Applied Mathematical Sciences, Vol. 6, 2012, no. 94, 4673 - 4684

Stocks Portfolio Optimization Using Classification and Genetic Algorithms

M. El hachloufi

Faculty of Science
Department of Mathematics, Rabat, Morocco
elhachloufi@yahoo.fr

Z. Guennoun

Faculty of Science
Department of Mathematics, Rabat, Morocco
guennoun@fsr.ac.ma

F. Hamza

Faculty of Law
Department of Economics and Management
Tangier, Morocco
fhamza2004@yahoo.fr

Abstract

In this paper we present an approach based on the classification and genetic algorithms to obtain an optimal stock portfolio of a reduced size from an initial portfolio, which leads to a financial gain surplus in terms of cost and taxes reduction, and a performance at reduced design loads. This approach takes place in two steps: The first step is to classify the actions of this portfolio into classes, known as under portfolios, with the expected returns and VaR's close to each other by using the K-Means method, then we apply an algorithm dynamic optimization called Min-VaRMaxVaL to the portfolio that has the highest expected return and the lowest average VaR obtained by this classification.
The algorithm proposed MinVaRMaxVaL for the selection of optimal actions portfolio is based on genetic algorithms and Value at Risk (VaR). The objective of our algorithm is to minimize risk and maximize portfolio value at the same time through two stages.
The first step is to minimize the risk measured by the value at risk (VaR) for a given value of portfolio. While the second step, a dynamically maximizes the value of portfolio as the result is greater than the

portfolio value set at the first stage and the risk resulting from the second stage is lower than that obtained from the first.
The proportions of shares in the portfolio are the optimal portfolio.

Keywords: Optimal Choice, Actions Portfolio, Value at Risk (VaR), Profitability, Classification, K-Means, Genetic Algorithms

1 Introduction

The choice of an optimal portfolio of shares has been for so long a subject of major interest in the field of financial mathematics. His theory is developed following the pioneering work of Harry Markowitz [1] which aims to answer the question: if one gives an amount to be invested in financial markets, what should be the structure of this investment, what are the titles that we should select and in which proportions?
Markowitz was the first to introduce an approach used to optimize a portfolio of financial assets by providing the variance of returns as a measure of risk. Several criticisms have been directed to this approach with the character of the quadratic objective function and calculating the variance-covariance matrix being the most important, making it used so often in practice. To simplify the difficulties associated to the design load of Markowitz model, several models have been proposed as alternative models to the mean-variance approach. Some authors have attempted to linearize the portfolio choice problem as Sharpe [2], [3] and Stone [4]. Rudd and Rosenbeg [5] showed that the Markowitz model in its classic formulation still far from meeting satisfying a professional investor and they proposed a realistic portfolio management. More recently, several linear programming models were developed for a portfolio selection. Konno[6], Konno and Yamazaki [7], Zenios and Kang [8] and Speranza [9] have proposed to calculate the portfolio risk using the mean absolute deviation. They also have proposed a model to include the criterion of asymmetric risk, which eliminates most problems associated to the optimization model. With the appearance of linear models, Yoshimoto [10] developed an optimization system based on the mean-variance criterion. Hamza and Janssen [11] generalized standard models, while retaining the dual risk-return as a backdrop proposing a new measure of risk defined as a convex combination of the two semi-variances. They proposed also a general linear programming model based on separable programming techniques. Recently, a new risk measure called Value at Risk VaR [12] has been implemented to quantify the maximum loss that might occur with a certain probability, over a given period. Although VaR is subjected to a number of limitations, it has been chosen by the Basel Committee as the standard method for measuring the risk in the market portfolio. There are several models to calculate the VaR, the choice depends on the nature of the portfolio and the data used to estimate these parameters [13], [14]. The different models presented in this literature act either at the

return (maximizing the return for a given risk) or at the risk (minimizing the risk for a given return) to determine the optimal choice of portfolio, but not both at the same time, an approach that we developed in this work through a dynamic algorithm called MinVaRMaxVaL to improve the optimal choice in the sense of having a highest return (portfolio value) and small risk as mush as possible. In addition, the proposed approach is to obtain an optimal stock portfolio reduced in size from an initial portfolio to achieve a financial gain surplus in terms of cost and tax reduction, and a performance at reducing design loads.
This work is organized as follows: Section 1 an introduction, Section 2 deals with the presentation of some important elements in portfolio theory. The classification is presented in Section 3. In Section 4, we present genetic algorithms. Then, in Section 5, we propose the optimization procedure. Finally, the experimental results are presented and discussed.

2 Elements of portfolio theory

2.1 Profitability and Value Portfolio

We call return r_t of an action obtained by investing in an action, the ratio between the share price at the moment t and its course at the moment $t-1$, plus income (dividends) received during the period $[t-1, t]$:

$$r_t = \frac{c_t - c_{t-1} + d_t}{c_{t-1}} \tag{1}$$

where :

- c_t : The course of action i at the end of the period t.
- d_t : The dividend income at the end of the period.

The expected return of a share for a period T is given by

$$\overline{r_i} = \frac{1}{T}\sum_{t=1}^{T} r_{it} \tag{2}$$

The profitability of a portfolio consisting of expected return of k shares $\overline{r_i}$, $i = 1, \ldots, k$:

$$\overline{R_p} = \sum_{i=1}^{k} x_i \overline{r_i} \tag{3}$$

where $x_1, \ldots, x_n$ are the proportions of the wealth of the investor placed respectively in the shares i $(i = 1, \ldots, n)$. Let a portfolio containing k shares, its value is given by:

$$V = \sum_{i=1}^{k} n_i.V_i \tag{4}$$

where $n_1, \ldots, n_k$ are the numbers of the investor wealth of placed respectively in the shares and V_i is i^{eme} the price action. The change in value will obey the following relationship:

$$\Delta V(n) = \sum_{i=1}^{k} n_i \Delta V_i \tag{5}$$

2.2 Risk portfolio

The risk of a financial asset is the uncertainty about the value of this asset in an upcoming date. Variance, the average absolute deviation, the semi-variance, VaR and CVaR are means of measuring this risk.

2.2.1 VaR of an individual asset

The VaR is defined as the maximum potential loss in value of a portfolio of financial instruments with a given probability over a certain horizon. In simple words, it is a number that indicates how much a financial institution can lose with some probability over a given time. According to Esch, Kieffer, Lopez and Jorion [15], the VaR of the assets, for a period t and a probability level q is defined as the amount of loss expected so that this amount period $[t-1, t]$, should not be larger than the VaR with a probability α.

$$P[P_{t,0} \leq VaR(t, \alpha)] = \alpha \tag{6}$$

where

- $P_{t,0} = P_t - P_0$: represents the loss ("Loss") and is a positive random variable or negative;
- t : is the horizon associated with the VAR that is 1 day or more a day;
- α : is the probability level. Typically is 95%, 98% or 99%.

VaR depends on three elements:

- Distribution of profits and losses of the portfolio is valid for the period of detention.
- The level of confidence.
- The holding period of assets.

If the distribution of assets during the following normal distribution yields the following result:

$$P[P_{t,0} \leq VaR(t,\alpha)] = \alpha \Rightarrow P\left[\frac{P_{t,0} - E(P_{t,0})}{\sigma(P_{t,0})} \leq \frac{VaR(t,\alpha) - E(P_{t,0})}{\sigma(P_{t,0})}\right] = \alpha$$

$$\Rightarrow \frac{VaR(t,\alpha) - E(P_{t,0})}{\sigma(P_{t,0})} = z_\alpha \Rightarrow VaR(t,\alpha) = E(P_{t,0}) + \sigma(P_{t,0}) \times z_\alpha \tag{7}$$

2.2.2 Portfolio risk

Portfolio risk is measured by one of the measuring elements mentioned above.It depends on three factors namely:

- The risk of each action included in the portfolio
- The degree of independence of changes in equity together
- The number of shares in the portfolio

For a portfolio of fewer k shares $n_1, \ldots, n_k$. If we assume that the distribution of the value of this portfolio V follows the multivariate normal, then the VaR of the portfolio is given by:

$$VaR(t,\alpha) = -N\mu_t + z_\alpha.\sqrt{N^t\Omega_t N} \tag{8}$$

where : $\mu_t = E(\Delta V)$, $\Omega_t = \sigma(\Delta V)$, $N = (n_1, \ldots, n_k)$ and z_α is the quantile of order of confidence level α.

3 Classification

Classification [17], [18] is a research area that constitutes the basis of several systems with decision support. The objective of classification is to assign a set of objects to a set of classes according to the properties it checks. There are two methods of classification [19] method of automatic classification and allocation method. The methods of automatic classification [20] are methods based on unsupervised learning and are used to organize objects in the same class unifying. The purpose of this method is to form homogeneous classes constructed from objective criteria. Allocation methods [21]: these are methods based on supervised learning using a set of examples where the classes of membership are already known and used to define standards of employment.

4 Genetic algorithms

A genetic algorithm was originally developed by John Holland [17]. It is an algorithm iterative for finding optimum, it manipulates a population of constant size. This population is composed of candidate points called chromosomes. The constant size of the population leads to a phenomenon of competition between chromosomes. Each chromosome represents the encoding of a potential solution to the problem to be solved, it consists of a set of elements called genes, which can take several values belonging to an alphabet which is not necessarily digital [18]. At each iteration, called generation, a new population is created with the same number of chromosomes. This generation consists of chromosomes better "adapted" to their environment as represented by the selective function. As in generations, the chromosomes will tend towards the optimum of the selective function. The creation of a new population base on the previous one is done by applying the genetic operators that are: selection, crossover and mutation. These operators are stochastic [19]. The selection of the best chromosomes is the first step in a genetic algorithm. During this operation the algorithm selects the most relevant factors that optimize the function. Crossing permits two chromosomes to generate new chromosomes "children" from two "parents" chromosomes selected. The mutation makes the inversion of one or more genes of a chromosome [19].
Figure 1 illustrates the various operations involved in a basic genetic algorithm:

Random generation of initial population
Calculation of the selective function
Repeat
Selection
Crossing
Mutation
Calculation of the selective function
Until stopping criterion satisfaction

Fig. 1: Basic genetic algorithm

5 Optimization method

5.1 Classification Algorithm

Consider a portfolio of k actions characterized by the following:

- The expected returns $\overline{r_i}$;
- Values at risk VaR_i;

These elements are arranged in a matrix $MaT = [(\overline{r_1}, VaR_1), \ldots, (\overline{r_n}, VaR_k)]$ Our approach is initially to classify these elements by using the K-Means as follows:

1. Initialization: Choose initial $m_j(1)$ arbitrary center of the Vector $Vect$.
2. Assignment: iteration i, the element x is assigned to w_j, if:

$$\|x - m_j(i)\| = min_{l-1}^{\iota}\|x - m_l(i)\| \tag{9}$$

   All samples are classified according to this rule (the closest center)

3. Update centers:

   - Design of new centers $m_j(i+1)$ to minimize the squared error:

$$J_j = \sum_{x \in W_j} \|x - m_j(i+1)\|^2 \tag{10}$$

   - By canceling the derivative of this expression with respect to yields:

$$\frac{\partial J_j}{m_j} = -2 \sum_{x \in W_j} (x - m_j) = 0 \tag{11}$$

   Hence the optimum value m_j for iteration (i+1):

$$m_j(i+1) = \frac{1}{n_j} \sum_{x \in W_j} x \tag{12}$$

4. Convergence test: If $\forall j, m_j(i+1) = m_j(i)$ the end. If no return to Step 2.

The process is thus repeated until a stable state where no improvement is possible. Once the classification of the matrix is complete, the selected class (under portfolio) that has the greatest expected return and the lowest average VaR is then applied to the optimization algorithm MinVaRMaxVal in this portfolio.

5.2 Optimization algorithm MinVaRMaxVal

Our algorithm for optimizing dynamic portfolio consists at first time minimize the risk to an initial value of portfolio under certain additional constraints and maximize the value of a portfolio dynamically so that the portfolio value obtained is greater than that obtained in the first stage, and the resulting VaR is less than that adopted in the first stage, in addition to some additional constraints, using genetic algorithms.

5.2.1 Initialization

The population is a set of chromosomes which are composed of k genes representing $n_i (i = 1, \ldots, k)$ numbers, which n_i is the number of wealth invested in the action i. This population is initially randomly using real code.

n_1	n_2	$\cdots\cdots$	n_k

Fig. 2: Structure of chromosome

5.2.2 Evaluation Function

The following operation is the evaluation of chromosomes generated by the previous operation by an evaluation function (fitness function), while the design of this function is a crucial point in using GA. The fitness functions using in this work are:

- In the case of minimization, we use: $f = VaR(\alpha, t)$
- In the case of maximization, we use: $g = \Delta V(n)$

5.2.3 Operations of selection

After the operation of the assessment of the population, the best chromosomes are selected using the wheel selection that is associated with each chromosome a probability of selection, noted, P_i.

- for minimisation problem:

$$P_i = \frac{1}{N-1}\left(1 - \frac{f_i}{\sum_{i \in Pop} f_i}\right) \tag{13}$$

- for maximisation problem :

$$P_i = \frac{g_i}{\sum_{j \in Pop} g_j} \tag{14}$$

Each chromosome is reproduced with probability. Some chromosomes will be "more" reproduced and other "bad" eliminated.

5.2.4 Operations crossing

After using the selection method for the selection of two individuals, we apply the crossover operator to a point on this couple. This operator divide each parent into two parts at the same position, chosen randomly. The child 1 is made a part of the first parent and the second part of the second parent when the child 2 is composed of the second part of the first parent and the first part of the second parent.

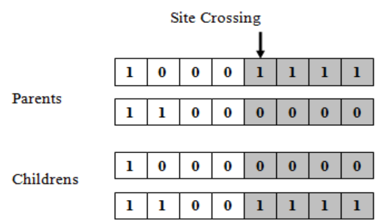

Fig. 3: operation at a crossing point

5.2.5 Operation of mutation

This operation gives to genetic algorithms property of ergodicity which indicates that it will be likely to reach all parts of the state-owned space, without the travel all in the resolution process. This is usually to draw a random gene in the chromosome and replace it with a random value.

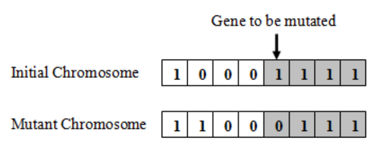

Fig. 4: Mutation Operation

5.2.6 Conditions for Convergence

At this level, the final generation is considered. If the result is favorable then the optimum chromosome is obtained. Otherwise the evaluation and reproduction steps are repeated until a certain number of generations, until a defined or until a convergence criterion of the population are reached.

5.2.7 Schema of dynamic optimization MinVaRMaxVal

Notes : $V_0 = \lambda C_0$ is the value of portfolio expected witch C_0 that the initial capital of the investor and λ is the minimum rate of return expected by investors.x' is the transposed of x and is the amount of maximum loss set in advance.

6 Results

Consider a portfolio containing 48 shares of the Casablanca exchange taken monthly from 01/01/2008 to 01/01/2010. The distribution of these actions is normal. After calculating the matrix of expected returns and VaR of these actions $MaT = [(\overline{r_1}, VaR_1), \ldots, (\overline{r_n}, VaR_k)]$, we move to the classification stage of matrix elements to extract the classes (in portfolios) homogeneous (containing class actions whose expected returns and VaR are very close together), then the class that holds the greatest expected return and the lowest average VaR, which contains 15 actions in this case. The following figure shows the classes obtained by this step.

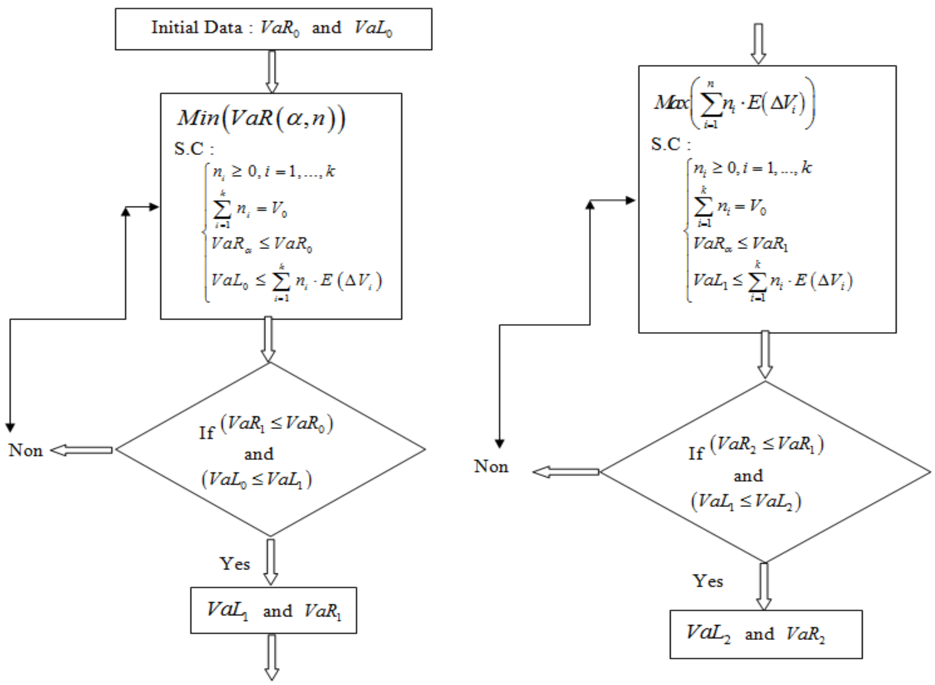

Fig. 5: Diagram of the algorithm MinVaRMaxVaL

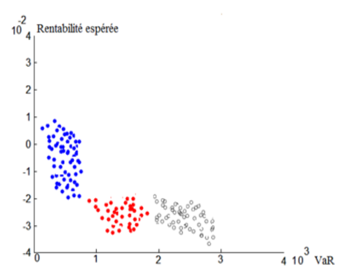

Fig. 6: Classes identified by the classification method K-Means.

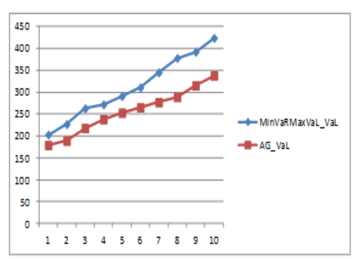

Fig. 7: Graphical representation of portfolio value with genetic algorithms and the algorithm MinVaRMaxVaL.

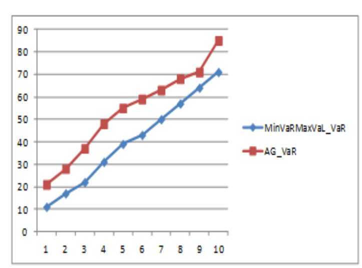

Fig. 8: Graphical representation of the VAR with genetic algorithms and the algorithm MinVaRMaxVaL.

Next, we apply the algorithm of dynamic optimization MinVaRMaxVaL on this retained class. This algorithm consists in first step on minimizing the VaR under certain constraints and in a dynamic way on maximizing the value of portfolio under other constraints as shown in the algorithm MinVaRMaxVaL. From Figure 6 and Figure 7, we see that the values of the VaR (Val, respectively) obtained by the algorithm MinVaRMaxVaL are lower (respectively higher) than those obtained by genetic algorithms for the same portfolio actions. Furthermore, according to Figure 8 the VaR obtained by MinVaRMaxVaL in portfolio are lower than initial portfolio and from Figure 9 the values obtained by holding MinVaRMaxVaL in portfolio are higher than initial portfolio. These comparisons are favorable to our algorithm and show the performance of it.

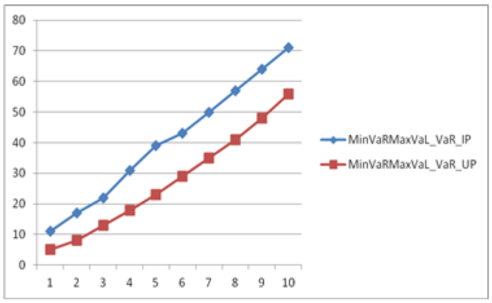

Fig. 9: Graphical representation of the initial portfolio (IP) VaR and under portfolio (SP) using MinVaRMaxVaL.

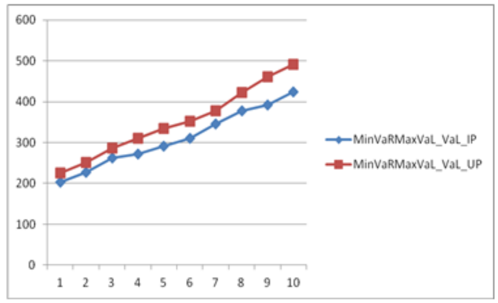

Fig. 10: Graphical representation of the initial portfolio (IP) value and under portfolio (UP) using MinVaRMaxVaL.

7 Conclusion

In this paper we presented an approach for an optimal choice of a reduced size portfolio. This approach consists in a first step on classifying the actions of classes known as under portfolio using the algorithm K-Means. In a second step, we apply the dynamic optimization algorithm MinVaRMaxVaL On the class (under portfolio) that has the highest expected return and the lowest average VaR. The algorithm MinVaRMaxVaL takes place in two stages: The first stage is to minimize the risk to a constraint limiting the portfolio value,in the second one we seeks to maximize the value of the portfolio dynamically so that the result shall be greater than the portfolios value fixed in the first stage and the risk arising in this step are lower than that obtained in the first step. The obtained simulation results were presented. They were generally satisfying and show the validity of our approach.

References

[1] Markowitz, H.(1952). Portfolio selection. Journal of Finance, fvrier 1952.

[2] Sharpe (1967), A Linear programming Algorithm for Mutual fund Portfolio Selection. Management Science,13,p.499-510.

[3] Sharpe (1971), A Linear Programming Approximation for General Portfolio Selection. Problem , J.Quantitative Anal ,6,p.1263-1275.

[4] Stone (1973) , A Linear Programming Formulation of the General Portfolio Selection. Problem , J.Quantitative Anal ,6,p.107-123.

[5] Rudd et Rosenbeg (1979), Realistic Portfolio Optimisation. TIMS Studies in the Management Science,11,p.21-46.

[6] Konno (1990), Piecewise Linear Risk Function and Portfolio Optimization. Journal of Operation Research, Society of Japan , 33,p.139-156.

[7] Konno et Yamazaki(1991), A Mean Aboluate Deviation Portfolio Optimisation Model and Application to Tokyo Stock Market. Management Science,37,p.519-531.

[8] Zenios et Pang (1993), A Mean Aboluate Deviation Portfolio Optimisation for Mortgage Backed Securities. Annals of Operations Research,45,9.433-450.
[9] Speranza(1993), Linear Programming Models for Portfolio Optimisation. Finance,14,p.107-123
[10] Yoshimoto (1996), The Means Approch to Portfolio Optimisation Subject to Transaction Costs. Journal of the Operation Research, Society of Japan,p.99-117
[11] Hamza. F. et Janssen. J. (1995) 'Portfolio Optimization Model Using Asymmetric risk Functions' Actuarial Approach for Financial Risks, Vol III Bis. p. 3-32.
[12] Jorion.P "Value at Risk"; The New Benchmark for Managing Financial Risk, 2nd Edition McGraw-Hill.(2000)
[13] Beder T.; "VaR : seductive but dangerous" Financial Analysts Journal 51, pages 12-24(1995).
[14] Hendricks D, "Evaluation of value at risk models using historical data" Economic Policy Review, Federal Reserve Bank of New York, Issue April, pages 39-69.(1969)
[15] Esch, Kieffer, Lopez. "Value at Risk , vers un risque management moderne" De Boeck universit.
[16] P. Thibaut. Rapport sur la classification des formes donde dans PISTACH. Technical report, Collecte Localisation Satellites, 2008.
[17] C. Saint-Jean and C. Frlicot. Une mthode paramtrique et robuste de classification semi-supervise avec rejet, pages 85_100. Dans les actes de la Confrence sur l'apprentissage (CAP' 2001), 2001.
[18] C. Ambroise. Approche probabiliste en classification automatique et contraintes de voisinage. PhD thesis, Universit de Technologie de compigne, France, 1996.
[19] C. Biernacki. Choix de modles en classification. PhD thesis, Universit de technologie de Compigne, 1997.
[20] M. Van Dang. Classification de donnes spatiales : Modles probabilistes et critres de partitionnement. PhD thesis, Universit Technologique de compiegne, Dcembre 1998.
[21] Hamad, D., El Saad, S. et Postaire, J.G. "Algorithmes d'apprentissage comptitif pour la classification automatique", TISVA98, Oujda, Maroc, 27-28 April 1998, pp. 159-164
[22] Lutton, E., Algorithmes gntiques et Fractales. Dossier d'habilitation diriger des recherches, Universit Paris XI Orsay, 11 Fvrier 1999.
[23] Cantu-Paz E., Efficient and Accurate Parallel Genetic Algorithms, Kluwer Academic Publishers, 2000.

Received: April, 2012