Understanding and Calculating Value at Risk

Philippe De Brouwer

holds degrees in Physics and Economics, and is currently Director at Fortis Investment Management.
Philippe De Brouwer actually graduated in theoretical physics (University of Brussels) and published some work in the field of non-linear dynamics and chaos theory. Afterwards he oriented his career to financial economics and promoted in commercial engineering (University of Brussels). Now after the recent merger with Fimagen, and the creation of FIM (Fortis Investment Management) he is Director and is responsible for strategy and co-ordination.

philippe.debrouwer@fimgroup.com

Abstract: *Value at risk is a relatively recent concept in the measurement of financial risks. Its universal existence and interpretation that is closely related to the perception of "risk" is the basis of its increasing success. Even the Basle Committee, suggests its use by banks for their risk control. This paper aims to explore a fundamental understanding of the concept "value at risk" (VaR), and provides insight in estimation methods.*

Keywords*: value at risk, varcov-matrix, variance-covariance method, historical simulation method, Monte-Carlo simulation, bootstrapping, extreme value theory.*

1 Table of Content

1 Table of Content

2 Introduction and Definition

2.1 Financial Risk in a historical context

2.2 Defining VaR

3 Elements of the basic calculus and some pitfalls

3.1 Which distribution law?

3.2 The Return of a Composite Portfolio

3.3 The Volatility of a composite portfolio

4 Estimation methods for VaR

4.1 Variance-Covariance methods

4.1.1 The Fixed Weight Moving Average Estimation (FWMA)

4.1.2 The Exponentially Weighted Moving Average (EWMA)

4.1.3 The multivariate GARCH

4.1.4 The Kernel Estimation

4.2 Historical Simulation methods

4.2.1 Fixed Weight Historical Simulation

4.2.2 Antithetic Historical Simulation

4.2.3 Simulated Covariances and Historical Probability Distribution

4.2.4 Exponential Historical Simulation

4.3 The Monte-Carlo approach

4.4 Boot-strapping approach

4.5 Extreme Value Estimation

5 Conclusions

6 Acknowledgement

7 References and Endnotes

2 Introduction and Definition

2.1 Financial Risk in a historical context

Since Markowitz [1 and 2], we have a rational theory to select a financial portfolio. This famous Mean Variance Criterion (MVC) is well studied and is in accordance with the theory of utility functions. The idea is to compare portfolios based on two parameters: "return" and "risk". The portfolio with the highest expected return, given a desired level of risk, should be chosen.

Return can be defined as the logarithmic return of the value of our portfolio (with price S, having $S(0) = S_0$ on a time horizon t = T):

(2.1) $$R(T) = \log\left(\frac{S(T)}{S_0}\right)$$

When Markowitz presented his theory, standard deviation was an obvious choice for the parameter "risk", since:

- it was possible to calculate,
- familiar to other scientists,
- it produced reasonable portfolio's,
- the Central Limit Theorem (CLT) seems to provide a justification for the use of a Gaussian probability density law.

The volatility or standard deviation equal the square root of the variance. The variance is defined as:

(2.2) $$VAR[R(T)] = E\left[(R(T) - E[R(T)])^2\right]$$

If $f_{R(T)}$ is the probability density function of R for the time horizon T, then the volatility $\sigma(T)$ is defined by:

(2.3) $$\sigma(T) = \sqrt{\int_{-\infty}^{+\infty} x^2 . f_{R(T)}(x)dx - \left(\int_{-\infty}^{+\infty} x . f_{R(T)}(x)dx\right)^2}$$

Since then, the MVC was very successful. But, the concept of volatility itself is questioned more and more, and with different arguments:

1. **Behavioural Finance** (investor psychology)[3]: an investor does not perceive risk as volatility. In contrast with volatility that treats losses and gains on an equal basis (is symmetrical for losses and gains), an investor is merely concerned with downside risk: an investor is loss-averse and not volatility averse.
2. Solid mathematical evidence that all financial markets do have a **finite variance** is absent. On the contrary, experiments suggest that the second moment does not exist in financial markets.
3. For a finite number of convolutions, the **Central Limit Theorem (CLT) only applies for the centre** of the distributions. The "tails" of the distributions however are not well described with a Gaussian law: the Gaussian law underestimates the probability of extreme events.
4. For a wide range of investments the Gaussian law is unacceptable (e.g. options, warrants, capital guaranteed funds, etc.).

5. Variance is a quite unstable measure, its calculation is permanently perturbed by large movements.

2.2 Defining VaR

It seems that a more suitable definition of risk is needed: value at risk is a measure that eliminates the previously mentioned problems, and today it is also in agreement with Markowitz' criteria.

Value at Risk can be defined as the amount of money that can be lost on a given financial portfolio (whose return is described by the stochastic variable R), with a given investment horizon (τ) and a given confidence level ($1 - \alpha$). So value at risk (VaR) is the solution to the following equation:

$$\alpha = \int_{-\infty}^{-VaR} f_{R(\tau)}(x)dx \tag{2.4}$$

With: $\alpha = P[R(\tau) < -VaR]$

Another interpretation is that in repeated sampling, there is a small probability equal to α of loosing more money than the VaR on the horizon τ.

It is important to understand that the value at risk for a given investment horizon τ does not imply anything for time horizons shorter than τ. It is possible to encounter larger losses with a probability even larger than α on these shorter horizons. Keep also in mind that in any case we are dealing with a confidence $1 - \alpha$ (say 98%), and that there is always a probability equal to α (say 2%) to encounter larger losses.

3 Elements of the basic calculus and some pitfalls

This chapter describes some basic elements that are important to calculate the VaR.

3.1 Which distribution law?

A lot of literature exists about the question whether returns of financial assets do follow a "Normal" (= "Gaussian") distribution. This paper does not aim to add to this discussion. Most authors do agree that the Gaussian law is not a good approximation, a very interesting study can be found in the book "Théorie des risques financiers"[4]. The authors propose a Lévy-distribution, truncated by an exponential law ("Truncated Lévy laws").

Close to the mean, the CLT learns us that the Gaussian law is a good approximation. But as you will understand, when calculating value at risk, there will be much interest in extreme returns far from the mean.

Some methods to calculate the value at risk that are described in this paper use explicitly the underlying distribution f_R, while others get around this problem and make no explicit assumption on the form of the distribution.

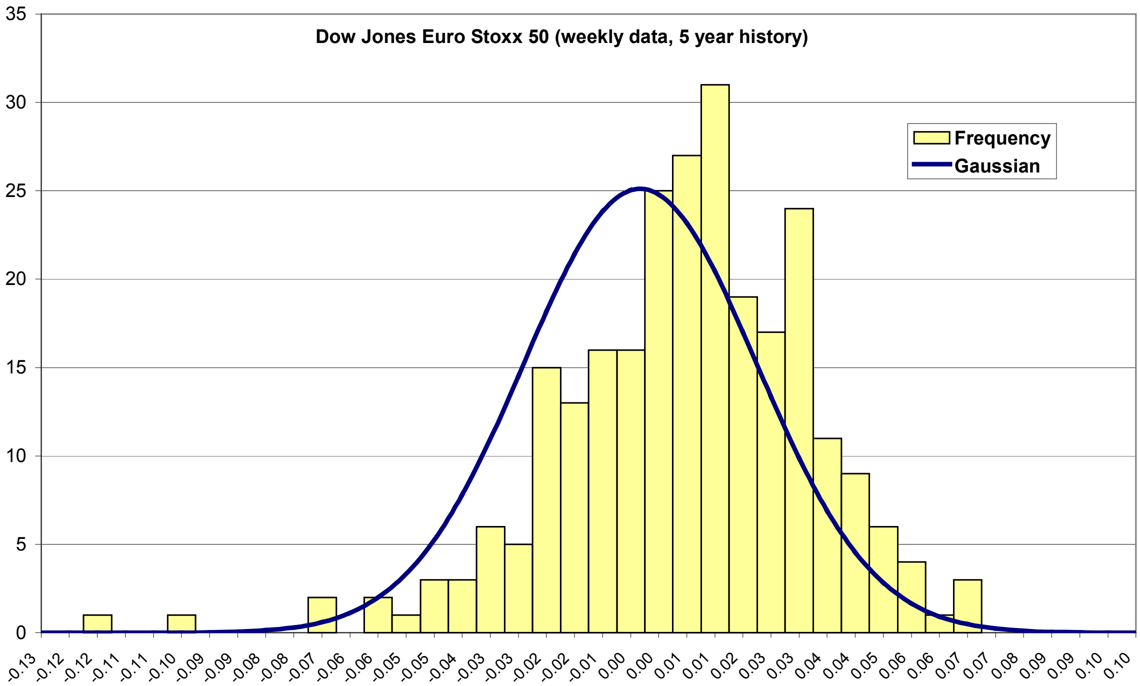

Figure 1. Comparing the Gaussian distribution with real market data for the DJ-Euro Stoxx 50. (Weekly data from 11/95 till 11/00, source: Datastream)

The above illustration shows us that to some extent the Normal distribution is a good proxy in financial markets. However, there exist some systematic bias.

➜ The Normal distribution underestimates extreme values of the return: financial market data show “fat tails” compared to the Gaussian law. This means that returns, that are so low or high that their observation is almost excluded according the Normal distribution, will actually occur.

➜ A Normal distribution underestimates the frequency of observations in the immediate neighbourhood of the mean, and underestimates frequency of occurrence in a second range.

3.2 The Return of a Composite Portfolio

It is a basic property of stochastic variables that $E\left[\sum_t X_t\right] = \sum_t E[X_t]$, and therefore the return of a portfolio (p) is defined by the weighted sum of the returns of its constituents.

(3.1) $$R_p = \sum_{i=1}^{N} w_i R_i \ ,$$

where w_i represents the weight of asset i in our portfolio, and R_i is the expected (log)-return of asset i over a small time interval.

Rewriting these formulae with the *use of matrices* will not only simplify the notation, but it will also be helpful for practical implementation, since many softwares are able to handle matrices. In order to differentiate between scalars, vectors and matrices we underline the symbol with a number of lines equal to its dimension.

(3.2) $$R_p = \underline{w}'.\underline{R}$$

3.3 The Volatility of a composite portfolio

For independently distributed variables, it is also well known[5] that $VAR\left[\sum_t X_t\right] = \sum_t VAR[X_t]$, however the returns of financial assets are on the contrary highly correlated. Taking this correlations into account, we obtain the following expression for the volatility of our composite portfolio (σ_p).

$$(3.3) \qquad \sigma_p = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \sigma_{ij}} = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \rho_{ij}\sigma_i\sigma_j}$$

$$\text{with} \quad \sigma_{ij} = E\left[(R_i - E[R_i]).(R_j - E[R_j])\right]$$

for which we can use the following estimator:

$$(3.4) \qquad \hat{\sigma}_{ij} = \sum_{t=1}^{M} \frac{(R_{i,t} - \mu_i).(R_{j,t} - \mu_j)}{M}$$

with: $R_{\nu,\tau}$ = the return of the component ν at observation t,

$\mu_\nu = \frac{1}{M}\sum_{t=1}^{M} R_{i,t}$ = the average return of component ν,

M = the number of observations,

i and j are dummy counters for the assets composing the portfolio, they run from 1 to N.

Rewritten with matrices, the standard deviation becomes:

$$(3.5) \qquad \sigma_p = \sqrt{\underline{w}'.\underline{\underline{\Sigma}}.\underline{w}}$$

with the following estimator:

$$(3.6) \qquad \begin{aligned}\hat{\sigma}_p &= \frac{1}{M}\sqrt{\underline{w}'.\hat{\underline{\underline{\Sigma}}}.\underline{w}} \\ &= \frac{1}{M}\sqrt{\underline{w}'\underline{\underline{\varepsilon}}'.\underline{\underline{\varepsilon}}\underline{w}}\end{aligned}$$

where $\underline{\underline{\varepsilon}}$ is the residual return matrix, and is defined as follows:

$$(3.7) \qquad \underline{\underline{\varepsilon}} = \begin{pmatrix} R_{1,1}-\mu_1 & R_{1,2}-\mu_1 & \cdots & R_{1,j}-\mu_1 & \cdots & R_{1,M}-\mu_1 \\ R_{2,1}-\mu_2 & R_{2,2}-\mu_2 & \cdots & R_{2,j}-\mu_2 & \cdots & R_{2,M}-\mu_2 \\ \vdots & \vdots & & \vdots & & \vdots \\ R_{i,1}-\mu_i & R_{i,2}-\mu_i & \cdots & R_{i,j}-\mu_i & \cdots & R_{i,M}-\mu_i \\ \vdots & \vdots & & \vdots & & \vdots \\ R_{N,1}-\mu_N & R_{N,2}-\mu_N & \cdots & R_{N,j}-\mu_N & \cdots & R_{N,M}-\mu_N \end{pmatrix}$$

4 Estimation methods for VaR

4.1 Variance-Covariance methods

The Variance-Covariance methods assume that returns of financial markets follow a multivariate Normal distribution.

Further in this chapter we will not consider complicated financial assets such as derivatives.

The estimator for the return of our portfolio is then given by:

$$(4.1) \qquad \hat{R}_p = \underline{\hat{R}}.\underline{w}'$$

and the volatility is estimated by:

$$(4.2) \qquad \hat{\sigma}_p = \sqrt{\underline{w}'.\underline{\underline{\hat{\Sigma}}}.\underline{w}}$$

This allows us to determine the value at risk in a very straightforward way:

$$(4.3) \qquad VaR = \hat{R}_p - z(\alpha).\hat{\sigma}_p$$

Where $z(\alpha)$ is the inverse of the cumulative distribution function.

This looks too easy to be true. Indeed, the main difficulty remains in estimating the parameters to be used: the expected return and the varcov-matrix $\underline{\underline{\Sigma}}$.

On short time intervals (seconds, minutes, hours, and maybe days), one could use zero as estimator for the returns, this will make our results more reliable and stable. On longer time scales, other estimations have to be made, but each and every trader or researcher will have his own estimate.

But as far as the varcov matrix is concerned, it becomes more complicated. The human brain is not in the least adapted to taking into account correlations[6], and therefore no trader will be able to give you his estimate of a correlation matrix. The rest of this chapter is devoted to the estimation of the varcov-matrix.

4.1.1 The Fixed Weight Moving Average Estimation (FWMA)

If it is true that financial markets can be described by a Normal distribution then the natural way to proceed is to take all available data and use the following estimator for $\underline{\underline{\Sigma}}$.

$$(4.4) \qquad \underline{\underline{\hat{\Sigma}}}_{t+1} = \frac{1}{M}\sum_{\nu=0}^{M-1} \underline{\underline{\varepsilon}}_{M-\nu}\,\underline{\underline{\varepsilon}}_{M-\nu}^{'}$$

where $\underline{\underline{\varepsilon}}_{M-\nu}$ is the residual return matrix as calculated (M - ν) time intervals ago (see also (3.6) and (3.7)).

Only the M most recent observations are taken into account, because there is evidence that old data should be ignored (see for example: Engel and Gizycki [7]). This estimator is called the Fixed Weight Moving Average Estimator (the FWMA-estimator).

4.1.2 The Exponentially Weighted Moving Average (EWMA)

Financial markets are characterised by periods of low volatility and periods of high volatility. A transition to a period with high volatility increases the value at risk dramatically, however the FWMA-method will only very gradually take this effect into account. Besides, if we lived a market crash, and M time units later this crash drops "out of sight" the volatility calculated by the FWMA-method will drop dramatically without any reasonable explanation.

Both effects can be taken into account by assigning more weight to more recent observations. The most popular way of doing this goes as follows.

$$\hat{\underline{\underline{\Sigma}}}_{t+1} = \frac{1-\lambda}{1-\lambda^M} \sum_{\nu=0}^{M-1} \lambda^\nu \underline{\varepsilon}_{M-\nu} \underline{\varepsilon}_{M-\nu}^{'} \tag{4.5}$$

with $\lambda \in \,]0, 1[$ (the rate of decay)

By calculating the expected value of this expression, we can easily verify that this is also an estimator for $\underline{\underline{\Sigma}}$. It is a common practice to neglect the term in λ^M, however there is no reason whatsoever to do so. The only effect is that our estimator becomes biased: our estimate of $\underline{\underline{\Sigma}}$ will underestimate $\underline{\underline{\Sigma}}$ with a factor equal to $(1 - \lambda^M)$. There is no reason why this would be a priori neglectable, if for example we use $\lambda = 0.99$ and use a sample of 60 observations, then we underestimate $\underline{\underline{\Sigma}}$ with 55%. However if one uses for example λ equal to 0.5 and 250 observations there will be no effect at all.

The value of λ does not emerge from a deeper theory, but it is possible to determine an optimal value. This can be accomplished by maximum likelihood estimations (ML) or by simply trying some values and comparing their results with historical backtesting. Trial and error will then lead to an optimal result.

It is also interesting to note that if we calculate the limit for $\lambda \rightarrow 1$, we find exactly the same formulae as in (4.4).

$$\lim_{\lambda \to 1} \frac{1-\lambda}{1-\lambda^M} \overset{(H)}{=} \lim_{\lambda \to 1} \frac{-1}{-M\lambda^{M-1}} = \frac{1}{M}$$

The parameter λ can be interpreted as a "rate of decay". The smaller λ, the faster old observations will lose their importance. This will decrease the effective sample size (smaller than M), which leads to an increased measurement error.

This is also illustrated if we see that the EWMA estimate of $\underline{\underline{\Sigma}}_{t+1}$ is the weighted average between the FWMA estimate of $\underline{\underline{\Sigma}}_t$ and the random-walk estimate of $\underline{\underline{\Sigma}}_t$.

$$\begin{aligned} \hat{\underline{\underline{\Sigma}}}_{t+1} &= \frac{1}{1-\lambda^M} \left[(1-\lambda) \underline{\varepsilon}_M \underline{\varepsilon}_M^{'} + \lambda \left(1-\lambda^{M-1}\right) \sum_{\nu=0}^{M-2} \lambda^\nu \underline{\varepsilon}_{M-1-\nu} \underline{\varepsilon}_{M-1-\nu}^{'} \right] \\ &= \frac{1}{1-\lambda^M} \left[(1-\lambda) \underline{\varepsilon}_M \underline{\varepsilon}_M^{'} + \lambda \left(1-\lambda^{M-1}\right) \hat{\underline{\underline{\Sigma}}}_t \right] \end{aligned} \tag{4.6}$$

This is a first order autoregressive structure for $\underline{\underline{\Sigma}}$, a special case of the more general ARCH models [[8]]. This structure is consistent with the idea that there are periods of high and low volatility (volatility clustering).

In order to illustrate the use of this very basic and easy approach, we will consider the exchange rate of the Korean Won.

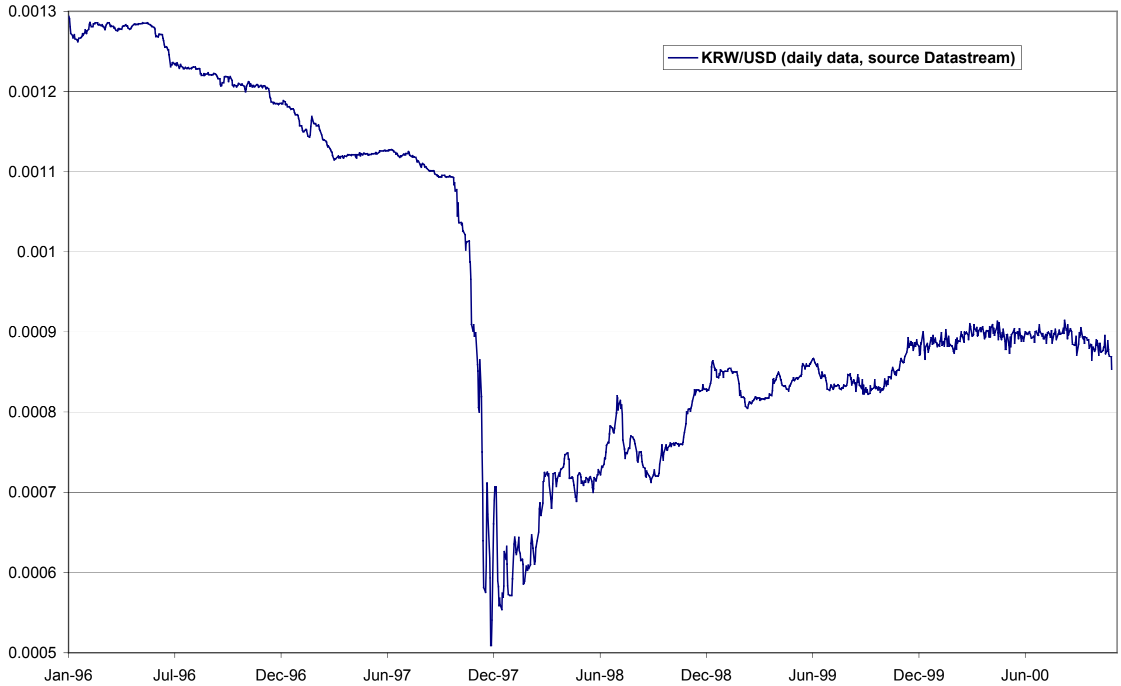

Figure 2. The evolution of the exchange rate of the Korean Won vs. the US Dollar. It is obvious that before the crisis in 1997 the distribution used to be completely different, and that in this case, the past is of no use to predict the movements after the crisis. (source Datastream, daily observations)

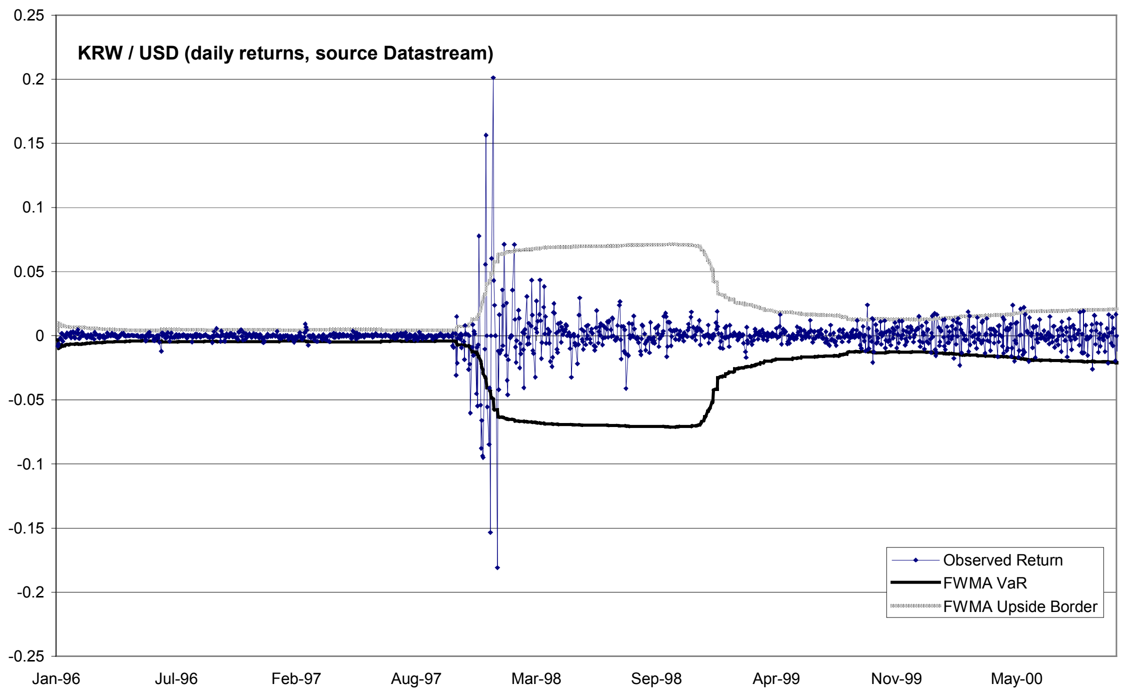

Figure 3. The return of the exchange rate of the Korean Won (KRW) versus the US dollar (USD); and the value at risk calculated according to the FWMA method with a one year window.

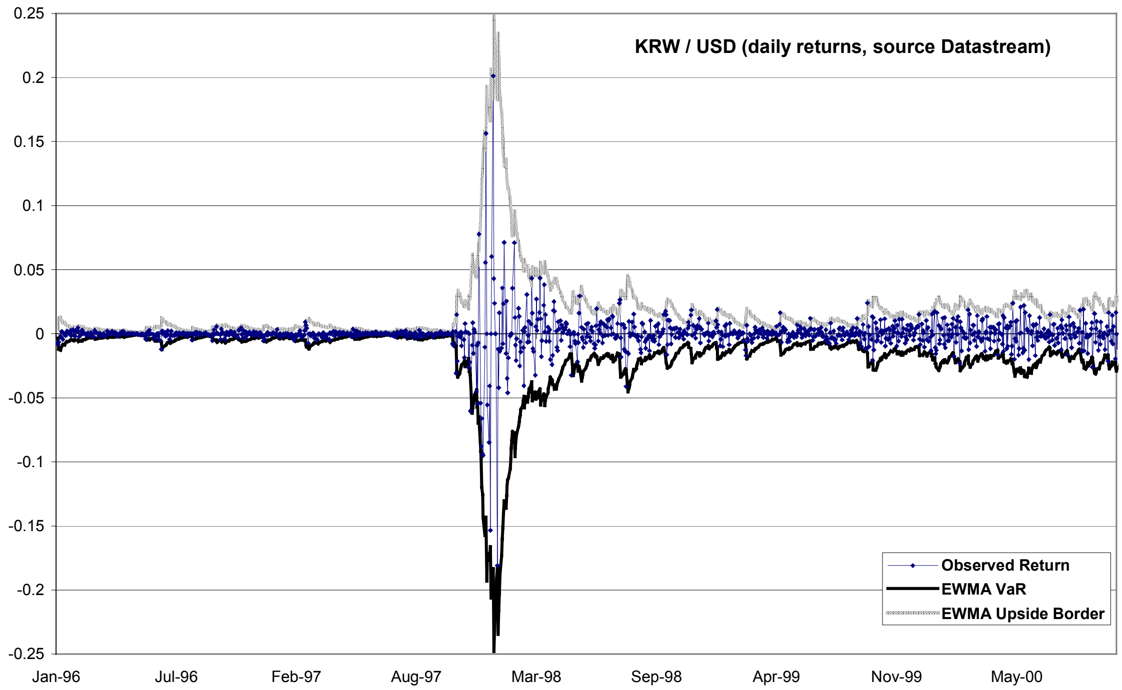

Figure 4. The return of the exchange rate of the Korean Won (KRW) versus the US dollar (USD); and the value at risk calculated according to the EWMA-method, with λ equal to 0.85.

In these calculations, we used $\lambda = 0.9$ and a one tail confidence level of 99%. It is clear that the EWMA method has some major advantages: the response to the crisis is much quicker and after the crisis the VaR will adapt promptly to the new level of volatility.

4.1.3 The multivariate GARCH

Where the EWMA estimation introduces a first order autoregressive structure, the Generalised AutoRegressive Conditional Heteroscedasticy (GARCH) model introduces a much richer structure for $\underline{\underline{\Sigma}}$. These models were proposed by Bollerslev[9].

For a start, let us have a look at the GARCH(1, 1) model. This model describes the volatility of a univariate stochastic variable.

(4.7) $$\hat{\sigma}_{t+1}^2 = \omega + \alpha\left(R_t - \mu\right)^2 + \beta\sigma_t^2$$

The main idea is that we estimate the volatility as a function of the previous estimation and the moving average. This is very close to the EWMA model, but there are two supplementary degrees of freedom. Again, one can determine the best values for the parameters ω, α, and β by quasi maximum-likelihood methods.

Since we are working with a financial portfolio that consists of multiple correlated assets, the GARCH(1,1) is not sufficient, and we will have to look at more complicated models.

(4.8) $$\hat{\sigma}_{ij,t+1} = \omega_{ij} + \alpha_{ij,1}\left(R_{i,t} - \mu_i\right)^2 + \alpha_{ij,2}\left(R_{j,t} - \mu_j\right)^2 + \beta_{ij}\sigma_{ij,t}$$

Again, it is possible to estimate all the parameters, however the number of parameters increases rapidly with the number of components in our portfolio. There are 4 N^2 parameters to be estimated, for ten components this equals 400. This rapidly increasing number of parameters makes practical implementation impossible. The following chapters will discuss some popular restrictions allowing to reduce the number of parameters to be estimated.

4.1.3.1 The Constant-Correlation GARCH

In 1990 Bollerslev proposed an approach, which is now called the “constant correlation GARCH”. The idea is that each diagonal element of $\underline{\underline{\Sigma}}_{t+1}$ (i.e. each volatility) is estimated by a univariate GARCH(1,1) model.

$$(4.9) \qquad \hat{\sigma}^2_{i,t+1} = \omega_i + \alpha_i \left(R_{i,t} - \mu_i\right)^2 + \beta_i \sigma^2_{i,t}$$

whereas the off-diagonal elements in $\underline{\underline{\Sigma}}_{t+1}$ are estimated in the framework of the “constant correlation assumption”, i.e. we assume that the correlation remains time-invariant.

$$(4.10) \qquad \hat{\sigma}^2_{ij,t+1} = \rho_{ij} \sigma^2_{i,t} \sigma^2_{j,t} \qquad (\forall\ i \neq j)$$

This scheme reduces the number of parameters to be estimated to:

(4.11) 3 N + N(N – 1) / 2

If the number of components in the portfolio is 10, then we will have to estimate 75 parameters (and not 400).

4.1.3.2 The BEKK-GARCH

One of the major drawbacks of the constant correlation approach is that this scheme does not guarantee that $\underline{\underline{\Sigma}}_{t+1}$ is positive definite. This problem is eliminated in the approach of Babba, Engle, Kraft, and Kroner (BEKK; see Engle and Kroner [10]). Their model is quadratic and not linear in its parameters, and this ensures the forecasts of $\underline{\underline{\Sigma}}_{t+1}$ to be positive definite. Their model is given by:

$$(4.12) \qquad \hat{\underline{\underline{\Sigma}}}_{t+1} = \underline{\underline{C}}'\underline{\underline{C}} + \underline{\underline{B}}'\underline{\underline{\Sigma}}_t\underline{\underline{B}} + \underline{\underline{A}}'\underline{\underline{\varepsilon\varepsilon}}'\underline{\underline{A}}$$

It is obvious that this model without any constraint does contain too many parameters (3 N^2). Usually one uses the estimation in which the matrices A, B and C are diagonal. This means that we assume that there is no cross-market influence in the time structure of the covariances.

This leads to the following results.

$$(4.13) \qquad \begin{aligned} \hat{\sigma}^2_{i,t+1} &= C_i^2 + B_i^2 \sigma^2_{i,t} + A_i^2 \varepsilon_i^2 & \quad i \in \{1,2,\ldots N\} \\ \hat{\sigma}_{ij,t+1} &= C_i C_j + B_i B_j \sigma_{ij,t} + A_i A_j \varepsilon_i \varepsilon_j & \quad i \neq j \end{aligned}$$

This looks very similar to the constant correlation approach, but here the correlations are allowed to vary. The number of parameters in the constrained BEKK approach is 3 N. If we compare this with (4.11) we notice that the number of parameters to be estimated is significantly less than in the constant-correlation approach.

In our example of a portfolio with 10 components, we only need to determine 30 parameters.

This approach is very elegant: it is less restrictive than the constant-correlation approach, it has less parameter to be estimated, while the positive definiteness of the varcov matrix is ensured.

4.1.3.3 The Orthogonal GARCH

In case we are dealing with a large number of components in our portfolio (risk management of a complete bank for example), even 3 N parameters can be too much. Therefore we can try to restrict ourselves to the most relevant risk factors, that describe the main part of the variance.

Such process is a standard statistical method called "principle component analysis" (PCA). First, we orthogonalyse the risk factors, and call them the "principle components". These principle components are linear combinations of the original risk factors, and in general they group similar risk factors together in one principle component.

Since the principle components are orthogonal, we do not need to estimate the covariances any more (they are not correlated). This is of course the same as assuming that the correlations are time invariant.

The basic ideas behind the process are as follows. Define $\underline{\underline{\Xi}}$ as the matrix of eigenvectors of $\underline{\underline{\varepsilon\varepsilon}}'$. Then, the principle components of our system can be identified as the columns of the matrix $\underline{\underline{P}}$.

(4.14) $\underline{\underline{P}} = (\underline{P}_1, \ldots, \underline{P}_N) = \underline{\underline{\varepsilon}}.\underline{\underline{\Xi}}$

Using the fact that $\underline{\underline{\Xi}}^{-1} = \underline{\underline{\Xi}}'$ we can rewrite this as:

$$\underline{\underline{\varepsilon}} = \underline{\underline{P}}.\underline{\underline{\Xi}}'$$

This allows us to write the change in each risk factor i as a linear combination of the principal components $\underline{P}_m$.

$$\varepsilon_i = \xi_{i1} P_1 + \xi_{i2} P_2 + \ldots + \xi_{iN} P_N$$

Then we can identify the estimate of $\underline{\underline{\Sigma}}_t$ as:

$$\hat{\underline{\underline{\Sigma}}}_t = \underline{\underline{\Xi}} \, \mathrm{VAR}(\underline{\underline{P}}) \, \underline{\underline{\Xi}}'$$

The number of parameters to be estimated has been reduced: we only have to determine the eigenvectors of $\underline{\underline{\varepsilon}}'\underline{\underline{\varepsilon}}$ and the diagonal elements of VAR($\underline{\underline{P}}$). And on top of that, the variance of each principal component can be modelled as a univariate GARCH.

4.1.4 The Kernel Estimation

All previously discussed methods are parametric estimations of $\underline{\underline{\Sigma}}$. The number of parameters is directly related to the richness of interactions described by the model, but the number of parameters is also directly related to the complexity of the model.

However it is possible to make estimations of $\underline{\underline{\Sigma}}_{t+1}$ by using non-parametric methods. The idea is that historical data should be weighted in order to calculate $\underline{\underline{\Sigma}}_{t+1}$:

(4.15) $$\hat{\sigma}^2_{ij,t+1} = \sum_{\nu=0}^{M-1} \omega_{ij,\nu,t} \varepsilon_{i,M-\nu} \varepsilon_{j,M-\nu}$$

These weights ($\omega_{ij,\nu,t}$) are then estimated, by a non-parametric kernel method.

4.2 Historical Simulation methods

Despite the complexity and level of sophistication, the previous theories rely on the fact that financial data are distributed according to a Normal law. As we have seen before, this is a weak point since returns of financial markets display fat tails and other deviations from the normal distribution. When using the Normal distribution, we will systematically underestimate the occurrence of extreme events.

Historical simulations completely eliminate the need for an underlying model that describes the stochastic behaviour of market prices. The only estimation made is that the past will be to some extent a good predictor for the future. Which is, as you will understand, also a "model" or even a paradigm.

4.2.1 Fixed Weight Historical Simulation

This is the most common and easy method: we take a set of sample data, and use the observed returns as the histogram of the underlying distribution. Ranking the returns allows determining the percentiles and calculating the VaR for the horizon that has been calculated and the desired confidence level.

An important point is that good modelling of the extreme events (corresponding to high confidence levels) requires a huge set of data. And, as mentioned in the previous chapter, one has to be careful when older data is used.

Of course, it is possible to end up somewhere between two returns. In such a case, we will need an interpolation method to get the exact value of the VaR. The linear interpolation is the most commonly used and the easiest to calculate.

4.2.2 Antithetic Historical Simulation

Another point that requires our attention is that asset prices often exhibit trending behaviour. For equities, cash and bonds this might be logical. But as far as exchange rate portfolios are concerned, there is no reason whatsoever to assume that the trend observed in the past would reproduce itself into the future.

This can be taken into account by removing every possible trend: for every observed return x_i we will add $-x_i$ to our data sample. This actually doubles the size of our data sample, reducing the possible measurement errors.

4.2.3 Simulated Covariances and Historical Probability Distribution

This approach uses the results of the previous chapter to estimate the varcov matrix (for example EWMA or BEKK model), but replaces in equation (4.3) $z(\alpha)$ by the historically observed value. This value will be estimated by selecting the percentile corresponding to the desired level of confidence from a set of historical returns.

This approach has the advantage of being a parsimonious calculation while taking into account the typical fat tails of returns in financial markets.

4.2.4 Exponential Historical Simulation

The problem with all previously mentioned simulation methods is the existence of two underlying contradictory trends. One trend tells us to take more data in order to make our estimations (and especially those of extreme events) more accurate, and another trend indicates that financial markets exhibit phenomena like volatility clustering and tells us not to use old data.

In the theory about the variance-covariance models we have solved this problem by using an exponentially weighting of observations. The exponential historical simulation approach does exactly the same.

We allocate a weighting factor to all our observations:

$$(4.16) \qquad \omega_{t-v} = \frac{\lambda^{v}(1-\lambda)}{1-\lambda^{M}}$$

Then we will rank all the returns in ascending order. To obtain the value at risk for a given probability level (α), we will then add all the weights starting with the weight of the smallest return till the sum of the weights equals $1 - \alpha$.

4.3 The Monte-Carlo approach

The Monte-Carlo method is in fact a simulation, it is not an experiment such as the historical simulation approach, nor a calculation like the variance-covariance method. A Monte-Carlo simulation consists mainly in drawing randomly our basic variables (the returns) from a theoretical distribution and calculating their outcome. This experiment is repeated thousands of times and all the outcomes lead to a final distribution of our resulting variable (the return of the portfolio). Ranking these "observations" we use the same method as the fixed weight historical simulations to determine the value at risk.

To some extent the Monte-Carlo approach can be considered as an enriched historical simulation. All the historical observations are used to determine a distribution law, from which our variable is drawn. Drawing a large number of "observations" from this distribution will also result in "observations" that were not really observed.

The Monte-Carlo method should be considered as a "method-of-last-resort", meaning that if we can analytically calculate the result (by the aid of a variance-covariance method for example), there is absolutely no need to use a Monte-Carlo technique. With the Monte-Carlo simulations we will only approach our analytical result.

But if we have (exotic) derivatives in our portfolio, then there is no other alternative than using the Monte-Carlo technique. Doing so we will randomly draw the returns of the basic market variables (interest rate and return of the equity market for example) and then calculate the price of the derivatives on equities accordingly.

4.4 Boot-strapping approach

This method lies somewhere in between the historical simulations and the Monte-Carlo method. The procedure is comparable to the one in the Monte-Carlo approach, however the returns are not drawn from a theoretical distribution, but from a historical set of data.

It is clear that we can use two approaches: (1) when a historical observation is drawn we remove it from our data set preventing it to be drawn a second time (since it appeared only once in reality). Or (2) we can just consider the historical data as a theoretical distribution and leave a drawn return in our data set.

Here again we can use variants like exponential weighting on historic data. The probability that it can be drawn will then be adjusted by the same factor as in (4.16).

4.5 Extreme Value Estimation

Since we are usually interested in the value at risk with relatively high confidence levels (like 99%) we are working in the tails of our distribution. As mentioned before, the Normal distribution is a very bad approximation as far as the tails are concerned. That is where the extreme value approach comes in.

Only the extreme events (the lowest returns if we are calculating VaR) are used to estimate a "tail-index", which describes the "fatness" of the tails. The theoretical attractiveness and the practical simplicity emerge from the fact that there exist only three classes of extreme value distributions. All other distribution laws will, under repeated convolution, tend to one of these three classes of extreme value distributions (EVD), one says that a particular distribution is in the "maximum domain of attraction" of a certain EVD.

(4.17)

	Cumulative Distribution Function	**Probability Density Function**	**Maximum Domain of Attraction (Examples)**

Fréchet	$\Phi_\gamma(x) = \exp(-x^{-\gamma})$ with $x > 0$ and $\gamma > 0$	$\varphi_\gamma = \gamma\, x^{-\gamma-1} \exp(-x^{-\gamma})$	Cauchy, Levy, Pareto, log-gamma.
Gumbel	$\Lambda(x) = \exp(-e^{-x})$ with $x > 0$	$\lambda_\gamma = e^{-x} \exp(-e^{-x})$	Exponential, gamma, normal, lognormal
Weibull	$\Psi_\gamma(x) = \exp(-(-x)^\gamma)$ with $x \leq 0$ and $\gamma > 0$	$\psi_\gamma = \gamma\, (-x)^{\gamma-1} \exp(-x^\gamma)$	uniform, beta

Experimental evidence shows that the Fréchet distribution is the most suitable law for modelling returns of financial assets. The fat tails predicted by the Fréchet distribution are indeed observed. The Gumbell distribution would occur in the case of "light tails". The Weibull distribution does not seem to be relevant, since it describes the tails of bounded distributions.

In all cases, one has to estimate two scaling and shifting parameters (comparable to the standardisation of a Normal distribution). In the case of fat tails, one will use the Fréchet distribution, and the additional parameter γ must be estimated. The most common estimator used is the Hill estimator:

$$(4.18) \qquad \hat{\gamma}_m = \frac{1}{\frac{1}{m-1}\sum_{i=1}^{m}\left(\log X_{(i)} - \log X_{(m)}\right)}$$

where $x_{(i)}$ is the i^{th} largest observation of the returns x, and m is the number of observations used to estimate γ.

This process is not easy, the smaller m, the higher the variance of the estimator, and on the other hand the larger m, the larger the bias in the estimator. There does not seem to be any general method of selecting an optimal m, but one can use an "arbitrary judgement approach". One plots the estimate of γ against m, and one can then select a value of m where $E[\gamma]$ seems to stabilise but where the bias is not yet important.

The value at risk is then obtained by taking the inverse of the Taylor series expansion of the Fréchet distribution function, and is usually called the "extreme quantile estimator".

$$(4.19) \qquad \hat{x}_{1-\alpha} = x_{m+1} \sqrt[\gamma]{\frac{m}{M(1-\alpha)}}$$

This approach has the advantage of keeping the correct probability mass under the tails and smoothing the tail of the empirical density function.

5 Conclusions

Where the return of a financial asset or portfolio is almost impossible to predict, the risk can be discussed and quantified to some extend. In this paper we just have had a look at the top of the iceberg, we defined value at risk and studied numerous methods to calculate it. Every specific application should be investigated in order to find out which method produces the best estimates of the VaR (and remains possible to calculate). Besides every situation should be evaluated in order to find out if value at risk on itself produces enough information. Lots of risk management applications will require other additional methods such as stress testing.

And, … this is only the beginning. Value at Risk is a very useful concept, and it can be successfully applied in various ways:

1. Communication with clients

2. Contracting with clients (using VaR instead of the symmetrical Tracking Error)
3. Optimising portfolios
4. Selecting portfolios
5. Classifying portfolios
6. Etc.

6 Acknowledgement

I would like to thank Prof. Dr. C. Vanderhoeft for carefully checking this text and providing many interesting remarks.

7 References and Endnotes

[1] Markowitz Harrry M., “Portfolio Selection”, Journal of Finance, 1952, vol. 6, pp. 77-91.

[2] Markowitz Harry M., “The Utility of Wealth”, Journal of Political Economy, 1952, vol. 60, pp. 151-158.

[3] See for example the book “Beyond Greed and Fear” of Hersh Sheffrin; or Shefrin Hersh, and Statman Meir, “Behavioural Portfolio Theory”, working paper Sancta Clara University, 1997 (completely revised in 1999).

[4] One of the best books about this subject is: Bouchard, J.P. and Potters M. (1997), “Théorie des risques financièrs”, Commisariat à l’énergie Atomique, ISBN 2-7272-0190-7. Another brilliant explanation can be found in the books of Edgar E. Peters.

[5] Please note the subtle difference in notation: VaR is “value at risk”, whereas VAR stands for “variance”.

[6] This feature is well described in the framework of behavioural finance.

[7] Engel, J. and M. Gizycki (1999). “Value at Risk: On the Stability and Forecasting of the Variance-Covariance Matrix”, Reserve Bank of Australia Research Discussion Paper, Forthcoming.

[8] Bollerslev, T. (1990), “Modelling the Coherence in Short-Run Nominal Exchange Rates: A multivariate generalised ARCH model”, *The Review of Economics and Statistics*, **72(3)**, pp. 498-505.

[9] Bollerslev, T. (1986), “Generalised Autoregressive Conditional Heteroscedasticity”, *Journal of Econometrics*, **31**, pp. 307-327.

[10] Engle, R. and K. Kroner (1995), Multivariate Simultaneous Generalised ARCH”, *Econometric Theory*, **11(1)**, pp. 122-150.