A multivariate FGD technique to improve VaR computation in equity markets

Francesco Audrino* and Giovanni Barone-Adesi
University of Southern Switzerland

Revised Version: September 2003

Abstract

It is difficult to compute Value-at-Risk (VaR) using multivariate models able to take into account the dependence structure between large numbers of assets and being still computationally feasible. A possible procedure is based on functional gradient descent (FGD) estimation for the volatility matrix in connection with asset historical simulation. Backtest analysis on simulated and real data provides strong empirical evidence of the better predictive ability of the proposed procedure over classical filtered historical simulation, with a resulting significant improvement in the measurement of risk.

Key words: Volatility estimation; Functional Gradient Descent; Filtered Historical Simulation; Value-at-Risk

1 Introduction

The measurement of market risk (the risk that a financial institution incurs losses on its trading book due to unexpected changes in prices or rates) has assumed a primary importance for regulators and for internal risk control, because of the growth in trading in most financial institutions. One of the most widely used risk measures is Value-at-Risk, or VaR (see Duffie and Pan, 1997, for a review of the early literature on VaR). A portfolio's (or an asset's) VaR is commonly defined as the maximum loss that will be incurred on the portfolio with a given level of confidence over a specified holding period, based on the distribution of price changes over a given observation period. Or, in other words, a VaR calculation amounts to a simple quantile estimation of the Profit-and-Loss distribution of a given portfolio over a prescribed holding period.

The main advantage of using VaR as a risk measure is that it is very simple and can also be used to summarize the risk of individual positions. Because of this, it has been adopted for regulatory purposes. More specifically, the BIS has stipulated that the minimum capital requirement for market risk should be based on a 10-day VaR at 99% confidence level.

A lot of different ways have been proposed so far to compute VaR with univariate methods: see for example Dowd (1998) or Jorion (2001). In this paper, we study whether the accuracy of VaR predictions for individual positions estimated with univariate techniques can be significantly

*The authors would like to thank Peter Bühlmann and two anonymous referees for some helpful comments. Financial support by the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK) is gratefully acknowledged.

improved using multivariate methods, which can also take into account the predictive contributions and interactions of other positions belonging to a common market segment. We present some simulations and results for two real data examples: the Swiss chemical/pharmaceutical and the US biotechnological equity market segment.

Although VaR is conceptually a simple measure of risk, computing it in practice using multivariate methods which allow to take into account the possible non-linear dependence structure across different assets in large equity markets can be very difficult, due to the well known curse of dimensionality when estimating high-dimensional conditional covariance matrices. Previous work on multivariate volatility models has been done by Bollerslev (1990), Engle et al. (1990), Lin (1992), Engle and Kroner (1995), Alexander (2001) and Engle (2002) in the framework of GARCH-type models, and by Harvey et al. (1994), Chib et al. (1999) and Aguilar and West (2000) within the stochastic volatility (state space model) framework. In the GARCH-type framework, only very simple models (Bollerslev, 1990; Engle, 2002) are feasible in high dimensions without resorting to variance reduction techniques, whereas with stochastic volatility models, only Chib et al. (1999) present an example with dimensionality as large as 40 which is still far lower than the number of assets which often occurs in practice.

We present here a non-parametric technique for constructing accurate daily VaR estimates for individual positions which is able to take into account all the possible non-linear dependence structure across different assets and which is still computationally feasible for multivariate problems in large dimensions. Our strategy is based on functional gradient descent (FGD) estimation for the multivariate conditional covariance matrix in connection with historical simulation. FGD is a recent technique from the area of machine learning introduced to solve the classification problem (Mason et al., 1999; Breiman, 1999; Friedman et al., 2000; Friedman, 2001). The FGD algorithm that we propose is the same as in Audrino and Bühlmann (2003), who have studied the statistical performance of FGD in the financial field. It is very general and can be further adapted to solve other multivariate problems dealing with high-dimensional (volatility) function estimation, such as asset allocation problems, involving the allocation of assets among several stocks whose returns are correlated, or risk management for large global trading portfolios with time-dependent weights.

The main advantage of our technique is its ability to construct reliable and powerful VaR predictions in a high-dimensional multivariate GARCH-type set-up. As we have already said above, it is not possible to use standard multivariate GARCH-type models, such as for example the BEKK models, to estimate the conditional covariance matrix in large dimensions, because we would have to face an intractable model-selection problem and most parameters would have to be set to zero in order to avoid overfitting. Using FGD this problem can be overcome: this technique can also be used in situations where we deal with more parameters than observations. Choosing reasonable starting functions (for example estimated by a very simple multivariate GARCH-type model), FGD tries to improve, often successfully, those components where the initial predictions are poorest. Clearly, as Audrino and Bühlmann (2003) have already shown, we can not expect to learn in all d dimensions when increasing d and keeping a fixed sample size. However, although the gain on average will generally decrease with the number of fitted assets, FGD still improves the worst cases.

Once that FGD yields accurate predictions for the conditional covariance matrix, we can use a model-based bootstrap (Efron and Tibshirani, 1993) to generate recursively pathways for future returns. This methodology can also be seen as a multivariate extension of the method proposed and backtested by Barone-Adesi et al. (1998), (1999) and (2002) based on filtered historical simulation. The main difference here is that we use a multivariate GARCH-type model in connection with FGD for filtering the residuals.

Our strategy contrasts well several critiques made about the use of filtered historical simulation for estimating VaR (Pritsker, 2001). First of all, our FGD technique allows for the use of cross-terms as predictor variables. This is a reasonable assumption if we consider that assets belonging to a common market segment show some dependence structure and it is conceivable that one asset can be influenced and predicted by past values of some other. This possibility has not been considered in the filtered historical simulation method proposed by Barone-Adesi et al. (1998) and (1999), where the volatility of an asset depends only on its own past lagged values and volatilities.

A second critique is related to the assumption of independent identically distributed (i.i.d.) innovations, which implies fixed conditional correlations in a multivariate setting. In our procedure, we only assume constant conditional correlations in a rolling (i.e. not fixed) time-window of about three years of data, using to model the dynamics of the multivariate return series the constant conditional correlation (CCC) model firstly proposed by Bollerslev (1990). Our method can be perhaps further improved by assuming dynamic conditional correlations (see for example Engle, 2002), but this is not in the spirit of this paper and it is left to future research.

Using different statistical and economical backtests, we collect empirical evidence of the better predictive power of our multivariate procedure over other univariate techniques, and in particular over the filtered historical simulation method of Barone-Adesi et al. (1998) and (1999). Through a simulation exercise and our two real data examples for individual assets belonging to a common equity market segment, we found that VaR estimates from our technique are more accurate, with a resulting improvement in the measurement of market risk.

The paper is organized as follows. We present and discuss our FGD algorithm in section 2. Section 3 is concerned with the description of the model-based bootstrap method used for the construction of daily VaR estimates. In section 4 we propose two simulation exercises to test the goodness of our multivariate procedure, also in comparison to the univariate method of Barone-Adesi et al. (1999). The results of the real data backtest analysis are summarized in section 5. Section 6 concludes the paper.

2 Volatility estimation with Functional Gradient Descent

2.1 Starting point

The multivariate real data of interest are in our case time series of asset prices $\{P_{t,i};\ t = 0, 1, \ldots, T,\ i = 1, \ldots, d\}$. Their (log)-returns (in percentages) are defined as the change in the logarithms of the individual prices

$$X_{t,i} = 100 \cdot \big(\log(P_{t,i}) - \log(P_{t-1,i})\big),\ t = 1, \ldots, n.$$

We assume stationarity of the returns (at least within a suitable time-window). In the empirical investigations of section 5, results are based on a rolling time-window of about three years, which seems to be consistent with the assumption of stationarity (Mikosch and Starica, 1999).

As Audrino and Bühlmann (2003) have already shown, the Functional Gradient Descent (FGD) technique (Friedman et al., 2000 or Friedman, 2001) is a powerful strategy to construct accurate predictions for the multivariate volatility matrix

$$V_t = \mathrm{Cov}_{d\times d}(\mathbf{X}_t|\mathcal{F}_{t-1}),\ \mathbf{X}_t = (X_{t,1}, \ldots, X_{t,d})^T, \tag{2.1}$$

where $\mathcal{F}_{t-1}$ denotes the information available up to time $t-1$, i.e. the σ-algebra generated by $\{\mathbf{X}_s;\ s \leq t-1\}$. As already mentioned in section 1, the importance of FGD is revealed

particularly in large dimensions (for example d in the dozens or hundreds) where predicting the multivariate volatility matrix raises huge challenges because of the well-known curse of dimensionality. In such a case FGD is one of the few feasible non-parametric techniques (if not the only one so far).

Our working model is a generalization of the classical constant conditional correlation (CCC) GARCH model firstly introduced by Bollerslev (1990),

$$\mathbf{X}_t = \mu_t + \Sigma_t \mathbf{Z}_t, \tag{2.2}$$

where we assume the following:

(A1) (innovations) $\{\mathbf{Z}_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. multivariate innovations with spherical distribution (e.g. multivariate normal) having mean zero and covariance matrix $\mathrm{Cov}(\mathbf{Z}_t) = I_d$. Moreover, $\mathbf{Z}_t$ is independent from $\mathcal{F}_{t-1} = \{\mathbf{X}_s; s \leq t-1\}$.

(A2) (CCC construction) The conditional covariance matrix $V_t = \mathrm{Cov}(\mathbf{X}_t|\mathcal{F}_{t-1}) = \Sigma_t \Sigma_t^T$ is almost surely positive definite for all t. The typical element of V_t is $v_{t,ij} = \rho_{ij}(v_{t,ii}v_{t,jj})^{1/2}$ $(i, j = 1, \ldots, d)$. The parameter $\rho_{ij} = \mathrm{Corr}(X_{t,i}, X_{t,j}|\mathcal{F}_{t-1})$ equals the constant conditional correlation and hence $-1 \leq \rho_{ij} \leq 1$, $\rho_{ii} = 1$.

(A3) (functional form) The conditional variances are functions of the form

$$v_{t,ii} = \sigma_{t,i}^2 = \mathrm{Var}(X_{t,i}|\mathcal{F}_{t-1}) = F_i(\{X_{t-j,k};\ j = 1, 2, \ldots, k = 1, \ldots, d\})$$

where F_i takes values in $\mathbb{R}^+$.

(A4) (conditional mean) The conditional mean μ_t is of the form

$$\mu_t = (\mu_{t,1}, \ldots, \mu_{t,d})^T = A\mathbf{X}_{t-1}$$

with A a diagonal $d \times d$ matrix (parametric vector AR(1) in mean).

Note that (A2) can be represented in matrix form as

$$V_t = \Sigma_t \Sigma_t^T = D_t R D_t,$$
$$D_t = \mathrm{diag}(\sigma_{t,1}, \ldots, \sigma_{t,d}),\ R = [\rho_{ij}]_{i,j=1}^d.$$

The functional form (A3) allows clearly for cross-terms, since the conditional variances of one series depends on past multivariate observations. This is one of the nice features of such a multivariate GARCH-type model and is motivated from the fact that in reality some instruments can be influenced and better predicted using past information from other risk factors.

2.2 Volatility estimation

FGD estimates the (squared) individual volatility functions $F_i(\cdot)$ in (A3), where $F_i(\cdot) : \mathbb{R}^{pd} \to \mathbb{R}^+$ is restricted to be a function of the last p lagged multivariate observations, with p finite. The main idea of FGD is to find the estimates for the functions $F_i(\cdot)$ which minimize a suitable loss function λ, under the constraint that the solutions $\widehat{F}_i(\cdot)$ are additive expansions of "simple estimates". These "simple estimates" are given from a statistical procedure $\mathcal{S}$, called the base learner, which is often constructed from a (constrained or penalized) least squares fitting; common examples of base learners are regression or decision trees, projection pursuit regressors,

neural nets and splines. For more details, we remand to Friedman et al. (2000), Friedman (2001) and, for a first application in the financial field, Audrino and Bühlmann (2003).

To proceed with the FGD technique, we have therefore to specify a suitable loss function which has to be minimized during the estimation. Assuming multivariate normality of the innovation variables $\mathbf{Z}_t$ in (2.2), the (multivariate) negative log-likelihood (conditional on the first p observations) is given by

$$\begin{aligned} & -\sum_{t=p+1}^{n} \log\left((2\pi)^{-d/2}\det(V_t)^{-1/2}\exp(-\xi_t^T V_t^{-1}\xi_t/2)\right) \\ = & \sum_{t=p+1}^{n}\left(\log(\det(D_t)) + \frac{1}{2}(D_t^{-1}\xi_t)^T R^{-1}(D_t^{-1}\xi_t)\right) + n' d\log(2\pi)/2 + n'\log(\det(R))/2 \end{aligned}$$

where $\xi_t = \mathbf{X}_t - \mu_t$, D_t is diagonal with elements $\sqrt{F_i(\mathbf{X}_{t-p}^{t-1})}$ and $n' = n - p$. For this reason a natural loss function is

$$\begin{aligned} & \lambda_R(\mathbf{Y},\mathbf{f}) = \log(\det(D(\mathbf{f})) + \frac{1}{2}(D(\mathbf{f})^{-1}\mathbf{Y})^T R^{-1}(D(\mathbf{f})^{-1}\mathbf{Y}) + \frac{1}{2}\log(\det(R)) + \frac{d}{2}\log(2\pi), \\ & D(\mathbf{f}) = \mathrm{diag}(f_1,\ldots,f_d), \end{aligned} \tag{2.3}$$

where the terms $d\log(2\pi)/2$ and $\log(\det(R))/2$ are constants and may be dropped. As pointed out with the subscript, the loss function λ_R depends on the unknown correlation matrix R. The FGD algorithm will be constructed iteratively by estimating R and using the loss function with the estimated R to get an estimate for all F_i's.

Estimation of the correlation matrix R can be easily done via empirical moments of residuals. Having (previous) estimates $\hat{\mathbf{F}} = (\hat{F}_1,\ldots,\hat{F}_d)$, we build the residuals

$$\hat{\varepsilon}_{t,i} = (X_{t,i} - \hat{\mu}_{t,i})/\hat{F}_i(\mathbf{X}_{t-1},\ldots)^{1/2},\ t = p+1,\ldots,n$$

and define

$$\hat{R} = (n-p)^{-1}\sum_{t=p+1}^{n}\hat{\varepsilon}_t\hat{\varepsilon}_t^T,\ \hat{\varepsilon}_t = (\hat{\varepsilon}_{t,1},\ldots,\hat{\varepsilon}_{t,d})^T. \tag{2.4}$$

As the name "functional gradient descent" suggests, we need to calculate the partial derivatives of the loss function λ_R. They are given (in the case of normality of the innovations $\mathbf{Z}_t$) by

$$\frac{\partial\lambda_R(\mathbf{Y},\mathbf{f})}{\partial f_i} = (\frac{1}{f_i} - \sum_{j=1}^{d}\frac{\gamma_{ij}y_i y_j}{f_i^{3/2}f_j^{1/2}})/2,\ i = 1,\ldots,d, \tag{2.5}$$

where $[\gamma_{ij}]_{i,j=1}^d = R^{-1}$. This will be used when computing negative gradients (see Step 2 in the following FGD algorithm) for every component $i = 1,\ldots,d$.

If the assumption of normality of the innovations $\mathbf{Z}_t$ in (2.2) is violated, the estimates may be consistent but inefficient and this can result in poor performance. As it is shown in the empirical investigations of section 5, an alternative can be to assume a fat-tailed distribution for the innovations (such as for example a scaled t_ν distribution with a fixed number of degrees of freedom ν), that is consistent with the belief that financial (log-) returns are leptokurtic. Another possibility could be to assume a normal inverse gaussian distribution, which seems to work quite well (Venter and de Jongh, 2001).

In analogy to Audrino and Bühlmann (2003), the FGD algorithm for multivariate volatility looks as follows.

FGD algorithm

Step 1 (initialization). Choose the starting function $\hat{F}_{i,0}(\cdot)$ and denote by

$$\hat{F}_{i,0}(t) = \hat{F}_{i,0}(\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \ldots),\ i = 1, \ldots, d.$$

Construct estimates $\hat{\mu}_t$ for the conditional mean from a starting model and compute $\hat{R}_0$ as in (2.4) using $\hat{\mathbf{F}}_0$. Set $m = 1$.

For every component $i = 1, \ldots, d$, do the following.

Step 2$_i$ (projection of component gradients to base learner). Compute the negative gradient

$$U_{t,i} = -\frac{\partial \lambda_{\hat{R}_{m-1}}(\mathbf{X}_t - \hat{\mu}_t, \mathbf{F})}{\partial F_i}|_{\mathbf{F}=\hat{\mathbf{F}}_{m-1}(t)},\ t = p+1, \ldots, n.$$

This is explicitly given in (2.5). Then, fit the negative gradient vector $U_i = (U_{p+1,i}, \ldots, U_{n,i})^T$ with a base learner, using always the first p time-lagged predictor variables (i.e. $\mathbf{X}_{t-p}^{t-1}$ is the predictor for $U_{t,i}$)

$$\hat{f}_{m,i}(\cdot) = \mathcal{S}_X(U_i)(\cdot),$$

where $\mathcal{S}_X(U_i)(x)$ denotes the predicted value at x from the base learner $\mathcal{S}$ using the response vector U_i and predictor variables X.

Step 3$_i$ (line search). Perform a one-dimensional optimization for the step-length,

$$\hat{w}_{m,i} = \text{argmin} \sum_{t=p+1}^{n} \lambda_{\hat{R}_{m-1}}(\mathbf{X}_t - \hat{\mu}_t, \hat{\mathbf{F}}_{m-1}(t) + w\hat{f}_{m,i}(\mathbf{X}_{t-p}^{t-1})).$$

($\hat{\mathbf{F}}_{m-1}(t) + w\hat{f}_{m,i}(\cdot)$ is defined as the function which is constructed by adding in the ith component only). This can be expressed more explicitly by using (2.3).
(Note that the line search guarantees that the negative log-likelihood is monotonely decreasing with every iteration.)

Step 4 (up-date). Select the "best component" as

$$i_m^* = \text{argmin}_i \sum_{t=p+1}^{n} \lambda_{\hat{R}_{m-1}}(\mathbf{X}_t - \hat{\mu}_t, \hat{\mathbf{F}}_{m-1}(t) + \hat{w}_{m,i}\hat{f}_{m,i}(\mathbf{X}_{t-p}^{t-1})).$$

Up-date

$$\hat{\mathbf{F}}_m(\cdot) = \hat{\mathbf{F}}_{m-1}(\cdot) + \hat{w}_{m,i_m^*}\hat{f}_{m,i_m^*}(\cdot).$$

Then, compute the new estimate $\hat{R}_m$ according to (2.4) using $\hat{\mathbf{F}}_m$.

Step 5 (iteration). Increase m by one and iterate Steps 2–4 until stopping with $m = M$. This produces the FGD estimate

$$\hat{\mathbf{F}}_M(\cdot) = \hat{\mathbf{F}}_0(\cdot) + \sum_{m=1}^{M} \hat{w}_{m,i_m^*}\hat{f}_{m,i_m^*}(\cdot).$$

The stopping value M is chosen by the following scheme: split the (in-sample) estimation period into two sets, the first of size $0.7 \cdot n$ used as training set and the second of size $0.3 \cdot n$ used as test set (this can also be used when the data are dependent). The optimal value of M is then chosen to optimize the cross-validated log-likelihood.

Remark 1. Initialization in Step 1 is very important here to achieve good estimates. As a starting function, we propose to use the fit from a AR(1)-CCC-GARCH(1,1) model (Bollerslev, 1990) which is of the form (2.2) with (A3) specified to

$$F_i(X_{t-1}, X_{t-2}, \ldots) = \sigma_{t,i}^2 = \alpha_{0,i} + \alpha_{1,i}\xi_{t-1,i}^2 + \beta_{0,i}\sigma_{t-1,i}^2, \ i = 1, \ldots, d. \tag{2.6}$$

We construct the estimates by maximum likelihood from the d individual series, ignoring the more general correlation structure in R. This causes some statistical decrease in efficiency, but it gains the advantage that the estimates remain quickly computable in high dimensions d.
Note that the starting estimates $\hat{\mu}_t$ for the conditional mean are kept fixed during the FGD estimation of the volatility functions.

Remark 2. The base learner in Step 2 obviously determines the FGD estimate $\hat{\mathbf{F}}_M(\cdot)$.This should be "weak" (not involving too many parameters to be estimated) enough not to immediately produce an overfitted estimate at the first iteration. The complexity of the FGD estimate $\hat{\mathbf{F}}_M(\cdot)$ is increased by adding further terms with every iteration (Bühlmann and Yu, 2003). We choose regression trees as base learners, since particularly in high dimensions, they have the ability to select variables by choosing few of the explanatory variables for prediction. This choice should not be regarded as exclusive: other base learners could be tried out and compared using some form of cross-validation.
As stated above, it is often desirable to make a base learner sufficiently "weak". A simple but effective way to reduce the complexity of the base learner is via shrinkage towards zero. The up-date in Step 4 of the FGD algorithm is then replaced by

$$\hat{\mathbf{F}}_m(\cdot) = \hat{\mathbf{F}}_{m-1}(\cdot) + \nu \cdot \hat{w}_{m,i_m^*} \hat{f}_{m,i_m^*}(\cdot), \ 0 < \nu \leq 1. \tag{2.7}$$

Obviously, this reduces the variance of the base learner by the factor ν^2.

Remark 3. Stopping in Step 4 is important. It can be viewed as a regularization device which is very effective in complex model fitting. We find empirically that estimating M by the simple 70%-30% cross-validation scheme works well.

Summarizing the above three remarks, the functional form that the individual (squared) volatility functions in (A3) can take in our simulations and real data examples is

$$\mathbf{F}(t) = \mathbf{F}_0(\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \ldots) + \nu \sum_{m=1}^{M} \sum_{k=1}^{L} \gamma_{i_m^*,k}^{(m)} \ I_{[\mathbf{X}_{t-p}^{t-1} \in \mathcal{R}_{i_m^*,k}^{(m)}]}, \tag{2.8}$$

where the starting functions $\mathbf{F}_0(\cdot)$ are given by (2.6) and the cells $\mathcal{R}_{i_m^*,k}^{(m)}$ are constructed when fitting by least squares a regression tree to the negative gradient vector U (see step 2 of the above FGD Algorithm). This produces partitions $\{\mathcal{R}_{i,1}^{(m)}, \ldots, \mathcal{R}_{i,L}^{(m)}\}$ of the predictor space $\mathbb{R}^{pd}$

$$\cup_{k=1}^{L} \mathcal{R}_{i,k}^{(m)} = \mathbb{R}^{pd}, \ \mathcal{R}_{i,k}^{(m)} \cap \mathcal{R}_{i,h}^{(m)} = \emptyset \ (k \neq h), \ i = 1, \ldots, d.$$

Standard optimal volatility parameters in (2.8) chosen by the FGD Algorithm are $L \in \{2, 3, 5\}$ (number of end nodes), $\nu \in [0.1, 0.5]$ (shrinkage factor) and $p \in \{1, 2\}$ (number of past lags

used as predictors). In general, the estimated optimal location parameters $\gamma^{(m)}_{i^*_m,k}$ are very small (i.e. $\left|\gamma^{(m)}_{i^*_m,k}\right| \leq 0.2$), not to have overfitting. Moreover, elements of the constant conditional correlation matrix assumed in (2.2) are in the most cases higher (in absolute terms) than the ones from a classical filtered historical simulation approach.

Past multivariate returns are used to estimate the optimal partition cells in (2.8) for the chosen asset i^*_m. We have observed in the real data examples of section 5 that most of the predictors chosen by the Algorithm to estimate the optimal partition and therefore the individual (squared) volatility functions *are not* past returns from the same series. For example, we find that in the estimation of our AMEX data *more than* 85% of the optimal predictors are other firms' past lagged returns. This result supports our assumption that there is information included in past lagged observation of other firms that can be used for a better estimation and forecast of risk.

A good feature of such a FGD procedure, particularly in connection with tree-structured base learners (see Remark 2), is that it is a computationally feasible, simple method aiming to improve the initial estimates. FGD traces out a one-dimensional sequence of estimated predictions, which is feasible to optimize via choosing a stopping value M. One can alternatively try to estimate predictions for the volatility matrix V_t in (2.1) with more complex multivariate GARCH models, but this becomes quickly an intractable model-selection problem in large dimensions d. For example, if we wish to fit a multivariate BEKK model (Engle and Kroner, 1995) with $d = 10$ individual series, many of the hundreds of parameters would have to be set equal to zero in order to avoid overfitting and more than 10^{73} models would have to be fitted and checked when using a classical strategy for selecting the best subset of non-zero parameters with a model-fitting criterion, such as the Akaike's information criterion (AIC).

A feasible extension, left to future research, of our FGD algorithm could be the use, instead of our working model (2.2), of a generalization of the dynamic conditional correlation (DCC) model, recently proposed by Engle (2002), using FGD for estimating the individual conditional variances. Note that in our model (2.2) we assume constant conditional correlations only in a rolling (i.e. not fixed) time-window consisting of the last 800 observations (i.e. at time t, the time-window contains the multivariate observations $\mathbf{X}_{t-800}^{t-1}$, about three years of data), and not in the full period.

3 VaR estimation

Our VaR estimation is based on a multivariate generalization of the filtered historical simulation procedure proposed by Barone-Adesi et al. (1998) and (1999). Our simulation is based on the combination of multivariate GARCH modelling, using the FGD technique introduced in section 2, and historical asset returns. As we have already explained in section 1, the use of a multivariate GARCH model (2.2) in connection with FGD as a filter for the estimation of the standardized residuals is needed to remedy the main criticisms made about the use of standard filtered historical simulation for estimating VaR (see Pritsker, 2001). For example, the working model (2.2) allows clearly for cross-terms and consequently the (squared) volatility function $F_i(\cdot)$ of an asset i can be influenced and predicted by *all* the p-past lagged multivariate observations. This is a realistic assumption if we consider (log-) returns of different assets belonging to a common market segment (in our empirical cases the chemical or the biotechnological one).

The complete methodology stands as follows. In a first step, we filter the multivariate

standardized innovations $\mathbf{Z}_t$ using our model (2.2)

$$\mathbf{Z}_t = (\Sigma_t)^{-1}(\mathbf{X}_t - \mu_t),$$
$$V_t = \Sigma_t \Sigma_t^T = D_t R D_t,\ t = 1, \ldots, n,$$

where the individual (squared) volatility functions $\sigma_{t,i}^2 = F_i(\cdot),\ i = 1, \ldots, d$ are estimated using the FGD technique presented in the algorithm of section 2. Under the assumption (A1), the standardized innovations are i.i.d. and independent from the past.

Now, the historical standardized residuals can be drawn randomly (with replacement) and may be used to generate pathways for future returns. In other words, we use a model-based bootstrap (Efron and Tibshirani, 1993): from an i.i.d. resampling of the standardized residuals we recursively generate a time series using the structure and the fitted parameters of the estimated optimal model (2.2). Thus, we choose randomly dates with corresponding standardized innovations

$$\mathbf{Z}_1^*, \mathbf{Z}_2^*, \ldots, \mathbf{Z}_x^*, \tag{3.1}$$

where x is the time horizon at which we want to estimate the VaR (in general from 1 up to 10 days), and we construct for each asset i pathways for (squared) volatilities and returns from $t+1$ up to $t+x$ using (2.2):

$$\begin{aligned}
&\widehat{v}_{t+b,ii}^* = (\widehat{\sigma}_{t+b,i}^*)^2 = \widehat{F}_i(\{X_{t+b-s,k}^*; s = 1, 2, \ldots, p,\ k = 1, \ldots, d\}) \\
&\widehat{v}_{t+b,ij}^* = \widehat{\rho}_{ij} \sqrt{\widehat{v}_{t+b,ii}^* \widehat{v}_{t+b,jj}^*} \\
&X_{t+b,i}^* = \mu_{t+b,i}^* + (\widehat{\Sigma}_{t+b}^* \hat{\mathbf{Z}}_b^*)_i,\ b = 1, \ldots, x,\ i, j = 1, \ldots, d.
\end{aligned} \tag{3.2}$$

Note that all quantities denoted by "$\widehat{\ }$" use the fitted structure and parameters from the FGD algorithm of section 2.

The "empirical" distribution of simulated, model-based returns at the chosen time horizon x for each asset $i, i = 1, \ldots, d$, is obtained replicating the above procedure a large number of times, e.g. 2000. An estimate of the VaR at time horizon x and at level q (q in general $\in \{0.05, 0.01, 0.005\}$) is given by the corresponding q-quantile of the "empirical" returns distribution.

An alternative way to calculate VaR could be the use of extreme value theory (EVT) in connection with the popular peaks over the threshold (POT) method. Such a strategy is well illustrated in McNeil and Frey (2000). If the assumption made in the FGD estimation of section 2 (i.e. normal or scaled t_ν distributed innovations) is violated, EVT can sometimes yield better VaR predictions than the simpler empirical quantiles.

4 Simulation results

In this section, we present a simulation exercise to study the accuracy of daily VaR predictions estimated with our FGD procedure. We compare our predictions with the ones from the classical filtered historical simulation method of Barone-Adesi et al. (1999, from now we will denote BAGV) for a normal data generating process. We focus our analysis on the case of 1-day and 10-day VaR estimates for 99% confidence level. This is of particular interest since the BIS capital requirements for market risk are based on VaR at these time horizons and confidence level.

We simulate series of sample size 1500 for 12 assets from the model (2.2) with standard normally distributed innovations and various individual (squared) volatility functions F_i in (2.2,

A3). One such function is the classical GARCH(1,1) volatility

$$\begin{aligned}
&\sigma_{t,i}^2 = F_i(X_{t-1,i}, \sigma_{t-1,i}^2) \text{ where} \\
&F_i(x, \sigma^2) = \alpha_0 + \alpha_1 x^2 + \beta\sigma^2, \text{ where} \\
&\alpha_0 \sim \text{Unif}([0, 0.2]),\ \alpha_1 \sim \text{Unif}([0.05, 0.15]),\ \beta \sim \text{Unif}([0.8, 0.84])
\end{aligned} \tag{4.1}$$

and $\alpha_0, \alpha_1, \beta$ mutually independent. Another function is from a threshold model

$$\begin{aligned}
&\sigma_{t,i}^2 = F_i(X_{t-1,i}, \sigma_{t-1,i}^2) \text{ where} \\
&F_i(x, \sigma^2) = \begin{cases} \alpha_1 + \alpha_2 x^2\ , & \text{if } x \le d_1 = 0, \\ 0.2 + \alpha_3 x^2 + \alpha_4 \sigma^2\ , & \text{if } x > d_1 = 0 \text{ and } \sigma^2 \le d_2 = 0.5, \\ 0.8 + \alpha_5 \sigma^2\ , & \text{if } x > d_1 = 0 \text{ and } \sigma^2 > d_2 = 0.5, \end{cases} \quad \text{where} \\
&\alpha_1 \sim \text{Unif}([0, 0.3]),\ \alpha_2 \sim \text{Unif}([0.4, 0.6]),\ \alpha_3 \sim \text{Unif}([0.1, 0.3]), \\
&\alpha_4 \sim \text{Unif}([0.6, 0.8]),\ \alpha_5 \sim \text{Unif}([0.4, 0.6])
\end{aligned} \tag{4.2}$$

($\alpha_1, \ldots, \alpha_5$ mutually independent). A third function, in which we also allow for one cross-term, is

$$\begin{aligned}
&\sigma_{t,i}^2 = F_i(X_{t-1,i}, X_{t-1,j}, \sigma_{t-1,i}^2) \text{ where} \\
&F_i(x, y, \sigma^2) = (\alpha_1 + 0.2\,|y| + \alpha_2 x^2) \cdot (0.8 \exp(\alpha_3\,|x|\,|\sigma|)) + (0.4x^2 + \alpha_4\sigma^2)^{3/4}, \\
&\alpha_1 \sim \text{Unif}([0.05, 0.15]),\ \alpha_2 \sim \text{Unif}([0.8, 0.95]), \\
&\alpha_3 \sim \text{Unif}([-1.6, -1.4]),\ \alpha_4 \sim \text{Unif}([0.4, 0.6])
\end{aligned} \tag{4.3}$$

($\alpha_1, \ldots, \alpha_4$ mutually independent), where the component $j \in \{1, \ldots, d\} \setminus i$ is chosen randomly. We choose the simple GARCH(1,1) volatility function (4.1) for 4 assets, the threshold function (4.2) for 3 assets and the general function (4.3) allowing also for cross-terms for the remaining 5 assets. Note that also the coefficients in these functions are randomly chosen. The constant conditional correlation matrix R is chosen to mimic the one of real log-returns. This model is "fairly close" to a CCC-GARCH(1,1) model since more than half of the volatility functions involve only auto-dependence (no dependence on a cross-series in (4.1) and (4.2)), a third of them actually being linear GARCH-type where the BAGV approach is correctly specified.

We backtest the accuracy of the VaR predictions on the last 500 observations using a rolling time-window of size 1000 to estimate the parameters for the simulated 12-dimensional system. We report results with the use of $p = 2$ (number of multivariate lagged returns used as predictors), $L = 3$ (number of end-nodes in the regression trees) and $\nu = 0.5$ (shrinkage factor) in the FGD Algorithm.

Analyzing as a first step the individual (squared) volatility functions chosen by our FGD procedure, we find the following. As expected, most of the times (13 out of 14 total FGD iterations, i.e $M = 14$ in the FGD algorithm) the "best component" chosen by the FGD Algorithm in Step 4 corresponds to individual series generated with volatility functions (4.2) and (4.3), i.e. not being of a GARCH(1,1) type. This is not surprising, since the initial starting functions used in the FGD Algorithm (i.e. CCC-GARCH(1,1) estimates) have already the correct volatility structure for individual series generated with volatility function (4.1) and they can not be improved by FGD. However, in the only exception we get, a truly univariate (i.e. depending only on the same asset's past returns) volatility function is chosen.

About 60% of the times (8 out of 14 total iterations) FGD optimize individual assets simulated using the volatility function (4.2). This is also reasonable, since it has already been shown that the starting GARCH(1,1) functions yield poor performance when the true data generating

process is of a threshold type. In such cases, our FGD estimates (2.8) improve the accuracy of the initial predictions. Moreover, most of the times (about 85%) FGD chooses the correct univariate volatility functions. Similarly, when the best components correspond to series simulated using the volatility function (4.3) also allowing for a (randomly chosen) cross-term, in most cases (about 60%) the final FGD volatility functions have the correct multivariate volatility structure.

To verify the accuracy of VaR predictions using the different approaches, we compute classical mean absolute errors (MAE) and mean squared errors (MSE) (averaged across assets simulated with each specific volatility function in a first time, and averaged across all assets in a second time) by comparing them with "true VaR" predictions based on a full Monte-Carlo simulation. Goodness-of-fit results for 99%-VaR predictions at 1-day and 10-day time horizons for the BAGV method and our FGD VaR procedure are summarized in Table 4.1.

TABLE 4.1 ABOUT HERE

Table 4.1 clearly shows that the accuracy of VaR estimates computed using our FGD procedure is better globally. The largest gains are realized at the 1-day time horizon, whereas at 10-day time horizon the improvements reduce to about one third, although they are still significant (more or less 10% depending on how we measure performance). Moreover, when considering the average gains for each different individual (squared) volatility function used in the simulation, we can see that, as expected, the largest improvements are realized in the cases where both approaches are mis-specified (functions (4.2) and (4.3)), meaning that our FGD procedure, allowing for a general and more flexible functional form given by (2.8) works better than the BAGV approach. These results confirm the ones found in Audrino and Bühlmann (2003).

To end this simulation exercise, in Table 4.2 we also report correlations of the levels and changes of 1-day VaR predictions at 99% confidence level with true for the FGD VaR procedure and the BAGV method (averaged across assets simulated with each specific volatility function in a first time, and averaged across all assets in a second time).

TABLE 4.2 ABOUT HERE

Table 4.2 shows that the average correlation of the VaR estimates with "true VaR" simulated values is significantly higher using our FGD VaR procedure for all type of functions used to model the individual (squared) volatilities. Average correlations of changes in the VaR estimates with changes in true VaR are for both approaches lower. However, our strategy responds faster to changes in risk than the BAGV method. As a result, as we have already seen in Table 4.1, the VaR estimates from our procedure are more accurate.

5 Backtesting VaR for two real data examples

We backtest here the non-parametric procedure for the estimation of VaR proposed in sections 2-3 on two different real data examples. We always use parameters $L = 3$, $p = 1$ and $\nu = 0.5$ in the FGD Algorithm. Our tests are essentially the same as in Barone-Adesi et al. (2002).

The analysis is based on two criteria: statistical and economical. The former investigate the frequency and the losses exceeding the VaR predicted by our strategy (violations); the latter examine the implications of these violations (or breaks) and of the structure of the estimated VaR in economic terms.

As VaR's and asset gains and losses are calculated consistently, they can be compared directly to each other, for the corresponding number of days ahead in the holding period. We define the following:

$$\text{a } \textit{violation} \text{ (or a } \textit{break}\text{) has occurred when } \big(\text{VaR} > \text{actual asset value}\big). \tag{5.1}$$

If the model to compute VaR is correct, the actual asset losses should exceed VaR a certain number of times which corresponds to the total number of testing days multiplied by the confidence level used. This means that sometimes the VaR estimated is not sufficient to cover the actual loss. For example, for 95% confidence and 1500 testing days, we should have 75 violations (or breaks).

We focus our empirical analysis on two particular market segments of two different countries: the Swiss chemical/pharmaceutical one and the US biotechnological one. In each of the following two backtests, we stored the risk measures for five different VaR horizons ($x = 1, 2, 3, 5, 10$ days) and three different probability levels ($q \in \{0.95, 0.99, 0.995\}$). We estimate daily VaR for a period of 300 days in both the Swiss and the US example. The results from our procedure are always confronted with the ones from the BAGV method.

5.1 The Swiss chemical/pharmaceutical case

We consider five assets of the SPI chemical and pharmaceutical segment among the ones with more liquidity with 1100 daily (log-) returns (in percentages): Novartis, Roche, Serono, Ciba Spez Chemie and Sika. The data are from the time period between June 4, 1997 and August 21, 2001. We always use a rolling window of 800 days for the estimation and the parameters are re-calculated every 10 days (about two business weeks).

We estimate daily VaR for each of these five assets for a backtesting period of 300 days using the strategy proposed in sections 2-3, where we assume normal distributed innovations in the FGD algorithm. These values are then compared to the actual ones and the number of violations is recorded.

The first tests are *overall frequency tests*. In Table 5.1, we show the number of violations of each individual asset for our backtesting period (total of 300 days). The number of violations recorded for the entire backtesting period are reported in each column, where 1-Day up to 10-Day are the 1,2,3,5 and 10-day VaR horizons. We record the violations at each of the three different confidence levels for our procedure (denoted by FGD VaR) and the standard one of Barone-Adesi et al. (1999) (denoted by BAGV). The backtest results marked with an asterisk show some significant difference from the following success criterium. Under the hypothesis of unconditional unbiasedness of the VaR estimates, the numbers of violations are binomially distributed around their expected values, with standard deviation ranging from 3.77 (95% level) to 1.22 (99.5% level). A two-standard deviation interval can be used as tolerance for testing the null hypothesis of unconditional unbiasedness.

TABLE 5.1 ABOUT HERE

Table 5.1 shows that in one case (i.e. violations for the Ciba spez. Chemie firm at the 95% confidence level) the BAGV method seems to be too conservative. In general, we can see that in all tests and especially at high confidence levels and long time horizons we record more violations than expected, although the values are not significantly different from our success criterium except for one particular case (Novartis' violations at 99.5% confidence level). Both methods yield similar results. Consequently, it seems that the risk is slightly underestimated. As we have already explained in section 2, this can be due to violations of the normal assumption leading to inefficient estimates. To confirm this, we assumed t_4 distributed innovations in (2.2), allowing for more fat-tails. At high confidence levels and long time horizons we recorded in general fewer violations than the ones given in Table 5.1, although the differences were small. Analogous results can be obtained using EVT.

A second type of tests that we perform are *individual firm tests*. These tests determine whether violations occur randomly in our sample or cluster for some firms for which risk may be

miss-specified. Under the hypothesis of randomness the number of violations in the two halves of our backtesting period are independent. Therefore a cross-sectional regression of the violations which each asset reports in the first half on the number of violations recorded in the second half, should have zero slope. The values of these tests are for all confidence levels and at all time horizons for the two method proposed not significant and therefore they are not reported here.

The next step is to search for a *time clustering effect*. We apply the well-known Ljung-Box test to the time series of the number of violations observed each day. The autocorrelations in this series detect whether days with large number of violations across all assets tend to be followed by other days with large number of violations, pointing to a miss-specification of the time series model of volatility. We found no significant serial correlation's (order 1 to 6) for any confidence level at 1-day VaR horizon using our procedure and the BAGV method. Note that this result can be due to the low power of such tests to detect errors in the VaR estimates, because the serial correlation in the VaR violations can be very low even if the VaR model is not correctly specified (see Pritsker, 2001).

To end this section, we focus our backtest analysis on some economical criteria. So far, the tests that we have performed have not shown a significant difference between our multivariate procedure and the BAGV technique from Barone-Adesi et al. (1999), and both methods seem to work well and yield good VaR predictions, with only one exception (Ciba company) where the BAGV method yields too conservative VaR estimates. Now, when focusing the attention a little more on the (absolute) size of the VaR estimates obtained using the different methods, we can observe some interesting differences.

The first one, well illustrated by Figure 5.1 and Table 5.2, is that the BAGV method particularly in the periods of low returns (in absolute values) yields too conservative VaR predictions and tends to overestimate the risk. In contrast, our approach is less conservative, capturing better the passage from stressed, high volatility periods to more stable periods and vice versa.

FIGURE 5.1 AND TABLE 5.2 ABOUT HERE

Table 5.2 clearly shows that the mean of VaR estimates is in four cases out of five lower when using our strategy. This result, in connection with the ones of Tables 5.1, yields additional empirical evidence that VaR estimates from the BAGV method are in some cases too conservative. Moreover, the mean absolute difference of consecutive VaR estimates is in the most cases also smaller using our FGD VaR method, indicating that VaR estimates from our procedure change more slowly. The only exception is the Serono company. In this case the magnitude of the returns in the backtesting period is considerably larger than it is the case for all other firms. Our strategy seems to be able to accurately estimate risk also during this period, in which we observe large positive and negative returns (up to 10%).

We may interpret the average estimated individual VaR as the average necessary risk capital. Our results show that it is lower for all assets at each time horizon using our procedure. This is a consequence of our assumptions in (2.2), which allow for cross-terms and the use of more conditioning information than the BAGV method. Information can then flow from one asset to an other causing a better reaction to changes in market conditions and a further reduction of the VaR predictions (in absolute terms) during the periods characterized by small returns. Therefore, our FGD VaR procedure can achieve the same VaR coverage with less capital on average.

A second difference appears clearly when we consider the largest daily violation recorded during our backtesting period. The example illustrated in Figure 5.2 for a 3-day VaR horizon and at 99% confidence level is obtained by aggregating (with equal weights) individual asset violations.

FIGURE 5.2 ABOUT HERE

Our procedure is able to remove some of the largest aggregated violations (although in some cases yielding a large number of small ones) which occur when VaR is estimated using standard filtered historical simulation. As the results of the overall frequency tests have already shown, in this case we can improve our FGD VaR procedure and also the BAGV method by changing the assumption of normality in (2.2) for the innovations distribution. As expected, assuming scaled t_4 distributed innovations we can further reduce the number of days and the size of large aggregated violations.

To conclude the analysis, we compare the intervals within the sum of VaR's (in %, computed aggregating with equal weights individual VaR estimates for assets belonging to the SPI segment of chemi/pharma) over the different days ranges with the maximal and the mean size of aggregate violations. The results are summarized in Table 5.3.

TABLE 5.3 ABOUT HERE

At high VaR confidence levels, the maximal sum of VaR's (in %, maximal value of aggregated individual VaR's) estimated with our FGD VaR procedure is considerably smaller than those from a classical filtered historical simulation. In comparison, the maximal and the mean size of aggregate violations between the methods are very similar, with a little improvement using our strategy. Moreover, the intervals, within the sum of VaR's ranges, are in the most cases also smaller and the risk estimated with our strategy seems to be less noisy. In particular, the standard deviation of estimated VaR measuring their variability through time is lower using our procedure for each asset and at all time horizons. This is also a consequence of the use of more conditioning information allowed by our FGD VaR procedure. Therefore, our strategy can achieve the same VaR coverage with less capital on average than the BAGV method.

5.2 The US biotechnological case

We consider here all 13 assets with enough liquidity belonging to the US AMEX Biotechnology Index with 1100 daily (log-) returns (in percentages): from the Affymetrix Inc., the Amgen Inc., the Biogen Inc., the Cephalon Inc., the Chiron Corporation, the Genzyme Corporation, the Gilead Sciences Inc., the Human Genome Sciences Inc., the IDEC Pharmaceuticals Corporation, the Medimmune Inc., the Millenium Pharmaceuticals Inc., the Protein Design Labs Inc. and the Vertex Pharmaceuticals Inc. The data are from the time period between June 7, 1996 and August 24, 2000. The analysis is made using for prediction a rolling time-window of 800 days and the parameters are re-calculated every 10 days.

We estimate daily VaR for each of these thirteen companies for a backtesting period of 300 days using our FGD algorithm with normal distributed innovations and BAGV method. The estimates are then compared to the actual values and the number of violations is recorded. The tests we perform are the same already introduced for the Swiss example of section 5.1.

Individual backtest results of overall frequency tests are summarized in the Table 5.4.

TABLES 5.4 ABOUT HERE

As we can see, a relevant number of times, especially for high confidence levels, test values are significantly different from our success criterium. For example, when considering the results for the 99% confidence level, we get 9 and 17 (out of 65 tests) rejections of the null hypothesis of unconditional unbiasedness using our FGD VaR strategy and the classical BAGV method, respectively. The two methods seem to underestimate risk, although our procedure is in general better for estimating VaR and lead to a consistently smaller number of rejections.

We try now with some other tests to understand the reason why our procedure and the BAGV method yield for some assets such poor daily VaR predictions. We perform individual firm tests to determine whether violations cluster for one or two companies for which risk may be miss-specified. The values of these tests are for all confidence levels and at all time horizons not significant and therefore we do not report them here. We also search for a time clustering effect applying the Ljung-Box test to the time series of the number of aggregated violations across all companies occurring each day. The resulting values for 1-day VaR at 99% confidence level are summarized in Table 5.5, Panel A.

TABLE 5.5 ABOUT HERE

The tests reject clearly for both methods the assumption of no autocorrelations in the time series of number of aggregated violations for orders bigger than 3. This result is true for all confidence levels. A detailed analysis of the time series of aggregate violations show that breaks tend to cluster for a short period in March 2000 (10 business days) in relation with the well-known US technological market crash, where all companies registered several consecutive big losses. This can be the reason why daily VaR predictions using both strategies for these days are poor and the risk tends to be underestimated.

The values for the same overall frequency tests at the 99% confidence level and clustering tests (1-day time horizon, 99% confidence level) on the violations recorded during the backtesting period without the dates between March 3, 2000 and March 21, 2000 are summarized in Table 5.6 and 5.5 Panel B, respectively. Similar results also hold for the other confidence levels.

TABLES 5.5 AND 5.6 ABOUT HERE

Without this short period in March 2000 there is no significant serial correlation's (order 1 to 6) for any confidence level at 1-day time horizon for the remaining dates. Moreover, the most values of the overall frequency tests are now turned to be not significant. The better potential in predicting daily VaR of our FGD procedure over the BAGV method is clearly shown by the results of Table 5.6. Particularly when considering daily VaR predictions at long-time horizons (5, 10 days) for all confidence levels, our strategy results to be more attractive for risk management than the standard filtered historical simulation BAGV method. The reason of this result can be explained with the fact that our FGD VaR procedure uses a larger number of predictor variables and the volatility estimates condition on more information than the ones from the BAGV method.

We also evaluate the same economical criteria already introduced for the Swiss example. The results for the largest daily (aggregate) violations, the intervals within the sum of VaR's (equally weighted) for all companies ranges, the value of the maximal aggregate violation and the mean size of aggregate violations are similar to the ones of section 5.1. One illustrative example for the largest daily violations at 10-day time horizon and for 99.5% confidence level is shown in Figure 5.3.

FIGURE 5.3 ABOUT HERE

Using our procedure to estimate daily VaR reduces some peaks with large (aggregate) violations when compared to the ones from the BAGV method. As expected, the period of time with the largest aggregate daily violation is March 2000, where we have seen that violations tend to cluster. In particular, for the 99.5% confidence level and at the 10-day time horizon, the maximal aggregate violation and the mean size of violations are significantly larger (17.21% vs. 14.88% and 3.56% vs. 3.24%) using the BAGV method. Analogously to the Swiss example, we found that BAGV tends to yield too conservative VaR predictions in the periods of low (absolute)

returns. The capital needed to cover possible losses is on average smaller when estimating daily VaR with our procedure.

6 Conclusions

We have presented a non-parametric technique to construct daily VaR estimates. Our strategy is based on a multivariate FGD algorithm, which is a method for estimating the conditional covariance matrix in (2.1), in connection with historical simulation. The use of multivariate GARCH-type models as a filter for historical simulation improves the BAGV method, based on filtered historical simulation (Barone-Adesi et al., 1998 or 1999). For example, our technique allows for cross-terms and the conditional correlation matrix is assumed to be constant only in a rolling (i.e. not fixed) time-window. So far, the use of multivariate GARCH-type models (for example BEKK models) for the estimation and the prediction of the conditional covariance matrix (2.1) for large dimensions was a huge challenge computationally and in most cases led to an intractable model-selection problem. Our FGD algorithm solves these problems: it is computationally feasible in multivariate set-ups with dozens up to hundreds of return series. This is the most attractive feature of FGD.

Our simulation exercise and our tests on two real data-sets belonging to the pharmaceutical and biotechnological market segments show that our technique produces accurate and powerful daily VaR estimates, significantly outperforming the VaR predictions from the BAGV method. The results of the backtests show that our multivariate FGD VaR technique, conditioning on more information, has the ability to correct the inaccuracies, which sometimes occur using the BAGV method, yielding better risk estimates. Moreover, we found that BAGV tends to overestimate risk during the periods of low volatility. Summarizing, there is empirical evidence that our procedure can achieve the same VaR coverage with less capital on average than the BAGV method.

In the backtests of section 5, it is shown that our procedure can be further improved if the daily VaR predictions from the standard FGD algorithm are not satisfactory (for example, with a modification of the assumption about the distribution of the innovations in (2.2) or allowing for more complex, time-varying conditional correlation dynamics). These extensions are left to future research.

References

Aguilar, O. and West, M. (2000). Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. *Journal of Business and Economic Statistics* **18**, 338-357.

Alexander, C. (2001). A primer on the orthogonal GARCH model. ISMA Centre, Mimeo.

Audrino, F. and Bühlmann, P. (2003). Volatility Estimation with Functional Gradient Descent for Very High-Dimensional Financial Time Series. *Journal of Computational Finance* **6**, No. 3, 65-89.

Barone-Adesi, G., Bourgoin, F. and Giannopoulos, K. (1998). Don't Look Back. *Risk* **11**, August, 100-104.

Barone-Adesi, G., Giannopoulos, K. and Vosper, L. (1999). VaR Without Correlations for Portfolio of Derivative Securities. *Journal of Futures Markets* **19** (April), 583-602.

Barone-Adesi, G., Giannopoulos, K. and Vosper, L. (2002). Backtesting Derivative Portfolios with FHS. *European Financial Management* **8**, 31-58.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The Review of Economics and Statistics* **72**, 498-505.

Breiman, L. (1999). Prediction games & arcing algorithms. *Neural Computation* **11**, 1493-1517.

Bühlmann, P. and Yu, B. (2003). Boosting with the L_2-loss: regression and classification. *Journal of the American Statistical Association*, forthcoming.

Chib, S., Nardari, F. and Shephard, N. (1999). Analysis of high dimensional multivariate stochastic volatility models. Preprint, University of Oxford.

Dowd, K. (1998). *Beyond value-at-Risk: The New Science of Risk Management*, Wiley.

Duffie,D. and Pan, J. (1997). An Overview of Value at Risk. *Journal of Derivatives* **4**, Spring, 7-49.

Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, London.

Engle, R.F. (2002). Dynamic Conditional Correlation: a simple class of multivariate GARCH models. To appear in the *Journal of Business and Economic Statistics*.

Engle, R.F. and Kroner, K.F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory* **11**, 122-150.

Engle, R.F., Ng, V.K. and Rothschild, M. (1990). Asset pricing with a factor ARCH covariance structure: empirical estimates for treasury bills. *Journal of Econometrics* **45**, 231-238.

Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics* **29**, 1189-1232.

Friedman, J.H., Hastie, T. and Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting. *Annals of Statistics* **28**, 337-407 (with discussion).

Harvey, A.C., Ruiz, E. and Shephard, N. (1994). Multivariate stochastic variance models. *Rev. Economic Studies* **61**, 247-264.

Jorion, P. (2001). *Value-at-Risk: the New Benchmark for Controlling Market Risk.* McGraw-Hill, Chicago.

Lin, W.-L. (1992). Alternative estimators for factor GARCH models – a Monte Carlo comparison. *Journal of Applied Econometrics* **7**, 259-279.

Mason, L., Baxter, J. Bartlett, P. and Frean, M. (1999). Functional gradient techniques for combining hypotheses. In *Advances in Large Margin Classifiers.* MIT Press.

McNeil, A.J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* **7**, 271-300.

Mikosch, T. and Starica, C. (1999). Change of structure in financial data, long-range dependence and GARCH modelling. Technical report. University of Groningen.

Pritsker, M. (2001). The Hidden Danger of Historical Simulation. Working Paper (April), University of California, Berkeley.

Venter, J.H. and de Jongh, P.L. (2001). Risk Estimation using the Normal Inverse Gaussian Distribution. *Journal of Risk* **4**, Number 2.

1-day time horizon

Function	Assets	Model	MAE	MSE
Function (4.1)	4	BAGV	0.2471	0.0954
		FGD VaR	0.2409 (2.5%)	0.0907 (4.9%)
Function (4.2)	3	BAGV	0.5292	0.4560
		FGD VaR	0.3580 (32.4%)	0.2402 (47.3%)
Function (4.3)	5	BAGV	0.5009	0.4398
		FGD VaR	0.4231 (15.5%)	0.3112 (29.2%)
Global	12	BAGV	0.4234	0.3290
		FGD VaR	0.3461 (18.3%)	0.2199 (33.2%)

10-day time horizon

Function	Assets	Model	MAE	MSE
Function (4.1)	4	BAGV	0.4508	0.3327
		FGD VaR	0.4438 (1.6%)	0.3301 (0.8%)
Function (4.2)	3	BAGV	0.8426	0.9525
		FGD VaR	0.7286 (13.5%)	0.7064 (25.8%)
Function (4.3)	5	BAGV	0.5683	0.5295
		FGD VaR	0.5448 (4.1%)	0.4757 (10.2%)
Global	12	BAGV	0.5977	0.5697
		FGD VaR	0.5571 (6.8%)	0.4849 (14.9%)

Table 4.1: Mean absolute errors (MAE) and mean squared errors (MSE) (averaged across assets simulated with each specific volatility function a first time, and averaged across all assets a second time) of VaR predictions at 99% confidence level for 1-day (top) and 10-day (bottom) time horizons computed using the FGD VaR procedure and the BAGV method for a 12-dimensional data set simulated using different individual (squared) volatility functions. Improvements over the classical BAGV method are given between parenthesis. "Assets" denotes the total number of assets for which we choose each specific volatility function to simulate the data.

Function	Assets	Model	Corr. levels	Corr. changes
Function (4.1)	4	BAGV	0.7307	0.5255
		FGD VaR	0.8245	0.6105
Function (4.2)	3	BAGV	0.6242	0.3572
		FGD VaR	0.7875	0.6166
Function (4.3)	5	BAGV	0.3650	0.2668
		FGD VaR	0.4924	0.4180
Global	12	BAGV	0.6026	0.4165
		FGD VaR	0.6956	0.5367

Table 4.2: Correlations of the levels and changes of 1-day VaR predictions at 99% confidence level with true VaR (averaged across assets simulated with each specific volatility function a first time, and averaged across all assets a second time) using the FGD VaR procedure and the BAGV method for the same simulated 12-dimensional data set of Table 4.2. "Assets" denotes the total number of assets for which we choose each specific volatility function to simulate the data.

95% confidence level

Asset	Expected	Model	1-Day	2-day	3-Day	5-Day	10-Day
Novartis	15	BAGV	17	17	16	15	14
		FGD VaR	17	18	15	16	13
Roche	15	BAGV	16	16	15	15	16
		FGD VaR	15	17	17	16	15
Serono	15	BAGV	20	19	21	20	21
		FGD VaR	17	19	20	18	19
Ciba	15	BAGV	8	7*	8	7*	7*
		FGD VaR	10	10	9	9	8
Sika	15	BAGV	19	19	18	18	20
		FGD VaR	19	20	19	20	19

99% confidence level

Asset	Expected	Model	1-Day	2-day	3-Day	5-Day	10-Day
Novartis	3	BAGV	4	5	5	6	5
		FGD VaR	5	6	5	6	6
Roche	3	BAGV	3	3	3	4	3
		FGD VaR	4	4	3	4	5
Serono	3	BAGV	6	6	6	4	4
		FGD VaR	5	4	5	6	5
Ciba	3	BAGV	3	3	3	3	2
		FGD VaR	3	3	2	3	2
Sika	3	BAGV	5	5	3	6	6
		FGD VaR	3	4	4	4	6

99.5% confidence level

Asset	Expected	Model	1-Day	2-day	3-Day	5-Day	10-Day
Novartis	1.5	BAGV	4*	4*	5*	5*	5*
		FGD VaR	3	4*	4*	3	3
Roche	1.5	BAGV	1	2	3	2	1
		FGD VaR	2	3	3	2	3
Serono	1.5	BAGV	3	3	2	2	2
		FGD VaR	3	3	3	3	3
Ciba	1.5	BAGV	2	2	2	2	1
		FGD VaR	2	3	2	2	2
Sika	1.5	BAGV	2	3	2	1	3
		FGD VaR	1	3	3	2	3

Table 5.1: Overall frequency tests: violations for five assets belonging to the SPI segment of chemi/pharma recorded during the backtesting period between June 28, 2000 and August 21, 2001, at the 95% (top), 99% (middle) and 99.5% (bottom) confidence levels. Backtest results marked with an asterisk show a rejection of the null hypothesis of independence.

Asset	Model	mean(VaR_t)	mean($\lvert\mathrm{VaR}_t - \mathrm{VaR}_{t-1}\rvert$)
Novartis	BAGV	-3.647	0.437
	FGD VaR	-3.525	0.403
Roche	BAGV	-3.302	0.452
	FGD VaR	-3.039	0.441
Serono	BAGV	-7.679	0.734
	FGD VaR	-8.052	0.891
Ciba	BAGV	-3.306	0.352
	FGD VaR	-3.038	0.350
Sika	BAGV	-5.399	0.842
	FGD VaR	-5.287	0.702

Table 5.2: Backtest analysis: mean of VaR estimates (in %) and of absolute differences of consecutive VaR estimates (in %) at 99% confidence level and 1-day time horizon of five firms belonging to the SPI segment of chemi/pharma obtained using the FGD strategy proposed in sections 2-3 and the BAGV method from Barone-Adesi et al. (1999) for the entire backtesting period between June 28, 2000 and August 21, 2001.

C. level	Horizon	Model	Sum of VaR's interval	Max viol.	Mean size of viol.
95%	1-Day	BAGV	2.20 to 3.91	6.53	1.68
		FGD VaR	2.00 to 3.64	6.76	1.61
95%	10-Day	BAGV	2.31 to 3.57	6.59	1.74
		FGD VaR	2.16 to 3.43	6.62	1.73
99.5%	1-Day	BAGV	3.82 to 7.24	7.02	2.70
		FGD VaR	3.94 to 6.80	6.86	2.62
99.5%	10-Day	BAGV	4.35 to 7.66	3.75	2.48
		FGD VaR	4.16 to 6.76	3.84	2.28

Table 5.3: Intervals for the sum of VaR's, the maximal aggregate violation and the mean size of aggregate violations (in %) estimated using our FGD VaR procedure and classical BAGV method from Barone-Adesi et al. (1999) for five assets belonging to the SPI segment of chemi/pharma (aggregated with equal weights) over the backtesting period between June 28, 2000 and August 21, 2001.

Asset	Expected	Model	1-Day	2-day	3-Day	5-Day	10-Day
Affymetrix	15/3/1.5	BAGV	25*/3/2	24*/4/1	27*/4/3	22/5/3	22/5/4*
		FGD VaR	21/4/2	22/5/3	22/5/3	20/6/4*	21/5/5*
Amgen	15/3/1.5	BAGV	23*/6/2	23*/6/2	24*/4/1	25*/6/4*	29*/9*/7*
		FGD VaR	21/6/2	21/3/1	19/6/1	21/6/3	22/4/3
Biogen	15/3/1.5	BAGV	18/3/3	15/3/3	14/3/3	17/3/3	20/3/3
		FGD VaR	17/3/3	16/3/3	17/3/3	14/4/3	18/3/3
Cephalon	15/3/1.5	BAGV	20/4/3	19/5/1	18/3/1	18/5/2	15/3/2
		FGD VaR	21/4/2	18/6/1	18/3/1	16/3/1	16/4/1
Chiron	15/3/1.5	BAGV	19/8*/5*	16/7*/7*	19/7*/5*	20/7*/6*	21/6/4*
		FGD VaR	16/6/3	16/6/6*	18/6/4*	19/6/3	22/3/2
Genzyme	15/3/1.5	BAGV	20/6/4*	21/8*/4*	18/6/5*	18/6/3	24*/7*/4*
		FGD VaR	18/6/5*	17/6/5*	17/6/3	16/5/2	22/7*/3
Gilead	15/3/1.5	BAGV	19/5/2	19/5/1	21/5/3	22/4/2	21/5/2
		FGD VaR	20/5/3	17/4/3	21/3/2	21/3/1	21/3/2
H. Genome	15/3/1.5	BAGV	18/5/3	18/6/6*	18/6/5*	20/7*/7*	19/6/5*
		FGD VaR	16/6/3	15/6/3	17/6/3	20/7*/5*	18/6/3
IDEC	15/3/1.5	BAGV	21/7*/4*	24*/5/3	23*/7*/2	19/5/3	25*/8*/5*
		FGD VaR	22/5/1	27*/5/3	26*/6/3	20/5/3	23*/7*/4*
Medimmune	15/3/1.5	BAGV	26*/9*/5*	30*/8*/6*	27*/8*/3	32*/10*/3	35*/10*/4*
		FGD VaR	23*/9*/3	27*/9*/3	24*/5/2	30*/7*/3	35*/8*/3
Millenium	15/3/1.5	BAGV	18/4/4*	17/4/4*	19/6/5*	19/6/3	17/6/4*
		FGD VaR	17/5/3	17/6/3	16/6/5*	19/5/3	18/6/4*
P. Design	15/3/1.5	BAGV	18/5/3	18/9*/5*	17/5/3	21/5/3	26*/5/5*
		FGD VaR	19/5/2	16/6/3	18/8*/3	21/6/3	21/7*/3
Vertex	15/3/1.5	BAGV	15/4/2	16/4/2	19/5/1	22/2/2	20/5/3
		FGD VaR	16/4/3	15/4/2	16/3/3	22/3/0	18/4/1

Table 5.4: Overall frequency tests: violations for the thirteen assets belonging to the AMEX Biotechnology Index recorded during the backtesting period between July 1, 1999 and August 24, 2000, at the 95%/99%/99.5% confidence levels. The values marked with an asterisk lead to a rejection of the null hypothesis of unconditional unbiasedness of VaR estimates.

Time clustering effect: A

	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6
FGD VaR	3.799	6.065*	6.950	22.57*	23.08*	23.49*
BAGV	3.823	5.611	8.116*	20.61*	20.62*	20.88*

Time clustering effect: B

	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6
FGD VaR	0.143	1.797	3.663	7.131	7.357	7.779
BAGV	0.003	2.199	3.764	7.681	8.602	8.961

Table 5.5: Time clustering effect: Ljung-Box tests for the number of aggregated violations for 99% confidence level and at 1-day time horizon for thirteen companies belonging to the AMEX Biotechnology Index recorded during the entire backtesting period between July 1, 1999 and August 24, 2000 (Panel A) and during the backtesting period without the dates between March 9, 2000 and March 22, 2000 (Panel B). The values marked with an asterisk are significantly different from our null hypothesis of no autocorrelations in the time series of number of aggregated violations.

Asset	Expected	Model	1-Day	2-day	3-Day	5-Day	10-Day
Affymetrix	3	BAGV	1	2	2	2	3
		FGD VaR	3	4	3	3	4
Amgen	3	BAGV	5	5	3	3	7*
		FGD VaR	5	2	5	4	4
Biogen	3	BAGV	3	3	3	2	3
		FGD VaR	3	3	3	3	3
Cephalon	3	BAGV	3	4	2	2	3
		FGD VaR	3	5	2	1	3
Chiron	3	BAGV	7*	6	6	7*	5
		FGD VaR	5	4	5	5	2
Genzyme	3	BAGV	4	6	4	4	7*
		FGD VaR	4	4	4	3	5
Gilead	3	BAGV	5	5	5	3	3
		FGD VaR	5	4	3	3	3
Human Genome	3	BAGV	3	4	4	7*	4
		FGD VaR	4	4	4	5	4
IDEC	3	BAGV	7*	5	7*	5	7*
		FGD VaR	5	5	6	5	5
Medimmune	3	BAGV	7*	6	6	8*	8*
		FGD VaR	6	6	4	5	4
Millenium	3	BAGV	3	3	5	4	4
		FGD VaR	3	4	4	3	6
Protein Design	3	BAGV	3	5	2	2	5
		FGD VaR	4	4	4	3	4
Vertex	3	BAGV	3	4	4	1	3
		FGD VaR	3	3	3	2	3

Table 5.6: Overall frequency tests: violations for the thirteen assets belonging to the AMEX Biotechnology Index recorded during the entire backtesting period without the dates between March 9, 2000 and March 22, 2000 (10 days), at the 99% confidence level. The values marked with an asterisk lead to a rejection of the null hypothesis of unconditional unbiasedness of VaR estimates.

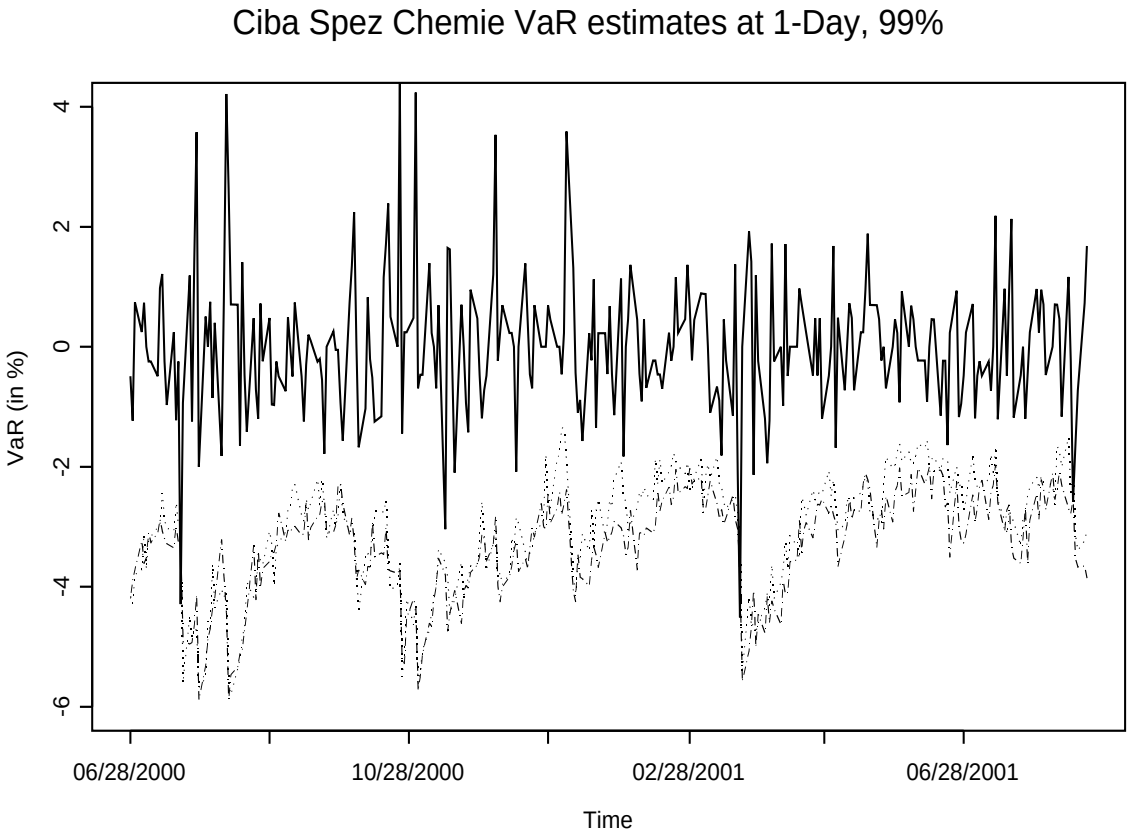

Figure 5.1: Backtest analysis: VaR estimates at 99% confidence level and 1-day time horizon obtained using the FGD strategy proposed in sections 2-3 (dotted line) and the BAGV method from Barone-Adesi et al. (1999) (dashed line) superimposed on the (log-) returns of the Ciba Spez Chemie firm for the entire backtesting period between June 28, 2000 and August 21, 2001.

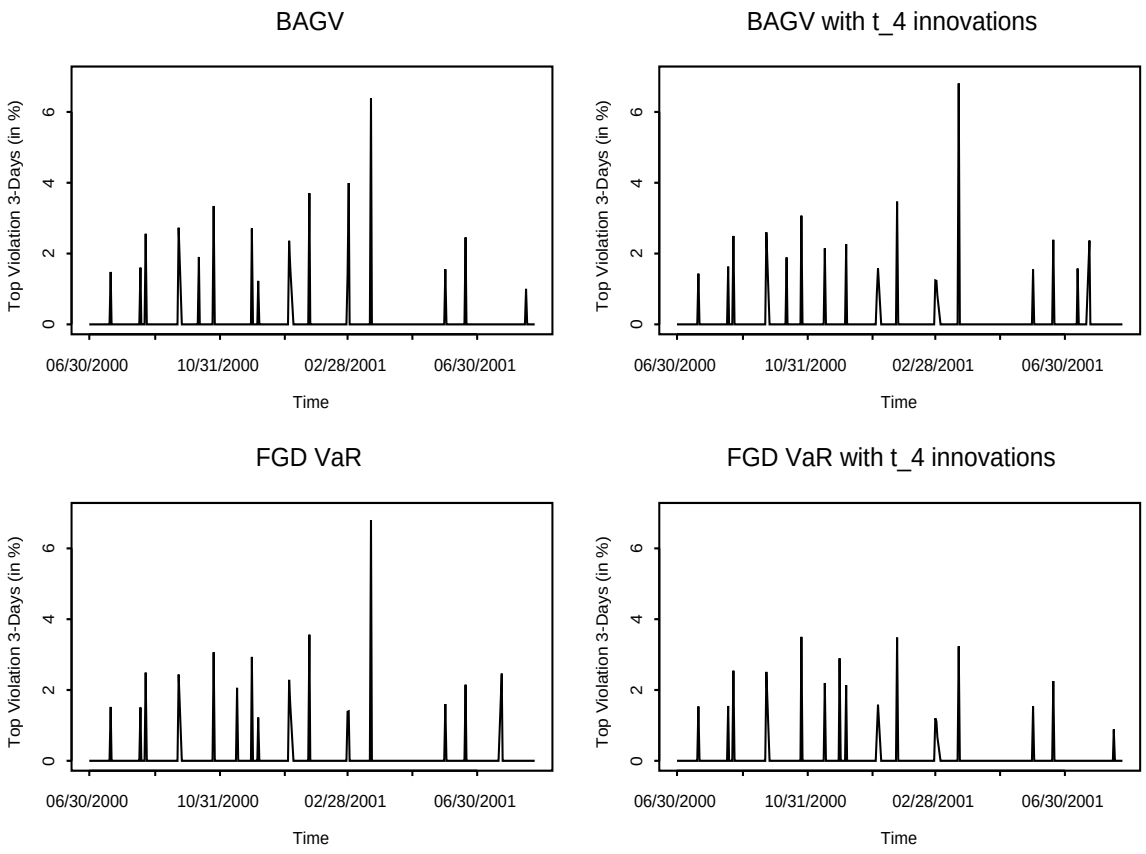

Figure 5.2: Backtest analysis for five assets belonging to the SPI segment of chemi/pharma: largest daily violations (equally weighted, in %) for a 3-day VaR horizon and at 99% confidence level. They are computed aggregating individual asset violations. Results are shown for our FGD VaR procedure with normal innovations (bottom left), our FGD VaR procedure with t_4 innovations (bottom right), the standard BAGV method from Barone-Adesi et al. (1999) with normal innovations (top left) and the BAGV method with t_4 innovations (top right). The backtesting period goes from June 28, 2000 to August 21, 2001.

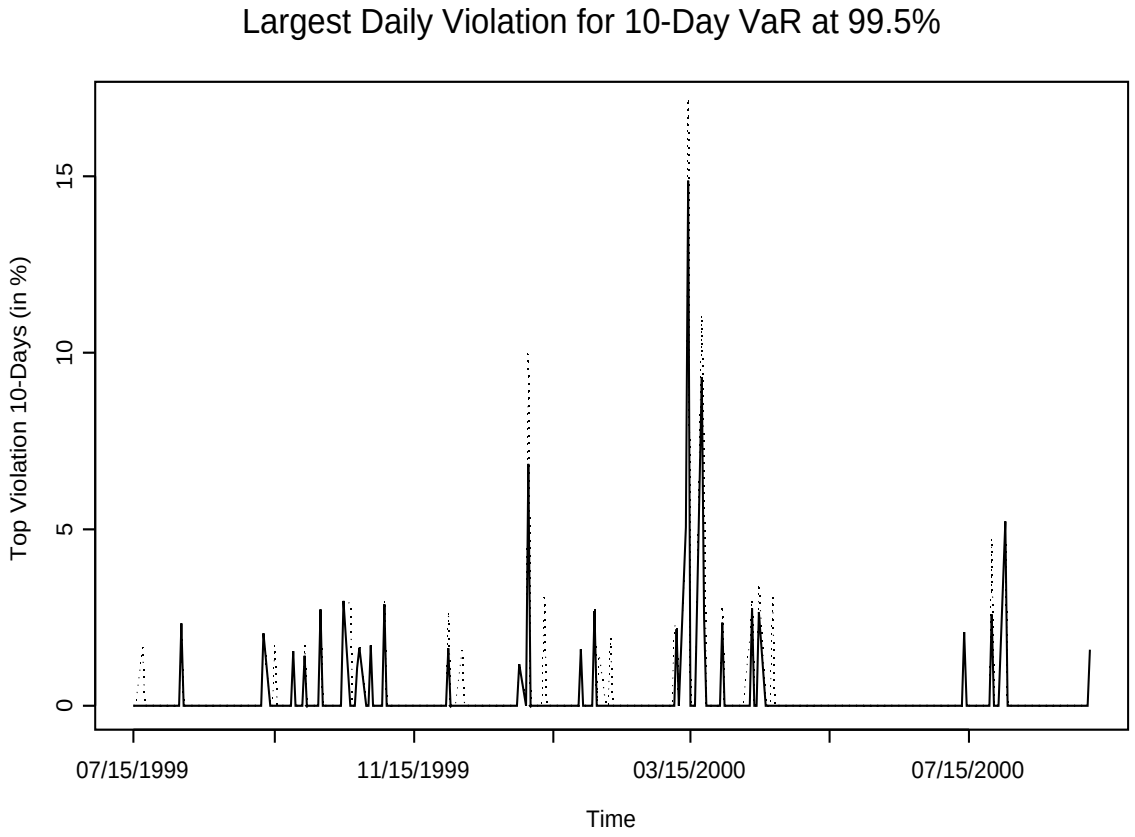

Figure 5.3: Backtest analysis for thirteen assets belonging to the AMEX Biotechnology Index: largest daily violations (equally weighted, in %) for a 10-day VaR horizon and at 99.5% confidence level. They are obtained aggregating individual asset violations. Results from our FGD VaR procedure and from the BAGV method are shown by solid lines and dotted lines, respectively. The backtesting period goes from July 1, 1999 to August 24, 2000.