# ANALYSIS AND SOLUTION OF CARDINALITY CONSTRAINED QUADRATIC PORTFOLIO OPTIMIZATION PROBLEM USING EIGEN PORTFOLIOS

by

Necdet Serhat Aybat

B.S., Industrial Engineering, Boğaziçi University, 2003

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Industrial Engineering
Boğaziçi University
2005

## ACKNOWLEDGEMENTS

I would like to take this opportunity to thank my thesis supervisor Prof. İlhan Or for his invaluable, directive and continuing support throughout this study. He has provided the necessary motivation to ensure that this thesis is accomplished.

I would like to thank Prof. Vedat Akgiray for taking part in my thesis committee, for his concern and valuable comments.

I would also like to thank Assoc. Prof. David Pinhas for taking part in my thesis committee, for his concern and valuable comments.

I owe an immeasurable debth to my parents, Sinan Aybat and İnci Aybat, and to my angel, my all, my very self, Özge. They provided me with great support and were very kind and tolerative to me throughout my life when I am so helpless and depressed. Today, therefore, I would like to share my happiness with them and with all my friends.

# ABSTRACT


## ANALYSIS AND SOLUTION OF CARDINALITY CONSTRAINED QUADRATIC PORTFOLIO OPTIMIZATION PROBLEM USING EIGEN PORTFOLIOS

The aim of this study is to simplify the classic Markowitz quadratic portfolio optimization model by transforming it into an equivalent but simpler optimization model so that when additional realistic constraints, such as cardinality and minimum trading constraints, are added to the model, the transformed model can be solved with widely available MIP solvers. The columns of the matrix that defines the linear transformation of variables are the eigen vectors of the sample covariance matrix found in the objective function of the Markowitz classic model.

The classic model and the transformed model with and without realistic additional constraints are compared. During the comparison, two directions are followed. The first one is to compare both models in terms of their solution times required by the exact optimization techniques. Second direction followed is to approximate the separable objective function of the transformed model with piecewise linear functions and to compare the classic model that has a quadratic objective function with the transformed model that has a linear objective function in terms of the approximation error and their solutions times.

To sum up, due to the simple structure of the transformed model and its separable objective function, cardinality constrained quadratic optimization model can be solved through solvers that are widely used by practitioners such as Excel Solver.

# ÖZET

## KÜME BOYU KISITI OLAN KARESEL PORTFÖY ENİYİLEME PROBLEMİNİN ÖZYÖNEY PORTFÖYLERİ KULLANILARARAK İNCELENMESİ VE ÇÖZÜMÜ

Bu çalışmanın ana hedefi Portföy Eniyilemesi konusundaki klasik Markowitz ortalama varyans karesel programlama modelinin daha basit, eşdeğer bir modele dönüştürülmesi ve bu sayede küme boyu ve en az alım-satım kısıtları gibi gerekçi tamsayı kısıtlar modele eklendiğinde optimizasyon yazılımı gereksinimi açısından daha mütevazi ve yaygın kullanımlı paketlerle çözülebilir alternatif bir modelin elde edilmesidir. Dönüşümü tanımlayan matrisin kolonlarnı Markowitz'in modelinindeki varyans kovaryans matrisinin özvektörleri oluş turmaktadır.

Problemi zorlaştıran gerçekçi tam sayı kısıtlarının bulunduğu ve bulunmadığı durumlarda da dönüştürülmüş model ile Markowitz'in klasik modelinin karşılaştırmalarında iki yol izlenmiştir. Birinci yolda, iki yaklaşımın da standard en iyileme teknikleri ile çözülüp harcadıkları işlemci zamanları karşılaştırılmış; ikinci izlenen yolda ise dönüşümden sonra, yeni modelin amaç fonksiyonunun ayrışabilir hale gelmesinden faydalanılmıştır. Yeni amaç fonksiyonunun parçalı doğrusal fonksiyonlarla yaklaşıklanması ile elde edilen doğrusal amaç fonksiyonlu model ile Markowitz'in karesel amaç fonksiyonlu modelinin çözümlerinin yaklaşıklama hatası ve işlemci zamanı yönlerinden karşılaştırılması ikinci yolu tanımlamaktadır.

Sonuç olarak, dönüştürülmüş modelin basit yapısı ve parçalı doğrusal fonksiyonlarla yaklaşıklanmaya açıklığı hem excel-solver gibi daha yaygın kullanımlı paketlerle çözülebilirliğini arttırmakta, hem de ilave tamsayı kısıtların klasik ortalama-varyans modelinin çözümünü çok zorlaştırması sorununu büyük ölçüde bertaraf etmektedir.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF SYMBOLS/ABBREVIATIONS

| | |
|---|---|
| $A_k$ | $k^{th}$ order principal minor of the hessian matrix, $H$ |
| $\mathbf{det}()$ | Determinant operation symbol |
| $E(\xi_0, ..., \xi_L)$ | Total approximation error function of variable grid points |
| $f_{Transformed}(W)$ | The objective function of Model 6 |
| $H$ | Hessian matrix of total approximation error function |
| $K$ | Number of assets with positive weight in the optimal portfolio |
| $L$ | Number of intervals used to partition $[-1, 1]$ |
| $N$ | Total number of assets available |
| $NVEC$ | $N \times N$ matrix, each column corresponds to an eigen-portfolio |
| $NVEC_i$ | $i^{th}$ column of $NVEC$ corresponding to asset weights of $i^{th}$ eigen-portfolio |
| $P$ | $T \times N$ observation matrix |
| $P_{ij}$ | An element of $P$ denoting the return of $j^{th}$ asset in period $i$ |
| $R$ | Minimum required portfolio return |
| $\bar{P}$ | $T \times N$ average return matrix |
| $\bar{\mathbf{p}}$ | A row of $\bar{P}$ |
| $R^N$ | N dimensional real space |
| $ResLim$ | Real time resource limitation of solver |
| $S$ | Diagonal matrix whose $i^{th}$ diagonal is equal to the reciprocal of $VEC_i$'s sum of components |
| $T$ | The number of periods return data collected for each asset |
| $VEC$ | An $N \times N$ matrix of which $i^{th}$ column is $VEC_i$ |
| $VEC_i$ | $i^{th}$ eigen-vector of $\widehat{\Sigma}$ |
| $W$ | $N$-vector, the proportion of the investor's wealth allocated to each eigen-portfolio |
| $w_i$ | The proportion of the investor's wealth allocated to $i^{th}$ eigen-portfolio |
| $X$ | $N$-vector, the proportion of the investor's wealth allocated to the assets |
| $Y$ | $N$-vector, binary decision variables |

| | |
|---|---|
| $\delta$ | $N$-vector, lower bound on asset weights, $\delta \geq \mathbf{0}$ due to short selling restrictions |
| $\gamma$ | $N$-vector, upper bound on asset weights |
| $\Lambda$ | Diagonal matrix equal to $VEC^T \; \widehat{\Sigma} \; VEC$ |
| $\Lambda'$ | $N \times N$ diagonal matrix, $i^{th}$ diagonal element of $\Lambda'$ is $\widehat{\sigma_i^p}$ |
| $\lambda_v$ | Continuous variable used in $\lambda$-form approximation |
| $\mu$ | Theoretical mean of multinormal distribution of monthly asset returns |
| $\widehat{\mu}$ | $N$-vector, estimated expected return vector of assets |
| $\widehat{\mu^p}$ | $N$-vector, $i^{th}$ element of the vector is $\widehat{\mu_i^p}$ |
| $\widehat{\mu_i^p}$ | Estimated expected return for $i^{th}$ eigen-portfolio |
| $\Sigma$ | Theoretical covariance matrix of multinormal distribution of monthly asset returns |
| $\widehat{\Sigma}$ | $N \times N$, sample covariance matrix of the asset returns |
| $\widehat{\sigma_i^p}$ | Estimated variance for $i^{th}$ eigen-portfolio |
| $\xi_v$ | $v^{th}$ grid point v goes from 1 to L+1 |
| $\Theta$ | Convex nonlinear function of $\xi$, defined over the closed interval $[a, b]$ |
| $\hat{\Theta}$ | Piecewise linear approximation of $\Theta$ |
| $\theta$ | Trade off parameter |

# 1. INTRODUCTION

Development and maintenance of sound portfolios consisting of a number of financial securities is a key issue in the financial management of corporations and individuals. This bunch of securities, however, cannot be chosen arbitrarily. Instead, the portfolio must be a balanced whole, providing the investor with protections and opportunities with respect to a wide range of contingencies [1].

When dealing with a portfolio problem, three major steps are to be considered. First, a data model that satisfactorily represents the data, mostly coming from asset returns, should be selected. Then the objective of the investor and the constraints that he is facing should be determined. Finally, in the third step a suitable optimization method should be selected [2].

Markowitz [1], in his now classic model assumed a multivariate normal distribution for the asset returns. However, in real life, individual asset returns show signs of a higher probability extreme values that is not consistent with the multivariate normality assumption. In order to capture this divergence from multivariate normality higher moments could be incorporated in the model. Furthermore, another complexity is that the distribution of parameters varies over time. Nevertheless, modeling and thorough analysis of such intricate data is a hard and complex task in itself; and whether its study justifies itself in terms of efficiency and accuracy is another question. Therefore, as in the Markowitz' model, a multivariate normal distribution for the asset returns is assumed in this study. This assumption, which is also undertaken by many other researchers of portfolio optimization, makes life easy since the sample mean vector and the sample variance-covariance matrix are jointly sufficient statistics for the joint normal distribution.[1]

The qualitative description of a portfolio selection study is to construct a portfolio that suits the needs of the investor most, from a feasible set of portfolios. Today, many

[1]See Anderson , (1971 page 77) for proof.

selection models are used to build up the "optimal" portfolio, such as the geometric mean return, stochastic dominance, lower partial risk measures and utility analysis, which is adopted in this study as a selection criterion. Utility analysis can be reduced to the well known mean-variance analysis when the investors are utility maximizers, prefer more to less, avoid risk; and either the asset returns are distributed joint normally or the utility function of the investor is quadratic. For instance, when the investor has a log utility function and the returns follow joint normal distribution, maximizing the expected utility and maximizing the geometric mean return lead to the same optimal portfolios [3]. Nevertheless, when the above conditions do not hold, maximizing the geometric mean return is not equivalent to maximizing the expected utility of the investor, and vice versa. In other words, in general, different selection criterion leads to different optimal portfolios while the selection criteria are totally up to the investor. In this study, maximization of the expected utility is taken to be the selection criterion, while asset returns are assumed to be joint normally distributed. In other words, the minimization of portfolio variance is adopted as the selection criterion.

Feasible set of portfolios is determined by the constraints that the investor is facing. In the Markowitz' model, there are simple non-negativity constraints, which prohibit short selling, and two linear constraints: the first one, the normalization constraint, which guarantees that the investor invests all of his money to the securities, and the second one, which enforces that the mean return of the portfolio satisfy a certain predetermined level. With only those above constraints, Markowitz' mean-variance portfolio problem becomes a quadratic optimization problem. However, modeling the optimal portfolio selection problem with only those constraints is not very realistic since it is lacking practical real life limitations, such as minimum and maximum buy and sell amounts and limits on the number of assets in the portfolio. When anyone of those realistic side-constraints is added, the problem becomes nonlinear (quadratic) mixed integer optimization problem, which is much harder to tackle with. In this study, cardinality side-constraint, which limits the number of assets in a portfolio, is added to the Markowitz' classic mean-variance model.

The aim of this study is to transform the Markowitz' mean variance model to an equivalent but simpler model (i.e. in terms of the structure of its objective function), whose objective function is of separable form. It is shown that due to some properties of the resulting new constraint set and the new objective function, the new transformed model allows one to solve the complex realistic model approximately with widely available MIP (mixed integer programming) solvers. Furthermore, the factors that affect the quality of approximation are also investigated.

The study is organized as follows: The literature survey on the single period portfolio optimization problem is given in Chapter 2. Chapter 3 introduces the classical model with cardinality constraints. Chapter 4 describes the transformation that, when applied to the classical model, makes the new objective function separable. Exploiting the structure of the covariance matrix, when its rank is low, and of the transformed model through piecewise linear approximation of its objective function are discussed in Chapter 5 and Chapter 6, respectively. Section 7.2 and Section 7.3 in Chapter 7 compare solution performances of both models under the original Markowitz environment and when cardinality and minimum buy-sell constraints are deployed, respectively. Chapter 8 presents the concluding remarks and possible future research.

# 2. LITERATURE SURVEY

Portfolio management is primarily a task of wealth allocation among numerous financial securities. The ultimate objective, which is to construct a portfolio that provides the investor with protections and opportunities with respect to inherent uncertainty about future outcomes, is the same for almost every investor. Nevertheless, the ultimate objective is highly dependent on the investor, thus the preferred strategies vary a lot from one to another. In order to start any discussion about the "optimal" portfolio construction, one should first make some assumptions about the general characteristics of investors' preferences [3]:

- Investors prefer more to less
- Investors are risk averse

Next, all possible alternatives that an investor faces, in other words, the opportunity set, which consists of all possible portfolios, should be defined. Since investors are assumed to have conflicting objectives that are high return versus low risk, the portfolios that cannot be dominated by any others in the opportunity set form the efficient frontier. Then from the opportunity set every investor selects the portfolio that suits best to his needs according to his preferences.

In 1952, Harry Markowitz [4], in his seminal paper "Portfolio Selection" argued the explicit recognition of risk and its quantification in terms of variance. And he introduced two problems that are to find the minimum variance portfolio for a given return level and to find the portfolio with highest expected return for a given variance. In 1959, H. Markowitz [1] discussed at length the construction of the optimal portfolio based on the means and variances of portfolio returns, which defines the classical mean-variance approach to the problem of portfolio optimization. According to mean-variance approach, means and variances of portfolios constitute sufficient information for the decision problem. Investors who choose mean-variance approach to solve their decision problem face an efficient frontier that matches every possible portfolio mean return with the lowest possible variance of portfolio return.

Although there are alternative approaches such as maximizing the geometric mean return, safety first, stochastic dominance and etc., mean-variance approach is the traditional and the most widely accepted method for portfolio management. Mean-variance approach has its roots in utility theory. It is shown that if investors behave according to some postulates, then their selection behavior among choices is indistinguishable from the others who behave according to expected utility theorem. According to the theorem, while making decisions, investors behave in such a way in order to maximize their expected utility. As stated in [3], mean-variance approach holds exactly when

1. investors are utility maximizers,
2. prefer more to less,
3. avoid risk
4. either the asset returns are distributed joint normally or the utility function of the investor is quadratic.

Even though the assumptions seem to be strict, H. Markowitz [1] has shown that violation of joint normality or quadratic utility function assumptions does not make the analysis invalid since most of the time its results holds approximately when assumption 4 is violated partially.

In the economics literature the properties of the utility functions, especially those of quadratic utility function, is a well studied subject. In terms of the mean-variance approach to portfolio management, examining the economic properties of the quadratic utility function is essential, for a quadratic utility function of an investor can be defined in terms of mean and variance of his final wealth regardless of the distribution of the returns [3]. Thus, if the joint normality assumption of the asset returns does not hold, then whether the investors' utility function is quadratic or not arises as an important question to be answered. To summarize the properties of the quadratic utility function briefly, a generic form of the utility function is given: $U(W) = W - bW^2$

[3], where W is the final wealth of the investor. It should be noted that for a single period all properties of their utility function with respect to final wealth are also true for all properties with respect to return. For mean-variance approach to be valid, investors should be risk averse, which implies that the utility function has negative second derivative, which holds when b is positive. Furthermore, having investors "prefer more to less" necessitates that the first derivative of their utility function be positive, which, for the quadratic utility function, holds when $W \leq \frac{1}{2b}$. Then, in the valid range of W, examining the absolute risk aversion and relative risk aversion properties of the quadratic utility function reveals that an investor with quadratic utility function shows increasing absolute risk aversion, which means as his wealth increases he reduces the investment on risky assets, and that as the amount of investment on risky assets decreases, the percentage of investment on risky assets decreases, in other words the investor shows increasing relative risk as his wealth increases.

The assumptions about the preferences of the investors that they prefer more to less and that they are risk averse are almost consistent with the empirical evidence. On the other hand, to justify the validity of the quadratic utility function, other properties of investors' preferences should be investigated. There are two important studies about the absolute risk aversion and relative risk aversion characteristics of investors. When Blume and Friend [5] examined the Federal Reserve Board survey of the financial characteristics of the investors, they found that as wealth increases the percentage of investment in risky assets stays the same when investment characteristics of investors with different wealth levels were investigated. Furthermore, since constant relative risk aversion implies decreasing absolute risk aversion, they concluded that investors also exhibit decreasing absolute risk aversion. Another study was conducted by Cohn, Lewellen, Lease, and Schlarbaum [6]; they examined the survey results of a brokerage firm and it was concluded that investors exhibit decreasing relative risk aversion, thus decreasing absolute risk aversion. The results of these two major empirical studies have shown that the characteristics of the investors' utility functions do not conform to the properties of quadratic one. Yet, it should be noted that in both studies, due to lack of data, instead of examining the change in investment characteristics of investors as their wealth changes, the investment characteristics of different people with different wealth were examined. In addition to above findings, Mossin [7] argued that approximating other types of utility functions with quadratic utility functions gives good approximation results.

To select the optimal portfolio, mean-variance analysis makes some strong assumptions about the distribution of returns or the preference function of the investor. As in the mean-variance analysis two basic assumptions are also assumed in the second degree stochastic dominance, which is based on an axiomatic model of risk-averse preferences. However, stochastic dominance does not make any assumptions other than these on the distribution of the returns or on the preference function of the investor. Besides, if the asset returns are joint normally distributed and short sales are not allowed, then second degree stochastic dominance and mean-variance analysis produce the same efficient set of portfolios, which renders the stochastic dominance a more general tool, though not practical, than mean-variance analysis. Wlodzimierz Ogryczak and Andrzej Ruszcynski [8] in their article showed that standard semideviation and absolute semideviation as risk measures make the mean-risk model consistent with the second degree stochastic dominance under certain assumptions, in other words significant portions of the efficient frontier found by mean-risk model using aforementioned two risk measures are also efficient under the second degree stochastic dominance rules. The main point is that if standard semideviation is used as a risk measure, then the result hold for every return distribution, which can be even nonsymmetric; otherwise if standard or absolute deviation is used as a risk measure, then the same result holds under the assumption that the return distribution is symmetric.

Furthermore, it can be argued that only means and variances of portfolio returns are not sufficient to define the opportunity set. At this point, other measures such as higher moments of the return distributions can be used. Third order moment is called skewness, where positive skewness means the number of observations above the mode is higher than the number of observations below the mode. Therefore, some authors claimed that maximization of the skewness of the portfolio return could be the other objective of the investors. While in the mean-variance approach opportunity set is a two dimensional region, incorporating the skewness as a selection criterion to mean-variance approach makes the opportunity set a three dimensional region. Furthermore, the efficient frontier becomes a region in the outer shell of a three dimensional opportunity set. For instance Konno and Suzuki [9] incorporated skewness into the mean-variance objective function.

After Markowitz proposed his influential work, many departures from the classical model have been observed. Some authors used other objective functions in order to redefine the risk that investors face or to adjust the rate of return of the portfolio by deducting the transaction costs of buying and selling.

It would be appropriate to start with William F. Sharpe's single index model [10], one of the significant reformulations of the classical model using alternative assumptions about the return structure of securities. Sharpe assumed that the return of each security can be separated mainly into two parts: security specific return and the return that can be explained by a common factor affecting all securities in some way. That common factor, having the most significant influence on the security returns, can be the level of stock market index, gross national product or any thing that has an effect on all the security returns. Moreover, security specific return part has two subparts, namely the constant and the random part, both depend on the specific security. Another thing about the return structure is that the common factor does not affect all the securities equally; therefore, the influence of the factor on each security is measured by a security specific parameter. Since all parameters of the model are estimated using regression analysis, first assumption of the model, the independence of both parts of its return for a given security is guaranteed. However, second strong assumption of the model, independence of firm specific random returns of all securities may or mat not hold, which is the weak part of the model. Sharpe enumerated the advantages of this reformulation that while the number of estimated parameters in the classical model is in the order of $n^2$, in the single index model that number drops down to $3n + 2$, n showing the number of assets in the problem, and that adding an auxiliary variable turns the covariance matrix into diagonal form, which increases the solution speed of the new problem on quadratic solvers.

Following Sharpe [10], in the sense of using factor models, A. Perold and H. Markowitz [11] obtained an alternative formulation based on defining the return structure of securities using multi-factor model. Afterwards, in the year 1984 A. Perold [12] exploited the structure of the covariance matrix of the multi-factor model sparsifying the matrix and solved the model parametrically in the risk-reward trade off parameter in the existence of transaction costs.

H. Konno [13] compared various lower partial risk models in order to show that when the distribution of returns is not symmetric then controlling downside risk can be achieved using some other risk measures such as lower semi-absolute deviation (LSAD), first order below target risk and conditional value-at-risk (CVaR). According to the author, the main reason for mean-variance approach being popular is that mean-variance models are consistent with the maximization of expected utility when the returns follow joint normal distribution, which was believed to be valid for many common stocks. However, various financial securities such as bonds and options are known to exhibit nonsymmetrical return distributions; furthermore there are some studies that revealed not all stocks follow normal distribution. Since both facts challenge the validity of mean-variance models, under the light of this information, author claims that using other measures of risk would be more appropriate. Thus, in the article optimal portfolios generated by alternative mean-risk models were compared using the 60 monthly data of 1100 stocks in the Tokyo Stock Exchange. In the computational study instead of VaR (Value-at-risk), the author chose to minimize CVaR due to the facts that VaR is not a convex function of the weights of the assets in the portfolio, which makes the optimization very difficult, and that CVaR is a convex function of asset weights and it gives an upper bound for VaR [2] . The author concluded that mean-CVaR model and mean-LSAD model can be used to control downside risk when asset return distribution is neither normal nor symmetric. Furthermore, he adds that using factor models makes both models solvable much faster.

In 1991 H. Konno and H. Yamazaki [14] proposed absolute deviation as a risk measure alternative to variance used in the mean-variance analysis. The motivation

[2] See Rockafellar , (2000 page 21) for proof.

behind their study is to alleviate the computational difficulty associated with solving the large scale quadratic programming problem with dense variance-covariance matrix while maintaining the advantages of the mean-variance analysis over the equilibrium models such as CAPM and APT. In the article authors enumerated the disadvantages of the mean-variance approach, which are the difficulty of estimating a vast number of constants, the excessive time required to solve large scale quadratic programming problems and some problems about the applicability of the results that the analysis generate. To give an example about the applicability problems, authors stated that large number of positive weights in the solution, which can be as high as the number of assets in the problem, can make the solution infeasible to be realized in the market due to minimum trading restrictions. Furthermore, they added that many small amounts of securities in the solution have an adverse effect on the portfolio return due to high transaction costs, which are not included in the classical model. According to the observations of the authors, to ease the problem many practitioners round the small weights to zero, which increases the variance of the final portfolio. Another disadvantage of the classical mean-variance approach according to the authors is that the definition of risk in the model does not fit to the definition of the risk in investors' mind. They claim that neither perceived risk nor the return distributions are symmetric around the mean. In their study, Konno and Yamazaki showed that under the condition of joint normally distributed asset returns, absolute deviation of the portfolio is equal to constant times the standard deviation of the portfolio. Therefore, using unbiased estimations of mean returns as input and an estimator of portfolio absolute deviation as the objective function, they formed a linear programming problem to find the efficient frontier. The authors listed some of the advantages of their formulation over the mean-variance formulation: first, the estimation of large covariance matrix is not needed for their formulation; second, when there are no upper bounds on the decision variables, the number of positive variables in the solution of the model is limited up to a number, which is a function of observation number used to estimate the asset returns and can be used as a control parameter; and third, by setting the upper bounds of the variables to integer multiple of minimum transaction limit, some variables are guaranteed to be integer multiples of the transaction limit. In the computational study authors used three datasets; each consisted of the rate of returns of 224 stocks for 60 months. For a given return level, they compared the market portfolios generated by three different methods, which are Konno and Yamazaki's mean-absolute deviation, Markowitz's mean-variance and Sharpe's single index model. The authors concluded that although the compositions of the two market portfolios generated by mean-absolute deviation and mean-variance model were similar, the portfolio generated by single index model was quite different than the others. Furthermore, they compared the behaviors of market portfolios when the ex-post data was employed and concluded that the mean-absolute deviation model can be used as an alternative to Markowitz's mean-variance model.

As a response to H. Konno and H. Yamazaki [14], Yusif Simaan [15] compared the mean-absolute deviation model and the mean-variance model from the point of estimation risk. In the article he argued that ignoring the covariance matrix in the mean-absolute deviation model results in estimation risk that offsets its benefits over mean-variance approach. The author noted that the equivalence of both models requires that true parameters of joint normal distribution are known and the models are solved using those parameters. On the other hand, when unbiased estimators are used instead of true parameters, then the equivalence of the two models is no longer valid. To compare the estimation risk of both models, Y. Simaan first formed the efficient frontier of 20 assets using true parameters that he set. Since solutions of both models are same when the true parameters are used, it does not matter which approach he selected to form the true efficient frontier. Then he generated random data coming from multivariate normal distribution using the true parameters and estimated the required statistics using the generated data to form both models. For different sample sizes of randomly generated data, generated to estimate the distribution parameters, he calculated opportunity costs of portfolios obtained using both methods. However, the opportunity cost of suboptimal portfolio due to estimation error was calculated only for the exponential utility function, which impairs the generality of the article. According to test results, the author concluded that the superiority of a model under perfect information is not maintained when the estimation risk is involved.

In 1992 H. Konno and K. Suzuki [16] proposed an alternative formulation for the Markowitz's mean-variance model. Since the computation time for solving quadratic programming problems depends on the density of the covariance matrix, the computation time required for solving the model proposed by Markowitz and Perold [12], which makes the efficient solution techniques exploiting the sparsity applicable, is less than that of classical model of the same size. H. Konno and K. Suzuki took it one step further by decreasing the number of quadratic terms in the objective function and making it separable. Basically, they exploited the special structure of the estimated variance-covariance matrix. They argued that the number of data points sufficient enough to estimate the covariance matrix of the asset returns is less than one hundred. Furthermore, if the number of assets in the problem is more than one thousand, then the unbiased estimator of the covariance matrix will be a low rank matrix. Applying Cholesky factorization to the low rank matrix makes the objective function separable, decreases the number of quadratic terms in the objective while adding equality constraints to the model, whose number is equal to the number of data points used to estimate the covariance matrix. R. J. Vanderbei and T. J. Carpenter [17] independently proposed the same factorization of the objective function, turning the model into a separable quadratic programming problem at a cost of adding some constraints.

The authors pointed some advantages of their formulation over the classical model and Markowitz and Perold's multifactor model [12]. The first advantage is that their formulation does not require any data preprocessing such as estimation of covariance matrix or any regression analysis to estimate parameters in the multi factor model. Let n be the number of assets and T be the number of data points over which the returns are observed. The other advantage is that the number of quadratic terms in the objective function of the proposed model is T while it is $n^2$ in the classical model and $n+T^2$ in the model proposed by Markowitz and Perold [12]. Moreover, the authors approximated the separable objective function by piecewise linear functions to turn the quadratic programming problem into a linear programming problem and according to the results of their computational studies reported in the article, the efficient portfolios found by using the classical model and their model with piecewise linear approximation were very similar. This study also explains the similarity of the results of mean-absolute-deviation model to that of mean-variance model reported in [14], as absolute deviation, a measure of risk, is a special piecewise approximation to variance.

Other authors also applied piecewise approximation to quadratic portfolio optimization problem to make problem easily solvable. In 1971 W. Sharpe [18] made a diagonalizing transformation of variables to convert the expression for portfolio variance into a diagonal form and applied piecewise linear approximation to terms that sum up to give the total variance. After his transformation new variables become unrestricted in sign; therefore, bounds for each variable should be found to obtain a precise approximation. However, Sharpe did not give explicit bounds for the new variables; instead he indicated in a footnote that restrictions on the original variables will imply implicit restrictions on the new free variables. In addition, two approximation errors were defined, namely the maximum error, which is the maximum deviation from the true curve, and the expected error, which gives the overall deviation of the approximating function from the true curve. In the article, he answered two questions about the approximation scheme: what should be the number of line segments in order to keep approximation error small? And how these line segments should be constructed to minimize the defined error measures? He stated that the optimal procedure for constructing the line segments is to keep their range same for all segments regardless of the number of the line segments used. On the other hand, it was shown that increasing the number of line segments decreases both type of approximation error.

Another author who proposed a piecewise linear approximation is B. K. Stone [19]. In 1973, after Sharpe [18] proposed his approximation scheme, Stone proposed another approximation based on the single-index model of Sharpe [10] while extending the mean-variance criterion to reflect the possibility of different attitudes to market and nonmarket risk. Therefore, in the objective function, the author included market standard deviation, instead of market variance, and nonmarket risk separately with their tradeoff coefficients. Since the market standard deviation is a linear term, it is not approximated. The only approximated part is the nonmarket variance of the portfolio. Thus, it can be argued that while the approximation scheme provided by

Sharpe [18] approximates all the variability in the return of the resulting portfolio, with this approach only the independent part of the variance, nonmarket variance, of the portfolio is approximated. Nevertheless, the weak sides of this approximation scheme are its dependence on the single-index model and its restriction on the feasible set of portfolios due to upper-bound constraints. Furthermore, in the article Stone suggested a mean-variance-skewness criterion and an approximation to resulting cubic objective function to obtain a linear programming problem that approximates the efficient surface provided by the mean-variance-skewness criterion.

In 1996, Daniel Bienstock [20] examined the same problem, which will also be discussed in this study, using exact optimization techniques. He proposed a branch and cut algorithm where disjunctive cuts are implemented in order to create cutting planes. The author also considered MIR (Mixed Integer Rounding) cuts in the article; however, since the selection of multipliers is a critical task in the MIR procedure and there is no clear way to choose them for this problem, he did not adopt MIR cuts. On the other hand, some authors adopted heuristics to deal with the cardinality constrained quadratic portfolio optimization problem. In 2000, Chang and Beasley [21] tailored various meta-heuristics to the portfolio optimization problem with cardinality constraints. In the article it is illustrated that the efficient frontier obtained in the existence of cardinality or minimal trading constraints become discontinuous and shown that some parts of the efficient frontier become inaccessible to the parametric optimization using tradeoff parameter suggested by H. Markowitz. Consequently, the authors attempted to find the portfolios on the efficient frontier using three heuristic algorithms based on genetic algorithms, tabu search and simulated annealing. The gist of approaching the cardinality constrained quadratic optimization problem with heuristics is that those methods can easily adopt different objective functions and handle integer constraints. According to the results produced by each heuristic, the authors suggested pooling the efficient portfolios obtained by different methods to form an efficient frontier. Other authors, Crama and Schyns [2], also applied simulated annealing meta-heuristic to obtain the efficient frontier of the portfolio selection problem. In addition to all type of constrains mentioned above, their problem involves other realistic side constraints such as turnover and trading constraints. The advantages claimed for this study are the same as ones elucidated in [21]. On the other hand, the disadvantages of using heuristics were identified as the difficulty of fine-tuning the parameters of the heuristic, while the performance of the algorithm significantly depends on the selection of those parameters.

# 3. THE CLASSIC QP AND MIQP MODELS

## 3.1. Definition of Symbols

Parameters:

| | |
|---|---|
| $N$ | total number of assets available |
| $\widehat{\mu}$ | $N$-vector, estimated expected return vector of assets |
| $\widehat{\Sigma}$ | $N \times N$, sample covariance matrix of the asset returns |
| $R$ | scalar, minimum required portfolio return |
| $\theta$ | scalar, trade off parameter |
| $K$ | scalar, number of assets with positive weight in the portfolio |
| $\delta$ | $N$-vector, lower bound on asset weights, $\delta \geq \mathbf{0}$ due to short selling restrictions |
| $\gamma$ | $N$-vector, upper bound on asset weights |

Decision variables:

| | |
|---|---|
| $X$ | $N$-vector, the proportion of the investor's wealth allocated to the assets |
| $Y$ | $N$-vector, binary decision variables |

### 3.2. The Classic Markowitz Portfolio Selection Model

**Model 1.** *The Classic Portfolio Selection Model (The Classic QP Model)*

Objective function:

$$min \qquad X^T \, \widehat{\Sigma} \, X \tag{3.1}$$

Constraints:

$$X^T \, \mathbf{1} = 1 \tag{3.2}$$

$$X^T \, \widehat{\mu} \geq R \tag{3.3}$$

$$X \geq \delta \tag{3.4}$$

$$X \leq \gamma \tag{3.5}$$

The objective function of the Markowitz' model, Equation 3.1, is portfolio variance, which is minimized while portfolio-return satisfying a minimum required return level, R. When the optimal solution of the Markowitz' model is parameterized over values of R, the so called "Efficient Frontier" is obtained. Another way of modeling this problem is to incorporate expected return to the objective function.

**Model 2.** *The Classic Portfolio Selection Model with Parametric Objective Function*

Alternative Objective function:

$$min \qquad X^T \, \widehat{\Sigma} \, X \; - \; \theta \, (X^T \, \widehat{\mu}) \tag{3.6}$$

This alternative formulation consists of Equation 3.6 as its objective function, Equation 3.2 and Equation 3.4 as its constraints. Instead of $R$, this model uses $\theta$ as the parameter to form the efficient frontier by changing its value between 0 and $\infty$. The efficient frontier obtained by solving parametric programming problem is a continuous and concave function of portfolio variance as in Figure 3.1. When the objective function with $\theta$ tradeoff parameter (i.e. Equation 3.6) is examined, it is obvious that varying $\theta$ corresponds to changing the slope of the line:

$$PortfolioVariance \; - \; \theta \; PortfolioReturn = Z \tag{3.7}$$

$$\Longleftrightarrow \; PortfolioReturn = \frac{1}{\theta} \; PortfolioVariance \; - \; Z \tag{3.8}$$

Therefore, given a $\theta$ value, minimizing $Z$ is equal to maximizing the intercept of Equation 3.8 while still being feasible, which occurs when Equation 3.8 is tangent to feasible region.

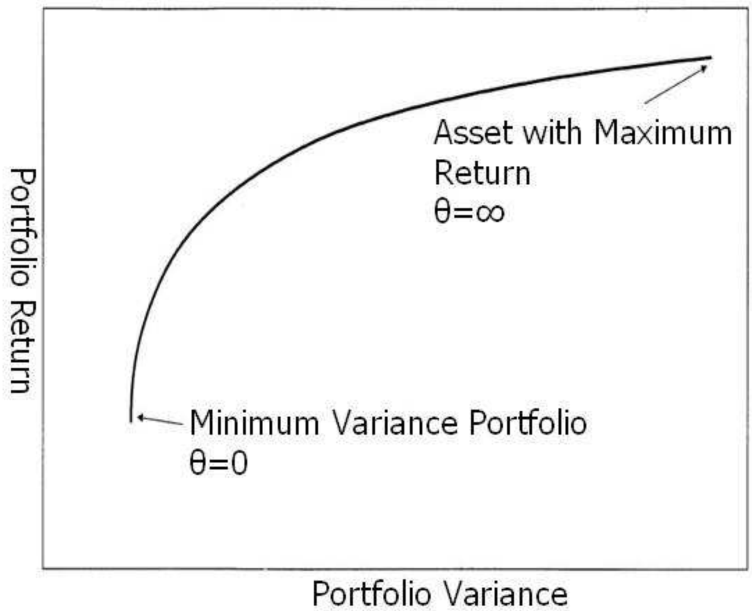


Figure 3.1. Efficient Frontier for The Classic QP Model with $\theta$ Tradeoff Parameter

### 3.3. The Mean-Variance Problem with Additional Realistic Constraints

When more realistic but more complex constraints are added, the efficient frontier becomes discontinuous and loses its concavity. Therefore, solving the parametric programming problem with objective function Equation 3.6 may not provide all the points on the efficient frontier. In other words, some of the points on the efficient frontier may become inaccessible to parametric optimization. The reason is that there could be more than one point corresponding to a single $\theta$ value and the model with objective function Equation 3.6 selects only one of those points. On the other hand, when "The Classic QP Model" is considered, due to concavity of efficient frontier, only one portfolio is efficient for each $\theta$ value.

The extended model involves mainly two different set of constraints. One set of constraint brings minimum trading restriction on the amount of investment in a particular asset that is either zero or bigger than or equal to some determined level for that asset. The second set guarantees that final portfolio will have at most $K$ assets invested. To model those realistic constraints binary variables are used in order to make the model more versatile; however inclusion of binary variables also makes the solution process more complicated. The extended model is as follows:

**Model 3.** *The Cardinality Constrained Portfolio Selection Model*

Objective function:

$$min \qquad X^T \, \widehat{\Sigma} \, X \;-\; \theta \; (X^T \, \widehat{\mu}) \tag{3.9}$$

Constraints:

$$X^T \, \mathbf{1} = 1 \tag{3.10}$$

$$X \geq Y \; \delta \tag{3.11}$$

$$X \leq Y \; \gamma \tag{3.12}$$

$$Y^T \, \mathbf{1} \leq K \tag{3.13}$$

$$Y \in \{0, 1\} \tag{3.14}$$

In order to give an example to how the efficient frontier becomes a discontinuous and non-concave function of portfolio variance, an example similar to one given in [21] will be presented. For the sake of simplicity, suppose there are four assets (i.e $N = 4$) and at most two assets can be found in the final portfolio (i.e. $K = 2$). Furthermore, $\delta$ is equal to $\mathbf{0}$ and $\gamma$ is equal to $\mathbf{1}$. In that case, the feasible region becomes union of possible convex combinations of four chooses two (i.e. 6) asset couples.

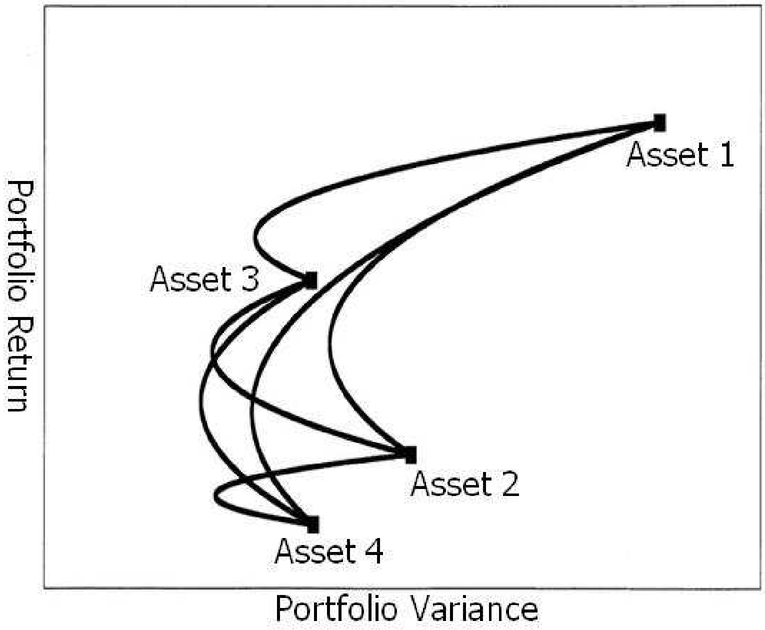


Figure 3.2. Feasible Portfolios for $N = 4$, $K = 2$

When dominated points are discarded, the efficient frontier Figure 3.3 is obtained. However, the parametric programming model with objective function Equation 3.6 can only find the points identified in Figure 3.4 because of reasons mentioned above.

In order to deal with the problem of inaccessible efficient portfolios, the parametric objective function Equation 3.6 is discarded; instead, Equation 3.1 is adopted as the objective function of parametric optimization model. Furthermore, the equality constraint Equation 3.3 is changed to grater than or equal to type constraint. All these changes account for avoiding missing efficient portfolios on the frontier since this model can detect jumps on the efficient frontier. The final model can be seen below:

**Model 4.** *Revised Cardinality Constrained Portfolio Selection Model (The Classic MIQP Model)*

Objective function:

$$min \qquad X^T \, \widehat{\Sigma} \, X \tag{3.15}$$

Constraints:

$$X^T \, \mathbf{1} = 1 \tag{3.16}$$

$$X^T \, \widehat{\mu} \geq R \tag{3.17}$$

$$X \geq Y \, \delta \tag{3.18}$$

$$X \leq Y \, \gamma \tag{3.19}$$

$$Y^T \, \mathbf{1} \leq K \tag{3.20}$$

$$Y \in \{0, 1\} \tag{3.21}$$

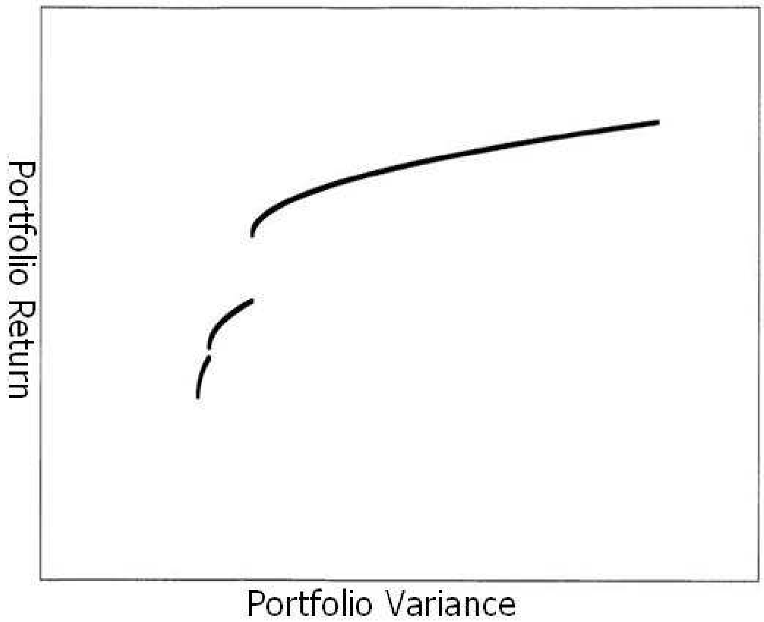


Figure 3.3. True Efficient Frontier for The Classic MIQP Model

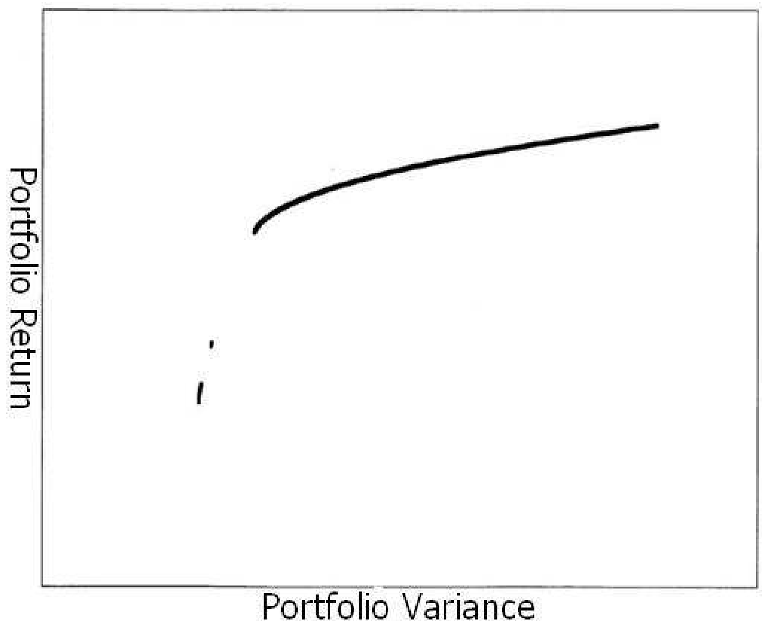


Figure 3.4. Efficient Frontier Found for The Classic MIQP Model

# 4. THE EIGEN PORTFOLIO TRANSFORMATION

First, in Section 4.1 the details of the transformation are discussed on "The Classic QP model" ( i.e. Model 1) for the sake of clarity. Subsequently, in Section 4.2, application of the transformation, defined in Section 4.1, is presented.

## 4.1. The Mechanics of Transformation

The optimization problem focused on is a quadratic programming problem due to its objective function, which includes cross-product terms because of nonzero covariance between asset returns and squared terms because of nonzero variance of the asset returns. In other words, when sample covariance matrix, $\widehat{\Sigma}$, is a dense matrix, the objective function of the "The Classic QP Model" contains many non-linear terms. By applying a linear transformation to all standard unit basis vectors in the N dimensional space, the "degree of nonlinearity" of the objective function can be decreased, at an expense of changing simple non-negativity constraints to general inequality constraints.

By the definition of sample covariance matrix, $\widehat{\Sigma}$, is positive semi-definite and symmetric, which means that even if the eigen values of the matrix are repeated, there can be found a distinct eigen vector for each eigen value of matrix $\widehat{\Sigma}$. (i.e. there is a full set of eigen vectors.) Since symmetric matrices have full set of orthonormal eigen-vectors, those vectors form a basis of $R^N$. In other words, the basis that is used in the "The Classic QP Model", which consists of N unit vectors, can be changed with a new basis, which comes from a defined linear transformation that is represented by a matrix, whose columns are "normalized" eigen-vectors of $\widehat{\Sigma}$ ( "Normalization" of an eigen-vector is done by dividing it by the sum of all its components). If an eigen-vector is divided by a scalar, then it is still an eigen-vector. Therefore, the matrix containing normalized eigen-vectors as its columns is called "normalized eigen vector matrix" and denoted by $NVEC$. Every column of $NVEC$ is a portfolio since its components sum up to one. Notice that those portfolios corresponding to the columns of $NVEC$ can be illegitimate (i.e. containing components that violate short-selling restrictions) because of possible negative values of their components; however, the feasibility (non-negativity of all components) of the optimal portfolio is still maintained. From now on, the portfolios corresponding to the columns of $NVEC$ are called "Eigen-Portfolios". In the Markowitz' model (i.e. Model 1), the decision variables are the weights of the assets; however, in the transformed model, the decision variables will be the weights of the "eigen-portfolios". It is important to note that eigen-portfolios are such formed that their covariance are zero. Thus, the objective function of the transformed model does not contain cross-product terms, and it consists of only squared (diagonal) terms for N eigen-portfolio.

#### 4.1.1. Definition of Symbols

Indices:

| | |
|---|---|
| $i$ | index for eigen-portfolios that changes from 1 to N |

Parameters:

| | |
|---|---|
| $\widehat{\sigma_i^p}$ | estimated variance for $i^{th}$ eigen-portfolio |
| $\widehat{\mu_i^p}$ | estimated expected return for $i^{th}$ eigen-portfolio |
| $NVEC_i$ | $i^{th}$ column of $NVEC$ corresponding to asset weights of $i^{th}$ eigen-portfolio |
| $R$ | Scalar which shows minimum required portfolio return |

Decision variables:

| | |
|---|---|
| $w_i$ | the proportion of the investor's wealth allocated to $i^{th}$ eigen-portfolio |

### 4.1.2. Applying The Transformation to The Classic QP Model

Consequently, the new optimization problem becomes:

**Model 5.** *The Transformed Classic Portfolio Selection Model (The Transformed QP Model)*

Objective function:

$$min \quad \sum_i w_i^2 \widehat{\sigma_i^p} \tag{4.1}$$

Constraints:

$$\sum_i w_i = 1 \tag{4.2}$$

$$\sum_i w_i \ \widehat{\mu_i^p} \geq R \tag{4.3}$$

$$\sum_i w_i \ NVEC_i \geq \delta \tag{4.4}$$

$$\sum_i w_i \ NVEC_i \leq \gamma \tag{4.5}$$

$$w_i \ \ free \qquad \forall i \tag{4.6}$$

It can easily be proven that Markowitz' model (i.e. "The Classic QP Model") and "The Transformed QP Model" are equivalent optimization problems. The objective function of "The Transformed QP Model" is the variance of the portfolio of eigen-portfolios. Constraint 4.2 is the normalization constraint that makes sure the investor invests all of his money to the eigen-portfolios and Constraint 4.3 enforces that the return of the final portfolio satisfies a certain required return level. Finally, Constraint 4.4 guarantees that the total weight of each asset (found as the weighted sum of all eigen-portfolios) be larger than zero, which avoids short selling.

To state the transformation differently, the matrix that defines the linear transformation will be derived below in the light of linear algebra. Let $VEC_i$ be the $i^{th}$ eigen-vector of $\widehat{\Sigma}$; since $\widehat{\Sigma}$ is a real symmetric $N \times N$ matrix, it is guaranteed that there exit $N$ eigen-vectors that are orthonormal to each other. Let $VEC$ be an $N \times N$ matrix of which $i^{th}$ column is $VEC_i$. Through the spectral theorem, $VEC^T \; \widehat{\Sigma} \; VEC$ is a diagonal matrix, $\Lambda$. ( According to spectral theorem it is known that, if full set of eigen vectors of a matrix exist, the matrix can be factorized. The factorization is a product of a matrix of which columns are eigen-vectors of the factorized matrix, a diagonal matrix and the inverse of the matrix of which columns are the eigen-vectors.)

$$VEC^T \; \widehat{\Sigma} \; VEC = \Lambda \tag{4.7}$$

Let $S$ be a diagonal matrix of which $i^{th}$ diagonal is equal to the reciprocal of $VEC_i$'s sum of components.

$$S = \begin{pmatrix} 1/s_1 & 0 & 0 & ... & 0 \\ 0 & 1/s_2 & 0 & ... & 0 \\ 0 & 0 & 1/s_3 & ... & 0 \\ ... & ... & ... & ... & 0 \\ 0 & 0 & 0 & 0 & 1/s_N \end{pmatrix} \quad \text{and} \quad \mathbf{1}^T \; VEC = (s_1 \; s_2 \; s_3 \; ... \; s_N)$$

Multiplying $VEC$ with $S$ from right yields $NVEC$, and every column of $NVEC$ has a property that their components sum up to 1.

$$VEC \; S = NVEC$$

Then, the transformation of decision variables can be briefly summarized as:

$$\begin{aligned} X &= NVEC \; W \\ &= VEC \; S \; W \end{aligned}$$

where W is an N-vector of which $i^{th}$ component is $w_i$. Therefore, if $VEC \; S \; W$ is substituted for X in Model 1, following model will be obtained:

Objective function:

$$min \qquad W^T\ S^T\ VEC^T\ \widehat{\Sigma}\ VEC\ S\ W \tag{4.8}$$

Constraints:

$$W^T\ S^T\ VEC^T\ \mathbf{1} = 1 \tag{4.9}$$

$$W^T\ S^T\ VEC^T\ \widehat{\mu} \geq R \tag{4.10}$$

$$S\ VEC\ W \geq \delta \tag{4.11}$$

$$S\ VEC\ W \leq \gamma \tag{4.12}$$

$$W free \tag{4.13}$$

Using the relation given Equation 4.7, the objective function of "The Transformed QP Model" can be rewritten by substituting $\Lambda$. The new objective function becomes:

$$min \qquad W^T\ S^T\ \Lambda\ S\ W \tag{4.14}$$

and since product of diagonal matrices is also a diagonal matrix, Equation 4.15 can be written, where $\Lambda^{'}$ is a diagonal matrix and each of its elements on the diagonal is the variance of corresponding eigen-portfolio. (i.e. $i^{th}$ diagonal element of $\Lambda^{'}$ is $\widehat{\sigma_i^p}$.)

$$S^T\ \Lambda\ S = \Lambda^{'} \tag{4.15}$$

Consequently, substituting Equation 4.15 into the transformed objective function (i.e. Equation 4.14), the final form of the transformed objective function is obtained.

$$min \qquad W^T\ \Lambda^{'}\ W \tag{4.16}$$

Furthermore, due to the definition of $S$, Equation 4.9 simplifies to

$$W^T \ \mathbf{1} = 1 \tag{4.17}$$

$S^T \ VEC^T$ in Equation 4.10 and Equation 4.11 can be written as $NVEC^T$. Moreover, since every row of $NVEC^T$ corresponds to an eigen-portfolio, product of $NVEC^T$ and $\widehat{\mu}$ is a vector of which components are expected returns of the eigen-portfolios. (i.e. $i^{th}$ element of the vector is $\widehat{\mu_i^p}$.) The final form of the transformed model is as below:

**Model 6.** *Revised Transformed Classic Portfolio Selection Model (The Transformed QP Model)*

Indices:

| | |
|---|---|
| $i$ | index for eigen-portfolios that changes from 1 to N |

Parameters:

| | |
|---|---|
| $\Lambda'$ | $N \times N$ diagonal matrix, $i^{th}$ diagonal element of $\Lambda'$ is $\widehat{\sigma_i^p}$ |
| $\widehat{\mu^p}$ | $N$-vector, $i^{th}$ element of the vector is $\widehat{\mu_i^p}$ |
| $NVEC$ | $N \times N$ matrix, each column corresponds to an eigen-portfolio |
| $R$ | Scalar which shows minimum required portfolio return |

Decision variables:

| | |
|---|---|
| $W$ | $N$-vector, the proportion of the investor's wealth allocated to each eigen-portfolio |

Objective function:

$$min \qquad W^T \; \Lambda' \; W \tag{4.18}$$

Constraints:

$$W^T \; \mathbf{1} = 1 \tag{4.19}$$

$$W^T \; \widehat{\mu^p} \geq R \tag{4.20}$$

$$NVEC \; W \geq \delta \tag{4.21}$$

$$NVEC \; W \leq \gamma \tag{4.22}$$

$$W free \tag{4.23}$$

## 4.2. Applying The Transformation to The Transformed MIQP Model

After having introduced Model 4 in Section 3.3 and explained the mechanics of the transformation in Section 4.1 on a simple model, Model 1; transforming model with additional constraints of Model 4 will be straight forward.

**Model 7.** *The Transformed Cardinality Constrained Portfolio Selection Model (The Transformed MIQP Model)*

Objective function:

$$min \qquad W^T \; \Lambda' \; W \tag{4.24}$$

Constraints:

$$W^T \ \mathbf{1} = 1 \tag{4.25}$$

$$W^T \ \widehat{\mu^p} \geq R \tag{4.26}$$

$$NVEC \ W \geq Y \ \delta \tag{4.27}$$

$$NVEC \ W \leq Y \ \gamma \tag{4.28}$$

$$Y^T \ \mathbf{1} \leq K \tag{4.29}$$

$$Y \in \{0, 1\} \tag{4.30}$$

$$W free \tag{4.31}$$

# 5. EXPLOITING THE LOWRANK SAMPLE COVARIANCE MATRIX

All models mentioned, in Chapter 3 and Chapter 4 above, incorporate the covariance relations between assets as parameters. Since true values of the covariance matrix are not known, one common method is to estimate the covariance matrix using an unbiased estimator. In this study sample covariance matrix, an unbiased estimator, is used; it has some special characteristic that the transformed models (i.e. "The Transformed QP Model" and "The Transformed MIQP Model") can exploit, whereas the classic models (i.e. "The Classic QP Model" and "The Classic MIQP Model") can not. In the literature, H. Konno and K. Suzuki in [16], R. J. Vanderbei, T. J. Carpenter in [17] and D. Bienstock in [20] have also utilized the low rank characteristic of the sample covariance matrix.

Suppose $P$ is a $T \times N$ observation matrix, where $T$ denotes the number of periods return data collected for each asset. $P_{ij}$, which is an element of $P$ ,observation matrix, denotes the return of $j^{th}$ asset in period $i$. Furthermore, $\bar{P}$ is a $T \times N$ average return matrix, each row of which is $\bar{\mathbf{p}}$.

$$\bar{\mathbf{p}} = (\mathbf{1}^T \; P) \; / \; T \quad \text{and} \quad \bar{P} = \begin{array}{c} 1 \\ 2 \\ \ldots \\ T \end{array} \begin{pmatrix} \bar{\mathbf{p}} \\ \bar{\mathbf{p}} \\ \ldots \\ \bar{\mathbf{p}} \end{pmatrix}$$

$\widehat{\Sigma}$ is found using $P$ and $\bar{P}$ as shown below:

$$\widehat{\Sigma} = \frac{1}{(N-1)} \; (P - \bar{P})^T \; (P - \bar{P})$$

Suppose that the rank of $P$ is equal to $T$, which is generally the case. Since $\bar{P}$ introduces a linear dependence among rows of $(P-\bar{P})$, rank of $(P-\bar{P})$ is equal to $T-1$. Therefore, rank of $\widehat{\Sigma}$ is equal to $T-1$. Due to the fact that the number of positive eigen values of a positive-semi-definite matrix is equal to the rank of the matrix, the number of positive terms in $\Lambda'$, in the objective function of Model 6 (i.e. Equation 4.16) is equal to $T-1$ regardless of the number of assets, $N$. Note that, although $\widehat{\Sigma}$ in the objective function of Model 1 is a dense matrix, after the transformation is applied, the objective function of the transformed model becomes sparse.

Moreover, when monthly returns of the assets are used, generally $T$ is around 60, which corresponds to 5 years of data; on the other hand, if the portfolio selection problem deals with high number of assets, where $N$ is around 1000, the values of $T$ and $N$ together imply that the diagonal of $\Lambda'$ has few positive elements. That is to say, without the transformation the number of quadratic terms in the objective function, Equation 3.1, of Model 1 is around $\frac{N^2+N}{2}$; whereas it is equal to $T-1$ in the objective function of the Model 6 (i.e. Equation 4.16).

Having the transformed model take advantage of the low-rank property of the estimated covariance matrix of asset returns in this fashion has two important ramifications: first, some quadratic programming solvers can benefit from the fewer quadratic terms in the objective function of the transformed model; second, if, due to separability of the objective function of the transformed model, piecewise-linear-approximation is employed to the quadratic terms, then only $T-1$ number of variables will need to be approximated.

# 6. APPROXIMATION WITH PIECEWISE LINEAR FUNCTIONS

After the transformation, described in Chapter 4, is applied to the decision variables of "The Classic QP Model" (i.e. Model 1), the new objective function of "The Transformed QP Model" (i.e. Model 6) becomes separable; in other words, there are no cross product terms in the objective function that prevent us from approximating the individual and additive quadratic terms with piecewise-linear functions. The objective function of Model 6 is as follows:

$$f_{Transformed}(W) = W^T \ \Lambda' \ W \tag{6.1}$$

$\Lambda'$ is a diagonal matrix for reasons explained in Section 4.1.2. Furthermore, the number of positive terms on the diagonal would be far smaller than $N$; ( as elucidated in Chapter 5): that number can be as small as $T-1$ (which is especially attractive when $T << N$).

It is stated in Section 4.1.2 that $i^{th}$ element on the diagonal of $\Lambda'$ is $\sigma_i^p$. Hence, Equation 6.1 can be rewritten in the following form:

$$f_{Transformed}(W) = \sum_{i=1}^{N} \ w_i^2 \widehat{\sigma_i^p} \tag{6.2}$$

The maximum number of $w_i^2$ functions to be approximated is equal to $N$; on the other hand, as pointed out above, if $T$ is less than $N$ (which is generally the case) then the number of quadratic functions to be approximated will be reduced significantly.

### 6.1. Description of $\lambda$-form Approximation

The approximation scheme used to approximate $w_i^2$'s is $\lambda$-form approximation. Following description of the approximation scheme is based on [22]. To summarize the $\lambda$-form approximation briefly, suppose $\Theta$ is a convex nonlinear function of $\xi$, defined over the closed interval $[a, b]$. First, the closed interval $[a, b]$ is partitioned into smaller intervals. Let the grid points chosen are as follows:

$$a \;=\; \xi_0,\; \xi_1,\; ...,\; \xi_L = b$$

Then, $\Theta$ is approximated for each interval; the resulting piecewise-linear function is called $\hat{\Theta}$, which approximates $\Theta$ in the interval $[\xi_v, \xi_{v+1}]$ as follows:

$$\hat{\Theta(\xi)} = \lambda\ \Theta(\xi_v)\ +\ (1-\lambda)\ \Theta(\xi_{v+1}) \qquad for\ some \quad \lambda\ \in\ [0,\ 1] \tag{6.3}$$

To generalize Equation 6.3 for interval $[a,\ b]$, the following equations and constraints are sufficient.

$$\hat{\Theta}(\xi) \;=\; \sum_{v=0}^{L} \lambda_v\ \Theta(\xi_v) \tag{6.4}$$

$$\sum_{v=0}^{L} \lambda_v \;=\; 1 \qquad \forall\ v \tag{6.5}$$

$$\lambda_v \;\geq\; 0 \qquad \forall\ v \tag{6.6}$$

$$\lambda_v \;\in\; SOS2 \qquad \forall\ v \tag{6.7}$$

SOS2 is called "special ordered set of type 2", which guarantees that at most two of its adjacent members can be positive. On the other hand, if the problem is minimization of $\Theta$ subject to some linear equalities and inequalities, then the optimal solution of the problem satisfies SOS2 conditions without Equation 6.7 explicitly being stated since $\Theta$ is a convex function. The reason it is important to avoid SOS2 constraint is that in order to model SOS2 constraint, integer variables are used.

### 6.2. Approximating The Transformed QP Model

Let us now reconsider the portfolio optimization problem. As stated above, after the transformation of variables is done, the transformed objective function is equal to the summation of convex $w_i^2$ functions. Since the problem involves the minimization of a convex function over a convex set, there is no need for the SOS2 constraint; thus, no need to introduce integer variables into the approximation model.

From now on, in this chapter the approximation will be described on Model 5. One important thing about Model 5 is that all $w_i$'s are free. Since in order to apply the $\lambda$-form approximation, it is required that closed intervals for each $w_i$ (i.e. $[a_i\ ,b_i]$) be defined, bounds for each interval that are implicitly implied by the constraint set should be determined. Therefore, selecting bounds for each $w_i$ while not imposing extra restrictions on them (other than those imposed by the constraints of Model 5) is an important task. The following lemma simplifies that task considerably.

**Lemma 1.** *All feasible $w_i$'s of Model 5 take values between -1 and 1.*

*Proof.* Equation 4.2, Equation 4.4 and Equation 4.6 of Model 5 are repeated below for the sake of clarity of the proof.

$$\sum_i w_i = 1$$

$$\sum_i w_i \ NVEC_i \geq \delta$$

$$w_i \ \ free \qquad \forall i$$

Suppose that $\delta$ is a zero vector(i.e. all its elements are zero). At the end of the proof, it will be clear to the reader that the following arguments can easily be extented to cover the case where $\delta > \mathbf{0}$.

Suppose that at least one variable is strictly greater than 1 and others are between -1 and 1. Without loss of generality, let $w_1 > 1$.

$$w_1 > 1 \Leftrightarrow w_1 = 1 - w_2 - w_3 \ldots - w_N > 1 \quad \textit{From Equation } 4.2 \tag{6.8}$$

$$\Leftrightarrow w_2 + w_3 \ldots + w_N < 0 \tag{6.9}$$

Furthermore, deploying Equation 4.4, the following statement can be written:

$$(1 - w_2 - w_3 \ldots - w_N)\ NVEC_1 + \sum_{i=2}^{N} w_i\ NVEC_i \geq 0 \tag{6.10}$$

$$\Leftrightarrow \quad NVEC_1 + \sum_{i=2}^{N}(NVEC_i - NVEC_1)\ w_i \geq 0 \tag{6.11}$$

Columns of $NVEC$, which are denoted by $NVEC_i$ (i=1...N), are orthogonal to each other since $\widehat{\Sigma}$ is a symmetric real matrix. Therefore, the row rank of $NVEC^T$ is equal to $N$; in other words, the rows of $NVEC^T$ are linearly independent since $N$ orthogonal row vectors in $N$ dimensional space are always linearly independent. Due to the fact that elementary row operations do not change the rank of the matrix, if first row of $NVEC^T$, which is $NVEC_i^T$, is subtracted from the other rows, all rows of the new matrix will also be linearly independent. Consequently, vectors in Equation 6.11 are linearly independent and no linear combination of linearly independent vectors will be equal to the **0** vector. Thus 6.11 will never hold as an equality. In other words, the following relation, 6.11 can be used instead of 6.11.

$$\Leftrightarrow \quad NVEC_1 + \sum_{i=2}^{N}(NVEC_i - NVEC_1)\ w_i > 0 \tag{6.12}$$

$$\Leftrightarrow \quad (1 - w_2 - w_3 \ldots - w_N)\ NVEC_1 + \sum_{i=2}^{N} w_i\ NVEC_i > 0 \tag{6.13}$$

For all $i$ other than 1, if Equation 6.13 is multiplied by $NVEC_i^T$, the following will be obtained:

$$\Leftrightarrow \quad NVEC_i^T \ NVEC_i \ w_i \ > 0 \qquad \forall i \neq 1 \tag{6.14}$$

$$\Leftrightarrow \quad \|NVEC_i\|^2 \ w_i \ > 0 \qquad \forall i \neq 1 \tag{6.15}$$

$$\Leftrightarrow \quad w_i \ > 0 \qquad \forall i \neq 1 \tag{6.16}$$

$$\Leftrightarrow \quad w_2 \ + \ w_3 \ ... \ + \ w_N \ > 0 \tag{6.17}$$

Since Equation 6.17 contradicts 6.9 (that is having at least one $w_i$ is greater than 1), it can be concluded that no $w_i$ can be grater than 1.

Now, suppose that at least one variable is strictly less than 1 and others are between -1 and 1. Without loss of generality, $w_1 \ < -1$.

$$w_1 \ < -1 \Leftrightarrow \ w_1 \ = \ 1 \ - \ w_2 \ - \ w_3 \ ... \ - \ w_N \ < \ -1 \quad \textit{From Equation } 4.2 \tag{6.18}$$

$$\Leftrightarrow \ w_2 \ + \ w_3 \ ... \ + \ w_N \ > \ 2 \tag{6.19}$$

Since all statements from Equation 6.10 to Equation 6.13 above still hold, Equation 6.13 is rewritten below in order to make reader follow rest of the proof easily.

$$\Leftrightarrow \quad (1 \ - \ w_2 \ - \ ... \ - \ w_N) \ NVEC_1 \ + \sum_{i=2}^{N} w_i \ NVEC_i > 0 \tag{6.20}$$

$$\Leftrightarrow \quad (1 \ - \ w_2 \ - \ ... \ - \ w_N) \ \|NVEC_1\|^2 > 0 \tag{6.21}$$

$$\Leftrightarrow \quad 1 \ - \ w_2 \ - \ ... \ - \ w_N \ > 0 \tag{6.22}$$

$$\Leftrightarrow \quad w_2 \ + \ w_3 \ ... \ + \ w_N \ < 1 \tag{6.23}$$

If Equation 6.20 is multiplied by $NVEC_1^T$, Equation 6.21 is obtained. At the final step of the proof, Equation 6.23 is obtained, which contradicts 6.19 (that is having at least one $w_i$ is less than $-1$). Thus, it can be concluded that no $w_i$ can be less than $-1$. $\square$

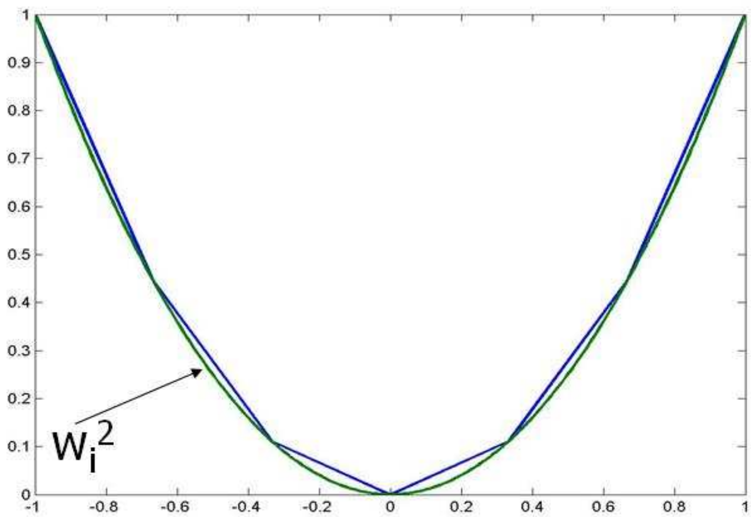


Figure 6.1. $\lambda$-form Approximation Applied to $w_i^2$

Using Lemma 1, all $w_i^2$ terms can be approximated through the $\lambda$-form approximation scheme as seen in Figure 6.1. Although implicit bounds are determined for any $w_i^2$, there are other important questions to be answered. One is about the selection of grid points in the interval $[-1,\ 1]$. The reason behind seeking "appropriate" grid points is to minimize the "approximation error" of deploying piecewise linear approximations to the quadratic terms. In this study "approximation error" refers to the area between $w_i^2$ and its piecewise linear approximation. It is clear that if the number of linear segments (of the piecewise linear approximation) is increased, then the approximation error will be reduced. On the other hand, increasing the number of linear segments of the piecewise linear function corresponds to increasing the number of variables in the model, thus increasing the complexity. Suppose that all $w_i^2$'s, (i=1...N), are to be approximated (i.e. low rank property of the sample covariance matrix can not be exploited, for instance $T > N$). Then, increasing the number of grid points from $L+1$ to $2L+1$ increases the number of variables in the model due to the approximation from $N(L+1)$ to $N(2L+1)$. So, especially the cases where $N$ is large, increasing the number of linear segments has a significant effect on the total number of variables in the model. Therefore, given that interval $[-1,\ 1]$ is to be partitioned into $L$ pieces, what would be the right set of grid points in order to minimize the approximation error is a vital question to be answered. The following lemma points at a simple answer to this important question.

**Lemma 2.** *Suppose that the quadratic function $w^2$ is to be approximated with a piece-wise linear function having L linear segments. Minimum total approximation error can be achieved by selecting grid points that partition $[-1,\ 1]$ interval into L subintervals with equal length.*

*Proof.* Suppose that $[-1,\ 1]$ interval is partitioned into $L$ subintervals by using the following grid points:

$$-1 \ = \ \xi_0,\ \xi_1,\ ...,\ \xi_L = 1$$

The equation of the line segment, $\hat{\Theta}_v(w)$, approximating $\Theta(w) = w^2$ between $\xi_{v-1}$ and $\xi_v$ is stated below:

$$\hat{\Theta}_v(w) = \frac{\xi_v^2 - \xi_{v-1}^2}{\xi_v - \xi_{v-1}}\ (w - \xi_{v-1})\ +\ \xi_{v-1}^2 \quad , \quad \xi_{v-1} \leq w \leq \xi_v \tag{6.24}$$

Then, the total approximation error, denoted by $E$, can be written in terms of $\hat{\Theta}_v$ as follows:

$$E(\xi_0, ..., \xi_L) = \sum_{v=1}^{L} \left( \int_{\xi_{v-1}}^{\xi_v} \hat{\Theta}_v(w)\ dw \right) \ - \ \int_{-1}^{1} w^2\ dw \tag{6.25}$$

In order to minimize total approximation error, Equation 6.25 needs to be minimized in terms of the vector of variable grid points, $(\xi_1, ..., \xi_{L-1})$ ($\xi_0$ and $\xi_L$ are not included in the vector of variable grid points because their values are fixed at $-1$ and 1, respectively). Furthermore, since the last term of Equation 6.25 is constant, it can safely be removed during this minimization process.

Let $\Delta_v(\xi_0, ..., \xi_L)$ be defined as follows:

$$\Delta_v(\xi_0, ..., \xi_L) = \int_{\xi_{v-1}}^{\xi_v} \hat{\Theta}_v(w)\ dw \tag{6.26}$$

$$= \int_{\xi_{v-1}}^{\xi_v} \left( \frac{\xi_v^2 - \xi_{v-1}^2}{\xi_v - \xi_{v-1}}\ (w - \xi_{v-1})\ +\ \xi_{v-1}^2 \right) dw \tag{6.27}$$

$$= (\xi_v - \xi_{v-1}) \left( \frac{\xi_v^2 + \xi_{v-1}^2}{2} \right) \tag{6.28}$$

Using the definition of $\Delta_v(w)$, total approximation error function can be rewritten:

$$E(\xi_0, ..., \xi_L) = \sum_{v=1}^{L} \Delta_v(\xi_0, ..., \xi_L) - \int_{-1}^{1} w^2\ dw \tag{6.29}$$

First order optimality conditions, FOC, of Equation 6.29 are as follows:

$$\frac{\partial E}{\partial \xi_v} = 0 \qquad \forall v \in \{1, 2, ..., L-1\} \tag{6.30}$$

$$\frac{\partial \left( (\xi_v - \xi_{v-1}) \left( \frac{\xi_v^2 + \xi_{v-1}^2}{2} \right) + (\xi_{v+1} - \xi_v) \left( \frac{\xi_{v+1}^2 + \xi_v^2}{2} \right) \right)}{\partial \xi_v} = 0 \qquad \forall v \in \{1, 2, ..., L-1\} \tag{6.31}$$

$$\frac{\xi_{v-1}^2}{2}\ -\ \xi_{v-1}\ \xi_v\ +\ \xi_{v+1}\ \xi_v\ -\ \frac{\xi_{v+1}^2}{2} = 0 \qquad \forall v \in \{1, 2, ..., L-1\} \tag{6.32}$$

$$\xi_v\ -\ \left( \frac{\xi_{v+1} + \xi_{v-1}}{2} \right) = 0 \qquad \forall v \in \{1, 2, ..., L-1\} \tag{6.33}$$

Equation 6.33 shows that in order to satisfy the FOC, grid points which are in the middle of the subinterval defined by their adjacent grid points, should be selected. In other words, if $L+1$ grid points are to be selected, then the points should be chosen such that they partition interval $[-1,\ 1]$ into $L$ equal length subintervals. Furthermore, this is the only extreme point of Equation 6.29 since the solution of Equation 6.33 for all $v$ from 1 to $L-1$ is unique.

In order to show that the vector of grid points obtained by solving the FOC is the global minimum, it should be proven that the hessian matrix is positive definite at that point. This is sufficient for the existence of the global minimum since there is only one extreme point; in other words, having a unique solution of the FOC and a positive definite hessian at that point at the same time implies that Equation 6.29 is convex. The second order partial derivatives of total approximation error, $E$, with respect to variable grid points are given below:

$$\frac{\partial^2 E}{\partial \xi_v^2} = \xi_{v+1} - \xi_{v-1} \qquad \forall v \in \{1, 2, ..., L-1\}$$

$$\frac{\partial^2 E}{\partial \xi_v\ \partial \xi_{v-1}} = \xi_{v-1} - \xi_v \qquad \forall v \in \{1, 2, ..., L-1\}$$

$$\frac{\partial^2 E}{\partial \xi_v\ \partial \xi_{v+1}} = \xi_v - \xi_{v+1} \qquad \forall v \in \{1, 2, ..., L-1\} \tag{6.34}$$

Hessian matrix, $H$, obtained using the second order optimality conditions (SOC) of Equation 6.29, is as follows:

$$
\begin{array}{c c}
 & \begin{array}{ccccccc} \xi_1 & \xi_2 & \xi_3 & \xi_4 & \cdots & \xi_{L-2} & \xi_{L-1} \end{array} \\
\begin{array}{c} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \vdots \\ \xi_{L-2} \\ \xi_{L-1} \end{array} &
\begin{pmatrix}
\xi_2+1 & \xi_1-\xi_2 & 0 & 0 & \cdots & 0 & 0 \\
\xi_1-\xi_2 & \xi_3-\xi_1 & \xi_2-\xi_3 & 0 & \cdots & 0 & 0 \\
0 & \xi_2-\xi_3 & \xi_4-\xi_2 & \xi_3-\xi_4 & \cdots & 0 & 0 \\
0 & 0 & \xi_3-\xi_4 & \xi_5-\xi_3 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \cdots & \xi_{L-1}-\xi_{L-3} & \xi_{L-2}-\xi_{L-1} \\
0 & 0 & 0 & 0 & \cdots & \xi_{L-2}-\xi_{L-1} & \xi_{L-1}-1
\end{pmatrix}
\end{array}
$$

Positive definiteness of $H$ at the unique point satisfying FOC will be shown using induction. Let $A_k$, where $k$ goes from 1 to $L-1$, be the $k^{th}$ order principal minor of the hessian matrix, $H$. Before introducing the induction technique applied, a relation between the principal minors of $H$ is given below:

$$\mathbf{det}(A_k) = (\xi_{k+1} - \xi_{k-1})\ \mathbf{det}(A_{k-1}) \ - \ (\xi_{k-1} - \xi_k)^2\ \mathbf{det}(A_{k-2}) \tag{6.35}$$

Equation 6.35 can be easily verified. To evaluate the determinant of $k^{th}$ order principal minor, $\mathbf{det}(A_k)$ is expanded in cofactors. From the structure of the hessian matrix, $H$, it is true that $k^{th}$ row (i.e. last row) of the $k^{th}$ order principal minor contains only two nonzero elements (which are found in (k,k-1) and (k,k) positions of the principal minor). When the expansion of $\mathbf{det}(A_k)$ in cofactors is carried out for the $k^{th}$ row, Equation 6.35 is obtained.

First, it will be shown that $\mathbf{det}\,(A_1)$ and $\mathbf{det}(A_2)$ are greater than zero. Furthermore, showing that Equation 6.36 is true completes the first step of induction.

$$\mathbf{det}(A_2) - (\xi_3 - \xi_2)\ \mathbf{det}(A_1) \ > \ 0 \tag{6.36}$$

The second step of the induction requires that validity of Equation 6.38 be verified given that Equation 6.37 is true.

$$\mathbf{det}(A_{k-1}) - (\xi_k - \xi_{k-1})\ \mathbf{det}(A_{k-2}) \ > \ 0 \tag{6.37}$$

$$\mathbf{det}(A_k) - (\xi_{k+1} - \xi_k)\ \mathbf{det}(A_{k-1}) \ > \ 0 \tag{6.38}$$

Trivially $\mathbf{det}\,(A_1)$ is equal to $\xi_2 + 1$, which is strictly greater than zero since $\xi_2$ is greater than $\xi_0 \ = \ -1$. Therefore, $\mathbf{det}(A_1)$ is greater than zero.

Subsequently, $\mathbf{det}\,(A_2)$ is calculated below:

$$\begin{aligned} \mathbf{det}\,(A_2) \ &= \ (\xi_2 + 1)\ (\xi_3 - \xi_1) \ - \ (\xi_1 - \xi_2)^2 & (6.39) \\ &= \ (\xi_2 - \xi_0)\ (\xi_3 - \xi_1) \ - \ (\xi_2 - \xi_1)\ (\xi_2 - \xi_1) & (6.40) \end{aligned}$$

It is true that $\xi_k \ < \xi_{k+1}$ since $\xi_k$ is between $\xi_{k+1}$ and $\xi_{k-1}$. Therefore, every term in Equation 6.40 is positive. Furthermore, both $(\xi_2 - \xi_0)$ and $(\xi_3 - \xi_1)$ are greater than $(\xi_2 - \xi_1)$. Consequently, $\mathbf{det}(A_2)$, Equation 6.40, is greater than zero.

To show that Equation 6.36 is true, both determinants in the expression are written explicitly.

$$\begin{aligned} &\mathbf{det}(A_2) \ - \ (\xi_3 - \xi_2)\ \mathbf{det}(A_1) & (6.41) \\ &= (\xi_2 + 1)\ (\xi_3 - \xi_1) \ - \ (\xi_1 - \xi_2)^2 \ - \ (\xi_3 - \xi_2)\ (\xi_2 + 1) & (6.42) \\ &= (\xi_2 + 1)\ (\xi_2 - \xi_1) \ - \ (\xi_2 - \xi_1)\ (\xi_2 - \xi_1) > 0 & (6.43) \end{aligned}$$

The same argument used in showing the positivity of $\mathbf{det}(A_2)$, also applies here.

Finally, the second step of the induction process is proved below:

$$
\begin{aligned}
& & \mathbf{det}(A_{k-1}) \; - \; (\xi_k - \xi_{k-1})\, \mathbf{det}(A_{k-2}) \; > \; 0 \qquad (6.44)\\
\Leftrightarrow & & (\xi_k - \xi_{k-1})\, \mathbf{det}(A_{k-1}) \; - \; (\xi_k - \xi_{k-1})^2\, \mathbf{det}(A_{k-2}) \; > \; 0 \qquad (6.45)\\
\Leftrightarrow & & (\xi_{k+1} - \xi_{k-1})\, \mathbf{det}(A_{k-1}) \; - \; (\xi_k - \xi_{k-1})^2\, \mathbf{det}(A_{k-2}) & \\
& & - \; (\xi_{k+1} - \xi_k)\, \mathbf{det}(A_{k-1}) \; > \; 0 \qquad (6.46)\\
\Leftrightarrow & & \mathbf{det}(A_k) \; - \; (\xi_{k+1} - \xi_k)\, \mathbf{det}(A_{k-1}) \; > \; 0 \qquad (6.47)
\end{aligned}
$$

Equation 6.45 is obtained by multiplying Equation 6.44 by $(\xi_k - \xi_{k-1})$. The direction of the inequality does not change since $(\xi_k - \xi_{k-1})$ is positive by definition. Then, by adding and subtracting $\xi_{k+1}\, \mathbf{det}(A_{k-1})$ to and from Equation 6.45, Equation 6.46 is obtained. Subsequently, using the relation given in Equation 6.35, Equation 6.47 is written as a compact form of Equation 6.46.

To complete the induction, $\mathbf{det}(A_{k+1})$ is shown to be positive.

$$
\begin{aligned}
& & \mathbf{det}(A_k) \; - \; (\xi_{k+1} - \xi_k)\, \mathbf{det}(A_{k-1}) \; > \; 0 \qquad (6.48)\\
\Leftrightarrow & & (\xi_{k+2} - \xi_k)\, \mathbf{det}(A_k) \; - \; (\xi_{k+1} - \xi_k)^2\, \mathbf{det}(A_{k-1}) \; > \; 0 \qquad (6.49)\\
\Leftrightarrow & & \mathbf{det}(A_{k+1}) \; > \; 0 \qquad (6.50)
\end{aligned}
$$

Equation 6.49 can be written from Equation 6.48 since $(\xi_{k+2} - \xi_k)$ is greater than $(\xi_{k+1} - \xi_k)$. Multiplying the first term and the second term with positive numbers does not change the sign of the equation since the first multiplier is greater than the second. Once Equation 6.49 is obtained, Equation 6.50 follows through the relation between the principal minors of $H$, given in Equation 6.35. □

**Model 8.** *Approximating Portfolio Selection Model (Approximating LP Model)*

Model 8 is obtained when objective function of Model 6 is approximated using the $\lambda$-form approximation. As explained in Chapter 5, the number of positive terms in $\Lambda^{'}$, in the objective function of Model 6 (Equation 4.16) will be equal to $T-1$ regardless of the number of assets, $N$. Therefore, the $\lambda$-form approximation is applied only to those variables whose coefficients in $\Lambda^{'}$ are positive. Due to Lemma 1 and Lemma 2, $T-1$ variables with positive coefficients in $\Lambda^{'}$ are approximated in the interval $[-1\ ,1]$, which is partitioned into $L$ equal length subintervals.

**Model 9.** *Approximating Cardinality Constrained Portfolio Selection Model (Approximating MIP Model)*

Model 9 is obtained when the $\lambda$-form approximation explained in the definition of Model 8 is applied to Model 7 which have positive coefficients in $\Lambda^{'}$.

# 7. COMPUTATIONAL EXPERIMENTS

This chapter starts with Section 7.1 describing the design of experiments conducted in order to compare different models under the same set of randomly generated problems. The section continues with explaining the rationales of some assumptions adopted in data generation. In Section 7.2.1, “The Classic QP Model” (i.e. Model 1) and “The Transformed QP Model” (i.e. Model 6) are compared in terms of their solution times using different commercial quadratic programming solvers, namely MINOS 5.51, CPLEX 9.0 and CONOPT 3.0. Then, in Section 7.2.2, the separable objective function of Model 6 is approximated with piecewise linear functions and the obtained linear programming models are solved using the linear programming solver CPLEX 9.0. The aim is to find how close the approximating solutions, obtained by solving the linear models, are to the true values. The classic and transformed models with additional integrality constraints (i.e. Model 4 and Model 7) are compared in Section 7.3.1 and Section 7.3.2. Comparisons in both sections are similar to those in Section 7.2.1 and Section 7.2.2.

## 7.1. Introduction

Throughout the experimental part of this study random instances of portfolio selection models (i.e. Model 1, Model 6, Model 4 and Model 7) are generated under the following parameter settings:

- $N$ is fixed at 300 for Model 1 and Model 6 (i.e. models without integrality constraints.) and at 140 for Model 4 and Model 7 (i.e. models with integrality constraints.)
- $T$, number of monthly observation points (i.e. assets return data), is fixed at 60.
- all the elements of $\delta$ are fixed at 0.001.
- all the elements of $\gamma$ are fixed at 0.3.
- R is fixed at 0.07.
- K is fixed at 20 in the cardinality constrained models.

Given total number of assets in the problem, $N$, in order to generate a random instance of any model, first, an $N \times N$ random covariance matrix, $\Sigma$, is created. Second, depending on variances of assets, a random mean vector, $\mu$, is generated. Throughout the computational study it is assumed that monthly returns of assets follow a multinormal distribution that is defined by $\mu$ and $\Sigma$. In other words, $\Sigma$ and $\mu$ are the theoretical covariance matrix and the theoretical mean vector, respectively, that define the behavior of asset returns. Furthermore, efficient market theorem implies that, given two assets, if mean return of the first asset is higher than that of second, then it should be expected that also variance of the first asset is higher than the second asset. Otherwise, it means that there is mismatch in the prices of the assets, which quickly disappears when arbitrageurs take advantage of the mismatch in the prices. Therefore, while allowing some randomness in data generation, it is also reasonable to assume some dependency between $\Sigma$ and $\mu$; this is accomplished by defining a linear relationship between the variances and the mean returns of assets, as shown in Figure 7.1. In this study risk-free mean retun is taken two per cent, while maximum return is limited to 8.5 per cent. The first step in the random problem generation is to create an $N \times N$ random covariance matrix, $\Sigma$, that is positive semi-definite. Furthermore, the distribution of entries is non-uniform; it is roughly symmetric about 0; all are in $[-1, 1]$. Subsequently, depending on the diagonal of covariance matrix, $\Sigma$, theoretical mean vector, $\mu$, is calculated. As implied by the risk-free return level and the maximum return restriction, entries of $\mu$ are in $[0.02, 0.085]$.

$$Mean\ of\ Asset\ Return = 0.02 + Variance\ of\ Asset\ Return * 0.065$$

After $\mu$ and $\Sigma$ values are created, a set of multi-variate normally distributed asset returns are generated for $T$ points in time. Then, as usually happens in practice, "estimated" mean and variance of return for each asset and covariances between asset returns are estimated using the $T$ observations, generated randomly. Those estimates, $\widehat{\mu}$, $\widehat{\Sigma}$, define a random instance of a model in this study. Other parameters used in the transformed models, $\Lambda'$, $\widehat{\mu^p}$ and $NVEC$, are obtained by applying the transformation, defined in Section 4.1, to $\widehat{\mu}$ and $\widehat{\Sigma}$.

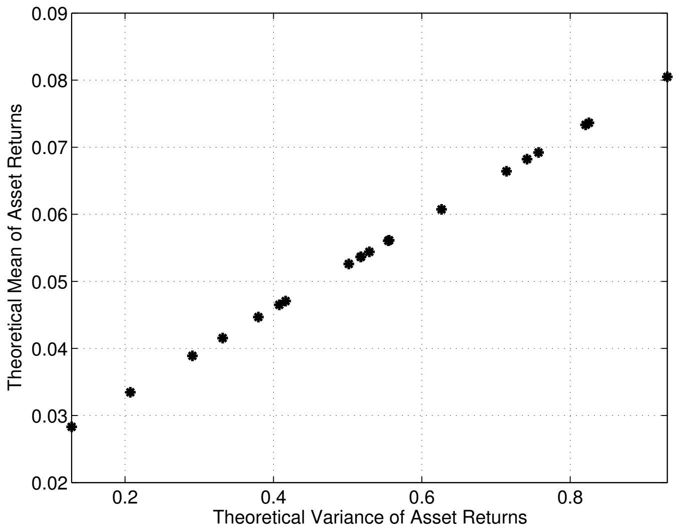


Figure 7.1. Theoretical Mean and Variance of Asset Returns for $N = 20$

Random problems, generated as explained above, are modeled using an algebraic modeling language and solved with above mentioned solvers by invoking them through GAMS 21.4. [23]. Finally, necessary output variables are written into a text file in order to calculate some statistics to compare the classic model with the transformed model. All the experiments, of which results are reported below, have been implemented through a computer program written in MATLAB 6.5 [24] code. This program first creates random problems and their GAMS model, then invokes the mentioned solvers, and finally reports the outputs. The whole package is run on a PC Pentium 3.0 GHz with 1 GB RAM under Windows XP Professional. All computation times reported in the following sections are approximate real times, not CPU times. It is important to note that in following chapters, "**P**" and "**X**" are used in any tabel or graph as abbreviations for the transformed models (i.e. Model 6 and Model 7) and the classic models (i.e. Model 1 and Model 4), respectively.

## 7.2. Comparison of The Transformed QP Model With The Classic QP Model

First, in Section 7.2.1, using different commercial quadratic solvers, “The Classic QP Model” (i.e. Model 1) and “The Transformed QP Model” (i.e. Model 6) have been compared only in terms of their solution times (not in terms of their objective value since they yield the same solutions due to their equivalence). Then, in Section 7.2.2, quality of the optimal solutions produced by the “Approximating MIP Model” (i.e. Model 8) have been examined, and their closeness to the true optimal values, obtained in Section 7.2.1, have been investigated.

### 7.2.1. The Classic QP Model vs The Transformed QP Model

This section compares Model 1 with Model 6. The total number of assets, $N$, is equal to 300, and $\delta$, $\gamma$ and $R$ are fixed at 0.001, 0.3 and 0.07, respectively. 20 random problems are created. For each problem, the number of monthly observations per each asset return, $T$, is set at 60 (in other words 5 years of time).

Each random problem has been modeled through both Model 1 and Model 6. Then, it is statistically tested whether there is difference between average solution times of Model 1 and Model 6 (in other words, from an experiment design point of view, solution time is the response variable and model type is the factor that affects the solution time). Furthermore, since the same set of 20 random problems have been solved with both models, random problems are the blocks of this design. To sum up, the best statistical model that fits this design is the paired t-test. The following three sections (i.e. Section 7.2.1.1, Section 7.2.1.2 and Section 7.2.1.3) contain a discussion of the solution times of both models with respect to the above explained experimental design.

7.2.1.1. Comparison With MINOS. The results reported in this section have been obtained using MINOS 5.51. In Table 7.1, column headings, “Time X” and “Time P”, refer to solution time of Model 1 and Model 6, respectively.

The underlying assumption of paired t-test is that the residuals follow a normal distribution. The residuals in this design are the differences in the solution times of Model 1 and Model 6 for each random problem. Since the statistical hypothesis is that Model 6 can be solved faster than Model 1 by MINOS, here the residuals are obtained by subtracting the solution time of Model 6 from the solution time of Model 1 for each random problem. In Figure 7.2, plotted residuals and the results of the Kolmogorov-Smirnov normality test can be seen. Using a 95 per cent confidence interval, the hypothesis that those residuals follow a normal distribution cannot be rejected with P-value greater than 0.15. Since residuals do not come from significantly nonnormal distribution, the results of paired t-test is relevant.

According to the results of the paired t-test (in which T-value is 14.65 and P-value is less than 0.0001), mean solution time of Model 1 is significantly greater than that of Model 6.

Furthermore, when Figure 7.3 is examined, it can be seen with plain eyes that as the optimal objective value decreases solution times of Model 1 increases, while solution times of Model 6 stay constant. This observation shows that solution time of Model 1 depends on the size of the feasible region (a smaller feasible region leading to a larger computation time). On the other hand, the size of feasible region is highly dependent on the value of $R$ (The smaller the value of $R$, the larger the feasible region).

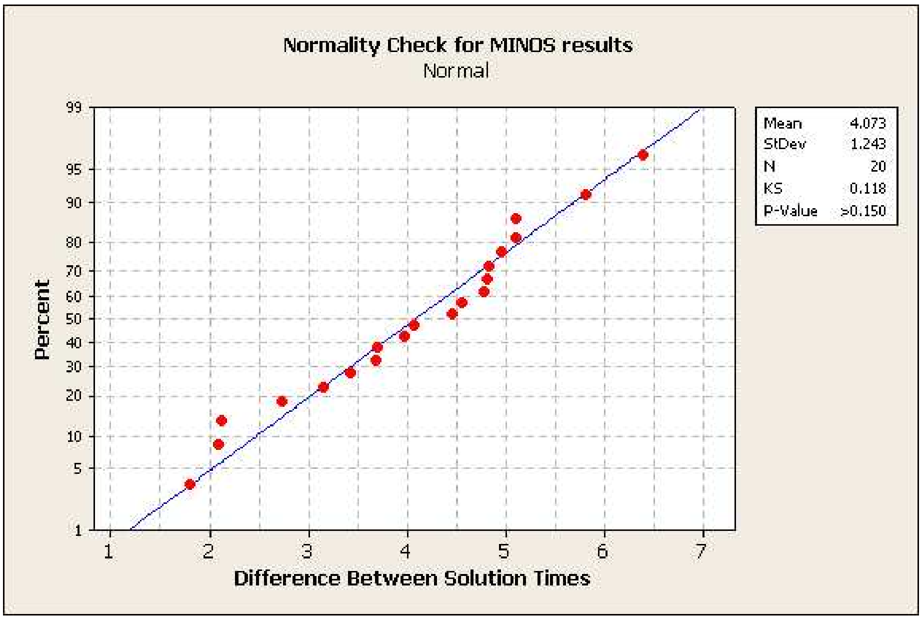


Figure 7.2. Normality Check for MINOS Solution Time Differences

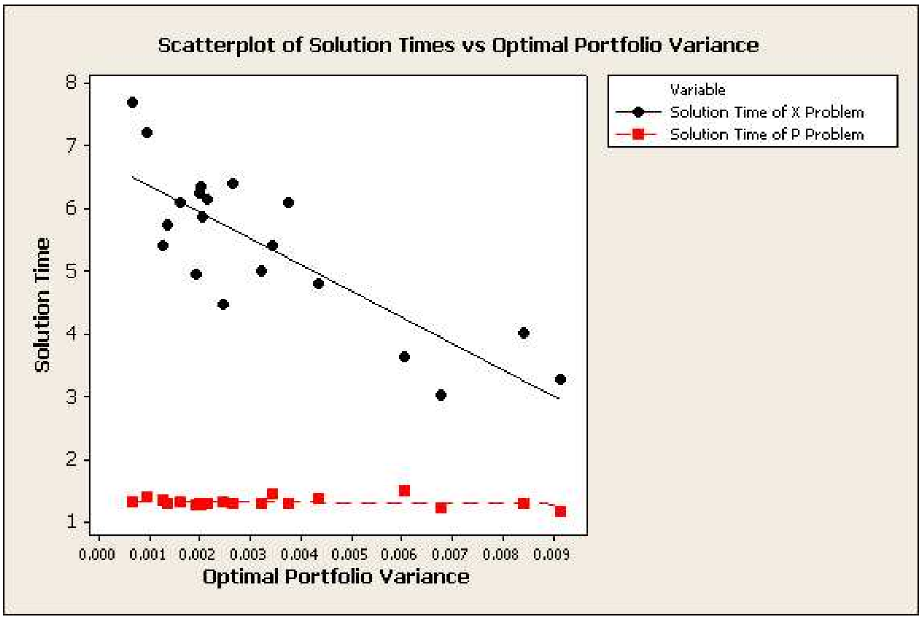


Figure 7.3. Dependency of Solution Times on Optimal Variance for MINOS

Table 7.1. Comparison of Two Models with MINOS 5.51

| | **MINOS** | | |
|---|---|---|---|
| **#** | **Variance** | **Time X** | **Time P** |
| **1** | 0.0091446 | 3.2891 | 1.1953 |
| **2** | 0.0019308 | 4.9688 | 1.2891 |
| **3** | 0.0067776 | 3.0469 | 1.2422 |
| **4** | 0.0060553 | 3.6484 | 1.5234 |
| **5** | 0.0012796 | 5.4141 | 1.3516 |
| **6** | 0.0024611 | 4.4766 | 1.3281 |
| **7** | 0.0013618 | 5.75 | 1.3047 |
| **8** | 0.0084351 | 4.0313 | 1.3047 |
| **9** | 0.0026655 | 6.4141 | 1.3125 |
| **10** | 0.0043682 | 4.8047 | 1.375 |
| **11** | 0.0037735 | 6.1094 | 1.3047 |
| **12** | 0.0032416 | 5.0234 | 1.3203 |
| **13** | 0.0006782 | 7.7109 | 1.3359 |
| **14** | 0.0009567 | 7.2188 | 1.4141 |
| **15** | 0.0034368 | 5.4297 | 1.4609 |
| **16** | 0.0020405 | 6.3672 | 1.2813 |
| **17** | 0.0020022 | 6.2578 | 1.3047 |
| **18** | 0.0020571 | 5.875 | 1.3203 |
| **19** | 0.0016249 | 6.0938 | 1.3281 |
| **20** | 0.0021550 | 6.1484 | 1.3203 |
| **mean** | **0.0033223** | **5.40392** | **1.33086** |
| **std** | **0.0024464** | **1.250491** | **0.071516** |

7.2.1.2. Comparison With CPLEX. The results reported in this section have been obtained using CPLEX 9.0. In Table 7.2, column headings, "Time X" and "Time P", refer to solution time of Model 1 and Model 6, respectively.

The underlying assumption of the paired t-test is that the residuals follow a normal distribution. The residuals in this design are the differences in the solution times of Model 1 and Model 6 for each random problem. Since the statistical hypothesis is that Model 1 can be solved faster than Model 6 by CPLEX, here the residuals are obtained by subtracting the solution time of Model 1 from the solution time of Model 6 for each random problem.

In Figure 7.4, plotted residuals and the results of the Kolmogorov-Smirnov normality test can be seen. Using a 95 per cent confidence interval, the hypothesis that those residuals follow a normal distribution cannot be rejected with P-value greater than 0.127. Since residuals do not come from significantly non-normal distribution, the results of paired t-test are relevant.

Furthermore, according to the results of the paired t-test (in which T-value is -30.54 and P-value is less than 0.0001), the mean solution time of Model 1 is significantly less than that of Model 6.

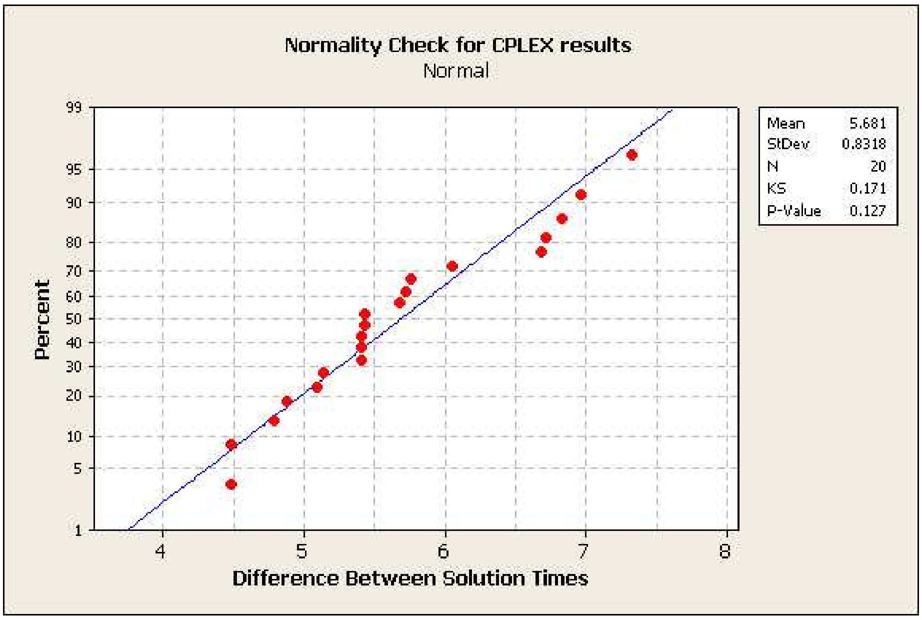


Figure 7.4. Normality Check for CPLEX Solution Time Differences

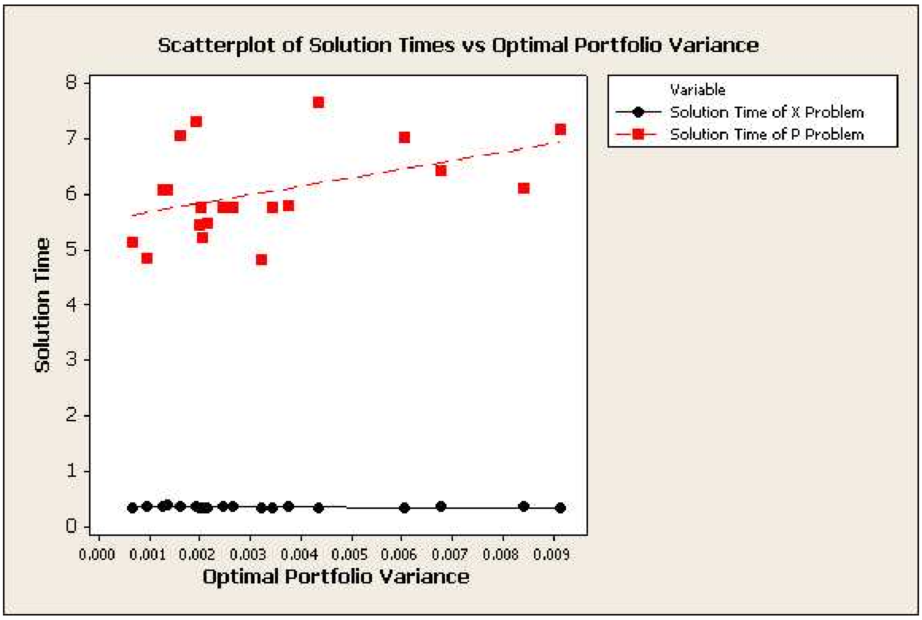


Figure 7.5. Dependency of Solution Times on Optimal Variance for CPLEX

Table 7.2. Comparison of Two Models with CPLEX 9.0

| | **CPLEX** | | |
|---|---|---|---|
| **#** | **Variance** | **Time X** | **Time P** |
| **1** | 0.0091446 | 0.344 | 7.169 |
| **2** | 0.0019308 | 0.36 | 7.318 |
| **3** | 0.0067776 | 0.376 | 6.426 |
| **4** | 0.0060553 | 0.329 | 7.04 |
| **5** | 0.0012796 | 0.36 | 6.08 |
| **6** | 0.0024611 | 0.36 | 5.77 |
| **7** | 0.0013618 | 0.391 | 6.068 |
| **8** | 0.0084351 | 0.36 | 6.113 |
| **9** | 0.0026655 | 0.36 | 5.768 |
| **10** | 0.0043682 | 0.344 | 7.665 |
| **11** | 0.0037735 | 0.36 | 5.785 |
| **12** | 0.0032416 | 0.344 | 4.828 |
| **13** | 0.0006782 | 0.344 | 5.127 |
| **14** | 0.0009567 | 0.376 | 4.857 |
| **15** | 0.0034368 | 0.344 | 5.754 |
| **16** | 0.0020405 | 0.344 | 5.768 |
| **17** | 0.0020022 | 0.344 | 5.437 |
| **18** | 0.0020571 | 0.344 | 5.218 |
| **19** | 0.0016249 | 0.376 | 7.053 |
| **20** | 0.0021550 | 0.329 | 5.469 |
| **mean** | **0.0033223** | **0.35445** | **6.03565** |
| **std** | **0.0024464** | **0.016356151** | **0.83243696** |

<u>7.2.1.3. Comparison With CONOPT.</u> The results reported in this section have been obtained using CONOPT 3.0. In Table 7.3, column headings, "Time X" and "Time P", refer to solution time of Model 1 and Model 6, respectively.

The underlying assumption of the paired t-test is that the residuals follow a normal distribution. The residuals in this design are the differences in the solution times of Model 1 and Model 6 for each random problem. Since the statistical hypothesis is that Model 6 can be solved faster than Model 1 by CONOPT, here the residuals are obtained by subtracting the solution time of Model 6 from the solution time of Model 1 for each random problem. In Figure 7.6, plotted residuals and the results of Kolmogorov-Smirnov normality test can be seen. Using a 95 per cent confidence interval, the hypothesis that those residuals follow a normal distribution cannot be rejected with P-value greater than 0.15. Since residuals do not come from significantly non-normal distribution, the results of paired t-test are relevant.

According to the results of the paired t-test (in which T-value is 5.36 and P-value is less than 0.0001), mean solution time of Model 1 is significantly greater than that of Model 6.

Furthermore, when Figure 7.7 is examined, it can be seen with plain eyes that as the optimal objective value decreases solution times of Model 1 increases, while solution times of Model 6 stay constant. This observation shows that solution time of Model 1 depends on $R$ since as $R$ decreases the optimal objective value decreases. Therefore, as $R$ decreases, solution time of Model 1 increases.

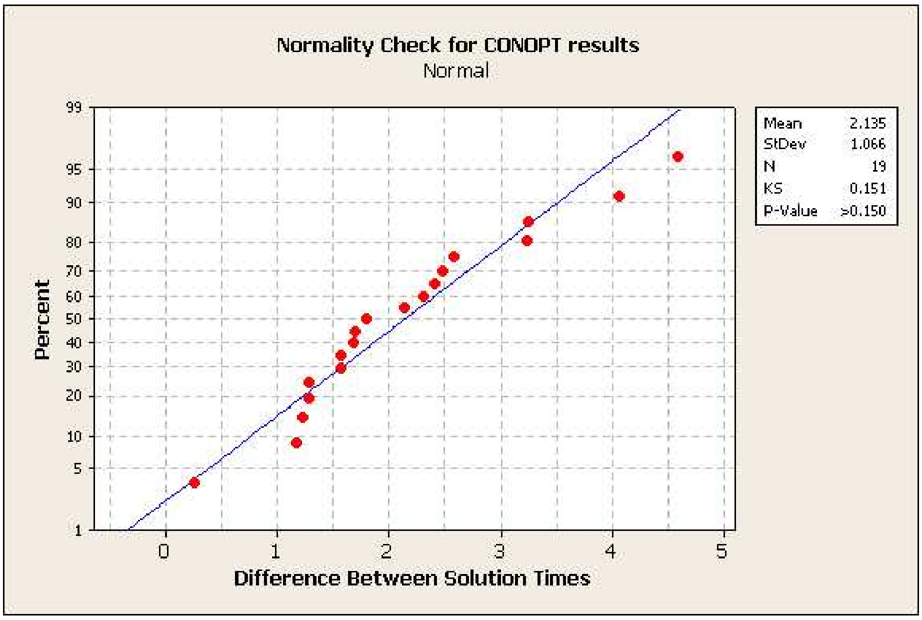


Figure 7.6. Normality Check for CONOPT Solution Time Differences

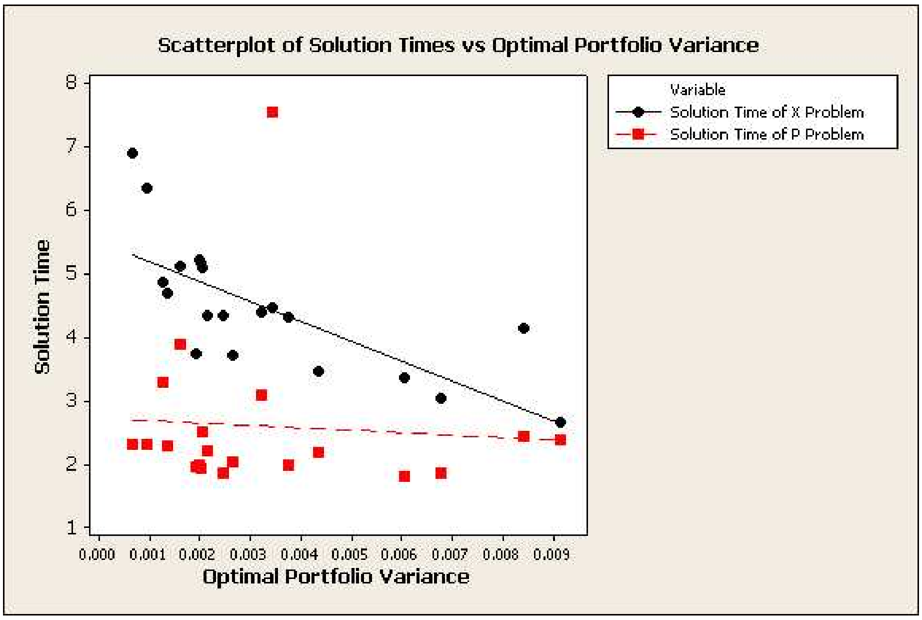


Figure 7.7. Dependency of Solution Times on Optimal Variance for CONOPT

Table 7.3. Comparison of Two Models with CONOPT 3.0

| | **CONOPT** | | |
|---|---|---|---|
| **#** | **Variance** | **Time X** | **Time P** |
| **1** | 0.0091446 | 2.6563 | 2.3984 |
| **2** | 0.0019308 | 3.7578 | 1.9609 |
| **3** | 0.0067776 | 3.0391 | 1.8672 |
| **4** | 0.0060553 | 3.3672 | 1.8047 |
| **5** | 0.0012796 | 4.8672 | 3.2969 |
| **6** | 0.0024611 | 4.3438 | 1.8672 |
| **7** | 0.0013618 | 4.7031 | 2.2891 |
| **8** | 0.0084351 | 4.1406 | 2.4453 |
| **9** | 0.0026655 | 3.7266 | 2.0391 |
| **10** | 0.0043682 | 3.4766 | 2.1953 |
| **11** | 0.0037735 | 4.3125 | 2 |
| **12** | 0.0032416 | 4.3906 | 3.1016 |
| **13** | 0.0006782 | 6.8984 | 2.3125 |
| **14** | 0.0009567 | 6.3672 | 2.3047 |
| **15** | 0.0034368 | 4.4688 | 7.5625 |
| **16** | 0.0020405 | 5.1719 | 1.9453 |
| **17** | 0.0020022 | 5.2344 | 1.9922 |
| **18** | 0.0020571 | 5.0938 | 2.5156 |
| **19** | 0.0016249 | 5.1328 | 3.9063 |
| **20** | 0.0021550 | 4.3594 | 2.2266 |
| **mean** | **0.0033223** | **4.475405** | **2.60157** |
| **std** | **0.0024464** | **1.035706124** | **1.283479603** |

### 7.2.2. QP Models vs Approximating LP Model

This section compares quality of approximate solutions and solution times of the "Approximating LP Model", Model 8, for different $L$ values. In this section, a problem modeled with a piecewise linear objective function is called the "approximate problem".

The total number of assets, $N$, is equal to 300, and $\delta$, $\gamma$ and $R$ are fixed at 0.001, 0.3 and 0.07, respectively. The same 20 random problems with those in Section 7.2.1 have been used in the computational experiments for this section. For each problem, the number of monthly observations per each asset return, $T$, is set at 60 (in other words 5 years of time). Each random problem has been solved for four different $L$ values: 2, 20, 40, 200. Since the objective function is linear, CPLEX 9.0 has been used to solve the random linear programming problem instances. For all random problems the true optimal portfolio variance values are known from the results obtained in Section 7.2.1. True optimal portfolio variance values have been employed as benchmarks for the quality of approximate solutions.

It is important to note that values reported under "Approximate Variance" headings in Figure 7.4 are not the optimal objective function values of the "approximate problems". Those values are obtained when optimal solutions of "approximate problems" are evaluated in the related quadratic objective functions. This is the reason why some approximate variance values do not change when $L$ increases (although it is known that optimal objective function values of "approximate problems" decrease as $L$ increases). The approximate solutions obtained by solving "approximate problems" have been evaluated in related quadratic function since this evaluation removes extra approximation error. The reason of extra approximation error when evaluation done in piecewise linear objective function is that quadratic objective function is always below the piecewise linear objective function. The definition of piecewise linear objective function, coming from the definition of $\lambda$-form approximation, guarantees that. Therefore, the quadratic portfolio variance functions have been used since the piecewise linear function overestimates the portfolio variance corresponding to the approximate solution, obtained by solving Model 8.

The row of Figure 7.4 named "mean" gives column means. Note that the increase in $L$ values from 2 to 40 has only improved the mean optimal approximate portfolio variance from 0.0072042 to 0.0070790, while the mean true optimal portfolio variance is 0.0033223. On the other hand, the mean optimal approximate portfolio variance for L value 200 is equal to 0.0052827, which is smaller than the means for other $L$ values, but still a lot higher than 0.0033223 (the mean optimal portfolio variance). Furthermore, if Table 7.1 is examined, it is seen that Model 6 has been solved for all 20 random problems with an average solution time of 1.33086, whereas the average solution time for Model 8, when $L$ is equal to 200, is 1.8044000. Since as L goes to infinity Model 8 becomes equivalent to Model 6, for a finite value of $L$ Model 6 always results better than Model 8 does. Moreover, since as $L$ increases the solution time of Model 8increases and since, when L is 200, the average solution time of Model 8 is more than that of Model 6, instead of solving Model 8, it is recommeded to solve Model 6 or Model 1.

Another thing worth considering, is the convergence behavior of approximate solutions to optimal solution. Although approximate solutions for $L$ values 2, 20, 40 are almost same, when $L$ is equal to 200, approximate solutions get much closer to optimal values, as seen in Figure 7.8.

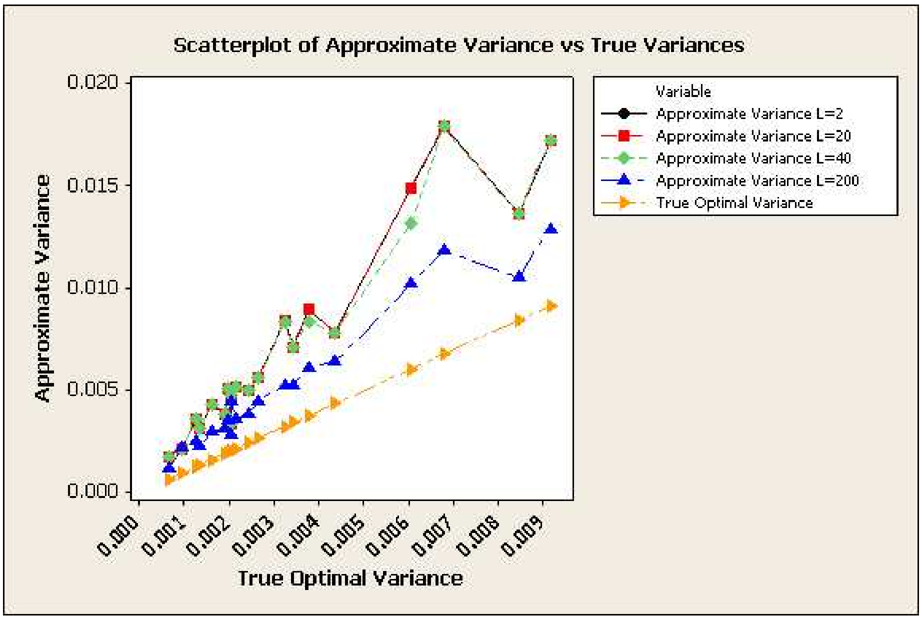


Figure 7.8. Dependency of Quality of Approximate Solutions on L

Table 7.4. Solution Times and Quality of Approximate Solutions for Some L Values

| # | True Optimal Variance | Approximate Variance, L=2 | Solution Time, L=2 | Approximate Variance, L=20 | Solution Time, L=20 | Approximate Variance, L=40 | Solution Time, L=40 | Approximate Variance, L=200 | Solution Time, L=200 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.0091446 | 0.0172280 | 1.237 | 0.0172280 | 1.237 | 0.0172280 | 1.362 | 0.0129072 | 1.871 |
| 2 | 0.0019308 | 0.0038468 | 1.127 | 0.0038468 | 1.237 | 0.0038468 | 1.111 | 0.0031202 | 1.481 |
| 3 | 0.0067776 | 0.0179050 | 1.206 | 0.0179050 | 1.19 | 0.0179050 | 1.33 | 0.0118640 | 2.028 |
| 4 | 0.0060553 | 0.0149190 | 1.127 | 0.0149190 | 1.19 | 0.0131730 | 1.19 | 0.0101959 | 1.715 |
| 5 | 0.0012796 | 0.0036388 | 1.221 | 0.0036388 | 1.377 | 0.0036388 | 1.377 | 0.0024998 | 1.871 |
| 6 | 0.0024611 | 0.0050444 | 1.237 | 0.0050444 | 1.174 | 0.0050444 | 1.314 | 0.0038764 | 1.622 |
| 7 | 0.0013618 | 0.0031450 | 1.206 | 0.0031450 | 1.205 | 0.0031450 | 1.236 | 0.0022969 | 1.731 |
| 8 | 0.0084351 | 0.0136310 | 1.221 | 0.0136310 | 1.174 | 0.0136310 | 1.19 | 0.0105253 | 1.856 |
| 9 | 0.0026655 | 0.0056214 | 1.112 | 0.0056214 | 1.127 | 0.0056214 | 1.205 | 0.0044670 | 1.637 |
| 10 | 0.0043682 | 0.0078416 | 1.268 | 0.0078416 | 1.268 | 0.0078416 | 1.299 | 0.0063933 | 3.369 |
| 11 | 0.0037735 | 0.0090002 | 1.19 | 0.0090002 | 1.237 | 0.0083509 | 1.268 | 0.0060990 | 2.074 |
| 12 | 0.0032416 | 0.0084483 | 1.221 | 0.0084483 | 1.127 | 0.0083407 | 1.19 | 0.0052325 | 1.747 |
| 13 | 0.0006782 | 0.0017797 | 1.065 | 0.0017797 | 1.252 | 0.0017797 | 1.314 | 0.0012259 | 1.637 |
| 14 | 0.0009567 | 0.0021634 | 1.049 | 0.0021634 | 1.111 | 0.0021634 | 1.111 | 0.0021990 | 1.544 |
| 15 | 0.0034368 | 0.0071201 | 1.143 | 0.0071201 | 1.252 | 0.0071201 | 1.174 | 0.0052837 | 1.747 |
| 16 | 0.0020405 | 0.0034027 | 1.221 | 0.0034027 | 1.19 | 0.0034027 | 1.174 | 0.0028593 | 1.669 |
| 17 | 0.0020022 | 0.0051105 | 1.19 | 0.0051105 | 1.19 | 0.0051105 | 1.237 | 0.0035050 | 1.684 |
| 18 | 0.0020571 | 0.0047740 | 1.159 | 0.0047740 | 1.205 | 0.0047740 | 1.283 | 0.0045059 | 1.56 |
| 19 | 0.0016249 | 0.0043089 | 1.112 | 0.0043089 | 1.096 | 0.0043089 | 1.22 | 0.0029839 | 1.607 |
| 20 | 0.0021550 | 0.0051547 | 1.127 | 0.0051547 | 1.143 | 0.0051547 | 1.174 | 0.0036139 | 1.638 |
| **mean** | **0.0033223** | **0.0072042** | **1.1719500** | **0.0072042** | **1.1991000** | **0.0070790** | **1.2379500** | **0.0052827** | **1.8044000** |
| **std** | **0.0024464** | **0.0049246** | **0.0611697** | **0.0049246** | **0.0648471** | **0.0047816** | **0.0775394** | **0.0034230** | **0.3988856** |

### 7.3. Comparison of Both Models in the Existence of Integrality Constraints

First, in Section 7.3.1, "The Classic MIQP Model" (i.e. Model 4) and "The Transformed MIQP Model" (i.e. Model 7) have been compared in terms of their solution times and their objective values using CPLEX MIQP (Mixed Integer Quadratic Programming) solver. Although both models are equivalent, it has been observed that they can yield different solutions with the default settings of CPLEX solver.

In Section 7.3.2, the quality of approximate solutions produced by "Approximating MIP Model", Model 9, have been examined. Their closeness to the true optimal values, obtained in Section 7.3.1, and number of assets in the approximate solution that are in common with the optimal portfolio have been investigated.

#### 7.3.1. The Classic MIQP Model vs The Transformed MIQP Model

This section compares Model 4 and Model 7. The total number of assets, $N$, is equal to 140, and $K$ is set at 20. In other words, 20 assets out of 140 are chosen while minimizing the portfolio variance. $\delta$, $\gamma$ and $R$ are fixed at 0.001, 0.3 and 0.07, respectively. 20 random problems are created. For each problem, the number of monthly observations per each asset return, $T$, is set at 60 (in other words 5 years of time). Finally, resource limit is set to 3600 sec. for CPLEX MIQP solver, which means that if at the end of 3600 sec. the solution at hand cannot be proven to be optimal, the search for optimal will stop and the solution at hand will be reported. Even though this requirement seems restrictive at the first sight, the solution process of only three models out of forty have been terminated because of resource limitation.

Each random problem has been modeled through both Model 4 and Model 7. Then, it is statistically tested whether there is difference between average solution times of Model 4 and Model 7. (In other words, from an experiment design point of view, solution time is the response variable and model type is the factor that affects the solution time). Furthermore, since the same set of 20 random problems have been solved with both models, random problems are the blocks of this design. To sum up, the best statistical model that fits this design is the paired t-test. On the other hand, according to the results, reported in Table 7.5, it is obvious that with above parameter settings Model 4 is superior to Model 7 in terms of solution times.

In Table 7.5, column headings, “Time X” and “Time P”, refer to solution times of Model 4 and Model 7, respectively. Furthermore, column headings, “Variance X” and “Variance P”, refer to optimal portfolio variances found by Model 4 and Model 7, respectively.

When optimal portfolio variances of both models have been compared for each random problem, it is seen that Model 4 has found better solutions in random problems 2,3,11,12,17 and 18; while Model 7 has found better solutions in random problems 4,5 and 16. In the other random problems both models have found the same optimal solution.

Table 7.5. Comparison of Cardinality Constrained Models with CPLEX

| | **CPLEX** | | | |
|---|---|---|---|---|
| **#** | **Variance X** | **Time X** | **Variance P** | **Time P** |
| **1** | 0.00719230766 | 0.547 | 0.00719230766 | 16.914 |
| **2** | 0.00351026395 | 591.268 | 0.00366374195 | 3599.205 |
| **3** | 0.00723946275 | 3.578 | 0.00733699105 | 44.954 |
| **4** | 0.00477057009 | 3.031 | 0.00473345031 | 25.944 |
| **5** | 0.00436925718 | 694.549 | 0.00436839129 | 2712.977 |
| **6** | 0.00307681032 | 194.426 | 0.00307681032 | 3600.348 |
| **7** | 0.00241501021 | 5.765 | 0.00241501021 | 82.916 |
| **8** | 0.00551517902 | 8.065 | 0.00551517902 | 117.966 |
| **9** | 0.00222856731 | 266.35 | 0.00222856731 | 1842.339 |
| **10** | 0.00823378489 | 1.25 | 0.00823378489 | 13.5 |
| **11** | 0.00389964657 | 3600.162 | 0.00397936055 | 3600.268 |
| **12** | 0.00440040702 | 23.564 | 0.00442152108 | 184.392 |
| **13** | 0.00302103544 | 118.663 | 0.00302103544 | 1065.48 |
| **14** | 0.00307894232 | 468.346 | 0.00307894232 | 2102.326 |
| **15** | 0.00298391967 | 62.37 | 0.00298391967 | 456.166 |
| **16** | 0.00604876943 | 0.75 | 0.00589863523 | 10.984 |
| **17** | 0.00652983945 | 3.906 | 0.00655311682 | 33.095 |
| **18** | 0.00310449323 | 2217.406 | 0.00316805372 | 3600.24 |
| **19** | 0.00254854042 | 172.55 | 0.00254854042 | 1382.534 |
| **20** | 0.00286151286 | 6.781 | 0.00286151286 | 71.157 |
| **mean** | **0.00435141599** | **422.16635** | **0.00436394361** | **1228.18525** |
| **std** | **0.00183130813** | **903.8554496** | **0.00182702814** | **1453.939476** |

### 7.3.2. MIQP Models vs Approximating MIP Model

This section compares the quality of approximate solutions of “Approximating MIP Model”, Model 9, for different $L$ values and different resource limitation times (ResLim: 1800 sec. and 3600 sec.). As before, a problem modeled with a piecewise linear objective function is called the “approximate problem”.

The total number of assets, $N$, is equal to 300, and $\delta$, $\gamma$ and $R$ are fixed at 0.001, 0.3 and 0.07, respectively. The same 20 random problems with those in Section 7.3.1 have been used in the computational experiments for this section. For each problem, the number of monthly observations per each asset return, $T$, is set at 60 (in other words 5 years of time). All random problems have been solved for four different $L$ values: 2, 20, 40, 200. Since the objective function is linear, CPLEX 9.0 has been used to solve the random linear mixed integer programming problem instances. For all random problems (with the exception of the one that could be solved in the given time limit) the true optimal portfolio variance values are known from the results obtained in Section 7.3.1. True optimal portfolio variance values have been employed as benchmark for the quality of approximate solutions.

The values under “Approximate Variance” headings in all the figures below are calculated as explained in Section 7.2.2.

A new issue in this section is the concept of “Fine-tuned Variance”. In the following figures, the values under “Fine-tuned Variance” headings are the optimal objective function values of a ”restricted” version of Model 4 applied to the related random problems. ”Restricted Model 4” of related random problem is formed as follows: Binary variables in Model 4 are fixed according to the solution of Model 9. In other words, the binary variables in Model 4 corresponding to assets that have positive weights in the optimal solution of Model 9 are set to one (and the remaining are set to zero). The reason is that for the same set of positive variables, quadratic optimization can easily find portfolios that have less variance than the portfolios produced by the approximating linear program. On the other hand, the solution time required to solve the additional quadratic programming problem is much less than the time needed to solve Model 9. Therefore, by speedily solving an additional (but far smaller) quadratic programming problem, the approximate solution obtained through Model 9 is considerably improved.

When following figures are investigated, a counterintuitive observation is made: solution times of problems that reached their optimal in the given time limit, have not increased significantly when $L$ is increased. On the other hand, the solution quality have improved significantly when $L$ value is changed from 40 to 200 (both in terms of the number of assets in common with the optimal portfolio calculated in Section 7.3.1 and the closeness of the fine-tuned variance to optimal portfolio variance). Then, in order to see whether better solutions can be found when the limitation on resource time is relaxed and set to 3600 sec., the same set of experiments has been carried out with resource limit changed to 3600 sec. Consequently, it has been observed that solutions do not differ significantly. As far as the limited experimentations indicate, if one needs to improve the quality of the "Approximating MIP Model", a policy of "refining the approximating function by increasing L" should be preferred to increasing computation time limit.

Table 7.6. Properties of Solutions Found in 1800 sec., L=2

| | | L=2 | | | | | |
|---|---|---|---|---|---|---|---|
| # | True Optimal Variance | Fine-Tuned Approximate Variance | Approximate Variance | Solution Time | # Common Assets with Optimal Portfolio | # Common Assets with X Portfolio | # Common Assets with P Portfolio |
| 1 | 0.00719230766 | 0.009347025 | 0.01266359786 | 12.844 | 16 | 16 | 16 |
| 2 | 0.00351026395 | 0.005924083 | 0.00825677550 | 1800 | 12 | 12 | 11 |
| 3 | 0.00723946275 | 0.011871709 | 0.02105408098 | 83.094 | 14 | 14 | 13 |
| *4 | 0.00473345031 | 0.006798514 | 0.00993839736 | 1014.3 | 14 | 15 | 14 |
| *5 | 0.00436839129 | 0.008429472 | 0.01079689083 | 1800.1 | 11 | 10 | 11 |
| 6 | 0.00307681032 | 0.005498048 | 0.00654982182 | 1800.1 | 10 | 10 | 10 |
| 7 | 0.00241501021 | 0.008736879 | 0.01517874515 | 1800.1 | 12 | 12 | 12 |
| 8 | 0.00551517902 | 0.011727365 | 0.01763641041 | 338.92 | 12 | 12 | 12 |
| 9 | 0.00222856731 | 0.004090986 | 0.00553942195 | 1800.1 | 16 | 16 | 16 |
| 10 | 0.00823378489 | 0.014486902 | 0.01957574246 | 34.781 | 13 | 13 | 13 |
| 11 | 0.00389964657 | 0.007120016 | 0.00932004099 | 1800.1 | 10 | 10 | 10 |
| 12 | 0.00440040702 | 0.012281836 | 0.01721486916 | 1800.1 | 11 | 11 | 11 |
| 13 | 0.00302103544 | 0.009029472 | 0.01290139102 | 1800.2 | 9 | 9 | 9 |
| 14 | 0.00307894232 | 0.007287068 | 0.01014895072 | 1800.1 | 8 | 8 | 8 |
| 15 | 0.00298391967 | 0.00676029 | 0.01013302051 | 1800.1 | 12 | 12 | 12 |
| *16 | 0.00589863523 | 0.009630425 | 0.01171125750 | 76.563 | 16 | 17 | 16 |
| 17 | 0.00652983945 | 0.011867509 | 0.01534673568 | 111.91 | 11 | 11 | 11 |
| 18 | 0.00310449323 | 0.006008306 | 0.00651874653 | 1800.1 | 8 | 8 | 8 |
| 19 | 0.00254854042 | 0.006952808 | 0.01072980728 | 1800.2 | 11 | 11 | 11 |
| 20 | 0.00286151286 | 0.004371525 | 0.00720678744 | 1800.1 | 14 | 14 | 14 |
| **mean** | **0.00434201000** | **0.00841101182** | **0.01192107456** | **1253.6906** | **12** | **12.05** | **11.9** |
| **std** | **0.00182383978** | **0.00286060713** | **0.00447642047** | **790.089071** | **2.449489743** | **2.62528194** | **2.425739171** |

Table 7.7. Properties of Solutions Found in 1800 sec., L=20

| | | L=20 | | | | | |
|---|---|---|---|---|---|---|---|
| # | True Optimal Variance | Fine-Tuned Approximate Variance | Approximate Variance | Solution Time | # Common Assets with Optimal Portfolio | # Common Assets with X Portfolio | # Common Assets with P Portfolio |
| 1 | 0.00719230766 | 0.009347025 | 0.01266359786 | 32.046 | 16 | 16 | 16 |
| 2 | 0.00351026395 | 0.007055018 | 0.01034284817 | 1800.2 | 11 | 11 | 11 |
| 3 | 0.00723946275 | 0.011789582 | 0.01903877416 | 104.56 | 13 | 13 | 12 |
| *4 | 0.00473345031 | 0.00614143 | 0.00842021808 | 1535.1 | 15 | 15 | 15 |
| *5 | 0.00436839129 | 0.009589201 | 0.01134632995 | 1800.1 | 11 | 10 | 11 |
| 6 | 0.00307681032 | 0.005975375 | 0.00754023363 | 1799.7 | 13 | 13 | 13 |
| 7 | 0.00241501021 | 0.006238693 | 0.01043115335 | 1800.2 | 15 | 15 | 15 |
| 8 | 0.00551517902 | 0.011899891 | 0.01634796780 | 1249.3 | 12 | 12 | 12 |
| 9 | 0.00222856731 | 0.006183044 | 0.00809683780 | 1800.2 | 15 | 15 | 15 |
| 10 | 0.00823378489 | 0.013643284 | 0.02056424365 | 84.626 | 13 | 13 | 13 |
| 11 | 0.00389964657 | 0.007220928 | 0.00786110444 | 1800.2 | 11 | 11 | 13 |
| 12 | 0.00440040702 | 0.010502158 | 0.01315267027 | 1800.1 | 12 | 12 | 12 |
| 13 | 0.00302103544 | 0.006829283 | 0.00867333538 | 1800.1 | 10 | 10 | 10 |
| 14 | 0.00307894232 | 0.008077304 | 0.00923931184 | 1800.2 | 11 | 11 | 11 |
| 15 | 0.00298391967 | 0.00648638 | 0.01128018128 | 1800.1 | 14 | 14 | 14 |
| *16 | 0.00589863523 | 0.009791268 | 0.01134338069 | 552.56 | 16 | 16 | 16 |
| 17 | 0.00652983945 | 0.014670874 | 0.02186352821 | 231.91 | 12 | 12 | 12 |
| 18 | 0.00310449323 | 0.006008306 | 0.00651874653 | 1800 | 8 | 8 | 8 |
| 19 | 0.00254854042 | 0.004318556 | 0.00634281796 | 1799.8 | 14 | 14 | 14 |
| 20 | 0.00286151286 | 0.005044554 | 0.00736884293 | 1799.6 | 15 | 15 | 15 |
| **mean** | **0.00434201000** | **0.00834060776** | **0.01142180620** | **1359.5301** | **12.85** | **12.8** | **12.9** |
| **std** | **0.00182383978** | **0.00292705611** | **0.00463399637** | **705.27588** | **2.158825217** | **2.214782869** | **2.125038699** |

Table 7.8. Properties of Solutions Found in 1800 sec., L=40

| | | L=40 | | | | | |
|---|---|---|---|---|---|---|---|
| # | True Optimal Variance | Fine-Tuned Approximate Variance | Approximate Variance | Solution Time | # Common Assets with Optimal Portfolio | # Common Assets with X Portfolio | # Common Assets with P Portfolio |
| 1 | 0.00719230766 | 0.007716349 | 0.01088803737 | 79.499 | 17 | 17 | 17 |
| 2 | 0.00351026395 | 0.006910558 | 0.00879543233 | 1799.7 | 11 | 11 | 11 |
| 3 | 0.00723946275 | 0.010900498 | 0.01547209877 | 276.6 | 14 | 14 | 13 |
| *4 | 0.00473345031 | 0.006107176 | 0.00775887065 | 366.82 | 13 | 13 | 13 |
| *5 | 0.00436839129 | 0.006518761 | 0.00744535176 | 1799.8 | 10 | 9 | 10 |
| 6 | 0.00307681032 | 0.005269046 | 0.00591928887 | 1800.2 | 10 | 10 | 10 |
| 7 | 0.00241501021 | 0.004554731 | 0.00654056063 | 1799.7 | 15 | 15 | 15 |
| 8 | 0.00551517902 | 0.010923087 | 0.01412124635 | 1799.7 | 12 | 12 | 12 |
| 9 | 0.00222856731 | 0.004087629 | 0.00535015869 | 1799.7 | 16 | 16 | 16 |
| 10 | 0.00823378489 | 0.012023983 | 0.01594554761 | 249.55 | 12 | 12 | 12 |
| 11 | 0.00389964657 | 0.006683103 | 0.00788794575 | 1799.7 | 12 | 12 | 14 |
| 12 | 0.00440040702 | 0.008947866 | 0.01190069762 | 1799.8 | 13 | 13 | 13 |
| 13 | 0.00302103544 | 0.006212392 | 0.00872916312 | 1799.6 | 11 | 11 | 11 |
| 14 | 0.00307894232 | 0.005236904 | 0.00693379603 | 1799.7 | 11 | 11 | 11 |
| 15 | 0.00298391967 | 0.004202451 | 0.00587200746 | 1799.8 | 13 | 13 | 13 |
| *16 | 0.00589863523 | 0.009786965 | 0.01225029157 | 150.8 | 15 | 15 | 15 |
| 17 | 0.00652983945 | 0.010026408 | 0.01358852145 | 663.89 | 12 | 12 | 12 |
| 18 | 0.00310449323 | 0.006035168 | 0.00680014480 | 1799.7 | 7 | 7 | 7 |
| 19 | 0.00254854042 | 0.004228156 | 0.00566990131 | 1799.8 | 14 | 14 | 14 |
| 20 | 0.00286151286 | 0.00512468 | 0.00676926547 | 1800.2 | 12 | 12 | 12 |
| **mean** | **0.00434201000** | **0.00707479560** | **0.00923191638** | **1349.21295** | **12.5** | **12.45** | **12.55** |
| **std** | **0.00182383978** | **0.00250522852** | **0.00346415180** | **713.965932** | **2.305599591** | **2.372540278** | **2.305028827** |

Table 7.9. Properties of Solutions Found in 1800 sec., L=200

| | | L=200 | | | | | |
|---|---|---|---|---|---|---|---|
| # | True Optimal Variance | Fine-Tuned Approximate Variance | Approximate Variance | Solution Time | # Common Assets with Optimal Portfolio | # Common Assets with X Portfolio | # Common Assets with P Portfolio |
| 1 | 0.00719230766 | 0.007651999 | 0.00881466137 | 12.016 | 17 | 17 | 17 |
| 2 | 0.00351026395 | 0.005229641 | 0.00599298691 | 1798.01 | 12 | 12 | 12 |
| 3 | 0.00723946275 | 0.008650694 | 0.00936919586 | 637.246 | 15 | 15 | 15 |
| *4 | 0.00473345031 | 0.00530412 | 0.00592359953 | 601.461 | 16 | 16 | 16 |
| *5 | 0.00436839129 | 0.005247435 | 0.00605313147 | 1800.252 | 15 | 13 | 15 |
| 6 | 0.00307681032 | 0.003594529 | 0.00399378235 | 1800.142 | 17 | 17 | 17 |
| 7 | 0.00241501021 | 0.003615397 | 0.00490322506 | 1800.289 | 17 | 17 | 17 |
| 8 | 0.00551517902 | 0.007111647 | 0.00933119843 | 1800.311 | 16 | 16 | 16 |
| 9 | 0.00222856731 | 0.003078273 | 0.00396824825 | 1800.203 | 19 | 19 | 19 |
| 10 | 0.00823378489 | 0.009081168 | 0.00953128081 | 92.553 | 18 | 18 | 18 |
| 11 | 0.00389964657 | 0.004901586 | 0.00576410686 | 1800.247 | 12 | 12 | 16 |
| 12 | 0.00440040702 | 0.005244683 | 0.00627769551 | 1800.209 | 18 | 18 | 18 |
| 13 | 0.00302103544 | 0.005053416 | 0.00591154116 | 1800.207 | 11 | 11 | 11 |
| 14 | 0.00307894232 | 0.003877039 | 0.00467730831 | 1799.816 | 14 | 14 | 14 |
| 15 | 0.00298391967 | 0.003756396 | 0.00473971515 | 1800.255 | 16 | 16 | 16 |
| *16 | 0.00589863523 | 0.006048769 | 0.00701080029 | 103.924 | 19 | 20 | 19 |
| 17 | 0.00652983945 | 0.006632018 | 0.00808079655 | 229.378 | 18 | 18 | 18 |
| 18 | 0.00310449323 | 0.003771834 | 0.00408714148 | 1800.237 | 16 | 16 | 19 |
| 19 | 0.00254854042 | 0.003492711 | 0.00410442548 | 1800.233 | 16 | 16 | 16 |
| 20 | 0.00286151286 | 0.003447206 | 0.00443934154 | 1800.139 | 17 | 17 | 17 |
| **mean** | **0.00434201000** | **0.00523952804** | **0.00614870912** | **1343.8564** | **15.95** | **15.9** | **16.3** |
| **std** | **0.00182383978** | **0.00179113343** | **0.00192327326** | **728.478836** | **2.258900524** | **2.425739171** | **2.154554539** |

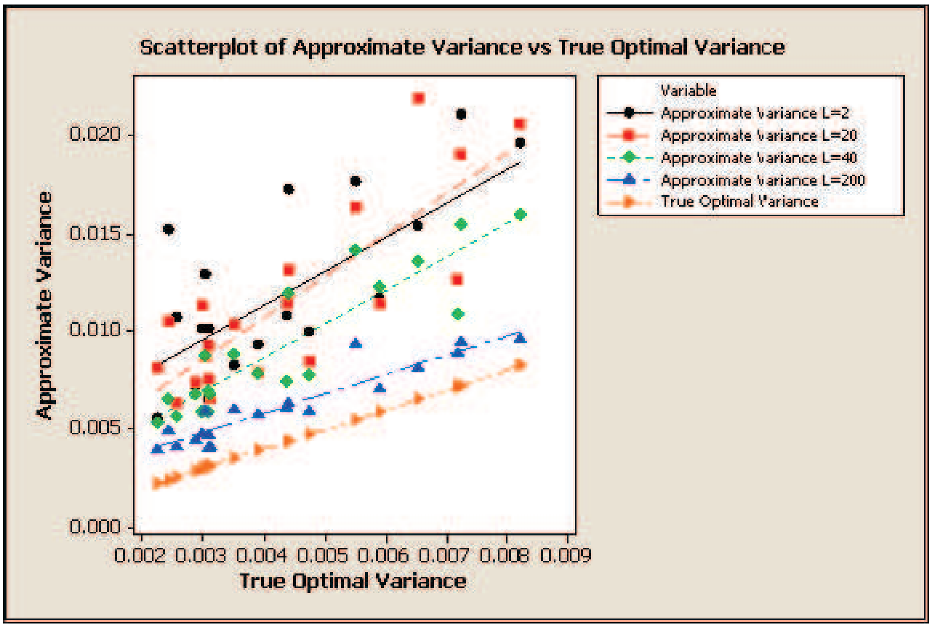


Figure 7.9. ResLim=1800 sec., Effect of L on Quality of Approximate Portfolio Variance

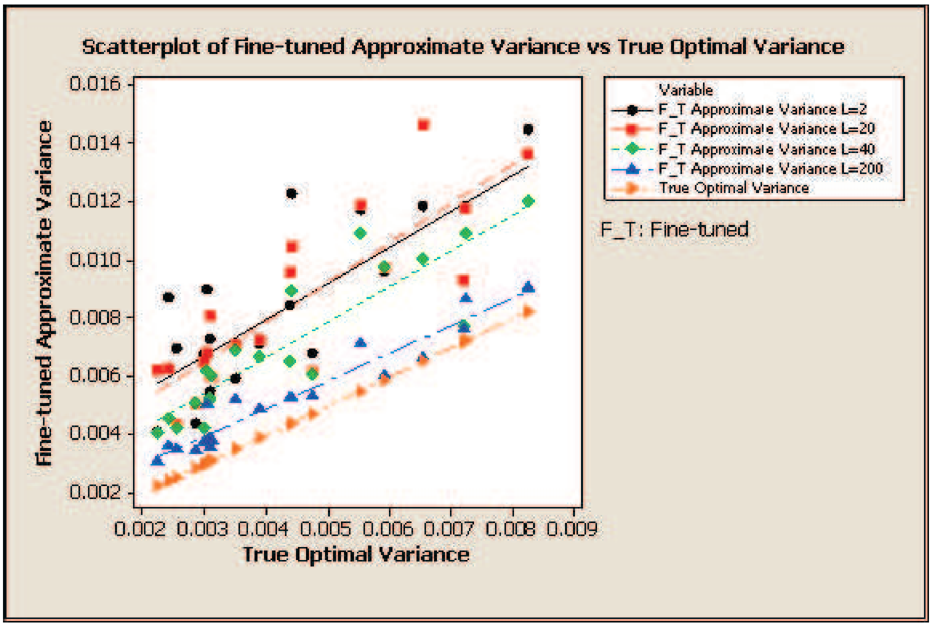


Figure 7.10. ResLim=1800 sec., Effect of L on Quality of Fine-tuned Approximate Portfolio Variance

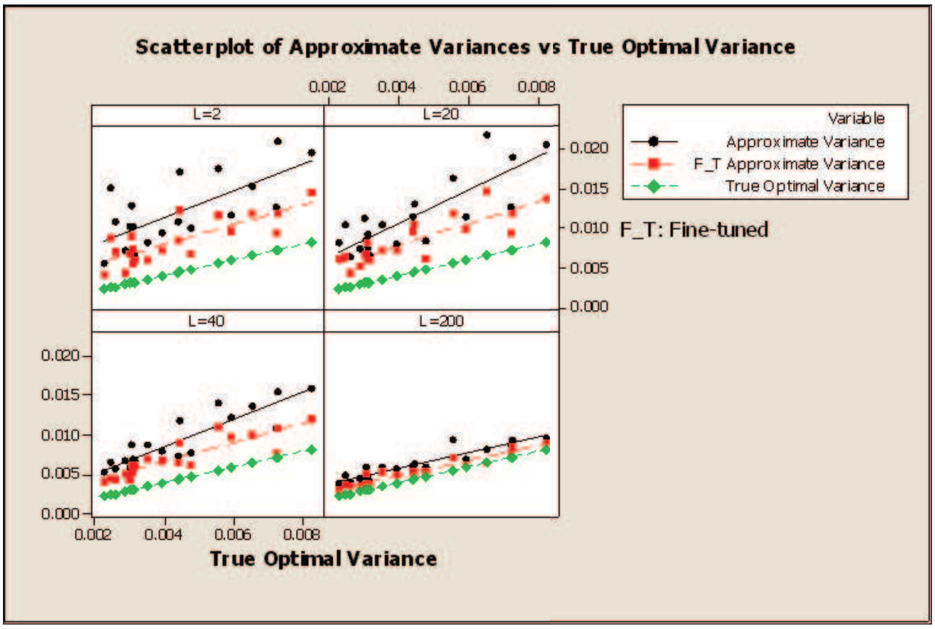


Figure 7.11. ResLim=1800 sec., Effect of L on Quality of Approximate and Fine-tuned Approximate Portfolio Variance

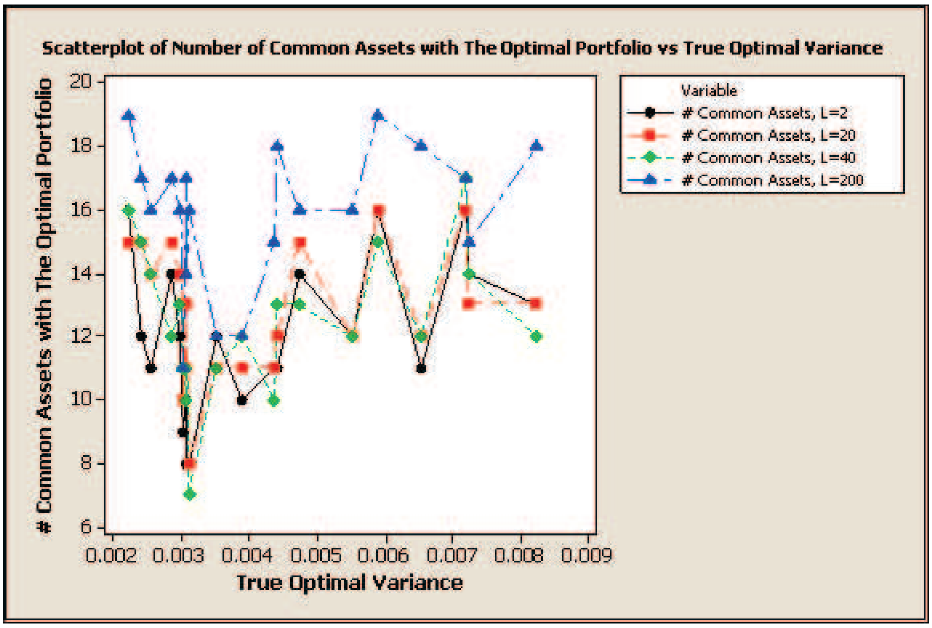


Figure 7.12. ResLim=1800 sec., Effect of L on The Number of Assets in Common with The Optimal Portfolio

Table 7.10. Properties of Solutions Found in 3600 sec., L=2

| | | L=2 | | | | | |
|---|---|---|---|---|---|---|---|
| # | True Optimal Variance | Fine-Tuned Approximate Variance | Approximate Variance | Solution Time | # Common Assets with Optimal Portfolio | # Common Assets with X Portfolio | # Common Assets with P Portfolio |
| 1 | 0.00719230766 | 0.009347025 | 0.01266359786 | 12.855 | 16 | 16 | 16 |
| 2 | 0.00351026395 | 0.005757313 | 0.00717287385 | 3609.636 | 12 | 12 | 12 |
| 3 | 0.00723946275 | 0.011871709 | 0.02105408098 | 84.082 | 14 | 14 | 13 |
| *4 | 0.00473345031 | 0.006798514 | 0.00993839736 | 1019.608 | 14 | 15 | 14 |
| *5 | 0.00436839129 | 0.008429472 | 0.01079689083 | 3599.096 | 11 | 10 | 11 |
| 6 | 0.00307681032 | 0.005498048 | 0.00654982182 | 3600.147 | 10 | 10 | 10 |
| 7 | 0.00241501021 | 0.008736879 | 0.01517874515 | 3600.112 | 12 | 12 | 12 |
| 8 | 0.00551517902 | 0.011727365 | 0.01763641041 | 344.734 | 12 | 12 | 12 |
| 9 | 0.00222856731 | 0.007247029 | 0.00863821396 | 3600.138 | 9 | 9 | 9 |
| 10 | 0.00823378489 | 0.014486902 | 0.01957574246 | 36.016 | 13 | 13 | 13 |
| 11 | 0.00389964657 | 0.00614835 | 0.01033951609 | 3600.162 | 9 | 9 | 10 |
| 12 | 0.00440040702 | 0.011275348 | 0.01546996224 | 3599.689 | 11 | 11 | 11 |
| 13 | 0.00302103544 | 0.008076814 | 0.00960465980 | 3600.162 | 10 | 10 | 10 |
| 14 | 0.00307894232 | 0.006477601 | 0.00843145518 | 3599.136 | 10 | 10 | 10 |
| 15 | 0.00298391967 | 0.006315974 | 0.01049897553 | 3599.303 | 12 | 12 | 12 |
| *16 | 0.00589863523 | 0.009630425 | 0.01171125750 | 77.109 | 16 | 17 | 16 |
| 17 | 0.00652983945 | 0.011867509 | 0.01534673568 | 109.277 | 11 | 11 | 11 |
| 18 | 0.00310449323 | 0.006100245 | 0.00785324091 | 3600.09 | 7 | 7 | 7 |
| 19 | 0.00254854042 | 0.007362366 | 0.01028592882 | 3600.178 | 11 | 11 | 11 |
| 20 | 0.00286151286 | 0.00583884 | 0.00685601804 | 3600.142 | 11 | 11 | 11 |
| **mean** | **0.00434201000** | **0.00844968641** | **0.01178012622** | **2424.5836** | **11.55** | **11.6** | **11.55** |
| **std** | **0.00182383978** | **0.00259938259** | **0.00425577900** | **1656.730756** | **2.258900524** | **2.458069418** | **2.163695669** |

Table 7.11. Properties of Solutions Found in 3600 sec., L=20

| | | L=20 | | | | | |
|---|---|---|---|---|---|---|---|
| # | True Optimal Variance | Fine-Tuned Approximate Variance | Approximate Variance | Solution Time | # Common Assets with Optimal Portfolio | # Common Assets with X Portfolio | # Common Assets with P Portfolio |
| 1 | 0.00719230766 | 0.009347025 | 0.01266359786 | 32.17 | 16 | 16 | 16 |
| 2 | 0.00351026395 | 0.007055018 | 0.01034284817 | 3600.161 | 11 | 11 | 11 |
| 3 | 0.00723946275 | 0.011789582 | 0.01903877416 | 103.309 | 13 | 13 | 12 |
| *4 | 0.00473345031 | 0.00614143 | 0.00842021808 | 1544.283 | 15 | 15 | 15 |
| *5 | 0.00436839129 | 0.007756995 | 0.00905378705 | 3600.02 | 11 | 11 | 11 |
| 6 | 0.00307681032 | 0.005975375 | 0.00754023363 | 3600.188 | 13 | 13 | 13 |
| 7 | 0.00241501021 | 0.006238693 | 0.01043115335 | 3600.155 | 15 | 15 | 15 |
| 8 | 0.00551517902 | 0.011899891 | 0.01634796780 | 1254.543 | 12 | 12 | 12 |
| 9 | 0.00222856731 | 0.006183044 | 0.00809683780 | 3600.11 | 15 | 15 | 15 |
| 10 | 0.00823378489 | 0.013643284 | 0.02056424365 | 84.952 | 13 | 13 | 13 |
| 11 | 0.00389964657 | 0.007220928 | 0.00786110444 | 3600.071 | 11 | 11 | 13 |
| 12 | 0.00440040702 | 0.009736026 | 0.01166896212 | 3600.198 | 12 | 12 | 12 |
| 13 | 0.00302103544 | 0.006829283 | 0.00867333538 | 3600.197 | 10 | 10 | 10 |
| 14 | 0.00307894232 | 0.006753855 | 0.00905088445 | 3600.351 | 10 | 10 | 10 |
| 15 | 0.00298391967 | 0.00648638 | 0.01128018128 | 3600.153 | 14 | 14 | 14 |
| *16 | 0.00589863523 | 0.009791268 | 0.01134338069 | 545.017 | 16 | 16 | 16 |
| 17 | 0.00652983945 | 0.014670874 | 0.02186352821 | 232.026 | 12 | 12 | 12 |
| 18 | 0.00310449323 | 0.006008306 | 0.00651874653 | 3600.131 | 8 | 8 | 8 |
| 19 | 0.00254854042 | 0.004318556 | 0.00634281796 | 3600.147 | 14 | 14 | 14 |
| 20 | 0.00286151286 | 0.00592302 | 0.00720682570 | 3600.178 | 11 | 11 | 11 |
| **mean** | **0.00434201000** | **0.00818844164** | **0.01121547142** | **2529.918** | **12.6** | **12.6** | **12.65** |
| **std** | **0.00182383978** | **0.00286363890** | **0.00465673025** | **1535.77724** | **2.186080366** | **2.186080366** | **2.158825217** |

Table 7.12. Properties of Solutions Found in 3600 sec., L=40

| | | L=40 | | | | | |
|---|---|---|---|---|---|---|---|
| # | True Optimal Variance | Fine-Tuned Approximate Variance | Approximate Variance | Solution Time | # Common Assets with Optimal Portfolio | # Common Assets with X Portfolio | # Common Assets with P Portfolio |
| 1 | 0.00719230766 | 0.007716349 | 0.01088803737 | 79.562 | 17 | 17 | 17 |
| 2 | 0.00351026395 | 0.006418471 | 0.00838603062 | 3600.207 | 12 | 12 | 11 |
| 3 | 0.00723946275 | 0.010900498 | 0.01547209877 | 278.038 | 14 | 14 | 13 |
| *4 | 0.00473345031 | 0.006107176 | 0.00775887065 | 364.402 | 13 | 13 | 13 |
| *5 | 0.00436839129 | 0.005933624 | 0.00718457184 | 3600.174 | 13 | 15 | 13 |
| 6 | 0.00307681032 | 0.005456889 | 0.00735305658 | 3599.883 | 11 | 11 | 11 |
| 7 | 0.00241501021 | 0.004283947 | 0.00678644556 | 3600.24 | 15 | 15 | 15 |
| 8 | 0.00551517902 | 0.010923087 | 0.01412124635 | 3600.204 | 12 | 12 | 12 |
| 9 | 0.00222856731 | 0.004087629 | 0.00535015869 | 3600.111 | 16 | 16 | 16 |
| 10 | 0.00823378489 | 0.012023983 | 0.01594554761 | 252.654 | 12 | 12 | 12 |
| 11 | 0.00389964657 | 0.006683103 | 0.00788794575 | 3600.208 | 12 | 12 | 14 |
| 12 | 0.00440040702 | 0.008734785 | 0.01070124797 | 3600.104 | 14 | 14 | 14 |
| 13 | 0.00302103544 | 0.006212392 | 0.00872916312 | 3600.183 | 11 | 11 | 11 |
| 14 | 0.00307894232 | 0.005236904 | 0.00693379603 | 3600.138 | 11 | 11 | 11 |
| 15 | 0.00298391967 | 0.005775564 | 0.00812314702 | 3600.174 | 9 | 9 | 9 |
| *16 | 0.00589863523 | 0.009786965 | 0.01225029157 | 152.353 | 15 | 15 | 15 |
| 17 | 0.00652983945 | 0.010026408 | 0.01358852145 | 655.618 | 12 | 12 | 12 |
| 18 | 0.00310449323 | 0.006008306 | 0.00651874653 | 3600.128 | 8 | 8 | 8 |
| 19 | 0.00254854042 | 0.004228156 | 0.00566990131 | 3600.177 | 14 | 14 | 14 |
| 20 | 0.00286151286 | 0.00512468 | 0.00676926547 | 3599.505 | 12 | 12 | 12 |
| **mean** | **0.00434201000** | **0.00708344584** | **0.00932090451** | **2609.20315** | **12.65** | **12.75** | **12.65** |
| **std** | **0.00182383978** | **0.00245141205** | **0.00330144534** | **1556.392975** | **2.207046133** | **2.268201235** | **2.230765741** |

Table 7.13. Properties of Solutions Found in 3600 sec., L=200

| | | L=200 | | | | | |
|---|---|---|---|---|---|---|---|
| # | True Optimal Variance | Fine-Tuned Approximate Variance | Approximate Variance | Solution Time | # Common Assets with Optimal Portfolio | # Common Assets with X Portfolio | # Common Assets with P Portfolio |
| 1 | 0.00719230766 | 0.007651999 | 0.00881466137 | 12 | 17 | 17 | 17 |
| 2 | 0.00351026395 | 0.005153335 | 0.00619004062 | 3599.249 | 11 | 11 | 11 |
| 3 | 0.00723946275 | 0.008650694 | 0.00936919586 | 639.162 | 15 | 15 | 15 |
| *4 | 0.00473345031 | 0.00530412 | 0.00592359953 | 599.478 | 16 | 16 | 16 |
| *5 | 0.00436839129 | 0.005215903 | 0.00608701798 | 3600.268 | 14 | 12 | 14 |
| 6 | 0.00307681032 | 0.003594529 | 0.00399378235 | 3600.342 | 17 | 17 | 17 |
| 7 | 0.00241501021 | 0.003399823 | 0.00440256207 | 2383.807 | 17 | 17 | 17 |
| 8 | 0.00551517902 | 0.007111647 | 0.00933119843 | 3600.354 | 16 | 16 | 16 |
| 9 | 0.00222856731 | 0.003078273 | 0.00396824825 | 3600.278 | 19 | 19 | 19 |
| 10 | 0.00823378489 | 0.009081168 | 0.00953128081 | 92.47 | 18 | 18 | 18 |
| 11 | 0.00389964657 | 0.00473171 | 0.00516094386 | 3600.181 | 12 | 12 | 15 |
| 12 | 0.00440040702 | 0.005244683 | 0.00627769551 | 3600.205 | 18 | 18 | 18 |
| 13 | 0.00302103544 | 0.005360499 | 0.00562312695 | 3600.269 | 10 | 10 | 10 |
| 14 | 0.00307894232 | 0.003877039 | 0.00467730831 | 3599.874 | 14 | 14 | 14 |
| 15 | 0.00298391967 | 0.003638664 | 0.00422191431 | 3600.207 | 15 | 15 | 15 |
| *16 | 0.00589863523 | 0.006048769 | 0.00701080029 | 104.033 | 19 | 20 | 19 |
| 17 | 0.00652983945 | 0.006632018 | 0.00808079655 | 229.89 | 18 | 18 | 18 |
| 18 | 0.00310449323 | 0.00352084 | 0.00403382107 | 3600.226 | 17 | 17 | 19 |
| 19 | 0.00254854042 | 0.003698788 | 0.00427150241 | 3600.277 | 16 | 16 | 16 |
| 20 | 0.00286151286 | 0.003447206 | 0.00443934154 | 3600.265 | 17 | 17 | 17 |
| **mean** | **0.00434201000** | **0.00522208532** | **0.00607044190** | **2543.14175** | **15.8** | **15.75** | **16.05** |
| **std** | **0.00182383978** | **0.00181096809** | **0.00197311612** | **1550.310171** | **2.525657809** | **2.712058841** | **2.459674775** |

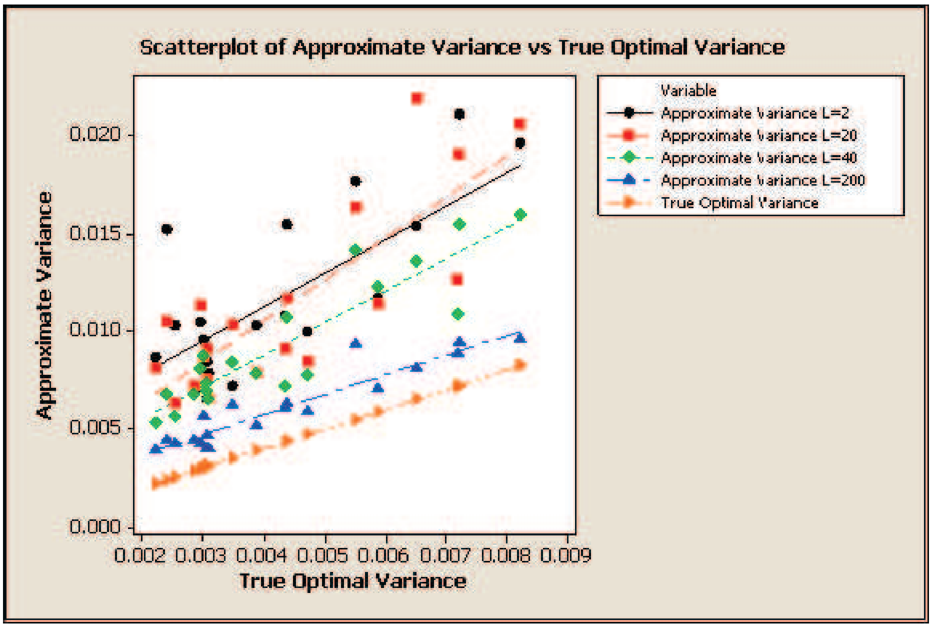


Figure 7.13. ResLim=3600 sec., Effect of L on Quality of Approximate Portfolio Variance

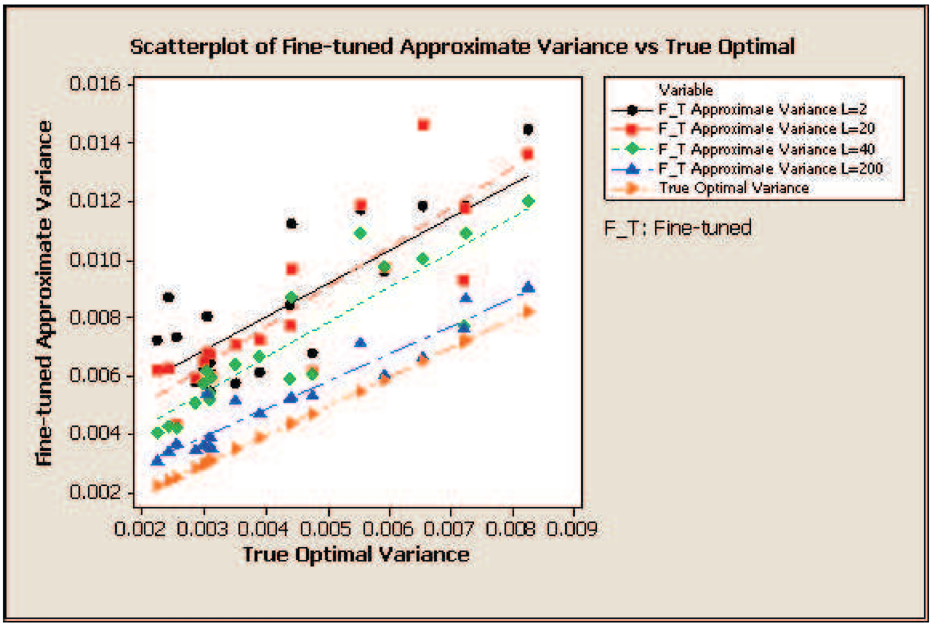


Figure 7.14. ResLim=3600 sec., Effect of L on Quality of Fine-tuned Approximate Portfolio Variance

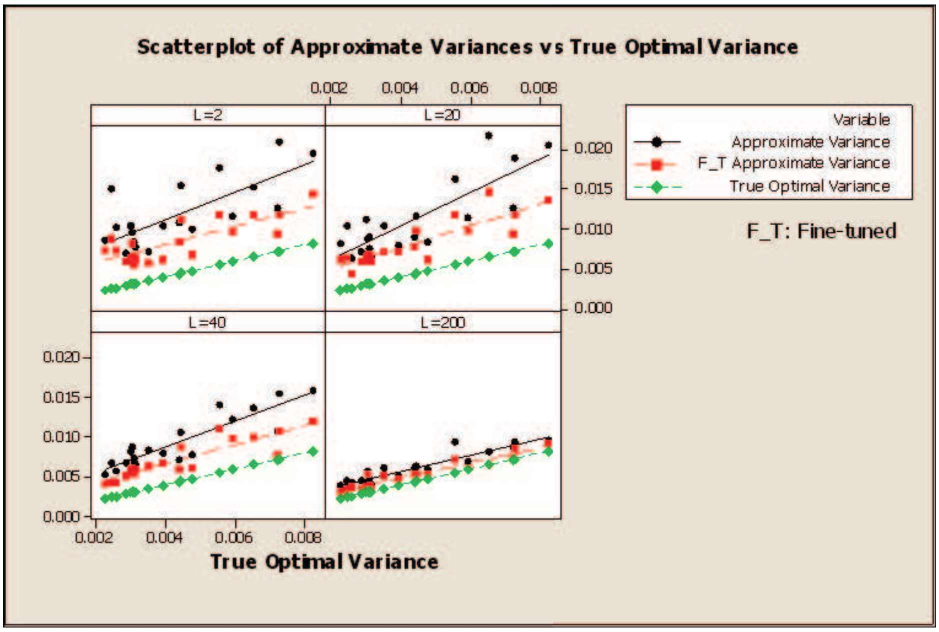


Figure 7.15. ResLim=3600 sec., Effect of L on Quality of Approximate and Fine-tuned Approximate Portfolio Variance

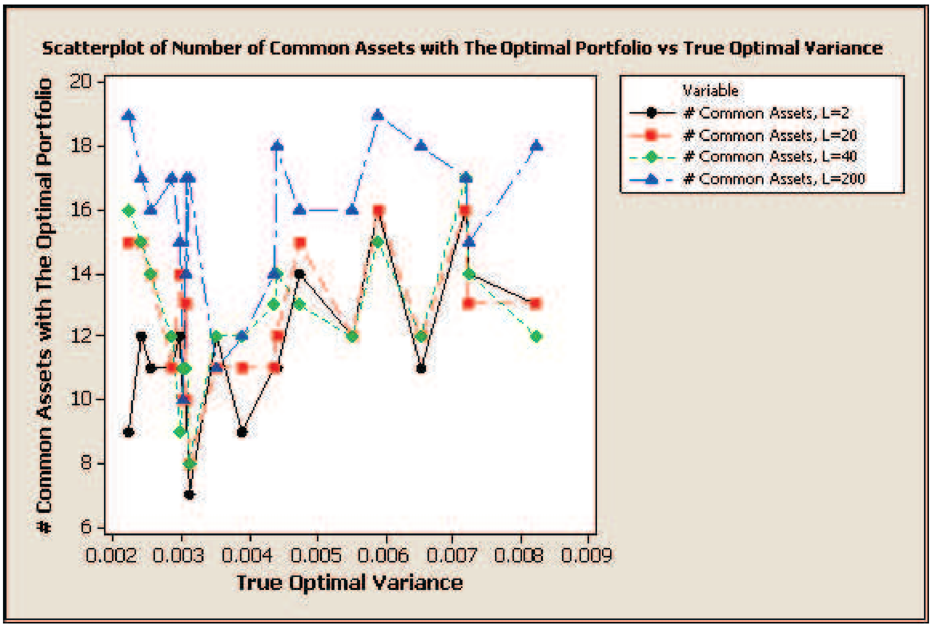


Figure 7.16. ResLim=3600 sec., Effect of L on The Number of Assets in Common with The Optimal Portfolio

# 8. CONCLUSIONS AND FUTURE WORK

In this study, a transformation for the Markowitz quadratic portfolio optimization is investigated. It is seen that MINOS 5.51 and CONOPT 3.0 have performed much more effectively on the transformed quadratic problems than they have performed on the classic quadratic problems. On the other hand, the performance of CPLEX 9.0 on these problems is completely the opposite: average CPLEX 9.0 solution times for the classic model, are the least among average solution times of all solver and model type combinations. CPLEX solves the classic model much faster than MINOS and CONOPT solve the transformed quadratic problem; and it solves the transformed quadratic model much slower than MINOS and CONOPT. On the other hand, it is observed that as the required minimum portfolio return, $R$, is decreased computation time for the classic model solved through MINOS and CONOPT increases, while it stays constant for the transformed problem.

After the transformation is applied to the classic quadratic model, the objective function of the transformed quadratic problem becomes separable. Therefore, it becomes open for piecewise linear approximation. An approximation scheme has been devised using the results of Lemma 1 and Lemma 2. According to the results of the experiments conducted, it is seen that piecewise linear approximation is not an efficient way for solving quadratic programming problems in terms of solution time. But it provides some benefits for MIQP portfolio optimization models.

Efficient MIQP solvers are rare, and many practitioners do not have access to these solvers, and many use Microsoft Office EXCEL. With the aid of the transformation and the approximation scheme defined in Chapter 4 and Chapter 6, respectively, the cardinality constrained portfolio selection problem, which is an MIQP problem, becomes solvable with MIP solvers (in a reasonable amount of time of 30 minutes). Efficient MIP solvers, in contrast to MIQP solvers, are widely available. Therefore, near optimal portfolios can be easily obtained by solving “Approximating MIP Model” with those efficient, available MIP solvers. Moreover, computational tests have indicated that increasing L from 40 to 200 significantly improves the quality of solutions obtained in 30 minutes (this is somewhat counterintuitive since increasing L also increases the required computational effort).

Future work can continue in two directions. One direction is to make "Approximating MIP Models" be solved more effectively using different cuts investigated in the integer programming literature. Therefore, instead of solving the problem with MIP solvers, a branch-and-cut algorithm can be developed and used to solve "Approximating MIP Models". Just to give some insight, during this study MIR (mixed integer rounding) cuts have been added to the constraint set of "Approximating MIP Model" (i.e. Model 9) and it has been seen that although relaxed solution did not increased after adding a set of valid MIR cuts, for some problems solution times have decreased. Since this topic was out of scope of the thesis, detailed study about Model 9 with MIR cuts was not done. On the other hand, using cuts to solve "Approximating MIP Model" seems to be a promising direction for future work. Furthermore, developing a branch-and-cut algorithm also requires a new function definition that provides lower bound for the original objective function of "The Transformed QP Model". Since current piecewise linear objective function does not provide lower bound, new approximation schemes can be used. For instance, instead of using $\lambda$-form approximation, one can approximate each term of the separable objective function with tangents drawn from the grid points. Finally, second direction would be comparing the quality of solutions found by solving "Approximating MIP Model" with commercial solvers and of solutions found by some heuristics suggested for this problem in the literature. Furthermore, it should be noted that throughout the thesis randomly generated normal data are used. On the other hand, it would be interesting to use real data, which is known to be non-normal, and to compare the classic Markowitz model and the approximating model with piecewise linear objective function in terms of their optimal portfolio solutions. Since real data are known to be non-normal, the results of Markowitz model are no longer the optimal according to utility maximization principle.

# REFERENCES


1. H. M. Markowitz, *Portfolio selection: efficient diversification of investments*, John Wiley & Sons, New York, 1959.

2. Y. Crama, M. Schyns, "Simulated annealing for complex portfolio selection problems", *European Journal of Operational Research*, Vol. 150, No. 3, pp. 546-571, November 2003.

3. E.J. Elton, M.J. Gruber, *Modern Portfolio Theory and Investment Analysis, fourth ed.*, John Wiley & Sons, New York, 1991.

4. H. Markowitz, "Portfolio selection", *The Journal of Finance*, Vol. 7, No. 1, pp. 77-91, March 1952.

5. M. Blume, I. Friend, "The asset structure of individual portfolios and some implications for utility functions", *The Journal of Finance*, Vol. 30, No. 2, pp. 585-603, May 1975.

6. R. Cohn, W. Lewellen, R. Lease, G. Schlarbaum, "Individual investor risk aversion and investment portfolio composition", *The Journal of Finance*, Vol. 30, No. 2, pp. 605-620, May 1975.

7. J. Mossin, *Theory of Financial Markets*, Prentice Hall, New Jersey, 1973.

8. W. Ogryczak, A. Ruszczynski, "From stochastic dominance to mean risk models: semideviations as risk measures", *European Journal of Operations Research*, Vol. 116, No. 1, pp. 33-50, July 1999.

9. H. Konno, KI. Suzuki, "A mean-variance-skewness portfolio optimization model", *Journal of the Operations Research Society of Japan*, Vol. 38, No. 2, pp. 173-187, June 1995.

10. W. F. Sharp, "A simplified model for portfolio analysis", *Management Science*, Vol. 9, No. 2, pp. 277-293, January 1963.

11. H. Markowitz and A. Perold, *In Sparse Matrices and Their Uses, Duff ed.*, Academic Press, New York, 1981.

12. A.F. Perold, "Large-scale portfolio optimization", *Management Science*, Vol. 30, No. 10, pp. 1143-1160, October 1984.

13. H. Konno, H. Waki, A. Yuuki, "Portfolio optimization under lower partial risk measures", *Asia Pacific Financial Markets*, Vol. 9, No. 2, pp. 127-140, December 2002.

14. H. Konno, H. Yamazaki, "Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market", *Management Science*, Vol. 37, No. 5, pp. 519-531, May 1991.

15. Y. Simaan, "Estimation risk in portfolio selection: the mean variance model versus the mean absolute deviation model", *Management Science*, Vol. 43, No. 10, pp. 1437-1446, October 1997.

16. H. Konno, K. Suzuki, "A fast algorithm for solving large scale mean-variance models by compact factorization of covariance matrices", *Journal of the Operations Research Society of Japan*, Vol. 35, No. 1, pp. 93-107, March 1992.

17. R. J. Vanderbei, T. J. Carpenter, "Symmetric indefinite systems for interior point methods", *Mathematical Programming*, Vol. 58, No. 1, pp. 1-32, January 1993.

18. W. F. Sharpe, "A linear programming approximation for the general portfolio analysis problem", *The Journal of Financial and Quantitative Analysis*, Vol. 6, No. 5, pp. 1263-1275, December 1971.

19. B. K. Stone, "A linear programming formulation of the general portfolio selection problem", *The Journal of Financial and Quantitative Analysis*, Vol. 8, No. 4, pp.

621-636, September 1973.

20. D. Bienstock, "Computational study of a family of mixed-integer quadratic programming problems", *Mathematical Programming*, Vol. 74, No. 2, pp. 121-140, August 1996.

21. T.J. Chang, N. Meade, J.E. Beasley, Y. M. Sharaiha, "Heuristics for cardinality constrained portfolio optimisation", *Computers and Operations Research*, Vol. 27, No. 13, pp. 1271-1302, November 2000.

22. Mokhtar S. Bazaraa, Hanif D. Sherali, C. M. Shetty, *Nonlinear Programming Theory and Algorithms, second ed.*, John Wiley & Sons, New York, 1993.

23. GAMS Development Corporation., http://www.gams.com, 2005.

24. The Mathworks, Inc., http://www.mathworks.com/products/matlab, 2005.

25. G.H. Golub, Charles F. Van Loan, *Matrix computations, third ed.*, Johns Hopkins University Press, Baltimore, 1996.

26. G. Strang, *Linear Algebra and Its Applications, third ed.*, Harcourt, Brace, Jovanovich, Publishers, San Diego, c1988.