

Testing the Existence of Lead-Lag Effects Between the US and Brazilian Stock Markets

Gustavo Rezende de Oliveira[†]
University of Brasilia

Otavio Ribeiro de Medeiros^Ω
University of Brasilia

ABSTRACT: This article examines the existence of lead-lag effects between the U.S. stock market (NYSE) and the Brazilian stock market (Bovespa), i.e., whether upward and downward price movements in the NYSE are followed, on average, by similar movements in Bovespa, which would enable predicting stock prices in the Brazilian market, thus providing arbitrage opportunities. The existence of this effect would indicate a relative segmentation between these two markets, which would violate the efficient market hypothesis, whereby stock prices are unpredictable. Cointegration between the two markets was identified as well as the existence of bi-directional causality (Granger test). The results obtained from VECM, TSLS and GARCH regressions showed that the two markets are segmented and that returns of the Bovespa Index (Ibovespa) are to a large extent explained by the stock price movements in the Dow Jones Index some minutes beforehand. However, the results also show that the practice of arbitrage based on the lead-lag effects is not economically feasible due to transaction costs.

Keywords: lead-lag effects, efficient market hypothesis, arbitrage.

Received in 06/30/2008; revised in 12/10/2008; accept in 03/18/2009..

Corresponding authors:

[†] *Professor of University of Brasília*

*Address: Av. Parque Águas Claras, 3825, Apto 808,
Brasília – DF – Brazil - CEP: 71930-000*

e-mail: gustavorezende@hotmail.com

Telephone: (61) 9999-6502

^Ω *Professor of University of Brasília*

*Address: SQN 205, Bloco C, Apto. 401, Asa Norte,
Brasília – DF – Brazil - CEP: 70843-030*

e-mail: otavio@unb.br

Telephone: (61) 9978-9503

Editor's note: *This paper was accepted by Antonio Lopo Martinez.*

1. INTRODUCTION

The advances in communications and information technology, dating from the middle of the twentieth century, have contributed decisively toward the integration of global stock markets. This integration means that the expected return on assets of a single risk class will be the same in different markets, according to Eiteman et al. (1994), cited in Costa Júnior and Leal (1997). Greater integration among international financial markets tends to make them move together. Therefore, a single relevant event can make the general prices of shares – general stock market indexes – rise or fall at the same time in markets throughout the world.

According to the efficient market hypothesis (EMH), stock prices are not predictable, because they behave like a random walk, hence not permitting arbitrage. The EMH focuses on the market as a single entity throughout the world (FAMA, 1970). If markets that are not completely integrated – i.e., segmented markets – are analyzed together, as if they were a single market, the EMH can fail for various reasons. However, there are markets that are more developed than others, causing them to have different characteristics, among them size differences. The New York Stock Exchange (NYSE), for example, is at least twenty times larger than the São Paulo Stock Exchange (Bovespa). The degree of concentration of various stock markets also varies. The Bovespa, for example, is very concentrated, because over 60% of the volume traded is concentrated in the shares of only 19 companies (ASSAF NETO, 2007). The existence of better mechanisms to protect minority investors in some markets in detriment to others also indicates that some markets are more developed. Besides this, the more advanced communications media in more developed countries provide information more quickly and efficiently to investors, so that those in less developed countries operate at a disadvantage. All these characteristics, which cause different degrees of development among markets, permit information asymmetry to occur.

If the EMH does not hold, the asset prices in a developed market can incorporate the effect of information more efficiently. In other words, asset price movements in a more efficient market – the United States for instance – can precede the price movements in less developed markets, such as Brazil and other emerging countries, if there is partial integration of the developed and emerging markets.

This is the lead-lag effect, whereby the asset price movements in one market lead and those in other markets lag behind. Such a situation runs counter to the EMH and makes it possible to predict, with a certain level of confidence, the price movements in the lagging markets in function of those in the leading market. This situation, if sufficiently pronounced, would provide opportunities for abnormal gains in the lagging markets through arbitrage.

In this work we investigate the existence of lead-lag effects between the American and Brazilian stock markets by determining the lags between price movements in the NYSE and Bovespa. We also examine the possibility of profiting from arbitrage as a result of the lead-lag effect, considering the transaction costs.

To do this, we use the Dow Jones Industrial Average (DJIA) as a proxy for the performance of the NYSE and the counterpart Bovespa Index (Ibovespa) for the Brazilian market. We also analyze the effects of variations in the DJIA on specific shares in the Brazilian market, specifically the twelve stocks with greatest weight in the Ibovespa, in light of their higher trading volume. These stocks represent over 50% of the index's weight.

The article is organized into five sections including this introduction. Section 2 reviews the empirical literature on the lead-lag effect; Section 3 presents the methodologies

employed; Section 4 reports the application of these methods to the practical case; and Section 5 concludes.

2. EMPIRICAL EVIDENCE

Various works have examined the existence of a lead-lag effect between correlated markets. For example, Brooks, Rew and Ritson (2001) analyzed the existence of relations between future contracts and spot prices of the FTSE 100, the index of the 100 most highly capitalized firms listed on the London Stock Exchange, using observations at ten-minute frequencies. They confirmed what various similar works had identified: the returns in the future market lead those in the spot market, confirming the hypothesis that information is first absorbed by the future market. The reason for this according to the authors is that

the futures price should quickly reflect all available information regarding events that may affect the underlying and respond quickly to new information. The index should respond in a similar fashion, but for the index to react to the new information completely the underlying stocks must all be revalued, i.e. every constituent stock must re-evaluate the new information and adjust accordingly. Because most stocks are not traded constantly every 10 min, the index will respond to new information with a lag. (BROOKS, REW AND RITSON, 2001, p. 34).

According to the authors, this lead-lag effect is not enough to allow profitable arbitrage because the transaction costs more than offset the abnormal returns.

Li, Greco and Chavis (2001) observed the existence of previous movements in the price of shares traded in Hong Kong, the so-called H shares, which are followed by stock price movements of the same firms traded in China, the so-called A shares. The authors attributed the lead-lag effect in these markets to the different level of information disclosure in the two markets. They noted that the lead-lag effect increased according to the volatility of the return on the asset.

Poshakwale and Theobald (2002) analyzed the presence of the lead-lag effect in the Indian stock market, which is segmented in terms of market capitalization. The authors observed that the share prices of firms with high capitalization led those of firms with low capitalization because the share prices of larger firms adjust faster to new information than those of smaller firms.

In a study of the New York Stock Market covering the period from August 1984 to September 1987, Chan (1992) concluded that the intra-day future index strongly influenced the spot market, as predicated by the theory, while there was weak evidence of the contrary effect. This influence occurred both in response to good and bad news.

To analyze the real flow of information, Jong and Nijman (1997) observed whether there was a lead-lag relationship of the S&P500 and the futures written on it, based on minute-by-minute information. They also detected that the information is absorbed by the future market minutes before the spot market, with an average lag of ten minutes. In the few situations where the spot index led the future one, the lag was at most two minutes.

In this same line of research, Kawaller, Koch and Koch (1987) observed the presence of lead-lag between the future and current S&P500. Using data gathered minute-by-minute, they concluded that the influence of the former on the latter takes between 20 and 45 minutes. In the few cases where the opposite was true, the lag was at most one minute.

Kofman and Martens (1997) examined the correlation of the prices of the future indexes of the London and New York stock markets. The results indicated that the lead-lag

effect of the American market was preponderant over the British market, although in some cases the opposite occurred. They also examined the possibility of gains by running simulations with real prices and considering transaction costs. They concluded that there is a possibility for gain by acquiring future contracts in the London exchange a few minutes after a price rise in the New York exchange.

3. METHODOLOGY

The section presents the data and methods used to check for the existence of lead-lag effects between the NYSE and Bovespa. In this effort, we used various econometric models: regression analysis; cointegration analysis, the VAR and VEC models, the Granger causality test and the GARCH model.

3.1. The Data

The data were obtained from the CMA system, with frequency of one minute, at closing prices. All the tests involving the DJIA and Ibovespa series were performed with data from July 19, 2006 to September 28, 2007. Each of these series contains 105,443 observations. We only considered price quotations in periods when the two exchanges were operating simultaneously. Since the data are made available according to local time, we had to adjust for the time difference, including considering the different periods of daylight savings time in the two countries.

The analyses involving the companies making up the Ibovespa cover the period from March 14 to September 28, 2007. We used the 12 companies with greatest weight in the Ibovespa: Petrobras PN (PETR4), Vale do Rio Doce PNA (VALE5), Bradesco PN (BBDC4), Usiminas PNA (USIM5), Banco Itaú PN (ITAU4), Telemar PN (TNLP4), Vale do Rio Doce ON (VALE3), Gerdau PN (GGBR4), Petrobras ON (PETR3), Siderúrgica Nacional (CSNA3), Itausa PN (ITSA4) and Unibanco UNT (UBBR11).

3.2. Regression Analysis

We started from the hypothesis that there are lead-lag effects between the NYSE and Bovespa, and that these effects can be captured by the regression:

$$P_t^{ib} = \beta_0 + \sum_{i=1}^n \beta_i P_{t-i}^{ib} + \sum_{j=0}^n \gamma_j P_{t-j}^{dj} + u_t \quad (1)$$

where P_t^{ib} is the prices of the shares making up the Ibovespa; P_t^{dj} is the prices of shares making up the DJIA; γ_j and β_i are parameters to be estimated in the equation and u_t are random residuals $\sim N(0, \sigma^2)$.

Since stock prices are non-stationary variables¹, it was necessary to use the stock prices given by:

$$R_t = \log \frac{P_t}{P_{t-1}} = \Delta \log P_t \quad (2)$$

¹ Regressions that use non-stationary variables are considered spurious regressions, where apparently significant relationships can appear that are in truth nonexistent. See Brooks (2002, p.367).

Therefore, equation (1) was transformed into:

$$R^{ib}_t = \beta_0 + \sum_{i=1}^n \beta_i R^{ib}_{t-i} + \sum_{j=0}^n \gamma_j R^{dj}_{t-j} + u_t \quad (3)$$

where R^{ib} is the return of the Ibovespa; R^{dj} is the return of the DJIA; γ_j and β_i are parameters to be estimated in the equation and u_t is the random error $\sim N(0, \sigma^2)$.

Equation (3), however, can cause two econometric problems. In the first place, if R^{dj} is also an endogenous variable, then there is a problem of simultaneity and equation (3) cannot be estimated by ordinary least squares (OLS), because this would result in biased and inconsistent coefficients. The second problem is that if P^{ib} and P^{dj} are cointegrated series, then equation (3) must include an error correction term.

The solution to the first problem is to estimate equation (3) by a method that takes this simultaneity into consideration, such as two-stage least squares (TSLS). To do this it is necessary to define the instrumental variables to be used. In the case here, these were the variables R^{ib} and R^{dj} themselves, lagged.

The solution to the second problem is to check for cointegration, and if it is found, to introduce an error correction term in equation (3). However, although the parameters estimated by OLS are inconsistent when the regressors are cointegrated, this does not happen when the estimation is obtained by TSLS: for construction of empirical models, in a structural approach, it is necessary to pay attention to the question of identification and simultaneity bias, but it is not necessary to worry about questions of non-stationarity and cointegration. In practice, even though the series are cointegrated, it is not necessary to include an error correction term when using TSLS (HSIAO, 1997, p.395).

3.3. Analysis of Cointegration

According to Harris (1995), cointegration indicates the existence of a long-term equilibrium to which the economic system converges. To test for the existence of cointegration in this work, we used the single-equation technique of Engle and Granger (1987), because there are only two stochastic variables and at most one cointegration vector.

The cointegration regression by the Engle-Granger method is given by:

$$\log P^{ib}_t = \beta_0 + \log \beta_1 P^{ib}_t + u_t \quad (4)$$

where P^{ib} is the Ibovespa; P^{dj} is the DJIA index; β_0 and β_1 are parameters to be estimated in the equation and u_t are the random errors $\sim N(0, \sigma^2)$.

To have cointegration, it is necessary for the two series to have the same order of non-stationarity and for the residuals of this regression to be stationary. We used the augmented Dickey-Fuller (ADF) test to check for the existence of non-stationarity.

According to Groppo and Bacchi (2003), the most widely used statistics to test for the existence of stationarity are intended for series that have at most one unit root. In the model developed by Dickey and Fuller, known as the DF model:

$$X_t = \rho X_{t-1} + u_t \quad (5)$$

where u_t is an error $\sim N(0, \sigma^2)$. Then $H_0: \rho = 1$ is tested against $H_1: \rho < 1$. If $|\rho| < 1$, then X_t is stationary and described by an AR(1) process. In contrast, if $\rho = 1$, then X_t is non-stationary and described by a random walk.

We used the ordinary least squares method to test for a unit root. Unit root tests do not employ Student's standard t -distribution, but rather statistics whose critical values were tabulated by Dickey and Fuller. For the variables to be cointegrated, the residuals must be $I(0)$,

that is, stationary. Hence, we estimated the regression between the natural logarithms of the variables of the two series of interest to test the stationarity of the residuals.

If cointegration is found between the two series, the error correction model (ECM) must be used, which here can be expressed as:

$$R_t^{ib} = \beta_0 + \beta_1 R_t^{dj} + \beta_2 (R_{t-1}^{ib} - \beta_3 R_{t-1}^{dj}) + u_t \quad (6)$$

where R^{ib} is the return of the Ibovespa; R^{dj} is the return of the DJIA; β_0 is the constant of the equation; β_1 represents the short-term relationship between the two variables; β_2 represents the speed of adjustment to equilibrium and u_t are the random residuals $\sim N(0, \sigma^2)$.

As seen previously, however, in cases where the regression is estimated by TSLS, it is not necessary to include an error correction term, even when the series are cointegrated.

3.4. The Vector Autoregression (VAR) Model

According to the VAR model, developed by Sims (1980), a variable's value is expressed as a linear function of the past, or lagged, value of this variable and of all the other variables included in the model. If each equation contains the same number of lagged variables in the system, it can be estimated by OLS without needing to resort to any method employing systems of equations, such as TSLS. The model estimated here is:

$$R_t^{ib} = \beta_0 + \sum_{i=1}^n \beta_i R_{t-i}^{ib} + \sum_{j=1}^n \gamma_j R_{t-j}^{dj} + u_t \quad (7)$$

$$R_t^{dj} = \alpha_0 + \sum_{i=1}^n \alpha_i R_{t-i}^{dj} + \sum_{j=1}^n \delta_j R_{t-j}^{ib} + v_t \quad (8)$$

where R^{ib} is the return of the Ibovespa; R^{dj} is the return of the DJIA; β_i , γ_j , α_i , and δ_j , are parameters to be estimated in the equation, and u_t and v_t are the random errors $\sim N(0, \sigma^2)$.

3.4.1. The Vector Error Correction (VEC) Model

The estimation of the vector error correction (VEC) model was popularized by Engle and Granger (1987) and aims to analyze the short-term adjustments that occur in cointegrated series. If two markets function perfectly, the price movements in these markets should be better described by the first-order VEC model, with the error correction term being the difference of the returns between the markets. The VEC model here is given by:

$$R_{t_{ib}} - R_{t_{ib}^1} = \beta_1 [R_{t_{dj}} - R_{t_{ib}^1}] + u_t \quad (9)$$

$$R_{t_{dj}} - R_{t_{dj}^1} = \beta_2 [R_{t_{ib}} - R_{t_{dj}^1}] + v_t \quad (10)$$

where R^{ib} is the return of the; R^{dj} is the return of the DJIA; β_1 and β_2 are parameters to be estimated in the equation, and u_t and v_t are the random errors $\sim N(0, \sigma^2)$.

3.5. Granger Causality

To check the direction of causality between the series, we performed a causality test according to Granger (1969). The concept of causality between two variables in the Granger sense refers to the time element, that is, if the variable R^{dj} causes the variable R^{ib} , then the lagged values of the variable R^{dj} help to predict the values of R^{ib} . We used the following equations of the VAR model to test the direction of causality between the two indexes:

$$R_t^{ib} = \beta_0 + \sum_{i=1}^n \beta_i R_{t-i}^{ib} + \sum_{j=1}^n \gamma_j R_{t-j}^{dj} + u_t \quad (11)$$

$$R_t^{dj} = \alpha_0 + \sum_{i=1}^n \alpha_i R_{t-i}^{dj} + \sum_{j=1}^n \delta_j R_{t-j}^{ib} + v_t \quad (12)$$

where R^{ib} is the return of the Ibovespa; R^{dj} is the return of the DJIA; β_i , γ_j , α_i , δ_j , are parameters to be estimated in the equation, and u_t and v_t are random errors $\sim N(0, \sigma^2)$.

The hypotheses to be tested are that the coefficients of the past values of the exogenous variable of the above equations are equal to zero, that is:

$$\gamma_1 = \gamma_2 = \dots = \gamma_p = 0 \quad (13)$$

$$\beta_1 = \beta_2 = \dots = \beta_p = 0 \quad (14)$$

We used the F-test to determine the direction of causality. If the two hypotheses are rejected, there is bi-directional causality, while if neither of the two hypotheses is rejected, there is no causality.

3.6. The GARCH Model

If the residuals of a regression show variance that is not constant, but rather variable with time, they are said to present autoregressive conditional heteroskedasticity (ARCH) effects, and it is necessary to specify an ARCH model to estimate the referred regression by the maximum likelihood method. Here we used a generalized autoregressive conditional heteroskedasticity (GARCH) model, expressed by:

$$R_t^{ib} = \alpha + \beta R_t^{dj} + u_t \quad u_t \sim N(0, \sigma_t^2) \quad (15)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i u_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \quad (16)$$

where R^{ib} is the return of the Ibovespa, R^{dj} is the return of the DJIA and σ_t^2 is the variance of the errors u_t . This model assumes that the variance of the error terms is not constant, making it more suitable to analyze financial time series whose error variances are not constant over time.

4. RESULTS

As expected, cointegration was identified between the two markets. Initially we checked for the existence of unit roots for the natural logarithm of the two series, as shown in Tables 1 and 2:

Table 1: Unit Root Test – Dow Jones, in Level

Null Hypothesis: LOG(DJ) has a unit root		
Lag Length: 1 (Automatic based on SIC, MAXLAG=68)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.559289	0.5035
Test critical values: 1% level	-3.430243	
5% level	-2.861377	
10% level	-2.566723	

The null hypothesis – that the natural logarithm of the DJIA, in level, has a unit root – was not rejected.

Table 2: Unit Root Test – Ibovespa, in Level

Null Hypothesis: LOG(IBOV) has a unit root		
Lag Length: 4 (Automatic based on SIC, MAXLAG=68)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-0.567703	0.8753
Test critical values: 1% level	-3.430243	
5% level	-2.861377	
10% level	-2.566723	

Likewise, the null hypothesis that the Ibovespa has a unit root in level was not rejected.

In the model in first difference, however, the null hypothesis of having a unit root was rejected in both cases, as shown in Tables 3 and 4 below. This is the first condition for there to be cointegration: that the two series have the same level of integration. We identified first-order integration (I(1)) in the natural logarithms of the two series.

Table 3: Unit Root Test – Dow Jones, first difference

Null Hypothesis: D(LOG(DJ)) has a unit root		
Lag Length: 0 (Automatic based on SIC, MAXLAG=68)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-316.7573	0.0001
Test critical values: 1% level	-3.430243	
5% level	-2.861377	
10% level	-2.566723	

Table 4: Unit Root Test – Ibovespa, first difference

Null Hypothesis: D(LOG(IBOV)) has a unit root		
Lag Length: 3 (Automatic based on SIC, MAXLAG=68)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-149.6146	0.0001
Test critical values: 1% level	-3.430243	
5% level	-2.861377	
10% level	-2.566723	

Therefore, in first difference of the natural logarithms, both the DJIA and Ibovespa series are stationary.

Next we isolated the regression residuals of the two series to check for the existence of a unit root, since the Engle-Granger method forecasts that if the regression residuals of two

non-stationary series of the same order do not have a unit root, that is, are stationary, the two series will be cointegrated. Table 5 shows the results of the unit root test of the regression residuals.

Table 5: Unit Root Test – Residuals, in level

Null Hypothesis: U has a unit root		
Lag Length: 8 (Automatic based on SIC, MAXLAG=68)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.043122	0.0311
Test critical values: 1% level	-3.430243	
5% level	-2.861377	
10% level	-2.566723	

Since the null hypothesis was rejected at 5% significance, we could conclude that the DJIA and Ibovespa series are cointegrated. The existence of cointegration implies the use of an error correction term to estimate the regression by OLS. However, as seen before, the inclusion of an error correction term is unnecessary when estimating the regression by means of TSLS.

To ascertain the level of dependence of the current value of the Ibovespa to the previous values of the two variables (R^{ib} and R^{dj}), it was necessary to use a VAR model. As explained before, when the relationship between two non-stationary variables is important, their transformation into stationary series by their difference is not enough for the result of their regression to be valid.

By definition, the variables converge to the same value and then do not change over the long run. Thus, the difference between their variables becomes zero (BROOKS, 2002). In this form, to eliminate this error we used the vector error correction (VEC) model.

The results revealed various important aspects of the relationship between the series, including the exchange rate.

The first of these is that the significant lags of the DJIA variable go to nine minutes before the current moment. The coefficients for the lags longer than this were zero. This result is very important since it demonstrates that past movements in the NYSE have an influence on the current value of the shares making up the Ibovespa.

Except for the period $t-6$, the lags of the Ibovespa itself that exercise an influence on this index's current value go back to minute $t-8$, but all these coefficients had low absolute values.

Another relevant conclusion is that the exchange rate variable was insignificant in nearly all the lags. Only for the lags $t-13$, $t-14$, $t-15$ and $t-18$ was the null hypothesis that the coefficients are equal to zero rejected, at 5%. Although significant, all these coefficients had extremely low values. For the $t-13$, $t-14$, $t-15$ and $t-18$ lags, the values of the coefficients were 0.0008, 0.0007, 0.0009 and 0.0007, respectively.

Because it turned out to have little influence on the variation of the Ibovespa, we eliminated the exchange rate variable from the model in the subsequent tests. For this same reason we once again estimated the VEC model without this variable.

The lags of the DJIA variable with significant coefficients did not vary much from those of the previous model with the exchange rate. The difference was that the interval of influence increased to ten minutes. For minutes $t-1$ to $t-10$, the coefficients – all significant – were 0.367, 0.225, 0.111, 0.046, 0.0364, 0.035, 0.026, 0.021, 0.013 and 0.021, respectively. It

is interesting to note that all these coefficients are positive. These numbers indicate to what proportion each movement at a previous instant of the DJIA is reflected in the level of the Ibovespa. Therefore, according to this model, if for example one minute beforehand there was a fall of 0.02% in the American exchange, this event will cause the Brazilian index to fall 0.074% ($0.02\% \times 0.36950$ – coefficient of lag $t-1$) in the present minute. Of course the value of the shares making up the Ibovespa is subject to other influences, including those of only local scope, but there is unquestionably a very strong relationship between the two markets.

These coefficients indicate that the information is not all reflected at the same moment in the price of all the stocks making up the Ibovespa basket, and would not be reflected in any other asset market. Relevant information that prompts a variation in the price of a determined asset or set of assets at the worldwide level is not available to all investors immediately. This lapse of ten minutes can be interpreted as the average time it takes for a certain piece of news to be incorporated in the value of a certain asset.

It is interesting to observe that, as expected, the coefficients of the lagged variables of the NYSE index decline in absolute value each minute that precedes the current value of the Ibovespa index. The three most recent minutes are crucial to assimilate information in the value of an asset, while the next seven minutes exercise slightly less influence, until after ten minutes the full effect of the information is reflected in the asset's price.

Only four of the twenty lagged coefficients of the Ibovespa variable tested were significant. They were minutes $t-1$, $t-2$, $t-3$ and $t-7$, with respective coefficients of 0.016, -0.009, -0.103 and -0.117. With the exception of the minute immediately preceding the current one, all the other significant coefficients were negative, denoting an inverse relationship between past and present minutes of the Ibovespa variable. Note that all four of these coefficients had low absolute values, the greatest being 11.74%.

To confirm the strong relationship between the two variables, we used the Granger causality model, with 20 lagged values. The results are presented in Table 6. The test was carried out with the difference of the logarithm of each element of the sample.

Table 6: Granger Causality Test

Pairwise Granger Causality Tests			
Date: 03/11/08 Time: 17:14			
Sample: 1 105443			
Lags: 20			
Null Hypothesis:	Obs	F-Statistic	Probability
DLOG(DJ) does not Granger Cause DLOG(IBOV)	105422	281.870	0.00000
DLOG(IBOV) does not Granger Cause DLOG(DJ)		17,3089	3.5E-61

The significance of the relations between the series is given by the F statistic, which tests the hypothesis that the coefficients of the equations' variables are zero.

Both null hypotheses – that the DJIA does not cause the Ibovespa and that the Ibovespa does not cause the DJIA – were rejected. This result, indicating bi-directional causality, confirms the strong relation between the two variables.

To confirm the conclusions obtained by means of the methods already presented, we performed two other regressions, this time using the two-stage least squares (TSLS) method. The results are presented below.

Subitem 2.2 above presented two problems in estimating equation (3) by OLS. One of these was that if R^{dj} is influenced by R^{ib} , meaning the variable R^{dj} is endogenous, the regression should be estimated by TSLS.

The tables below show that in the regressions where R^{dj} is the dependent variable, the series R^{ib} was significant. With the use of a dummy variable for positive returns, both the contemporary variable R^{ib} and its values lagged by up to four minutes were significant.

Table 7: OLS Regression of the variables R^{ib} (dependent) and R^{dj} with a dummy for positive returns

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.71E-06	7.08E-07	10.88269	0.000000
DLOG(DJ(-1))	0.004417	0.002063	2.140630	0.032300
DLOG(DJ(-2))	-0.010404	0.002078	-5.006047	0.000000
DLOG(DJ(-3))	-0.008022	0.002081	-3.855710	0.000100
DLOG(DJ(-4))	-0.009918	0.002077	-4.776043	0.000000
DLOG(IBOV)	0.011604	0.001086	10.68425	0.000000
DLOG(IBOV(-1))	0.014103	0.001126	12.52486	0.000000
DLOG(IBOV(-2))	0.007130	0.001125	6.337884	0.000000
DLOG(IBOV(-3))	0.003526	0.001117	3.157902	0.001600
DLOG(IBOV(-4))	0.004800	0.001093	4.392987	0.000000
R-squared	0.003958	Mean dependent var	7.86E-06	
Adjusted R-squared	0.003873	S.D. dependent var	0.00023	
S.E. of regression	0.00023	Akaike info criterion	-13.91801	
Sum squared resid	0.005571	Schwarz criterion	-13.91710	
Log likelihood	733753.7	F-statistic	46.54395	
Durbin-Watson stat	2.012031	Prob(F-statistic)	0.00000	

With the use of a dummy for negative returns, only the R^{ib} variable lagged by four minutes was not statistically significant at 5%, as shown in the table below.

Table 8: OLS Regression of the variables R^{ib} (dependent) and R^{dj} with a dummy for positive returns

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.21E-05	6.61E-07	-48.59469	0.00000
DLOG(DJ(-1))	0.008849	0.001927	4.592801	0.00000
DLOG(DJ(-2))	-0.004084	0.001941	-2.104309	0.03540
DLOG(DJ(-3))	-0.002276	0.001943	-1.171414	0.24140
DLOG(DJ(-4))	-0.002620	0.001939	-1.351224	0.17660
DLOG(IBOV)	0.011994	0.001014	11.82640	0.00000
DLOG(IBOV(-1))	0.013078	0.001051	12.43855	0.00000

(continued on next page)

Table 8: OLS Regression of the variables R^{ib} (dependent) and R^{dj} with a dummy for positive returns (continued)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLOG(IBOV(-2))	0.004216	0.001051	4.013318	0.00010
DLOG(IBOV(-3))	0.002336	0.001043	2.239933	0.02510
DLOG(IBOV(-4))	-0.000209	0.001020	-0.204878	0.83770
R-squared	0.004335	Mean dependent var		-3.20E-05
Adjusted R-squared	0.00425	S.D. dependent var		0.000215
S.E. of regression	0.000215	Akaike info criterion		-14.05503
Sum squared resid	0.004858	Schwarz criterion		-14.05412
Log likelihood	740969.9	F-statistic		50.99688
Durbin-Watson stat	1938121	Prob(F-statistic)		0.00000

The results of these regressions indicated that the DJIA series is endogenous. For this reason, we performed new regressions, this time using the TSLS method.

Of the fifteen variables of the regression using a dummy for positive returns, only five were not statistically significant at 5%, as shown in the table below:

Table 9: TSLS regression of the variables R^{ib} (dependent) and R^{dj} with a dummy for positive returns

Variable	Coefficient	Std. Error	t-Statistic	Prob.	Non-signific. at 5%
C	0.000149	8.64E-07	172.8133	0.0000	
DLOG(IBOV(-1))	-0.000484	0.001377	-0.351568	0.7252	*
DLOG(IBOV(-2))	-0.007845	0.001377	-5.698065	0.0000	
DLOG(IBOV(-3))	-0.007536	0.001377	-5.472939	0.0000	
DLOG(IBOV(-4))	-0.002102	0.001377	-1.526518	0.1269	*
DLOG(DJ)	0.037001	0.002383	15.5286	0.0000	
DLOG(DJ(-1))	0.193812	0.002472	78.41562	0.0000	
DLOG(DJ(-2))	0.114543	0.002521	45.43227	0.0000	
DLOG(DJ(-3))	0.050621	0.00254	19.92839	0.0000	
DLOG(DJ(-4))	0.024446	0.002543	9.614649	0.0000	
DLOG(DJ(-5))	0.013615	0.002457	5.542427	0.0000	
DLOG(DJ(-6))	0.009642	0.002406	4.008208	0.0001	
DLOG(DJ(-7))	0.009991	0.002386	4.186484	0.0000	
DLOG(DJ(-8))	0.002969	0.002382	1.246399	0.2126	*
DLOG(DJ(-9))	-0.002996	0.002381	-1.258213	0.2083	*
DLOG(DJ(-10))	-0.003171	0.00238	-1.332615	0.1827	*

Only the variables of the first and last lags of the Ibovespa, along with the variables of the last three lags of the DJIA, were not statistically significant. All the significant coefficients of the lagged variables of the DJIA were positive. This positive correlation fell progressively until minute $t-7$, with results similar to those obtained by the VEC model.

The main difference between this result and that obtained by VEC, for positive returns, is that the coefficient of the variable at the current minute in the TSLS model is smaller than that of the variable lagged by one minute.

The sum of all the coefficients is 0.454, indicating that the positive return of the current minute of the NYSE and the seven preceding minutes explain 45.36% of the return in minute $t-0$ of the Bovespa.

The lagged coefficients of the Ibovespa itself have little explanatory power on its present value. As seen, only the coefficients of minutes $t-2$ and $t-3$ were significant, with values of -0.008 and -0.007 respectively, for an unimpressive total of -1.53%. As will be seen shortly, the lagged coefficients of the Ibovespa for negative returns presented positive signs, meaning that for positive returns the relation of the lags of the Ibovespa is inverse and for negative returns it is direct.

The asteristics in the table below indicate the coefficients that were not significant at 5%:

Table 10: TSLS regression of the variables R^{ib} (dependent) and R^{dj} with a dummy for negative returns

Variable	Coefficient	Std. Error	t-Statistic	Prob.	Non-signific. at 5%
C	-2.64E-05	9.78E-07	-26.96551	0.0000	
DLOG(IBOV(-1))	0.005501	0.001559	3.529149	0.0004	
DLOG(IBOV(-2))	-0.001928	0.001559	-1.237119	0.216	*
DLOG(IBOV(-3))	-0.00173	0.001559	-1.109718	0.2671	*
DLOG(IBOV(-4))	0.004351	0.001559	2.790832	0.0053	
DLOG(DJ)	0.042383	0.002698	15.71064	0.0000	
DLOG(DJ(-1))	0.172321	0.002798	61.58221	0.0000	
DLOG(DJ(-2))	0.107115	0.002854	37.5268	0.0000	
DLOG(DJ(-3))	0.060432	0.002876	21.0136	0.0000	
DLOG(DJ(-4))	0.019266	0.002879	6.693029	0.0000	
DLOG(DJ(-5))	0.017923	0.002781	6.444348	0.0000	
DLOG(DJ(-6))	0.005979	0.002723	2.195263	0.0281	
DLOG(DJ(-7))	0.013378	0.002702	4.951372	0.0000	
DLOG(DJ(-8))	0.012126	0.002697	4.496686	0.0000	
DLOG(DJ(-9))	0.004597	0.002696	1.705191	0.0882	*
DLOG(DJ(-10))	0.006779	0.002694	2.516146	0.0119	

The coefficients with lags of two and three minutes of the Ibovespa, as well as of nine minutes of the DJIA, were the only cases where the null hypothesis of coefficients equal to zero was not rejected. All the other coefficients showed positive results.

The sum of all the coefficients is 0.458, very similar to the sum of the coefficients of the NYSE for positive returns. This value suggests that a negative return at the current minute of the NYSE and the ten previous minutes explains 45.77% of the return at minute $t-0$ of the Bovespa.

Some premises have to hold for the TSLS to be valid. We conducted tests of autocorrelation, normality and heteroskedasticity. The graphs of the residuals of the TSLS regressions are presented below.

Autocorrelation occurs in analyses of time series when the errors associated with observations in one period of time hold up by transfer in future periods. The Durbin-Watson test, of the most popular ones, cannot be used for regressions in which lags of the dependent variable are regressors. For this reason we used the Breusch-Godfrey test. As can be seen below, the regressions with a dummy for both negative and positive returns showed autocorrelation of the residuals, because the null hypothesis of no autocorrelation was rejected in both cases.

Table 11: Breusch-Godfrey serial correlation test of the residuals for the TSLS regression with a dummy for positive returns

Breusch-Godfrey Serial Correlation LM Test:			
Obs*R-squared	4,058.581	Probability	0.000000

Table 12: Breusch-Godfrey serial correlation test of the residuals for the TSLS regression with a dummy for negative returns

Breusch-Godfrey Serial Correlation LM Test:			
Obs*R-squared	1,312.707	Probability	0.000000

The presence of autocorrelated residuals indicates that the estimated coefficients are inefficient, i.e., they do not have the least possible variance. However, they are unbiased (BROOKS, 2002, p. 166).

One of the properties of least squares estimators is that the errors must follow a normal distribution. Therefore, it is reasonable to assume this premise based on the meaning of the error term: a sum of factors that were not included in the model. If these factors are many, their sum must be normally distributed, according to the central limit theorem.

To check for the normality of the residuals, we performed the Jarque-Bera test, which uses the results of the moments of the normal distribution, in particular the coefficient of skewness – which is zero for the normal distribution – and the coefficient of kurtosis – which is equal to 3 in the normal distribution (SARTORIS, 2007). In both cases, i.e., for the regressions with a dummy variable for positive and negative returns, the null hypothesis that the residuals are normally distributed was rejected. Nevertheless, the violation of this premise of normality is virtually inconsequential for large samples (BROOKS, 2002, p. 182).

Another property of least squares estimators is that the variance of the errors must be constant, which is known as homoskedasticity, which means absence of heteroskedasticity. The consequence of violating this hypothesis is that even though OLS estimators are not skewed, they do not have the minimum variance.

The null hypothesis that the errors are homoskedastic was rejected for the regressions with both a dummy variable for positive and for negative returns.

Table 13: White heteroskedasticity test of the residuals for the TSLS regression with a dummy for positive returns

White Heteroskedasticity Test:			
F-statistic	107.6836	Probability	0.000000
Obs*R-squared	3,135.362	Probability	0.000000

Table 14 : White heteroskedasticity test of the residuals for the TSLS regression with a dummy for negative returns

White Heteroskedasticity Test:			
F-statistic	127.4389	Probability	0.000000
Obs*R-squared	3,690.430	Probability	0.000000

Because of the heteroskedasticity of the regressions, we carried out other regressions by the GARCH methodology, which incorporates the inconstancy of the variance in the model itself.

The answer to the research question will only be possible if the number of correctly predicted positive returns, based on the returns of the DJIA, is greater than the number of errors. To test for this, we randomly chose an interval of thirty days (March 14 to April 12, 2007) with real data for the twelve stocks with greatest weight in the Ibovespa. For each company we used 7,598 observations.

We only used five lags in this test. We weighted the current values of the DJIA and of the four previous minutes by the respective coefficients to identify an upward return for the next minute. The prediction for two minutes ahead was based on the current quotation of the American index and in the three preceding minutes, weighted by their coefficients, and so on successively until the prediction for five minutes in the future, when the only piece of information available was the current DJIA.

The results confirmed the hypothesis that it is possible to predict, with a reasonable degree of certainty, positive returns of the Bovespa based only on the movements of the NYSE.

As can be seen in the table below, for the interval from one to four minutes, the maximum percentage of correct predictions of positive returns reached over 60% in all cases. For the fifth minute, the highest correct percentage was 59.40% and the lowest was 54.53%. We did not compute the returns equal to zero, which was the case when there was no variation in the two quotations or when one figure was not available for a determined minute.

Table 15: Forecast returns for the twelve stocks of the Ibovespa, by time interval

Shares	1 minute			2 minutes			3 minutes			4 minutes			5 minutes		
	correct	incorrect	% corr.	correct	incorrect	% corr.	correct	incorrect	% corr.	correct	incorrect	% corr.	correct	incorrect	% corr.
Petrobras PN (PETR4)	1,755	1,238	58.64%	1,854	1,288	59.01%	1,846	1,412	56.66%	1,870	1,432	56.63%	1,870	1,430	56.67%
Vale do Rio Doce PNA (VALE5)	1,488	936	61.39%	1,607	933	63.27%	1,613	991	61.94%	1,597	1,042	60.52%	1,570	1,073	59.40%
Bradesco PN (BBDC4)	1,181	779	60.26%	1,301	838	60.82%	1,387	898	60.70%	1,392	956	59.28%	1,391	998	58.23%
Usiminas PNA (USIM5)	885	616	58.96%	960	653	59.52%	1,005	652	60.65%	982	745	56.86%	994	755	56.83%
Banco Itaú PN (ITAU4)	921	623	59.65%	986	678	59.25%	1,003	705	58.72%	1,039	742	58.34%	1,021	763	57.23%
Telemar PN (TNLP4)	717	498	59.01%	771	505	60.42%	824	553	59.84%	822	582	58.55%	818	577	58.64%
Vale do Rio Doce ON (VALE3)	828	538	60.61%	886	552	61.61%	855	595	58.97%	867	566	60.50%	851	585	59.26%
Gerdau PN (GGBR4)	1,018	677	60.06%	1,118	713	61.06%	1,191	755	61.20%	1,178	852	58.03%	1,161	854	57.62%
Petrobras ON (PETR3)	729	446	62.04%	741	515	59.00%	738	542	57.66%	741	581	56.05%	753	548	57.88%
Cia. Siderúrgica Nacional (CSNA3)	929	583	61.44%	997	598	62.51%	1,027	657	60.99%	1,021	706	59.12%	1,022	721	58.63%
Itausa PN (ITSA4)	554	419	56.94%	634	450	58.49%	735	506	59.23%	736	567	56.49%	792	558	58.67%
Unibanco UNT (UBBR11)	657	527	55.49%	806	543	59.75%	810	625	56.45%	801	667	54.56%	813	678	54.53%
Maximum correct			62.04%			63.27%			61.94%			60.52%			59.40%
Minimum correct			55.49%			58.49%			56.45%			54.56%			54.53%
Average correct			59.54%			60.39%			59.42%			57.91%			57.80%

It remains, however, to test the possibility of using arbitrage to gain from this lead-lag effect. This will only be possible if the transaction cost is less than the gain provided. For each purchase or sale transaction the Bovespa charges 0.019% to cover the cost of settlement plus a fee of 0.006%. Since arbitrage involves buying and then selling the same asset, the total transaction cost is 0.05% of the value of the trade. There is also the brokerage fee (an average of R\$ 10.00 per transaction). At large trading volumes this cost becomes less significant, so we disregarded it in the analysis.

Based on the positive returns predicted by the above test (period from March 14 to April 12, 2007), we obtained the average of the positive returns. None was greater than 0.05%. The possibility of engaging in arbitrage is thus made unprofitable just by the transaction costs, which outweigh the expected profits from predicting positive returns based on the variation of the NYSE.

Since the TSLS regressions show heteroskedasticity, we carried out other regressions by the GARCH method, a model that does not assume constant variance.

We also used the GARCH model to check for the influence of the lagged variables on the current value of the Ibovespa. In this phase, the analysis was divided into two parts: determination of the coefficients for the upward and for the downward variations. For this purpose we created dummy variables to annul the effects of the upward and downward intervals, alternately.

Another important observation refers to the high volatilities found between the stock prices at market closing and opening the next day. To exclude this effect, we created another dummy variable to offset the result of these abnormal returns.

Since the result of the VEC model demonstrated that only the first ten lags of the DJIA variable were significant, we used the GARCH regression with this number of lags. For the Ibovespa variable, we use four lags.

For upward returns, all the coefficients of the four lags of the Ibovespa variable were statistically significant, with the following values from $t-1$ to $t-4$, respectively: 0.038, -0.038, 0.006 and 0.039. For downward returns, only lags $t-1$ and $t-4$ were significant, with respective coefficients of 0.011 and 0.003. For positive returns, the influence of the DJIA was statistically significant for all ten lags.

For the negative returns, only one of the coefficients was not statistically significant – that lagged by nine minutes. An important observation is that the coefficient of minute $t-1$ was higher than that of minute $t-0$, which urges a certain degree of investor caution about negative movements. Unlike what occurred with the dummy for positive returns, the first five lags of the regression with the dummy for downward returns turned out positive.

5. CONCLUSIONS

The objective of this study was to test for the existence of lead-lag effects between the NYSE and Bovespa. The existence of these effects was verified by use of various tests.

First, the strong relation between the two stock markets was shown by the identification of cointegration (Engle-Granger test) between the first-order non-stationary series that compose these markets.

The Granger causality test detected bi-directional causality between the markets, confirming the relationship between the series.

The integration identified, however, is not total since the lagged movements of both series have an influence on minute $t-0$ of the Ibovespa. If the integration is not total, there is segmentation between the markets, which opens a margin for arbitrage.

To investigate to what intensity the lagged periods of the two series explain the current Ibovespa, we used some models in natural logarithm of the return of these series: a vector autoregression (VEC) model with error correction, with sixty lags; a GARCH model; and a simultaneous equations model estimated by TSLS. For the last two we used ten lags for the DJIA series and four for the Ibovespa series. In all the tests we observed an influence of the lags of the NYSE on the current minute of the Ibovespa.

Based on the results obtained in all the tests, it was possible to affirm there is significant evidence it is possible to make some prediction about the Ibovespa from the information about the preceding behavior of the DJIA.

These results reveal that the return of the Bovespa is to a large extent explained by the movement of the NYSE in previous minutes (and also, to a lesser scale, by its own lagged movements).

Although obvious, it should be recalled that the value of the assets making up the Ibovespa is not only subject to the movements of the NYSE. There are purely local events that only affect the Brazilian market. Besides this, the reaction of investors to a determined news item, even one with global scope, is attenuated by local circumstances. For example, a rise in the price of oil in the international market may not be reflected in the same proportion in the local market because of its different level of external dependence or the existence of alternative fuels, etc.

Even though the reaction to events may not be exactly the same, the fact is that the movements of the two markets analyzed are extremely similar, with a certain precedence of the American market.

Therefore, it can be inferred that one of the assumptions of the efficient market hypothesis (EMH) is questionable: information is not being incorporated in all asset prices instantaneously.

If it is possible to partly predict the movement of one stock market from the preceding performance of another market, it may also be possible to profit from arbitrage. If arbitrage is possible, the market is definitely not efficient.

Our intention here was to test the efficient market hypothesis by examining the interrelationship of two markets, to add evidence to help with future studies of the theme.

With respect to arbitrage, the transaction costs, consisting of the settlement costs and fees charged by the Bovespa and brokerage fees, more than offset the profits that could be gained by arbitrage. The returns of assets for periods less than ten minutes are on average less than the transaction costs, ruling out arbitrage gains.

REFERENCES

ASSAF NETO, A.; LIMA F. G.; DE ARAÚJO, A. M. P. **A realidade da concentração do capital no Brasil: Um estudo comparativo com duas outras economias emergentes.**

Available at <http://www.fipecafi.com.br/palestras/assaf/pdf/Artigo_Concentracao_Capital_Brasil_Dividendos.pdf>. Accessed on February 20, 2008.

BOVESPA– <www.bovespa.com.br>. Accessed on March 20, 2008.

- BROOKS, C. **Introductory Econometrics for Finance**. 1st Edition, Cambridge, Cambridge University Press. 2002.
- BROOKS, C. REW, A. G.; RITSON, S. A Trading Strategy Based on the Lead-lag Relationship Between the Spot Index and Futures Contract for the FTSE 100. **International Journal of Forecasting**, V. 17, 2001.
- CAMPBELL, J. Y., LO, A.W.; MACKINLAY, A. C. **The Econometrics of Financial Markets**. Princeton University Press. New Jersey. 1997.
- CARNEIRO, FRANCISCO GALRÃO. **A Metodologia dos Testes de Causalidade em Economia**. Available at < <http://www.unb.br/face/eco/textos/causal.pdf>> Accessed on January 10, 2008.
- CHAN, K. A Further Analysis of the Lead--Lag Relationship Between the Cash Market and Stock Index Futures Market. **The Review of Financial Studies**, Vol. 5, No. 1 (1992), pp. 123-152
- COSTA JUNIOR, N.C.A.; LEAL, R.P.C. Mercosul e a Globalização dos Mercados de Capitais: Testes de Causalidade. IN: **Revista de Administração**, vol. 32, 1997.
- CUNHA, M. S. **Raiz Unitária e Co-integração: Três Aplicações**. Available at <<http://www.unb.br/face/eco/textos/cointegracao.pdf>> Accessed on September 30, 2007.
- ENGLE, R. F.; GRANGER, C. W. E. Cointegration and error correction: Representation, estimation and testing. **Econometrica**, V.55. 1987.
- FAMA, E. F. Efficient capital markets: A review of empirical work, **Journal of Finance**, 417. 1970.
- GRANGER, C. W. J. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. **Econometrica**, V. 37, 1969.
- GROPPO, G. S.; BACCHI, M. R. P. Causalidade entre mercados acionários: estudo sobre Ibovespa, Dow Jones e Merval. In: **Resenha BM&F** n. 166. 2005.
- GUJARATI, D. N. **Econometria Básica**. São Paulo. Makion Books. 2000.
- HARRIS, R. I. D. **Using Cointegration Analysis in Econometric Modelling**. London: Prentice Hall, 1995.
- HSIAO, C. Statistical Properties of the Two-Stage Least Squares Estimator under Cointegration. **Review of Economic Studies**, Vol. 64, n. 3. 1997
- JONG, F.; NIJMAN, T. High frequency analysis of lead-lag relationships between financial markets. **Journal of Empirical Finance**, Volume 4, Issues 2-3, June 1997, pages 259-277.
- KAWALLER, I.G.; KOCH, P.D.; KOCH, T.W. The Temporal Price Relationship Between S&P 500 Futures and the S&P 500 Index. **The Journal of Finance**, Vol. 42, No. 5 (Dec., 1987), pp. 1309-1329.
- KOFMAN, P.; MARTENS, M. Interaction between stock markets: an analysis of the common trading hours at the London and New York stock exchange. **Journal of International Money and Finance**, Volume 16, Issue 3, June 1997, pages 387-414.
- LI, Y.; GRECO, J.F.; CHAVIS, B. **Lead-lag relations between A shares and H shares in the Chinese stock markets**. California State University of Fullerton Working Paper, 2001.

POSHAKWALE, S.; THEOBALD, M. Market capitalisation, cross-correlations, the lead/lag structure and microstructure effects in the Indian stock market. **Journal of International Financial Markets, Institutions and Money**, Volume 14, Issue 4, October 2004, Pages 385-400.

SARTORIS, A. **Estatística e Introdução à Econometria**. São Paulo: Saraiva, 2007.

SIMS, C. A. Macroeconomics and Reality. **Econometrica**, n. 48. 1980.