Hedging with value-at-risk[1]

Claudio Albanese[2] Ken Jackson[3] Petter Wiberg[4]

January 29, 2001

[1]Parts of this work was presented at the Workshop on Mathematical Physicists in Finance and Industry, CRM 2000, Montreal; Quantitative Risk Management in Finance, Carnegie Mellon University; and the International Conference on Numerical Mathematics, BIT Anniversary, Lund University. We thank the participants for valuable comments. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada.

[2]Department of Mathematics, University of Toronto, 100 St. George Street M5S 3G3, Toronto, Canada, and Math Point Ltd., Toronto (www.math-point.com). e-mail: albanese@math.toronto.edu, phone: 416-978-5237

[3]Department of Computer Science, University of Toronto, 10 King's College Road, Toronto, Ontario, Canada M5S 3G4. e-mail: krj@cs.toronto.edu, phone: 416-978-7075

[4]Departments of Computer Science, University of Toronto, 10 King's College Road, Toronto, Ontario, Canada M5S 3G4. e-mail: wiberg@cs.toronto.edu, phone: 416-978-3905

Abstract

Methods to computing value-at-risk gradients with respect to portfolio positions have many applications. They include calculation of capital/reward efficient frontiers, hedging of derivative portfolios and optimal replication. We present a new algorithm for computing value-at-risk and its gradients. If the return can be decomposed as a sum of independent portfolio marginals, the pay-off distribution can be computed with multiple convolutions. This principal-component-type decomposition is also useful for calibrating fat-tailed distributions. We conclude with two applications of hedging with value-at-risk.

1 Introduction

Value-at-risk (VaR) is the established standard for measuring market risk. It is endorsed by regulators [3] and by financial institutions [11, 9]. However, VaR has conceptual limitations, as noted by Artzner et al. [2]. Moreover, modeling and algorithm choices may influence the computed measure significantly [4, 8]. Nevertheless, VaR improves previously used risk measures, such as standard deviation. Moreover, because of its wide spread, it is worth while to revisit the questions of hedging and portfolio replication, traditionally studied in a return/variance framework. To this end, we present a new method for computing VaR and its gradients, and we explore the calculation of optimal hedge ratios for VaR.

In general, algorithms to solve optimization problems require repeated calculations of the objective function. For VaR optimization problems, fast VaR algorithms are essential. Exceptions are the one-step simulation based methods proposed by Andersson et al.[1] and Duarte [5]. These methods can handle generic pay-off functions, but the Monte Carlo simulation is computationally expensive.

Many different computational methods for VaR have been proposed. They include simple analytic approximations, e.g. using Pearson, Cornish-Fisher and Johnson distributions, see El-Jahel et al. [6]; Monte Carlo methods that are unbiased statistical estimates, but often require huge computational resources; and, in our experience, the difficult to implement fast Fourier methods proposed by Mina and Ulmer, see [10], [15]. Fat-tailed distributions are important for realistic models of return distributions, see for instance Jorion [8], Pownall et al. [13] or Stambaugh [14]. Both analytic approximations and Monte Carlo methods can handle non-normal return distributions. In addition to the ability to handle non-normal returns, the algorithm we propose can compute the VaR gradients efficiently. The method is related to the fast Fourier method of Mina and Ulmer [10]. However, unlike Mina and Ulmer, we do not assume that returns are normal, and, instead of finding an analytic form for the Fourier transform, we use the fast Fourier transform. The return distribution is expressed as the convolution of several functions and we propose the name *fast convolution method*. The terms of the convolution can be interpreted as uncorrelated 1-dimensional $\Delta - \Gamma$ approximations.

As Ihle points out [7], hedge ratios computed in the traditional way sometimes increase, instead of decreasing, risk measured in terms of VaR. We conclude the paper with some numerical examples to illustrate the performance of the method. In particular, we discuss two applications of VaR gradients to hedge derivative portfolios. In both examples, the gammas are negative and the return distributions are fat-tailed. We show that the Black-Scholes hedge ratios can differ substantially from the optimal VaR hedge ratios when the re-hedging frequency is around one business day.

Acknowledgements. C.A. profited from an earlier collaboration with Smaranda Păun [12], now at Deloitte and Touche. C.A. thanks Raphael Douady for discussions. All errors in this article are our own.

2 Portfolio dependent estimations

Value-at-risk (VaR) is defined by the equation

$$\mathrm{Prob}_0\big[\Pi(T) - E_0[\Pi(T)] \leq -VaR\big] = 1 - p, \tag{1}$$

where Π is a random variable giving portfolio returns, p is the confidence level (typically 99%), T is the time horizon and $E_0[\Pi(T)]$ is the expected return. Consider a portfolio depending on N risk factors. In the $\Delta - \Gamma$ approximation, portfolio values are computed as

$$\Pi(T) - E_0[\Pi(T)] = \Delta \cdot x + \frac{1}{2} x \cdot \Gamma \cdot x. \tag{2}$$

Here x is the N dimensional random vector of risk factor shocks, for which historical time series $x_{1t}, \ldots, x_{Nt}$ are known. The calendar time is t, where $t = 0$ is today's date. Estimate the covariance matrix

$$C_{ij} = E\big[x_i x_j\big] \tag{3}$$

using historical time series. The matrix C is positive definite and can be diagonalized simultaneously with the matrix Γ. Let R be a matrix satisfying $C = R^T R$. The matrix R can be computed by the Cholesky decomposition[1]. Let Q be the orthogonal matrix and Λ the diagonal matrix giving the Schur decomposition $R\Gamma R^T = Q\Lambda Q^T$. Then, we have

$$\Pi(T) - E_0[\Pi(T)] = \sum_{i=1}^{N} \overline{\Delta}_i y_i + \frac{\lambda_i}{2} y_i^2$$

where $y = Q^T R^{-T} x$, $\overline{\Delta} = Q^T R \Delta$ and the λ_i are the diagonal elements of Λ. The new risk factors y are uncorrelated and have unit volatility as

$$E\big[y_i y_j\big] = \delta_{ij}. \tag{4}$$

To continue the statistical analysis, we postulate that the transformed risk factors y_i are mutually independent. This assumption gives

$$\Pi(T) - E_0[\Pi(T)] = \sum_{i=1}^{N} \pi_i \tag{5}$$

where the portfolio marginals

$$\pi_i = \overline{\Delta}_i y_i + \frac{\lambda_i}{2} y_i^2 \tag{6}$$

are independent random variables.

The probability distribution function $\phi_i(\eta)$ for y_i can be estimated from the historical time series $y_t = A \cdot x_t$, where A is the matrix $Q^T R^{-T}$. Several alternatives are conceivable. In a non-parametric

[1] The Cholesky decomposition is perhaps the simplest choice, but another alternative is for example the singular value decomposition. The choice of decomposition may have statistical importance for interpreting the principal components, and it may also be important for numerical and algorithmic reasons.

approach, a possible choice for the function $\phi_i(\eta)$ is the historical distribution. In a parametric approach, the parameters α that control the probability distribution function $\phi(\eta;\alpha)$ are chosen based on an appropriate goodness-of-fit criterion. Maximum likelihood estimation is not always a robust procedure in situations where models are potentially misspecified, and the first priority is to reproduce the tails of the distribution correctly. We find that a robust method is to use (possibly asymmetric) Student-t distributions. The number of degrees of freedom is set so that the values of VaR of the portfolio marginals π_i are correctly reproduced. We emphasize that the implementation of the fast convolution method described below can accommodate any model for the individual risk factors y.

3 VaR sensitivities

In equation (5), the portfolio return is represented as a sum of independent portfolio marginals π_i given by equation (6). The profit-and-loss distribution, i.e. the probability distribution function for Π, is given by the convolution product

$$\Psi(\xi) = (\psi_1 * \ldots * \psi_N)(\xi), \tag{7}$$

where $\psi_i(\xi)$ is the probability distribution function of the i-th portfolio marginal π_i. Recall that convolutions are computed according to the rule

$$(f * g)(\xi) = \int f(\eta) g(\xi - \eta) d\eta. \tag{8}$$

The most efficient way of evaluating convolutions is with the fast Fourier transform method. The Fourier transform is defined as

$$(\mathcal{F}f)(\lambda) = \int_{-\infty}^{\infty} f(\xi) e^{i\xi\lambda} d\xi, \tag{9}$$

and the Fourier transform of a convolution is the product of the Fourier transforms of the terms, i.e.

$$\mathcal{F}(f * g) = \mathcal{F}f \cdot \mathcal{F}g. \tag{10}$$

To make the continuous integrals amenable to numerical evaluation, an approximation scheme is needed to discretize the integrals. The solution proposed in [10] is to assume that risk factor returns obey a multivariate normal distribution, evaluate the Fourier transform analytically, compute products and then evaluate the inverse Fourier transform by discretization.

In this article, we proceed differently. For value-at-risk applications, it is important to select an appropriate discretization rule, for which the resulting probability distribution function is positive. It is also important to allow for non-normal risk factor distributions. To achieve these goals, we first discretize marginal portfolio return distributions and then use discrete Fourier transforms to compute convolutions.

It is not difficult to discretize a marginal portfolio distribution $\psi_i(\xi)$. It is easy to find a rough estimate for VaR using for instance a Monte Carlo method. This estimate does not need to be precise and serves only to identify an appropriate interval $[a, b]$ that comfortably accommodates the total P&L distribution function and to which we can safely restrict our return distributions. Notice that since the portfolio marginals π_i are mutually independent in our VaR model, an interval which comfortably accommodates the total portfolio profit-and-loss distribution, also accommodates the marginal distributions. Consider the interval between two contiguous grid points $[\xi_j, \xi_{j+1}]$. In this interval, the distribution $\psi_i(\xi)$ can be approximated by the constant

$$\psi_i(j; \bar{\Delta}_i, \lambda_i) = \mathrm{Prob}\Big[\pi_i \in [\xi_j, \xi_{j+1}]\Big], \tag{11}$$

which can be readily computed from the distribution function $\phi_i(\eta)$ for the i-th factor y_i. Introducing the Fourier series

$$\hat{\psi}_i(\omega, \bar{\Delta}_i, \lambda_i) = (\mathcal{F}\psi_i)(\omega, \bar{\Delta}_i, \lambda_i) = \sum_k e^{i\tau_k \cdot \omega} \psi_i(k, \bar{\Delta}_i, \lambda_i), \tag{12}$$

where τ_k is the mid point of the interval $[\xi_k, \xi_{k+1}]$. The profit-and-loss distribution on the lattice is then given by the inverse Fourier series

$$p(j, \bar{\Delta}, \Lambda) = (\mathcal{F}^{-1} \prod_i \hat{\psi}_i)(j, \bar{\Delta}_i, \lambda_i) \tag{13}$$

The discretization scheme we use has a simple financial interpretation. Real world currency is not a continuous variable, as it is usually assumed in mathematical finance, but is discrete, as fractions of pennies are rarely traded. In the proposed approximation, it is assumed that the discretization unit is the distance between the grid points. The interval $[a, b]$ is approximated by a grid, and the portfolio marginals π_i are approximated by a constant on each sub-interval.

The gradients for the probability distribution function can be computed with the chain rule. The Δ gradients are given by

$$\frac{\partial}{\partial \Delta_i} p(j, \Delta, \Gamma) = \sum_k \frac{\partial \bar{\Delta}_k}{\partial \Delta_i} \frac{\partial}{\partial \bar{\Delta}_k} p(j, \bar{\Delta}, \Lambda). \tag{14}$$

where

$$\frac{\partial \bar{\Delta}_k}{\partial \Delta_i} = (Q^T R)_{ki}. \tag{15}$$

The derivative of the probability distribution function is given by

$$\frac{\partial}{\partial \bar{\Delta}_k} p(j, \bar{\Delta}, \Lambda) = \mathcal{F}^{-1} \sum_l \frac{\partial \hat{\psi}_l}{\partial \bar{\Delta}_k}(j, \bar{\Delta}_l, \lambda_l) \prod_{m \neq l} \hat{\psi}_m(j, \bar{\Delta}_m, \lambda_m) \tag{16}$$

Notice that much of the work needed to evaluate the portfolio distribution functions can be re-used to compute the derivatives.

The modified delta vector $\bar{\Delta}$ also has a non-zero derivative with respect to the matrix Γ, because the rotation matrix R depends on Γ. Hence the Γ gradients involve two terms

$$\frac{\partial}{\partial \Gamma_{mn}} p(j, \Delta, \Gamma) = \sum_k \frac{\partial \lambda_k}{\partial \Gamma_{mn}} \frac{\partial}{\partial \lambda_k} p(j, \bar{\Delta}, \Lambda) + \frac{\partial \bar{\Delta}_k}{\partial \Gamma_{mn}} \frac{\partial}{\partial \bar{\Delta}_k} p(j, \bar{\Delta}, \Lambda) \tag{17}$$

Elementary linear algebra shows that

$$\frac{\partial \lambda_i}{\partial \Gamma_{j_1, j_2}} = (v_i, Z_{j_1, j_2} v_i) \tag{18}$$

where v_i is the $i-$th eigenvector of the matrix $R^T Q \Lambda Q^T R$ and Z_{j_1,j_2} is the matrix with zero entries everywhere, except for the matrix element in position (j_1, j_2) which is unity. We also have

$$\frac{\partial \bar{\Delta}_i}{\partial \Gamma_{ij}} = \frac{\partial Q^T}{\partial \Gamma_{ij}} \cdot R\Delta. \tag{19}$$

and the matrix derivative $\frac{\partial Q^T}{\partial \Gamma_{ij}}$ can also be computed using elementary linear algebra. Finally, the derivative of the probability distribution function with respect to the eigenvalues is given by an expression similar to (16)

$$\frac{\partial}{\partial \lambda_k} p(j, \bar{\Delta}, \Lambda) = \mathcal{F}^{-1} \sum_l \frac{\partial \hat{\psi}_l}{\partial \lambda_k} (j, \bar{\Delta}_l, \lambda_l) \prod_{m \neq l} \hat{\psi}_m (j, \bar{\Delta}_m, \lambda_m). \tag{20}$$

The VaR gradients can then be computed by implicit differentiation of (1). Using the derivatives (14) and (17), the VaR gradients are

$$\frac{\partial VaR}{\partial \Delta_i}(-VaR, \Delta, \Gamma) = \frac{1}{p(-VaR)} \frac{\partial}{\partial \Delta_i} p(-VaR, \Delta, \Gamma), \tag{21}$$

$$\frac{\partial VaR}{\partial \Gamma_{ij}}(-VaR, \Delta, \Gamma) = \frac{1}{p(-VaR)} \frac{\partial}{\partial \Gamma_{ij}} p(-VaR, \Delta, \Gamma). \tag{22}$$

Numerical errors have two different origins: roundoff errors in the calculation of derivatives for the discretized model and discretization errors. The round-off errors are small compared to the discretization errors. A comparison between a finite difference calculation of derivatives in a 2-dimensional

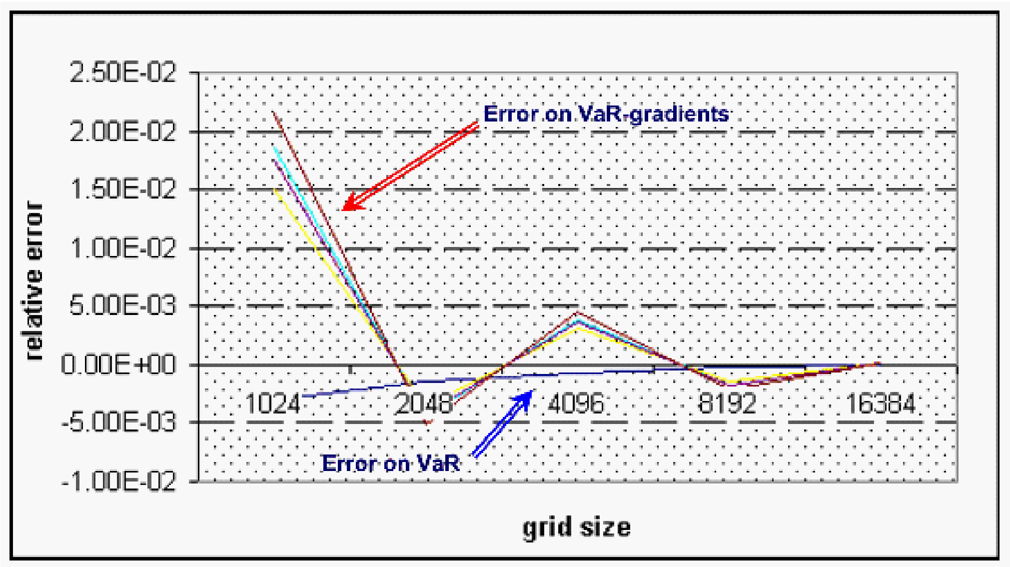

Figure 1: *The figure shows the errors in VaR and its derivatives as functions of the lattice size. The two jagged lines are the errors in the derivatives and these are larger than the errors in VaR itself, the line closer to the axis. The relative error is less than 0.1% for lattice sizes larger than approximately* $10,000$ *elements.*

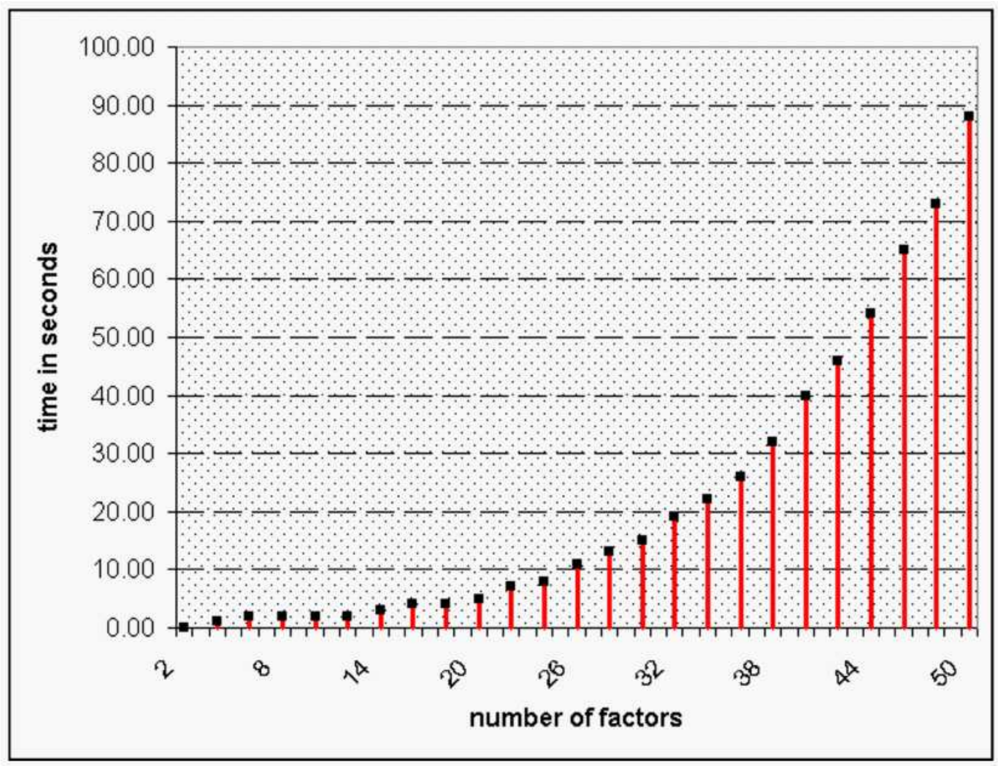

Figure 2: *CPU time on a notebook as a function of the number of risk factors on randomly generated data. The time switches from being almost linear in the number of factors for the first part, to being cubic for more than 25 risk factors.*

benchmark example complemented with Richardson extrapolation gives the following:

$$\frac{\partial VaR}{\partial \Delta} = 10^{-2} \begin{pmatrix} -8.874151177 \\ -10.21365093 \end{pmatrix}, \qquad \frac{\partial VaR}{\partial \Gamma} = 10^{-3} \begin{pmatrix} -3.27086146 & 3.63958829 \\ -3.63958829 & 4.23760128 \end{pmatrix} \tag{23}$$

and

$$\frac{\partial VaR}{\partial \Delta} = 10^{-2} \begin{pmatrix} -8.874151168 \\ -10.21365092 \end{pmatrix}, \qquad \frac{\partial VaR}{\partial \Gamma} = 10^{-3} \begin{pmatrix} -3.27086146 & 3.63958828 \\ -3.63958828 & 4.23760127 \end{pmatrix} \tag{24}$$

We believe that the obtained accuracy is satisfactory, and above the resolution of the model. Figure 1 shows the dependence of the relative error as a function of the number of lattice points in the grid. In this example, the errors in the VaR derivatives are larger than the errors in VaR itself. This is to be expected since derivatives tend to magnify errors. The relative errors are less than 0.1% for lattice sizes of more than $\approx 10,000$ elements. Finally, Figure 2 shows the CPU time on a notebook computer as a function of the number of risk factors. The number of operations is proportional to $N^3 + NM \log M$ where N is the number of risk factors and M is the size of the lattice. The figure shows that as the number of factors N increase the time is eventually dominated by the cubic term, this happens around $N = 25$. The location is dependent on the implementation. It is worth noting that for smaller problems the bottleneck for the algorithm is in the special functions used in the discretization. Computing time is about a minute on a PC for the case of about 50 risk factors.

4 Capital reward optimization

The convolution method is suitable for VaR optimization problems, as it provides not only VaR but also the direction of steepest descent of VaR, as a function of the portfolio composition and the hedge ratios. Here we discuss two simple cases to illustrate the possible uses of the proposed algorithm.

These two examples are no substitute for a thorough empirical study, but they simply demonstrate how the mathematics in our VaR model can be used to support practical applications

Consider the problem of delta-hedging a short option position in an at-the-money IBM option using the underlying IBM stock. Black-Scholes theory, based on the hypotheses of continuous trading and normal return distributions, shows that a delta-neutral position would minimize risk. However, since the gamma is negative and returns are fat-tailed, the optimal delta-ratio is not zero. As Figure 3 shows, the delta-neutral hedge from the Black-Scholes model corresponds to a local maximum of VaR. The two local minima nearby provide a more efficient hedging strategy against large moves. Notice that this follows because gamma is negative, so that gamma risk causes losses for moves of the underlying in both directions.

A second example is about hedging in incomplete markets. Suppose that we want to hedge a short position in a Microsoft stock option using a position in Nasdaq futures. In this example, we consider two scenarios to demonstrate the effect of fat-tailed returns. In the first, we postulate that return distributions are normal. In the second, we estimate return distributions using asymmetric Student-t distributions. The second calculation is more precise and gives rise to only 3 exceptions in 250 days, as compared to the 6 exceptions obtained under the assumption of normality. For model to be accepted by regulators, the number of losses exceeding VaR should be less than 4 over 250 days. Furthermore, the optimal hedge ratios are different, as Figure 4 shows. For fat-tailed distributions, the optimal delta hedge ratio is higher than the optimal hedge ratio with normal returns. For normal returns, the risk in positions with smaller with a smaller delta is underestimated and the risk for larger delta is overestimated. The misspecified normal model suggests a hedge ratio which provides less protection against down movements, which in this would cause relatively more severe losses. A possible explanation of this result is that the Microsoft stock and the Nasdaq index are more correlated than predicted by the normal distribution hypothesis whenever deviations are large. Hence, accounting for fat-tailed distributions, incomplete hedging appears to be more effective than it would be otherwise.

5 Concluding remarks

For derivative portfolios, the non-linearity of the return distribution has significant impact on the value-at-risk. The $\Delta - \Gamma$ approximation provides an accurate approximation with correct convexity. The portfolio value can, subject to some assumptions of independence, be decomposed into portfolio marginals. The method presented uses multiple convolutions to compute both VaR and its gradients from the portfolio marginals. The method can handle non-normal returns, and, as illustrated by the examples, it is fast and accurate.

VaR gradients have interesting applications to the optimization problem for hedge ratios, and we present two simple examples. In this article, we have focused on describing the algorithm. We defer a more detailed empirical study to a forth coming paper.

References

[1] F. Andersson and S. Uryasev. Credit risk optimization with conditional value-at-risk criterion. Technical Report #99-9, Center for Applied Optimization, Department of Industrial and Systems Engineering, University of Florida, 1999.

[2] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

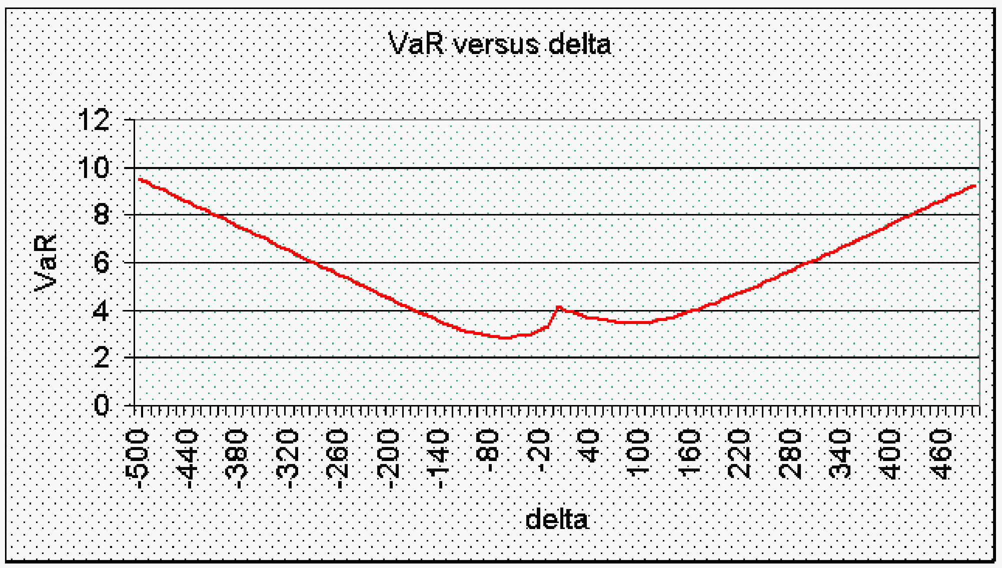

Figure 3: *VaR as a function of the portfolio total delta. The delta-neutral position is a local maximum, surrounded by two local minima.*

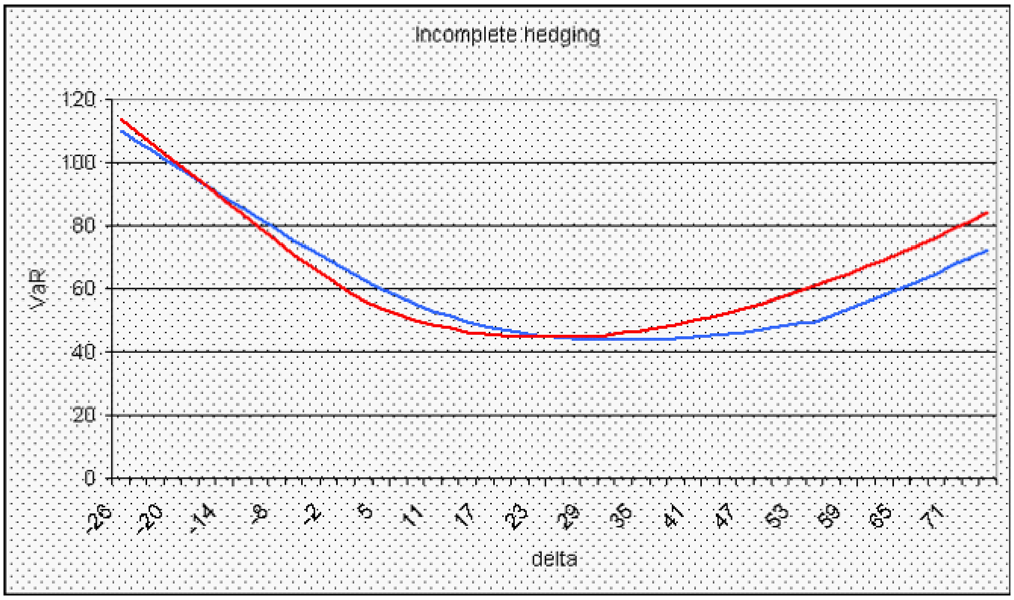

Figure 4: *VaR as a function of the portfolio total delta. The optimal delta ratio accounting for fat-tailed distribution is higher than the delta-ratio obtained under the assumption of normality.*

[3] Basle Committee on Banking Supervision. Amendment to the Capital Accord to incororate market risks. Technical report, Bank for International Settlements, 1996.

[4] T.S. Beder. VAR: Seductive but dangerous. *Financial Analysts Journal*, 51(5):12–24, 1995.

[5] A.M Duarte, Jr. Optimal value at risk hedge using simulation methods. *Derivatives Quarterly*, 5(2):67–75, 1998.

[6] L. El-Jahel, W. Perraudin, and P. Sellin. Value at risk for derivatives. *Journal of Derivatives*, 6(3):7–26, 1999.

[7] G. Ihle. Forward hedges that increase value at risk. *Derivatives Quarterly*, 4(4):67–72, 1998.

[8] P. Jorion. Risk^2: Measuring the risk in value at risk. *Financial Analysts Journal*, 52(6):47–56, 1996.

[9] P.H. Kupiec. Risk capital and VaR. *Journal of Derivatives*, 7(2):41–52, 1999.

[10] J. Mina and A. Ulmer. Delta-Gamma four ways. Technical report, RiskMetrics Group, 1999.

[11] J.P. Morgan/Reuters. *RiskMetrics—Technical Document*. J.P. Morgan, 4th edition edition, 1996.

[12] S. Păun. Sensitivities of value at risk for non-linear portfolios. Master's thesis, Department of Mathematics, University of Toronto, Toronto, Canada, 1999.

[13] R.A.J. Pownall and K.G. Koedijk. Capturing downside risk in financial markets: the case of the Asian Crisis. *Journal of International Money and Finance*, 18:853–870, 1999.

[14] F. Stambaugh. Risk and value at risk. *European Management Journal*, 14(6):612–621, 1996.

[15] A. Ulmer. Picture of risk. Technical report, RiskMetrics Group, 1999.