A CVaR Model for ETF Portfolio Selection

Ciro Chiappini
Imperial College London

Flavia Cortelezzi[*] [†]
Insubria University

Gaia Serraino
Risk Management MBSD, GSD, NYPC - DTCC - London

March 20, 2012

Abstract

In recent years, the ETFs market has seen an impressive growth among investors. In this paper, we propose a mixed integer linear programming (MILP) model dealing with the ETF portfolio selection in a single investment periods. Our approach allows to optimize the portfolio minimizing CVaR as a measure of risk for a given required return, considering the level of transaction costs and minimum round lots. The model has been tested on several different realizations of Monte Carlo simulated scenarios. Results show a coherent behaviour of the model applied to ETFs and consistency in portfolio selection.

Key Words: ETF, Portfolio Optimization, Integer Programmig, Minimum Transaction Lots.

1 Introduction

In the last few years, the Exchange-Traded Funds (or ETFs) market has seen an impressive growth among investors, stimulating the creation of new benchmark. ETFs are investment funds, traded as shares on most global stock exchanges, such as the main European and USA Stock Exchanges. ETFs are characterized by the presence of an issuer that builds the basket of commodities, bonds or stocks and issues certificates each of which represents a percentage of the basket itself. Typically, ETFs try to replicate a stock market index such as the S&P 500 or DJ Index[1]. Moreover, the diversification of the underlying is

[*] *We would like to thank Giuseppe Colangelo, Giovanni Marseguerra, and PierCarlo Nicola for many helpful comments. The usual disclaimer applies. This paper has been realised within the research program "Nonlinear Models in Economics and Finance: Interactions, Complexity and Forecast". Financial support from the Italian Ministero dell'Università e della Ricerca Scientifica e Tecnologica - COFIN-MIUR 2004-2006 is gratefully acknowledged.*

[†] *Corresponding author. E-mail: flavia.cortelezzi@uninsubria.it. Address: c/o Università dell'Insubria - Via Garibaldi, 61 - 22100 - Como.*

[1] In the ETFs we consider there is no active management involved.

usually large and it can also be representative, for example, of a sector of the market (e.g., utilities, pharmaceutical companies, energy sector, blue chips..), of a geographical region, of a set of bond maturities. Most of the ETFs pay dividends periodically and are characterized by competitive management fees with respect to the ones required by mutual funds[2]. Because ETFs are traded on stock exchanges, they can be bought and sold at any time during the day (unlike most mutual funds). Their price will fluctuate from time to time, just like any other stock price, and they are attractive to investors because they offer the diversification of mutual funds with the features of a stock. Thus, popularity of ETFs is increasing, given their flexibility, low costs and the diversification they offer. Various aspects differentiate ETFs from stocks and from mutual funds. In what follows, we try to identify some of them, which justify a separate analysis[3]. First, unlike mutual funds, ETFs are continuously priced throughout the trading day, whereas mutual funds sales take place at the end of trading day. Then, they both need to adjust their holdings in response to changes in the companies included in the underlying index they track. However, fund managers are also confronted with the need to provide liquidity to buyers and sellers of their fund's shares, which requires them to hold a percentage of their assets in cash. ETF managers do not have to face with this problem, because purchases and sales of their funds' shares only take place in the secondary market[4]. Moreove, as long as they do not have to hold cash to provide liquidity, they are able to track an index more closely than a mutual fund. Furthermore, an investor can employ a wider range of trading techniques using them, such as, e.g., stop loss, limit orders and short sales, because ETFs are traded like a stock. Finally, the operating expenses on many ETFs tend to be lower than on mutual funds which track the same index, because ETFs do not provide the same level of service to their owners as those provided by mutual funds (e.g., telephone service centers, free fund transfers, check writing privileges, etc.).

In this paper, we analyse the problem of selecting an ETFs portfolio in a buy and hold strategy, in order to minimize the portfolio Conditinal Value at Risk (CVaR), given a minimum required return and an initial capital endowment. As proved by Artzner et al.(1999), the CVaR[5] is a "coherent risk measure". Although CVaR has not yet become a standard in the financial industry, it is widely accepted in the insurance industry. In our model the investor faces constraints on minimum round lots, and on fixed and proportional costs[6]. Ideally, we can image a provider of one of these trading system that decides to

[2] They require only management fees in the range 0.2%-0.6% according to the capital invested.

[3] For a detailed description of ETFs, see, e.g., www.borsaitaliana.it.

[4] ETFs are closed end funds.

[5] A simple description of the approach for CVaR minimization and optimization problems with CVaR constraints can be found in Uryasev (2000), Krokhmal et al. (2001), Alexander (2004).

[6] As a proxy for the transaction costs, we consider the average of the commission fees required by the main Italian trading online system.

offer to its clients a software that allow them to find the optimal portfolio composition in order to minimize the CVaR over a given horizon and to obtain a minimum required return. The software mentioned could be implemented using the suggested model.

The general methodology adopted in the paper is that of the general portfolio selection problem in a buy and hold strategy (this is the typical strategy used by small investors), whilst our basic assumptions can be contrasted with those of Angelelli et al. (2004), Mansini et al. (1999,2001), Kellerer et al. (2000) and Chang et al. (2000).

The paper is organized as follows. In Section 2 the general model is described. In Section 3 the problem is described. In Section 4 the experiments and the results on historical data relative to 142 ETFs are discussed. Section 5 concludes and remarks on future research are discussed.

2 The Problem Analysis

In this section, we analyse the problem of selecting a portfolio consisting of ETFs in a buy and hold strategy, i.e. the portfolio is bought and held without investments' modification. In this sense, the number of selected ETFs is fixed over a single period. The problem is that of deciding which ETF to select and how much to invest in each of them in order to minimize the risk, measured by CVaR, given a required rate of return, r. The inclusion of transaction costs and minimum lots in the portfolio selection of ETFs constitute essential issues to trackle this kind of problem.

Different aspects should be considered to model ETFs portfolio selection. First, it is usually possible to invest in ETFs every desired amount regardless any initial threshold. However, to avoid trivial solution, we require that a minimum percentage of the total capital must be invested in ETFs. Then, each ETF in the portfolio can be bought in multiples of the minimum lot (1 ETF). Moreover, the capital invested in ETFs is divided into shares quoted from time to time on the market. The quotation of a share depends on the quotation of the underlying stocks composing the ETF. Finally, since commissions constitute a relevant aspect of the final performance, we consider both the fixed and the proportional transaction costs.

3 The Problem Formulation

In this section we provide the mathematical formulation of the problem above described.

Let $J = \{1, 2, ..., N\}$ be the set of ETFs available for the investment. Let $\mathbf{x} = (x_1 ..., x_n) \in \Re^n$ be the decision vector which represents a portfolio of ETFs, where x_j is the number of the minimum lot position in ETF j. Thus, it must than be:

$$x_j \geqslant 0, \quad x_j \in \mathbb{N}$$

Let $\mathbf{q} = (q_1..., q_n) \in \Re^n$ be the price vector of the ETF portfolio $\mathbf{x}$. Thus, the inner product $\mathbf{q}^T\mathbf{x}$ represents the portfolio value. The expected rate of return for each ETF $j \in J$ is represented by a random variable y_j with density $p(y_j)$. Thus, the gross expected returns of a portfolio over a given period is represented by $\mathbf{x} \cdot \mathbf{y}^T$, where the vector $\mathbf{y} = [y_1...y_n] \in \Re^n$ represents the collection of expected return of the ETF j. As the analytical representation of the density funcion $p(\mathbf{y})$ is not available, we can have M scenarios, $\mathbf{y}_k$, $k = 1, ...m$, sampled from the density $f(\mathbf{x}, \mathbf{y})$[7]. The model also takes into account the presence of fixed costs[8] $(f_j \geq 0)$, operational costs (g_j) and proportional costs (c_j) for each transaction over a given ETF lot. Thus, if a number $x_j > 0$ of lots of ETF j is selected, the investor is required to pay $f_j + c_j \cdot x_j + q_j \cdot x_j$; operational costs g_j are impicit in the quoted price q_j.

In order to take such a cost structure into account we introduce a vector of binary variables $\mathbf{z} = [z_1...z_n] \in \Re^n$, defined as:

$$z_j = \begin{cases} 1 & \text{if } x_j > 0 \\ 0 & \text{otherwise} \end{cases}$$

This allows us to define the loss function over a period for each possible portfolio[9]:

$$f(\mathbf{x}, \mathbf{y}) = -\sum_j [(y_{jk} - c_j)\, q_j x_j - f_j z_j] \quad \text{for each } k \tag{1}$$

It is worth noticing that for each $\mathbf{x}$ the loss function $f(\mathbf{x}, \mathbf{y})$ is a random variable having a distribution in $\mathbb{R}$ induced by that of $\mathbf{y}$. The probability of $f(\mathbf{x}, \mathbf{y})$ not exceeding a threshold α is given by

$$\Psi(\mathbf{x}, \alpha) = \int_{f(x,y) \leq \alpha} p(\mathbf{y}) d\mathbf{y}$$

Thus, $\Psi(\mathbf{x}, \alpha)$ is a cumulative distribution function for the loss associated with $\mathbf{x}$[10] as a function of α. It completely determines the behavior of this random variable and it is fundamental in defining CVaR. Given the quantile size β, $0 < \beta \leq 1$,the performance function considered is defined as follows[11]:

[7] The existence of the density funcion is not critical for the considered approach. It is enough to have a code which generates random sample for $p(\mathbf{y})$. A two step procedure can be used to derive analytical expression for $p(\mathbf{y})$ or construct a Monte Carlo simulation code for drawing samples from $p(\mathbf{y})$, as, for instance, in RiskMetrics (1996). See Rockafellar et al. (1999) for a detailed exposition.

[8] It captures the situation in which an individual investor asks a broker to invest money on the stock exchange, paying a fixed sum for the service. The payment might include brokerage fees, information processing costs, or taxes, which are independent of the amount invested in each security.

[9] For a detailed description of the loss function, see Uryasev (2000).

[10] See Rockafellar et al. (1999) for a detailed exposition.

[11] See Uryasev (2000) for a detalied exposition of the crucial features of $F_\beta(\mathbf{x}, \alpha)$.

$$F_\beta(\mathbf{x},\alpha) = \alpha + (1-\beta)^{-1}\int_{y\in\Re^n} [f(\mathbf{x},\mathbf{y})-\alpha]^+\, p(\mathbf{y})dy \tag{2}$$

where $[t]^+ = t$ when $t > 0$ but $[t]^+ = 0$ when $t \leq 0$. Following Rockafellar et al. (2000), we now consider the approximation of $F_\beta(\mathbf{x},\alpha)$ obtained by sampling the probability distribution in $\mathbf{y}$. A sample set $\mathbf{y}_1, \mathbf{y}_2 ... \mathbf{y}_m$ yields the approximate function:

$$\tilde{F}_\beta(\mathbf{x},\alpha) = \alpha + \frac{1}{m(1-\beta)}\sum_{k=1}^{m} [f(\mathbf{x},\mathbf{y}_k)-\alpha]^+ \tag{3}$$

The minimization of (3) is performed in order to get a solution. Expression (3) represents a Mixed Integer Problem (MIP) that can be easily reduced to a Linear Problem (LP)[12] by introducing the auxiliary real variables u_k for $k = 1; ...; m$. The minimization of (3) is therefore equivalent to minimize the following linear expression[13]

$$\tilde{F}_\beta(\mathbf{x},\alpha) = \alpha + \frac{1}{m(1-\beta)}\sum_{k} u_k \tag{4}$$

subject to the linear constraints

$$u_k \geq 0 \tag{5}$$

$$\sum_j [(y_{j,k}-c_j)\, x_j q_{j,}] - \sum_j (f_j z_j) + \alpha + u_k \geq 0 \quad \forall k \tag{6}$$

$$\sum_j [(y_j - c_j)\, x_j q_j] - \sum_j (f_j z_j) - \sum_j (g_j z_j) \geq r \sum_j x_j q_j, \tag{7}$$

$$C_1 \leq \sum_j x_j q_j + \sum_j c_j x_j + \sum_j (f_j z_j) \leq C \tag{8}$$

$$\sum_j z_j \leq M \tag{9}$$

$$z_j l_j \quad \leq \quad x_j \leq z_j u_j \qquad \forall j, \tag{10}$$

$$x_j \quad \in \mathbb{N} \tag{11}$$

$$z_j \in \{0,1\} \tag{12}$$

[12] LP approaches are routinely used in portfolio optimization with various criteria such as mean absolute deviation, maximum deviation and mean regret (see, among the others, Konno et al. (1991), Young (1998), Dembo et al. (1992)).

[13] Several case studies have demonstrated that this formulation provides a very powerful and numerically stable technique which can solve problem with a large number of instruments and scenarios.

The objective function (4) and the first group of constraints (6) determine the minimization of risk. Constraint (7) imposes that the net portfolio mean return has to be greater or equal to the required return (minimum return constraint), where y_j is defineded as the mean of $y_{j,k}$, $k = 1, ...m$. Note that the proportional costs c_j, $j \in J$, defined as percentages of the amount invested in each security j, directly influence the portfolio return. The constraint (8) enstablishes that the capital invested in the portfolio of ETFs must be included between the lower bound, C_1, and upper bound, C. Constraint (9) expresses the "cardinality condition" imposing that the number of ETFs selected in the portfolio cannot be grater than $M \in N$. If an ETF j is selected in the portfolio, then constraint (10) forces the number of stock units for ETF j to be included between l_j and u_j, i.e. respectively the minimum an maximum purchaseable lots for ETF j. Finally, constraint (12) means that the binary variable z_j is set equal to 1 whenever the corresponding security is selected in the portfolio and 0 otherwise[14].

4 Computational Results on Stock Market Data

This section is devoted to the computational analysis of the described model[15]. We have tested[16] the heuristics model on historical data of 142 ETFs[17] quoted with continuity on the international financial market over the time period december 2003- december 2005. The data consist of 13 ETFs from Italy, 11 ETFs from France, 15 ETFs from Germany and 103 ETFs from US Stock Exchanges.

Since in the common financial practice the ETFs performance in the short horizon is assumed to be misleading, we have tested the a portfolio selection in the long run: the quotations of ETFs span a period of two years for a total of 728 daily returns for each ETF in the sample. For the same period of time, we consider the evolution of Eur/Usd foreign exchange rate, that will then be simulated to convert dollar returns into euro. The CVaR model with integer constraints has been solved for different values of β, i.e. $\beta = (0.99, 0.95, 0.9)$. We use scenarios generation[18] to find the optimal portfolios that minimize CVaR on a 1-month trading horizon.

As first, data time series have been analyzed using PcGive descriptive statistics package. Jacque Bera test statistics rejects the null hypothesis of normality of returns for the majority of time series[19]; at the 5% confidence level the null hypothesis has been accepted for 15% of the ETFs, while at 1% confidence level,

[14] Notice that all values are homogeneous monetary values.

[15] The model has been solved by using MATLAB 7.1 and GLPK mixed integer linear programming solution engine on a 3.00 Ghz Pentium 4 with 2 Gb DDR RAM. The maximum computational time has been set equal to 3 hours.

[16] See Jamshidian et al (1997), Glasserman (2003) and Jorion (2005) for methodological issues and implementation of the model.

[17] Data Source: Reuters.

[18] See Glasserman (2003) for details.

[19] The Jarque Bera statistic has a X-squared distribution with two degrees of freedom. Its critical values at the 5% and 1% confidence levels are 5.991 and 9.210 respectively. Therefore, the normality hypothesis is rejected when the JB statistic has a higher value than the corresponding critical value at the respective confidence level. See Riskmetrics (1996).

it has been accepted for 21% of the ETFs. For each time series of returns, we then derive an estimate of expected return, standard deviation and covariance matrix. We finally generate correlated monthly returns of the 142 ETFs and of proportional variations of EurUsd exchange rate. Stochastic processes for ETFs and EurUsd has been approximated under the assumption of constant drift and volatility; the obtained returns are correlated and charatcterized by a normal distribution, as a quite common market practice[20]. Asset returns over an interval of length dt are given by:

$$dS/S = \mu dt + \sigma\epsilon\sqrt{dt} \tag{13}$$

where S is the asset price, μ is the expected rate of return, σ is the volatility of the asset price and ϵ represents a random draw from a standard normal distribution. The optimization problem has been solved for an italian investor who pays lower commisions when buying on Italian market with respect to foreign ones. Fixed transaction costs have been set to 0 Euro for Italian ETFs ($f_j = 0$), 9 Euro for European ETFs($f_j = 9$), and 19 Euro for US ones($f_j = 19$), while proportional costs c_j are 0.2% for Italian ETFs, 0.25% for European ETFs and 0% for US ETFs in order to reflect the actual average propotional costs for the different kinds of ETFs[21]. We use these level of costs as an approximation of the real costs requested by Italian online trading systems. We have assumed the investor to dispose of a capital of $100,000$ Euro of which at least 90% has to be invested. We have considered different values for minimun required monthly return, setting r to 1.0%, 1.5%, 2.0% respectively; the number of simulated scenarios has been set equal to $1,000$ and to $5,000$. In the default setup, lower and upper bounds on the number of each selected ETF have been set equal to 20 and 500 respectively (see Fig. 1). The obtained results show that the more relevant constraint is, as expected, the one on minimum portfolio return, while the portfolio selection appears to be sensitive to capital requirement and lower or upper bounds (see Fig. 2 below). The problem becomes infeasible for $r = 3\%$, nonetheless this is a high monthly return. In fact, in the historical period considered, we observe realized average monthly returns in the range $[-0.10\%; 3.65\%]$ and standard deviations[22] in the range $[0.39\%; 8.41\%]$.

Figure 1 shows that no Italian ETFs have been selected; USA ETFs always constitute the biggest part of the portfolios. This is a quite interesting result considering that we are taking into account the foreign exchange risk. Fixed costs are thus almost irrelevant in the portfolio selection, given that the highest costs are associated with USA ETFs and the lowest with the Italian ones. The number of each selected ETF requires an investment such that the fixed cost represent a negligible part of the necessary capital, thus fixed costs cannot significantly decrease the return on the investment. In addition, USA ETFs have null proportional costs, thus favouring the selection of a big number of the

[20] See Riskmetrics Tec. Doc. for a VaR detailed description.

[21] These assumptions are consistent with the time period considered, namely december 2003 - december 2005.

[22] The availability of longer time series would have made possible a better estimate of the drift and volatility parameters used as inputs in the scenario generation.

same USA ETF. In the default setup, the model built portfolios made by, at most, 7 different ETFs, i.e. at most 5% of the 142 ETFs has been selected. Given that ETFs represent diversified indexes, a selection of portfolios made by at most 7 ETFs and at least 5 is considered acceptable. By requiring a lower value for upper bound u, by lowering u from 500 to 100, it is possible to obtain a more diversified portfolio with up to 9 different ETFs and, coherently, the CVaR and VaR raise since a stricter constraint to the model is posed. Such a restriction on u is not realistic with the considered value of C (see Fig. 3) since the investor is limited in purchasing only up to 100 ETF of a selected kind. When we increased the number of generated scenarios, for given β, r, u_j and l_j, we found portfolios with lower CVaR and VaR, but with very similar chosen ETFs (see Fig. 1). As expected, increasing the minimum return r leads to greater VaR and CVaR (see Fig. 3).

Figure 4 shows that also incrementing β always leads to an increment in the value of VaR and CVaR.

Finally incrementing Computational time for $5,000$ simulations can be greater than 90. Thus the model can be computational expensive but it is well known that GLPK engine is as efficient as CPLEX for linear programming problems, while its efficiency is equal to $1/10$ of the Cplex one for MILP problems.

5 Conclusions

In this paper we consider the problem of the initial selection of an ETF optimal portfolio, minimizing CVaR as a measure of risk. The contribution of this paper is two fold. First, from a methodological point, we solve the problem in a mixed integer linear problem framework Then, we implement the procedure on market data and over a single period analysis. Results show a coherent behavior of the model applied to ETFs and consistency in portfolio selection.

This paper open several possible future directions of research. On the one hand, it would be interesting to consider the portfolio rebalancing problem. On the other hand, the model presented, opportunely modified, can be used to build a more reasonable portfolio, based not only on ETFs, but also on shares, bonds, mutual funds and derivatives, which implies a higher degree of reality as well as a higher difficulty of implementation.

References

[1] Alexander C. (2004). *Risk Measurement*, Wiley Eds.

[2] Angelelli E., Mansini R., Speranza M.G. (2004). A comparison of Mad and Cvar with side constraints. Technical Report n. 238, Department of Quantitative Methods. University of Brescia.

[3] Artzner P., Delbaen F.,Eber J.M., Heath D., (1999). Coherent measures of risk. *Mathematical Finance*, 9, 203-228.

[4] Chang, T.J., Meade, N., Beasley, J.E., and Saraiha, Y.M. (2000). Heuristics for Cardinality Constrained Portfolio Optimisation. *Computers and Operations Res.*, 27, 1271-1302.

[5] Glasserman P. (2003). *MonteCarlo Methods in Financial Engineering*, Springer Verlag New York Inc.

[6] Kellerer H., Mansini R., Speranza M.G., (2000). Selecting portfolios with fixed costs and minimum transaction lots. *Annals of Operations Research* 99, 287-304.

[7] Konno, H. and Yamazaki, H., (1991). Mean-Absolute Deviation Portfolio Optimization Model and its Application to Tokyo Stock Market. , *Management Science,*37, 5, 233-250.

[8] Krokhmal P., Palmquist J., Uryasev S. (2001). Portfolio optimization with conditional value at risk objective and constraints. Research report Ise Dept, Univ of Florida.

[9] Jamshidian, F., ZhuY. (1997). Scenario Simulation: Theory and methodology. *Finance and Stochastic*, 1, 43–67.

[10] Jorion, P. (2005), *Financial risk manager handbook.* Wiley (third edition).

[11] Mansini R., Speranza M.G. (1999). Heuristic algorithms for the portaolio selection problem with minimum transaction lots. *European Journal of Operational Research*, 114, 219-233.

[12] Mansini, R., Ogryczak W., and Speranza, M.G. (2001). LP solvableModels for Portfolio Optimization: a Classification and Computational Comparison. Technical Report TR 01-25, Warsaw University of Technology, Institue of Control and Computational Engineering, Poland.

[13] Markowitz H. (1952). Portfolio selection. *Journal of Finance*, 7, 77-91.

[14] RiskMetrics (1996). *Technical document 4-th edition*. JpMorgan.

[15] Rockafellar R.T., Uryasev S. (2000). Optimization of conditional value at risk. *Journal of Risk*, 2, 21-41.

[16] Uryasev, S. (2000). Conditional Value at Risk: optimization algorithms and applications, *Financial engineering News*.