Identifying Stress Test Scenarios*

Thomas Breuer
Fachhochschule Vorarlberg
Achstraße 1, A-6850 Dornbirn, Austria

Gerald Krenn
Oesterreichische Nationalbank
Otto Wagner-Platz 3, A-1090 Wien, Austria

Abstract

The usefulness of stress tests as risk management tool crucially depends on the choice of stress scenarios: if the scenario is too implausible, the stress test results will be ignored by management. Unfortunately, there is no generally accepted standard how to choose scenarios. Many institutions apply some standardized scenarios derived from extreme market moves in the past. A substantial drawback of such scenarios is the fact that they usually ignore the characteristics of the portfolio under consideration. Here we make two points: (1) We argue for the importance of taking into account portfolio characteristics when choosing stress scenarios, and (2) we compare four different methods in terms of the loss incurred by the resulting scenarios *and* the plausibility of these scenarios. These methods are: the standard scenarios proposed by the Derivatives Policy Group, naive modelling of historical crisis, portfolio-dependent modelling of historical crisis, and Monte Carlo search for worst case scenarios. Portfolio-dependent methods of scenario identification – in particular Monte Carlo search – produce scenarios which are at the same time more severe and more plausible. In passing we introduce a new measure of plausibility for stress scenarios.

* The statements made herein are the authors' opinions, and do not necessarily reflect the views of the Fachhochschule Vorarlberg or of the Oesterreichische Nationalbank.
e-mail: Thomas.Breuer@fh-vorarlberg.ac.at
e-mail: Gerald.Krenn@oenb.co.at

Table of Contents

1 Importance of Portfolio Specific Stress Tests

1.1 Importance of Stress Tests in Risk Management

Both the Basle Committee's market risk paper of January 1996 and the EU Directive (CAD II) stipulate that the use of an internal model be subject to approval by banking supervisory authorities. What is more, both papers spell out stress testing as one of the prerequisites for model approval. In other words, bank regulators consider stress tests to be an effective and necessary tool that complements statistical models for quantifying and monitoring risk.

The need for stress testing is justified by the Basle Committee (1995) as follows:

> "Banks that use the internal models approach for meeting market risk capital requirements must have in place a rigorous and comprehensive stress testing program. Stress testing to identify events or influences that could greatly impact banks is a key component of a bank's assessment of its capital position.
>
> Understanding and protecting against the vulnerabilities of a financial company's risk-taking activities is of course one of the major responsibilities of its board of directors and senior management. Banks' stress scenarios need to cover a range of factors that can create extraordinary losses or gains in trading portfolios, or make the control of risk in those portfolios very difficult. These factors include low-probability events in all major types of risks, including the various components of market, credit, and operational risks. Stress scenarios need to shed light on the impact of such events on positions that display both linear and non-linear price characteristics (i.e. options and instruments that have options-like characteristics).
>
> Banks' stress tests should be both of a quantitative and qualitative nature. Quantitative criteria should identify plausible stress scenarios to which banks could be exposed. Qualitative criteria should emphasise that two major goals of stress testing are to evaluate the capacity of the bank's capital to absorb potential large losses and to identify steps the bank can take to reduce its risk and conserve capital. This assessment is integral to setting and evaluating the bank's management strategy and the results of stress testing should be routinely communicated to senior management and, periodically, to the bank's board of directors."

Thus, stress tests should provide financial institutions with answers to three kinds of questions:

1. What will be the loss be in the event of scenario X?
2. What are our institution's worst-case scenarios?
3. What can we do to limit the losses incurred in the worst-case scenarios?

In this paper we will concentrate on the second kind of question.

1.2 Stress Tests and Value-at-Risk

Regulators require stress testing as a complement to VaR models. Why do VaR models call for such complementary measures, and how come stress tests fit the bill?

The VaR methodology is fairly well-known: A *holding period* of t days and a *confidence level* of p% are given. The VaR is a statistical measure of the loss of a portfolio – as measured in monetary units – which will not be exceeded with a probability of p% given the portfolio remains constant throughout the holding period. Losses in excess of the VaR only occur with a low probability [(1-p)%]. A VaR model does not shed light on the dimension of such heavy losses. This is the first reason why stress testing is required as a complementary measure: stress tests serve to capture potential extreme losses.

The second important reason why VaR calculations shall be combined with stress tests lies in the somewhat skeptical attitude towards the assumptions on which most VaR calculations are based. In the same vein, the multiplication factor applied to the VaR in computing the regulatory capital requirement helps absorb the remaining uncertainty about the accuracy of the model.

There are first and foremost two assumptions whose validity is debatable. For one, the market characteristics are assumed not to change in the future. Only in the event that future market movements resemble those of the past can models produce reliable results. Yet, there have always been breaks in market movements. They may be attributable to various causes, for instance, to full-blown crises, such as wars or environmental catastrophes, changes in the interest rate or exchange rate policies pursued by central banks, speculative attacks on currencies and the like. In this context, a stress situation means a break in the temporal constancy of a market. The objective of stress tests is, among other things, to assess the potential loss resulting from such breaks.

Furthermore numerous VaR models assume that changes in risk factors are normally distributed. However, changes in financial time series are, as a rule, not normally distributed. Instead, such time series are marked by fat tails. It follows that extreme changes in the risk factors are considerably more likely than is accounted for under the assumption of a normal distribution. The slump in stock prices triggered by the equity crash of 1987, for example, amounted to something between 10 and 20 standard deviations. Considerng that under normality a 7 standard deviations change should on average occur at one day in a billion years, the assumption of normality seems inadequate. Stress tests are not based on statistical assumptions on how the changes in risk factors are distributed. This is why the results of stress tests are not distorted by fat tails.

As stress tests do not quantify the probability of occurrence of the individual scenarios, they lend themselves to verifying and complementing statistical risk measures such as VaR. While stress tests do not put exact figures on the probability of scenarios, scenarios still need to be plausible – in some sense still to specify. The evaluation of scenario plausibility calls for, at least, a rough idea of the probability with which given scenarios will occur (Berkowitz 1999).

1.3 Shortcomings of Conventional Methods for Identifying Stress Scenarios

1.3.1 Standardized Stress Tests

Many banks conduct periodic stress tests involving a revaluation of their current portfolio against certain standard scenarios. These are often standard scenarios in a dual sense: the choice of the scenarios depends neither on the bank nor on the timing of the stress test.

Thus, stress testing with standard scenarios has the advantage of guaranteeing comparability in two respects. Firstly: when several banks look at the same scenarios one can compare the outcome of stress tests of different banks. This allows the supervisor to assess the banks' exposure to those risk categories whose risk factors are changed in the standard scenarios. Secondly: when a bank always looks at the same scenarios, it can compare the results of stress tests conducted at different points in time. This enables it to monitor how its exposure to the risk categories in the standard scenarios changes over time (exposure monitoring).

Many banks use standard scenarios similar to the stress scenarios proposed by the Derivatives Policy Group (DPG) as in their "Framework for Voluntary Oversight." The DPG recommends the performance of stress tests to measure the exposure of a portfolio to certain core risk factors. The scenarios proposed by the DPG include: parallel yield curve shifts of 100 basis points up and down, a change in slope of the yield curves, a combination of shifting and changing the slope, a change in yield volatilities by 20 percent of prevailing levels, a change in equity index values by 10 percent etc.

A comparison of these DPG standard scenarios with actual maximum changes shows that some of the DPG scenarios are far removed from the maximum changes observed in the past. Therefore, they should not be regarded as reconstructions of historical crises nor as worst-case scenarios.

1.3.2 Stress Scenarios Modelled after Historical Crises

The Basle Committee (1996; section B.5 no 6) requires the construction of stress scenarios on the basis of historical crises:

> "Banks should subject their portfolios to a series of simulated stress scenarios and provide supervisory authorities with the results. These scenarios could include testing the current portfolio against past periods of significant disturbance, for example, the 1987 equity crash, the ERM crises of 1992 and 1993 or the fall in bond markets in the first quarter of 1994, incorporating both the large price movements and the sharp reduction in liquidity associated with these events."

One may ask now, why use reconstructions of historical crises? After all, VaR models also use historical data. How can stress tests supply additional information if they are based on the same data? One major difference between the two methods is that VaR models usually include only data from a relatively short previous period – e.g. the previous year – while stress testing can be used to reconstruct exceptional market situations which occurred at a more distant point in the past. And because VaR models use all data from the recent past, including calm market periods, peaks tend to be smoothed out. Conversely, when historical crises are modeled, only periods of dramatic market movements are taken into account while data from calm periods is discarded. As a result, the peaks of market movements can be modeled in full force.

Stress scenarios which use historical data model stress events after extreme historical market movements. But this approach involves some danger: extreme scenarios are not necessarily worst-case scenarios; the maximum changes are not necessarily the worst that can happen. Certain portfolios are more dramatically affected by slight changes than by major movements. And if one pursues a straddle strategy, for example, the worst that can happen is that the market does not move at all.

1.4 Portfolio Specific Worst-Case Scenarios

The Basle Committee on Banking Supervision (1996; section B.5 no 4) requires that the composition of portfolios be taken into account in selecting stress scenarios:

> "Banks should combine the use of supervisory stress scenarios with stress tests developed by banks themselves to reflect their specific risk characteristics."

and goes on to say (1996; section B.5 no 7):

> "In addition to the scenarios prescribed by supervisory authorities [...], a bank should also develop its own stress tests which it identifies as most adverse based on the characteristics of its portfolio (...)."

The search for worst-case scenarios differs from the construction of historical scenarios in two main aspects. *Firstly, past crises or scenarios constructed on the basis of historical maximum movements are not necessarily worst-case scenarios*. There may well be potential market movements which have not yet occurred, but which would result in worse consequences for the bank than the historical crises which did occur. Neither are maximum movements necessarily worst-case scenarios, for certain portfolios may suffer the greatest damage when risk factors move only slightly. In an attempt to identify worst-case scenarios, one does not only consider events which occurred at some point in the past, but also all potential future scenarios. For this reason, hypothetical scenarios are also called "forward-looking scenarios".

Secondly, the construction of scenarios using historical data pays little attention to the characteristics of the bank's portfolio. At best, the current portfolio plays a role in the selection of the risk factors subjected to change, or when portfolio-related measures are applied to assess the severity of risk factor changes – as, for example, the use of sensitivities, or of P&Ls (Shaw 1997). Apart from these exceptions, the portfolio of the bank is of minor importance in the construction of scenarios from historical data. Conversely, the portfolio plays a central role in defining worst-case scenarios. What may be a worst-case scenario for one portfolio, may result in profits for another.

Neither national nor international supervisors include any provisions on how to identify portfolio-specific worst-case scenarios. There are two fundamental options: A bank may rely on the experience and economic expertise of staff from as wide a range of fields as possible, who use their knowledge of the market, of the portfolio and of the trading and hedging strategies of the bank in an attempt to identify those market situations which could lead to particularly high losses of the bank. This approach may be termed the *subjective search* for worst-case scenarios. But a bank may also use its computers to search systematically for worst-case scenarios. This may be called a *systematic search* for worst-case scenarios.

Stress tests which use historical or subjectively presumed worst-case scenarios may overlook fatal stress scenarios. They determine the potential loss only at very few points within the high-dimensional space of scenarios. One difficulty with historical and suspected worst-case scenarios is that knowing which losses a portfolio can be expected to suffer under a few selected scenarios may give the bank a false sense of security, if the projected losses are manageable. The sense of security may be false because the bank does not know whether there are other scenarios which are equally plausible and result in much heavier losses. Even in a subjective search for suspected worst-case scenarios, one cannot know whether the scenarios found are actually the worst ones.

2 Systematic Search for Portfolio-Specific Worst-Case Scenarios

2.1 Description of the Problem

The concept of stress testing is based on the notion that the value of a portfolio depends on *market risk factors (risk factors)*. Let us call the risk factors with an impact on the portfolio $r_1, r_2, ..., r_n$ and the function determining the value of the portfolio when the values of all risk factors are given, P. The values of the risk factors $r_1, r_2, ..., r_n$ characterize the market situation as far as it is of relevance to the portfolio. The risk factors may be combined into one single vector $\boldsymbol{r} := (r_1, r_2, ..., r_n)$ describing the market situation. In a market situation $\boldsymbol{r}$, the value of the portfolio is $P(\boldsymbol{r})$. Below, $\boldsymbol{r}_{MM}$ will stand for the vector representing the current values of the risk factors, i.e. the current market situation. $P(\boldsymbol{r}_{MM})$ therefore represents the current value of the portfolio.

The choice of risk factors depends on the portfolio. Different portfolios are, in general, influenced by different risk factors. The number of risk factors must be chosen so as to include all parameters likely with an impact on the value of the portfolio. The function P depends on the portfolio: a different portfolio has a different valuation function.

Stress tests answer the question of "What would happen if a market situation $\boldsymbol{r}$ suddenly occurred?" The scenario in this case is the sudden emergence of a market situation $\boldsymbol{r}$. Scenarios may therefore be identified with market situations and represented by vectors $\boldsymbol{r}$. In general language, a "scenario" is a potential future development. In connection with stress testing, a scenario is a possible future market situation. In this context, the term scenario therefore does not stand for a process but only for its outcome.

In most of the cases, this change in meaning makes sense. Disturbances in financial markets are characterized by a sudden confrontation of market participants with a changed market situation. This may have been caused, for example, by a dramatic rise in volatilities: when prices move so rapidly that market participants are unable to restructure their portfolios within the reaction time available, the portfolios have to be revalued on the basis of changed market conditions. The same effect occurs in liquidity crises: for a market participant, only those prices are relevant at which he can rebalance his positions to the extent desired. In illiquid markets, trading close to quoted market prices is impossible. Therefore, a portfolio can be restructured only at a later time and at dramatically different prices.

On the other hand, reducing stress tests to instantanous changes in risk factors can be questionable in some cases. The value of path dependent options not only depends on some future market situation but also on how this situation comes about. Stress scenarios for portfolios containing path dependent options should be characterized by paths representing the evolution of markets instead by a single future market situation. In the following we will disregard this complication.

For stress testing, a scenario $\boldsymbol{r}_{stress}$ is selected according to specific criteria and the value of the current portfolio under this scenario is calculated. Denote it by $P(\boldsymbol{r}_{stress})$. By comparing them with the current value of the portfolio $P(\boldsymbol{r}_{MM})$ one can assess the losses that would be incurred if the market suddenly moved from $\boldsymbol{r}_{MM}$ to $\boldsymbol{r}_{stress}$ without allowing a chance for rebalancing the portfolio.

2.2 Plausibility Standards

In general it will be impossible to find a market state in which the portfolio has its *smallest* value, since the loss potential of a portfolio is usually unlimited. A simple example is that of a portfolio which consists only of a short call: its value will fall without limit as long as the value of the underlying instrument rises. For this reason, not all scenarios will be admitted; rather, the search will be for the minimum among those scenarios which meet certain plausibility standards.

The plausibility of scenarios matters for the interpretation of stress test results. Stress test results which show heavy losses for a bank will more readily lead to counter-measures if decision-makers tend to regard the scenario as plausible. Plausibility standards should therefore exclude scenarios which are next to impossible and could for this reason undermine the credibility of stress test results. For this reason, correlations between risk factors are to be taken into account when identifying stress scenarios (Kupiec 1998). A scenario in which risk factor movements straightly violate correlations is highly unplausible, even if every individual risk factor movement is fairly plausible. For that reason, correlations have to be taken into account when defining plausibility conditions.

The well-known tendency of correlations to change abruptly in stress events is no valid argument against the inclusion of correlations in the formulation of plausibility standards. For the plausibility standards can be based on crisis correlations as well as correlations in calm periods. (Finger and Kim 2000).

How can we include correlations in the definition of admission conditions for scenarios? Assume a variance-covariance matrix of risk factor changes,

$$S := \begin{pmatrix} s_1^2 & s_{12} & \cdots & s_{1n} \\ s_{21} & s_2^2 & \cdots & s_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ s_{n1} & s_{n2} & \cdots & s_n^2 \end{pmatrix},$$

with the variances σ_i^2 of the risk factor changes on the diagonal, and the covariances $\sigma_{ij} = \sigma_{ji} = \sigma_i \sigma_j \rho_{ij}$ of the risk factor changes outside the diagonal. A normal distribution of n variables with a density of

$$P(\Delta r_1,...,\Delta r_n) = const \cdot \exp(-\tfrac{1}{2}(\Delta r_1,...,\Delta r_n)^T \cdot \Sigma^{-1} \cdot (\Delta r_1,...,\Delta r_n))$$

results exactly in these correlations. If the risk factor changes were normally distributed, and given these correlations, then the scenarios $\mathbf{r}$, to which a leap from the present scenario $\mathbf{r}_{MM}$ is equally probable, would form an n-dimensional ellipsoid

$$(\mathbf{r}_{MM} - \mathbf{r})^T \cdot \Sigma^{-1} \cdot (\mathbf{r}_{MM} - \mathbf{r}) = k^2 .$$

The lengths of the major axes of the ellipsoids are k times the eigenvalues of the matrix Σ^{-1}. If the risk factor changes are normally distributed, and given covariances Σ, the probability that the market state $\mathbf{r}$ lies within the ellipsoid is determined by the value of the χ^2 distribution function with n degrees of freedom at k^2,

$$F_{\chi^2_n}(k^2) = \frac{1}{2^{n/2}\,\Gamma(n/2)} \int_0^{k^2} s^{\frac{n}{2}-1} e^{-\frac{s}{2}} ds .$$

Admission criteria for scenarios may now be specified as in the following example.

1. A confidence level p is set, for example, $p = 95\%$.
2. Determine the k^2 for which $F_{\chi^2_n}(k^2) = p$.
3. The admissibility domain is defined as the set of all scenarios $\mathbf{r}$ which fulfil

$$(\mathbf{r}_{MM} - \mathbf{r})^T \cdot \Sigma^{-1} \cdot (\mathbf{r}_{MM} - \mathbf{r}) \le k^2 .$$

Let us denote this filled ellipsoid by E.

Only if the risk factor changes are normally distributed with covariance matrix Σ, is it justified to state that with a probability p one of the scenarios will be situated within the k times enlarged ellipsoid with covariances Σ. It must be doubted, however, that the same correlations will apply in stress periods and in untroubled periods, and that the risk factor changes are indeed normally distributed, instead of showing fat tails, for example. For this reason, one cannot as a rule say that a scenario will be situated in the ellipsoid with a probability p. Even so, the ellipsoids can serve as suitable admissibility domains for scenarios

It should be noted again that present correlations are not the only choice for the covariance matrix Σ. It may also be useful to use stress correlations for the correlation matrix. Stress correlations can be estimated, for example, on the basis of historical stress event data (see Finger and Kim 2000).

2.3 Methods for the Systematic Search for Worst-Case Scenarios

2.3.1 Factor Push Method

The factor push method is a relatively simple way of getting a rough idea of worst-case scenarios. The basic process is to change each individual risk factor by a given value in the direction that will most reduce the portfolio value. More specifically, one proceeds as follows:

1. For each risk factor r_i, the portfolio values which result from a positive and a negative risk factor change by a defined value, are determined. The value of the change is usually defined as a multiple of the standard deviation of the risk factor change under consideration, e.g. k times the standard deviation. So the two values $P(r_{MM,1},...,r_{MM,i}(1+ks_i),...,r_{MM,n})$ and $P(r_{MM,1},...,r_{MM,i}(1-ks_i),...,r_{MM,n})$ are calculated.
2. A plus or minus sign $VZ(i)$ is assigned to each risk factor, using the formula

$$VZ(i) := \operatorname{sgn}\left(P(r_{MM,1},...,r_{MM,i}(1+ks_i),...,r_{MM,n}) - P(r_{MM,1},...,r_{MM,i}(1-ks_i),...,r_{MM,n})\right).$$

   $VZ(i)$ is 1 if the upward change of the i-th risk factor results in a higher portfolio value than the downward change. Otherwise, $VZ(i)$ is -1.
3. The new stress scenario can be written as:

$$\boldsymbol{r}_{WC} := \left(r_{MM,1}\cdot[1-VZ(1)\cdot ks_1],..., r_{MM,n}\cdot[1-VZ(n)\cdot ks_n]\right).$$

One of the main advantages of this method is that computation is easy. Only $2n$ evaluations of the valuation function are necessary to find the new scenario.

Two drawbacks of the method should be mentioned: firstly, it supplies only scenarios which are situated on the surface – or, more precisely, at a corner – of the n-dimensional cuboid

$$\left\{\boldsymbol{r}\in R^n : r_i = r_{MM,i}(1\pm ks_i)\right\}.$$

This may be of little consequence for portfolios with linear valuation functions. But for portfolios with non-linear valuation functions, the valuation function minimum may be situated within the n-dimensional cuboid. This may easily be the case for portfolios which contain derivative instruments.

For this reason, the simple factor push method is unsuitable for portfolios containing derivatives. This can partly be remedied by applying the factor push method not only for one, but for several values of k, e.g. k = 1/2, 1, 3/2, 2, ..., 10. This produces minima on the surfaces of 20 cuboids which are stuck within each other. The shorter the distance between one cuboid surface and the next, the more precisely can we localize minima within the biggest cuboid. The amount of computation required increases in linear proportion to the number of cuboids considered.

Another drawback of the method is that it cannot even be safely assumed that it will find the minimum on the cuboid surface. There are portfolios whose minimum over the cuboid surface is not situated at a corner.

2.3.2 Monte Carlo and Quasi-Monte Carlo Methods

Monte Carlo and quasi-Monte Carlo methods provide an approximation for the minimum of the portfolio value. They are relatively simple, but require substantial computation power. To find an approximate minimum within the admissibility domain defined by the plausibility condition, one could, for example, proceed as follows:

1. Generate a series $\boldsymbol{x}_j$ of random points or quasi-random points in the n-dimensional cuboid Q centered at the origin with edges of length $2k/\sqrt{EW(i)}$ parallel to the coordinate axes. Here $EW(i)$, $i=1,\dots,n$, denote the eigenvalues of Σ^{-1}.
2. Determine the unitary matrix U with the eigenvectors of Σ^{-1} as columns. This maps the cuboid Q onto the smallest cuboid containing the ellipsoid E of points fulfilling $(\boldsymbol{r}_{MM} - \boldsymbol{r})^T \cdot \Sigma^{-1} \cdot (\boldsymbol{r}_{MM} - \boldsymbol{r}) = k^2$. The main axis of E are of length $k/\sqrt{EW(i)}$.
3. Discard those transformed random points $U\,\boldsymbol{x}_j$ which lie outside E. Denote by S the sequence of remaining transformed random points.
4. An approximation for the minimal portfolio value in E is

$$m_N(P,S) := \min_{1\le i\le N} P(\boldsymbol{r}_i).$$

   N is the number of runs of the Monte Carlo search. As worst case scenario one can choose a point $\boldsymbol{r}_j$ with $P(\boldsymbol{r}_j) = m_N(P,S)$.

Compared to the factor push method, the most important advantage of Monte Carlo methods and quasi-Monte Carlo methods is that minima can be found not only on the surface of an n-dimensional cuboid, but anywhere. This is absolutely necessary for portfolios with strongly non-linear elements, in particular option portfolios. As a refinement it is possible to gradually shrink the search domain. This prevents that the algorithm spends too much time evaluating P in regions far from the minimum and thus increases efficiency.

The generation of the random vectors is the chief point on which an efficient implementation of these methods depends. The more uniformly and densely the n-dimensional unit cube is filled by the random vectors, the more reliable is the approximation of the minimum portfolio value. The more irregular the surface of the valuation function as a function of the risk factors, the more densely must the interior of the n-dimensional unit cube be filled by the random vectors to produce a reasonably reliable approximation of the minimum.

The efficiency of Monte Carlo methods and quasi-Monte Carlo methods is determined by the number of random vectors required. The extent of computation power required is caused by the fact that the valuation function has to be computed separately for each random vector. It is not necessary to compute the series of random vectors afresh every time; rather, it can be saved and used again for any problem with the same dimension – i.e. with the same number of risk factors.

Once the desired density of coverage is fixed, the required number of random vectors is determined by how uniformly they fill the unit cube. Discrepancy is a mathematical measure for the deviation from the greatest possible uniformity of filling. The lower the discrepancy of a series of random vectors, the smaller the series may be to reach a given coverage density and thus, a given accuracy of approximation.

What is decisive for the efficiency of Monte Carlo methods and quasi-Monte Carlo methods is not the randomness of the random vectors, but the filling density. It will therefore be more efficient to select series of random vectors with as little discrepancy as possible, even though these series may contain only quasi-random numbers, rather than genuine random numbers or pseudo-random numbers. A more detailed discussion of the efficiency of Monte Carlo methods and quasi-Monte Carlo methods can be found in Niederreiter (1992).

2.3.3 Other Loss Maximization Algorithms

Monte Carlo methods and quasi-Monte Carlo methods fill the admissibility domain defined by the plausibility condition as uniformly as possible with points at which the portfolio is then evaluated. A drawback of these methods – if applied without shrinking the search domain – is that many points are situated in parts of the search domain where no valuation function minimum is expected. Other methods promise to be more efficient in minimizing the valuation function. In practice, however, most valuation functions are so complicated that it would require too much computation power or be simply impossible to calculate valuation function derivatives with respect to risk factors. This leaves banks with minimization algorithms which require only evaluation of the function, but not of its derivatives. Descriptions of the following algorithms, including programming instructions, can be found in Press et al. (1992)

The *multidimensional simplex method* was first described by Nelder and Mead. (It should not be confused with the simplex process which is used in linear programming to find extreme values of a linear function.) A simplex in an n-dimensional space consists of a vertex and n linearly independent vectors. The simplex is the n-dimensional domain which is created if the n vectors act at the vertex. Beginning with a start simplex, the algorithm determines a series of wandering simplexes of shirnking size which approach a domain in which a local minimum of the valuation function is situated. The series can be halted if the distance between a new simplex and the preceding one gets smaller than a certain tolerance, or if the value of the valuation function diminishes by less than a given tolerance from one step to the next. The resulting scenario is the vertex of the last simplex.

The multidimensional simplex method is relatively simple, but it requires a rather high number of function evaluations. A more efficient method, but one which is more complicated to implement, is the multidimensional Powell method. This method consists of steps whereby in each step, n one-dimensional minimizations in n directions are performed. The crucial point is the determination of the n directions for the next step of n minimizations. In this, one of two strategies may be followed: One either searches for directions which correspond as closely as possible to the directions of the valleys of the valuation function, or one searches for directions with the characteristic that the minimization in one direction is not destroyed by the subsequent minimization in another direction. Implementations of both strategies can be found in Press et al. (1992; pp. 413-420).

The *simulated annealing method* has received much attention because it can be used to solve optimization problems which are notorious for their high computation requirements. ("Annealing" is a term used for the slow cooling-off process of metals which leads to a state of minimum energy.) The particular strength of this method lies in dealing with cases where the desired global minimum is hidden among many small local minima. The special feature of the method is that it proceeds from one scenario to the next not by a deterministic, but by a stochastic process. On the basis of one scenario, a candidate for a new scenario is randomly selected. Assume that the difference between the valuation function values of the would-be scenario and the old scenario is ΔP. If the valuation function has a lower value in the would-be scenario, it is realized; if the valuation function has a higher value in the would-be scenario (i.e. $\Delta P > 0$), then it is realized only with a probability of $e^{-\Delta P/T}$. The parameter T corresponds to temperature and determines the inclination of the system to go into a market state with a higher portfolio value. As the search process continues, T – and thus, the inclination of the system to realize market states with higher portfolio values – is gradually reduced. The number of searches and the extent by which the parameter T is reduced are determined in an "annealing schedule". The selection of the annealing schedule is vital for the efficiency of the algorithm.

The simulated annealing method carries a lower risk of getting stuck in a local minimum than other minimization algorithms, because the process can also move to market states with higher portfolio values. The step-by-step reduction of the parameter T corresponds to the gradual shift from rough searches to fine-tuned searches.

As the risk factors have a continuous domain, the simulated annealing method is more difficult to implement in this case than in the minimization of functions with a discrete domain. An implementation is given in Press et al. (1992, pp. 451-455). For the purposes of a search for worst-case scenarios, the risk factor domain may also be discretized, provided that a sufficiently fine-meshed grid is selected. The sharper the peaks of the valuation function, the more fine-meshed the grid to be selected.

2.4 Reporting Stress Test Results

How can the results of a search for portfolio-specific worst-case scenarios be presented in a concise and readily understandable manner? It is certainly not enough to simply report the values of the risk factors in the worst-case scenario that has been found. For example, listing 500 risk factors of the worst-case scenario would hopelessly overtax the capacity of any recipient of the report. Consequently, reports should include only the *most important* risk factors in the worst-case scenario.

What are the "most important" risk factors of a worst-case scenario? Sensitivities are certainly not an appropriate indicator of the importance of a risk factor: sensitivities in the present market state are completely unrelated to the worst-case scenario to be characterized; and all sensitivities will be zero in the worst-case scenario if it is a local minimum.

The following approach appears more useful: The search for the key risk factors is a search for a subset of risk factors which have a certain explanatory power, i.e.which explain the loss under the worst-case scenario up to a previously defined degree. For example, an explanatory power of 80% means that we are looking for a subset of the risk factors which will be able to explain at least 80% of the loss under the worst-case scenario. This means: Let us assume that, instead of the complete worst-case scenario $\boldsymbol{r}_{WC} = (r_{WC,1},\dots,r_{WC,n})$, only the values of a subset of w risk factors $r_{i_1}, r_{i_2},\dots,r_{i_w}$ are reported. This corresponds to a simplified report scenario

$$\boldsymbol{r}_{report} := (r_{MM,1},\dots,r_{WC,i_1},\dots,r_{WC,i_2},\dots,r_{WC,i_w},\dots,r_{MM,n}),$$

where the risk factors $r_{i_1}, r_{i_2}, ..., r_{i_w}$ have their worst-case values $r_{WC,i_1}, ..., r_{WC,i_w}$, and all other risk factors have their actual values. The subset of risk factors will explain 80% of the loss suffered from the worst-case scenario if

$$P(\boldsymbol{r}_{MM}) - P(\boldsymbol{r}_{report}) \geq 0.8(P(\boldsymbol{r}_{MM}) - P(\boldsymbol{r}_{WC}))$$

applies. The task is to find a fairly small w and a set of w risk factors which has a high explanatory power. This can be solved by optimization algorithms in a discrete w-dimensional space, e.g. by simulated annealing.

3 A Comparison of Methods for Identifying Stress Scenarios

Now we compare four different methods for identifying stress scenarios on two sample portfolios. Portfolio 1 contains European options on the Dow Jones Industrials (DJI) and the DAX equity indices and Portfolio 2 contains exclusively equity positions proportional to their representation in these two indices, as well as cash in USD and EUR. We restrict ourselves to three risk factors: the DJI, the DAX and the USD/EUR exchange rate. Other risk factors for the option portfolio (EUR and USD interest rates, DJI and DAX volatilities) were assumed to be constant. All options were taken to have the same expiration date.

Figures 1 and 2 show the behaviour of the values of Portfolios 1 and 2 in dependence of the relative changes of the DJI and the DAX indices when the USD/EUR rate is kept constant at its present value. In the Figures it is easy to locate the worst case scenarios, namely the points at which the portfolio value is minimal. Unfortunately such intuitive methods are not available for more than two dimensions.

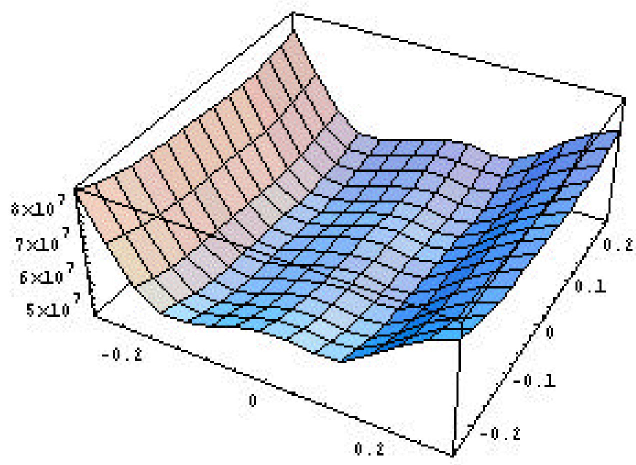

Figure 1: Value of Portfolio 1 as a function of relative changes in DJI and DAX

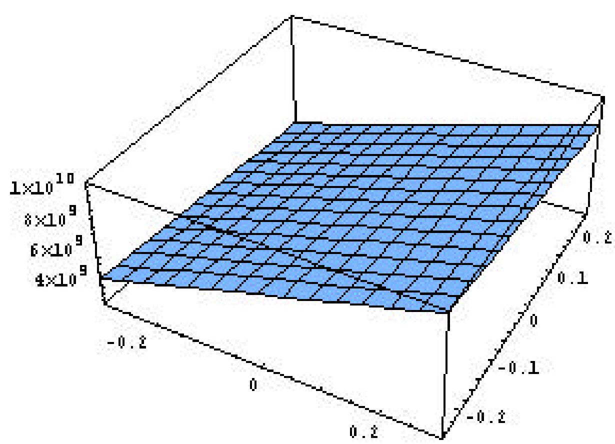

Figure 2: Value of Portfolio 2 as a function of relative changes in DJI and DAX

We used the following four methods to identify stress scenarios for Portfolios 1 and 2. Method 1, in the sequel referred to as "DPG Scenarios", applies all the relevant scenarios proposed by the DPG to the portfolio: changes of the DAX and the DJI by +/-10% while leaving the exchange rate unchanged, and a move in the exchange rate by +/-6% while leaving the equity indices unchanged. Of the four resulting scenarios we picked as stress scenario the one leading to the worst loss.

Methods 2 and 3 model the 1987 equity crash with different degrees of sophistication. Method 2 ("Black Friday") models the crisis as the move of the risk factors on Black Friday, where the DJI experienced its worst slump, 22.9%. It might seem interesting to ask what is the loss of an equity or an equity option portfolio if Black Friday happened again tomorrow. Method 3 ("Historical Simulation") is much more portfolio-specific. It calculates the relative daily changes of the risk factors during fourty trading days around the crisis (October and November 1987), and then picks as stress scenario the relative change which would cause the greatest harm if applied to the current portfolio. This method strongly resembles the technique of historical simulation in calculating VaR, but in the end one picks the worst change rather than the worst change of probability greater than say 5%.

Method 4 uses Monte Carlo simulation with 500 runs to determine the worst case scenario within the ellipsoid defined by the worst historical scenario determined by Method 3 together with the covariance matrix. We roughly proceed as described in Section 2.3.2, but without gradually focusing the search domain. The result of Method 4 will therefore be a worst case scenario which is at least as plausible as the scenario determined by "Historical Simulation". (Alternatively, it would have been possible to search for the most plausible scenario yielding a worse loss than the scenario determined by "Historical Simulation".)

It turned out that conventional Monte Carlo simulation is too crude an optimisation algorithm. Even in the highly simplified framework of three risk factors the speed of convergence is far from satisfactory. The need for more sophisticated optimisation algorithms is evident.

For each of the methods we determined the following quantities: stress scenario, relative loss in the stress scenario, plausibility of the stress scenario. The notion of "plausiblity of scenario $\boldsymbol{r}_{stress}$" needs some explanation. The scenario $\boldsymbol{r}_{stress}$ together with the covariances Σ determines an n-dimensional ellipsoid of scenarios $\boldsymbol{r}$ satisfying

$$(\boldsymbol{r}_{MM} - \boldsymbol{r})^T \cdot \mathrm{S}^{-1} \cdot (\boldsymbol{r}_{MM} - \boldsymbol{r}) = (\boldsymbol{r}_{MM} - \boldsymbol{r}_{stress})^T \cdot \mathrm{S}^{-1} \cdot (\boldsymbol{r}_{MM} - \boldsymbol{r}_{stress}) \,.$$

This ellipsoid can be interpreted as the set of all scenarios which are equally probable as $\boldsymbol{r}_{stress}$ under the assumption that relative changes in risk factors are distributed normally with zero mean and covariances Σ. Now we take as "plausiblity of scenario $\boldsymbol{r}_{stress}$ " the mass of the normal distribution contained in this ellipsoid. It follows that the plausibility of $\boldsymbol{r}_{stress}$ can be interpreted as the probability that – under the assumption of normal distribution – a market situation is reached that is equally far away as $\boldsymbol{r}_{stress}$ from the current market situation, or further.

By adopting this measure of plausibility we do not want to imply that we accept the assumption of normal distribution of relative changes. In fact we do not. This is why we refrained from calling this number "probability of $\boldsymbol{r}_{stress}$ ". Still, as an intuitive measure of plausibility the number seems acceptable.

The question arises which volatilites and correlations among risk factors, as reflected in Σ, should be applied. For our purposes we simply use an equally weighted estimate of Σ with an observation period of one year. All market data are drawn from Primark/Datastream. The cutoff-date for our calculations is 1 February 2000. Of course, more refined methods for estimating Σ could be applied. Moreover, presumed stress correlations could be used instead of correlations prevailing at cutoff date (see Section 2.2 and Finger and Kim 2000).

The use of different covariance matrices would have yielded different portfolio specific worst case scenarios and different plausibilities of worst case scenarios for the respective methods. Nevertheless, as our focus here is the comparison of methods of identifying stress scenarios, we do not investigate the effect of choosing different covariance matrices.

Now the results of using the four methods to identify stress scenarios for Portfolios 1:

	Relative Loss	Plausibility	Change DJI	Change DAX	Change USD/EUR
DPG Scenarios	-7.4%	$8.9 \quad 10^{-28}$	-10%	-10%	0%
Black Friday	+2.7%	$1.8 \quad 10^{-113}$	-23%	-9.4%	-1.2%
Historical Simulation	-6.2%	$8.2 \quad 10^{-17}$	-8.0%	-7.7%	-0.32%
Monte Carlo Search	-6.5%	$3.5 \quad 10^{-10}$	-4.7%	-1.7%	-3.3%

Table 1: Results of identifying stress scenarios for Portfolio 1 with four different methods

Several aspects of this result are interesting: First, the naive "Black Friday" model of the 1987 equity crash is not only highly implausible, but it even leads to a profit. This scenario is useless as stress scenario for Portfolio 1. Methods 1 ("DPG Scenarios"), 3 ("Historical Simulation"), and 4 ("Monte Carlo") yield scenarios with losses of comparable size. But Monte Carlo search fares by far best because it leads to a scenario which is much more plausible than the others. Under the assumption of normal distribution the scenario produced by the Monte Carlo search is more than 20 million times more probable than the scenario produced by "Historical Simulation" and more than 2.10^{18} times more probable than the worst DPG scenario.

The three scenarios leading to losses for Portfolio 1 do not involve large moves of the DJI and the DAX. Although Black Friday involves much bigger moves it does not lead to great losses. This is due to the non-linear character of Portfolio 1 displayed in Figure 1.

And what about the results for Portfolio 2? For linear portfolios big losses must be due to big moves. Since the scenario identified by "Historical Simulation" is by construction at the edge of the search domain for "Monte Carlo", one would expect that for Portfolio 2 "Monte Carlo" leads to less severe scenarios than "Historical Simulation". This expectation is does not come true:

	Relative Loss	Plausibility	Change DJI	Change DAX	Change USD/EUR
DPG Scenarios	-10%	$8.9 \quad 10^{-28}$	-10%	-10%	0%
Black Friday	+2.1%	$1.8 \quad 10^{-113}$	-23%	-9.4%	-1.2%
Historical Simulation	-6.6%	$8.2 \quad 10^{-17}$	-8.0%	-7.7%	-0.32%
Monte Carlo Search	-8.1%	$1.4 \quad 10^{-16}$	-5.3%	-1.4%	-4.4%

Table 2: Results of identifying stress scenarios for Portfolio 2 with four different methods

First note that, again, the "Black Friday" scenario does not lead to a great loss, although it is highly implausible. The greatest loss is produced by the worst DPG scenario, and it is much more plausible than the Black Friday scenario. The losses produced by "Historical Simulation" and by "Monte Carlo Search" are slighty smaller than the loss in the worst DPG scenario. But both are roughly a trillion times more plausible than the DPG scenario.

The big surprise is that "Monte Carlo" produces a scenario which is more severe than the scenario produced by "Historical Simulation" – and also more plausible. This comes as a surprise given the fact that for linear portfolios bigger losses must be due to bigger moves. This counterintuitive effect can be explained in the following way. The covariance matrix Σ determines a family of concentric ellipsoids. The scenario produced by "Historical Simulation" – $\boldsymbol{r}_{HS}$ – picks one of these ellipsoids as search domain for "Monte Carlo". $\boldsymbol{r}_{HS}$ lies on the surface of the search domain. Since the portfolio value is a linear function of the risk factors it will take its minimum value over the filled ellipsoid on its surface. (This fact is at the root of the saying that for linear portfolios big losses must be due to big moves.) But it would be a formidable coincidence if the minimum were taken exactly at $\boldsymbol{r}_{HS}$. This coincidence did not happen in the case of Portfolio 2. And close to the minimum there are points, which are in the inner of the ellipsoid (and thus more plausible than $\boldsymbol{r}_{HS}$) and at which the portfolio value is smaller than in $\boldsymbol{r}_{HS}$. One such point was found by the Monte Carlo search.

By comparing the stress scenarios produced by Monte Carlo Search for Portfolios 1 and 2 we observe that the results of– which is the stress scenario and what is the loss in the stress scenario? - crucially depend on the portfolio.[1] One the one hand this confirms the general

[1] "Historical Simulation" is also a portfolio-specific method of identifying stress scenarios. It is mere chance that the

considerations in Sections 1 and 2: banks have to take account of their portfolio when identifying stress scenarios. Scenarios which qualify as stress scenarios for one bank might be lucky strikes for another. On the other hand portfolio specificity raises the question how reliable the conclusions are which we drew about the merits of "Monte Carlo Search". Certainly there are portfolios for which Monte Carlo Search is even farther ahead of the other methods. Surely there will be portfolios for which Monte Carlo Search is not far enoug ahead to justify the computational efforts. More efficient optimisation algorithms in high dimensional spaces are needed. But given any stress scenario identified with any method, optimisation algorithms can find a scenario which is at least as plausible and as severe as the given scenario.

References

Basle Committee on Banking Supervision (1995): *An Internal Model-Based Approach to Market Risk Capital Requirements*, proposal for consultation.

Basle Committee on Banking Supervision (1996): *Amendment to the Capital Accord to Incorporate Market Risks.*

Berkowitz J. (1999): A coherent framework for stress-testing. *Finance and Economic Discussion Series* 1999-29, Board of Governors of the Federal Reserve Board.

CAD: *Council Directive 93/6/EEC of 15 March 1993 on the capital adequacy of investment firms and credit institutions*, Official Journal of the European Communities No. L 141 of 11 June 1993, pp. 1-26.

CAD II: *Directive 98/31/EC of the European Parliament and of the Council of 22 June 1998 amending Council Directive 93/6/EEC on the capital adequacy of investment firms and credit institutions*, Official Journal of the European Communities No. L 204 of 21 July 1998, pp. 13-25.

Derivatives Policy Group (1995): *Framework for Voluntary Oversight*. To be accessed at //risk.ifci.ch/137790.htm

Finger. C. C., Kim J. (2000): A Stress Test to Incorporate Correation Breakdown, Working Paper No. 99-08, The RiskMetrics Group. To be accessed at http://www.riskmetrics.com.

worst historic moves in the equity crash 1987 coincide for Portfoilios 1 and 2.

Kupiec, P. H. (1998): Stress Testing in a Value at Risk Framework, *Journal of Derivatives*, Fall 1998, pp. 7-24.

Niederreiter H. (1992): *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia PA.

Press W. H., Teukolsky S. A., Vetterling W. T. and Flannery B. P. (1992): *Numerical Recipes in C: The Art of Scientific Computing,* 2nd edition, Cambridge University Press, Cambridge, UK.

Shaw J. (1997): Beyond VAR and Stress Testing. In: *VAR – Understanding and Applying Value-at-Risk*, KPMG/Risk Publications, pp. 221-224.