1}{\theta}\left(E\left[r_{p,t}|r_{p,t} \leq -VaR_{\theta,t}\right] - VaR_{\theta,t}\Pr\left[r_{p,t} \leq -VaR_{\theta,t}\right] - \theta\right).$$

8.1.2 Models for portfolio returns

To summarize, computing the VaR or ES of a portfolio requires the following elements:

- the probability θ,
- the horizon of the investment k,
- the value $W_t(\alpha_t)$ of the portfolio at date t,
- the *cdf* of the portfolio return.

The first three elements are given in practice, and the main task consists in estimating the *cdf* of the portfolio return.

As argued before, there is a huge literature on how to compute the VaR of a portfolio. One reason is probably that VaR involves several dimensions that can be dealt with using completely different approaches.

A first issue is the *aggregation level*. For a mere measure of the VaR, using a time series of portfolio returns is in general enough. In contrast, if we are interested in active portfolio management, it is more appropriate to evaluate the VaR for asset returns.[3] In such a case, however, we face a problem of dimensionality, because actual portfolios may include several hundreds of assets.

A second issue is the *choice of the model* to estimate the conditional distribution of portfolio returns. For instance, the VaR is a high quantile. So it may be natural to estimate it using an approach that specifically focuses on the tails of the distribution (such as the EVT). But at the same time, it is known that the distribution of returns varies over time, in particular because of changes in volatility. Therefore, it is also of importance to correctly describe how the return distribution evolves through time. There are four broad categories of VaR models that we will discuss below: Historical simulation, semi-parametric models (such as Extreme Value Theory and CAViaR), parametric models (such as RiskMetrics and GARCH models), and finally non-linear techniques that are designed to compute the VaR in presence of derivatives.

Another issue is the fact that the portfolio value may be non-linearly affected by changes in asset prices, for instance when derivative assets are included in the portfolio. The difficulty is that in many cases, historical data is not available. In such cases, non-linear methods, based on Taylor's approximation of the portfolio value or on Monte Carlo simulations, may be used. See Section 8.5.

For large-scale portfolios, such as those managed in financial institutions, simultaneously modeling the joint dynamic of all asset returns may simply be impossible. In such cases, it is preferable to work with a reduced number of base assets (or risk factors) that are thought to drive risks. Adding risk factors

[3] The portfolio-level approach is not appropriate for measuring the effect on the VaR of a change in the portfolio weights.

is analytically more demanding, because it requires an additional layer in the modeling step, but it is likely to reduce the eventual computational burden very significantly. Obviously, in such an approach, the main tasks consist in identifying the factors that may capture the various sources of risk and then in modeling their joint dynamics.

8.2 Historical simulation

Historical simulation is probably the simplest and most widely used approach for computing VaR and ES. It is fundamentally non-parametric, in the sense that it does not require any assumption about the distribution of returns. The method works as follows: Assume that a sample of T past realizations $\{r_1, \cdots, r_T\}$ is available. We define a window size, N, that is used to construct subsamples of size N. Then, $T-N+1$ overlapping subsamples $\{r_1, \cdots, r_N\}$, $\cdots$, $\{r_{T-N+1}, \cdots, r_T\}$ are available. Each of these subsamples is used to approximate the *cdf* of the series. To do this, take one of these subsamples, for instance the tth one, $\{r_{t-N+1}, \cdots, r_t\}$, and sort this subsample in increasing order. Define the sorted data as $\{\tilde{r}_{t-N+1,t}, \cdots, \tilde{r}_{t,t}\}$ where $\tilde{r}_{t-N+1,t} \leq \cdots \leq \tilde{r}_{t,t}$. Now, the VaR with probability $\theta\%$ is defined as the θ-quantile of the subsample. This is therefore the (θN)th order statistic $\tilde{r}_{\theta N,t}$ if θN is an integer. If θN is not an integer, the quantile is defined using a linear interpolation between $\tilde{r}_{\lfloor \theta N \rfloor,t}$ and $\tilde{r}_{\lfloor \theta N \rfloor+1,t}$. To simplify, assume that θN is an integer and that $\tilde{r}_{\theta N,t}$ is the θ-quantile of the subsample. Then, the VaR at date t for date $t+1$ is[4]

$$VaR_{\theta,t} = -\tilde{r}_{\theta N,t}.$$

Finally, the ES is defined as the average of the realizations that are below this level

$$ES_{\theta,t} = -\frac{1}{\lfloor \theta N \rfloor} \sum_{i=1}^{\lfloor \theta N \rfloor} \tilde{r}_{i,t}.$$

The method has two main advantages: First, it is very easy to implement. Second, it allows for non-normal returns. Indeed, it does account for fat tails. Since it uses the actual realizations of returns, it is able to capture most of the empirical features of this series.

However, this approach also raises several difficulties. First, although the historical simulation approach is said to be non-parametric, it is based on a strong underlying assumption: the return process is supposed to be *iid*. Empirical evidence clearly suggests that this is not the case. For this reason, the choice of the window size N is crucial in practice. Longer samples increase the accuracy of VaR estimates, but at the same time increase the probability of using irrelevant data, in particular if there are some changes in the underlying

[4] Given that small quantiles of returns are negative, the VaR measure will be positive.

Table 8.1. *Five smallest and largest daily returns*

SP500	min	−22.87	−8.68	−7.15	−7.08	−7.04
	max	8.67	5.54	5.23	5.09	4.95
DAX	min	−13.74	−9.90	−9.89	−8.91	−8.41
	max	7.52	7.26	7.14	7.05	6.95
FT-SE	min	−11.95	−10.43	−7.46	−5.55	−5.39
	max	5.66	5.63	5.06	4.68	4.61
Nikkei	min	−16.14	−7.24	−6.87	−6.83	−6.14
	max	12.42	8.88	7.65	7.54	7.27

process. On the other hand, it would be meaningless to use a small window size if we are interested in a very small probability θ, because the historical simulation approach cannot produce a quantile that would be smaller than the minimum return observed within the given sample. If we want to compute the 1%-quantile, we need at least 100 observations in the subsample! Thus there is a trade-off between longer and shorter sample sizes.

Another undesirable feature is that the VaR obtained with this method does not vary often but when it does, it varies sharply. In fact, the only source of variation in this approach is the shift of the window over time. As a consequence, we observe jumps in the reported VaR, when an extreme (negative) return is introduced or dropped from the subsample. This problem is mainly due to the discreteness of the empirical distribution of returns. In particular, in the tails, the interval between adjacent returns can be rather large, as Table 8.1 illustrates. We display the 5 smallest and largest daily returns on the four indices at hand over the period from 1980 to 2004.

The problem is also illustrated in Figures 8.1 and 8.2, where we display the 1% and 5% VaR and ES of a portfolio constituted of the SP500, the DAX, the FT-SE, and the Nikkei (with equal weights). Since the window size is $N = 500$, they correspond to the 5th and the 25th order statistics of each subsample. One explanation for this pattern is that all realizations in the subsample have the same weight. To partially cope with this problem, we may, for instance, introduce declining weights to past observations to smooth computations.

8.3 Semi-parametric approaches

The problem with the historical simulation approach stems from the fact that returns are not *iid* and from the discreteness of the tails of the empirical distribution. The two semi-parametric approaches presented now put some structure on the tails of the distribution by using parametric models but without estimating the complete distribution of returns. To some extent, they are also able to cope with the temporal dependence of returns. The first one is based on the EVT presented in detail in Section 7.1.4. The second approach is

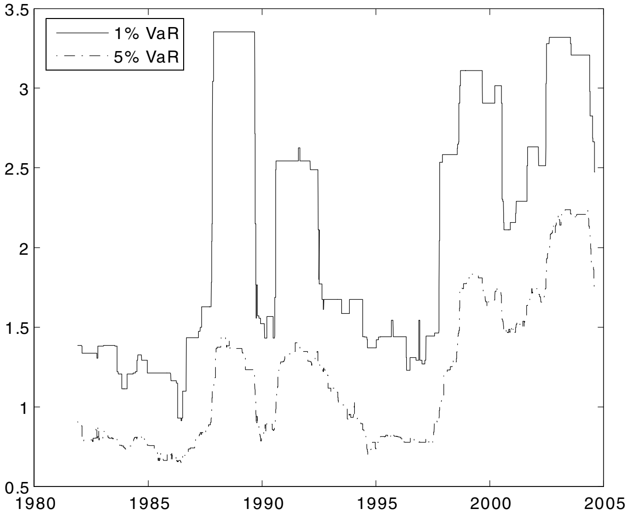

Fig. 8.1. *VaR of portfolio computed with the historical simulation approach.*

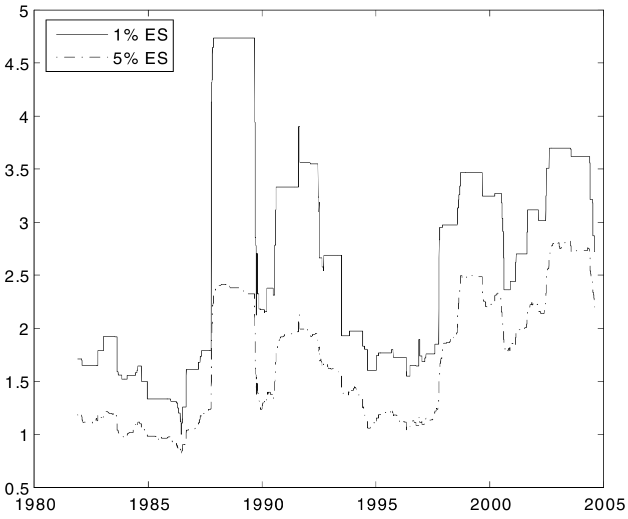

Fig. 8.2. *ES of portfolio computed with the historical simulation approach.*

based on a regression quantile technique as developed by Engle and Manganelli (2004) and Chernozhukov and Umantsev (2001).

8.3.1 Extreme Value Theory (EVT)

The EVT is designed to model the specific behavior of very large (positive or negative) returns. It provides a parametric representation of the distribution of the extremes (or of the tails). This representation is able to give smoothed estimates of the VaR and is designed to produce very high quantiles, possibly smaller than the minimum of the sample distribution. Such an approach has been adopted for VaR evaluation by several authors: McNeil (1997) estimates a gpd to evaluate the tail index; Danielsson and de Vries (1997) uses a semi-parametric estimate of the tail index; McNeil and Frey (2000) introduce a GARCH–EVT model that incorporates the temporal evolution of volatility; Longin (2000) adopts a multivariate approach to capture the joint extreme behavior of risk factors.

In Section 7.1.4, we described how the quantile of a univariate distribution can be computed using the various approaches developed in the context of the EVT: the distribution of extremes, the distribution of the tails or the semi-parametric estimation of the tail index. We briefly recall the estimation of high quantiles in the case of the tail approach.

Unconditional EVT

The main idea of the tail approach to EVT is that the distribution of the lower tail (i.e., the returns that are below a given threshold u) can be approximated, when u is sufficiently large (in absolute value), by the so-called generalized Pareto distribution (gpd). The gpd is defined as

$$G_{\xi,u,\psi}(r) = \begin{cases} 1 - \left(1 + \frac{\xi}{\psi}(r-u)\right)^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1 - \exp\left(-\frac{(r-u)}{\psi}\right), & \text{if } \xi = 0, \end{cases}$$

where the tail index ξ characterizes the shape of the tails of the distribution and ψ is a scaling parameter.

Assume that the estimation of the gpd is performed on the absolute value of the lower tail observations, as detailed in Section 7.1.2. Once the parameters of this gpd are estimated, we can deduce the θ-quantile of the actual distribution as

$$q_\theta = \begin{cases} u + \frac{-\hat{\psi}}{\hat{\xi}}\left(\left(\frac{T}{N_u}\theta\right)^{-\hat{\xi}} - 1\right), & \text{if } \xi \neq 0, \\ u + \hat{\psi}\log\left(\frac{T}{N_u}\theta\right), & \text{if } \xi = 0, \end{cases}$$

where N_u is the number of exceedances below the threshold. Evidently, since it is computed using results that are valid for extreme returns only, the quantile

q_θ is valid for very small probability θ. Then, the VaR at date t for date $t+1$ is

$$VaR_{\theta,t} = -q_\theta,$$

so that the estimated VaR is actually constant over time.

As outlined by McNeil and Frey (2000), the ES is strongly related to the notion of mean excess function, because we have the relation

$$\begin{aligned} ES_{\theta,t} &= E\left[-r \mid -r \geq VaR_{\theta,t}\right] \\ &= VaR_{\theta,t} + E\left[-r - VaR_{\theta,t} \mid -r \geq VaR_{\theta,t}\right]. \end{aligned}$$

Recall that, if $r - u | r > u \sim G_{\xi,\psi}$, the mean excess function above level u is defined as

$$e(u) = E\left[-r - u \mid -r \geq u\right] = \frac{\psi + \xi u}{1-\xi}, \qquad \psi + \xi u > 0.$$

Therefore, the expected shortfall is equal to the sum of the VaR and the mean excess funcion above the VaR. Now, since $VaR_{\theta,t} > u$, we may write

$$\begin{aligned} &E\left[-r - VaR_{\theta,t} \mid -r \geq VaR_{\theta,t}\right] \\ &\qquad = E\left[(-r-u) - (VaR_{\theta,t} - u) \mid (-r-u) \geq (VaR_{\theta,t} - u)\right], \end{aligned}$$

with $-r - VaR_{\theta,t} | -r \geq VaR_{\theta,t} \sim G_{\xi,\psi+\xi(VaR_{\theta,t}-u)}$. Finally, we deduce the estimated ES

$$ES_{\theta,t} = VaR_{\theta,t} + \frac{\hat{\psi} + \hat{\xi}\left(VaR_{\theta,t} - u\right)}{1-\hat{\xi}} = \frac{VaR_{\theta,t}}{1-\hat{\xi}} + \frac{\hat{\psi} - \hat{\xi} u}{1-\hat{\xi}}.$$

This approach is called unconditional EVT, because it is based on the assumption that returns are *iid*. It therefore produces an unconditional VaR, i.e., pertaining to the unconditional distribution of returns. This implies that the quantile q_θ will not vary over time, even if there is a sudden change in market conditions. Recently, McNeil and Frey (2000) have proposed a model that provides conditional measures of the VaR based on EVT. This approach is described below.

Another drawback of the tail approach is that we have to select a threshold u below which we consider that returns belong to the tail. Although some tools have been proposed to select this threshold (such as bootstrap techniques), it is still an open question how to define the optimal threshold. Figure 8.3 displays the 1%-quantile $q_{0.01}$ as a function of the threshold u for the SP500. This plot suggests that the choice of the threshold does not necessary play a crucial role for the value of the quantile, given that the VaR varies only between 3.77 and 3.87 as u varies between 0 and 0.1.

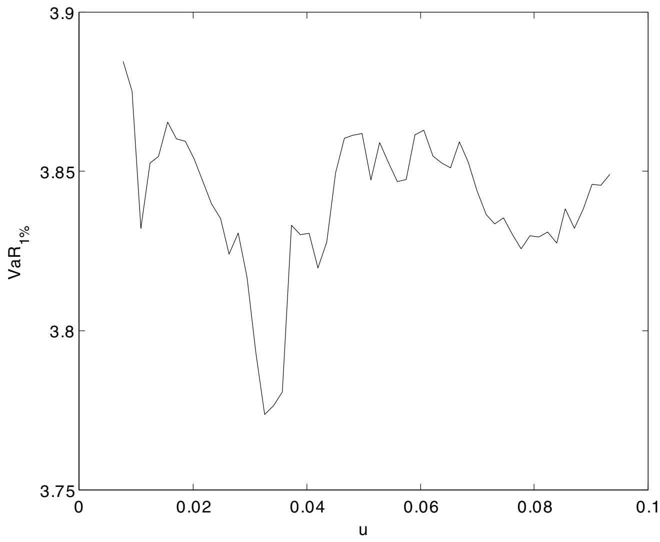

Fig. 8.3. *1%-quantile of the SP500 as a function of the threshold u.*

The GARCH–EVT model

The GARCH–EVT model proposed by McNeil and Frey (2000) consists in modeling the conditional volatility and the distribution of the tails separately. In fact, only the left tail is needed, because the right tail is not relevant for VaR computation.

For this purpose, McNeil and Frey (2000) proceed as follows: In the first step, they filter the dependence in the return series by computing the residuals of a GARCH model, which should be *iid* if the GARCH model correctly fits the data. In the second step, they model the extreme behavior of the residual using the tail approach developed in Section 7.1.2. Finally, in order to produce a VaR estimate for the original return, they trace back the steps by first producing the θ-quantile estimate for the GARCH-filtered residuals and convert the θ-quantile estimate to the original return using the conditional volatility forecast for the required horizon.

For a given return series $\{r_1, \cdots, r_T\}$, the model adopted to filter out the first- and second-order dynamics is of the form

$$\begin{aligned} r_t &= \mu_t + \varepsilon_t, \\ \varepsilon_t &= \sigma_t z_t, \\ \mu_t &= \mu + \varphi r_{t-1}, \\ \sigma_t^2 &= \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned}$$

with $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, and $\alpha + \beta < 1$ to ensure a positive volatility and a covariance stationary process. To avoid specifying an arbitrary distribution for the innovation process z_t, they resort to the Quasi Maximum Likelihood (QML) estimation, which consists in maximizing the normal log-likelihood of the model even though the true generating process of z_t is not Gaussian. This technique has been shown to provide consistent estimates of the model parameters, provided the conditional mean and variance equations are correctly specified (Gouriéroux, Monfort, and Trognon, 1984). The standardized residual $\hat{z}_t$ is then estimated by $\hat{z}_t = (r_t - \hat{\mu}_t)/\hat{\sigma}_t$, where $\hat{\mu}_t$ and $\hat{\sigma}_t^2$ are the fitted mean and variance, respectively. We also use this model to produce forecasts of the expected return $\mu_t(1) = \hat{\mu} + \hat{\varphi} r_t$ and variance $\sigma_t^2(1) = \hat{\omega} + \hat{\alpha}\hat{\varepsilon}_t^2 + \hat{\beta}\hat{\sigma}_t^2$.

The next step involves estimating the gpd to all exceedances, i.e., all realizations $\hat{z}_t$ that are below a given high threshold u.[5] Adopting a similar approach to the one in Section 7.1.2, we define $N_u = \sum_{t=1}^{T} 1_{\{\hat{z}_t < u\}}(\hat{z}_t)$ the number of exceedances and $\{\hat{z}_{1,T}, \cdots, \hat{z}_{T,T}\}$ the vector of standardized residuals sorted by increasing order, such that $\hat{z}_{1,T} \leq \cdots \leq \hat{z}_{T,T}$. Finally, we define the $(N_u, 1)$ vector of exceedances as $\{\hat{z}_{1,T}, \cdots, \hat{z}_{N_u,T}\}$. Then, we estimate the parameters $(\xi, \psi)'$ of the gpd to the exceedances $\{\hat{z}_{i,T}\}_{i=1}^{N_u}$.

Once the parameters ξ and ψ are estimated, the θ-quantile is obtained by inverting the *cdf* of exceedances

$$q_\theta = \begin{cases} u + \frac{-\hat{\psi}}{\hat{\xi}} \left(\left(\frac{T}{N_u} \theta \right)^{-\hat{\xi}} - 1 \right), & \text{if } \xi \neq 0, \\ u + \hat{\psi} \log \left(\frac{T}{N_u} \theta \right), & \text{if } \xi = 0. \end{cases}$$

We are now ready to evaluate the aggregate VaR and ES from t to $t + 1$

$$VaR_{\theta,t} = -\left(\mu_t(1) + q_\theta \sigma_t(1)\right),$$
$$ES_{\theta,t} = \frac{VaR_{\theta,t}}{1 - \hat{\xi}} + \frac{\hat{\psi} - \hat{\xi} u}{1 - \hat{\xi}}.$$

It is worth emphasizing that the GARCH–EVT approach incorporates the two ingredients required for an accurate evaluation of the conditional VaR, i.e., a model for the dynamics of the first and second moments, and an appropriate model for the conditional distribution. An obvious improvement of this approach as compared to the unconditional EVT is that it incorporates in the VaR changes in expected return and in volatility. For instance, if we assume a change in volatility over the recent period, the GARCH–EVT is able to incorporate this new feature in its VaR evaluation, whereas the unconditional EVT remains stuck at the average level of volatility over the estimation sample.

McNeil and Frey (2000) also provide a back-testing experiment, in which they compare the performances of various methods to correctly reproduce

[5] Notice that we focus in this section on the lower tail, whereas all developments in Chapter 7 were based on the upper tail.

the quantiles of several asset returns. They show that the GARCH–EVT performs much better than the unconditional EVT, suggesting that the ability to capture changes in volatility is crucial for VaR computation.

Multivariate EVT

The VaR computation described above is useful for a single asset or for cases where there is only one risk factor. The cases for multi-asset and multi-risk-factor are a lot more complex, because they resort to multivariate EVT. We have shown in Section 7.2 how to characterize and measure extremal dependence. The measures introduced are based on the evaluation of the properties of the joint distribution at equal probability marginal quantiles. To use these measures to characterize extremes of a portfolio, i.e., of a linear combination of assets, the full joint distribution needs to be estimated in the tail region. For the case where returns are asymptotically dependent or exactly independent, such methods exist, see Coles and Tawn (1994), and de Haan and de Ronde (1998). Poon, Rockinger, and Tawn (2004) follow the approach of Ledford and Tawn (1997) for handling asymptotic independence via non-parametric and parametric approaches. VaR and ES computations rely on Monte Carlo simulations.

8.3.2 Quantile regression technique

In the quantile regression approach, another route is taken. Instead of focusing on the modeling of the tail distribution as in the EVT, this approach focuses on its dynamic component. The technique of quantile regression has been introduced in the statistical literature by Koenker and Bassett (1978).[6] It has been applied to VaR computation by Engle and Manganelli (2004) and Chernozhukov and Umantsev (2001). The basic idea consists in modeling a given quantile of the distribution through time. Such an approach is justified by that empirical evidence that volatility tends to cluster, so that the distribution itself is serially correlated. Engle and Manganelli (2004) described a new regression quantile model, called CAViaR, standing for Conditional AutoRegressive Value-at-Risk, which allows modeling the dynamics of the VaR. In this section, we start by reviewing the concept of quantile regression before turning to the one of CAViaR.

Quantile regression

The θth quantile q_θ of a time series $\{y_t\}_{t=1}^{T}$ can be defined as the solution of the following minimization problem[7]

---

[6] A general presentation can be found in Koenker and Hallock (2001).

[7] A well-known special case is the median ($q = 1/2$) in which the median β is defined as

$$\min_{q\in\mathbb{R}} \sum_{t\in\{t:y_t\geq q\}} \theta\left|y_t - q\right| + \sum_{t\in\{t:y_t<q\}} (1-\theta)\left|y_t - q\right|, \qquad 0<\theta<1,$$

or, equivalently

$$\min_{q\in\mathbb{R}} \sum_{t=1}^{T} w_\theta\left(y_t - q\right),$$

where

$$w_\theta\left(z_t\right) = \begin{cases} \theta z_t, & \text{if } z_t \geq 0, \\ (1-\theta)\, z_t, & \text{if } z_t < 0. \end{cases}$$

The estimated q is the unconditional θ-quantile of $\{y_t\}_{t=1}^{T}$. Now, in a regression context, $y_t = x_t'\beta + u_t$, where x_t is a $(k,1)$ vector of regressors and u_t is the error term with *cdf* F, the θth regression quantile of $u_t = y_t - x_t'\beta$ conditional on x_t is obtained as the solution of the minimization problem

$$\min_{\beta\in\mathbb{R}^k} \sum_{t\in\{t:y_t\geq x_t'\beta\}} \theta\left|y_t - x_t'\beta\right| + \sum_{t\in\{t:y_t<x_t'\beta\}} (1-\theta)\left|y_t - x_t'\beta\right|, \qquad 0<\theta<1. \tag{8.1}$$

The purpose of this regression is to find the vector of parameters β that will ensure that the θ-quantile of u_t will be as close to 0 as possible. A well-known quantile regression is the one associated with $\theta = 1/2$. In this case, β is optimized in order to obtain a median equal to 0.

The estimation technique and the asymptotic properties of the estimator of β are developed in Koenker and Bassett (1978). This class of *robust estimators* includes some that have similar efficiency to the least-square estimator under normality of the error term but that out-perform this estimator under non-normal errors.

Koenker and Bassett (1978) also provide some useful asymptotic results on regression quantiles. In particular, the solution $\hat{\beta}_T(\theta)$ of the optimization problem 8.1 for quantile θ is shown to be consistent and asymptotically normal. This result is extended by Engle and Manganelli (2004) to the case of non-linear regressions.

CAViaR

Engle and Manganelli (2004) proposed a model based on quantile regressions, called CAViaR. The idea is to directly model the evolution of the quantile q_θ instead of considering the entire distribution of returns. The general specification they propose has the form

$$q_{\theta,t} = \beta_0 + \beta_1 q_{\theta,t-1} + \sum_{j=1}^{r} \beta_{j+1} g\left(x_{t-j}\right),$$

$$\min_{\beta\in\mathbb{R}} \sum_{t=1}^{T} \left|y_t - \beta\right|.$$

where x_t is a set of regressors and $g(.)$ a possibly non-linear function of the regressors. Some of the suggested specifications are particularly appealing:

Symmetric absolute value model

$$q_{\theta,t} = \beta_0 + \beta_1 q_{\theta,t-1} + \beta_2 \left| r_{t-1} \right|.$$

Asymmetric slope model

$$q_{\theta,t} = \beta_0 + \beta_1 q_{\theta,t-1} + \beta_2 1_{\{r_{t-1}>0\}} + \beta_3 1_{\{r_{t-1}<0\}}.$$

Indirect GARCH model

$$q_{\theta,t}^2 = \beta_0 + \beta_1 q_{\theta,t-1}^2 + \beta_2 r_{t-1}^2.$$

The parameters of the quantile regressions are then estimated using non-linear techniques to solve the optimization problem

$$\min_{\beta \in \mathbb{R}^k} \sum_{t \in \{t: y_t \geq q_{\theta,t}(\beta)\}} \theta \left| y_t - q_{\theta,t}(\beta) \right| + \sum_{t \in \{t: y_t < q_{\theta,t}(\beta)\}} (1-\theta) \left| y_t - q_{\theta,t}(\beta) \right|$$

with $0 < \theta < 1$. This approach is qualified as semi-parametric, because it does not specify the distribution of returns. It can therefore be applied to non-*iid* returns as well as to time-varying volatility. It should be mentioned that the estimation of such a model is far from trivial. Engle and Manganelli (2004) also propose some test procedures for evaluating VaR models. This issue is addressed in detail in Section 8.6.1.

8.4 Parametric approaches

While the historical simulation and the unconditional EVT approaches are by nature unable to capture changes in the behavior of returns, the GARCH–EVT technique and CAViaR incorporates some stylized facts of returns concerning the first- and second-order dynamics. We now turn to parametric models of VaR that hold under a complete set of assumptions concerning the dynamics and the conditional distribution of returns.[8]

We recall here that the main empirical features for asset returns are the following:

1. Returns may be serially correlated, even if this correlation is not very large in practice.
2. Return volatility is serially correlated and possibly asymmetric. These features can be captured by the well-known GARCH family.
3. The conditional distribution of returns is probably non-normal. Typical characteristics of this distribution are skewness and fat-tailedness.

[8] Evidently, these assumptions can sometimes be excessively strong.

These three components have not been introduced systematically in previous approaches. For instance, the widely used RiskMetrics methodology assumes normality of returns. As long as such an assumption yields correct forecasts of VaR and ES, it may be relevant. A thorough test of the implications of the assumptions and a test for the validity of the model appears as crucial.

8.4.1 RiskMetrics – J.P. Morgan

RiskMetrics is a methodology developed by J.P. Morgan to compute VaR. It has played an important role in the increasing popularity of VaR as a risk measure. See J.P. Morgan's RiskMetrics Technical Document (1996).

The RiskMetrics methodology uses historical return data to forecast future volatility. More precisely, the basic RiskMetrics model is based on the following assumptions:

1. The return r_t is modeled as $r_t = \mu_t + \varepsilon_t$ with $\varepsilon_t = \sigma_t z_t$.
2. Daily log-returns are supposed to be centered (or are preliminary demeaned), so that $\mu_t = 0$.
3. The dynamic of volatility is modeled using an exponentially weighted moving average (EWMA), with

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1-\lambda) r_{t-1}^2, \qquad \text{for } t = 2, \cdots, T, \tag{8.2}$$

   with $0 < \lambda < 1$. This model may be viewed as a special case of the Integrated GARCH model. In the first versions of RiskMetrics, the decay factor λ was chosen to be equal to 0.94. The recursion can be initialized by the sample variance ($\sigma_1^2 = \hat{\sigma}^2$) or by the square of the first observation ($\sigma_1^2 = r_1^2$).
4. The innovation z_t is supposed to be distributed as an *iid* $\mathcal{N}(0,1)$.

Under these various assumptions, the conditional distribution of r_t at date t is $\mathcal{N}\left(0, \sigma_t^2\right)$. If we are interested in the one-step-ahead VaR, we need the conditional distribution of r_{t+1}. Conditionally on the information at time t, it is $\mathcal{N}\left(0, \sigma_t^2(1)\right)$ where $\sigma_t^2(1) = \lambda \sigma_t^2 + (1-\lambda) r_t^2$. For the $\mathcal{N}(0,1)$ distribution, we denote $q_\theta = \Phi^{-1}(\theta)$ the quantile for a probability of loss equal to θ. For instance, for $\theta = 1\%$, we have $q_\theta = -2.326$. Now, the daily VaR is given by

$$VaR_{\theta,t} = -q_\theta \times \sigma_t(1). \tag{8.3}$$

In addition, we have the following expression for the ES

$$\begin{aligned} ES_{\theta,t} &= -E[r_t | r_t \leq -VaR_{\theta,t}] \\ &= -E\left[\frac{r_t}{\sigma_t(1)} | \frac{r_t}{\sigma_t(1)} \leq \frac{-VaR_{\theta,t}}{\sigma_t(1)}\right] \times \sigma_t(1) \\ &= ES_{\theta,t}^z \times \sigma_t(1) \end{aligned}$$

where

$$ES_{\theta,t}^{z} = -\frac{1}{\Phi(q_\theta)} \int_{-\infty}^{-q_\theta} z \times \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) dz = \frac{1}{\theta}\varphi(q_\theta)$$

where $\varphi(z)$ denotes the *pdf* of the $\mathcal{N}(0,1)$ distribution. Finally, we obtain

$$ES_{\theta,t} = \frac{\varphi(q_\theta)}{\theta}\sigma_t(1).$$

Multi-period VaR

One interesting property of the RiskMetrics approach is that a multi-period VaR can be very easily computed. Assume we have to compute the VaR over the following k periods. We observe that the log-return between t and $t+k$ is simply defined as $r_t[k] = r_{t+1} + \cdots + r_{t+k}$. We deduce that the volatility forecast for the k-period log-return is

$$\sigma_t^2[k] = V_t[r_t[k]] = V_t\left[\sum_{i=1}^{k} r_{t+i}\right].$$

As we have seen in Section 4.3, for an IGARCH model, the κ-step ahead volatility forecast is

$$\sigma_t^2(\kappa) = \cdots = \sigma_t^2(2) = \sigma_t^2(1) = \lambda\sigma_t^2 + (1-\lambda)r_t^2, \qquad \text{for } \kappa \geq 1,$$

because the innovation process is *iid*. Therefore, the volatility forecast for the k-period log-return is

$$\sigma_t^2[k] = k\sigma_t^2(1),$$

so that the volatility forecast of $r_t[k]$ is proportional to the horizon k. We finally obtain that the k-day VaR is given by

$$VaR_{\theta,t:t+k} = -q_\theta \times \sqrt{k}\sigma_t(1).$$

This expression is known as the square-root-of-time rule. This rule has been recommended by the Basle Committee's Amendment to the Capital Accord to Incorporate Market Risks (1996) in order to compute the 10-day VaR from daily estimates.

Multiple position

Above, we computed the VaR of a single asset, a case that is not very interesting in practical applications. Notice that we may alternatively view r_t as the return of the portfolio, not as the return of an asset. In such a case, consistently with the portfolio-level approach, the computations correspond to the aggregate VaR. Notice that this interpretation also raises some new difficulties. Since the portfolios of financial institutions are likely to change daily, it implies that each day, the risk manager would have to compute a historical time series of the new portfolio and to estimate the aggregate VaR using this approach.

Another way, consistent with the asset-level approach, has been suggested for the latter. It is based on the observation that the aggregate VaR can be rewritten as

$$VaR_{\theta,t} = -q_\theta \times \sigma_{p,t}(1) = -q_\theta \times \sqrt{\alpha' \Sigma_t(1)\alpha},$$

where $\Sigma_t(1)$ is the one-step-ahead forecast of the covariance matrix of asset returns. It is computed assuming that all variances and covariances are driven by the same model (8.2)

$$\begin{aligned} \sigma_{i,t}^2 &= \lambda \sigma_{i,t-1}^2 + (1-\lambda) r_{i,t-1}^2, && \text{for } i = 1, \cdots, n, \\ \sigma_{ij,t} &= \lambda \sigma_{ij,t-1} + (1-\lambda) r_{i,t-1} r_{j,t-1}, && \text{for } i, j = 1, \cdots, n, \end{aligned}$$

using the same parameter λ for all assets.

It is clear that such an approach is equivalent to computing the VaR of each asset in the portfolio and then deducing the VaR of the portfolio. To see this, we denote $VaR_{\theta,t}^i$ the VaR of asset $i = 1, \cdots, p$. Then, we observe that, for a given asset, we have the relation between variance and VaR

$$\sigma_{i,t}^2(1) = \left(\frac{1}{q_\theta} VaR_{\theta,t}^i\right)^2,$$

so that the VaR of the portfolio is now

$$\begin{aligned} VaR_{\theta,t} &= -q_\theta \sqrt{\alpha' \Sigma_t(1)\alpha} \\ &= -q_\theta \sqrt{\sum_{i=1}^n \alpha_i^2 \sigma_{i,t}^2(1) + 2\sum_{i=1}^n \sum_{j=i+1}^n \alpha_i \alpha_j \sigma_{ij,t}(1)}. \end{aligned}$$

If we use $\sigma_{ij,t}(1) = \sigma_{i,t}(1)\sigma_{j,t}(1)\rho_{ij,t}(1)$ where $\rho_{ij,t}(1)$ denotes the one-period forecast of the correlation between assets i and j, we obtain after simplification (see, for instance, Longin, 2000)

$$(VaR_{\theta,t})^2 = \sum_{i=1}^n \alpha_i^2 \left(VaR_{\theta,t}^i\right)^2 + 2\sum_{i=1}^n \sum_{j=i+1}^n \rho_{ij,t}(1)\alpha_i\alpha_j VaR_{\theta,t}^i VaR_{\theta,t}^j.$$

The squared VaR of the portfolio is simply the quadratic form of the VaR of returns, weighted by the correlation matrix between returns. Computing the VaR in this way does not even require the knowledge of the portfolio weights.

Advantages and limitations

The main advantage of RiskMetrics is its simplicity of implementation. If we are willing to accept the value of the decay factor adopted by RiskMetrics, no estimation is needed, and the update of the VaR of any portfolio is extremely fast. In addition, computing multi-period VaR or multi-position VaR does not raise any additional difficulty. Obviously, this simplicity has a cost. Some of the underlying assumptions are in fact overly strong.

First, the assumption of normality of innovations is simply untenable. Most asset returns are characterized by a distribution with fat tails and/or asymmetry. Such assumption will lead in general to an underestimation (in absolute value) of the quantile q_θ to be used in the VaR formula.

Second, the dynamics of volatility is too simplistic. Although the IGARCH model has the advantage that the multi-period VaR can be computing using the simple square-root-of-time rule, it also yields some undesirable properties, such as the lack of mean-reversion in the variance process. This issue has been investigated by Diebold et al. (1998). They show that the square-root-of-time rule produces overestimates of the variability of long-horizon volatility.

A last, less stringent, assumption is the absence of dynamics for the expected return. This is probably too strong an assumption, but in practice, the effect of altering the conditional mean equation on the VaR measures is barely noticeable. The reason is that for most asset returns, the VaR computation is largely dominated by volatility, rather than expected return, considerations.

8.4.2 The portfolio-level approach

The variance-covariance method, also known as the correlation method, is essentially a parametric approach in which the VaR is measured from the variances and covariances of the constituents of a portfolio. A simple version of this approach is the RiskMetrics methodology. The main task of this approach is to model the different components of the dynamic of returns that are useful for VaR computation. As seen above, these components are: The dynamic of expected returns, the dynamic of volatility and finally the conditional distribution of the innovation process.

We assume that returns are possibly autocorrelated, that volatility σ_t is driven by a GARCH$(1,1)$ model and that innovations z_t are distributed as an *iid* $\mathcal{N}(0,1)$. The conditional distribution of r_t at date t is then $\mathcal{N}\left(\mu_t, \sigma_t^2\right)$. The main change as compared with the RiskMetrics method is a conceptual one: the parameters of this model have to be estimated, rather than calibrated. This estimation is more demanding in terms of computational burden, in particular if a set of asset returns must be modeled.

The choice of a GARCH$(1,1)$ model with parameters α and β estimated without the assumption $\alpha + \beta = 1$ is very likely to provide a more realistic mean-reverting behavior of volatility forecasts. The 1-step ahead forecasts for μ_{t+1} and σ_{t+1}^2 are

$$\mu_t(1) = \mu + \varphi_1 r_t,$$
$$\sigma_t^2(1) = \omega + \alpha\varepsilon_t^2 + \beta\sigma_t^2.$$

Conditionally on the information at time t, the conditional distribution of r_{t+1} is $\mathcal{N}\left(\mu_t(1), \sigma_t^2(1)\right)$. The daily VaR is given by

$$VaR_{\theta,t} = -\left(\mu_t(1) + q_\theta \times \sigma_t(1)\right),$$

and the ES is

$$\begin{aligned} ES_{\theta,t} &= \frac{1}{\theta}\varphi\left(\frac{-VaR_{\theta,t} - \mu_t(1)}{\sigma_t(1)}\right)\sigma_t(1) - \mu_t(1) \\ &= \frac{\varphi(q_\theta)}{\theta}\sigma_t(1) - \mu_t(1). \end{aligned} \tag{8.4}$$

Conditional distribution

Although the model above is able to capture the dynamic in expected returns as well as in volatility, the conditional distribution is still assumed to be normal. Empirical evidence suggests that distributions allowing fat tails and asymmetry should be used for modeling the innovations. In Chapter 5, we described several alternative distributions that may be used in such a context. For instance, the Student t distribution may be very easily used in place of the normal distribution, because procedures to compute the inverse of its cdf are available in most econometric software. The skewed Student t distribution also appears as an obvious alternative.[9]

For instance, assume now that the innovation process is drawn from a standardized Student t distribution with ν degrees of freedom. It is worth emphasizing that in the context of a GARCH model, the innovation process z_t is supposed to have zero mean and unit variance. Consequently, the appropriate distribution is not the usual, but the standardized t, defined as

$$t(z_t|\nu) = c\left(1 + \frac{z_t^2}{\nu - 2}\right)^{-\frac{\nu+1}{2}},$$

with $\nu - 2$ in place of ν and $c = \Gamma\left(\frac{\nu+1}{2}\right) / \left(\sqrt{\pi(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)\right)$. Since this is the usual t that is available in most software, the quantiles have to be appropriately corrected.[10] Finally, the θ-VaR for r_{t+1} is given by

$$VaR_{\theta,t} = -\left(\mu_t(1) + \tilde{q}_\theta \times \sigma_t(1)\right),$$

[9] Since the VaR computation only involves the tails of the distribution, the quantiles of the skewed Student t distribution can be computed using the procedure built for the standard Student t, provided the asymmetric component is properly taken into account.

[10] More precisely, if $\breve{z}_q$ is the q-quantile of the usual t distribution with ν degrees of freedom (in general given by econometric software), then the quantile $\tilde{z}_q$ of

where $\tilde{q}_\theta = t_\nu^{-1}(\theta)$ is the quantile for a probability of loss equal to θ from the standardized t distribution. For the ES, we have to evaluate the same expression as for the normal case

$$ES_{\theta,t} = ES_{\theta,t}^{z} \times \sigma_t(1) - \mu_t(1),$$

where

$$\begin{aligned} ES_{\theta,t}^{z} &= -\frac{1}{t_\nu(\tilde{q}_\theta)} \int_{-\infty}^{-\tilde{q}_\theta} z \times c\left(1+\frac{z^2}{\nu-2}\right)^{-\frac{\nu+1}{2}} dz \\ &= \frac{c}{\theta}\frac{\nu-2}{\nu-1}\left(1+\frac{\tilde{q}_\theta^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}, \end{aligned}$$

so that we have eventually

$$ES_{\theta,t} = \frac{c}{\theta}\frac{\nu-2}{\nu-1}\left(1+\frac{\tilde{q}_\theta^2}{\nu-2}\right)^{-\frac{\nu+1}{2}} \times \sigma_t(1) - \mu_t(1). \tag{8.5}$$

Multi-period VaR

Computing the VaR over k periods requires the cumulative expected return and volatility forecasts over k-periods. Using, abusively, the definition of the multi-period log-return, the k-period expected return is

$$\mu_t[k] = \mu_{t+1} + \cdots + \mu_{t+k} = k\mu + \frac{1-\varphi_1^k}{1-\varphi_1}(\mu_t(1) - \mu), \qquad \text{for } k > 1.$$

Moreover, the k-step ahead volatility forecast of a GARCH$(1,1)$ is given by

$$\sigma_t^2(k) = \sigma^2 + (a+b)^{k-1}\left(\sigma_t^2(1) - \sigma^2\right), \qquad \text{for } k > 1.$$

Summing these volatilities, we obtain the volatility forecast for the k-period log-return

$$\sigma_t^2[k] = k\sigma^2 + \frac{1-(a+b)^k}{1-a-b}\left(\sigma_t^2(1) - \sigma^2\right), \qquad \text{for } k > 1.$$

Contrary to what we obtained with RiskMetrics, the volatility forecast of $r_t[k]$ is now mean-reverting, meaning that the dynamic of volatility estimated over the sample plays a role in the forecast process. Finally, the k-day VaR is given by

---

the standardized t distribution is deduced using the relation: $\tilde{z}_q = \sqrt{\frac{\nu-2}{\nu}}\breve{z}_q$. For instance, for $\nu = 5$, the 1%-quantile of the usual t distribution is $\breve{z}_q = -3.3649$, and the corresponding quantile of the standardized t distribution is $\tilde{z}_q = -2.6065$. As it can be noticed, the difference if very significant for small degrees of freedom.

$$VaR_{\theta,t:t+k} = -\left(\mu_t[k] + q_\theta \times \sigma_t[k]\right),$$

where

$$\mu_t[k] = k\mu + \frac{1-\varphi_1^k}{1-\varphi_1}\left(\mu_t(1) - \mu\right), \qquad \text{for } k > 1.$$

This expression for the VaR is less intuitive than the square-root-of-time rule. But it is also more consistent with the observed dynamic of volatility, in that it displays mean-reversion.

8.4.3 The asset-level approach

The asset-level estimation of the aggregate VaR has been advocated as allowing for a better control on the VaR estimation. As already argued, it allows measuring the effect on the aggregate VaR of a change in portfolio weights (the portfolio-level approach would require a complete re-estimation of the model). However, this advantage comes at a cost. Since we are interested in the modeling of the joint dynamic of asset returns, we have to turn to a multivariate GARCH-type model. As we have seen in Section 6.1, this approach raises a dimensionality problem, even for a moderate number of assets. Even though we emphasized the importance of non-Gaussian distributions for returns, the multivariate modeling in a non-Gaussian setting still represents a challenge. For this reason, at the asset-level approach, it is customary to assume a multivariate Gaussian distribution.

Once the multivariate GARCH model is estimated, computing the aggregate VaR is easy, since the distribution of the portfolio return is

$$r_{p,t} \sim \mathcal{N}\left(\mu_{p,t}, \sigma_{p,t}^2\right),$$

with $\mu_{p,t} = \alpha_t' \mu_t$ and $\sigma_{p,t}^2 = \alpha_t' \Sigma_t \alpha_t$. We deduce the aggregate VaR as

$$VaR_{\theta,t} = -\left(\mu_{p,t}(1) + q_\theta \times \sigma_{p,t}(1)\right),$$

where $q_\theta = \Phi^{-1}(\theta)$ is the quantile for a probability of loss equal to θ, from the univariate normal distribution $\mathcal{N}(0,1)$. The ES is given by expression (8.4).

Conditional distribution

The main limitation of this approach is the maintained assumption of multivariate normality. Although the estimation of the DCC model can be performed under normality with a reasonable computational burden, another distributional assumption would dramatically increase the burden for large-dimension portfolios. The reason is that in such case the log-likelihood cannot be broken in separate components anymore. To be more precise, two cases have to be considered:

- For elliptical distributions (such as the Student t distributions) but the Gaussian distribution, the n univariate GARCH processes cannot be estimated separately anymore, because these different components interact in the log-likelihood. Yet, the DCC part of the model can still be estimated separately. In Section 6.2.4, we have seen that the log-likelihood of elliptical distributions only involves $(r_t - \mu_t)' \Sigma_t (r_t - \mu_t)$, that can be rewritten as $z_t' z_t$, therefore justifying the preliminary estimation of the covariance matrix dynamic.
- For other distributions (including the skewed t distribution), the estimation of the full model has to be performed in one step. Even for moderate-scale portfolios, such an estimation would be simply unmanageable.

An additional issue has to be addressed in the context of non-Gaussian distributions. Once the multivariate GARCH model is estimated, we need to compute the quantile of the distribution of the portfolio. The difficulty is that, in general, the distribution of the aggregate return cannot be deduced from the multivariate distribution of asset returns. The exception is once again the elliptical distribution family. In this case, the aggregate VaR is computed in exactly the same way as for the Gaussian case. Assume for instance that asset returns are distributed as a multivariate Student t distribution (as defined in (6.15) in Section 6.2.1) with ν degrees of freedom. The aggregate VaR is therefore given by

$$VaR_{\theta,t} = -\left(\mu_{p,t}(1) + \tilde{q}_\theta \times \sigma_{p,t}(1)\right),$$

where $\tilde{q}_\theta = t_\nu^{-1}(\theta)$ is the quantile for a probability of loss equal to θ from the univariate Student t distribution with ν degrees of freedom. The ES is given by (8.5).

In the non-elliptical cases, no analytical solution for the distribution of the portfolio return is available. Therefore, we have to turn to alternative techniques, such as numerical integration or Monte Carlo simulation.

First, numerical integration would not be possible in most applications of interest in VaR computation. The reason is that, even for moderate-scale portfolios, the computation burden is excessive. The difficulty is accentuated by the fact that the part to be integrated lies in the lower tail of the distribution.

Consequently, Monte Carlo simulation seems to be the only promising way of evaluating the aggregate VaR when asset returns are modeled through a multivariate distribution whose inverse *cdf* is not known analytically. Giot and Laurent (2003) adopted such a simulation-based approach for the estimation of the VaR of a portfolio with asset returns distributed as a multivariate skewed t distribution. To illustrate how this approach works, we assume that the distribution of the innovation process is $F(z|\eta)$ where η is the vector of shape parameters. The general procedure is the following for computing the aggregate VaR between dates t and $t+1$:

1. At simulation j, simulate a sample of innovations $z^j = (z_1^j, \cdots, z_n^j)'$, drawn from the multivariate distribution $F(z|\hat{\eta})$. Deduce the sample of

asset returns $r_{t+1}^{j} = \mu_t(1) + \Sigma_t(1)^{-1/2} z^j$. Compute the implied portfolio return $r_{p,t+1}^{j} = \alpha_t' r_{t+1}^{j}$.

2. Iterate step 1 for $j = 1, \cdots, J$, where J should be large enough to provide accurate estimates of the desired quantile.
3. Sort the sample $\{r_{p,t+1}^{j}\}_{j=1}^{J}$ in increasing order and compute the desired θ-quantile as the $\lfloor \theta J \rfloor$-th value of the sample.

Needless to say that, for large-scale portfolios, the use of such Monte-Carlo simulations at a daily basis would be quite heavy.

Dealing with large-scale portfolios

Some multivariate GARCH models are well designed for large-scale portfolios. Some of them have been described in Section 6.1.2. They include Factor GARCH models (and their generalizations, such as Orthogonal GARCH models) and the Flexible GARCH model. The former approach relies on reducing the dimensionality of the problem by selecting a reasonably small number of factors to which the multivariate GARCH model gets adjusted. The latter approach decentralizes the estimation task, by estimating the dynamic covariance matrix using univariate and bivariate GARCH models only.

It should be emphasized that these models have been designed in a Gaussian context. For the Flexible GARCH model, it is not clear how it may be extended to a non-Gaussian distribution. Ledoit, Santa-Clara, and Wolf (2003) suggest a trick to circumvent this difficulty. Once the large-dimensional covariance matrix Σ_t is estimated (as described in Section 6.1.2), it is used to estimate the conditional distribution of the portfolio return. They first estimate the variance of the portfolio return as $\hat{\sigma}_{p,t}^2 = \alpha_t' \hat{\Sigma}_t \alpha_t$. Then, they adjust a (standardized) Student t distribution to the standardized innovations evaluated as $\hat{z}_t = \alpha_t' (r_t - \hat{\mu}) / \hat{\sigma}_{p,t}$, where $\hat{\mu}$ is the sample mean of the vector of asset returns. They are then able to compute the quantile of the portfolio return.[11]

Copula functions

It is clear from the discussion above that the main limitation of the use of the asset-level approach for computing VaR is the difficulty to deal with a multivariate non-elliptical distribution, in particular for the estimation of the complete model. This approach has recently benefited from the development of the copula approach (see Section 6.3). One definite advantage of this approach

[11] A comparison of multivariate GARCH models designed for large-scale portfolios has been performed by Ledoit, Santa-Clara, and Wolf (2003). They compare RiskMetrics, the diagonal BEKK model, the DCC model, and the Flexible GARCH models in terms of their ability to estimate the quantile of some empirical portfolios correctly. They do not find significant differences between the various models.

is that the estimation of the marginal distributions (typically, the univariate GARCH models) and the dependence structure can be performed separately. This is true whatever the (possibly different) marginal distributions adopted for the asset returns and whatever the dependence structure. For instance, it is possible to model the univariate distribution of each asset return using a GARCH model with skewed t innovations and then to join these various margins through a Student t copula or any other copula function. Of course, this makes the use of copula much less constraining than the use of multivariate distributions.

The design of a conditional copula in this context can be the following:

1. Each asset return $r_{i,t}$ has its own marginal model given by $r_{i,t} = \mu_{i,t} + \varepsilon_{i,t}$ with $\varepsilon_{i,t} = \sigma_{i,t} z_{i,t}$. Expected returns $\mu_{i,t}$ may be assumed to be constant, so that returns can be demeaned in an initial step. Volatility is modeled as a GARCH(1, 1) model $\sigma_{i,t}^2 = \omega_i + a_i \varepsilon_{i,t-1}^2 + b_i \sigma_{i,t-1}^2$. The standardized innovation $z_{i,t}$ is *iid* with zero mean and unit variance with distribution function $F_i\left(z_{i,t}\right)$. For instance, it may be a skewed Student t distribution, denoted t_{ν_i,λ_i}, where λ_i denotes the asymmetry parameter.
2. The margin of each univariate distribution is given by $u_{i,t} = t_{\nu_i,\lambda_i}^{-1}\left(z_{i,t}\right).$ Then, the copula that links the various margins is

$$H\left(z_{1,t}, \cdots, z_{n,t}\right) = C\left(F_1\left(z_{1,t}\right), \cdots, F_n\left(z_{n,t}\right)\right).$$

Assuming, for instance, a Student t copula with ν degrees of freedom would yield

$$H\left(z_{1,t}, \cdots, z_{n,t}\right) = T_{R,\nu}\left(t_{\nu_1,\lambda_1}^{-1}\left(z_{1,t}\right), \cdots, t_{\nu_n,\lambda_n}^{-1}\left(z_{n,t}\right)\right),$$

where $T_{R,\nu}$ is the *cdf* of the multivariate Student t distribution

$$\begin{aligned}
&T_{R,\nu}\left(u_{1,t}, \cdots, u_{n,t}\right) \\
&= \int_{-\infty}^{u_{1,t}} \cdots \int_{-\infty}^{u_{n,t}} \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\pi\nu)^n\left|R\right|}} \left(1 + \frac{y' R^{-1} y}{\nu}\right)^{-\frac{\nu+n}{2}} dy,
\end{aligned}$$

with R the (n, n) correlation matrix of $u_t = \left(u_{1,t}, \cdots, u_{n,t}\right)'.$

This example illustrates why the copula approach may appear so promising at first sight. The estimation of this model only requires the estimation of $(n+1)$ models (the univariate components and the dependence structure), with each time only a few number of parameters.

It should be noticed however that the great generality allowed by copula functions has also a cost. When we are ultimately interested in VaR computation, we need once again to compute the θ-quantile of the portfolio return distribution. The only available approach appears to be the Monte Carlo simulation. A similar procedure to the one presented in Section 8.4.3 may be adopted. A further cost is the difficulty to use copula functions for the case of more than two or three assets.

8.5 Non-linear models

Other recent research focuses on estimating the VaR of portfolios containing options or other positions with non-linear price behavior. VaR methods employing a linear approximation to the relation between instrument values and the underlying risk factors are unlikely to be robust when applied to non-linear portfolios. Britten-Jones and Schaefer (1999) proposed a VaR framework that is based on a second order "delta-gamma" approximation and recognizes the impact that this will have, not only on variance, but on the form of the distribution.

Let us now describe this approach. As seen in Section 8.1, the change in the value of the portfolio is $\Delta W_{t+1}(\alpha_t) = \alpha_t' \Delta p_t = \sum_{i=1}^{n} \alpha_{i,t} \Delta p_{i,t}$, where $\Delta p_{i,t}$ denotes the change in the value of asset i. Clearly, the change in the portfolio value is linear in the change of the asset prices. In some cases, however, the portfolio may include some derivatives that would introduce a non-linear relation between asset prices and portfolio value. To cope with this non-linear relation, two main approaches have been proposed: the "delta-only" method uses a linear approximation; the "delta-gamma" method involves a linear-quadratic approximation. We investigate these various approaches in turn. An alternative approach is based on a Monte Carlo simulation of a large number of market scenarios.

8.5.1 The "delta-only" method

We assume that the assets depend on a set of K risk factors. For instance, in a structured product such as an insured portfolio, containing the underlying asset and a put option, the risk factor would be the underlying asset. The first-order approximation of the value of the portfolio is given by

$$\begin{aligned} \Delta^{\delta} W_{t+1} &= \sum_{i=1}^{n} \alpha_{i,t} \frac{\partial p_i(f,t)}{\partial t} \Delta t + \sum_{i=1}^{n} \alpha_{i,t} \sum_{k=1}^{K} \frac{\partial p_i(f,t)}{\partial f_k} \Delta f_{k,t} \\ &= \mu_\delta + \sum_{k=1}^{K} \delta_k \Delta f_{k,t} = \mu_\delta + \frac{\partial W'}{\partial f} \Delta f_t, \end{aligned}$$

where μ_δ denotes the change in portfolio value due to time, and δ_k is the aggregate effect on the portfolio value of factor k. The last equation introduces notations to be used below. In some cases, this approach may be viewed as a valid approximation for evaluating the portfolio value. However, in general, as exemplified by Britten-Jones and Schaefer (1999), it may yield large approximation errors.

8.5.2 The "delta-gamma" method

If we incorporate the second-order effects of the risk factors, we then obtain

$$\Delta^{\gamma} W_{t+1} = \mu_{\gamma} + \frac{\partial W'}{\partial f} \Delta f_t + \frac{1}{2} \Delta f_t' \frac{\partial^2 W}{\partial f \partial f'} \Delta f_t.$$

If we define $\delta = \partial W / \partial f$ and $\Gamma = \partial^2 W / (\partial f \partial f')$, the $(K, 1)$ vector and (K, K) matrix of aggregate delta and gamma, respectively, we have

$$\Delta^{\gamma} W_{t+1} = \mu_{\gamma} + \delta' \Delta f_t + \frac{1}{2} \Delta f_t' \; \Gamma \; \Delta f_t. \tag{8.6}$$

Completing the square in (8.6), we obtain

$$\Delta^{\gamma} W_{t+1} = \mu_c + \frac{1}{2} \left(\Delta f_t + \Gamma^{-1} \delta \right)' \Gamma \left(\Delta f_t + \Gamma^{-1} \delta \right),$$

with $\mu_c = \mu_{\gamma} - \frac{1}{2} \delta' \Gamma^{-1} \delta$.

Assume now that the factor vector is distributed as a multivariate normal distribution $\Delta f_t \sim \mathcal{N}\left(\mu_f, \Sigma_f\right)$. Then, we have that $\left(\Delta f_t + \Gamma^{-1} \delta\right) \sim \mathcal{N}\left(\mu_f + \Gamma^{-1} \delta, \Sigma_f\right)$ so that, defining $y_t = \Sigma^{-1/2} \left(\Delta f_t + \Gamma^{-1} \delta\right)$, we obtain $y_t \sim \mathcal{N}\left(\mu_f + \Gamma^{-1} \delta, I_K\right)$. If we define $A = \Sigma^{-1/2} \Gamma \Sigma^{-1/2}$, we have

$$\Delta^{\gamma} W_{t+1} = \mu_c + \frac{1}{2} y_t' A y_t,$$

which shows that the change in portfolio value is a linear combination of uncorrelated non-central χ_1^2 variables.

This result is useful for computing the VaR of such a portfolio. It is possible to evaluate the moment generating as well as the characteristic functions of $(\Delta^{\gamma} W_{t+1} - \mu_c)$. See for instance Johnson, Kotz, and Balakrishnan (1995, vol. 2, p. 447). The characteristic function can be numerically inverted to compute probabilities such as $\Pr\left[\Delta^{\gamma} W_{t+1} - \mu_c < x\right]$ and then to evaluate the VaR of this portfolio.[12]

Glasserman, Heidelberger, and Shahabuddin (2000) propose an extension of the delta-gamma method to the case where risk factors are assumed to be t distributed rather than normally distributed. In such case, the difficulty comes from the fact that, although uncorrelated, risk factors are not independent anymore.

8.6 Comparison of VaR models

Several contributions present and compare the main approaches adopted for computing VaR. They include the work of Hsieh (1993), van den Goorbergh and Vlaar (1999), Christoffersen, Hahn, and Inoue (2001), Giot and Laurent (2003), and Ledoit, Santa-Clara, and Wolf (2002). Many tools have been used to compare these various techniques. Various statistical methods for evaluating VaR models have been suggested by Kupiec (1995), Christoffersen (1998), Lopez (1998), and Engle and Manganelli (2004).

---

[12] Britten-Jones and Schaefer (1999) propose an alternative approach based on the approximation of the distribution of a sum of independent non-central χ^2 variables.

8.6.1 Evaluation of VaR models

A natural way to evaluate VaR models is the *"hit" test* developed by Christoffersen (1998) already described in Section 5.3.3. The test is designed to evaluate if a given model is able to provide interval forecast that have the same coverage as in the data. Here, the central object in the approach is the Hit_t variable, defined in the case of VaR evaluation as

$$Hit_{t+1} = \begin{cases} 1 & \text{if } r_{p,t+1} < VaR_{\theta,t}, \\ 0 & \text{if } r_{p,t+1} > VaR_{\theta,t}. \end{cases} \tag{8.7}$$

Lopez (1998) proposes a supplementary evaluation based on ad hoc loss functions. The loss function is specified as the cost of the various outcomes

$$C_{t+1} = \begin{cases} f\left(r_{p,t+1}, VaR_{\theta,t}\right), & \text{if } r_{p,t+1} < VaR_{\theta,t}, \\ g\left(r_{p,t+1}, VaR_{\theta,t}\right), & \text{if } r_{p,t+1} \geq VaR_{\theta,t}, \end{cases}$$

Since this is a cost function and because prevention of VaR exceedances is of paramount importance, $f(x,y) \geq g(x,y)$ for a given y. The best VaR model is the one that minimizes the total cost over the N last days, $\sum_{i=0}^{N-1} C_{t-i}$.

There are many ways to specify f and g depending on the concern of the decision maker. For instance, choosing $f(x,y) = 1$ and $g(x,y) = 0$, we obtain Christoffersen's hit test. If the exceedances as well as the magnitude of the exceedances are of importance, we may choose

$$C_{t+1} = \begin{cases} 1 + \left(r_{p,t+1} - VaR_{\theta,t}\right)^2, & \text{if } r_{p,t+1} < VaR_{\theta,t}, \\ 0, & \text{if } r_{p,t+1} \geq VaR_{\theta,t}. \end{cases}$$

For the bank that implements the VaR model and that has to set aside capital reserves, $g = 0$ is not appropriate because liquid assets do not provide good returns. So one cost function that will take into account the opportunity cost of money is

$$C_{t+1} = \begin{cases} \left|r_{p,t+1} - VaR_{\theta,t}\right|^{\gamma}, & \text{if } r_{p,t+1} < VaR_{\theta,t}, \\ \left|r_{p,t+1} - VaR_{\theta,t}\right| \times i, & \text{if } r_{p,t+1} \geq VaR_{\theta,t}, \end{cases}$$

where γ reflects the seriousness of large exception and i is a function of interest rate.

8.6.2 Comparison of methods

In this section, we present some guidelines on some empirical studies that appeared in the literature.

Van den Goorbergh and Vlaar (1999) compare several approaches for computing VaR, including historical simulation, unconditional EVT, and various GARCH models with normal (RiskMetrics) and t innovations. Using data from the Dutch stock index and the Dow Jones, they show that (i) conditional methods (GARCH models) out-perform unconditional ones, suggesting that the main characteristic of returns for evaluating VaR is volatility clustering; (ii) using a conditional t distribution provides a better fit than a normal distribution, implying that capturing distribution fat-tailedness is also crucial for an accurate measure of VaR. The GARCH model with t innovations is the only model found to perform well for all the probabilities considered. Other techniques such as historical simulation and EVT tend to underestimate the actual VaR. Importantly, this empirical evidence suggests that the unconditional EVT approach is unable to capture the consequences of the time-variability of volatility.

McNeil and Frey (2000) compare the conditional EVT method with GARCH models that have either normal or t innovations. They show that the conditional EVT model provides more accurate estimates of the VaR than the GARCH with normal innovations. Their coverage test does not reject the two approaches that are able to capture both the volatility clustering and the fat-tailedness of the distribution, i.e., the GARCH–EVT method and the GARCH model with t innovations.

Giot and Laurent (2003) more specifically investigate the GARCH approach and consider several distributional assumptions. They highlight that, at least for some return series, the asymmetry of the distribution should be taken into account for capturing VaR. Indeed, they obtain for NASDAQ and Nikkei indices, that a GARCH model with t innovations fails to measure the VaR accurately. In contrast, the model with skewed t innovations performs very well. When extended to the multivariate set-up, the model with skewed t innovations provides very accurate measures as well.

Ledoit, Santa-Clara, and Wolf (2002) compare the performances of various techniques in the context of a large-scale portfolio. More precisely, working with a large number of asset returns, they compute the covariance matrix using different approaches and then compute the VaR assuming a t distribution for portfolio returns. They find that the various techniques (including the CCC, BEKK and Flexible GARCH models) perform broadly equally well in terms of unconditional coverage.

8.6.3 10-day VaR and scaling

It is well-known that the variance of a Gaussian variable follows a simple scaling law. Indeed, the Basel Committee, in its 1996 Amendment, states that it will accept a simple $\sqrt{T}$ scaling of 1-day VaR for deriving the 10-day VaR required in calculating market risk related risk capital.

The stylized facts of financial market volatility and research findings have repeatedly shown that a 10-day VaR is not likely to be the same as $\sqrt{10} \times$ 1-day VaR. First, the dynamic of a stationary volatility process suggests that if the current level of volatility is higher than unconditional volatility, the subsequent daily volatility forecasts will decline and converge to unconditional volatility, and vice versa for the case where the initial volatility is lower than the unconditional one. The rate of convergence depends on the degree of volatility persistence. In the case where initial volatility is higher than unconditional volatility, the scaling factor will be less than $\sqrt{10}$. In the case where initial volatility is lower than unconditional volatility, the scaling factor will be more than $\sqrt{10}$. In practice, due to volatility asymmetry and other predictive variables that might be included in the volatility model, it is always better to calculate $\hat{\sigma}^2_{t+1}, \hat{\sigma}^2_{t+2}, \cdots, \hat{\sigma}^2_{t+10}$ separately. The 10-day VaR is then produced using the 10-day volatility estimate computed from the sum $\sum_{i=1}^{10} \hat{\sigma}^2_{t+i}$.

Second, financial asset returns are not normally distributed. Danielsson and de Vries (1997) show that the scaling parameter for quantile derived using the EVT method increases at the approximate rate of T^{ξ}, which is typically less than the square-root-of-time adjustment. For a typical value of $\xi\,(= 0.25)$, we have $T^{\xi} = 1.778$, which is less than $10^{0.5}$ $(= 3.16)$. McNeil and Frey (2000) on the other hand dispute this finding and claim the exponent to be greater than 0.5. The scaling factor of $10^{0.5}$ produced far too many VaR violations in the back-test of five financial series, except for returns on gold. In view of the conflicting empirical findings, one possible solution is to build models using 10-day returns data. This again highlights the difficulty due to the inconsistency in the rule applying to VaR for calculating risk capital and the one applying to VaR for back-testing.

8.6.4 Illustration

We consider once again the four market indices, SP500, DAX, FT-SE, and Nikkei over the period from January 1980 to December 2004. We create a portfolio composed of the four indices with an equal weight of 25%. Then, we compute the 1% and 5% VaR using different approaches developed in the previous sections. We consider the historical simulation (based on subsamples of size $N = 500$), RiskMetrics (with $\lambda = 0.94$), the GARCH–EVT approach of McNeil and Frey (2000) (with u corresponding to the 10% lower tail), a GARCH$(1,1)$ model with t innovations, and finally a GARCH$(1,1)$ model with skewed t innovations.

Notice that the parameter estimates in the gpd for the GARCH–EVT approach are found to be equal to $\hat{\xi} = 0.0504$ (with a standard error of 0.0324) and $\hat{\psi} = 0.5929$ (with a standard error of 0.2993), suggesting that the lower tail of standardized residuals is not very fat and is actually quite close to the attraction domain of the Gumbel distribution. Given that the univariate series produce fat-tailed residuals, this result suggests that extreme risks diversify away in a portfolio. Table 8.2 reports the parameter estimates for the three GARCH$(1,1)$ models estimated (with standard errors in parentheses). The QML estimation of the McNeil and Frey model assumes normality of innovations, whereas the two other GARCH$(1,1)$ models assume t and skewed t innovations, respectively. We notice that estimates of the parameters pertaining to the volatility evolution are not significantly affected by the change of conditional distribution. We also observe that the degree-of-freedom parameter ν is too small to be consistent with the normality assumption.[13] Finally, the asymmetry parameter λ is strongly significant, suggesting that allowing fat tails without asymmetry would not be sufficient in order to evaluate the VaR accurately.

Table 8.3 reports information concerning the computation of the VaR and ES. First, since we consider conditional VaR and ES, for each day, we have a different estimate of these statistics. For this reason, we therefore present the average of the estimates obtained from the various methods. Second, the number of exceedances is the number of dates t when the observed return exceeds the theoretical VaR. The expected exceedance is given by θT where θ is the confidence level and T the number of observations in the sample. Then, we present the three tests developed by Christoffersen (1998) (see Section 8.6.1). They allow identifying where the possible rejection of the model comes from (p-values are in parentheses).

At the 1% confidence level, two methods perform very well: the GARCH–EVT method and the GARCH model with skewed t innovations. In both cases, the actual number of exceedances is very close to the expected number. The unconditional coverage is not rejected for the two methods, and the conditional coverage is not rejected for the GARCH–EVT method only. At the 5% confidence level, the only method able to satisfy the unconditional coverage test is the GARCH model with skewed t innovations. The independence of $Hits$ is rejected for all methods.

To sum up, it appears that the conditional EVT method performs very well for very small confidence levels.[14] In contrast, for larger confidence levels,

Table 8.2. *Parameter estimates of GARCH(1,1) models*

	Normal distribution	Student t distribution	Skewed t distribution
ω	0.0122	0.0084	0.0085
	(0.0017)	(0.0017)	(0.0017)
α	0.1159	0.0917	0.0904
	(0.0084)	(0.0098)	(0.0095)
β	0.8661	0.8948	0.8960
	(0.0089)	(0.0109)	(0.0105)
ν	–	8.0622	8.1545
		(0.7245)	(0.7463)
λ	–	–	−0.0943
			(0.0170)

[13] Remember that the Student t distribution converges to a normal distribution when $\nu \to \infty$.

[14] Unreported results indicate that it is the only approach to accurately estimate the VaR at the 0.5% level.

the GARCH model with skewed t innovations provides accurate estimates of the VaR, as well.

Table 8.3. *VaR computation for various methods*

	Var ES (avg)	Number of exceed.	Uncond. coverage	Indep. of *Hits*	Cond. coverage
1% conf. level		Exp.: 59			
Hist. simulation	2.093	83	8.462	37.409	45.871
	2.609		(0.004)	(0.000)	(0.000)
RiskMetrics	1.649	119	46.857	17.878	64.735
	1.952		(0.000)	(0.000)	(0.000)
GARCH–EVT	1.897	55	0.332	2.570	2.902
	2.557		(0.565)	(0.109)	(0.234)
GARCH-t	1.768	71	2.170	6.384	8.554
	2.256		(0.141)	(0.012)	(0.014)
GARCH-skewed t	1.866	59	0.002	8.941	8.943
	2.393		(0.963)	(0.003)	(0.011)
5% conf. level		Exp.: 296			
Hist. simulation	1.192	357	12.105	83.593	95.698
	1.689		(0.001)	(0.000)	(0.000)
RiskMetrics	1.157	381	23.166	33.818	56.984
	1.518		(0.000)	(0.000)	(0.000)
GARCH–EVT	1.156	340	6.335	7.750	14.085
	1.777		(0.012)	(0.005)	(0.001)
GARCH-t	1.125	368	16.761	13.703	30.464
	1.590		(0.000)	(0.000)	(0.000)
GARCH-skewed t	1.167	325	2.740	11.921	14.661
	1.667		(0.098)	(0.001)	(0.001)