Stochastic evaluation of robust portfolios based on hierarchical clustering and worst-case scenarios

Paulo Rotela Junior[1*], Luiz Célio Souza Rocha[2], Giancarlo Aquila[2], Edson de Oliveira Pamplona[2], Pedro Paulo Balestrassi[2] and Anderson Paulo de Paiva[2]

[1]Departamento de Engenharia de Produção, Universidade Federal da Paraíba, Campus Universitário I, s/n., 58051-900, João Pessoa, Paraíba, Brazil. [2]Instituto de Engenharia e Gestão de Produção, Universidade Federal de Itajubá, Itajubá, Minas Gerais, Brazil. *Author for correspondence. E-mail: paulo.rotela@gmail.com

ABSTRACT. The objective of this paper is to present a proposal to form robust portfolios using a stochastic efficiency analysis of assets from companies in the Sao Paulo Stock Exchange, focusing on the worst market state. In order to do this, information about the market in all of its phases and information from low market periods were employed in a stochastic efficiency analysis using the Chance Constrained Data Envelopment Analysis method, along with a Hierarchical Clustering approach. Then, the portfolios underwent a capital allocation model to obtain the ideal participation of each share. The portfolios formed in both scenarios were analyzed and compared. The joint application of the approaches supplied with information about the worst market state was able to form robust portfolios that lead to a higher accumulated return in the validation period than portfolios optimized from information about the entire period, and still resulted in portfolios with smaller beta values.

Keywords: stochastic optimization, chance constrained programming, Sharpe approach, worst state, robust optimization.

Otimização de portfólios: avaliação estocástica associada ao agrupamento hierárquico e à abordagem de baixa do mercado

RESUMO. O objetivo deste artigo é apresentar uma proposta para formação de portfólios robustos a partir da análise estocástica da eficiência de ações de empresas negociadas na Bolsa de Valores, Mercadorias e Futuros de São Paulo. Para isso, informações dos ativos em períodos de baixa do mercado foram agrupadas por meio do agrupamento hierárquico, e então submetidas a uma análise estocástica de eficiência por meio do modelo *Chance Constrained Data Envelopment Analysis*. Por fim, para obter a ideal participação dos ativos, estes foram submetidos a um modelo clássico da alocação de capital. Os portfólios formados com o método proposto foram analisados e comparados a outros, formados por diferentes modelos. A utilização do método proposto, abastecido de informações de pior estado do mercado, permitiu a formação de portfólios robustos que apresentaram um maior retorno acumulado no período de validação, o que resultou em portfólios com menores valores beta.

Palavras-chave: otimização estocástica, programação restringida por chances, modelo de Sharpe, período de baixa do mercado, otimização robusta.

Introduction

The Markowitz model was developed in 1952. More than sixty years later, the pioneering classic approach of average-variance is still the main model used in the practice of allocation of assets and portfolio management. It has led to the creation of new academic proposals (Zopounidis, Doumpos, & Fabozzi, 2014). As with investors and in the academic environment, the selection process for investments in risky assets remains a challenge for financial management.

Charnes, Cooper, and Rhodes (1978) developed the concept of Data Envelopment Analysis, DEA. It

is used to evaluate and compare organizational units that use multiple inputs to produce different outputs in a certain period of time (Kao & Liu, 2014). Data Envelopment Analysis is a non-parametric method, which stands out among quantitative modelling that help to make decisions. It is used by managers in many areas, including the financial area (Kao, 2014, Azadi, Jafarian, Saen, & Mirhedayatian, 2015, Rotela Junior et al., 2017).

This concept has been widely discussed and today new variations of the classic DEA models are being created. In some of these variations, uncertain and approximate reasoning can already be considered, as proposed by Azadi et al. (2015) in a

DEA model with Fuzzy coefficients, or the model proposed by Sengupta (1987), which associated Chance Constrained Programming, CCP, proposed by Charnes and Cooper (1963) with the DEA model (Jin, Zhou, & Zhou, 2014, Rotela Junior, Pamplona, Rocha, Valerio, & Paiva, 2015).

According to Kim, Kim, and Fabozzi (2014), and Kim, Kim, Mulvey, and Fabozzi (2015), classic models for portfolio optimization such as those proposed by Markowitz (1952) and Sharpe (1963) cannot be considered robust, since they are very sensitive to small input variations. Thus, researchers and scholars have more recently started to incorporate uncertainties by estimating errors directly in the portfolio optimization process, using mathematical techniques for robust optimization (Fabozzi, Huang, & Zhou, 2010, Kim et al., 2014, Kim et al., 2015).

However, even though some studies such as Bertsimas and Sim (2004) have already confirmed a relationship between the increase of return from a worst-case portfolio and the increase of robustness, Kim et al. (2015) believe that robustness in models of robust portfolios is probably reached by betting systematically on information about worst-case market periods. That is, in the formation of a robust portfolio, bear markets (low volume trading periods) are more relevant than bull markets (high volume trading periods). Thus, information regarding returns of assets on days with worse performance are extremely important to create a robust portfolio (Kim et al., 2015).

The aim of this paper is to present a proposal for the formation of robust portfolios using a stochastic efficiency analysis of assets from companies in the Sao Paulo Stock Exchange, focusing on the worst state market.

This paper has as specific objectives the presentation and use of the Chance Constrained Data Envelopment Analysis model to reduce the search space, considering randomness and uncertainty in the variables, testing its applicability in optimizing portfolios. Then, the CCDEA model along with the Hierarchical Clustering system have been used to optimize a robust portfolio, using input data from low trading periods (a bear market) and comparing it to the same model supplied with complete information. Finally, it presents the importance of market worst-case information to reach a robust performance.

Chance constrained DEA

Most DEA models used in the literature are deterministic and do not consider random errors of input variables and outputs. According to Jin

et al. (2014), generalized randomness in evaluation processes come from errors in data collection.

Sengupta (1987) incorporated stochastic variables in a DEA CCR formulation and then transformed it into an equivalent deterministic model. The formulation of a stochastic DEA model, obviously a non-linear program, is presented in Equation 1, 2, 3 and 4. The *i-th* DMUs, $\hat{x}_i = (\hat{x}_{i1}, \hat{x}_{i2},..., \hat{x}_{ia})^T$ and $\hat{y}_i = (\hat{y}_{i1}, \hat{y}_{i2},..., \hat{y}_{ib})^T$, represent stochastic variables for input and output vectors, respectively, and $i = 1, ..., n$.

$$\max E = \sum_{q=1}^{b} u_q \hat{y}_{oq} \text{ Subject to:} \tag{1}$$

$$E(\sum_{p=1}^{a} v_p \hat{x}_{op}) = 1 \tag{2}$$

$$P_r \left(\frac{\sum_{q=1}^{b} u_q \hat{y}_{iq}}{\sum_{p=1}^{a} v_p \hat{x}_{ip}} \leq \beta_i \right) \geq 1 - \alpha_i \quad i = 1, 2, ..., n \tag{3}$$

$$u_q, v_p \geq 0 \tag{4}$$

where:

$u_1, ..., u_b, v_1, ..., v_a$ are weights to be estimated by the model. The symbols u_q and v_p represent the weights of multipliers related to the *q-th* output and *p-th* input, respectively. P_r represents a probability and the superscript `^` indicates that $\hat{x}_{ip}$ and $\hat{y}_{iq}$ are random variables. For the constraints, the model formulates a proportion of being less than or equal to β_i, which represents the expected efficiency level for the *i-th* DMU, which, according to Jin et al. (2014), has a variation [0,1] that is defined as an aspiration level. α_i is a risk criterion referring to the utility of a decision maker. $1-\alpha_i$ indicates the probability of reaching the demand, which is considered one level of confidence (Jin et al., 2014). As with β_i, the risk criterion (α_i) is a value measured in the interval between 0 and 1.

In order to obtain a computer-viable model, the formulation must be rewritten according to the proposal by Charnes and Cooper (1963), which considers randomness. The stochastic variable $\hat{x}_{ip}$ for each input can be expressed as $\hat{x}_{ip} = \bar{x}_{ip} + a_{ip} \xi$, in which p has variation [1,b] and i has variation [1,n]. $\bar{x}_{ip}$ is the expected value for $\hat{x}_{ip}$ and a_{ip} is the standard deviation. In a similar manner, the

stochastic variable $\hat{y}_{iq}$ for each output can be expressed as $\hat{y}_{iq} = \bar{y}_{iq} + b_{iq}\,\xi$, in which q has variation [1,a] and i has variation [1,n]. $\bar{y}_{iq}$ is the expected value and b_{iq} is the standard deviation. Jin et al. (2014) argue that, since part of the stochastic disturbance indicates that errors come from data collection, it is natural to suppose that the random variable ξ follows a normal distribution (N $(0,\sigma^2)$).

The equivalent deterministic formulation of the model should be derived in order to facilitate the resolution of the model. The formulation of the objective function, represented by Equation 1, can be rewritten as Equation 5:

$$E\sum_{q=1}^{b}u_q\,\hat{y}_{oq} = \sum_{q=1}^{b}u_q\,\bar{y}_{oq} \tag{5}$$

The constraints represented in Equation 2 and 3, including the stochastic process, can be rewritten in Equation 6 and 7:

$$E(\sum_{p=1}^{a}v_p\,\hat{x}_{op}) = (\sum_{p=1}^{a}v_p\,\bar{x}_{op}) = 1 \tag{6}$$

$$P_r\left(\frac{\sum_{q=1}^{b}u_q\,\hat{y}_{iq}}{\sum_{p=1}^{a}v_p\,\hat{x}_{ip}} \le \beta_i\right) = P_r(\sum_{q=1}^{b}u_q\,\hat{y}_{iq} - \beta_i\sum_{p=1}^{a}v_p\,\hat{x}_{ip} \le 0) \tag{7}$$

$$\ge 1-\alpha_i \quad i=1,2,...,n$$

Equation 7 can be written in its equivalent form, according to Equation 8:

$$P_r\left(\frac{(\sum_{q=1}^{b}u_q\,\hat{y}_{iq} - \beta_i\sum_{p=1}^{a}v_p\,\hat{x}_{ip} \le 0) - E_i}{\sqrt{V_i}} \ge \frac{E_i}{\sqrt{V_i}}\right) \tag{8}$$

$$\ge 1-\alpha_i \quad i=1,2,...,n$$

In which E_i and V_i indicate the mean and variance of each random variable. Thus, they can be represented according to Equation 9 and 10:

$$E_i = \sum_{q=1}^{b}u_q\,\bar{y}_{iq} - \beta_i\sum_{p=1}^{a}v_p\,\bar{x}_{ip} \tag{9}$$

$$V_i = (\sum_{q=1}^{b}u_q\,\bar{y}_{iq} - \beta_i\sum_{p=1}^{a}v_p\,\bar{x}_{ip})^2\sigma^2 \tag{10}$$

In this manner, the random variable $\dfrac{(\sum_{q=1}^{b}u_q\,\hat{y}_{iq} - \beta_i\sum_{p=1}^{a}v_p\,\hat{x}_{ip} \le 0) - E_i}{\sqrt{V_i}}$ follows a normal distribution of average zero and variance one. Equation 8 can be presented as Equayion 11:

$$\frac{-E_i}{\sqrt{V_i}} \ge \Phi^{-1}(1-\alpha_i) \quad i=1,2,...,n \tag{11}$$

The equivalent form is presented as Equation 12:

$$\sum_{p=1}^{a}v_p\,\beta_i(\bar{x}_{ip} + \Phi^{-1}(1-\alpha_i)a_{ip}\,\sigma) - \sum_{q=1}^{b}u_q \tag{12}$$

$$(\bar{y}_{iq} + \Phi^{-1}(1-\alpha_i)b_{iq}\sigma) \ge 0_i \quad i=1,2,...,n$$

Finally, the model of multipliers can be addressed, differentiating from the proposal of Jin et al. (2014), which deals with undesirable outputs in DEA stochastic model. In here, Φ represents a function of standard normal distribution, and Φ^{-1} is the inverse of the function. Thus, the original model could be reformulated as the linear programming model, the equivalent of which is presented as Equation 13, 14 and 16 (Rotela Junior et al., 2015):

$$\max \sum_{q=1}^{b}u_q \cdot \bar{y}_{oq} \quad \text{Subject to:} \tag{13}$$

$$\sum_{p=1}^{a}v_p\,\bar{x}_{op} = 1 \tag{14}$$

$$\sum_{p=1}^{a}v_p\,\beta_i(\bar{x}_{ip} + \Phi^{-1}(1-\alpha_i)a_{ip}\,\sigma) - \sum_{q=1}^{b}u_q \tag{15}$$

$$(\bar{y}_{iq} + \Phi^{-1}(1-\alpha_i)b_{iq}\sigma) \ge 0_i \quad i=1,2,...,n$$

$$u_q, v_p, w_r \ge 0 \tag{16}$$

Worst market state approach

It has been reported that the correlation between financial return of assets increases in a bear market. In addition, the stock market is not different, and to make matters worse, the correlation within the stock market has been increasing recently (Kim et al., 2015).

Since returns of assets shares during periods of crisis are more positively correlated, models such as Markowitz (1952) may not be appropriate to protect investors.

In the study proposed by Kim et al. (2015), the authors make the assumption that the market can be divided into several states, each one with its own characteristics. This is not a new assumption and was already made by Turner, Startz, and Nelson (1989) and Schaller and Van Norden (1997). However, the author was not interested in detailing the behavior of each one of the possible states, but in finding an ideal static portfolio.

The proposition suggested by Kim et al. (2015, p. 4) is that "[…] when several states exist in a market, the stochastic portfolio of ideal mean-variance for a risk-averse investor is a robust portfolio, in which its expected return is constant in all states".

Kim et al. (2015) found evidence that an emphasis on extreme left-tailed events (below-average results) results in the construction of portfolios with more robust performance than portfolios built for best case scenarios, as well as average-variance portfolios without state information.

Beta is defined as a measure of volatility or systematic risk, of a security or a portfolio in comparison with the market as a whole (Ross, Westerfield, & Jaffe, 2012). Aware of it, another valuable contribution from the research developed by Kim et al. (2015) is the affirmation that during market crash periods, assets with a low beta reduce the general portfolio risk and offer better returns than the assets that have a higher beta.

Materials and methods

The objective of this research is to analyze the behavior of portfolios set up using efficiency evaluation when there is risk and uncertainty considering only information regarding worst-case market conditions, and compare them to results obtained through the same model when considering information for any other period. In order to do this, the Chance Constrained DEA model will be employed, supplied with information from a case market, along with models such as Hierarchical Clustering and the model proposed by Sharpe (1963), as shown on Figure 1.

In order to build up the sample, the assets had to have information from a long period of time. Then, 61 assets were chosen, which are traded in the Sao Paulo Stock Exchange and are part of the Bovespa Index (Ibovespa).
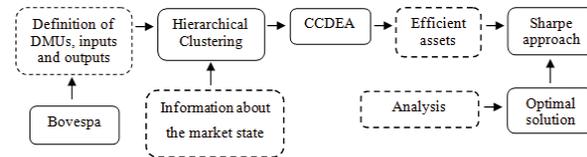
Figure 1. Research flowchart.

Once the assets that made up the sample were selected, a set of input and output indicators to use in the efficiency analysis had to be chosen. As this step, which can be considered independent from the others, the variables presented in the literature were analyzed again. The studies by Powers and McMullen (2000) and Rotela Junior, Pamplona, and Salomon (2014) were used for this purpose. The affirmation by Kim et al. (2015) was also taken into consideration, which stated that in periods of crisis, assets with low beta reduce the portfolio risk and lead to better returns than the assets with high beta.

Then, for this step of the research, model output variables were chosen, such as: use of return, asset profitability, and profit-price (LP). Input variables were adopted, such as beta, price-profit (PL), and volatility. Information collected with the software Economática® correspond to daily data from the months November 2009 to November 2014.

The same strategy proposed by Kim et al. (2015) was adopted to better identify the crisis period. In this study, there was a comparison between portfolios optimized with information for which market states were not defined and portfolios that had been optimized considering a worst-case market. To define when the market was in its worst state, n was defined as four.

Considering information from the whole period, for each DMU (asset), it was possible to calculate the average and variance for each of the variables adopted for efficiency analysis, for each of the scenarios determined, with complete information for the market and worst market state.

According to Johnson and Wichern (2012), grouping is performed based on similarities or dissimilarities (distances). The inputs required are similarity measures or data from which similarities can be computed. Hierarchical clustering techniques could proceed by means of successive mergers. Agglomerative hierarchical methods start with the individual items. The most similar objects are first grouped, and these initial groups are merged according to their similarities (Johnson & Wichern, 2012).

Hierarchical Clustering was used, grouping DMUs by degree of similarity, taking into consideration, in this research, the averages and variances for all variables selected for the model.

This initiative helped to form the group, significantly increasing the degree of similarity. Thus, the restriction about the minimum number of DMUs required by CCDEA model has been attended. The literature recommends, however, that DMUs number is equal to three times the total number of input and output variables (Rotela Junior et al., 2014). Figure 2 shows the grouping of the DMUs when considering all information from assets between 2009 and 2014. Figure 3 shows the grouping of the DMUs when information collected in the worst case market is analyzed.
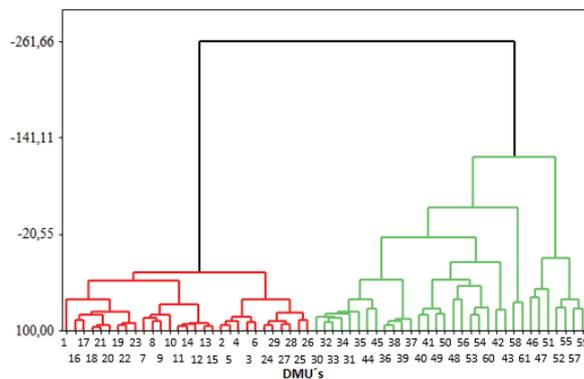
Figure 2. Dendogram of grouping using Hierarchical Clustering, considering all market information.
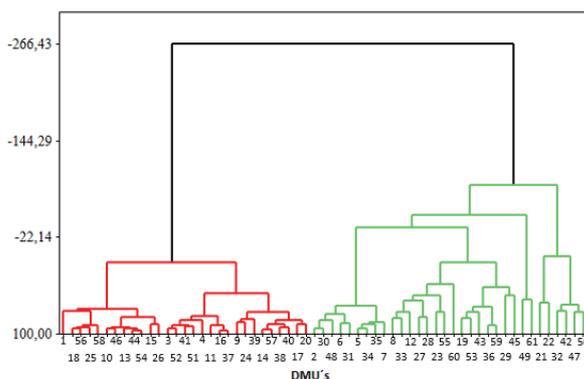
Figure 3. Dendogram using Hierarchical Clustering considering only information from the worst market state.

Therefore, for each of the proposed scenarios, the whole market and the worst state market, two groupings were made. Based on this analysis, they behaved in a more similar manner than when considered to be whole group, which is in accordance to Johnson and Wichern (2012).

Considering information from the whole market, Table 1 presents the descriptive statistics for DMUs that make up group 1 and group 2, respectively. In this step, the codes were presented for assets involved in the efficiency analysis. In addition, considering information from the worst case market, Table 2 presents the descriptive statistics for DMUs that make up group 1 and 2 respectively.

Table 1. Descriptive statistics for group 1 and group 2 from the whole market.

	Return		Liquidity		LP		Beta		PL		Volatility	
	μ_1	σ_1^2	μ_2	σ_2^2	μ_3	σ_3^2	μ_4	σ_4^2	μ_5	σ_5^2	μ_6	σ_6^2
Group 1												
Mean	0.07	3.52	0.65	0.09	7.03	9.19	0.63	0.01	19.11	132.21	1.87	0.09
Median	0.07	3.39	0.59	0.03	6.31	1.89	0.61	0.01	19.49	18.88	1.77	0.08
Stdev	0.03	0.78	0.26	0.22	3.13	18.73	0.18	0.01	9.55	327.64	0.25	0.06
Minimum	0.01	2.17	0.29	0.00	2.55	0.18	0.30	0.00	1.00	1.98	1.45	0.01
Maximum	0.13	4.79	1.30	1.19	14.42	92.35	0.99	0.02	42.94	1706.57	2.31	0.27
Group 2												
Mean	-0.02	6.02	1.62	0.27	3.36	286.57	1.05	0.01	12.27	3365.30	2.39	0.19
Median	-0.02	5.41	1.04	0.10	6.32	31.91	1.06	0.01	10.24	1393.75	2.33	0.14
Stdev	0.06	2.29	1.63	0.59	10.38	650.30	0.22	0.01	15.98	4451.30	0.45	0.15
Minimum	-0.17	3.10	0.35	0.01	-26.73	0.54	0.59	0.00	-30.46	2.07	1.70	0.03
Maximum	0.08	10.77	7.05	2.59	17.39	3173.20	1.50	0.05	47.87	14978.73	3.29	0.60

Table 2. Descriptive statistics for group 1 and group 2 from the worst case market.

	Return		Liquidity		LP		Beta		PL		Volatility	
	μ_1	σ_1^2	μ_2	σ_2^2	μ_3	σ_3^2	μ_4	σ_4^2	μ_5	σ_5^2	μ_6	σ_6^2
Group 1												
Mean	-0.98	3.15	0.65	0.09	7.12	9.46	0.63	0.01	19.39	123.06	1.86	0.09
Median	-0.97	3.12	0.60	0.03	6.32	2.12	0.60	0.01	19.27	18.35	1.77	0.07
Stdev	0.29	0.70	0.26	0.22	3.13	20.24	0.18	0.01	8.48	318.48	0.24	0.06
Minimum	-1.52	1.95	0.29	0.01	2.62	0.20	0.29	0.00	7.77	2.14	1.45	0.01
Maximum	-0.46	4.53	1.31	1.19	14.61	102.62	0.98	0.02	40.20	1686.56	2.29	0.23
Group 2												
Mean	-1.86	4.42	1.62	0.31	3.25	308.54	1.05	0.01	12.70	3152.21	2.37	0.20
Median	-1.78	4.12	1.04	0.10	6.44	35.82	1.06	0.01	10.61	1127.46	2.31	0.14
Stdev	0.39	1.93	1.63	0.69	10.92	692.02	0.22	0.01	18.73	4112.77	0.44	0.15
Minimum	-2.79	2.01	0.36	0.01	-29.27	0.58	0.60	0.00	-30.61	1.89	1.70	0.03
Maximum	-1.14	9.85	7.10	3.17	18.03	3312.91	1.51	0.05	65.25	13506.93	3.25	0.64

Negative data were transformed, with a value being added that changed the most negative value of the series into a positive number without altering the efficiency analysis. This strategy was adopted by Cook and Zhu (2008) and Rotela Junior et al. (2015). For achieving such transformation, for each variable or column the positive value must be added, which makes positive the most negative value in the series.

It is worth mentioning that in this step of the research, input and output variables were considered independent.

A value equal to 1 was used for efficiency level (β_i). For the data analyzed, a good discrimination range for units of analysis is obtained when the risk criterion (α_i) varies between 0.5 and 0.6. This range may vary according to data evaluated by the CCDEA model. To estipulate the risk criterion range, the efficiency analysis through the CCDEA model was performed, and revealed that when the risk criterion is greater than 0.6, all assets are given as efficient;

and when using a smaller value than 0.5, no asset is given as efficient.

The variation within the range set up in the previous step exists in order to insert the risk-averse investor. Specifically for this step of the research, a variation of 0.02 was defined as ideal for processing the variation of probability of restriction $(1-\alpha_i)$; so, due to this variation, six portfolios will be generated for each state.

To analyze the results, Capital Asset Pricing Model (CAPM) was used. CAPM was presented by Sharpe (1964) to identify the existence of abnormal returns (revenue). In addition, the Sharpe Ratio (S_R) was used, which is the most-used metric. Sharpe ratio can be defined as the ratio between the mean and the standard deviation of the expected excess return of investment opportunity (Schuster & Auer, 2012). Therefore, defined as a ratio of reward-to-risk, it has been adopted by many authors to evaluate portfolio performance (Auer & Schuhmacher, 2013).

The software Economática® was also used to validate daily information between the periods of November 2014 and April 2015. In order to do that, accumulated return was calculated in the validation period for each portfolio according to the participation defined by the optimization models.

Results and analysis

Efficiency for the proposed groups was analyzed considering the risk criteria (α_i) previously adopted in Equation 15. Table 3 presents the results of descriptive statistics of efficiency for groups 1 and 2, respectively. The CCDEA model represented by Equation 13, 14 and 16 was applied, with different probability levels $(1-\alpha_i)$ to fulfill the restrictions of the model (see Equation 15), supplied with information from the whole state of the market in the stipulated period.

Table 3. Descriptive statistics of Efficiency for groups 1 and 2 from the whole market.

Group 1						
$(1-\alpha_i)$	40%	42%	44%	46%	48%	50%
Mean	1.25	1.13	1.03	0.96	0.90	0.85
Median	1.24	1.11	1.02	0.96	0.92	0.87
Standard Deviation	0.17	0.15	0.14	0.13	0.12	0.11
Minimum	0.97	0.88	0.80	0.73	0.68	0.66
Maximum	1.56	1.40	1.27	1.17	1.08	1.00
Group 2						
Mean	2.25	1.93	1.69	1.34	1.10	0.82
Median	1.80	1.59	1.37	1.11	0.96	0.82
Standard Deviation	2.12	1.84	1.71	1.16	0.81	0.15
Minimum	1.10	0.96	0.83	0.66	0.55	0.47
Maximum	13.53	11.70	10.81	7.52	5.39	1.00

In the same manner, Table 4 presents the results of descriptive statistics of efficiency for groups 1 and 2, respectively. The CCDEA model was used with different probability levels $(1-\alpha_i)$ to fulfill the restrictions of the model (see Equation 15). However, in this step of the research, the CCDEA model (see Equation 13, 14, 15 and 16) was supplied with information from the worst-case market. It is worth mentioning that a reduction of risk criterion (α_i) leads to an increase in the probability of processing the optimization model restrictions, making the model more rigorous. Thus, a fewer assets are classified as efficient.

Table 4. Descriptive statistics of Efficiency for groups 1 and 2 from the worst case market.

Group 1						
$(1-\alpha_i)$	40%	42%	44%	46%	48%	50%
Mean	0.96	0.94	0.91	0.89	0.87	0.85
Median	0.97	0.93	0.90	0.87	0.87	0.85
Standard Deviation	0.15	0.14	0.14	0.13	0.12	0.12
Minimum	0.70	0.69	0.68	0.66	0.66	0.65
Maximum	1.22	1.17	1.13	1.08	1.04	1.00
Group 2						
Mean	1.24	1.10	0.99	0.89	0.81	0.75
Median	1.08	0.97	0.91	0.86	0.81	0.76
Standard Deviation	1.00	0.77	0.57	0.39	0.25	0.21
Minimum	0.39	0.38	0.37	0.32	0.35	0.31
Maximum	6.28	4.94	3.66	2.47	1.37	1.00

The hierarchical grouping analysis was performed in order to group assets that had a certain degree of similarity so that there would be better discrimination of DMUs. When such assets are grouped, divergence is decreased among the CCDEA model restrictions. However, the assets found to be efficient in the groups for each market state are gathered and optimized according to a proposal by Sharpe (1963).

Considering information from the whole market for each risk criterion adopted (α_i), the assets considered to be efficient in Table 3 were submitted to the proposal by Sharpe (1963). However, some of them were not considered efficient by the CCDEA model, and therefore did not undergo Sharpe (1963) optimization. Six portfolios were proposed from the variation of risk criterion, which were identified for ease of discussion as Portfolios A, B, C, D, E and F.

As discussed previously, the more rigorous is the CCDEA model, the fewer are the assets that will be considered to be efficient, because there is an increase in probability for processing restrictions. This fact shows that smaller α_i values correspond to higher aversion to risk. Thus, with reduction of risk criterion (α_i), the portfolio tends to be made up of a smaller number of assets.

Six other portfolios (U-Z) were proposed according to the variation of risk criterion (α_i), when the model is supplied only with information from the worst market state. As shown in Table 5 and 6, it

is interesting to observe that when the model was supplied with information from the worst case market, a small number of assets were considered to be efficient. Moreover, when submitted to the Sharpe proposal, few of them were selected for the portfolios, which were identified as Portfolios U, V, X, W, Y, and Z.

Table 5. Results of optimized portfolios by the risk criterion considering the whole market.

	Portfolio					
	A	B	C	D	E	F
Risk criterion (α_i)	60%	58%	56%	54%	52%	50%
Portfolio beta (β)	0.702	0.691	0.703	0.696	0.668	0.616
Expected return	1.07%	1.06%	1.07%	1.07%	1.05%	1.03%
Standard deviation	5.70%	5.92%	5.56%	5.93%	5.42%	5.67%
Return	1.21%	1.30%	2.33%	3.81%	3.24%	3.20%
Sharpe ratio (S_R)	0.024	0.219	0.419	0.642	0.597	0.564
Number of assets	57	53	46	33	19	12
Accumulated return	2.30%	2.49%	5.25%	5.26%	6.49%	6.63%

Table 6. Results of optimized portfolios by the risk criterion considering the worst case market.

	Portfolio					
	U	V	X	W	Y	Z
Risk criterion (α_i)	60%	58%	56%	54%	52%	50%
Portfolio beta (β)	0.458	0.446	0.444	0.432	0.429	0.429
Expected return	0.96%	0.96%	0.96%	0.95%	0.95%	0.95%
Standard deviation	5.22%	5.47%	5.34%	5.75%	5.67%	5.52%
Return	2.10%	1.25%	2.63%	3.70%	3.63%	3.64%
Sharpe ratio (S_R)	0.218	0.228	0.492	0.643	0.640	0.659
Number of assets	11	11	10	9	8	8
Accumulated return	4.97%	4.13%	5.75%	8.46%	8.64%	8.65%

Table 5 presents the adopted risk criterion (α_i), portfolio beta, return results, standard deviation, S_R obtained and number of assets that make up the portfolio for each optimized portfolio from information about the whole market.

Table 6 shows the same information for optimized portfolios considering the worst case market.

The most important at this step of the research is to show the importance of information about the worst-case market scenario for robust optimization of portfolios when optimized through the CCDEA model in conjunction with other techniques. Table 5 and 6 will allow analyses and comparisons to be made between the optimized portfolios from the whole market (portfolios A-F) and those that were optimized from information of the bear market (portfolio U-Z). To make this comparison, portfolios A-F and U-Z will be compared in pairs according to the adopted risk criterion (α_i) (see Equation 15).

A comparison of the S_R results obtained by portfolios A and U shown in Table 5 and 8 reveals that portfolio A had an S_R of 0.024, while portfolio U had a value of 0.218. In both cases, the asset

efficiency was evaluated, considering α_i equal to 60%. After allocation by the Sharpe model, 57 assets were used to make up portfolio A. For portfolio U, only 11 were selected from the initial sample at the end of the optimization process. The same comparison could be made for the others pairs of portfolios. It is noted that portfolios optimized from bear market information have better performance when measured by the S_R, and are composed of less assets.

It is worth mentioning that optimized portfolios from historical data of bear market periods had better results in the Sharpe Ratio (S_R) than those optimized from whole market information for different adopted risk criterion values (α_i).

Table 5 and 6 also present values for expected portfolio return (restitution), which were calculated according to what was shown previously. In order to do this, it was necessary to calculate beta values (β) for each portfolio, also presented in the tables. For portfolios optimized from information about the whole market (A-F), the expected return varied between 1.03 and 1.07% per month. For portfolios optimized from information from bear market periods (U-Z), the expected return was concentrated in the range between 0.95 and 0.96% per month. The non-parametric Mann-Whitney test was applied. Then, the results obtained in such tests, for all pairs of portfolios, were values of P smaller than 0.05. Therefore, it can be affirmed that the abnormal accumulated return of the optimized portfolios with information from worst market state is statistically higher than the accumulated return obtained with optimized portfolios from total market state.

Regarding effective profitability, values of 1.21, 1.30, 2.33, 3.81, 3.24 and 3.20% were found for portfolios A to F, respectively. For portfolios optimized from bear market periods, U to Z, the effective profitability values were 2.10, 1.25, 2.63, 3.70, 3.63 and 3.64%, respectively. This fact demonstrates the existence of abnormal returns when compared to expected profitability, also presented in Table 5 and 6.

The accumulated return of the portfolios could be obtained from information collected within the period selected for validation. A comparison of portfolio pairs was made in which one of them is always optimized from information about the whole market and the other only with information of the worst market state. Figure 4 shows the accumulated return in pairs formed according to the probability level $(1-\alpha_i)$ of processing of restrictions in the CCDEA model.
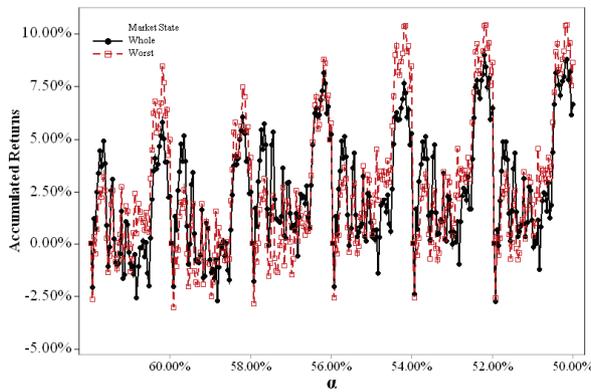
Figure 4. Accumulated return of portfolios defined by pairs using risk criterion.

It is interesting to observe that optimized portfolios from information about bear markets (U-Z) had better S_R values. Nevertheless, beta values for the portfolios, independent of the adopted risk criterion, were smaller than those of portfolios A-F. Kim et al. (2015) affirmed that robust portfolios optimized through stochastic models reach such robustness because they concentrate specifically on information from a crisis period in the financial market. Kim et al. (2015) also believe that robust portfolios tend to be made up of assets that have a low beta value, and such assets tend to behave better than assets that have high beta value, in any classification period.

Conclusion

This proposal does not claim to replace established approaches such as Markowitz (1952) and Sharpe (1963), but to promote a reduction of search space for assets considered to be efficient through stochastic data from different variables.

The variation of the probability level for processing of restrictions $(1-\alpha_i)$ of the CCDEA model meets investor requirements with different attitudes towards risk, from the most conservative investor to the risk taker.

The employment of Hierarchical Clustering allowed the grouping of assets with the same behavior towards the adopted variables, taking into consideration average values as well as variance values.

Another fact worth mentioning is that for each risk criterion adopted, a smaller beta value was obtained by portfolios that were optimized from information about a bear market state than for those supplied with information from the whole period. Portfolios resulting from robust optimization tend to be made up of assets with low beta values, which perform well regardless of market conditions.

Finally, the application of Hierarchical Clustering gave the CCDEA model better data discrimination, even when there were fewer risk criterion, since a reduction was already obtained in the degree of contradiction between model restrictions supplied by the data.

Acknowledgements

References

Auer, B., & Schuhmacher, F. (2013). Performance hypothesis testing with the Sharpe Ratio: The case of hedge funds. *Finance Research Letters, 10*(4), 196-208.

Azadi, M., Jafarian, M., Saen, R. F., & Mirhedayatian, S. M. (2015). A new fuzzy DEA model for evaluation of efficiency and effectiveness of suppliers in sustainable supply chain management context. *Computers & Operations Research, 54*, 274-285.

Bertsimas, D., & Sim, M. (2004). The price of robustness. *Operations Research, 52*(1), 35-53.

Charnes, A., & Cooper, W. W. (1963). Deterministic equivalents for optimizing and satisfying under chance constraints. *Operations Research, 11*(1), 18-39.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research, 2*(6), 429-444.

Cook, W., & Zhu, J. (2008). *Data envelopment analysis: modelling operational processes and measuring productivity*. Worcester, MA: Create Space Independent Publishing Platform.

Fabozzi, F., Huang, D., & Zhou, G. (2010). Robust portfolio: contributions from operations research and finance. *Annals of Operations Research, 176*(1), 191-220.

Jin, J., Zhou, D., & Zhou, P. (2014). Measuring environmental performance with stochastic environmental DEA: The case of APEC economies. *Economic Modelling, 38*, 80-86.

Johnson, R. A., & Wichern, D. W. (2012). *Applied multivariate statistical analysis* (6th ed.). Wisconsin, NJ: Pearson – Prentice Hall.

Kao, C. (2014). Efficiency decomposition for general multi-stage systems in data envelopment analysis. *European Journal of Operational Research, 232*(1), 117-124.

Kao, C., & Liu, S. T. (2014). Multi-period efficiency measurement in data envelopment analysis: The case of Taiwanese commercial banks. *Omega, 47*, 90-98.

Kim, J., Kim, W., & Fabozzi, F. (2014). Recent developments in robust portfolios with a worst-case approach. *Journal of Optimization Theory and Applications, 161*(1), 103-121.

Kim, W., Kim, J., Mulvey, J., & Fabozzi, F. (2015). Focusing on the worst state for robust investing. *International Review of Financial Analysis, 39*, 19-31.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance, 7*(1), 77-91.

Powers, J., & McMullen, P. (2000). Using data envelopment analysis to select efficient large market cap securities. *Journal of Business and Management, 7*(2), 31-42.

Ross, S., Westerfield, R., & Jaffe, J. (2012). *Corporate finance* (10th ed.). New York City, NY: McGraw-Hill Irwin.

Rotela Junior, P., Pamplona, E. O., & Salomon, F. R. (2014) Otimização de portfólios: análise de eficiência. *RAE – Revista de Administração de Empresas, 54*(4), 405-413.

Rotela Junior, P., Pamplona, E. O., Rocha, L. C., Valerio, V. E., & Paiva, A. P. (2015). Stochastic portfolio optimization using efficiency evaluation. *Management Decision, 53*(8), 1698-1713.

Rotela Junior, P., Rocha, L. C., Aquila, G., Balestrassi, P. P., Peruchi, R. S., & Lacerda, L. S. (2017). Entropic data envelopment analysis: a diversification approach for portfolio optimization. *Entropy, 19*(9), 352-362.

Schaller, H., & Van Norden, S. (1997). Regime switching in stock market returns. *Applied Financial Economics, 7*(2), 177-191.

Schuster, M., & Auer, B. (2012). A note on empirical Sharpe Ratio dynamics. *Economics Letters, 116*(1), 124-128.

Sengupta, J. (1987). Data envelopment analysis for efficiency measurement in the stochastic case. *Computers and Operational Research, 14*(2), 117-129.

Sharpe, W. F. (1963). A simplified model for portfolio analysis. *Management Science, 9*(2), 277-293.

Sharpe, W. F. (1964). Capital assets prices: a theory of market equilibrium under conditions of risk. *Journal of Finance, 19*(3), 425-442.

Turner, C., Startz, R., & Nelson, C. (1989). A Markow model of heteroskedasticity, risk, and learning in the stock market. *Journal of Financial Economics, 25*(1), 3-22.

Zopounidis, C., Doumpos, M., & Fabozzi, F. (2014). Preface to the special issue: 60 years following Harry Markowitz's contributions in portfolio theory and operations research. *European Journal of Operational Research, 234*(2), 343-345.