COLEGIO UNIVERSITARIO DE ESTUDIOS FINANCIEROS

GRADO EN ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESA – PROGRAMA BILINGÜE

Trabajo Fin de GRADO

"Econometrics models for the weekly high and low prices forecast of Iberdrola S.A"

Econometrics Department

Author: Cristina Rivero Navarro
Mentor: Rafael Flores De Frutos

Madrid, May 8th, 2020

Index.

1. Abstract.

This paper contributes to the theory of technical analysis for the forecasting of stock yields, analysing the past of the stocks to forecast and predict the future yields. Also, to investigate the efficiency of markets, proving the weak and semi-strong efficiency. The study is based on the paper carried out by Caporin et al. (2013) of stock predictability. It is proposed to extend the research of Caporin et al (2013) and prove predictability for Iberdrola with a vector error-correction model.

Key words:

- Market weak and semi-strong efficiency.
- Maximum, minimum, closing and NYSE composite price.
- Yield performance.
- ARMA Model.
- Co-integration and correlation relationships.
- Vector Error-Correction Model.

2. Introduction.

The research carried out by Caporin et al. (2013) looks forward to proving the predictability of Dow Jones daily high and low prices, with the conclusion that prices are predictable. These authors also investigate if the technical analysis used is improvable.

The objective of this paper is to extend the distinguished study accomplished by Caporin et al. (2013) creating a multivariate stochastic model, specifically a vector error-correction model.

First, we must say this paper is based on the theory that Iberdrola stock yields can or cannot be predicted using only past information about its stock prices, rather than any other information. Therefore, the information analysed is the historical data of high, low and closing prices of the company from the IBEX 35 stock market for a period of ten years. Moreover, this paper includes data from the NYSE composite Index to prove if the information is useful trading rules.

So far, the study solely collects the information as the article Caporin et al (2013). Whereas, as said before, the present paper aims to improve this article, enlarging the information by adding a variable, closing price, as it is directly related to stock maximum and minimum prices and the presented information is consistent and relevant. In addition, instead of considering daily data, with the objective of diminishing volatility and market excessive fluctuations, weekly data is considered.

The study is structured in four main areas. First, at the section 2, it is discussed different opinions referring to the predictability of market performance and market efficiency. Then, section 3, where an empirical evidence of this model is determined. It is followed by section 4, an explanation of the trading strategy to be performed. Finally, section 5 concludes the paper.

3. Survey: The predictability of stock prices.

3.1 Eugene Fama, Lars Peter Hansen and Robert Shiller.

This section aims to emphasize the studies made by the 2013 Nobel prize winners, Eugene Fama, Lars Peter Hansen and Robert Shiller, for their empirical study on the calculations of yields forecasting.

Fama and Shiller are economists with contrasting ideologies in this field. Hansen is an econometrician. Fama is a professor at The University of Chicago, he is a true supporter of the theory of "Efficiency Rational Expectations", placing him in opposition to market bubbles, as he claimed for the New Yorker in 2010: "It's easy to say prices went down, it must have been a bubble, after the fact. I think most bubbles are twenty-twenty hindsight. (…) They have to be predictable phenomena." See: John Cassidy (2010). On the other hand, Shiller, professor at Yale University, states on the idea where markets are driven by "Behavioural Economics", promoting market bubbles. Finally, the econometric Hansen, professor at the University of Chicago, criticises proves the rejection for the basic CCAPM (Consumption-Capital Asset Pricing) model.

Eugene Fama promoted the belief of being very challenge the forecast of yield performance in the short term, stating that prices follow a random walk. See Malkiel, B. G. (2007).

Whereas, this author also shares that the most accurate prediction for the price of a stock is for the period "t+1", conditioned to all know until "t (Ωt)", the price at "t". The prices forecast implies a consequence of lack of efficiency.

The author refers to “weak efficiency” of the information when the error (at) is not auto-correlated. When the error (at) is auto-correlated with the public information at “t” exists “semi-strong efficiency”. And, when the error(at) is auto-correlated with private information there is “strong efficiency”. Therefore, when the market yields are unforeseeable and follows a random walk, there is efficiency. See Fama (1988)

On the other hand, **Robert Shiller** explaining long term forecast of stock yields has a total different perspective than Fama. This author believes in “Behavioural Economics” where the economic agents are not rational or behave in an efficient way, instead, they not only miscalculate on how to measure risk or performance of prices, but they follow them sense and feelings.

Furthermore, this author proved that the variances of the forecasted prices were higher than the variances of the forecasted dividends discounted at a constant interest rate. Concluding that dividends do not have enough volatility to explain stock performances fluctuations.

Shiller provided numerous empirical analysis proving the prediction of stock performances. Identifying financial ratios has been one of his most famous contributions. It is worth noting “dividend/price” ratio or “price/mean profit” ratio as essential for speculative bubble prediction. A high price compared to historical mean profits may be a warning signal. The prediction of market bubbles as 90s crises or 2008 real estate crises by Shiller, are based on these ratios.

Although Fama and Shiller may seem to follow contrary tendencies, both agree on the forecast of yield performances. They also share the idea of the increase in the predictability when the prevision horizon increases. Shiller defends that performances may be forecast on the long run, while Fama and EER defenders

do not deny long-term forecast. These authors disagree as to why of that forecast. To conclude, predictability is found whereas it is inaccurate. See Flores (2014)

3.2. Caporin, Ronaldo and The Magistris.

This area aims to illustrate the study carried out by Caporin et al. (2013), in which paper is based on the analysis of the historical data for a period of eight years about the maximum and minimum prices for the Dow Jones index, concluding with its usefulness in future stock prices and providing better strategies rather than buy/hold.

To prove the target goal, these authors propose three questions: are the high and low prices of stocks predictable; can we build a model based on that statement? are the highs and lows predictions useful when improving technical analysis techniques for trading?

For these authors, it is essential using historical highs and lows, because they find it is a relevant and useful information source. Also, because they are collected and can be used. High and low prices are co-related and it is common for these two to fluctuate in the long run. Thus, the answer to the first question is yes, these performances are foreseeable.

To answer the second question, we must count with a vector error-correction model, as the information has more than one variable and the information provides co-relation relationships. Therefore, is it possible to build a model which foresees stocks.

Finally, these authors manage to provide a model with strong capacity of predictability. The model provided a period of efficiency, whereas not total nor irrational efficiency.

The presented paper aims to build a VECM model comparable to the one proposed by Caporin et al. (2013), whereas it will be broadened.

Thus, this paper will consider Caporin et al. (2013) for the following purposes. It can be highlight the employment of three variables, the maximum, minimum and an index prices. These variables will be proved to also be predictable and present correlation. Besides, this project will provide a VECM model, as it counts with correlated relationships and the group of information demonstrates a long-term range. Finally, this study will suggest a trading strategy.

The techniques this paper follows in order to expand the article of Caporin et al. (2013) are the following ones. First, it will include the variable of the closing price, a relevant and with high degree of information variable. Also, the group of information will consist on weekly historical data instead of daily, this way the unpredictability and volatility of a market will decrease.

4. Empirical analysis.

To test the weak efficiency and the semi-strong efficiency an empirical analysis is computed on the data of Iberdrola S.A.

4.1. Data

To conduct the model and the study of the possible yield forecast of the shares it is essential mentioning that the data gathered for the paper consist of a total of 527 observations. This data is composed by the historical public information of the high, low and closing prices of the company from the IBEX 35 stock market. It corresponds to the weekly period from January 2010 to January 2020. All data are collected from http://www.bolsamadrid.es/. This paper also counts with data from the NYSE composite, a stock market index and measures the performance of all the stocks that perform on the New York market.

After the information is gathered, the data is used in the form of a logarithm to induces normality. The variables maximum, minimum and closing price are non-stationary and present co-integration.

Chart 1: Logarithms of maximum, minimum and closing prices.
Source: Own elaboration with Eviews.

It is essential when estimating a model knowing if all the variables are relevant or not. For this reason, the variables are transformed into stationary variables implementing a difference to the logarithms.

Variable	**VAR**	**Log**	**Interest**
Maximum	MAX	LMAX	∇LMAX
Minimum	MIN	LMIN	∇LMIN
Closing	ULT	LULT	∇LULT
Composite	DJ	LDJ	∇LDJ

Chart 2: Logarithms of maximum, minimum and closing prices with an increment obtaining them yield.

Source: Own elaboration with Eviews.

4.2. ARMA Model.

ARMA model will allow us to check if the variables present a weak efficiency. This tool is the appropriate for explaining generator process of the set of information presented, one variable and its pasts, because this family can identify the generator process or approach any stochastic process if the generator does not reside in the ARMA family. This model has the following general structure:

$$W_t = \delta + \varepsilon_1 W_{t-1} + \varepsilon_2 W_{t-2} + \cdots + \varepsilon_p W_{t-p} + a_t$$

First, it is essential to identify P, checking the correlograms of the variable in the partial correlation. Also, it is necessary that the error "a_t" is white noise, as the autocorrelation coefficients must be equal to zero, to test the hypothesis.

$$W_t = \delta + \varepsilon_1 W_{t-1} + \varepsilon_2 W_{t-2} + \cdots + \varepsilon_p W_{t-p} + a_t$$

Estimation of the ARMA Model					
	δ	ε_1	ε_2	R^2	P
∇LMAX	-	0.0875	-	0.68%	1
∇LMIN	-	0.1551	-0.1451	3.83%	2
∇LULT	-	-0.0789	-0.0771	1.07%	2
∇LDJ	-	-	-	1.23%	-

To conclude, the prices of a stock follow a univariate process as R^2 presents a low value for all the variables. This process is very close to a random walk, therefore weak efficiency cannot be excluded.

Once the weak efficiency has been tested, it is necessary to test the semi-strong efficiency, since the necessary data are presented and, therefore, leaving the paper this way would be insufficient.

4.3. VECM Model.

The semi strong efficiency is going to be investigated as the set of information is not only constitute by one variable, but it is composed by four variables: maximum, minimum, closing and index prices.

From the econometric point of view, the most suitable instrument to investigate the generator process of this four-variable vector will be the Vector Error-Correction Model. This vector has the following general structure:

$$\nabla Y_t = \beta_{Z_{t-1}} + \gamma_1 \nabla Y_{t-1} + \gamma_2 \nabla Y_{t-2} + \gamma_3 \nabla Y_{t-3} + \cdots + \gamma_{p-1} \nabla Y_{t-(p-1)} + a_t$$

Each segment of the general formula is explained as follows:

∇Y_t

Is the vector 4x1 of the variable.

$$\beta = \begin{pmatrix} \beta_{1,1} & \beta_{1,2} \\ \beta_{2,1} & \beta_{2,2} \\ \beta_{3,1} & \beta_{3,2} \\ \beta_{4,1} & \beta_{4,2} \end{pmatrix}$$

Is the matrix 4x2 of coefficients associated to the co-integration relationships.

$$z_{t-1} = \begin{pmatrix} CR1_{t-1} \\ CR2_{T-1} \end{pmatrix}$$

Is the vector 2x1 of the co-integration variables

$$\gamma_{p-1} = \begin{pmatrix} \gamma_{1\,1,p-1} & \gamma_{1\,2,p-1} & \gamma_{1\,3,p-1} & \gamma_{1\,4,p-1} \\ \gamma_{2\,1,p-1} & \gamma_{2\,2,p-1} & \gamma_{2\,3,p-1} & \gamma_{2\,4,p-1} \\ \gamma_{3\,1,p-1} & \gamma_{3\,2,p-1} & \gamma_{3\,3,p-1} & \gamma_{3\,4,p-1} \\ \gamma_{4\,1,p-1} & \gamma_{4\,2,p-1} & \gamma_{4\,3,p-1} & \gamma_{4\,4,p-1} \end{pmatrix}$$

Is the matrix 4x4 of coefficients associate to the lag p-1 of the vector ∇Y_t.

Once the general model is explained, the next step is to determine the order of the VAR in levels with the objective of finding the order to be used in the VECM model. For this reason, the VAR model is estimated. As the coming table demonstrates, following the AIC method, P = 4. It is important to prove and confirm that the residuals are not white noise.

VAR Lag Order Selection Criteria
Endogenous variables: LMAX LMIN LULT LDJ
Exogenous variables: C
Date: 04/22/20 Time: 11:41
Sample: 1/04/2010 2/03/2020
Included observations: 521

Lag	LogL	LR	FPE	AIC	SC	HQ
0	2842.107	NA	2.18e-10	-10.89485	-10.86217	-10.88205
1	5463.940	5193.341	9.87e-15	-20.89804	-20.73467*	-20.83405
2	5506.266	83.19005	8.92e-15	-20.99910	-20.70504	-20.88391*
3	5521.597	29.89817	8.94e-15	-20.99653	-20.57178	-20.83015
4	5538.863	33.40388*	8.90e-15*	-21.00139*	-20.44594	-20.78382
5	5549.118	19.68413	9.10e-15	-20.97934	-20.29319	-20.71057

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

Chart 3: Table of VAR lag order.

Source: Own elaboration with Eviews.

The co-integration is defined as a strong relationship between two variables. It is said that when two variables present a correlation relationship they fluctuate in parallel.

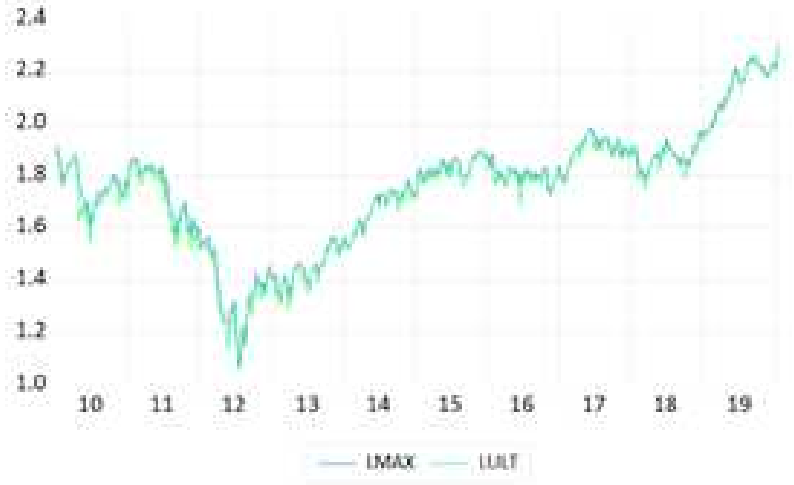

Chart 4: Maximum and last prices of the company fluctuation

Source: Own elaboration with Eviews.

The theorem of Causality of Granger (Wiener-Granger, 2003) affirms that for a vector of four variables there is a maximum of three correlation relationships. Also, via Johansen's method (Johansen and Juselius, 1990) we find two relationships of co-integration. Johansen's resolve the constraints imposed by the co-integration of the series included in an unrestricted VAR model.

Johansen's test has been carried out to find out which and how many CRs exist in this vector.

The Correlation Relationship founded are:

$$CR1_t = \nabla \mathrm{LMAX}_t - 0.9441 * \nabla \mathrm{LMIN}_t - 0.1423$$
$$(0.0043) \qquad (0.0074)$$

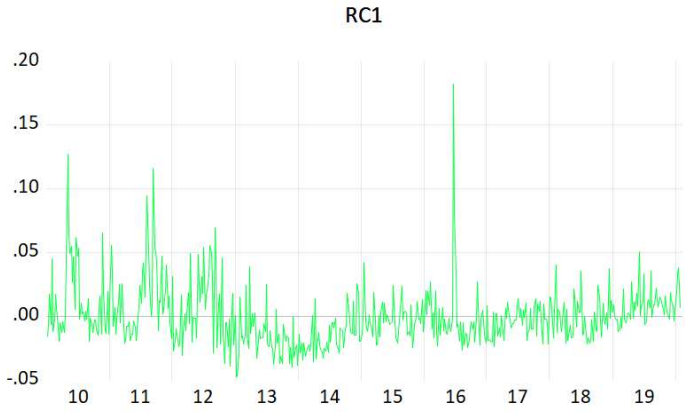

Chart 5: Co-integration relationship 1.
Source: Own elaboration with Eviews.

$$CR2_t = \nabla \mathrm{LMAX}_t - 0.9719 * \nabla \mathrm{LULT}_t - 0.7015$$
$$(0.0039) \qquad (0.0069)$$

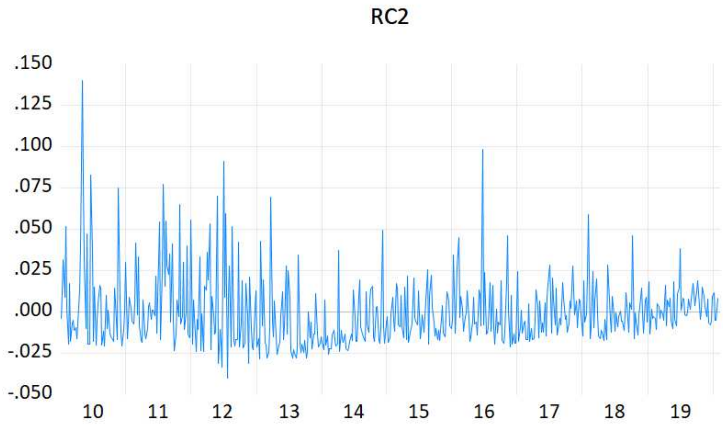

Chart 6: Co-integration relationship 2.
Source: Own elaboration with Eviews.

$CR1$ and $CR2$ will be included in the general model as additional variables to the vector. The information now it is wider and related to the stock prices.

Also, the study of an $CR3$ for the variables $\nabla\mathrm{LMIN}$ and $\nabla\mathrm{LULT}$ has been carried out and it has been determined not to be possible to use it since it is a linear combination of the others, it is not independent.

Moreover, after studies of co-integration of $\nabla\mathrm{LDJ}$ with the other variables, it is determined that there is no relationship.

To conclude, there are two clear groups. On the one hand $\nabla\mathrm{LMIN}$, $\nabla\mathrm{LMAX}$ and $\nabla\mathrm{LULT}$, and on the other $\nabla\mathrm{LDJ}$. The two co-integration relationships explained before are the only ones that exist.

The estimation for model is presented in the following chart.

Estimation of the Vector Error-Correction Model				
	$\nabla\mathrm{LMAX}$	$\nabla\mathrm{LMIN}$	$\nabla\mathrm{LULT}$	$\nabla\mathrm{LDJ}$
C	-	-	-	-
RC1(-1)	0.2305	0.6692	-	-
RC2(-2)	-0.9113	-0.8800	0.2123	-
$\nabla\mathrm{LMAX}$ (-1)	-0.0877	0.2961	-	-
$\nabla\mathrm{LMAX}$ (-2)	-	0.1147	-	-
$\nabla\mathrm{LDJ}$ (-1)	-0.1395	-	-	-
$\nabla\mathrm{LDJ}$ (-2)	-	-	-	-
$\nabla\mathrm{LDJ}$ (-3)	-	-	-	-0.1238
$\nabla\mathrm{LMIN}$ (-1)	-	-0.1533	-	-
$\nabla\mathrm{LMIN}$ (-2)	-	-0.1309	-	0.0947
$\nabla\mathrm{LULT}$ (-1)	-	-	-	-0.0918
$\nabla\mathrm{LULT}$ (-2)	-	-	-	-0.0649
R^2	37.04%	39.69%	17.65%	37.31%
Adjusted R^2	36.67%	39.11%	17.65%	31.73%
S.D dependent variable	2.51%	3.05%	3.34%	1.97%
Q Statistic (6)	2.93	2.04	3.91	3.17
P-value (6)	0.818	0.916	0.689	0.787

The prices of the stock present a high value of R^2 the variables of the maximum, minimum and index prices. On one hand, it is possible to state that the past affects the present not accepting the semi-strong efficiency for these two variables. And on the other hand, it is possible to accept semi-strong efficiency in the closing price because of having a low value for R^2 it could be said that this variable present a random walk.

6. Conclusion

To conclude, the questions presented in Caporin et al. (2013), and motivated in the introduction section, are responded.

Referring to the forecast of yield performance for Iberdrola stocks, it can be concluded that the maximum and minimum prices can be predicted with a 37.04% and a 39.69%, respectively. Whereas, the closing price has a poor degree of predictability.

When attributing efficiency to markets, it can be confirmed the existence of weak efficiency as all the variables almost follow a random walk. On the other hand, it be can also confirm the rejection to semi-strong efficiency for maximum and minimum prices, therefore, present a predictable Price.

7. Bibliography.

Campbell, J y Shiller, R. (1998) The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *Review of Financial Studies.*

Caporin, M., Ronaldo, A., & De Magistris, P. (2013). On the predictability of stock prices: A case for high and low prices." *Journal of Banking & Finance.*

Cassidy, J. (2010). "Interview with Eugene Fama." The New Yorker.

Fama, E.F. and K.R. French (1988), "Dividend yields and expected stock returns," *Journal of Financial Economics.*

Flores de Frutos, R. (2014). Eficiencia versus exuberancia irracional. *Economistas.*

Flores de Frutos, R. "Autoregressive Vector Processes". Curse *notes.*

Jafari-Mamaghani M. (2014), "Robust non-linear Wiener-Granger causality for large high-dimensional data." *Journal of Mathematical Statistics.*

Johansen, S., & Juselius, K. (1990). Maximum likelihood estimation and inference on co-integration—with applications to the demand for money. *Oxford Bulletin of Economics and statistics.*

Malkiel, B. G. (2007). A random walk down Wall Street: the time-tested strategy for successful investing. *WW Norton & Company.*

Shiller, R. J. (1990). Market volatility and investor behaviour. The American Economic Review.