Chinese Business Review, April 2015, Vol. 14, No. 4, 192-200
doi: 10.17265/1537-1506/2015.04.003

Analysis of the Bovespa Futures and Spot Indexes With High Frequency Data

Edimilson Costa Lucas, Danilo Braun Santos, Bruno Nunes Medeiro, Vinicius Augusto Brunassi Silva
EAESP-FGV, Sao Paulo, Brazil
Luiz Carlos Monteiro
Universidad Nacional de Misiones, Misiones, Argentina

Data from the World Federation of Exchanges show that Brazil's Sao Paulo stock exchange is one of the largest worldwide in terms of market value. Thus, the objective of this study is to obtain univariate and bivariate forecasting models based on intraday data from the futures and spot markets of the BOVESPA index. The interest is to verify if there exist arbitrage opportunities in Brazilian financial market. To this end, three econometric forecasting models were built: ARFIMA, vector autoregressive (VAR), and vector error correction (VEC). Furthermore, it presents the results of a Granger causality test for the aforementioned series. This type of study shows that it is important to identify arbitrage opportunities in financial markets and, in particular, in the application of these models on data of this nature. In terms of the forecasts made with these models, VEC showed better results. The causality test shows that futures BOVESPA index Granger causes spot BOVESPA index. This result may indicate arbitrage opportunities in Brazil.

Keywords: econometric models, arbitration, stock exchange, vector autoregressive (VAR), vector error correction (VEC), Granger causality

Introduction

According to the World Federation of Exchanges, the Sao Paulo stock exchange (Bovespa) in Brazil is one of the largest stock exchanges globally in terms of market value. It is well known that the possibility of forecasting the value of economic assets is very important for investment activities, risk management, and identification of price lags among financial markets for arbitration purposes. Therefore, understanding the relationship between spot and futures markets could be crucial for investors to take positions in the market. If it was possible to identify a precedence of the Bovespa futures index over the Bovespa spot index, the strategies that make it possible to obtain abnormal returns in these markets could be adopted. Considering the use of high frequency data, where investors select and maintain portfolios for short periods of time, a successful strategy could result in the identification of momentary arbitrage opportunities.

Edimilson Costa Lucas, master degree in statistics, EAESP-FGV, Sao Paulo, Brazil.
Danilo Braun Santos, master degree in applied mathematics, UNIFESP/EPPEN, EAESP-FGV, Sao Paulo, Brazil.
Bruno Nunes Medeiro, master degree in business administration, EAESP-FGV, Sao Paulo, Brazil.
Vinicius Augusto Brunassi Silva, master degree in business administration, EAESP-FGV, Sao Paulo, Brazil.
Luiz Carlos Monteiro, master degree in economics, UNAM—Universidad Nacional de Misiones, Misiones, Argentina.
Correspondence concerning this article should be addressed to Edimilson Costa Lucas, Av. 9 de Julho, 2029, Bela Vista, 01313-902, São Paulo, Brazil. E-mail: costalucas@yahoo.com.

Bachelier (1964) and Cowles (1993) proposed the concept of market efficiency (the market efficiency hypothesis). A market can be considered efficient, if price changes cannot be predicted, that is, if all the information available to investors is already incorporated in the price. In such a circumstance, it is impossible to obtain abnormal returns. Fama (1970) distinguished three forms of market efficiency: weak, semi-strong, and strong. If the market has a weak form of efficiency, trading tactics based on past prices will not be useful. If it has a semi-strong form of efficiency, the publicly available information will not help in obtaining gains above the market. If the valid hypothesis was that of a market with the strong form of efficiency, the use of securities analysis would be put into question. Oliveira, Nobre, and Zárate (2013) mentioned that asset prices are unpredictable, because they follow a random walk, and thus, it is not possible to obtain abnormal returns. Rittler (2012) found that the futures market is the first to incorporate information, which it subsequently transfers to the spot market. In this paper, it does not intend to identify the market form valid in Brazil. Rather, it seeks to analyze intraday series and find relationships between the spot market and the futures markets.

This approach will make it possible to build and analyze forecasting models for the returns series and test the causality among them. This type of study shows that it is important to identify arbitrage opportunities in financial markets and, in particular, in the application of these models on data of this nature.

Theoretical Framework

Stoll and Whaley (1990) verified the causal relationship between the returns of the S&P 500 futures and spot markets. The results indicated that the S&P 500 futures index seemed to anticipate the spot index at an average time interval of five minutes. Tse (1995) studied the behavior of the Nikkei average index and its futures contracts. Using an error correction model, the researcher found that changes in futures price lags affect short-term adjustments of the underlying asset's future price. Regarding arbitrage, arbitrageurs try to obtain gains by taking advantage of the difference in prices in one or more markets. Thus, in partially segmented markets, there would be an opportunity for arbitrage, which would be an error correction mechanism for the possible variance that the market may show in relation to the intrinsic value of the assets.

According to Lien and Tse (1999), the possibility of arbitrage between the spot price and the future price is a determinant for the formation of future prices. If the relationship between prices is broken, it will be possible to earn risk-free gains. If the spot and futures markets are efficient, the price of the futures contracts of an underlying index in the period t will be a function of the price level of that index in the period t, considering the net loading cost of the stock index until expiration of the contract and the time to expiration of the futures contract. Therefore, in line with the efficient market hypothesis, returns in the spot and futures markets should have a perfect simultaneous correlation and can not be correlated over time.

Brooks, Rew, and Ritson (2001) mentioned that it is possible to observe changes in the stock indexes that can anticipate changes in future prices, since the index can be considered as a minor set of data affecting the futures market. Thus, the futures market can anticipate the spot market and market changes will generally affect both the futures market and the spot market. The researchers studied the UK financial market, looking for relationships between the FTSE100 index and its index futures using the ARIMA (autoregressive integrated moving average), vector autoregressive (VAR), and vector error correction (VEC) models. In this paper, the best forecasting model obtained was the VEC, and the authors inferred that changes in the price lag of futures could help in anticipating changes in the spot price.

Silva (2006) used Johansen's cointegration and found that the Brazilian futures market anticipates the spot market. Fonseca, Lamounier, and Bressan (2012) sought to identify lucrative trading strategies based on the discrepancies between the futures and spot markets for high frequency data. The researchers built the ARIMA, ARFIMA, VAR, and VEC prevision models and tested the liquid trading strategy and the purchasing and position maintenance strategy for the August 2006 to October 2009 period. They could identify possibilities for earning abnormal returns with trading strategies using the VAR model and taking into consideration the lead-lag effects between the Bovespa spot index and the Bovespa futures index. J. Yang, Z. Yang, and Zhou (2012) investigated the intraday relationship in the Chinese futures and spot markets, showing a bidirectional relationship in both markets.

Conrad, Rittler, and Rotfuß (2012) modeled intraday prices in the American market with a GARCH model. They concluded that the prices of American stocks increased in response to better-than-expected news about future economic performance.

Finally, Kang, Cheong, and Yoon (2013) studied the futures and spot markets in Korea empirically, using high frequency data. Their results showed a strong bidirectional causal relationship between the futures and spot markets, which suggested that the volatility of returns in the spot market could influence the futures market and vice versa.

Empirical Strategy

The method used in this study focuses on the construction of univariate and bivariate forecasting models with high frequency data (interval of 15 minutes, due to its liquidity in the Brazilian market). It was chosen to use the ARFIMA, VAR, and VEC models. The variables used are the closing prices of the Bovespa spot index and the Bovespa futures index.

The sample used was collected from Bloomberg's site, taking into consideration intraday data for the period of 05/02/2014 through 10/19/2014, which corresponds to 119 working days, using a sample with 3,332 observations. Only data generated during the opening times of the stock exchange, between 10 am and 5 pm, were used. Trading in the "after-market" period was excluded. Some researchers consider these trades to have a different pattern from the trading during regular opening hours of the stock exchange (Zhang, Russel, & Tsay, 2001). In addition to the models cited above, the causal relationship between the Bovespa futures and spot indexes was verified through the Granger causality test.

ARFIMA Model

The ARIMA model was proposed by Box and Pierce (1970) and divided into three phases: (a) model identification/selection; (b) estimation; and (c) verification. In the ARIMA approach (p, d, q), authors transformed a non-stationary series into a stationary one through d differentiations, considering autoregressive and moving average components in the series.

It says that the variable xt follows the ARIMA (p, d, q) model, if the variable $y_t = \Delta^d x_t$ (x_t differentiated d times) follows an ARMA (p, q) model:

$$y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \sigma_\epsilon \epsilon_t + \theta_1 \sigma_\epsilon \epsilon_{t-1} + \ldots + \theta_q \sigma_\epsilon \epsilon_{t-q}$$

with $\epsilon_t \sim NI\,(0, 1)$.

The ARFIMA model is used, when the series shows significant autocorrelation for long intervals. That is, the decay of the autocorrelation function is hyperbolic.

If the fractionally differentiated series $(1-L)^d x_t$ follows an ARMA (*p, q*), then x_t is called an ARIMA (*p*, *d*, *q*) process.

VAR and VEC Models

The VAR model is an econometric model used to capture the evolution of the interdependence relations of multiple temporal series. VAR models generally define restrictions among the equations of the model. A fundamental question of the VAR model is whether, from the reduced form, it is possible to recover information in the structural form. Thus, a VAR model can be understood as a generalization of an AR (autoregressive) univariate model. The process generated by the VAR model is given by:

$$\Phi_0 y_t - \sum_{j=0}^{q} \Phi_j y_{t-j} = \varepsilon_t$$

where y_t is a vector $N \times 1$; Φ_j are matrices $N \times N$ with $\Phi_0 = I_e \epsilon_t \sim NI_N(0, \Omega)$.

In addition to the VAR models, the VEC model incorporates a long-term relationship into the VAR relationship, which may improve the estimations.

Results

Economic agents are very interested in forecasting the future value of assets. Researchers and market participants try to find ways to use the historical prices of assets to obtain information useful for forecasting their future prices.

Figures 1 and 2 present the logarithm of both series studied. The graphs show that the series are not stationary and their evolution is apparently very similar, which shows a possible long-term relationship among them.

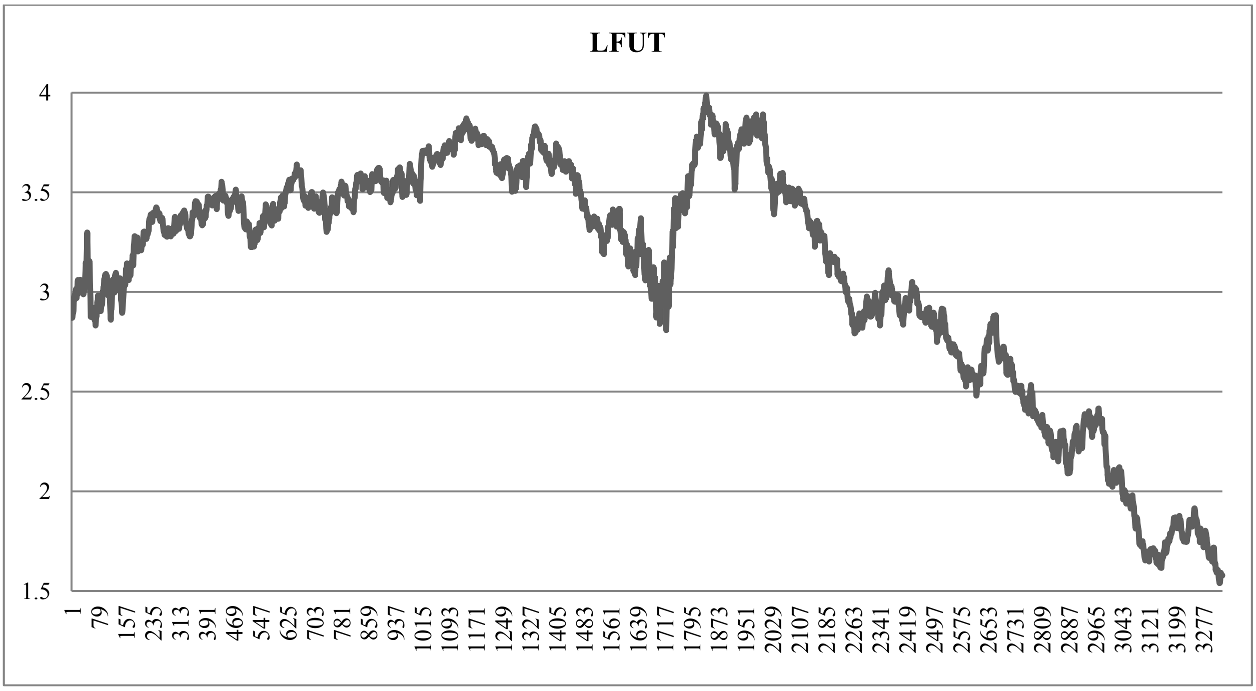

Figure 1. Series of the futures Ibovespa logarithm.

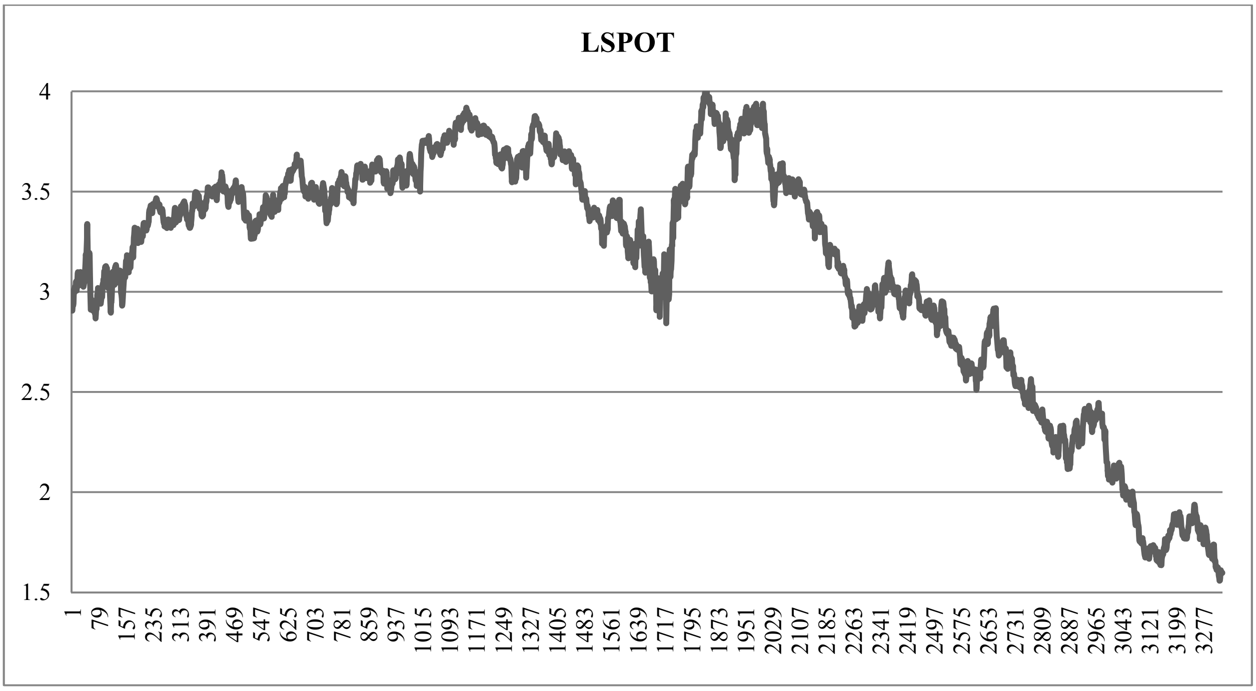

Figure 2. Series of the spot Ibovespa logarithm.

Table 1 shows the descriptive statistics of both univariate series. The descriptive statistics show that both series are negatively a symmetric. The Jarque-Bera test has been used to reject the null hypothesis of normality for each of them, which was already expected in the financial series. All positioning measurements and dispersion measurements are very close across both series.

Table 1
Descriptive Statistics

	LFUT	LSPOT
Mean	10.95790	10.95323
Median	10.96202	10.95495
Maximum	11.06037	11.05586
Minimum	10.86828	10.86558
Std. Dev.	0.047910	0.047264
Skewness	-0.020506	-0.004797
Kurtosis	1.896665	1.875511
Jarque-Bera	169.2421	175.5642
Probability	0.000000	0.000000
Observations	3,332	3,332

Stationarity of the Series

The augmented Dickey-Fuller (ADF) test for a unit root was individually performed on each of the series, and showed that the null hypothesis of a unit root was not rejected for each of the series (p-value = 0.1363 (Ibov Spot); p-value = 0.1361 (Ibov Fut)).

The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) unit root test, whose null hypothesis is that the series are stationary, was also performed, with the objective to obtain additional confirmation of the results obtained with the ADF test. The KPSS tests rejected the null hypothesis that the series are stationary, with 1% significance (p-value < 0.001). Thus, the results obtained with the ADF test are confirmed by the KPSS test. Therefore, both series are integrated of order 1.

In order to verify that the series became I(0), ADF unit root tests were performed for the first difference of both series (p-values < 0.001), indicating that both integrated series are stationary.

Figures 3 and 4 show the graphs of both integrated series, indicating their stationarities.

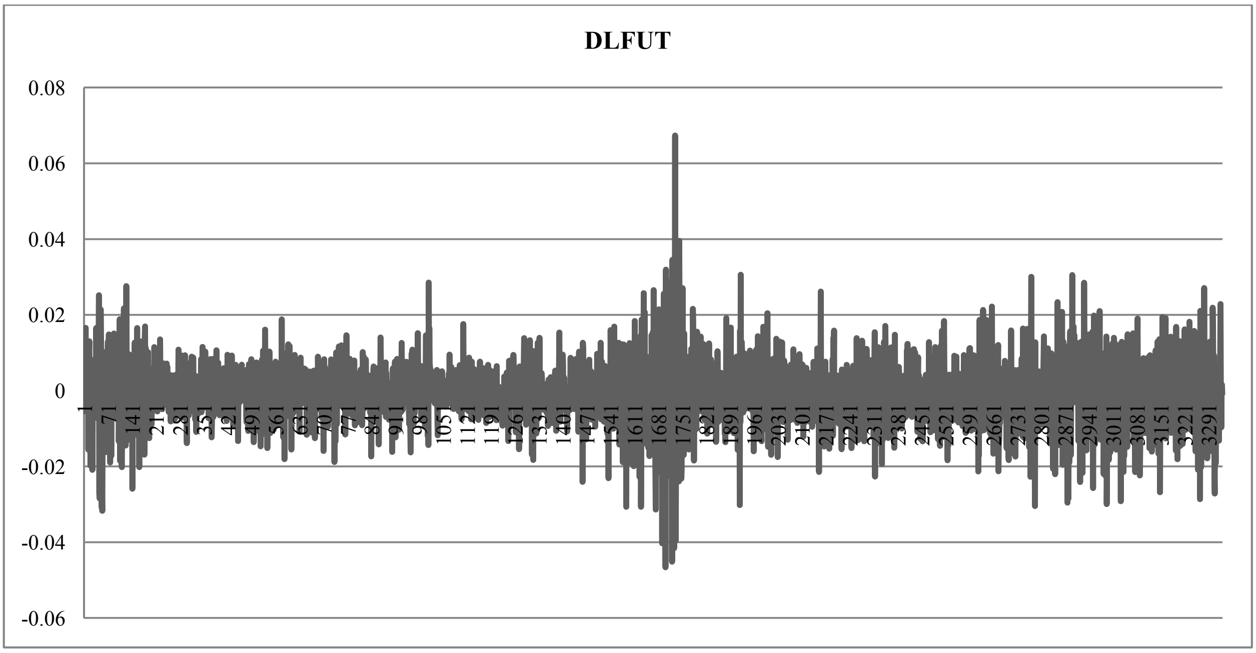

Figure 3. Graphs of the differentiated series (futures).

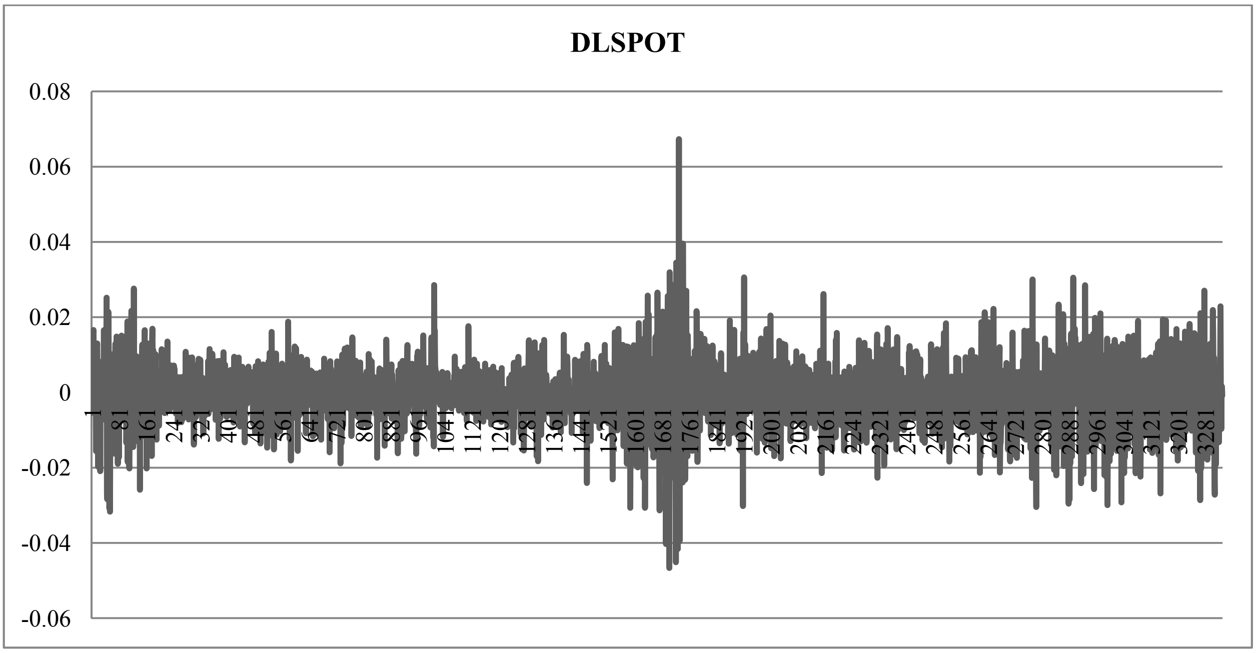

Figure 4. Graphs of the differentiated series (spot).

VAR

The first model to be estimated will be the VAR, whose objective is to confirm the relationship among the current values of the Bovespa index, its previous values, and the values of the Bovespa futures index. The first condition to estimate a VAR model is that the variables in the model as previously shown must be stationary, the variables *d* (log) of both series are I(0), and they will be used to estimate the model.

Initially, as per the Akaike information (AIC), the lag order chosen for the VAR model is 4, and without the constant, with 1% significance. Given that the coefficients for lag 4 were not significant, the VAR model was estimated again using 3 lags. It improved the model in terms of reducing the AIC.

All the coefficients were shown to be significant (with 10% significance) for the equation relative to the spot market. Hence, the Bovespa futures index and the Bovespa index lags can forecast the current values of the Bovespa index, as had already been observed for other financial markets (Stoll & Whaley, 1990). With regard to the equation of the Bovespa futures index, several coefficients were not significant, with 10% significance.

Thus, the model found was as follows:

$$\Delta X_{1,t} = -0.3244\Delta X_{1,t-1} - 0.2995\Delta X_{1,t-2} - 0.1203\Delta X_{1,t-3} + 0.3295\Delta X_{2,t-1} + 0.3329\Delta X_{2,t-2} + 0.1084\Delta X_{2,t-3}$$

$$\Delta X_{2,t} = 0.1286\Delta X_{1,t-1} - 0.1191\Delta X_{2,t-1}$$

where $X_{1,t}$ = Ibovespa(spot) and $X_{2,t}$ = Ibovespa(future).

After estimating the coefficients, the next stage consists of testing the normality of the residuals and determining whether they are autocorrelated.

The residuals normality hypothesis is rejected by the Jarque-Bera test (p-value < 0.001). However, according to Brooks et al. (2001), the violation of this premise does not have any major consequences for large samples. The residuals did not present significant autocorrelations.

VEC

The Granger causality test verified that DLFUT Granger causes DLSPOT (p-value < 0.001) and non-Granger DLSPOT causes DLFUT (p-value = 0.1055), as determined by the earlier model.

Using the Johansen cointegration test, along with AIC (AIC = -20.53344*), the null hypothesis of cointegration of order 1 was not rejected and the chosen model had no constant and no tendency.

After imposing a restriction of the LSPOT coefficient equal to 1 and LFUT equal to -1, this hypothesis was rejected with 1% significance (p-value < 0.001). Thus, the estimated cointegration coefficients are valid.

The restriction that the cointegration coefficient of LSPOT was equal to 1 and the load coefficient of DLSPOT was equal to zero was also imposed. This restriction was not rejected with 1% significance (p-value = 0.6441). In the same vein, the restriction that the cointegration coefficient of LSPOT was equal to 1 and the load coefficient of DLFUT was equal to zero was also imposed. This restriction was rejected with 11% significance (p-value = 0.1067).

Therefore, the model found was as follows:

$$\Delta X_{1,t} = -0.3029\Delta X_{1,t-1} - 0.2517\Delta X_{1,t-2} + 0.3098\Delta X_{2,t-1} + 0.2864\Delta X_{2,t}$$

$$\Delta X_{2,t} = 0.0182(X_{1,t-1} - 0.9996X_{2,t-1}) + 0.1132\Delta X_{1,t-1} - 0.1045\Delta X_{2,t-1}$$

The autocorrelations of the residuals are controlled and the Jarque-Bera normality test was significant (p-value < 0.001), rejecting the normality of the residuals hypothesis. As has been previously mentioned, the rejection of this premise does not have a great influence on the results for large samples.

ARFIMA

Following Fonseca et al. (2012), the ARFIMA model will be used. The last model to be used will be an ARFIMA model for the univariate series of the Bovespa spot index. This model will be built so as to facilitate comparison with the forecasts produced using the earlier models.

The correlogram (not shown here) of the spot series shows the presence of long memory in the process, suggesting the use of fractional difference in the ARIMA model. It can be tested with the R/S test, where the value of H was significant (p-value < 0.001), indicating that the series presents long memory. The parameter d was estimated with the Geweke and Porter-Hudak (1983) method. The d value estimated with this method was 0.92.

Thus, the best model found was ARFIMA (1, 0.92, 1) shown in Table 2.

Table 2

ARFIMA Model

Dependent variable: Ibov-spot			
Coefficient	Standard error	z	p-value
Phi1 = 0.99711	0.0040897	243.8103	0.00001
Theta1 = 0.189649	0.0524204	36.179	0.00030
Dependent var. average	35.21632	D.P. dependent var.	1,088.413
Innovations average	2.940327	D.P. of innovations	1,005.842
Log-likelihood	-27,766.72	AIC	55,539.44
Schwarz criterion	55,557.78	Hannan–Quinn criterion	55,546

Forecast

The VAR, VEC, and ARFIMA models were used to make forecasts for the following 10 working days, with a total of 280 observations. Using the same criterion as Fonseca et al. (2012) proposed, the mean absolute percentage error (MAPE) was used to measure the performance of these three forecasting models.

Table 3

MAPE

	VAR	VEC	ARFIMA
MAPE	0.17%	0.15%	0.19%

Table 3 shows that the VEC model presented a lower MAPE, indicating that it is the model whose forecasts are closer to the observed value, that is, with a lower mean percentage error, in relation to the other two models used.

Conclusions

The objective of this study was to apply the VAR and VEC multivariate models using high frequency data from the Ibovespa spot and futures markets. The ARFIMA model was also used to model the univariate series of the Ivobespa spot market for comparing forecasts.

The Ibovespa future Granger series was found to cause the Ibovespa spot, while the non-Granger Ibovespa spot causes the Ibovespa future. In terms of the forecasts produced, the VEC model proved to be superior to others in terms of mean absolute percentage error, showing the lowest mean percentage error of 0.15%.

Based on the results presented above, the possibility of arbitrage in the market was verified, considering the spot and futures indexes. Although the models presented are focused on forecasting, such results provide the basis for the adoption of trading strategies, as proposed by Fonseca et al. (2012).

This study may be extended by increasing the data collection period; despite the successful estimations of the model, the period analyzed includes only six months of observations, which could introduce a bias in the results obtained. Furthermore, using other information criteria, like those of Schwarz or Hannan-Quinn, may lead to substantially different models, which could corroborate or even refine the present findings.

References

Bachelier, L. (1964). Theory of speculation. In P. Cootner (Ed), *The random character of stock market prices*. Cambridge: MIT Press.

Box, G. E. P., & Pierce, D. (1970). Distribuition of residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Statistical Association, 65,* 1509-1526.

Brooks, C., Rew, A. G., & Ritson, S. A. (2001). Trading strategy based on the lead-lag relationship between the spot index and futures contract for the FTSE 100. *International Journal of Forecasting, 17,* 31-44.

Conrad, C., Rittler, D., & Rotfuß, W. (2012). Modeling and explaining the dynamics of European Union allowance prices at high-frequency. *Energy Economics, 34*(1), 316-326.

Cowles, A. (1933). Can stock market forecasters forecast? *Econometrica, 1,* 309-324.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance, 25,* 383-417.

Fonseca, N. F., Lamounier, W. M., & Bressan, A. A. (2012). Abnormal returns in the Ibovespa using models for high-frequency data. *Brazilian Review of Finance, 10*(2), 243-265.

Geweke, J., & Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis, 4,* 221-238.

Kang, S. H., Cheong, C., & Yoon, S. M. (2013). Intraday volatility spillovers between spot and futures indices: Evidence from the Korean stock market. *Physical A: Statistical Mechanics and Its Applications, 392*(8), 1795-1802.

Lien, D., & Tse, Y. K. (1999). Forecasting the Nikkei spot index with fractional cointegration. *Journal of Forecasting, 18,* 259-273.

Oliveira, F. A., Nobre, C. N., & Zárate, L. E. (2013). Applying artificial neural networks to prediction of stock price and improvement of the directional prediction index—Case study of PETR4, Petrobras, Brazil. *Expert Systems With Applications, 40*(18), 7596-7606.

Rittler, D. (2012). Price discovery and volatility spillovers in the European Union emissions trading scheme: A high-frequency analysis. *Journal of Banking and Finance, 36*(3), 774-785.

Silva, D. T. (2006). *O Conteúdo Informacional Dos Contratos Futuros de IBOVESPA* (Doctoraldissertation, Tese (Doutorado em Contabilidade), Departamento de Contabilidade e Atuária, Universidade de Sao Paulo, Sao Paulo).

Stoll, H., & Whaley, R. (1990). Stock market structure and volatility. *Review of Financial Studies, 3,* 37-71.

Tse, Y. K. (1995). Lead-lag relationships between spot index and futures price of the Nikkei stock average. *Journal of Forecasting, 14,* 553-563.

Yang, J., Yang, Z., & Zhou, Y. (2012). Intraday price discovery and volatility transmission in stock index and stock index futures markets: Evidence from China. *The Journal of Future Markets, 32*(2), 99-121.

Zhang, M. Y., Russel, J. R., & Tsay, R. S. (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics, 104,* 179-207.