Revista Eletrônica de Estratégia & Negócios

ISSN 1984-3372
http://portaldeperiodicos.unisul.br/index.php/EeN/index

LINEAR AND NONLINEAR ASSOCIATION MEASURES WITH INTRADAY/HIGH FREQUENCY DATA FOR ALL IBOVESPA STOCKS

MEDIDAS DE ASSOCIAÇÃO LINEAR E NÃO LINEAR COM DADOS DE ALTA FREQUÊNCIA INTRADIÁRIA PARA TODAS AS AÇÕES IBOVESPA

MEDIDAS DE ASOCIACIÓN LINEAL Y NO LINEAL CON DATOS DE ALTA FRECUENCIA EN UN DÍA PARA TODAS LAS ACCIONES IBOVESPA

Alexandre Costa
Doutorando no PPGA da Universidade Federal de Santa Maria (UFSM)
Mestre em Administração pela UFSM
Endereço: Av. Roraima, n. 1000, Prédio 74C, Sala 4303, Santa Maria, RS, Brasil
Telefone (55) 3220-9258
E-mail: xandicosta@gmail.com

Alexander Souza Block
Administrador pela Universidade Federal do Pampa
Mestre em Administração pela UFSM
Endereço: R. Luiz Joaquim de Sá Britto, s/n, Itaqui, RS, Brasil
Telefone: (55) 3432-1854
E-mail: block.as@outlook.com

Paulo Sérgio Ceretta
Professor na Universidade Federal de Santa Maria (UFSM)
Doutor em Engenharia de Produção pela Universidade Federal de Santa Catarina (UFSC)
Endereço: Av. Roraima, n. 1000, Prédio 74C, Sala 4303, Santa Maria, RS, Brasil
Telefone (55) 3220-9258
E-mail: ceretta10@gmail.com

Artigo recebido em 18/03/2015. Revisado por pares em 13/07/2015. Reformulado em 23/12/2015. Recomendado para publicação em 23/12/2015 por Ademar Dutra (Editor Científico). Publicado em 30/12/2015. Avaliado pelo Sistema *double blind review*.

Programa de Pós-graduação em ADMINISTRAÇÃO

UNISUL

RESUMO

Analisamos quatro tipos de medidas de associação, a Correlação de Pearson, o Coeficiente de Determinação, a Associação Multidimensional e o Coeficiente Máximo de Informação, os dois primeiros lineares e os outros dois não lineares. Utilizamos 10 minutos de dados intradiários de todas as ações Ibovespa, que conta para o principal índice de ações no mercado do Brasil. Na literatura financeira não há muito sobre os métodos utilizados neste trabalho, de onde surgiu a motivação para o estudo. A metodologia é significativa para comerciantes, assim como algumas ações são altamente correlacionadas com o principal índice, um pode ser a base estratégica quando estão operando independentemente em um dado dia, como este padrão poderia reverter o significado. Um dos mais importantes achados deste trabalho é que o tratamento dos dados como não lineares produziram resultados mais fortes.

Palavras-chave: Associação Multidimensional; Coeficiente Máximo de Informação; Coeficiente de Determinação; Ibovespa; Intradiário.

ABSTRACT

We analyze four types of association measurements, the Pearson's Correlation, the Coefficient of Determination, the Multidimensional Association and the Maximal Information Coefficient, the two first linear, and the other two, nonlinear. We utilize 10 minutes intraday data of all Ibovespa stocks, that account for the main stock market index in Brazil. Not much of the methods used in this paper have been seen in financial literature, hence the motivation for this study. The methodology is meaningful to traders, as some stocks are highly correlated to the main index, one can base a strategy when they are trading independently in a given day as this pattern should revert to the mean. One of the most important findings of this work is that treating data as nonlinear yielded stronger results.

Keywords: Multidimensional Association; Maximal Information Coefficient; Coefficient of Determination; Ibovespa; Intraday.

RESUMEN

Analizamos cuatro tipos de medidas de asociación, la Correlación de Pearson, el Coeficiente de Determinación, la Asociación Multidimensional y el Coeficiente Máximo de Información, los dos primeros, lineal, y los otros dos, no lineal. Utilizamos 10 minutos de datos dentro de un día, de todas las acciones Ibovespa, que cuenta para el principal índice de acciones en el mercado de Brasil. En la literatura financiera no hay mucho sobre los métodos utilizados en este trabajo, de dónde viene la motivación para el estudio. La metodología es significativa para negociantes, así como algunas acciones son altamente correlacionadas con el principal índice, un puede ser la base estratégica cuando están operando independientemente en un dado día, como este estándar podría reverter la significación. Uno de los más importantes resultados de este trabajo es que el tratamiento de los dados como no lineal produjeron resultados más fuertes.

Palabras-clave: Asociación Multidimensional; Coeficiente Máximo de Información; Coeficiente de Determinación; Ibovespa; Dados dentro de un día.

1 INTRODUCTION

Measuring association between assets is a key in modern finance, academics seek explanations and relationships between stochastic processes and market professionals are in the lookout for optimal asset placement strategies and arbitrage opportunities. As almost everything in finance the fundamentals for association measures were laid by Markowitz in 1952 in his Portfolio Selection theory. From there many other forms of association measurements were created. This work focuses on four different correlation indicators, two linear: Pearson's Correlation and the Coefficient of Determination; and two nonlinear: MIC (Maximal Information Coefficient) and MA (Measuring Association). As the subject of this study is the stock market, we used intraday data in an interval of 10 minutes of the main stock market index in Brazil, the Ibovespa and all the stocks which form this benchmark, in a total of 71 companies that make up for the index. The idea is to provide valuable information for *daytraders* when designing their strategy. We conducted this research subcategorized by economic sectors, aiming to facilitate the scan of possible trades and to compare stocks in the same group.

Many economics and financials theories are based by associating different variables and studying the correlation between many assets. Longin and Solnik (2001); Embrechts, McNeil and Straumann (2002) and Highman (2002), Engle and Granger (1987) with their work *Co-integration and error correction: representation, estimation, and testing* were awarded with a Nobel Prize in 2003. Christopher A. Sims (2012) has also been awarded with a Nobel Prize in 2011, analyzing association, cause and effect between variables in his work *Statistical modeling of monetary policy and its effects*. With far less ambition, this work tries to lead to a better understanding of the intraday relationships between the main index in Brazil and all the stocks that comprise that index.

There are many ways to correlate variables and their applications are also various and all types of studies, being medical, financial, engineering and so on. In any field of study, in this case, finance and economics, it is important to understand that not always relationships mean cause and effect, the so-called spurious correlation or either nonsense correlation mentioned respectively by Granger and Newbold (2001), and Hassler (2003). There are two more features of this work that we believe justifies it, and brings contemporaneity to its application: they are the nonlinear treatment of the variables, and the use of 10 minutes intraday data.

One of the biggest mistakes in finance is treating data as linear. Many economists blame the subprime meltdown on Gaussian functions (Salomon, 2009), even with extensive mathematical input, the presumption was wrong, financial data are not linear. If they were, a drop in 400 points in the Dow Jones Average would happen once in every four hundred years or so. But they happen quite a lot due to the nonlinearity of the data source. The other is the use of intraday data which is hard to compile, get and treat but due to the advance of computer statistics, its usage has become more clear and accessible to academics and market professionals. Moreover, most of the methods in usage assume normal distributions; financial data is far from normal, therefore the failure of many methods to prevent risk loss during recent financial crises.

One last thing this paper tries to do is to verify which sectors have a higher association with the index Ibovespa, for that we divided the companies among the following sectors: Financials, Real Estate, Basic Materials, Industrials, Oil, Mining, Holdings, Public Services, Consumption, Telecom and Transport. After this brief introduction, this paper is structured as follows: Literature Review, Methodology and data, Results and lastly, Conclusion. In the Literature Review chapter, a comprehensive review of several methods to assess association is laid. The Data chapter brings a full explanation on the intraday data used to compile the results found in the following part, Results. Lastly the paper is concluded with the findings and suggestion for future studies.

2 LITERATURE REVIEW

A good and effective risk management strategy has to come from the study of correlation between assets as exposed by many works, such as Laloux, Cizeay, Bouchad, and Potters (1999), Engle (2002) Embrechts, Lindskog, and McNeil (2003). In simple words correlation is the degree to which one variable moves in line with another, and the associations can be multiple, direct or even can make no sense at first sight, as the relationship between rain in Brazil and Starbucks stocks. Navarro (2002), as an illustration is that if it is raining in Brazil, coffee prices go down because of production surplus of therefore stock prices will go up. And the associations take many facets and many dynamics; this work tries to understand a simple and easily noted relationship, index prices and stock prices.

Many works and several authors have treated and dealt with the association between assets in finance and economics, as a matter of fact, association is one of the core studies of these fields, and here some of them are set as examples. Dumas, Harvey and Ruiz (2003) study stock returns correlations and changes in national outputs. Mayasami, Lee and Hamzah (2005) study the co-integration between macroeconomic variables and stock market index in Singapore. Walti (2005) studies the relationship between stock market returns for fifteen industrialized countries over a period of twenty-six years and macroeconomic fundamentals.

Kawaller, Koch and Koch (2012) empirically examine the intraday price relationship between S&P 500 futures and the S&P 500 index using minute-to-minute data. They chose a three-stage least-squares regression to estimate lead and lag relationships with estimates for expiration days of the S&P 500 futures compared with estimates for days prior to expiration. The results suggest that futures price movements consistently lead index movements by twenty to forty-five minutes while movements in the index rarely affect futures beyond one minute.

Forbes and Rigobon (2002) study stock market co-movements in several before-and-after crashes in world markets. Eichengreen, Rose and Wyplosz (1996) utilize the collapse of fixed exchange rates to compute the probability of one country´s crises affect the probability of other countries facing a speculative attack. King, Santana and Wadhwani (1990) analyze the volatility relationship between sixteen international markets.

Kwon, Choi, and Moon (2005) propose a system to predict stock movements based on financial correlation between companies. Barndorf-Nielsen, Ole and Shephard (2004) analyze multivariate high frequency financial data using realized covariation, regression and correlation analysis. Chiang, Jeon, and Li (2007) apply a dynamic conditional correlation model to nine Asian stock market returns from 1990 to 2003. Lundin, Dacorogna and Müller (1998), as this work does, study correlation of high-frequency (intraday) financial time series.

3 METHODOLOGY AND DATA

Four types of association measurement are used in this paper: the usual Pearson´s correlation, the R^2, MIC and MA, being the first two linear, and the other two nonlinear. The most common and simple form is the Pearson's correlation, or *Pearson's correlation coefficient*, commonly called simply *the correlation coefficient* (CROXTON; COWDEN; BOLCH*,* 1963). It is obtained by dividing the covariance of the two variables by the product of their standard deviations. The population correlation coefficient $\rho X,Y$ between two random variables X and Y with expected values μX and μY and standard deviations σX and σY is defined as: $\rho_{X,Y} = cov(X,Y)/\sigma_X\sigma_Y = E[(X-\mu X)(Y-\mu Y)]/\sigma_X\sigma_Y$.

The usage of the traditional form of correlation has already been discarded for advanced economic studies (FORBES; RIGOBON, 2002), because the nonlinearity association between the variables. In recent times many other forms of assessing association have been applied such as dynamic correlation, functions of copulas, quantile regressions and etc. This work will bring the ancient and regular type of correlation, as to make a comparison with two new approaches, the MIC (maximum information coefficient) and the MA (multidimensional association).

The coefficient of determination, denoted R^2 or r^2 and pronounced R squared, indicates how well data points fit a line or curve (DRAPER; SMITH, 1998). It is a statistic measurement as a form of a coefficient from 0 to 1, used in the context of statistical models whose main purpose is either the prediction of future outcomes or the testing of hypotheses, on the basis of other related information. It provides a measure of how well observed outcomes are replicated by the model, as the proportion of total variation of outcomes explained by the model. One use of the coefficient of determination is to test the goodness of fit of the model. It is expressed as a value between zero and one. A value of one indicates a perfect fit, and therefore, a very reliable model for future forecasts. A value of zero, on the other hand, would indicate that the model fails to accurately model the dataset. The coefficient of determination for a linear regression model is the quotient of the variances of the fitted values and observed values of the dependent variable. If we denote y_i as the observed values of the dependent variable, y as its mean, and y_i as the fitted value, then the coefficient of determination is: $R^2 = {}_{i=1}^{N}(y_i - y)^2 / {}_{i=1}^{N}(y_i - y)^2$. Where *N* is the number of observations used to fit the model.

Reshef *et al.* (2011) firstly introduced the concept of Maximal Information Coefficient (MIC). MIC in its essence is able to capture a wide range of associations both functional and not, and for functional relationships provides a score that roughly equals the coefficient of determination (R^2) of the data relative to the regression function. MIC belongs to a larger class of maximal information-based nonparametric exploration (MINE) statistics for identifying and classifying relationships. In that initial work, MIC was applied to datasets in global health, gene expression, major-league baseball, and the human gut microbiota. The basic idea is that if there is a relationship between two random variables, then a grid can be drawn on the scatterplot to encapsulate that relationship. So the next thing is to determine the maximal grid resolution, computing for every pair (*x,y*) and the largest possible mutual information achievable by any *x*-by-*y* grid applied to the date. After that, there is a need to normalize the information regarded to *x* and *y*, to ensure that that different dimensions are compared under the same assumption.

Under a different perspective, for a grid *G*, let I_G denote the mutual information of the probability distribution induced on the boxes of *G*, where the probability of a box is proportional to the number of data points falling inside the box. The (*x*,*y*)-th entry $m_{x,y}$ of the characteristic matrix equals max{I_G}/log min{*x*,*y*}, where the maximum is taken over all *x*-by-*y* grids *G*. MIC is the maximum of $m_{x,y}$ over ordered pairs (*x,y*) such that $xy < B$, where *B* is a function of sample size; we usually set $B = n0.6$.

Every entry of *M* falls between zero and one, and so MIC does as well. MIC is also symmetric (i.e. MIC(*X*, *Y*) = MIC(*Y*, *X*)) due to the symmetry of mutual information, and because I_G depends only on the rank order of the data, MIC is invariant under order-preserving transformations of the axes. Importantly, although mutual information is used to quantify the performance of each grid, MIC is not an estimate of mutual information.

The authors of the MIC methodology state they have tested it over several simulations and confirm the mathematical result that noiseless relationships such an r^2 equals 1.0 receive scores of 1.0 as per the MIC what makes easy to interpret and compare results from both methods. As more noise is added to the function, and more associations not perfectly modeled by a function, the MIC performs in an intuitive manner.

There have been some criticism over this methodology, Murrel, Murrel and Murrel (2013) and Noah and Tibshirani (2011), both present reasoning over the efficiency of the MIC but more related to large datasets and large scale exploratory analysis as it tends to produce many false positives. As this paper is not based on a large scale dataset, it is believed in the efficiency of the method, moreover it produced a very similar result to the next method, as it will be seen on the Results chapter.

The Multidimensional Association (MA) method utilized in this paper is based on the works of Murrel, Murrel and Murrel (2013); and the prior motivation of picking this method is due to its ability to treat nonlinear data. When two variables are related by a known function, the coefficient of determination measures the proportion of the total variance in the observations that is explained by that function. That could be translated as the strength of the relationship, by describing what proportion of the variance is signal as opposed to noise. For linear relationships, this is equal to the square of the correlation coefficient, the R^2, however is unlikely to find linear association among financial data. The objective of this method is to how to directly estimate a generalized R^2 when the form of the association between the variables is unclear.

As the name suggests, leave-one-out is a type of cross-validation (LOOCV) and it involves using a single observation from the original sample as the validation data, and the remaining observations as the training data. This is repeated such that each observation in the sample is used once as the validation data. This is the same as a K-fold cross-validation with K being equal to the number of observations in the original sampling (KEARNS; RON, 1999).

So to demonstrate the MA method, consider two vector valuable variables such as (*X*,*Y*), with *n* observations $(x_1,\ldots,x_n)$ and $(y_1,\ldots,y_n)$, each of the *x* may or may not be a vector $x^{\alpha}_{i},\ldots,$ x^{z}_{i}, the same goes to the other vector, y_i. Then, consider kernel distributions, K_X(x), K_Y(y) and K_{XY}(xy), symmetric, non-negative and integrate to 1, and where angle brackets indicate vector concatenation. The null model assumes that *X* and *Y* are independent and it is defined the leave-one-out cross validation likelihood as the product of the marginal kernel density estimates:

$$L_{CV}\ null\ =\ \mathop{}_{i=1}^{n}\ P\ \ x_i\ x_{\forall j\neq 1}\ \ P\ \ y_i\ y_{\forall j\neq 1}\ \ \approx\ \mathop{}_{i=1}^{n}\ \mathop{}_{\forall j\neq 1}\ \frac{K_X(x_j-x_i)}{n-1}\times\frac{K_Y(y_j-y_i)}{n-1}$$

The alternative model allows *Y* to depend on *X* for a proportion of points, *w*, with a leave-one-out cross validation likelihood defined as:

$$L_{CV}\ alt\ =\ \mathop{}_{i=1}^{n}\ \ w\times P\ \ x_i,y_i\mid x_{\forall j\neq 1},y_{\forall j\neq 1}\ \ +\ \ 1-w\ \ P\ \ x_i\mid x_{\forall j\neq 1}\ \ P\ \ y_i\mid y_{\forall j\neq 1}$$

$$\approx\ \mathop{}_{i=1}^{n}\ \mathop{}_{\forall j\neq 1}\ \frac{K_{XY}(<x_j-x_i,y_j-y_i)}{n-1}+(1-w)\ \mathop{}_{\forall j\neq 1}\ \frac{K_X(x_j-x_i)}{n-1}\ \mathop{}_{\forall j\neq 1}\ \frac{K_Y(y_j-y_i)}{n-1}$$

In this particular case, the values of (*X*, *Y*) are replaced with their ranks and the kernels are isotropic Gaussians, with K_x and K_y sharing and independent kernel variance parameter σ_I^2, and K_{xy} having a dependent variance parameter, σ_D^2. The null model $L_{CV}\ null$ has a single parameter, σ_I^2, and the alternative model $L_{CV}\ alt$ has three parameters: σ_I^2, σ_D^2 and w.

Murrel, Murrel, and Murrel (2013) ., state that this method has advantages over the previous, the MIC. According to the authors, the MA approach behaves equitably, has more power than MIC and converges faster with increasing sample size. Moreover, it states that the MA generalizes to higher dimensions, allowing to estimate the strength of multivariate relationships (not the case of this paper), and to measure relationships while controlling covariates as raised by Speed (2011). Anticipating the results chapter, association between the variables of this paper was found stronger when utilizing the MA, over the MIC and the R^2.

Intraday data of returns in a 10 minutes interval is the base for the dataset. The trading days chosen span from January to March 2013, in a total of 1770 observations per stock. The source of this data was the Thompson Reuters Eikon Software® which is extremely reliable and one of the most trustworthy trading software in the market today. As a remainder, intraday data from every stock that makes up the index is extracted and its relationship to the index is the core of this work, so they are analyzed pairing up, the index and each and every stock. The companies are aggregated by sectors as following: Financials, Real Estate, Basic Materials, Industrials, Oil, Mining, Holdings, Public Services, Consumption, Telecom and Transport.

As common in finance, returns were calculated in a logarithmic base, to make the time series stationary and what was compared to extract association were the index returns in a logarithmic base and the stock returns also in a logarithmic base in the same time. Whenever intraday data is in usage, some major drawbacks come to life, for an example when compiling 10 minutes interval of less liquid stocks, there are times that that particular stock have not been traded in a few of the 10 minutes along the entire trading day, but the index was. So what is seen is a long record of index data but not so long of the less liquid stocks. To correct and unsettle that, all blank spaces were filled with the last traded value. In this way, the index and the stocks have equal number of observations and more importantly, the stock is set at its real price in time, at the appropriate index price, considering the same 10 minutes interval.

In this work the R^2 is only displayed to be and accessory of the linear correlation, the idea is to show it as base in the interpretation.

4 THE IBOVESPA INDEX

Ibovespa is the main indicator of the Brazilian stock market's average performance. Ibovespa reflects the variation of the main stock market in Brazil and its most traded stocks. There have been no methodological changes to the index since its inception in 1968, when it has been attributed a base value of 100 points as of a hypothetical investment. The participation of each stock in the portfolio has a straight relation with its representativity in the cash market, measured in terms of number of trades and financial value, adjusted to the sample size. From time to time less traded stocks give place to others that obtained greater numbers in a set time frame. In this work, the stocks that pertain to the index in the beginning of the sample size will be evaluated, if there were changes in the period, these changes will not affect the results of this work.

One of the main goals is to study if there is any more conspicuous pattern in co-movements amongst all the economic sectors. These sectors and the interdependence between them and the index will be explained in depth in the next chapter.

5 RESULTS

Results of this work are summarized in Table 1 and in Figure 1, which bring all sectors and the average of the coefficients of all methods used and tests to verify the difference between measurements. Friedman test rejects the hypothesis that the association coefficients (R2, MIC, MA) are equal and the Wilcoxon signed ranks test rejects equality for paired measurement (R2, MIC) and (R2, MA).

In a nutshell, Mining and Basic Materials have a greater association with the index. The case of Mining is easily explained because the two assets that contribute to that higher association are both pertaining to Cia. Vale do Rio Doce, VALE3 and VALE5. These two stocks comprehend roughly 11% of the index. This behavior should be similar in the Oil sector as PETR3 and PETR4 sum up to be around 11% of the index, plus the fact that there are not many other companies in this sector, but it does not. The MA produced more significant results than the linear correlation and this fact is detailed in the remaining of the paper.

Table 1 - Association measurement of all sectors with the Ibovespa Index

Sector	**Number of Companies**	**Correlation**	**R²**	**MIC**	**MA**
Basic Materials	5	0.537	0.289	0.237	0.302
Consumption	5	0.287	0.090	0.154	0.126
Financials	8	0.394	0.171	0.189	0.203
Holdings	3	0.433	0.226	0.218	0.251
Industrials	13	0.320	0.109	0.156	0.133
Mining	3	0.626	0.394	0.286	0.387
Oil	5	0.420	0.222	0.214	0.235
Public Services	12	0.298	0.091	0.152	0.127
Real Estate	8	0.441	0.203	0.193	0.220
Telecom	4	0.307	0.095	0.150	0.136
Transport	5	0.351	0.130	0.150	0.127
Test for difference	Statistics	p-value			
Friedman test					

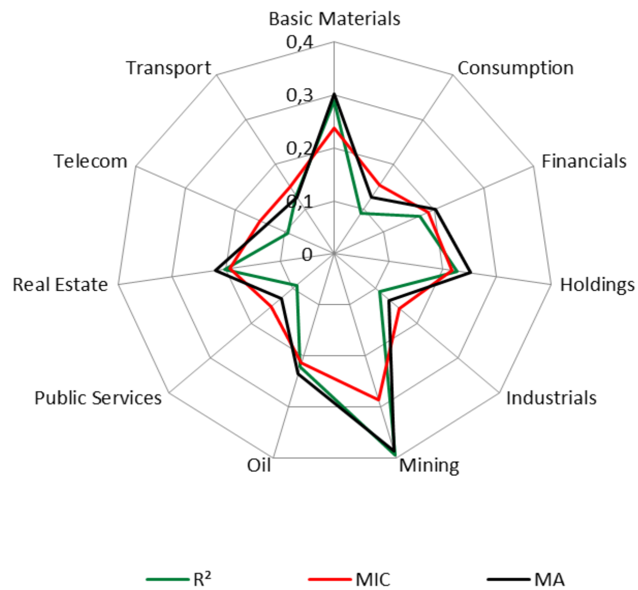
(R2, MIC, MA)	χ^2 = 22.290	0.000
Wilcoxon signed ranks test		
(R2, MIC)	z = -2.263	0.024
(R2, MA)	z = -4.954	0.000
(MIC, MA)	z = -0.027	0.979

Source : Authors (2015).

The Basic Materials sector comprehends the companies that produce steel, iron and aluminum and the major players in this sector are: Gerdau, Usiminas and Companhia Siderúrgica Nacional. This sector achieved the second highest degree of association in many of the measurements chosen such as the second highest MA (0.302) and the second highest R^2 (0.289). *Consumption* includes the stocks in this sector are supermarkets chains such as (PCAR4), retail stores (LREN3 and LAME4) and internet based sales (BTOW3). The degree of relationship that these stocks achieved with the main index was below average. With the exception of PCAR4, they all drifted around the same levels of integration.

The Financial sector brings Banks such as BBAS3 (Banco do Brasil), BBDC3 and BBDC4 (Bradesco), ITUB4 (Itaú-Unibanco); credit cards providers (CIEL3) and the company that controls the Bovespa and the mercantile trading operations, (BVMF3). There is no pattern of association among these stocks; some high degrees are easily noted in some stocks such as ITUB4 and BBDC4, and some weak forms in SANB11 (Banco Santander Brasil). *Holdings* sector is made up by 3 companies that belong to other major economic players in Brazil, for example, ITSA4 is a holding that invests in several sectors such as technology, gas, real estate and others, and it is controlled by the ITUB4 (ITAÚ-Unibanco). The same goes to BRAP4 (Bradespar), controlled by BBDC4 (Bradesco). Also an erratic behavior is seen among the stocks in this sectors, some very high degrees in some cases (BRAP4) and some very weak (UGPA3).

Figure 1 - Visualization of association measurement of all sectors with the Ibovespa Index



Industrials sector, as the name explains, this sector is formed by large industrial complexes such as AMBV4 (Ambev), one of the world´s largest beer producer; EMBR3 (Embraer) a competitive player in jet industry; some paper and cellulose producers SUZB5 (Suzano) and KLBN4 (Klabin); some of the world´s largest cattle slaughters JBBS3 (JBS) and MRFG4 (Marfrig). The Mining sector includes the biggest company in Brazil, Cia. Vale do Rio Doce (VALE3 and VALE5), as mentioned before, this company accounts for over 11% of the index formation. This fact shall be taken in consideration when analyzing that this company has the highest degree of MA association 0.430 and 0.459 for VALE3 and VALE5 respectively. The other company in this sector is MMXM3 (MMX) experience also a high degree of association.

In the Oil sector is the second largest company in Brazil, PETR3 and PETR4 (Petrobras). Other oil company is OGXP3 (OGX Oil), and other two produce biodiesel VAGR3 (VA Agropecuária) and CSAN3 (Cosan) which is the country´s biggest producer of Ethanol, a substitute for oil and gasoline. Among the oil producers, a high degree of association was found; the other two companies (not oil related) present weak linkage to the main index. Public Services sector are subdivided into three: Electric Power CESP6, CPFL6, CMIG4, CPFE3, ELET3, ELET6, ELPL4, ENBR3, LIGT3 and TRPL4; Health care providers DASA3 and lastly, Sanitation SBSP3. A week degree in association is seen all across the board in this sector.

As per Real Estate, the companies that belong to this group are somewhat new to the index, although they represent a large share of the theoretical portfolio. This type of company has not been listed in the stock market for very long as the real estate market, was less developed and organized as most of the financial markets throughout the globe. This paper was initially bound to make a comparison of the intraday data and daily returns of the major stocks in the index, unfortunately due to the recent entrance of all these stocks, it had to be adjusted only to intraday. Nevertheless, the companies in this group are large constructors, incorporators such as RSDI3 (Rossi), CYRE3 (Cyrela) and shopping mall administrators BRML3 (BR Malls). Along this group it is perceived moderate to high association, but all together, just above average.

The Telecom sector comprehends the following companies: OIBR3 and OIBR4 (Oi), TIMP3 (Tim) and VIVT4 (Vivo); this sector presents a very small degree of linkage to the main variable. Another group is Transports, which is a diverse block and does not come out as having a strong relationship with the index behavior. ALL3 (America Latina Logística) is a railroad company; CCRO3 (CCR Rodovias) runs toll sites all over the country, GOLL4 (Gol) is an airline, LLXL3 (LLX Logística) is a transportation and logistics company and RENT3 (Localiza) is a car rental.

6 CONCLUSION

After thorough analysis of all Ibovespa portfolio one thing can be assured: relationships are stronger when nonlinearity comes to the equation. The MIC and the MA, both nonlinear measures of association have performed stronger than the Coefficient of Determination (R^2). The MA, in average, is the one which found stronger relationships between the variables.

Mining and Basic Materials sectors had a stronger R^2, MIC and MA, and this fierce relationship can lead to a day trade or swing trade investment strategy, arbitraging the stock or the index, with long or short positions. The MA captured a strongest degree of association, the runner up was the MIC and in last the coefficient of determination. These results suggest that even though the nonlinear measures were more expressive, most of the dependence captured is linear.

The original idea was to compare high frequency data (intraday) of returns of the index stocks with the index itself, and then do the same analysis with daily closing prices over time. As mentioned before, many stocks have debuted recently in the stock market and many of them play a special role in the index nowadays, but did not play before. Hence as a suggestion for future studies is that, given the index some time to ripe up, and then do the same analysis with daily closing prices, but this only time will allow, as the moment this paper is being written, there are just not enough observations.

REFERENCES

BARNDORFF-NIELSEN, O. E.; SHEPHARD, N. Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. **Econometrica** 72, 3, 2004, 885-925.

CHIANG, T. C.; JEON, B. N.; LI, H. Dynamic correlation analysis of financial contagion: Evidence from Asian markets. **Journal of International Money and Finance***,* 26, 7, 2007, 1206-1228.

CROXTON, F. E.; COWDEN, D. J.; BOLCH, B. W. **Practical business statistics,** Prentice-Hall New York, 1969.

DRAPER, N. R.; SMITH, H. **Applied regression analysis.** Wiley series in probability and statistics, 1998.

DUMAS, B.; HARVEY, C. R.; RUIZ, P. Are correlations of stock returns justified by subsequent changes in national outputs? *Journal of International Money and Finance* 22, 6, 2003, 777-811.

EICHENGREEN, B.; ROSE, A. K.; WYPLOSZ, C. **Contagious currency crises**. National Bureau of Economic Research, 1996.

EMBRECHTS, P.; LINDSKOG, F.; MCNEIL, A. Modelling dependence with copulas and applications to risk management. **Handbook of heavy tailed distributions in finance,** v. 8, n. 1, 2003, 329-384.

EMBRECHTS, P.; MCNEIL, A.; STRAUMANN, D. Correlation and dependence in risk management: properties and pitfalls. **Risk management:** *value at risk and beyond,* 2002, 176-223.

ENGLE, R. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. **Journal of Business & Economic Statistics**,v. 20, n.3, 2002, 339-350.

ENGLE, R. F.; GRANGER, C. W. Co-integration and error correction: representation, estimation, and testing. **Econometrica:** *journal of the Econometric Society,* 1987, 251-276.

FORBES, K. J.; RIGOBON, R. No contagion, only interdependence: measuring stock market comovements. **The Journal of Finance** 57, 5, 2002, 2223-2261.

GRANGER, C.; NEWBOLD, P. Spurious regressions in econometrics. **Essays in Econometrics***:* Collected Papers of Clive WJ Granger*,* 2, 2001, 109.

HASSLER, U. Nonsense regressions due to neglected time-varying means. **Statistical Papers** 44, 2, 2003, 169-182.

HIGHAM, N. J. Computing the nearest correlation matrix—a problem from finance. **IMA journal of Numerical Analysis** 22, 3, 2002, 329-343.

KAWALLER; KOCH; KOCH. The Temporal Price Relationship between S&P 500 Futures and the S&P 500 Index.

KEARNS, M.; RON, D. Algorithmic stability and sanity-check bounds for leave-one-out cross-validation. **Neural Computation** 11, 6, 1999, 1427-1453.

KING, M.; SENTANA, E.; WADHWANI, S. **Volatility and Links Between National Stock Markets**. National Bureau of Economic Research, 1990.

KWON, Y.-K.; CHOI, S.-S.; MOON, B.-R. **Stock prediction based on financial correlation.** ACM, 2005, 2061-2066.

LALOUX, L.; CIZEAU, P.; BOUCHAUD, J.-P.; POTTERS, M. Noise dressing of financial correlation matrices. **Physical Review Letters** 83, 7, 1999, 1467.

LONGIN, F.; SOLNIK, B. Extreme correlation of international equity markets. **The Journal of Finance.** 56, 2, 2001, 649-676.

LUNDIN, M. C.; DACOROGNA, M. M.; MÜLLER, U. A. Correlation of high frequency financial time series. *Available at SSRN 79848,* 1998.

MARKOWITZ, H.; Portfolio Selection*. **The Journal of Finance.** 7, 1, 1952, 77-91.

MAYSAMI, R. C.; LEE, C. H.; HAMZAH, M. A. Relationship between macroeconomic variables and stock market indices: Cointegration evidence from stock exchange of Singapore's All-S sector indices. **Jurnal Pengurusan***, v.*24, 2005, 47-77.

MURRELL, B., MURRELL, D. and MURRELL, H. Discovering general multidimensional associations. **arXiv preprint arXiv***:1303.1828,* 2013.

NAVARRO, P. *If It's* **Raining in Brazil, Buy Starbucks**: The Investor's Guide to Profiting from Market-moving Events, McGraw-Hill, 2002.

RESHEF, D. N.; RESHEF, Y. A.; FINUCANE, H. K.; GROSSMAN, S. R.; MCVEAN, G.; TURNBAUGH, P. J.; LANDER, E. S.; MITZENMACHER, M.; SABETI, P. C. Detecting novel associations in large data sets. **Science.** 334, 6062. 2011, 1518-1524.

SIMS, C. A. Statistical modeling of monetary policy and its effects. **American Economic Review V.** 102, n. 4, p.1187, 2012.

SPEED, T. A Correlation for the 21st Century. *Science* 334, 6062, 2011, 1502-1503.

WALTI, S. The macroeconomic determinants of stock market synchronization. **Journal of International Banking Law** . V. 11, N. 10, 2005, pp. 436-441.