Predictive output validation of almond prices.

Author: Rubén Adell Sales

Degree: Economics

Group: E

Email: Al258381@uji.es

Professor: Gabriele Tedeschi

-Abstract

The almond market is especially important in Castellón area and a large part of the farmland in the province is used for its production. The new consumer trends and new production techniques have meant an improvement in the sector in recent years, but taking into account that it is a fairly traditional sector there is a resistance to change.

There are not specific quantitative studies about Spanish almond market, so this paper aims to analyse the prices of almonds at the national level.

To carry out this study, we have used the techniques “in sample” and “out sample” calibration, consisting on the development of several models. In the first test we are going to take the whole data set (in sample) and in the second calibration we will take data out of the sample with "t-1" to "t-5".

The object of the study is to analyse two of the most representative varieties of almonds nationwide: Marcona and Comuna. We have created two models for each one: on the one hand OLS and on the other AR (1) using the same explanatory variables for both.

The results after both calibrations show that, as a general rule, in this market, AR models have a better adjustment to reality, and in terms of the difference between varieties, in both calibrations the Marcona extracts more robust models.

Keywords: Calibration, Validation, Forecasting, Commodity, Farming, Prices

Acknowledgments:

-To Frusema and Importaco companies for the good treatment. And for provide me with all the information I have requested, I would not have been possible this paper without your help.

-To my family, who has always supported me even though sometimes I don't deserve it. Especially my sister Aloma.

- My tutor Gabriele. I am grateful to him for accepting me as his student for this project. Whenever I have needed advice he has taken care of me. It has been a pleasure learn with he and I am very grateful.

- To my family in Exeter. Rebbeca, Lester, Christopher and Holly. They have given me the courage to write the project in English myself.

- Index

1. Introduction

Agriculture is a traditional sector in the province of Castellón. It has been estimated that around 377,000 hectares are dedicated to agricultural exploitation in 2016. Its main crops are citrus fruits such as tangerine or orange, olives and almond cultivation among others.

Among these crops, the almond sector is especially important in the province's interior. A few years ago, this area was mainly dedicated to the olives cultivation and oil production, however due to the great frost occurred in 1956 olive crop was devastated and plunged the sector into poverty.

The need to repopulate the fields and the fear of other frosts introduced the almond tree as an alternative. The main reasons why the almond tree was chosen were the almond economic performance which is higher than olives, and the tree resistance to frosts.

In the following years, the industrialization of the economy did the primary sector goes to the background. The population in rural areas started to dwindle to the benefit of the cities, so many of the agricultural field were abandoned.

In spite of this, in the next years the almond sector is being reinvested. There are being improvements in productive techniques and carrying out studies for the development of new varieties, which are more resistant to low temperatures and more productive. As a consequence the almond production has not stopped increasing.

The almond has been associated with snacks or the manufacture of sweets, such as nougat or other pastries, so its market has not been too broad. This trend has been changing as a consequence of many studies carried out to highlight the healthy properties of nuts (see Rune Blomhoff, Monica H. Carlsen, Lene Frost Andersen and David R. Jacobs, 2006; Emilio Ros and José Mataix, 2007).

In a society increasingly concerned about health, almonds have become common nutrient in most diets. This fact has become an opportunity in the sector that has reopened a path that seemed closed; agriculture can be a sustainable way of life in our region.

At global level, among world almond producers Spain ranks third. This position may suggest that the country has a strong influence on the sector, but in absolute terms only represent 7% of total production. On the top of the ranking, the USA with an approximated 80% of the almond market share, has the greatest impact; the second place is for Australia with 8% of the total. As we have said, the market force is to the USA so it is reasonable that its prices mark the path to follow for the rest of the economy.

This study intends to perform an analysis of the evolution of almond market prices. Although there are a lot of studies about the nuts benefits and its properties, very few investigate the behaviour of its market prices.

With a series of explanatory variables, two models are presented (OLS and autoregressive) for the almond varieties Marcona and Comuna. The project will replicate previous studies of calibrations. (see Maria Cristina Recchioni, Gabriele Tedeschi and Mauro Gallegati, 2015; Carlo Bianchia, Pasquale Cirillo, Mauro Gallegati and Pietro A.Vagliasindi, 2008) For both models using calibrations "in sample" and "out sample", we compare and determine which model adjusts better to reality, and which of the varieties are able to perform with a minor error.

The original idea was to make this comparison using OLS as a traditional econometric tool, and make a comparison with new self-apprehensive techniques such as neural networks. This idea was discarded, since this type of models requires large amounts of data to work optimally, so during the creation of the dataset we realise that there was not enough data in our time series to train a model of these characteristics.

The project is developed with the following steps:

1- **Creation of the database:** The object of the study is to analyse the behaviour of the prices of the two most relevant varieties at the national level. On the one hand the Marcona, this is the most representative variety in the Valencian region, its production is scarce, however it is the most appreciated in the markets for its properties and its price is usually higher than any other variety. Our dataset includes the temporary series of monthly prices from 2013 to 2018, as well as the production during the same period.

   On the other hand, the Comuna, it is a mix of varieties and is not exclusive to Spain. It is produced in large amount throughout the national territory, however for its analysis we have the same time series as the Marcona.

   Two time series of prices of different varieties are introduced as explanatory variables: Largueta and Standad. On the one hand Largueta, it is another national variety and its production characteristics are similar to the Marcona. On the other hand the Standard variety. It is the index variety in the USA, as the main producer has incorporated the time series of its price as an indicator of international prices.

   The last variable taken into account is the price of Hazelnuts as a possible substitute.

   Once the database has been defined, a first analysis is made. Starting from the temporary price series of the marcona and the comuna, the series is broken down into: trend, seasonality and noise. Moreover a simple holt winters regression is performed in order to see its predictions based on its price.

   This first study is compared with the production data and also contrasted with the economic theory. We continue defining the correlations of the explanatory variables with those explained. The conclusions drawn from the dataset are:

   The scarce production of the marcona and the non-perishable property, makes it a highly speculative product, because its correlation with international and substitute prices is very weak or not significant. However it is strongly correlated with the variety Largueta. This fact is attributed to their similar characteristics and also in both cases prices encourage speculation, so their movements correlate.

As for the comuna, it follows a more normal pattern; it correlates positively with the international prices and with Largueta, and negatively with hazelnuts.

2- **In sample Model:** Once the data has been reviewed, the model is created and we perform the in sample calibration:

   The results show that in both models the R2 is higher for the case of Marcona, and its errors are minor. Although the comuna has a higher number of significant variables; we continue associating this with the strong correlation that Marcona and Largueta keeps.

   Regarding the difference between the models, the autoregressive have better adjusted their results to the real series than the OLS.

3- **Out sample Calibration**: The last step is calibration out sample. To its perform, we proceed to make estimates outside the sample for several periods in order to see which of the models and which variety predicts better. In this case the models will make predictions from t-1 (March 2018) through t-5 (November 2017 to March 2018).

The results in the out sample show that in both models the Marcona variety is better adjusted to reality. In practically all periods its error is lower than for Comuna, which shows an atypical behaviour in this type of calibrations.

Regarding to the models, again the AR shows the best results: in the case of the Marcona, the results are very adjusted to reality, while in the case of the Comuna, although both models continue to show behaviour outside the expected, in all periods the ARs are closer to the real values.

To finish the introduction, I would like to comment on the motivations that led me to propose this topic.

The rural exodus for the benefit of the cities is a fact more and more usual. Agricultural sector has been abandoned by the young people since it was not profitable, nevertheless the new production techniques and the new varieties research are achieving that almond crop reaches historical levels. Moreover new consumer trends have made almonds a very attractive product due to its health benefits. I believe that it can be an opportunity that makes the agricultural sector profitable again.

Depopulation in the interior area of Castellón is probably the biggest challenge we have faced. From my point of view it is the obligation of those who live in, to try to reverse it, so I would like that the objective of this study is to contribute our grain of sand to achieve it.

2. Literary Review

Once the field of study is defined, we will proceed to carry out the literary review with the aim of finding techniques that allow us to face the project. For this we have focused on the market of almonds from two perspectives: firstly qualitative, in order to understand the growth of consumer habits, and secondly quantitative, which will allow us to carry out the project.

From the quality point of view, there are many studies that demonstrate the beneficial properties of almonds in health, Spiller GA (1992) Performed dietetic experiments with a group of 26 people. The baseline diet was modified in a similar way for all subjects by limiting meat, fatty fish, high-fat milk products, eggs, and saturated fat. Grains, beans, vegetables, fruit, and low-fat milk products were the foundation of the diet. During the almond diet period, raw almonds (100 mg/day) supplied 34 g/day of monounsaturated fatty acid (MUFA), 12 g/day of polyunsaturated fatty acid, and 6 g/day of saturated fatty acid. Almond oil was the only oil allowed for food preparation. There was a rapid and sustained reduction in low-density lipoprotein cholesterol without changes in high-density lipoprotein cholesterol. This was reflected in a total plasma cholesterol decrease.

Emilio Ros (2006) defends in his thesis that healthy fats in nuts contribute to the beneficial effects of frequent nut intake observed in epidemiological studies (prevention of coronary heart disease, diabetes, and sudden death) and in short-term feeding trials (cholesterol lowering, LDL resistance to oxidation, and improved endothelial function.

We also find articles that talk about the properties of some derivatives, Zeeshan Ahmad (2009) defends that the almond oil has long been used in complementary medicine circles for its numerous health benefits. Although no conclusive scientific data exists currently, almonds and almond oil have many properties including anti-inflammatory, immunity-boosting and anti-hepatotoxicity effects. Further, associations between almond oil and improved bowel transit have been made, which consequently reduces irritable bowel syndrome symptoms.

Although our study does not has a qualitative basis, the growing social concern about healthful eating habits, together with the studies mentioned, partly explains the increase in consumption and production of this nut. It also gives us more knowledge when we come to drawing non-numerical conclusions.

The project explores the time series of prices to make forecast models that can help us to analyse the effects of external varieties on the typical varieties of national production. For that purpose, we come back to literature to see forecasting methodology.

Paresh KumarNarayan (2010) in his publication, creates a model through OLS for modelling the impact of oil prices on Vietnam's stock prices:

$$\ln SP_t = \alpha_0 + \alpha_1 \ln OILP_t + \alpha_2 \ln ER_t + \varepsilon_t$$

Where SP is the natural record of the stock price series, in OILP it is the natural record of the crude price series, and ln ER it is the natural record of the series of nominal exchange rates.

Paresh KumarNarayan (2010). His article investigates the behaviour of United States stock prices using an autoregressive two-regime threshold model (ART) without restrictions with an autoregressive unit root. The TAR model is applied to the monthly stock price data (Common Stocks of the New York Stock Exchange) for the United States for the period 1964: 06 to 2003: 04. Among the main results, they find that the stock price in The United States is a non-linear series characterized by a unitary root process, consistent with the efficient market hypothesis.

Finally, to make the comparisons. Gabriele Tedeschi (2010) introduce a calibration procedure for validating of agent based models, which shows how a proper calibration allows the model to describe time series of prices.

With this information, we will calibrate the OLS and autoregressive models and compare both by calibrations.

3. The dataset description.

This part explains the variables that are collected to analyse and explain the variations in the almond price. It should be pointed that all data has been collected individually and the database is self-created. The tools used for this analysis have been R studio, Gretl and Knime Analytics.

3.1 The data

-**National almond prices:** Historical series of prices from 2013 to 2018. Table 3.1 contains the data of Marcona, Comuna and Largueta, the three most representative varieties nationwide.

-Marcona is the most representative variety in the Valencian region, its production is scarce, however it is the most appreciated in the markets for its properties and its price is usually higher than any other variety.

Date	Marcona Price	Largueta Price	Comuna Price
2013-01-01	5.2	4.9	4.75
2013-02-01	6.3	6.3	5.55
2013-03-01	6.6	6.5	5.5
2013-04-01	6.65	6.65	5.8
2013-05-01	7	6.75	6
2013-06-01	7.9	6.75	6.5
2013-07-01	7.9	6.75	6.55
2013-08-01	8.5	7	6.3
2013-09-01	8.5	7	5.8
2013-10-01	9.25	8.6	6.8
2013-11-01	9.4	8.6	6.7
2013-12-01	9.5	8.6	6.8
2014-01-01	9.4	8.4	6.8
2014-02-01	9.4	8.4	8
2014-03-01	9.2	8.2	7.6
2014-04-01	9.1	8.1	7.3
2014-05-01	8.75	7.75	6.8
2014-06-01	8	7	6.5

Table 3.1: National Prices

Source: DESCALMENDRA (https://www.descalmendra.com/es/)

- Comuna, is a mix of varieties and is not exclusive to Spain. It is produced in large amount throughout the national territory.

- Largueta, another variety of Spanish production, like Marcona, it has a lower production index, it is more elongated and has a beautiful appearance, for these reasons Largueta is usually sold natural.

Prices are published every three weeks; to adapt them, we have transformed the time series into months. For this purpose we have used the means whenever two prices have been published in the same month. The measure is € per Kg.

The source of the data is Descalmendra, an association of national producers of almond who share the production data and prices together. The data is only available to association members, so we appreciate the collaboration of Frutos Secos del Maestrazgo, SA an association member that has let me access to the database.

-International almond prices: USA prices data are also included, since the country is the main world producer and its prices are expected to have relevance on the national market.

As a representative sample of the international market, we chose the Standard variety, since it is the indicator variety in the USA. Its high production rate and low appreciated quality in the market means that it has lower prices than national varieties analysed.

In this case, the data is a historical series between 2013 and 2018 with monthly inputs, expressed in $/Lb correctly transformed to € / Kg.

Date	Standard Price
2013-01-01	4.589
2013-02-01	4.708
2013-03-01	5.069
2013-04-01	5.239
2013-05-01	5.202
2013-06-01	4.978
2013-07-01	4.975
2013-08-01	5.115
2013-09-01	4.838
2013-10-01	4.97
2013-11-01	5.135
2013-12-01	5.355
2014-01-01	5.624
2014-02-01	5.683
2014-03-01	5.773

Table 3.2: International Prices

Source: https://www.usda.gov/

The prices shown in table 3.2 correspond exclusively to export prices, since we do not consider that internal market is relevant for our research.

The source is the open data agricultural accounts of the US government.

-Hazelnut exportation prices: We have taken into consideration the hazelnut exportation prices because of it belongs to the same family and we forecast that could have a correlation.

Historical data series with the same characteristics as the Standard variety, is the USA General Indicator for exports. For the study it has been transformed from $ / Lb to € / Kg

The source is the open data agricultural accounts of the US government.

Year	Hazelnut Price
2013-01-01	3.678
2013-02-01	3.565
2013-03-01	4.804
2013-04-01	3.19
2013-05-01	3.209
2013-06-01	4.372
2013-07-01	3.866
2013-08-01	3.284
2013-09-01	2.871
2013-10-01	2.852
2013-11-01	2.871

Table 3.3: Hazelnut Prices

Source: https://www.usda.gov/

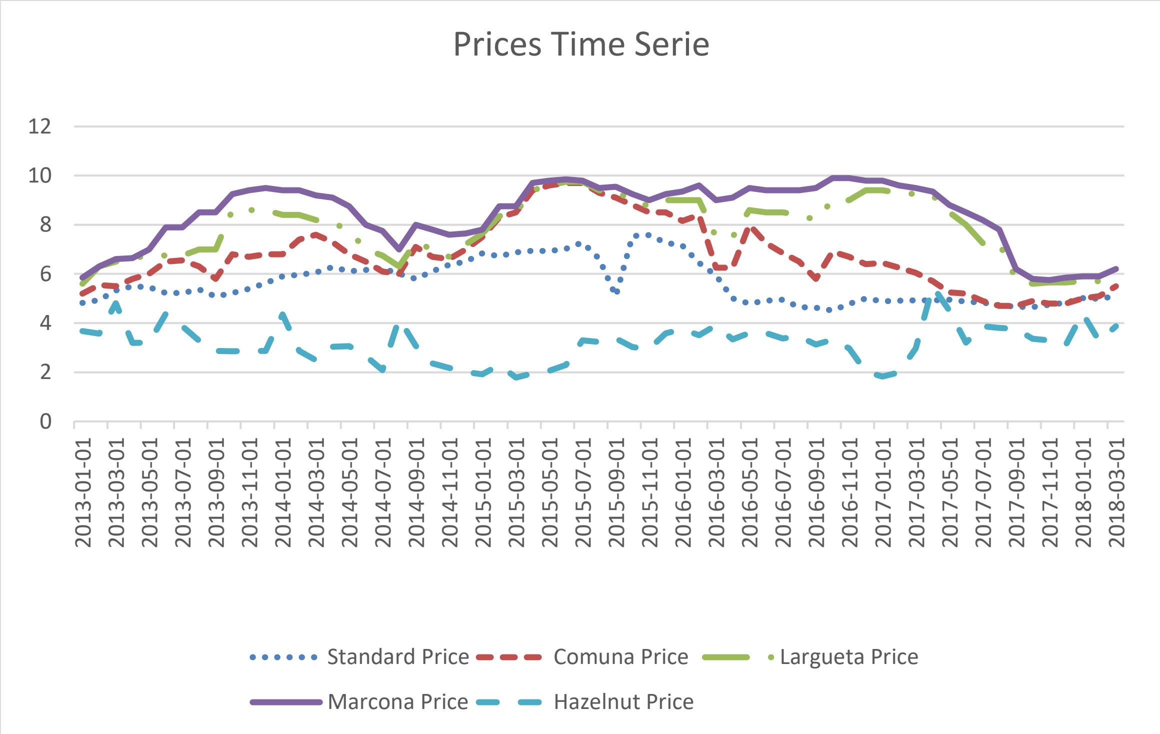


Figure 3.1: Prices Time Series

-Production: Historical series of production from 2013 to 2018. This part contains the data of Marcona, Comuna and Largueta, the three most representative varieties nationwide.

The prices are published every month, and the production is expressed in metric tons. The seasonality of the almond (only produces once a year) makes the analysis of production more difficult to perform.

Row ID	D Comuna	D Largueta	D Marcona
2013-01-01	1,096.667	442.222	247.778
2013-02-01	1,070	454.444	266.667
2013-03-01	928.889	437.778	143.333
2013-04-01	452.5	271.25	227.5
2013-05-01	460	141.176	216.471
2013-06-01	570	121.429	181.429
2013-07-01	366.154	155.385	138.462
2013-08-01	571.25	200	81.25
2013-09-01	2,831.25	168.75	570
2013-10-01	2,342.5	236.25	553.75
2013-11-01	1,430	428.75	435
2013-12-01	860	228.75	161.25
2014-01-01	780	542.5	197.5
2014-02-01	685	268.75	107.5
2014-03-01	643.75	143.75	111.25
2014-04-01	373.75	43.75	148.75
2014-05-01	261.25	153.75	82.5
2014-06-01	306.25	142.5	105
2014-07-01	413.75	135	96.25
2014-08-01	1,251.111	180	180

Table 3.4: National Production

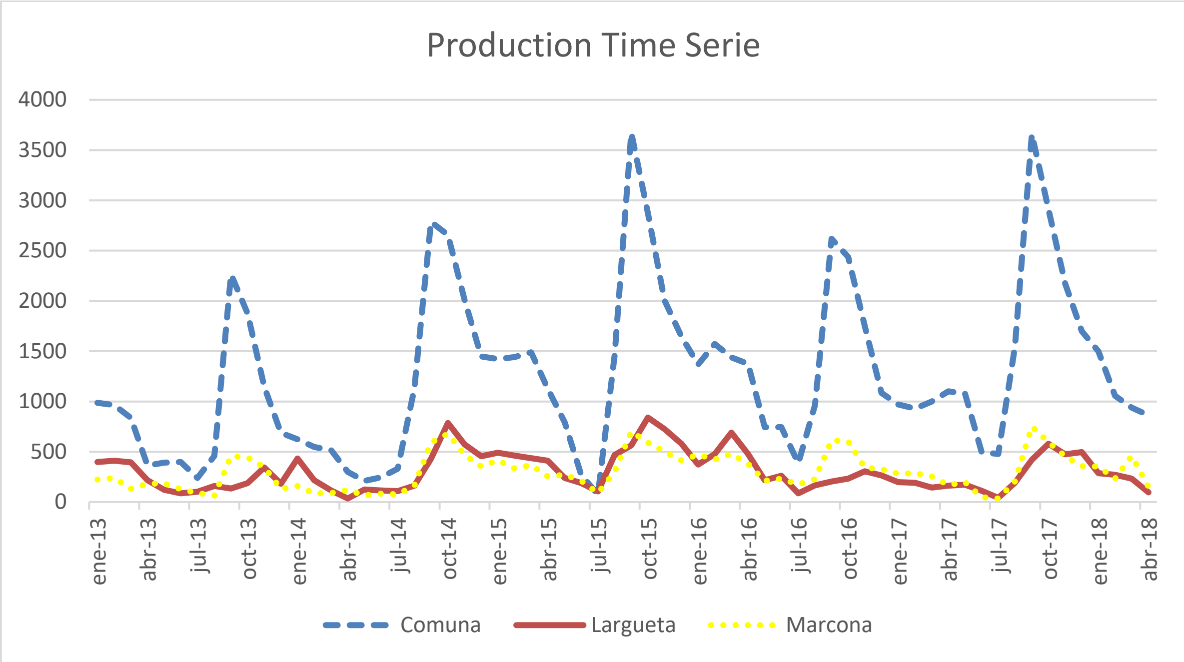


Figure 3.2: National Production Time Series

In the upper graph (Figure 3.2) clearly shows the seasonality mentioned, the production time is during the harvest. As we previously mentioned, Comuna has the highest production index, while Marcona and Largueta move onto quite similar stripes.

3.2 Data analysis.

This part focuses on three key points:

1- Firstly, to carry out an individual analysis about the time series of the two varieties that we want to study (Comuna and Marcona).

2- Secondly, we break down the production series into seasonality and noise, and compare them with the prices in terms of economic theory.

3- And finally, in order to create the models in the next point we define the correlations with the other variables of the described dataset.

3.2.1 Individual analysis

We observe the price charts of both varieties, as well as draw their trend and see the seasonality and noise of the data over time to analyse and understand the data series.

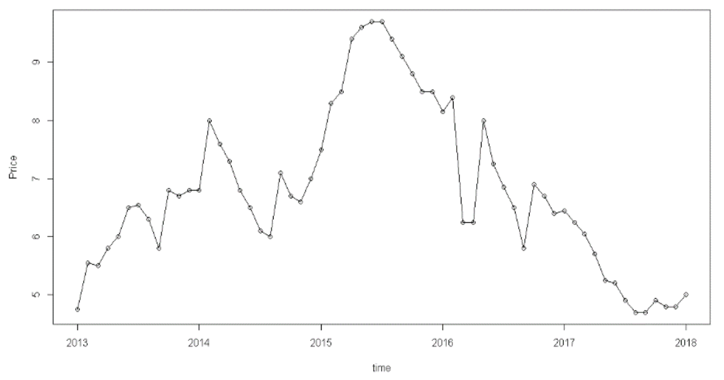
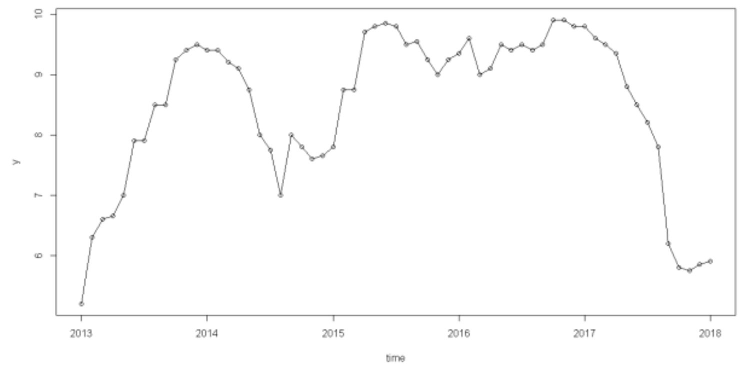
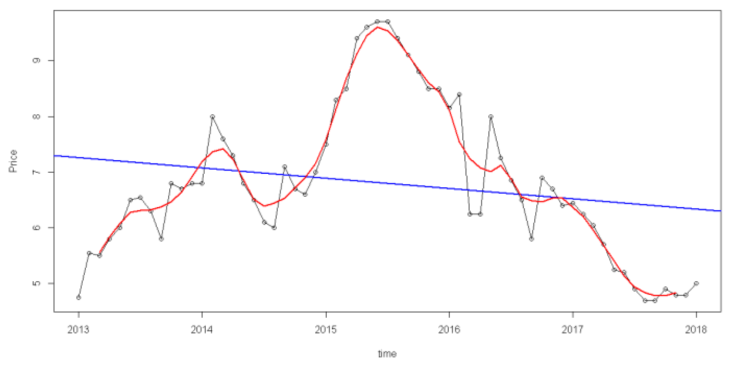
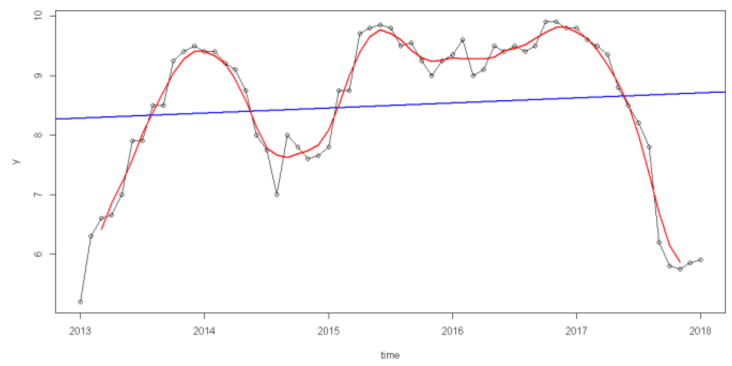
Comuna	*Marcona*
	
	

Table 3.5: Prices Trend

The graphs illustrated in the table 3.5, show us the time series as well as its tendency and linearity of the Marcona and Comuna varieties.

In the following two graphics in the table 3.6, the series are broken down into: observations, tendency, stationary and noise.

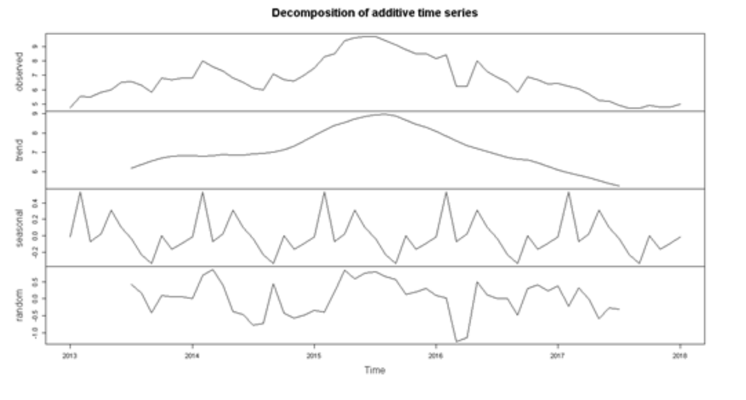
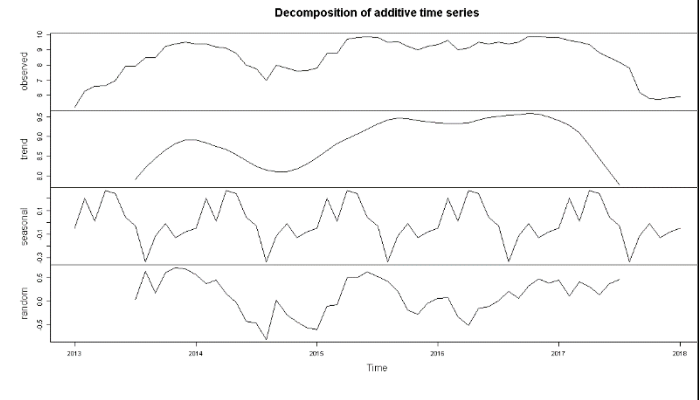
Comuna	Marcona
	

Table 3.6: Prices Noisy and Seasonality

To conclude the analysis, we resort to the Holt Winter method. The Holt-Winters method, is one of the many methods or algorithms that can be used to forecast data points in a series, provided that the series is "seasonal", i.e. repetitive over some period, to perform a simple regression and see how it behaves. This method uses the mean, the trend and the seasonality to carry out regressions using the price as a parameter.

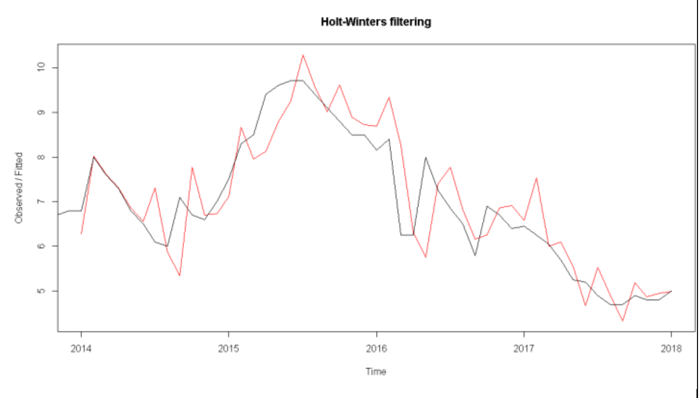
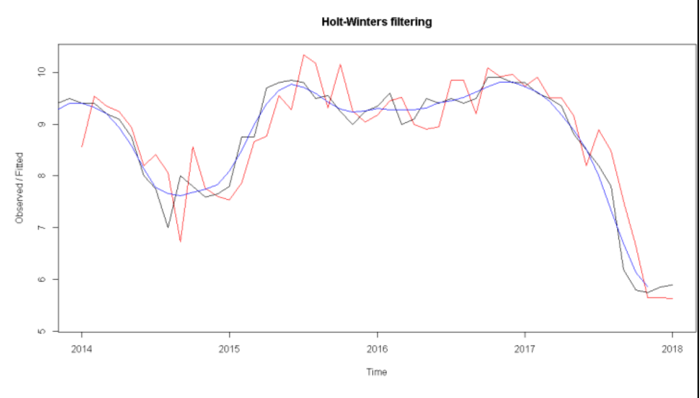
Comuna	Marcona
	

Table 3.7: Holt Winter Regression.

The conclusion of this analysis is that the price trend is different for the two varieties (Table 3.5). While it seems that the prices of the Comuna tend to fall, the Marcona shows a positive sign.

The production series shows a clear seasonality. On the other hand, the prices (see Table 3.6) express seasonality although it is not so visible since it also contains quite unstable noise.

The Holt Winters regression (Table 3.7) manages to approximate the reality, however the trend rarely corresponds to reality. It should be added that the errors seem quite high, so it is convenient to look for adjustment alternatives.

3.2.2 Production analysis

We analyse the production to observe the effect on the price. The seasonality, we previously mentioned, makes that the peaks and downs of production do not coincide with those of the prices. Therefore, it has been decided to use the general trend of the time series of price and production, to observe if the trends coincide with the economic theory.

The economic theory tells us that an increase in production should have a negative effect on the price, since it increases the supply and the market is regulated with a price depression. However we must not forget the demand in the equation, since an increase in demand would increase the price of the product. Although we do not have empirical data on consumption, a research have been made among different companies in order to know their opinions and market vision. The data obtained is confidential so we do not have the permission to publish.

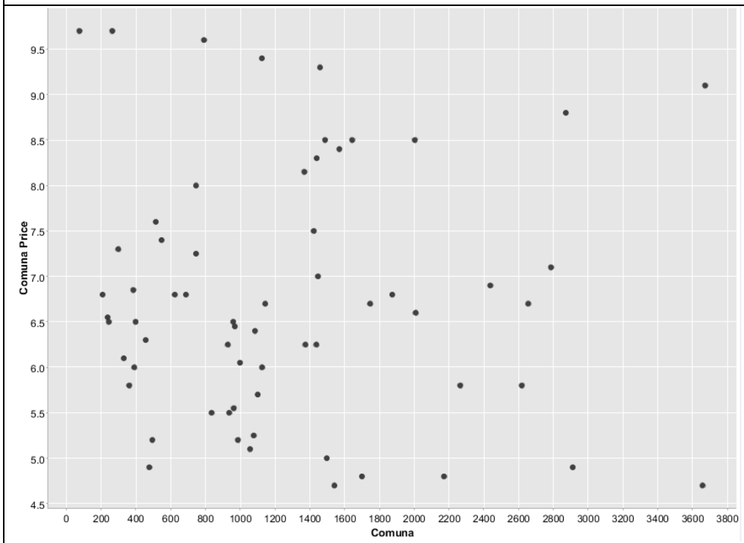
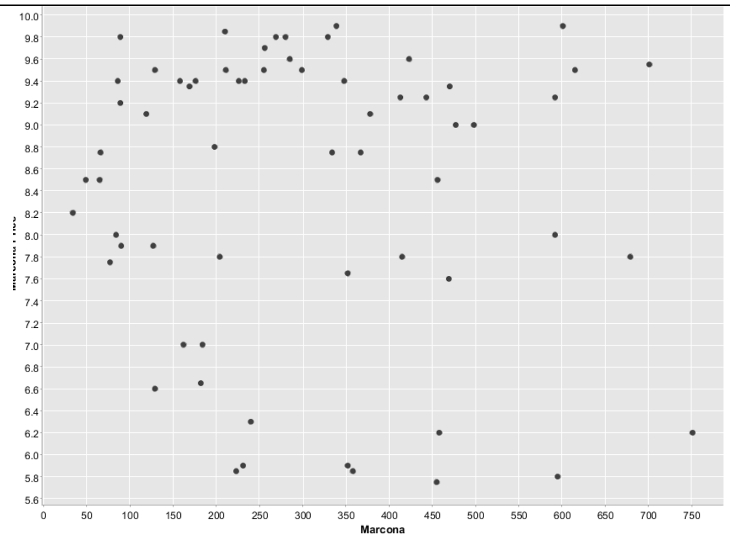
<table>
<tr><th>Comuna</th><th>Marcona</th></tr>
<tr><td>
</td><td>
</td></tr>
</table>

Table 3.8: Scatter plots whit prices in y axis and production in x (axis)

Table 3.8 shows scatter plots, but dispersed and none of the two varieties show a clear pattern between price and production. This is due to the periodicity of the production, only performed during the harvest while the prices are maintained throughout the year.

The graphs included in the following table 3.9 compare the production and prices of the variables to analyse: in black, we have the real series, in red the trend, and in blue the slope expressed linearly.

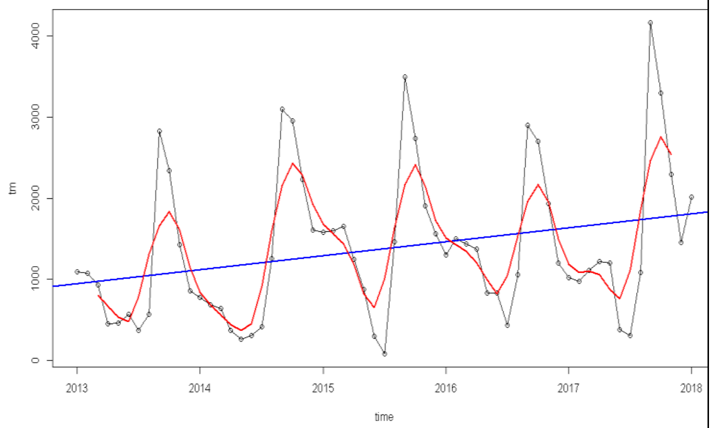
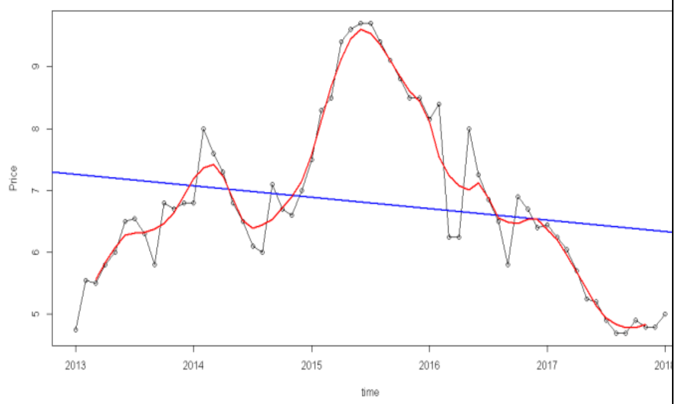
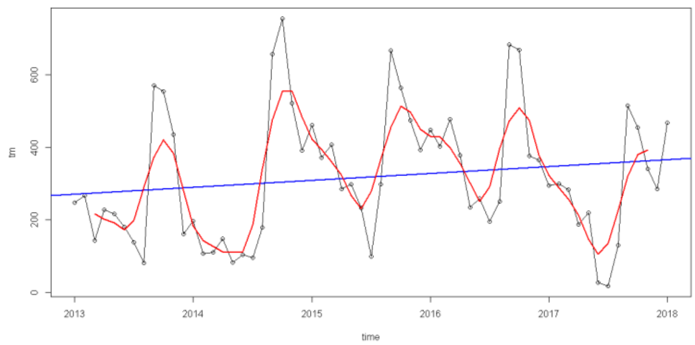
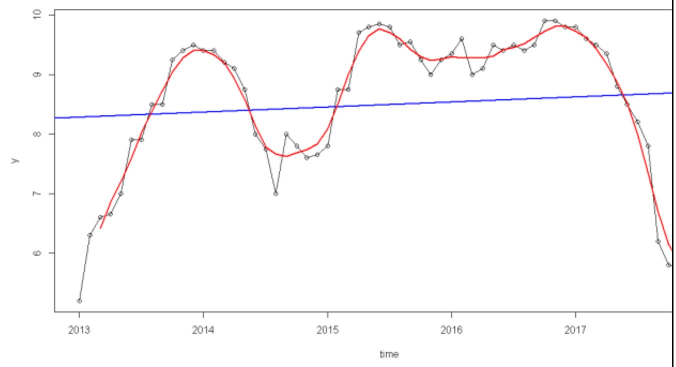
Variedad	Producción	Price
Comuna		
Marcona		

Table 3.9: Price and production trend

The first thing that can be highlighted from the table is the positive slope of the production in both graphs, which indicate that the sector is growing. In the interview conducted with the General Manager of Frutos Secos del Maestrazgo, SA highlighted that the amount of land exploited and the evolution of new production techniques in agriculture, have made that the production increase exponentially in recent years.

Regarding the price there are differences between varieties, while the price of the Comuna goes down, the Marcona has a positive tendency, although very weak.

To explain this phenomenon, the economic theory suggests that a greater increase in consumption could cause this effect. Although we do not have consumption data, reality seems to be different. Returning to the interview, when we asked about this topic, the answer was that there are many problems to quantify the real production of Marcona, since it is a variety that only is produced in Spain.

Marcona production index is low and the fact that it is not a perishable good, incentives to hide the product waiting for a higher price. Therefore the production data we collected probably do not correspond to the reality of each year.

Regarding to Comuna, the positive trend of its production is related in an expected way to its price. As this variety not only is produced in Spain the price speculation is more difficult.

3.2.3 Correlations with the explanatory variables.

Correlation is used to test relationships between quantitative variables or categorical variables. In other words, it is a measure of how things are related. The study of how variables are correlated is called correlation analysis.

-<u>Correlations with Comuna almond</u>.

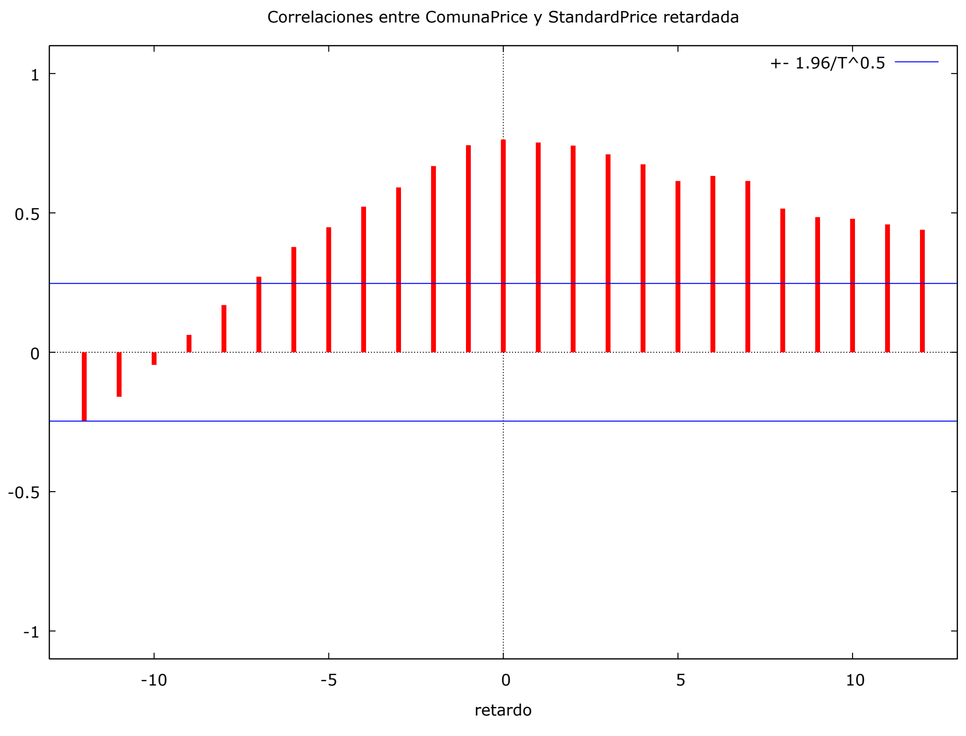


Figure 3.3:Comuna and Standard correlation

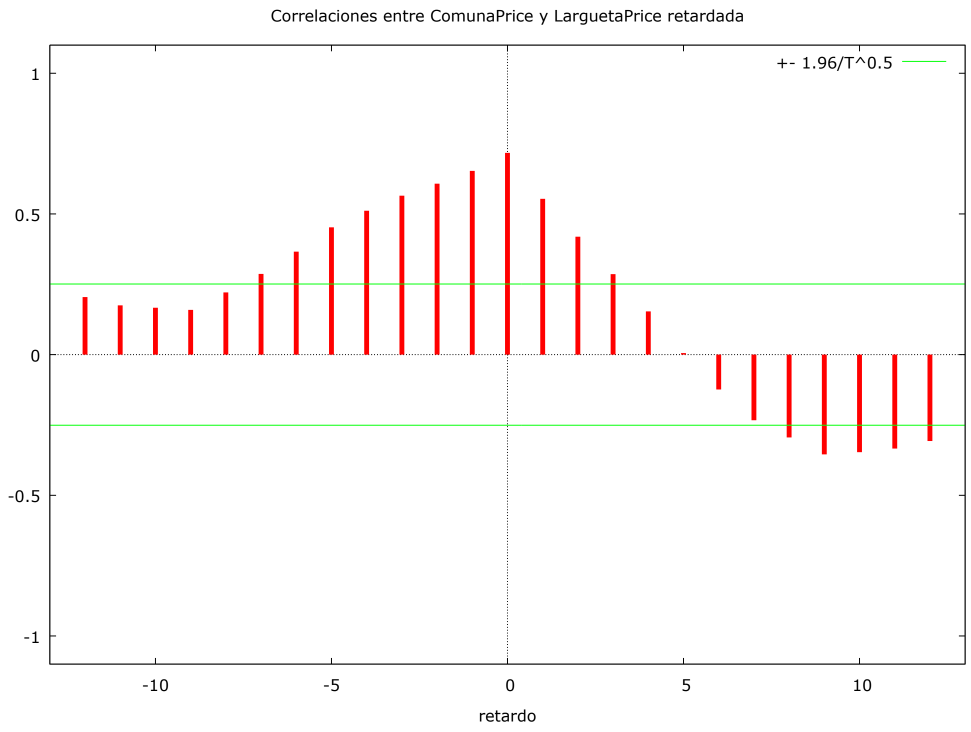


Figure 3.4: Comuna and Largueta correlation

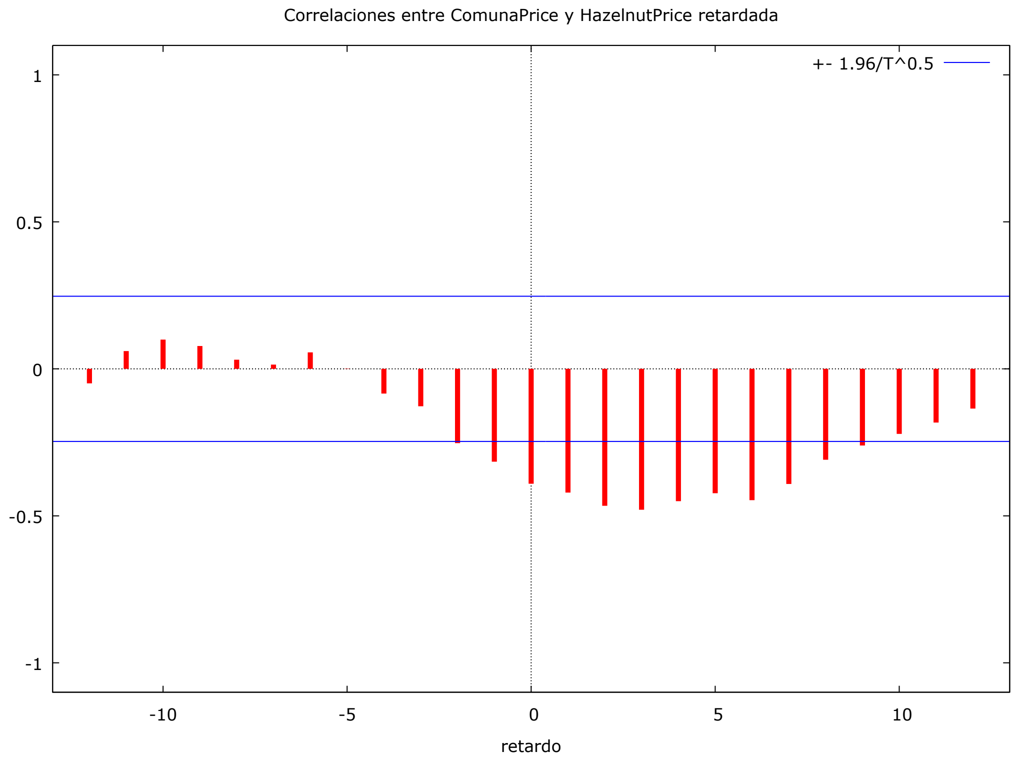


Figure 3.5: Comuna and hazelnut correlation

This study of the variables shows us that the Comuna is positively correlated in a robust way with the other national variety (Largueta) and with the variation in international prices (Standard).

On the other hand it expresses a clear negative correlation with the price of hazelnuts, although weaker than with almond varieties. We do not know the exact reason for this negative correlation. While investigating we found that the harvest season starts approximately one month after that of the almond season, as well as a natural substitute for the almond since both have similar properties. That is why we expected a correlation between both prices, although this is negative.

-<u>Correlations with Marcona almonds.</u>

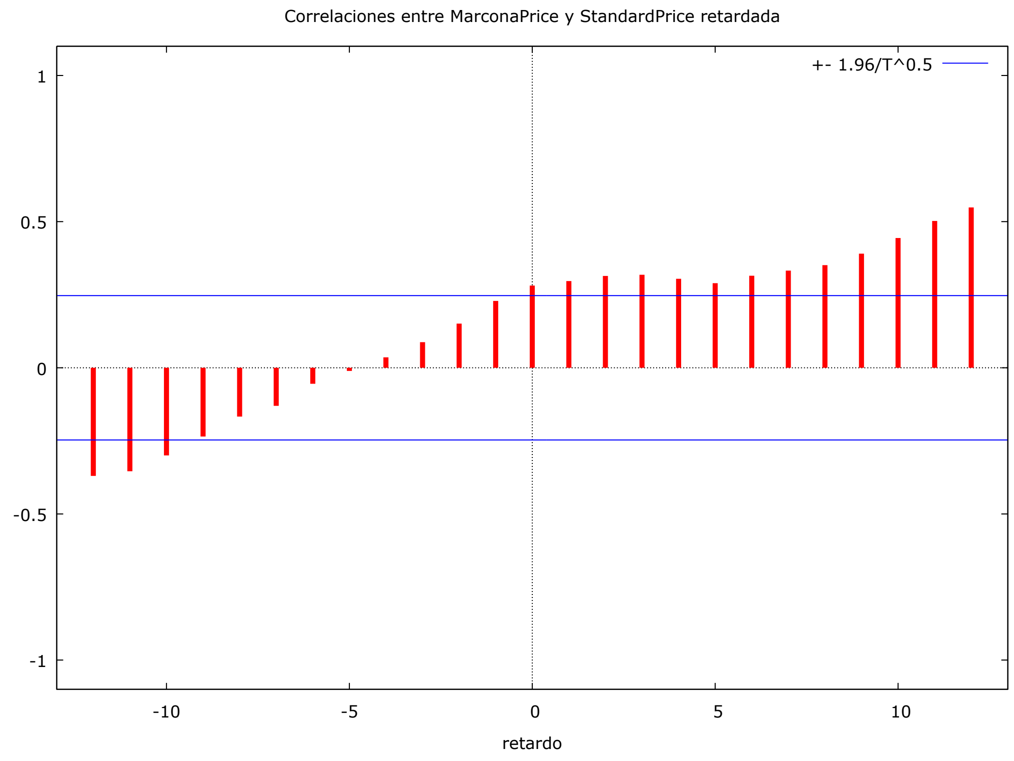


Figure 3.6: Marcona and Standard correlation

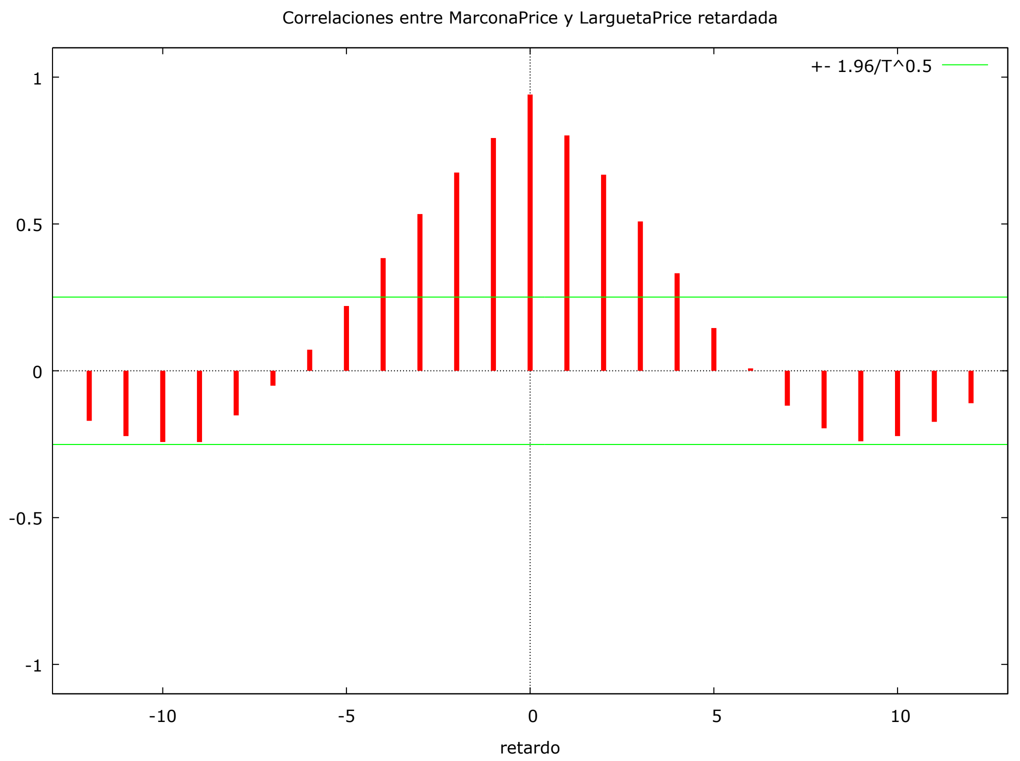


Figure 3.7: Marcona and Largueta correlation

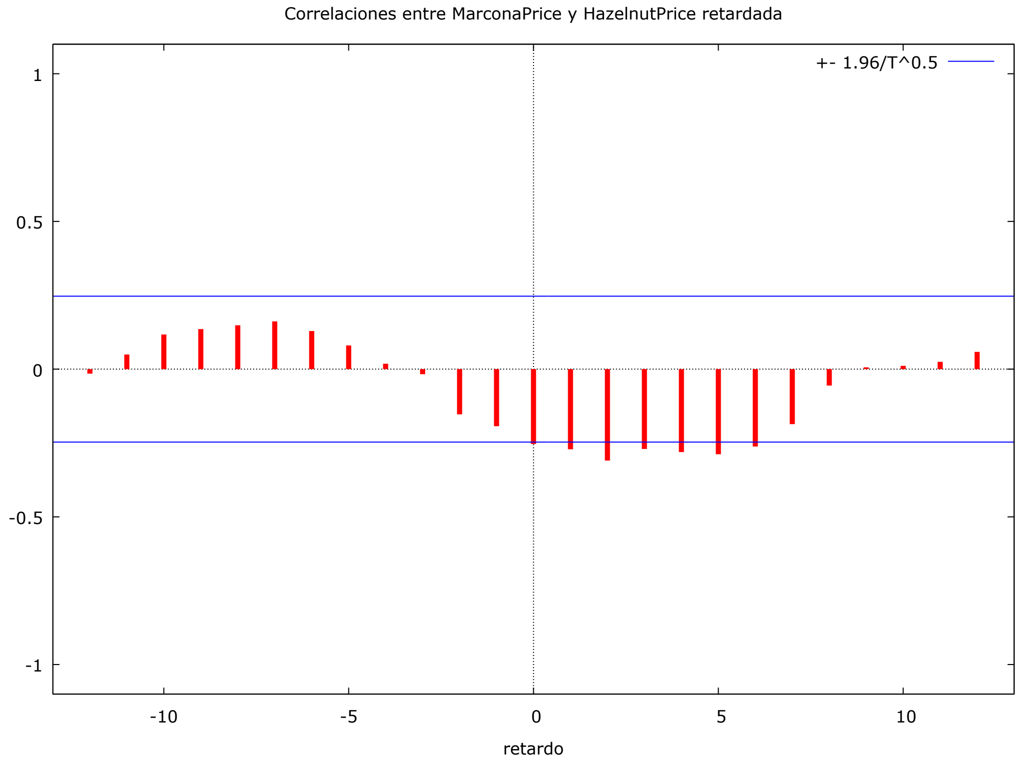
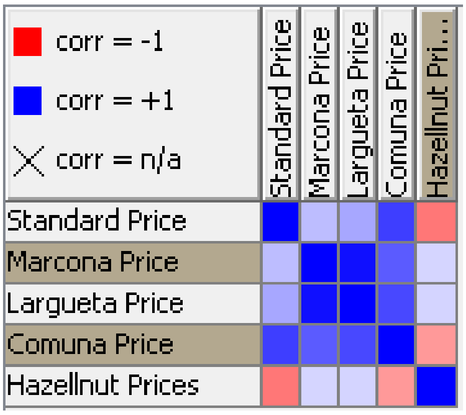


Figure 3.8: Marcona and hazelnut correlation

Unlike the Comuna variety, the Marcona only shows a strong correlation with Largueta, and a much weaker correlation with international prices. Which makes us wonder why.

“Marcona is a variety with little production, which is only produced in Spain, as the product does not expire in the short term, small growers can store the merchandise for longer periods of time waiting for a better price.”

Balma Boix (CEO, Frutos Secos del Maestrazgo SA)



This quote tells us about a variety with speculative power that does not respond effectively to market movements. What qualifies to our correlations and explains why the variations of its price do not move with the international market.

-The correlalogram shows how effectively the international variations and the price of hazelnuts have a minor impact on the largueta and marcona.

Figure 3.9: Correlogram between prices

It is also significant to see that only a strong correlation is observed between them, so we assume that both varieties follow a similar speculative pattern.

4. In Sample Validation.

4.1 The model

For the study, we will try to make two models that estimate the price of the Marcona and Comuna varieties. We have focused on two types of econometric models, the classic OLS models and those of autoregressive time series.

We make a comparison between both models to analyse which criterion best fits the real ones. As well as if the same model is better for both varieties.

4.2 The OLS Model.

In statistics, ordinary least squares (OLS) or linear least squares is a method for estimating the unknown parameters in a linear regression model. OLS chooses the parameters of a linear function of a set of explanatory variables by minimizing the sum of the squares of the differences between the observed dependent variable (values of the variable being predicted) in the given dataset and those predicted by the linear function. Geometrically, this is seen as the sum of the squared distances, parallel to the axis of the dependent variable, between each data point in the set and the corresponding point on the regression line – the smaller the differences, the better the model fits the data. The resulting estimator can be expressed by a simple formula, especially in the case of a single regressor on the right-hand side.

4.2.1 Comuna OLS

After doing several tests and trying different models, we have chosen with the following expression for the Comuna.

$$Y = \beta 0 + \beta 1SP + \beta 2LP + \beta 3HP + \beta 4SP^2 + \beta 5LP^2 + \beta 6tt + \beta 7Y_{t-1}$$

Where Y represents the price of the Comuna, PS, PL and PH the prices of Standard, Largueta and hazelnuts respectively and tt correspond to seasonal tendency. We also include the Y variable itself with a delay (Yt-1).

Modelo 16: MCO, usando las observaciones 2013:02-2018:03 (T = 62)
Variable dependiente: ComunaPrice

	Coeficiente	*Desv. Típica*	*Estadístico t*	*valor p*	
const	6.38978	3.05016	2.095	0.0409	**
StandardPrice	0.365899	0.121710	3.006	0.0040	***
LarguetaPrice	−1.81768	0.760756	−2.389	0.0204	**

HazelnutPrice	0.651818	0.480941	1.355	0.1810	
sq_LarguetaPrice	0.139519	0.0503664	2.770	0.0077	***
sq_HazelnutPrice	−0.111582	0.0710771	−1.570	0.1223	
time	−0.0113560	0.00403798	−2.812	0.0068	***
ComunaPrice_1	0.481847	0.102203	4.715	<0.0001	***

Media de la vble. Dep.	6.771774	D.T. de la vble. Dep.	1.368247
Suma de cuad. Residuos	11.01683	D.T. de la regresión	0.451681
R-cuadrado	0.903529	R-cuadrado corregido	0.891023
F(7, 54)	72.25037	Valor p (de F)	3.96e-25
Log-verosimilitud	−34.41516	Criterio de Akaike	84.83033
Criterio de Schwarz	101.8474	Crit. De Hannan-Quinn	91.51166
rho	0.209617	h de Durbin	2.780469

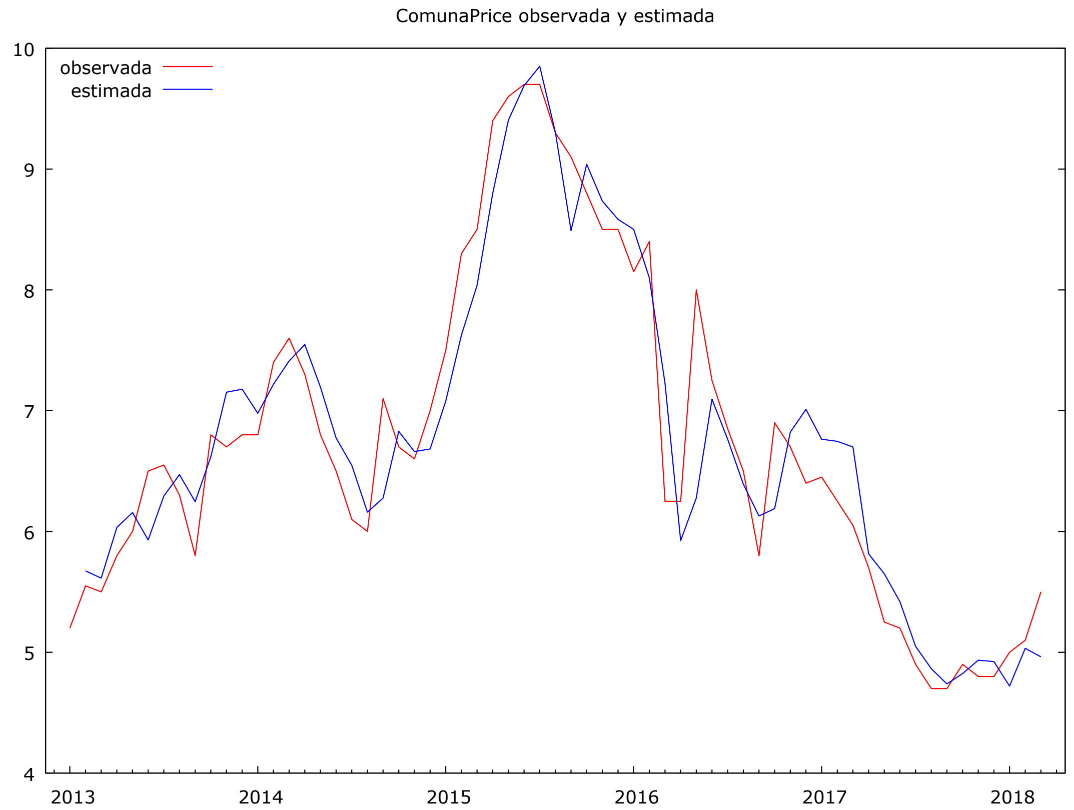


Figure 4.1: OLS Estimated and Observed for Comuna

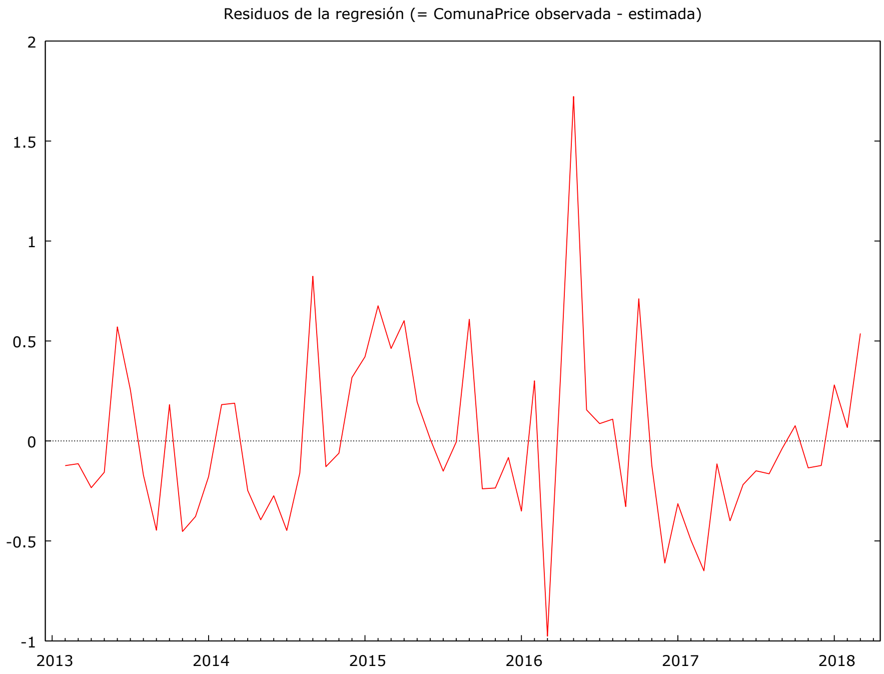


Figure 4.2: OLS Error picture for Comuna

-R2 is 0.9035 which indicates a pretty good fit with the variables.

-The sum of the squares of the remainder is 11.01683

- The graph of errors already suggests that there is no heterocedasticity problem, to check it a contrast is made Breusch Pagan Statistical contrast: LM = 11.128555, with value p = P (Chi-square (7)> 11.128555) = 0.133109. So we reject the heteroscedasticity hypothesis.

4.2.2 Marcona OLS.

As it is indicated above, both models use the same variables for our analysis to be consistent, so the only variation with respect to the model of the comuna will be the variable explained (Marcona Price) and the explanatory Yt-1.

Modelo 17: MCO, usando las observaciones 2013:02-2018:03 (T = 62)
Variable dependiente: MarconaPrice

	Coeficiente	*Desv. Típica*	*Estadístico t*	*valor p*	
const	−5.66939	1.78339	−3.179	0.0024	***
StandardPrice	−0.112901	0.0427680	−2.640	0.0108	**
LarguetaPrice	2.27801	0.475810	4.788	<0.0001	***
HazelnutPrice	0.328342	0.242286	1.355	0.1810	
sq_LarguetaPrice	−0.106313	0.0292989	−3.629	0.0006	***
sq_HazelnutPrice	−0.0541819	0.0358947	−1.509	0.1370	
time	−0.00505668	0.00217261	−2.327	0.0237	**
MarconaPrice_1	0.389126	0.0582627	6.679	<0.0001	***

Media de la vble. dep.	8.470161	D.T. de la vble. dep.	1.277884
Suma de cuad. residuos	3.101775	D.T. de la regresión	0.239667
R-cuadrado	0.968862	R-cuadrado corregido	0.964825
F(7, 54)	240.0270	Valor p (de F)	2.57e-38
Log-verosimilitud	4.875769	Criterio de Akaike	6.248461
Criterio de Schwarz	23.26554	Crit. de Hannan-Quinn	12.92979
rho	0.371783	h de Durbin	3.294567

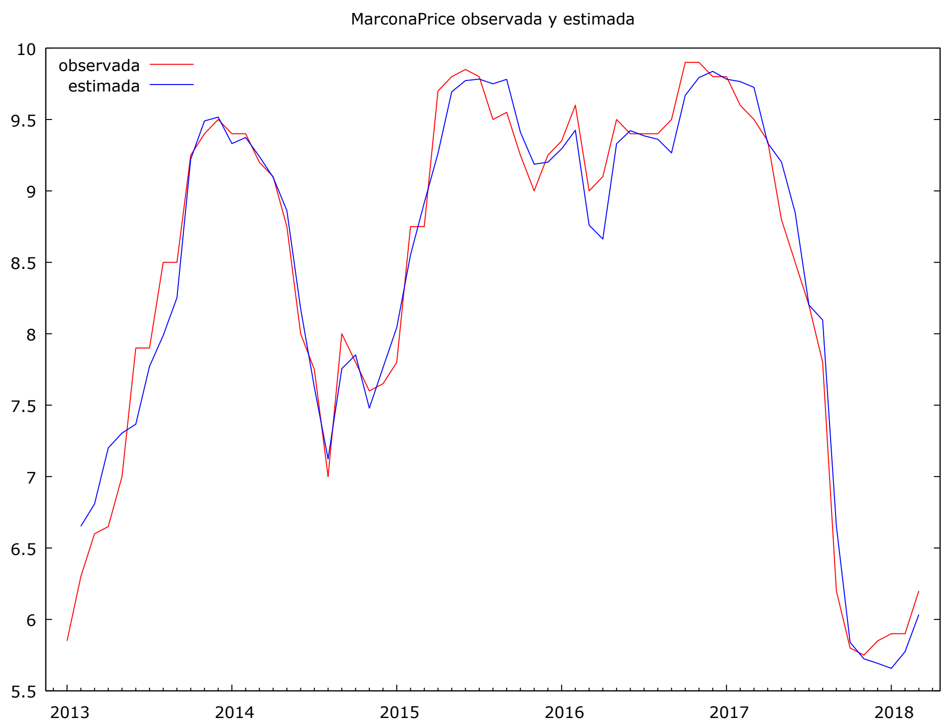


Figure 4.3: OLS Estimated and Observed for Marcona

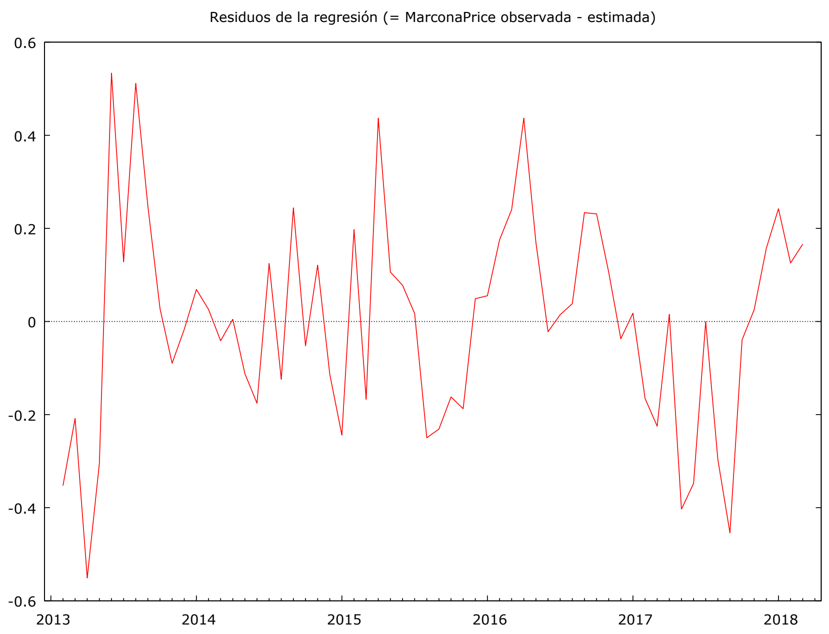


Figure 4.2: OLS Error picture for Marcona

-R2 is 0.968 which indicates a pretty good fit with the variables.

-The sum of the squares of the remainder is 3.101775

- The graph of errors already suggests that there is no heterocedasticity problem, to check it a contrast is made Breusch Pagan Statistical contrast: LM = 9.510096, with value p = P (Chi-square (7)> 9.510096) = 0.218076. So we reject the heteroscedasticity hypothesis

4.2.3 Comuna Vs Marcona using OLS.

Initially observing the statistics of the R2, we can see how the model explains better the data with Marcona than with Comuna 0.968 > 0.903. Also the residue is lower in the case of the Marcona, and the limits of the error graphics are smaller than in the case of the Comuna.

It is curious to observe that if we look at the significance level, the Comuna variety has more relevant variables than the Marcona. Both varieties have the same number of significant variables, but in the case of the Comuna the number of significant at 99% is greater than in the Marcona. It is also observed that for both models the hazelnuts variable is not significant at a relevant level.

Sig.	COMUNA	MARCONA
99%	SP,LP^2, tt, Yt-1	LP, LP^2, Yt-1
95%	LP	SP,tt
90%	-	-
-90%	HP,HP^2	HP,HP^2

Table 4.1: Significance OLS

This phenomenon can be explained due to the strong correlation between Marcona and Largueta that explains quantitatively a large part of the price variations in Marcona variety.

4.3 Autoregressive model

In statistics and signal processing, an autoregressive (AR) model is a representation of a type of random process; as such, it is used to describe certain time-varying processes in nature, economics, etc. The autoregressive model specifies that the output variable depends linearly on its own previous values and on a stochastic term (an imperfectly predictable term); thus the model is in the form of a stochastic difference equation.

For autoregressive models the same variables have been used as for OLS.

4.3.1 AR(1) Comuna

Modelo 3: Cochrane-Orcutt, usando las observaciones 2013:03-2018:03 (T = 61)
Variable dependiente: ComunaPrice
rho = 0.919297

	Coeficiente	*Desv. Típica*	*Estadístico t*	*valor p*	
const	11.8957	0.321565	36.99	<0.0001	***
StandardPrice	0.0639504	0.102854	0.6218	0.5368	
LarguetaPrice	−2.06945	0.852347	−2.428	0.0186	**
HazelnutPrice	0.594806	0.355889	1.671	0.1006	
sq_LarguetaPrice	0.186194	0.0546200	3.409	0.0013	***
sq_HazelnutPrice	−0.0872717	0.0484241	−1.802	0.0772	*
time	−0.0355407	0.0318613	−1.115	0.2697	
ComunaPrice_1	−0.0564771	0.0886707	−0.6369	0.5269	

Estadísticos basados en los datos rho-diferenciados:

Media de la vble. Dep.	6.791803	D.T. de la vble. Dep.	1.370407
Suma de cuad. Residuos	5.571086	D.T. de la regresión	0.324214
R-cuadrado	0.951495	R-cuadrado corregido	0.945089
F(7, 53)	15.64326	Valor p (de F)	6.34e-11
rho	0.143474	h de Durbin	1.553373

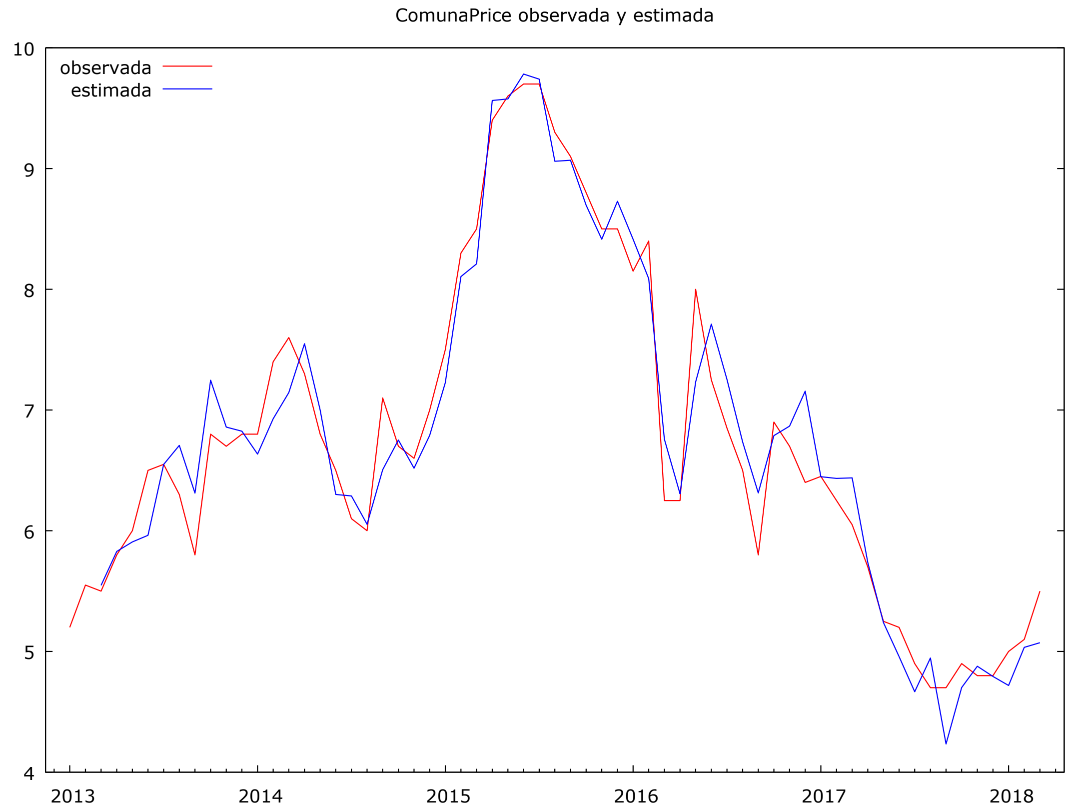


Figure 4.5: Autoregressive estimated and observed for Comuna

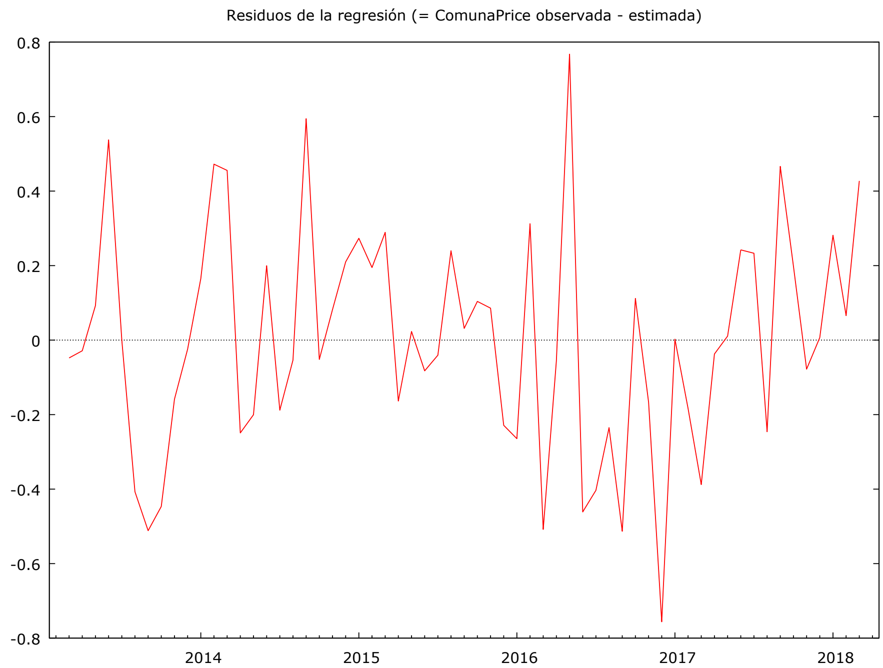


Figure 4.6: Autoregressive error picture for Comuna

-R2 is 0.951495 which indicates a pretty good fit with the variables.

-The sum of the squares of the remainder is 5.571086

4.3.2 AR(1) Marcona

Modelo 19: Cochrane-Orcutt, usando las observaciones 2013:03-2018:03 (T = 61)
Variable dependiente: MarconaPrice
rho = 0.620746

	Coeficiente	*Desv. Típica*	*Estadístico t*	*valor p*	
const	−4.67252	0.820287	−5.696	<0.0001	***
StandardPrice	−0.134206	0.0604406	−2.220	0.0307	**
LarguetaPrice	2.41461	0.543493	4.443	<0.0001	***
HazelnutPrice	0.136317	0.239037	0.5703	0.5709	
sq_LarguetaPrice	−0.108441	0.0347241	−3.123	0.0029	***
sq_HazelnutPrice	−0.0173167	0.0329435	−0.5256	0.6013	
time	−0.00860895	0.00463679	−1.857	0.0689	*
MarconaPrice_1	0.217541	0.0655295	3.320	0.0016	***

Estadísticos basados en los datos rho-diferenciados:

Media de la vble. dep.	8.505738	D.T. de la vble. dep.	1.257149
Suma de cuad. residuos	2.183807	D.T. de la regresión	0.202987
R-cuadrado	0.977003	R-cuadrado corregido	0.973966
F(7, 53)	63.57539	Valor p (de F)	1.60e-23
rho	−0.093239	h de Durbin	−0.847653

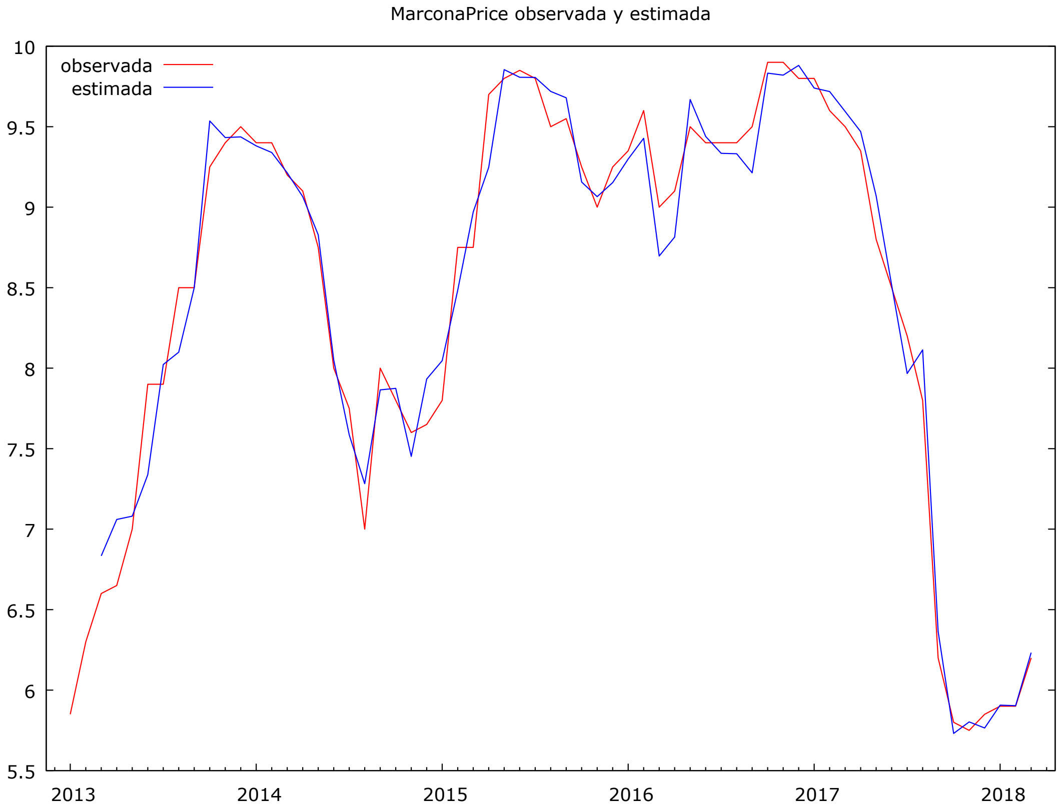


Figure 4.7: Autoregressive estimated and observed for Marcona

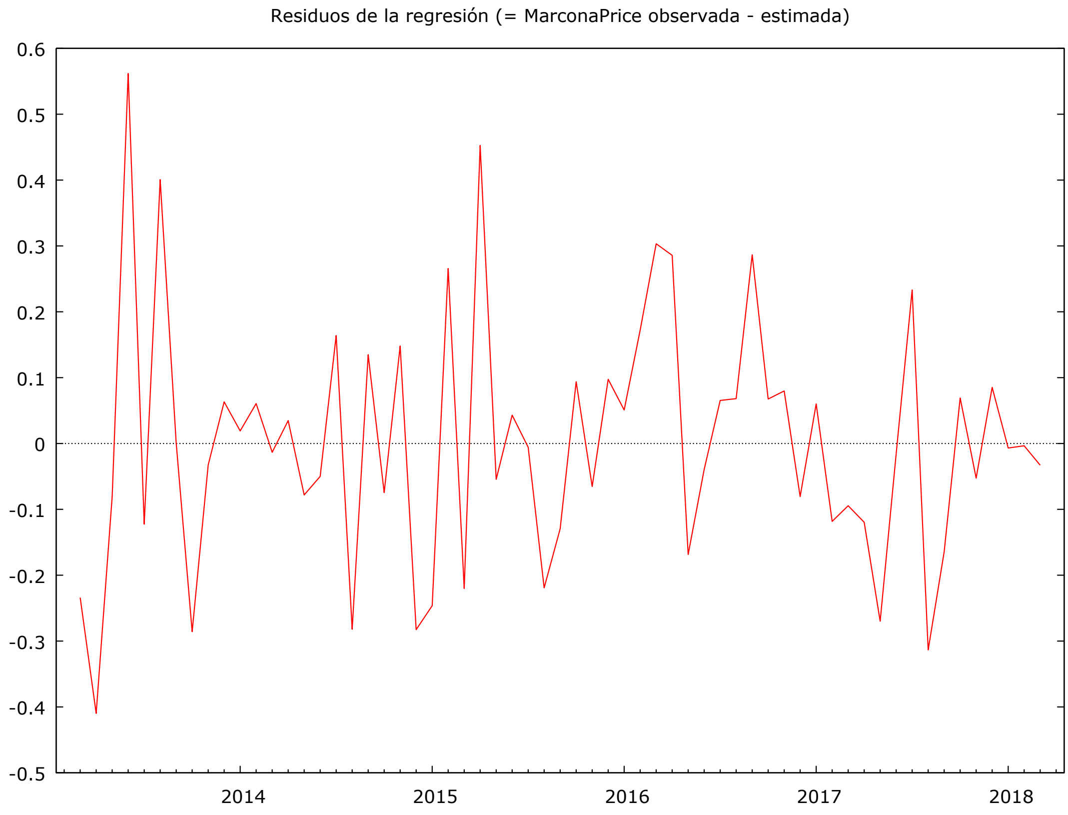


Figure 4.8: Autoregressive error picture for Marcona

-R2 is 0.977003 which indicates a pretty good fit with the variables.

-The sum of the squares of the remainder is 2.183807

4.3.3 Comuna Vs Marcona using autoregressive.

This case is better specified Marcona with R2 = 0.977 that the Comuna with R2 = 0.951.

Sig.	COMUNA	MARCONA
99%	LP^2	LP, LP^2, Yt-1
95%	LP	SP
90%	HP^2	tt
-90%	SP,HP,Yt-1,tt	HP,HP^2

Table 4.2: Autoregressive significance

In terms of significance, on this occasion marcona shows better levels than the comuna. The first correlates at a significant level with 5 of the 7 variables, while the comuna only does so with 3.

As with OLS, the errors are minor in the case of the marcona, it seems capable of making approximations closer to reality.

4.4 OLS Vs AR in sample.

Once the two models in sample validation have been calibrated, the observations indicate that the autoregressive models are better adjusted to the real time series.

	OLS		Autoregressive	
	Comuna	Marcona	Comuna	Marcona
R2	0.903	0.965	0.951	0.977
Medium error	1.4039e-015	-6.8242e-015	-1.2522e-015	1.7764e-015
Absolute medium error	0.31827	0.16555	0.23535	0.14319

Table 4.3: Models comparative

As we observe in the table 4.3, the autoregressive model have a higher R2 specification than the OLS models for both cases and reduces the error for the two models.

With the obtained results we can highlight two points:

- First, the analysis in sample adjusts with more precision the model in the Marcona variety.
- And secondly in the comparison of models, the Autoregressive are more efficient than the OLS in all cases for the problem we are facing.

These results are explained by the fact that the OLS models responds to a linear model while the autoregressive model specifies that the output variable depends linearly on its own previous values and on a stochastic term, so it generally explains time series better than the OLS models.

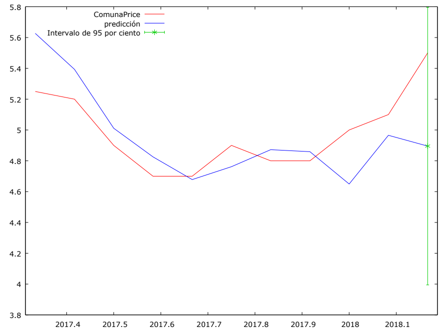
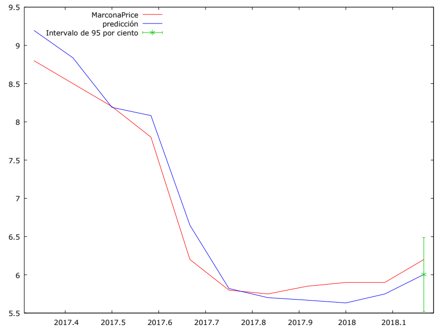
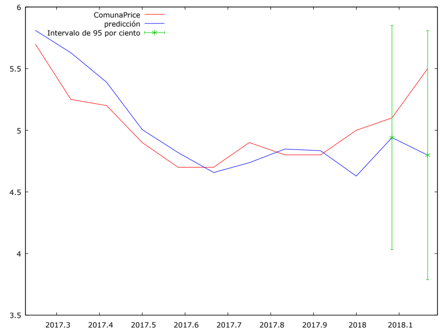
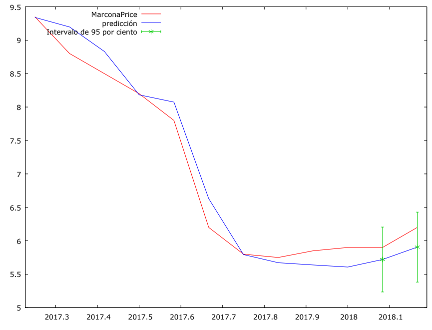
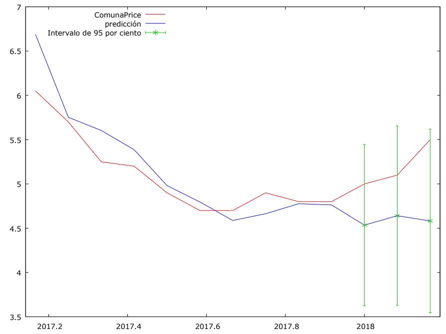
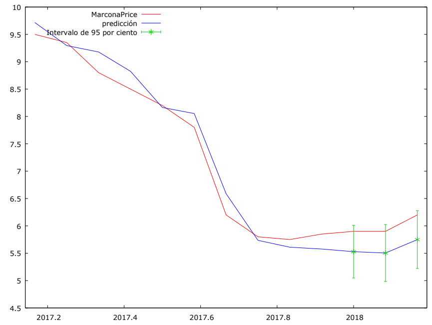
It would be necessary to know the real deviations in sample, to reproduce the time series with the same standard deviation (Bootstrap) a fixed number of times and to apply the models to them.

5. Out sample calibration.

This part of the project aims to create off-sample predictions using the two models described above, to see which is more reliable and if it is the same for the two varieties studied.

Different tests are carried out with the existing models and with different samples that range from t-1 (January 2018) to t-5 (November 2017 to March 2018).

5.1 Out sample using OLS

	Comuna		Marcona	
p	Prediction	Error	Prediction	Error
1		0.6045		0.1954
2		0.43056		0.23727
3		0.61397		0.40615

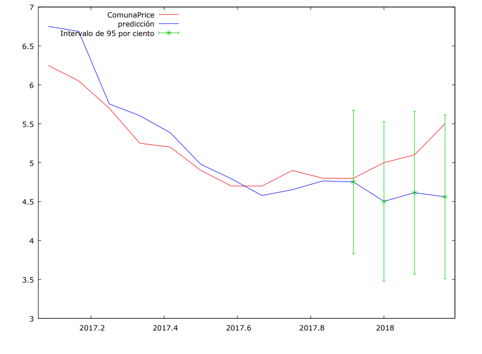
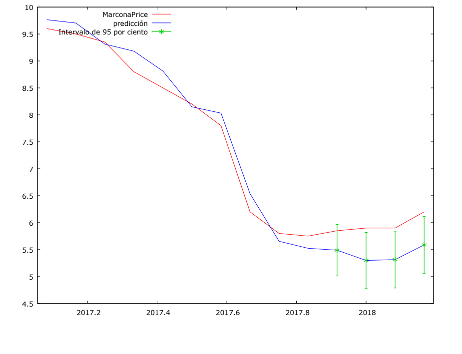
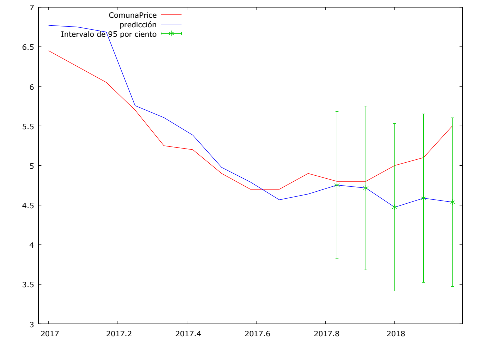
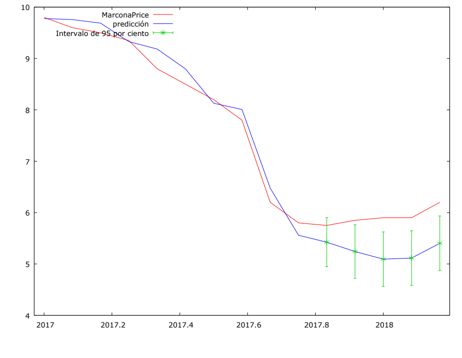
4		0.49233		0.5400
5		0.4264		0.6647

Table 5.1: OLS out sample estimated and observed

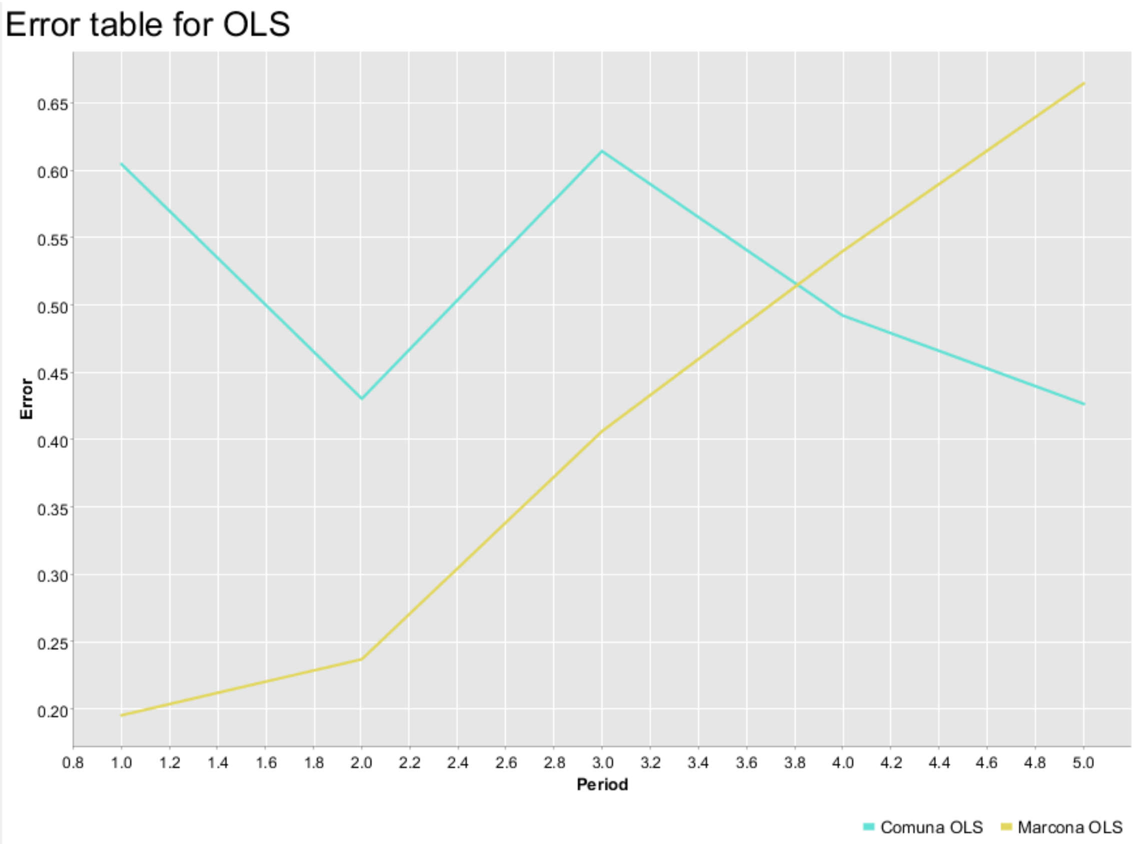


Figure 5.1: OLS Error

It is curious to observe how in the first period the estimator of the comuna responds with a much higher error than we expected at the beginning, 0.6045. However it is even more curious the random behaviour it shows when advancing in time. The error has been reduced for the second, in the third it increases and decreases again for the next two. The expected behaviour would be a low error in the first period that increases over time. But this case shows us a higher error in period one than for the 5 that follow 0.6045>0.4264

Periods	1	2	3	4	5
Error Comuna	0.6045	0.43056	0.61397	0.49233	0.4264
Error Marcona	0.1954	0.23727	0.40615	0.5400	0.6647

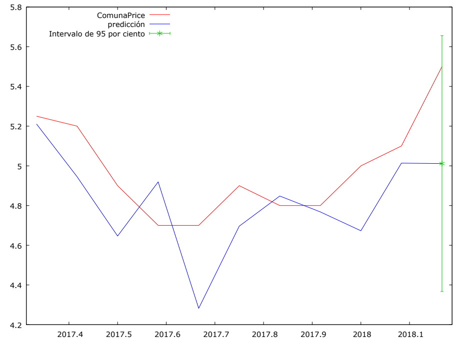
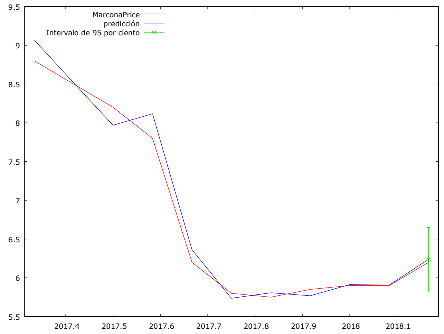
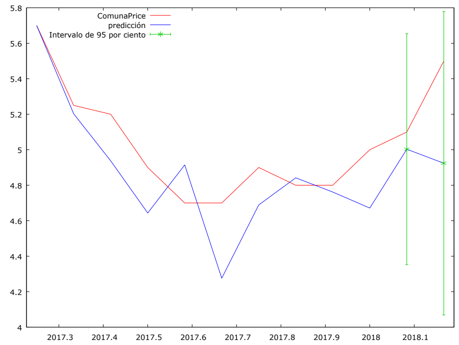
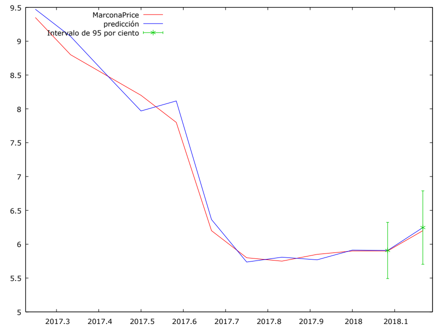
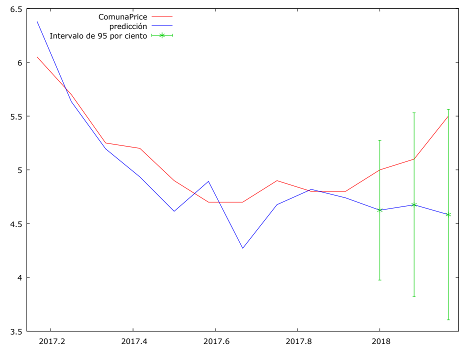
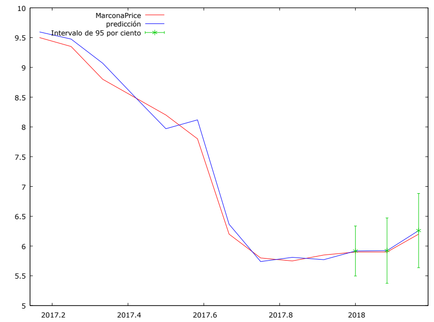
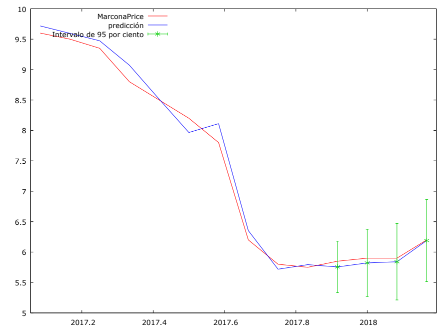
Table 5.2: Comparative OLS error table

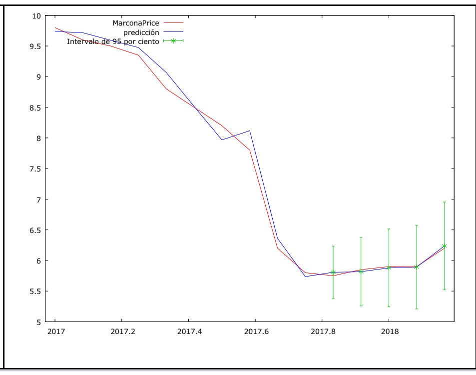
The errors of the Marcona out sample are more expected. There is a low error in the first period 0.19545 which increases with the passage of the next and reaches its maximum in the fifth with 0.6647.

Regarding the adjustment difference, the Marcona is better adjusted for all periods except the last one, so it seems that is more simple to forecast the behaviour of the Marcona time series.

5.2 Out sample using Autoregressive

Table 5.3: AR Out sample estimated and observed

p	Comuna		Marcona	
	Prediction	Error	Prediction	Error
1		0.48857		0.038794
2		0.33653		0.026747
3		0.57181		0.033323
4		0.48921		0.060678

5		0.38288		0.031432

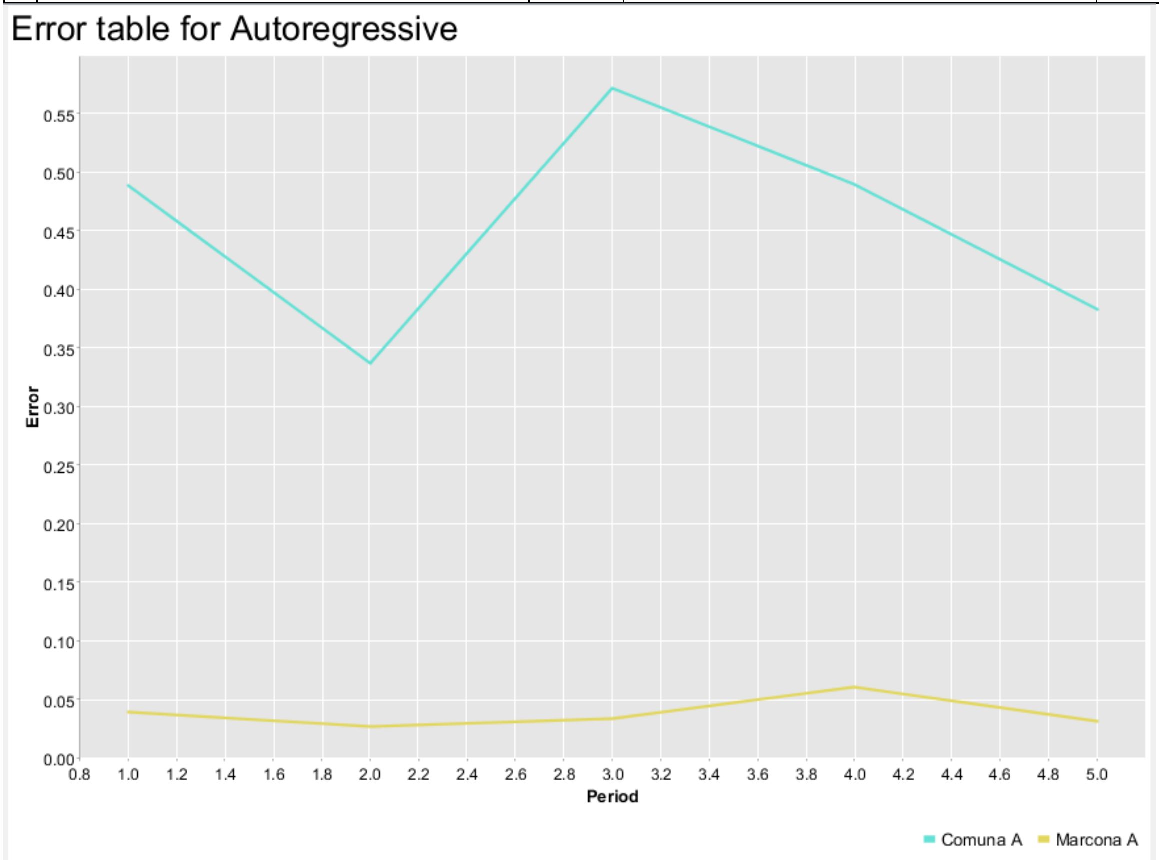


Figure 5.2: AR Error

As happened with the OLS model, the Comuna still shows a few quite random errors with respect to time. It follows the same previous pattern: a large error in the first period that reduces in the second, makes its maximum in the third and returns to go down for the last two. The error again is higher in the first period than the in last 0.48857>0.38288.

Periods	1	2	3	4	5
Error Comuna	0.4885	0.33653	0.57181	0.48921	0.3828
Error Marcona	0.0387	0.02674	0.03332	0.06067	0.0314

Table 5.4: Comparative AR error table

In the case of the Marcona the autoregressive has considerably adjusted to reality in all the periods. It starts with an error of 0.0387 and in any case in the whole series its error is higher than 0.06067, so it demonstrate that out sample has given some great results.

As for the comparison between varieties for the Autoregressive, the Marcona variety has adjusted much better than Comuna. Its errors are inferior for all cases and not even after 5 periods it explodes.

5.3 Out sample AR VS OLS

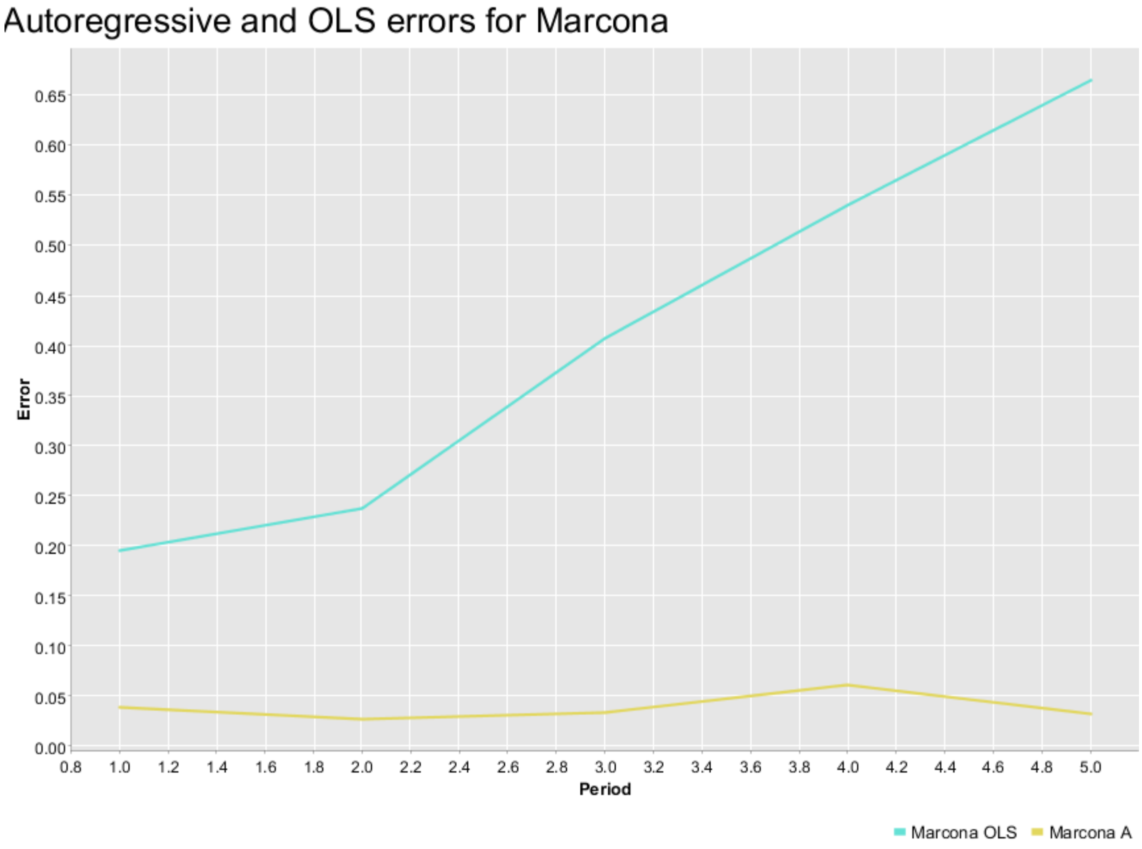


Figure 5.3: Errors AR and OLS for Marcona

Regarding the comparison of models, first we made the comparison between both models for Marcona. As shown in the upper graph, the errors are much lower in the Autoregressive than in the OLS model. Therefore, we come to the conclusion that with the same variables and for this time series the autoregressive works better.

Periods	1	2	3	4	5
Error OLS	0.1954	0.23727	0.40615	0.5400	0.6647
Error AR	0.0387	0.02674	0.03332	0.06067	0.0314

Table 5.4: Comparative OLS and AR (Marcona)

As the table shows, the OLS model has its minimum error in 0.1954, while the Autoregressive has its maximum in 0.6647.

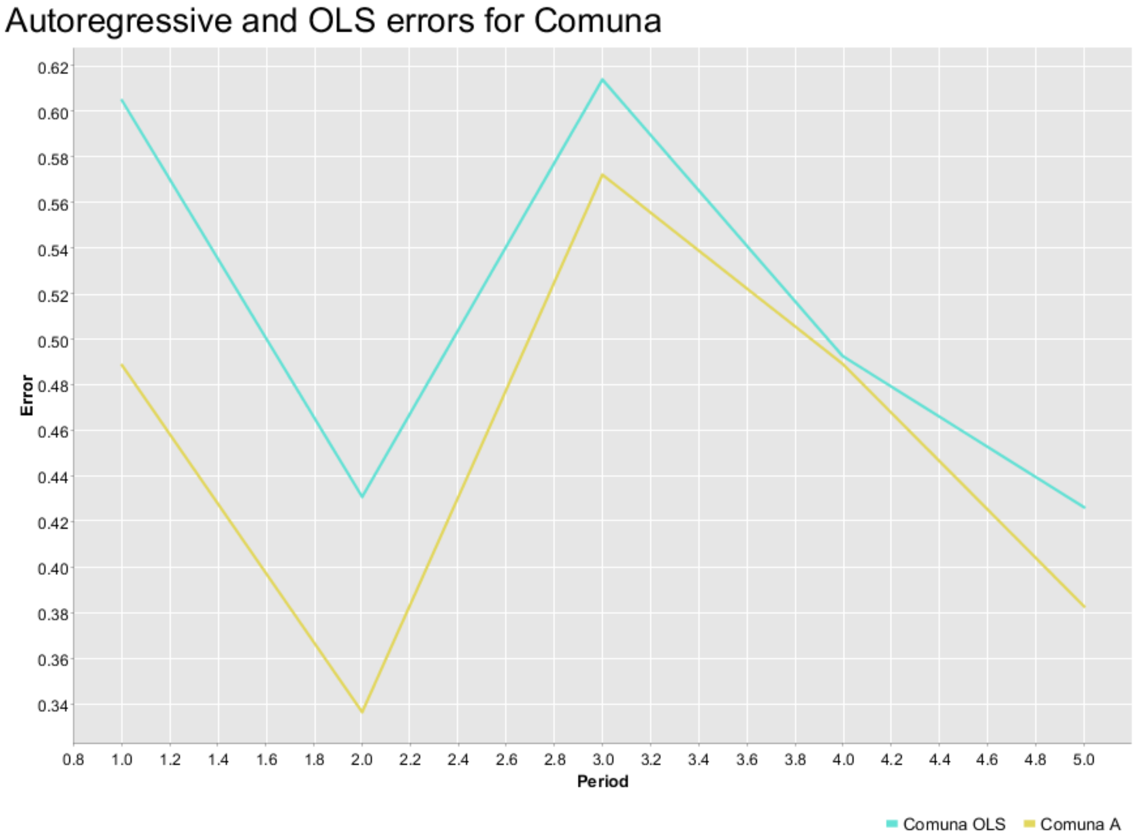


Figure 5.4: Errors AR and OLS for Comuna

As we have seen, the models for the Comuna out sample have had a strange behaviour within the expected. The error, far from increasing, is reduced in the last periods. This case can be explained simply with the sample. If the point of the sample has shown an unusual behaviour, for instance in a bubble shape, it can explain the behaviour of the predictions in our calibration.

Periods	1	2	3	4	5
Error OLS	0.6045	0.43056	0.61397	0.49233	0.4264
Error AR	0.4885	0.33653	0.57181	0.48921	0.3828

Table 5.4: Comparative OLS and AR (Comuna)

Both lines of error follow the same trend in the table 5.4 with almost identical slopes. Although in this case, by minimum error, we also conclude that the AR model fits the reality better than the OLS model. Although both contain a significant error, the AR is lower in all periods.

In conclusion, for both models the calibration out sample has shown that the AR models are more efficient approaching the real result. Both of them in Marcona and in Comuna the results have been better than those obtained by OLS.

6. Conclusions.

The conclusions drawn from this project have been:

In the creation and analysis of the dataset, we observe that the production and price trends agree with the economic theory in the case of the Commune, but not with the Marcona, which responds to an increase in production over time with a price increase.

The correlations show that both Comuna and Marcona are positively correlated with the other national variety (Largueta) and with the international price indicator (Standard). However its correlation with the price of hazelnuts is negative (very weak in the case of the Marcona).

From this part, taking into account the information obtained from the interview, we conclude that Marcona and Largueta are varieties with a high speculative power. As a consequence the production is very difficult to quantify and also get the real price.

In the calibration sample, the OLS model has had a better adjustment for the Marcona than for the Comuna with an R2 0.968> 0.903 respectively.

In the AR models this trend has been repeated. Marcona is closer to reality with a higher R2 and smaller errors ($R^2=0.977>0.951$).

Regarding the difference between in-sample models, the AR models show a lower number of significant variables in the case of the comuna, and reproduce the series with more fidelity than OLS.

The out sample calibration gives us the following conclusions:

For the OLS models, except for the test for the 5 periods, Comuna adjusts with a higher error range, which is the reason why Marcona offers the best results.

For the AR, we find a Marcona model that has adjusted surprisingly well, with some errors much lower than those adjusted in the other variety.

Finally in the comparison between models in out sample, the autoregressive models have shown lower errors for all the periods that the calibrated models of OLS. So we can conclude that the AR models are the bests to adjust this type of time series.

In future studies, it would be necessary to reproduce the time series with the same standard deviation a fixed number of times (Bootstrap) and to apply the models to them with the intention of extracting their average and real deviations from the model.

I would try to compare these models with self-raising techniques. With a large enough amount of data, they are able to do forecast with well-adjusted results. So it would be interesting to do the study.

Bibliography

-Hansen, L.P., Heckman, J.J., 1996. The empirical foundations of calibration. J. Econ. Perspect. 10.1, 87–104.

-Blomhoff, R., Carlsen, M. H., Andersen, L. F. and Jacobs, D. R. (2006) “Health benefits of nuts: potential role of antioxidants,” British Journal of Nutrition. Cambridge University Press, 96(S2), pp. S52–S60. doi: 10.1017/BJN20061864.

-Ros, E. and Mataix, J. (2006) “Fatty acid composition of nuts – implications for cardiovascular health,” British Journal of Nutrition. Cambridge University Press, 96(S2), pp. S29–S35. doi: 10.1017/BJN20061861.

-Recchioni, Maria Cristina, Tedeschi, Gabriele, Gallegati, Mauro, (2015)” A calibration procedure for analysing stock price dynamics in an agent-based framework,” Journal of Economic Dynamics and Control.

-Bollerslev, T. (1987). A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. The Review of Economics and Statistics, 69(3), 542-547. doi:10.2307/1925546

-Kendall, M., & Hill, A. (1953). The Analysis of Economic Time-Series-Part I: Prices. Journal of the Royal Statistical Society. Series A (General), 116(1), 11-34. doi:10.2307/2980947

- Spiller, GA, Jenkins, DJ, Cragen, LN, Gates, JE, Bosello, O, Berra, K, Rudd, C, Stevenson, M and Superko, R. (1992) Effect of a diet high in monounsaturated fat from almonds on plasma cholesterol and lipoproteins: Journal of the American College of Nutrition, 2 (11), 126-130.

- Zeeshan Ahmad. (2010) The uses and properties of almond oil: Complementary Therapies in Clinical Practice, 1 (16), 10-12.

- Paresh Kumar Narayan, Seema (2010) Modelling the impact of oil prices on Vietnam’s stock prices, 87 (1), 356-361.

- Paresh Kumar Narayan (2006) The behaviour of US stock prices: Evidence from a threshold autoregressive model, Mathematics and Computers in Simulation, 71 (2), 103-108.

Apendix.

This part, is outside the official study, but was part of the original idea, which I will be interested in continue developing in the future.

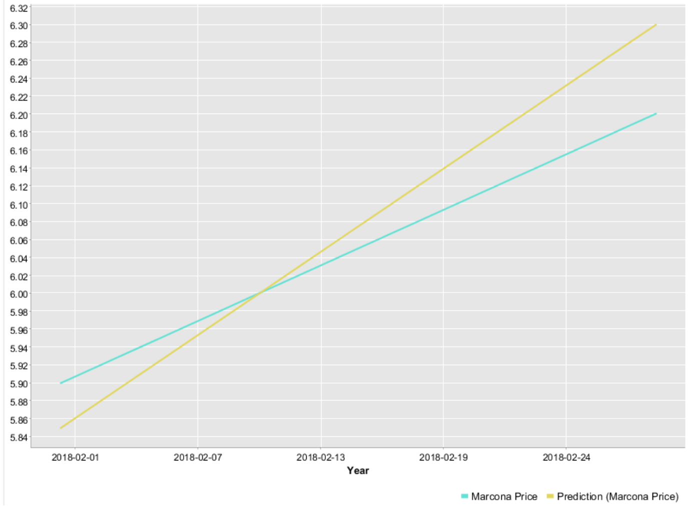
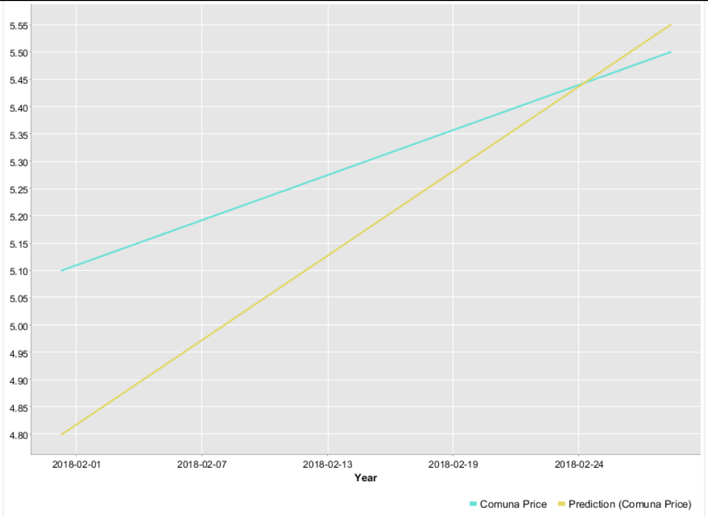
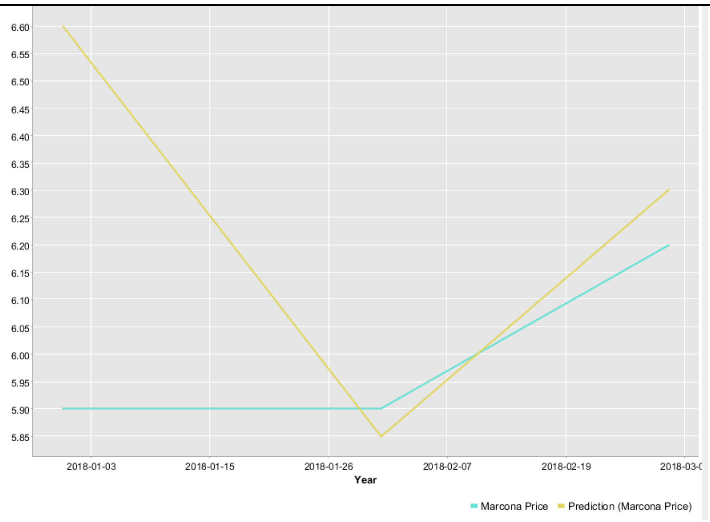
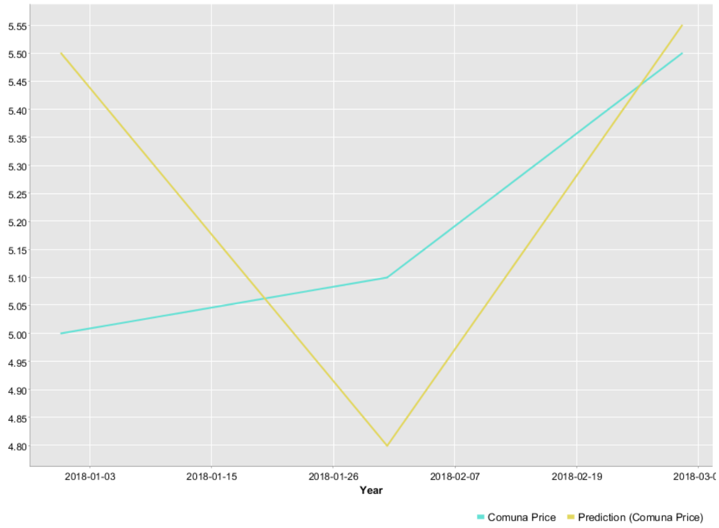
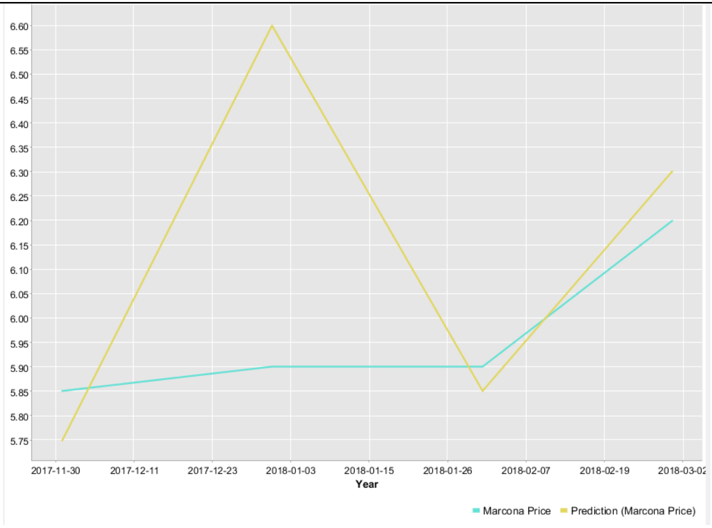
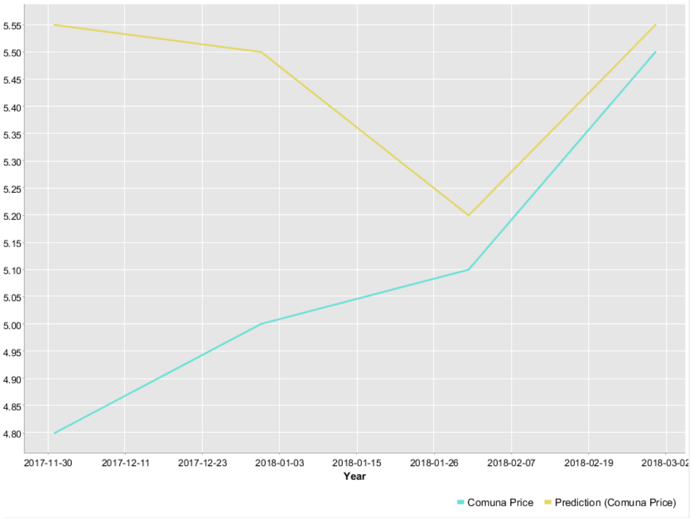
-Out sample calibration using Neuronal Net.

A neural network is a series of algorithms that attempts to identify underlying relationships in a set of data by using a process that mimics the way the human brain operates. Neural networks have the ability to adapt to changing input so the network produces the best possible result without the need to redesign the output criteria.

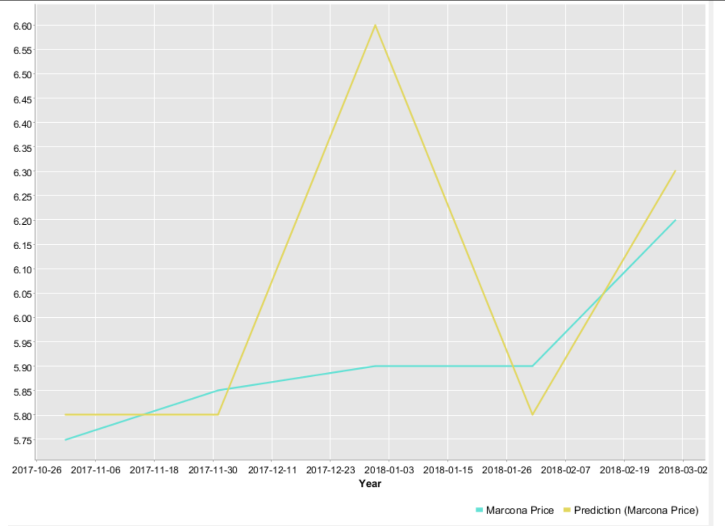
The original idea of the project was to compare between classic OLS models and Machine Learning models.

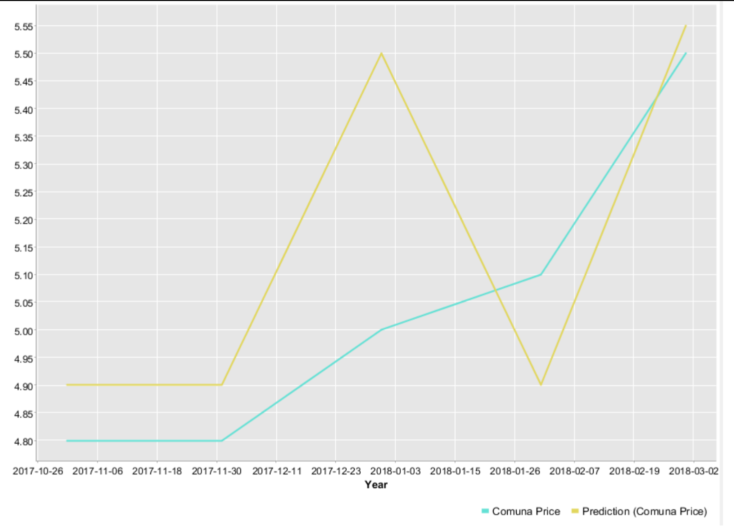
P	Marcona		Comuna	
1	R^2:	-∞	R^2:	-∞
	Mean absolute error:	0,1	Mean absolute error:	0,05
	Mean squared error:	0,01	Mean squared error:	0,002
	Root mean squared error:	0,1	Root mean squared error:	0,05
	Mean signed difference:	0,1	Mean signed difference:	0,05
2	R^2:	0,722	R^2:	-0,156
	Mean absolute error:	0,075	Mean absolute error:	0,175
	Mean squared error:	0,006	Mean squared error:	0,046
	Root mean squared error:	0,079	Root mean squared error:	0,215
	Mean signed difference:	0,025	Mean signed difference:	-0,125
3	R^2:	-7,375	R^2:	-1,446
	Mean absolute error:	0,283	Mean absolute error:	0,283
	Mean squared error:	0,167	Mean squared error:	0,114
	Root mean squared error:	0,409	Root mean squared error:	0,338
	Mean signed difference:	0,25	Mean signed difference:	0,083
4	R^2:	-5,667	R^2:	-2,173
	Mean absolute error:	0,237	Mean absolute error:	0,35
	Mean squared error:	0,128	Mean squared error:	0,206
	Root mean squared error:	0,358	Root mean squared error:	0,454
	Mean signed difference:	0,162	Mean signed difference:	0,35
5	R^2:	-3,558	R^2:	0,059
	Mean absolute error:	0,2	Mean absolute error:	0,19
	Mean squared error:	0,103	Mean squared error:	0,062
	Root mean squared error:	0,321	Root mean squared error:	0,25
	Mean signed difference:	0,14	Mean signed difference:	0,11

Apendix Table 1: R squared and errors with NN

P	Marcona	Comuna
2		
3		
4		

5





Apendix Table 2: Estimated and observed NN

Periods	1	2	3	4	5
Error OLS	0.1954	0.23727	0.40615	0.5400	0.6647
Error AR	0.0387	0.02674	0.03332	0.06067	0.0314
Error NN	0.1	0.075	0.283	0.237	0.2

Appendix Table 3: Marcona errors comparative

In my opinion there are few data to affirm that the NN is stable, but the calibrated error table shows that although it does not improve the results of the AR models, if they qualitatively exceed the OLS models for the Marcona.

Periods	1	2	3	4	5
Error OLS	0.6045	0.43056	0.61397	0.49233	0.4264
Error AR	0.4885	0.33653	0.57181	0.48921	0.3828
Error NN	0.05	0.175	0.283	0.35	0.19

Appendix Table 4: Comuna Errors comparative

Regarding the Comuna, the neuronal network improves the forecast for all the periods than the AR and OLS models.

We have not included the model in the project because unlike the models of which we know the function, the neural network is based on self-limiting by minimum error. Therefore the function will not be the same as that used with the models (OLS and AR) seen in the project, so a comparison between the two would have no basis.

In short, as we cannot say that the models created in the project are the most efficient in the use of explanatory variables, it is not possible to make a realistic comparison between self-limiting models and traditional econometric models.