

# A multivariate FGD technique to improve VaR computation in equity markets

by

Francesco Audrino\*  
University of Southern Switzerland

and

Giovanni Barone-Adesi  
University of Southern Switzerland

**Abstract.** We present a multivariate, non-parametric technique for constructing reliable daily VaR predictions for individual assets belonging to a common equity market segment, which takes also into account the possible dependence structure between the assets and is still computationally feasible in large dimensions. The procedure is based on functional gradient descent (FGD) estimation for the volatility matrix (see Audrino and Bühlmann, 2002) in connection with asset historical simulation and can also be seen as a multivariate extension of the filtered historical simulation method proposed by Barone-Adesi et al. (1998). Our FGD algorithm is very general and can be further adapted to other multivariate problems dealing with (volatility) function estimation. We concentrate our empirical investigations on the Swiss pharmaceutical and the US biotechnological equity market and we collect, using statistical and economical backtests, strong empirical evidence of the better predictive potential of our multivariate strategy over other univariate techniques, with a resulting significant improvement in the measurement of risk.

**Keywords:** Volatility estimation; Filtered Historical Simulation; Value-at-Risk.

**JEL codes:** C14, C19.

\*Corresponding author: USI, Institute of Finance, Via Buffi 13, 6900 Lugano, Switzerland.  
e-mail: audrinof@lu.unisi.ch

# A multivariate FGD technique to improve VaR computation in equity markets

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

September 2002

## Abstract

We present a multivariate, non-parametric technique for constructing reliable daily VaR predictions for individual assets belonging to a common equity market segment, which takes also into account the possible dependence structure between the assets and is still computationally feasible in large dimensions. The procedure is based on functional gradient descent (FGD) estimation for the volatility matrix (Audrino and Bühlmann, 2002) in connection with asset historical simulation and can also be seen as a multivariate extension of the filtered historical simulation method proposed by Barone-Adesi et al. (1998). Our FGD algorithm is very general and can be further adapted to other multivariate problems dealing with (volatility) function estimation. We concentrate our empirical investigations on the Swiss pharmaceutical and the US biotechnological equity market and we collect, using statistical and economical backtests, strong empirical evidence of the better predictive potential of our multivariate strategy over other univariate techniques, with a resulting significant improvement in the measurement of risk.

Key words: Volatility estimation; 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 derivative 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 historical observation period. Or, in other words, a VaR calculation amounts of course 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.

---

\*I thank Peter Bühlmann for some interesting remarks.

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 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 here the results found for the Swiss chemical/pharmaceutical and the US biotechnological equity market segment.

Although VaR is conceptually a simple measure of risk, computing it in practice with multivariate methods for large equity markets can be very difficult when taking into account also the dependence structure between the different assets. We present a non-parametric technique for constructing accurate daily VaR estimates for individual assets, which is still computationally feasible for multivariate problems in large dimensions, where so far it does not exist any alternative competitive method. Our strategy is based on functional gradient descent (FGD) estimation, a recent technique for the classification problem from the area of machine learning (Mason et al., 1999; Breiman, 1999; Friedman et al., 2000; Friedman, 2001), for the multivariate conditional covariance matrix of the individual assets in connection with historical simulation. The FGD algorithm that we propose is the same as in Audrino and Bühlmann (2002), 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 (volatility) function estimation, such as for example asset allocation problems or risk management for global trading portfolios of large trading banks with time dependent weights. In this instance VaR has been so far computed only using approximations and univariate models (Berkowitz and O'Brien, 2002).

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. So far, it was not possible to use 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 the most parameters would have to be set to zero in order to avoid overfitting. Using FGD this problem can be overcome: this technique can be also 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 try to improve, often successfully, those components where the initial predictions are poorest. Clearly, as Audrino and Bühlmann (2002) have already shown, we can not expect to learn in all  $d$  dimensions when increasing  $d$  and keeping sample size fixed. However, although the gain on average will generally decrease, FGD still improve 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 recursively generate 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) and (2002) based on filtered historical simulation, where we use a multivariate GARCH-type model in connection with FGD for filtering the residuals.

Our strategy contrasts well most 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 the estimation cross-terms as predictor variables. This is a reasonable assumption if we consider that assets belonging to a common market segment (like in our case) show some dependence structure and it is conceivable that one asset can be influenced and predicted by the past values of some other. This possibility have not been considered in the filtered historical simulation method proposed by Barone-Adesi et al. (1998), 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 a constant conditional correlation (CCC)-type 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 is left to future research.

The third and last critique on the use of the filtered historical simulation method by Barone-Adesi et al. (1998) for estimating VaR is an empirical one: Pritsker found in his investigations that this method is not able to accurately estimate VaR for long time horizons and at high confidence levels (for example for 10-day VaR at 99%) using 2 years of historical data. This is due to a lack of extreme outliers in the filtered data set. One solution can be the use of a longer span of historical data, which may improve the VaR predictions by allowing for more extreme observations. Unluckily, this will not be consistent with the use of fixed conditional correlations, assumption which is clearly violated if we consider longer periods of historical data. On the other side, we found through a simulation exercise and in our real examples that our multivariate procedure yields more accurate VaR predictions by allowing the information to flow through the different series; the lack of extreme outliers in a particular series is filled up by multivariate modelling.

Using different statistical and economical backtests, we collect empirical evidence of the better predictive potential of our multivariate procedure over other univariate techniques, and in particular the filtered historical simulation method of Barone-Adesi et al. (1998). The VaR estimates for the individual assets belonging to a common equity market segment are more accurate using our technique 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 a simulation exercise which proves that multivariate modelling can correct the possible inaccuracies of daily VaR predictions estimated with the univariate method of Barone-Adesi et al. (1998). The results of our real empirical investigations and the backtest analysis are summarized in section 5. Section 6 concludes the paper.

## 2 Volatility estimation with Functional Gradient Descent

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

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

We assume stationarity of the returns (at least within a suitable time-window). In the empirical investigations of section 5, we always report results using 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 (2002) have already shown, a slightly modified (from the generic one of Friedman et al., 2000 or Friedman, 2001) Functional Gradient Descent (FGD) technique is a powerful strategy to construct computable and good predictions for the multivariate volatility

matrix

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

where  $\mathcal{F}_{t-1}$  denotes the information available up to time  $t-1$ , i.e. the  $\sigma$ -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 hundreds) where predicting the multivariate volatility matrix raises huge challenges in computational and modelling issues due to the well-known curse of dimensionality. In such a case FGD is one of the few non parametric techniques (if not the only one so far) which are feasible.

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

$$\mathbf{X}_t = \mu_t + \Sigma_t \mathbf{Z}_t, \quad (2.2)$$

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  $\text{Cov}(\mathbf{Z}_t) = I_d$ . 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  $V_t = \text{Cov}(\mathbf{X}_t | \mathcal{F}_{t-1}) = \Sigma_t \Sigma_t^T$  is almost surely positive definite for all  $t$ . The typical element of  $V_t$  is  $v_{t,ij} = \rho_{ij} (v_{t,ii} v_{t,jj})^{1/2}$  ( $i, j = 1, \dots, d$ ). The parameter  $\rho_{ij} = \text{Corr}(X_{t,i}, X_{t,j} | \mathcal{F}_{t-1})$  equals the constant conditional correlation and hence  $-1 \leq \rho_{ij} \leq 1$ ,  $\rho_{ii} = 1$ .
- (A3) (functional form) The conditional variances are of the form

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

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

- (A4) (conditional mean) The conditional mean  $\mu_t$  is of the form

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

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

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

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

The functional form (A3) allows clearly for cross-terms, since the conditional variances of one series depends on the 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 the past information from other risk factors.

We propose FGD for estimating the (squared) individual volatility functions  $F_i(\cdot)$  in (A3), where we restrict  $F_i(\cdot) : \mathbb{R}^{pd} \rightarrow \mathbb{R}^+$  with  $p$  finite, i.e. involving the first  $p$  lagged multivariate observations. The main idea of FGD is to find the estimates for the functions  $F_i(\cdot)$  which minimize a suitable loss function  $\lambda$ , under the constraint which requires 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 trees, projection pursuit or neural nets. For more details, we remand to Friedman et al. (2000), Friedman (2001) and, for the applications in the financial field, Audrino and Bühlmann (2002).

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), we have that 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

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

where the terms  $d \log(2\pi) / 2$  and  $\log(\det(R)) / 2$  are constants and could be dropped. As pointed out with the subscript, the loss function  $\lambda_R$  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  $\hat{\mathbf{F}} = (\hat{F}_1, \dots, \hat{F}_d)$ , we build the residuals

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

and define

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

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

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

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  $i = 1, \dots, d$ .

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, a possible alternative can be to assume a fat-tailed distribution for the innovations (such as for example a scaled  $t_\nu$  distribution with fixed degrees of freedom parameter  $\nu$ ), 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).

The FGD algorithm, which is the same FGD algorithm for multivariate volatility as in Audrino and Bühlmann (2002), looks as follows.

---

### FGD algorithm

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

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

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  $i = 1, \dots, d$ , do the following.

*Step 2<sub>i</sub> (projection of component gradients to base learner).* Compute the negative gradient

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

This is explicitly given in (2.5). Then, fit the negative gradient vector  $U_i = (U_{p+1,i}, \dots, U_{n,i})^T$  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<sub>i</sub> (line search).* Perform a one-dimensional optimization for the step-length,

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

$(\hat{\mathbf{F}}_{m-1}(t) + w \hat{f}_{m,i}(\cdot))$  is defined as the function which is constructed by adding in the  $i$ th 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

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

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 with 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 in the financial field 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

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

We construct the estimates with maximum likelihood from the  $d$  individual series, ignoring the more general correlation structure in  $R$ , causing some statistical decrease in efficiency, but gaining 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 conditional 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, 2001). We choose decision trees as base learners, since particularly in high dimensions, they have the ability to do variable selection 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.

For the reasons explained 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

$$\hat{\mathbf{F}}_m(\cdot) = \hat{\mathbf{F}}_{m-1}(\cdot) + \nu \cdot \hat{w}_{m,i_m^*} \hat{f}_{m,i_m^*}(\cdot), \quad 0 < \nu \leq 1. \quad (2.6)$$

Obviously, this reduces the variance of the base learner by the factor  $\nu^2$ .

*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.

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 may 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 for example 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 working model (2.2) we assume constant conditional correlations only in a rolling (not fixed) time-window, which contains the last 800 days of data (about three years) used for the estimation (i.e. at time  $t$ , the time-window contains the multivariate observations  $\mathbf{X}_{t-800}^{t-1}$ ), and not in the full period. This brings us to the next section, where we present the technique we use for estimating VaR.

### 3 VaR estimation

We calculate VaR following a multivariate generalization of the filtered historical simulation procedure proposed by Barone-Adesi et al. (1998). 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 the multivariate GARCH model (2.2) in connection with FGD as a filter for the estimation of the standardized residuals is motivated by the need to overcome the main criticisms about VaR estimation using standard filtered historical simulation (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  $\mathbf{Z}_t$  with our model (2.2)

$$\begin{aligned}\mathbf{Z}_t &= (\Sigma_t)^{-1}(\mathbf{X}_t - \mu_t), \\ V_t &= \Sigma_t \Sigma_t^T = D_t R D_t, \quad t = 1, \dots, n,\end{aligned}$$

where the individual (squared) volatility functions  $\sigma_{t,i}^2 = F_i(\cdot)$ ,  $i = 1, \dots, d$  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

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

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):

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

Note that all quantities denoted by “ $\widehat{\phantom{x}}$ ” 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  $i, i = 1, \dots, d$ , 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) and, if we believe that the assumption made in the FGD estimation of section 2 (i.e. normal or scaled  $t_\nu$  distributed innovations) is violated, it can sometimes behave like a correction and yield better VaR predictions than the simpler empirical quantiles.

## 4 A simulation exercise

In this section, we present a simulation exercise to study the differences of daily VaR predictions estimated with the filtered historical simulation method of Barone-Adesi et al. (1998), which from now on we denote with BAGV (deriving from the names of the authors), in comparison with the ones obtained from our multivariate procedure. We will confront our results with the ones found by Pritsker (2001) in his investigations. We concentrate our analysis on the case of 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 this horizon and confidence level.

To focus exclusively on the differences of the two approaches, the simulations are conducted under ideal conditions for using the BAGV method. In particular, exactly as in Pritsker (2001), we assume that the parameters of the GARCH processes are estimated exactly, the filtered innovations that are used in the simulations are the true filtered innovations, and that the starting conditional volatility is known, and that all pricing models are correct.

To examine the accuracy of the 10-day VaR estimates from our procedure and the BAGV method, we generated 800 days of random data for five assets among the ones of the AMEX Biotechnological Index, and then used the simulated data with the true GARCH parameters to estimate VaR at a 10-day horizon using our FGD strategy and BAGV. The VaR estimates from these methods are then evaluated by comparing them with VaR predictions based on a full Monte-Carlo simulation.

The data are generated exactly with the same methodology already given in Pritsker (2001), Appendix B. We assume that the primitive shocks are independent scaled  $t_4$  random variables (i.e. with variance equal to one) and that the beginning conditional volatilities for each asset are chosen by picking days between June 1996 and June 1999 and starting the process with that day’s conditional volatilities. To keep the results from depending on any particular set of filtered innovations, the results are based on 100 independent simulations.

To verify the accuracy of the 10-day VaR predictions for 99% confidence level from the BAGV method and our FGD procedure, we consider differences between the true VaR's and the estimates from the two different strategies and use the concept of hypothesis testing.

For each individual asset, we consider the differences between the true 10-day VaR and the 10-day VaR estimates from the two methods for 100 independent simulations

$$\widehat{D}_t = \text{VaR}_{k;\text{true}} - \widehat{\text{VaR}}_{k;\text{method}}, \quad k = 1, \dots, 100 .$$

We are now testing the null hypothesis that the differences  $\widehat{D}_k$  have mean zero against the alternative of mean less or bigger than zero, i.e. we have a risk under- or over-estimation, respectively. For this purpose, we use t-tests for the independent observations  $\widehat{D}_k$ .

We also perform sign-type tests based on the number of positive differences

$$\widehat{W}_k = I_{\{\widehat{D}_k \geq 0\}}, \quad k = 1, \dots, 100 ,$$

for the null hypothesis that the positive differences  $\widehat{W}_k$  have mean  $\frac{1}{2}$  against the alternative of mean less (or bigger, respectively) than  $\frac{1}{2}$ .

The number of positive differences and the results of the t-type and sign-type tests are summarized in Table 4.1.

TABLE 4.1 ABOUT HERE

Considering the t-tests, we find that in two cases out of five at the 10% level and in one out of five at the 5% level BAGV significantly underestimate the risk. Moreover, if we consider the sign-type tests, which are robust against deviations from Gaussianity, this results becomes more evident: in four cases out of five at the 10% level and in two out of five at the 5% level the null hypothesis is rejected for the 10-day VaR estimates computed with BAGV and the risk tends to be under-estimated (only in one case we have a overestimation of the risk). On the other side, our multivariate approach yields better and more accurate 10-day VaR estimates and in no case we have a rejection of the null hypothesis for both t-type and sign-type tests at the 10% or 5% level.

This result corresponds to one of Pritsker (2001). Using the BAGV method, we are not able to compute accurate 10-day VaR predictions for 99% confidence level, due to a lack of extreme observations in the filtered data set. Our multivariate strategy, on the other side, allowing to predict one particular series future returns also using the information coming from others, yields more accurate risk management. The lack of extreme observations in one series seems to be filled up by multivariate modelling.

## 5 Backtesting VaR for two real data examples

We backtest here the non-parametric procedure for the estimation of VaR proposed in the last two sections on two different real data examples. The tests we make 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:

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

The objective is to exceed the actual asset losses 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 concentrate our empirical investigations 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 5 and 13 asset for a period of 300 days in the Swiss and US example, respectively. 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): from the Novartis company, the Roche holding, the Serono company, the Ciba Spez Chemie company and the Sika company. The data are from the time period between June 4, 1997 and August 21, 2001. We always report with the use of 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 we make are *overall frequency tests*. In the following Table 5.1, we show the number of violations across all five assets for our backtesting period (total of 1500 daily assets). 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 used in the backtest for our procedure (denoted by FGD VaR) in confront with the standard one of Barone-Adesi et al. (1998) (denoted by BAGV). The backtest results marked with an asterisk show some significant difference from the following success criterium. Under the hypothesis of independence, the numbers of violations are binomially distributed around their expected values, with standard deviation ranging from 8.44 (95% level) to 2.73 (99.5% level). A two-standard deviation interval can be heuristically augmented to account for the dependence across assets, leading to a tolerance of 3-4 standard deviations.

TABLE 5.1 ABOUT HERE

From the results summarized in Table 5.1 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 case. This happens for both the methods we consider, which yield similar results. Consequently, it seems that the risk is slightly underestimated. As we have already explained in section 2, this can be given by the fact that the normal assumption is violated leading to inefficient estimates. We try to check

this assuming  $t_4$  distributed innovations in (2.2) and allowing for more fat-tails. The results are summarized in the following Table 5.2.

TABLE 5.2 ABOUT HERE

Confronting the results of Table 5.1 with the ones of Table 5.2 we can see the effect given by the new distribution assumed for the innovations: at high confidence levels and long time horizons we record less violations and the risk seems to be better estimated. 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 and 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 methods proposed not significant and therefore 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 aggregate number of violations occurring each day, since by autocorrelations in this series we can 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. The results for 1-day VaR at 99% are summarized in Table 5.3.

TABLE 5.3 ABOUT HERE

To end this section, we want to concentrate 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. (1998), and both methods seem to work well and yield good VaR predictions. Now, if we concentrate a little more on the accuracy of the VaR estimates obtained using the different methods, we can observe some interesting differences.

The first one, well illustrated by Figure 5.1, 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. On the other side, our approach is less conservative and captures better the passage from stressed, high volatility periods to more stable periods and vice versa.

FIGURE 5.1 ABOUT HERE

This result is confirmed when we look at the average estimated individual VaR: the average capital employed when estimating daily VaR with our procedure is for all assets at all time horizon lower than the one required by BAGV. This is a consequence of our model assumptions in (2.2), which allow for cross-terms. Information can then flow from one asset to an other causing a better reaction to changes in the market conditions and a further reduction of the VaR predictions (in absolute terms) during the periods characterized by small magnitude of returns.

A second difference appears clearly when we consider the largest daily violation recorded. The example illustrated by 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). As we expect, 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 %) over the different days for all assets (aggregated with equal weights) ranges with the maximal and the mean size of aggregate violations. The results are summarized in Table 5.4.

TABLE 5.4 ABOUT HERE

At high confidence levels, the maximal sum of VaR's needed as reserve estimated with our FGD VaR procedure is considerably smaller (from 0.4% up to 1%, depending on the chosen horizon and confidence level) respect to the ones from a classical filtered historical simulation. On the other side, the maximal and the mean size of aggregate violations between the method 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 is lower using our procedure for each asset and at all time horizons. Consequently the uncertainty over the amount of capital required to cover unexpected losses is reduced. This is also a consequence, as we have already explained above, of the more information that we use in our method for prediction.

## 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 last section 5.1.

The backtest results of the overall frequency tests are summarized in the Table 5.5.

TABLE 5.5 ABOUT HERE

As we can see, most values, including all values from the BAGV method, are significantly different from our success criterium. The two methods seem to underestimate the risk, although our procedure is better for estimating VaR at short time horizons.

We try now with some other tests to understand the reason why our procedure and the BAGV method yield such poor predictions for daily VaR, in particular at long-time horizons. 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 aggregate number of

violations across all companies occurring each day. The resulting values for 1-day VaR at 99% confidence level are summarized in Table 5.6.

TABLE 5.6 ABOUT HERE

The tests reject clearly for both methods the assumption of no autocorrelations in the time series of aggregate number of 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 contemporarily have 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 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 5.7 and 5.8 respectively.

TABLES 5.7 AND 5.8 ABOUT HERE

We found that without this short period in March 2000 we have 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 5.7. Particularly when considering daily VaR predictions at long-time horizons 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 (similarly to Pritsker, 2001) could be the fact that, in this particular case, the time-window of 800 days used for predictions may not contain enough extreme outliers to accurately compute VaR at long time horizons and high confidence levels using the BAGV method. On the other side, our procedure, using a larger number of predictor variables and allowing the information available for each asset to depend on all past multivariate observations, yields more precise VaR estimates (see also the simulation exercise of section 4).

To conclude the section, we concentrate on 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 sample 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 we expect, the period of time with the largest aggregate daily violation is March 2000, where we have seen that the violations tend to cluster. In particular, for 99.5% confidence level and at 10-day VaR 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 and that the reserve needed to cover the eventual 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 estimation of 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 and contrasts the most criticisms about the use of the BAGV method based on filtered historical simulation (Barone-Adesi et al., 1998) for predicting VaR. 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 in computational issues and in most cases an intractable model-selection problem. Our FGD algorithm solves this problem: it is computationally feasible in multivariate set-ups with dozens up to hundreds of return series. Choosing a reasonable model for estimating the starting functions in FGD (in our case the standard CCC-GARCH(1,1) model), we try to improve the initial predictions with some iterations. This is the most attractive feature of FGD, which is therefore not necessarily restricted to the framework of CCC models, but can be extended to other more complex multivariate models, for example extending our working model by assuming dynamic conditional correlations (Engle, 2002).

We have demonstrated through one simulation exercise and on two real data-sets belonging to the pharmaceutical and biotechnological market segment that our technique produces accurate and powerful daily VaR estimates, significantly outperforming the VaR predictions from the BAGV method. The results of the backtests provides empirical evidence of the fact that our multivariate FGD VaR technique is able to correct inaccuracies, which sometimes occurs when daily VaR is estimated at long time horizons (5 to 10 days) and for high confidence levels (99% or 99.5%) using the BAGV method, yielding better risk estimation and capital allocation. We found that BAGV tends to overestimate risk during the periods of low volatility. Moreover, for high confidence levels the maximal sum of (aggregate) VaR's needed as reserve to cover the eventual losses is considerably smaller and the number and magnitude of the largest aggregate violations are significantly reduced when using our procedure instead of the BAGV method.

In the backtests of section 5, we have also shown how 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)). This is left to future research. The proposed FGD algorithm is very general and can be further adapted to other multivariate problems dealing with volatility estimation, such as computing VaR for global trading portfolios of large trading banks.

## References

- Audrino, F. and Bühlmann, P. (2002). Volatility Estimation with Functional Gradient Descent for Very High-Dimensional Financial Time Series. To appear in the *Journal of Computational Finance*.
- Barone-Adesi, G., Giannopoulos, K. and Vosper, L. (1998). 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.
- Berkowitz, J. and O'Brien J. (2002). How Accurate are Value-at-Risk Models at Commercial Banks. *Journal of Finance* **57**, Issue 3, 1093-1111.
- 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. (2001). Boosting with the  $L_2$ -loss: regression and classification. Preprint, ETH Zürich.
- 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.
- 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).
- Jorion, P. (2001). *Value-at-Risk: the New Benchmark for Controlling Market Risk*. McGraw-Hill, Chicago.
- 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.

	Asset 1		Asset 2		Asset 3		Asset 4		Asset 5	
Method	BAGV	FGD	BAGV	FGD	BAGV	FGD	BAGV	FGD	BAGV	FGD
# pos. diff.	58	49	40	56	41	51	43	49	51	47
t-tests	0.994 (0.161)	0.865 (0.195)	-1.515 (0.066)	-0.675 (0.251)	-0.253 (0.400)	1.141 (0.128)	-2.203 (0.015)	0.589 (0.279)	1.074 (0.143)	0.400 (0.345)
sign-tests	1.613 (0.053)	-0.199 (0.421)	-2.031 (0.021)	1.203 (0.114)	-1.821 (0.034)	0.199 (0.421)	-1.407 (0.080)	-0.199 (0.421)	0.199 (0.421)	-0.598 (0.275)

Table 4.1: t-type and sign-type tests for the differences between true VaR's and 10-day VaR estimates computed with BAGV and our multivariate FGD strategy for 100 independent simulations. P-values are given between parentheses.

C. level	Expected	Model	1-Day	2-day	3-Day	5-Day	10-Day
95%	75	BAGV	80	78	78	75	78
		FGD VaR	78	84	80	79	74
99%	15	BAGV	21	22	20	23	20
		FGD VaR	20	21	19	23	24
99.5%	7.5	BAGV	12	14	14	12	12
		FGD VaR	11	16*	14	13	14

Table 5.1: Overall frequency tests: violations across five assets belonging to the SPI segment of chemi/pharma recorded during the backtesting period between June 28, 2000 and August 21, 2001.

C. level	Expected	Model	1-Day	2-day	3-Day	5-Day	10-Day
95%	75	BAGV with $t_4$	79	80	80	77	75
		FGD VaR with $t_4$	81	82	80	81	72
99%	15	BAGV with $t_4$	17	20	22	21	20
		FGD VaR with $t_4$	20	21	20	21	17
99.5%	7.5	BAGV with $t_4$	13	13	12	8	11
		FGD VaR with $t_4$	12	14	13	11	10

Table 5.2: Overall frequency tests: violations across five assets belonging to the SPI segment of chemi/pharma recorded during the backtesting period between June 28, 2000 and August 21, 2001. The estimates are constructed assuming  $t_4$  distributed innovations in (2.2).

	$Q_1$	$Q_2$	$Q_3$	$Q_4$	$Q_5$	$Q_6$
FGD VaR	0.187	0.957	1.009	1.190	1.907	2.632
BAGV	0.110	0.989	1.083	1.188	2.010	2.842

Table 5.3: Time clustering effect: Ljung-Box tests for the aggregate number of violations at 99% confidence level and 1-day horizon for five assets belonging to the SPI segment of chemi/pharma recorded during the backtesting period between June 28, 2000 and August 21, 2001.

C. level	Time 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 5.4: 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. (1998) for all 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.

C. level	Expected	Model	1-Day	2-day	3-Day	5-Day	10-Day
95%	195	BAGV	260*	260*	264*	275*	294*
		FGD VaR	247	244	249	259*	276*
99%	39	BAGV	69*	74*	69*	71*	78*
		FGD VaR	68*	68*	66*	66*	67*
99.5%	19.5	BAGV	42*	45*	40*	44*	52*
		FGD VaR	35	39*	36	35	39*

Table 5.5: Overall frequency tests: violations across thirteen assets belonging to the AMEX Biotechnology Index recorded during the backtesting period between July 1, 1999 and August 24, 2000. The values marked with an asterisk show some significant difference from our success criterium.

	$Q_1$	$Q_2$	$Q_3$	$Q_4$	$Q_5$	$Q_6$
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*

Table 5.6: Time clustering effect: Ljung-Box tests for the aggregate number of violations for 99% confidence level and at 1-day time horizon for thirteen companies belonging to the AMEX Biotechnology Index recorded during the backtesting period between July 1, 1999 and August 24, 2000. The values marked with an asterisk are significantly different from our null-hypothesis of no autocorrelations in the time series of aggregate number of violations.

C. level	Expected	Model	1-Day	2-day	3-Day	5-Day	10-Day
95%	188.5	BAGV	229	227	230	237	255*
		FGD VaR	216	211	216	224	236
99%	37.7	BAGV	54	58*	53	50	62*
		FGD VaR	53	52	50	45	50
99.5%	18.85	BAGV	31	33*	29	33*	37*
		FGD VaR	26	28	27	22	24

Table 5.7: Overall frequency tests: violations across thirteen assets belonging to the AMEX Biotechnology Index recorded during the entire backtesting period without the dates between March 3, 2000 and March 21, 2000. The values marked with an asterisk show some significant difference from our success criterium.

	$Q_1$	$Q_2$	$Q_3$	$Q_4$	$Q_5$	$Q_6$
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 5.8: Time clustering effect: Ljung-Box tests for the aggregate number of violations for 99% confidence level and at 1-day time horizon for thirteen companies belonging to the AMEX Biotechnology Index recorded during the backtesting period without the dates between March 3, 2000 and March 21, 2000. The values marked with an asterisk are significantly different from our null-hypothesis of no autocorrelations in the time series of aggregate number of violations.

### Ciba Spez Chemie VaR estimates at 1-Day, 99%

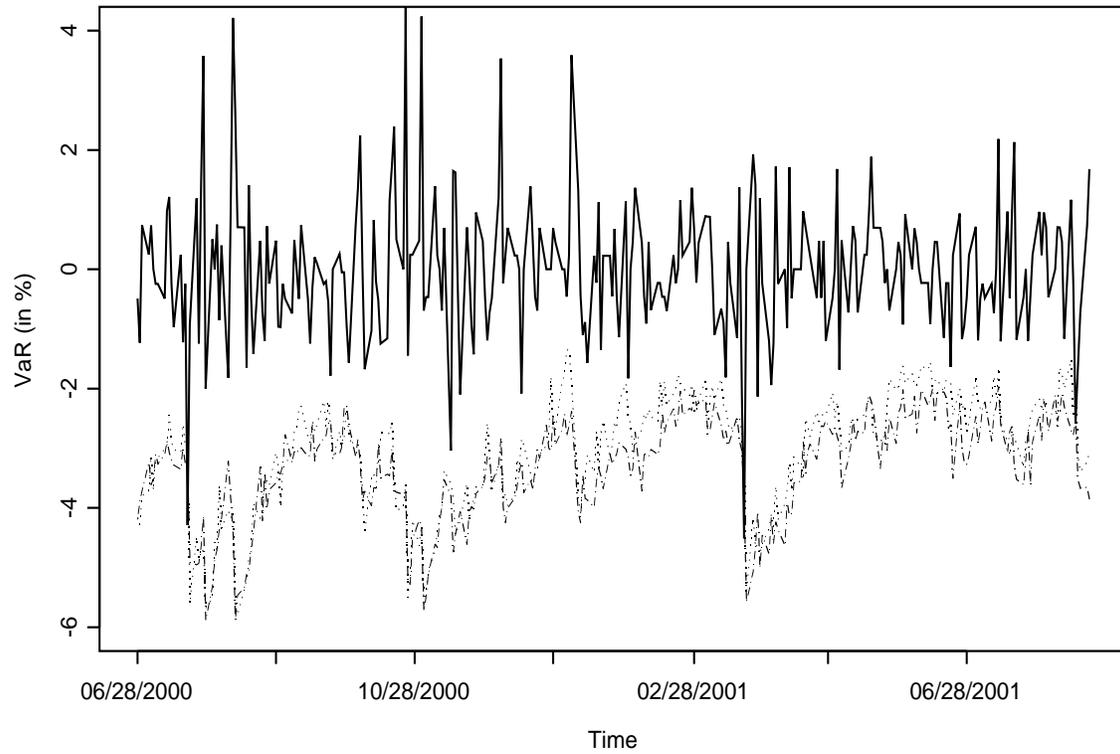


Figure 5.1: Backtest analysis: VaR estimates at 99% confidence level and 1-day time horizon obtained using the FGD strategy proposed in sections 2-3 (dotted line) and the BAGV method from Barone-Adesi et al. (1998) (dashed line) superimposed on the (log-) returns of the Ciba Spez Chemie firm for the entire backtesting period between June 28, 2000 and August 21, 2001.

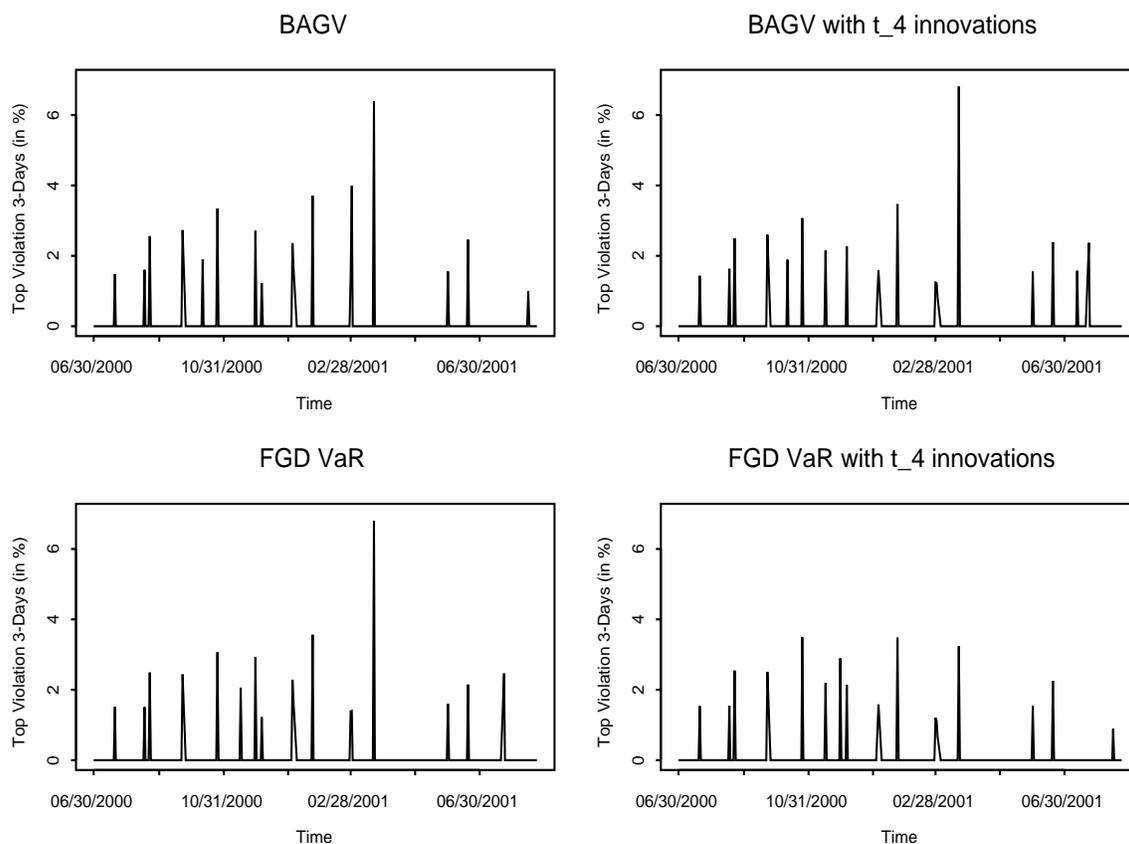


Figure 5.2: Backtest analysis: largest daily violation (equally weighted, in %) for a 3-day VaR horizon and at 99% confidence level obtained aggregating individual asset (belonging to the SPI segment of chemi/pharma) violations from our procedure FGD VaR (bottom left), our strategy FGD VaR with  $t_4$  distributed innovations (bottom right), the standard BAGV method from Barone-Adesi et al. (1998) (top left) and the BAGV method with  $t_4$  distributed innovations (top right) during the backtesting period between June 28, 2000 and August 21, 2001.

### Largest Daily Violation for 10-Day VaR at 99.5%

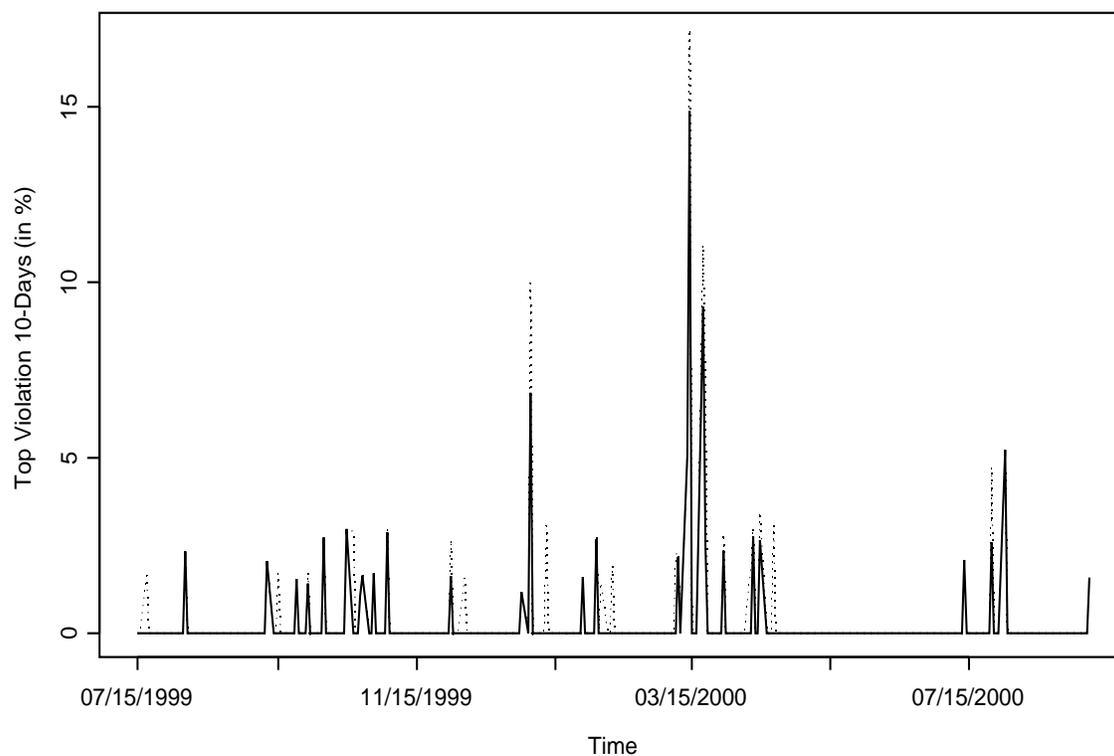


Figure 5.3: Backtest analysis: largest daily violation (equally weighted, in %) for a 10-day VaR horizon and at 99.5% confidence level obtained aggregating individual asset (belonging to the AMEX Biotechnology Index) violations from our procedure FGD VaR (solid line) and the BAGV method (dotted line) during the backtesting period between July 1, 1999 and August 24, 2000.