Revista de Educação e Pesquisa em Contabilidade
Journal of Education and Research in Accounting
Revista de Educación e Investigación en Contabilidad

REPeC, Brasília, v. 8, n. 3, art. 3, p. 267-283, Jul./Sep. 2014
Available online at *www.repec.org.br*
DOI: http://dx.doi.org/10.17524/repec.v8i3.1109

Quarterly journal, digital and free published by Brazilian Academy of Science in Accounting

ABRACICON
ACADEMIA BRASILEIRA DE CIÊNCIAS CONTÁBEIS

ISSN 1981-8610

Application of Extreme Value Theory and Fundamental Analysis in Long-Short Strategies: an Analysis of Pair Tradings in The Brazilian Market

Abstract

In recent decades, the number of funds has increased which aim to explore market inefficiencies through arbitrage strategies, among which the long-short strategy stands out. A large part of the analyses used to obtain the pair tradings, however, does not consider the extreme deviations that exist in the interdependence process between the assets involved and the firms' operational quality indicators. The Extreme Value Theory and Fundamental Analysis were used in this study to model the series of the asset pair price indices obtained based on the accounting indicator structure proposed by Piotroski (2000). These approaches permitted considering companies with positive signs of profitability, an operational capital structure and efficiency, besides distributions that are capable of capturing the extreme co-movements associated with the selected pair tradings. Based on this model, a new quantitative approach was created for the long-short strategy, called the GEV Long-Short. The obtained results suggest that the best adjustment of the extreme quantiles through the extreme value distribution can provide more refined probabilistic support for the return to the average to justify the possibility of long-short arbitrage.

Key Words: Long-Short Strategies; Extreme Value Theory; Arbitrage; *Pair* Trading; Fundamental Analysis.

Danilo Soares Monte-Mor
M.Sc. in Economics (UFES), Ph.D. candidate in Accountancy and Business Administration (Fucape) and Professor at Fucape Business School. **Contact:** Av. Fernando Ferrari, 1358, Boa Vista, Vitória (ES), CEP 29075-505.
E-mail: *danilo@fucape.br*

Marco Aurélio dos Santos Sanfins
Ph.D. in Statistics (UFRJ) and Professor at Universidade Federal Fluminense (UFF). **Contact:** Rua Mário Santos Braga, S/N, Valonguinho, Niterói (RJ), CEP 24.020-140.
E-mail: *marcosanfins@automata.uff.br*

Silvania Neris Nossa
M.Sc. in Accountancy (Fucape) and Ph.D. candidate in Accountancy and Business Administration (Fucape). **Contact:** Av. Fernando Ferrari, 1358, Boa Vista, Vitória (ES), CEP 29075-505.
E-mail: *silvania@fucape.br*

Aridelmo José Campanharo Teixeira
Ph.D. in Controllership and Accounting (USP) and Professor at Fucape Business School. **Contact:** Av. Fernando Ferrari, 1358, Boa Vista, Vitória (ES), CEP 29075-505.
E-mail: *aridelmo@fucape.br*

Published in Portuguese, English and Spanish. Original Version in Portuguese.

Received in 04/20/2014. Ask to Revise on 07/24/2014. Resubmitted on 07/28/2014. Accepted on 07/29/2014 by Dr. Valcemiro Nossa (Editor). Published on 09/30/2014. Organization responsible for the journal: Abracicon.

1. Introduction

In recent decades, the number of funds that aim to explore market inefficiencies through arbitrage has increased. The exploitation of such inefficiencies can happen through strategies like the long-short strategy, in which both long and short positions in stocks or stock indices are assumed, whether in the spot or derivatives market. For many years, however, most of the pair tradings used in these strategies were constructed based on analyses that do not consider the extreme deviations that exist in the interdependence process between the assets involved and the firms' operational quality indicators. The objective in this study is to verify, using the fundamental analysis proposed by Piotroski (2000) and the Extreme Value Theory, whether a new quantitative approach can be established for the long-short strategy, so as to consider, besides companies with positive signs of profitability, capital structure and operational efficiency, distributions that can capture extreme co-movements associated with the selected pair tradings.

Studies, including Gatev, Goetzmann and Rowwenhorst (2006), use approaches in which the co-movements between two assets are measured by the sum of the square differences between their respective standard price series. Nath (2006) and Do and Hamza (2006) adopt similar approaches. The probabilistic modeling of the co-movement series of the pair trading prices, however, require heavy-tailed distributions, as a large part of the arbitrage opportunities derives from these extreme movements.

Other studies, such as Alexander (1999), Alexander and Dimitriu (2005a), Dunis and Ho (2005) and Caldeira and Portugal (2010), have used the co-integration technique, which aims to explore the existing long-term dependence between time series in the context of index tracking strategies. Herlemont (2000), Alexander and Dimitiu (2005b) and Lin, McRae and Gulati (2006) also applied this method in the context of strategies that involve pair tradings. It is perceived, however, that even pairs of assets without long-term interdependence can create arbitrage windows in the short term. In that sense, more appropriate long-short quantitative techniques need to be developed to analyze the extreme co-movements of the asset pairs, whether co-integrated or not. In that context, the Extreme Value Theory plays a fundamental role because of its ability to model extreme data.

With a view to checking the possibility of using the extreme theory to detect more proper timings for arbitrage, Winner companies were selected in the Brazilian market with higher liquidity and market value. According to Piotroski (2000), when considering portfolios with such characteristics, abnormal buy-and-hold returns of up to 23% could be obtained in the American market. The results by Piotroski (2000) were also confirmed in the Brazilian market by Lopes and Galgi (2007), especially in the second year after the construction of the portfolio. Based on these results, it is expected that the use of the Fundamental Analysis in line with the Extreme Value Theory permits considering, besides companies with positive signs of profitability, capital structure and operational efficiency, distributions that can capture the extreme co-movements associated with the selected pair tradings.

Stock pairs from important companies in the Brazilian scenario were used to apply the new quantitative approach to be proposed, which was called the GEV Long-Short. The results obtained were compared with the results obtained when the adjustments are made based on the normal distribution. The GEV Long-Short strategy was implemented through the Maximum Block approach. The parameters were estimated through maximum likelihood. The estimation of the parameters and other analyses were done using the statistical software R-Project based on the package fExtremes, which specifically serves to analyze extreme values.

The obtained results suggest that the choice of pair tradings, through the fundamental analysis, and the modeling of the series of price quotients, through the distribution of extreme values, can provide a better adjustment of the extreme probability quantiles, so as to detect more proper timings to revert the downward and upward trends in the pricing ratio between asset pairs more precisely.

The obtained results also suggest that, despite the apparent synchronism between the stock prices, the co-integration condition is not necessarily a requisite for the construction of long-short strategies, like in the case of the proposed GEV Long-Short. This fact finds grounds, as the probabilistic modeling of the pricing ratio, through the extreme value distribution, does not aim to establish a relation of dependence between the triggering processes of each price series, but to probabilistically verify, based on the existence of each triggering process, moments when the pricing ratio is positioned in extreme quantiles, beyond the normality standards.

This paper contributes to the finance literature by proposing the use of the extreme data modeling as a tool that can provide information for an ex-ante monitoring of the gradual dissemination of private information and the respective relative adjustments of the stock pair prices. This monitoring gains importance as it allows the players to reconsider their expectations (Fundamental Analysis) and perceive with a shorter time lag the appropriate timing for them to operate as traders (Extreme Value Theory). The extreme event modeling reveals the proposed approach as an advance in terms of arbitrage opportunities in relation to the method that uses co-integration, as the fact that price series are not co-integrated does not exclude the possibility that arbitrage windows exist in the short term.

This paper is structured as follows: sections 2 and 3 present short introductions on the Long-Short strategy and the Extreme Value Theory; in section 4, the strategy GEV Long-Short is introduced and the comparative results are presented between the adjustments made by the extreme value distribution and the normal distribution. Section 5 presents the conclusion and other comments.

2. Hedge Funds and Long-Short Strategies

Among the funds that use the long-short strategy, the Hedge Funds stand out. Created in 1949 by Alfred W. Jones, the first Hedge Fund was a fund whose strategy was based on the purchase of underpriced stocks and the short selling of overpriced stocks with a view to protecting the portfolio against market risks, i.e., to reduce the portfolios' risk through positions sold in other stocks (Fothergill & Coke, 2000). Thus, by selling stocks which he believed had a good potential of returns and by holding a short position in stocks he was pessimistic about, Jones ended up reducing his exposure to the stock market risk.

By working with these positions, Jones reduced his risk as, if the market as a whole performed negatively, the fund would win at least in part of its positions (in this case the sold or long positions). His strategy, however, once the perspectives outlined had been confirmed, permitted the fund to grant profit in both operations, that is, through the rise of long stocks and the drop of short ones, as his position permitted purchasing them for a lower amount. As a result of this strategy, despite charging a remuneration fee of about 20% of the net profit, the fund Jones administered showed a higher return than any other fund, which aroused great interest. At the end of the 1960's, it attracted both investors and market professionals, who created hundreds of new hedge funds (Fothergill & Coke, 2000).

As a result of the crisis in the 1970's, some Hedge Funds were extinguished and it was only in the 1990's that they were reintroduced among the funds the investors took interest in (Anjivel, Boudreu, Perskin & Urias, 2000). If, theoretically, the investment strategies this type of fund adopts reduce the risk, then why do some investors associate the Hedge Funds with high-risk investments and why did they suffer that much from the crisis in the 1970's? The answer rests in the fact that the objective of the above mentioned strategy cannot be the only one a Hedge Fund adopts.

To give an example, many Hedge Funds adopt strategies that involve leverage, i.e. the use of derivatives to multiply the fund's performance. It was exactly the adoption of more aggressive strategies, with extremely levered positions, that made many Hedge Funds incur heavy losses at the end of the 1960's and early 1970's in the USA. This accumulation process of losses took until 1986, when the performance of the Tiger Fund brought the Hedge Funds back as an option for large investors (Branco & Franco, 2004).

In Brazil, there are no rules for the Hedge Funds, but for funds that sometimes act as Hedge Funds. According to the new classification by the Brazilian Central Bank and based on risk (introduced in March 2005), these portfolios figure among the generic funds. Funds of this category are allowed to maintain levered positions (Branco & Franco, 2004). In the USA, the Hedge Funds are limited liability companies in terms of the number of stockholders, with little regulation and exempt from the control of the 1940 Company Act (restrictions on leverage, short sales, risk concentration etc.), applied to the mutual funds (Liang, 2003). The lack of requirements regarding the transparency of the North American Hedge Funds' balance sheets and operations leads to particular audits. According to Liang (2003), however, these audits are limited, as 40% from a large sample is not appropriately audited. In Brazil, the regulation of the Hedge Funds' operations is much stricter, as these funds receive the same legal treatment as other funds, like fixed interest funds for example. Although this regulation in some cases presents restrictions in trade terms, it grants the investors more safety in the pricing of the respective fund quota.

Nowadays, although the Hedge Funds present other, very diversified investment strategies in terms of the degree of specialization and categorization (*opportunistic, event driven, futures and currencies arbitrage, market timing, market neutral, global, equity hedge*, etc.), the importance of Alfred W. Jones should be admitted as, based on his hedge fund, the long-short strategy was established as it is known in the stock market today.

In the next section, the main foundations of the Extreme Value Theory and the Fundamental Analysis are presented, which were used in the construction of a new quantitative approach for the long-short strategy.

3. Fundamental Analysis and Extreme Value Theory (EVT)

In the quantitative universe of the detection of arbitrage windows, it is not convenient to consider that all investors agree with the probability distribution of the asset prices. What the investors do is approximate, in function of the available information, the empirical distribution of the data to known distributions and that better adapt to the type of analysis required. Thus, in the short term, the prices may not reflect all existing information, but the set of interpretations each investor reaches regarding the publicly available information (short-term inefficiencies). Consequently, arbitrage windows are created, given each investor's different expectations in relation to the strategy and the best portfolio to allocate his resources.

The information asymmetry between the manager and the stockholders who do not directly participate in decision making, for example, is one of the factors that can justify the arbitrage possibilities. In the finance literature, the use of accounting to reduce this information asymmetry has become more frequent. Some studies aimed to relate the stock prices with financial indicators, built based on accounting information (fundamental analysis), so as to verify the existing relation between the prices and information contained in the companies' balance sheets: Ball and Brown (1968), Fama and French (1992), Baruch and Thiagarajan (1993), Fama e French (1995), Fama and French (1996), Abarbanell and Bushee (1997), Abarbanell and Bushee (1998), Ali and Hwang (2000), Bird, Gerlach and Hall (2001), Piotroski (2000), and Lopes and Galdi (2007). These studies suggest that the transparency and quality of the management process as well as the balance sheet information are factors associated with the companies' stock market performance. Based on these characteristics, Piotroski (2000) built portfolios in the American market based on companies that were classified as winners and losers.

In the next section, it will be verified that, based on the fundamental analysis proposed by Piotroski (2000) and the Extreme Value Theory, an approach can be established that permits the market players to reconsider their expectations (Fundamental Analysis) and perceive the proper timing for them to operate as traders with a shorter time lag.

3.1 Univariate Modeling of Maxima

EVT is a branch of the probability theory that studies the stochastic behavior of extremes associated with a set of random variables with common distribution F. The distribution characteristics of these extremes (distribution of maximum or minimum) are determined by the causes of the F distribution.

Some authors refer to the study by Bortkiewicz (1922) as the initial milestone in the development of the Extreme Value Theory (EVT). In that paper, Bortkiewicz discusses the distribution of the interval size between the maximum and the minimum, but in a sample with normal distribution. In more general terms, the basic foundations of the EVT were initially established by Fisher and Tippett (1928), who introduced the three possible types of asymptotic distribution of the extreme values, known today as the distributions of Gumbel, Fréchet and Weibull. One of the first to study and formalize the statistic application of this theory was Gumbel (1954), whose method will be presented next.

When analyzing a sample as a whole, only one absolute maximum (or minimum) is found, associated with a finite number of observations. To make the arrangement of the observations feasible for analysis based on the EVT, a sample can be divided in distinct sub-periods of the same size, with a view to extracting the maxima from each sub-period.

Therefore, let us consider random variables $X_1, X_2, \ldots$ i.i.d. with distribution function F and a set $X=\{X_1, X_2, \ldots, X_n\}$ constituted by n of these variables. Based on this set, a new sequence can be obtained when its terms are rearranged in rising order of magnitude, that is

$$(X_{(n)}) = \{X_{(1)}, X_{(2)}, \ldots, X_{(n)}\}$$

where $X_{(1)} = \min X$ and $X_{(n)} = \max X$.

Definition 3.1: The function $F_{X_{(n)}}(x) = P\left(X_{(n)} \leq x\right) = [P(X \leq x)]^n = F^n(x), x \in \mathbb{R}, n \in \mathbb{N}$ is the so-called maximum distribution function.

In the definition above, $P(X_{(n)} \leq x) = [P(X \leq x)]^n$, given the independence of the variables $P(\cap_{i=1}^{n} [X_i \leq x] = \prod_{i=1}^{n}[P(X_i \leq x)]$.

Observation 3.1 Although, in most cases, EVT is treated based on an approach of the maximum observations, the same results can be used in the approach of the minima, given the immediate conversion

$$\text{Min}\,(X_1, X_2, \ldots X_n) = -\,\text{Max}\,(-X_1, -X_2, \ldots -X_n)$$

Based on the distribution of the maximum and interested in the analysis of the tail of the F distribution, it is observed that the maximum converges in probability towards the upper limit x_F of the support of the F distribution, that is

$$X_{(n)} \xrightarrow{p} x_F$$

when $n \to \infty$ and for $x_F \leq \infty$, being $x_F = sup\{x \in \mathbb{R}/F(x) < 1\}$, as, for a sufficiently large n:

- if $x < x_F$, then $P(X_{(n)} \leq x) = F^n(x) \to 0$
- if $x \geqslant x_F$, with $x_F < \infty$, then $P(X_{(n)} \leq x) = F^n(x) \to 1$

For small n values, however, previous knowledge of the F distribution is needed (generally unknown) to establish the distribution of the maximum and, for sufficiently large n values, the distribution function of the maximum becomes degenerate, that is, $P(X_{(n)} \leq x) \to 0$.

Fisher and Tipet (1928), however, established a result in which the distribution of the standard maxima per sequence of constants (a_n) and (b_n) converges towards certain borderline distributions, today called extreme distributions or Extreme Value (EV) distributions. This result is known today as the Fisher-Tippett theorem.

Theorem 3.1 (Fisher-Tippett Theorem) – $(x_n)_{n \in \mathbb{N}}$ is a sequence of random variables i.i.d with distribution function F. If a sequence (a_n) of positive terms exists, a real sequence (b_n) and a non-degenerate distribution function H, so that

$$P[X_{(n)} \leq a_n x + b_n] = F^n(a_n x + b_n) \xrightarrow{d} H(x)$$

then the only possible forms of H are the Gumbel, Fréchet or Weibull distributions, also called type I, II and III distributions, respectively.

These borderline distributions are associated with a λ parameter. Next, the borderline distributions for maxima are presented, as well as the respective graphs of the extreme distributions with a shape parameter equal to 0, 1 and -1, respectively:

- Gumbel $\lambda = 0, x \in \mathbb{R}$): $F_\lambda(x) = \exp[-\exp(-x)]$
- Fréchet $\lambda > 0, x \in \mathbb{R}$): $F_\lambda(x) = \begin{cases} \exp(-x)^{-\lambda}, & se\ x > 0 \\ 0 \quad , & se\ x \leq 0 \end{cases}$
- Weibull ($\lambda < 0, x \in \mathbb{R}$): $F_\lambda(x) = \begin{cases} \exp[-(-x)^{-\lambda}], & se\ x \leq 0 \\ 1 \quad , & se\ x > 0 \end{cases}$

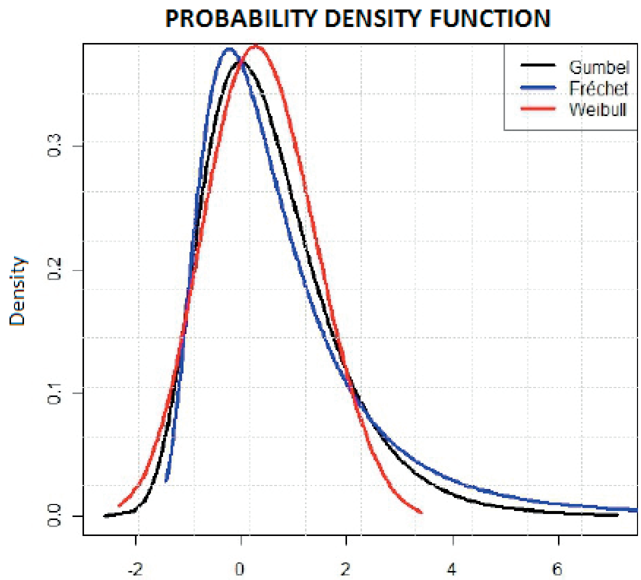

Figure 1. Graph of extreme distributions with shape parameters equal to 0, 1 and -1, respectively

The Fisher-Tippett Theorem specifies the borderline extreme distributions for which there exists a possibility of convergence of the standard maximum distribution $\frac{X_n - b_n}{a_n}$, although it does not establish conditions F should comply with for this convergence to occur (Embrechts, Klippelberg Mikosch, 1997).

The EV distributions can be estimated without the need to use the F distribution. After estimating the Borderline distribution, however, inferences about the Empirical distribution F can be made as, based on $H(x) = F^n(x)$, the following is obtained:

$$F(x) = [H(x)]^{1/n}$$

In studies previous to those published by Von Mises (1936) and Jenkison (1955), however, the λ parameter, which specifies the borderline distribution, was estimated by supposing the convergence of the theoretical distribution F for each of the borderline distributions, with further application of tests to verify the extreme distribution with better adjustment quality. The theoretical advance these authors introduced was very significant as, under the ξ-parametrization, or representation of Jenkison-van Mises (1936) and Jenkison (1955), the Gumbel, Frichét and Weibull distributions can be generalized in a form called Generalized Extreme Value (GEV), which represents a family of distribution functions of a single parameter ξ. In other words, the parametrization of the extreme functions permitted the estimation of the parameters of true borderline distribution. Below is the Generalized Extreme Value:

$$G_\xi\big(T(x)\big) = G_{\xi,\mu,\sigma}(x) = \begin{cases} \exp\left(-\left(1+\xi\dfrac{x-\mu}{\sigma}\right)^{\frac{-1}{\xi}}, \quad \xi \neq 0, \quad 1+\xi\dfrac{x-\mu}{\sigma} > 0\right) \\ \exp\left[-\exp\left(-\dfrac{x-\mu}{\sigma}\right)\right], \quad \xi = 0 \end{cases}$$

For the borderline case ($\xi \to 0$), G_ξ corresponds to the Gumbel distribution. Also, if $\xi < 0$, H_ξ corresponds to the Weibull distribution and, if $\xi > 0$, H_ξ corresponds to the Fréchet distribution.

The probability density function (fdp) of the generalized value $G_{\xi\mu\sigma}$ can be obtained by differentiation and application of the chain rule, as follows:

$$G_{\xi,\mu,\sigma}(x) = \begin{cases} \left(1+\xi\dfrac{x-\mu}{\sigma}\right)^{\left(\frac{-1}{\xi}-1\right)} \exp\left[-\left(1+\xi\dfrac{x-\mu}{\sigma}\right)^{\frac{-1}{\xi}}\right], \quad \xi \neq 0, \quad \mu \in R, \quad \sigma > 0 \\ \exp\left[-\exp\left(-\dfrac{x-\mu}{\sigma}\right)\right] \exp\left(-\dfrac{x-\mu}{\sigma}\right)\dfrac{1}{\sigma}, \quad \xi = 0, \quad x \in \mathbb{R}, \quad \mu \in \mathbb{R}, \quad \sigma > 0 \end{cases}$$

Figure 2 shows the graphs of the fdp associated with each of the three distributions.

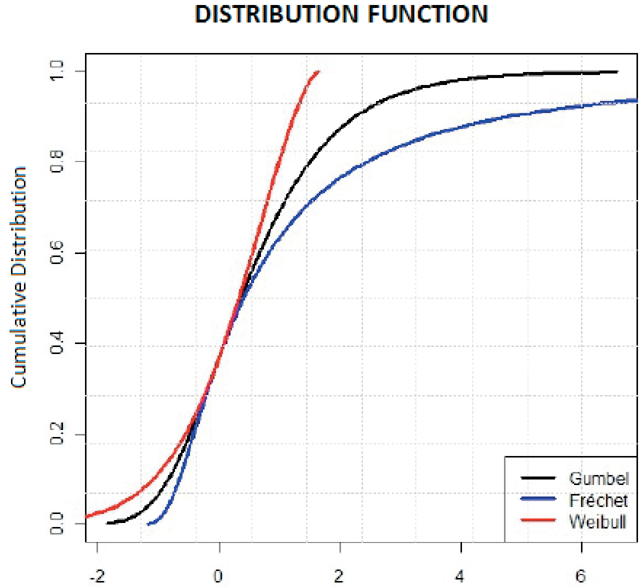

Figure 2. fdp graph of extreme values with shape parameter equal to 0, 1 and -1, respectively

4 . New Quantitative Long-Short Proposal

In this section, the new quantitative long-short approach is presented, constructed based on an empirical research in which data were used from companies listed on BM&FBovespa for the period 2005-2008. Besides the prices, accounting information was used obtained from the database Economática to select the winner companies, according to the method proposed by Piotroski (2000). As the focus is on verifying the possible use of the extreme value theory to detect the most appropriate timing for arbitrage, only winner companies with greater liquidity and market value were selected. According to Piotroski (2000), when considering such portfolios, abnormal returns of up to 23% could be obtained in buy-and-hold strategies. The results by Piotroski (2000) were also confirmed in the Brazilian market by Lopes and Galdi (2007), especially in the second year after the construction of the portfolio. Table 1 presents the 9 indicators used to construct the score proposed by Piotroski (2000).

Among the companies in the upper quintile (winners), the companies with the highest liquidity and market (Lopes & Galdi, 2007). Based on the modeling established with the Extreme Value Theory of the selected companies' price indices, the selected winners' stocks were monitored to verify the moment when the price ratio reached the extreme quantiles provided by the modeling of the extremes. The selected companies were monitored during the two years after the constitution of the winner portfolio, according to the proposal by Lopes and Galdi (2007).

4.1 GEV Long-Short strategy

The simplest long-short strategy the players adopt, also called strategy number one, is a type of pair trading strategy and is based on the hiring of a certain stock that is overpriced in relation to the price of another specific stock (Nicholas, 2000). The objective of this strategy is to obtain an additional return and which is not related to the valuation or devaluation of these assets, but to the relative performance between the long and short positions.

In this operation, the hired stock is immediately sold and the amount obtained is fully invested in the purchase of another stock in the pair trading. At the end of the maturation period of the rent, it is expected that the pair trading has a pricing ratio (k) in relation to the initial lower position, sufficient to cover the general transaction costs and generate the desired additional returns, that is, it is expected that the sale of the long stocks is higher than the value needed to purchase the hired stocks and which should be returned together with the trading fee, generally 1% (Nicholas, 2000). Similarly, if k is obtained through the price ratio of the undervalued and overvalued assets, respectively, then it is expected that the k ratio is high enough to be able to reach the desired results.

Even in a falling market, positive profitability can be obtained. It is sufficient for the positions to be established so that the short asset is further devalued than the long asset. That is, among others, one of the important benefits the long-short strategies can provide as, besides permitting positive profitability rates, they protect the portfolio against systemic risks of a market decline.

The main question involving the new quantitative long-short implementation proposed is based on the modeling (through the EVT) of the series obtained based on the quotient of the pair tradings that is to be arbitrated, so as to detect, for example, with a 95% confidence level, the moment when this ratio reaches extreme quantiles of the adjusted GEV, that is, in order to *a priori* determine arbitrage opportunities between asset pairs.

Therefore, in this study, the strategy proposed by Piotroski (2000) to select winner companies was adopted, using accounting information from 2005 and 2006. From the winner portfolio, the companies with the highest liquidity levels were selected to construct the pair tradings, which were monitored over the two years after the constitution of the winner portfolio (2007 and 2008), according to the proposal by Lopes and Galdi (2007).

Table 1
Accounting indicators and variables used in the construction of the F-Score

Indicador	Variable	Score
ROA_{it}	LL_{it}/AT_{it-1}	↑ ROA > 0 (1)
		↓ ROA < 0 (0)
CF_{it}	$(CXECX_{it} - CXECX_{it-1})/AT_{it-1}$	↑ CF > 0 (1)
		↓ CF < 0 (0)
ΔROA_{it}	$ROA_{it} - ROA_{it-1}$	↑ Δ ROA > 0 (1)
		↓ Δ ROA < 0 (0)
$ACCRUAL_{it}$	$[LL_{it} - (CXECX_{it} - CXECX_{it-1})]/AT_{it-1}$	↑ CF > ROA (1)
		↓ CF < ROA (0)
$\Delta LIQUIDEZ_{it}$	$(AC_{it} / PC_{it}) - (AC_{it-1}/PC_{it-1})$	↑ ΔLIQ > 0 (1)
		↓ ΔLIQ < 0 (0)
$\Delta ENDIVID_{it}$	$[(PC_{it} + PELP_{it})/AT_{it}] - [(PC_{it-1} + PELP_{it-1})/AT_{it-1}]$	↑ ΔENDIVID > 0 (1)
		↓ ΔENDIVID < 0 (0)
STOCK $OFFER_{it}$	Stocks offered in the year before the construction of the portfolio	↑ OFER = 0 (1)
		↓ OFER > 0 (0)
$\Delta MARGEM_{it}$	$(LB_{it} / REC_{it}) - (LB_{it-1}/REC_{it-1})$	↑ ΔMARG > 0 (1)
		↓ ΔMARG < 0 (0)
$\Delta GIRO_{it}$	$(REC_{it}/AT_{it}) - (REC_{it-1}/AT_{it-1})$	↑ ΔGIRO > 0 (1)
		↓ ΔGIRO < 0 (0)

Obs.: The F-Score (Piotroski, 2000) is an indicator constructed based on the sum of scores related to accounting indicators regarding profitability, capital structure and operational efficiency. According to this classification, each company i can have a minimum score of zero and a maximum score of nine in each fiscal year. The definitions of the variables that make up the indicators in the F-score are presented next:

- CXECX = Cash and cash equivalents of company i in year t;
- AC = Current assets of company i in year t;
- PC = Current liabilities of company i in year t;
- LB = Gross profits of company i in year t;
- PELP = Long-term liabilities of company i in year t;
- REC = Sales revenues of company i in year t;
- LL = Net profit of company i in year t;

Source: Nossa, Teixeira and Lopes (2010, p. 8), adapted from Piotroski (2000) and Lopes and Galdi (2007) (TRAD: STOCK OFFER – stocks offered in the year before the construction of the portfolio – indicator-variable)

To be able to analyze the price series based on the extreme theory, a period needs to be chosen in which both assets exist, that is, so as to have available information on the asset prices for all data in the period chosen. In this analysis, the period chosen to obtain the extreme quantiles does not incorporate the financial crisis in 2008, as this crisis caused singular impacts in the behavior of many assets. Also, the number of data collected should be sufficiently large to be able to divide the series of the price quotient of the pair trading to be arbitrated in blocks (Maximum Block approach).

To be able to verify the best adjustments of the GEV and guarantee the stability of the adjusted distribution, a comparative analysis was elaborated through models with blocks of 5, 10, 15 and 20 observations, so as to obtain four series with maximum and minimum indices for periods of 1, 2, 3 and 4 weeks, respectively.

Kolmogorov-Smirnov statistics were used to test whether the supposition that the distribution of the maximum and minimum data analyzed converge towards the GEV distribution. A residue analysis through QQ-Plot was elaborated to verify the quality of the GEV adjustment for the different block sizes considered.

After adjusting the GEV for the series of maxima and minima and after verifying for which block size the stability is the highest, the 95% and 5% quantiles were calculated for the series of maxima and minima related to that block size, respectively. This adjustment, through the GEV, leads to the definition:

Definition 4.1 (GEV Long-Short quantile) – $\{X_1, X_2, ..., X_n\}$ is a set of random variables that represent the pricing ratio between assets. The 95% (or 5%) quantile obtained through the extreme distribution for which the distribution of the standard maxima (or minima) converges is called the GEV Long-Short quantile.

The GEV Long-Short quantiles, together with the quantiles provided by the normal distribution, were overlapped with the series of indices for a period subsequent to the analysis period. As inferences on the actual F distribution of the series can be made based on the adjusted GEV, the true quantiles of the series were also overlapped in this graph, both obtained based on the estimated GEV. These quantiles were used to verify the moments when the pricing ratio occupied extreme quantiles. The search for the most appropriate arbitrage timing leads to the definition:

Definition (Relative disequilibrium saturation) – Two assets are in a relative disequilibrium saturation situation when the pricing ratio of these assets lies beyond the interval defined by the GEV Long-Short quantiles.

Based on the detection of these windows (relative disequilibrium saturation points), arbitrage sub-strategies were structured that start in long and short positions and that settle or increase the positions, depending on the behavior of the returns offered by the arbitrated asset pair. A 3% profitability level for the operation was set as the stop point for the operation. It should be highlighted that the profitability level to stop the operation should not necessarily be 3%, but can vary according to the risk the investor is willing to incur and the desired profitability percentage.

Based on the information obtained and the elaboration of the sub-strategies, the profitabilities provided by the GEV approach and by the approach that uses the Normal distribution were calculated. The comparison of these strategies is essential to confirm the supposition that the better adjustment of the extreme quantiles through the GEV can provide more refined probabilistic grounds for the return to the average to justify the possibility of long-short arbitrage.

The winner asset pairs PETR4 x VALE5 and VIVO4 x TCSL4 were used for the application of the new quantitative approach proposed. The first case is an example of arbitrage between companies from different sectors, which is very representative of all negotiations in the Brazilian market. The second pair refers to arbitrage between winner assets in the same sector, which at the time was going through an important stage of growth in Brazil.

Despite the existence of synchronism between the asset pair prices considered in both cases, this synchronism is not necessarily a requisite to be able to construct long-short strategies, like in the case of the proposed GEV Long-Short strategy. This fact is sustained as the probabilistic modeling of the pricing ratio through the GEV does not aim to establish a relation of dependence between the triggering processes of each asset price series, but to probabilistically verify, based on the existence of each triggering process, moments when the pricing ratio occupies extreme quantiles and goes beyond the usual patterns. This interpretation turns the analysis method into an advance in terms of arbitrage opportunities in comparison with the method that uses, for example, co-integration, as the fact that asset price series are not co-integrated does not exclude the possibility of existing short-term arbitrage windows. Based on the Engle-Granger test, it was verified that the asset pair PETR4 x VALE5 is not co-integrated and that the pair VIVO4 x TCSL4 is co-integrated.

4.2 PETR4/VALE5 and VIVO4/TCSL4

Next, a comparative model is presented of the ratio between the prices of PETR4 and VALE5, based on the adjustments of the GEV and Normal distributions. The collected data constitute a sample of 651 observations of the daily asset prices for the period from 1/1/2005 till 7/2/2007.

Based on the maximum likelihood method, the estimates were obtained for the parameters ξ, μ and σ of the GEV in the maxima and minima modeling, respectively, for blocks of 5, 10, 15 and 20 days.

Figures 3 and 4 refer to the QQ-plot of the residues for the models related to the maxima and minima for the different block sizes. In the minima model, the blocks of 20 days guarantee greater stability to the GEV, as the residue plot demonstrated a better adjustment. In the maxima model, the blocks of 15 days guaranteed greater stability. Next, the estimates of the parameters ξ, μ and σ are presented for the maxima and minima models of 15 and 20 days, respectively.

$$\xi = -0{,}3945844 \qquad \mu = 1{,}6959661 \qquad \sigma = 0{,}2533173$$

$$\xi = -0{,}5216825 \qquad \mu = 1{,}5910817 \qquad \sigma = 0{,}2604460$$

The Kolmogorov-Smirnov statistics were used to test the supposition that the maximum and minimum distribution of the analyzed date converges to the adjusted GEV distribution. The p-values obtained in the modeling of the maxima and minima corresponded to 0.9429 and 0.9373, respectively, sufficient to guarantee the convergences.

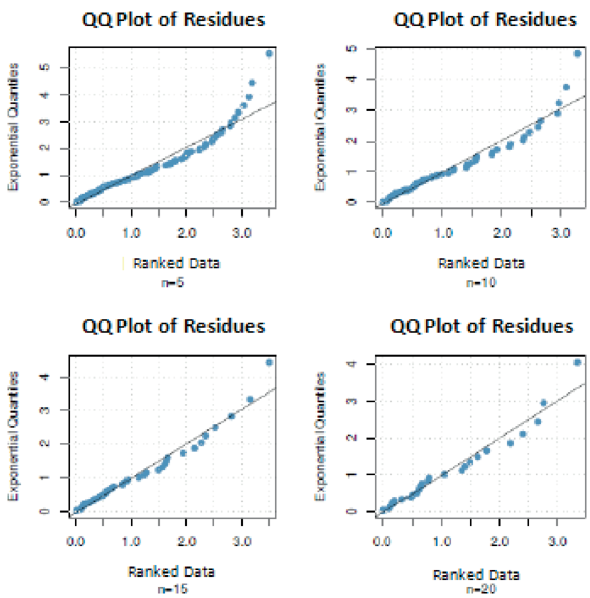

Figure 3. QQ-PLot of residues for maxima modeling

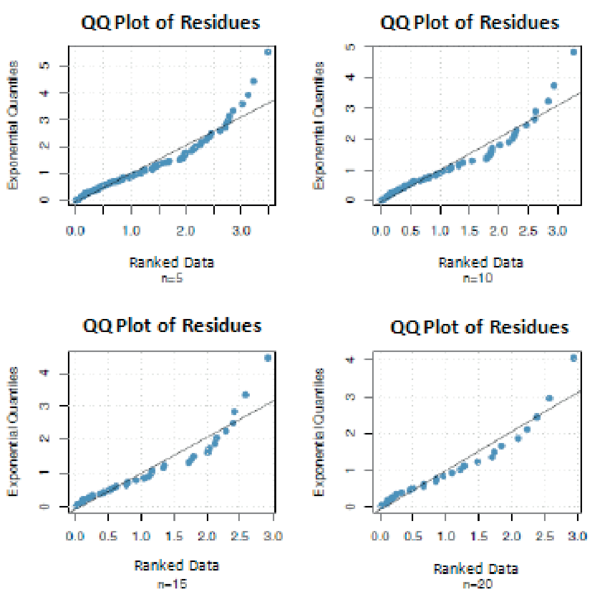

Figure 4. QQ-PLot of residues for minima modeling

Based on the parameters obtained, the Adjusted GEV for minima and maxima provides 1.183819 and 2.139097 as 5% and 95% quantiles for the index k = PETR4.SA/VALE5.SA, respectively. When using a Normal distribution, for the same significance level, the values 1.299182 and 2.109515 are obtained, respectively. Figure 5 compares the probability quantiles obtained as a result of the performance of the k ratio for the 131 days subsequent to the analysis period. If the players knew the quantiles estimated by the Normal and GEV distributions, based on the daily monitoring of the evolution in the asset price index, the following arbitrage strategy could be constructed:

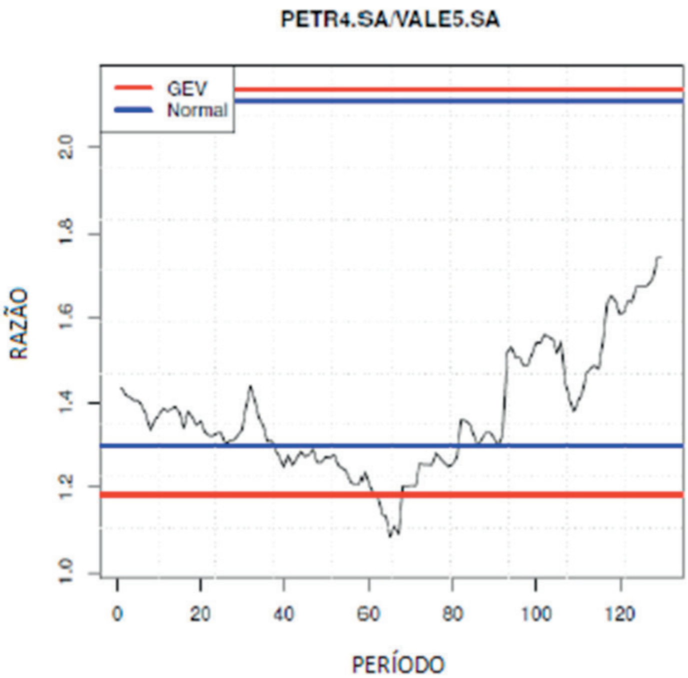

Figure 5. Comparison of probability quantiles obtained through the performance of the index k = PETR4.SA/VALE5.SA for the 131 days subsequent to 07/02/2007.

1. When monitoring the evolution of the index k = PETR4.SA/VALE5.SA, it is verified that, on 8/8/2007 (period 26), long positions could be acquired in PETR4 and short positions in VALE5 as, on that day, the k index reaches the 5% quantile estimated by means of the Normal distribution;
2. When setting an accumulated profitability of 3% as the stop point for the operation, these positions would stop on 8/15/2007 (period 32), five workdays after the start of the operation (see Figure 6);
3. As the modeling based on the Normal distribution overestimates the lower quantile, instead of simply stopping the position, the players could invert the positions, that is, acquire a long position in VALE5 and a short position in PETR4 as, in that period, the k ratio did not reach the inferior support provided by the GEV;
4. The new position could be maintained until the index reached the inferior quantile indicated by the GEV, which happened on 9/28/2007 (period 61), 31 workdays after the inversion. The graph in Figure 7 shows the accumulated profitability as from the inversion of the positions acquired.

The analysis of the strategy as a whole shows a joint profitability of 12.19% for a period of 36 workdays.

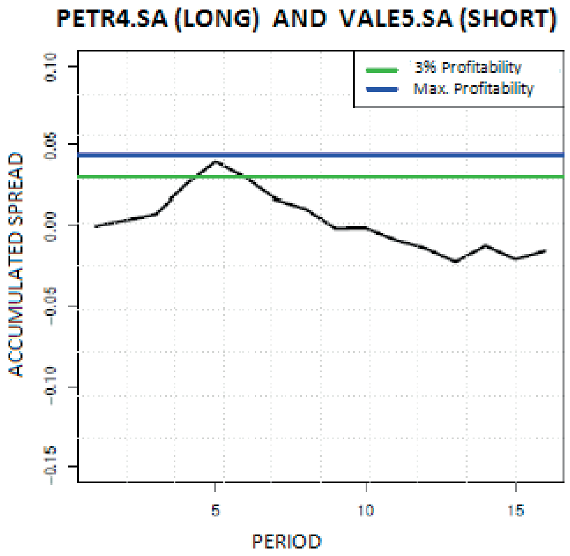

Figure 6. Accumulated profitability of positions based on the normal quantile

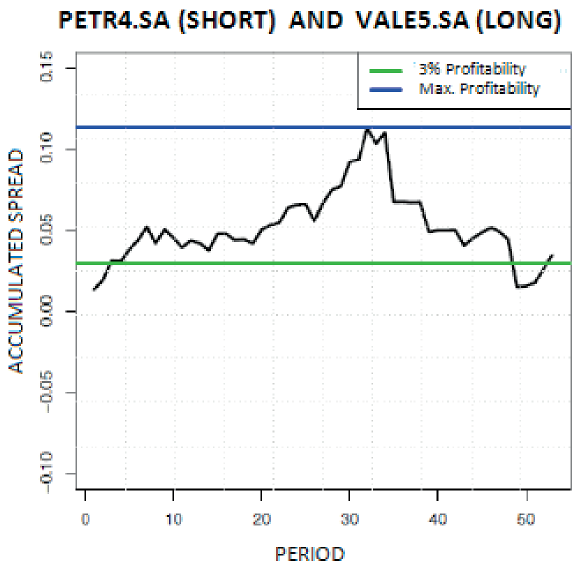

Figure 7. Accumulated profitability of positions based on the inverted positions

The same analysis was developed for the pair VIVO4 and TCSL4, whose collected data constituted a sample of 396 observations of the daily asset prices for the period from 1/6/2006 till 12/15/2007. In this case, both in the maxima and minima models, the blocks of 5 guaranteed greater max-stability to the GEV.

Based on the parameters obtained, the adjusted GEV for minima and maxima provides 0.9042828 and 1.455642 as 5% and 95% quantiles for the index k = VIVO4/TCSL4, respectively. When using a normal distribution, for the same significance level, the values obtained are 0.9285019 and 1.424783, respectively. In Figure 8, the probability quantiles are compared that were obtained based on the performance of the k index for the 54 days subsequent to the analysis period. Figure 8 also includes the inferences about the actual probability quantiles, i.e. 1.052407 and 1.505114, respectively.

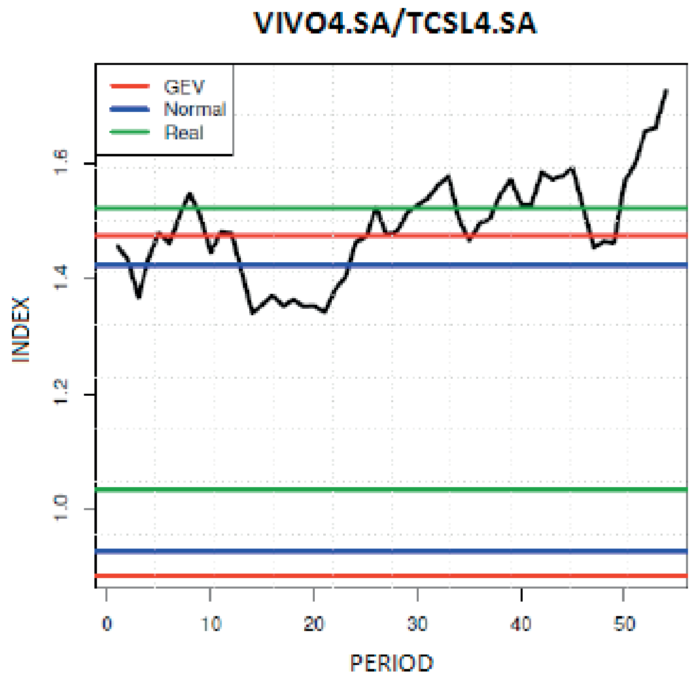

Figure 8. Performance of index k = PETR4.SA/VALE5.SA for the 54 days subsequent to the period between 06/01/2006 and 12/15/2007

If the players knew the quantiles estimated by the normal and GEV distributions, based on the daily monitoring of the evolution in the asset price index, the following arbitrage strategy could be set up:

1. When monitoring the evolution of the index k = VIVO4/TCSL4, it is verified that, in period 2, the pricing index k maintains a downward trend and its value very closely approximates the 95% quantile provided by the normal distribution. Although the index remains below the superior supports of the GEV, the k index is close to these supports, which reinforces the expected continuing drop in the pricing index. Thus, long positions could be acquired in TCSL4 and short positions in VIVO4, setting an accumulated profitability of 2% to invert the positions. Based on that limit, it is verified that this inversion would happen at the end of the same day, i.e. one workday after the start of the operation (see Figure 9), based on the achievement of an accumulated profitability of 2.04%.
2. When reaching the desired profitability, a long position is acquired in VIVO4 and a short position in TCSL4. The inversion of the positions is justified by the market's natural reaction to the continuous downward tend the k index was submitted to.
3. After acquiring the new positions, the stop point of the operation can be determined when k reaches the actual superior support of the data distribution or when an accumulated profitability of 2% is obtained, counted as from the point when k breaks the superior limit provided by the GEV. In this case, that moment happens 5 workdays after the start of the inversion, when the pricing index reaches the actual value of the distribution support. For that period, the operation provides an accumulated profitability of 5.38% (see Figure 10), equal to the maximum profitability that could be obtained in that period. This fact is yet another factor that proves the better adjustment of the extreme quantiles based on the GEV.

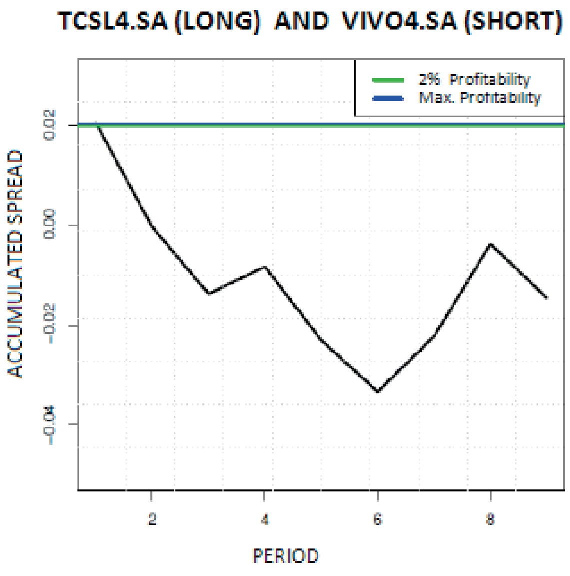

Figure 9. Accumulated profitability of positions based on the normal quantile

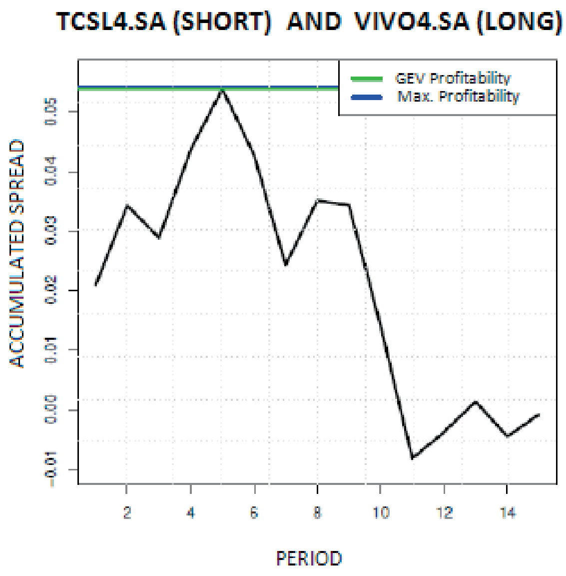

Figure 10. Accumulated profitability of positions based on the inverted positions

When analyzing the strategy as a whole, a joint profitability of 7.42% is obtained for a period of 6 workdays. The strategy used could also be expanded to again invert the positions, as the superior support of the actual data distribution was reached. The stop point of the operation could be determined if an accumulated profitability of 2% is obtained, counted as from the point when k breaks the upper limit provided by the GEV. If this inversion were used, 3.99% would be incorporated into the joint profitability of the operation after 5 workdays.

More aggressive arbitrators could adopt the point when an accumulated profitability of 2% is obtained as the stop of the above strategy, counted as from the moment when k breaks the superior limit provided by the normal distribution (and not by the GEV). If these arbitrators' expectations were confirmed, the market would return to the initial situation at the start of the entire strategy, in view of the downward trend, breaking the superior quantile provided by the normal distribution. This fact is confirmed for the analysis period.

5. Conclusion

In the quantitative universe of the detection of arbitrage windows, it cannot be considered that all investors agree with the probability distribution of the asset prices. That is so because, in most cases, the actual distribution of the prices, and the respective returns, is unknown. What the investors do, in fact, is to approximate the available data distribution to known distributions that better adapt to the type of analysis required, in function of their available information. As a result of this fact, in practice, we cannot guarantee the existence of a set of portfolios and strategies that is efficient for all investors (which does not imply the non-theoretical existence of the efficient border and of profitability maximizing strategies). Consequently, arbitrage windows are created, given each investor's different expectations regarding the best portfolio to allocate his resources, the evolutionary trends of financial asset prices and the strategies that are to be used.

The speed and efficiency with which the investors interpret the information cannot be considered homogeneous either, even if considering that the information is instantly available to all investors.

As a result of the time lag caused by the processing factor and the interpretation quality caused by the efficiency factor, one cannot consider the unlimited rationality and perfect information hypotheses yet, as modern finance does. In that sense, the financial asset prices do not reflect all available information at all times, but the set of interpretations each investor reaches based on that information.

In that sense, the search for analysis methods that better incorporate the investors' reactions to the changes in the market's behavior become increasingly necessary. As verified based on the examples, the modeling (using the EVT) of the series obtained, based on the arbitrated pair trading price quotients, showed that the GEV better adjusted the extreme probability quantiles. This result was obtained when detecting, with a 95% confidence level, a more appropriate timing to revert the downward and upward trend in the pricing index between the asset pairs chosen.

The asset pair PETR4 and VALE5 is very representative of all negotiations in the Brazilian market. Despite an apparent synchronism between the prices of the two assets considered, it was verified based on this example that the condition of co-integration is not a necessary requisite to construct long-short strategies, like in the case of the GEV Long-Short Strategy proposed. This fact is sustained as the probabilistic modeling of the pricing ratio through the GEV does not aim to establish a relation of dependence between the triggering processes of each asset price series, but to probabilistically verify, based on the existence of each triggering process, moments when the pricing ratio occupies extreme quantiles and goes beyond the usual patterns. This interpretation turns the analysis method into an advance in terms of arbitrage opportunities in comparison with the method that uses, for example, co-integration, as the fact that asset price series are not co-integrated does not exclude the possibility of existing short-term arbitrage windows.

Besides the positive profitabilities obtained for the asset pairs considered, the results suggest that the modeling based on the GEV can provide information on short-term arbitrage windows, even for asset pairs that are not interdependent in the long term. The results also suggest that the better adjustment of the extreme quantiles through the GEV can provide more refined probabilistic support for the return to the average to justify the possibility of long-short arbitrage. It should be observed, however, that although the data modeling based on the normal distribution overestimated the inferior quantiles and underestimated the superior ones, these quantiles signaled secondary growth trends for the asset pairs considered.

The use of the theoretical quantiles of the GEV can provide information for an ex-ante monitoring of the gradual dissemination of private information and the respective relative price adjustments. This monitoring is important as it allows the players to reconsider their expectations and perceive the proper timing for them to operate as traders with a short time lag. The high performances obtained through the monitoring of the two most important stocks in the Brazilian market and two important stocks in the communication sector confirm the important of the new quantitative long-short strategy proposed as an analysis method in the arbitrage market.

6. References

Abarbanell, J. S. & Bushee, B. J. (1998). Abnormal return to a fundamental analysis strategy. The Accounting Review. *73* (1), pp. 19-45.

Abarbanell, J. S. & Bushee, B. J. (1997). Fundamental analysis, future earnings, and stock prices. Journal of Accounting Research, *35*(1), pp. 1-24.

Alexander, C. (1999). Optimal Hedging Using Cointegration. Philosophical Transactions of the Royal Society, London, Series A, *1*(357), pp. 2039-2058. Recuperado em 11/04/2014 de http://www.icmacentre.ac.uk/pdf/cointegration.pdf.

Alexander, C. & Dimitriu, A. (2005a). Indexing and statistical arbitrage: tracking error or cointegration? Philosophical Transactions of the Royal Society, Journal of Portfolio Management, *31*(1), pp. 5063-5078.

Alexander, C. & Dimitriu, A. (2005b). Indexing, cointegration and equity market regimes. International Journal of Finance and Economics, *10*(3), pp. 213-231.

Ali, A. & Hwang, L. (2000). Country-specific factors related to financial reporting and the value relevance of accounting data. Journal of Accounting Research, *31*(1), pp. 1-21.

Anjivel, S.I., Boudreu, B.E., Perskin, M.W. & Urias M.S. (2000). Why hedge funds make sense? Quantitative Strategies Research Memorandum, Morgan Stanley.

Ball, R. & Brown, P. (1968). An Empirical evaluation of accounting income numbers. Journal of Accounting Research, *6*(2), pp. 159-178.

Baruch, L. & Thiagarajan, S. R. (1993). Fundamental information analysis. Journal of Accounting Research, *31*(2), pp. 190-215.

Bird, R., Gerlach, R. & Hall, A. D. (2001). The prediction of earnings movements using accounting data: an update and extension of Ou and Penman. Journal of Asset Management, *2*(2), pp. 180–195. doi:10.1057/palgrave.jam.2240043

Bortkiewicz, L. V. (1922). Variationsbreite und mittlerer Fehler. Sitzungsber, Berli. Math. Gess, *3*(21), pp. 3-11.

Branco, G.C. & Franco, D. (2004). Risco e retorno nos hedge funds brasileiros. 4º Encontro da Sociedade Brasileira de Finanças – SBFIN. Rio de Janeiro.

Caldeira, J.F. & Portugal, M.S. (2010). Estratégia Long-Short, Neutra ao Mercado, e Index Tracking Baseadas em Portfólios Cointegrados. Revista Brasileira de Finanças, *8*(4), pp. 469-504.

Dunis, C.L. & Ho, R. (2005). Cointegration portfolios of european equities for index tracking and market neutral strategies. Journal of Asset Management, 1(1), pp. 33-52.

Do, B.F. & Hamza, K. (2006). A new approach to modeling and estimation for pairs trading. Working Paper Series. Monash University. Recuperado em 11/04/2014 de https://statarb.googlecode.com/files/0PairsTrading_BinhDo.pdf.

Embrechts, P., Kluppelberg, C. & Mikosch, T. (1997). Modelling extremal events for insurance and finance. Springer-Verlag: Berlin.

Fama, E. F. & French, K. R. (1992). The cross section of expected stock returns. The Journal of Finance, *47*(2), pp. 427-465.

Fama, E. F. & French, K. R. (1995). Size and book-to-market factors in earnings and returns. The Journal of Finance, *50*(1), pp. 131-155.

Fama, E. F. & French, K. R. (1996). Multifactor explanations of asset pricing anomalies. Journal of Finance, *51*(1), pp. 55-84.

Fisher, R. A. & Tippett, L. H. C. (1928). Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. Proccedings of the Cambridge Philosophical Society, *1*(24), pp. 180-190.

Fothergill, M. & Coke, C. (2000). Funds of Hedge Funds. An Introduction to Multi-manager Funds. Deutsche Bank.

Gatev, E., Goetzmann, W. & Rouwenhorst, K. (2006). Pairs trading: Performance of a relative-value arbitrage rule. Review of Financial Studies *3*(19), pp. 797-827.

Gumbel, E. J. (1954). Statistics Theory of Extreme Values and Some Pratical Applications. Nat. Bureau of Standards Applications Mathmatics Series, *2*(33), pp. 1-51.

Gumbel, E. J. (1958). Statistics of Extremes. Columbia University Press, New York Journal of Business, *1*(63), pp. 383-408.

Herlemont, D. (2000). Pairs Trading, Convergence Trading, Cointegration. YATS Working Papers. YATS Finances Technologies.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quart. J. Roy. Meteo. Soc. *1*(81), pp. 158-171.

Liang, B. (2003). Hedge fund returns: Auditing and accuracy. The Journal of Portfolio Management *29*(3), pp. 111-122.

Lin, Y., McRae, M. & Gulati, C. (2006). Loss Protection in Pairs Trading through Minimum Profit Bounds: A Cointegration Approach. Journal of Applied Mathematics and Decision Sciences, *6*(16), pp. 114-130.

Lopes, A. B. & Galdi, F. C. (2007). Financial statement analysis generate abnormal returns under adverse conditions? In: Annual Meeting of the American Accounting Association. Chicago, Illinois.

Loomis C. J. (1966). The Jones Nobody Keeps Up With. Fortune Magazine, pp. 237-248.

Nath, P. (2006). High Frequency Pairs Trading with U.S Treasury Securities: Risks and Rewards for Hedge Funds. Working Paper Series. London Business School.

Nicholas, J. G. (2000). Market-Neutral Investing: long/short hedge funds strategies. Bloomberg Professional Lybrary, New York, ed. 1.

Piotroski, J. D. (2000). Value investing: the use of historical financial statement information to separate winners from losers. Journal of Accounting Research, *38*, pp. 1-41.

Sanfins, M. A. S. (2009). Copulas para distribuições generalizadas de valores extremos multidimensionais. Rio de Janeiro: UFRJ/IM.

Von Mises, R. (1936). La distribution de la plus grande de n valeurs. Reprinted in Selected Papers Volumen II, American Mathematical Society, Providence, R.I., 1954, pp. 271-294.