ESSAYS ON FINANCIAL OPTIMIZATION

by

Anders Buvik

A dissertation submitted for the degree of dr.oecon.

Acknowledgements

Contents

1. Introduction.

1.1. The dissertation.

This dissertation contains work from the field of financial optimization. As the name suggests, financial optimization is optimization or operations research applied to financial problems. In general, optimization is applicable to many different problems in science, ranging from physics to management science, but the perspective adopted in this dissertation is optimization in finance.

The history of optimization is long, but several milestones may be mentioned, for example the development of differentation 300 years ago, the work on constrained optimization by Lagrange, the development of linear programming by Dantzig and von Neuman and the work on discrete programming. In the field of finance, a natural starting point would be the work of Markowitz.

The applicability of optimization, expecially for large programs, has always been dependent on existing computing possiblities. As we know, development in computer hardware and software has accelerated in the last few decades and today compitutional possibilities exist that were not available 20-30 years ago. Thus, the possibilities of optimization have increased rapidly.

In chapter 2 of this dissertation, we have a look at the mean absolute deviation model and first perform a traditional sensitivity analysis. The sensitivity analysis is performed by using actual data from the Oslo Stock Exchange. The parameters under investigation are the estimated expected return for each asset. The results from the traditional analysis are compared to a traditional analysis on the mean variance model. Also in chapter 2, we perform a simultaneous sensitivity analysis on the expected return estimates in the mean absolute deviation using the approach suggested by Wendell (1992) and by Wodolowski (1991). In Chapter 3 we will study portfolio optimization problems with the use of discrete decision variables. We first study the problem in general, and then we procede to look at two specific problems that could arise in practice. It is shown that even if the problem is only slightly different from the traditional portfolio problem, the algorithms differ considerably. In Chapter 4 we consider portfolio optimization using three dimensions. The dimensions are mean, variance and the third central moment. As the resulting problem is nonconvex a nonconvex solution procedure must be utilized. The procedure adopted utilized special ordered sets of integer variables. This procedure is tested using actual data from the Oslo Stock Exchange. In Chapter 5, we present a multiperiodic model that

considers optimal decisions with periodic penalty for nonfullfillment of certain requirements. These requirements are related to capital and return requirements often faced by insurance companies. The model is illustrated by constructiong an example and then solving the problem by using a progressive hedging algorithm presented by Wets and Rockafellar (1991). In Chapter 6, which is written together with Kurt Jørnsten, Ronnie Aboudi and Rainer Leisten, includes consideration of an alternative risk measure. The risk measure is the mean Gini difference. In this chapter we present a solution procedure for this portfolio problem utilizing an iterative aggregation and reclustering algorithm. The algorithm is tested by using actual data from the CRSP (University of Chicago Center for Research on Security Prices) database. Chapter 7, also written together with Aboudi, Jørnsten and Leisten, presents a comparison of the mean variance, mean absolute deviation and mean Gini model using data from the Oslo Stock Exchange. Finally, Chapter 8 presents a fast solution algorithm based upon least element theory for a type of complementarity problems that arises in the field of American option pricing. The remainder of this chapter provides a brief introduction to the field of optimization.

Although each of the chapters in this dissertation is meant to be self–contained, readers may benefit from reading the papers refered to at the end of each chapter. The symbols used in the equations could change from chapter to chapter.

1.2. Models in financial optimization.

We will now take a closer look at the various models used in financial optimization. One problem that arises in discussing different models in this field is the set of criteria that should be used to distinguish between the models that can be said to belong to this class and the models that do not. In my opinion, there is no clear distinction between financial optimization models and other optimization models, for example those used in accounting. The reader may be of another opinion as regards the classification of the models. In the remainder of this chapter, I will consider the different models as classified into different types of programming.

First we shall consider linear programs, and then procede to a survey through the class of quadratic programs. We will then take a look at integer programming, in the sense of mixed integer linear programming. Finally we will look at stochastic programming. Stochastic integer programming will not be discussed in this dissertation.

In each section, we will take a look at the general formulation of such problems, discuss

solution methodologies for the problems and review some of the literature in financial optimization that uses the optimization model in question. We will also present a simple model from the literature in each of the programming categories.

1.3. Linear programming.

A linear program is an optimization problem in which the aim is to maximize or minimize a linear objective function subject to linear constraints. In general, the optimization problem is of the type[1]:

[1a]   minimize $C^T X$

subject to:

[1b]   AX=B
[1c]   X≥0

Here X is the variables vector, C is a cost vector, A is a matrix representing the linear constraints in the model and B is the resource constraints. We have, in the formulation above, included slack variables in the problem, such that when the restriction in a linear program is an inequality, the general formulation above allows for this by adding a slack variable such that the equality holds.

Linear programs have been extensively studied in the literature, and the theory about this class of optimization problems was developed by Dantzig and Von Neuman after the second world war. Included in this work was the development of a solution algorithm. This algorithm, the simplex algorithm, yields a relatively efficient way of solving the class of linear optimization problems. The algorithm seeks the optimal solution of the problem [1] by moving on the surface of the convex region defined by the restrictions in [1b] and [1c] and moving towards a optimal basis, by the movement between different bases, until the minimum of [1a] is reached. In the solution to the linear problem above, the user also obtains a set of dual prices, which give the user more information on the optimal solution. For further information about the simplex method, see Dantzig (1963), Hadley (1971) or Gass (1975), among others.

There is also a class of solution methods for linear programs which take into account the duality theorem (given below) and perform the simplex procedure on this formulation.

Such solution techniques are called primal–dual and dual algorithms.

Another interesting approach to solving linear programming problems is given by Karmarkar's interior point algorithm, Karmarkar (1984). This algorithm performs a search through the interior of the set defined by the constraints in [1], as opposed to searching through the surface of the set given by the simplex method. The method starts at an initial point X^0, and makes a projection towards the optimal solution. By imposing penalties on moves towards the surface of the set, the projection is guaranteed to be in the interior of the set defined by the restrictions. The algorithm seems to be more efficient than the older, though still extensively used, simplex method. It also seems to avoid some undesirable numerical properties that occur when using the simplex method in large scale linear programming.

In addition to the two algorithms described above, there are others, for more information on these see Murty (1983).

Dantzig and Von Neuman's work on linear programming is the first investigation of the problem in what we today call linear programming. The history of the mathematics on related topics includes Gauss's work on linear equations in the 1820's. Gauss discovered that linear equations could be solved by what we today call Gaussian elimination. Sets of large scale linear equations have never been trivial to solve. The recent introduction of computers able to handle such large systems makes solving these problems easier.

Leontief's input–output models were a step towards an application of linear programming. Leontif's model consisted of 44 variables and equations, but at that time the solution to such a system required enormous calculations, and due to the lack of computational equipment in Russia, (or anywhere else), the problem was not solvable. Leontief's aggregated version with 10 variables and equations was more within reach, and at that time this was the largest solvable system. Another Russian, Kantorovich, solved some industrial type optimization problems roughly, by the using dual variables.

The Karush–Kuhn–Tucker conditions, derived in the late thirties and in the early fifties, developed the field of optimization further, including linear programming as a subclass, by stating optimality conditions for optimization problems. These conditions are essential in the study of mathematical programming.

Together with each linear program on the form above, there is a related problem, the dual. The dual of the general problem above is given by[2]:

[2a]   maximize $U^T B$

subject to:

[2b] $A^T U \leq C$
[2c] $U \geq 0$

The duality theorem reads as follows:

Theorem:

If [1] and [2] have solutions, then max $U^T B$=min $C^T X$. If either of them has an unbounded solution, the other has no solution.

This means that the optimal objective function value of either of the problems, if it exists, is equal to the optimal solution of the other. Another interesting aspect of the duality theory is that the solution vector of the dual, given by U^*, is the dual prices for the constraints in the primal problem and vice versa.

In this dissertation, we will study an optimization model that utilizes aggregation techniques. Below we will give a short introduction to aggregation in linear programming.

Aggregation is a procedure that groups the variables in a large program into fewer aggregate variables. The main advantage of the method is that the problem becomes smaller and easier to solve. In brief, aggregation consists of three steps. The first step is to determine clusters which consist of one or more of the original variables. The allocation of the variables into clusters could be based upon different criteria. Variables that have similar characteristics are typical candidates for the same cluster. Linear program variables with approximately the same objective function coefficient and coefficient in the restrictions would be candidates for the same cluster. The number of clusters to be be created depends upon the users tradeoff between compitutional burden (few clusters) and accuracy (many clusters).

After we have grouped the variables in the original problem into clusters, we have to assign a weight to each of the variables in the cluster. These weights, which add up to unity for each cluster, could be constructed on the basis of several criteria. One approach uses their relative importance in the objective function as the criterion. Another naive method is to give the variables equal weight. One could also select the most representative of the variables in the cluster and give this variable weight one.

After we have solved the aggregated problem, we have to disaggregate the obtained solution. The disaggregation could be done by using the weights used in the aggregation procedure, but other promising ways to do this have also been discussed. For more on this problem, see Rogers (1986) and Zipkin (1977).

When we use aggregation techniques, we obtain an easier model to solve, at the cost of accuracy. Thus, it is important to quantify the errors obtained by solving an aggregated version of the model as compared to the original model. We therefore need knowledge on error bounds in aggregation procedures. These bounds could be divided into two groups, a priori bounds and a posterori bounds. The former are computed after we have aggregated the problem, but before we have solved it. The latter are computed after we have solved the aggregated problem. A priori bounds are typically larger than a posterori bounds. For further information on aggregation, see the excelent survey by Rogers et al. (1991).

In the rest of this section, we will look at some practical implementations of linear programming in financial optimization, and at some of the different models presented in the literature.

Linear programming, as with all optimization, has many possible applications. Many of the applications appear in the literature, but there are also many that are held in-house in different commercial firms, ranging from small banks using an LP model to match input output payments, to larger models used by large financial institutions such as Goldman Sachs and others. Many of the models, have characteristics in common, so the search for the original model, and the original scientist behind it, may be a difficult task.

Below we shall consider a bond investment model currently used by Salomon Brothers and Bear Stearns, among others, in advising institutional investors how on how to take advantage of subjective predictions of yields (a measure of effective interest rate on goverment bonds). The objective function [3a] maximizes the excess return above a predefined index for a portfolio in a base case scenario given by the user. In adition to this, the user provides a set of yield curve movements he believes is probable, but not dominant. The program ensures that the optimal portfolio gives better return than the index if the scenarios are properly selected. This is handled by the constraints [3b].

The linear programming problem is[3]:

$$[3a] \quad \text{maximize} \quad \sum_{i=1}^{n} R_i^s x_i - R_I^s$$

subject to:

$$[3b] \quad \sum_{i=1}^{n} R_i^j x_i - R_I^j \geq 0 \quad j \in 1,2,...,T.$$

$$[3c] \quad \sum_{i=1}^{n} x_i = 1.$$

$$[3d] \quad x_i \geq 0 \quad i \in 1,2,.....,n.$$

The variables and symbols are:

T = Number of scenarios (not including the base case used in the objective function).

n = Total number of bonds (union of bonds in index and in investors universe).

x_i = Fraction invested in bond i.

R_i^j = Total return measured by change in market value and accrued interest on bond i in yield curve scenario j on a percentage basis (until a predefined date).

R_I^j = Total return, measured as above, on a predefined index in yield curve scenario j.

R_i^s = Total return, as measured above, on bond i in yield curve basis scenario on a percentage basis.

R_I^s = Total return, as measured above, on a predefined index in the yield curve basis scenario.

The model is used in practice to find a portfolio that is tilted towards a base case scenario but at the same time give returns above the index in the other scenarios. If the number of scenarios is large, the optimal portfolio converges to the index used in the analysis. The program, if used by a bond brokerage firm on the basis of input from an investor, also gives the broker first order information regarding the customers`s needs. From a financial economics perspective, the model is poor, since the state space specified by the user must permit arbitrage in order to obtain a objectve function value $\neq$ 0.

In general, there are several models in the literature, which are characterized by an application of linear programming to financial modeling. Sharpe (1967), assuming a market model for generating returns, formulates a linear program for mutual fund portfolio selection. Sharpe (1971a), uses linear programming to estimate beta in the CAPM model. The LP model minimizes the absolute deviation of the regression errors, instead of the standard deviation of the errors, thus resulting in a linear programming model. Sharpe (1971b) presents a portfolio model, equal to the original Markowitz mean variance model, but he makes an linear approximation of the objective function, giving an linear programming problem instead of a quadratic one.

Another portfolio optimization model that is formulated by the use of linear programming, is the mean absolute deviation model. The first version of this model was presented by Hazzel (1971) in the agricultural economics field. He formulated a portfolio model based upon the risk measure absolute deviation in decision under uncertainty in farm planning. Konno and Yamazaki (1991) reinvented this model.

Linear programs for portfolio selection are often characterized by the use of linear or piece-wice linear risk measures. Stone (1973a), (1973b) was among the first to study this and in his articles he formulated linear programs that use linear or piece-wice linear risk measures. Ang (1975) formulated a linear programming problem which was meant to be an approximation of the traditional MV problem, by using a linear risk measure that penalizes deviations below the mean. As if we look at this problem today, the Ang portfolio problem is analogous to the mean absolute deviation model. The linear programming problems based upon these linear risk measures, whether they are deviations below the expected return of the portfolio or return below some predefined constant, fall into the category of risk measures in which financial theory is developed by Bawa et al. (1975), (1977), (1978) and Harlow and Rao (1989).

1.4. Quadratic programming.

Quadratic optimization is among the most studied nonlinear optimization problems. As its name suggests, a quadratic optimization problem involves an objective function which is a quadratic form. A general quadratic optimization problem with linear constraints is written[4]:

[4a]   minimize $f(X) = \frac{1}{2}X^TQX + CX$

subject to:

[4b]   AX ≥ B

[4c]   X ≥ 0

where X is a vector consisting of the n×1 variables in the problem, Q is a n×n matrix, C is a vector with n elements and A is an n×m matrix.

The quadratic program above has a continuous homogeneous function of degree 2 as the objective and a closed set as the feasible region. The Bolzano-Weierstrauss theorem assures the existence of a global minimum or maximum if the feasible region is bounded and compact. The Frank-Wolfe theorem gives us a neccesary and sufficient condition if the feasible region is bounded.

The Karush-Kuhn-Tucker conditions give us the optimality conditions for the quadratic programming problem [4].

If there exists a vector U, such that[5]:

[5a] $C + QX - A^TU = 0$

[5b] $U \geq 0$

[5c] $U^T(AX - b) = 0$

then X is a stationary point for the quadratic program [4].

The conditions are neccesary, but not sufficient for a global optimum for the quadratic programming problem.

There are many classifications of the quadratic problem above. Many of the classifications deal with the various forms of Q. One classification of the matrix Q is that Q belongs to the class of positive definite matrices. This means that the quadratic form:

[6] X^TQX

is positive for any selection of the vector X. There are several ways to investigate this property. One way to do this is to compute the eigenvalues of the matrix Q and investigate whether all of these eigenvalues are positive. If all these eigenvalues are strictly positive, the matrix is positive definite. To find the n eigenvalues of the matrix Q we have to compute the n eigenvalues λ, given by the solution to:

[7] $\det(Q - \lambda I) = 0$

The conditions in [5] are necessary but not sufficient conditions for a global optimum. These conditions are only sufficient when the quadratic form is convex. In this case this means that Q is positive semidefinite and in that case all the eigenvalues of Q, given by solving [7] are greater than or equal to zero. In quadratic programs in financial optimization, the matrix Q is often a covariance matrix, or strongly related to a covariance matrix. Since this matrix is positive definite, conditions for global optimum are satisfied by the Karush-Kuhn-Tucker conditions.

Quadratic programming is used extensively in financial optimization. One reason for this is the central role covariance structures play in financial theory. Listing all the applications of quadratic programming in this dissertation is too ambitious. But we will in this chapter look at an example where quadratic programming is used to match an index. Following this example, we will have a look at the mainstream applications of quadratic programming in the financial literature.

Indexation or benchmark tracking has in the last two decades become one of the most popular asset allocation techniques. Indexation is defined as constructing a portfolio that has the same risk and return characteristics as a predefined benchmark. Thus, indexation does not imply outperforming any predefined index, but rather matching one. Indexation means index matching in any asset category, not necessarily stocks.

One popular rating concept is to evaluate fund managers and their managed portfolios on how they performed relative to a predefined index. Using indexation means to make the portfolio's return and behavior as equal as possible to this index.

There are several reasons for a fund manager to match an index. One reason is that it gives the fund manager a possibility to expose himself to nationwide factors without buying the whole composite index. For instance in the stock market, linking a portfolio to the Wilshire 5000 index, a broad index for the US equity market, means that one has to buy 5000 different shares. Indexation separates market selection skills from stock selection skills and it gives the fund manager the possibility to avoid non−market risk when there is a penalty involved for portfolio returns less than the index return for some predefined period.

The most popular tracking efficiency measure is the standard deviation of the residual between the benchmark and the indexed portfolio. We denote this tracking error measure by $\sigma_{\varepsilon I}$. [8]:

[8a] $\sigma_{\varepsilon I} = \sqrt{E\{(\tilde{r}_p - \tilde{r}_I)^2\}}$

where:

$\tilde{r}_p$ = The return of the portfolio.
$\tilde{r}_I$ = The return of the index.

The estimator for $\sigma_{\varepsilon I}$ is given by:

[8b] $\hat{\sigma}_{\varepsilon I} = \sqrt{\dfrac{\displaystyle\sum_{t=1}^{T}(r_{pt} - r_{It})^2}{T - 1}}$, T = Number of observations.

Above, r_{pt} and r_{It} are portfolio and index returns in period t respectively. To form a tracking portfolio by the use of a subset of n equities we could solve the following quadratic programming problem[9]:

[9a] minimize $\dfrac{1}{T-1}\displaystyle\sum_{t=1}^{T} y_t^2$

subject to:

[9b] $\displaystyle\sum_{i=1}^{n} x_i r_{it} - r_{It} = y_t$ $t \in 1,2,...,T.$

[9c] $\displaystyle\sum_{i=1}^{n} x_i = 1$

[9d] $x_i \geq 0$ $i \in 1,2,...,n.$

The problem above, where r_{it} and y_t are asset i`s return and residual return in period t respectively, minimizes the estimated tracking error given by [8b] for a sample of T observations.

After the realization of the MV model in the 1950s, empirical tests of the model were not

easy to carry out until the computer software and hardware were developed with the required speed and size to solve such problems. In the `60s and `70s, the solution of quadratic programming problems of some size were not trivial, and empirical testing of the model may have suffered because of this. Due to the revolution in the micro computer market in the early 80´s until now, the possibility of performing a MV analysis is considerably enlarged. Today there is readily accessible software for MV analysis.

Empirical testing of the MV model was done by Farrar (1962). In his sample, he used mutual funds as the assets. Using monthly observations and an arbitrary one-year investment horizon, he concluded that mutual funds lie close to the efficient frontier on a ex post basis.

One disadvantage of the mean variance model is the sensitivity of the optimal solution to the change in the input data in the model. In their article, Grauer and Best (1984) point out this fact. The optimal solution is especially sensitive to changes in the expected returns of the assets. With this in mind, practioners often claim that the model lacks investment value, and consequently reject the approach. Michaud (1989) termed this effect the "Markowitz OptimizationEnigma". As the optimal solution is affected by errors in the expected returns of the assets, researchers have turned their attention to portfolios on the efficient frontier that are not heavily influenced by errors in the estimated expected returns. One of these portfolios is the minimum variance portfolio and Robert Haugen, University of California (Irvine) in a set of lectures, claims that picking this portfolio, under some set of restrictions varying from application to application, gives a portfolio that performs better than for example a capital weightened index. The list of advantages and disadvantages of the MV model is long and the debate will probably continue further as better estimation technique develop.

Another use of the MV model is as an ex post evaluation of different portfolios. By plotting different portfolios into the MV diagram, one could see their position relative to the efficient frontier ex post. This approach was used by Jennergren and Sørensen (1986) in their study of Danish mutual funds and in the study of the French stock market by Chatry, Jennergren and Szala (1990).

Sharpe (1963) presented a quadratic optimization problem for portfolio selection using the single index model. Technically, the optimization model in Sharpe´s article turned out to be a minimization of diagonal quadratic form, which, in the light of the computational resources at that time, was very convenient. The model's strongest assumption was that of a single factor or index that systematically influenced the returns of the single assets.

Alternative risk measures, such as the semivariance, also require help from quadratic programming. Markowitz (1959) suggests this risk criterion, since one could argue that investors care more about deviations below the expected return than about the dispersion around the mean. Harlow (1991) also considers this risk measure and performs an empirical investigation of it using historical data. An optimization algorithm for the problem is discussed in Markowitz, Todd, Xu and Yamane (1993).

1.5. Models involving discrete variables.

So far, we have concentrated on models in which the variables used are continuous. Such models are, in the convex case, typically easy to solve and many solution algorithms exist handling their relatively simple nature. In practice, situations often arise which require the introduction of discrete decision variables. This means that we must consider a model with such variables. In this dissertation we will concentrate on the set of discrete variables allowed to take any value among the nonnegative integer numbers. Thus the variables take values in the range:

[10] $\{0,1,2,3,....\}$

There are several situations in which integer variables need to be introduced. If we build a model that must consider logical constraints of the type "no more than four assets are allowed to be a member of the optimal portfolio", we could simply solve this by introducing integer variables.

It is possible to distinguish between linear programming extended with integer variables and nonlinear problems extended with integer variables. For the former, algorithms are extensively studied in the literature and are easily available. For the latter, the amount of research is somewhat more limited and the research also more complex. The papers presented in this dissertation are based upon mixed integer linear programming.

Consider the extended linear programming problem[11]:

[11a] minimize $C^T X$

Subject to:

[11b] $AX = B$
[11c] $X \geq 0$

[11d] $X \subset S$
[11e] $X^Z \subset S$
[11f] $X^Z \subset \{0,1,2,3,....\}$

Above S is a st and X is a vector. The problem above differs from the traditional linear programming problem because we have defined a subset, X^Z, of the solution vector to be a member of the class of nonnegative integer numbers. The problem above is a mixed integer linear programming problem.

As an example of such a model, we could consider the linear model [3] extended with the following constraints[12]:

[12a] $\sum_{i=1}^{n} \delta_i \leq k$

[12b] $\delta_i - x_i \geq 0$
[12c] $\delta_i \in \{0,1\}$

The extension of the model is done by the introduction of n binary variables, δ_i. The new restrictions ensure that the optimal solution contains no more than k assets with a positive weight.

Integer linear programming was investigated in an article by Gomory (1958), where a method called cutting planes was investigated. The key idea behind this method is first to solve the integer programming problem as an ordinary linear programming problem, thus ignoring the integer constraints imposed on the problem. If one or more of the integer variables are fractional, we restrict one of these variables to be less than or equal to the nearest lower integer value. Suppose one of the integer variables in a program turned out with a value of 16.8. We then add the restriction that the value of the variable should be less than or equal to 16 and solve the problem again. For each variable this procedure is

repeated until all the integer constraints are fulfilled. This procedure is often referred to as Gomory cuts. Even if the method is simple, it is unstructured and may exhibit slow convergence. Algorithms based upon Gomory cuts that are more structured are discussed in Gomory (1958), (1963), Young(1965), (1968) and Glover (1965).

The branch and bound algorithm, based upon the work of Land and Doig (1960), Bertier and Roy (1964) among others, reduces the integer programming problem to a problem involving binary variables only. These variables are only allowed to take the value 0 or 1. The conversion of integer variables to binary variables could be done by replacing the integer variable x_j by:

$$[13] \qquad x_j = \sum_{l=1}^{\rho_j} 2^l z_j^l \qquad (\rho_j \text{ sufficient large and } z_j^l \text{ binary}).$$

The identity above states that each integer variable could be expressed as a sum of a product of binary numbers z_j^l and 2^l. One example is when x_j is the integer 7. In that case $z_j^0 = z_j^1 = z_j^2 = 1$ and all the other binary variables are set to zero. By imposing the binary constraint on the new variables (z_j^l) and adding restrictions of the type [13], we have converted the mixed linear integer program to a mixed binary linear program.

An obvious, but very time–consuming, way of solving such a problem is to consider all possible combinations of the binary variables and solve all these problems. After that we select the solution with the lowest objective function value.

With m integer variables the number of combinations is:

$$[14] \qquad \phi = 2^{\sum_{j=1}^{m} \rho_j}$$

The number of combinations could be a very large number, and the time used to solve the problem this way is typically large. It is not surprising that there exist other, more efficient ways of solving problems with discrete decision variables. In the following, a brief description of one of these will be provided.

The branch and bound method partions the binary variables in the converted binary linear program into different levels. Each level fixes one of the binary variables to either one or zero, and at the bottom level the program is fully partioned with nodes consisting of each possible combination of the binary variables. The partitioning can be visualized by a tree, where at the top node no binary variable is fixed, at level one, one variable is fixed and at

level p, there are p variables fixed. The deepest level in the tree is level ϕ, where all the binary variables are fixed. Fortunately, when we search for an optimal solution to the problem, we do not have to visit all the nodes in the tree. Suppose that we have a feasible, but not optimal solution to the problem. The corresponding objective function value of this non optimal solution could be used as a upper bound to the optimal solution. So, if we fix the binary variables, by adding equality restrictions to the linear program, such that there is a correspondence between the values of the binary variables in one particular node, we know that searching in nodes below this node is not necessary if the optimal solution to the linear program is greater than the upper bound. If the optimal solution is less than the upper bound, a further search down the tree is neccecary until a valid and new upper bound is discovered. The process is stopped until the whole tree is investigated this way. The procedure increases the speed of solving the mixed integer linear programming problem.There are several issues addressing the procedure for searching the different nodes discussed in the literature, and the reader is advised to consult Minoux (1986) for a description of and references to these procedures.

In the financial literature models involving discrete decision variables are not discussed as much as linear and quadratic programming. Nevertheless, a brief list of some of them is presented below.

A bond portfolio model that incorporates transaction costs by the use of integer programming is discussed by Nauss in Zenios (1993). This model uses binary variables to restrict the minimum or maximum amount allowed sold or purchased in a particular security. A relatively complex integer program utilizing binary variables is a model for structuring Collateralized Mortage Obligations (CMO) discussed by Dahl, Meeraus and Zenios in Zenios (1993). In this model, the binary variables are used to control the amortization of the different tranches.

1.6. Stochastic programming.

Traditional mathematical models assume that the coefficients in the objective function and in the constraints are known with certainty. These coefficients are often based upon estimation, and thus exposed to estimation error. So, if we work with a mathematical model, parameters that are assumed constant may be stochastic. As financial optimization could use parameters that are asumed to be stochastic we need optimization procedures that handle stochastic parameters. This leads us to the area of stochastic programming. A stochastic programming problem has the general form [15]:

[15a]   minimize $E\{f_0(X,\xi)\}$

subject to:

[15b] $E\{f_i(X,\xi)\} \leq 0$ $i \in 1,2,...,m$.

Where:

ξ = Random vector with support $\Xi \in \mathfrak{R}^n$ and probability distribution function P.

f_0 = Function in which expected value have to be minimized.

f_i = Function in which expected value is restricted.

X = decision vector.

$E\{\bullet\}$ = Expectation operator.

In principle, the expected value of a function with several stochastic parameters could be evaluated using multidimensional integration or summation. For typical problems, where the number of stochastic variables is not too few, this is a difficult task and robust and time–consuming algorithms are necessary. In order to handle stochastic mathematical optimization models, one needs knowledge from the theory of expectations functionals and from convex analysis.

In the literature stochastic models are divided into two categories, anticipative and adaptive models. The anticipative models do not depend upon the future realization of the stochastic parameters. The objective function or constraint in such a model can be probabilistic, such as the type:

[16] $P(f_i(X,\xi) \leq 0) = 0.05$

or involving the type:

[17] $E\{f_i(X,\xi)\} + \beta\sqrt{Var(f_i(X,\xi))} \leq 0$, β = constant.

Examples of such models are asset allocation models with shortfall risk (chance) constraint.

In the other category, adaptive models, the uncertainty in the model is time–dependent, such that some of the decisions can be made after realization of some or all the stochastic variables (multi stage problems). Uncertainty is modeled by the use of an information structure. These models are often of a type where the decision maker makes a decision,

waits for a realization of some or all the parameters in the model and takes corrective action on the basis of this new information. An example of an adaptive model from the financial literature is the class of multiperiodic financial models.

Mathematical models with probabilistic constraints were discussed in the article by Charnes and Cooper (1959). The stochastic program with recourse was discussed as early as in Dantzig (1955). The categorization of stochastic programs can be done in the same way as we characterized the traditional deterministic programs, as linear, quadratic, nonlinear and integer stochastic programs. Another way of describing the models is by means of numbers of stages, or number of batches with information realizations in the model.

As an example of a stochastic programming problem, we consider the following model[18].

[18a] $\quad$ maximize $E\{\tilde{V}\} = \displaystyle\sum_{s=1}^{S} \pi_s V_s$

subject to:

[18b] $\quad \displaystyle\sum_{i=1}^{n} x_i v_{is} = V_s \quad S \in 1,2,...,m.$

[18c] $\quad \displaystyle\sum_{i=1}^{n} x_i v_i^0 = V_0$

[18d] $\quad P(\tilde{V} \le V_0) \le 0.05$

[18e] $\quad x_i \ge 0 \quad i \in 1,2,...,n.$

Where:

$\tilde{V}$ = Value of the portfolio.

V_s = Value of the portfolio in state s.

V_0 = Initial value of portfolio.

v_i^0 = Value of asset i at time 0.

v_{is} = Value of asset i in state s.

π_s = Probability of state s.

x_i = Fraction held of asset i.

S = Number of states.

The one period model above maximizes the expected value of a portfolio given that the probability of zero return on the portfolio should be less than or equal to 0.05.

Other stochastic models from the literature include a dynamic bank portfolio management model by Bradley and Crane (1972) and the portfolio management model by Mulvey and Vladimirou (1988). Zenios (1991) presented a stochastic model applicable for mortgage-backed securities financing. There is also a stochastic bond dedication model which was presented by Hiller and Schaack (1990).

As stochastic programming allows the specification of stochastic parameters, its use in financial optimization is well founded. Thus it may be expected that such models will be applied more frequently.

For some of the stochastic programs, such as linear stochastic programs, it is possible to formulate other and much larger equivalent deterministic problems in which the traditional solution techniques are applicable. These problems, as in the case with linear stochastic problems, utilize large scale linear programming techniques such as Benders decomposition. For further reference to such methods, see Minoux (1986). Wets and Rockafellar (1991) presented an approach to be used on relatively general types of stochastic programming problems where the stochastic environment is specified with a discrete probability function. The method, to be described and utilized later in this dissertation, is advantageous in that it keeps the problem in its original and simple form and is easy to implement.

1.7 The data from the Amadeus Database.

The dissertation uses data from the Oslo Stock Market. The dataset is total returns on a monthly basis from at most 21 stocks which was selected on the basis of liquidity and the requirement that the stock was listed during the whole period under investigation. The dataset was provided by the Norwegian School of economics and business administrations database Amadeus. This database performs the standard corrections to asset returns as adjustments for dividend payout, splits and emmisions. Under is a description of the use of the dataset:

Chapter	Period	#obs	#Assets	Freq.	Comments
2	1/87–12/92	72	21	M	Aritm. mean.
3	1/87–12/92	72	21	M	Aritm. mean.
3	1/93 – 12/94	24	21	M	Backtesting
4	1/88 – 12/89	24	10	M	Comp. intensive

Table 1.1. The OSE data set.

1.8. Conclusions.

In this introduction we have presented the dissertation and briefly had a look at different optimization methods, their applications in the literature and some of the solution algorithms of these problems. We will now turn our attention to the main part of the dissertation, and present the main body of research done by the author.

References:

Ang, J. (1975) "A Note on the E, SL Portfolio Selection Model." Journal of Financial and Quantitative Analysis Des pp. 849-857.

Bawa, V.S. (1975) "Optimal Rules of Ordering Uncertain Projects." Journal of Financial Economics. No. 1. pp. 95-121.

Bawa, V.S., Lindeberg, E.B. (1977) "Capital Market Equilibrium in A Mean-Lower Partial Moment Framework." Journal of Financial Economics. vol 5. pp. 189-200.

Bawa, V.S. (1978) "Safety First, Stochastic Dominance, and Optimal Portfolio Choice." Journal of Financial and Quantitative Analysis, June pp. 255-271.

Bertier, P., Roy, B. (1964) Une Prodédure de Résolution pour une Classe de Problémes Pouvant avoir un Caractére Combinatoire." Cahiers du Centre D´ Études de Recherche Opérationnelle, vol. 6, pp. 202-08.

Bradley S.P., Crane, D.B. (1972) "A Dynamic Model for Bond Portfolio Management." Management Science, 19(2) pp. 139-51.

Bradley, S. P., Hax, A. C., Magnati, T.L. (1977) Applied Mathematical Programming. Reading, MA. Addison Wesley.

Broadie, M. (1993) "Computing Efficient Frontiers using Estimated Parameters."

Annals of Operational Research Vol 45. Nos 1-4. pp. 21-58.

Charnes,A., Cooper, W.,W.(1959) "Chance Constrained Programming." Management Science 5, pp. 73-79.

Chatry, P., Jennergren, L.P., Szala, A. (1990) "Two Investigations of Portfolio Selection and Mean − Standard Deviation Performance Evaluation in the French Capital Market: A Research Note." EFI − Stockholm School of Economics.

Dantzig, G.B., (1955) "Linear Programming under Uncertainty." Management Science Vol 1 pp. 197-206

Dantzig, G.B. (1963) Linear Programming and Extensions. Princeton University Press.

Farrar, D.E. (1962) The Investment Decision Under Uncertainty. Prentice Hall, Englewood Cliffs, NJ.

Gal, T. (1979) Postoptimal Analyses, Parametric Programming, and Related Topics. McGraw Hill, New York.

Gass, S. I. (1975) Linear Programming, Methods and Applications. McGraw Hill Ltd.

Glover, F. (1965) "A Multiphase-Dual Algorithm for the Zero-One Integer Programming Problem." Operations Research, 13 pp. 879-919.

Gomory, R. E. (1958) "Outline of an Algorithm for Integer Solutions to Linear Programs." Bulletin Am. Math, Soc. 64 pp. 275-78.

Gomory, R.E. (1963) "An Algorithm for Integer Solutions to Linear Programs." In: Recent Advances in Mathematical Programming. Eds. R.L. Graves and P. Wolfe, pp. 193-302.

Grauer,R.R., Best, M.J. (1984) "Sensitivity Analysis for Mean Variance Portfolio Problems." Working Paper. Burnaby, B.C.: Department of Economics, Simon Fraser University.

Hadley, G. (1971) Linear Programming. Addison Wesley Publishing Company.

Harlow, W.V. (1991) "Asset Allocation in a Downside-Risk Framework." Financial Analyst Journal September-October 1991 pp. 28-40.

Harlow, W.V., Rao, R.K.S. (1989) "Asset pricing in a Generalized Mean Lower Partial Moment Framework." Journal of Financial and Quantitative Analysis vol 24 pp.285-312.

Hazzel, P. (1971) "A linear Alternative to Quadratic and Semivariance Programming for Farm Plans under Uncertainty." American Journal of Agricultural Economics 53 pp 53-62.

Hiller, R.S., Schaack, C. (1990) "A Classification of Structured Bond Portfolio Modeling Techniques." Journal of Portfolio Management, 17(1).

Jennergren, L. P., Sørensen, B.G. "Investeringsforeningenes Avkastning og Risiko" Fagskrift for Bankvesen, Vol 45 no: 3, Sept 1986.

Karmarkar, N. (1984), "A new Polynominal Time Algorithm for Linear Programming", Combinatorica, 4, No. 4. pp. 373–395.

Konno, H. Yamasaki, (1991) "Mean-Absolute Deviation Portfolio Optimization

Model and its Application to the Tokyo Stock Market." Management Science Vol 37 May pp.519-531

Land, A.H., Doig, A.G. (1960) "An Automatic Method for Solving Discrete Programming Problems." Econometrica, 28, pp. 497-520.

Levy, H. & Markowitz, H. (1979) "Approximating Expected Utility by a Function of Mean and Standard deviation." The American Economic Review. Vol 69 pp. 308-317.

Markowitz, H.M. (1952) "Portfolio Selection." Journal of Finance. pp 77-91.

Markowitz, H.M. (1959) Portfolio selection. Efficient Diversification of Investments. Basil Blackwell.

Markowitz, H.M.,Todd, P., Xu, G.and Yamane, Y. (1993) "Computation of Mean-semivariance Efficient Sets by the Critical Line Algorithm." Annals of Operations Research. Vol 45. Nos 1-4. pp. 307-317

Michaud, R.,O. (1989) "The Markowitz Optimization Enigma: Is "Optimized" Optimal?" Financial Analyst Journal. January/February 1989.

Minoux, M.(1986) Mathematical Programming- Theory and Algorithms. John Wiley & Sons.

Mulvey, J.M., Vladimirou, H. (1988) "Solving Multistage Stochastic Networks: An Application of Scenario Aggregation." Report SOR 88-1, Princeton University.

Murty, K.G. (1983), Linear Programming. John Wiley & Sons.

Pratt, J. (1964) "Risk Aversion in the Small and in the Large." Econometrica 32. pp. 122-136.

Rogers, D.F. (1986) "Aggregation Theory in Optimization and Markovian Theory: A Survey and an Analysis of Aggregation in Markov Chains." P.hD. Dissertation, Krannert Graduate School of Management, Purdue University, West Lafayette.

Rorgers, D.F., Plante, R.D., Wong, R.T.,Evans, J.R. (1991) "Aggregation and Disaggregation Techniques and Methodology in Optimization." Operations Research, Vol. 39 No.4 pp. 553-582.

Sharpe, W.F. (1963) "A Simplified Model for Portfolio Analysis." Management Science, January 1963 pp. 277-293.

Sharpe, W.F. (1967) "A Linear Programming Algorithm for Mutual Fund Portfolio Selection." Management Science pp. 499-510.

Sharpe, W.F. (1971a) "Mean Absolute Deviation Lines for Securities and Portfolios." Management Science. Vol 18 No. 2. B1-B13.

Sharpe, W.F., (1971b) "A Linear Programming Approximation for the General Portfolio Analysis Problem." Journal of Financial and Quantitative Analysis. December pp1263-1275.

Stone, B.K. (1973a) "A General Class of Three Parameter Risk Measures." Journal of Finance Vol. 28 pp. 675-685.

Stone, B.K. (1973b) "A Linear Programming Formulation of the General Portfolio

Selection Problem." Journal of Financial and Quantitative Analysis.pp. 621-636.

Von Neumann, J., Morgenstern, O. (1953) "Theory of Games and Economic Behavior." Princeton University Press. Princeton. New Jersey.

Wendell, R. E. (1985) "The Tolerance Approach to Sensitivity Analysis in Linear Programming." Management Science. pp. 564–578.

Wendel, F. R. (1992) "Sensitivity Analysis Revisited and Extended." Decision Sciences. Vol 23. pp. 1127–42.

Wets, R. J. B., Rockafellar, R.T. (1991) "Scenarios and Policy Aggregation in Optimization under Uncertainty." Mathematics of Operations Research. Vol 16 No.1 pp. 119–147.

Wodolowski, F. R. (1991) "A Generalization of Wendell's Tolerance Approach Analysis in Linear Programming." Descision Sciences Vol. 22. pp. 792–810.

Young, R.D. (1965) "A Primal (all Integer) Integer Programming Algorithm." Journal Res. Nat. Bur. Standards, 69B pp. 213-250.

Young, R.D. (1968) "A Simplified Primal (all Integer) Integer Programming Algorithm." Operations Research, 16, pp. 750-782.

Zenios, S. A. (1991) "Massively Parallel Computations for Financial Modeling Under Uncertainty." In J. Merisov eds: Very Large Scale Computing in the 21st Century. SIAM, Philadelphia, PA, pp. 273-94.

Zenios, S. A. (1993) Eds: Financial Optimization, Cambridge University Press.

Zipkin, P.H. (1977) "Aggregation in Linear Programming. Ph.D. Dissertation, School of Organization and Management, Yale University, New Haven, Conn., USA.

2. Risk Return Models and Sensitivity Analysis.

Abstract: In this chapter we consider a linear risk return model and a quadratic risk return model. In the first model the investment decision is made by the use of the expected return and the expected absolute deviation of the portfolio. In the second, traditional model, the investment decision is based upon the expected return and variance of return for the portfolio. For the linear model, we present the model formulated by Konno and Yamazaki (1991) and Hazzel's formulation (1972). The quadratic optimization model is the model formulated by Markowitz (1952). For the linear model, we will first look at traditional sensitivity analysis. For the quadratic model we will first look at a quadratic parametric programming approach presented by Grauer and Best (1984) for investigating variation in the estimated parameters for the MV model. After that, we will investigate how single perturbations in the expected returns can be handled in this model. For the linear model we will consider a technique introduced by Wendell (1985) and extended by Wodolowski (1991). We then perform a sensitivity analysis based upon individual, single perturbations of the estimated expected returns. Further, we perform simultaneous sensitivity analysis for the linear model based upon the theory of Wendell and Wodolowski. The analysis is based upon the use of actual data from the Oslo Stock Exchange. In the appendix we have a closer look at the risk measure absolute deviation, and some of its properties.

2.1. Introduction.

When we use optimization models, we often use historical data to find estimates of the different moments for the distribution of the assets return. How good these estimates are depends upon several factors. If we use a large time interval when we estimate the expected return, variance and covariance for a particular asset, we may include irrelevant history, i.e. from a period when the underlying company activitiy was of a different nature than it is today. The debt structure of the company may also have changed. If we select a shorter time interval, we may end up with too few observations; hence the estimated value is not a good estimator due to lack of information. Alternatively, the estimator may be influenced by some extreme observations or outliers. Errors in the estimated values in the optimization model may also have been caused by extraordinary events in the sample period that the decision maker thinks are particular to this period and not relevant to future events in the economy.

Sensitivity analysis is an optimization technique that allows us to study how an original optimal solution changes as we change the inputh to the model. The behaviour of the optimal solution when we change the coefficients in the model is interesting, since we may not be sure of the correct values of the coefficients in the model. In this chapter we will

consider different approaches for performing sensitivity analysis.

The chapter is organised as follows. In Section 2.2. we will look at a linear risk return model that uses the absolute deviation as the measure of risk. In Section 2.3., we will look at the traditional MV problem. In Section 2.4. we will present the traditional way of conducting sensitivity analysis in linear programming. Single perturbation analysis for the quadratic model will be covered in Section 3.5. In Section 3.6. we will compare the two single perturbation approaches using data from the Oslo Stock Exchange. In Section 2.7. and 2.8. we will investigate optimality for the linear model further, by using the sensitivity approach of to Wendell (1984), (1985) and Wodolowski (1991). These two approaches will be investigated using a data set in Section 2.9. Finally, in section 2.10. we will draw conclusions.

2.2. A linear risk return model.

In this Section we will describe the mean absolute deviation model. It is assumed that the investor has preferences for expected return and preferences against absolute deviation of the return for the portfolio. Although no explicit Von Neuman Morgenstern utility function, where expected utility is a function of the portfolio's expected return and absolute deviation, has been presented, the model has been used in portfolio selection. The reason is that by doing distributional assumptions about the return generating process, i.e. by assuming a multivariate normal distribution, it can be shown that the portfolios characteristics are uniquely determined by the parameter's expected return and the absolute deviation of the return of the portfolio.

The general formulation of the problem is [1]:

[1a]   minimize $E\{ |\tilde{r}_p - \mu_p| \}$

subject to:

[1b] $E\{\tilde{r}_p\} \geq \bar{\mu}$

[1c] $\sum_{i=1}^{n} x_i = 1$

[1d] $x_i \geq 0$ $i \in 1,2,....,n$.

Here $\tilde{r}_p$ is the stochastic return of the portfolio, x_i is the fraction invested in security i and $E\{\bullet\}$ is the expectation operator. The problem [1] will determine the portfolio with the

minimum absolute deviation subject to the constraint that the expected return of the portfolio is greater than or equal to some predefined constant $\bar{\mu}$. If we let the stochastic return of asset i be $\tilde{r}_i$, then the expected return of asset i is:

[2] $\quad E\{\tilde{r}_i\} = \mu_i$

The expected return of the portfolio is then given by:

[3] $\quad E\{\tilde{r}_p\} = \sum_{i=1}^{n} x_i E\{\tilde{r}_i\} = \sum_{i=1}^{n} x_i \mu_i$

Model [1] is a general version of the mean absolute deviation problem. When we use the model we use estimates for the expected return and absolute deviation. The formulation of the problem below is based upon the estimation of these parameters.

The programming technique used to solve this risk return model is linear programming, for which efficient algorithms are easily available. The model uses the empirical data directly, so there is no need for preprocessing the data by, for example, estimating a covariance matrix or another measure of comovements between asset prices. The effect of comovements between the assets is handled by the optimization model. The estimated absolute deviation of a portfolio is given by:

[4] $\quad \hat{\upsilon}_p = \sum_{t=1}^{T} \frac{|r_{pt} - \hat{\mu}_p|}{T-1} = \sum_{t=1}^{T} \frac{|a_{pt}|}{T-1}$

[4a] $\quad a_{pt} = r_{pt} - \hat{\mu}_p$

[4b] $\quad r_{pt} = \sum_{i=1}^{n} x_i r_{it}$

[4c] $\quad \hat{\mu}_p = \sum_{i=1}^{n} x_i \hat{\mu}_i$

[4d] $\quad \hat{\mu}_i = \sum_{t=1}^{T} \frac{r_{ti}}{T}$

Where:

r_{it} = Return of asset i in period t.

T = Number of observations.

n = Number of assets.

$\hat{\mu}_i$ = Estimated mean return on asset i.

$\hat{\mu}_p$ = Estimated mean return on the portfolio.

$\hat{\upsilon}_p$ = Estimated absolute deviation of portfolio.

If we assume that the return generating process (ex ante) is multivariate normal, Peters (1856) showed that:

[5] $E\{\hat{\upsilon}_p\} = E\{|\tilde{r}_p - \mu_p|\}$

Thus, Peter's results show that $\hat{\upsilon}_p$ is an unbiased estimator for the absolute deviation of the portfolio if the returns of the asset are distributed multivariate normal.

Absolute deviation is an L_1 measure of dispersion, devations are not squared as in the case with variance; instead the length of the deviation from the mean is used as a measure of risk. Such a measure has nice properties as regards the existence of the risk measure. It places weaker assumptions on the underlying distribution, since one could present a class of distribution functions where the risk measure variance does not exist, but where the L_1 measure exists[1].

Here we will present an optimization model that finds efficient portfolios under the risk measure absolute deviation. The formulated problem is a linear programming problem.

When we study this problem we will minimize risk or absolute deviation subject to the constraint that expected return of the portfolio is greater than or equal to a predefined constant. Konno's and Yamazaki's (1991) formulation of the problem is [6]:

[6a]   minimize $\dfrac{1}{T-1}\displaystyle\sum_{t=1}^{T} y_t$

subject to:

---

[1] Such distributions as stable Paretian with characteristic exponent less than two.

$$[6b] \quad y_t - \sum_{i=1}^{n} a_{it} x_i \geq 0 \quad t \in 1,2,\ldots.T.$$

$$[6c] \quad y_t + \sum_{i=1}^{n} a_{it} x_i \geq 0 \quad t \in 1,2,\ldots.T.$$

$$[6d] \quad \sum_{i=1}^{n} x_i \hat{\mu}_i \geq \bar{\mu}$$

$$[6e] \quad \sum_{i=1}^{n} x_i = 1$$

$$x_i \geq 0, \, y_t \geq 0 \quad i \in 1,2,\ldots.n. \quad t \in 1,2,\ldots.T.$$

The problem minimizes the sum of the deviations for each time period. The deviation for each period is given by one of the equations [6b] or [6c] above, and the objective function gives us the estimator described above. For each t, one of the constraints [6b] or [6c] is active. The constraint [6d] gives the lower bound of the expected return of the portfolio, and constraint [6e] reflects that the fraction invested should total one. However, it can be shown that the expected absolute deviation is twice the expectation of the deviations below the mean (υ_p^-). The derivation of this relation is given below:

Theorem:

$$[7] \quad \begin{cases} \upsilon_p = 2w_p^- = 2w_p^+ \text{ where } w_p^- = E(\,|\tilde{r}_p - \mu_p|^-) \text{ and } w_p^+ = E(\,|\tilde{r}_p - \mu_p|^+). \\ \text{where } |g|^- = 0 \text{ if } g \geq 0 \text{ and } -g \text{ if } g < 0, \text{ and } |g|^+ = 0 \text{ if } g < 0 \text{ and } g \text{ if } g \geq 0. \\ g \in \Re. \end{cases}$$

Proof:

We define the variables $\tilde{w}_p^-$ and $\tilde{w}_p^+$ as [8]:

$$[8a] \quad \tilde{w}_p^- = |\tilde{r}_p - \mu_p|^-$$
$$[8b] \quad \tilde{w}_p^+ = |\tilde{r}_p - \mu_p|^+$$

The following equation is satisfied by $\tilde{w}_p^-$ and $\tilde{w}_p^+$:

[9] $\tilde{r}_p + (\tilde{w}_p^- - \tilde{w}_p^+) = \mu_p$

Solving for $\tilde{w}_p^-$ and $\tilde{w}_p^+$ gives [10]:

[10a] $\quad \tilde{w}_p^+ = \tilde{r}_p + \tilde{w}_p^- - \mu_p$

[10b] $\quad \tilde{w}_p^- = \mu_p - \tilde{r}_p + \tilde{w}_p^+$

If we take expectations of $\tilde{w}_p^-$ and $\tilde{w}_p^+$ we obtain[11]:

[11a] $\quad E(\tilde{w}_p^-) = E(\tilde{w}_p^+)$

[11b] $\quad E(\tilde{w}_p^+) = E(\tilde{w}_p^-)$

Using $E(\tilde{w}_p^-) + E(\tilde{w}_p^+) = E(\tilde{\upsilon}_p)$ completes the proof.

□

The distribution of the return $\tilde{r}_p$ is not specified to be, for example symmetric. The result above holds as long the expected value, μ_p is finite. We then have:

[12] $\quad \upsilon_p^- = E\{ |\tilde{r}_p - \mu_p|^- \} = -1*E\{\min(0, \tilde{r}_p - \mu_p)\} = \dfrac{1}{2}\upsilon_p$

This relation also holds for the estimators, so:

[13] $\quad \hat{\upsilon}_p = \displaystyle\sum_{t=1}^{T} \dfrac{|r_{pt} - \hat{\mu}_p|}{T-1} = \dfrac{-2\displaystyle\sum_{t=1}^{T}\min\{0, r_{pt} - \hat{\mu}_p\}}{T-1}$

The implication of this is that the expected absolute deviation is twice the expected value of a portfolio's return, given that it falls below the mean. The equivalence between the risk measure absolute deviation and the risk measure expected value below the mean, imply using these two risk measure should give the same optimal portfolios. This means that the

commonly used assumption, that υ_p^- gives portfolios with positively skewed densities, is wrong and that in fact υ_p^- is a symmetric risk measure, giving equal weights to observations or outcomes above and below the expected return of the portfolio. The result above also addresses the risk measure absolute deviation as one special case of the lower partial moment risk measures discussed by Bawa and Lindeberg (1977). The results above also hold for the traditional estimators.

Using the information outlined above, the Konno and Yamazaki (1991) formulation can be simplified, and we are left with the formulation presented by Hazzel, a model originally used in farm planning under uncertainity (1972)[14]:

[14a]   minimize $\hat{\upsilon}_p = \dfrac{1}{T-1} \displaystyle\sum_{t=1}^{T} y_t$

subject to:

[14b] $y_t + \displaystyle\sum_{i=1}^{n} a_{it}x_i \geq 0 \quad t \in 1,2,....T.$

[14c] $\displaystyle\sum_{i=1}^{n} x_i \hat{\mu}_i \geq \mu$

[14d] $\displaystyle\sum_{i=1}^{n} x_i = 1$

[14e] $x_i \geq 0, \; y_t \geq 0, \; i \in 1,2,...,n., \; t \in 1,2,....,T.$

The objective function here is one half of the objective function in the formulation of Konno and Yamazaki (1991).

We note that we now have a formulation with n + T variables and T + 2 constraints. Thus, an optimal solution exists with at most T + 2 positive variables. For each time period, the value of y_t is zero if the return of the portfolio is above its expected return, and positive if the return of the portfolio is below its expected return.

There have been several attempts to use the model outlined above with actual data. Konno and Yamasaki (1991), in their study, used the 225 assets included in the Nikkei index. They performed a comparison of the mean absolute deviation model against the traditional

MV model, and one of their conclusions was that the portfolios generated by the two approaches did not differ much. My own experiences with the model, using data from the Oslo Stock Exchange, confirm these results.

Below we have calculated the ex post efficient frontier for the mean absolute deviation model for 21 stocks listed on the Oslo Stock Exchange in the period from 1.jan 1987 to 31. Des. 1992 (72 observations), using the LINDO optimization software.
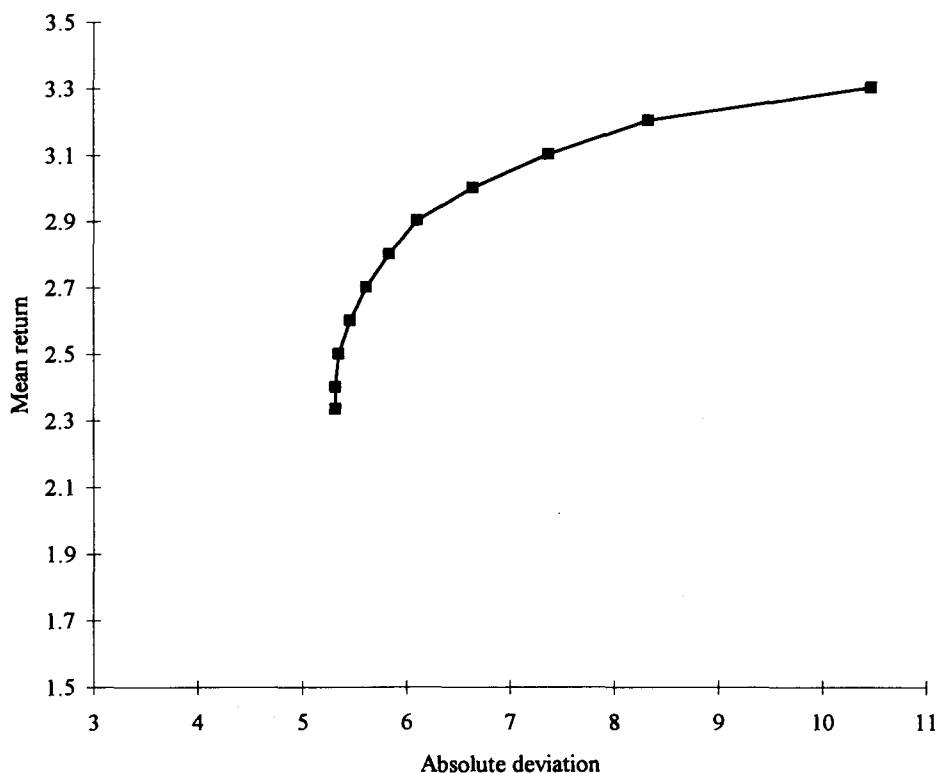
Figure 2.1. Mean Absolute deviation ex post frontier.

We now turn our attention to the quadratic portfolio problem.

2.3. The quadratic portfolio problem.

In this Section we will begin with a brief presentation of the traditional mean variance model, as formulated by Markowitz (1952), and then investigate a related, perturbed problem. In section 2.5. we will look at a method which may be used when we investigate the sensitivity of the optimal solution to changes in the estimated expected returns for the individual assets in the MV problem.

The traditional MV model can be formulated as follows[15]:

[15a]   minimize $\dfrac{1}{2}\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \sigma_{ij}$

subject to:

[15b] $\displaystyle\sum_{i=1}^{n} x_i \mu_i \geq \bar{\mu}$

[15c] $\displaystyle\sum_{i=1}^{n} x_i = 1$

[15d] $x_i \geq 0$

[15e] $i \in 1,2,...n.$

Here, x_i is the weight associated to asset i in the portfolio, μ_i is the expected return to asset i (ex ante), $\bar{\mu}$ is some predetermined level of return and σ_{ij} is the covariance between asset i and j.

When we solve this problem using actual data, we compute estimators. The estimators used here[2] are, for the elements in the variance covariance matrix, given by:

[16] $\hat{\sigma}_{ij} = \dfrac{1}{T-1}\displaystyle\sum_{t=1}^{T} a_{it} a_{jt}$

Above, a_{it} is given by $r_{it} - \hat{\mu}_i$. T is the number of observations. For μ_i we use the estimator referred to earlier, $\hat{\mu}_i$.

We have computed the ex post MV efficient frontier for the data set using the GAMS–MINOS optimization software. The results are shown in the figure below.

---

[2] The estimator used here is the traditional one. There also exist other estimators for the variance covariance matrix, such as e.g. Bayes-Stein estimators. See e.g. Efron and Morris (1973).
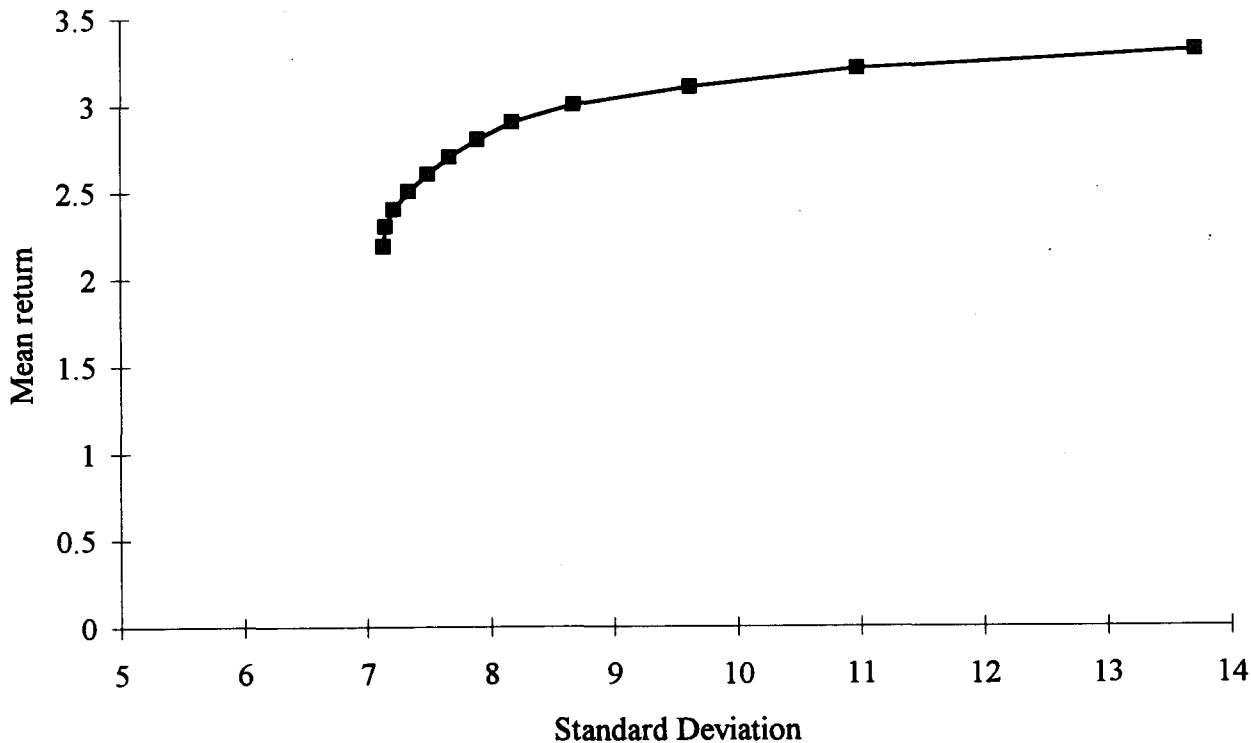
Fig. 2.2. Ex post MV frontier.

We now turn to the sensitivity analysis.

2.4. Methodologies of sensitivity analysis in linear programming.

In this Section, we will present the tools necessary to perform a sensitivity analysis of the mean absolute deviation model using an approach that perturbes the coefficients in the model. To do this, we will first formulate the linear programming problem as a parametric programming problem. We then state the theorem which gives bounds on the individual variations in the expected return vector, such that the original optimal basis remains optimal. In Section 2.6., we will look at the computational results of using the perturbation techniques outlined in this and the next section.

One common approach to sensitivity analysis is parametric programming. Parametric linear programming means reformulation of the linear programming problem to a problem where the parameters under investigation are perturbed.

In the expected return, absolute deviation model this means replacing the expected return

constraint [14c] by:

[17] $$\sum_{i=1}^{n} x_i(\mu_i + d_i\varphi) \geq \bar{\mu}$$

We have here perturbed the expected return vector, by a parameter φ, multiplied by the vector **d**. The n components of **d** are given by $d_1, d_2, \ldots, d_n$. This formulation leads us to a revised linear programming problem. The procedure involving such a program is termed parametric linear programming.

If we, for convenience[3], reformulate problem [14] to a maximization problem, and use the estimators, we have[18]:

[18a]   maximize $$\sum_{i=1}^{n} x_i(\hat{\mu}_i + d_i\varphi)$$

subject to:

[18b] $$\frac{1}{T-1}\sum_{t=1}^{T} y_t \leq \bar{\upsilon}$$

[18c] $$y_t + \sum_{i=1}^{n} a_{it}x_i \geq 0 \qquad t \in 1,2,\ldots.T.$$

[18d] $$\sum_{i=1}^{n} x_i = 1$$

[18e] $x_i \geq 0, \; y_t \geq 0$

Here we have formulated the problem such that the expected return of a portfolio is maximized subject to the constraint that its estimated absolute deviation should be less than or equal to some predefined level, given by $\bar{\upsilon}$, elsewhere the problem is equal to the original one.

We now have a parametric linear programming problem. We can solve the problem for different values of φ. An interesting problem is for what values of φ the original solution, or the solution where $\varphi = 0$, is still optimal. We will later answer this question, but first

---

[3] It is easier to perform a sensitivity analysis with coefficients in the objective function than the coefficients in the constraint matrix of a linear programming problem.

we must define certain sets.

In a mathematical linear program, there are variables that have a positive value in the optimal solution, and there are variables that are equal to zero. In our linear problem, there are two kinds of variables. The variables x_i $i \in 1,2,...,n$, are the decision variables and indicate the optimal fractions to invest. The variables y_t $t \in 1,2,...,T$, are intermediate auxiliary variables in the mathematical program. We will define the index set I_x for those decision variables that have a positive value in the optimal solution when $\varphi = 0$, and we will define the index set J_x to those decision variables that have the value zero in the optimal solution for $\varphi = 0$. In addition we define the index set I_y to those intermediate auxiliary variables (y_t) that have a strictly positive solution, and we define J_y as those of these variables that are zero. We will also define, for further use later, the union of the index sets I, J, as [19]:

[19a] $I = I_x \cup I_y$
[19b] $J = J_x \cup J_y$

In general, J will be the set of nonbasic variables, and I will be the set of basic variables in the optimal solution.

It is clear that if we change the value of the cofficients in which the corresponding variables are positive for $\varphi=0$, the objective function value will change in some direction, although the basis, in some interval, will remain unchanged. If we perturb the members of I_x enough, the basis may also change. Perturbations on the members of J_x do not change the solution if they do not change the optimal basis of the problem.

In the linear model we are then left with the following classifications of the perturbations:

Elements perturbed	Comments
$\in I_x$	Changes optimal obj. func. value, may change basis.
$\in J_x$	Has to change optimal basis to change optimal solution
$\in I_x \cup J_x$	Changes optimal obj. func. val., may change basis.

Table 2.1. Classifications of perturbations.

If we perturb the expected returns on members of I_x without changing the optimal basis,

the change will occur in the objective function and is given by [18a] minus the original optimal value. Below we will discuss changes that may affect the optimal basis. Specifically we will present a qualitative criterion on the scalar, given by the literature, such that the original basis remains optimal.

To clarify the generality of the results, we will assume for the linear program that the constraints are written in a general form. By this, we mean that the inequalities are converted to equalities by introducing slack variables when appropriate. The constraints in the general problem are written on the form:

[20] $Ax = b$

This notation is introduced to keep the overall notation simple. For convenience we will use parameter symbols without subscript as vectors. Thus, μ is a $n \times 1$ vector with the expected returns of the assets, and μ_I is a vector with the expected returns corresponding to the assets that are basic in the optimal solution, ($i \in I$). This notation will also be used for d and y.

We will denote the basis matrix of the problem [18] by B and $A_{\bullet j}$ as the j´th column of the generalized constraint matrix to the linear program. Conversely, $A_{i\bullet}$ is the i´th row of the matrix A. Further, we let c be the generalized cost vector, i.e. a vector that also consist of the zero coefficients of the auxiliary and slack variables.

Using results from linear programming we have the following theorem:

Theorem:

The initial optimal solution, the optimal solution with $\varphi = 0$, remains optimal as long as φ $\in [0,\overline{\varphi}]$ where $\overline{\varphi}$ is given by:

[21] $\overline{\varphi} = \min_{k \in J^*} \left\{ \dfrac{c_k - c_I B^{-1} A_{\bullet k}}{d_I B^{-1} A_{\bullet k} - d_k} \right\}$

[22] $J^* = \{ J : d_I B^{-1} A_{\bullet k} - d_k < 0 \}$

Proof [23]:

We investigate the optimal solution for the problem when $\varphi = 0$. The solution is given by:

[23a] $x_I^* = B^{-1}b$

Our perturbed cost vector is given by:

[23b] $c + \varphi d$

To obtain the maximum value of φ given by $\bar{\varphi}$ in the theorem we investigate the reduced cost vector of the problem. This vector is given by[4]:

[23c] $\xi = c_I B^{-1}A_J - c_J$

The reduced cost of the perturbed problem is given by:

[23d] $\xi^+ = [c_I + d_I\varphi]B^{-1}A_J - [c + d\varphi] = \xi + \varphi[d_I B^{-1}A_J - d_J]$

For a particular perturbation, given by the changes in φ, ξ^+ should stay positive. If it becomes negative, it violates the optimality conditions. If it turns out in our model that the vector given by $d_I B^{-1}A_J - d_J$ is positive, then φ can have any positive number. Still, the original basis will be optimal. If one or more elements of this vector are negative, then φ is restricted inside some interval, with upper bound $\bar{\varphi}$. The parameter ξ_k^+ is given by:

[23e] $\xi_k^+ = \xi_k + \varphi[d_I B^{-1}A_{\cdot k} - d_k] \geq 0$

If the inequality [23e] holds for all k, then the optimality conditions are satisfied. We must be sure that picking a number $\bar{\varphi}$ ensures that the condition holds. This means that $\bar{\varphi}$ is determined by the lowest value, for all k, such that the condition holds for all k. Using [23e] and [23c] gives us [21]:

□

We notice that if we want to perform an individual, single asset sensitivity analysis, this could be done by defining the vector **d** as the vector with only one element different from

---

[4] How one defines the reduced cost vector is a matter of definition. We have here used ξ, but we could also have used $-\xi$.

zero. This element corresponds to the particular asset under investigation, and the value of the element is one. The parametric programming approach discussed above is also applicable for right–hand side perturbations. As a example of this approach, we turn back to the mean absolute deviation model and we study perturbations in $\bar{v}$. The effect of changing this parameter is to move along the efficient frontier. If we establish a range for variations in this coefficient, we have established the range or interval on the efficient frontier that has the same basis as the optimal one. To do this, we formulate the problem as [24]:

[24a]   maximize $\displaystyle\sum_{i=1}^{n} \hat{\mu}_i x_i$

subject to:

[24b] $\displaystyle\frac{1}{T-1}\sum_{t=1}^{T} y_t \le \bar{v} + \beta$

[24c] $\displaystyle y_t + \sum_{t=1}^{T} a_{it} x_i \ge 0$

[24d] $\displaystyle\sum_{i=1}^{n} x_i = 1$

[24e] $x_i \ge 0$

We then have a perturbed problem with one new parameter.

We add a parameter, β, to the right–hand side of the constraint [24b]. The effect of this perturbation would illustrate the risk return tradeoff between the mean return and the absolute deviation of the optimal portfolio. From linear programming theory we know that the optimal solution, x_I^*, given a basis B, is given by:

[25] $x_I^* = B^{-1}b$

Here b is the vector, described in the previous section, consisting of zeros everywhere, except in two positions, where it has the values $\bar{v}$ and 1. We now define the vector $\hat{b}$ as b, except that $\bar{v}$ is replaced by β and the number 1 is replaced with 0. The vector $\hat{b}$ is a perturbation vector used in sensitivity analysis on $\bar{v}$. In this case, the perturbation vector only consists of one element different from zero.

Allowing the perturbation coefficient β to have values different from zero, but values such that the original basis remains optimal gives:

[26] $\quad x_I^* = B^{-1}b + B^{-1}\hat{b}$

Theorem:

The basis remains optimal as long as:

$B^{-1}b + B^{-1}\hat{b} \geq 0$

Proof:

The proof is trivial since the solution to a linear program is given by [26] when β is the optimal basis. If the perturbation above drives either X or Y to zero, new constraints become effective and a new basis enters the optimal solution.

☐

We will now look at sensitivity analysis for the quadratic model.

2.5. Sensitivity analysis in the quadratic model.

In this Section we discuss approaches to use in performing a sensitivity analysis on the quadratic model. We discuss the study by Grauer and Best (1984), and an approach where we perturbe the estimated expected return of the assets on a one at a time basis. This approach utilizes known properties of the efficient frontier to obtain a linear inequality between the covariance between the optimal portfolio – nonbasic assets and allowable perturbation upwards in the estimated return of the nonbasic asset. This relationship is usable when we compare the result for this procedure for the quadratic model to that of single perturbations in the linear model.

In their article Grauer and Best (1984) investigates the sensitivity of the solutions to

changes in the (estimated) expected returns in the traditional MV problem. We recall that the MV problem is a portfolio problem where the variance of the portfolio is minimized, subject to the condition that the expected return should be over some predefined level. If one allows the expected return to vary, one will have a tradeoff between expected return and variance in the expected return variance space. Grauer and Best studied a parameterized version of the problem, in which they allowed for perturbations in the expected returns of the assets. If we write the parametric MV problem in compact form, we have:

$$[27] \quad \text{maximize} \left\{ \lambda(\mu + \theta q)x^T - \frac{1}{2}x^T\Sigma x \mid x \geq 0, x^T 1 = 1 \right\}$$

Here λ is a parameter reflecting the investor's tradeoff between risk and return. The parameter θ is a parameter that is allowed to vary, q is a vector and indicates the direction of the changes in the expected returns. If θ or q is equal to zero, this is the ordinary parameterized MV problem. Denoting nonzero values to θ and q gives us a parametric programming problem where we are able to investigate the effects of changes in the estimated values of μ_i (for instance from μ_i to $\mu_i + \theta q_i$) on the optimal solution. If we assume that the problem [27] has been solved for some specific value λ, we can investigate how the optimal solution depends on different values of θ. We observe the similarity between the perturbed version of the quadratic problem given by [27] and the parametric version of the linear model given by [18].

Let us now assume that we have performed a mean variance optimization with the original expected return parameters, then assume that we perform a perturbation on the expected returns as indicated in problem [27].

Expected return
portfolio
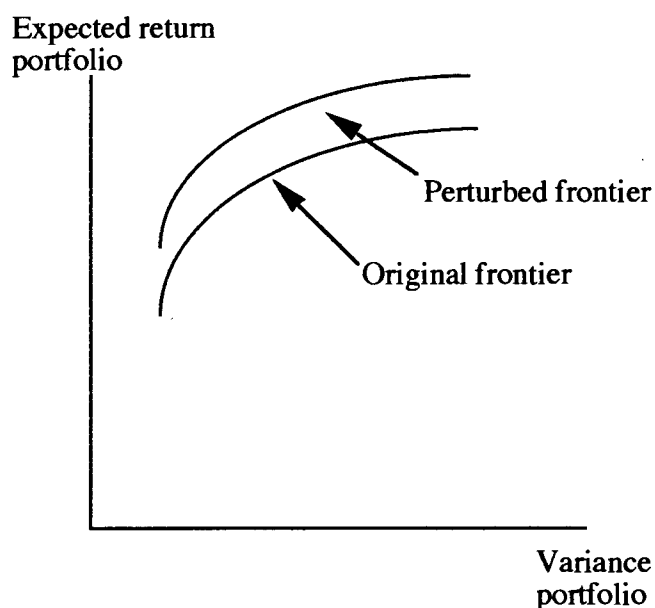
Figure 2.3. Perturbed frontier.

Above we have illustrated the effect on the parameterized quadratic programming problem by drawing the original portfolio frontier and the frontier obtained by a perturbation of the expected return vector. The figure illustrates that for one arbitrary perturbation, the new perturbed frontier changes; but since the variance covariance matrix is unchanged, the variance of the minimum variance portfolio frontier is unchanged. In the figure above the perturbation is done in a direction that could imply overall higher expected returns. If we interpret the figure above as a uniform shift of the efficient frontier upwards, the perturbation could be equal for all the expected returns.

There are several methods for solving the perturbed problem above. One possible approach is to derive the optimal weights as a function of the perturbation parameters. Another approach is to solve the problem iteratively, by varying the perturbation parameters until the requested information is revealed. The sensitivity analysis above can be done for a single security, by setting all components in q equal to zero except for one, corresponding to the asset under investigation. Or it can be viewed as a multiple perturbation, in one direction, by letting two or more of the elements in q be nonzero. For the first part of the analysis in this chapter, we will look at single perturbations, following a simpler, or a special case, compared with the approach of Grauer and Best.

We will now consider single perturbations of the nonbasic assets, and by the use of efficient frontier mathematics (discussed later), we can state the following theorem:

~41~

Theorem:

Consider an optimal solution with return r_p^* and index set I_x associated with the basic variables in the problem [15]. The index set I_x is no longer optimal if a nonbasic asset's expected return, with original expected return μ_k, is replaced by μ_k^* where μ_k^* satisfies:

$$[28] \quad \mu_k^* \geq \frac{Cov(r_p^*, r_k) - \lambda_1}{\lambda_E}$$

Proof:

Assume an optimal solution, X^*, to the problem where $k \in I_x$. In addition, assume that the optimal weight for this solution is given by, $x_k = Inf\{x_k : k \in I_x\} = 0$. To clarify this, the expected return of asset k is set such that it lies on the border between being a member of I_x and J_x, which means that it is a member of the basis and nonbasic assets. This is what we would call a degenerate optimal solution.

If this is true, the optimality condition associated with this asset is:

$$[29] \quad \sum_{i=1}^{n} x_i \sigma_{ik} - \mu_k \lambda_E - \lambda_1 = 0$$

Equation [29] holds for all $k \in I_x$. We recall that the other optimality conditions are not affected by variations in the nonbasic asset's (indexed by k) expected return. Solving for μ_k gives us:

$$[30] \quad \frac{\sum_{i=1}^{n} x_i \sigma_{ik} - \lambda_1}{\lambda_E} = \frac{Cov(r_p^*, r_k) - \lambda_1}{\lambda_E} = \mu_k^*$$

Since $k \in I_x$ for $\mu_k \geq \mu_k^*$, the equation [30] holds with inequality and μ_k^* is the value for asset k in which asset k enters the basis.

□

We will now perform an analysis of the characteristics for the line given by [30]. To be able to do this, we have to work with the assets which are in the basis in the original

optimal solution. This means that we analyse a subproblem. This subproblem considers only the assets which are basic in the original optimal solution. To use Markowitz words we study the critical line given by the basic assets with index set I. The analysis lies close to the analysis performed by Merton (1973) and Huang and Linzenberger (1988). The problem is [31]:

[31a]   minimize $\frac{1}{2}\mathbf{x}^T \Sigma_I \mathbf{x}$

subject to:

[31b] $\mathbf{x}^T \mu_I = \bar{\mu}$

[31c] $\mathbf{x}^T \mathbf{1} = 1$

[31d] $\mathbf{x} \in \mathfrak{R}^n$

Here $\mathbf{1}$ is a vector $n \times 1$ consisting of ones and Σ_I is the covariance matrix corresponding to the basic assets. The Lagrangian is:

$$L = \frac{1}{2}\mathbf{x}^T \Sigma_I \mathbf{x} - \lambda_1(\mathbf{x}^T \mathbf{1} - 1) - \lambda_e(\mathbf{x}^T \mu_I - \bar{\mu})$$

If we differentiate L with respect to the variables $\mathbf{x}$, λ_1 and λ_e we get the optimality conditions[32]:

[32a] $\dfrac{\partial L}{\partial X} = \Sigma_I \mathbf{x} - \lambda_e \mu_I - \lambda_1 \mathbf{1} = 0$

[32b] $\dfrac{\partial L}{\partial \lambda_e} = \mathbf{x}^T \mu_I - \bar{\mu} = 0$

[32c] $\dfrac{\partial L}{\partial \lambda_1} = \mathbf{x}^T \mathbf{1} - 1 = 0$

It can be shown by the use of the conditions stated above that [33]:

[33a] $\lambda_e = \dfrac{C\bar{\mu} - A}{D}$

[33b] $\lambda_1 = \dfrac{B - A\bar{\mu}}{D}$

The parameters A, B, C and D are defined as follows [34]:

[34a] $A = 1^T \Sigma_I^{-1} \mu_I$

[34b] $B = \mu_I^T \Sigma_I^{-1} \mu_I$

[34c] $C = 1^T \Sigma_I^{-1} 1$

[34d] $D = BC - A^2$

If we insert the expressions given by [33] and [34] above into [28], we obtain the relation between the covariance and the cutoff rate. Utilizing some algebra, this excess cutoff rate, defined as the level of expected return that separate those cases where the asset is a member of the optimal basis with those cases in which it is not a member of the optimal basis, can be formulated as:

$$[35] \quad \mu_k^* \geq \frac{A\overline{\mu} - B}{C\overline{\mu} - A} + \frac{D}{C\overline{\mu} - A} Cov(r_p^*, r_k)$$

The relation given by [35] is useful for several reasons. First, it gives us a nice linear relationship between the expected return necessary for a particular nonbasic asset, with the expected return on the other asset held fixed, to be a member of the basic assets. Second, it gives us a relation such that we can compare the effects of single perturbations in the MV model with the effects of single perturbations in the mean absolute deviation model. In the mean absolute deviation model, the analogous relation was obtained by theorem [21]. We will utilize these two relations in the next section.

2.6. Results for the traditional sensitivity analysis on data from Oslo Stock Exchange.

In this Section we will present results of using the two parametric programming approaches outlined in the previous chapters for the linear and quadratic models.

In our study we will use 21 stocks listed on the Oslo Stock Exchange. The data is taken from the period of 1 Jan. 1987 to 31.Des 1992 (72 monthly observations). We have selected the stocks on the basis of several criteria. First, we avoid stocks that have low liquidity, such that stocks traded in fewer than 150 trading days during a year on average are excluded from the set. Second, the stock has to be a member of the trading list over the

whole period. If mergers have taken place,

we have adjusted for this by adjusting the new stock with the switch ratio offered to the old stockholders in the two companies. For sectors with special events like the bank and insurance sector, we have used the bank and insurance index instead. All return series are adjusted for splits, emissions and dividends to preserve the true return of each asset. The data was delivered by the database Amadeus at the Norwegian School of Economics and Business Administration. This data set will be used for all the tests of the models described in this chapter.

The analysis is done by the use of the LINDO linear programming optimization software package. An ASCII file is constructed with the problem in a format readable by LINDO. The single perturbations sensitivity analysis is a part of the LINDO optimization package.

We will consider three optimal portfolios, which we will denote by the low, medium and high risk portfolios. Their returns for the period under investigation are 2.50, 3.05 and 3.25 respectively.

As is often the case when using historical data to perform a mean variance analysis, only a few assets have positive weight in the optimal solution. Among the 21 assets used in the analysis, only 7, 4 and 2 had positive weights in the three portfolios described above.

We will first look at a table which gives us the results for the sensitivity analysis on the linear model. Then we combine the results obtained from the linear model with those from the mean variance model utilizing the relation [28].

TICKER	$\hat{\mu}_i$	μ_{Low}^-	μ_{Low}^+	μ_{Medium}^-	μ_{Medium}^+	μ_{High}^-	μ_{High}^+
AKE	0.435	−Inf	3.169	−Inf	3.051	−Inf	3.150
DNL	−0.021	−Inf	4.022	−Inf	2.913	−Inf	3.073
DYN	1.580	−Inf	2.308	−Inf	2.838	−Inf	3.099
ELK	0.762	−Inf	8.414	−Inf	3.524	−Inf	3.159
HNA	2.789	2.357	3.080	2.789	2.810	−Inf	3.030
KVE	1.782	−Inf	3.678	−Inf	2.971	−Inf	3.057
NHY	1.781	1.447	2.234	−Inf	2.774	−Inf	3.098
ORK	2.251	−Inf	4.570	−Inf	3.050	−Inf	3.128
SAG	3.311	−Inf	6.326	3.223	3.315	3.040	9.970
SIM	1.798	1.209	2.153	−Inf	2.885	−Inf	3.074
TOM	2.092	1.403	2.490	−Inf	2.804	−Inf	3.086
UTO	3.036	2.649	4.115	3.036	3.102	2.808	3.31
BEA	1.980	1.733	2.247	−Inf	2.695	−Inf	3.049
BON	2.387	−Inf	7.774	−Inf	3.128	−Inf	3.131
NAL	2.717	2.450	3.337	2.695	2.724	−Inf	3.058
STO	2.177	−Inf	4.283	−Inf	2.887	−Inf	3.096
VAR	1.104	−Inf	5.06	−Inf	3.246	−Inf	3.141
WWI	2.281	−Inf	6.171	−Inf	3.158	−Inf	3.182
BNK	−2.255	−Inf	2.455	−Inf	2.888	−Inf	3.083
FOR	−0.192	−Inf	1.606	−Inf	2.967	−Inf	3.096
NSI	1.013	−Inf	1.532	−Inf	2.918	−Inf	3.065

Table 2.2. Results single perturbations in the linear model.

In the table above, we have listed the different assets and their range of allowable variations. The sensitivity analysis is performed on the estimated expected returns. The first column is the ticker code for the asset. A table explaining these ticker codes can be found in the appendices. In column two we have the estimated expected returns for the

assets, which are simply averages of the returns in the periods under investigation. At a first glance at these estimates, one could pick several assets that have an estimated expected return that makes them invalid for implementation of a mean risk model. An example is the BNK asset, an index representing the returns of the banking industry in Norway in the period, which has an estimated expected return of -2.25 %. One could hardly argue that this estimate is representative for the expected return in the forthcoming periods. This case also points out the relevance of a sensitivity analysis, to investigate what happens with the optimal solution if we change this value.

We have in the analysis above evaluated three portfolios. These portfolios are the low risk, medium risk and high risk portfolios. The six other columns in the table above correspond to the three portfolios. There are two columns for each portfolio denoted by μ^- and μ^+. In the table above, μ^- is the lowest value to which we could change the expected return of the asset without changing the optimal basis. Conversely, μ^+ is the highest value we could assign to the expected return of the asset without changing the optimal basis of the optimal portfolio.

Assets which are not in the optimal basis may not be attractive because they have a low expected return. Changing of this expected return then has no effect, such that μ^- for these assets is $-\infty$. For example one particular asset, ELK (Elkem), a typical high risk asset which is not in the optimal basis in the optimal low risk portfolio, must have a high estimated expected return to be a member of the low risk portfolio (8.414). This is also the case with two other risky assets, the shipping companies Bonheur (BON) and Wilhemsen (WWI).

In general, not surprisingly, it seems that risk is a factor that determines the cutoff rate (μ^+). The risk measured by the variance of the asset or the absolute deviation of the asset could only explain a part of the cutoff rate. Covariability with the other assets is also a central point.

We will now look at the relation between single perturbations for the mean variance and the mean absolute deviation model. In the mean variance model we use the derived relationship in [29]. The equation for our portfolios is computed as[5]:

---

[5] The numbers are computed in a spreadsheet by using the relevant submatrices and average returns.

Portfolio	$\dfrac{A\bar{\mu} - B}{C\bar{\mu} - A}$	$\dfrac{D}{C\bar{\mu} - A}$
Low risk	−2.9357	0.05078
Medium risk	2.0939	0.00577
High risk	2.8044	0.00153

Table 2.3. Single perturbations in MV model.

Below we have illustrated the three portfolios by single perturbation of the nonbasic assets.
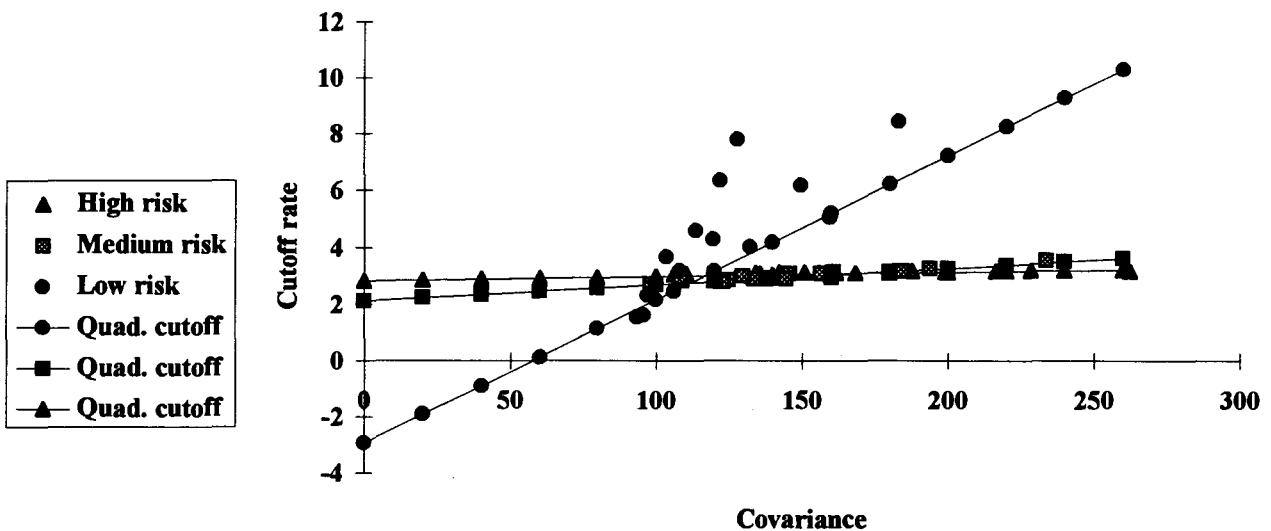
Figure 2.4. Cutoff rates MAD versus MV implied cutoff rates.

It seems that the allowable range for variations upwards for the high risk portfolio when using the MV model and the mean absolute deviation model does not differ much. As the risk in the portfolio is reduced there seems to be a difference between the two models. This holds specially for the low risk portfolio. This is reasonable since the difference between the two models lies in the way they consider the risk of the portfolio. For portfolios with low risk, typically selected by investors that emphasize the risk element in the portfolio, the two models differ, reflecting the difference in the risk measure used.

We will now look closer at multiple perturbations in the linear model.

2.7. Wendell's tolerance approach.

The analysis described in the previous sections in general, specifies a direction, given by either q or d, and the perturbation is done in the specified direction. One disadvantage of this approach is that it requires a specification of the vector d or the parameter q. We may not have any idea of how these parameters should be specified.

In this section we will present the analysis tools necessary to perform individual perturbations of the expected return vector of the linear model. The technique is based on the tolerance approach to sensitivity analysis in linear programming, presented by Wendell (1984), (1985) and extended by Wodolowski (1991) and Wendell (1992).

First we will look at the problem from a general linear programming view, and thereafter look at the theoretical consequences for our linear model. Consider the general perturbed linear programming problem[36]:

[36a]   maximize $\sum_{i=1}^{n} (c_i + d_i \gamma_i) x_i$

subject to:

[36b] $\sum_{i=1}^{n} a_{ij} x = b_j + \delta_j \beta_j \quad j \in 1,2,...,m.$

[36c] $x_i \geq 0 \quad i \in 1,2,...,n.$

This problem differs from the traditional parametric linear program in the sense that we allow for independent perturbations on the cost vector and right–hand side variables. As in the case with parametric programming, we specified a direction, given by the vector d, in which we forced the coefficients in the objective function by the use of the scalar φ. In the approach above we liberate these movements by not assigning such a direction. We refer to these as independent perturbations. In the model above these perturbations are in the model above also incorporated in the right–hand side of the constraints.

If we let $c_i = d_i$ and $b_j = \delta_j$ for all i and j, we can interpretet γ_i and β_j as percentage deviations of the parameters c_i and b_j. If d_i and δ_j are equal to one, the interpretation of γ_i and β_j is additive deviation from the original parameters. If we in a sensitivity analysis want to incorporate sensitivity analysis to one parameter but not to another, we can do this by giving the value 0 to d_i if c_i is to be held fixed and we give δ_j the value 0 if b_j is to be held fixed.

Another application of the sensitivity analysis above, without extending applications of this approach on the models in this chapter, is to define a region in which the parameters may vary. If we let Δ be such a subset, assuming that Δ is a polotype, i.e. $\Delta = \{\delta_1 : \delta_1 \leq \delta_2 \leq \delta_1 + 1\}$, then we could perform a sensitivity analysis given the polotype Δ. Of course, such definitions are also allowable for the perturbation parameters (d_i) in the objective function.

We will now study in more detail the sensitivity analysis based upon Wendell´s approach. We will only study variations in the objective function coefficients of the linear programming problem [36]. Variations in the right–hand side coefficients in the general linear programming problem will not be investigated. For further information see Wendell (1984), (1985) and (1992).

In the Wendell analysis we are interested in a single tolerance value, denoted by τ^*, which gives a general bound for the allowable variations in the coefficients in the objective function of the linear programming problem. In the presentation below we give the results in a general form. We do not yet restrict ourselves if the variations are additive or percentage deviations. In addition, we will assume that the tolerance value τ^* is finite. This means that we assume that the problem under investigation is not degenerate.

The original, unperturbed solution, remains optimal as long as the perturbation is inside the following interval:

[37] $R_\gamma = [\gamma : (-\tau^* \leq \gamma_i \leq \tau^*)]$

The number τ^* is given by[38]:

[38a] $\tau^* = \{\min(\tau_k : k \in J_x)\}$

Here τ_k is a parameter computed for each nonbasic variable. In order to describe this parameter, we first look at the reduced cost in the problem with multiple perturbations.

The traditional reduced cost vector is given by:

[39] $\xi = c_I B^{-1} A_J - c_J \geq 0$

For one particular nonbasic variable, k, the reduced cost is given by:

[40] $\xi_k = c_I B^{-1} A_{\bullet k} - c_k \geq 0$

When we allow for multiple perturbations, we have to incorporate this into the equation [40]. The reduced cost, for variable k, with multiple perturbations is given by:

[41] $\xi_k^* = \{c_B B^{-1} A_{\bullet k} - c_k\} - \gamma_k d_k + \gamma_B \begin{bmatrix} d_{h_1} B_{1\bullet}^{-1} A_{\bullet k} \\ \cdot \\ \cdot \\ d_{h_m} B_{m\bullet}^{-1} A_{\bullet k} \end{bmatrix} \geq 0 \ \forall \ k \in J$

The term in the bracket is the reduced cost obtained from the original solution and $\gamma_k d_k$ is the direct increase in the coefficient in the objective function of the perturbed problem. Above h_i is an index set for the basic variables. The last term of the right–hand side in [41] correspond's to the term $\varphi d_I B^{-1} A_{\bullet k}$ in the reduced cost of the traditional parametric linear program. Since we now study multiple perturbations, the scalar φ is replaced with the vector γ_B. The right–hand side of equation [41] is the reduced cost in the problem with individual perturbations. As long as this inequality holds, the original solution remains optimal.

We recognize that the traditional reduced cost is extended in [41] by the incorporation of the multiple perturbations. The equation [41] defines a halfplane. If the inequality is replaced with an equality, it defines a hyperplane. It may be shown that the parameter τ_k is the optimal solution to the following program[6] [42]:

[42a]   minimize $\|\gamma\|_\infty$

subject to:

[42b] $\gamma \in H_k^\tau(\gamma)$

---

[6] Here the norm $\|\gamma\|_\infty$ is the max norm of the vector γ, or the maximum value of the absolute values of the elements in the vector.

Here $H_k^=(\gamma)$ is the set:

$$[43] \quad H_k^=(\gamma) = \left\{ \gamma : c_k + \gamma_k d_k - c_B B^{-1} A_{\bullet k} - \sum_{j=1}^{m} \gamma_{h_j} d_{h_j} B_{j\bullet}^{-1} A_{\bullet k} = 0 \right\}$$

A useful result, to be used below, is given in Wendell (1985). The results give an optimal solution to problems of the type[44]:

[44a] minimize $\|z\|_\infty$

subject to:

[44b] $g z = g_0$

Above g is a vector with constants, z is a variable vector and g_0 is a constant. The solution to the problem is given by:

$$[45] \quad z^* = \frac{g_0 \text{sgn}(g)}{g \text{sgn}(g)}, \quad \|z^*\| = \frac{g_0}{g \text{sgn}(g)}$$

We notice that [43] is similar to [41], except that we have replaced the inequality with an equality. In order to obtain an intuitive understanding of the approach, we notice that the solution to the program [42], given by the vector γ^*, consists of elements with equal absolute value. This follows, since if the absolute values were different, one could decrease the variable in the vector γ with the largest absolute value by altering another variable in γ such that [42b] is fullfilled. Since such a modification that gives us a lower objective function [42a] is possible, γ can not be optimal until the absolute values in γ are equal.

Further we recognize that as long as the absolute values of the elements in γ are less than or equal to $\|\gamma^*\|_\infty$, the reduced cost for variable k given by [41] remains positive. Thus perturbations of the objective function of the program with any of the γ_i in the range $\pm \|\gamma^*\|_\infty$, would not change the the sign of [41]. As long as[7] $\Delta \in \Re^n$ the solution to the problem [42] could be obtained by the use of [44]. By the use of [45] we obtain:

---

[7] Here $\Delta \in \Re^n$ means that we have not restricted the variations in the objective function to be member of a subset of $\Re^n$.

[46] $\quad \tau_k = \dfrac{c_B B^{-1} A_{\cdot k} - c_k}{|d_k| + \displaystyle\sum_{j=1}^{m} |d_{h_j} B_{j\cdot}^{-1} A_{\cdot k}|}$.

Once we have computed the different τ_k, the value of τ^* is obtained by selecting the minimum value of the τ_k's computed. By using τ^* as an upper bound of the allowable variations in the parameters γ_i, we are assured that $\xi_k^* \geq 0 \ \forall \ k$. This means that the original basis remains optimal.

We now turn to the linear mean risk model. For the mean absolute deviation we could incorporate the approach by formulating the linear programming problem with perturbations in the objective function [47]:

[47a] $\quad$ maximize $\displaystyle\sum_{i=1}^{n} (\hat{\mu}_i + \gamma_i d_i) x_i$

subject to:

[47b] $\quad \dfrac{1}{T-1} \displaystyle\sum_{t=1}^{T} y_t \leq \bar{\upsilon}$

[47c] $\quad y_t + \displaystyle\sum_{t=1}^{T} a_{it} x_i \geq 0$

[47d] $\quad \displaystyle\sum_{i=1}^{n} x_i = 1$

[47e] $\quad x_i \geq 0$

We then have a perturbed problem with 2n new parameters. Note that we do not perturb the auxiliary variables y_t, which have zero as the coefficient in the objective function in the original problem.

We will consider additive shifts in the expected returns in the objective functions of the problem. We could then compute the different τ_k as:

$$[48] \quad \tau_k = \frac{\mu_I B^{-1} A_k - \mu_k}{1 + \sum_{j=1}^{m} |I_j B_{j\bullet}^{-1} A_k|}$$

Here I_j is given by:

$$[49] \quad I_j = \begin{cases} 1 : j \in I_x \\ 0 : \text{else} \end{cases}$$

The indicator variable, I_j, reflects the fact that we do not perturbe the objective function coefficient, (given by zero), to the variables y_t. The variable τ^* is then found, by following [38a].

The maximum tolerance approach is the infinum of the calculated parameters τ_k. Wendell´s approach, with simultaneous changes of the parameters in the objective function, could be illustrated, in the two parameter case, as in the figure below:
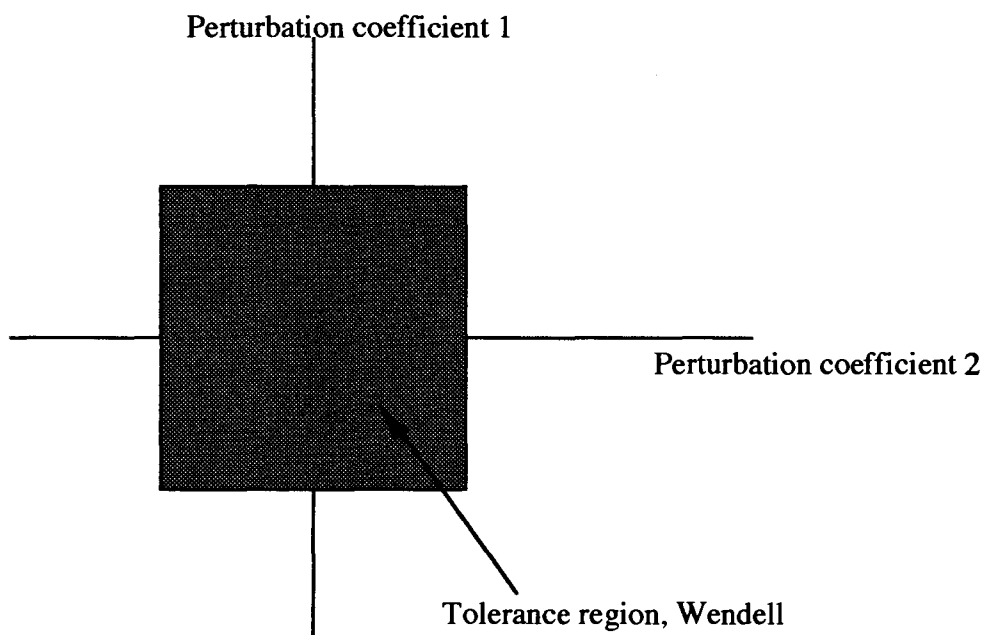
Figure 2.5. Wendels single tolerance value with two variables.

Above we have illustrated the concept of the tolerance parameter τ^*, by using an example with two variables with perturbations on their cofficients in the objective function. The detemination of the parameter τ^*, defines the shaded area in the figure above. As long as

the perturbation of the two cofficients lies inside the shaded box, the original optimal solution is still optimal. In our particular problem, the shaded area will turn up as an n dimensional hypercube.

Wendell's method tells us how much each coefficient is allowed to vary, independent of the others, so that the solution remains optimal. This approach, however, may result in conservative ranges for the uncertain parameters in question, since the selected tolerance is the infinum for the coefficients in which variation does not lead to a new optimal basis. Further, the range is based on the assumption that the direction of change in the coefficient is the direction in which the optimal basis changes first. With this in mind, Wondolowski presented an approach in which those objections were considered.

2.8. Expanding the tolerance region.

In this Section we will widen the band given by the tolerance approach proposed by Wendell. The expansion is based upon the work of Wodolowski (1991)[8].

In the tolerance approach above, we determined a parameter τ^* that is a value in which the estimated expected returns could vary without changing the optimal basis. The approach has been criticised for not providing as wide bands on the variation in the objective function parameters as it could. Wodlowski claimed that it is possible to expand the tolerance region by calculating a vector of parameters that gives the allowable variation of the coefficients upwards and downwards. In our model this means allowable ranges in downwards and upwards moves for the expected return coefficients.

We will denote those vectors by ψ^- and ψ^+ respectively, and an interval for objective function coefficient i, is given by $[\psi_i^-, \psi_i^+]$. The elements of this vector are, by the use of the notation for the general linear programming problem, defined by [50]:

[50a] $\quad \psi_i^- = \begin{cases} c_i - \min_{k \in J} \{\tau_k : B_{i\bullet}^{-1} A_{\bullet k} > 0\} \\ -\infty \quad \text{otherwise.} \end{cases}$

[50b] $\quad \psi_i^+ = \begin{cases} c_i + \min_{k \in J} \{\tau_k : B_{i\bullet}^{-1} A_{\bullet k} < 0\} \\ c_i + \tau_i \quad \text{otherwise} \end{cases}$

---

[8]This approach was also presented by Wendell (1992).

To clarify this result, we have to consider the modified reduced cost vector. The elements in this vector are given by [41]. If we notice that the vector $\mathbf{d}$ is the unit vector, which means that we investigate additive perturbations, element k in the modified reduced cost vector becomes:

$$[51] \quad \xi_k^* = \{c_B B^{-1} A_{\bullet k} - c_k\} - \gamma_k + \gamma_B \begin{bmatrix} B_{1\bullet}^{-1} A_{\bullet k} \\ \cdot \\ \cdot \\ B_{m\bullet}^{-1} A_{\bullet k} \end{bmatrix} \geq 0 \;\; \forall \; k \in J$$

We first consider positive perturbations variables given by γ_B. If:

$$[52] \quad B_{i\bullet}^{-1} A_{\bullet k} \geq 0 \;\; \forall \; k \in J$$

then positive perturbations in the cost coefficient in the objective function for the basic variables would not change the sign of ξ_k^*. The allowable upward movement for the cost coeficient i is then given by τ_i. If variable i is nonbasic, this value is given by τ_k, as computed above. If any of the values given by [52] is negative, it would limit the allowable variation of the cost coefficient upwards, and we have to select the tolerance value upwards as the minimum of the computed parameters τ_k, given that the corresponding value of $B_{i\bullet}^{-1} A_{\bullet k}$ is negative. Conversely, in the case with downside variations, a situation where:

$$[53] \quad B_{i\bullet}^{-1} A_{\bullet k} \leq 0 \;\; \forall \; k \in J$$

means that perturbations where γ_B is negative do not make ξ_k^* negative. If one or more of the values given by [53] is positive,this would limit the variation downwards, and we have to use the minimum value of τ_k, given that the corresponding value of $B_{i\bullet}^{-1} A_{\bullet k}$ positive.

The result above is converted to our portfolio model by letting $c_i = \mu_i$, such that if the expected return coefficient changes from μ_i to $\hat{\mu}_i$ and is inside the interval $[\psi_i^-, \psi_i^+]$, the original solution remains optimal.

The ranges for individual perturbations specified by the interval $[\psi_i^-, \psi_i^+]$ are completely included in the interval obtained by the use of the τ^* parameter in Wendell's analysis. The approach of Wodolowski is then a widening of the bands given by the use of the parameter τ^*. We illustrate the expansion of the bands by using the figure below, where there are

two variables.
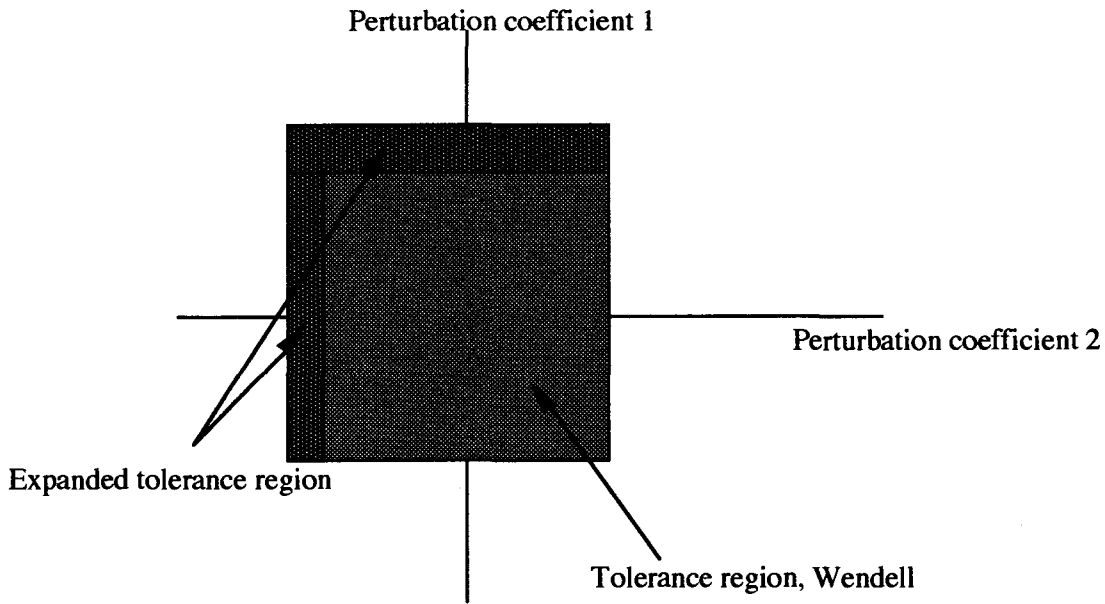
Figure 2.6. Wodolowski's individual tolerance ranges.

We see that Wodolowski's extension of the tolerance approach expands the set in which the coefficient of the objective function is allowed to vary. The expanded region is illustrated in the figure above by the dark shaded area.

If we compare the sensitivity analysis obtained by Wodolowski's approach to that of single perturbations or traditional sensitivity analysis presented earlier in this chapter, the Wodolovski approach gives us more freedom when we vary the expected returns of the assets. This is so because we are allowed to vary all the expected returns in this approach, compared with one variation at a time in traditional sensitivity analysis. As in the case with parametric linear programming, the sensitivity analysis is performed in one specified direction.

2.9. Results – individual perturbations in the mean absolute deviation model.

In this Section we will look at the extensions of the sensitivity analysis presented in the last two sections applied to our portfolio problem. We will conduct an analysis of the mean absolute deviation portfolio problem by studying the sensitivity of the solution when we perform perturbations on the expected returns of the assets and determine which perturbations the original solution remains optimal for. Individual perturbation will decrease the allowable varying range in the parameters, and to cope with this we will study

Wodolowski's (1991) extension of the tolerance approach, to widen the bands, such that our sensitivity analysis will be more complete. A similar approach in the quadratic case is not performed since the theory of multiple perturbations in this case is, to our knowledge, not fully developed.

We use the data from the Oslo Stock Exchange to perform a sensitivity analysis based upon the approaches of Wendell and Wodolowski. The sensitivity analysis is based upon additive perturbations.

To be able to do this we use the LINDO linear programming software, which has the ability to provide us with the optimal simplex tableau. We have constructed an ASCII file with the problem. This file is read by LINDO and solved for the three portfolios studied. The optimal simplex tableau is converted into an ASCII file which is read into Microsoft Excel, where the necessary calculations are done. We start by looking at the single tolerance approach given by Wendell. The results for the parameter τ^* are as follows:

Portfolio	Low	Medium	High
τ^*	0.138	0.0039	0.116

Table 2.4. Single tolerance values.

The number τ^* is the allowable variation in the expected return. If for example the estimated return is 1.78%, the range of variations given by the low risk portfolio is (1.642,1.918). As indicated above, the parameter τ^* gives very conservative ranges for the allowable variations in the estimated expected returns. This is especially the case for the medium risk portfolio. If we look at Table 2.2., we see that the ranges for the basic asset for this portfolio, compared with the other portfolios, are also conservative. We will try to widen these bands by using the approach suggested by Wodolowski (1991).

Below we have listed the results for the Wodolowski approach using the same type of table as in the analysis performed on the data using traditional sensitivity analysis earlier in this chapter.

TICKER	μ_i	μ^-_{Low}	μ^+_{Low}	μ^-_{Medium}	μ^+_{Medium}	μ^-_{High}	μ^+_{High}
AKE	0.435	–Inf	1.133	–Inf	0.865	–Inf	1.791
DNL	–0.021	–Inf	0.345	–Inf	0.301	–Inf	1.527
DYN	1.580	–Inf	1.733	–Inf	2.111	–Inf	2.340
ELK	0.762	–Inf	1.059	–Inf	0.969	–Inf	1.960
HNA	2.789	2.751	2.930	2.789	2.794	–Inf	2.906
KVE	1.783	–Inf	2.046	–Inf	2.213	–Inf	2.418
NHY	1.781	1.640	2.046	–Inf	2.02	–Inf	2.439
ORK	2.251	–Inf	2.527	–Inf	2.424	–Inf	2.689
SAG	3.311	–Inf	3.513	3.306	3.314	3.309	3.426
SIM	1.798	1.660	1.840	–Inf	2.319	–Inf	2.437
TOM	2.093	1.950	2.129	–Inf	2.344	–Inf	2.588
UTO	3.035	3.001	3.180	3.036	3.044	2.924	3.0401
BEA	1.980	1.840	2.120	–Inf	2.316	–Inf	2.515
BON	2.387	–Inf	2.656	–Inf	2.585	–Inf	2.7605
NAL	2.717	2.580	2.870	2.716	2.724	–Inf	2.889
STO	2.176	–Inf	2.492	–Inf	2.374	–Inf	2.638
VAR	1.104	–Inf	1.474	–Inf	2.044	–Inf	2.121
WWI	2.281	–inf	2.571	–Inf	2.440	–Inf	2.731
BNK	–2.255	–Inf	–0.856	–Inf	–1.424	–Inf	0,416
FOR	–0.192	–Inf	0.153	–Inf	0.308	–Inf	1.453
NSI	1.013	–Inf	1.150	–Inf	1.268	–Inf	2.037

Table 2.5. Individual tolerance ranges.

The table above provides the results for Wodolovski's individual tolerance range approach, applied for the mean absolute deviation model. If we consider the sensitivity analysis obtained by the use of the Wodolowski approach to traditional sensitivity analysis presented earlier, we see that the band is tightened. This is not surprising, since we now allow for multiple perturbations. Typically, large variation in the bands as allowed by the traditional sensitivity analysis, e.g. for the ELK security with traditional μ^+_{Low} =8.414, is replaced by the more moderate 1.059. This reduction in the band width is also the case for BON and WWI.

If we compare the results with the single tolerance approach of Wendell, there is a considerable widening of the bands for the expected return of the assets, except for the basic assets in the medium risk portfolio, which is still very tight.

2.10. Conclusions.

In this chapter we have studied two mean risk models: the classical mean variance model and the mean absolute deviation model. First, we presented the two models and stated a theorem such that the formulation of the mean absolute deviation model of Konno and Yamazaki (1991) could be simplified. Then, we presented standard sensitivity analysis for the linear model. For the MV model we had a cursory look at the sensitivity analysis presented by Grauer and Best (1984). Thereafter, we discussed single perturbations of the MV model by working with the critical line of the original optimal basis. Utilizing the theory presented for the two models for single perturbations in the expected returns of the assets, we performed an empirical analysis and compared the two models. One of the results was, not surprisingly, that the allowable range of variations differed for low risk portfolios, but was almost identical for the high risk portfolio. We then turned towards multiple perturbations of the linear model and discussed the theory of Wendell´s tolerance approach and Wodolowski´s extension. Finally, we performed a multiple perturbation analysis using the two approaches on actual data from the Oslo Stock Exchange. We discovered that the single tolerance approach by Wendell resulted in conservative ranges, as opposed to the individual tolerance ranges obtained using of Wodolowski´s procedure.

Appendix I: Properties of the risk measure absolute deviation.

Abstract: This appendix gives a brief summary of the geometric interpretation of the risk measure absolute deviation and it's properties when the distribution of the returns on the assets are multivariate normal.

A.1.1. Absolute deviation - geometric interpretation.

If we let $\tilde{r}_p$ be a random variable with probability density function $f(\tilde{r}_p)$, using the result [7], the absolute deviation is given by:

[A.1.] $$v_p = E(|r_p - \mu_p|) = 2\int_{-\infty}^{\mu_p} (\mu_p - \tilde{r}_p)f(\tilde{r}_p)d\tilde{r}_p = 2F_2(\mu_p)$$

Here we have defined F and F_2 as:

[A.2.] $$F(x) = \int_{-\infty}^{x} f(y)dy$$

[A.3.] $$F_2(x) = \int_{-\infty}^{x} F(y)dy$$

A geometric interpretation of this result is that v_p is twice the area under the cumulative distribution function in the interval $(\infty, \mu_p]$.

A.1.2. Normality - and estimators.

When $\tilde{r}_p$ is normally distributed with parameters (μ_p, σ_p^2), the mean absolute deviation is given by:

[A.4.] $$v_p = E(|r_p - \mu_p|) = 2\int_{-\infty}^{\mu_p} (\mu_p - \tilde{r}_p)\frac{1}{\sqrt{2\pi}\,\sigma_p} \text{Exp}\left(\frac{(\tilde{r}_p - \mu_p)^2}{2\sigma_p^2}\right)d\tilde{r}_p = \sqrt{\frac{2}{\pi}}\,\sigma_p \approx 0.8\sigma_p$$

Thus there is a linear relation between the expected absolute deviation and the standard deviation of a portfolio.

When we have a return generating process from the multivariate normal distribution, an

estimator often used for the standard deviation is:

$$[\text{A.5.}] \quad \hat{\sigma} = \sqrt{\frac{\sum_{t=1}^{T}(r_{pt}-\bar{r}_p)^2}{T-1}}$$

The variance[9] of this estimator is given by Fisher (1920):

$$[\text{A.6.}] \quad \text{Var}(\hat{\sigma}) = \left[\frac{T-1}{2} + \frac{\Gamma^2\left(\frac{T-1}{2}\right)}{\Gamma\left(\frac{T}{2}\right)} - 1\right]\sigma^2 = \frac{1}{2T}\sigma^2 + O\left(\frac{1}{T^2}\right)$$

Peter (1856) showed that:

$$[\text{A.7.}] \quad E(\tilde{\sigma}) = \sqrt{\frac{2}{\pi}}\,\frac{\sum_{t=1}^{T}|r_{pt}-\bar{r}_p|}{T-1} = \sigma$$

Thus we have an alternative estimator ($\tilde{\sigma}$) for the population´s standard deviation.

If we take the variance of this estimate, we have, from Helmert (1876):

$$[\text{A.8.}] \quad \text{Var}(\tilde{\sigma}) = \left[\frac{\pi}{2} + \arcsin\left(\frac{1}{T-1}\right) - T + \sqrt{T(T-2)}\right]\frac{\sigma^2}{T} = \frac{\pi-2}{2T}\sigma^2 + O\left(\frac{1}{T^2}\right)$$

If we compare the variance of the two estimators by dividing one by the other, we have, by letting $T \rightarrow \infty$:

$$[\text{A.9.}] \quad \frac{\text{Var}(\hat{\sigma})}{\text{Var}(\tilde{\sigma})} \approx 88\,\%$$

Thus, the alternative estimator gives us an efficiency loss. In other words, we need approximately 14% more observations, when T is large, to get as accurate an estimate using the absolute deviation as that obtained using the traditional one.

---

[9] Here Γ is the gamma function.

Appendix II.   List of companies used.

Ticker	Name
AKE	Aker
DNL	Det Norske Luftfartsselskap (SAS)
DYN	Dyno Industrier
ELK	Elkem
HNA	Hafslund Nycomed
KVE	Kværner Ind.
NHY	Norsk Hydro
ORK	Orkla
SAG	Saga
SIM	Simrad
TOM	Tomra
UTO	Unitor
BEA	Bergesen dy.
BON	Bonheur (Fred. Olsen & Co.)
NAL	Den Norske Amerika Linje
STO	Storli
VAR	Vard
WWI	Wilhelm Wilhemsen Ltd.
BNK	Oslo Stock Exchange Bank Index
FOR	Oslo Stock Exchange Insurance Index
NSI	Norske Skogindustrier

References:

Bawa, V.S., Lindeberg, E.B. (1977) "Capital Market Equilibrium in A Mean-Lower Partial Moment Framework." Journal of Financial Economics. vol 5. pp. 189-200.

Effron, B., Morris, C., (1973) " Stein's Estimation Rule and Its Competitors – An Empirical Bayes Approach." Journal of the American Statistical Association, Vol 68. pp. 117–130.

Fisher, R.A. (1920) "A Mathematical Examination of the Methods of Determining the Accuracy of an Observation by the Mean Error, and by the Mean Square error." Royal Astronomical Society 80 pp. 758-69.

Grauer,R.R., Best, M.J. (1984) "Sensitivity Analysis for Mean Variance Portfolio Problems." Working Paper. Burnaby, B.C.: Department of economics, Simon Fraser University.

Hazzel, P. (1971) "A Linear Alternative to Quadratic and Semivariance Programming for Farm Plans under Uncertainty." American Journal of Agricultural Economics 53 pp 53-62.2.

Helmert, F.R. (1876) "Die Genaugigkeit der Formel von Peters zur Berechnung des Wahrscheinlichen Beobachtungfehlers Directer Beobachtungen Gleicher Genauigkeit." Astronomiche Nachrichten 88 pp. 112-31.

Huang, C. & Litzenberger, R.H. (1988) Foundations for Financial Economics. North Holland.

Konno, H. Yamasaki, H., (1991) "Mean-absolute Deviation Portfolio Optimization Model and its Application to the Tokyo Stock Market." Management Science Vol 37 May pp.519-531.

Markowitz, H.M. (1952) Portfolio Selection. Journal of Finance. pp 77-91.

Merton, R. (1972) "An Analytic Derivation of the Efficient Portfolio Frontier." Journal of Financial and Quantitative analysis.

Peters, C.A.F., (1856) "Ueber die Bestimmung des Wahrsceinlichen Fehlers einer Beobachtung aus den Abweichungen der Beobachtungen von ihrem Aritmetishen Mittel." Astronomishe Nachrichten, 44, pp. 30–31.

Wendell, R. E., (1984) "Using Bounds on the Data in Linear Programming: The Tolerance Approach." Mathematical programming 29. pp. 304–322.

Wendell, R. E., (1985) "The Tolerance Approach to Sensitivity Analysis in Linear Programming." Management Science Vol 31. No.5. pp. 564–578.

Wendel, F. R. (1992) "Sensitivity Analysis Revisited and Extended." Decision Sciences. Vol 23. pp. 1127–42.

Wodolowski, F. R. (1991) "A Generalization of Wendell's Tolerance Approach Analysis in Linear Programming." Decision Sciences Vol. 22. pp. 792–810.

3. Portfolio Optimization Models and Imposing Constraints Involving Integer Programming.

Abstract: This chapter investigates the MV problem when, in addition to the linear constraints imposed on such problems, such as sector constraints and upper and lower bounds, one imposes constraints that involve the use of discrete variables. The formulated problem is used to derive the ex post efficient frontier. Often, in practical situations, situations arise in which one has to consider portfolio models that use discrete decision variables. This chapter considers several cases where we formulate the portfolio selection problem as a problem with discrete decision variables. Among the cases studied are models with transaction costs, basis constraints and intervaling constraints. We formulate a linear program which is an approximation of the quadratic model. By the inclusion of the discrete variables, this problem becomes a mixed integer or binary linear programming problem. We construct two test problems, in which one is a formulation that captures the current regulation imposed on fund managers in Norway today. By using historical data we compute the ex post efficient frontier for the period 1987–92 based upon the use of data from the Oslo Stock Exchange. In addition we evaluate the generated portfolios in the period 1993–94.

3.1. Introduction.

Traditional portfolio theory, which has its origin in the Markowitz (1952) model, uses optimization techniques such as quadratic programming to solve the programming problem. This model has been extensively investigated and capital asset pricing models such as the Sharpe (1964), Lintner (1965) and Mossin (1965) CAPM are based upon this model.

When an investor designs his portfolio he is more or less restricted by the means in which he would select a portfolio that accomplishes the goals that he has set for his investments. Traditionally, these constraints dictate that no asset can be held short during the investment period. Further, the investor may be faced with sector–wise restrictions such that the assets invested in one sector, whether this sector is geographical, markets for firms' outputs or cap-size sectors, should or should not exceed some predefined fraction. This class of sector–wise constraints is linear, and the inclusion of these restrictions in a well defined optimization model does not complicate the model.

Another type of restriction, which involves the use of more complicated optimization tools like integer programming, has not received so much attention in the literature. One example of a restriction that requires the introduction of discrete decision variables is that the fund

manager is allowed to invest in a maximum of two assets from one specific sector. Another example is if one includes transaction cost in the model, where the transaction cost is fixed up to some level and above this level varies at a linear rate with the amount invested.

The introduction of integer variables makes an investment problem more complicated in the sense that one can not rely on the optimization algorithm used for continuous optimization. The set of allowable decisions is no longer continuous. To comply with this, we have to use the branch and bound algorithm or other alternative methods for handling discrete decision variables to sort out the optimal solutions.

Chapter 3 is organized as follows. In Section 3.2., we will reformulate the traditional MV problem as a problem with a separabable objective function. This is neccesary in order to perform the approximation outlined in Section 3.3. In Section 3.4., we discuss the use of discrete decision variables in general. In Section 3.5., we discuss basis constraints and interval constraints. Transaction cost is covered in Section 3.6. The effect of the introduction of discrete decision variables on the shape of the tradeoff between risk and return is discussed in Section 3.7. In Section 3.8., we present two test problems and the result of an optimization using actual data. Improvement of the solution procedure for one of these problems is discussed in Section 3.9. Finally, in Section 3.10. we draw some conclusions.

3.2. Expected return - Variance of return - Direct formulation.

The introduction of integer constraints into a model makes the model, in an operational research sense, quite different from its non–integer counterpart. When integer variables are introduced, the solution algorithm necessary to solve the problem has to take into account that the set of allowable solutions is nonconvex in $\mathfrak{R}^m$, where m is the number of variables in the problem. So, although from a user point of view the model is only slightly modified, the procedure to solve it differs considerably.

We assume that the basic model used by the investor in this chapter is the mean variance model. The models with discrete variables and solution procedures discussed in the chapter may also be used on other quadratic models, such as models which consider semivariance and index matching models. The mean variance model in its simple form, with short sale restrictions, takes the form[1]:

$$[1a] \quad \text{minimize} \quad \sum_{i=1}^{n} \sum_{i=1}^{n} x_i x_j \sigma_{ij}$$

subject to:

$$[1b] \quad \sum_{i=1}^{n} x_i \mu_i \geq \bar{\mu}$$

$$[1c] \quad \sum_{i=1}^{n} x_i = 1$$

$$[1d] \quad x_i \geq 0$$

$$[1e] \quad i \in 1,2,\dots,n.$$

The problem above is the traditional portfolio problem ex ante. Above, x_i is the fraction invested in security i and μ_i is the expected return of asset i.

The above problem is a quadratic programming problem, and could be solved using a suitable algorithm, such as Lemke's (1965) algorithm or the reduced gradient method.

In a portfolio model, we often use historical data to estimate the parameters in the model. In our formulation of the problem, we will formulate the problem in a direct way. In other words, we mean to formulate the model in such a way that we can use the observed data directly.

We define a_{it} to be $r_{it} - \hat{\mu}_i$, where $\hat{\mu}_i$ is the estimated expectation of the random return assets for asset i, and r_{it} is the realized return of asset i in period t. As estimates we use the arithmetic averages of the return series. We know that $E(a_{it}) = 0$. We also have:

$$[2] \quad \frac{\displaystyle\sum_{t=1}^{T} a_{it}^2}{T - 1} = \hat{\sigma}_i^2$$

where $\hat{\sigma}_i^2$ is the traditional estimator of asset i's variance.

We can now define a_{pt} as:

$$[3] \quad a_{pt} = \sum_{i=1}^{n} x_i a_{it}$$

Thus, to obtain the estimated variance for a portfolio we can calculate this as:

$$[4] \quad \hat{\sigma}_p^2 = \frac{1}{T-1} \sum_{t=1}^{T} \left[\sum_{i=1}^{n} x_i a_{it} \right]^2 = \frac{1}{T-1} \sum_{t=1}^{T} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j a_{it} a_{jt} =$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \sum_{t=1}^{T} \frac{a_{it} a_{jt}}{T-1} = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \hat{\sigma}_{ij}$$

Thus we can obtain a portfolio´s estimated variance as the sum from t=1 to T of the square of the variables in [3] above:

$$[5] \quad \frac{1}{T-1} \sum_{t=1}^{T} a_{pt}^2 = \hat{\sigma}_p^2$$

We have above defined the empirical variance of the portfolio by using the observations directly.

Using the relation [5], we can formulate the MV model on empirical data as[1] [6]:

$$[6a] \quad \text{minimize} \quad \frac{1}{T-1} \sum_{t=1}^{T} z_t^2$$

subject to:

$$[6b] \quad \sum_{i=1}^{n} x_i a_{it} = z_t \quad t \in 1,2,\ldots\ldots T.$$

---

[1] The formulation of the mean variance model given here could, with slight modifications, be turned to a model that uses other risk measures such as absolute deviation, semivariance, partial moments, etc.

[6c] $\displaystyle\sum_{i=1}^{n} x_i \hat{\mu}_i \geq \bar{\mu}$

[6d] $\displaystyle\sum_{i=1}^{n} x_i = 1$

[6e] $x_i \geq 0 \quad i \in 1,2,...,n.$

[6f] $z_t \in \mathfrak{R} \quad t \in 1,2,....,T.$

We see that the square of the deviations for each period $t \in 1,2,......T$ is added and minimized in the objective function. This means that we minimize the empirical variance of the portfolio. Although the formulation [6] is quite different from [1] at first glance, and although they have different numerical properties, they provide us with the same solution. The objective function in [1] is a quadratic form $x^T \Sigma x$, where Σ is the n×n variance covariance matrix of the assets return. In [6] the objective function is the quadratic form $z^T I z$, where I is the T×T identity matrix, which is very sparse. The MV problem is now in a form in which where the variables in the objective function are separable. Note that the variable z_t can take any value in the formulation above. We also note that the above problem has n+T variables and T+2 restrictions. If we want to restrict our use to nonnegative variables only[2], we could do this by defining:

[7] $z_t = z_t^+ - z_t^- \quad t \in 1,2,...,T.$
[8] $z_t^+ \geq 0, z_t^- \geq 0 \quad t \in 1,2,...,T.$

Here z_t^- takes a positive value if z_t is negative and z_t^+ takes a positive value if z_t^+ is positive. If we use [7] the program will be extended by T variables. Thus, the problem will have n+2T variables. The relation [7] also has to be substituted into the objective function.

3.3. Approximated version of direct formulation.

In this section we will present the approximated version of the direct formulation of the MV model. This approximation is done using the same principles as in Chapter 4 of this dissertation. In order to do this we define a set of meshpoints, denoted by:

[9] $S^{jt} \quad j \in 1,2,...,m_t. \ t \in 1,2,....,T.$

---

[2]This could be useful if we want the approximated version of the problem (LP) to be a LP with only nonnegative variables.

The meshpoints S^{jt} take values such that [10]:

[10a] $S^{max,t} = \max_{j}\{S^{jt}\} \geq \max_{i}\{a_{it}\}\ t \in\ 1,2,.....T.$

[10b] $S^{min,t} = \min_{j}\{S^{jt}\} \leq \min_{i}\{a_{it}\}\ \ t \in\ 1,2,.....T.$

The other meshpoints, in principle arbitrary choosen, lie in the interval:

[11] $S^{jt} \in [S^{min,t}, S^{max,t}]$

The meshpoints will work as approximation points to our original nonlinear function. In the case using the direct version of the MV problem, this function is:

[12] $f(S^{jt}) = (S^{jt})^2$

In the problem we will formulate the objective function given by:

[13] $\displaystyle\sum_{t=1}^{T} z_t^2$

We define a weight associated with each meshpoint S^{jt}. We denote this weight by w^{jt}. In order to assign two properties to these weights, we have to define the concept of neighbour as follows. Two weights w^{jt} and $w^{j^{*}t}$ are neighbours if $j = 1$ and $j^{*} = 2$ or, $j = m_t-1$ and $j^{*} = m_t$ or $j^{*} = j + 1$ or $j^{*} = j - 1$. The second property is [14]:

[14a] $\displaystyle\sum_{j=1}^{m_t} w^{jt} = 1\ \ \ t \in\ 1,2,...,T.$

The two properties we impose on the weights w^{jt} are then:

– The weights assigned to each t must be neighbours.

– The condition [14a] must hold for each t.

These two conditions are imposed on the vectors w^t, one for each t, with elements w^{jt}.

They simply says that in each of these vectors at most two neighbouring elements could be nonzero. The usefulness of these vectors becomes apparent when we study the linear programming problem[15]:

[15a]   minimize $\displaystyle\sum_{t=1}^{T}\sum_{j=1}^{m_t} w^{jt}(S^{jt})^2$

subject to:

[15b] $\displaystyle\sum_{j=1}^{m_t} w^{jt}S^{jt} - \sum_{i=1}^{n} a_{it}x_i = 0 \quad t \in 1,2,...,T.$

[15c] $\displaystyle\sum_{j=1}^{m} w^{jt} = 1 \quad t \in 1,2,...,T.$

[15d] $\displaystyle\sum_{i=1}^{n} x_i\mu_i \geq \bar{\mu}$

[15e] $\displaystyle\sum_{i=1}^{n} x_i = 1$

[15f] $x_i \geq 0 \; i \in 1,2,...,n.$

[15g] $w^{jt} \geq 0 \; j \in 1,2,...,m., \; t \in 1,2,...,T.$

This problem solves an approximated version of the traditional mean variance problem. The restriction [15b] assigns values to the weights w^{jt}. These weights will, according to [15c] for each time period, or observation, add up to one. As a consequence of the characteristics of the problem studied, at maximum and most probably two w^{jt} for each time period will be positive. If two of these weights are positive in one particular time period t, these two weights are neighbours. The reasons for this are that the approximated function is convex, and that we have a minimization problem. The fulfilment of these criterias is handled by the form of the objective function. As in the case with non convex functions, this condition is not neccesarly fullfilled. An example of such a case can be found in Chapter 4 of this dissertation. Nonconvex problems are more difficult to solve and we have to apply additional logical programming by introducing special ordered sets of type 2.

The objective function in the above problem will give an approximation to the sum of the squares in the problem. The goodness of the approximation is determined by the number of mesh points used.

We will illustrate the approximation for one particular time period t .
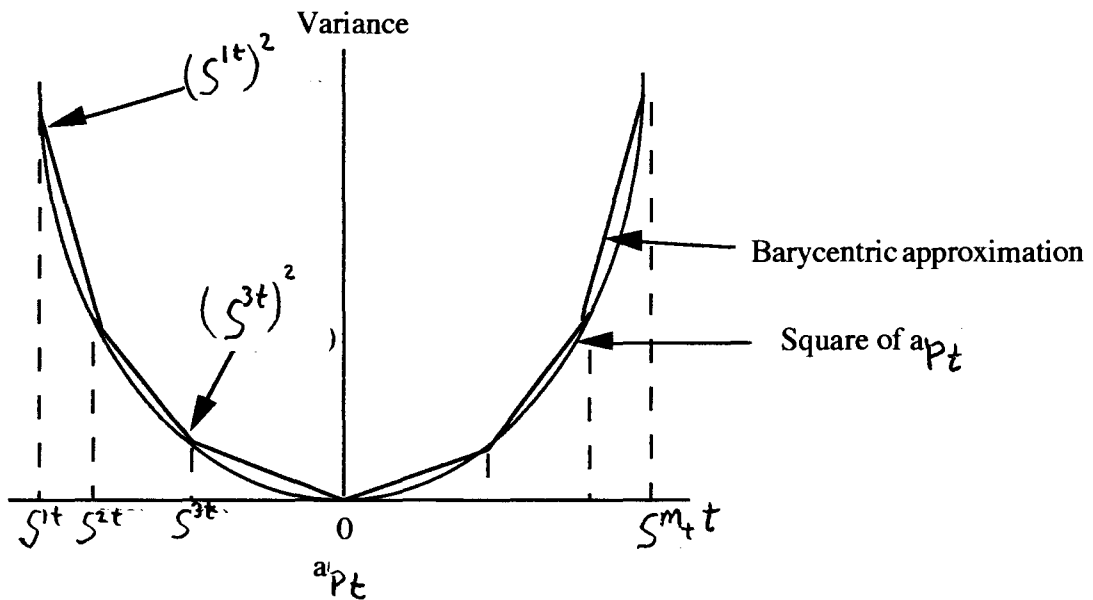
Figure 3.1. Approximation of the quadratic term – period t.

The above figure shows how the approximation works for one particular observation of the time series. Along the horizontal axis we have the deviation a_{pt}. In the direct MV formulation we want to minimize the sum of the squares of a_{pt}. In the approximated problem, the square of a_{pt} is approximated by the piece–wise linear function in the figure above. The weights w^{jt} are determined by the distance from a_{pt} to the two nearest meshpoints. If a_{pt} lies near one mesh point we will have an almost perfect approximation of a_{pt}^2. A value of a_{pt} exactly between two mesh points will have a somewhat poorer approximation of a_{pt}^2. The approximation is performed by a convex combination of the mesh points S^{jt}.

The constraints [15f], [15e] and [15d] are from the original mean variance problem.

The degree of approximation is determined by the number of and the selection of meshpoints S^{jt}. The user of such a model is to a large degree able to tune this

approximation such that the approximated objective function value and the associated solution to the linear programming problem [15] are close to the solution obtained if the original quadratic programming problem is solved.

The approximated linear programming problem above, with the inclusion of the integer constraints presented later, constitutes the problem we intend to solve to obtain a solution to the portfolio problem with discrete variables. Such problems are termed mixed binary linear programming problems.

3.4. General integer programming constraints.

In this section we will discuss, in a general framework, model formulations where we use integer constraints. Introducing integer constraints into a model clearly expands the set of possible restrictions in a mathematical model. In contrast to continuous programming, where the restrictions put into the model typically regulate the amount to invest in assets or a sector of assets, integer constraints expands the concept further by allowing us to formulate constraints on a logical form.

Integer constraints involve the use of integer variables which can be classified into several groups. One group comprises binary variables which are only allowed a value of 0 or 1. Another group consists of positive variables which are allowed to take positive integer values, such as $\{0,1,2,3,4,..\}$. A third classification is the set of all integer variables such as $\{-4,-3,-2,-1,0,1,2,3,..\}$. If we denote the variables by δ_i, we can construct the following table for a classification of the integer variables:

Variable	Set	Symbol
δ_i	0 or 1	$Z_{0,1}$
δ_i	0,1,2,3,4,......	Z^+
δ_i	-3,-2,-1,0,1,2,3	Z

Table 3.1. Different classes of discrete variables.

The models used in this chapter will only use one category of the integer variables discussed above. This category is $Z_{0,1}$.

The discrete decision variables can be used to form different types of conditions, a general discussion of which follows. The first are logical conditions which are of the type:

[16]   if A_1 then A_2 else A_3

Here A_1, A_2 and A_3 are events. A typical situation in optimization using such a condition is the use of models with transaction costs. The fee structure is often of a form such that a fixed fee is imposed if the transaction is below some fixed amount. If the transaction is above this level the fee is proportional to the transaction.

The next condition is that of set covering. If we let B be a set and B_1, B_2,..,B_L be subsets of B, we can cover the whole set by the use of B_1, B_2,..,B_L, such that the union of the subsets selected is B. An example of a situation where covering conditions could be imposed is for investments in some risky assets where the management requires that such investments should be covered by put options.

Set packing and set partitioning are conditions related to the set covering mentioned above. But, in addition, we require that the subsets selected, B_1, B_2,..,B_L, should be disjoint. Such a condition could be imposed if we want to guarantee no duplication of coverage in the option coverage example above.

With these general examples in mind we see that discrete optimization gives the user a new toolbox of restrictions to use in the asset allocation models. In order to understand the introduction of the binary variables into our problem, we now give a short introduction to the graphical representation of binary variables. The presentation of binary variables is often done by the help of a graph. If we consider a vector $\overline{A}$ with the three binary variables δ_1, δ_2 and δ_3, a possible ordering of these three variables is given in the figure below.

Root $[\delta_1,\delta_2,\delta_3]$

$[0,\delta_2,\delta_3]$ $[1,\delta_2,\delta_3]$

$[0,0,\delta_3]$ $[0,1,\delta_3]$ $[1,0,\delta_3]$ $[1,1,\delta_3]$

$[0,0,0]$ $[0,0,1]$ $[0,1,0]$ $[0,1,1]$ $[1,0,0]$ $[1,0,1]$ $[1,1,0]$ $[1,1,1]$
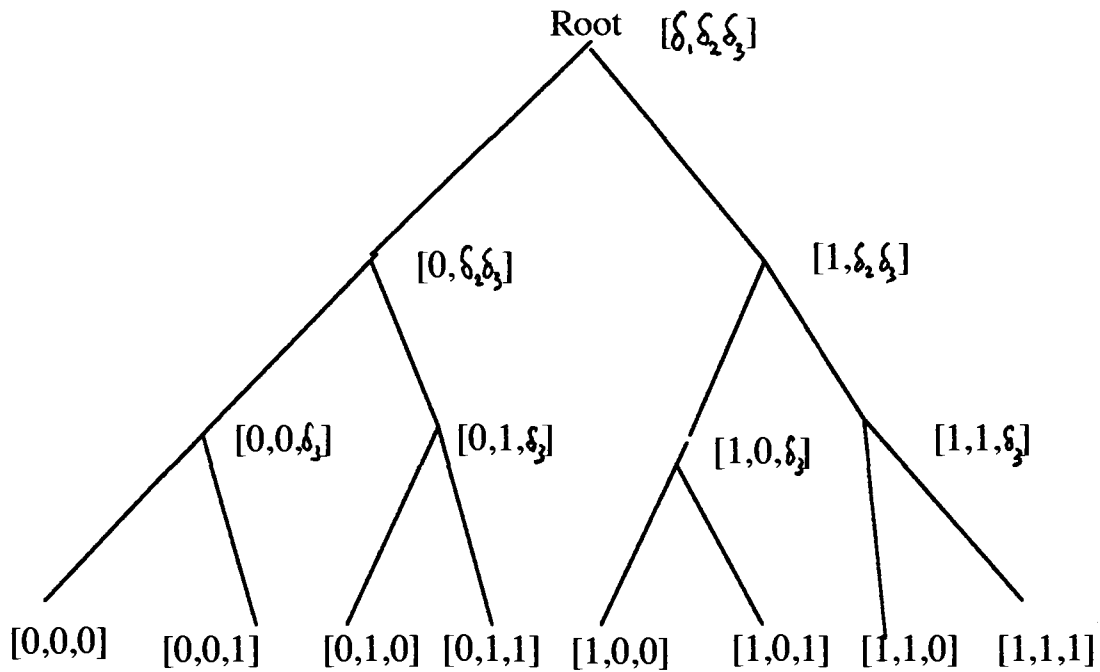
Figure 3.2. Binary search tree for three variables.

The set of possible combinations of the binary variables in the vector $\overline{\Delta}$ is partitioned, and in the leaves of the tree, or terminal vertices, the partition is complete. The procedure of moving down into finer and finer partitions is termed branching. By considering one variable for each level in the three, the different combinations of the vector $\overline{\Delta}$ are obtained by following one path.

There are four characteristics of the representation above. First, the graph does not contain any circuit. Second, there is one and only one vertex, termed root, that is not the terminal extremity of any arc. All the other vertices are the terminated extremities of a single arc. From the figure above we also note that the different nodes at each level only differ in one element. The term used in this case is that the Hamming[3] distance between the two vectors is one.

The general presentation shown above is fairly general, but we now turn to some specific cases.

3.5. Basis constraints and intervaling constraints.

In an optimization model we may demand that a finite number of the assets we are able to select from should have positive weights. If we had an optimization model with 1000

---

[3] The general Hamming distance between two binary vectors $\overline{\Delta}_1$ and $\overline{\Delta}_2$ is given by $h = |\overline{\Delta}_1 - \overline{\Delta}_2|$.

assets to select from, it would be an enormous task to handle small proportions in all of these assets, if a run by an ordinary optimization model could select 600-700 assets. Let us restrict the number of assets to be selected to a fixed number, say k. Then if we extend the original optimization model by the following equations, we will have a solution that contains a maximum of k assets[17]:

[17a] $\quad \delta_i - x_i \geq 0$

[17b] $\quad \sum_{i=1}^{n} \delta_i \leq k$

[17c] $\quad \delta_i \in Z_{0,1}$

[17d] $\quad i \in 1,2,3..,n.$

The restrictions of [17a] give the variable δ_i a value of 1 if asset i has a positive weight. The control [17b] assures that the number of assets with positive weights is less than or equal to k. We refer to such constraints as basis constraints. An interesting question is the effect on the objective function of increasing or decreasing k. One should expect that the dual price of the restriction [17b] should decrease as a function of k. If k, for example, is held at 3, the effect of increasing this number to 4 should be larger than the effect of increasing k from 50 to 51. The number of possible combinations of binary variables with the restrictions [17] is given by 2^n, where n is the number of assets.

By the use of the restrictions shown above, the decision maker controls the number of assets in the portfolio. Now we assume that the investor for some reason only wants to invest in amounts or fractions that are between a lower and upper bound. Otherwise he will not invest at all. Now assume that we put interval restrictions on the fraction to invest. By this we mean that the assets are restricted to be inside an interval or they have to be set at zero. We denote the interval for asset i as $[\zeta_i^d , \zeta_i^u]$. If asset i is inside this interval, the parameter δ_i should take the value 1. To obtain such a problem we extend the original optimization model with the following restrictions [18]:

[18a] $\quad x_i \leq \zeta_i^u \delta_i \quad i \in 1,2,....,n.$
[18b] $\quad x_i \geq \zeta_i^d \delta_i \quad i \in 1,2,....,n.$

[18c] $\quad \sum_{i=1}^{n} \delta_i \leq k$

[18d] $\quad \delta_i \in Z_{0,1} \quad i \in 1,2,....,n.$

The restrictions [18a] and [18b] are restrictions that give the parameter δ_i the value 1 if the optimal fraction happens to fall inside the interval. Restriction [18a] gives δ_i a value of one if x_i is strictly positive. Since δ_i is binary this restriction also assures an upper bound, ζ_i^u, of x_i. Since δ_i is one for strictly positive values of x_i, the lower bound, ζ_i^d, is only activated when $x_i > 0$. Thus x_i has to be in the interval $[\zeta_i^d, \zeta_i^u]$, or set to zero. The restriction [18c] is a counter for the number of cases falling inside its interval. We can then control the number of assets in the interval by using the restriction [18c]. It is important to note that this formulation is not equal to traditional restrictions imposed on MV problems such that each asset should be inside some interval given by an upper and a lower bound on the fractions to invest. In our case the assets are forced to be inside the interval $[\zeta_i^d, \zeta_i^u]$ if this is optimal compared to setting x_i to zero.

3.6. Transaction cost and integer programming.

In this section we will study the impact transaction costs have in a model formulation. To carry out a trade, a broker charges a fee which implicitly affects the return on the portfolio. There are several ways transaction cost can occur when purchasing assets. We will study one particular scheme, which is as follows:
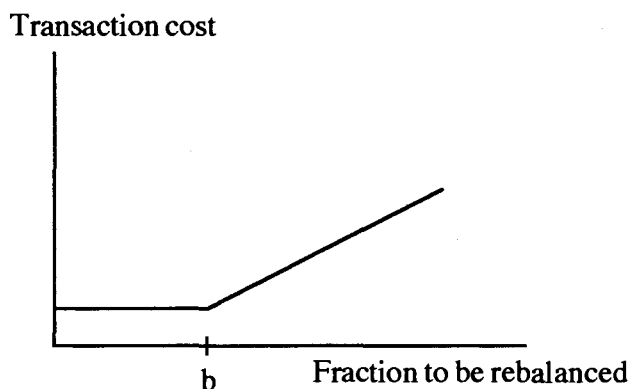
Figure 3.3. Transaction costs.

Here b is related to the minimum amount purchased of one share before the investor has to pay a fixed fee to the broker instead of a variable fee[4]. The minimum amount is given by bM_0, where M_0 is the value of the portfolio today. We will denote the fraction of security i, with investment weight x_i to be rebalanced with Δx_i. If security i has the weight $\bar{x}_i$ in the

---

[4] The parameter b is equal for all securities such that the minimum amount for a variable fee charge is equal for all assets.

portfolio now, and optimization shows that the optimal value to be x_i, the fraction to be rebalanced is given by:

[19] $\Delta x_i = |\bar{x}_i - x_i|$

If $\Delta x_i \leq b$ the investor has to pay a fixed amount in brokerage fee. If $\Delta x_i \geq b$, the brokerage fee is a fraction (α) of the value of the shares bought or sold. This particular transaction reduces the expected return of the optimal portfolio by $\Delta x_i \alpha$. Now consider a quadratic portfolio optimization model with fractions x_i to be invested in the securities and extended with the following restrictions [20]:

[20a] $y_i^- + (x_i - \bar{x}_i) \geq 0$ $i \in 1,2,...,n.$

[20b] $y_i^+ - (x_i - \bar{x}_i) \geq 0$ $i \in 1,2,...,n.$

[20c] $\Delta x_i = y_i^- + y_i^+$ $i \in 1,2,...,n.$

[20d] $z_i - \Delta x_i + b \geq 0$ $i \in 1,2,...,n.$

[20e] $\delta_i - \Delta x_i \geq 0$ $i \in 1,2,...,n.$

[20f] $y_i^-, y_i^+, \Delta x_i, z_i \geq 0$ $i \in 1,2,...,n.$

[20g] $\delta_i \in Z_{0,1}$ $i \in 1,2,...,n.$

Here y_i^- and y_i^+ are variables that will reflect the change in the fraction held of asset i. If we sell asset i , y_i^- will be positive and if we buy asset i, y_i^+ will be positive. The restrictions [20a] and [20b] also ensure that the product of y_i^- and y_i^+ is zero. In other words, they are complementary. The variable Δx_i is defined by [20c] to be the absolute change in the fraction held of security i, whether it is a net selling or net buying of the position. Restriction [20d] assigns the excess value, if any, of Δx_i over the cutoff parameter b to the variable z_i. Restriction [20e], gives δ_i the value one if Δx_i is positive. The total transaction costs are discounted forward to the end of the period by a predefined discount rate ρ. The transaction costs are then incorporated by using the following restriction[21]:

[21a] $TC = (1 + \rho) \sum_{i=1}^{n} (\delta_i \eta + z_i \alpha)$

[21b] $\eta = \dfrac{\text{Fixed minimum fee}}{M_0}$

The adjusted expected return of the portfolio is then given by:

$$[22] \quad \sum_{i=1}^{n} x_i \mu_i - TC \geq \bar{\mu}$$

We will now discuss convexity and the inclusion of integer variables in portfolio problems.

3.7. Integer constraints and nonconvexity.

The efficient frontier in the MV model is for the standard assumptions imposed upon the model, such as no shortsale and upper and lower bound of the fraction to invest, smooth and convex in the expected return variance space.

To see this, we first define the function $f^*(\bar{\mu})$.

$$[23] \quad f^*(\bar{\mu}) = \min_{x} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \sigma_{ij} : \sum_{i=1}^{n} x_i \mu_i \geq \bar{\mu}, \, x \in C \right\}$$

Here C is a convex closed set, and we have for simplicity denoted the set of all investments **x**. The MV problem is a minimization of a convex function of a convex domain. In selecting two solutions x_a and x_b, weighted by α_a and $(1 - \alpha_a)$, the following condition must hold. From the convexity of the objective function and the convex restrictions it follows for a convex combination of two different f^*, with expected return $\bar{\mu}_a$ and $\bar{\mu}_b$, that:

$$[24] \quad \alpha_a f^*(\bar{\mu}_a) + (1 - \alpha_a) f^*(\bar{\mu}_b) \geq f^*(\alpha_a \bar{\mu}_a + (1 - \alpha_a) \bar{\mu}_b)$$

The condition [24] follows simply from the fact that the restrictions in the problem are convex and the objective function is convex, and a linear combination of two solutions is admissable.Thus, $f^*(\bar{\mu})$ is convex in $\bar{\mu}$.

The result can be visualized by looking in a figure. If there is a kink on the efficient frontier, this kink could be removed by forming a portfolio consisting of two other portfolios, such that this new portfolio dominates the portfolios, in a mean variance sense, where the kink occurs.

Expected return
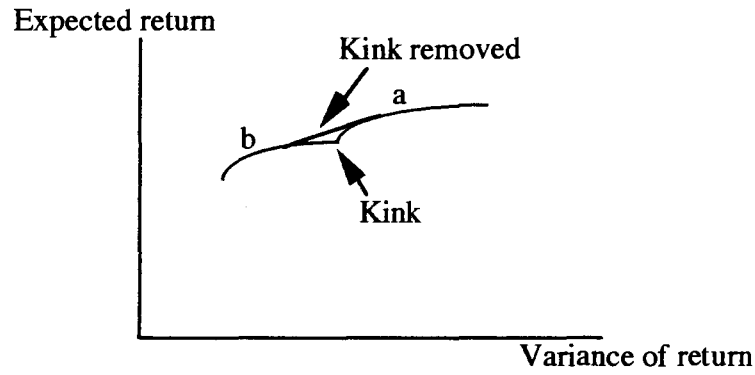
Figure 3.4. Kink in frontier.

In the figure above, we have the efficient frontier with a kink. By forming a portfolio of two other portfolios, i.e. a and b with weights α_a and $(1 - \alpha_a)$, the risk return relationship when varying α_a would connect these two portfolios by the line indicated above. This line would be a straight line if the correlation between the two portfolios is one. If the correlation is less, it would bend upwards. In all instances the possibility of constructing such a portfolio would remove the kink on the frontier.

Now, let us introduce integer constraints into the problem, and we define $f^*_\Lambda(\bar{\mu})$ as:

$$[25] \quad f^*_\Lambda(\bar{\mu}) = \min_{x} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \sigma_{ij} : \sum_{i=1}^{n} x_i \mu_i \geq \bar{\mu}, \, x \in C, \, \Lambda \in D \right\}.$$

We have here defined the set of integer variables as Λ, and the set formed by the integer constraints as D.

Theorem:

The function $f^*_\Lambda(\bar{\mu})$ is not necessarily convex in $\bar{\mu}$.

Proof:

In order to have convexity of the function $f^*_\Lambda(\bar{\mu})$ we must for an arbitrary selection of two portfolios a and b on the efficient frontier have:

$$[26] \quad \alpha_a f^*_\Lambda(\bar{\mu}_a) + (1 - \alpha_a)f^*_\Lambda(\bar{\mu}_b) \geq f^*_\Lambda(\alpha_a \bar{\mu}_a + (1-\alpha_a)\bar{\mu}_b) \quad \forall \, \mu_p, \, p \in [a, b]$$

Since we have integer constraints in the problem, the set formed by the constraint is no longer convex; thus linear combinations of the two solutions corresponding to $\bar{\mu}_a$ and $\bar{\mu}_b$, are not neccesary admissable. Then the relation [26] no longer holds.

☐

When we introduce integer variables into the model, the formation of the portfolios by combinations of a and b above may violate some of the integer constraints. An example of this is the interval constraint described earlier. Let us say that all assets have to be invested in fractions between 5 and 10 percent or set to zero. We assume that one particular asset in portfolio b has weight 5 percent and that its weight in portfolio a is zero. All combinations of a and b, except $\alpha_a = 0$ or $\alpha_a = 1$, would violate the constraints. It is possible to construct analog examples for the other integer constraints discussed in this chapter. This does not mean that kinks will occur when we introduce integer constraints into the problem, but that they may occur.

3.8. Two formulations tested on historical data.

In this section we will present two test problems which we solve using historical data from the Oslo Stock Exchange. We start with a presentation of the two problems[5].

Test problem one – basis constraints.

Assume now that the investor is not allowed to invest more than 10% of his total amount available in one particular asset. In addition, assume that the investor can only have four assets in his optimal portfolio with a market value between 5% and 10% of the value of his total portfolio. To put these restrictions into our model, the problem has to be restated with these new constraints. The constraints for upper bounds on investment in assets of 10% is easily handled by linear constraints. To incorporate the second constraint, we have to define a set of n new binary variables, which means variables only allowed to have the value 0 or 1.

With these new restrictions in mind, we now add the following restrictions to our

---

[5] Test problem number one is the restrictions imposed on Norwegian fund managers today.

problem[27]:

[27a] $x_i \leq 0.1$ $i \in 1,2,\ldots,n.$

[27b] $0.05(\delta_i+1) - x_i \geq 0$ $i \in 1,2,\ldots,n.$

[27c] $\sum_{i=1}^{n} \delta_i \leq 4$

[27d] $\delta_i \in Z_{0,1}$ $i \in 1,2,\ldots,n.$

Restriction [27a] assures that the maximum proportion allowed in each asset is 10%. Restriction [27b] lets the fraction invested vary between 0 and 10%. In addition, it has the following properties: if the fraction invested in asset i is less than 5%, δ_i is allowed a value of zero, if it lies between 5% and 10%, δ_i will have a value of 1. The properties mentioned here are assured by restriction [27d]. Restriction [27b] together with [27d] also ensures that no asset can have a value above 0.1 or 10%, so these two restrictions together make restriction [27a] redundant. In our problem we do not allow that more than four securities can have a value between 5% and 10%. Restriction [27d], which counts the number of securities with values between 5 and 10%, ensures this by restricting the sum of δ_i to be less or equal to 4. The original problem [1], with the new constraints formulated above, then solves our extended approximated mean variance problem.

Test problem two – Interval constraints.

In this problem we assume that the investor only invests in assets with weights between 5 and 10 percent. If it is not advantageous for the investor to invest in this range, the invested amount in one particular security is set to zero. In order to formulate such a problem, we must extend the original approximated problem [15] with the following restrictions[28]:

[28a] $x_i \leq 0.1\delta_i$ $i \in 1,2,\ldots,n.$
[28b] $x_i \geq 0.05\,\delta_i$ $i \in 1,2,\ldots,n.$
[28c] $\delta_i \in Z_{0,1}$ $i \in 1,2,\ldots,n.$

As indicated in Section 3.5., these constraints give us the desired result.

The problems discussed above are mixed integer linear programming problems. When we now do not have the possibility to rely on quadratic solution procedures alone, we have to discuss special solution procedures. If the universe consisted of 21 assets, we could calculate the maximum expected return possible to obtain under the restriction mentioned above. This is an easy task, since we simply would invest 10% in the four assets with highest expected return, and invest 5% in the securities ranking from fifth to 16'th. In test problem two, we would simply invest 10% in the ten assets with the highest expected return. Any other asset would have weight zero. The solution we obtain is a starting point for the efficient frontier generated by the problems we have formulated.

When it comes to the point where we want to minimize the variance of the portfolio subject to the restrictions above, the problem becomes more complicated. The problem is a mixed integer programming problem with an linear objective function. This problem is solvable by, for example, the branch and bound algorithm.

In our study we use 21 stocks listed on the Oslo Stock Exchange. The data ranges from the period of 1 Jan. 1987 to 31 Des. 1992 (72 monthly observations). Selection of the stocks was based upon several criteria. First, we avoided stocks with low liquidity, such that stocks traded in fewer than 150 trading days during a year, on average, are excluded from the set. Second the stock had to be on the trading list over the whole period. If mergers had taken place, we during the period adjusted for this by correcting with the switch ratio offered to the old stockholders in the two companies. For sectors with special events like the bank and insurance sector, we used the bank and insurance index instead. All return series were adjusted for splits, emissions and dividends to preserve the true return of each asset. The data was provided by the database Amadeus at the Norwegian School of Economics and Business Administration.

The analysis is performed by constructing an ASCII problem file readable by the mixed integer linear programming optimizer XPRESS-MP. The problem is run on the micro machine MOSES at The Norwegian Institute of Technology, Department of Economics. The operation environment of this machine is UNIX.

The overall problem is solved by iteratively solving the problem for different values of the expected return.

Results Ex.Post - Oslo Stock Exchange 1987-1992
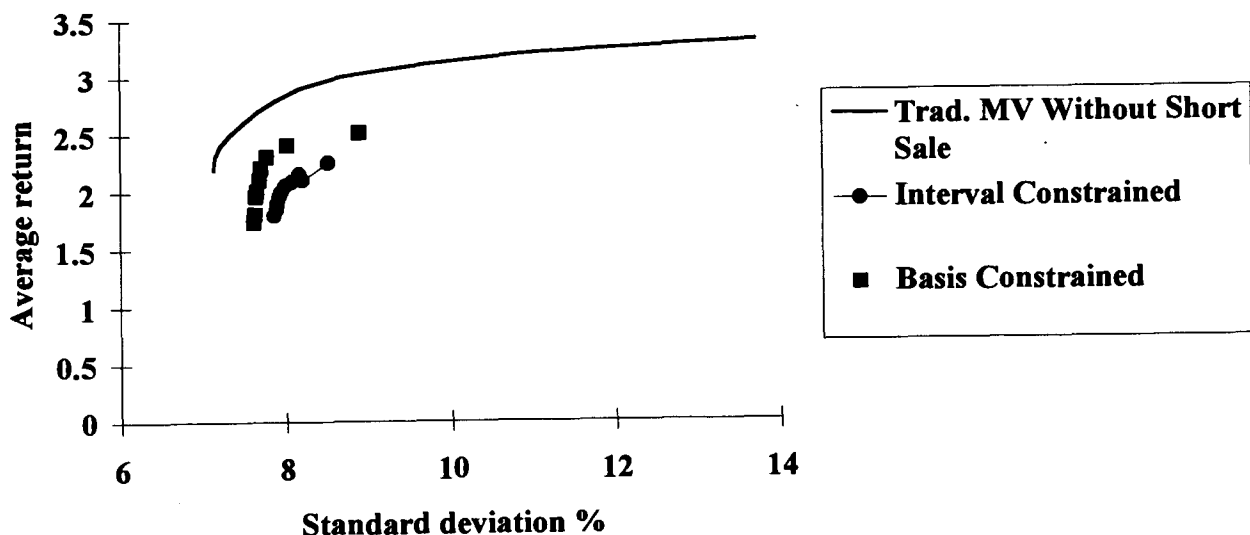
Fig. 3.5. Ex post frontiers

We have plotted the ex post efficient frontier under the set of restrictions given by [27] and [28]. The basis constrained portfolio seems to be nonconvex for large values in the estimated expected return. This is confirmed by forming a linear combination of the two neighbour portfolios, where a combination of these portfolios violates some of the integer constraints. Such observations are also the case for test problem two. The interval constrained portfolio is more efficient than the basis constrained portfolio. As compared with the no short sale allowed frontier, the new constrained frontiers are inside the efficient frontier corresponding to the traditional MV problem

3.9. Improving the binary search – test model I.

The procedure of finding an optimal solution, a minimum, to the mixed integer linear programming problems discussed earlier in this chapter relies on procedures such as the branch and bound algorithm. This procedure will find a solution if it exsists, but the efficiency of this algorithm depends upon the way the selection of branching variables is done[6]. Software packages for mixed integer programming are often able able to give some of the variables in the model preference in the branching process. A question which then arises is whether it is possible to use some apriori information such that we could give some variables preference, in order to obtain a optimal solution faster. In test model two

---

[6] For a introduction to the Branch and Bound algorithm, one could consult Minoux (1986).

this could mean to have apriori information of one or more of the variables δ_i. Another issue is whether it is possible to reformulate the problem, including new variables such that the search could be done more efficiently. Consider a reformulated version of equation [17b]:

[29] $\quad \sum_{i=1}^{n} \delta_i + \beta = k, \ \delta_i \in Z_{0,1}, \ \beta \in Z^+, i \in 1,2,...,n.$

Above we have reformulated equation [17b] by adding a slack variable, β, which could take integer values only. If we use restriction [29] instead of [17b], and give the variable β preference in the branch process by branching along nodes where β is set in the interval:

[30] $\quad \beta \in [0,1,2,....,k]$

we exclude the nodes where constraint [17b] is not fulfilled. We have illustrated this in the figure below.
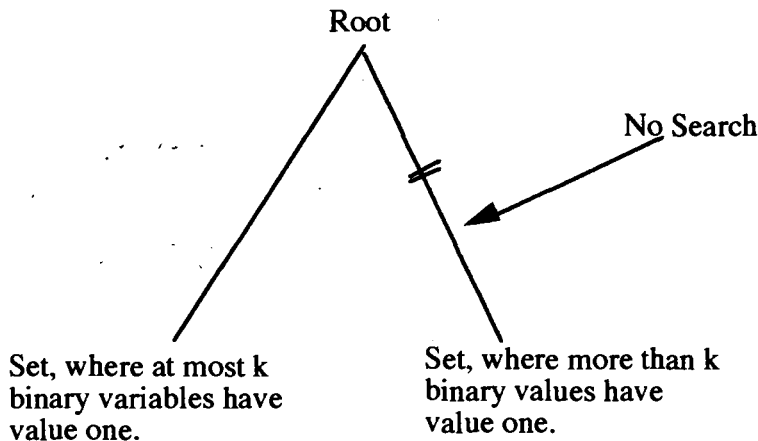
Figure 3.6. Partition of binary tree.

By using the reformulation above and by giving preference to the variable β in the branching process, we are likely to reduce the time necessary to solve the problem on a computer dramatically.

To give an illustration of the reduction in the tree by the use of the procedure above, we could compute the number of end leaves in the part of the tree where at most k binary variables are nonzero and compare the size with the size of the original binary tree. Although there is not necessary a direct relation between the time it takes to solve a mixed

binary programming problem by the use of the branch and bound algorithm and the number of leaves in the tree, this gives us a indication of the reduced complexity of the search.

The number of possible combinations, or end leaves in a tree of binary variables, is given by 2^n. The number of end leaves in the tree branched using the procedure described above is given by:

$$[31] \quad \gamma(n,k) = \sum_{i=1}^{k} \binom{n}{i} = \sum_{i=1}^{k} \frac{n!}{(n-i)!i!}$$

In the test case above, there are 21 binary variables and k is equal to 4 such that γ is equal to 7446, which is considerably smaller than 2^{21}.

3.10. Results – using data 1993–94.

We computed the means and standard deviation for the optimal portfolios generated earlier in this chapter using a dataset from the period 1993–94. In addition, we computed the ex post portfolio frontier using the same dataset. The results are illustrated in the figure below.
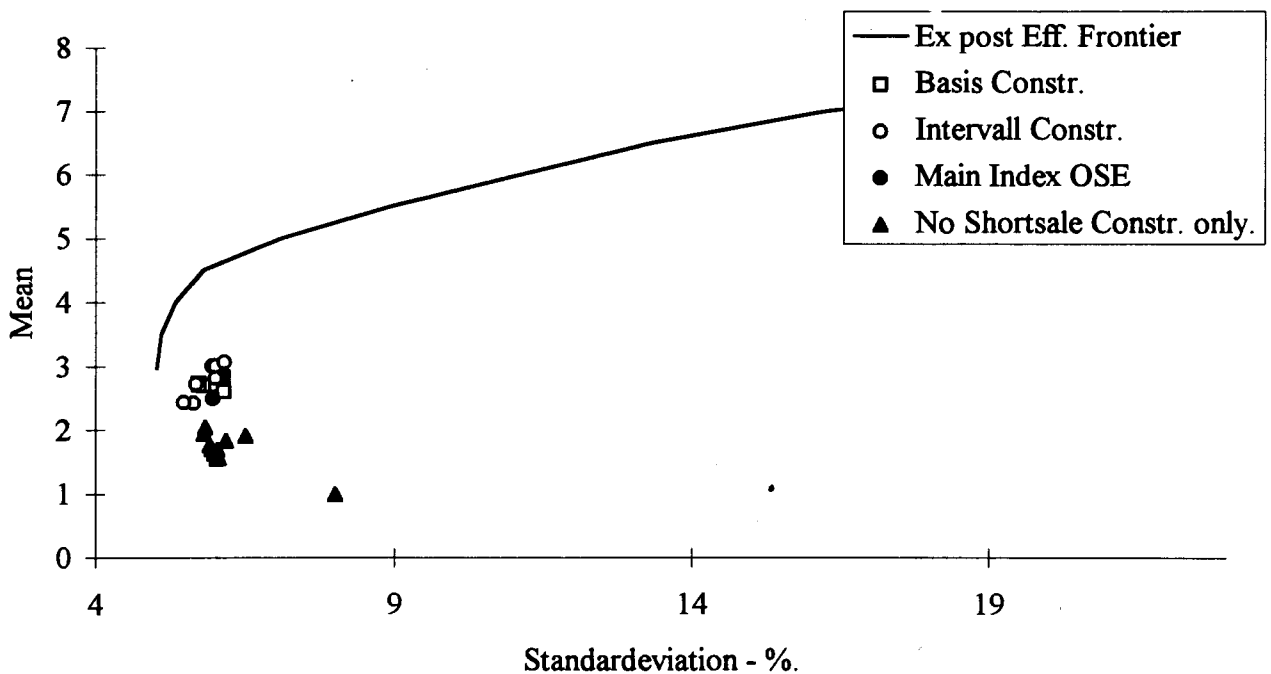

Fig. 3.7. The portfolios 1993–94.

The figure above indicates that the use of forced diversification as in basis constrained or interval constrained portfolio optimization gives us more efficient portfolios than the traditional model that uses no shortsale constraints only. This result was also pointed out in the paper of Chatry, Jennergren and Szala (1990) in their study of the French Stock market. The highly constrained portfolios lie in a cluster near the main index, although the composition of these portfolios can differ significantly from the composition of the main index. This indicates that highly diversified portfolios in the Oslo Stock Market tend to have the same risk as the main index.

3.11. Conclusions.

In this chapter we have discussed how one could incorporate discrete decision variables into a portfolio model. The incorporation of such variables into a portfolio model enlarges the set of possible restrictions one could formulate. We have discussed several constraints, among these, basis constraints, interval constraints and transaction costs. By an approximation of the quadratic problem we arrived at a mixed binary/integer linear programming problem. Using data from the Oslo Stock Exchange, we performed an optimization involving two test problems. One of these test problems was one that incorporated the regulation imposed on Norwegian fund managers today. The generated frontier is useful, as mentioned by Jennergren (1992), as a graphical tool for evaluating the performance of portfolios. We also discussed how we could increase the efficiency of the solution procedure for one of these test problems. Finally, we performed an ex. post analysis using data from 1993–94, where we evaluated the portfolios generated from the models. The analysis indicated that portfolios generated by forced diversification had a better risk return performance than the portfolios generated by the traditional Markowitz model.

References:

Chatry, P., Jennergren, L.P., Szala, A. (1990) "Two Investigations of Portfolio Selection and Mean – Standard Deviation Performance Evaluation in the French Capital Market: A Research Note." EFI – Stocholm School of Economics.

Jennergren, L.P. (1992), "Ex Post Efficiency and Mutual Fund Evaluation." OMEGA International Journal of Management Science." Vol. 20, pp. 249–255.

Lemke, C.E. (1965) " Bimatrix Eqilibrium Points and Mathematical Programming." Management Science 11 pp. 681–689.

Lintner, J. (1965). "The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." Review of Economics and Statistics. 47: pp. 13–37

Markowitz, H.M. (1952) Portfolio Selection. Journal of Finance. pp 77-91.

Markowitz, H.M. (1959) Portfolio Selection. (John Wiley & Sons NY)

Markowitz, H.M. (1987) Mean Variance Analysis in Portfolio Choice and Capital markets (Basil Blackwell)

Minoux,M. (1986), Mathematical Programming, Theory and algorithms. John Wiley Publications.

Mossin, J. (1965) "Equilibrium in a Capital Asset Market." Econometrica vol 34 pp 768-783.

Sharpe, W. (1964) "Capital Asset Prices: A Theory of Capital Equilibrum under Conditions of Risk." Journal of Finance 19: pp.425–442.

4. Three Dimensional Portfolio Analysis.

Abstract: This chapter studies a three dimensional portfolio optimization model where in addition to the expected return and standard deviation, the investor takes the third central moment into account. We propose an approximate problem to solve the optimization problem which arises from such an analysis, and perform an analysis of the model using historical data from the Oslo Stock Exchange.

4.1. Introduction.

Historically, in classical portfolio theory, the parameters the investor uses for selecting optimal portfolios are the expected return of the portfolio, and the variance or standard deviation of the portfolio's return. This model has its orgins in the classical mean variance model of Markowitz (1952). In the literature there has been some work evaluating the impact on the portfolio selection problem by taking the third central moment into account. Arditti (1969) carried out an empirical investigation of asset's return involving the third central moment. Jean (1971), (1973) investigated the problem further by discussing the optimal asset mixes in models involving the third central moment. Levy (1969) presented a cubic utility function, which implies that the investor's preference ordering is completely determined by the expected return, variance of return and the third moment of return. Kraus and Linzenberger (1976) derived a capital asset pricing model in the CAPM framework where the third moment was also considered. Some preliminary work on this model was also done by Jean (1973). Empirical testing of this model was done by Friend and Westerfield (1980). Konno, Shirakawa and Yamazaki (1993) developed an optimization model which was based upon a linearization of the risk parameter's standard deviation and skewness.

In Section 4.2. we will define the third central moment and discuss the relevance of incorporationg the third central moment into the portfolio analysis. In Section 4.3. we will review some of the theory concerning the third central moment in relation to expected utility maximization. In Section 4.4. we outline the general description of the model. In Section 4.5. we discuss the general optimization problem and the assumption made on the preferences. In the literature, a version of the model was presented by Konno and Yamasaki (1993). Konno's approach is discussed in Section 4.6. Section 4.7. presents the discrete version of the model, which is modified to give us the empirical version in Section 4.8. The empirical model is solved by an approximation of the nonlinear objective function. How this approximation is done is presented in Section 4.9. The approximated problem is tested ex post in Section 4.10. using historical data from the Oslo Stock

Exchange. In Section 4.11. we discuss the results obtained from the analysis in section 4.10. in the light of some commonly used utility functions. Finally, a comparison between the model presented in this chapter and the model presented by Konno (1993) is done in Section 4.12., and in Section 4.13. we draw some conclusions.

4.2. Definition of the third central moment.

The third central moment is defined by:

[1] $\quad \gamma_p = E((\tilde{r}_p - \mu_p)^3)$

It is also possible to standardize the third moment several ways, i.e. by m_3^1 or m_3^2 [2]:

[2a] $\quad m_3^1 = \gamma_p^{\frac{1}{3}}$

[2b] $\quad m_3^2 = \dfrac{\gamma_p}{\sigma_p^3}$

Here μ_p is the expected return of the portfolio and σ_p is the standard deviation of the portfolio.

Another way of expressing the third moment of a portfolio is given by:

[3] $\quad \gamma_p = E((\tilde{r}_p - \mu_p)^2(\tilde{r}_p - \mu_p))$

Practically, there are several objections to involving the third moment in an analysis. One important issue is that the estimated third moment is dependent on the length of the interval between the observations. Hawawini (1980) studied this, and derived the effect of the time difference between the observations on the estimated third moment. Investigations of stationarity of the skewness of asset price returns, such as by Singleton and Wingender (1986), also tend to conclude that the estimated third moment is not stable over time.

Another aspect of portfolio analysis is leverage with limited responsibility for the investor. The returns of the portfolio may then have a considerable skewness and a third moment analysis may be adequate. Finally, the introduction of derivative securities as options into a portfolio may, in undiversified portfolios, give us a portfolio with nonsymmetric returns.

These advantages and disadvantages are acknowledged in this paper, and imply that asset allocation with the third moment is not so easy a task. We will not consider capital market theory or the introduction of a risk free asset into the model. In this chapter we only consider the optimization procedure related to portfolio problems involving the third central moment.

Computationally, portfolio optimization involving the third moment may cause a problem for us since we have an optimization problem with an objective function that is concave in some sets and concave in anothers. Common techniques used in convex optimization are no longer directly usable. Jean (1973) pointed out that using the third moment around zero may make the problem convex, but we regard that as a separate problem. The intention of this paper is to present a solution technique for financial optimization problems involving the third central moment. The method uses an linear approximation scheme for the nonlinear functions. The approximation of the nonconvex function is supported with the use of special ordered set in accordance with Beale and Thomlin's (1969) special ordered sets II (SOS2) principle.

4.3. Preferences and the third central moment.

As mentioned in the introduction of this dissertation, we can define a utility function U(W) to represent the preferences of the investor.

If we make a Taylor expansion of the utility function around the expected value of wealth, and take the expectation, we obtain[4]:

[4a] $\quad E(U(W)) = U(E(W)) + \dfrac{\partial^2 U(E(W))}{\partial W} \dfrac{1}{2} E((W - E(W))^2) +$

$\quad \dfrac{\partial^3 U(E(W))}{\partial W} \dfrac{1}{6} E((W - E(W))^3) + R$

[4b] $\quad R = \displaystyle\sum_{i=4}^{\infty} \dfrac{\partial^i U(E(W))}{\partial W} \dfrac{1}{i!} E((W - E(W))^i)$

We see that the third moment plays a role in the determination of expected utility. Loistl (1976) argued that the Taylor expansion performed above may lead to errors of non−negligable size, and that such a series should be handled with care. In our own experience with the expansion for the power utility function, (see Table 4.1. for a list of the utility functions), with α considerably negative, the errors could be very large. For

approximation of the most common utility functions with reasonable parameters, the Taylor expansion [4a] seems adequate.

The debate on this issue was initiated by the article by Tsiang (1972), who argued that for practical purposes, moments of higher order than the variance have relatively little impact on expected utility. Among the opponents to this view were Levy (1974) and Borch (1974). Kroll, Levy and Markowitz (1984) investigated utility maximization by use of the mean variance approach. In their analysis the mean variance model seems to be a good approximation, when they used actual stock returns.

Levy and Hanoch (1970) and Levy (1969) presented a cubic utility function of the form:

[5] $U(W) = W + bW^2 + cW^3$

Above b and c are constants. This is a polynomial of the third degree and therefore is not concave for all values of W. If we put restrictions on the parameters a and b, it can be shown that the expected utility is a function of expected wealth, variance of wealth and the third central moment of wealth.

Scott and Horwath (1980) investigated the relation between preferences and the different moments of the underlying distribution. They presented results concerning the different directions on expected utility the different moments could have. Among their results was the following theorem:

Theorem (Scott and Horwath)[6]:

If $U^1(W) > 0$ $\forall$ W and $U^2(W) < 0$ $\forall$ W and the decision maker is strictly consistent in preference direction, viz:

[6a] $\begin{cases} U^n(W) > 0 \ \forall \ W \ \text{or,} \\ U^n(W) = 0 \ \forall \ W \ \text{or,} \\ U^n(W) < 0 \ \forall \ W \end{cases}$

then the decision maker will have positive preference for positive skewness.

Proof[7]:

We assume the contradictory $U^3(W) < 0$ or $U^3(W) = 0$

Lagrange's mean value theorem states that for two wealth levels, W_1 and $W_2 \; \exists \; \overline{\overline{W}} \in (W_1, W_2)$ such that:

[7a] $U^1(W_2) - U^1(W_1) = U^2(\overline{\overline{W}})(W_2 - W_1) \leftrightarrow$

[7b] $U^1(W_2) = U^1(W_1) + U^2(\overline{\overline{W}})(W_2 - W_1)$

since we have assumed that $U^3(W) < 0$ or $U^3(W) = 0$, it follows that $U^2(W_1) \geq U^2(\overline{\overline{W}})$.

Then by [7b] it follows that:

[7c] $U^1(W_2) \leq U^1(W_1) + U^2(W_1)(W_2 - W_1)$

If we evaluate W_2 for the values:

[7d] $W_2 = W_1 + \dfrac{U^1(W_1)}{-U^2(W_1)} + C$

where $C \geq 0$, we will have, by the use of [7c], that:

[7e] $U^1(W_2) \leq C U^2(W_1) \leq 0$

If $C \geq 0$, and $U^2(W_1) \leq 0$, we will have $U^1(W_2) < 0$

This contradicts the assumption $U^1(W) > 0$ thus $U^3(W) > 0$.

☐

The derivation of expected utility by the expected moments above indicates that the sign of $U^3(W)$ determines the direction of the preferences for the third central moment. If $U^3(W)$ is positive, a large third central moment is preferred ceteris paribus. If $U^3(W)$ is negative, the decision maker prefers a lower third central moment ceteris paribus. If $U^3(W)$ is nil, the decision maker's expected utility is not affected by the third central moment. The theorem of Scott and Horwath (1980) means that preference for the third central moment is a relevant assumption to impose on models which incorporate this moment.

Below is a table with the most common utility functions, which allows us to investigate the

effect on expected utility of the third central moment. A, a, b, c are positive constants, and the constant $\alpha \in (-\infty, 1)$. The constant $\lambda > 0$.

Utility Func.	U(W)	$U^3(W)$	Sign
Neg. Exp.	$A(1 - e^{-\lambda W})$	$A\lambda^3 e^{-\lambda W}$	Pos
Logarithmic	ALn(W)	$\dfrac{A}{W^3}$	Pos.
Quadratic	$aW - \dfrac{b}{2}W^2$	0	0
Power	$\dfrac{1}{\alpha}W^\alpha$	$(\alpha-2)(\alpha - 1)W^{\alpha - 3}$	Pos.
Cubic	$W + bW^2 + cW^3$	6c	Sgn(c)

Table 4.1. Common utility functions and the third central moment.

The table above indicates that the most commonly used utility functions imply a preference for the third central moment. In the optimization model developed in the next sections, we will assume that the investor has preferences for the third central moment.

4.4. The general information structure in the model.

In this section we will describe the general information structure in the model. The general model assumes that the decision maker makes decisions on the investments to be made. After this selection of assets is completed, the investor observes a return on his portfolio.

The first interpretation is a model that assigns a discrete probability distribution to a set of possible states. In general, the model assumes that n assets have T possible outcomes or events ω_t. We have indexed the T outcomes by t. The random variables then follow the following filtering process[8]:

[8a] $f_0 = \{\omega_1, \omega_2, ..., \omega_T\}$

[8b] $f_1 = \{\{\omega_1\}, \{\omega_2\}, ..., \{\omega_T\}\}$

Each state ω_t occurs with probability P_t. We could illustrate the process by the following figure:
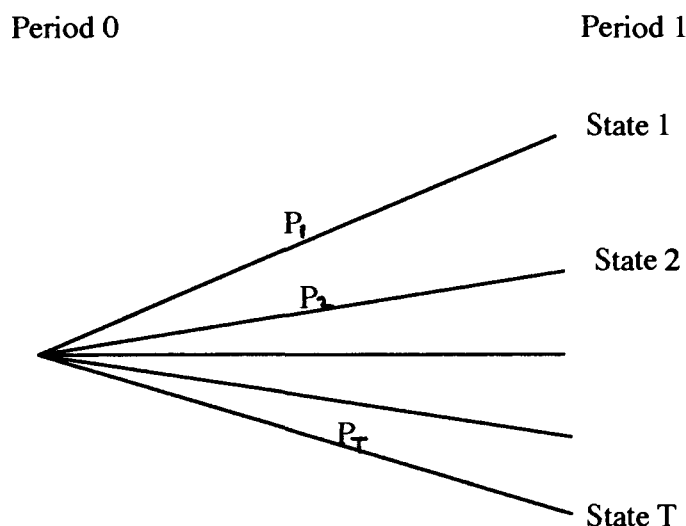
Fig. 4.1. General model and states.

The states are connected to period 0 by an arch. Each state is assigned a probability. In a particular state, an outcome of the n random variables is realized consisting of the return or value of the assets.

The model above is general in the sense that it allows the user of the model to specify the outcomes in each state and a probability, or more generally, a weight associated with that state. These parameters could be selected on the basis of purely subjective grounds, or the user is given them by a "neutral" agent.

We could also use historical data, which leads us to the second interpretation of the model. If we use historical data, a reasonable value to give to the probabilities or weights for each state is a probability such that the expected value of the different moments of the random variables is in harmony with the common estimators of that moment. An example of such probabilities or weights is 1/T and 1/(T-1). In such a case, the expected moments of, for example a portfolio of assets constitute the traditional estimator of the moment for that portfolio. In the analysis of the model in this chapter, we will use historical data and the second interpretation of the model.

4.5. Portfolios and the third moment.

In this section we will list some results concerning calculations of the third moment of a portfolio. We define the cross third moment between the assets i,j,k as:

[9] $E((r_i - E(r_i))(r_j - E(r_j))(r_k - E(r_k))) = \zeta_{ijk}$

We note that if i=j=k, ζ_{ijk} is the third central moment of asset i. If we have n assets the total number of ζ_{ijk} is equal to n^3. The third central moment of a portfolio is given by:

[10] $\gamma_p^3 = E((r_p - E(r_p))^3) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} x_i x_j x_k \zeta_{ijk}$

If we want an estimate of the portfolio´s third moment, we have to estimate n^3 parameters, and calculate the value of [10] above. The number of estimated parameters increases sharply as a function of n, and if we have 1000 assets for example, we have to estimate 1 billion parameters, a number of parameters that, even with today´s technology, is difficult to calculate.

In this chapter we will look at portfolio analysis with three moments, we will assume that the expected return and the third central moment are characteristics that the investor has preferences for. We will also assume that the investor dislikes variance of return.

The expected utility to the investor is then assumed to be a function of the expected return, variance of return and the third central moment of return for the portfolio[11].

[11a] $E(U) = f(\mu_p, \sigma_p, \gamma_p)$

[11b] $\dfrac{\partial f}{\partial \mu_p} > 0, \quad \dfrac{\partial f}{\partial \sigma_p} < 0, \quad \dfrac{\partial f}{\partial \gamma_p} > 0.$

The above assumption on the partial derivatives of the utility function with respect to the different moments is well suited for example when the utility function representing the preferences to the decision maker is logarithmic or negative exponential. The higher moments than the moments of the third degree are also persistent in an examination of expected utility for the two utility functions mentioned earlier, but one could assume that their influence on the expected utility is neglictable, although the model presented could

also be modified to take those moments into account.

We will calculate an efficient set of possible investments. In our terminology we will then refer to a portfolio as efficient if:

i) For any joint level of expected return and variance, it is not possible to obtain a higher third central moment without increasing the variance of the portfolio or decreasing the expected return of the portfolio.

With this definition in mind, the set of $(\mu_p, \sigma_p^2, \gamma_p)$ combinations defines a set of investments from which the decision maker is able to select a portfolio.

To obtain the set of efficient portfolios, we could solve an optimization problem given by[12]:

$$[12a] \quad \text{maximize} \quad \left\{ \gamma_p^3 = E((r_p - E(r_p))^3) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} x_i x_j x_k \zeta_{ijk} \right\}$$

subject to:

$$[12b] \quad \mu_p = \sum_{i=1}^{n} x_i \mu_i \geq \bar{\mu}$$

$$[12c] \quad \sigma_p^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \sigma_{ij} \leq \bar{\sigma}^2$$

$$[12d] \quad \sum_{i=1}^{n} x_i = 1$$

$$[12e] \quad x_i \geq 0 \quad i,j,k \in 1,2,\ldots\ldots n.$$

Here $\bar{\mu}$ is some predefined level of the expected return, and $\bar{\sigma}^2$ is a predefined level of the variance.

If we compare this optimization problem with the traditional MV problem, we see that the problem has a nonconvex objective function. Thus, this problem is impossible to solve

with traditional convex optimization techniques. In the literature there have been arguments about whether the problem above is convex for a proper selection of the centrality of the moments. Jean (1973) argued that if one analyses the third moment around zero one will have a problem that is concave and solvable. From a mathematical programming point of view, the model is clearly nonconvex, even if the nonconvexity is located around zero. Hence the arguments by Jean (1973) are questionable. Thus we will deal with a nonconvex optimization problem and global optimization techniques have to be considered.

4.6. Konno's approximation to a linear program.

Konno and Yamazaki (1993) present an optimization model that finds a suitable portfolio in the mean variance skewness space by formulating a linear programming problem. The problem approximates the actual moments by use of linear functions. The first moment is not approximated, so here one uses the mean of the portfolio's return. The dispersion is represented by using the mean absolute deviation. So the second (dispersion) moment is given by:

[13] $\upsilon_p = E\{ |\tilde{r}_p - \mu_p| \}$

The third moment is approximated by considering the lower semi third moment. The definition of the lower semi third moment, γ_p^- [14] is:

[14a] $\gamma_p^- = E((\min(0, \tilde{r}_p - \mu_p))^3)$

Further, Konno approximates the lower semi third moment by a piecewise linear function. The approximation could be explained by the use of the figure below where we have illustrated the functions in which expectations are used as approximations of the third central moment in the Konno and Yamazaki (1993) model.

Different moments

Original third moment

$\rho_2 \ \rho_1$

Return - expected return

Lower semi third moment

Approximated lower semi third moment
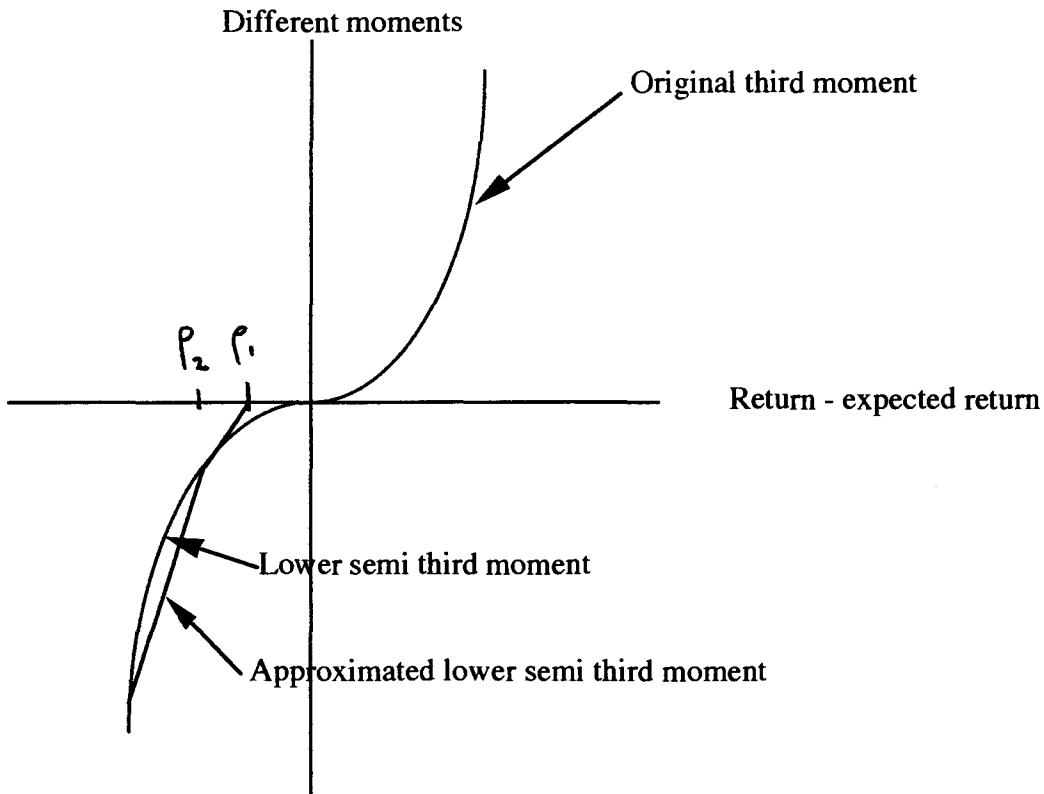
Fig. 4.2.   Third moment – shape and lower third central moment.

The center in the figure, where the two axes cross, is origo. Along the horizontal axis we have the return of the portfolio minus the expected return of the portfolio. If we take the expectation of the cube of this value we will have the third moment of the portfolio. So the third central moment is the expected value of the function appearing in the first and third quadrant in the figure above. The part of this curve lying in the southwest quadrant defines the semi lower partial third moment and the expected value of this function is the lower partial third moment. The piecewise linear function in the southwest quadrant is the function which Konno uses as the approximated third moment. This approximation will be discussed below.

We refer to Chapter 2 or Konno and Yamazaki (1991) for information about the risk measure absolute deviation. We will look at the formulation that is directly related to the third moment approximation.

The analytical form of the linear piecewise approximation is [15]:

[15a] $G(\tilde{r}_p - \mu_p) = - \left| \tilde{r}_p - \mu_p - \rho_1 \right| - \alpha \left| \tilde{r}_p - \mu_p - \rho_2 \right|$ $\rho_2 < \rho_1 < 0$

[15b] $|u|^- = 0$ if $u \geq 0$ and $-u$ if $u < 0$.

Below, we have a figure with the G–function used by Konno and Yamazaki (1993).
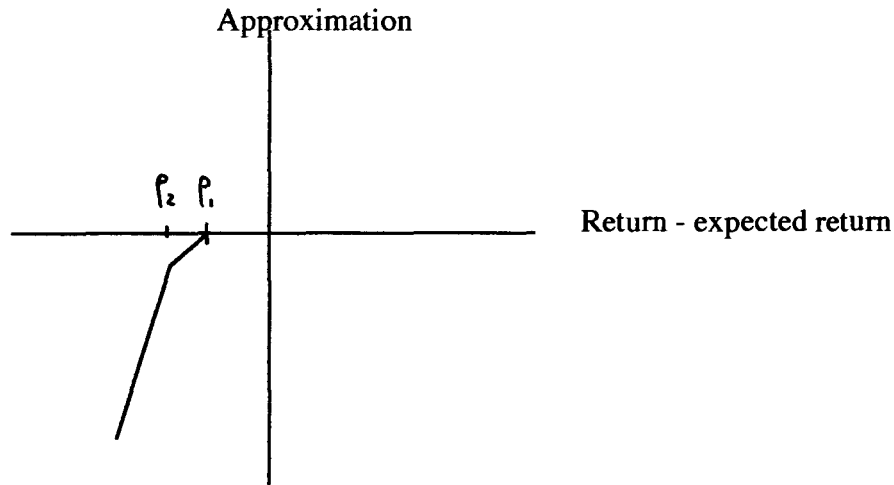

Fig. 4.3. Konno´s G – Function.

By inspection we see that the function G gives a piecewise linear function of the type shown in the figure above. If we move to the left from origo, the function first has a value 0 until we reach ρ_1, then the slope of the function will be 1 until we reach ρ_2, where the slope will be $1 + \alpha$.

The approximation done by Konno, presented above, may seem very rough. First he uses the absolute deviation of the portfolio as a surrogate for the variance of the portfolio. This risk measure is an alternative to the variance. As an approximation it is very rough. Second, he uses the piecewise linear function above as an approximation of the skewness of the portfolio. In his article he advocates minimizing the G function as not if the same as, but nearly the same as maximizing the third moment of the portfolio. How good this approximation is depends upon the selection of α, and the actual data used in the analysis. How the measure distinguishes between low variance portfolios and high third central moment portfolios is unclear. It may seem that adding all these approximations together may invalidate the model. The advantage of the model is that it is a purely linear programming problem where easily usable algorithms and software packages are available.

The formulation of the Konno mean - risk - skewness model will, with estimated values, then be [16]:

[16a]   minimize $\dfrac{1}{T-1}\left(\displaystyle\sum_{t=1}^{T}u_t + \alpha\displaystyle\sum_{t=1}^{t}v_t\right)$

subject to:

[16b] $u_t + \displaystyle\sum_{i=1}^{n}r_{it}x_i \geq \rho_1 \quad t\in 1,2,...,T.$

[16c] $v_t + \displaystyle\sum_{i=1}^{n}r_{it}x_i \geq \rho_2 \quad t\in 1,2,...,T.$

[16d] $y_t - \displaystyle\sum_{i=1}^{n}a_{it}x_i \geq 0 \quad t\in 1,2,....T.$

[16e] $y_t + \displaystyle\sum_{i=1}^{n}a_{it}x_i \geq 0 \quad t\in 1,2,....T.$

[16f] $\dfrac{1}{T-1}\displaystyle\sum_{t=1}^{T}y_t \leq \bar{\upsilon}$

[16g] $\displaystyle\sum_{i=1}^{n}x_i\hat{\mu}_i \geq \bar{\mu}$

[16h] $\displaystyle\sum_{i=1}^{n}x_i = 1$

$x_i \geq 0. \quad i\in 1,2,...,n.$

$y_t \geq 0, \ v_t \geq 0, \ u_t \geq 0. \quad t\in 1,2,....T.$

The objective function minimizes the approximated semi lower third moment. The restrictions [16f] and [16g] define an upper bound for the absolute deviation for the portfolio and a lower bound for the expected return of the portfolio. The restrictions [16c] and [16b] give values to the approximated semi lower third moment in the objective function. In the next section we will discuss the three moment portfolio model in a discrete distribution framework. We will not discuss the optimization technique for this case, but we do that for the empirical version of the model presented in the following section.

4.7. The model revisited using a discrete distribution.

If we assume a discrete distribution, the different central moments of the return of the assets and portfolio are known. The moments are given by[17]:

[17a] $$\mu_i = \sum_{t=1}^{T} P_t r_{it}$$

[17b] $$\mu_p = \sum_{t=1}^{T} P_t \sum_{i=1}^{n} x_i r_{it}$$

[17c] $$\sigma_p^2 = \sum_{t=1}^{T} P_t \left(\sum_{i=1}^{n} x_i (r_{it} - \mu_i) \right)^2$$

[17d] $$\gamma_p = \sum_{t=1}^{T} P_t \left(\sum_{i=1}^{n} x_i (r_{it} - \mu_i) \right)^3$$

We have above defined a_{it} as $a_{it} = r_{it} - \mu_i$, conversely $a_{pt} = r_{pt} - \mu_p$. The variable r_{pt} is portfolio return corresponding to event t.

The formulation of the model [12] in the discrete distribution case will then be[18]:

[18a] $$\max \sum_{t=1}^{T} P_t a_{pt}^3$$

subject to:

[18b] $$\sum_{i=1}^{n} x_i a_{it} = a_{pt} \quad t \in 1,2,\ldots,T.$$

[18c] $$\sum_{t=1}^{T} P_t a_{pt}^2 \le \bar{\sigma}_p^2$$

[18d] $$\sum_{t=1}^{T} P_t \sum_{i=1}^{n} x_i r_{it} \ge \bar{\mu}_p$$

[18e] $\displaystyle\sum_{i=1}^{n} x_i = 1$

[18f] $x_i \geq 0 \quad i \in 1,2,...,n.$

The model above does not use estimators; it uses the computable actual central moments, σ_p^2 and γ_p, and the first moment μ_p, of the distribution. We will not make any closer investigation of the model with discrete probabilities here, but we note that the optimization tools used on the upcoming empirical version can be used on the model above as well.

4.8. The principles of the model using historical data.

In a portfolio model, we often use historical data to estimate the parameters in the model. We will do the problem formulation in a direct way, which means that we do not have to preprocess the data by calculating a covariance matrix or cross third central moments. We define a_{it} to be $r_{it} - \hat{\mu}_i$, where $\hat{\mu}_i$ is the estimated expectation of the random return for asset i, and r_{it} is now the realized return of asset i in period t. We also have that:

[19] $\displaystyle \frac{\sum_{t=1}^{T} a_{it}^2}{T-1} = \hat{\sigma}_i^2$

where $\hat{\sigma}_i^2$ is the traditional estimator of assets i's variance.

We can now define a_{pt} as:

[20] $\displaystyle a_{pt} = \sum_{i=1}^{n} x_i a_{it}$

Thus, to obtain the estimated variance for a portfolio we can calculate this as:

[21] $\displaystyle \hat{\sigma}_p^2 = \frac{1}{T-1}\sum_{t=1}^{T}\left[\sum_{i=1}^{n} x_i a_{it}\right]^2 = \frac{1}{T-1}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j a_{it} a_{jt} =$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \sum_{t=1}^{T} \frac{a_{it}a_{jt}}{T-1} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \hat{\sigma}_{ij}$$

Thus we can obtain a portfolio's estimated variance as the sum from t=1 to T of the a_{pt}^2, such that:

$$[22] \quad \frac{1}{T-1}\sum_{t=1}^{T} a_{pt}^2 = \hat{\sigma}_p^2$$

We can continue in the previous fashion with the third central moment[23]:

$$[23a] \quad \frac{1}{T-1}\sum_{t=1}^{T} a_{pt}^3 = \frac{1}{T-1}\sum_{t=1}^{T}\left[\sum_{i=1}^{n} x_i a_{it}\right]^3 = \frac{1}{T-1}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} x_i x_j x_k a_{it} a_{jt} a_{kt} =$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} x_i x_j x_k \sum_{t=1}^{T} \frac{a_{it}a_{jt}a_{kt}}{T-1} = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} x_i x_j x_k \hat{\zeta}_{ijk} = \hat{\gamma}_p^3$$

From the expression above, we see that the second and third moment can be calculated directly by use of the sums from t=1 to T of the square of, and the cube of, a_{pt}. Following is a formulation of our optimization problem, which be viewed as the starting point of our portfolio analysis[24]:

$$[24a] \quad \text{maximize } \hat{\gamma}_p^3 = \frac{1}{T-1}\sum_{t=1}^{T} a_{pt}^3$$

subject to:

$$[24b] \quad \frac{1}{T-1}\sum_{t=1}^{T} a_{pt}^2 \leq \bar{\sigma}_p^2$$

$$[24c] \quad \sum_{i=1}^{n} x_i \hat{\mu}_i \geq \bar{\mu}$$

[24d] $\displaystyle\sum_{i=1}^{n} x_i = 1$

[24e] $\displaystyle\sum_{i=1}^{n} a_{it}x_i = a_{pt} \quad t \in 1,2,.....T.$

[24f] $x_i \geq 0 \quad i \in 1,2,...n.$

We now have an equivalent program to that of [12]. But it differs in some respects. First, we do not have to calculate the n^3 cross third central moments. Or do we have to compute the estimated covariance matrix. The only estimators that we need to calculate are the n expected return estimators. In one respect the problem is still equal to [12]. The problem is still nonconvex.

4.9. Approximation schemes.

In order to overcome the problem of nonconvexity of the optimization problem [24] we first have to approximate the nonlinear functions involved by using linear functions. There are two possible approaches to use when we approximate a nonlinear function. Tangential approximation, as opposed to barycentric approximation in general, replaces the nonlinear function f(x) in the neighbourhood $x = x^0$ with:

[25] $\hat{f}(x) = f(x^0) + \nabla f^T (x - x^0)$

Here, ∇f^T is the gradient of the objective function. Since we will not use tangential approximation, we are not going to discuss it further here.

We have in our analysis used barycentric approximation, which approximates the nonlinear function by using linear segments, connected by points lying on the original nonlinear function. We will now have a short introduction to the general principle of this approximation. Let the general programming problem be [26]:

[26a] min f(z)

subject to:

[26b] $z \in C$

Here z is a vector or a single variable and C is the set formed by the constraints to the

problem. If z is a vector with T elements indexed by t, and if we for simplicity assume that f(z) could be written in a separable form, we have:

$$[27] \quad f(z) = \sum_{t=1}^{T} f_t(z_t)$$

The method requires a set of meshpoints, S^{jt}, ordered in increasing order for each t, and associated weights w^{jt}. The points S^{jt} are selected such that all possible values of z_t lie inside the interval:

$$[28] \quad \left\lceil \min_j(S^{jt}), \; \max_j(S^{jt}) \right\rceil$$

This condition is imposed on each t. The number of meshpoints could in general differ for each t. We denote the number of meshpoints for each t by m_t. The selection of the values S^{jt} depends upon the problem we have under investigation. The approximated problem is given by [29]:

$$[29a] \quad \text{minimize} \; \sum_{t=1}^{T} \left\{ \sum_{j=1}^{m_t} w^{jt} f(S^{jt}) \right\}$$

subject to:

$$[29b] \quad \sum_{j=1}^{m_t} w^{jt} S^{jt} = z_t$$

$$[29c] \quad \sum_{j=1}^{m_t} w^{jt} = 1 \quad t \in 1,2,...,T.$$

$$[29d] \quad w^{jt} \geq 0 \; j \in 1,2,...,m. \quad t \in 1,2,...,T.$$

$$[29e] \quad x \in C$$

The key issue in this program is the linear weighting of different values of the objective function by the use of the points S^{jt}. The constraint [29b] is the link between the values of z_t and the approximation used in the objective function. The weights w_{jt} are, for one particular value of z_t, given values such that the equality [29b] holds. If the nonlinear objective function in [29] is convex, the weights would satisfy the properties described below.

In order to assign two properties to these weights w^{jt}, we have to define the concept of

neighbour as follows. Two weights w^{jt} and w^{j^*t} are neighbours if $j = 1$ and $j^* = 2$ or, $j = m_t - 1$ and $j^* = m_t$ or $j^* = j + 1$ or $j^* = j - 1$. The second property is [30]:

$$[30a] \qquad \sum_{j=1}^{m_t} w^{jt} = 1 \quad \forall\, t$$

These two conditions are imposed on the vectors w^t, one for each t, formed by the elements w^{jt}. This simply means that in each of these vectors, at most two neighbouring elements could be nonzero, and for each t, the sum of the weights should be equal to one.

We will now turn back to our optimization problem and illustrate the approximation for our nonlinear functions. If the problem has a separable objective function, the approximation is easier to do compared to a model where the objective function is nonseparable. To start with, we will consider the approximation of a_{pt}^2, where the sum of these variables over t is equal to $(T-1)\hat{\sigma}_p^2$.

This is illustrated for one particular value of a_{pt} in the figure below.
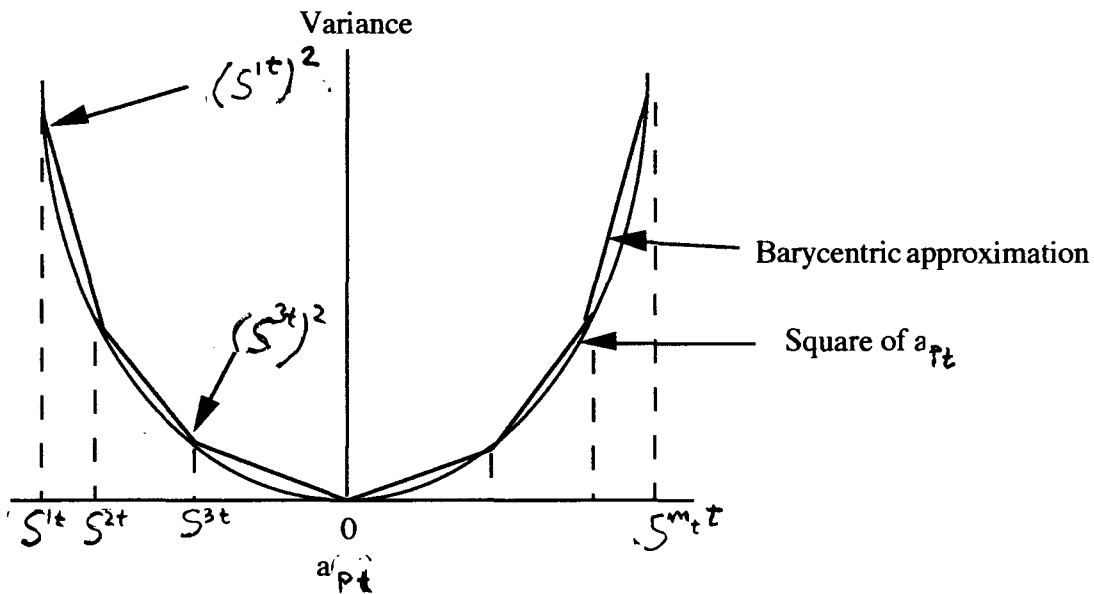
Figure 4.4. Approximation of the Variance of the portfolio.

On the horizontal axis we have the value a_{pt}. Along the vertical axis we have the square of a_{pt}. The smooth line represents the quadratic function or a_{pt}^2 approximated. The line segments show the linear approximation. The approximation is done by selecting m_t mesh

points $S^{1t}, S^{2t}, \ldots, S^{m_t t}$, computing the functional value of these points and connecting these points by linear segments. The functional value that replaces the value of a_{pt} is given by the piecewise linear function in the figure. In our problem this function is of a type such that property 1 is satisfied.

Property 1 is not automatically satisfied for the next nonlinear function. This function is a_{pt}^3. To comply with this, we have to put on a restriction on the weight to be a member of the special ordered sets 2. We will first have a look at the approximation in diagram form. The special ordered sets 2 principle will be explained later in this section.
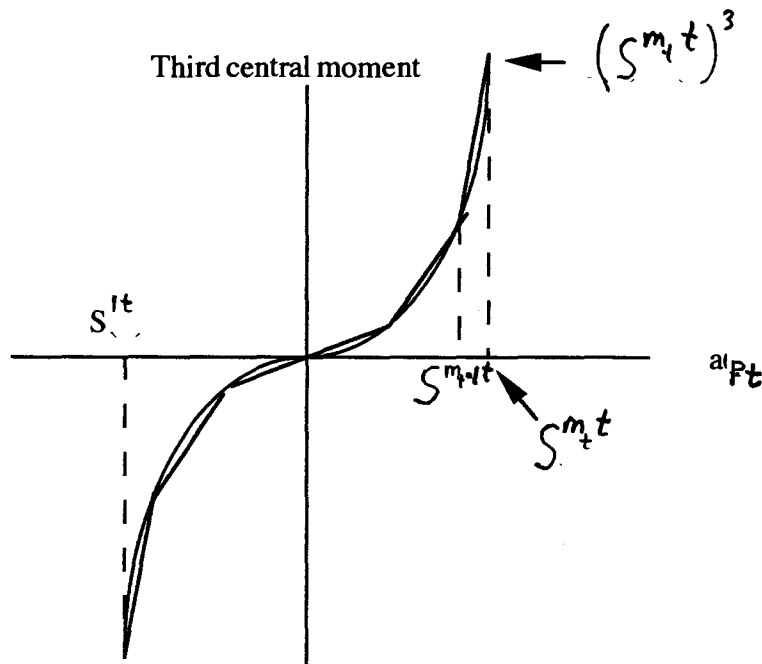

Figure 4.5. Approximation of the third central moment of the portfolio.

The approximation for the third central moment is done in a similar fashion, but we now instead use the cube of the mesh point as the functional value. In contrast to the Konno model, we approximate the function for all values of the cubed deviations from the mean.

For each time period t we have to perform the approximation above. So in the model we have T approximations of this type. A critical issue in the problem is, of course, which and how many meshpoints we use in the analysis. In order to have a good analysis and to keep the objective function as close as possible to the actual one, one should have many meshpoints. To keep the problem simple and not too time–consuming to compute, one has to keep the number of meshpoints small, or iteratively change the approximations of the functions included in the model.

The approximation outlined above can, as mentioned earlier, be operationalized by

introducing m_t weights(w^{jt}), for each time period t, that sum to one. The value of a_{pt}, and the approximated value of a_{pt}^2 and a_{pt}^3 is then given by [31]:

[31a] $\quad a_{pt} = \displaystyle\sum_{j=1}^{m_t} w^{jt} S^{jt}$

[31b] $\quad a_{pt}^2 = \displaystyle\sum_{l=1}^{m_t} w^{jt} (S^{jt})^2$

[31c] $\quad a_{pt}^3 = \displaystyle\sum_{l=1}^{m_t} w^{jt} (S^{jt})^3$

We then have one problem left. During optimization, it is required that the line connecting the functional values of two neighbouring meshpoints is the effective one. This means that a maximum of two adjacent w^{jt} for all t must be nonzero. In the case with a portfolio model where we want to minimize the approximated variance, this is no problem: the structure of the optimization problem ensures that this is the case. In the case of the model we study here, in which the approximated third central moment is to be maximized, this is no longer guaranteed. To ensure this we have to put a restriction on the w^{jt} variables. We do this by saying that each set w^{jt} for all t is an SOS set of type II. We will denote this here by $w^{jt} \in S_2 \ \forall \ t$. The imposition of this constraint moves the problem from a linear programming problem to a mixed integer linear programming problem. The complexity of the problem, and the computer time necessary to solve it, then increase.

It is now possible to formulate the model which we want to study [32]:

[32a] $\quad \max \dfrac{1}{T-1} \left\{ \displaystyle\sum_{t=1}^{T} \gamma_t^+ - \sum_{t=1}^{T} \gamma_t^- \right\}$

subject to:

[32b] $\quad \displaystyle\sum_{i=1}^{n} x_i a_{it} = z_t^+ - z_t^- \quad t \in 1,2,...T.$

[32c] $\quad \displaystyle\sum_{j=1}^{m_t} w^{jt} S^{jt} = z_t^+ - z_t^- \quad t \in 1,2,...T.$

[32d] $\quad \displaystyle\sum_{j=1}^{m_t} w^{jt}(S^{jt})^2 = s_t \quad t \in 1,2,...T.$

[32e] $\quad \displaystyle\sum_{j=1}^{m_t} w^{jt}(S^{jt})^3 = \gamma_t^+ - \gamma_t^- \quad t \in 1,2,...T.$

[32f] $\quad \dfrac{1}{T-1} \displaystyle\sum_{t=1}^{T} s_t \le \bar{\sigma}_p^2$

[32g] $\quad \displaystyle\sum_{i=1}^{n} \hat{\mu}_i x_i \ge \bar{\mu}$

[32h] $\quad \displaystyle\sum_{i=1}^{n} x_i = 1$

[32i] $\quad w^{jt} \in S_2 \; t \in 1,2,...,T.$

[32j] $\quad x_i \ge 0 \quad i \in 1,2,...n.$

[32k] $\quad s_t, \; \gamma_t^-, \gamma_t^+, z_t^-, \; z_t^+ \ge 0 \; t \in 1,2,...,T.$

[32l] $\quad w_{jt} \ge 0 \; j \in 1,2,...m. \;, \; t \in 1,2,...m.$

The restrictions [32b] define the deviation of the portfolio. There are T such restrictions, one for each time period. The variables z_t^+, and z_t^- are defined by [32k] to be positive, so if the return of the portfolio is above the mean return, z_t^+ attains a positive value. The variables z_t^- and z_t^+ are complementary to each othe, which means if one of them is positive the other is zero. If the return is below the mean return of the portfolio, z_t^- takes a positive value. The right–hand side of the constraint [32b] is the value of a_{pt}. The restriction [32c] performs the transformation of a_{pt} such that a_{pt} could be expressed by meshpoints weighted by the weights w^{jt}. For a particular time period t, this restriction finds a proper vector w^t with elements w^{jt} such that the value of a_{pt} is given by:

[33] $\quad a_{pt} = \displaystyle\sum_{j=1}^{m_t} w^{jt} S^{jt}$

The restrictions [32d] and [32e] transform the deviations a_{pt} to the approximated values of a_{pt}^2 and a_{pt}^3 respectively. The approximated value of a_{pt}^2 is for each time period given by s_t. We note that the right hand side of restriction [32d] consists of only one variable, since a_{pt}^2 and the approximated value of a_{pt}^2 for each t, only take positive values. In a similar fashion

~110~

we obtain the approximated value of a_{pt}^3 by the restriction [32e]. We here have to use two variables for each time period since a_{pt}^3 could take positive and negative values. If a_{pt}^3 is negative γ_t^- takes a positive value, and if a_{pt}^3 is positive γ_t^+ takes a positive value. The variables γ_t^- and γ_t^+ are also complementary. The restriction [32f] is the restriction ensuring that the estimated approximated variance of the portfolio is under the predefined level required. Restriction [32g] is the traditional MV restriction imposed on the estimated means or expected return of the portfolio. The SOS II constraints of the weights w^{jt} are given by [32i], and the restrictions [32j], [32k] and [32l] are ordinary nonnegative restrictions imposed on the variables in the problem.

The objective function [32a] maximizes the estimated third central moment of the portfolio. Since γ_t^- takes a positive value when a_{pt}^3 is negative, γ_t^- is subtracted in the objective function. The estimated third central moment is obtained by adding all the approximated cubes of the deviations from the mean and dividing by $T-1$.

The model above has $(m + 5)T + n$ variables and $4T + 3$ constraints.

We have now formulated the model, and we will proceed test it by running the model on actual data from the Oslo Stock Exchange.

4.10. A test of the model on actual data.

In this section we will test the model using actual data from the Oslo Stock exchange. The model is tested by the use of the XPRESS – MP optimization package. This package has the facility of allowing the user to impose constraints like SOS II on a group of variables. The problem is modelled in a specific language for the XPRESS – MP modelling module. After a model is formulated and checked for errors, the module generates a matrix file. This matrix file is read by MP–OPT, another module in XPRESS – MP. This module first solves the problem and gives us the continuous solution to the problem. Then the user is allowed to search for a global optimum by the GLOBAL commands. It is here the optimizer considers the SOS II constraints. When this search is done, the optimizer provides us with the optimal solution to the problem. The package is run on the Norwegian Institiute of Technology, Department of Economic´s micro machine MOSES, which uses the operative system UNIX.

When testing some examples, it appeared that in problems with many observations, i.e. when T is large, the time to solve the problem on the computer was large. It seems that the number of observations, rather than the number of assets, had most influence on the time spent solveing the problem. To give an idea of the time spent on the different problems we

could look at the table below:

Time periods – T	Time to generate one point on the 3D frontier
72	Over 3 days
48	Approximately 8 hours
24	Approximately 0.5 Hour

Table 4.2. Computer time and number of observations.

When we first formulated the problem, we used 72 time periods. If we wanted to generate 50 points on the three dimensional frontier, the time to do this would be more than 150 days, or approximately 5 months. It turned out that the time to solve the problem was too large and the size of the problem was reduced. By testing different time periods it was decided that the number of periods in the model should be 24, such that the tests of the model could be done during one week. We will now describe the data used and the results. In our study we use 10 stocks listed on the Oslo Stock Exchange. The data ranges from the period of 1 January 1988 to 1 January 1990. The stock had to be listed over the whole period. If mergers had taken place,we allowed for this by adjusting the new stock with the switch ratio offered to the old stockholders in the two companies. All return series (arithmetic) are adjusted for splits, emissions and dividends to preserve the true return of each asset. The data was delivered by the Amadeus database at the Norwegian School of Economics and Business Administration. We now show a three dimensional graph of the obtained (μ_p, σ_p^2, γ_p) efficient points.
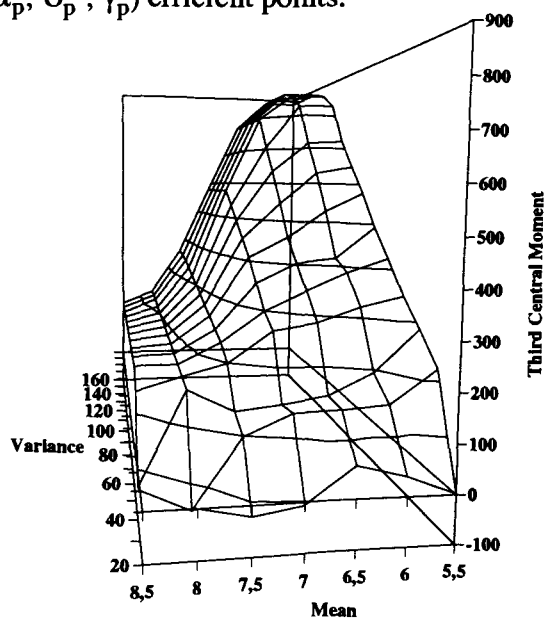

Fig. 4.6. Ex post 3D frontier.

~112~

If we start in the nearest southwest corner of the diagram and move to the right, the restricted lower limit of the mean return of the portfolio decreases. If we move into the diagram, we obtain portfolios with higher variances. The vertical axis denotes the estimated third central moments of the portfolios. In the front of the diagram there are portfolios with a low third central moment. These portfolios correspond to the MV efficient portfolios. The portfolios lying in the east in the diagram with the highest skewness for each level of variance are obtained by maximizing the third central moment and constraining the variance of the portfolio at different levels. We see from the diagram that the maximum third central moment is obtained at a relatively moderate return. To obtain a higher mean, we have to decrease the third central moment of the portfolio. For the other portfolios generated, we have to increase the variance or decrease the mean to obtain a higher third central moment.

To study the $(\mu_p, \sigma_p^2, \gamma_p)$ plane further, we also provide a graph showing the tradeoff between variance and the third central moment. In addition, we provide the MV efficient frontier.
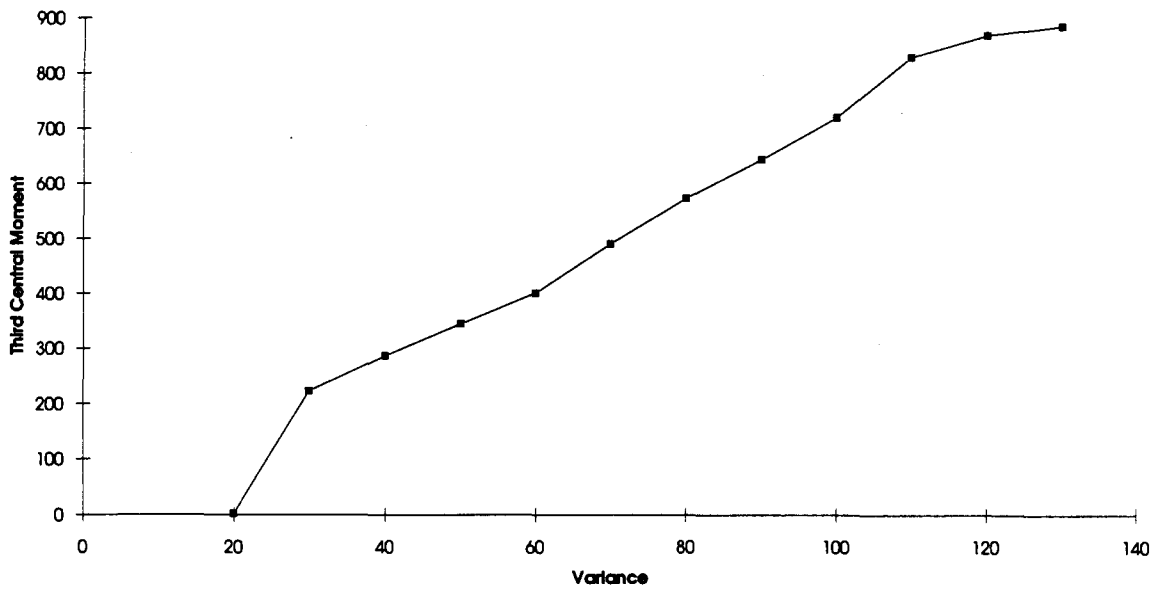

**Fig. 4.7. Tradeoff variance and third central moment.**
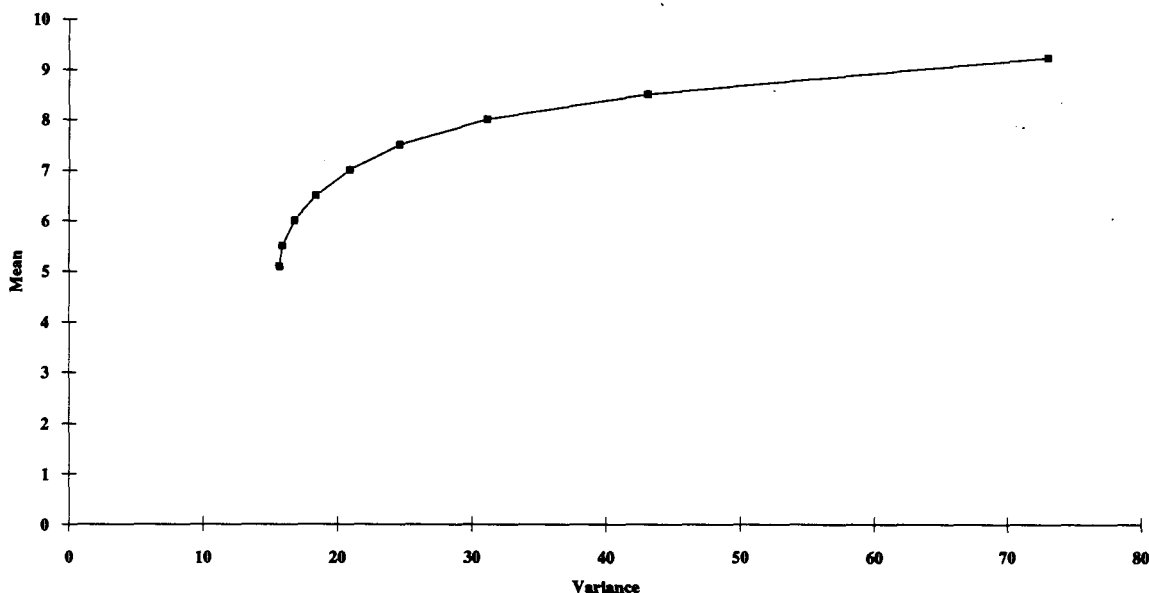
Fig. 4.8. Traditional ex post MV frontier.

4.11. Expected utility considerations.

In section 4.9. we computed different three moment efficient portfolios. The optimizer provided us with the optimal weights for the different portfolios, and we computed theex post mean, variance and third central moment for these portfolios. A question often raised in financial economics by many researchers, among them Tsiang(1972), Kroll, Levy and Markowitz(1984) and Grauer (1986) is whether mean variance models provide the decision maker with enough information, such that information about the third central moment and higher moments do not change the decisions of the investor. The results we obtained in Section 4.9. could be used to compute the expected utility of the different generated portfolios and to see which of the portfolios would be selected by investors with different utility functions. It is not possible to do this without adding some restrictions. This restrictions are:

We assume that the 24 observations used in the analysis form the set of possible outcomes ex ante.

This means that we move from the empirical ex post version of the model to a model with

24 events, each with the probability of 1/24 of occuring. If we had not made this assumption, we would have had to assume that the estimated moments were the actual ones, and thus ignore estimation risk. Formally, here Ω is the set of the 24 observations. Each of the events is denoted by:

[34] $\omega_t : t \in 1,2,...,24.$

In using a finite sample space with 24 outcomes, the variance and the third central moment calculated have to be adjusted. The central moment should be divided by T instead of T-1. The factor used in the adjustment is then given by F and is:

[35] $F = \dfrac{T-1}{T}$

such that the variance and the third central moment in the discrete model are given by the estimated variance multiplied by F. The means or averages of the observations are now the correctly expected returns of the assets. The means in the model do not have to be adjusted. Further we assume that:

The generated approximated points in the model constitute the set of allowable investments.

This means that we only consider the points in the $(\mu_p, \sigma_p^2, \gamma_p)$ plane generated in the analysis in Section 4.9. Further, we assume that the investor, before allocating his wealth to the assets, possesses one unit of wealth. Since the data used is on a percentage basis we have to convert it to a fractional basis. This means that an investment with an expected return of 5% would give the investor an expected end of period wealth of 1.05. Conversely, the variance and the third central moment of wealth could be obtained by dividing the variances and third central moments after the adjustment in [35] by 10000 and 1000000 respectively.

We have performed the analysis by using the Taylor expansion in [4]. We have especially looked for cases where the optimal portfolio does not lie on the MV frontier. For the logarithmic utility function, the negative exponential utility function and the power utility function with exponent positive and less than or equal to one, the results indicated that all optimal portfolios, which means portfolios giving the maximum value of approximated expected utility by the Taylor expansion [4], were MV efficient portfolios. Typically the portfolios selected were those with high expected returns. To obtain an optimal solution not on the mean variance efficient frontier, we have to assume that the preferences to the investor could be represented by a power utility function with exponent considerably negative. As the Taylor expansion for such a utility function is somewhat poor, we have

chosen not to focus on that result.

4.12. Konno and Yamazaki model – comparisons.

An interesting question is how the model presented in this chapter performs relative to the approach suggested by Konno and Yamazaki (1993). In order to answer this, we tested Konno et. al. model on the same data set as used above. Since the two models put great emphasis on the formulation of the third central moment, we used the Konno and Yamazaki model to find the maximum third central moment portfolio. Unfortunately, this portfolio gave us a small third central moment compared with the maximum third central moment computed by the model presented in this chapter. By using $\alpha=1$, $\rho_1 = \hat{\mu}_p - 5\%$ and $\rho_2 = \hat{\mu}_p - 15\%$, we obtained an estimated third central moment of 113,475, which is small compared with the estimated third central moment of 886,487 obtained by our model. Typically, we obtained small variance portfolios with moderate third central moment. One reason for this could be that the G function used in the Konno et. al. model attracts low variance portfolios more than it attracts high third central moment portfolios, since the Konno and Yamazaki (1993) G function implies no preference for outcomes in the upper tail of the distribution.

4.13. Conclusions.

In this chapter we have studied a three moment portfolio model. We have discussed the theory concerning three dimensional portfolio analysis, and we have performed an empirical analysis using actual data from the Oslo Stock Exchange. One of the results was that the model, for small time series, worked well as compared with the Konno and Yamazaki (1993) approach. However, for large time series the model was, from a computing perspective, very time consuming.

References:

Arditti, F. D. (1967), "Risk and the Required Return on Equity." Journal of Finance, Vol. 22., No. 1. pp. 19–36.

Beale, E.M.L., Tomlin, J.A. (1969) Special facilities in a General Mathematical Programming System for non Convex Problem Using Ordered sets of Variables, in J.

Lawrence (Ed.), Proceedings of the 5th International Conference on Operations Research, Tavistock, London.

Borch, K.A. (1973) "Uncertainty and Indifference Curves - A Correction." Review of Economic studies. Vol 40 pp141-

Borch, K.A. (1974) "The Rationale of the Mean Standard Deviation Analysis: Comment." The American Economic Rewiew Vol 64 N0 3. pp 428-30.

Friend, I. and Westerfield, R. (1980) "Co–Skewness and Asset Pricing." Journal of Finance. Vol 35 pp 897–913

Jean, W. (1971) " The Extension of Portfolio Analysis to Three or More Parameters." Journal of Financial and Quantitative Analysis, Vol. 6. No. 1. pp. 505–515.

Jean, W. (1973), "More on Multidimensional Portfolio Analysis." Journal of Financial and Quantitative Analysis, Vol. 8, No. 3, pp. 475–490.

Grauer, R.R. (1986), "Normality, Solvency and Portfolio Choice." Journal of Financial and Quantitative Analysis. Vol 21. pp. 265–278

Hawawini, G.A. (1980) "An Analytic Examination of the Intervalling Effect of Skewness and Other Moments." Journal of Financial and Quantitative Analysis. Vol 15. pp 1121–1127.

Kroll, Y. Levy, H. & Markowitz. H. "MV versus Direct Utility Maximation." Journal of Finance March 1984 pp 47-61.

Kariya, T. et al. "Distribution of Stock Prices in the Stock Market of Japan." Toyo Keizai Publishing Co., Tokyo, 1989.

Konno, H., Yamazaki, H. (1991) "Mean Absolute Deviation Portfolio Optimation Model and its Application to the Tokyo Stock Market." Management Science Vol 37 May pp.519-531.

Konno, H. & Yamazaki, H. (1993) "A Mean Absolute Deviations Skewness Portfolio Optimization Model." Annals of Operations Research vol. 45 pp. 205–220.

Kraus, A., Litzenberger, R. (1976) "Skewness Preference and the Valuation of Risky Assets." Journal of Finance, Vol. 31, No. 4. pp. 1085 – 1100.

Levy , H. (1969) " A Utility Function Depending on the Three First Moments." The Journal of Finance.September 1969.

Levy, H.(1974) "The Rationale of the Mean Standard Deviation Analysis : Comment." The American Economic Rewiev. Vol 64. No 3. pp 434-441.

Levy, H. & Markowitz, H. (1979) "Approximating Expected Utility by a Function of Mean and Standard Deviation." The American Economic Review. Vol 69 pp. 308-317.

Loistl, O. (1976) "The Erroneous Approximation of Expected Utility by Means of a Taylor Series Expansion." Analytic and Computational Results. American Economic Rewiew.Vol 66. No. 5. pp. 904–10.

Markowitz, H.M. (1952) "Portfolio Selection." Journal of Finance. pp 77-91.

Pang, J.S. (1980) "A New Efficient Algorithm for a Class of Portfolio Selection

Problems." Operations Res. 28 pp. 754-767.

Perold, A. (1984) "Large Scale Portfolio Optimizations." Management Science 30 pp. 1143-1160.

Pratt, J.W. (1964) "Risk aversion in the Small and in the Large." Econometrica vol 32 No 1-2. pp. 122-136.

Scott, R. C., Horvath, P. A. (1980) " On The Direction of Preference for Moments of Higher Order Than The Variance." Journal of Finance Vol 35 pp. 915-919.

Singleton, J.C. and Wingender, J. (1986) "Skewness Persistence in Common Stock Returns." Journal of Financial and Quantitative Analysis. Vol. 21. pp. 335-341

Tsiang, S.C. (1972) "The Rationale of the Mean Standard Deviation Analysis, Skewness Preference, and the Demand for Money." The American Economic Rewiev. Vol 62. pp. 354-371.

Tsiang, S.C. "The Rationale of the Mean Standard Deviation Analysis : Reply and Errata for Original Article." The American Economic Rewiew Vol 64. No 3. pp 442-451.

5. A Discrete Multiperiodic Asset Allocation Model With Periodic Penalty for Nonfullfillment of Certain Obligations.

Abstract: This paper discusses a multiperiodic model in a discrete time framework. The uncertain environment is given by an event tree. The investor´s aim is to maximize the expected utility of the terminal value for a portfolio by trading securities in a finite number of periods. The terminal value of the portfolio is equal to the capital gain accumulated by the trading of n securities less the accumulated penalties. The penalties involved are connected to not fullfilling certain capital and return requirements at the end of the given time periods $t \in [1,2,..,T]$. The penalties are incorporated by assigning of a penalty fee for each sub period which has to be paid at time T. The resulting problem is a nonlinear stochastic program which is solved by separating the different paths in the event tree and solving these subproblems independently. The nonantipacivity constraints are taken care of by forming the objective function in the sub problem as an augumented Lagrange function, a method suggested by Wets and Rockafellar (1987). By iteratively solving these subproblems and computing a convex combination of the solution, one obtains a solution that converges to the solution of the original problem. The paper will also show some computational results. The model is an abstraction of the regulations imposed on insurance companies in Norway and other countries.

5.1. Introduction.

Models that determine the amount to be invested in different assets are often of the type where the decision maker determines an optimal asset mix at the beginning of a period, and after that, the investor observes a stochastic return on his portfolio. No corrective action is taken, and the model stops there. The first asset allocation models, originated by the work of Markowitz (1952), were of this type, and there exists a large block of literature and research on these models.

Multiperiodic models allow for the decision maker to make corrective actions during the lifetime of the model. We say corrective actions, since, during the lifetime of the model, information becomes known to the decision maker, and decisions are made as a response to this new information.

Typically, a multiperiodic model is a model where there are several points in time between what we would call today and a future terminal date, T, the horizon of the model. We

denote the sequence of points in time starting at time 0 as $t \in [0,1,2,..,T]$. From the notation, it follows that there are T+1 such points in time.

According to the Von Neumann Morgenstern expected utility theorem, uncertain investments are ranked according to the value of the expected utility. In the case of single period models such as the Markowitz (1952) model mentioned above, an approximation of expected utility is made by the use of the expected return and variance of return of the portfolio.

When it comes to multiperiodic models, the decision is not as simple. Information in a multiperiodic model increases with time and decisions made today have to incorporate the effect of information realizations through future periods. When we consider a multiperiodic model, solutions could, under certain circumstances, be the same as those obtained from a partial, one step analysis. By this we mean that the correct decision could be taken by just considering the stochastic asset prices one period into the future. Mossin (1968) studied a multiperiodic model with two assets and presented a class of utility functions which possess this property. He divided the utility functions into two categories. The categories were utility functions that imply partial myopic solutions, and utility functions that imply complete myopic solutions. The former category of utility functions requires only the knowledge of the future risk free interest rate. In this case, the investor with preferences represented by a utility function that belongs to this category could simply allocate the wealth by a one period into the future analysis, as if the entire resulting wealth would be invested in the risk free asset. As in the case with complete myopia, the decision could be based upon an expected utility analysis only one period into the future.

We define $U(v^T)$ to be a utility function representing the preferences of the decision maker where v^T is wealth. In addition we assume as usual that the utility function is monotonically increasing and concave. The Mossin analysis is based upon characteristics about the fraction M:

$$[1] \quad M = -\frac{\dfrac{dU}{dv^T}}{\dfrac{d^2U}{dv^T}}$$

The results were:

Economy	Complete myopia	Partial myopia
One riskless asset with zero yield	$M = a + bv^T$	•
One riskless asset with nonzero yield	$M = bv^T$	$M = a + bv^T$
Both asset risky	$M = bv^T$	•

Table 5.1. Characteristics M and myopic decision rules.

Over, v^T is terminal wealth and a and b are constants.

If one considers the Bernoulli, $\ln(v^T)$, utility function, and an investor who maximizes the expected utility of terminal wealth T periods into the future, one should, in the model setting according to Mossin with yield distribution identical for all periods, base the decision on a one-period model. As Samuelson (1969) also points out, the optimal behaviour of an investor that maximizes the expected value of a Bernoulli utility function does not change during the lifetime of the model, as long as the distribution of the asset return is stationary.

As real situations are often complex, and distributions may not be stationary over time, the introduction of a multiperiodic analysis is relevant. The purpose of this chapter is to study a model where we implement actual regulations appearing in the economy and allow for the possibility of nonstationary distribution of the asset prices. The chapter is organised as follows.

In Section 5.2., we outline the price system for the securities. In Section 5.3. we describe the decision variables. Probabilities and stochastics in the model are discussed in Section 5.4.. The model is presented in Section 5.5. and some general considerations concerning solution procedures are covered in section 5.6. The progressive hedging algorithm is discussed in Section 5.7. In connection with the progressive hedging algorithm, one has to select start values and step size for Lagrange multipliers and penalty parameters. How this is done is covered in Section 5.8. Section 5.9. covers the single period formulation of the model and the corresponding deterministic equivalent program. In Section 5.10. we illustrate the model by a numerical example. Finally, in Section 5.11, we draw some conclusions.

5.2. Price system.

The model presented will consist of a selection of n assets. The model runs through T time periods that end for $t \in [1,2,...,T]$, and has a finite sample space with K elements:

[2] $\Omega = \{\omega_1, \omega_2, ... \omega_K\}$

The probability of a sample event occurring is positive. In this chapter, we will use the terms scenario and path in the same meaning as event. We use the index j for the K events in the model, thus:

[3] $j \in \{1,2,...,K\}$

The process starts with no information at time 0, and ends up with full information at time T. Formally, we state this by letting f_t represent the information structure at time t, hence:

[4] $f_0 = \{\{\omega_1, \omega_2, ... \omega_K\}\}$, $f_T = \{\{\omega_1\}, \{\omega_2\}, ..., \{\omega_K\}\}$

The information structure (f_t) , for any $t \in [0,T]$ is finer than that of the preceding period (f_{t-1}). In other words, every set in f_t is a subset of a set in f_{t-1}. Formally, this is:

[5] $f_t \in f_s \ \forall \ s \leq t$

Thus the process is measurable, and the sequence of random variables is adapted to the information structure f_t. An event is a path of relieved asset prices and dividends through time. An event can be viewed as a scenario. At one particular period in time, we denote the prices of the assets by a vector $\tilde{v}^t$:

[6] $\tilde{v}^t = [\tilde{v}_1^t, \tilde{v}_2^t, ..., \tilde{v}_n^t]$

At each point in time there are dividends. We denote the dividends at time t as:

[7] $\delta^t = [\delta_1^t, \delta_2^t, ..., \delta_n^t] \ \ \delta^t \geq 0 \ \forall \ t$

The assets are claims to take part in a dividend process.

With these definitions in mind, we have that the total return from asset i in period t is given

by:

[8] $v_i^t - v_i^{t-1} + \delta_i^t.$

This definition implies that the owner of the asset receives an appreciation or depreciation of the asset plus a nonnegative dividend. We will assume that the dividends are not paid directly to the investor, but assume that the dividends are kept inside the firm which is the issuer of the asset. This means that gains from dividends are represented by movements in the asset prices $\tilde{v}^t$.

To make the concept of an information structure more tangible, we illustrate the information structure in the following figure.
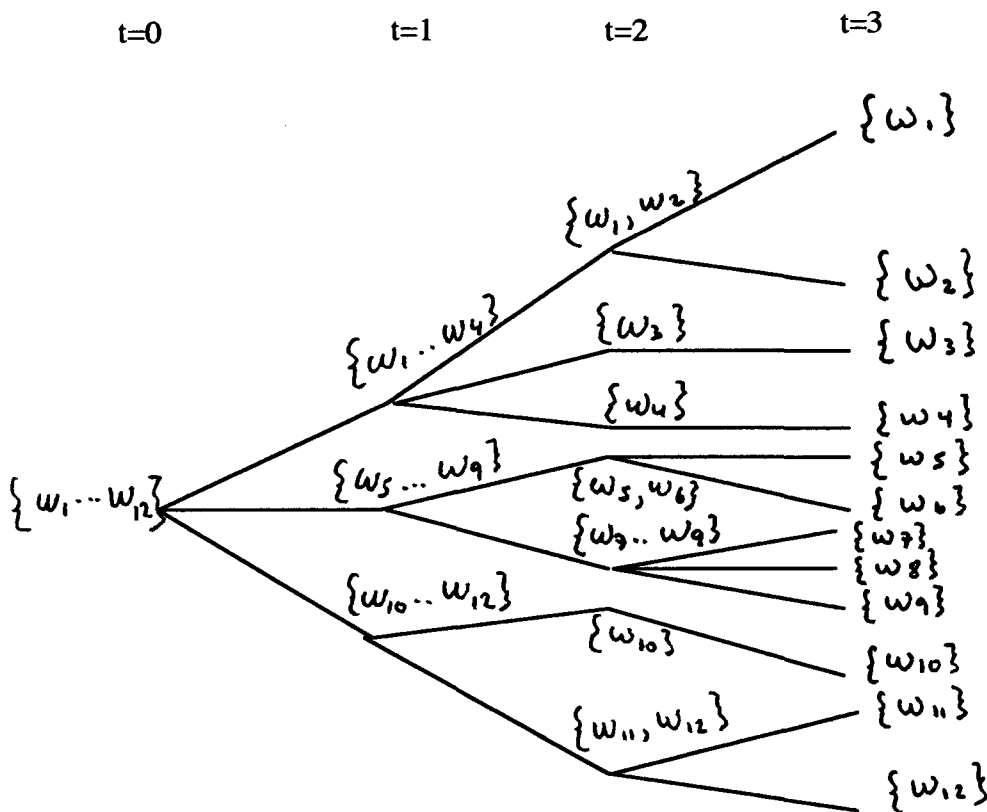

Fig.5.1.   An information structure.

Above we have illustrated the concept of an information structure by using an event tree. The event tree is a pair, consisting of a set of vertices denoted by Ξ, and a set of edges, given by A, connecting the different vertices. The vertices represent a particular time and state in the event tree. At a particular node or vertex, the future state or the next node is

uncertain. We denote the current node at time t by ξ_t. For any $s > t$, ξ_s is stochastic. An event or scenario in the model is then a sequence of nodes given by $\{\xi_1, \xi_2, ..., \xi_T\} \in \Xi$ such that a walk through an edge between two subsequent nodes is given by $(\xi_t, \xi_{t+1}) \in$ A.

We have now described the stochastic environment of the asset prices and in the next section we will consider the decision variables.

5.3. The decision variables.

A trade strategy $\mathbf{x}$, is a sequence of investments in assets. We denote the number held of assets i at the beginning of time period t as x_i^t. The set $\mathbf{x}$ is then defined as follows:

[9] $\mathbf{x} = [x_i^t \mid i \in 1,2,...,n.; t \in 0,1,2,...,T-1.]$

The process $\mathbf{x}$ is an adapted process in $\mathfrak{R}^n$. In other words the sequence, x_i^1, x_i^2,..., x_i^T, $\forall$ i, is a random variable with respect to (Ω, f_t). This means that the adjustments in the portfolio made as time passes depend upon the evolution of the asset prices.

We will define the trades or holdings at a particular point in time with $\mathbf{x}^t$.

[10] $\mathbf{x}^t = [x_i^t \mid i \in 1,2,...,n.]$

In section 5.2. we defined a a price system, and in this section we define variables x_i^t which denote holdings of the different assets at points in time. We now move a step further and look at a property that is reasonable to impose on the asset price and dividend system described above. The common assumption imposed on economic models is that the decision maker and actors in the market prefer more for less. A necessary condition to put on the system above is then that it should not be possible to create something out of nothing. We refer to this condition by stating that no arbitrage opportunity should exist in the model. To study this property, we introduce certain definitions:

A consumption plan is an adapted process in $\mathfrak{R}$ given by $\mathbf{c} = [\mathbf{c}^t \mid t \in 1,2,...,T]$. The consumption plan satisfies the following equation[1]:

---

[1]The letter T outside the brackets is the transpose operator here.

[11] $[x^{t+1}]^T v^t = [x^t]^T v^t + c^t$

The equation [11] simply states that c^t is the difference between the value of a portfolio before and after adjustments for the portfolio holdings in one particular time period. We then could state the no arbitrage criterion.

Definition of no arbitrage opportunity:

We say that there is an arbitrage opportunity if there exists a trading strategy **x** where **c** is nonnegative and there exists at least one t where c^t is positive for all $\omega \in \Omega$. In addition we require:

[12] $[x^0]^T [v^0 + \delta^0] \leq 0$

Here x^0 are the trades done at t = 0. By assumption $\delta^0 = 0$.

An arbitrage trading strategy is then a strategy that is always nonnegative, and strictly positive for at least one event, acquired at a nonpositive cost. We will assume that such arbitrage opportunities does not exist.

5.4. Probabilities and stochastics in the model.

If we consider an arbitrary realization of the stochastic process given by event three (Ξ, A), and consider the process at time period t, the current vertex is given by ξ_t. We denote the corresponding sub-tree with root ξ_t to the pair $(\Xi(\xi_t), A_t)$. At this node there are $m_t(\xi_t)$ edges connected to, at the moment unknown, ξ_{t+1}. We denote such an edge by $\zeta_q(\xi_t)$. The edges are indexed by q. The probability of a walk through edge $\zeta_q(\xi_t)$ is given by $P(\zeta_q(\xi_t))$. These probabilities are nonnegative and satisfy:

[13] $\sum_{A_t} P(\zeta_q(\xi_{t+1})) = 1$

These probabilities could have been formed by the use of a priori information, given by a neutral agent or based upon subjective judgments.

5.5. The model.

Due to the central role they play in economies, financial institutions are often subject to more regulation than non financial firms. Due to this central role, credit authorities continuously monitor specific capital requirements and investigate whether these are satisfied by the institution. The reason for this is that the solvency of those institutions is important not only for the monetary part of the economy, but also the economy as a whole. These regulations are often implemented such that the financial institution is required to have specific funds available at fixed dates. These requirements are often complicated, but could, for example be linked to the total amount outstanding for the institution. In the case of banks, the Bank of International Settlements (BIS) defines these rules.

The model we will now present is multiperiodic, and the main purpose of the decision maker is to maximize the expected utility of terminal wealth. Since the decision maker is restricted by a set of capital requirements, he also has to take into consideration the effect these requirements have on the return on the portfolio or the value of the portfolio at a fixed date. How this effect is incorporated will be shown below.

We will now list the assumptions used in the model.

1. The decision maker maximizes the expected utility of terminal wealth.

This assumption implies that the investor ranks investments according to the expected value of a utility index. This is in accordance with the expected utility theorem of John Von Neuman and Oscar Morgenstern (1952). We will use U($\bullet$) to represent the utility index. In addition, we assume that if v^T is a positive variable[14]:

[14a] $\quad \dfrac{\partial U(v_T)}{\partial v_T} > 0$

[14b] $\quad \dfrac{\partial^2 U(v_T)}{\partial v_T} < 0$

These assumptions imply that the decision maker has a preference for wealth [14a] and is risk averse [14b]. We do not consider a multiperiodic additive utility function to keep the model simple, but we recognize that the model presented could, with slight modifications, be reformulated in order to capture such a utility function.

2. All returns on the assets from one period to another are reinvested.

We then do not allow for consumption until period T. All flow of wealth is then reinvested into the assets in the model.

3. Assets are infinitely divisible.

This assumption enables us to buy any fraction or any number of shares in the asset, not integer values, as common for practical purposes. If the value of the portfolio is large, this assumption makes no significant difference for practical purposes.

4. There is a penalty for not fulfilling certain obligations.

As the return of the portfolio is realized from one period to another, there is a possibility that this return could be below some required level, if it is so the institution has to cover the return shortfall by the use of funds. If there have been several periods with an unfavorable return, these funds may vanish. In the model we will have two types of fund that it is possible to use for cover in the case of unfavorable returns. We will denote these funds B_1 and B_2. We will model these funds such that if it is necessary to cover unfavorable returns, one covers first using fund 1 until it is empty, and then one continues covering from fund 2. We will penalize withdrawals from fund 2 harder than withdrawals from fund 1.

There is a critical point given by k, which is the critical return from time t to t+1 in which returns below k imply withdrawals from funds. If the return from one period to another is larger than k, the funds are increased. The filling up of these funds is done on a LIFO basis (Last In First Out). If fund 2 is not at its maximum level, indicated by k_2, it is increased until it reaches k_2. If there are still excess returns available, fund 1 is increased.

We will illustrate the flow into funds by the following figure:
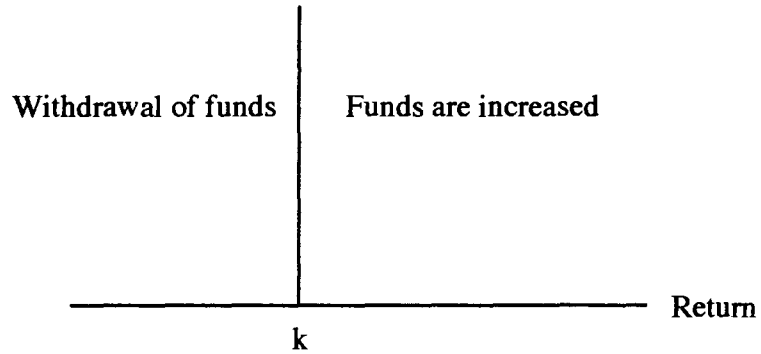
Figure 5.2. Funds and the shortfall constant k.

The figure illustrates that there are withdrawals from the funds if the return is below the predefined return k. The dynamics of fund 1 and fund 2 can be illustrated in the following figure:
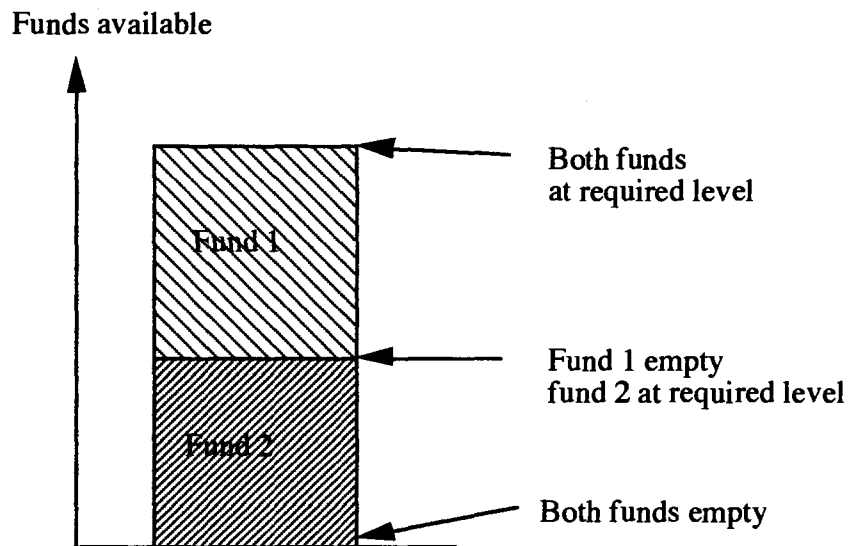

Fig. 5.3. The funds in the model.

Above we have illustrated the dynamics of the two funds in the model. Keeping fund 1 intact has lower priority, and low returns are first covered by the use of funds from fund 1. If fund 1 is empty one covers unfavorable returns from fund 2.

In the model, it is also possible to have excellent returns. If this is the case, and both fund

1 and fund 2 are at the required level, the excess funds are just added to a fund 3, which is a buffer for the use of funds from fund 1 and 2 in the next period. If it turns out that there are funds available in the excess fund 3, and withdrawals from funds are necessary, no penalty is performed on such covers. Only withdrawals from fund 1 and fund 2 are penalized.

The penalization takes the form a fee, charged in each time period, proportional to the amount of missing funds in fund 1 and fund 2. This fee is not paid directly, but is a debt to the legal authority. This debt, with accrued interest, is paid in period T. The interest charged on this debt is fixed for all periods. We will represent the charging fee´s fraction of missing funds in fund 1 and fund 2 with β and α respectively. A practical interpretation of the above penalization is seen in capital regulations imposed on life insurance companies.

We will also impose a fee to be paid if the actual return from period t to t+1 is below the predefined level k. This fee is not connected to the level of the funds but is contingent on the return of the portfolio in a single period. The penalty connected to the return of the portfolio is activated independent of the funds available. The penalty for shortfall return is given by γ. A practical interpretation of this penalty is the return requirement guaranteed by life insurance companies. We will now formulate the problem.

The problem is[15]:

$$[15a] \quad \text{maximize } E\left(U\left(\tilde{v}_p^T - \sum_{t=1}^{T} d_t \rho_t\right)\right)$$

subject to:

$$[15b] \quad \tilde{v}_p^t = \sum_{i=1}^{n} x_i^{t-1} \tilde{v}_i^t \quad t \in 1,2,..,T$$

$$[15c] \quad \tilde{v}_p^t = \sum_{i=1}^{n} x_i^t \tilde{v}_i^t \quad t \in 0,1,2,..,T-1.$$

$$[15d] \quad \Delta F_t^- - \tilde{v}_p^{t-1}(1+k) + \tilde{v}_p^t \geq 0 \quad t \in 1,2,..,T.$$

$$[15e] \quad F_t = F_{t-1} - \Delta F_t^- + \Delta F_t^+ \quad t \in 1,2,..,T$$

$$[15f] \quad \tilde{v}_p^t + \Delta F_t^- - \Delta F_t^+ - \tilde{v}_p^{t-1}(1+k) = 0 \quad t \in 1,2,..,T$$

[15g] $y^t + F_t - k_1 \geq 0$ $t \in 1,2,...,T$

[15h] $z^t + F_t - k_2 \geq 0$ $t \in 1,2,...,T$

[15i] $u^t - \Delta F_t^- \geq 0$ $t \in 1,2,...,T$

[15j] $\rho_t = y^t \beta + z^t \alpha + u^t \gamma$ $t \in 1,2,...,T$.

[15k] $\Delta F_t^-, \Delta F_t^+, F_t, y_t, z_t, u_t, \rho_t$ $t \in 1,2,...,T$.

where:

$\tilde{v}_p^t$ = value of portfolio at time t

ΔF_t^- = withdrawals from funds at time t

ΔF_t^+ = increase in funds at time t

F_t = funds at time t

y_t = amount below requirement in fund 1

z_t = amount below requirement in fund 2

u_t = difference between actual and required return – absolute basis

k_1 = level at which we make withdrawals from fund 1

k_2 = level at which we make withdrawals from fund 2

k = required return level

d_t = discount factor for penalties at time t

F_0 = initial level of total funds available

α = penalization of withdrawals from fund 2

β = penalization of withdrawals from fund 1

γ = penalization of shortfall return

The number of variables in the model is for each time period, in addition to the stochastic asset prices, given by n+7. In total there are (n+7)T variables. There are 9T restrictions in the model.

The objective function [15a] maximizes the expected utility of terminal value of the portfolio, less compounded penalty. The compounding is done by the predefined constants d_t. The constraint [15b] controls the value of the portfolio. The funds available at time t are F_t, and the increase and decrease in F_t are given by ΔF_t^+ and ΔF_t^- respectively. The restriction [15c] assures that all investments done in period t are equal to the funds available at time t. The flows from the funds are controlled by the restriction [15d], which holds

because we penalize withdrawals from the funds. The equality [15f] defines the inflow of wealth to the funds as a function of portfolio value and outflow of funds. The restrictions [15g], [15h], [15i], and [15j] are related to the penalization of withdrawal of funds such that y^t takes a positive value if the fund is below the critical value k_1. The value of y^t is then the difference $k_1 - F_t$. Conversely, z^t takes a positive value if the fund is below the critical value k_2. Analogous to y^t above, z^t would take the value $k_2 - F_t$. Finally, u^t takes a positive value if the return in period t is below the predefined level. This is because ΔF_t^- in that case has a positive value.

The problem [15] is a stochastic programming problem with a concave objective function and linear constraints. Any deterministic version of the problem is easily solved by nonlinear programming. The introduction of stochastics into the model complicates it and a relevant algorithm has to be found to solve this problem.

5.6. Solution techniques – General considerations.

The stochastic program [15] maximizes the expected value of a concave function. The expected value of the objective function is given by:

$$[16] \quad E\left(U\left(\tilde{v}_T - \sum_{t=1}^{T} d_t\rho_t\right)\right) = \sum_{\Omega} P(\omega_j)U\left(\tilde{v}_T - \sum_{t=1}^{T} d_t\rho_t\right)$$

The expected value of the objective function is a linear combination of concave functions, weighted with the probability for each event. This function is also concave, since linear combinations of concave functions are also concave. In principle, we could solve programs with an objective function of the form [16] by nonlinear optimization techniques. In addition, we have to modify the restrictions in the problem and add the nonantipacivity constraints that occur in stochastic programs. If the number of states or scenarios is large, the nonlinear programming problem becomes considerably large, and a more specially designed optimization algorithm is preferable.

In the stochastic programming literature, there are several algorithms available for use. We will use an approach suggested by Wets and Rockafellar(1987). This approach, the progressive hedging algorithm or scenario aggregation algorithm, has the advantage that it keeps the problem in its original form, and that it is applicable to, in principle, all stochastic problems which have a solvable deterministic counterpart. The algorithm, to be described below, requires only the presence of a nonlinear optimization package with the ability to

handle quadratic penalization.

5.7. The progressive hedging algorithm and the multiperiodic model.

To simplify the notation we will write the problem in a more compact form. We will define Ψ to be the set of all variables in the model. This set of variables consists of the $n \times T$ decision variables x, and the other $7 \times T$ variables in the model. Further we define Ψ^t to be the set of these variables corresponding to time period t. Since the variables in the model depend upon the development in the asset prices, we will denote the conditional values of the variables given the event ω_j as Ψ_{ω_j} and x_{ω_j}. The set defined by the restrictions in the model is given by C. Since the restrictions in the model are stochastic, we define C_{ω_j} to be the restrictions if event ω_j occurs.

When the random asset prices evolve through time, there may be several paths that are equal up to some point t. A necessary assumption is that the trading strategy up to that point must agree. If we let φ_t be the coarsest partition of Ω in sets at time t, and A_p^t be a collection of events, which is a member of such a set, and in addition have two states ω_1 and ω_2 such that:

$$\omega_1, \omega_2 \in A_p^t$$

Then this condition can be stated as:

[17] $x_{\omega_1}^t = x_{\omega_2}^t \ \forall \ \omega_1, \omega_2 \in A_p^t$ or $\Psi_{\omega_1}^t = \Psi_{\omega_2}^t \ \forall \ \omega_1, \omega_2 \in A_p^t$

The latter conditions follow from the former since the auxiliary variables in the model are indirectly defined by the decision variables. Condition [17] guarantees that the solution provided by the model does not depend upon information not yet available.

The set of implementable policies defines a subspace N or function from Ω to $\mathfrak{R}^{T(n+7)}$ that satisfies the condition [17]. Formally this is:

[18] $N = \{\Psi : \Psi \to \mathfrak{R}^{T(n+7)} \, | \, \Psi_{\omega_1}^t = \Psi_{\omega_2}^t \ \forall \ \omega_1, \omega_2 \in A_p^t \} \ t \in \{0,1,...,T-1\}$

This criterion simply states that all decisions formed by the trading strategy x must agree in time for paths that have similar evolutions. This condition could be imposed by modeling

it as an explicit constraint. Such a formulation makes the problem nonseparable in the events. Another approach is the one suggested by Wets and Rockafellar, where we model these constraints by using an augumented Lagrange function as the objective function for the different subproblems corresponding to the different events ω_i. This makes it possible to separate the optimization for the different scenarios.

A policy or trading strategy is admissible if:

[19] $\quad \mathbf{x} \in C = \{\mathbf{x} \mid \Psi_{\omega_j} \in C_{\omega_j} \; \forall \; \omega_j \in \Omega\}$

This means that a policy is admissible if it satisfies the explicit constraints in the model.

If a policy is both implementable and admissible, we say that it is feasible.

With these definitions in mind, and an optimization where these constraints are imposed,we obtain a problem with the necessary conditions. Formally this is an optimization where we impose:

[20] $\quad \Psi \in N \cup C$

The difference between this constraint set and the original one is the imposition of the nonanticipativity constraints.

The criterion stating that a policy or trading strategy should be implementable is inescapeable in the optimization procedure. These constraints ensure that the solution we end up with through running the optimization procedure is independent of the different events in the model. This is reasonable since, at time 0, we have no information, except in a probabilistic sense, about the future development of the asset prices.

Using this compact notation the problem is[21]:

[21a] $\quad$ Max $E(U(\mathbf{x}))$

subject to:

[21b] $\quad \Psi \in C \cap N$

We will define the solution $x^*(\omega_j)$ as the optimal solution to the following problem:

[22] $\quad x^*(\omega_j) = \text{argmin}\{-U^*(x) \mid \Psi \in C_{\omega_j}\}$

Here $U^*(x)$ is a modified objective function that penalizes violation of the nonantipacivity constraints. How this penalization is done is described in the algorithm under. To further exploit the details we assign $x^{t*}(\omega_j)$ to the decision variables corresponding to time t in the solution of [22].

If no penalization is done $U^*(x)$ is equal to $U(x)$. Over, $x^{t*}(\omega_j)$ is the optimal, ex post value of the variables, associated with time t, if the scenario or path followed by the asset turned out to be ω_j. We can interpret $x^*(\omega_j)$, to be a choice or policy, through the selection of an optimal trade strategy, given the event ω_j. We note that the choice of a trade strategy in one particular scenario determines Ψ_{ω_j}, since the selection of investment proportions, x, determines, under one particular state, the other variables in the model. We will now define:

[23] $\quad x^{tP} = E(x^t \mid A_p^t)$

Here x^{tP} can be interpreted as the expected value of x^t given that the process leads us to A_p^t. In other words, x^{tP} is the conditional expectation of the ex post optimal trade strategy given A_p^t. The probability of reaching A_p^t is given by:

[24] $\quad P(A_p^t) = \sum_{\omega_j \in A_p^t} P(\omega_j)$

Using this, x^{tP} is given by:

[25] $\quad x^{tP} = \sum_{\omega_j \in A_p^t} P(\omega_j) \frac{x^{t*}(\omega_j)}{P(A_p^t)}$

An implementable policy is therefore reached by requiring:

[26] $\quad x^t(\omega_j : \omega_j \in A_p^t) = x^{tP}$

The condition [26] is the nonantipacivity constraint for the model.

We recognize that [26] defines a linear transformation of the variables $\mathbf{x}^t$ to $\mathbf{x}^{tP}$. It is termed the aggregation operator relative to the given information structure f_t.

We have now defined a number of variables and explained certain concepts. We will subsequently turn our attention to the algorithm. In order to utilize the algorithm we need to define a sequence of the Lagrange vector λ^l, which is the penalty we give to the nonfulfillment of the nonantipacivity constraints. Here, l is the iteration number in the algorithm. In addition we introduce a perturbation term, denoted by $W_t^l(\omega_j)$. The algorithm works as follows:

Step 0:

Initialization is done here. We set $W_t^0(\omega_j)$ $\forall$ t, equal to zero, in addition we select a value for λ^0. After the initialization is done we start with the main block of the algorithm by setting l=1.The variable l is then a counter for the iterations in the algorithm.

Step 1:

For each $\omega_j \in \Omega$ we solve the following subproblem:

$$[27] \quad \mathrm{argmin}\left\{-U(\mathbf{x}) + \sum_{t=1}^{T}\left(W_t^{l-1}(\omega_j)\mathbf{x}^t + \frac{1}{2}c^l\left\|\mathbf{x}^t - \mathbf{x}^{tP,(l-1)}\right\|^2\right): \Psi \in C_{\omega_j}\right\}$$

Here $\|\bullet\|$ is the Euclidian or L_2 norm, and $\mathbf{x}^{tP,(l-1)}$ is the conditional expectation obtained from the individual event solutions in the previous iteration, obtained by the use of [25]. In the first iteration we compute $\mathbf{x}^{tP,0}$ by solving a problem for each event without any penalization and by using [25]. The parameter c^l is a penalty corresponding to the quadratic penalty term in iteration l.

We also compute new values for $\mathbf{x}^{tP,l}$.

Step 2:

Here we update the perturbations in the problem. The updating is given by:

$$[28] \quad W_t^l(\omega_j) = W_t^{l-1}(\omega_j) + \lambda^l(\mathbf{x}^{t*,l}(\omega_j) - \mathbf{x}^{tP,l})$$

Above $x^{t*,l}$ is the optimal solution obtained in the current iteration (l).

Return to step 1.

☐

The algorithm presented above is utilized by using the step procedure iteratively. This raises the question of a stop criterion for the algorithm. One suggestion of a distance measure between each iteration is:

$$[29] \quad \kappa_l = \| x^{t\,p,l} - x^{t\,p,l-1} \|$$

This measure takes the L_2 norm of the difference of the variables $x^{t\,p}$ in the last two iterations. A possible stop criterion is then:

$$[30] \quad \text{Stop if } \kappa_l \leq \bar{\kappa}.$$

Here $\bar{\kappa}$ is a predefined constant. The selection of $\bar{\kappa}$ is a matter of convenience.

The above is a discussion of the general setting of the optimization procedure. The key issue in the procedure is first the condition that restricts the decision variables to be indistinguishable for different events with similar movements up to period t. Second, a probability weighted average, when this condition is fulfilled, is equal to the true optimal solution of the original stochastic programming problem. A solution of the original problem could be obtained by the help of a macro program, where the objective function is a probability weighted utility function and constraint imposed such that the solution is admissible for all the different events. However, such a problem which, in addition, requires the imposition of the nonantipacivity constraints by e.g. using ordinary constraints, could be very large. A decomposition of this problem by solving individual augumented subproblems for each scenario in an iterative manner and combining these by the use of probability weights is easier.

The algorithm above has some interesting features. A common way of handeling complicated constraints in nonlinear programming is to put them into the objective function and penalize them, which is done in the above algorithm. As the iterations start it can be viewed as a battle between the independent scenario solutions and the average solutions. Average here means probability weighted solutions. As we start, the scenario solutions

dominate. By the updating of the penalty parameter $W_t^j(\omega_j)$, the scenario solutions, obtained from [27], converge to a compromise between the scenario solutions and the average solution, a solution where the nonantipacivity constraints are fulfilled.

5.8. Lagrange multiplier generation.

A problem in the procedure above is to find a proper value of the Lagrangian penalty parameter λ^1. The Wets and Rockafellar algorithm incorporates the nonantipacivity constraints by forming a modified objective function for the different scenarios. In this objective function we have two custom set parameters: the initial Lagrange vector λ^0 and the inital parameter c^0. In order to do this, we have to study the theory of augmented Lagrange functions.

The initialization of λ^0 and the scalar c^0 is somewhat arbitrary, but one has to consider several issues. First, if λ^0 is set too far away from its true optimal value, one may have to perform many iterations before converging to the optimal solution, such that a selection of λ^0 close to its optimal solution is preferred. When it comes to the selection of the scalar c^0, this scalar should not be too large such that ill-conditioning, or no interior solution at the first iteration becomes apparent. Or should the sequence c^l be chosen too small such that the algorithm converges slowly. When the algorithm is initiated we select a sequence λ^l and c^l. The sequence c^l should not increase too rapidly such that ill conditioning could be apparent at any iteration. This is also the case for λ^l. Thus we see that the choice of λ^0 and the scalar c^0 is a an issue in its own, and careful consideration should be taken in selecting these parameters. For an in depth discussion of Lagrange multiplier generation, see Bertsekas (1976).

5.9. Single period formulation.

The model discussed in this paper is multiperiodic. As opposed to a one-period model, it considers the effect of stochastic asset prices over several periods. One may ask whether these extensions add any value to the analysis. Is it so that if we only consider the first time period and construct a model corresponding to the first period of the information structure mentioned earlier, would we obtain a solution such that we end up close to the solution for the multiperiodic case? In order to answer this question, we must formulate the model as a one period model. In this case, there will be distinctly fewer scenarios or paths for the asset prices. If the investor only looks ahead one period into the future, the stochastic process is somewhat simpler. The information structure is given by:

[31] $f_0 = \{\{\omega_1, \omega_2, \ldots \omega_K\}\}$, $f_T = \{\{\omega_1\}, \{\omega_2\}, \ldots, \{\omega_K\}\}$, $T = 1$.

Using the same notation as before, the problem becomes[32]:

[32a] $\max E(U(\tilde{v}_1 - \rho_1))$

subject to:

[32b] $\tilde{v}_p^1 = \sum_{i=1}^{n} x_i^0 v_i^1$

[32c] $\Delta F_1^- - v_p^0(1+k) + v_p^1 \geq 0$

[32d] $F_1 = F_0 - \Delta F_1^- + \Delta F_1^+$

[32e] $\tilde{v}_p^1 + \Delta F_1^- - \Delta F_1^+ - v_p^0(1+k) = 0$

[32f] $y^1 + F_1 - k_1 \geq 0$

[32g] $z^1 + F_1 - k_2 \geq 0$

[32h] $\rho_1 = y^1\beta + z^1\alpha + u^t\gamma$

[32i] $u^1 + \Delta F_1^- \geq 0$

The problem above is still stochastic, so we could solve it by use of the Wets and Rockafellar algorithm. The problem above could also be reformulated to give us the deterministic equivalent program. We have done this below, and the variables in the model will be functions of the different events. This is handled in the notation below by letting the different variables be functions of the different events using parantheses.The program is given by[33]:

[33a] $\max \sum_{j=1}^{K} P(\omega_j)U(v_p^1(\omega_j) - \rho(\omega_j))$

subject to:

[33b] $v_p^1(\omega_j) = \sum_{i=1}^{n} x_i^0 v_i^1(\omega_j) \quad j \in 1,2,...,K.$

[33c] $\Delta F_1^-(\omega_j) - v_p^0(1+k) + v_p^1(\omega_j) \geq 0 \quad j \in 1,2,...,K.$

[33d] $v_p^1(\omega_j) + \Delta F_1^-(\omega_j) - \Delta F_1^+(\omega_j) - v_p^0(1+k) = 0 \quad j \in 1,2,...,K.$

[33e] $F_1(\omega_j) = F_0 + \Delta F_1^+(\omega_j) - \Delta F_1^+(\omega_j) \quad j \in 1,2,...,K.$

[33f] $y^1(\omega_j) + F_1(\omega_j) - k_1 \geq 0 \quad j \in 1,2,...,K.$

[33g] $z^1(\omega_j) + F_1(\omega_j) - k_2 \geq 0 \quad j \in 1,2,...,K.$

[33h] $u^1(\omega_j) + \Delta F_1^-(\omega_j) \geq 0 \quad j \in 1,2,...,K.$

[33i] $p_1(\omega_j) = y^1(\omega_j)\beta + z^1(\omega_j)\alpha + u^t(\omega_j)\gamma \quad j \in 1,2,...,K.$

The number of variables has increased, since we have to assign one variable to each event. Since the problem is now deterministic, we can solve it by the use of a nonlinear optimization algorithm. The deterministic equivalent problem above consists of 7×K constraints, and may be solved by using a nonlinear optimization package.

10. A numerical example.

In order to illustrate the model presented in this chapter, we will look at an example. We will look at a three–asset problem, and we will first set T equal to three. We will also consider a one–period version of the problem using the formulation [33]. The event three will split into three nodes at times 1,2, and 3 such that the total number of events or end nodes in the model is 27. We consider three utility functions. These utility functions are the logarithmic, negative exponential and quadratic utility functions. The three assets in the model are a low risk, medium risk and high risk asset. Below are listed the movements in the asset prices for one period.

	Up	Medium	Down
v_1	1.06	1.07	1.07
v_2	1.1	1.15	1.00
v_3	1.35	1.10	0.85

Table 5.2. Asset price returns – single period.

The most risky asset is asset number three, the asset with lowest risk is asset number one. Asset number two has medium risk. The numbers in the table above are on a return basis, such that if we have an up market in the forthcoming period, asset number three gives us a return of 35 %. In a rising market, the asset with lowest risk gives us a return of 7% etc. The table above gives the return from one period to another such that if the outcomes after two periods are two down movements, the value of asset one is 1.07^2. The expected return of each of the asset depends upon the probability of an up, medium or down movement in the asset prices. These probabilities can change from one period to another. In the example we have used, these probabilities are nonstationary, such that the probability of an up movement in period two depends upon the returns in the previous periods. The

probabilities assigned to the different events are selected such that the expected return from one period to another reflects the risk associated with the asset, which means that asset three always has the highest expected return and asset one always has the lowest return. The paths or events used are assigned the following probabilities:

Event	probability	Event	Probability	Event	Probability
uuu	0.05333	uum	0.05333	uud	0,05333
umu	0.048	umm	0.036	umd	0.036
udu	0.06	udm	0.03	udd	0.03
muu	0.05	mum	0.05	mud	0.05
mmu	0.0375	mmm	0.01875	mmd	0.01875
mdu	0.0375	mdm	0.01875	mdd	0.01875
duu	0.06	dum	0.045	dud	0.045
dmu	0.0375	dmm	0.01875	dmd	0.01875
ddu	0.045	ddm	0.015	ddd	0.015

Table 5.3. Probability and events in the example.

The table gives the different probabilities in the model. The probability of an up movement three times is 0.05333 and the probability of an up, down and up movement is 0.06. Since the probabilities for an up, medium or down movement depend upon the order of the movements, the process is nonstationary. The parameters used in the model are given the following values:

Parameter	Value
k	4%
k_1	0.035
k_2	0.025
α	0.2
β	0.2
γ	0.2
F_0	0.04
d_t	1.07^t
v_p^0	1

Table 5.4. Parameters in the model.

Above, the initial value of the portfolio is given by unity; consequently the interpretation of the invested amount in the first period is the fraction invested in the securities. The required return for each period is set at 4%. This is in harmony with the regulation of life insurance companies in Norway today. The penalty for shortfall of funds is set to 0.2, which is discounted to the horizon at a rate of 7% each year.

We will study three different utility functions. These utility functions are given by:

U_1	$Ln(v^T)$
U_2	$-Exp(-v^T)$
U_3	$v^T - \dfrac{1}{4}(v^T)^2$

Table 5.5. The utility functions considered.

The utility functions listed above are the logarithmic, negative exponential and quadratic utility functions. The event three studied in the example is trinomial, which means that each node, except for the terminal nodes, splits into three new nodes in the next time

~141~

period. Using the notation from [13]:

[34] $q \in \{up, medium, down\} = \{\zeta_1, \zeta_2, \zeta_3\}$

The stochastics in the event three represented by the probabilities are then:

Case	$P(\zeta_1)$	$P(\zeta_2)$	$P(\zeta_3)$
1	1/3	1/3	1/3
2	0.4	0.3	0.3
3	0.5	0.25	0.25
4	0.6	0.2	0.2

Table 5.6. The stochastics for the single periods used in the example.

For example, the root node in the event tree connects the nodes at time 1 using the probabilities in case 2 in Table 5.6. If the process evolved with two up movements, the probabilities are given by case 1 above.

If we consider the one–period version of the model, the optimal solutions, x^0, are given by the vector $[x_1^0, x_2^0, x_3^0]$ in the following table:

Case	U_1	U_2	U_3
1	[0,1,0]	[0,1,0]	[0,1,0]
2	[0.210,0.691,0.098]	[0.218,0.680,0.101]	[0.225,0.669,0.105]
3	[0.682,0,0.318]	[0.682,0,0.318]	[0,682,0,0.318]
4	[0.682,0,0.318]	[0.682,0,0.318]	[0.682,0,0.318]

Table 5.7.   Single period version – optimal solutions.

We see that the optimal solution for the one period formulation is sensitive to changes in the probabilities. However, it seems that the selection of the utility function does not affect the optimal solution that much. For all the portfolios in Table 5.7., the worst case return is nil. The probability of obtaining zero return ranges from 0.33 to 0.2, so the bankrupcy risk is

accepted to be quite high.

To test the model, we need a nonlinear optimization package able to handle nonlinear optimization. A package with this ability, when the number of variables in the model is not too large, is the Microsoft Excel Solver optimization package. The Solver, with the inclusion of the Microsoft Excel macro language, gives us the ability to implement the progressive hedging algorithm. After preliminary experiments with the one–period version of the model, we decided to set λ^l constant and equal to the unit vector. Since the squared term of [29] rapidly converges to zero as the nonanticipativity constraints turn active, we decided to increase the sequence c^l rapidly. In the algorithm we let $c^l = 0.5 \times 2^t$. Solving one problem, required approxemately 15 iterations.

In the three–period example, the first stage, or time zero, optimal solutions are given in the table below[2]:

U	Optimal first stage solution
U_1	[0.7538,0.211875,0.0343]
U_2	[1,0,0]
U_3	[1,0,0]

Table 5.8. First stage optimal solutions – Example.

The results in the table above differ significantly from the solution obtained by solving the one–period version of the model. In the case with the utility functions U_2 and U_3, the optimal solution is to invest everything in the asset with least risk. For the logarithmic utility function, the results point in the same direction since approximately 75 % should be invested in the least risky asset. The results in Table 5.8. show that a multiperiodic consideration of the investment process generates portfolios with much lower risk than in the single period case. The portfolio generated will give a worst case return, for all the utility functions under consideration, above the requirement of 4%.

An obvious question now is how to explain the difference between the solution obtained in the one–period version of the model, and that obtained from the three–period version of the model. One possible explanation is that unfavourable returns in the first period affect

---

[2] Here, $\bar{\kappa}$ is set to 0.0005.

penalties in the forthcoming periods. In order to avoid a withdrawal from the funds in an early period that is likely to be penalized in the next periods, it is optimal to purchase a low risk portfolio at time zero.

The results from the numerical example above indicate that complex models that run over many periods and nonstationary distributions for the asset prices should not be based upon a myopic decision rule.

5.11. Conclusions.

In this chapter, we have studied a discrete multiperiodic asset allocation model. The key issue in the model is the penalty imposed if the specified capital requirements are not fullfilled at certain points in time. We have presented a multiperiodic version and a one period version of the model. By considering an example, we computed the optimal solution to the problem and we discovered that the optimal solution, even for the Mossin (1968) myopic decision rule optimal utility function, could differ considerably between the one period formulation and the three–period formulation. Topics for future research are formulations of the model above in an infinite horizon setting and of models where the stochastics are given by a Markow transition probability matrix.

References:

Bertsekas, D.P. (1976) "Multiplier Methods: a survey.". Automatica, 12. 133–45.

Markowitz, H.M. (1952) Portfolio Selection. Journal of Finance. pp 77-91.

Mossin, J. (1968) "Optimal Multiperiod Portfolio Policies." Journal of Business. Vol 41. pp. 215–229.

Samuelson, P.A. (1969) "Lifetime Portfolio Selection by Dynamic Stochastic Programming." The Review of Economics and Statistics. pp. 239–246

Von Neuman, J. & Morgenstern, O. (1953) Theory of games and economic behaviour. (John Wiley & Sons)

Wets, R. J. B., Rockafellar, R.T. (1991) "Scenarios and Policy Aggregation in Optimization under Uncertainty." Mathematics of Operations Research. Vol 16 No.1 pp. 119–147

6. Efficient Generation of Mean-Gini Efficient Sets through Aggregation.

Abstract: This paper presents an efficient way to generate the mean-Gini efficient set. The Mean-Gini approach to analyzing risky prospects and to constructing optimum portfolios was introduced by Yitzhaki (1982) as an alternative to the traditional Markowitz (1952) approach. The Mean-Gini analysis provides necessary conditions for second degree stochastic dominance. Hence, by using the mean Gini approach we are prevented from choosing portfolios considered to be inferior by all risk averse investors. The generation of the mean-Gini efficient set requires the solution of a number of large scale linear programming problems. In this paper we present a way to solve these large scale linear programs efficiently by the use of an iterative aggregation method. The method is tested on real life data from the CRSP (University of Chicago Center for Research on Security Prices) data base.

6.1. Introduction.

The mean-Gini approach to analyzing risky prospects and to constructing optimal portfolios was introduced by Yitzaki (1982) as an alternative to the traditional Markowitz Mean-Variance analysis. The advantage of the mean-Gini approach is that it has the simplicity of the mean variance model, as it is an alternative two-parameter model. Furthermore, the mean-Gini approach also has the main features of stochastic dominance efficiency. Hence, the mean Gini difference can be shown to be more adequate than the mean variance for evaluating the variability of a prospect. The reason for this is that the mean-Gini approach to analyzing risky investments and generating efficient portfolios circumvents the problems inherent in the mean variance approach that are specific to the choice of probability distributions. Yitzhaki (1982) showed that the use of the mean and mean-Gini difference as summary statistics for a risky investment yields the derivation of necessary conditions for stochastic dominance. Hence the use of the mean-Gini difference as opposed to the variance enables an investor to discard from the efficient set portfolios which are considered inferior by all risk averse investors. Okunev (1988) and Shalit and Yitzhaki (1989) have done comparative studies between mean-Gini and mean-variance efficient portfolios. Both studies conclude that mean-Gini and mean-variance analyses provide analogous portfolios for the data set used in the series. This leads Okunev to suggest the use of the mean variance model as an approximation of the mean-Gini model. The reason for this suggestion is that the mean-Gini model requires a substantial amount of extra computation as compared with the computational requirements for the traditional mean-variance model. However, as stated by Yitzhaki and by Shalit and Yitzhaki, the advantages of the mean-Gini over the variance as a measure of risk and dispersion are substantial. The reason for this is that the mean-Gini method allows for the construction of efficient

-145-

portfolios that are in the set of stochastic dominant portfolios regardless of the probability distribution of the returns. Also, as pointed out by Shalit and Yitzhaki, since the Gini index is the expected absolute difference between two realizations of the value of a portfolio, it is a an intuitive measure of investment risk. This means that the mean-Gini approach would be preferable given that the efficient set of portfolios could be calculated in an efficient way. In this chapter we will make use of the theory of aggregation in linear programming to generate an efficient algorithm for the calculation of efficient mean-Gini portfolios. The paper is arranged as follows. In Section 6.2. we present the linear programming formulation for the generation of efficient mean-Gini portfolios and the corresponding dual program. Section 6.3. describes the basics in linear programming aggregation theory with special emphasis on an iterative reclustering method which has been shown to provide an efficient way of solving large scale linear programs as a sequence of small linear programs. In Section 6.4 we adapt this aggregation method to the mean-Gini portfolio problem and present an application of the method on a problem with ten stocks and thirty time period observations. We also present results for a real data set extracted from the CRSP (University of Chicago Center for Research in Security Prices) data base. Finally, in Section 6.5. we give our conclusions and suggestions for future research.

2. The linear programming formulation of the Mean-Gini Portfolio Problem.

In this paper we will concentrate our attention on the model where we use the estimated mean-Gini coefficient using time series where the returns for an asset are given on a daily, weekly, monthly or annual basis. For discrete observations the Gini mean difference is:

[1] $$\Gamma = \frac{1}{T^2} \sum_{t=1}^{T} \sum_{k>t}^{T} |r_t - r_k|$$

Above T is the total number of time periods in the sample. The estimator above represents the average of the absolute differences between all the possible pairs of random variables.

In order to construct the mean-Gini efficient frontier, without short sales, we need to solve the following programming problem [2] parametrically in the return.

[2a] $$\min \left\{ \frac{1}{T^2} \sum_{t=1}^{T} \sum_{q>t}^{T} |y_t - y_q| \right\}$$

subject to:

[2b] $y_t = \sum\limits_{i=1}^{n} r_{it} x_i \quad t \in 1,2,...,T.$

[2c] $\sum\limits_{i=1}^{n} x_i \hat{\mu}_i \geq \bar{\mu}$

[2d] $\sum\limits_{i=1}^{n} x_i = 1$

$x_i, y_t \geq 0 \quad i \in 1,2,...,n. \quad t \in 1,2,...,T.$

Here x_i are the nonnegative weights for the alternative assets. The return of a portfolio with weights x_i of the asset in period t is given by the equation [2b], where r_{it} is the return on asset i in period t. Hence the objective function in the mathematical programming problem is the discrete version of the mean-Gini difference given above. The average or mean return for the prospect i is given by $\hat{\mu}_i$ and is calculated from the observed values r_{it}.

It is easy to observe that this mathematical programming problem can be reformulated as a linear programming problem by introducing auxiliary variables, z_{jk}^+ and z_{jk}^-, by representing:

[3] $|y_j - y_q| = z_{jq}^+ + z_{jq}^-$

Hence the reformulated mathematical programming problem is[4]:

[4a] $\min \dfrac{1}{T^2} \sum\limits_{t=1}^{T} \sum\limits_{q>t}^{T} (z_{jq}^+ + z_{jq}^-)$

subject to:

[4b] $\sum\limits_{i=1}^{n} x_i(r_{it} - r_{iq}) - z_{jq}^+ + z_{jq}^- = 0 \quad t \in 1,2,...,T \text{ and } q > t.$

[4c] $\sum\limits_{i=1}^{n} x_i \hat{\mu}_i \geq \bar{\mu}$

[4d] $\sum\limits_{i=1}^{n} x_i = 1$

[4e] x_i, z_{tq}^+, $z_{tq}^- \geq 0$

The linear programming problem above has $T(T-1) + n$ variables and $T(T-1)/2 + 2$ constraints. Thus for a problem in which the number of observations T is large this is a very large scale linear program. For instance, for the German DAX index including 30 stocks the problem to be solved parametrically would consist of size 2682 variables and 1328 constraints given weekly observations for one year. However, if we use a longer time horizon and a finer observation interval, for instance an observation series of ten years on a daily basis, we will get a linear program with 6247530 variables and 3123752 constraints. Here we have used the assumption that a year has 250 stock exchange days. Thus the problem size grows incredibly fast with the number of observations. Also, we know that the number of observations used in the analysis should be substantially larger than the number of stocks under consideration in order to get a fair representation of the efficient set independent of whether the mean-Gini or the mean-variance approach is used.

Before we proceed, we state the dual to the reformulated linear programming model of the mean-Gini portfolio problem. The dual program is[5]:

[5a] $\max \mu V + W$

subject to:

[5b] $$\sum_{t=1}^{T} \sum_{q>t}^{T} (r_{it} - r_{iq}) V_{jq} + \mu_j V + W \leq 0 \quad i \in 1,2,...,n.$$

[5c] $-1 \leq V_{jq} \leq 1 \quad j = 1,2,...,n. \quad q>j$
[5d] $V \geq 0$
[5e] W free.

The problem above consists of n constraints and $T(T-1)$ constraints of upper and lower bound type. The number of variables is $T(T-1)/2 + 2$. However, we could convert this problem into a problem with only upper bounded variables. This reformulated dual can be stated as[6]:

[6a] $\max \quad \bar{\mu}(V^+ - V^-) + (W^+ - W^-)$

subject to:

[6b] $$\sum_{t=1}^{T} \sum_{q>t}^{T} (r_{it} - r_{iq}) \xi_{tq} + \bar{\mu}_i(V^+ - V^-) + (W^+ - W^-) \leq \sum_{t=1}^{T} \sum_{q>t}^{T} (r_{it} - r_{iq}) \quad i \in 1,2..,n.$$

[6c] $0 \leq \xi_{tq} \leq 2$ $t \in 1,2,...,T$ and $q > t$.

[6d] ξ_{tq} $t \in 1,2,...,T$. and $q > t$.

[6e] V^+, V^-, W^+, $W^- \geq 0$

Above we defined V as V^+ - V^- and W as W^+ - W^-. The problem has n constraints and T(T-1)/2 constraints of upper and lower bound type. The number of variables is T(T-1)/2 + 2. The portfolio weights will in this formulation be the dual prices of [6b]. The reformulated dual also has a perfect structure for the iterative reclustering and aggregation method to be presented in the next section. Hence this is the problem formulation of the mean-Gini portfolio problem that we will use in the remaining part of this chapter.

6.3. Aggregation in Linear Programming with Iterative Reclustering.

The concept of aggregation and disaggregation has been used in many scientific disciplines. An excellent survey by Rogers et. al (1991) on the use of aggregation techniques and methods in optimization has recently been published. One of the most predominant reasons for using aggregation/disaggregation techniques in optimization is the computational aspect. Even though the fast development of computers has enabled increasingly larger models to be solved, size limitations still exist and it might be more practical to solve optimization problems by using aggregation techniques. This is especially true for the computation of the mean-Gini efficient frontier where the number of variables can be very large. We will summarize the relevant elements of aggregation theory as applied to large, and possibly intractable original problems. We consider the standard linear program [7].:

[7a]   max c x

subject to:

[7b]   ax $\leq$ b
[7c]   x $\geq$ 0

Above c and x are n-vectors and an is a m×n matrix. Here x are the primal decision variables and u is the vector of dual variables. We use lower case letters for original or disagreggated parameters and variables, whereas upper case letters are used for aggregated parameters and variables.

Rogers et. al. (1991) emphasize that little research has been conducted on how to apply cluster analysis to linear programs. One of the exceptions is the work by Taylor (1983)

and Shetty and Taylor (1987) who examine several aspects of clustering. In a resent paper Aboudi, Jørnsten and Leisten (1996) present a method where the level of aggregation expressed by the number of clusters is constant throughout. The main feature of their method is that the clusters of variables or constraints change during the iterative procedure. This procedure will be summarized after the general notation for aggregation for general linear programming.

The main steps in most iterative schemes for clustering/disclustering variables in linear programming focus on the iterative change of aggregation weights within given clusters of aggregation. The methods are mainly based on the fact that there exist a set of optimal variable weights g_k^* for every cluster k with the property that the aggregate problem generated with these weights will yield the optimal solution to the original problem when disagregated. The main problem is that in order to generate g_k^* for all clusters k, we need to know the optimal solution (x^*, u^*) of the original problem. This solution is, ofcourse, not known in advance. The cluster notation is similar to that of Rogers et. al. (1991)[8]:

[8a] S_k = The set of the indices of variables in cluster k, k=1,2,...,K.

[8b] $S_k \cap S_{k'} = \emptyset$ for $k \neq k'$, $\bigcup_{k=1}^{K} S_k = N: = \{1,2,....,n\}$

[8c] $n_k = |S_k|$, where n_k is the number of variables in cluster k.

[8d] $g^k = [g_j]$, non negative n_k - vector whose elements sum up to unity, indicating the

variable weights g_j of the (disaggregated) variables in cluster k.

[8e] $g = [g^k] = [g_j]$, non negative n-vector, indicating the weights of all variables $c^k =$

$[c_j], j \in S_k$.

[8f] $c^k = [c_j] j \in S_k$.

[8g] $C_k = c^k g^k$, the aggregate objective function coefficient of the variables in cluster k.

[8h] $C = [C_k]$, the vector of coefficients of the aggregate objective function.

[8i] $a_j = [a_{ij}] j \in 1,2...,n.$, the j'th column of the disaggregated matrix a.

[8j] $a^k = [a_j], j \in S_k$, the disaggregate columns belonging to cluster k.

[8k] $A^k = a^k g^k$, the aggregate column belonging to cluster k.

[8l]　$A = [\,A^k\,]$ the aggregate coefficient matrix.

[8m]　$x^k = [\,x_j\,]$, $j \in S_k$, the disaggregate variables belonging to cluster k.

[8n]　X_k = the decision variable for cluster k.

We have used a makron for sets, variables and weights to indicate if a reclustering is done such that S_{k_1} and $\overline{S}_{k_1}$ denote sets before and after reclustering respectively.

The column aggregated problem is[9]:

[9a]　Max CX

subject to:

[9b]　$AX \leq b$
[9c]　$X \geq 0$

Once the aggregated linear program is solved, the solution to the original problem can be obtained either by fixed weight disaggregation where $x_j = g_j x_k^*$ where x_k^* is the solution to the aggregate problem. In the case when only columns have been aggregated this yields a feasible solution. Hence a lower bound on the optimal objective function value is obtained. On the other hand, in order to derive a better lower bound, optimal disagregation is used. See Zipkin (1977) and Liesegang (1980).

Here it is necessary to solve K smaller linear programs, one for each cluster of variables[10]:

[10a]　$\max c^k x^k$

subject to:

[10b]　$a^k x^k \leq A^k X_k^*$
[10c]　$x^k \geq 0$

Note that Zipkin (1980), Mendelsohn (1980) and Taylor (1987) provide upper bounds, a posterori error bounds, for the column aggregated linear programs. Aboudi, Jørnsten and Leisten (1993) have introduced an iterative reclustering algorithm for large linear programs

where all the variables have upper bounds. This general method can be applied very successfully to the problem of computing the mean-Gini optimal portfolio as every variable in the formulation has an upper bound. The motivation for this method arises from the theory of linear programming. Consider the problem[11]:

[11a]   max cx

subject to:

[11b] $ax \leq b$
[11c] $0 \leq x \leq w$

Above c, b and a are defined as usual, w is an n-vector of upper bounds, x is the vector of primal variables and u is the vector of dual variables. From linear programming theory, we know that there exists an optimal solution x^* where at most m variables are not at the lower or upper bounds. Hence, if we aggregate the large scale linear programming problem into m+2 clusters in which m clusters consist of a single variable, the m+1´st cluster contains the variable with value 0 in the optimal solution and the m+2nd cluster contains all variables that are at their upper bound. In addition, we know that at the optimal solution the variables in clusters 1,2,....,m. have reduced costs zero, whereas the reduced costs of the lower bound variables are nonpositive and the reduced costs of the upper bounds are nonnegative. We assume that the variables are renumbered so that the basic variables in the optimal solution are the first m variables so $S_k = \{k\}$ k=1,2,...,m. We assume that m+1 and m+2 contain the zero variables and upper and lower bound variables respectively. It is easy to derive weights for each of the variables so that the aggregate and the original problem have the same optimal solution.

Since typically the optimal solution is not known in advance, the procedure that was investigated improves the quantity of the aggregated solution iteratively by changing the assignment of variables to different clusters. The summary of the procedure follows as by Aboudi, Jørnsten and Leisten (1996). Further details, examples and computational reports are given in Aboudi, Jørnsten and Leisten (1996).

1. Initialization:

Set iteration counter l to zero. Generate m+2 approximately equal-sized clusters by assigning every variable to one cluster, e.g. by using their index or their objective function coefficient. Weight every variable within its cluster, e.g. by equal weights $g_j^{(l+1)} = 1/n_k$ or by their relative upper bound variable:

$$g_j^{l+1} = \frac{w_j}{\displaystyle\sum_{j \in S_k^{l+1}} w_j} \quad \text{for } j \in S_k^{l+1}$$

2. Construction and solution of the aggregate problem:

Suppose that we are in iteration l. Aggregate the problems (the variables) as described above and solve this problem. The primal and dual solution is $(X^{*(l)}, U^{*(l)})$.

3. Disaggregation.

Disaggregate the aggregate solution in each cluster k. The primal and dual solution of subproblem k is named ($x_j^{k(l)}, u_i^{k(l)}$)

Here we can choose between fixed weight or optimal disaggregation. If some stopping criterion is fulfilled stop. Otherwise go to 4.

4. Reclustering.

Either omit reclustering, i.e. use a fixed cluster assignment of variables (then go to 5.) or recluster the variables according to the disaggregate primal and dual variable values x_j and $c_j - U^*a_j$, as follows:

4.1. Choose those variables for the "upper bound cluster" whose reduced cost indicates an increase (i.e. they are > 0) and whose primal values are only a certain percentage given by β away from their upper bound.

4.2. Do the same for the "lower Bound Cluster" with variables with negative reduced costs and with primal values only a maximum percentage β away from the lower bound.

4.3. If fewer than m are unassigned yet, withdraw those with the largest distance from the relevant bound from the extreme clusters to get (at least) m nonassigned variables.

4.4. Cluster all remaining variables into m equal-sized clusters relative to their reduced cost values. (The clusters should be chosen in a way that in no cluster should variables with positive and negative reduced cost appear. So the remaining variables should be separated into those with negative and those with positive reduced cost.)

4.5. Weight transformation:

We now have to assign weights to the variables within their clusters in the next iteration.We will look at a transformation that only considers reclustering of one variable

between two clusters. The procedure can be used iteratively to recluster more than one variable. If we require that a disaggregation using the new weights will give us the same objective function value (althoug not optimal), and the same disaggregated variable solutions as using the old weights, we are assured that the new optimal solution obtained by solving the new aggregated problem has a nondecreasing objective function value from iteration to iteration.

We consider only reclustering between two sets given by S_{k_1} and S_{k_2}. The other sets remain unchanged. As defined above, the disaggregated solution within set k is obtained by setting $x_j = g_j X_k$. We have to find a new weighting $\bar{g}$ such that the solution found by fixed disaggregation using these new weights is the same as using the old disaggregation weights. For the clusters this condition can be stated as:

[12] $\quad x_j = g_j X_k = \bar{g}_j \bar{X}_k \quad \forall \; k \in 1,2,.....,K. \quad j \in S_k$

For the clusters S_{k_1} and S_{k_2} the conditions are:

[13] $\quad A_{ik_1} X_{k_1} + A_{ik_2} X_{k_2} = \bar{A}_{ik_1} \bar{X}_{k_1} + \bar{A}_{ik_2} \bar{X}_{k_2}$

[14] $\quad A_{ik_1} C_{k_1} + A_{ik_2} C_{k_2} = \bar{A}_{ik_1} \bar{C}_{k_1} + \bar{A}_{ik_2} \bar{C}_{k_2}$

In using a transformation which satisfies [12],[13] and [14] the problem is still feasible and the objective function value is unchanged but not neccesarily optimal. We now describe the conditions for the one variable reclustering procedure in more detail. For the two clusters $\bar{S}_{k_1}$ and $\bar{S}_{k_2}$ the following constraints have to be fulfilled:

For cluster $\bar{S}_{k_1}$ [15]:

[15a] $\quad x_j = g_j X_{k_1} = \bar{g}_j \bar{X}_{k_1} \quad j \neq j_1$

[15b] $\quad x_{j_1} = g_{j_1} X_{k_2} = \bar{g}_{j_1} \bar{X}_{k_1}$

[15c] $\quad \sum_{j \in S_{k_1}} \bar{g}_j = 1$

For cluster $\bar{S}_{k_2}$ [16]:

[16a] $\quad x_j = g_j X_{k_2} = \bar{g}_j \bar{X}_{k_2} \quad j \neq j_1$

[16b] $\quad \sum_{j \in S_{k_2}} \bar{g}_j = 1$

-154-

The system given by [15] and [16] is nonlinear and consists of $n_{k_1} + n_{k_2} + 2$ variables and $n_{k_1} + n_{k_2} + 2$ constraints. A solution to this system of equations is provided in Aboudi Jørnsten and Leisten (1993).

5. Weight updating:

Omit additional weight updating or use one weight-updating scheme. One could for example update the weights by using the relative value of the disaggregated variables after a possible reclustering. Hence the weights are[17]:

$$[17a] \qquad g_j^{t+1} = \frac{x_j^t}{\displaystyle\sum_{j \in S_k^{t+1}} x_j^t} \quad \text{for } j \in S_k^{t+1}$$

$$[17b] \qquad g_j^{t+1} = \frac{1}{n_k^{t+1}} \quad \text{if } \sum_{j \in S_k^{t+1}} x_j^t = 0$$

6.4. Application of the Aggregation Procedure to the Mean-Gini Portfolio Problem.

As can be seen from the reformulated dual in Section 6.2., the mean-Gini problem fits perfectly into the iterative reclustering and aggregation framework. The problem is a large scale linear programming problem with most of the variables having upper and lower bounds. Also the number of variables is significantly greater than the number of constraints.

We will first apply the aggregation procedure to a small sample with ten assets and thirty time periods with financial observations. The sample contains the following observations:

Security	Mean Return	Gini´s mean difference
S1	7.62%	3.385
S2	7.88%	3.531
S3	8.68%	3.630
S4	9.15%	3.803
S5	10.11%	4.406
S6	10.73%	4.596
S7	11.60%	5.002
S8	12.06%	5.546
S9	12.45%	5.926
S10	12.80%	7.204

Table 6.1. Observations in sample.

It should be observed that given the unbounded variables V and W are placed in separate clusters in the aggregation procedure, the optimal dual to the aggregated linear programming problem yields a feasible solution to the original primal problem: hence a feasible portfolio solution and a corresponding upper bound.

When we start the procedure we perform an initial clustering of the variables through creating five equally-sized clusters according to the lexicographic order of the variables. Also in the subsequent iterations we perform reclustering as described in Section 6.3. with the exception that we always restrict the number of clusters to five: one cluster for the variables at their upper bound and increasing, one for the variables at their lower bound and decreasing, one cluster for the basic variables, two clusters for increasing variables at their upper bound, and finally one cluster for the decreasing variables not at their lower bound. The variables V and W do not take part in the aggregation procedure, but are left as they are in the original problem. We also use fixed weight disaggregation and we calculate bounds in every iteration. The results are very promising, since by solving a number of problems with seven variables and ten constraints we were able to obtain a very good approximation of the Mean-Gini efficient set by solving very few small linear programs. The table and figures below give the results in compact form:

It.	Error Bound			(Real) Error		
	Av.	Max.	Mean Quadr. Dev.	Av.	Max.	Mean Quadr. Dev.
2	11.60 %	28.85 %	58.97	2.09 %	6.16 %	2.98
5	2.10 %	6.69 %	4.15	0.51 %	2.51 %	0.52
10	0.70 %	3.28 %	0.78	0.14 %	1.02 %	0.04
15	0.44 %	2.70 %	0.46	0.09 %	1.02 %	0.04
20	0.31 %	2.07 %	0.29	0.07 %	1.02 %	0.04

Table 6.2. Iterations and Error Bound.

Above is a table describing some of the numerical results from using the iterative reclustering aggregation algorithm on the 10 asset problem. A sufficient low error bound is obtained after 20 iterations.
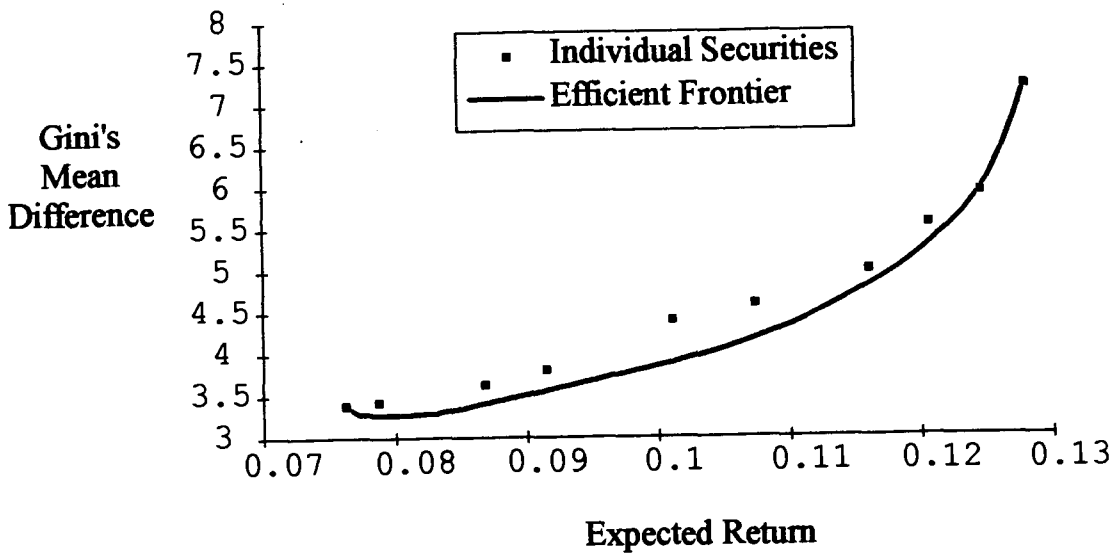
Fig. 6.1. Generated ex post efficient frontier.
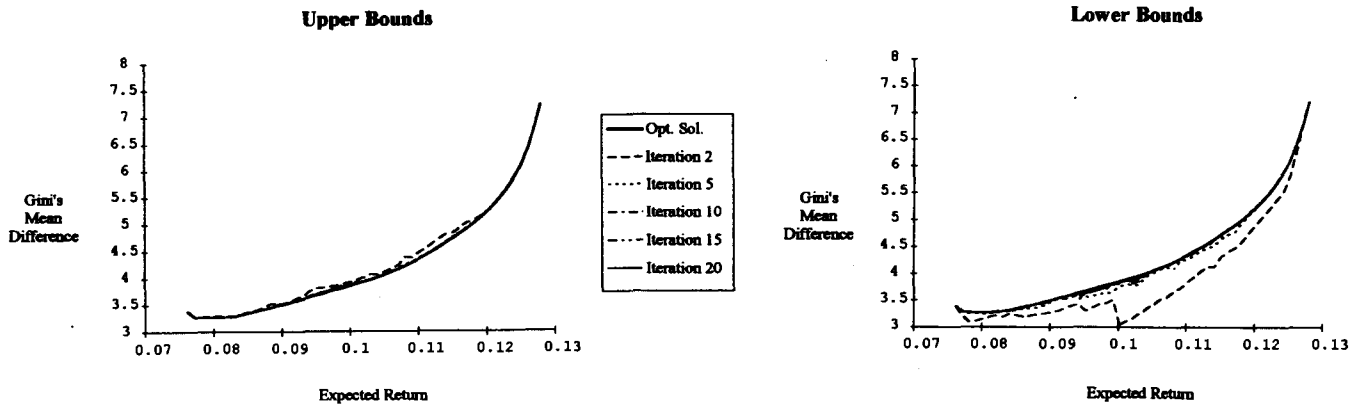
Fig. 6.2. Error bounds.

Above we have plotted the efficient frontier in the ten asset example. We have also plotted the error bounds for the frontier after some of the iterations.

We have also conducted experiments with an extracted financial data set taken from the CRSP data base. This set has 300 stocks and 300 periods of observations. Below we have plotted the stocks used in the analysis in the mean - mean-Gini difference diagram.
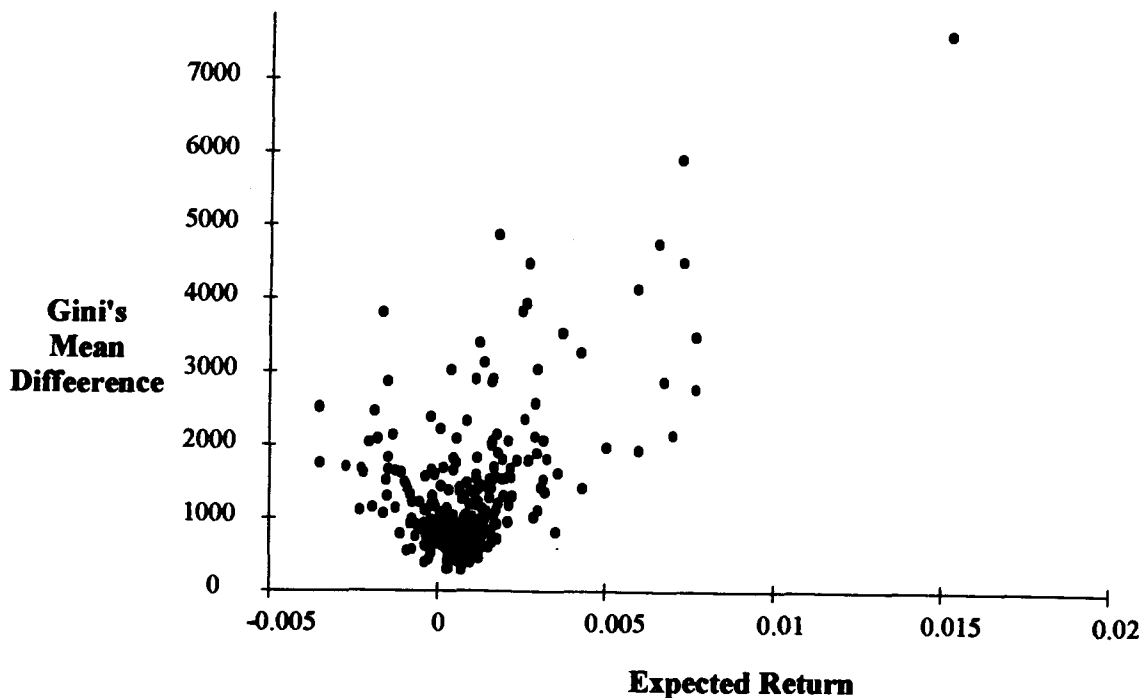
Fig. 6.3. Plot of the individual assets.

This leads to a reformulated Mean-Gini dual problem with 44852 variables and 300 constraints. The constraint matrix is very dense with almost 13.5 million nonzero matrix elements, thus representing a problem that is very difficult to tackle with standard linear programming procedures. We considered ten different equidistant expected daily return values ranging from the minimum value -0.00353 to the maximum value 0.015221 of the data and performed 20 iterations of the iterative aggregation procedure. The aggregation procedure was adapted as follows. For all ten test problems we start with a clustering of the variables according to their initial ordering and cluster into 46 equally sized clusters. The V and W variables are kept out of the aggregation. During the first three iterations we consider all 300 stocks, i.e. we solve Linear Programs of size 300*48. Reclustering is done as described in Section 6.3. From iteration four on, we consider only the stocks that have appeared in an optimal solution in at least one of iterations two and three. Clustering is now done into almost 500 clusters. This is done in order to obtain an LP problem with at most 15000 nonzero elements. Since this means that we have relaxed the original LP problem from iteration four on, we need to check for feasibility. If we detect infeasibility for the original problem in two consecutive iterations, we solve the problem with all 300 stocks in the next iteration. The results are presented in the figure and tables below.
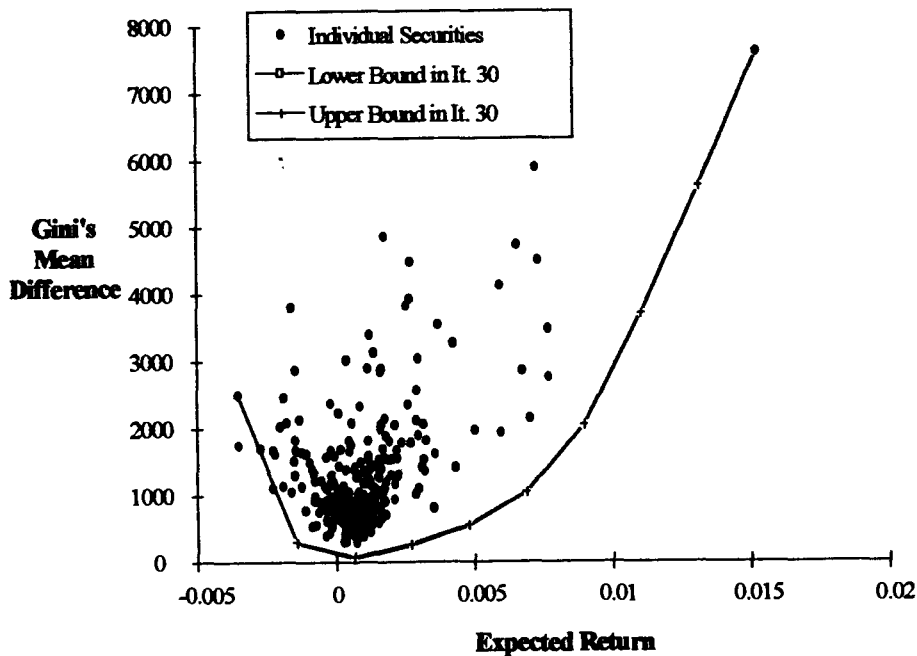
**Fig. 6.4. Ex post efficient frontier.**
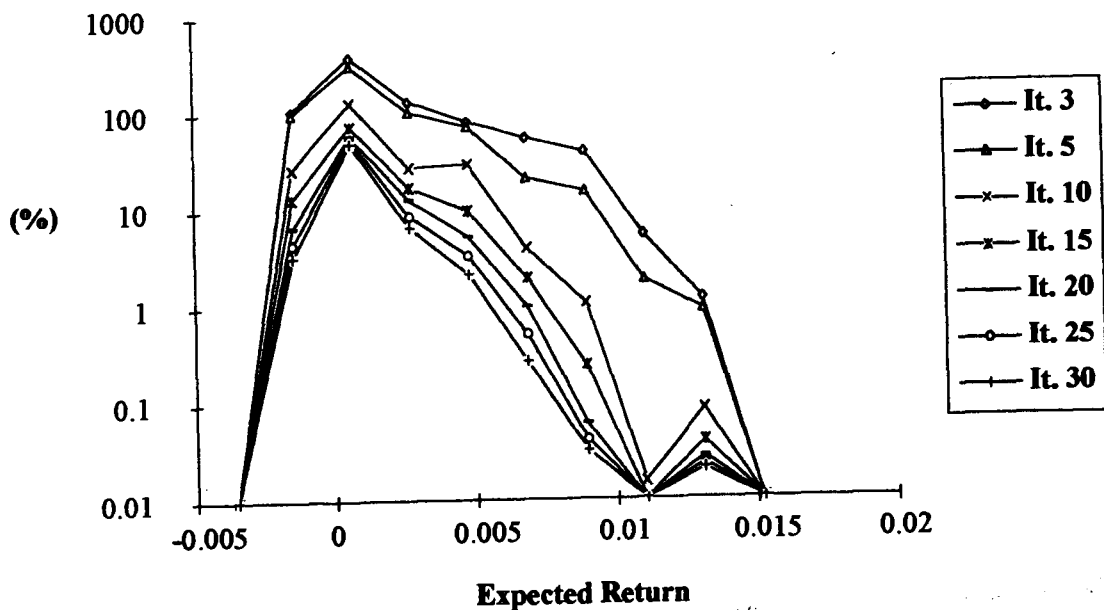
Fig. 6.5. Error bounds.

The optimal portfolios were generated using 30 iterations as seen by the efficient frontier and from the figure showing the error bounds for each iteration, the algorithm assures the convergence to the optimal solution.

6.5. Conclusions.

In this chapter we have studied a variant of the portfolio selection problem where the risk measure used is the Gini difference. As the formulation of the portfolio problem using this risk measure results in a large scale linear programming problem, we present an iterative aggregation and clustering algorithm to solve the problem. By first using a small sample and then using a lager sample of historical observations from the US stock market, we have illustrated the algorithm. The procedure seems to work well, as the optimal solution is typically obtained after 20-30 iterations. Possible areas of future research include empirical efficiency of the mean-Gini difference as compared to traditional measures of risk as standard deviation.

References:

Aboudi, R, Jørnsten, K., Leisten, R.(1996) "Solving Large-Scale Linear Programs with Bounded Variables by Iterative Aggregation." Submitted for publication in Operations Research, subject to revision.

Bey, R., Howe, K. (1984) "Gini Mean Difference and Portfolio Selection." Journal of Financial and Quantitative Analysis, Vol. 3. pp. 329-338.

Liesegang, D.G. (1980) "Aggregation bei Linearen Optimierungsmodellen.",

Habilitationsschrift, Universität zu Köln.

Markowitz, H.M. (1952) "Portfolio Selection." Journal of Finance. pp 77-91.

Mendelsohn, R. (1980) " Improved Bounds for Aggregated Linear Programs." Opsn. Res. 28, pp. 1450-1453

Okunev, J. (1988) "A Comparative Study of Ginis Mean Difference and Mean Variance." Accounting and Finance May pp. 1-15.

Okunev, J. (1989) "Mean Gini CAPM: Some Empirical Evidence", Accounting anf Finance Vol. 29.1, pp. 63-72.

Okunev, J. (1990) " An Alternative Measure of Mutual Fund Performance." Journal of Business, Finance and Accounting Vol. 17. 2.

Okunev, J. (1991) "The Generation of Mean Gini Efficient Sets." Jornal of Business Finance and Accounting Vol 18. pp. 209-218.

Okunev, J. (1992) "Is the Mean Gini CAPM Testable ?" JFBA vol. 19.2 pp. 271-76.

Rogers, D.F., Plante, R.D., Wong, R.T., Evans, J.R. (1991) "Aggregation and Disaggregation Techniques and Methodology in Optimization." Operations Research Vol. 39 pp. 553-82.

Shalit H., Yitzhaki S. (1984) "Mean Gini Portfolio Theory and the Pricing of Risky Assets." Journal of Finance Vol. 34 Des. pp. 1449-1467.

Shalit H., Yitzhaki S. (1989) "Evaluating the Mean Gini Approach to Portfolio Selection", The International Journal of Finance, Vol 1 No. 2 pp.15-31.

Shetty, C.M., Taylor, R.W. (1987) "Solving Large Scale Linear Programs by Aggregation." Comput. Opsn. Res 14, pp. 385-393.

Taylor, R.W. (1983) "Aggregate Programming in Large Scale Linear Systems. " Ph.D. Dissertation, Georgia Institute of Technology.

Yitzhaki S. (1982) "Stochastic Dominance, Mean Variance and Gini´s Mean Difference." American Economic Review Vol. 72, pp.178-85.

Zipkin, P.H. (1977) "Aggregation in Linear Programming." PhD. Dissertation, School of Organization and Management, Yale University, New Haven, Conn.

Zipkin, P.H. (1980) "Bounds on the Effect of Aggregation Variables in Linear Programs." Opns. Res. 28 pp. 403-418.

7. A Comparison Between Mean Variance, Mean Gini and Mean Absolute Deviation Portfolios.

Abstract: This paper investigates the relationships between mean variance, mean-Gini and mean absolute deviation portfolios. We present the three models and discuss their relations. We also point out the differences in the models with respect to their use on empirical data. Furthermore, we give a short summary of the history of the different models. Finally, we illustrate the differences in the ex post efficient portfolio frontiers for the three model concepts. The data used in the examples are from the Oslo Stock Exchange.

7.1. Introduction.

One of the most well known models in finance is the Markowitz portfolio optimization model, Markowitz (1952). This model is included in almost every textbook in finance, and is also a natural part of elementary courses in finance. However, it is often pointed out that the Markowitz model, in spite of its major influence on the development in the field, has not been used extensively in practice. One of the reasons for this is the large computitional burden involved with large quadratic programs. With the evolution of the modern computers, the computitional difficulties are at present not a severe problem for moderate size portfolio optimation models. It should be mentioned, as pointed out by Konno(1991), that for very large scale portfolio optimation model with more than 1000 assets, the computational requirement for the traditional Markowitz approach is still high. The reason for this computational difficulty is the density of the variance covariance matrix, which is the main component in the quadratic program.

The revival of the classical Markowitz model was partly a result of a paper by Markowitz and Perold (1981). In this paper the authors demonstrated that a large scale mean variance model can be solved by using a multifactor model, which has a spare matrix representation. Hence the parametric optimation model that must be solved in order to generate the efficient frontier can be efficiently handled by the use of sparse matrix techniques.

Due to the computational problems with the mean variance model in the 1960s and '70s, a number of attempts were made to approximate the quadratic function in the classical Markowitz model using linearisation techniques. Over the years, several linearisation techniques have been suggested. Most linearisation attempts have been constructed for computational reasons. Linearisation models based on separable programming techniques,

grid teqniques, succesive linearisation and matrix diagonalisation have been suggested. Among the more important suggestions are the linearisation approximations suggested by Sharpe(1967) and the linear programming model formulated by Stone (1973).

In parallel to the approximate models presented in the finance literature, are the suggestions made by researchers in agricultural economics. The most popular model within risk modelling in farm planning seems to be the MOTAD model, based on the alternative risk measure absolute devation. This model was originally developed by Hazzel (1971) as an approximation to the Markowitz mean variance model. In the literature on agricultural economics, the MOTAD and Target MOTAD models have been extensively used, see for instance Hazzel, Norton (1986). Also Thompson and Hazzel (1972) have made an extensive comparison of the differences between the MOTAD and mean variance model solutions. Furthermore there have been a number of articles in the agricultural economics literature that have focused on whether or not quadratic programming models, arising from risk models, should be approximated. The articles by McCarl and Tice (1982) and the article by Mc Carl and Onal (1989) give a good overview of this subject. It is interesting to observe that the suggestion made by McCarl and Tice (1982), that problems with more than 100 variables should be approximated, is loosened in the 1989 article by McCarl and Onal. However, when analysing their results, McCarl and Onal agree with Konno that very large scale quadratic programming problems are still computational difficult if they are dense.

In the finance literature, the mean absolute deviation aproach has not received the same attention as in the agricultural economics literature. The exception is the paper by Ang(1975). However, recently the topic has been the subject of a renewed interest, and it seems that it has been rediscovered. The set article by Konno (1991) uses a mean absolute devation model. It should also be noted that the focus in this article is completely different from the focus in the MOTAD literature. Konno et. al. suggest that the mean absolute deviation should be used as an alternative measure of risk, and hence the model developed is not to be seen as an approximation of the traditional Markowitz mean variance model. To mention one of the features of the mean absolute deviation risk measure, we note that the model is not based upon any probability distribution about the return. Furthermore, the mean absolute deviation model uses the historical time series data "as it"; hence the estimation procedure is obsolete. As pointed out by Hazzel and Thompson (1986), "The measure of income variance needed in quadratic programming is only a statistical estimate of the true variance". As such, there is no reason why alternative estimates of variance should not be used, particulary those that can be calculated from linear estimators. It is also observed that the portfolio generated by using the mean absolute deviation model is similar to the portfolios generated by using the traditional Markowitz model. Also, it has been observed that the number of stocks in an efficient portfolio is smaller in the mean

absolute deviation approach than the number of stocks in an efficient portfolio generated by the mean variance approach. This is of practical interest since its more difficult to manage a portfolio consisting of several hundred assets, some of which should be held in very small amounts. Also, the existence of transaction costs motivates an investor to hold few stocks in a portfolio and in relatively large amounts.

The Mean-Gini approach to analyzing risky prospects and to constructing optimal portfolios was introduced by Yitzaki (1982) as an alternative to the traditional Markowitz mean variance analysis. The advantage of the mean-Gini approach is that it has the simplicity of the mean variance model as it is an alternative two parameter model. Furthermore, the mean-Gini approach also has the main features of stochastic dominance efficiency. Hence the mean Gini difference can be shown to be more adequate than the mean variance for evaluationg the variability of a prospect. The reason for this is that the mean-Gini approach to analyzing risky investments and generating efficient portfolios circumvents the problems inherent in the mean variance approach that are specific to the choice of probability distributions. Yitzhaki (1982) showed that the use of mean and mean Gini difference as summary statistics for a risky investement yields the derivation of necessary conditions for stochastic dominance. Hence the use of the mean-Gini difference as opposed to the variance enables an investor to discard from the efficient set portfolios of projects which are considered inferior by all risk averse investors. Okunev (1988) and Shalit and Yitzhaki (1989) have done comparative studies of mean-Gini and mean–variance efficient portfolios. Both studies conclude that mean-Gini and mean-variance analyses provides analogous portfolios for the data set used in the series. This leads Okunev to suggest the use of the mean–variance model as an approximation of the mean-Gini model. The reason for this suggestion is that the mean-Gini model requires a substantial amount of extra computation as compared with the computational requirements for the traditional mean-variance model. However, as stated by Yitzhaki and by Shalit and Yitzhaki, the advantage of the mean-Gini over the variance as a measure of risk and dispersion are substantional. The reason for this is that the mean-Gini method allows for the construction of efficient portfolios that are in the set of stochastic dominant portfolios regardless of the probability distribution of the returns. Also, as pointed out by Shalit and Yitzhaki, since the Gini index is the expected absolute difference between two realizations of the prospects outcomes it is an intuitive measure of investement risk. This means that the mean-Gini approach would be preferable given that the efficient set of portfolios could be calculated in an efficient way. In this chapter we point out the relations and differences in the various models and present some experimental results. In Section 7.2. we give the mathematical formulation of the three models. Section 7.3. contains the result of using the models on data from the Oslo stock exchange.

7.2. Mathematical formulation of the models.

In this section we present the mathematical formulation of the classical Markowitz model, the mean absolute deviation model and the mean-Gini model.We start with the MV model. Let there be N securities indexed by i, j = 1,2....N. Let $\tilde{r}_i$ be a random variable representing the rate of return on security S_j. The variables x_j is the fraction of the total amount of money invested in security i. Here the total amount invested is normalized to one. The expected return for the investment made in a portfolio of the available assets is:

$$[1] \quad E\left(\sum_{i=1}^{n} \tilde{r}_i x_i\right) = \sum_{i=1}^{n} \mu_i x_j$$

where $E(\bullet)$ is the expectation operator of a random variable. We denote the expected return on asset i by μ_i. In the classical Markowitz approach the model is based on the following assumptions.

i) The investor has a desire for high expected returns.

ii) The investor considers variability of returns to be undesirable, so the investor prefers less variability to more. Variability is associated with risk and is measured by the variance of the portfolio.

iii) The investor bases his portfolio decision on the principle of utility maximation.

iv) The investors follow the rules of rational behaviour.

v) All assets have strictly positive variance.

vi) Investors make decisions about investments in the assets at the begining of the period.

vii) The fraction invested in the assets can be any $\{x \in \Re \mid x \in [0,1]\}$.

viii) At least two assets have unequal expected return.

ix) The assets have finite expectation and variance.

The variance of return for the portfolio is:

$$[2] \quad \sigma_p^2 = E\left[\left(\sum_{i=1}^{N} \tilde{r}_i x_i - \sum_{i=1}^{N} \mu_i x_j\right)^2\right]$$

Furthermore, let us denote the variance covariance matrix by $V = |\sigma_{ij}|$.

The classical Markowitz model without short sales can now be formulated as[3]:

$$[3a] \quad \min \sum_{j=1}^{n}\sum_{i=1}^{n} x_i \sigma_{ij} x_j$$

subject to:

$$[3b] \quad \sum_{j=1}^{n} \mu_i x_j \geq \mu$$

$$[3c] \quad \sum_{j=1}^{n} x_j = 1$$

$$[3d] \quad x_j \geq 0$$

This model has been much studied in the past and its mathematical properties have been thoroughly investigated. See Huang and Litzenberger (1988) and the book by Markowitz (1987) for details.

In order to be able to use the model we must have input data. As these parameters are unknown, it is neccesary to obtain estimates for them. In practice these estimates are obtained using time series or cross-sectional data of observed returns. If needed, these estimated values can be adjusted based on other available subjective information. It should also be observed that in order to be able to use the model, we must compute $n(n+2)/2$ constants σ_{ij}, $i,j = 1,2...n$, from the historical or projected data. Since this matrix is dense, it means consist of nonzeros almost everywhere, and the computational burden of solving this large scale quadratic programs is substantial even when the fast development of computer hard and software is taken into account.

As an alternative to the mean variance model the use of another measure of uncertainty has been suggested, i.e. the mean absolute deviation. In the literature in agricultural economics this was suggested by Hazzel (1971). Hazzel´s MOTAD model became very popular in agricultural economics as an approximation to expected value - variance analysis. The main reason for the popularity of the MOTAD model is that it leads to a linear program. Even large size models could be solved with the mean absolute value as a risk measure in portfolio models as early as twenty years ago. Also LP-codes were very easily available

for standard use. Recently, the mean absolute deviation model has undergone a revival in finance through the studies by Konno. Konno's motivation for using the mean-absolute deviation as the measure of uncertainity is also based on the fact that it leads to a linear program instead of a quadratic program (Konno 1991). This also means, as pointed out by Konno, that integer constraints associated with real transaction costs can easily be incorporated in the model without making the model unsolvable. The MOTAD formulation presented by Hazzel is as follows[4]:

$$[4a] \quad \min \sum_{t=1}^{T} z_t^+ + z_t^-$$

subject to:

$$[4b] \quad \sum_{i=1}^{n} (r_{it} - \hat{r}_j) x_j - z_t^+ + z_t^- = 0$$

$$[4c] \quad \sum_{i=1}^{n} \hat{\mu}_i x_j \geq \bar{\mu}$$

$$[4d] \quad \sum_{i=1}^{n} x_j = 1$$

$$[4e] \quad x_i \geq 0 \quad i = 1,2,....N$$

Also since:

$$[5] \quad \sum_{t=1}^{T} z_t^+ = \sum_{t=1}^{T} z_t^-$$

we have the alternative formulation[6]:

$$[6a] \quad \min \sum_{t=1}^{T} z_t^-$$

subject to:

$$[6b] \quad \sum_{i=1}^{n} (r_{it} - \hat{\mu}_i) x_j + z_t^- \geq 0$$

$$[6c] \quad \sum_{i=1}^{n} \hat{\mu}_i x_j \geq 0$$

$$[6d] \quad \sum_{i=1}^{n} x_j = 1$$

$$[6e] \quad x_i \geq 0 \quad i = 1,2,....n. \quad z_t^- \geq 0 \quad t = 1,2,....T.$$

Konno uses a third formulation of the mean absolute deviation model which is[7]:

$$[7a] \quad \min \sum_{t=1}^{T} y_t$$

subject to:

$$[7b] \quad y_t - \sum_{i=1}^{n} x_i (r_{it} - \bar{r}_i) \geq 0 \quad t \in 1,2,......T.$$

$$[7c] \quad y_t + \sum_{i=1}^{n} x_i (r_{it} - \hat{\mu}_i) \geq 0 \quad t \in 1,2,......T.$$

$$[7d] \quad \sum_{i=1}^{n} x_i \hat{\mu}_i \geq \bar{\mu}$$

$$[7e] \quad \sum_{i=1}^{n} x_i = 1$$

$$[7f] \quad x_i \geq 0 \quad y_t \geq 0$$

In the two first models it is easily seen that if a feasible solution exists, there is an optimal solution with at most T+2 of the variables with positive values. We also notice that if we square the variables z_t^- in the objective function of [6] and allow z_t^- to take negative values, [6] is converted to the sample mean variance model. In this model, one obtains the sample variance by dividing the optimal objective function value by T − 1. Also by looking carefully at the sample mean variance model and the mean absolute devation model it becomes clear that the model structure is the same, and it is thus not remarkable that the solution is of the same structure. Konno (1991) performed a comparative study of the mean variance and the mean-absolute devation model using data from the Tokyo Stock Exchange. One of the results was that the mean absolute deviation model and the mean variance model generated similar portfolios.

We now turn our attention to the Mean-Gini model. In this model we assume that the risk measure representing the preferences to the investor is the Mean-Gini difference. For discrete observations the Gini mean difference is:

$$[8] \quad \Gamma = \frac{1}{T^2} \sum_{t=1}^{T} \sum_{q>t}^{T} |r_t - r_k|$$

Above, T is the total number of time periods in the sample. The estimator above represents the average of the absolute differences between all the possible pairs of random variables.

In order to construct the Mean-Gini efficient frontier, without short sales, we need to solve the following programming problem [9] parametrically in the return.

$$[9a] \quad \min \left\{ \frac{1}{T^2} \sum_{t=1}^{T} \sum_{q>t}^{T} |y_t - y_q| \right\}$$

subject to:

$$[9b] \quad y_t = \sum_{i=1}^{n} r_{it} x_i \quad t \in 1,2,...,T.$$

$$[9c] \quad \sum_{i=1}^{n} x_i \hat{\mu}_i \geq \bar{\mu}$$

$$[9d] \quad \sum_{i=1}^{n} x_i = 1$$

x_i, $y_t \geq 0$ $i \in 1,2,...,n$. $t \in 1,2,...,T$.

Here x_i are the nonnegative weights for the alternative prospects. The return of a portfolio with weights x_i of the asset in period t is given by the equation [9b], where r_{it} is the return on asset i in period t. Hence the objective function in the mathematical programming problem is the discrete version of the mean-Gini difference given above. The average or mean return for the prospect i is given by μ_i and is calculated from the observed values r_{it}.

It is easy to observe that this mathematical programming problem can be reformulated as a linear programming problem by introducing auxiliary variables, z_{jk}^+ and z_{jk}^-, by representing:

$$[10] \quad |y_j - y_q| = z_{jq}^+ + z_{jq}^-$$

Hence the reformulated mathematical programming problem is[11]:

$$[11a] \quad \min \frac{1}{T^2} \sum_{t=1}^{T} \sum_{q>t}^{T} (z_{jq}^+ + z_{jq}^-)$$

subject to:

$$[11b] \quad \sum_{i=1}^{n} x_i(r_{it} - r_{iq}) - z_{jq}^+ + z_{jq}^- = 0 \quad t \in 1,2,...,T \text{ and } q > t.$$

$$[11c] \quad \sum_{i=1}^{n} x_i \mu_i \geq \bar{\mu}$$

$$[11d] \quad \sum_{i=1}^{n} x_i = 1$$

$$[11e] \quad x_i, z_{tq}^+, z_{tq}^- \geq 0$$

The linear programming problem above has $T(T-1) + n$ variables and $T(T-1)/2 + 2$ constraints. Thus for a problem in which the number of observations T is large this is a very large scale linear program. For instance, for the German DAX index including 30 stocks the problem to be solved parametrically would contain 2682 variables and 1328 constraints, given weekly observations for one year. However, if we use a longer time horizon and a finer observation interval, for instance an observation series of ten years on a

daily basis, we will get a linear program with 6247530 variables and 3123752 constraints. Here we have used the assumption that a year has 250 stock exchange days. Thus the problem size grows incredibly fast with the number of observations. Also, we know that the number of observations used in the analysis should be substantially larger than the number of stocks under consideration to get a fair representation of the efficient set, independent of whether the mean-Gini or the mean variance approach is used.

Before we proceed, we state the dual to the reformulated linear programming model of the mean-Gini portfolio problem. The dual program is[12]:

[12a]   max $\mu V + W$

subject to:

[12b] $\displaystyle\sum_{t=1}^{T}\sum_{q>t}^{T}(r_{it} - r_{iq})V_{jq} + \mu_j V + W \le 0$ $i \in 1,2,...,n.$

[12c] $-1 \le V_{jq} \le 1$ $j = 1,2,...,n.$ $q>j$

[12d] $V \ge 0$

[12e] W free.

The problem above consists of n constraints and T(T-1) constraints of upper and lower bound type. The number of variables is T(T-1)/2 + 2. However, we could convert this problem into a problem with only upper bounded variables. This reformulated dual can be stated as[13]:

[13a]   max $\overline{\mu}(V^+ - V^-) + (W^+ - W^-)$

subject to:

[13b] $\displaystyle\sum_{t=1}^{T}\sum_{q>t}^{T}(r_{it} - r_{iq})\xi_{tq} + \overline{\mu}_i(V^+ - V^-) + (W^+ - W^-) \le \sum_{t=1}^{T}\sum_{q>t}^{T}(r_{it} - r_{iq})$ $i \in 1,2..,n.$

[13c] $0 \le \xi_{tq} \le 2$ $t \in 1,2,...,T$ and $q > t.$

[13d] ξ_{tq} $t \in 1,2,...,T.$ and $q > t.$

[13e] $V^+, V^-, W^+, W^- \ge 0$

Above we defined V as $V^+ - V^-$ and W as $W^+ - W^-$. The problem has n constraints and

T(T-1)/2 constraints of upper and lower bound type. The number of variables is T(T-1)/2 + 2. The portfolio weights will, in this formulation, be the dual prices of [13b]. The reformulated dual also has a perfect structure for the iterative reclustering and aggregation method. For this problem a column generation strategy can be successful. For a description of this procedure, we refer to Aboudi, Jørnsten and Leisten (1996). The reformulated dual also has a perfect structure for the iterative reclustering and aggregation method described earlier in this dissertation. Hence this is the problem formulation of the Mean-Gini portfolio problem that we will use in the remaining part of this chapter.

7.3. Some empirical results from the three alternative models.

In our study we use 21 stocks listed on the Oslo Stock Exchange. The data ranges from the period of 1 Jan. 1987 to 31 Des. 1992, the same dataset described earlier in the thesis. Selection of the stocks was based upon several criteria. First, we avoided stocks with low liquidity, such that stocks traded infewerthan 150 trading days during a year on average were excluded from the set. Second, the stock had to be a member of the trading list over the whole period. If mergers had taken place, we allowed for this by adjusting the new stock with the switch ratio offered to the old stockholders in the two companies. For sectors with special events like the bank and insurance sector, we have the bank and insurance index instead. All return series were adjusted for splits, emissions and dividends to preserve the true return of each asset. The data was taken from the database Amadeus at the Norwegian School of Economics and Business Administration.
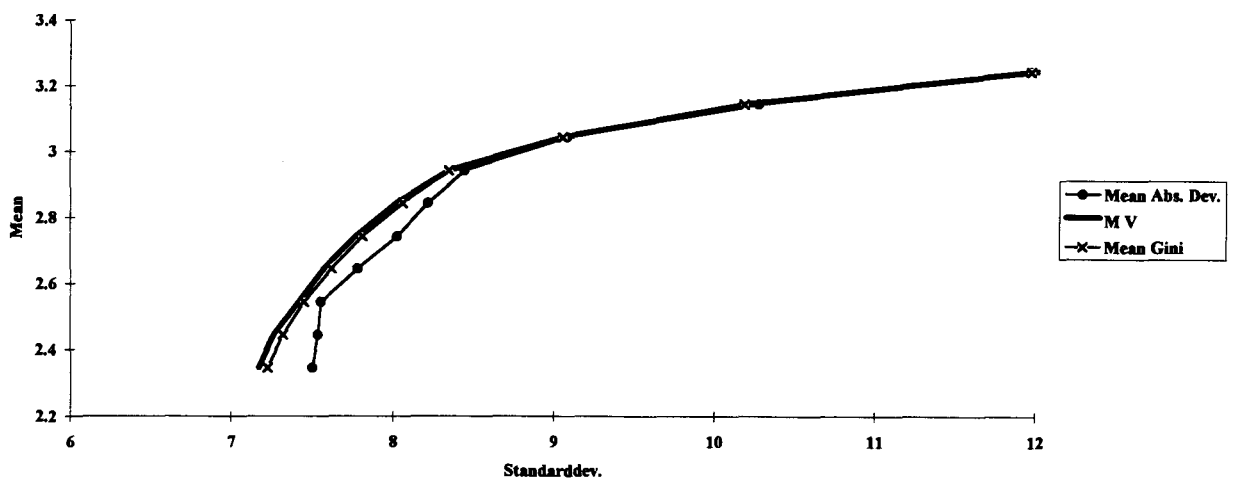
Fig. 7.1. Ex post portfolio frontiers.

Above, we have a graph with the results from the three models. We have computed the standard deviation for all portfolios, and plotted the generated portfolios into the mean

standard deviation diagram. As shown by the graph, it seems that the difference between the different models is most apparent between the mean absolute deviation model and the mean variance model. The mean Gini model and the mean variance model seem to generate similar portfolios as measured by the risk measure standard deviation. We will now look at the quality differences between the models.

As a measure of quality differences between two portfolios, p_1 and p_2, we use the measure:

$$[14] \quad C = \frac{1}{2} \sum_{i=1}^{n} | x_i^{p_1} - x_i^{p_2} |$$

where:

$x_i^{p_1}$ = Weight of asset i in portfolio p_1.

$x_i^{p_2}$ = Weight of asset i in portfolio p_2.

Below is a graph showing the qualitative differences between the portfolios.
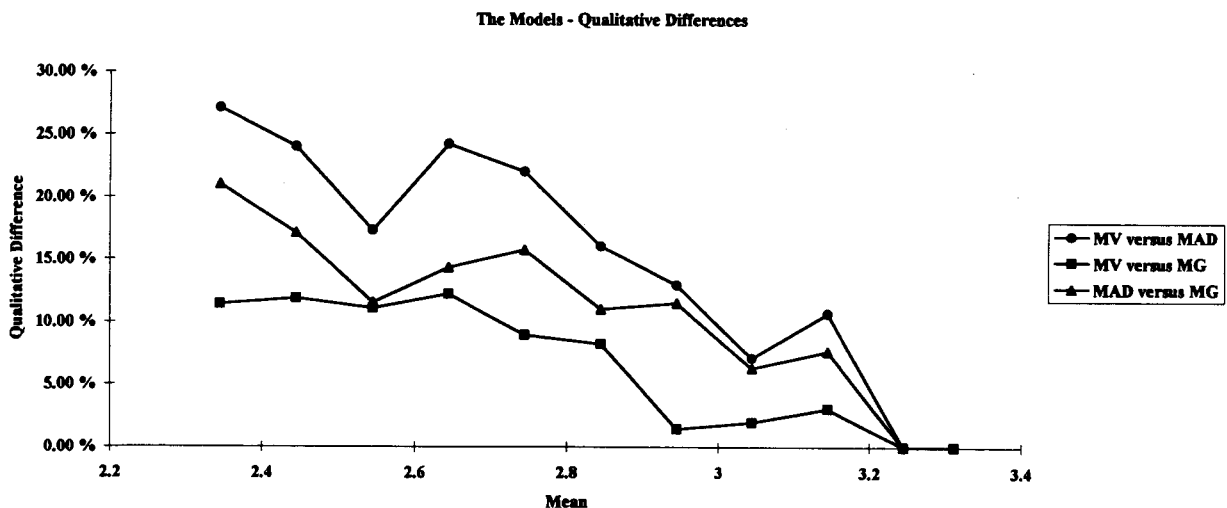
Fig. 7.2. Quality differences.

It seems that the quality difference between the models decreases as a function of the risk level, although in the case with the mean absolute deviation model versus the mean variance model, the quality difference is for many values of the expected return above 20%. The

difference between mean–Gini and mean–variance is smaller, whereas the difference between the mean absolute deviation and the mean Gini is typically in the range from 10-20 %. The fact that mean–Gini and mean–variance portfolios are similar seems to be in line with the observations of Okunev (1988) and Shalit and Yitzhaki (1989).

7.4. Conclusions.

In this paper we have given an overview of the history of mean absolute deviation optimation model and discussed similarities and differences between this model and the traditional model. Furthermore, we have had a look at the direct MV model and discussed the relations between this model and the traditional MV model. We have also presented the Mean Gini model. Further, we have discussed advantages and disadvantages of the models under different assumptions about the underlying distribution, and finally we have presented some empirical results of the models using data from the Oslo Stock Exchange.

References:

Aboudi, R, Jørnsten, K., Leisten, R.(1996) "Solving Large-Scale Linear Programs with Bounded Variables by Iterative Aggregation." Submitted for publication in Operations Research, subject to revision.

Ang, J. (1975) "A Note on the E, SL Portfolio Selection Model. Journal of Financial and Quantitative Analysis Dec 1975 pp. 849-857.

Hazzel, P. (1971) "A Linear Alternative to Quadratic and Semivariance Programming for Farm Plans under Uncertainty." American Journal of Agricultural Economics 53 pp 53-62.

Hazzel, P.Thompson, J. (1986) Mathematical Programming for Economic Analysis in Agricultural Economics.

Huang, C., Litzenberger, R., H., (1988) Foundations for Financial Economics. North Holland.

Konno, H. (1991) "Mean-Absolute Deviation Portfolio Optimation Model and its Application to the Tokyo Stock Market." Management Science Vol 37 May pp.519-531.

Konno, H. (1990) "Piecewice Linear Risk Functions and Portfolio Optimization." Journal of the operations Research of Japan Vol 33. No 2. pp 139-156.

Okunev, J. (1991) "The Generation of Mean Gini Efficient sets." Journal of Business Finance & Accounting. 18. pp.209-218.

Perold, A. (1984) "Large Scale Portfolio Optimations." Management Science 30 pp. 1143-1160.

Markowitz, H.M. (1952) Portfolio Selection. Journal of Finance. pp 77-91.

Markowitz, H.M. (1959) Portfolio Selection. (John Wiley & Sons NY)

Markowitz, H.M. (1987) Mean Variance Analysis in Portfolio Choice and Capital Markets (Basil Blackwell).

Shalit H., Yitzhaki S., (1989) "Evaluating the Mean Gini Approach to Portfolio Selection." The International Journal of Finance, Vol. 1 No. 2 pp. 15–31.

Sharpe, W.F. (1967) "A Linear Programming Algorithm for Mutual Fund Portfolio Selection." Management Science Vol 13. pp. 499-510.

Stone, B.K. (1973) "A linear Programming Formulation of the General Portfolio Selection Problem." Journal of Financial and Quantitative Analysis.pp. 621-636.

Yitzaki, S., (1982) "Stochastic Dominance, Mean Variance and Gini´s Mean Difference." American Economic Review, Vol 72. pp. 178-85.

8. A Least Element Interpretation of a Forward Backward Recursive Finite Difference Based Algorithm for the Pricing of American Derivatives.

Abstract: This chapter studies numerical solution techniques that are used in the pricing of American derivatives. The pricing problem for certain American options could, by finite difference discretization, be shown to result in, for each time step, a linear complimentary problem. By using results provided by Mangasarian (1978), this problem could be converted to a ordinary linear program. For one particular case, when the explicit finite differences are used, we give a simple evaluation formula for the solution to the problem. In cases where the approximation method is the Crank-Nicholson, implicit finite differences, or more generally the θ-method, we present the simple Forward Backward Recursive algorithm (FBR) for the problem. Similar algorithms have been presented for the traditional American put option problem by Brennan and Schwartz (1977). One of the advantages of this algorithm is that it provides a relatively fast means of solving free boundary problems, even in the case one have involving the implicit finite difference approximation. Finally, we discuss the efficiency of the algorithm as compared to the PSOR (Projective Succesive Over Relaxation) and linear programming algorithms.

8.1. Introduction.

This paper concerns solution procedures to discrete versions of the free boundary problem. In finance, one particular application of the problem is in the pricing of American derivatives. It can be shown that by applying a set of transformations, the pricing of American derivatives could be reduced to a problem to be discussed below.

If we let $u(x,\tau)$ and $g(x,\tau)$ be two functions of the variables x and τ, and if we let $u(x,\tau)$ be twice differentiable, the problem under investigation is to find a function $u(x,\tau)$ that satisfies the following equations[1]:

[1a] $\quad \dfrac{\partial u}{\partial \tau} - \dfrac{\partial^2 u}{\partial x^2} \geq 0$

[1b] $\quad u(x,\tau) - g(x,\tau) \geq 0$

[1c] $\quad \left(\dfrac{\partial u}{\partial \tau} - \dfrac{\partial^2 u}{\partial x^2} \right) (u(x,\tau) - g(x,\tau)) = 0$

We term the problem above the continious complimentarity problem (CCP). This problem is typically extended by a set of boundary conditions which could, as in the case with an American put option, have the form:

[1d] $u(x,0) = \max(e^{\frac{1}{2}(c_1 + 1)x} - e^{\frac{1}{2}(c_1 - 1)x}, 0)$

[1e] $\lim_{x \to \infty} u(x,\tau) = 0$

In [1d] c_1 is a constant. Other American options that could be reduced to the form above include lookback options (continious sampling), barrier options (where boundary conditions are handled by the method of images), cash or nothing options and calls with dividends. For further details, see Wilmott, Dewynne and Howison (1993).

Finite difference methods are well known as a solution technique to the partial differential equation that arises in derivative pricing. The explicit finite difference method is easily implemented. The implicit finite difference method, on the other hand, is somewhat trickier, but in the literature the method is implemented on American options using an iterative algorithm such as the PSOR or by "equation feasibility testing", as in the case of Brennan and Schwartz (1977). For further details see Wilmott, Dewynne and Howison (1993). One advantage of the implicit difference method as compared to the explicit finite difference method is that the explicit finite difference method has limitations, since the number of mesh points allowed along the x dimension is limited as related to the number of mesh points along the τ-axis. Thus, the need for algorithms applicable on implicit formulations of the problem above is present.

In som recent papers, Dempster and Hutton (1995a, 1995b) present a linear programming-based algorithm for the pricing of American derivatives. The key idea in their articles is to convert an abstract ordered complementarity problem to an abstract linear program. This conversion is based upon the work of Cryer and Dempster (1980). Furthermore, the abstract linear program is discretisized, and they use linear programming techniques to solve it. Our approach is somewhat different, but the answer is the same: we end up solving a sequence of linear programs. These linear programs have a simpler structure than those of Dempster and Hutton.and the reason for this difference is that the partial differential equation used in Hutton and Dempster is not completely transformed to the heat equation.

In this paper will we first discretisize problem [1], and by using the results of Mangasarian (1978), we will show how the discretisized version of the complimentarity problem could be solved by means of a sequence of linear programs. This derivation is based upon least

elements theory. Further, we present a Forward, Backward Recursive algorithm (FBR) which is applicable to the problem. The algorithm uses only one forward and one backward iteration for each time step. In the case of explicit finite differences, the method uses only one iteration. This is not surpricing since the explicit finite difference method is easily implemented. See eg Hull (1993).

Another alternative approach to the valuation of American derivatives involves Lattice methods, including bionominal methods. The implementation in this case is done by formulating the problem as a dynamic programming problem and solving it using recursive techniques. The bionominal method can also handle exotic options by introducing one or more state variables. How this is done in particular varies between the different derivatives. In a comparison of tree approaches versus finite difference methods, Geske and Shastri (1985) conclude: "researchers computing a smaller number of option values may prefer bionominal approximation, while practioners in the business of computing a larger number of option values will generally find that finite difference methods are more efficient". It is not our purpose to discuss the appropriateness of using finite difference methods, but we recognize that finite difference approaches give option values versus spotprices as output, which is not a facility directly inherit in the bionominal approach. Further, if implicit finite differences or a related approximation is used, the user has the freedom to select the number of meshpoints along the τ dimension independent of the number of mesh points along the x dimension.

In Section 8.2 we discretisize the problem by using finite difference methods. The discretization is general and includes both explicit, implicit, Crank-Nicolson and more generally, the θ finite difference method. In Section 8.3, we study linear complimentarity problems and least elements, which in Section 8.4 are shown to be applicable to the problem under investigation. In Section 8.5 we show the solution procedure for one particular simple case (explicit finite differences) and in Section 8.6 we present a forward backward recursive algorithm for the general case. In Section 8.7 we discuss the efficiency of the algorithm, and in Section 8.8 we draw some conclusions and discuss topics for future research. In this paper we use bold symbols for vectors. For matrices we use uppercase letters, and the optimal solution to for example a linear program, is denoted by an asterix. All other notation is explained as it occurs.

8.2. Discretization of the problem.

In this section we present a discretized version of the complimentary problem above in the (x,τ) plane. The approximation technique is similar to that of Wilmott, Dewynne and Howison (1993). We divide the horizontal axis into $\bar{m}$ mesh points with step size δ_τ and we divide the vertical axis into $\bar{n}+2$ mesh points with step size δ_x. The upper and lower bounds of x and τ are set by defining the least value of x in which the function $u(x,\tau)$ is approximated as $-k_1\delta_x$ and the largest value of x in which $u(x,\tau)$ is approximated as $k_2\delta_x$. A similar procedure is also done along the horizontal axis. The least value of τ is 0. Further, the largest value of τ is equal to $\bar{m}\delta_\tau$. This means that we consider functional values of $u(x,\tau)$ for $x \in [-k_1\delta_x, k_2\delta_x]$ and $\tau \in [\delta_\tau\bar{m}, 0]$. In total, there will be $\bar{m} \times (\bar{n}+2)$ mesh points and by definiton we have $k_1 + k_2 = \bar{n}+2$. The mesh point $(\hat{x},\hat{\tau})$ is given by $(\hat{n}\delta_x, \hat{m}\delta_\tau)$ for an appropriate selection of $\hat{n}$ and $\hat{m}$. The directional derivative of the first term is approximated using:

$$[2] \quad \frac{\partial u}{\partial \tau} \approx \frac{u(\hat{n}\delta_x,(\hat{m}+1)\delta_\tau) - u(\hat{n}\delta_x,\hat{m}\delta_\tau)}{\delta_t}$$

The second term is approximated using:

$$[3] \quad \frac{\partial^2 u}{\partial x^2} \approx \theta\left(\frac{u((\hat{n}+1)\delta_x,(\hat{m}+1)\delta_\tau) - 2u(\hat{n}\delta_x,(\hat{m}+1)\delta_\tau)+u((\hat{n}-1)\delta_x,(\hat{m}+1)\delta_\tau)}{\delta_x^2} \right) +$$

$$(1-\theta)\left(\frac{u((\hat{n}+1)\delta_x,\hat{m}\delta_\tau) - 2u(\hat{n}\delta_x,\hat{m}\delta_\tau)+u((\hat{n}-1)\delta_x,\hat{m}\delta_\tau)}{\delta_x^2} \right)$$

The selection of the parameter θ is a matter of convenience, but values of θ at 0, 0.5 and 1 imply explicit finite differences, Crank–Nicolson finite differences and implicit finite differences respectively.

The function $g(x,\tau)$ is approximated using:

$$[4] \quad g_n^m = g(\hat{n}\delta_x,\hat{m}\delta_t)$$

We now define the variable $v_{\hat{n}}^{\hat{m}}$, as the approximated value of the function $u(x,\tau)$ when $x=\hat{n}\delta_x$ and $\tau = \hat{m}\delta_\tau$. Since $\hat{n}$ could attain negative values, it could, for notational convenience when discussing linear programs later on, be convenient to rescale the parameter $\hat{n}$ to n, ($n \in 1,2,...,\bar{n}$.) such that $v_1^{m+1} = v_{k_2-1}^{m+1}$, $v_2^{m+1} = v_{k_2-2}^{m+1}$ and so on. Finally, $v_{-k_1+1}^{m+1} = v_{\bar{n}}^{m+1}$. Recall that $v_{k_2}^{m+1}$ and $v_{-k_1}^{m+1}$ are given by the boundary conditions, as is v_n^0.

The parameter $\hat{m}$ could only attain positive values, but for notational equivalence we replace $\hat{m}$ by m. Using the expressions [2] and [3], the approximated version of the inequality [1a] multiplied by the positive parameter δ_τ, gives us the inequality [5]:

[5a] $a_n^{m+1} \geq b_n^m$ $n \in 1,2,....,\bar{n}., \ m \in 1,2,...,\bar{m}.$

where:

[5b] $a_n^{m+1} = v_n^{m+1} - \alpha\theta(v_{n+1}^{m+1} - 2v_n^{m+1} + v_{n-1}^{m+1})$ $n \in 1,2,....,\bar{n}-1., \ m \in 1,2,...,\bar{m}.$

[5c] $b_n^m = v_n^m + \alpha(1-\theta)(v_{n+1}^m - 2v_n^m + v_{n-1}^m)$ $n \in 1,2,....,\bar{n}-1., \ m \in 1,2,...,\bar{m}.$

[5d] $\alpha = \dfrac{\delta_t}{\delta_x^2}$

When we use the finite difference approach, the stability of the approximation is affected by the choice of α . In the case where $\theta < 0.5$, α is restricted to lie in the interval:

[6] $0 < \alpha \leq \dfrac{1}{2(1-2\theta)}.$

For $\alpha \geq 0.5$, there are no such restrictions. Note that at the boundaries, the approximated value of $u(x,\tau)$ is fixed, such that for any m > 0, we have $\bar{n}$ inequalities of the type [5a]. At m=0 the value of $u(x,\tau)$ is given by the boundary condition.

The complimentarity constraint [1c] is approximated using:

[7] $(a_n^{m+1} - b_n^m)(v_n^{m+1} - g_n^{m+1}) = 0$ $n \in 1,2,....,\bar{n}-1., \ m \in 1,2,...,\bar{m}.$

The boundary conditions require special attention but before we discuss that, it could be an advantage to formulate the discretisized problem on a compact form. In order to do this we have to define a set of vectors and matrices. At time step m, we define the $\bar{n} \times 1$ vector v^m as:

[8] $v^m = \left\{ \begin{array}{c} v_1^m \\ \vdots \\ v_{\bar{n}}^m \end{array} \right\}.$

The vector [8] contains the approximated values of $u(x,\tau)$ at $\tau = m\delta_\tau$. Further, we define the

$n \times 1$ vector $\mathbf{g}^m$ as the approximated values of $g(x,\tau)$ at $\tau = m\delta_\tau$, thus:

$$[9] \quad \mathbf{g}^m = \begin{Bmatrix} g_1^m \\ \vdots \\ g_{\bar{n}}^m \end{Bmatrix}.$$

Finally, we define $\mathbf{b}^m$ as the $\bar{n} \times 1$ vector:

$$[10] \quad \mathbf{b}^m = \begin{Bmatrix} b_1^m \\ \vdots \\ b_{\bar{n}}^m \end{Bmatrix}.$$

The elements of $\mathbf{b}^m$ are defined as in [5c], except for $b_{\bar{n}}^m$ and b_1^m, which have to be adjusted to cope with any predefined boundary conditions in the problem. This means that we have to insert the functional values of $g(x,\tau)$ at the boundaries and rearrange [5a] so that $b_{\bar{n}}^m$ and b_1^m contain all determined values at the current step in the procedure. Such a procedure results in[11]:

$$[11a] \quad b_{\bar{n}}^m = v_{\bar{n}}^m + \alpha(1-\theta)(g_{\bar{n}+1}^m - 2v_{\bar{n}}^m + v_{\bar{n}-1}^m) + \alpha\theta g_{\bar{n}+1}^{m+1}$$

$$[11b] \quad b_1^m = v_1^m + \alpha(1-\theta)(g_0^m - 2v_1^m + v_2^m) + \alpha\theta g_0^{m+1}$$

Note that this modification removes the terms v_0^{m+1} and $v_{\bar{n}+1}^{m+1}$ from the equations for a_0^{m+1} and $a_{\bar{n}+1}^{m+1}$. We can now formulate the approximated version as a complimentarity problem as follows[12]:

$$[12a] \quad (C\mathbf{v}^{m+1} - \mathbf{b}^m) \geq 0$$

$$[12b] \quad (\mathbf{v}^{m+1} - \mathbf{g}^{m+1}) \geq 0$$

$$[12c] \quad (C\mathbf{v}^{m+1} - \mathbf{b}^m)(\mathbf{v}^{m+1} - \mathbf{g}^{m+1}) = 0.$$

In [12], C is the tridiagonal matrix:

$$[13] \quad C = \begin{bmatrix} 1+2\alpha\theta & -\alpha\theta & 0 & \cdots & & 0 \\ -\alpha\theta & 1+2\alpha\theta & -\alpha\theta & & & \vdots \\ 0 & -\alpha\theta & \ddots & & \ddots & 0 \\ \vdots & & & \ddots & 1+2\alpha\theta & -\alpha\theta \\ 0 & \cdots & & 0 & -\alpha\theta & 1+2\alpha\theta \end{bmatrix}$$

We note that C is an $\bar{n} \times \bar{n}$ matrix and b^m, g^{m+1} and v^{m+1} are vectors of dimension $\bar{n} \times 1$.

The problem is solved recursively starting at $\tau = \delta_\tau$, moving backwards to $\tau = \bar{m}\delta_\tau$, in order to obtain a solution to [1].

We have now arrived at the discrete linear complimentarity formulation of the free boundary problem. Since the problem above is a linear complimentarity problem, we will have a closer look at such problems and least elements.

8.3. Linear complementarity problems and least elements.

In this section we discuss problems of the type shown in[14]:

[14a] $Mz + q \geq 0$
[14b] $z^T(Mz + q) = 0$
[14c] $z \geq 0$

Above, M is an $\bar{n} \times \bar{n}$ matrix, z is the $\bar{n} \times 1$ solution vector and q is an $\bar{n} \times 1$ vector of constants. To keep the notation the same throughout the paper, we use n as the index for the variables in the problem above. We assume that the solution to the problem in [14] is unique.

We now turn our attention to possible solution procedures for the system in [14]. In order to do so, we have to understand some concepts related to least element theory. We will first define the term meet semi sublattice.

Definition 1: For a given set $S \in \mathfrak{R}^{\bar{n}}$, we define the vectors x, $y \in S$ with elements x_n and y_n and $d = \min(x,y)$ with elements d_n. If $\forall$ $x,y \in S \rightarrow d \in S$ then S is a meet semi sublattice.

Definition 2: A symetric matrix M is said to belong to the class Z if all off diagonal elements of M are nonpositive. (There are no restrictions on the diagonal elements.)

With these two definitions in mind we have the following proposition:

Proposition 1:

If $S = \{z : Mz + q \geq 0, z \geq 0\}$ and $M \in Z$, then S is a meet semi sublattice.

Proof:

Let $x, y \in S$, where S is the set defined in the theorem above. Since $x \geq 0$ and $y \geq 0$, $d = \min(x, y) \geq 0$. Let h_n be the n'th element of the vector $Md + q$. We can then, without loss of generality, set $x_n = d_n$ for an arbitrary n. Then:

$$[15] \quad h_n = m_{nn}x_n + \sum_{i \neq j} m_{nj}d_j \geq (q + Mx)_n \geq 0$$

This inequality holds since all offdiagonal elements of M are nonpositive. Conversely, if we let $y_n = d_n$ for an arbitrary n, then:

$$[16] \quad h_i = m_{nn}y_n + \sum_{i \neq j} m_{nj}d_j \geq (q + My)_n \geq 0$$

This means for any $x, y \in S \rightarrow d = \min(x, y) \in S$. This completes the proof.

□

We can now define the concept of a least element.

Definition 3: If S is a meet semi sublattice and there exists a vector $u \in S$ such that for any vector $x \in S$ $x \geq u$, then u is termed the least element of S.

The relation $x \geq u$ here means that each element of the vector x is greater than or equal to the corresponding element in u.

We are now able to state the following theorem provided by Mangasarian (1978):

Theorem 1:

If $S = \{z : Mz + q \geq 0, z \geq 0\}$ and $M \in Z$, then the solution to problem [14] could be found by solving the linear programming problem [17]:

$$[17a] \quad \min p^T z$$

subject to:

[17b] $Mz + q \geq 0$
[17c] $z \geq 0$

Where **p** is any vector satisfying **p** > 0.

Proof:

We note that the solution to [17] is the least element of S given by **u**, since any other feasible solution, such as $x \geq u$ implies $p^T x \geq p^T u$. The vector **u** is feasible since $Mx + q$ is a meet semi sublattice. If $(Mu + q)_n > 0$, and n is arbitrary, u_n has to be 0. If not, we could decrease u_n (and improve the objective function value) until $(Mu + q)_n = 0$ or $u_n = 0$. This operation would not affect the feasibility of the other variables. The latter follows since M is a Z-matrix. Since the solution of [17] clearly satisfies restriction [14a] and since $(Mu + q)_n u_n = 0$, **u** must satisfy [14].

$\square$

The implication of Mangasarian's theorem, is that certain linear complimentarity problems could be solved by means of linear programming. We utilize this approach in what follows.

8.4. The linear complimentary formulation of the free boundary problem and linear programming.

In Section 8.3. we performed a discretization of the free boundary problem and arrived at a linear complimentarity formulation. The linear complimentary formulation was, as shown in [18]:

[18a] $(Cv^{m+1} - b^m) \geq 0$
[18b] $(v^{m+1} - g^{m+1}) \geq 0$
[18c] $(Cv^{m+1} - b^m)(v^{m+1} - g^{m+1}) = 0.$

We are now able to state the following theorem:

Theorem 2:

The linear complimentary problem [18] may be solved by the linear programming problem in [19]:

[19a] $\min \mathbf{p}^T z$

subject to:

[19b] $Cz^{m+1} - q^{m+1} \geq 0$

[19c] $z^{m+1} \geq 0$

where:

$\mathbf{p} \geq 0$

[19d] $z^{m+1} = v^{m+1} - g^{m+1}$

[19e] $q^{m+1} = b^m - Cg^{m+1}$.

Proof:

Problem [18] could, by using [19d] and [19e], be converted to the linear complimentarity problem below [20]:

[20a] $Dz^{m+1} - q^{m+1} \geq 0$

[20b] $(Dz^{m+1} - q^{m+1})z^{m+1} \geq 0$

[20c] $z^{m+1} \geq 0$

Since $D \in Z$, it follows from Mangasarian's theorem, shown in Section 8.4 that [20] is solvable by linear programming. After we have solved [20] by linear programming, the solution to system [18] could be obtained by the relations in [20]. The interpretation of the optimal solution z^{m+1^*} is that the elements of z^{m+1^*} are equal to zero if early excersise is optimal.

☐

8.5. Solution procedure for $\theta = 0$.

It is now necessary to discuss whether the reformulations of the problem outlined in the previous sections could be solved by any effective custom made algorithm, or whether they must be solved by traditional algorithms.

Utilizing relations [19d] and [19e], the linear programming problem in [19] could be reformulated as [21]:

[21a] $\min \mathbf{p}^T \mathbf{v}^{m+1}$

subject to:

[21b] $\mathbf{C}\mathbf{v}^{m+1} \geq \mathbf{b}^m$

[21c] $\mathbf{v}^{m+1} \geq \mathbf{g}^{m+1}$

Let us now evaluate this program for $\theta = 0$. As this means that the method used is the explicit finite difference , it is no surprise that the solution has a simple structure. When $\theta = 0$, we have the following theorem:

Theorem 3:

If $\theta = 0$, the solution to the linear program in [21] is given by:

[22] $\mathbf{v}^{m+1*} = \max[\mathbf{b}^m, \mathbf{g}^{m+1}, \ \mathbf{0}]$

where $\mathbf{0}$ is a vector consisting of zeros everywhere.

Proof:

If we set $\theta = 0$, then problem [21] becomes[23]:

[23a] $\min \mathbf{p}^T \mathbf{v}^{m+1}$

subject to:

[23b] $v^{m+1} \geq b^m$

[23c] $v^{m+1} \geq g^{m+1}$

[23d] $v^{m+1} \geq 0$

This problem has the trivial solution given by [22].

☐

Typically, in many applications $g^{m+1} \geq 0$. If this is so, then restriction [23d] is redundant. In the case where $\theta \neq 0$, the solution to the problem may not be obtained so easily; thus we have to study the problem in [19] in more detail.

8.6. Solution procedures $\theta \neq 0$.

In this section, will discuss solution procedures for the problem when $0 < \theta \leq 1$. We recall that when $\theta = 0.5$ and $\theta = 1$, the approximations done are the Crank-Nicholson and the implicit finite differences, respectively. For an arbitrary selection of θ in the range $0 < \theta < 1$, the approximation method is generally termed the θ method. For convenience, we list the problem under investigation where we have put q^{m+1} on the right-hand side of the constraint equations in problem[24]:

[24a] $\min p^T z$

subject to:

[24b] $Cz^{m+1} \geq q^{m+1}$

[24c] $z^{m+1} \geq 0$

As mentioned earlier, v^{m+1} could be obtained by the relation in [19d]. The vector q^{m+1} is given by [19e]. The elements of the vector q^{m+1} are denoted by:

[25] q_n^{m+1} $n \in$ $1,2,...,\bar{n}$.

In order to obtain a usable algorithm which may be applied to the problem, we have to have a closer look at some results. In the following we will denote the optimal solution of problem [24], for any particular value of m, by z^{m+1^*}. This vector has elements $z_n^{m+1^*}$.

Proposition 2[26]:

[26a]   If $q_n^{m+1} > 0$ then $z_n^{m+1}{}^* > 0$.

Proof[27]:

Since constraint j is given by :

[27a] $-\alpha\theta z_{n-1}^{m+1} + (1 + 2\alpha\theta)z_n^{m+1} - \alpha\theta z_{n+1}^{m+1} \geq q_n^{m+1}$, $q_n^{m+1} > 0$ and $z^{m+1}{}^* \geq 0$,

the solution $z_j^{m+1}{}^*$, has to be positive for the to be feasible. By inspection we can also see this if n=1 or $\bar{n}$.

☐

Proposition 3[28]:

If:

[28a] $z_n^{m+1}{}^* > 0$ and $q_{n+1}^{m+1} > - z_n^{m+1}{}^*\alpha\theta$,

then the solution:

[28b] $z_{n+1}^{m+1}{}^* = 0$

is infeasible. This also holds if:

[28c] $z_n^{m+1}{}^* > 0$ and $q_{n-1}^{m+1} > -z_n^{m+1}\alpha\theta$.

Proof:

If:

[28d] $z_{n+1}^{m+1} = 0$

then constraint j+1 becomes:

[28e] $-\alpha\theta z_n^{m+1}{}^* - \alpha\theta z_{n+2}^{m+1}{}^* < q_{j+1}^{m+1}$

Thus, constraint j+1 is violated. By inspection, we also see that if:

[28f] $\quad z_n^{m+1^*} > 0$ and $q_{n-1}^{m+1} > -z_n^{m+1} \alpha \theta,$

then constraint j-1 is violated.

□

Theorem 4 [29]:

If there exists any feasible solution with $z_n^{m+1} = 0$ then $z_n^{m+1^*} = 0.$

Proof:

Suppose $z_n^{m+1} = 0$ is a feasible solution to the problem. Let $z_n^{m+1} > 0$ and let p_n be the n´th element of **p**. In addition, let the other elements of **p** be 1. By Mangasarian´s result, the optimal solution for the problem should not change for any value of $p_n \in (0,\infty)$. If we set p_n at a sufficiently large number, we could improve the solution by letting $z_n^* = 0.$

□

In the light of the theorem above we must try for a solution with $z_n^{m+1^*} = 0$, as long as this does not mean that the solution is infeasible. Thus, if we could construct an efficient procedure that simultaneously checks for feasibility and solves the equations in the basis of the optimal solution, we would end up with the optimal solution.

An efficient way of doing this is by using a modified Upper-Lower decomposition algorithm for the linear system. The procedure should simultaneously be able to detect which of the variables that in the optimal basis. In what follows we present a modified algorithm based upon LU tridiagonal decomposition.

Before presenting the algorithm, it would be helpful to have a brief introduction to the LU method of solving matrix equations. Assume for simplicity that we know the optimal basis for the problem. We let the index set I denote the optimal basis. From standar linear programming theory, we know that the optimal solution is the solution to the linear system:

[30] $\quad C_I z_I = q_I$

If we let $C_I = LU$, where L is a lower tridiagonal matrix and U is an upper tridiagonal

matrix, with L and U of the same dimension as C_I, we could make the substitution:

[31] $Ly = q_I$

[32] $y = Uz_I$

The matrices L and U are given by:

[33] $$L = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \lambda_1 & 1 & 0 & & \vdots \\ 0 & \lambda_2 & \ddots & \ddots & 0 \\ \vdots & & \ddots & 1 & 0 \\ 0 & \cdots & 0 & \lambda_{n^*} & 1 \end{bmatrix}$$

[34] $$U = \begin{bmatrix} \mu_1 & \omega_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \omega_2 & & \vdots \\ 0 & 0 & \ddots & \ddots & 0 \\ \vdots & & \ddots & \mu_{n^*-1} & \omega_{n^*} \\ 0 & \cdots & 0 & 0 & \mu_{n^*} \end{bmatrix}$$

The variables ω_n, μ_n and λ_n ($n \in 1,2,...n^*$) are intermediate variables in the algorithm. The constant n^* is the maximal value of n, given that z_n^{m+1} is a member of the optimal basis. The interpretation of n^* is described in more detail in the description of the algorithm below.

Thus, we can solve system [31] by forward substitution and system [32] by backward substitution. The solution of the linear equations is done in a simple and efficient manner.

The key idea of the algorithm is to solve [31] forward by selection, using as the initial basis all equations where $q_n^{m+1} > 0$. As we know from the proposition earlier in this section $q_n^{m+1} > 0 \rightarrow z_n^{m+1^*} > 0$. We can further check for feasibility on the equations where $q_n^{m+1} < 0$. This feasibility check requires knowledge of $z_{n-1}^{m+1^*}$ given that all remainding variables (n and above) are nonbasic in the optimal solution. Conveniently, the LU algorithm gives this value immediately.

The equations in the LP [24] are satisfied with equality for $n = 1,2,...,n^*$. In economic terms this simply means that the option has a positive value when there is positive time left to expiry, even if the payoff from the derivative, if exercised immediately, is zero.

The algorithm works on problems where we know the structure of the optimal solution in advance. The solution should have the following property[35]:

For $n \in 1,2,\ldots\bar{n}$. exist an n^* in which:

[35a] $\quad z_n^{m+1} > 0 \quad n \in 1,2,\ldots n^*-1$.

[35b] $\quad z_n^{m+1} = 0 \quad n \in n^*, n^*+1,\ldots,\bar{n}$.

If the problem has the following structure:

[35c] $\quad z_n^{m+1} = 0 \quad n \in 1,2,\ldots n^*-1$.

[35d] $\quad z_n^{m+1} > 0 \quad n \in n^*, n^*+1,\ldots,\bar{n}$.

the problem could easily be converted to a problem with the solution structure given by [35a] and [35b] by renumeration of the variables as $\bar{n} = 1, \bar{n}-1=2,\ldots,1=\bar{n}$.

For the derivatives discussed earlier one should construct the problem such that v_1^{m+1} is such that $g_1^{m+1} = 0$. Thus, by increasing n one moves towards the region in which early exercise is optimal.

Solution structures of the type above are common in financial applications. For example, when applied to American put option pricing, $z_{n^*}^{m+1} + g_n^{m+1}$ is the (approximated and transformed) value of the option price at which exercize is optimal for values including and below, and not optimal for values above.

The procedure to be outlined in detail below, will only use one iteration and one backward substitution, which, compared with the projective succesive overrelaxation algorithm (PSOR), for example, is a improvement in efficiency. The algorithm below assumes that q_1^{m+1} is positive. The procedure can be modified to capture more complex structures for q^{m+1}.

The procedure for each time step is as follows[36]:

Step 1.

Initialisation.

[36a] $\quad \mu_1 = 1+2\alpha\theta$

[36b] $\quad \rho_1 = q_1$

[36c] $n = 1$.

The parameter ρ_n , $(n \in 1,2,...,n^*)$ is an intermediate computing variable in the algorithm.

Step 2.1

Increase n by 1:

If $g_n^{m+1} < 0$, goto 2.2.

Check for feasibility:

[36d]   If $-\alpha\theta \dfrac{\rho_{n-1}}{\mu_{n-1}} > q_n$, go to Step 3.

Note that :

[36e] $\dfrac{\rho_{n-1}}{\mu_{n-1}}$

is the optimal value of z_{n-1}^{m+1}, given that z_{n-1}^{m+1} is the last variable that enters the basis. Thus, test [36d] checks whether z_n^{m+1} could be set at zero. If so, it should be set at zero, and the search stops.

Step 2.2

Compute:

[36f] $\mu_n = 1 + 2\alpha\theta - \dfrac{(\alpha\theta)^2}{\mu_{n-1}}$

[36g] $\rho_n = q_n + \dfrac{\alpha\theta}{\mu_{n-1}}\rho_{n-1}$

Repeat step 2.

Step 3

[36h] $n^* = n$

Solution given by:

[36i] $z_n^{m+1^*} = 0$ $n \in n^*, \ldots, \bar{n}.$

[36j] $z_{n^*-1}^{m+1^*} = \dfrac{\rho_{n-1}}{\mu_{n-1}}$

[36k] $z_n^{m+1^*} = \dfrac{(\rho_n + \alpha\theta z_{n+1}^*)}{\mu_n}$ $n \in n^*-2, \ldots, 1.$ (Backward substitution.)

$\square$

The algorithm needs to be explained in more detail. In Step 1, we initialize the parameters μ_1 and ρ_1. In Step 2, which is the main part of the algorithm, we check whether z_n^{m+1} should be in the basis. This is done by criterion [36d]. As long as the criterion is not fullfilled, we are not allowed to set z_{n+1}^{m+1} at zero. We then have to do the feasibility check once more on variable z_{n+2}^{m+1} and so on. Recall that this check is not neccesary for values of x in which the approximated value of $g(x,\tau) = 0$. When [36d] is fullfilled, we know that z_{n+1}^{m+1} could be set to zero, which is optimal according to the theorem above, and the algorithm goes to Step 3. In Step 3 we just assign the optimal values to z^{m+1} by backward substitution for the basic variables and set the nonbasic variables to zero. Finally, we compute the values of $v^{m+1} = z^{m+1} + g^{m+1}$ and start again on the next time step. Note also that we can reformulate the procedure starting by the variables z_n^m, which apriori will be equal to zero. In that case we start with the equation:

[37] $(Cz^{m+1} - q^{m+1})_{\bar{n}} \geq 0$

insert the value $z_n^{m+1} = 0$, and move backwards until the corresponding equation is not fullfilled, and assigne $n+1 = n^*$. The remaining equations are solved by the LU algorithm. This approach utilizes Theorem 4 and is similar to the method of Brennan and Schwartz (1977) in the American put option problem case.

8.7.Discussion of algorithmic efficiency.

We have now presented the algorithm and it is necessary to discuss what kind of efficiency improvement one should expect when using the algorithm. T discussion implies an evaluation of the algorithm´s performance as compared to its alternatives. How such a comparison in practice should be done is an open question. Implementation of algorithms typically means making a computer program of two different algorithms and comparing the time used to solve the problem under investigation by each. However, this assumes that both algorithms are programmed in an efficient manner, which may not always be the case. Another measure is the number of iterations used before the optimal solution is reached. In the case with the FBR algorithm, this is known in advance and is one forward and one backward iteration for each time step, independent of the type of derivative, time step, and parameters of the derivative and underlying asset. Hutton and Dempster (1995) perform comparative studies of the PSOR and the LP formulations of the problem for two types of American derivatives; namely, the ordinary American Put and an American lookback put. They used the Crank-Nicholson aproximation method. For the American put, the reported number of iterations was in the range from 13-17 for a sparse grid ($\bar{n} = 75$) and from 554-831 iterations for a very fine grid ($\bar{n} = 9600$). However, when using the PSOR algorithm, the user must specify a relaxation parameter in the range of (1,2) and the selection of the value for this parameter is also a subject on it´s own. For the LP variant, which was implemented using "hot-starts" from the previous iterations, the number of iterations reported was smaller and ranged from 0-3 iterations for the sparce grid and from 36-468 iterations for the fine grid. In their study, Dempster and Hutton (1995) used 1000 time steps, which means that the computed value of x at the free boundary could be very close to the boundary computed for the previous time step. In that case, the number of iterations in a "hot-started" simplex procedure would be very low, especially if the number of mesh points along the state variable is small. In the case of the American lookback put option they did not report the number of iterations, but the computer time used. The time used was higher than for the ordinary put options, such that, preassumably, the number of iterations was higher. The algorithm presented in this paper also allows for incorporation of subjective information which may be used in "hot-start" variants. Suppose that we have apriori information about the free boundary, i.e. that the free boundary is a decreasing function of τ. The algorithm can utilize this by excluding the feasibility test for values known to be far from the boundary.

The algorithm presented in this paper gives accurate answers to the complimentarity problem that arises in the pricing of American options when utilizing explicit or implicit (or any combination) finite difference methods. The accuracy of the discretization itself is a

separate subject, but since the algorithm presented is relatively efficient, the accuracy of the finite difference approximation can be adjusted by simply selecting enough mesh points in the grid. The "amount of work", as measured by number of iterations, is linear in the number of time steps and constant (2) along the state variable dimension.

8.8. Conclusions.

In this paper, we have studied a solution procedure that is applicable to the pricing of American options. We started by discretiing a complimentarity problem which was shown to have some crucial least element properties. Next, we showed how the problem could be solved by means of a sequence of linear programs. We then presented the trivial solution for the explicit finite difference case, and further we presented a forward Backward recursive algorithm which may be used in approximations such as the implicit finite differences, Chrank-Nicholson and, in more general, the θ method. Finally, we discussed the advantage of the method as compared to the LP formulation and the PSOR algorithm.

References:

Black, F., Scholes, M. (1973) "The Pricing of Options and Corporate Liabilities." Journal of Political Economy 82. pp. 637-654.

Brennan, M.J., Schwartz, E.S. (1977) "The Valuation of American Put Options." Journal of Finance, Vol. 32 pp. 449-462.

Cottle R.W., Pang, J.S. (1978) "On Solving Linear Complementarity Problems as Linear Programs." Mathematical Programming Study 7. pp. 88-107.

Cryer, C.W., Dempster, M.A.H. (1980) "Equivalence of Linear Complimentarity Problems." Maths of OR 14 pp. 534-558.

Dempster, M.A.H. , Hutton, J.P. (1995a) "Fast Numerical Valuation of American, Exotic and Complex Options. (Submitted to Applied Mathematical Finance.)

Dempster, M.A.H. , Hutton, J.P. (1995b) "Fast Numerical Valuation of American Options by Linear Programming." To be submitted to Mathematical Finance.

Geske, R., Shastri,K. (1985) "Valuation by Approximation: A Comparison of Alternative Option Valuation Techniques." Journal of Financial and Quantitative Analysis, 20 pp. 45-71.

Hull, J.C. (1993) Option, Futures and Other Derivative Securities. Prentice Hall International Editions.

Mangasarian, O.L. (1976) "Linear Complementarity Problems Solvable by Linear

Programming." Mathematical Programming 10 pp. 263-70.

Wilmott, P. Dewynne, J. , Howison, S. (1993) Option pricing - Mathematical Models and Computation.

Time series Oslo Stock Exchange

	AKE	DNL	DYN	ELK	HNA	KVI	NHY	ORK	SAG	SMD	TOM	UTO	BEA	BON	NAL	STO	VAR	WWI	BANK	FORSIK	NSI
					Data 1987-92																
870130	-4.27	0.71	7.21	5.07	9.86	7.08	-2.21	0.76	-0.92	0.02	-2.17	5.10	-5.49	15.54	26.83	13.06	0.26	-0.82	1.02	3.46	6.57
870227	-0.89	11.94	17.54	-2.14	8.91	-0.83	10.36	5.47	2.80	2.63	-9.71	-7.27	1.92	5.21	-7.69	3.41	4.30	-5.00	-1.38	0.32	-1.09
870331	3.60	2.70	5.22	4.92	9.35	2.78	10.36	10.51	1.82	-1.28	8.60	9.80	5.66	-0.90	-2.08	10.38	3.37	-3.95	-3.79	5.54	3.85
870430	10.43	11.70	0.71	5.73	7.18	7.03	10.14	4.47	41.96	16.13	22.77	0.00	5.36	21.82	2.13	7.25	-3.99	27.40	5.93	6.38	11.11
870529	3.73	-9.69	8.92	4.43	5.49	6.89	9.81	-2.35	15.09	6.82	9.68	-8.93	22.71	-2.61	-4.17	3.61	1.85	11.83	-4.23	0.28	9.38
870630	0.72	7.25	7.48	-3.77	6.27	5.06	4.96	2.99	12.02	9.87	-11.76	-3.92	3.59	23.54	26.09	-3.88	25.23	24.04	-3.87	-0.70	0.00
870731	5.71	3.24	13.92	22.06	27.18	3.90	13.95	11.87	22.93	4.90	12.50	8.16	10.67	0.63	48.28	4.03	12.43	8.53	10.29	3.99	-3.57
870831	8.78	39.79	8.33	12.45	24.19	13.02	6.02	19.64	-0.40	30.84	5.19	13.21	-5.42	2.48	-5.81	-0.78	-1.05	-10.00	7.23	10.83	6.48
870930	14.29	1.12	13.85	6.43	-1.79	16.99	-1.37	6.60	10.36	1.43	35.21	31.67	8.92	5.15	13.58	31.25	27.66	14.29	-9.90	6.37	28.70
871030	-29.89	-28.70	-22.52	-44.63	-35.54	-22.37	-27.58	-29.20	-13.36	-36.62	-1.04	-21.52	-21.64	-16.43	-8.70	-26.79	-46.25	-32.64	-20.49	-18.85	-29.32
871130	-17.83	-17.92	-13.95	-39.39	-6.41	-13.55	-25.75	-12.50	-35.83	-13.33	-26.32	-22.58	-2.99	-17.24	-22.62	-18.70	-27.91	-25.77	-19.37	-11.66	-9.18
871231	-9.43	-17.30	-3.16	8.64	17.76	-0.58	4.18	-5.06	0.43	-3.42	-1.91	0.00	4.09	-17.78	3.58	8.32	-4.30	-12.97	-7.55	-7.43	18.24
880129	4.17	-10.08	-8.60	39.91	7.02	5.09	10.51	-6.70	18.97	26.11	26.71	18.75	19.72	31.76	-13.85	20.01	34.27	21.29	-10.57	-12.92	-24.33
880229	5.00	12.77	-2.29	13.82	5.22	5.71	12.18	8.06	-2.72	13.68	-6.90	22.81	6.79	15.38	-6.90	-1.92	11.30	7.89	4.11	10.46	5.88
880331	-0.24	-13.31	13.09	34.54	7.59	10.07	10.93	22.53	6.28	1.85	0.46	4.29	10.11	-1.84	-1.39	12.73	-18.42	-4.57	0.49	13.89	11.11
880429	-4.53	-6.42	2.25	-5.05	6.58	-4.85	8.67	1.59	3.55	-2.73	10.60	6.85	4.99	-8.32	8.92	11.32	3.23	4.79	-3.15	-7.48	-9.00
880531	-9.62	3.26	-2.35	6.33	14.08	-2.76	-4.88	-4.82	-8.12	-7.08	-5.56	-2.56	14.34	-6.75	3.45	7.68	-12.50	-2.44	-7.37	-18.62	-1.55
880630	0.00	16.10	9.42	12.34	5.71	3.59	15.51	16.00	-11.60	0.00	9.41	17.11	4.95	7.20	3.33	5.96	2.04	25.00	1.00	-4.01	5.50
880729	-2.33	1.21	9.93	13.64	-3.42	6.06	2.55	3.45	7.50	4.08	16.13	10.11	9.09	0.75	17.74	3.31	16.50	14.00	-1.26	-0.23	2.17
880831	-9.52	4.00	-9.64	-6.10	-8.77	-11.02	-7.22	-13.33	-19.77	-7.84	-6.02	12.24	-8.49	-19.26	10.96	-3.85	18.03	-9.65	-7.75	-12.40	-4.26
880930	-6.58	2.31	-3.33	0.73	6.75	11.24	3.65	9.34	6.52	-12.77	-2.46	6.36	7.93	13.76	2.47	2.00	7.64	-0.97	-3.60	-2.51	-4.44
881031	15.49	-5.26	3.45	20.00	22.32	3.09	-4.69	-0.75	0.68	4.88	9.09	2.56	10.94	25.81	10.84	22.55	5.07	44.12	-6.35	-7.72	10.47
881130	15.85	3.17	8.00	-1.82	6.77	-4.00	4.43	11.39	25.00	0.00	-1.85	18.33	13.80	7.69	7.61	9.87	13.83	2.04	-3.52	1.68	0.00
881230	10.53	19.23	11.11	4.32	12.68	4.37	10.38	19.32	8.11	9.30	-1.89	9.86	3.96	26.19	6.06	18.93	5.08	8.67	25.53	29.76	26.32
890131	-0.95	32.26	16.67	15.09	7.81	20.56	22.22	21.14	21.50	38.30	8.65	16.67	1.67	34.91	16.19	18.98	19.62	23.93	24.68	16.71	37.50
890228	23.08	6.34	-3.33	4.11	-11.88	0.66	4.20	-0.31	-2.06	12.31	-5.75	1.10	3.51	1.40	-8.20	8.06	2.70	5.94	6.86	0.42	21.21
890331	14.06	12.39	10.34	30.37	8.55	6.58	12.75	6.15	41.60	9.59	2.82	13.59	-2.04	8.28	1.79	6.67	13.79	14.49	14.04	10.89	-27.50
890428	10.96	-2.04	1.82	42.05	-0.80	8.64	5.51	6.98	-5.64	3.75	4.57	29.60	14.92	1.91	-2.63	0.71	6.54	-10.20	11.32	17.07	12.41
890531	3.70	18.33	-6.82	-4.29	-13.50	1.91	-6.63	0.87	-12.82	3.52	4.80	-2.24	25.25	25.00	27.93	17.44	10.95	16.36	3.63	2.88	4.80
890630	0.60	-3.65	-6.34	-14.20	1.77	1.02	3.70	-7.56	0.00	-4.93	10.42	6.87	2.26	0.50	-8.45	-0.91	6.25	-1.12	-8.91	-5.62	-13.69
890731	11.83	15.93	1.56	1.32	9.76	13.42	-10.71	10.61	17.65	6.67	4.91	17.14	4.10	16.92	3.85	13.15	-0.74	6.22	2.92	8.44	10.34
890831	19.05	10.06	0.51	-5.88	-1.90	-6.51	5.00	12.33	18.75	16.67	5.04	2.13	0.00	-7.23	-0.74	11.08	22.22	37.24	-3.58	12.96	11.88
890929	11.56	-14.95	-2.04	-7.29	-6.80	3.48	-5.71	2.44	28.42	1.19	11.30	9.25	-0.91	15.60	20.15	13.38	-11.52	-6.10	10.75	1.75	7.26
891031	-23.51	-19.80	-11.46	-34.08	-3.82	-15.60	-5.39	-9.52	-15.57	-11.76	-13.85	-8.20	-0.61	-9.72	-13.66	-13.63	-26.03	-7.79	-22.89	-17.72	0.00
891130	-3.12	-2.13	4.71	5.68	-1.08	19.57	8.54	10.53	-0.97	6.67	12.86	11.31	8.00	18.68	2.16	-5.59	27.78	5.99	0.62	9.46	-5.18
891229	18.28	18.48	25.84	1.61	-1.46	12.12	9.51	12.14	26.47	3.75	-0.32	8.56	9.69	18.52	4.93	2.63	-8.70	6.31	5.54	8.48	6.75
900131	3.64	-5.32	19.64	13.76	22.22	20.54	6.29	7.43	18.60	2.41	8.57	7.39	0.00	3.12	5.37	-2.56	6.35	6.25	3.85	2.46	4.22
900228	3.95	-3.88	-6.72	6.05	-1.21	18.83	10.70	1.78	3.27	21.18	22.81	14.91	7.01	30.15	35.67	-7.89	22.76	8.82	7.73	15.65	-8.09
900330	-3.80	2.82	-4.00	8.55	0.00	7.55	2.29	8.52	-0.63	0.00	-5.95	-2.59	6.19	9.66	9.39	14.29	-8.81	-1.35	-1.35	-7.21	-1.89
900430	-11.40	-9.41	1.21	-3.84	-1.84	-5.26	-3.71	-16.34	-7.64	-2.91	-6.33	14.60	-0.44	-6.58	-6.44	-14.78	-11.33	-5.21	-5.62	-0.83	-4.49
900531	13.25	3.90	27.59	20.51	8.01	15.04	8.18	19.29	11.12	6.71	2.70	12.60	4.15	8.55	4.54	4.12	2.63	5.20	1.31	14.69	8.78
900629	-12.11	-4.17	6.08	0.91	-0.29	-7.42	-5.85	-4.00	-4.40	4.25	2.63	-5.59	-11.50	-11.95	-4.93	-7.91	4.03	-14.54	1.17	-0.97	-2.52
900731	9.18	-5.22	7.64	13.69	0.00	5.40	10.88	8.33	23.03	11.31	2.56	-0.74	10.50	10.71	7.55	18.10	0.70	7.10	0.33	-2.78	13.55
900831	-12.62	-27.06	-9.17	-20.13	-4.09	-13.72	5.37	-11.54	13.90	-4.07	-5.00	-24.63	-27.15	-29.03	-10.09	-22.08	-19.58	-23.80	-14.17	-13.11	-17.61
900928	-19.79	-24.53	-14.01	-16.67	-14.63	-19.54	-0.22	-16.09	-2.35	-13.56	-6.37	-23.76	-13.04	-7.27	-20.98	-25.00	-18.26	-12.25	-18.11	-10.54	-10.34
901031	0.00	-11.67	-17.42	-29.52	1.07	-6.67	-9.78	1.04	-14.42	-1.96	-1.63	11.04	-12.14	-17.65	14.81	11.11	0.00	-8.11	1.52	8.35	-3.85
901130	10.00	-14.15	-0.92	-8.78	0.71	12.76	-3.69	5.13	-9.55	-4.00	2.86	-3.51	2.44	-13.69	-3.23	12.00	-24.47	-21.57	-0.82	5.00	1.60
901231	-6.26	-9.34	4.63	-17.36	4.33	-7.22	-8.02	1.64	-8.49	5.87	9.01	1.61	-9.67	-6.11	-11.86	-18.02	-16.44	-5.84	-17.93	-8.97	-1.05
910131	-0.43	-11.52	1.77	-4.09	-0.45	2.41	-9.91	-16.01	-17.19	-11.45	9.57	1.40	7.62	8.71	3.58	4.55	-9.82	1.56	-17.49	-10.82	0.27
910228	18.18	50.68	6.96	40.19	13.51	-5.71	12.65	5.71	22.13	22.22	13.95	9.41	11.02	21.62	5.88	13.33	57.01	35.29	6.03	10.99	14.68
910329	-1.38	-6.40	0.00	10.00	5.06	-3.79	4.38	7.29	7.38	0.00	11.22	7.52	0.55	3.89	0.00	22.06	0.59	14.49	5.62	1.90	3.11

910430	-1.95	11.70	-1.71	1.82	3.12	-10.76	-2.54	2.52	3.75	-4.55	0.92	-1.00	-2.08	-8.55	3.45	-2.41	-16.57	-11.39	-13.72	-6.20	12.75
910531	1.62	-8.70	9.40	-1.19	11.44	21.59	12.95	13.81	10.81	10.37	10.91	11.11	18.80	12.89	-8.89	2.47	-3.55	15.93	-15.55	0.10	-5.39
910628	-18.60	10.48	-4.69	-2.41	-3.48	5.17	-8.54	-14.47	0.55	6.96	0.00	3.18	-0.63	-5.73	5.42	0.90	7.35	-9.09	-5.92	-5.82	-8.65
910731	1.43	7.76	1.64	-13.58	9.28	8.67	-1.07	1.54	8.74	-3.25	-1.64	-10.57	8.28	7.73	9.65	-2.09	2.74	10.45	-0.27	0.16	5.26
910830	2.82	8.00	-1.61	-6.43	8.02	0.43	5.39	-0.51	0.00	5.88	-1.67	4.43	2.65	9.49	2.07	-5.49	16.67	1.65	-10.28	-0.47	-10.00
910930	-12.33	-8.89	3.28	-19.85	-2.84	-4.51	-5.63	-11.17	-11.56	-4.76	6.78	-3.77	-2.29	-9.60	-0.96	-12.90	-4.00	-13.36	-25.74	-5.45	-10.00
911031	-12.50	-20.33	-3.97	-34.29	6.52	0.22	-10.57	-7.43	7.39	-4.17	-6.35	-5.39	-1.17	-1.04	0.00	3.70	3.57	-3.74	-17.39	-15.38	-16.05
911129	-16.96	-18.37	-23.55	-42.03	-5.49	-18.39	-21.21	-15.43	-20.11	-17.39	-11.86	-18.13	-16.91	-15.18	-11.29	-17.86	-27.01	-22.82	-30.99	-27.32	-23.53
911231	22.94	5.41	21.44	33.33	5.80	19.05	5.89	23.60	1.99	11.49	0.64	8.86	-10.97	6.28	5.05	-2.32	-17.89	-2.75	-10.48	20.12	23.93
920131	3.21	27.48	2.37	19.07	13.08	0.62	6.42	0.40	3.25	8.58	5.10	6.98	5.90	-4.16	5.00	22.85	34.25	4.77	36.43	5.25	8.63
920228	-11.02	-7.91	-12.17	-4.72	-6.53	-7.80	-6.48	-11.76	-10.69	-2.17	0.00	-8.15	-15.91	-15.15	-5.49	-12.32	2.86	-13.58	-14.74	-5.61	-15.24
920331	3.81	29.80	11.88	15.70	-1.40	6.47	13.50	16.67	2.11	2.22	14.55	12.43	-4.50	-20.00	-1.16	-17.36	9.72	-17.14	5.73	7.05	15.73
920430	0.92	-2.72	8.70	50.00	-2.02	-6.07	13.31	5.14	7.59	-0.87	-2.22	-10.53	12.28	2.68	4.12	7.00	13.92	12.07	5.31	0.14	13.59
920529	7.50	13.60	1.75	-7.62	7.56	4.47	2.32	1.80	2.27	-7.11	-3.57	-1.58	-7.20	-11.32	-3.95	7.48	-4.44	9.23	7.42	4.66	-1.54
920630	-11.50	-13.38	-2.16	-5.67	-10.30	-14.24	-12.18	-3.59	-12.74	-4.76	-3.70	-8.34	-16.44	-19.90	-7.06	-11.30	-8.14	-21.13	-7.74	-20.32	-4.35
920731	6.00	-9.76	-3.08	-15.85	4.39	-9.71	-5.16	-10.87	-6.57	2.00	-30.77	-11.49	-3.83	-4.35	5.06	-2.94	-34.18	5.36	-0.14	-16.60	-16.36
920831	-26.42	-43.24	-13.64	-47.40	-10.68	-18.35	-8.16	-28.92	-14.84	-16.67	-8.33	-16.03	-23.86	-19.48	-12.65	-27.27	-45.19	-28.81	-21.94	-31.87	-30.43
920930	-35.90	-9.52	-13.68	-12.35	-4.71	4.65	3.70	10.29	4.59	-11.76	-15.15	-9.09	12.69	4.84	3.45	1.39	17.54	3.98	1.47	8.60	-17.19
921030	34.00	12.28	-2.44	-55.63	5.32	7.41	-5.71	28.89	7.02	-13.33	10.71	14.00	17.22	33.85	13.33	6.85	1.49	39.53	2.11	12.82	7.55
921130	20.90	4.69	0.00	0.32	1.08	-17.93	11.36	1.38	11.48	3.08	-9.68	-7.89	5.08	-2.30	-2.35	5.13	-23.53	13.33	-8.62	1.87	7.89
921231	-5.79	-17.66	6.67	43.45	-1.43	26.32	3.17	12.93	-0.01	0.50	1.18	13.64	-1.44	-9.81	1.60	0.18	28.01	9.31	4.39	2.52	15.16
Average:	0.435	-0.021	1.580	0.762	2.788	1.783	1.781	2.251	3.311	1.798	2.093	3.035	1.980	2.387	2.717	2.177	1.104	2.281	-2.255	-0.192	1.013

Data 1993-94

Date	AXE	DNL	DYN	ELK	HRA	KVI	NBY	ORK	SAG	SMD	TOM	UTO	HEA	BON	NAL	STO	VAR	WWI	BANK	FORBK	DEI
jan.93	-0.41	7.86	4.3	8.09	3.62	-0.21	4.51	1.81	-4.41	5.45	-4.69	7.28	-8.9	-3.47	9.69	3.47	8.17	0.9	22.29	6.14	-1.16
feb.93	-10.53	15.97	-9.55	-8.16	-5.24	8.33	2.21	11.54	-4.62	28.17	18.52	15.63	5.39	4.05	16.22	-5.88	-12.5	1.33	21.74	12.1	22.86
mar.93	45.59	-2.9	10.56	13.33	8.49	13.54	4.32	9.28	16.94	0	-8.75	-1.35	17.61	31.17	1.4	10	-4.76	24.34	15.79	-3	6.98
apr.93	20.2	38.06	7.87	25.49	0.68	4.07	3.63	-3.4	8.28	12.09	-5.27	16.84	4.91	15.84	14.29	8	-16.67	15.76	23.13	2.32	8.7
mai.93	-6.03	-7.31	-6.88	2.34	-15.51	0.91	4.8	-3.71	-7.07	-4.89	-2.19	8.64	12.56	6.91	0.42	18.88	-18.02	10.65	-7.14	-8.16	6.49
jun.93	7.5	9.63	-4	37.41	-3.24	8.61	-2.36	7.45	3.16	7.3	-11.94	-0.57	4.13	21.46	-9.7	-0.67	-9.73	17.99	6.72	-9.04	-7.03
jul.93	13.04	35.11	7.14	18.89	-5.88	18.05	9.09	-8.42	10.2	-4.85	16.95	6.29	21.43	0.67	13.55	35.14	54.05	12.06	5.76	16.52	31.31
aug.93	12.31	-20.47	21.67	10.28	15.18	4.34	4.69	10.31	4.32	4.95	18.48	4.3	-3.92	-7.28	-4.53	-9.33	78.95	-6.96	12.42	25.18	0
sep.93	-6.85	-5.94	-7.76	-11.02	-0.78	-9.11	-2.74	9.02	-8.88	-14	-0.31	-10.31	-10.88	-5.36	-18.1	5.88	-16.67	-4.76	-5.98	-7.15	-15.38
okt.93	28.68	7.37	7.43	28.57	6.64	18.52	12.28	17.9	8.44	29.07	15.34	17.24	13.74	0.38	9.72	11.68	12.94	8.21	10.86	12.37	29.09
nov.93	-9.14	-14.71	4.15	-2.22	-19.78	0.55	-3.87	-5.5	-13.77	0.9	-9.57	0.98	-9.73	-10.53	-0.48	-3.8	-5.21	-14.19	-7.21	-7.74	4.93
des.93	11.32	8.57	8.38	29.53	11.18	22.94	1.3	11.14	-1.74	6.47	22.82	22.57	6.3	3.78	-5.35	15.46	27.47	9.23	-2.77	-1.11	17.45
jan.94	14.13	67.28	20.43	22.83	13.36	2.31	19.54	3.88	20.85	9.01	0.98	9.31	9.81	18.22	9.24	11.68	2.59	-1.41	19.8	4.22	14.29
feb.94	-4.90	-4.43	1.35	-1.90	5.42	8.12	3.12	0.69	-6.43	8.62	-6.65	0.00	-4.13	-3.40	-8.92	8.59	-25.00	0.00	-3.71	-4.35	-5.85
mar.94	6.19	1.99	-5.33	-13.04	-8.90	-3.49	-12.12	-14.83	-7.50	6.67	1.69	5.07	3.31	-7.04	-7.22	-3.72	0.00	-1.43	-9.42	-3.59	-12.70
apr.94	-9.71	31.82	0.35	-3.33	-9.77	-6.17	6.28	-5.67	2.70	-0.80	-8.33	-0.69	2.56	6.82	5.00	2.90	-27.78	2.17	-1.23	-5.41	1.91
mai.94	-5.37	0.49	13.68	8.05	-1.32	0.72	-0.79	6.90	4.61	-3.35	-1.47	-4.44	3.77	-2.45	10.05	8.57	29.22	-1.43	-7.45	2.29	0.59
jun.94	-1.78	-6.86	0.63	-6.38	5.26	-16.08	-7.40	-9.80	1.94	-6.37	5.97	-11.85	-3.03	-8.09	-5.00	-5.26	-7.15	0.00	1.33	-4.82	-8.24
jul.94	6.02	25.26	14.60	-0.57	5.00	8.01	16.89	10.86	1.90	8.84	3.17	9.24	8.75	0.00	0.00	0.93	6.41	2.17	2.30	3.60	14.10
aug.94	-4.55	-3.36	-3.52	-9.14	1.98	0.97	2.15	-5.51	1.86	-8.75	4.10	-5.38	-8.62	-6.40	3.16	-1.83	-4.81	-1.42	0.47	3.13	9.83
sep.94	-19.05	-15.22	1.40	-9.43	-12.84	-9.90	-4.78	-15.77	-9.15	-5.48	-1.64	0.41	-7.86	-5.98	-7.14	-2.80	-29.10	-2.88	-9.60	-8.78	-3.32
okt.94	10.29	4.62	-0.83	-8.33	5.80	-4.96	6.02	-1.28	3.36	-4.35	4.00	-4.45	-3.07	-9.09	3.30	0.00	-12.51	-5.19	7.33	6.92	-3.70
nov.94	6.67	7.84	-1.12	9.85	9.71	4.48	-1.89	12.73	0.00	25.76	0.64	-8.90	12.68	5.00	2.13	10.58	-24.46	12.50	2.29	10.68	-3.30
des.94	1.25	-3.18	11.86	14.11	11.53	13.93	2.70	15.21	-3.90	8.43	-1.91	6.05	3.13	-4.76	2.08	2.61	-19.48	11.11	5.06	8.72	7.95
Average:	4.37	7.40	4.03	6.47	0.86	3.77	2.82	2.31	0.88	4.54	2.08	3.41	2.91	1.69	1.38	4.96	-0.59	3.70	4.28	2.10	4.83