

## Research Article

# Model to Estimate Monthly Time Horizons for Application of DEA in Selection of Stock Portfolio and for Maintenance of the Selected Portfolio

José Claudio Isaias, Edson de Oliveira Pamplona, and José Henrique de Freitas Gomes

*Institute of Production Engineering and Management, Federal University of Itajubá, 37500-903 Itajubá, MG, Brazil*

Correspondence should be addressed to José Claudio Isaias; [joseclaudio.isaias@yahoo.com.br](mailto:joseclaudio.isaias@yahoo.com.br)

Received 24 March 2015; Revised 10 June 2015; Accepted 25 June 2015

Academic Editor: Yan-Jun Liu

Copyright © 2015 José Claudio Isaias et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the selecting of stock portfolios, one type of analysis that has shown good results is Data Envelopment Analysis (DEA). It, however, has been shown to have gaps regarding its estimates of monthly time horizons of data collection for the selection of stock portfolios and of monthly time horizons for the maintenance of a selected portfolio. To better estimate these horizons, this study proposes a model of mathematical programming binary of minimization of square errors. This model is the paper's main contribution. The model's results are validated by simulating the estimated annual return indexes of a portfolio that uses both horizons estimated and of other portfolios that do not use these horizons. The simulation shows that portfolios with both horizons estimated have higher indexes, on average 6.99% per year. The hypothesis tests confirm the statistically significant superiority of the results of the proposed mathematical model's indexes. The model's indexes are also compared with portfolios that use just one of the horizons estimated; here the indexes of the dual-horizon portfolios outperform the single-horizon portfolios, though with a decrease in percentage of statistically significant superiority.

## 1. Introduction

The investing in stocks requires taking into account the expected returns. To minimize the probability of loss, one should select stocks based on a prediction of their likely future earnings. The stock market does permit investors to obtain satisfactory returns, though the market can often be hostile and deliver sometimes severe losses and this return is accompanied by a certain level of risk [1–3].

The problem with capital losses in a stock market is the oscillating nature of a stock's price trajectory. That is, investors are able to occasionally obtain satisfactory returns, over time, even with little knowledge. At other times, however, investors are unable to achieve their target returns even when possessing a high degree of knowledge about the stock or its market sector [4, 5].

To make lucrative investments in the stock market, one can benefit from the support of a properly structured tool.

One such tool is Data Envelopment Analysis (DEA), which has a record of producing good results in the selection of stock portfolios [2, 4, 5].

According to the research that forms the basis of Institute Scientific Information (ISI) Web of Knowledge, no mathematical models, methods, or procedures have been found to create monthly time horizon estimates. Such a tool would be used to collect data to be used in applying a DEA model to select a stock portfolio and to maintain a stock portfolio so selected.

Hence, this study's general objective is to better estimate monthly time horizons. Such an estimate could take two forms: simultaneous and individual. The study aims to first propose a mathematical programming model to estimate three monthly time horizons. Second, the model will estimate, simultaneously, both the monthly time horizons. Third, it will estimate, individually, monthly time horizons, collecting data to be used in the application of a DEA model

as it selects a stock portfolio. And fourth the model will estimate, also individually, the monthly time horizons in the maintenance of a stock portfolio so selected.

## 2. Data Envelopment Analysis

Data Envelopment Analysis (DEA) is a set of mathematical programs that verify efficiency, comparing production units according to their proximity to an efficiency frontier. DEA, first proposed by Charnes et al. in 1978 [6], consists of a technique used to calculate the efficiency of decision maker units (DMUs). DMUs must be components of homogeneous production and it uses equal set of inputs and set of products. The differences between the DMUs consist only in the intensity and magnitude of utilization of inputs and of production of outputs [7].

In 1973, The International Journal of Management Science published the first study on partial productivity measure or efficiency. The article analyzed the productivity of manufacturing units in the USA. The study found problems in analyzing productivity indicators that considered an output type per an input type [8].

According to Charnes et al. [8, 9], in modeling the original DEA problem, the optimal point occurs determining the vector values of the input and output weights, represented by  $v$  and  $u$ . By linearizing the original model, the authors gave rise to the CCR Primal model, which is oriented to inputs according to (1)–(4). It is this model that has been adopted in the current research.

In the set of equations *Effo* is the efficiency of DMU *O* in analysis;  $v_i, u_j$  are the weights of inputs,  $i = 1 \dots r$ , and outputs,  $j = 1 \dots s$ , respectively;  $x_{ik}$  and  $y_{jk}$  are the inputs  $i$  and outputs  $j$  of the DMUs  $k = 1 \dots n$ ;  $x_{i0}$  and  $y_{j0}$  are the inputs  $i$  and outputs  $j$  of the DMU *O* in analysis [10, 11]. Consider

$$\text{Max } Effo = \sum_{j=1}^s u_j \cdot y_{j0} \quad (1)$$

$$\text{s.t. } \sum_{i=1}^r v_i \cdot x_{i0} = 1 \quad (2)$$

$$\sum_{j=1}^s u_j \cdot y_{jk} - \sum_{i=1}^r v_i \cdot x_{ik} \leq 0, \quad \forall k \quad (3)$$

$$v_i, u_j \geq 0, \quad \forall i, \forall j. \quad (4)$$

In the system of equations, (1) maximizes the efficiency of DMU *O* under analysis. Equation (2) constrains the results of the virtual input of DMU *O* under analysis to 1, meaning that it defines the consumption of DMU *O* as a reference for all the others. Equation (3) constrains the difference between virtual outputs and virtual inputs to zero at the maximum. So as to impose the constraint, the efficiency results are equal, at their maximum, to 1. Equation (4) relates to the nonnegativity of vectors  $v_i$  and  $u_j$  [10, 11].

Data Envelopment Analysis encompasses two classical models, the CCR and the BCC. The CCR, named after Charnes, Cooper, and Rhodes in 1978, is also known as

the constant return scale (CRS). This model estimated, using multiple inputs and multiple outputs, the technical efficiency of schools. One characteristic of the model is that it admits only positive data in its applications. The BCC model, named after Banker, Charnes, and Cooper in 1984, is also known as the variable return scale (VRS). This model differs mainly from the CCR in its utilization of the return variable scale [11–13].

## 3. Gaps of DEA Applications in Stocks

DEA has yielded good results in picking stock portfolios, although it is known to have weaknesses. These weaknesses concern its capacity to estimate the monthly time horizon's extension for data collection and to estimate the monthly time horizon's extension of maintenance of the selected portfolio.

According to research in the Institute Scientific Information (ISI) Web of Knowledge, Lamb and Tee [4] came up with a new theory. This theory utilizes the DEA in constructing an index of risk-returns for investment funds. To identify an appropriate form of returns to scale, the authors explored the so-called production possibility set of investment funds. They proposed that measures of risk and return can be combined in a justifiable manner. The theory indicates how to deal with downside risks and identify appropriate sets of measures to avoid them. The research identifies the problem as a result of not treating the diversification in investment funds. Finally, the authors proposed an iterative process to address the problem, a process that in practice proves its efficiency at simulating the utilization.

Pätäri et al. [14] examined the applicability of DEA in using multiple criteria to select stock portfolios. The stock portfolios formed in their study were obtained using three variants of DEA models: the pure CCR, the CCR with super efficiency, and the CCR with cross-efficiency. The results of this study showed that the DEA approach is really able to add value to the selection of stock portfolios. The study compared the superior performance of the main portfolios with both a portfolio of comparable funds from the market in question and the stock market average. In both comparisons, over all the performance measures used, the superiority of the main portfolios was statistically significant. The paper concluded that the DEA is particularly useful for multicriteria applications in the stock market. Therefore DEA holds interesting implications for the practice of portfolio management.

A new methodology for selecting the variables of inputs and outputs for DEA approaches was presented by Edirisinghe and Zhang [1]. Their paper proposed an iterative model of optimization constructed in such a way that it is executed in two stages. The first stage addresses the benefit of utilizing expert information to determine a performance evaluation. The second stage maximizes the correlation with the metric of return of DMUs under analysis. The methodology was applied in fundamental analyses of publicly traded US companies. The aim was to determine a single fundamental indicator based on the proposed DEA model. More than 800 companies from all major sectors of the US stock market were used in the empirical evaluation of the proposed model. The simulation of results of the model's application in companies

proved the methodology's superiority as a decision tool for investment.

According to the research that forms the basis of Institute Scientific Information (ISI) Web of Knowledge, no mathematical models, methods, or procedures have been found to generate monthly time horizon estimates. The purpose of such an approach would be to collect data to be used in applying a DEA model to select a stock portfolio and to maintain a stock portfolio so selected. The lack of such an approach is a result of the fact that when the aforementioned researchers applied some DEA model to the stock market, they left arbitrarily the time horizons of the collection variables and the maintaining of the portfolio. Or they simply utilized third-party extensions.

#### 4. Conception of the Proposed Mathematical Model

The step-by-step procedure followed in this chapter is intended to detail the design of this paper's model. As the paper's main contribution, this model estimates monthly time horizons of data collection to select a stock portfolio and monthly time horizons to maintain a portfolio so selected. The paper's research problem is determining how to estimate these time horizons.

The model, consisting of (5) through (17), comprises two sets of equations. The first set, (5)–(12), is made up of equations reformulated from the DEA CCR primal model, oriented to inputs so as to adapt to paper data. This is because the DEA CCR primal model oriented to the inputs is selected to be applied by research. The second set, (13)–(17), is binary programming and minimization of square errors, and it is made up of original equations to solve the research problems presented in Section 3.

In this section we take into consideration a number of factors. These include the selection of the DEA model used and the redefinition of its terms, the matrices of DEA efficiency that are planned to be generated, the formulations for minimizing square errors, the formulations for minimizing square errors, the matrices that aid in the solution in the minimizing square errors which are defined, and finally the proposed mathematical model for the phase of minimization of square errors.

The model's purpose is to estimate the monthly time horizons that were obtained by research and that, when used together, are most representative of one another. Such time horizons would offer greater maximizing tendencies of the indices of return of a portfolio that utilizes them. After all, the portfolio is formed under the constraint of a boundary that accords with the results of the DEA efficiency adopted by the model. Furthermore, the only output of this DEA model is the index of stock returns. Consequently, the greatest representability between the monthly time horizons utilized will increase the efficiency of the cutting line regarding the return generated. This cutting line is described in Section 6.2.

*4.1. DEA Model Utilized.* The DEA model applied in this research is the CCR Oriented Inputs. This model, when it is utilized in the stock market for the formation of portfolios,

generally has superior results to other models [14]. Also this research conducted several tests to verify among the classical DEA models and their variants the one that best represented the oscillations of stock returns. The CCR Oriented Inputs had the best results.

According to Charnes, Cooper, and Rhodes, DEA models should be oriented either to inputs or to outputs. The orientation of the CCR model in this study is to inputs. The objective of this orientation is that the model maintains the input level and maximizes the only output—the returns of the stocks studied. Thus, the stocks with higher indices of return also have higher levels of efficiency [9, 15].

*4.1.1. Redefinitions of the DEA Model Terms.* The model utilized for this research, the CCR Oriented Inputs, requires a redefinition of their terms. The purpose of redefining is clear when the model uses stock data related to the past and related to the future of the collection's eight sections. This is in line with sections of collection adopted in other researches.

When the DEA modeling uses only the historical past of the eight time frames from the collection, the CCR Model Oriented Inputs described in Section 2 will have to replace the terms found in the literature review. What follows are redefinitions of those terms:

- (a) The term *Effo* is renamed  $a_o$  and represents the efficiency of the stocks  $O$  in analysis of each application of the model. The result set  $a_o$  generates the matrix (M1) which is described in Section 4.1.2.
- (b) The terms  $x'_{ik}$  are the data relating to inputs  $i = 1 \dots r = 1 \dots 3$ . The 3 variables of the set  $i$  are market value, price for sales, and price per equity value. The terms  $y'_{jk}$  are the data relating to outputs  $j = 1 \dots s = 1 \dots 1$ . They are only the return variables of each stock.
- (c) The terms  $v'_i$  and  $u'_j$  are the weights of the inputs  $i = 1 \dots r = 1 \dots 3$  and the output  $j = 1 \dots s = 1 \dots 1$ . The variables are the same as those cited in (b) item for sets  $i$  and  $j$ .
- (d) The terms  $x'_{io}$  and  $y'_{jo}$  are related to inputs  $i = 1 \dots r = 1 \dots 3$  and outputs  $j = 1 \dots s = 1 \dots 1$  of stock  $O$  in analyses. The variables are the same as those cited in (b) item for sets  $i$  and  $j$ .
- (e) The set is  $k = 1 \dots n = 1 \dots 35$ . The set ranges up to 35 because this is the number of stocks that were used in the study that applies the model proposed by (5) through (17).

Thus, the model is redefined by (5)–(8) to measure efficiency based on only the historical past of the collection's eight sections:

$$\text{Max } Effo = \sum_{j=1}^s u'_j \cdot y'_{jo} = a_o \tag{5}$$

$$\text{s.t. } \sum_{i=1}^r v'_i \cdot x'_{io} = 1 \tag{6}$$

$$\sum_{j=1}^s u'_j \cdot y'_{jk} - \sum_{i=1}^r v'_i \cdot x'_{ik} \leq 0, \quad \forall k \quad (7)$$

$$v'_i, u'_j \geq 0, \quad \forall i, \forall j. \quad (8)$$

Now, when the modeling uses future data in relation to the collection's eight sections, the model will have to replace the terms of the review with the following redefinition:

- (a) The term *Effo* is renamed  $b_o$  and represents the efficiency of the stocks  $O$  in analyzing each application of the model. The result set  $b_o$  generates the matrix (M2), which is described in Section 4.1.2.
- (b) The terms  $x''_{ik}$  are the data relating to inputs  $g = 1 \cdots q = 1 \cdots 3$ . The 3 variables of set  $g$  are also market value, price for sales, and price per equity value. The terms  $y''_{hk}$  are the data relating to outputs  $h = 1 \cdots z = 1 \cdots 1$ . They are also only the return variables of each stock.
- (c) The terms  $v''_g$  and  $u''_k$  are the weights of the inputs  $g = 1 \cdots q = 1 \cdots 3$  and the outputs  $h = 1 \cdots z = 1 \cdots 1$ . The variables are the same as those cited in (b) item for sets  $g$  and  $h$ .
- (d) The terms  $x''_{go}$  and  $y''_{ho}$  are related to inputs  $g = 1 \cdots q = 1 \cdots 3$  and outputs  $h = 1 \cdots z = 1 \cdots 1$  of stock  $O$  in analyses. The variables are the same as those cited in (b) item for sets  $g$  and  $h$ .
- (e) Set  $k = 1 \cdots n = 1 \cdots 35$ . The set ranges up to 35 because this is the number of stocks that were used in the study applying the model proposed by (5) through (17).

In this way the model is redefined by (9)–(12) to measure efficiency using only future data from the collection's eight sections:

$$\text{Max} \quad \text{Effo} = \sum_{h=1}^z u''_h \cdot y''_{ho} = b_o \quad (9)$$

$$\text{s.t.} \quad \sum_{g=1}^q v''_g \cdot x''_{go} = 1 \quad (10)$$

$$\sum_{h=1}^z u''_h \cdot y''_{hk} - \sum_{g=1}^q v''_g \cdot x''_{gk} \leq 0, \quad \forall k \quad (11)$$

$$v''_g, u''_h \geq 0, \quad \forall i, \forall j. \quad (12)$$

**4.1.2. Matrices of Efficiency Generated by DEA Model.** The DEA model calculates the results of efficiency starting from data adjusted in the last of the collection's  $t$  sections. The model's objective is to assemble a set of matrices starting from the efficiency  $a_o$  of each stock of the set  $k$  in all the tested monthly time horizons  $l$ .

The number of efficiency matrices is equal to the number of the collection's  $t$  sections, where  $t = 1 \cdots f = 1 \cdots 8$ . The eight sections are equivalent to monthly closing of BM&F Bovespa stock exchange from September 2011 to April 2012.

These are the exact points in time from which the data will be collected. Monthly time horizons from the past will be used to collect data for portfolio selection; sometimes future monthly time horizons will be used to collect data for stock portfolio maintenance.

The number of monthly time horizons from the past is represented by the set of tested  $l$  horizons. The amount of these monthly time horizons is fourteen, where  $l = 1 \cdots m = 1 \cdots 14$ . This dimension sets the number of columns for each of the eight arrays.

The number of tested  $k$  stocks that define the vertical dimension of the matrices is 35, where  $k = 1 \cdots n = 1 \cdots 35$ .

This yields a set of eight matrices  $a_{lk}$  of efficiency from the results of the stocks' history:

$$\begin{bmatrix} a_{11} & \cdots & a_{m1} \\ \cdots & \cdots & \cdots \\ a_{1n} & \cdots & a_{mn} \end{bmatrix} = f. \quad (M1)$$

Similarly, when applying the DEA model to find results of efficiency using future data or to maintain portfolios initiated in sections or in the exact collection points, the set of matrices is equal to  $t = 1 \cdots f = 1 \cdots 8$ .

The other dimensions of the set of efficiency matrices  $b_o$  are the number of sets,  $k = 1 \cdots n = 1 \cdots 35$ , and the number of future monthly horizons, or portfolio maintenance, in relation to the size of the sections  $e = 1 \cdots w = 1 \cdots 6$ . The first dimension defines the number of rows and the second dimension defines the number of columns.

In this case, the die assembly is the generated  $b_{ek}$  of stock efficiency results from  $t$  sections. This uses future data in applying the adopted DEA model. So the set of matrices  $b_{ek}$  can be defined as

$$\begin{bmatrix} b_{11} & \cdots & b_{w1} \\ \cdots & \cdots & \cdots \\ b_{1n} & \cdots & b_{wn} \end{bmatrix} = f. \quad (M2)$$

**4.2. Formulations for Minimization of Squares Errors.** Presented below are definitions of matrices that aid in the solution of the problem and the proposed mathematical model both in phase minimization of errors squares. These matrices and equations consider the phase minimization of error squares.

**4.2.1. Matrices of Assistance to Solution.** At this point it is necessary to define which matrices facilitate the solution. The first two are binary matrix lines  $c_l$  and  $d_e$ .

Matrix  $c_l$  aids in the estimation of the monthly time horizon of data collection for the formation of the portfolio. In this—as occurred in the matrix of efficiency of results from the past  $a_{lk}$ —each column  $l$  represents a monthly time horizon from the past with the eight sections, where  $l = 1 \cdots m$ .

The matrix  $d_e$  aids in the estimation of the monthly time horizon of maintenance of the portfolio selected. In this—as in the matrix of efficiency of future  $b_{ek}$ —each column  $e$  represents a monthly time horizon of maintenance portfolio, where  $e = 1 \cdots w$ .

Thus, the two matrices are

$$\begin{aligned} & [c_1 \ \cdots \ c_m], \\ & [d_1 \ \cdots \ d_w]. \end{aligned} \tag{M3}$$

These matrices' binary lines are defined so as to be used in constraints. This is because the binary matrices' lines have the same number of columns as the number of monthly time horizons tested, when associating each of its columns to a monthly time horizon and restricting the sum of the columns of each matrix to a value equal to 1. Thus, in a model for minimizing square errors, it is possible to determine the monthly time horizons optimum.

The following describes the constraints that were included in the model of minimization of error squares that was elaborated starting from binary matrices' rows:

$$\sum_{l=1}^m c_l = 1, \quad \forall l, c_l \in \{0, 1\}, \tag{13}$$

$$\sum_{e=1}^w d_e = 1, \quad \forall e, d_e \in \{0, 1\}. \tag{14}$$

After determining that the constraints defined by (13) and (14) are still necessary, the two matrices are defined regarding the tested monthly time horizons in this study.

When the matrix refers to monthly time horizons of data collection to form the portfolio, it is represented by  $p'_l$  and it has the same dimension  $l = 1 \cdots m$  as the matrix binary row  $c'_l$ .

When the matrix refers to monthly time horizons of maintenance of the selected portfolio, it is represented by  $p''_e$  and has the same dimension  $e = 1 \cdots w$  as the matrix binary row  $d'_e$ . These matrices are

$$\begin{aligned} & [p'_1 \ \cdots \ p'_m], \\ & [p''_1 \ \cdots \ p''_w]. \end{aligned} \tag{M5}$$

With matrices  $p'_l$  and  $p''_e$  defined, they may be equated with the two scalars  $P'$  and  $P''$  to find the result. This is done first for the results estimated for the monthly time horizon of data collection to create the portfolio. It is done next for the results estimated for the monthly time horizon for maintenance of the selected portfolio. To do this, just insert the following equations in the model of minimization of square errors:

$$P' = \sum_{l=1}^m p'_l \cdot c_l, \tag{15}$$

$$P'' = \sum_{e=1}^w p''_e \cdot d_e. \tag{16}$$

**4.2.2. Model Proposed for Minimizing of Square Errors.** The estimation of the monthly time horizons for this search is done by looking for the best representability between efficiency results starting from the past, represented by matrix

$a_{lk}$ , with the results of the efficiency of future, represented by matrix  $b_{ek}$ .

Based on concepts from the method to minimization of errors squares, "which is well known for being the main concept in studies using variations of Regression Analysis," [16] the model can be proposed to estimate the monthly time horizon searches.

The proposal of the model of minimization of error squares is described starting from the definitions of the matrices of efficiency calculated from sections of the collection to the past,  $a_{lk}$ , of the matrices of efficiency calculated from the sections of the collection to the future,  $b_{ek}$ , and of the binary matrices lines  $c_l$  and  $d_e$ . With these matrices, it is possible to calculate the scalars  $P'$  and  $P''$ .

The result is then inserted into the model by objective function to minimize square errors described by (17). In the model, (17) minimizes the square of the product of binary matrices  $c_l$  and  $d_e$ . This product still multiplies the errors between the results of efficiency of matrices  $a_{lk}$  with the efficiency results of matrices  $b_{ek}$ .

The intention is to identify the columns—one from the past and one from the future—that best represent each other. Equations (13) and (14) constrain the sums of the binary matrices' rows  $c_l$  and  $d_e$  to extract a value equal to 1. Thus, as these binary matrices rows vary in sets 0 and 1, only one column of  $c_l$  and one column of  $d_e$  will be equal to 1.

Consequently, the column of matrix  $c_l$  that will be equal to 1 represents the monthly time horizon estimated from the data collection of the past to the portfolio selection. The column of matrix  $d_e$  that will be equal to 1 defines the monthly time horizon estimated to maintain the selected portfolio:

$$\begin{aligned} \text{Min} \quad & \text{Effo} = \sum_{h=1}^z u''_h \cdot y''_{ho} = b_o \\ \text{s.t.} \quad & (13) \\ & (14) \\ \text{f.c.} \quad & (15) \\ & (16). \end{aligned} \tag{17}$$

It is important to be clear that (15) and (16) did not constrain the model but are calculated only for scalars  $P'$  and  $P''$ .

## 5. Collection and Treatment of Data

This chapter explains how the collecting and treatment of data was planned in this study. Such data make up the inputs and outputs utilized in the initial applications of the equations of the proposed mathematical model in Section 4.

**5.1. Decision Maker Units.** The object of study was a set of stocks traded on the São Paulo exchange, the BM&F Bovespa. As of November 2014, the stock exchange offered over 500 investment options in stocks. To utilize the data of all these stocks would make an analysis of this research unfeasible due to the extensive calculations that would need to be executed. The solution is to utilize, as first filters, the so-called Market Indices.

The most important of the Brazilian stock markets is the stock exchange of São Paulo—the Bovespa Index. Among the indexes, this one represents the average of prices and the profile of the negotiations of the cash market for stocks with high degrees of liquidity [17].

The current study intends to scrutinize the behavior of stocks traded on the BM&F Bovespa. The Bovespa Index represents well the BM&F Bovespa. Thus the first filter in the selection of stocks here was that the stocks will be components of the BM&F Bovespa Index portfolio—the Ibovespa.

The research is defined to simulate an investment starting in April 2013. In this scenario, the fictitious investor intends to build a portfolio of BM&F Bovespa and to keep it for a maximum of one year before reassessing the market. Thus, the last BM&F Bovespa Index portfolio before the investment simulation period is of the first quarter of 2013.

The stocks that this study looks at, consisting of stocks from the Bovespa Index in the first quarter of 2013, must have data available for the period of application of the proposed mathematical model, from January of 2009 to April of 2013, relating to fundamental indices utilized.

For the period of applying this model, it was observed that, of the 69 stocks comprising the Bovespa Index in the first quarter of 2013, only 42 were available in the period outlined by the fundamental indices. The third constraint was that the number of stocks to be studied must be at least 80% of the 42 stocks that are available on the fundamental indices used.

Thus, we randomly selected 33 stocks in the code: BISA3, BRFS3, BRKM5, CESP6, CMIG4, CPLE6, CSNA3, CYRE3, DASA3, DTEX3, ELET3, ELET6, ELPL4, EMBR3, GGBR4, ITSA4, LAME4, LIGT3, LREN3, MRFG3, MRVE3, OIBR3, OIBR4, PCAR4, PETR3, RENT3, RSID3, SUZB5, TIMP3, USIM3, VAGR3, VALE3, and VIVT4.

*5.2. Delimiting of Scale.* To obtain the correct monthly time horizons for this research, there needed to be an imposition of a scale of inputs and outputs homogeneous in their maximum and minimum limits. This is before applying the DEA model.

The need to impose these limits is that tests show that when in a table for application of the model DEA, CCR Oriented Inputs exist only for DMUs with low values of inputs and outputs; the other table exists only for DMUs with high values of inputs and outputs, even if in each of the tables there is a DMU with the same values of inputs and outputs. The result is that these DMUs have different efficiency values. The divergence of values occurs due to differences in the limitation of scales generated by the DEA model used for each data set.

Tests also show that the problem can be prevented using, in each table of input and output, the values of two fictitious DMUs. Around the period of application of the proposed mathematical model, a monthly return index value that was the lowest among them all and also a monthly return index value that was the largest among them all were estimated.

Thus, the first fictitious DMU will have results of the output return equal to or lower than those found during the entire modeling period, and the second fictitious DMU will have results of the output return equal to or greater than those found during the entire modeling period. In both fictional

DMUs, the three inputs used had to have average values, which in this case were utilized values equal to 1. Thus, each table in the application of the model had 33 preselected stocks, increased by two more fictitious stocks from the delimitation of scale.

Between the data from monthly returns encountered during the whole period of application of the proposed mathematical model, it was observed that the lowest value was  $-19.66\%$  and the highest value was  $10.25\%$ .

*5.3. Tested Horizons and Sections of Collection.* The first boundary of this research is in relation to the monthly time horizons of data collection of the past from points or sections of that collection. In this case, the maximum monthly time horizon is, according to surveys, 32 months or the approximate average between two and three years. Within this maximum period, the relation of four months of the Bovespa Index portfolio is maintained. So in the initial stage eight horizons are tested, starting at 4 months and ending at 32 months.

The second boundary relates to the monthly time horizons of data collection of the future from points or sections of that collection, which represent the time horizons of maintaining the stock portfolio. In this case, the study uses the typical maximum monthly time horizon, according to research, which is 12 months. Within this maximum period, again the relationship of four months to the Bovespa Index portfolio is maintained. Hence, three horizons are tested in an initial phase, beginning at 4 months and ending at 12 months.

After applying the model to estimate the local monthly time horizons according to the preliminary scope, the study searches for, within the limits of this scope, the global optimal time horizons.

In practice, the horizons of data collection for the portfolio selection tested were, in months, 4, 8, 12, 16, 20, 24, 25, 26, 27, 28, 29, 30, 31, and 32. The horizons for the maintenance portfolio that had already been tested were, in months, 4, 8, 9, 10, 11, and 12.

This is because the results of the first application of the model proposed identified 28 months as the optimal monthly time horizon for collecting data for portfolio selection and 12 months as the optimum monthly time horizon of maintenance of the selected portfolio. The global optimum within the boundaries of the search scope is presented in Section 6.1.

The third definition of search time relates to the collection sections. These sections are exact points in time from which the data were collected. Sometimes the section is from the past, representing the variation of monthly time horizons for portfolio selection; sometimes the section is for the future, representing the variation monthly time horizons of portfolio maintenance. In total there are eight sections. The first section occurs at the close of the stock exchange in September 2011, and the last section occurs at the close of April 2012 of the BM&F Bovespa exchange. In this range, each monthly closing corresponds to a collection section.

The main constraint for the extensions of the two monthly time horizons of the first two delimitations, as well as the section number of the collection, was not to use 2008 data, because of the distortion from that year's economic crisis.

The fourth boundary for the times of research concerns the simulation period of the portfolio that used the time horizons estimated by the proposed mathematical model. The planning was that, 12 months after the last point of data collection, the simulation of the stock portfolio would be initiated with the time horizons estimated. This was chosen because the major monthly time horizon of maintenance portfolio tested was also 12 months. In this way, the period of application of the proposed mathematical model to estimate the time horizons and the period of simulation of the results of the portfolio with the estimated horizons were completely independent, giving more credibility to the study. The simulation period took place from the beginning of May 2013 to the end of June 2014, a total of 14 months.

**5.4. Data Utilized and Their Definitions.** The research presupposes that an investor at the end of April 2013 intends to build a portfolio of stocks from the BM&F Bovespa and to keep it at least one year, before assessing the need to change it. It is assumed then that the investor has access to the historical index of stocks on the BM&F Bovespa.

The fictitious investor intends to obtain a set of data to measure the efficiency of stocks solely to obtain the return index. In this way, the only data of output utilized by the DEA model is the monthly return of each stock.

Regarding other dates to be used by the DEA model as inputs, Nagano et al. [18] argued that “analyzing fundamental indices applying regression techniques to see which have the most power to explain the return of stocks BM&F Bovespa, its concluded among them those indexes the which show strong relationship with return are: Profit per price, market value, price per equity value, liquidity, beta and price per sales.”

However, modular correlation coefficients ranging between 0.10 and 0.30 are able to explain a response variable, while values lower than 0.10 cannot [19]. Thus the study analyzed which fundamental indices had modular correlation coefficients with the monthly stock returns. The study then studied which was greater than or equal to 0.20, which is the average of the variation of correlation that, according to Moore [19], is the least bit representative.

“The stock market is dynamic and their trends constantly change in short periods” [3]. To minimize the possibility of errors by changes in trends, data was used up to 32 months before situating the fictitious investor at the end of April of 2013. The extension of 32 months was chosen also to represent the maximum time extension tested in the survey to collect data on the portfolio formation.

Thus, the correlation calculations were executed with data from September 2010 to April 2013. That is, for this period we calculated the correlation between each of the fundamental indices selected by Nagano et al. [18], for the returns of all stocks studied here. Subsequently, a total average for the index was obtained.

The values that correlated between the indexes were as follows: profit per price = -0.17, market value = 0.26, price per equity value = 0.26, liquidity = -0.07, beta = 0.01, and price per sales = 0.25. In this way, the fundamental indices selected as inputs for the DEA model were the following: market value, price per equity value, and price for sale.

**5.5. Inputs Adjustments and Parameters.** The fundamental indices that were used as inputs in this research need to be adjusted to improve their representability in relation to the stock returns studied. This adjustment is calculated as the monthly percentage variation that the indices fell or rose in one unit.

For example, if in a certain section of data collection we wanted to calculate the adjusted input to be used, then we assigned this input the value of 1 plus the percentage monthly variation  $i$  for the horizon of data collection for the portfolio selection at a distance of  $n$  months. The calculation must agree that the value of crude input after collecting the month section is  $x_n$  and that the same input to  $n$  previous months has the value of  $x_o$ . So to calculate the percentage of monthly variation  $i$ , this index number should solve

$$i = \sqrt[n]{\frac{x_n}{x_o}} - 1. \tag{18}$$

To calculate the interference in one unit of the monthly percentage variation  $i$ , the formula to use is

$$\text{Adjusted input} = (1 + i) = 1 + \left( \sqrt[n]{\frac{x_n}{x_o}} - 1 \right) = \sqrt[n]{\frac{x_n}{x_o}}. \tag{19}$$

To calculate the result of a particular section for the future of that same section, representing the direction of maintaining the portfolio, the variable of the section must be  $x_o$  and of the future at  $n$  months should be  $x_n$ .

The justification for the adjustment is that the fundamental indices change significantly in size from one stock to another, which would make it an unfair or incorrect use of modeling in its own index number. If we want to see the maximum efficiency of stocks at producing a return, this way seeks to minimize any alteration in the results of efficiency caused by alterations in the inputs.

The database used in this research is from the *Econômica* Software. The collection parameters for each input index number are as follows:

- (a) Market value = enterprise value at the monthly close, in the country’s currency, to three decimal places.
- (b) Price per equity value = index value at the consolidated monthly close, in the country’s currency, to three decimal places.
- (c) Price for sales = index value at the consolidated monthly close (based on the demonstrations of monthly returns for the last year), in the country’s currency, to three decimal places.

**5.6. Output Adjustments and Parameters.** If for a certain section of data collection we want to calculate the adjusted value of output return, according to the percentage monthly variation  $i$  for a horizon of data collection for a portfolio selection, it is done in extension  $n$  months. Here the value of the index number of the return for the section’s month is  $y_n$ , and the same number index at  $n$  previous months is  $y_o$ .

The value of the adjusted output to be used by the DEA model can then be obtained by

$$\begin{aligned} &\text{Adjusted output} \\ &= \left[ \left( \sqrt[n]{\frac{y_n}{y_o}} - 1 \right) \times 100 + (2 \times 19,657) \right]^2 \end{aligned} \quad (20)$$

When the adjusted output from the percentage monthly variation  $i$  initiated in the future section needs to be calculated, the index number in the section is  $y_o$  and in the future  $n$  months is  $y_n$ .

The first portion of the equation is multiplied by 100 so as to alter the percentage values obtained for integer values with three decimals. This is mainly to maximize the interference of variations of return on the results of efficiency.

The value of  $2 \times 19,657$  is necessary due to the constraint of the DEA model CCR Oriented Inputs being unable to use negative data. As the period of application of the model, the lowest value of  $(\sqrt[n]{y_n/y_o} - 1) \times 100$  is  $-19.657$ ; then, as recommended in the literature, this value is summed twice in the adjustment of the data to eliminate the negative.

The equation is squared to maximize the differences of stock return indexes. The motivation is to make the DEA model be able to absorb the minimum difference in stock return indexes and reflect this difference in the efficiency results.

The collection parameters of this index are as follows: value of return on monthly closing calculated based on the previous month, in the original currency of the country (adjusted for dividends), with integer values obtained by multiplying the percentage by 100 and to three decimal places.

This index was collected in monthly percentage values. Before applying the calculation formula of (20) to the data, all stocks were attributed, in December 2008, an index number equal to 1, that is, one month before the period of application of the model. Subsequently, this index number was adjusted for all stocks and in all months. The index number represents the respective value of the monthly return for each stock.

**5.7. Negative Inputs or Outputs Adjustments.** The DEA CCR model oriented to inputs must treat those variables presenting negative data. As this is the model used in the research, it should verify the possible presence of negative data on inputs and outputs.

The formulas used in (19) and (20) for the adjustment of inputs and outputs were designed to eliminate negative data.

Equation (19), which is utilized for treatment of inputs, also has a structure that eliminates negative data.

An alternative to eliminating negative data in models DEA to each data of particular variable must be added a constant. This constant must have a positive value (+) greater than or equal to lowest data negative (-) that was found in the variable data set [20].

This is the procedure adopted for the elimination of negative data by (20) when adding up  $2 \times 19.657$ .

## 6. Results and Discussion

This chapter presents the results used to validate the proposed mathematical model. First, the extension is estimated in months for the time horizon of utilizing the DEA model in selecting stock portfolios. At the same time, the extension is estimated in months for the time horizon of maintaining the selected portfolio.

We validate the model by simulating the estimated annual indexes of return, both for the portfolio that utilizes the two horizons found and the others that do not.

To validate the proposed mathematical model for hypotheses testing, we compare the estimated annual indexes of returns. Here the study verifies the possible existence of statistically significant differences between the indexes of the portfolio formed using the two horizons estimated and the indexes of those not utilizing them.

It is important to clarify that the estimation period of the monthly time horizons is independent of the simulation period of the indices of return of portfolios. This is to highlight the conceit that the simulation was executed by a fictitious investor situated at a point in time after obtaining the horizons, with no knowledge of the stock market's future.

The period during which the model was applied to estimate the time horizon was from January of 2009 to April of 2013. Hence, the simulation period of the portfolio with the determined monthly time horizons, as well as all the other portfolios formed for comparison, begins in early May of 2013 and finishes at the end of June 2014.

**6.1. Results of Minimization of Squares Error.** At this point, to assist calculations Microsoft *Excel* software can be used. In the software the matrices efficiency results  $a_{ik}$  and the matrices efficiency results  $b_{ie}$  must be mounted. This is done for each of the eight sections in the collection.

This way, it is possible to execute the modeling of this stage to obtain the horizons sought. This is done according to the objective function given by (17), constraints by (13) and (14), and scalars by (15) and (16) of the system proposed in Section 4.

After applying the model using *Excel Solver* so that the model's equations are applied to the minimization of square error over the matrices  $a_{ik}$  and  $b_{ie}$ , we obtain the first response of scalar  $P'$ .

The value of the scalar  $P'$  is obtained by multiplying the matrix  $p'_i$ , of the time horizons of data collection, by results of the binary matrix line  $c_i$  of variables of decision.

Table 1 provides the matrix of options for the horizons of data collection for the formation of portfolio  $p'_i$ , the binary matrix of decision variables  $c_i$ , the results of the scalars  $P'$ , and also the result of the objective function of the model in the phase of minimization of errors square.

In Table 1, it can be observed that 26 months is the value of the scalar  $P'$ , which represents the time horizon estimated for the data collection of the past to the portfolio formation using the CCR DEA Model Oriented Inputs.

Also, with the application of the proposed model using *Excel Solver*, we obtain the response of scalar  $P''$ . Its value is

TABLE 1: Estimated result of the monthly horizon of data collection.

Line matrix of tested monthly horizons of the past						=	$p'_l = 1, \dots, m$						
32	30	31	29	28	27	<b>26</b>	25	24	20	16	12	8	4
Matrix binary line of decision variables						=	$c_l = 1, \dots, m$						
0	0	0	0	0	0	<b>1</b>	0	0	0	0	0	0	0
Scalar results of data collection						=	$P' = \sum_{l=1}^m p'_l c_l$						
<b>26</b>													
Objective function	$\text{Min} \sum_{t=1}^f \left( \sum_{l=1}^m \sum_{e=1}^w \sum_{k=1}^n [(c_l d_e) (a_{lk} b_{ek})]^2 \right)$											=	<b>2.1499</b>

TABLE 2: Estimated result of the maintenance horizon.

Line matrix of tested monthly horizons of the future				=	$p''_e = 1, \dots, w$				
4	8			9	10	11	<b>12</b>		
Matrix binary line of decision variables				=	$d_e = 1, \dots, w$				
0	0			0	0	0	<b>1</b>		
Scalar results of portfolio maintenance				=	$P'' = \sum_{e=1}^w p''_e d_e$				
<b>12</b>									
$\text{Min} \sum_{t=1}^f \left( \sum_{l=1}^m \sum_{e=1}^w \sum_{k=1}^n [(c_l d_e) (a_{lk} b_{ek})]^2 \right)$									<b>2.1499</b>

estimated by multiplying the matrix of options of horizons to maintain the portfolios tested  $p''_e$  by the resulting values of the Solver to the decision variables of the matrix binary line  $d_e$ .

Table 2 provides the matrix of options of horizons to maintain portfolio  $p''_e$  the binary matrix of decision variables  $d_e$ , the results of scalar  $P''$ , and also the result of the objective function of the model in the phase of minimization of errors squares.

Table 2 shows that the value of scalar  $P''$ , which represents the horizon estimated to maintain the portfolio formed using the CCR DEA Model Oriented Inputs, is 12 months.

It can be seen from the results that the objective functions of Tables 1 and 2 are equal. This is because they deal with the same value; that is, the proposed mathematical model estimates the monthly time horizons in the same application.

6.2. *Formation of the Portfolio with Horizons Estimated.* To execute the planned comparisons of the indexes of return, a portfolio must be designed that simulates the application of extensions of time horizons estimated by the proposed mathematical model.

The cutting line for the formation of portfolios done in this study should be discussed here. The DEA CCR model oriented to inputs gets great results in selecting stock portfolios. This great results is reached if it used a cut line which is only the stock within of the fourth quartile of best efficiency results—approximately 25%—form a portfolio [14].

So, according to the efficiency results found, this paper should search for the stocks' ranked matches. The ranking accounts for only the top 16 stocks classified, which are entered into the portfolio of horizons estimated by the proposed mathematical model.

There were 16 stocks because, among all the portfolios of the Bovespa Index in the period of model application, the minimum observed were 66 stocks. According to the theory of Pätäri et al. [14], 25% of the minimum observed determines the cut line, which in this case approximately represented 16 stocks.

It is also necessary to define the weight that will reflect the attributed participation of the stocks, both for the portfolio of horizons estimated by the model proposed and for all the other portfolios used to compare. To facilitate comparisons, the participation of each stock will receive a standardized form of weighting, according to the same weightings that each stock had in the Bovespa Index during the first quarter of 2013.

For example, in the case of the portfolio of horizons estimated by the proposed mathematical model, the weights should be attributed to the stocks for the formation of the portfolio in two allocations, the first from May of 2013 and the second from May of 2014.

After all, the estimated horizon optimal for maintaining a portfolio is 12 months and the period of simulation results of stock portfolios is 14 months. So the design procedure of stock portfolios should be repeated, as the simulation period goes two months longer than the optimal monthly time horizon estimated by the proposed model.

Starting from the planning described above and the results of the estimated horizons applied to the fictitious portfolio, it is possible to find the statistical parameters of the portfolio that utilize the horizons estimated by the proposed mathematical model, according to what is available in Table 3.

The table presents the results of monthly return, cumulative return, monthly standard deviation, average monthly

TABLE 3: General results of the portfolio with horizons estimated by the proposed mathematical model.

(a)	
Month	Monthly return
2013/May	-0.44%
2013/June	-7.13%
2013/July	1.93%
2013/Aug.	0.08%
2013/Sep.	5.20%
2013/Oct.	4.09%
2013/Nov.	-1.69%
2013/Dec.	-1.48%
2014/Jan.	-4.73%
2014/Feb.	-2.56%
2014/Mar.	8.47%
2014/Apr.	3.65%
2014/May	-0.97%
2014/June	4.47%

  

(b)	
Other parameters	Value
Average monthly return	0.64%
Monthly standard deviation	4.23%
Cumulative return	8.02%
Estimated annual return	7.90%
Annual standard deviation estimate	14.66%
<i>P</i> value of normality	0.869

return, estimated annual return starting from the average monthly return, and annual standard deviation estimated starting from the monthly standard deviation.

In Table 3, the monthly return is the percentage oscillation of the variation total obtained. The cumulative return is in relation to the monthly return. The monthly standard deviation is calculated over 14 monthly return samples. The estimated annual return is monthly plus 1, after being elevated to the 12th potency, and finally subtracting 1.

The annual estimated standard deviation is the monthly standard deviation of the samples, multiplied by the square root of 12—the number of months in a year—or the sum total of the amount of monthly variances of the period to which it wishes to estimate [21].

Table 3 also brings the *P* value of the test of normality by Anderson Darling. This follows the hypothesis that their value must be greater than 0.05. For statistical parameters, it can be applied. Its calculation was executed for the 14 monthly return samples with the *Minitab* 14 software.

**6.3. Hypothesis Testing.** All tables from this section bring the results of the monthly return, cumulative return, monthly standard deviation, average monthly return, estimated annual return starting from the average monthly return, and annual standard deviation estimated starting from the monthly standard deviation. The tables also bring the beta relative to the Bovespa index and the *P* value of

the Anderson Darling Normality Test. This latter one is for the authentication of the application of statistical parameters.

Most importantly, the following tables also bring the results of *P* values of hypothesis testing. These tests verify whether the estimated annual return of the portfolio that utilizes the horizons estimated by the model has differences greater than or equal to zero in relation to various portfolios formed for comparison. The analysis is complemented with the parameter of the expected difference between the indices of estimated annual returns of the portfolio with the estimated horizons with the other portfolios.

The test carried out for two samples of equal size to 14 is

$$\begin{aligned}
 H_0 : \mu (\text{Portfolio with horizons found}) \\
 - \mu (\text{Other portfolios}) = 0, \\
 H_1 : \mu (\text{Portfolio with horizons found}) \\
 - \mu (\text{Other portfolios}) > 0.
 \end{aligned} \tag{21}$$

In the hypothesis testing, the averages and standard deviations were used for the estimated annual values starting with the monthly values that actually occurred. The values are annuals because, generally, that is the value utilized for comparison. Namely, the indexes of interest to the government, to the market, and to research are usually shown on an annual basis.

Due to the high unpredictability of the stock market, the confidence level adopted is 90%. Thus, a null hypothesis is accepted for values of *P* value greater than 10%. The calculations are executed with Minitab 14 software by the tool 2 sample-t, where it was not assumed that the samples had equal variances.

The first comparison indicates the efficiency of the model in proposing, simultaneously, the two monthly time horizons in question. The portfolios subjected to tests for these first comparisons vary with both the horizons of data collection for formation of portfolio, as the horizon of maintained portfolio formed.

The premise here is to vary the range of the horizon data collection by approximately one year to two and a half years. Also, as the horizon of data collection for portfolio selection increases, the horizon of maintenance of the portfolio switches between four and eight months—the horizons of maintenance tested in the modeling, namely, within the scope, and also different from the optimum estimated. This is so that we can test the time horizons commonly used by investors based on research on the stock market. Also, this maintains the relation of four months of maintenance with the Bovespa Index portfolio.

Table 4 shows that the portfolio that uses the two horizons estimated by the model presents a superior performance to all other portfolios, in relation to all other portfolios with choices of different monthly time horizons. In the table, it can be observed that the cumulative return in the simulation period by the portfolio using the estimated horizon is greater than all six portfolios with random horizons.

TABLE 4: General results of the portfolio with the estimated and random horizons.

Statistical parameters	Portfolio Esti. Hz	Portfolio 08 × 08	Portfolio 12 × 04	Portfolio 16 × 08	Portfolio 20 × 04	Portfolio 24 × 08	Portfolio 28 × 04
Average monthly return	0.64%	-0.31%	0.23%	-0.53%	-0.23%	0.12%	-0.08%
Monthly standard deviation	4.23%	3.85%	4.69%	3.81%	4.29%	4.09%	4.09%
Cumulative return	8.02%	-5.24%	1.74%	-8.03%	-4.30%	0.57%	-2.15%
Estimated annual return	7.90%	-3.71%	2.76%	-6.16%	-2.68%	1.44%	-0.92%
Annual standard deviation estimate	14.66%	13.35%	16.25%	13.20%	14.87%	14.16%	14.18%
Beta	0.767	0.603	0.853	0.659	0.781	0.744	0.727
<i>P</i> value of normality	0.869	0.375	0.904	0.180	0.279	0.348	0.813
Estimated annual difference	—	11.61%	5.14%	14.06%	10.58%	6.46%	8.82%
<i>P</i> value hypothesis	—	0.019	0.194	0.007	0.035	0.123	0.059

TABLE 5: General results of the model portfolio and random horizons of data collection.

Statistical parameters	Portfolio Esti. Hz	Portfolio 08 × 12	Portfolio 12 × 12	Portfolio 16 × 12	Portfolio 20 × 12	Portfolio 24 × 12	Portfolio 28 × 12
Average monthly return	0.64%	0.40%	0.07%	-0.20%	0.29%	0.38%	0.52%
Monthly standard deviation	4.23%	3.55%	4.86%	4.20%	4.76%	4.17%	4.16%
Cumulative return	8.02%	4.91%	-0.53%	-3.82%	2.68%	4.28%	6.33%
Estimated annual return	7.90%	4.93%	0.88%	-2.33%	3.60%	4.66%	6.41%
Annual standard deviation estimate	14.66%	12.29%	16.82%	14.54%	16.49%	14.43%	14.42%
Beta	0.767	0.644	0.893	0.691	0.881	0.758	0.752
<i>P</i> value of normality	0.869	0.542	0.407	0.144	0.583	0.548	0.913
Estimated annual difference	—	2.97%	7.02%	10.23%	4.30%	3.24%	1.49%
<i>P</i> value hypothesis	—	0.283	0.125	0.038	0.236	0.280	0.394

The next comparisons are between the portfolio formed with both monthly time horizons estimated and the portfolios formed using the horizons of data collection different from what was estimated but with the monthly time horizon of maintenance equal to the estimated one.

Table 5 presents, beyond the results of the portfolio with estimated horizons, the results of the six portfolios that were tested in this case. Here the portfolios for comparisons vary only in terms of the horizon of data collection.

In this case, the premise is to vary the horizon for data collection from approximately one year to two and a half years. This is so that we can test horizons commonly used by investors and research on the stock market. However, to the extent that the horizon of collecting data for portfolio selection increases, the maintenance is kept at 12 months, which was the optimum estimated. Here again this maintains the relation of four months of maintenance with the Bovespa Index portfolio.

Table 5 shows that the portfolio that uses the horizons estimated by the proposed mathematical model presents a performance that is again superior to all portfolios formed for comparison in this case. In the table, it can be observed that the cumulative return in simulation by the portfolio that uses the estimated horizons is greater than all six portfolios with random horizons of data collection.

The final comparisons executed are among the portfolios formed with the two monthly time horizons estimated and those formed utilizing the monthly time horizon of data collection estimated but with the horizons for maintenance of portfolio different from that estimated.

Table 6 shows, beyond the results of the portfolio with horizons estimated by the proposed mathematical model, the results of the four portfolios tested. Here the compared portfolios vary only in their horizons of maintaining the portfolio.

The idea here is to vary the horizon for maintaining the portfolio from 4 to 10 months. The motivation is to test

TABLE 6: General results of the model's portfolio and random horizons to maintain.

Statistical parameters	Portfolio Esti. Hz	Portfolio 26 × 04	Portfolio 26 × 06	Portfolio 26 × 08	Portfolio 26 × 10
Average monthly return	0.64%	0.01%	0.29%	0.00%	0.18%
Monthly standard deviation	4.23%	4.01%	4.08%	4.04%	3.72%
Cumulative return	8.02%	-0.92%	2.97%	-1.12%	1.65%
Estimated annual return	7.90%	0.11%	3.50%	-0.05%	2.20%
Annual standard deviation estimate	14.66%	13.89%	14.15%	13.98%	12.89%
Beta	0.767	0.711	0.729	0.721	0.676
<i>P</i> value of normality	0.869	0.762	0.938	0.947	0.915
Estimated annual difference	—	7.79%	4.40%	7.95%	5.70%
<i>P</i> value hypothesis	—	0.081	0.213	0.077	0.143

horizons commonly used by investors and research on the stock market and also to seek to maintain the relation of four months of maintenance with the Bovespa Index. To the extent, however, that the horizon of maintaining the portfolio increases, the data collection is maintained at 26 months, the optimum horizon estimated in this case.

Table 6 shows that the portfolio using the horizons estimated by the model again outperformed all the others. In the table, it can be observed that the cumulative return in the simulation period of the portfolio that used both estimated horizons is greater than that of all four portfolios.

## 7. Conclusions

The main contribution of this study is the model it has proposed of binary mathematical programming of minimization of square errors to estimate monthly time horizons. This contribution is significant because the simulation using monthly time horizons estimated by the mathematical model proposed in a fictitious stock portfolio showed an increase of annual indices of return. This increase is in comparison with other stock portfolios with an identical methodology of formation, differing only in the extensions of the monthly time horizons used in their management.

The general results of estimated annual indices of return show that the portfolio formed with the two found horizons estimated annual return indices superior to all portfolios generated for comparison purposes, where the general difference between the indices is, on average, 6.99% and ranges from 1.49% to 14.06%.

The hypothesis tests in this case confirmed the superiority of the results of the estimated annual return indices of the portfolio with the two estimated horizons. The results were at statistically significant levels in approximately 50% of all comparisons executed. This affirms the achievement of the study's first specific objective, that of programming a mathematical model to estimate the monthly time horizons. Indeed, in spite of all the difficulties involved in improving the prediction of

stock markets, the model was able to increase the estimated annual return indices of the fictitious stock portfolio.

The study's second specific objective was to estimate, simultaneously, the monthly time horizons in question. The comparisons, in this case, are between the portfolio with the two estimated monthly time horizons and those not using any of the monthly time horizons estimated. These comparisons verified the capacity of the model to estimate simultaneously the monthly time horizons.

The hypothesis tests regarding the second specific objective confirm that the proposed mathematical model was statistically significant. In maximizing the estimated annual return of the portfolio with the two found horizons, the difference was on average 9.45% and ranged from 5.14% to 14.06%. The results of the hypothesis testing in this case are also confirmed by the model's superiority in four of the six comparisons executed, and it has a close result in the confirmation of the other one.

The third analysis of specific objectives was executed by comparing the portfolio with the two estimated monthly time horizons with portfolios using random horizons of data collection as a means for selection, but with the horizon estimated for the maintenance of the selected portfolio. Such comparisons allow us to check the capacity of the model to estimate solely the monthly time horizon of data collection in the portfolio selection.

These comparisons relating to the third specific objective, the results of the estimated annual return of the portfolio with the two monthly time horizons, found the model to be, again, superior to all other portfolios, with a difference that averaged 4.88% and ranged from 1.49% to 10.23%. However, the analysis of the *P* value of the hypothesis tests confirmed the superiority, in this case, of only one of the six comparisons, while having a result close to confirming another one.

The fourth analysis of specific objectives was executed by comparing the portfolio with both monthly time horizons estimated with those utilizing the horizon from data collection to portfolio selection but along with random horizons

of maintenance for the portfolio. These comparisons allow us to verify the capacity of the proposed mathematical model to estimate only the monthly time horizon of maintaining the portfolio.

In these comparisons relating to the fourth objective, the portfolio with both monthly time horizons showed the results of indices of estimated annual return that were superior to all other portfolios, where the difference averaged 6.46% and ranged from 4.40% to 7.95%. However, the analysis of  $P$  value of the hypothesis tests in this case confirmed the superiority of the portfolio with both the found horizons for only two of the four comparisons executed, along with a close result of confirming another one.

For future research, the proposed mathematical model could undergo alterations in cases where the investor knows or prefers a monthly time horizon for maintaining the stock portfolio that has been selected using the DEA model adopted here. Also, an investor could know or prefer a monthly time horizon of data collection from the past to use in the DEA model for an eventual portfolio formation. In both cases, the necessary adaptations just require the inactivating of the model stages related to the estimation of known or preferred monthly time horizons.

The proposed mathematical model has a structure, where prediction models can replace the DEA in selecting the stock portfolio. An example of replacement could be the use of Regression Analysis to predict return results. This varying of the data collection horizon for the prediction, while at the same monthly time collecting the return results, would vary the maintenance monthly time of the portfolio. Thus, the model with the correct adequacies would find the best fit by minimization of error squares, in this case, between the determined monthly time horizon to forecast using the Regression Analysis, towards determining the monthly time horizon to maintain the portfolios formed with the tool.

Among the results of portfolios formed for this study, the beta value was calculated, representing the risk of a stock according to the Capital Asset Pricing Model, CAPM. The calculation gave birth to the realization that although the portfolio that uses the horizons estimated by the proposed mathematical model returned results superior to all the others, their beta index indicated no increased or decreased risk for it. Namely, beyond offering monthly time horizons that allow the obtaining of a higher return, the proposed mathematical model also seems capable of offering only a medium risk compared to the other portfolios formed by research. Other studies could confirm or disconfirm this conjecture.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

The authors would like to express their gratitude to the Brazilian agency FAPEMIG (Foundation for the Promotion

of Science of the State of Minas Gerais), which have been supporting the efforts for the development of this research.

## References

- [1] N. C. P. Edirisinghe and X. Zhang, "Input/output selection in DEA under expert information, with application to financial markets," *European Journal of Operational Research*, vol. 207, no. 3, pp. 1669–1678, 2010.
- [2] P. Rotela Jr., E. O. Pamplona, A. F. Silva, F. L. R. Salomon, V. E. M. Valerio, and L. A. Carvalho, "Data envelopment analysis and fuzzy theory: efficiency evaluation under uncertainty in portfolio optimization," *WSEAS Transactions on Business and Economics*, vol. 12, no. 4, pp. 74–87, 2015.
- [3] J. L. Ticknor, "A Bayesian regularized artificial neural network for stock market forecasting," *Expert Systems with Applications*, vol. 40, no. 14, pp. 5501–5506, 2013.
- [4] J. D. Lamb and K.-H. Tee, "Data envelopment analysis models of investment funds," *European Journal of Operational Research*, vol. 216, no. 3, pp. 687–696, 2012.
- [5] R. Rodríguez, M. Luque, and M. G. Patari, "Portfolio selection in the Spanish stock market by interactive multi-objective programming," *Sociedad de Estadística e Investigación Operativa*, vol. 19, no. 1, pp. 213–231, 2011.
- [6] P. Rotela Junior, E. De Oliveira Pamplona, and F. L. R. Salomon, "Otimização de portfólios: análise de eficiência," *Revista de Administração de Empresas*, vol. 54, no. 4, pp. 405–413, 2014.
- [7] Muren, Z. Ma, and W. Cui, "Generalized fuzzy data envelopment analysis methods," *Applied Soft Computing*, vol. 19, no. 1, pp. 215–225, 2014.
- [8] B. Gold, "Technology, productivity and economic analysis," *The International Journal of Management Science*, vol. 1, no. 1, pp. 5–24, 1973.
- [9] C. K. Wei, L. C. Chen, R. K. Li, and C. H. Tsai, "Exploration of efficiency underestimation of CCR model: based on medical sectors with DEA-R model," *Expert Systems with Applications*, vol. 38, no. 4, pp. 3155–3160, 2011.
- [10] H. Y. Kao, C. Y. Chan, and D. J. Wu, "A multi-objective programming method for solving network DEA," *Applied Soft Computing*, vol. 24, pp. 406–413, 2014.
- [11] A. H. Marbini, A. Emrouznejad, and P. J. Agrell, "Interval data without sign restrictions in DEA," *Applied Mathematical Modelling*, vol. 38, no. 7-8, pp. 2028–2036, 2014.
- [12] S. Sarkar, "Prediction of a CRS frontier function and a transformation function for a CCR DEA using EMBEDDED PCA," *Data Envelopment Analysis and Decision Science*, vol. 2013, Article ID dea-00016, 15 pages, 2013.
- [13] R. D. Banker, A. Charnes, and W. W. Cooper, "Some models for the estimation of technical and scale inefficiencies in data envelopment analysis," *Management Science*, vol. 30, no. 9, pp. 1078–1092, 1984.
- [14] E. Pätäri, T. Leivo, and S. Honkapuro, "Enhancement of equity portfolio performance using data envelopment analysis," *European Journal of Operational Research*, vol. 220, no. 3, pp. 786–797, 2012.
- [15] M. Z. Angiz, A. Mustafa, and M. J. Kamali, "Cross-ranking of decision making units in data envelopment analysis," *Applied Mathematical Modelling*, vol. 37, no. 2, pp. 398–405, 2013.
- [16] J. Ku, "Least-squares solutions as solutions of a perturbation form of the Galerkin methods: interior pointwise error

- estimates and pollution effect,” *Journal of Computational and Applied Mathematics*, vol. 251, no. 1, pp. 67–80, 2013.
- [17] D. D. D. Pinto, J. G. M. S. Monteiro, and E. H. Nakao, “An approach to portfolio selection using an ARX predictor for securities’ risk and return,” *Expert Systems with Applications*, vol. 38, no. 12, pp. 15009–15013, 2011.
- [18] M. S. Nagano, E. M. Merlo, and M. C. Silva, “As Variáveis Fundamentalistas e seus impactos na taxa de retorno de ações no Brasil,” *Revista FAE*, vol. 6, no. 2, pp. 13–28, 2003.
- [19] D. S. Moore, *The Basic Practice of Statistics*, Freeman, New York, NY, USA, 4th edition, 2007.
- [20] W. W. Cooper, L. M. Seiford, and K. Tone, *Introduction to Data Envelopment Analysis and Its Uses: With DEA-Solver Software and References*, Springer, New York, NY, USA, 2006.
- [21] J. A. Dilello, R. Hesse, and D. J. Stanley, “Portfolio performance with inverse and leveraged ETFs,” *Financial Services Review*, vol. 23, no. 2, pp. 123–127, 2014.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

