Univerzita Karlova v Praze
Fakulta sociálních věd

Institut ekonomických studií

BAKALÁŘSKÁ PRÁCE

Výběr portfolia: řešení pomocí shlukovacích algoritmů

Vypracoval: Mgr. Petr Jenček
Vedoucí: PhDr. Jozef Baruník
Akademický rok: 2009/2010

I would like to thank to PhDr. Jozef Baruník for his support and valuable advices provided to me during my work on this bachelor thesis.

I declare that I wrote this thesis myself and used only the literature listed in References.

In Prague on 6th December 2009 Petr Jenček

UNIVERSITAS CAROLINA PRAGENSIS
založena 1348

Univerzita Karlova v Praze
Fakulta sociálních věd

Institut ekonomických studií

Opletalova 26
110 00 Praha 1
TEL: 222 112 330,305
TEL/FAX:
E-mail:
ies@mbox.fsv.cuni.cz

http://ies.fsv.cuni.cz

Akademický rok 2008/2009

TEZE BAKALÁŘSKÉ PRÁCE

Student:	Petr Jenček
Obor:	Ekonomie
Konzultant:	PhDr. Jozef Baruník

Garant studijního programu Vám dle zákona č. 111/1998 Sb. o vysokých školách a Studijního a zkušebního řádu UK v Praze určuje následující bakalářskou práci

Předpokládaný název BP:

Výběr portfolia: řešení pomocí shlukovacích algoritmů

Charakteristika tématu, současný stav poznání, případné zvláštní metody zpracování tématu:

Ceny akcií obchodovatelných na burze jsou závislé na jistých ukazatelích, které mohou, ale nemusejí být obecně známy, což způsobuje výraznou korelaci vybraných párů akcií. Při výběru akcií v portfoliu investora je zapotřebí vybírat do málo korelovaných aktiv. Moje bakalářská práce si klade za cíl představit alternativní metodu pro výběr portfolia za použití shlukovacích algoritmů a vzájemných korelací cen jednotlivých aktiv obsažených v portfoliu. Tuto metodu dále srovná s ekonometrickým přístupem přístupem pomocí Monte Carlo simulací provedených nad historickými daty vybraných akcií.

Struktura BP:

1. Úvod (jaké jsou doposud dostupné metody, definice dat, která použiji pro nalezení závislostí a otestování predikce)
2. Představení vlastní metody výběru portfolia
3. Popis a provedení Monte Carlo simulací
4. Vyhodnocení jednotlivých simulací
5. Závěr

Seznam základních pramenů a odborné literatury:

Ait-Sahalia Y. and Brandt M. W. (2001): Variable Selection for Portfolio Choice, The Journal of Finance, Vol. 56, No. 4, pp. 1297-1351
Markowitz H. (1952): Portfolio Selection, The Journal of Finance, Vol. 7, No. 1, pp. 77-91

Datum zadání:	Červen 2008
Termín odevzdání:	červen 2009

Podpisy konzultanta a studenta:

V Praze dne

Abstract

Prices of assets (stocks, commodities etc.) are dependent on many economic factors. These factors may be explicitly known but most of them are hidden. This dependency causes that price of an asset influences prices of another assets which makes it quite complicated to select optimal portfolio. Portfolio management is usually based on various mathematic models in conjunction with Value-at-Risk model. The aim of this thesis is to provide an alternative approach for optimal portfolio selection with mutual assets' prices correlation consideration using cluster analysis.

Title: Portfolio Selection: Clustering Algorithm Approach

Author: Mgr. Petr Jenček

Author's e-mail: jencek@atlas.cz

Supervisor: PhDr. Jozef Baruník

Academic year: 2009/2010

Keywords: portfolio selection, clustering algorithm, investment

Abstrakt

Ceny aktiv (akcie, komodity atd.) jsou závislé na mnoha ekonomických faktorech. Tyto faktory mohoy být explicitně známy, ale většina z nich zůstává ekonomům skryta. Tyto závislosti způsobují, že cena jednoho aktiva ovlivnňuje ceny dalších aktiv, což velmi ztěžuje výběr optimálního portfolia. Metody portfolio managementu jsou většinou založeny na různých matematických modelech v kombinaci s modelem Value-at-Risk. Cílem této práce je poskytnout alternativní postup pro výběr optimálního portfolia vzhledem k vzájemným korelacím cen jednotlivých aktiv v portfoliu.

Název práce: Výběr portfolia: řešení pomocí shlukovacího algoritmu

Autor: Mgr. Petr Jenček

E-mail autora: jencek@atlas.cz

Vedoucí: PhDr. Jozef Baruník

Akademický rok: 2009/2010

Klíčová slova: výběr portfolia, shlukovací algoritmy, investice

Contents

1 Introduction

Stock exchange traders and investors (hereafter only “investor”) are trying to find out ways how to maximize return of investment and minimize risk of their investment. Investors use several techniques for minimizing investment risk including regular investments and diversifying investments among several assets. This thesis concerns the issue of diversifying investment risk by selecting optimal quantity of several assets to investor’s portfolio.

Portfolio selection with purpose of investment risk diversification is subject of many research papers. Pioneer paper on this topic was published by Markowitz (1952). This thesis is based mainly on papers published by Brandt (1999) and Ait-Sahalia in cooperation with Brandt (2001). Most of the papers about portfolio selection use mathematical methods to determine optimal portfolio composition. Usually it is solved as optimization problem by determining the composition of portfolio which maximizes the estimated return. This thesis attempts to provide an alternative way of determining the composition of portfolio of assets with significant mutual correlation while maximizing minimum level of expected return.

The aim of chapter 2 is to formalize the problem of selecting optimal portfolio from a given set of assets. Chapter 3 provides a brief explanation of currently used portfolio selection methods. Whole process of portfolio selection according to methodology presented by this thesis will be described in chapter 4. Experiments with purpose of comparing our methodology with methodology presented by Ait-Sahalia and Brandt (2001) are described in chapter 5. In the final chapter you may find user’s manual for the application developed for the purpose of selecting the portfolio according to the process described in this thesis.

2 Problem Formalism

Until now we've used terms like securities, portfolio etc. without any precise definition, just by understanding its meaning intuitively. This will be changed in this chapter by implementing several precise definitions.

Let's assume that we wish to invest wealth W_t for one period and maximize the objective function $u(W_{t+1})$ with absolute risk aversion γ as defined by Formula 1 and . This formula basically gives minimum expected return on a certain level of risk adversity which may be considered as confidence. This means that in this thesis minimal return on a certain level of confidence will be maximized. This definition is used by Brandt and Ait-Sahalia in (2001).

$$Eu(W_{t+1}) = EW_{t+1} - \frac{\gamma}{2}\mathrm{var}(W_{t+1}^2)$$

$$\gamma = \frac{\frac{\partial^2 v(W)}{\partial W^2}}{\frac{\partial v(W)}{\partial W}}$$

Formula 1

The first step of determining the optimal portfolio composition is selecting N ($N > 1$) assets. Selection may be performed according to Markowitz's (1952) process described briefly in chapter 3. Even though Markowitz's (1952) process is recommended for this initial step, the methodology proposed by this thesis doesn't require it and any other process of selecting N assets for future investments may be used.

Denote ER_i ($i = 1 .. N$) their expected returns based on historical data (history should be longer than our investment horizon) and σ_{ij} ($i = 1 .. N, j = 1 .. N$) their mutual covariance. We'll refer to matrix of (σ_{ij}) at time t as Σ_t. Notice that σ_{ii} ($i = 1 .. N$) is variance of i^{th} security (based on the same historical data).

The result of the calculation process should be a vector $X = X_1, X_2, \ldots X_n$ where X_i denotes fraction of W_t which should be invested into i^{th} security. This means that $\sum_i X_i = 1$ (investor has to invest exactly all of his wealth at time t) and amount of money invested into i^{th} security is equal to $X_i \times W_t$. Transaction costs connected with buying or selling

securities will be disregarded in this thesis. This simplification will be commented in chapter 4.5. Because of this simplification the resulting vector X should not depend on W_t. We'll denote vector X as "portfolio".

3 Literature on Portfolio Selection

A lot of work studying optimal portfolio selection has been done. Several important studies about this topic are presented further in this chapter.

Markowitz may be considered as a founder of theory of portfolio selection. Markowitz (1952) suggests diversifying portfolio among more securities within "optimal set of securities". Markowitz's optimal set may be demonstrated as a set of all securities for which no other security provides the same expected return with lower volatility and no other security provides the same volatility with higher expected return. The optimal set may be observed in Figure 1[1] (marked as "efficient combinations") where axis marked as "v" stands for variance of return (volatility) and axis marked as "e" means expected return.

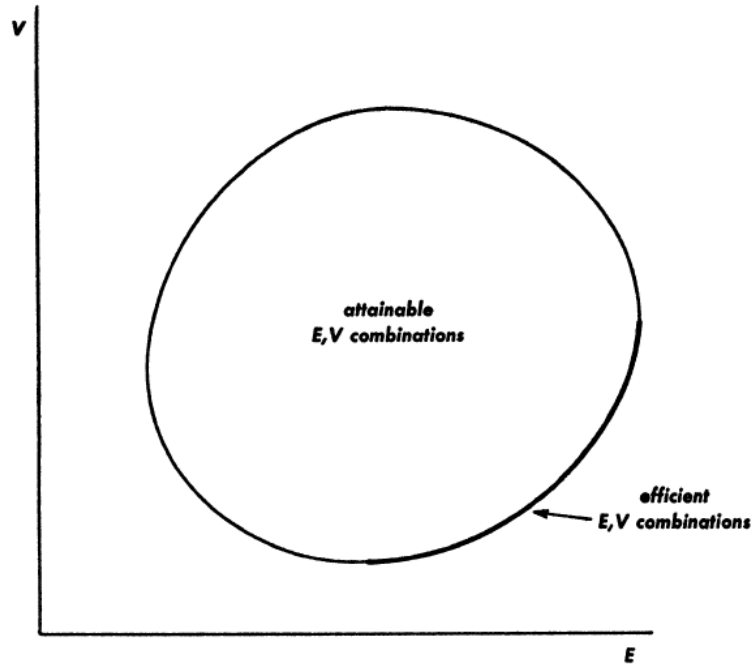


Figure 1

He also suggests investing into securities within different lines of business since companies within one line of business are too inter-correlated. However he doesn't

[1] Source: Markowitz (1952)

provide an exact process of determining the securities nor the quantity of money in which a trader should invest to them.

More mathematical and formal theoretical framework for portfolio selection problem is provided by Treynor and Black (1973). They have also presented calculation of optimal portfolio selection using statistical method. However, on contrary to approach presented in this thesis, they fix the expected return of the portfolio and then minimize the variance of return of the portfolio using the method of Lagrangian multiplier. They arrive to optimal portfolio composition defined by Formula 2.

$$X_i = \frac{ER_i}{ER}\frac{\text{var}(R)}{\text{var}(R_i)}, i = 1..N$$

Formula 2

In Formula 2 the following notation is used:

- X_i fraction of i^{th} security in the resulting portfolio
- R_i return of i^{th} security (hereafter considered as random variable)
- R return of whole portfolio (hereafter considered as random variable)
- N number of securities within the portfolio

Brandt (1999) considers model of investor with utility function $u(W)$ who allocates a fraction α of his wealth W for 1 period into portfolio which yields uncertain return $\tilde{R}^e_{t+1}$. The rest of his wealth is invested into a riskless security with return of R^f.

$$\hat{\alpha}_T = \{\alpha : \frac{1}{T}\sum_{t=1}^{T} u'(R^f + \alpha R^e_{t+1})R^e_{t+1} = 0\}$$

Formula 3

In Formula 3 and Formula 4 the following notation is used:

- $\hat{\alpha}_T$ set of possible fractions of wealth to be invested into portfolio with uncertain return
- T number of time units (usually days) the investment is planned for
- $u`$ derivative of investor's utility function
- Σ_t^{-1} inverse matrix to securities covariance matrix counted at time t (time of investment decision)
- γ investor's risk aversion
- W_t investor's wealth at time t

- ER_{t+1} vector of securities‘ expected returns
- ι vector of ones of the dimension equal to number of securities

Formula 3 uses law of iterated expectations to estimate a set of all investment actions where the investor expects zero marginal utility which is basically a set of optimal investment actions α at time T.

Brandt and Ait-Sahalia in (2001) generalized this framework for more securities. They arrived to Formula 4 using generalized moment method provided by Hansen (1982). The result of this formula is estimate of optimal weights of the assets within the investor's portfolio.

$$\hat{\alpha}_T = \Sigma_t^{-1}\iota \frac{\gamma W_t - \iota' \Sigma_t^{-1} ER_{t+1}}{\gamma W_t \iota' \Sigma_t^{-1} \iota} + \frac{\Sigma_t^{-1} ER_{t+1}}{\gamma W_t}$$

Formula 4

Brandt's approach defined by Formula 4 will be used as a benchmark and will be compared to our proposed approach for selection of portfolio.

Goldfarb and Iyengar (2002) propose a way of portfolio choice for investor with high risk aversion. Basically they minimize value at risk (hereafter VaR – explained in details in chapter 5.2) at first and then maximize expected returns. The optimal portfolio is then chosen using econometric calculations. The process of portfolio choice described in this thesis provides more flexibility than Goldfarb's and Iyengar's (2002) approach, because risk aversion may be chosen according to investor's preferences.

4 Method Description

In the following chapter we'll provide full description of methodology of portfolio selection method proposed by this thesis. Optimal portfolio selection for simple data containing 2 assets will be derived using derivative of the investor's utility function. This framework will be generalized to any number of securities by transforming the original problem into a set of equations which may be solved using Lagrange multipliers. Further in this chapter we'll present our algorithmic approach of portfolio selection using a modification of agglomerative clustering algorithm. Properties of this algorithm, its results and implementation of transaction costs will be commented in the final phase of this chapter.

4.1 Case of 2 Securities

Let's assume model of 2 securities, $N = 2$. Since one of the 2 securities may be a risk free security this assumption is general enough.

$$EW_{t+1} = (1-\alpha)ER_1 + \alpha ER_2 + W_t$$

Formula 5

Formula 5 characterizes expected wealth of the investor at the end of the investment period.

$$Eu(W_{t+1}) = EW_{t+1} - \frac{\gamma}{2}\operatorname{var}(W_{t+1}^2) = (1-\alpha)ER_1 + \alpha ER_2 + W_t - \frac{\gamma}{2}\left((1-\alpha)^2\sigma_{11} + \alpha^2\sigma_{22} + 2(1-\alpha)\alpha\sigma_{12}\right)$$

Formula 6

In Formula 5, Formula 6, Formula 7 and Formula 8 the following notation is used:

- EW_{t+1} expected investor's wealth at time t+1 (usually the next day)
- $Eu(W_{t+1})$ expected value of investor's utility function of wealth at time t+1 (usually the next day)
- γ investor's risk aversion
- ER_i expected return of i^{th} security
- α fraction of wealth invested into 2nd security
- δ_{ii} variance of return of i^{th} security
- δ_{12} covariance of returns of 1st and 2nd security or (from symmetry of

covariances) covariance of returns of 2nd and 1st security

Formula 6 defines the objective function of the investor. This function is to be maximized with regards to parameter $0 \le \alpha \le 1$. Maximizing this function using derivation brings us to Formula 7.

$$\alpha = \begin{cases} 0 \quad for \quad \dfrac{-ER_1 + ER_2 + \gamma\sigma_{11} - \gamma\sigma_{12}}{\gamma(\sigma_{11} + \sigma_{22} - 2\sigma_{12})} < 0 \\ \dfrac{-ER_1 + ER_2 + \gamma\sigma_{11} - \gamma\sigma_{12}}{\gamma(\sigma_{11} + \sigma_{22} - 2\sigma_{12})} \quad for \quad 0 < \dfrac{-ER_1 + ER_2 + \gamma\sigma_{11} - \gamma\sigma_{12}}{\gamma(\sigma_{11} + \sigma_{22} - 2\sigma_{12})} < 1 \\ 1 \quad for \quad 1 < \dfrac{-ER_1 + ER_2 + \gamma\sigma_{11} - \gamma\sigma_{12}}{\gamma(\sigma_{11} + \sigma_{22} - 2\sigma_{12})} \end{cases}$$

Formula 7

Using Formula 7 we arrive to the optimal portfolio selection according to Brandt and Ait-Sahalia's (2001) objective function $u(W_{t+1})$ (Formula 8).

$$X_1 = \min\left(\max\left(1 - \frac{-ER_1 + ER_2 + \gamma\sigma_{11} - \gamma\sigma_{12}}{\gamma(\sigma_{11} + \sigma_{22} - \sigma_{12})}, 0 \right), 1 \right)$$

$$X_2 = \min\left(\max\left(\frac{-ER_1 + ER_2 + \gamma\sigma_{11} - \gamma\sigma_{12}}{\gamma(\sigma_{11} + \sigma_{22} - \sigma_{12})}, 0 \right), 1 \right)$$

Formula 8

Notice that the quantities to be invested into the 2 securities do not depend on initial wealth W_t.

4.2 Generalization to N Securities (mathematical version)

We could generalize this context to generic N in the way which shows Formula 9.

$$Eu(W_{t+1}) = \sum_{i=1}^{N} X_i ER_i + W_t - \frac{\gamma}{2}\left(\sum_{i=1}^{N} \sum_{j=1}^{N} X_i X_j \sigma_{ij} \right)$$

Formula 9

Formula 9 provides us with utility function of investor's wealth at the end of the next period. Since the investor is required to invest all of his wealth condition of $\sum_{j=1}^{N} X_i = 1$ has to be met. The condition of non-negative investment (the investor can't borrow money)

can be described by $0 \leq X_i \leq 1$ inequation.

Solving this problem using Lagrange multipliers would give us a vector of optimal portfolio diversification. This approach would require too many calculations and therefore we'll build an algorithmic generalization.

4.3 Generalization to N Securities (algorithmic version)

In this chapter generalization of framework provided in chapter 4.1 using an algorithmic approach will be given. Since the algorithm uses a tree we'll call this algorithm Investment Tree Algorithm (hereafter ITA).

ITA is inspired by agglomerative clustering algorithm. This algorithm was published by Ward (1963) and it has many applications in data analysis (e.g. US Air Force's Comprehensive Occupational Data Analysis Programs - CODAP), text retrieval systems (Jenček et al. (2009)) and many other fields. Agglomerative clustering algorithm will be described only briefly in this thesis, better explanation may be found in different sources (e.g. Ward (1963)).

Agglomerative clustering algorithm takes desired number of clusters and matrix of mutual distances of multiple object as input and produces a set of clusters containing the objects (each cluster contains objects with low mutual distance). Let's demonstrate work of this algorithm on a simple example. In order to keep the demonstration simple no precise mutual distances will be defined for the objects. No formal definition of the algorithm is given here because of the same reason (formal definition of its modification for portfolio choice will be given in further in this chapter).

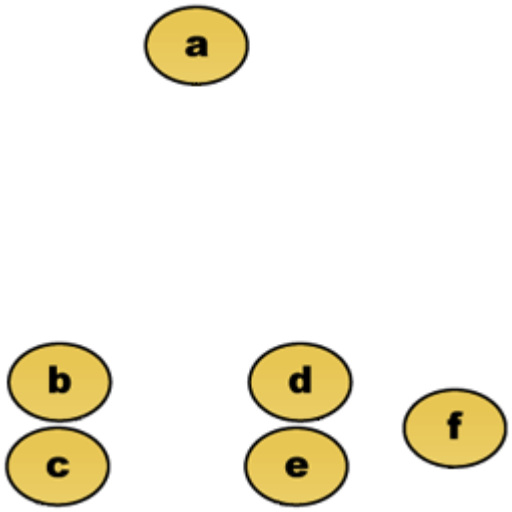


Figure 2

Suppose that agglomerative clustering algorithm runs on a set of objects displayed in Figure 2[2] with desired finishing number of clusters equal to 1 where mutual distance are defined as Euclidean distances of the centers of the objects according to their placement in Figure 2. In the picture it is obvious that $d(b,c) \approx d(d,e) < d(c,e)$ etc. This algorithm considers each object to be one cluster at the beginning. During each iteration it takes 2 nearest clusters and creates 1 cluster from these 2 clusters. As displayed in Figure 3[3] in the first iteration 2 iterations it creates clusters from objects (b,c) and (d,e), because their mutual distance is the smallest among all mutual distances. At the end the algorithm outputs tree displayed in Figure 3.

[2] Source: Wikipedia (http://en.wikipedia.org/wiki/Cluster_analysis)

[3] Source: Wikipedia (http://en.wikipedia.org/wiki/Cluster_analysis)

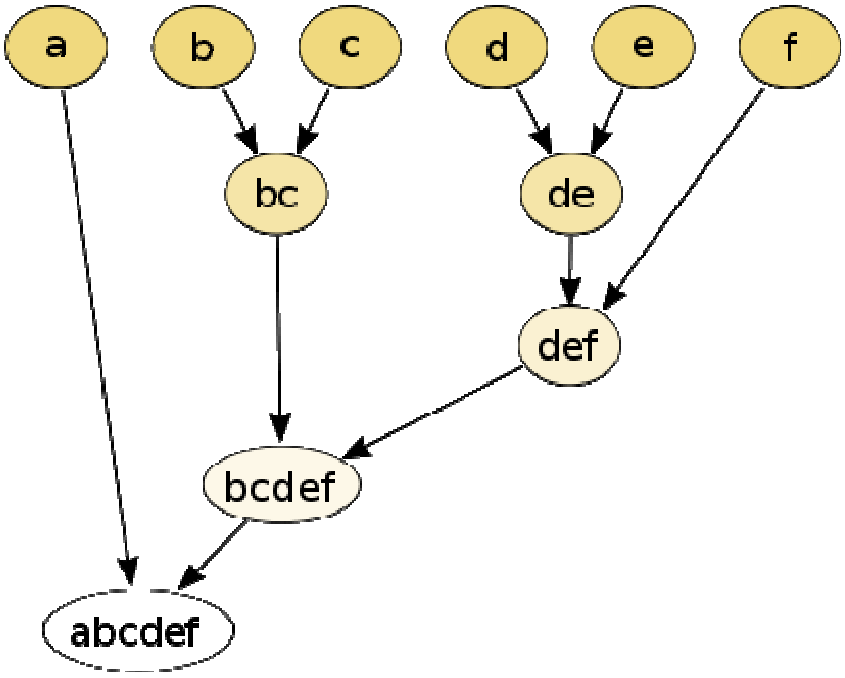


Figure 3

If we denote random variable expressing expected return of a group of i^{th} and j^{th} security with quantities X_i and X_j R_{ij}, then the following equalities hold:

$$\text{var}(R_{ij}) = X_i^2\sigma_{ii} + X_j^2\sigma_{jj} + 2X_iX_j\sigma_{ij}$$
$$ER_{ij} = X_iER_i + X_jER_j$$

We may consider this set of 2 securities as a new security (let's call it "virtual" security) which has correlation with other securities defined by the following formula:

$$cor(R_{ij}, R_k) = \min(cor(R_i, R_k), cor(R_j, R_k))$$

Let's denote (non-empty) set of all securities *S*. Then the process of creating virtual security from 2 securities with the strongest correlation could be generalized by the following algorithm based on agglomerative hierarchical clustering:

```
(1)      while count(S) > 1 do
(2)        let s1 and s2 are securities with greatest correlation;
(3)        create new virtual security sv from s1 and s2 with
   quantities as defined in Formula 8;
(4)        set sv as parent of s1 and s2;
(5)        remove s1 and s2 from S;
(6)        add sv to S;
(7)      end while;
```

At the end of this algorithm there is only 1 security in *S* (virtual security if more than 1

security was in *S* at the beginning of the algorithm run).

Let's show the result of this algorithm on the following data:

$$ER_1 = 0.1$$
$$ER_2 = 0.05$$
$$ER_3 = 0.15$$
$$cor(R) = \begin{pmatrix} 1 & 0.5 & 0.3 \\ 0.5 & 1 & 0.7 \\ 0.3 & 0.7 & 1 \end{pmatrix}$$

It will construct a tree similar to the one in Figure 4. In Figure 4 S1, S2 and S3 denote securities and VS… virtual securities. Let's call this tree "securities tree".

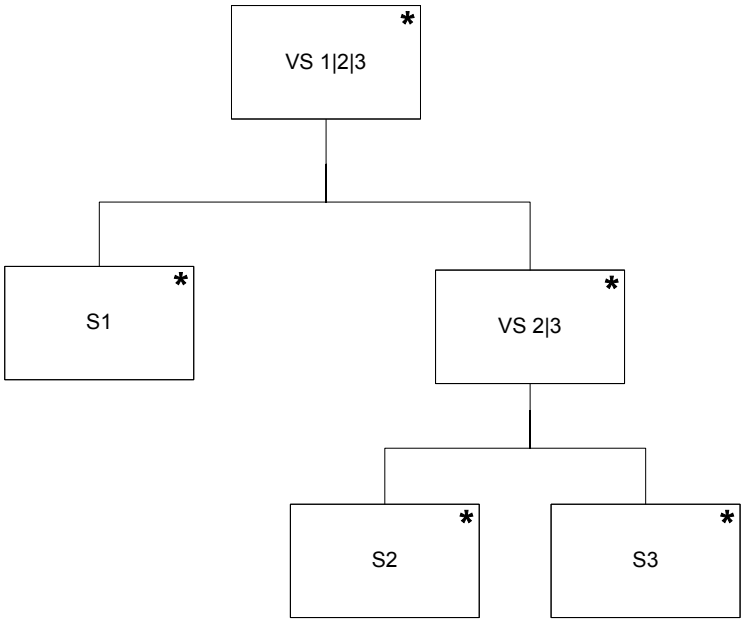


Figure 4

The output of the algorithm will be the following:

$$VS2|3\begin{cases} X_2 = 0.25 \\ X_3 = 0.75 \end{cases}$$
$$VS1|2|3\begin{cases} X_1 = 0.4445 \\ X_2 = 0.1338 \\ X_3 = 0.4167 \end{cases}$$

Final portfolio is contents of the only security in *S* (root node of tree in Figure 4). Notice that strong correlation of S2 and S3 caused lower fraction of these 2 securities in the output portfolio.

4.4 Properties of Investments Tree and ITA

4.4.1 Finiteness

Theorem: This algorithm always stops.

Proof: During each iteration of the main cycle (lines (1) – (7)) the number of items within *S* is decreased by 1.

4.4.2 Sum of the Portfolio Fractions

If we put quantities of securities in leaf nodes (nodes without any child) 1 then the following theorem is valid:

Theorem: Sum of quantities in all security nodes is 1.

Proof: When creating virtual security the quantities are defined by Formula 8. We can easily calculate that the sum of these quantities is equal to 1 for all values of expected return. Since quantity in leaves is (by definition) equal to 1 the proof is completed.

4.5 Transaction Costs

In this thesis we suppose that transaction costs are low in comparison with amount of money the trader trades with. If the condition about low transaction costs doesn't hold the output of ITA may be adjusted in the following way. If we denote costs of joint transaction (buying and selling of a security) by *C* then the investor should buy security *i* only if the additional costs of buying security *i* is lower than the expected outcome (according to definition in chapter 2). Expressed mathematically the investor should buy security *i* only if Formula 10 holds.

$$W_t X_i ER_i > C$$

Formula 10

If ER_i in Formula 10 is replaced by $\left(\frac{\chi k}{100T}\right)^{th}$ largest daily outcome of security *i* from the history the formula would be adjusted for level of confidence equal to χ%.

If the investor doesn't buy some of the securities due to too high transaction costs he should split his wealth to the remaining securities which he buys anyway according to Formula 11. This formula ensures that whole wealth will be invested, i.e. sum of all X_i's is equal to 1.

$$X_i' = \frac{X_i}{\sum_j X_j}$$

Formula 11

The algorithm itself would not be affected by existence of transaction costs, additional steps described in this chapter would have to be performed at the end of the portfolio selection process. In order to keep the description of the algorithm provided by this thesis and experiments simple we'll suppose that the condition of low transaction costs holds and therefore we'll disregard them without loss of generality.

5 Experiments

In this chapter we are going to test our ITA against Brandt's approach in several different scenarios in order to determine strengths and weaknesses of both approaches. We are going to perform Monte Carlo simulation based on historical data of 5 different sets of 3 different stock prices set during 3 historical periods. The Monte Carlo simulation was performed 100 times for each of these data sets. Since the underlying data were taken from the same period in each experiment the investment strategies (portfolio vectors) were the same for all simulations within a single experiment. In order to make our test as general as possible the stocks' belonging into efficient combinations of expected return and variance as stated in chapter 3 was not tested. In the first test both approaches will be applied on selection of 7 highly inter-correlated stocks of IT industry companies during 3rd quarter of 2009. World economic crisis gave us a great opportunity to test the algorithm during two main parts of economic cycle – expansion and crisis. We have selected 10 most liquid stocks from New York Stock Exchange and 10 most liquid stocks from London Stock Exchange and applied both our ITA and Brandt's approach on historical data of prices of these stocks taken from 2 periods (1st August 2006 – 31st October 2007 and 15th August 2008 – 14th August 2009).

The calculations for all 5 experiments were performed for 3 risk aversions (values of γ according to definition given in chapter 2):

- 10 (low risk aversion)
- 20 (middle risk aversion)
- 50 (high risk aversion)

The results of experiments will be commented in the beginning of each experiment. Tables of portfolios calculated by both approaches will follow. The following charts will be provided for each experiment:

- Histograms of daily returns for each stock in the experiment
- Arithmetic average of outcomes of all simulations for both portfolios
- Variance of outcomes of all simulations for both portfolios
- Number of better results of the 2 compared approaches
- Value-at-Risk of the portfolio selected by both approaches

At the end of this chapter conclusion of all experiment's results will be provided.

5.1 Data Preparation

For each period daily returns of individual stocks were calculated according to Formula 12, where closing price of i^{th} security on day t is denoted by p_t^i.

$$r_t^i = \frac{p_t^i - p_{t-1}^i}{p_{t-1}^i}$$

Formula 12

Vectors of daily returns were constructed for each selected company (this vector had T-1 items – number of daily returns for period of T days). Their mutual covariances were calculated according to Formula 13, where $\bar{r}^i$ and $\bar{r}^j$ are arithmetic average returns of i^{th} and j^{th} security respectively.

$$\sigma_{ij} = \frac{1}{T}\sum_{t=1}^{T}\left[(r_t^i - \bar{r}^i)(r_t^j - \bar{r}^j)\right]$$

Formula 13

In the following calculations $\bar{r}^i$ is considered to be equal to ER_i for all securities. After the suggested investments were calculated using both methods n days from the observed time interval were selected randomly. Let's denote return of i^{th} security on j^{th} future day $\tilde{r}_{T+j}^i$. The overall return of one particular security within our portfolio is then defined by Formula 14

$$R_i = X_i \sum_{t=T+1}^{T+n} \tilde{r}_t^i$$

Formula 14

Sum of these returns gives us total return of the portfolio.

5.2 Value at Risk

In order to express how much money is risked Value-at-Risk model (hereafter VaR) will be used. VaR basically means maximal loss which might occur in a given time period with given probability. VaR model is described in more details by Schachter (1997).

There are several ways to calculate VaR. Since we don't know the statistical distribution of returns of the stocks in our portfolios historical data have to be used to find VaR_α. VaR_α is calculated according to Formula 15, where $r_i^{\{(T-1)\alpha\}}$ is $(T-1)\alpha^{th}$ smallest return of i^{th} stock within the corresponding historical time period and I is investment horizon. VaR is therefore explained as fraction of W_t.

$$VaR_\alpha = \sum_{i=1}^{N} X_i r_i^{\{(T-1)\alpha\}} \sqrt{I}$$

Formula 15

We've put $\alpha = 5\%$ in all experiments. VaRs for each experiment are calculated and the values are shown in charts as a part of each experiment's results.

5.3 ITA Calculation

In order to calculate investment suggestions using ITA software which can be found on the attached CD was used. It takes additional information about the structure of the tree. Format of the tree is described (as well as other instructions for use of this software) in chapter 8 of this thesis. The calculation is processed according to description given in chapter 4 of this thesis.

5.4 Brandt's Approach Calculation

Brandt's investment strategy is calculated according to Formula 4. This formula always returns vector of suggested investment shares with sum of all of its elements equal to 1, however some of its elements may be negative. Since negative investments are not considered in ITA adjustments according to Formula 16 were performed on the result of Formula 4.

$$\hat{\alpha}_i' = \begin{cases} \hat{\alpha}_i & for\ \hat{\alpha}_i > 0 \\ 0 & for\ \hat{\alpha}_i \le 0 \end{cases}$$

Formula 16

After this modification the condition of investing exactly whole investor's wealth was usually not met and therefore normalization using Formula 11 was performed.

5.5 IT Industry Short Run Simulation Results

Data set used for the first set of experiments consists of data of several companies from IT industry traded on New York Stock Exchange (NYSE) were selected. Companies from a single line of business were selected on purpose, because strong correlations were expected. IT industry was selected, because companies within this line of business are on the top positions according to liquidity at NYSE. This fact ensures the prices of these stocks to behave in manner quite close to effective market.

For this Monte Carlo simulation the following companies were chosen:

- Google
- IBM
- Microsoft
- Sun Microsystems
- Yahoo
- Cisco
- Apple

The period from which the data were taken is 1st June 2009 – 14th October 2009. During this period world economies were about 1.5 years after 2008 economic crisis and they were growing. This means that we could expect the returns to be mostly positive. Shares of individual stocks in portfolio suggested by Brandt approach is given in Table 1 the one suggested by ITA is given in Table 2. Since the investor invests whole portfolio the sum of the shares (sum of all numbers within any row) is equal to 1.

risk aversion	google	ibm	microsoft	sun	yahoo	cisco	apple
10	0,04846	0	0,09731	0	0	0,225456	0,628774
20	0,110346	0,033248	0,130396	0	0	0,155229	0,57078
50	0,133473	0,606836	0,101374	0	0	0	0,158318

Table 1

risk aversion	google	ibm	microsoft	sun	yahoo	cisco	apple
10	0	0,107918	0,264855	0	0,024255	0,056636	0,546337
20	0,035459	0,213994	0,377925	0	0,073503	0,010241	0,288877
50	0,092814	0,218564	0,428095	0,051374	0,094898	0	0,114256

Table 2

Average returns for the simulations are given in Figure 6, variance of the outcomes in

Figure 7 and number of better performing simulations is given in Figure 8.

As we can see in Figure 6 ITA performed better in average only for risk aversion of 20 but it achieved greater number of better results than Brandt's approach for risk aversion of 10 and 20 as displayed in Figure 8. The explanation of this phenomenon is provided in Figure 7 – although variance of the outcomes should decrease with increasing risk aversion variance of outcomes of portfolio calculated according to Brandt's approach was greater for risk aversion of 50 than for 20 and even for 10. More risky portfolio selected for risk aversion of 50 therefore performed better than the less risky one selected by ITA.

In order to understand better the portfolio composition and results histograms of all stocks' daily returns are provided in Figure 5.

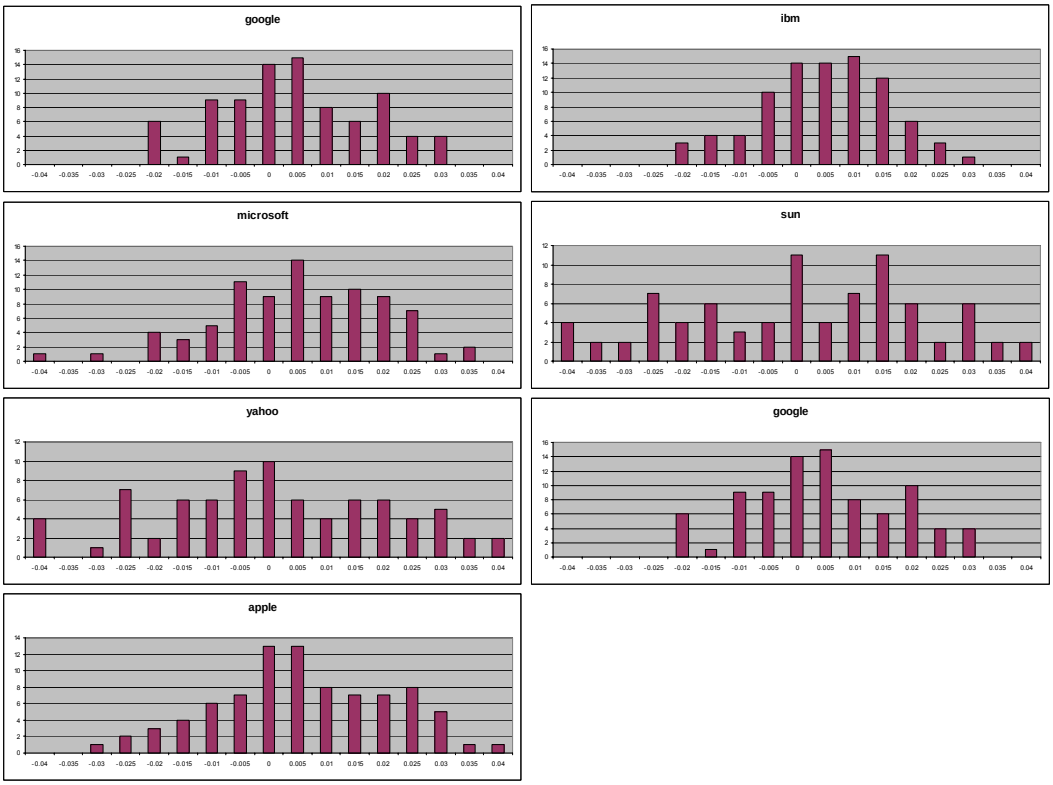


Figure 5

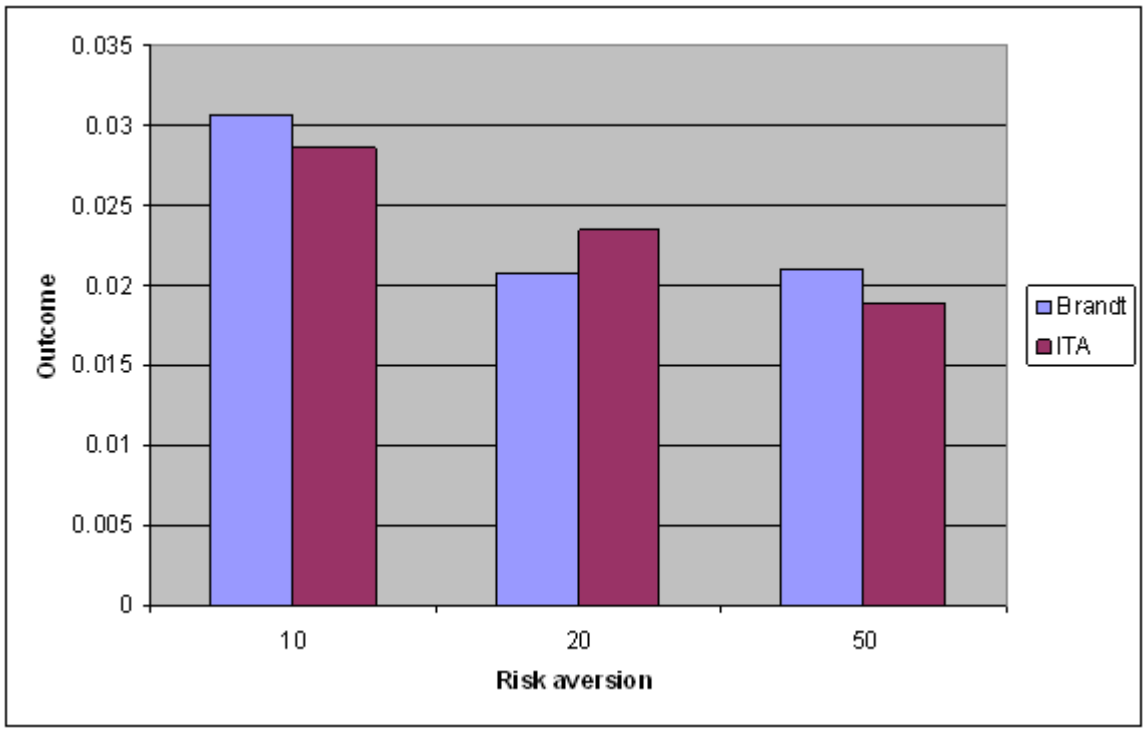


Figure 6

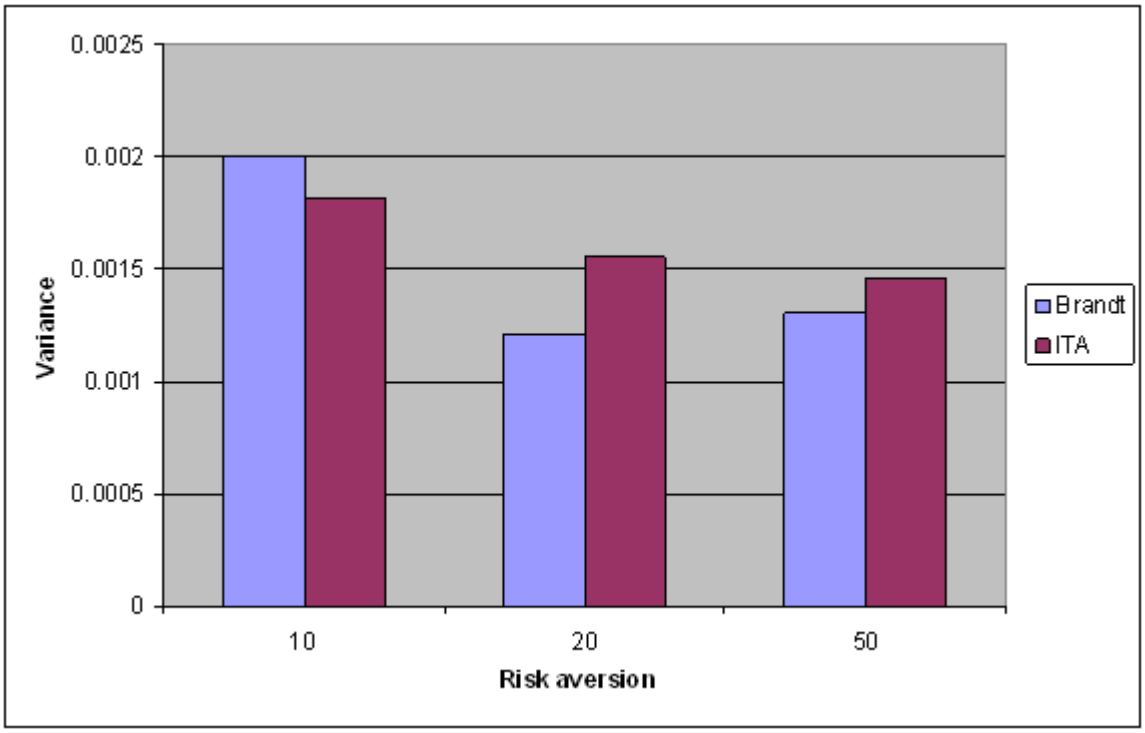


Figure 7

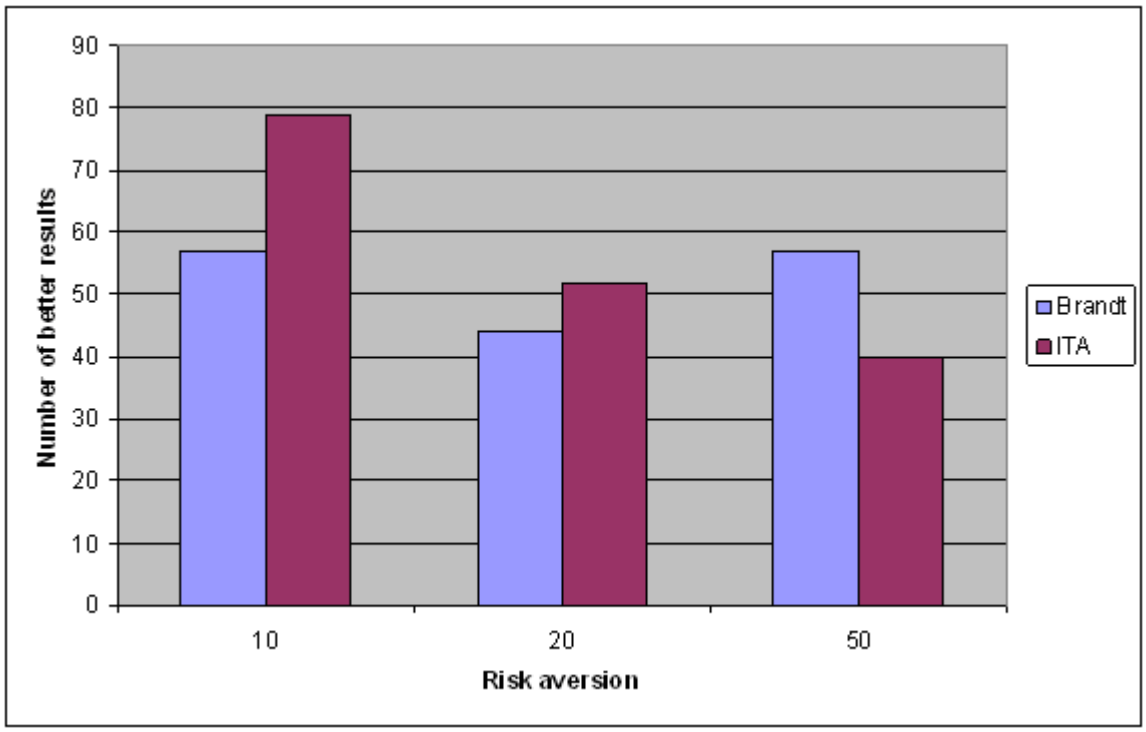


Figure 8

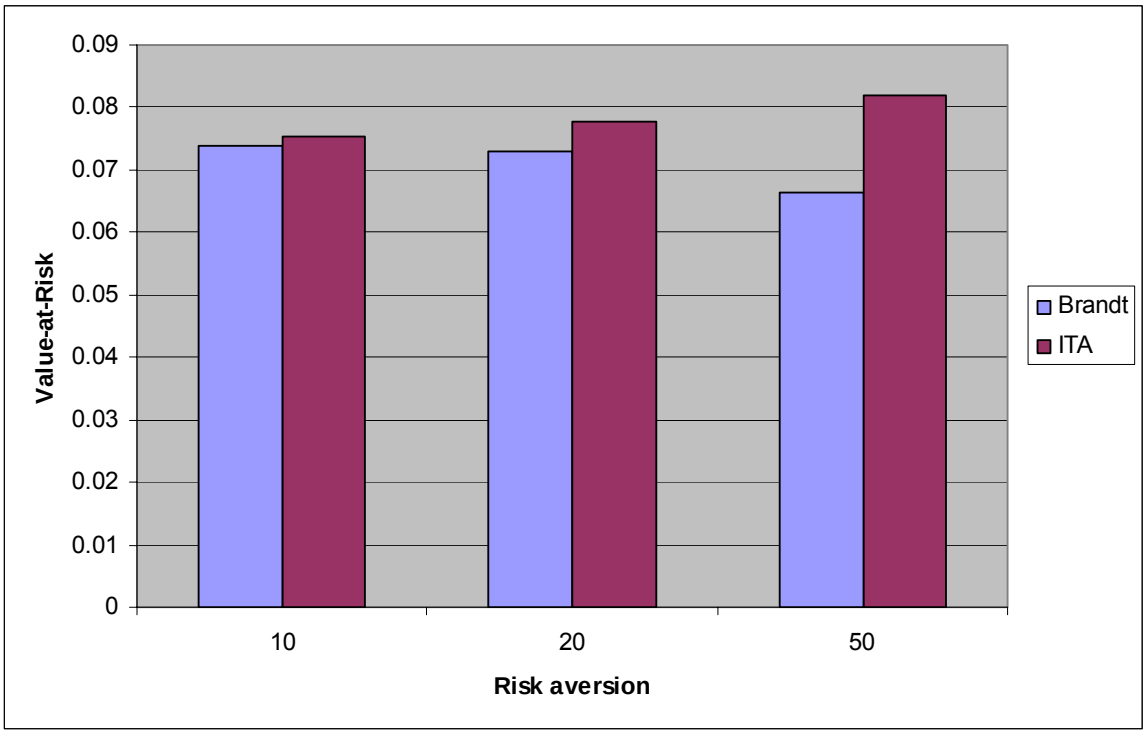


Figure 9

5.6 Middle Run Experiments Results

Performance of ITA in comparison to Brandt's approach was tested also for middle run (data from about 1 year with investing horizon of 30 days). In order to test its performance in various phases of economic cycle period before economic crisis (1st August 2006 – 31st October 2007) and period after economic crisis (15th August 2008 – 14th August 2009) were used as 2 reference periods for the Monte Carlo simulation which went according to the same framework as the one described in chapter 5.1. The simulation was performed for investment horizon of 30 days 100 times. Only aggregated data will be shown in the next chapters in order to keep the results clear.

5.6.1 Top 10 NYSE Stocks before Crisis

In order to perform the following experiments 10 most liquid stocks on NYSE were selected[4]:

- Citigroup Inc.
- EMC Corporation
- Exxon Mobil Corporation
- General Electric Company
- Hewlett-Packard Company
- Motorola, Inc.
- Pfizer Inc.
- Texas Instruments Incorporated
- Wal-Mart Stores, Inc.
- Time Warner Inc.

In Table 3

we can see portfolio selection determined by ITA. Notice especially high share of emc in case of risk aversion of 10 which (as we'll see in results of this experiments) was a very good choice. Optimality of high share of emc is also confirmed by its histogram provided in Figure 10.

risk aversion	citi	emc	exxon	ge	hp	motorola	pfizer	tex_ins	wal_mart	warner
10	0	0.575624	0.168828	0.074213	0.163146	0	0	0	0	0.018189
20	0	0.274552	0.205808	0.119315	0.126759	0.022573	0.030348	0.085526	0.047897	0.087222

[4] Most Active NYSE Stocks in Share Volume (http://www.infoplease.com/ipa/A0104607.html)

50	0	0.12644	0.164739	0.129439	0.090619	0.053587	0.116833	0.095438	0.086635	0.136269

Table 3

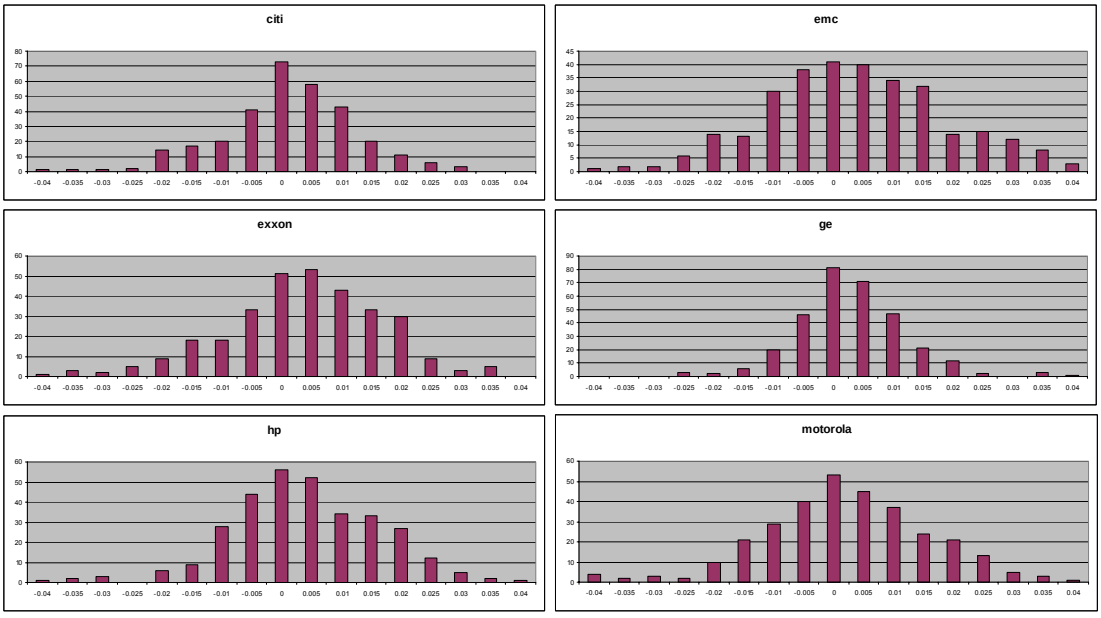
Portfolio selection determined according to Brandt's approach may be found in Table 4

.

risk aversion	citi	emc	exxon	ge	hp	motorola	pfizer	tex_ins	wal_mart	warner
10	0	0.304091	0.201947	0.191443	0.302519	0	0	0	0	0
20	0	0.217346	0.183051	0.257202	0.242687	0	0	0	0.042318	0.057396
50	0	0.105921	0.140388	0.284482	0.153258	0	0.067576	0.0101	0.102868	0.135407

Table 4

Average outcomes of all 100 Monte Carlo simulations may be seen in Figure 11. Average outcome of portfolio determined by ITA performed in average much better than the one selected by Brandt's approach for risk aversion of 10 (probably because of high share of emc with high volatility as we can see in Figure 12). In cases of risk aversions of 20 and 50 average outcomes and their variances are very similar. This doesn't correspond to results displayed in Figure 13 – number of simulations where the portfolio determined by ITA performed better than the one determined by Brandt's approach. Significant difference for risk aversion of 10 corresponds with average outcome for this aversion but significant difference of better performing portfolio for risk aversion of 50 doesn't. This might be explained by more risky stocks used and "wrongly" selected days during the simulation (random days selection).



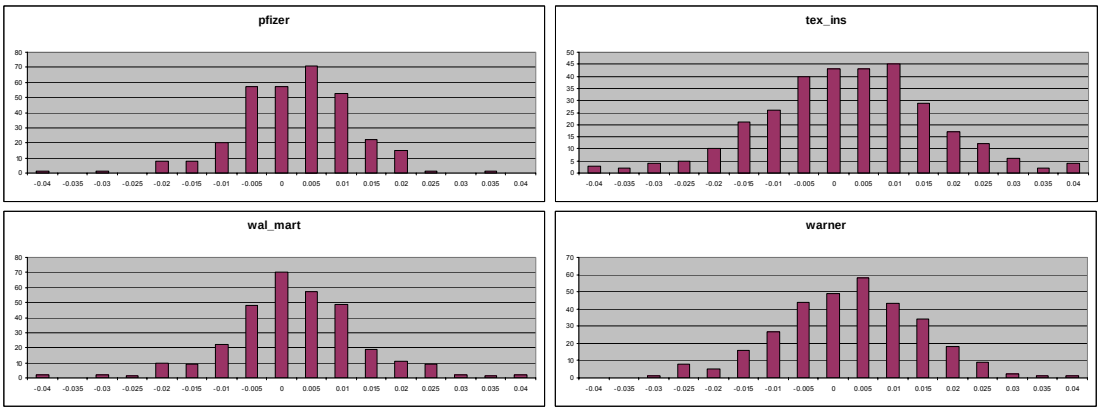


Figure 10

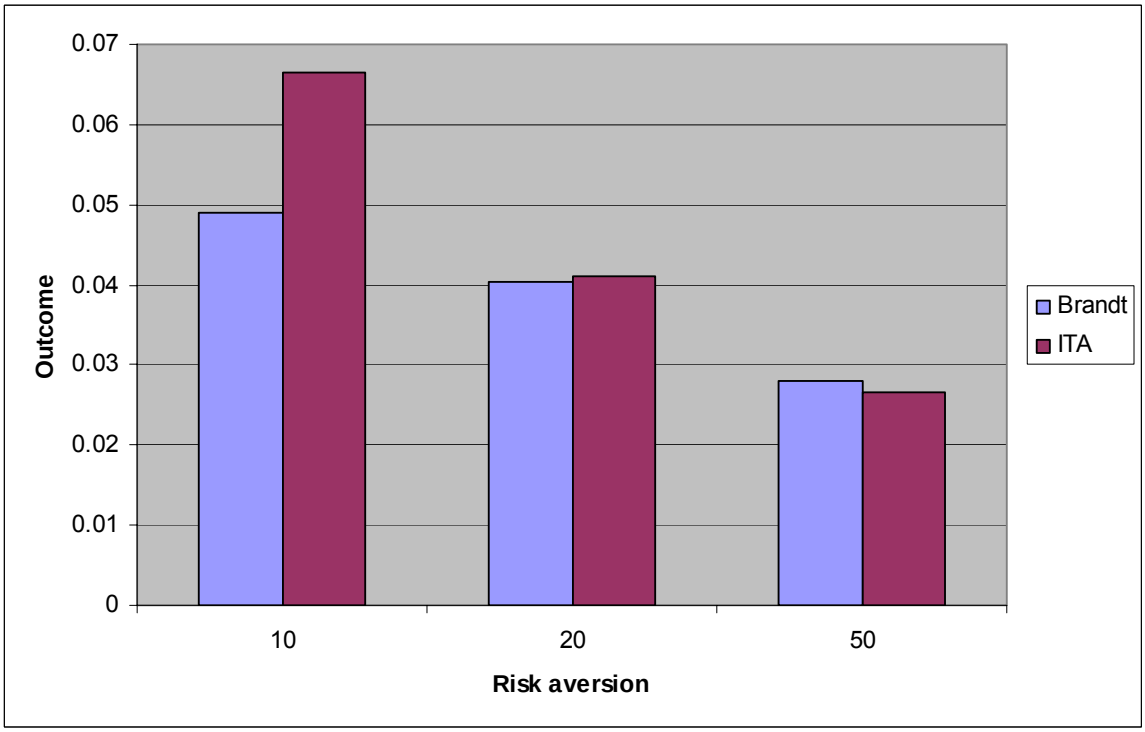


Figure 11

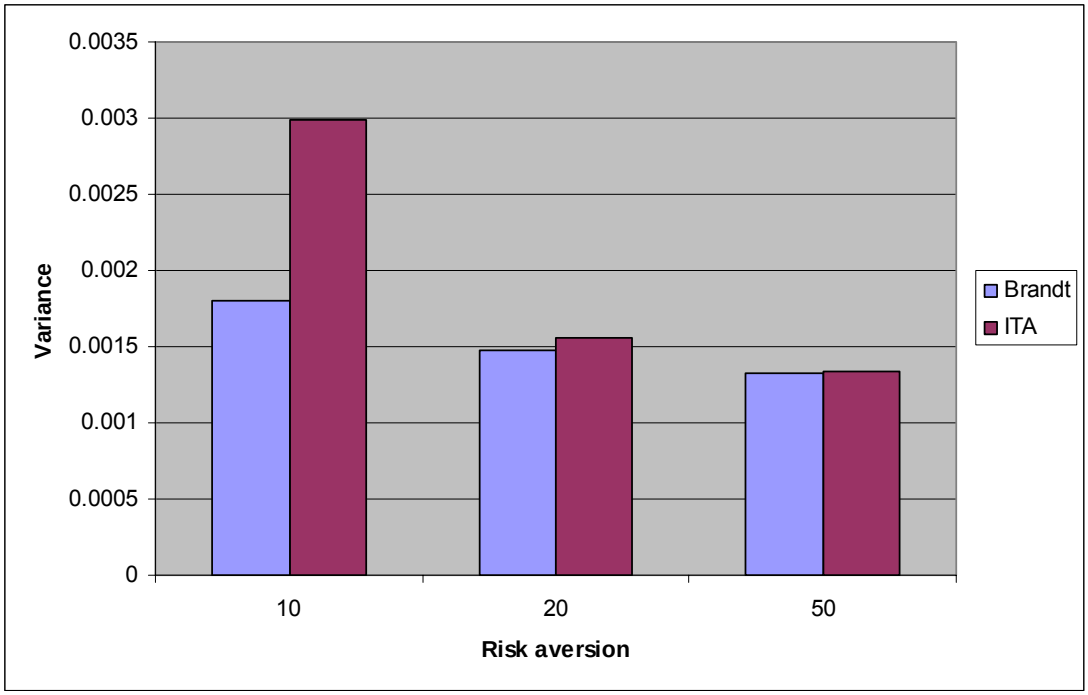


Figure 12

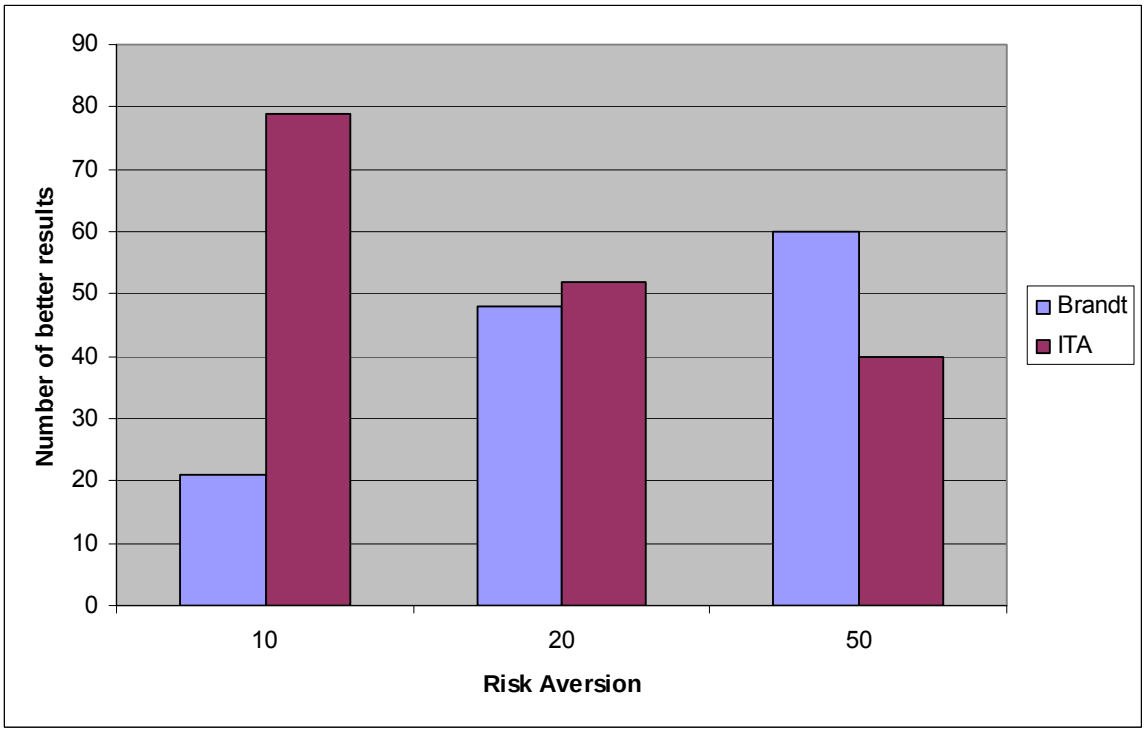


Figure 13

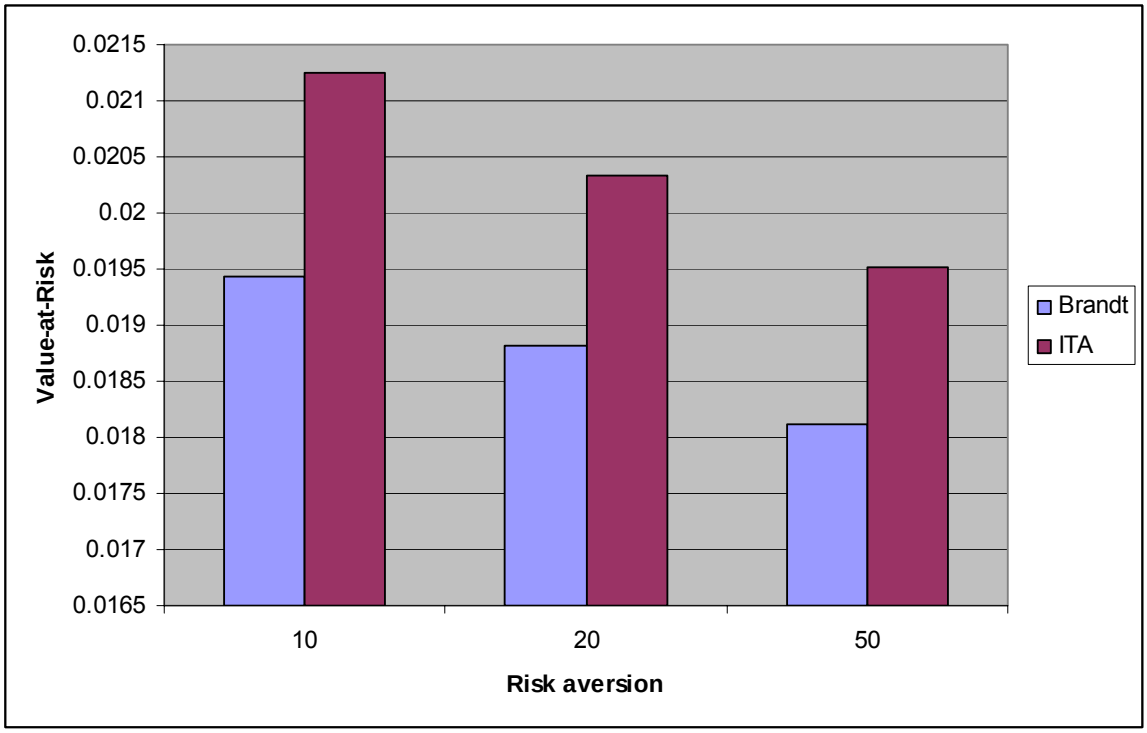

0.0215
0.021
0.0205
0.02
0.0195
0.019
0.0185
0.018
0.0175
0.017
0.0165
Value-at-Risk
10
20
50
Risk aversion
Brandt
ITA

Figure 14

5.6.2 Top 10 NYSE Stocks during Crisis

The same simulation for the same stocks in portfolio but for different time period (15th August 2008 – 15th August 2009) was performed. Resulting portfolio for ITA is in Table 5 and for Brandt's approach in Table 6. Histograms of all stocks' returns are represented by Figure 15.

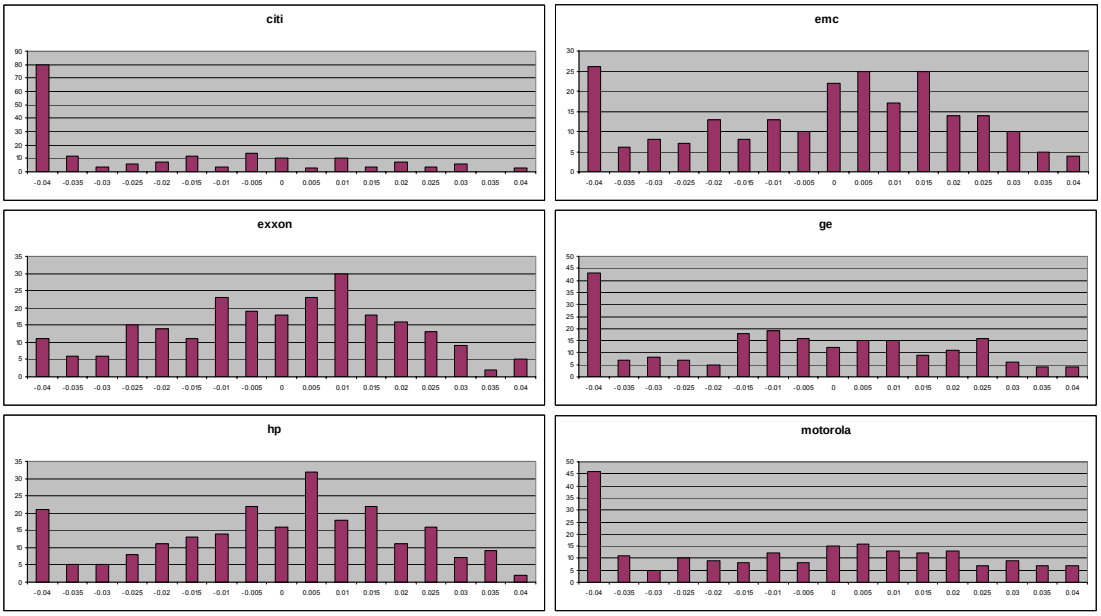
risk aversion	citi	emc	exxon	ge	hp	motorola	pfizer	tex_ins	wal_mart	warner
10	0.044543	0.065945	0.189218	0.038483	0.090666	0.047132	0.075801	0.116579	0.298853	0.032782
20	0.045937	0.058508	0.189772	0.044286	0.085486	0.044714	0.090979	0.109357	0.296776	0.034185
50	0.046711	0.054462	0.189584	0.04812	0.082559	0.043284	0.100345	0.105303	0.29469	0.034943

Table 5

risk aversion	citi	emc	exxon	ge	hp	motorola	pfizer	tex_ins	wal_mart	warner
10	0	0.108171	0.044536	0	0.036451	0	0.186402	0.082303	0.510333	0.031804
20	0	0.099521	0.04449	0	0.029279	0	0.195715	0.077245	0.520257	0.033494
50	0	0.094268	0.044461	0	0.024923	0	0.201371	0.074174	0.526284	0.034519

Table 6

Even though we expect loss because of the crisis ITA has selected portfolio which produced positive outcome for risk aversions of 10 and 20 as we can see in Figure 16 while the portfolio produced by Brandt's approach lost value in average for all selected risk aversions. The differences between average outcomes are quite small (the difference is not significant with regards to the corresponding variances) which corresponds to data displayed in Figure 18 – number of better results.



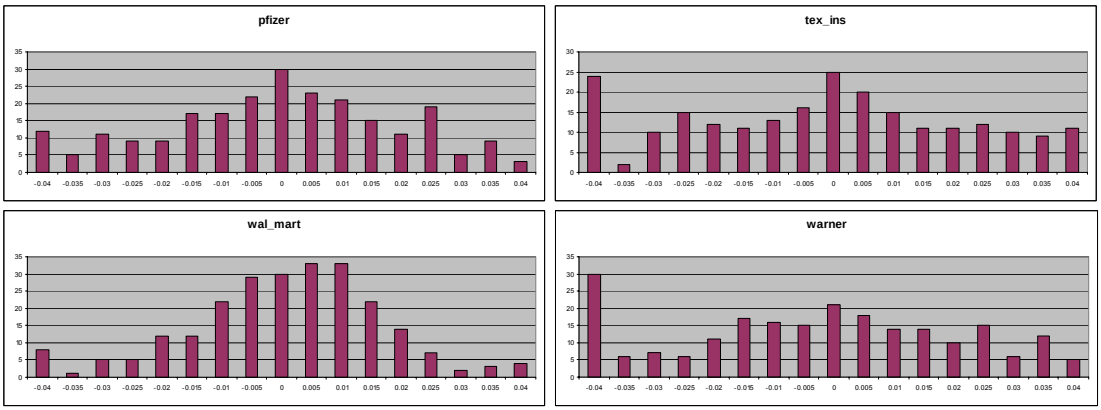


Figure 15

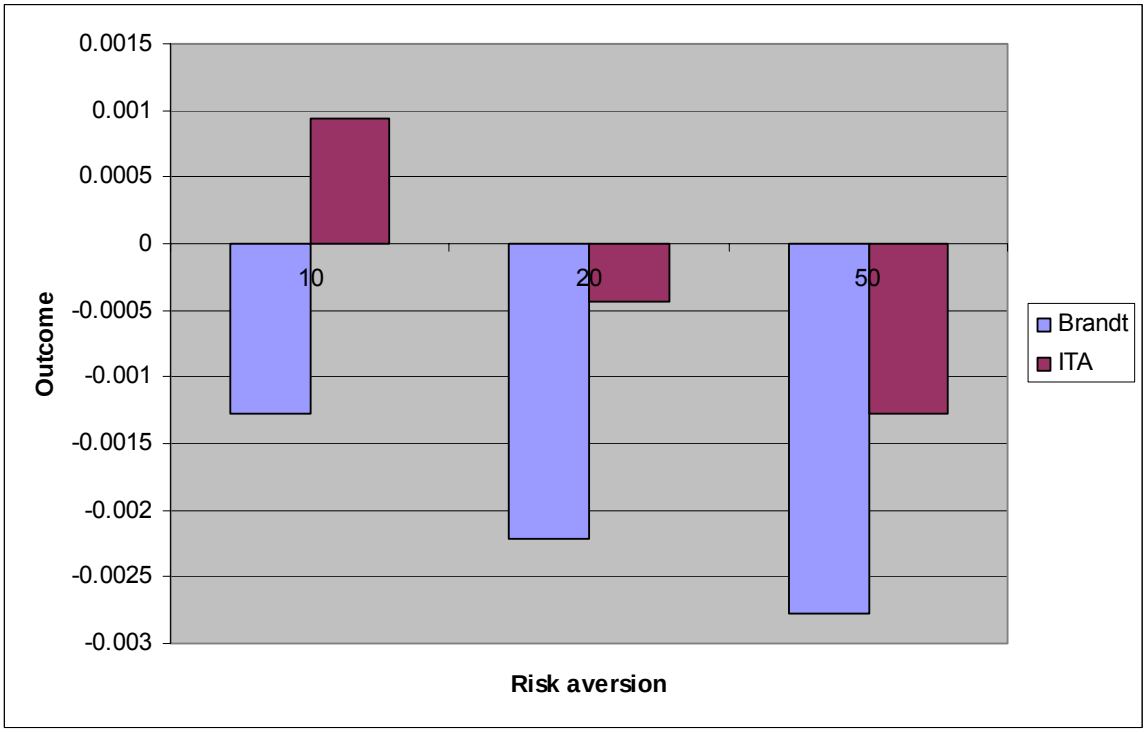


Figure 16

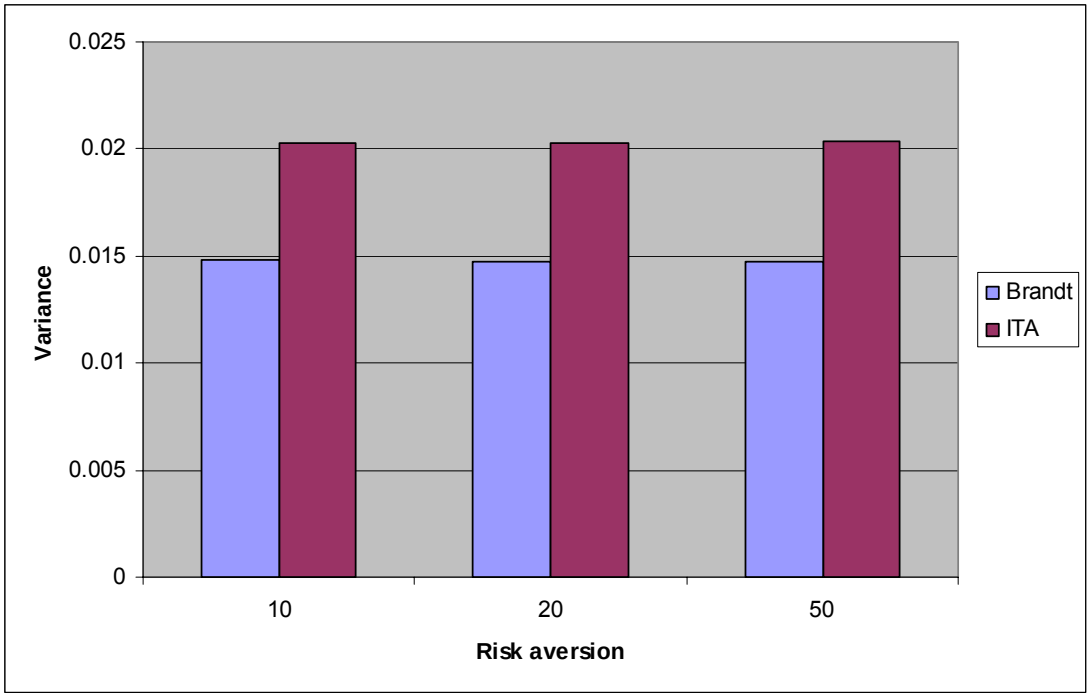


Figure 17

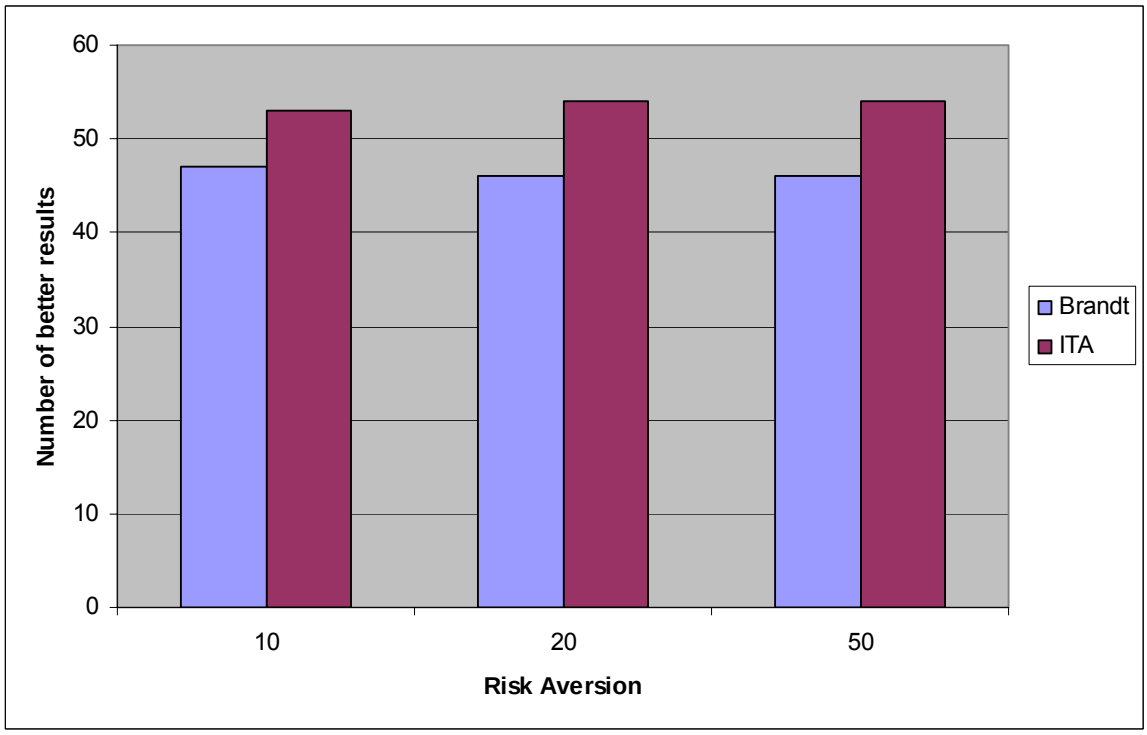


Figure 18

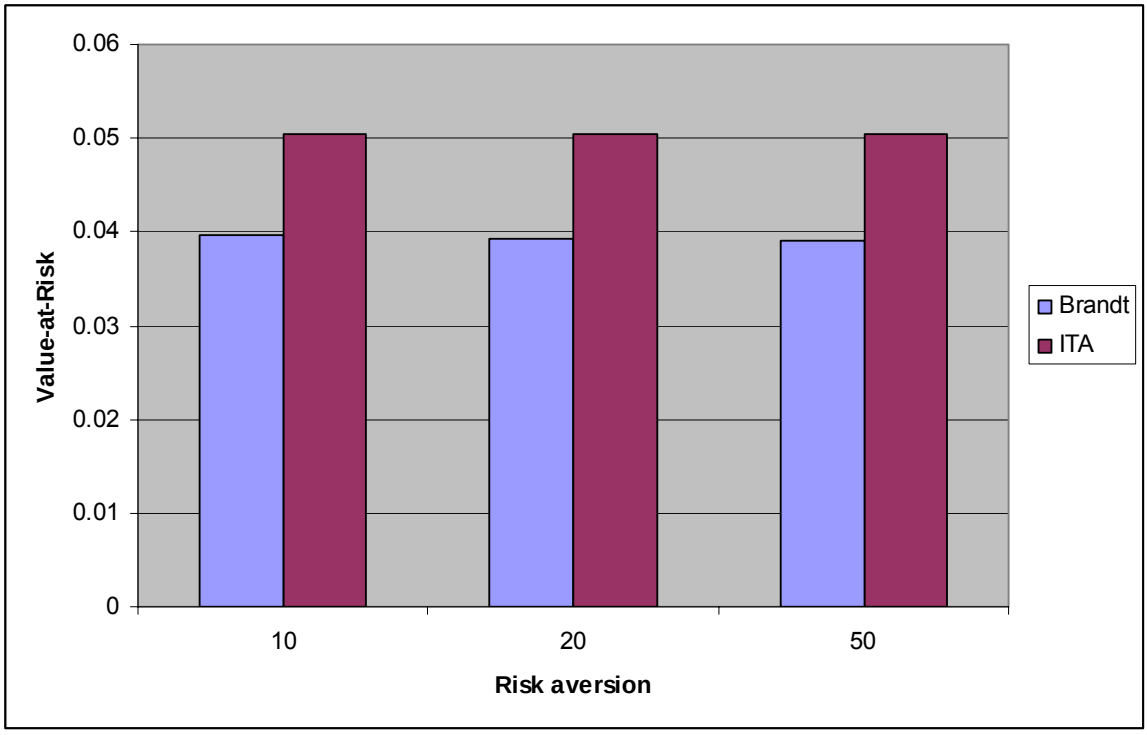

0.06
0.05
0.04
0.03
0.02
0.01
0
Value-at-Risk
10
20
50
Risk aversion
Brandt
ITA

Figure 19

5.6.3 Top 10 London Stock Exchange Stocks before Crisis

London Stock Exchange was founded in 1801 and now it the biggest purely European stock exchanges according to total value of share trading during year 2008[5]. Its index FTSE100 is a share index of 100 most capitalized UK companies listed in this stock exchange. 10 most traded companies from FTSE100 (according to volume of trade on 18th November 2009[6]) were chosen for the portfolio selection. These companies were:

- Vodafone Group PLC
- Lloyds Banking Group PLC
- Royal Bank of Scotland Group PLC
- Barclays PLC
- BT Group PLC
- WM Morrison Supermarkets PLC
- Marks & Spencer Group PLC
- HSBC Holdings PLC
- Centrica PLC
- Tesco PLC

Portfolio selection determined by ITA is in Table 7 and by Brandt's approach in Table 8.

risk aversion	vodafone	lloyds	rbsg	barclays	bt	wmms	msg	hsbc	Centrica	tesco
10	0.307204	0.012723	0.002466	0.003702	0.169925	0.064217	0.051039	0.033999	0.254131	0.100594
20	0.178249	0.030797	0.007933	0.008974	0.194314	0.039025	0.061163	0.090113	0.303512	0.085921
50	0.112361	0.048058	0.014587	0.013969	0.18396	0.029605	0.06926	0.147605	0.298886	0.08171

Table 7

risk aversion	vodafone	lloyds	rbsg	barclays	bt	wmms	msg	hsbc	Centrica	tesco
10	0	0.096307	0	0	0.148534	0	0.003162	0.535587	0.11344	0.10297
20	0	0.082097	0	0	0.145626	0	0.009459	0.540554	0.117504	0.104759
50	0	0.07124	0	0	0.143404	0	0.01427	0.544349	0.12061	0.106127

Table 8

Histograms of all stocks' returns are displayed in Figure 20. Charts displaying outcomes and number of better performing results are displayed in Figure 21 and Figure 23 respectively.

[5] Source: World Federation of Stock Exchanges (http://www.world-exchanges.org/statistics/ytd-monthly)

[6] Source: Bloomber (http://www.bloomberg.com/markets/stocks/movers_index_ukx.html)

In this case ITA was performing much better than Brandt's approach. Explanation of this good result of ITA may be the fact that Brandt's investment advices were mostly negative which were set to 0 according to Formula 16.

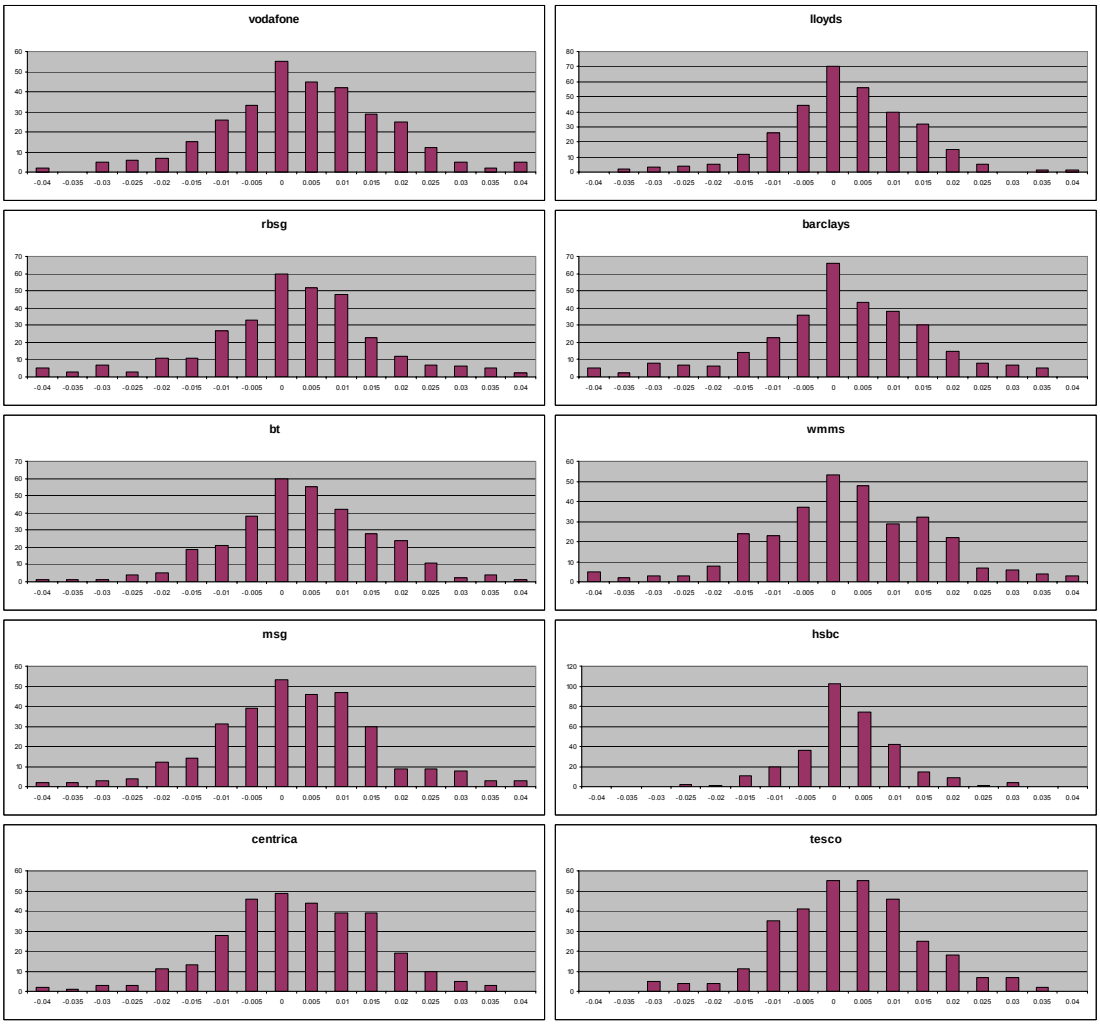


Figure 20

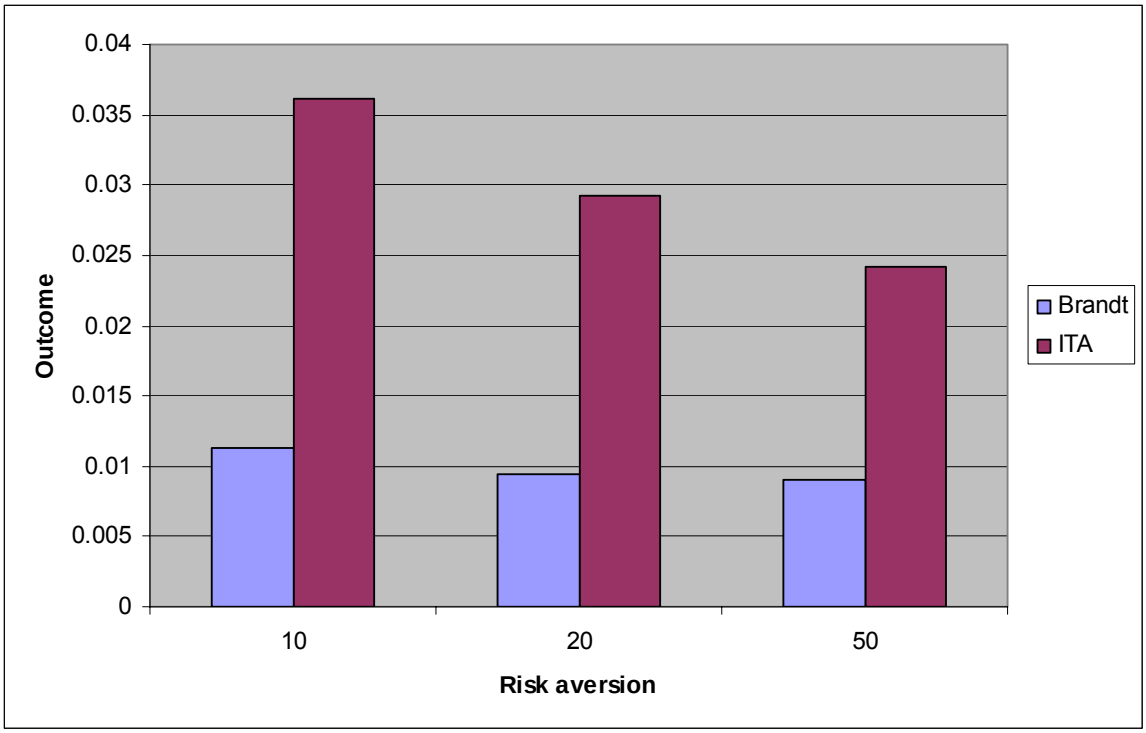


Figure 21

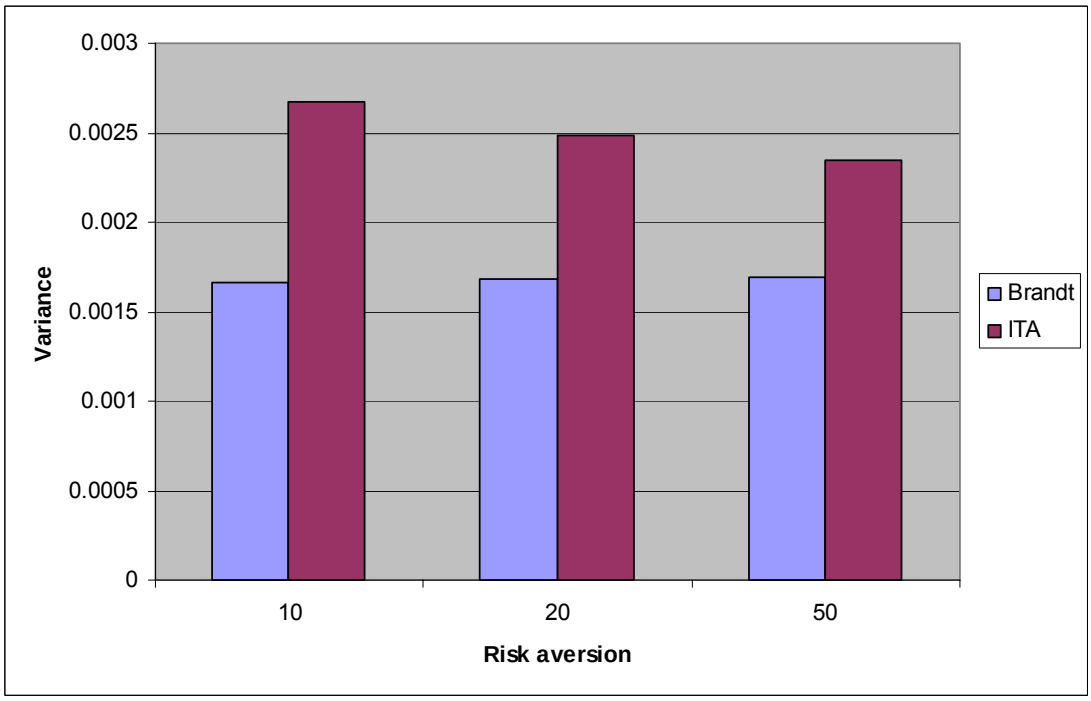


Figure 22

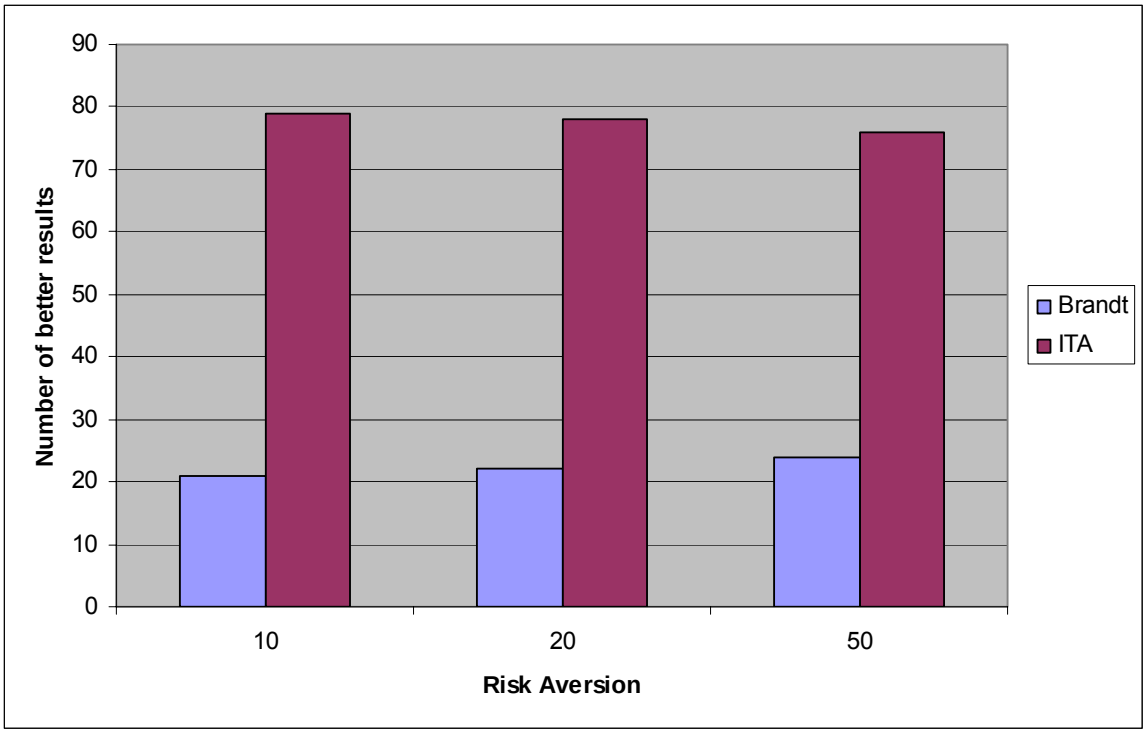


Figure 23

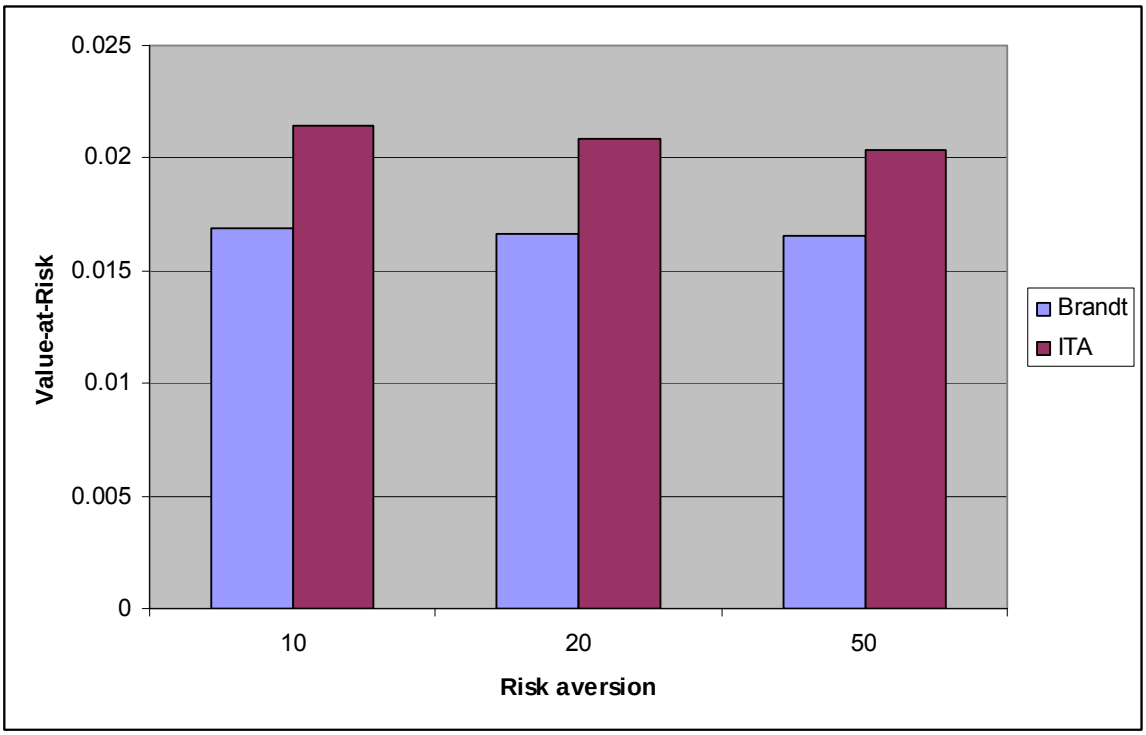


Figure 24

5.6.4 Top 10 London Stock Exchange Stocks during Crisis

This experiment takes into account the same stocks but different time interval (15th August 2008 – 14th August 2009). Since world economies were in crisis during this period negative outcomes are expected. The aim is therefore to minimize loses.

The portfolio selection made by ITA is written in Table 9 and by Brandt's approach in Table 10.

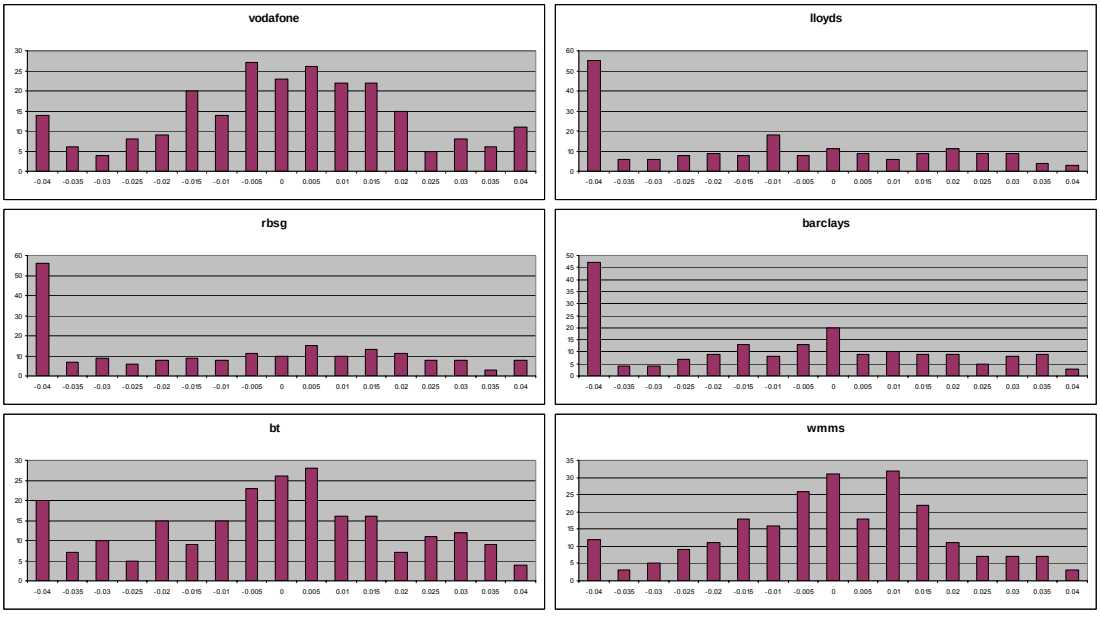
risk aversion	vodafone	lloyds	rbsg	barclays	bt	wmms	msg	hsbc	centrica	tesco
10	0.178882	0.01014	0.00387	0.016252	0.137464	0.076887	0.258587	0.13212	0.091555	0.094243
20	0.18655	0.008968	0.003789	0.014078	0.161516	0.076663	0.223124	0.128919	0.102537	0.093856
50	0.190449	0.008286	0.003713	0.012849	0.176486	0.076294	0.202925	0.126433	0.109231	0.093335

Table 9

risk aversion	vodafone	lloyds	rbsg	barclays	bt	wmms	msg	hsbc	centrica	tesco
10	0.192552	0	0	0.001736	0.040304	0.119588	0.212849	0.152122	0.114037	0.166812
20	0.190619	0	0	0	0.057353	0.110746	0.169317	0.17417	0.135631	0.162164
50	0.18945	0	0	0	0.066112	0.106077	0.14667	0.185407	0.146668	0.159616

Table 10

The charts displaying average outcome and number of better results for each process are Figure 26 and Figure 28 respectively. The outcomes are (as expected) negative but portfolio produced by ITA performed better for approx. 60% (measured by outcome) of 30 days Monte Carlo simulations.



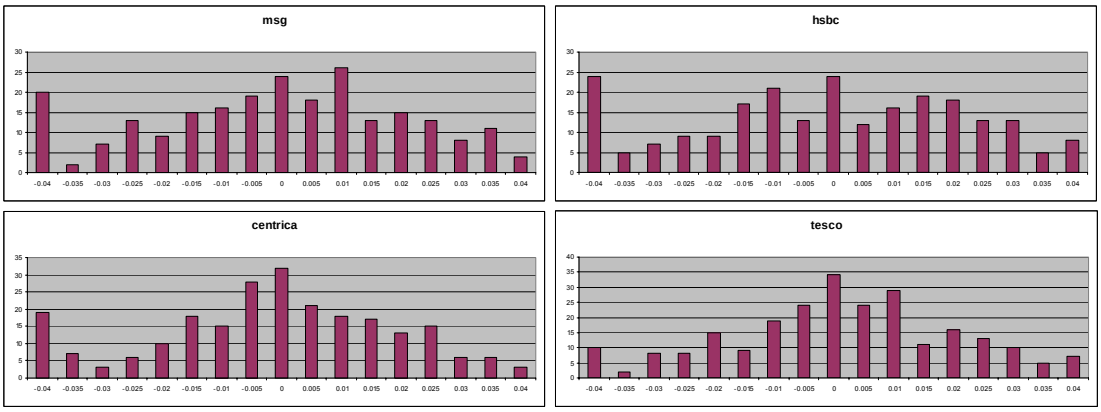


Figure 25

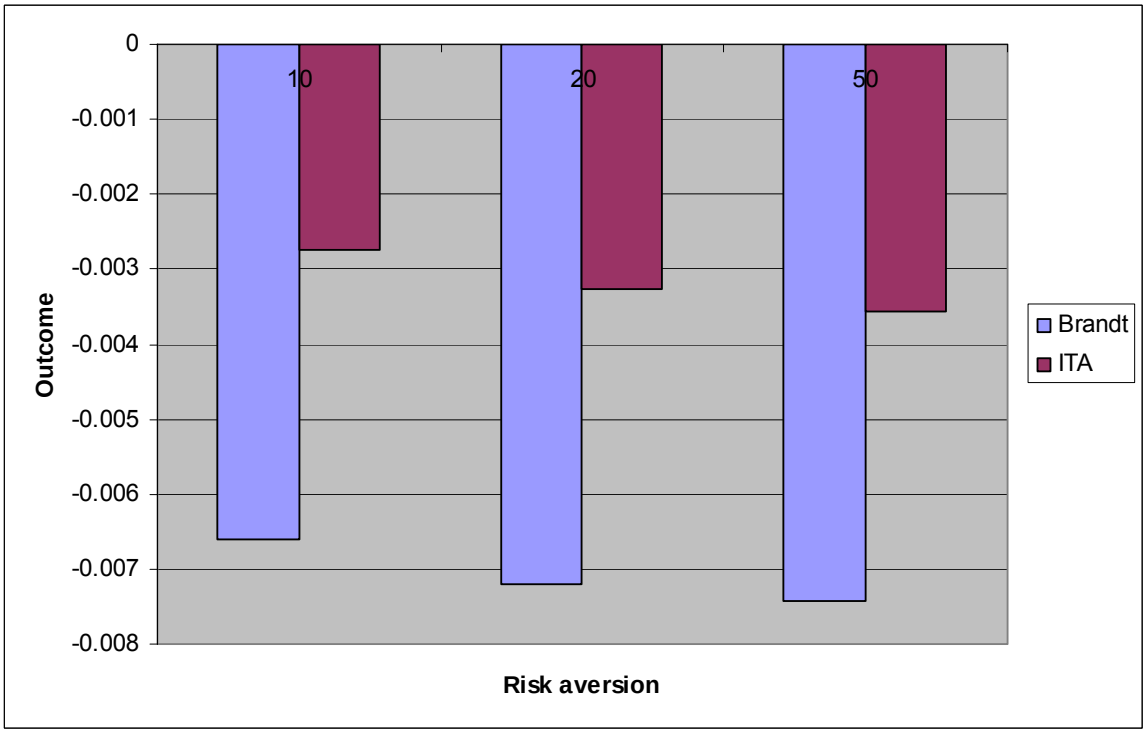


Figure 26

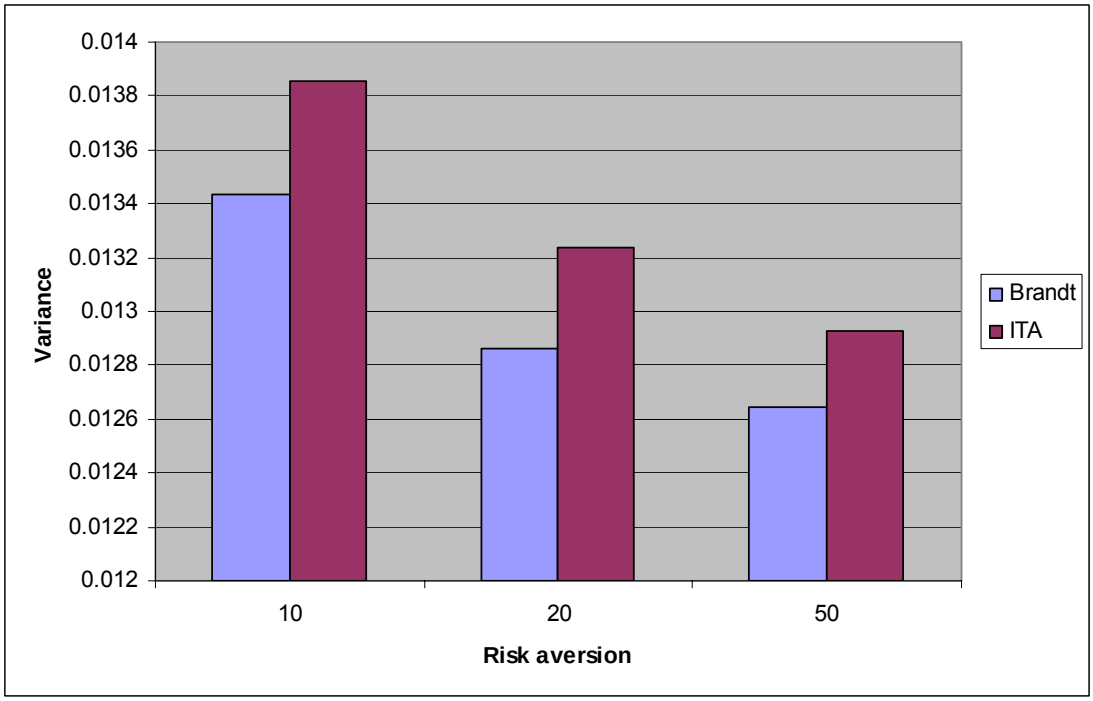


Figure 27

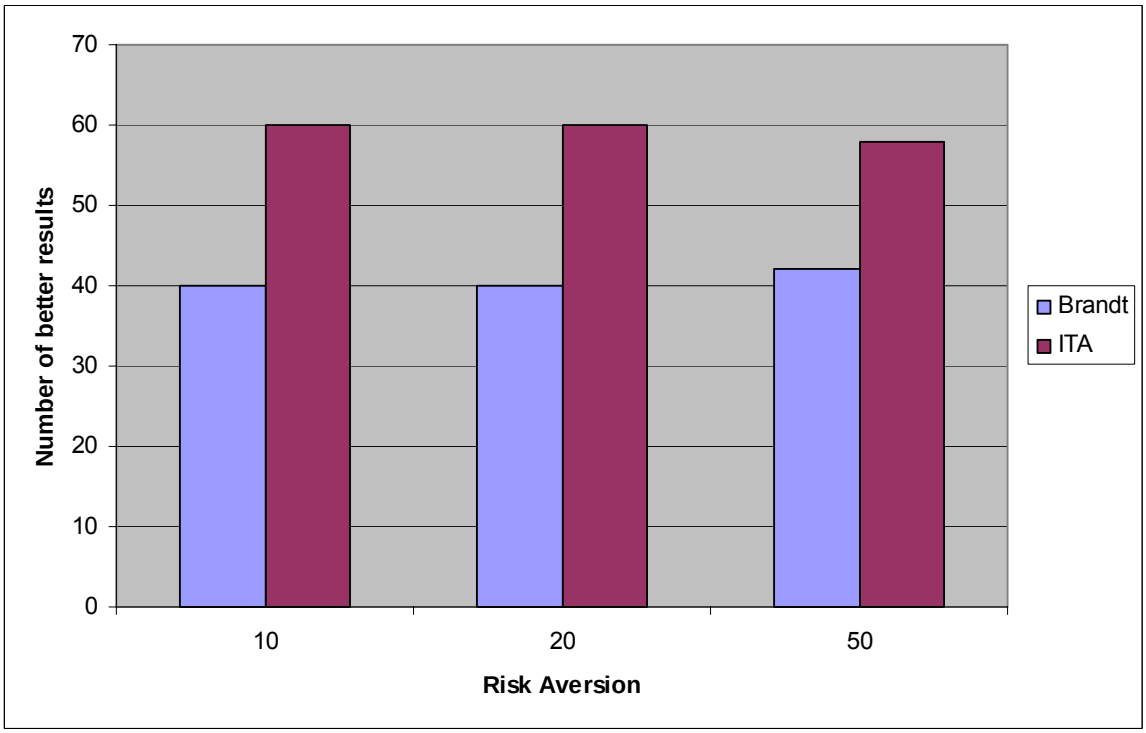


Figure 28

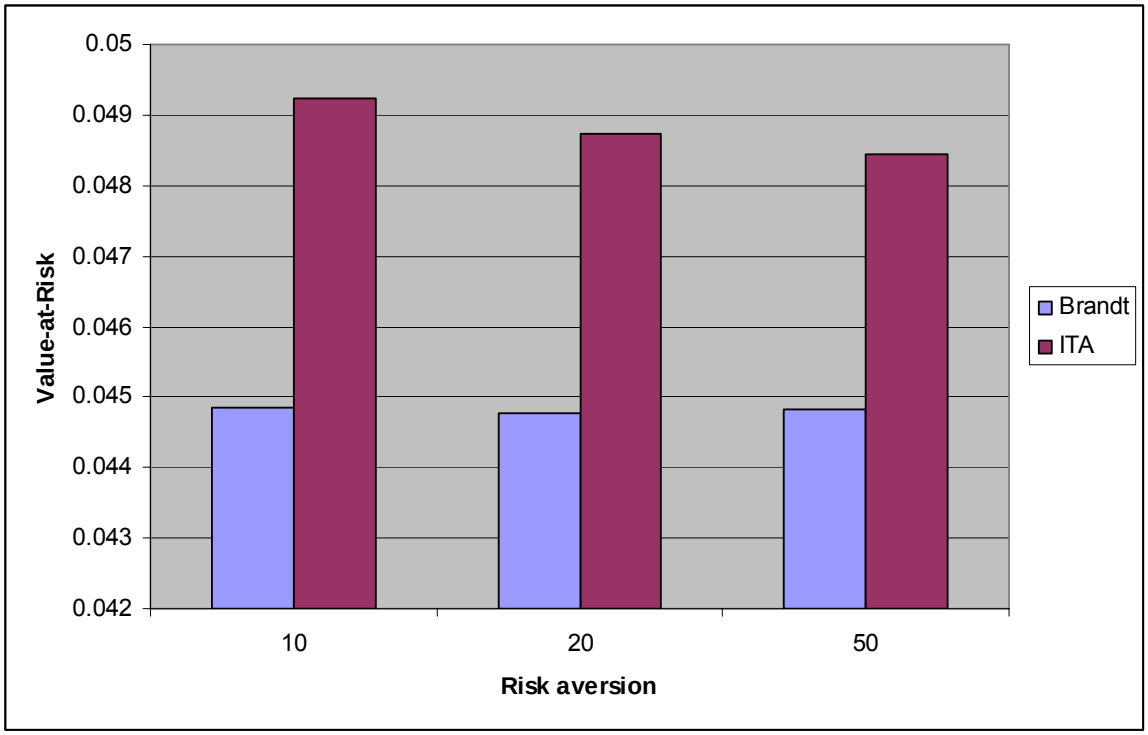


Figure 29

5.7 Conclusion of Experiments results

As we saw in the experiments ITA performs better on securities with very different inter-correlation (the ones in chapter 5.6) than on securities with similar (high) inter-correlation (the ones in chapter 5.5). This is probably caused by its nature of tree algorithm. Usually the portfolio generated by ITA was more volatile than the one generated by Brandt's approach (had greater variance and VaR than the portfolio calculated using Brandt's approach) but the ITA portfolio provided higher average outcomes and greater number of better results in simulations than Brandt's portfolio. The risk aversion factor works in most cases for both approaches well (reduces VaR and variance of the portfolio). However in several cases (e.g. in chapter 5.6.3) variance and VaR were almost the same for all 3 risk aversions in both cases (ITA and Brandt).

As a summary of the results it may be stated that on 2 out of 5 data sets ITA and Brandt's performance was quite similar (chapters 5.5 and 5.6.1) and on the remaining 3 data sets ITA performed better on all 3 levels of aversion.

6 Conclusion and Suggestions for Future Work

This thesis provided a short overview of several portfolio selection pieces of literature. In chapter 4 an alternative way of determining optimal portfolio composition based on agglomerative clustering algorithm according to expected returns and mutual covariances of securities was provided. The algorithm for portfolio selection calculation is based on a model with 2 securities in which the investor has to invest his wealth. This simple model is generalized to more securities using an algorithm described in chapter 4.3.

The approach suggested by this thesis was then compared with Brandt's approach using different datasets and different time periods during economic expansion and economic crisis in chapter 5. Highly correlated companies from IT industry were used in the first experiment. In other experiments most liquid stocks from NYSE and LSE were used. 2008 world economic crisis provided us with a greate opportunity to compare both approaches during expansion and crisis. Monte Carlo simulation (each with 100 steps) was performed for each data set (5 in total). In 3 of these simulations approach suggested by this thesis provided better results than Brandt's approach. In the remaining 2 simulations performances of both approaches were quite similar.

Use of informatics methods to solve economic problems probably provides large space for future research. The performance of this algorithm might be improved in several ways. One of them might be use of more sophisticated calculation of expected returns and their variance. The initial securities selection may be performed in a way which ensures efficient combination of expected return and variance according to Markowitz (1952). The set characterized by the "efficient combination of expected return and variance" may be modeled as a fuzzy set and therefore applying knowledge of fuzzy sets theory in our ITA might lead to better results as well.

7 References

Ait-Sahalia Y. and Brandt M. W. (2001): Variable Selection for Portfolio Choice, The Journal of Finance, Vol. 56, No. 4, pp. 1297-1351

Brandt M. W. (1999): Estimating Portfolio and Consumption Choice: A Conditional Euler Equations Approach, The Journal of Finance, Vol. 54, No. 5, pp. 1609-1645

Goldfarb D., Iyengar G. (2002): Robust portfolio selection problems, CORC Technical Report TR-2002-03

Hansen, L. P. (1982): Large sample properties of generalized method of moments estimators, Econometrica, Vol. 50, pp. 1029-1053

Jenček P, Vojtáš P., Kopecký M., Höschl C. (2009): Sociomapping in Text Retrieval Systems, Flexible Query Answering Systems 2009, LNAI 5822, 122-133.

Markowitz H. (1952): Portfolio Selection, The Journal of Finance, Vol. 7, No. 1, pp. 77-91

Schachter B. (1997): An Irreverent Guide to Value at Risk, Financial Engineering News, vol. 1 no. 1

Treynor J.L., Black F. (1973): How to Use Security Analysis to Improve Portfolio Selection, The Journal of Business, Vol. 46, No. 1 (Jan., 1973), pp. 66-86

Ward J. H., Jr. (1963): Hierarchical grouping to optimize an objective function, Journal of the American Statistical Association, 58, 236-244.

8 User's Manual

8.1 Installation and Prerequisities

In order to install TreeInvest copy content of \TreeInvest directory to any directory on your hard drive (e.g. C:\TreeInvest). You may run the software from the CD since it doesn't write anything to any disk drive.

Since the application is written using C# .NET framework 2.0 (or later) is required to be installed on Windows operating system.

8.2 Usage

After executing TreeInvest.exe main (and the only) window of the application appears (Figure 30). In the upper part of the window there is "Coherence tree" text box. Paste there coherence tree created using the correlation matrix of the securities in the format described by the following formal grammar:

<TreeString> = <Node>

<Node> = <SecurityName> | (<Node>, <Node>)<CorrelationOfSubtrees>

An example of such string is:

((S1, S2)0,805, S3)0,567

Paste the table consisting of security name, expected return and covariance matrix (use TAB as separator) into the multiline text box "Returns". You may copy these values from MS Excel sheet. The structure of this table for 3 securities (S1, S2 and S3) is as follows:

$$\begin{array}{llll} S1 & ER_1 & \sigma_{11} & \sigma_{12} & \sigma_{13} \\ S2 & ER_2 & \sigma_{21} & \sigma_{22} & \sigma_{23} \\ S3 & ER_3 & \sigma_{31} & \sigma_{32} & \sigma_{33} \end{array}$$

You can change the risk aversion in the corresponding text box.

In order to calculate suggested investments click on the Calculate button. Suggested investments of your portfolio will appear in the corresponding text field together with

their expected return and variance.

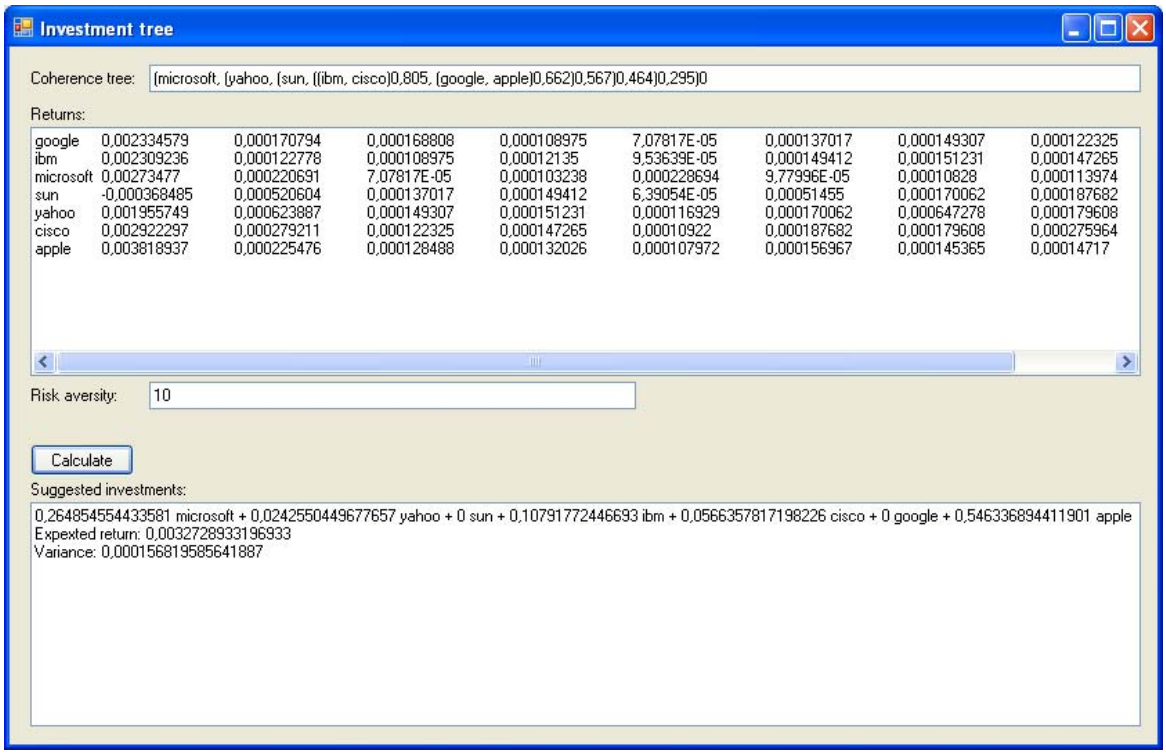


Figure 30