LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

MASTER OF SCIENCE IN
APPLIED ECONOMETRICS AND FORECASTING

MASTER'S FINAL WORK
DISSERTATION

CRYPTOCURRENCY PRICE FORECASTING – AN EMPIRICAL APPLICATION

RODRIGO MOREIRA SERRA

NOVEMBER - 2020

MASTER OF SCIENCE IN

APPLIED ECONOMETRICS AND FORECASTING

MASTER'S FINAL WORK

DISSERTATION

CRYPTOCURRENCY PRICE FORECASTING – AN EMPIRICAL APPLICATION

RODRIGO MOREIRA SERRA

SUPERVISION:

PROF. NUNO RICARDO MARTINS SOBREIRA

NOVEMBER - 2020

Acknowledgements

Firstly, I would like to express my gratitude to Professor Nuno Sobreira, my supervisor, who was extremely patient. He always supported me during the elaboration of this research project.

I would also like to thank all my family for their support. Particularly, I am thankful to my parents, for their effort and for the opportunity that they gave me to deepen my studies; to my brother, for his friendship and for the example he is to me; and to my girlfriend, who has been my motivation through these years, for her love.

Abstract

This dissertation is developed within the scope of the Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing models, with the main objective of comparing forecasting methods. In particular, forecasts will be made using these different classes of methods and cross-validation exercises will be performed to find the most suitable forecast model. Financial assets will be object of study; specifically, five (crypto) cryptocurrencies – Bitcoin, Ether, Litecoin, XRP and Bitcoin Cash – chosen based on their importance and representativity. The price data used are weekly.

The tests to be carried out on the cryptocurrency ‘s logarithm of prices and returns were presented, in order to prove that some stylized facts of the financial series are fulfilled. After showing the test results and the characterization of each asset, a demonstration of the *R* code used during the work is done.

The models that proved to be more adequate to predict the prices of the cryptocurrencies under analysis were ARIMA models of different orders, for each asset.

Keywords: Cryptocurrencies; Cryptoassets; ARIMA; Exponential Smoothing; Forecasting; Autocorrelation; Econometrics.

JEL: C12, C22, C52, C53, C88.

Resumo

A presente dissertação é desenvolvida no âmbito dos modelos Autorregressivo Integrado de Médias Móveis (ARIMA) e de Alisamento Exponencial, tendo como principal objetivo realizar uma comparação de métodos de previsão. Em particular, as previsões serão feitas usando essas diferentes classes de métodos e serão realizados exercícios de validação cruzada para encontrar o modelo de previsão mais adequado. O objeto de estudo serão ativos financeiros; especificamente, cinco criptomoedas (criptoativos) – Bitcoin, Ether, Litecoin, XRP e Bitcoin Cash – escolhidas com base na sua importância e representatividade. Os dados de preços utilizados são semanais.

Foram apresentados os testes a ser efetuados ao logaritmo dos preços e dos retornos de cada criptomoeda, de modo a provar que se cumprem alguns factos estilizados das séries financeiras. Após a demostração dos resultados dos testes e da caracterização de cada ativo, é feita uma demonstração do código de *R* utilizado durante o trabalho.

Os modelos que demonstraram ser mais adequados para prever os preços das criptomoedas em análise foram ARIMA de diferentes ordens, para cada ativo.

Palavras-Chave: Criptomoedas; Criptoativos; ARIMA; Alisamento Exponencial; Previsão; Autocorrelação; Econometria.

JEL: C12, C22, C52, C53, C88.

Contents

List of Abbreviations

ACF - Autocorrelation Function

AIC - Akaike Information Criterion

AICc - Corrected Akaike Information Criterion

ARIMA - Autoregressive (**AR**) Integrated Moving Average (**MA**)

BCH - Bitcoin Cash

BIC - Bayesian Information Criterion

BLCA - Bitcoin Like Crypto-assets

BP - Box-Pierce

BTC - Bitcoin

DES - Damped Trend Exponential Smoothing

ECB - European Central Bank

EMH - Efficient Market Hypothesis

ETH - Ether

ETS - Error, Trend and Seasonality (Exponential Smoothing)

IMF - International Monetary Fund

JB - Jarque-Bera

KPSS - Kwiatkowski-Phillips-Schmidt-Shin

LB - Ljung-Box

LTC - Litecoin

MAE - Mean Absolute Error

MASE - Mean Absolute Scaled Error

MLE - Maximum Likelihood Estimator

P2P - Peer-to-Peer

PACF - Partial Autocorrelation Function

RMSE - Root Mean Squared Error

RW - Random Walk

SARIMA - Seasonal Autoregressive Integrated Moving Average

SES - Simple Exponential Smoothing

TES - Holt’s Trend Exponential Smoothing

List of Figures

List of Tables

1. INTRODUCTION

In the last decades, the World Economy has been experiencing several innovations at a pace never seen before. And even though the scale of today's economic growth can be not as evident as in the period of, e.g., the Industrial Revolution, these are still challenging times, noticeably focused on automatization and digitalization, with aim at the increase of productivity and personal well-being levels. Of course, one might easily identify two key elements as important role-players on boosting this growth nowadays: data and the internet.

With this being said, new realities often come hand-by-hand with new threats, and one of the main concerns of the 21th century society has shown to be, until now, the increasing hazard of privacy violation. This fact inflicted the appearance of privacy-safeguarding technologies, mainly based on cryptography – a field widely mastered by David Chaum, also known as the inventor of digital cash. After this first step being taken, many developments have taken place in this innovative field, until the onset of a crypto-based asset, who has been given the (after all) controversial name of cryptocurrencies.

The main goal of this creation was that the issuance and control of currency was not a responsibility of a singular central authority. The great majority of cryptocurrencies use decentralized systems instead of typical banking systems, and work through a blockchain, which can be simply referred to as a public transaction database, held by a peer-to-peer network. Purportedly, this new type of currency and transaction must function as a way of mitigating some of the problems of the other conventional instruments (namely, lack of transparency, lack of operational accuracy, etc.)

The curiosity and growing scrutiny about cryptocurrencies and its potentially anxious behaviour have been an undeniable occurrence over the last decade, since the birth of Bitcoin (Nakamoto, 2008). Consequently, the research efforts and investigation around forecasting and predicting of future cryptocurrency prices grows on a day-by-day basis, as it can be a big challenge to foresee what will eventually happen in such a volatile market. Nowadays, it is undeniable that exponential smoothing (ETS) and autoregressive integrated moving average (ARIMA) models can prove to be very useful tools when the objective is to search for a way of predicting the future value of any financial asset. Therefore, and even though this market can hardly be considered similar to other ones, due to the characteristics and features of these assets, these methods are valid alternatives for the purpose.

The motivation of this dissertation is to forecast the quote of five different cryptocurrencies – Bitcoin (BTC), Ether (ETH), Litecoin (LTC), XRP, and Bitcoin Cash (BCH) – using the widely known and trustable methods mentioned above. These assets were chosen for the analysis according to their representativity, taking into account some indicators, as the market capitalization, the circulating supply and the price.

The applied method derives from Hyndman and Athanasopoulos (2019), a textbook providing a procedure for the application of these methods. The work goes through the most relevant points and uses many examples, with different types of data. For data collection, data manipulation and modelling, *R* will be the software to use, where three main packages will be installed: *fpp3* – which will provide us almost every function we need –, *summarytools* – that serves mainly to descriptive statistics calculation – and *crypto* – for the purpose of collecting cryptoassets data.

Section 1 (the present one) is designed to be the introduction, where the objectives, the methodology and structure of the thesis were subject of a first approach and explanation.

Section 2 is reserved for the literature review, that intends to go through the definition of cryptocurrency and its evolution, as well as mentioning and referring to the researches regarding forecasting and price behaviour.

Section 3 seeks to present and fully specify the methodology and data used in this work, and the reasons for that. A descriptive analysis is to be performed also, where some technical features and statistics of the series will be analysed. Moreover, an explanation of the *R* script used will be provided.

In Section 4 we will compare the models within each category and after choosing the best options, do a forecast competition between them, and produce the corresponding predictions.

Ultimately, Section 5 is composed by the conclusion, presenting also the limitations of the methods used for this particular case, and further possible investigations in the field.

2. LITERATURE REVIEW

There is a vast number of cryptocurrency definitions that we can find by doing a simple internet search, and it is fair to say that there is not a general agreement on how to name and treat this type of asset. It will eventually be a long road until we fully understand its characteristics and what are they comparable with. Perhaps the best approach is to address the question going through its roots. The word "cryptocurrency" first appeared in the early 21st century and it was originated by joining "crypto" – which refers to cryptography – with "currency". Actually, and assuming there is general consensus in affirming that cryptocurrency is deeply bonded to digital encryption techniques, the second part of the word is highly controversial. It is well known that for any given asset to be considered a currency, there are at least three main properties that need to be fulfilled: it must be used as a unit of account, a medium of exchange and a store of value. Jevons (1875) argued that it should also be acknowledged as a standard of deferred payment, even though most present-day articles prefer to omit this function.

As Bariviera et al. (2017) point out, cryptocurrencies are hardly satisfying all of these features. Baur et al. (2018) argues that they are, in fact, fulfilling these three essential requirements whenever one uses them as money. As the world experiences its evolution and acquires knowledge about this type of asset, institutions like the International Monetary Fund (IMF) and the European Central Bank (ECB) start referring to cryptocurrencies as cryptoassets instead, the later defining them as a new type of asset recorded in digital form and enabled by the use of cryptography (ECB, 2019). The IMF (IMF, 2019), while defining the term in an identical way, extended its analysis even deeper and proposed the existence of two different crypto groups: BLCA's (e.g., Bitcoin) – designed to be general-purpose mediums of exchange for P2P payments – and digital tokens (e.g., Litecoin) – that are intended to have other functions. We will, however, use the broad sense definition instead of entering in this level of detail.

Moreover, in 2019, during a conference about the subject, the IMF officially recognized cryptoassets as a store of value, even though this organization continues to reject the idea of a future replacement of conventional currencies. To help explaining this view, we might consider Chiu, J & Koepp, T. (2017), who took the example of Bitcoin and showed that using it instead of traditional currency can be until 500 more costly, if we consider low inflation conditions. One may argue, however, that this is a consequence of the poor design of the Bitcoin system and that other cryptoassets have more efficient characteristics.

Further investigations have been done with the intention of demystifying the crypto market. Cheah et al. (2018) and Kristoufek (2018), among others, conclude that the EMH is violated and that the Bitcoin market is inefficient – while other studies argue the opposite. Hu, Y. et al. (2019) go further and do panel unit root tests to investigate the efficiency of 31 cryptoassets, and show evidence of inefficiency.

We may then conclude that there is still not a widely accepted truthfulness around cryptocurrencies and one of the few characteristics we can undoubtfully attribute to them is their speculative nature. Their value (usually shown as per USD) is mainly driven by demand and supply, but also shows very high volatility to changes in financial agents' sentiments and fears, who tend to overreact to every stimulus. Accordingly, cryptocurrencies and its behaviour show, for the greater time, more similarities with a speculative asset than with a fiat currency; a fact usually related to the increasing possibility of bubble emergences (when the price of a given asset abnormally diverges from its underlying fundamental value) and price uncertainty.

With respect to cryptocurrency modelling and forecasting, Bakar & Rosbi (2017) used ARIMA models to predict exchange rates, concluding that the results are satisfactory, but that volatility decreases the accuracy of the forecasts. By running one-step ahead forecasts of three cryptocurrencies' returns (BTC, LTC and ETH), Hotz-Behofsits et al. (2018) compare univariate and multivariate time series models, finding that it is helpful to allow for flexible error distribution and time-changing parameters. Catania, Grassi & Ravazzolo (2019) demonstrate that using combinations of univariate models for point forecasting, and a selection of multivariate models for density forecasting shows good results. Still among model comparison frameworks, Bohte & Rossini (2019) defends that a combination of stochastic volatility and a student-*t* distribution shows the best results, between the options used.

Alahmari (2019) uses machine learning ARIMA models with weekly re-sampling to predict cryptoassets prices, and explains that these outperform other options, in terms of error measurements as RMSE and MAE. Kumar (2019) presents a comparison study between ARIMA and Neural Network approaches, concluding that the first is better than the second for a shorter time-horizon, and vice-versa. Considering only BTC, Rebane et al. (2018) defend that Recurrent Neural Networks generated throughout most of the price history show superior results comparing to ARIMA models.

3. METHODOLOGY AND DATA

After the above disclosure of the theoretical framework, a specification and explanation of the used methodology and tests will follow. Also, an approach to the data which will serve as input for this research will be done.

3.1. METHODS AND ESTIMATION

Following the literature review that was done and the available options for the purpose of this study, the choice was to use the guidelines and framework from Hyndman & Athanasopoulos (2019). This textbook provides theoretical and practical guidance for the use of time series forecasting techniques. These include exponential smoothing – either simple or accounting for trend and seasonality – and autoregressive integrated moving average models (ARIMA), and we are using those ones through this research. Both methods can be used in time series forecasting applications, such as for our purpose, of studying cryptocurrency quote evolution.

After running a comparison with the intention of choosing which ETS model type explains better the log price of each cryptocurrency, and doing the same for ARIMA models, a comparison between ETS and ARIMA models will be done too, as far as one-step ahead forecasting accuracy is concerned, so we can verify which method is the most suitable. A RW and a RW with a drift will also join the forecast competition. It is possible that the results will diverge, on the grounds that there are substantial attitude differences from asset to asset. These are special currencies – to the point that we are not quite sure if they should be called that way – and have special characteristics.

The forecast competition exercise will occur in the following way. A rolling window cross-validation procedure will be done to acknowledge the accuracy of each selected method. Here, we choose the initial window size of a training set (in our case, 10 observations), and the test set is composed by the next observation (the eleventh observation). After this, the sample window will grow and the training set will be then composed by the first 11 observations. This will continue until the last observed value. The predictions are always done taking only previously observations into consideration, and ultimately, the forecast accuracy is obtained by averaging over all the one-observation test sets.

3.1.1. STYLIZED FACTS

In order for us to apply these methods to our time series quotes, there are several stylized facts of financial assets that should be tested. As we are dealing with prices, the logarithm can be applied to reduce the size of the variations, but it is not enough for us to make our variable stationary. It is widely known that prices are non-stationary and usually a transformation (differencing) is performed in order to make its process stationary, and help stabilise the mean of a time series. The quote series analysed in this research may therefore be non-stationary and only their first differences (logarithmic returns) should be stationary.

By looking to the autocorrelation function plot of a time series, one can notice the non-stationary behaviour, as it tends to have large significant spikes for all lags (with correlations close to one) and slowly decrease. Even the plot of the price itself can be quite suggestive. Nonetheless, to be sure of what we are dealing with, there is a wide variety of methods that enable us to test for stationarity. In this dissertation we are going to use the KPSS test (Kwiatkowski, Phillips, Schmidt, & Shin, 1992), to test for the existence of a unit root in the log quotes and log returns.

$$KPSS = \frac{(T^{-2} \sum_{t=1}^{T} \hat{S}_t^2)}{\hat{\lambda}^2} \tag{1}$$

Above is the test statistic for this test, where $\hat{S}_t$ is a cumulative residual function and $\hat{\lambda}^2$ is a long-term variance of the errors. The null hypothesis of this test is that the underlying process to the data, say y, is stationary (H_0: y stationary). A sufficiently high value for this statistic, bigger than the critical value, suggests that the stationarity hypothesis should be rejected. After applying this test, we will apply a function to our data which runs a sequence of KPSS tests, with the function indicating the order of differencing that should be applied so we get a stationary series.

Subsequently, we will apply two types of portmanteau tests to the data, so we can assess if there are any symptoms of autocorrelation in the transformed series, which will happen to be the log returns of the quotes. As recommended by the textbook we followed, the tests we are going to use are the BP (Box and Pierce, 1970) and the LB (Box et. al, 1978) tests for the autocorrelation. For both tests, the null hypothesis is that there is no autocorrelation (H_0: no autocorrelation) in the data, and such null is rejected if low p-values are found, indicating, precisely, that the series exhibit autocorrelation (we use a 1% critical value).

The first test (BP) has the following statistic. Let r_k be the sample autocorrelation for lag k, T be the number of observations in our sample and l the maximum lag considered in the test. Then, the statistics is based on

$$Q = T \sum_{k=1}^{l} r_k^2 \tag{2}$$

where small values of Q suggest that there is no significant autocorrelation in our series. As we are going to see further ahead, our data shows no signs of seasonality, and thus we should be using ten lags to do the test ($l = 10$). The second test (LB) is often preferred to the previous one and it uses the statistic

$$Q^* = T(T+2)\sum_{k=1}^{l}(T-k)^{-1}r_k^2 \tag{3}$$

that follows the same notation.

After the two tests applied above, we only miss presenting the normality test, which as the scope of clarifying if the data follows a Normal distribution. One of the stylized facts of the returns of financial series is that their distributions tend to present either a positive or negative skewness and, thus, the aim of the application of this test is to understand if our data respects that hypothesis, as expected.

It is important to observe the histograms of the returns. From them we might take some conclusions or suggestions regarding normality of the distribution, but it is crucial to root the study in statistical evidence. Hence, we will use the JB statistic (Jarque & Bera, 1980), which is based both on the skewness of our distribution, and on its kurtosis. This is particularly relevant because, as far as financial returns are concerned, this type of assets often displays high kurtosis values. The test statistic follows below.

$$JB = \frac{n}{6}(S^2 + \frac{1}{4}(K-3)^2) \tag{4}$$

Here, S represents the skewness coefficient and K the kurtosis coefficient, calculated from the sample. The alternative hypothesis for this test is that the logarithmic returns of the asset of interest, say y, do not follow a normal distribution (H_0:

$y \sim Normal$ vs. H_1: $y \sim F$, where F is a distribution other than normal). Then, for this test, we expect that there is evidence suggesting the rejection of the null.

3.1.2. Exponential Smoothing

Exponential smoothing, also named in some literature as ETS, is a technique which consists in attributing weights to each observed value, in order to predict upcoming values. Newer observations have a higher weight, and there is a decay until the older one, which will have the lower impact on the predictions. Depending on the subject, these methods are usually capable of producing well-grounded forecasts. It can be considered a simple procedure, comparing to other alternatives, though it usually shows very decent results.

The method was firstly introduced in the late 60's, by Robert Goodell Brown, and it is still broadly used, namely in the field of economics and finance. With the objective of obtaining reasonable forecasts of the target variable *y*, it relies on the idea that the last observed value should explain better the variable's future value, than older ones. Gardner & McKenzie (1985) and Holt (1957) were responsible for two important developments in this field. The second one presented a method to deal with time series that embody a linear trend, whereas the first supplemented this work and added a dampening effect in form of a parameter, to correct the usual overestimation of Holt's method. The framework is presented as follows:

$$\begin{aligned} \hat{y}_{t+h|t} &= l_t + (\phi + \phi^2 + \cdots + \phi^h)b_t, \\ l_t &= \alpha y_t + (1-\alpha)(l_{t-1} + \phi b_{t-1}), \\ b_t &= \beta^*(l_t - l_{t-1}) + (1 - \beta^* \phi b_{t-1}) \end{aligned} \tag{5}$$

where y_t is, in our case, the log quote of the cryptocurrency at time t, and $1 > \alpha > 0$ is called the smoothing parameter. Note that α controls the weight given to each observation in predicting the future value of y, based on the available information at a specific moment in time. The conjunction of the equations above can be referred to as the Damped Exponential Smoothing (DES) framework, which incorporates a softened trend. If we do $\phi = 1$, we get what can be called a Trend Exponential Smoothing (TES – Holt's method) – which accommodates the trend but without any inhibition. Moreover, if in addition, we establish $\hat{y}_{t+h|t} = l_t$ and forget b_t and its past values, we get the equation of the SES model, used when there are no signs of existence of seasonal or trend components in the historical data, which would be an unexpected result as we deal with log prices.

In this research, we will perform a cross-validation procedure in the open-source software *R* to compare the accuracy of these options' forecasts when applied to the assets under investigation, and conclude about its effectiveness. The parameters will be estimated using MLE. It should not be left to mention that further works were conducted with the goal of incorporating not only the trend but the seasonal component in the methods. However, cryptoassets price series do not show evidence of seasonality, and thus we will not explore this field.

3.1.3. AUTOREGRESSIVE INTEGRATED MOVING AVERAGE

In respect to ARIMA, it is a model used to make predictions while considering the lagged values of the predicted variable as well as a moving average component, whilst accommodating for non-stationarity. For time series forecasting, this is one of the most popular models and it consists in a combination of AR and MA factors.

This model is a valid option for our purpose, since price variables are usually considered to be unit root processes – somewhere in the following sections, results will be provided regarding the stationarity of the series. In the subsection above, we mentioned the possibility of accommodating for seasonality of data. The same holds for this case (SARIMA), but as explained before, neither the series nor the tests, that will be exposed at some point below, show any seasonal patterns.

Let y_t be the log price series, computed from the quotes of each crypto asset.

For the sake of simplicity, let us use the backward operator (B) to work with time series lags. Then, the ARIMA(p, d, q) model shall be written as:

$$(1 - \phi_1 B - \cdots - \phi_p B_p)(1 - B)^d y_t = c + (1 + \theta_1 B + \cdots + \theta_q B_q)\varepsilon_t \qquad (6)$$

where ϕ_i denotes the autoregressive coefficients for each lag i, θ_j denotes the moving average coefficients for each lag j and ε represent the error. This can be considered the general representation of any ARIMA(p, d, q).

To do the price forecasts with these models, we will use MLE and apply a function which enables us to automatically choose the correct order for our model, based on information criteria (AIC, BIC, AICc, etc.). For this purpose, AICc will be the most relevant decision factor.

3.2. DATA AND FEATURES

This section intends to make a detailed approach to the data used in this research. We will therefore go through the quote time series of all cryptocurrencies under analysis, for the manners of growing understanding about the nature of each one of them. It will be noticeable that part of the evolution of the prices of these assets may be explained by the same factors, as some variations of the plots seem to be related (similar variations, with similar magnitudes and almost at the same period).

We will analyse five crypto assets, by the respective order: Bitcoin (BTC), Ether (ETH), Litecoin (LTC), XRP and Bitcoin Cash (BCH). They were selected taking some criteria into account – mostly market capitalization, but also price, volume and representativity. All periods of analysis will have different start dates, depending on the crypto assets launch date, and August 30th, 2020 as end date. The start date will hence be adjusted to make sure that we have full weeks, considering we will aggregate daily data by week and calculate weekly averages of the quotes.

As broadly recognized in the scientific community, typically, a price series is non-stationary and has specific characteristics which demand the application of some transformations to the variable. Although we will have a look at the linear scale for each cryptocurrency, for the matters of this technical research, a logarithmic (log) transformation to the prices will be done, with the intention of softening severe level variations. Moreover, a computation of log price's first difference will be performed – in order to obtain log returns – so we can work around the existence of non-stationarity, and analyse relevant statistical features.

3.2.1. BITCOIN (BTC)

Fig. 1 shows a plot of the trajectory of BTC quote overtime, using daily data, in the period between April 29th, 2013 and August 30th, 2020. When analyzing daily numbers, we are able to recognize high volatility in the data, suggesting that BTC prices behavior is much more alike the one of a stock than of a conventional currency.

By 2013's latest quarter that investors became aware of the inflammable essence of Bitcoin – between the 4th of October and December, the price of one BTC went from 129.01 USD to 1,151.17 USD, hitting its maximum close value until that date. Nonetheless, this turned out not to be a permanent increase and only by 2017, and after a significant gradual drop of more than 70%, the price showed a new expanding trend.

This is indeed the largest jump ever seen in BTC, probably boosted, among other factors, by the increasing acceptance of Bitcoin in a large number of businesses, reaching 19,497.40 USD by the end of the year (16th December). As expected, after a huge increase in BTC quote, finance professionals started to warn about the possibility of a burst in what seemed to be a price bubble, which eventually started to show. Some events (namely, rumors about the ban of crypto in South Korea) are thought to be the cause of what is known has the Great Crypto Crash.

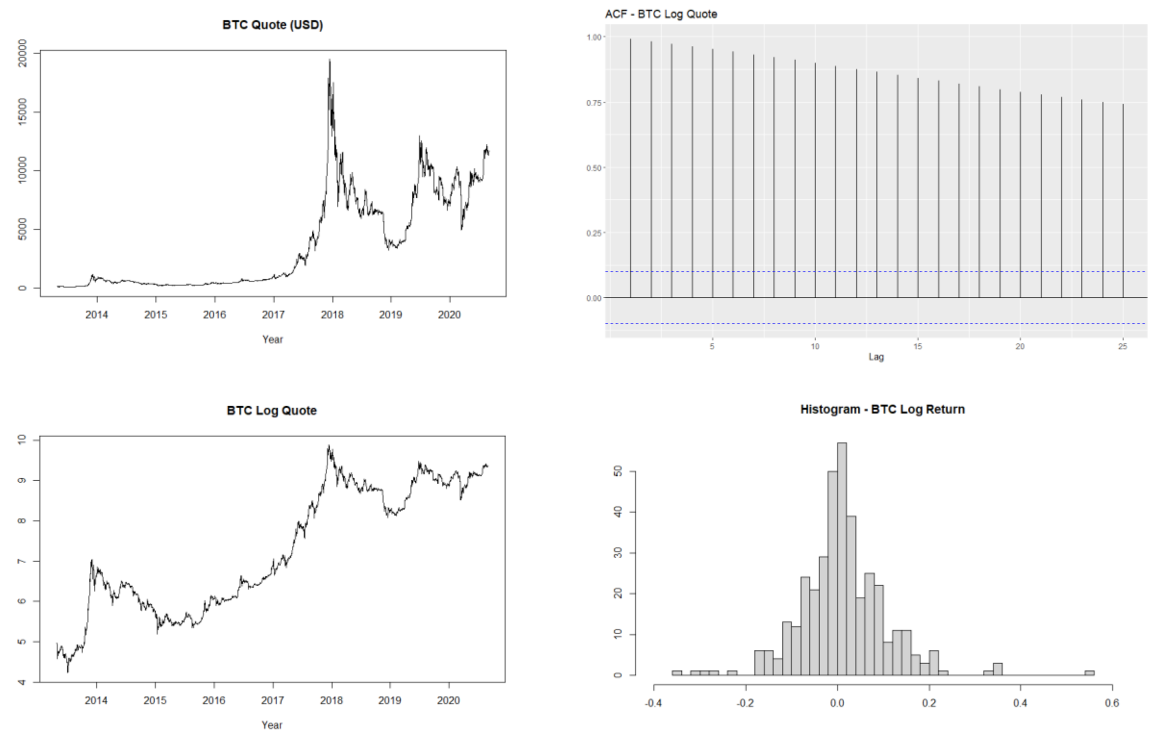

Figure 1 - BTC quote in USD overtime (top left), logarithmic quote (bottom left) and its autocorrelation function (top right), and the histogram of the logarithmic return (bottom right).

After this and during 2019, BTC value went through a similar evolution, but in a much smaller scale, and we can verify that with simple visual inspection of the plot. With this being noticed, it is inevitable to say that when it started spreading, COVID-19 seemed to be acting as a big trigger to what could be a large drop in this quote. However, the drop stopped soon and even though the world is still living this pandemic, looking to the most recent numbers suggests that it is not having a considerable and permanent impact in this market.

It can be noticed by the autocorrelation function plot (Fig. 1) of the log quote (already using weekly data) that the series shows signs of having a trend. When this is the case, the coefficients for small lags tend to be positive and high because observations that are nearby in time are also similar in size. Also, slowly decreasing positive values are observed. Therefore, we can conclude that the series has a non-stationary pattern, what must not be considered a great novelty as we are dealing with price series which seem to act as financial assets.

3.2.2. ETHER (ETH)

Even though Ethereum is usually referred to as the cryptocurrency itself, it is actually a decentralized, open-source platform which allows the development of consensus-based applications (in words, a blockchain). Ether (ETH), the cryptocurrency, was created to function with all the applications within this blockchain.

Fig. 2 shows the price evolution of ETH since August 10th, 2015, one of the first days for which our source (*CoinMarketCap*) has data – it is not the first, because we fixed the end date of the period, which is August 30th, 2020 for all the crypto assets, and collected entire weeks. Although Ether was initially released in July 30th, 2015, this year's first day was the date of the stable (also called "production") release of this cryptocurrency. A simple way to translate this is saying that only from 2020 on, all the remaining bugs of this platform were considered to be acceptable; a detail that can be suggestive with respect to the volatility of the asset.

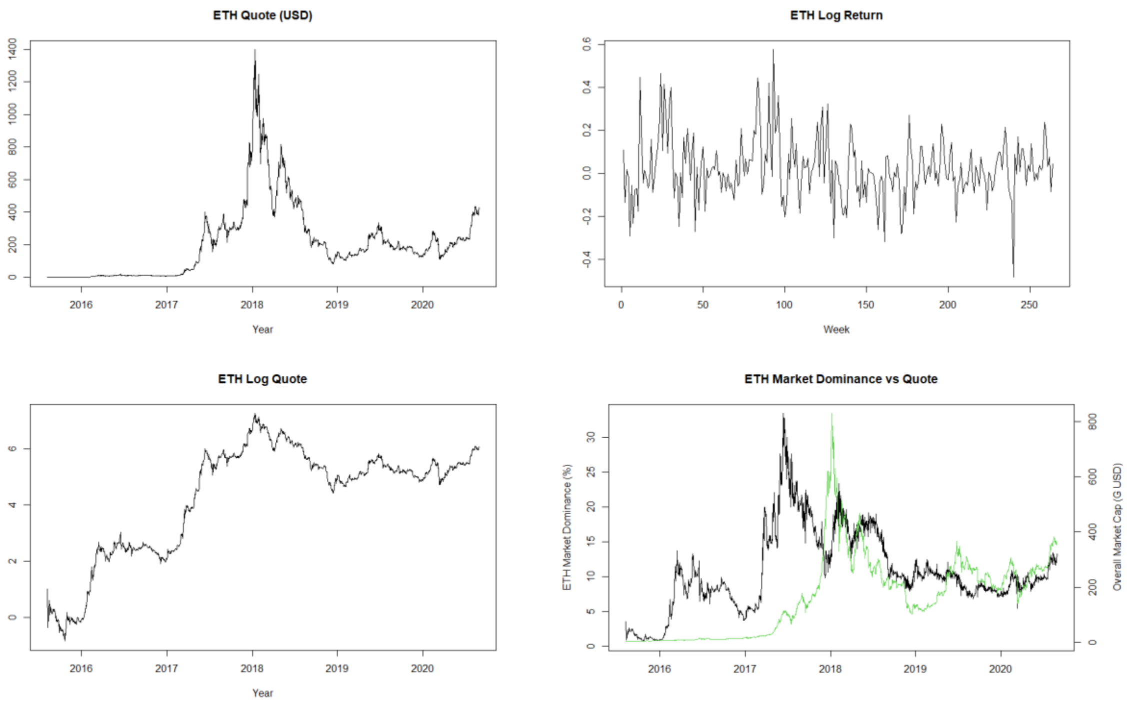

Figure 2 - ETH quote in USD overtime (top left), logarithmic quote (bottom left), the first difference of logarithmic quote – log return – (top right), and the evolution of the market dominance of ETH, when compared to its quote (bottom right).

Looking to the price evolution, at the beginning of the analysis period the closing price was 2.77 USD. This was the only day until January 24th, 2016 where the price was above 2.00 USD. However, thenceforward, it never got below. In 2017, the price of Ether started from a minimum of 8.17 USD and reached 756.73 USD, after a lot of peaks and valleys, and in just 13 days it has grown approximately 84.5% reaching 1,392.42 USD by January 13th, 2018. This growth seems to be largely leveraged by Bitcoin's momentum – investors started to diversify their cryptocurrency portfolios due to the growing success of BTC. This success stimulated the crypto market and its overall capitalization initiated a period of high rising; much of this rise can be attributed to ETH. Cryptocurrency traders putted their faith mostly in this asset when choosing to extend their sight, and its market capitalization dominance surpassed 30%, due to the increase of both price and circulating units.

We can then conclude that 2017 was definitely ETH's most thrilling year since its creation. Regarding the following years, the behaviour of ETH can be compared to the one of BTC, and even the COVID-19 impact has shown to be similar, until today.

3.2.3. LITECOIN (LTC)

Charlie Lee (2011) was the responsible for the creation of LTC, which is a BTC derivative. However, despite the root basis being similar, this financial asset is considered to be safer than its relative and works better as a mean of payment and transaction. In terms of historical data, we may detect that, clearly, the attitude of LTC price overtime is almost a copy of the one from BTC. Of course, in a much shorter scale. The period of analysis is exactly the same for both, we start in April 29th, 2013. Fig. 3 can be used for deeper analysis, but regarding the important marks of this evolution, the maximum closing price of LTC was 358.34 USD (also in December 2017), and the minimum was 1.16 USD early during 2015.

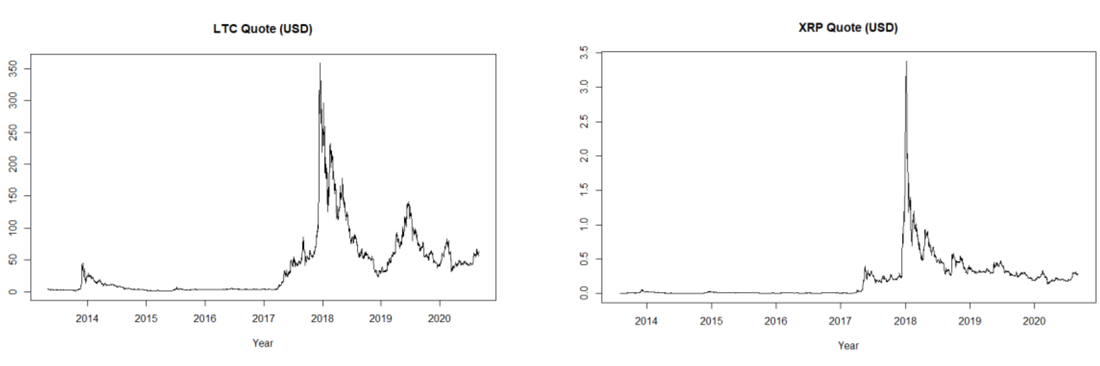

Figure 3 - LTC (left) and XRP (right) quotes in USD overtime.

This was also the year where a halving occurred in LTC market. A halving is an event that has the purpose of reducing by 50% the rewards per block, with the objective of preserving the purchasing power of the currency. Otherwise, eventually, LTC value would start to continuously deflating, because there is a limit of LTC in circulation.

After a halving, a gradual increase of the quote should be in march. And this expectation seems to be the best reason for the rise since the beginning of 2019 – during 2018, the price went down, just like BTC price – since it was known, or at least expected, that the next Litecoin halving would take place in 2019, somewhen in the summer. Following another price drop and a further smaller increase, the coronavirus impact is now being roughly the same as for BTC.

Nowadays, LTC is only behind BTC, Tether (USDT) and ETH, respectively, in what regards to the trading volume, i.e., the number of units of a crypto traded during a certain time.

3.2.4. XRP

This cryptocurrency is one of the most representatives of this market, as in terms of capitalization, it is 4th, only behind BTC, ETH and Tether (USDT). The asset is intimately connected to Ripple Labs, a blockchain company founded in 2004 which had the objective of creating a secure and quick option for digital payments. XRP dues its existence to this technological company, even if nowadays it is separate from its network, being an independent asset.

Our data period begins in 5th August, 2013 and ends at the same date as all the other assets under analysis. It is a fact that when BTC is showing high inflation, all the other altcoins tend to follow the same trend, because investors' expectations are similar between the most traded cryptocurrencies in the market. However, from Fig. 3, it can be seen that regardless the momentums being in pair with the ones of BTC, after 2018, XRP price grows much less and seems to be in the shade of the mother of digital coins. The price variations can be noticed at nearly the same spans, except they are in smaller magnitudes.

Nevertheless, almost all cryptocurrencies show the same explosive nature and, in this case, that could be remarked between the 7th of November and December, 2017, when XRP quote went from 0.22 USD to 3.38 USD (its maximum). To have an idea of the size of this variation, this corresponds to an increase of 1436% in this asset's price. This was the higher place where the price stand, and from that moment on it kept following a decreasing tendency. In what refers to 2020, and even in times of uplifted uncertainty, the expectations are that 2021 can be XRP's year.

3.2.5. BITCOIN CASH (BCH)

This subsection intends to present the last of the cryptocurrencies which will be object of research through this work. Bitcoin Cash, as the name suggests, was created from BTC; most specifically, from a split – a hard fork – where the tokens of every investor were duplicated, being these duplicates units of BCH. However, these new tokens cannot be considered clones of their primitives, considering that BCH has different technical properties which, among other details, increases the number of transactions per second and permits that new registrations in the blockchain are done in a faster and more efficient way.

The currency was founded in 2017, and we collected data since 24th of July, giving us a sample of 162 weekly prices to run our models. As in the subsections above, we choose to visually present the daily prices (Fig. 4), so a we can better probe the volatility and difficulty of predicting such assets.

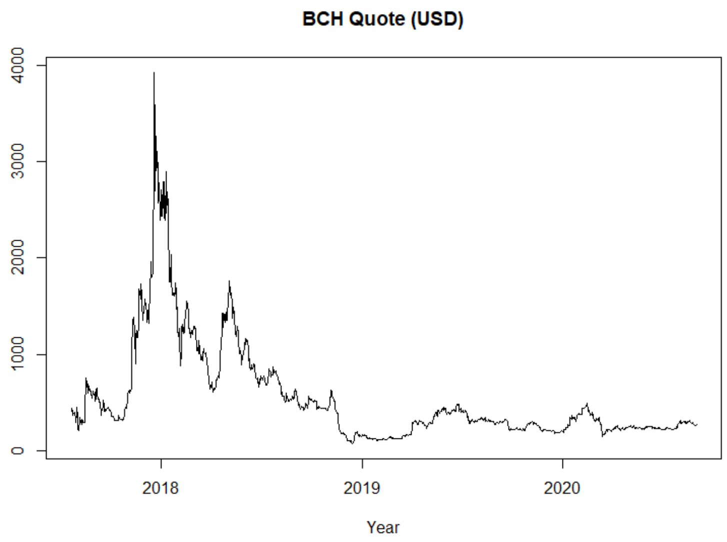

Figure 4 - BCH quote in USD overtime.

3.2.6. TIME-SERIES FEATURES AND TESTS

This section has the intention of assessing the realization of the stylized facts referred to in section 3.1., by presenting and showing the results of the tests that were run. It should be reminded that all test and forecasting methods are applied to weekly data.

With respect to the stationarity of our data, using *R*, we applied KPSS statistical tests to the log quotes, which we expected to be $I(1)$ (integrated of order 1) time series. For all cryptocurrencies under research, we found evidence pointing for the existence of at least a unit root in all cryptocurrency quote series, and thus, there are significant signs that our variables comply with at least one of the stylized facts of financial series. The test results can be verified in Appendix A – Table 2. The test statistics with most significant results were the ones from BTC and XRP, with a larger distance to the critical value of 1% than the others. Additionally, and to make sure that none of these series displays $I(2)$ behaviour, we run a sequence of KPSS statistics that will suggest the number of unit roots of each crypto asset log quote. The results suggest there is a single unit root in each of our crypto price series.

However, when applying the same tests to log returns – differenced log quotes –, one of the crypto assets (ETH) still showed evidence of $I(1)$*ness*. When we review the series of log returns itself (A – Fig. 7) we are tempted to believe that the series is stationary, but that is not what the test supports. This statistical evidence seems to be, still and all, weak considering that the test for the number of unit roots in log quote encouraged the idea of an $I(1)$ succession, and hence stationary after differencing. The conclusion regarding these tests is that the models we will apply to the log prices (ETS and ARIMA) can prove to be interesting forecasting techniques, since they accommodate for non-stationary behaviours.

In relation to the autocorrelation hypothesis, first of all, there must be a plot review. We know that there are clues that can lead us to answers about the presence of autocorrelation in a certain series. Fig. 5 (Appendix A) represents the autocorrelation functions (ACF) of the log prices of the cryptocurrencies subject to this study. There is no gain for us to observe the PACF plots of the non-differenced series, because only by looking to the ACF's, we notice the absence of stationarity supported from the KPSS tests presented before. There are large and relevant autocorrelation coefficients until high lag orders.

Nonetheless, the ACF's and PACF's of the differenced series (log returns) are of particular importance and, as advised by Hyndman & Athanasopoulos (2019), the signs we can try to spot are as follows: if we notice that the ACF has some significant coefficients and decays exponentially, and the PACF has only a few (p) significant spikes, then an ARIMA$(p, d, 0)$ is potentially a good option for modelling the variable. Beyond doubt, p will always be a positive number if we find that the PACF has the shape described above, which supports the materialization of autocorrelation in the target variable. Fig. 8 and 9 represent the ACF and PACF log return plots for each cryptocurrency, respectively. The main feature of both autocorrelation functions is that all of them have few relevant lags, even though there is always at least one in every graph. BTC and Ether ACF's present the exponentially decaying pattern in a transparent way, whereas for XRP and BCH it is very difficult to perceive that behaviour. As to the partial autocorrelations, and except for some statistically significant values in distant lags, in general, only the first lag is relevant. For ETH and XRP, the two first lags appear to be meaningful yet by a small margin. The conclusion we get from the observation of the autocorrelation plots is that an ARIMA$(1, d, 0)$ might be a valid option for modelling the prices. The fact that some cryptos have two significant spikes does not necessarily imply that there is a second order autocorrelation in y.

In this way, we may advance and present the results for the autocorrelation tests to the logarithmic returns of the assets. Fig. 10 shows the test statistics for the BP and LB. If the tests had some level of ambiguity, the one we would consider the most would be the second one, since it is well known that, from Monte Carlo studies, it usually displayed better finite sample properties. All tests point in the direction of the existence of autocorrelation, at the 1, 5 and 10% significance levels. The p-values of all tests are way below the thresholds and hence there is evidence supporting the existence of autocorrelation in the weekly returns of our crypto assets. Hence, the null hypothesis of absence of autocorrelation between different observations of the same series close in time is strongly rejected. All in all, this reinforces the idea that ARIMA and Exponential Smoothing models can be a convincing selection when dealing with cryptocurrency prices, due to their nature of exploring the inherent temporal dependence that is present in time series variables.

Moving forward to the last hypothesis testing performed in this research – normality test –, the Jarque-Bera test results confirm the non-normality of the series, and thus, all the tests supported the compliance of our data with these three stylized facts of financial . The function we used in *R – JarqueBera.test()*, from the *tsoutliers* package – provides three outcomes: the overall JB test, with $\mathrm{H_0}$: normality, a skewness test ($\mathrm{H_0}: S = 0$) and a kurtosis test ($\mathrm{H_0}: K = 3$). The test statistics can be checked in Fig. 11 – Appendix A. On the whole, the null hypothesis was rejected with a significance level of 1%, being all the p-values very close to zero. This defends the non-normality of the weekly returns, the skewness of the distributions and the non-mesokurtic nature of them, as we expected. The only somewhat weaker evidence to this general statement is BCH's skewness test, as it only rejects the null at 5% level. Anyhow, we consider it to be sufficient evidence that the distribution of BCH's weekly returns is skewed.

If we observe the summary statistics in Table 3, we may notice that it is a positive skewness that can be observed in all distributions of our data. LTC shows the higher value, with $S \approx 2.61$, which suggests exactly what can be verified in the histogram of the log returns – the distribution exhibits a long right tail. As predicted, the lower skewness coefficient is the one from Bitcoin Cash ($S \approx 0.25$), explaining the difficulty in rejecting the hypothesis of non-skewness. All the remaining values lie between these two extremes, and the reason why there is this general characteristic in financial series is the rationality of investors. In what regards to the kurtosis of the distributions, we find some unanticipated results. The expectation was that all distribution of the different log return series would present a leptokurtic form (which is characterized by having fatter tails, i.e., more extreme values far distant from the mean of the series). Although, surprisingly, ETH and BCH have their kurtosis coefficients below three ($2.5 < K < 3$). A fact that can be related to this output is the size of our samples, since comparing with the other cryptocurrencies under analysis, we have fewer weekly observations of the returns of these two. After running the same features for the daily returns, it was understood that this is the cause as shown in Fig. 11. If our analysis was based in daily data, with a larger sample, different results would be obtained.

Regarding the mean of our return series, the values are in the interval of [-0.002;0.02], and the minimum corresponds to BCH, with a negative mean of weekly returns, whereas the maximum is Ether's mean, the higher of the five under analysis. The standard deviation values variate within a small range, from around 0.1 to 0.18, corresponding respectively to BTC and XRP.

3.3. PROGRAMMING WALKTHROUGH

In this section we will present the steps used to get all the information and analysis of this research. Appendix B has the path to the repository where the *R* script can be found. The creation process of summaries that include all five assets is also provided. For the sake of simplicity, we will go through BTC part of the code, as an example. We start by installing and call all necessary *R* packages, each with its functionalities. One of the most important among these is *crypto*, which enable us to collect cryptocurrency data from *CoinMarketCap* website. Then, general crypto data (used mostly to do market dominance calculations) and BTC daily data are collected. For this crypto specific part, logarithms are applied to the close price, which is the one that is typically used in financial markets. After plotting the daily time-series, we prepare an identical table – now with weekly data – without losing the last date of each week after aggregation, so we are able to plot by date and not by week. We draw the plots, create a table with descriptive statistics and draw the histogram of log returns. Afterwards, KPSS stationarity tests, BP and LB autocorrelation tests and JB normality tests are computed, to get conclusions about the stylized facts of financial series.

We dive now into ETS section, where we begin by comparing one-step forecast accuracy between ETS types. Taking this results in consideration, we continue by running the *ETS()* function to find the model that minimizes the AICc, and gather report information of each of the other alternatives. This information table not only includes the estimated parameters and the information criteria, but also the p-values of BP and LB tests applied to the residuals. The *components()* function is visually useful as it draws four plots: explained variable, level, slope and remainder time-series. Finally, we create the residuals graphs, including the ACF and PACF, histogram and time-series in essence.

Thereafter comes the ARIMA part of the *R* script. The ARIMA model selection was done in a different manner. A model was fitted, where the specifications are as follows: there should be no seasonal component at the model, the integration order should be between 0 and 2, and the AR and MA orders should be between 0 and 5. After the first execution, a model minimizing AICc is returned. To obtain the second-best model, we remove one order possibility (of AR or MA component), and this is done again to get the third output. Subsequently, and as the information of these models is collected, we add the residual autocorrelation test p-value to create the final report table. A similar process to the one used for ETS.

Ultimately, and as we now have the models that were chosen to join the forecast competition, we run the cross-validation recursive window procedure, to compare the precision of the models for each currency. The forecast is computed and we get our final plots and predictions. The bottom of the code is left for the binding of some summaries that were split by currency.

4. Modelling and Comparison

We can now proceed with the application of the estimation methods, in order to do a comparison between all models and use the one with better properties to fit and forecast future weekly quotes of the crypto assets.

Regarding the Exponential Smoothing (ETS) approaches, cross-validation procedures will be run to compare the accuracy of the forecasts of each type of non-seasonal (as our data does not show signs of having seasonality) models – SES, TES or DES –, taking into account measures like the RMSE, MAE or MASE. Moreover, we will also execute a method that chooses the model with a smaller corrected Akaike Information Criterion (AICc), to see if the choices match. The selected method for each cryptocurrency should also, to a certain extent, be based at our economical judgement. As to ARIMA models, the *R* function executes the Hyndman-Khandakar algorithm, which does some of the work needed to choose the better model. It differentiates data when there is a unit root, analyses ACF and PACF to select the orders of the ARIMA and uses AICc to search the best fitting model. Before this, we will already employ the log transformation to each variable in order to stabilize the variance.

After selecting the most appropriate models of each type and before doing the ETS vs. ARIMA comparison, the residuals will be analysed, and portmanteau tests will be applied to spot any possible manifestation of autocorrelation in the errors. Only after this we will be able to compute predictions of future values of each variable.

Beginning with the ETS models, and specifically by analysing the results of the $\text{ETS}(A, N, N)$ – Simple ETS – we can notice that, for every cryptocurrency under research, α is approximately equal to 1, which means that if we base our predictions in this model, the results would be similar to the ones of the Naïve method, of simply assuming that the one-step future value of y equals the present value ($\hat{y}_{t+1|t} = y_t$). Fig. 12 represents the BP and LB tests for the residuals, and the statistics do not support the absence of autocorrelation in the errors of the model, for the five assets. The null hypothesis is relevantly rejected, as all p-values are nearly zero. Having the outcome of the estimations, we know an $\text{ETS}(A, N, N)$ model does not fit well to our data, and hence other options must be tested. As said before, there are no signs of seasonality, and thus the remaining estimations will be based on two methods that incorporate only a trend – $\text{ETS}(A, A, N)$ and $\text{ETS}(A, Ad, N)$.

For Bitcoin (the most representative crypto), recurring to Table 3 and 4, a comparison between all methods can be done. Still, for all of them, $\widehat{\alpha} \approx 1$, which shows the importance of the last observed value for our prediction. However, this is not the only component of the estimation, since in both Holt's trend-accommodating models there is the slope component, whose level at time t, in the particular case of the DES is estimated to be negative ($b < 0$). We concentrate on this one, as not only it is the one with the lowest information criteria but also, it is the one selected by the cross-validation procedure which compares the accuracy of the one-step forecasts of all the methods (comparing, e.g., RMSE, MSE and MASE). This $\text{ETS}(A, Ad, N)$ method, incorporates a dampening parameter (ϕ), whose estimate is relatively high, around 0.8, meaning the trend is reduced to a flat line during a not so close horizon – $\phi = 1$ corresponds to the Holt method (ETS(A, A, N)). Although this is the type of ETS showing better results, the p-values of both the BP and LB tests show that there is still significant autocorrelation in the residuals, at all usual significance levels (1, 5 and 10%). This can also be verified by observing the ACF and PACF of the residuals.

Regarding the model selection recurring to the minimization of the AICc, there are some general results that should be referred. Using this procedure, for all the cryptoassets, except for BCH, the chosen method was ETS(A, Ad, N), the one with the dampening trend. With respect to BCH, there are ambiguous results when we compare the forecast accuracy of each method, as error measures point out to contrasting decisions. Nevertheless, both MASE and MAE show smaller values in the case of the DES method, and even knowing the AICc is slightly lower for SES, this would not be sufficient to go for a non-trend approach to do our forecasting exercise, and the choice would fall on the trended one, as for all the remaining cryptocurrencies.

With this being said, and looking to all the components of the ETS(A, Ad, N), $\hat{\phi}$ is around the same for all the assets (≈ 0.8) and the estimation of the smoothing parameter of the level, $\hat{\alpha}$, is around 1. The only oddity is ETH, for which our estimates result in $\hat{\alpha} \approx 0.94$, indicating that the last observed value is still very relevant to predict future levels, but not entirely, as there is some weight left to other lags. With respect to β, the trend smoothing coefficient, the estimations lie in the interval [0.154, 0.446], with the extremes corresponding to BCH and ETH, respectively. The somewhat high values suggest that the trends change recurrently. Notwithstanding, as for BTC, Table 3 demonstrates there is statistical evidence supporting the alternative hypothesis of serial correlation in the residuals, with the exceptions of LTC, to which we almost have support of H_0 (at 1%), and ETH. For this crypto, the null hypothesis is not rejected at the 1% level, recurring to both tests, what is not a strong evidence.

Stepping into ARIMA modelling, like explained in section 3.1.5., we used an *R* function containing an algorithm which selects the best orders for the three components of the model. However, we will compare the most appropriate options to sense the data and understand if we would prefer a different approach. Fig. 8 and 9 show the ACF and PACF of BTC log returns. We know that if the suiting ARIMA model only has either an AR or a MA part, it is possible to try and identify the orders of the model by inspecting the autocorrelation plots of the series. The first three lags of BTC's ACF are exponentially decreasing, and only the first lag of the PACF is statistically significant, which is a symptom that an $\text{ARIMA}(1,1,0)$ can probably be a good choice to fit our data. After running the algorithm that compares all possibilities by minimizing AICc, the output returned this option with the addition of a drift was added. Table 5 reports the characteristics of this model, together with two possible options. By looking through them we verify that, and this serves to all, the constant does not appear to be relevant at the 10% significance level. However, we find that removing it deteriorates the AICc, what suggests the constant may be relevant to explain the variable. If we would begin our analysis with an $\text{ARIMA}(2,1,0)$ w/drift, we would certainly find that the second order AR coefficient is not significant. However, if we would begin by considering an $\text{ARIMA}(3,1,2)$ w/drift, we could be tempted to accept it, because even though there is evidence of non-relevance for the first MA coefficient (at 5 and 10% levels), this should stay in the model due to the second order component. Regardless, it is clear that our selection will rely on the ARIMA$(1,1,0)$ w/drift as it "beats" the others in all the information criteria. Also, the residuals pass the autocorrelation tests, and we have statistical support not to reject the null of no serial correlation.

Regarding LTC, the automatic procedure selected ARIMA$(1,1,0)$ as the best alternative. In fact, the ACF and PACF review allow to notice that this could be a plausible result. It is a similar model to BTC, but this time without a constant term. The autoregressive coefficient is statistically significant (p-value $= 0$), and around 0.336, whereas if the model chosen was an ARIMA$(2,1,0)$, the second AR component would be negative but not statistically relevant.

As per ETH, BCH and XRP cases, the alternatives are really distinct, as we are suggested to add MA components. Beginning with the first two assets, one option for both was to use an ARIMA$(1,1,5)$, with a strong moving average part, even though this is not straightforward when observing the sample AC functions. Yet, for ETH, the negative fifth order coefficient is not significant at any level, and for the fourth order coefficient, H_0 is only rejected at 5 and 10% criticalness. The removal of the MA(5) results in the significance of all coefficients, and we select the ARIMA$(1,1,4)$, with a strong negative AR(1) element (≈ -0.928). Meanwhile, for BCH log quotes modelling, having a MA(5) generates uncertainty with respect to the relevance of all MA coefficients, except for the first order. It turned out that the best model is an ARIMA$(2,1,3)$, even though we accept the relevance of the negative MA(1) only at 5% and 10% significance levels. Of substance to acknowledge the negativity of AR(2) too. Concerning XRP, we will choose the ARIMA$(0,1,1)$, which has a MA coefficient of approximately 0.414. All the other alternatives do not show to be appropriate. For the rest, for all cryptocurrencies, the resulting residuals passed the BP and LB tests and as the null hypothesis was not rejected, indicating the absence of autocorrelation in the residuals of each considered model.

Now that we covered the model selection for both ETS and ARIMA, we will present the model comparison results, by measuring one-step ahead forecasting accuracy. This cross-validation is done in a rolling window manner. A RW and RW w/drift will also join the forecasting competition, so we can verify if there is any gain in applying our methods, instead of considering a *naïve* model for the quotes. The error measures are the RMSE, MAE and MASE, but special relevance is given to the first.

As usual we begin with the most representative crypto asset, BTC. Table 6 shows the comparison for all assets. The results for all measures are unequivocal, and evidence that the most precise method is the ARIMA$(1,1,0)$ w/drift. As seen before, the ETS model residuals show evidence of autocorrelation through the tests, and thus this would not be a valid option. Anyhow, ARIMA model of BTC log quotes appears to have superior properties not only compared to ETS, but also to any RW model. The 5-step forecast was produced and the plots can be reviewed in Fig. 14. Using a confidence of 80%, we expect that the quote will lie in the interval $[10{,}267; 12{,}900[$ USD and the forecast mean shows a gradual increase, after a small drop at the one-step forecast. The volatility of the asset price is noticeable by the amplitude of the produced interval forecast.

With respect to ETH, if we consider RMSE as the most important decision factor, the ARIMA alternative is once more our choice. We eventually could take into consideration that the ETS and RW options show lower MAE and MASE values, but we easily come to realize this is not a good choice. As we know, the results regarding the residuals' autocorrelation of the ETS do not strongly suggest their absence and, furthermore, the ACF plot of RW's residuals show a smooth decay and three significant spikes at the first lags. The ARIMA$(1,1,4)$ is, therefore, the best option to model the log quote of ETH. After forecasting, the statistics point out that the expectation is that the price will decrease by around 18 USD at the first two weeks and then stabilize at approximately 379 USD.

The results of LTC and XRP are the less ambiguous. Beginning with the first, all error measures point ARIMA$(1,1,0)$ forecast as the most precise and by running the prediction we find a similar result to ETH, with a small downward aptitude in the line through the five-week period. After these weeks the price is not expected to be out of the [24.81; 48,50[range, with a 95% confidence. As far as XRP log quote is concerned, the selected model is again the ARIMA, in this case with a MA(1) part instead of the AR(1). Taking the forecast values in consideration, we do not expect that after five weeks the price of XRP will exceed 0.71 USD or shrink below 0.10 USD. The mean is around 0.27 USD during the forecasting timespan.

At last, BCH forecast competition show evidence that either the RW (supported only by RMSE) or the ETS(A, Ad, N) should be the most accurate when forecasting the log price for the next period. However, as stated previously, the exponential smoothing residuals exhibit signs of autocorrelation, and as presented in Fig. 13, RW "residuals" too. The next most precise methods would be the ETS(A, N, N) and the RW w/drift, but their residuals suffer from the same intrarelationship. Thus, such as for the remaining the assets, an ARIMA will be used to predict the log price of BCH, in this case with AR and MA components of orders 2 and 3, respectively, differentiated once to achieve stationarity. This prediction result is exceptional comparing to the other ones. The price is estimated to drop until reaching around 253 USD, by the third week, and then rise to approximately 266 USD in the fifth week. This value is still below the last observed price.

5. CONCLUSION

The cryptocurrency market is one of the most recent among the financial assets' markets. Consequently, the great majority of investigators and researchers that share curiosity – and that have the purpose of learning more about this market – always take into consideration that there is a short variety of cryptoassets with a sufficiently large historical data timespan. Most of them did not even exist in a five-year-ago past. Due mostly to this fact and to its representativity, Bitcoin is the crypto for which more studies can be found. However, some of the newer assets have different behaviours and features. Some were even created with the objective of correcting drawbacks of BTC itself. Thus, one must not generalize the whole market only by investigating one asset, even if its market dominance is between 60 and 70%. This was the head motivation for this research – to enrich the study about some of the most relevant cryptocurrencies.

The goal was to analyse and predict price and returns of five assets: Bitcoin, Ethereum, Litecoin, XRP and Bitcoin Cash – a brief introduction to the currencies and its historical evolution was done. For this purpose, a forecast competition between Exponential Smoothing and ARIMA models was performed, to select the best option to fit data and do a five-step ahead forecast exercise. Regarding ETS models, three possibilities (SES, TES and DES) were considered, and error measurements and information criteria were used to choose the best fitting ETS model. A similar procedure was done with ARIMA models: we analysed the goodness-of-fit of three ARIMA of different autoregressive and moving average orders. In the end, the forecast comparison was computed between ARIMA, ETS and Random Walk predictions, that were also added to supplement the analysis.

Having this ultimate goal in mind, we started by analysing if the crypto assets under research fulfil some of the financial series stylized facts. Namely, non-normality, serial correlation and non-stationarity. The performed tests defended these three facts, what supported the models that were used through the study. Depending on the ARIMA orders chosen and ETS model type, these two alternatives can be capable of accommodating for explosive behaviours and also for autocorrelation between the target variable. In general, they can prove to be good options for financial time series modelling and prediction.

Moreover, a step by step manual of the *R* script that was used is provided, from data collection until prediction part. The procedure followed the framework provided in Hyndman & Athanasopoulos (2019), added to the needed adaptations and adjustments to our data and to the specific field of cryptocurrencies.

The main conclusion about this investigation is that ARIMA models outperform ETS in explaining price behaviours of all the assets. This is due the lower error measurements of the one-step ahead forecasts, but also to the signs of residual autocorrelation in all the ETS model types (weak evidence only for ETH and LTC). Within the ARIMA models selected, the cryptocurrency for which we observed a lower RMSE (around 9%) was BTC. LTC was the one with the second lower value, and this fact seems to be associated with the period of analysis of the assets, since these were two of the assets for which we had larger samples.

For further investigation, it would be interesting to extend the scope and study the behaviour of more recent crypto assets. As we know, nowadays the number of cryptos in circulation grows almost overnight, and some of the newer assets are created with different characteristics and innovating mechanisms that intend to prevent, for example, high volatility, which is one of the disadvantages of these assets for investors. This will also, probably, provoke that different kinds of models should show better results, depending on each asset features.

Furthermore, to achieve better results it is important that historical data periods are larger. Also, different approaches can be done with respect to the data preparation and use. Instead of weekly average prices, using daily data would be another possibility. ARIMA modelling appears to show reasonable results for cryptocurrency price prediction, but many studies are now combining the use of these with machine learning techniques like Neural Networks, which can show to be better in some cases and for some assets.

References

Alahmari, S. A. (2019). Using Machine Learning ARIMA to Predict the Price of Cryptocurrencies. *ISeCure-The ISC International Journal of Information Security*, *11*(3), 139-144.

Bakar, N. A., & Rosbi, S. (2017). Autoregressive integrated moving average (ARIMA) model for forecasting cryptocurrency exchange rate in high volatility environment: A new insight of bitcoin transaction. *International Journal of Advanced Engineering Research and Science*, *4*(11), 237311.

Bank of England (2020). *What are cryptoassets (cryptocurrencies)?* [Online]. Available at: https://www.bankofengland.co.uk/knowledgebank/what-are-cryptocurrencies

Bariviera, A. F. (2017). The inefficiency of Bitcoin revisited: A dynamic approach. *Economics Letters*, *161*, 1-4.

Bariviera, A. F., Basgall, M. J., Hasperué, W., & Naiouf, M. (2017). Some stylized facts of the Bitcoin market. *Physica A: Statistical Mechanics and its Applications*, *484*, 82-90.

Baur, D. G., Hong, K., & Lee, A. D. (2018). Bitcoin: Medium of exchange or speculative assets?. *Journal of International Financial Markets, Institutions and Money*, *54*, 177-189.

Bohte, R., & Rossini, L. (2019). Comparing the forecasting of cryptocurrencies by Bayesian time-varying volatility models. *Journal of Risk and Financial Management*, *12*(3), 150.

Bouveret, A., & Haksar, V. (2018). What Are Cryptocurrencies. *Finance and Development*, *55*(2), 26-29.

Box, G. E., & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, *65*(332), 1509-1526.

Buterin, V. (2014). A next-generation smart contract and decentralized application platform. *white paper*, 3(37).

CoinMarketCap (2019). Crypto-Currency Market Capitalizations [Online]. Available at: https://coinmarketcap.com.

Catania, L., Grassi, S., & Ravazzolo, F. (2019). Forecasting cryptocurrencies under model and parameter instability. *International Journal of Forecasting*, *35*(2), 485-501.

Chaum, D. L. (1979). *Computer Systems established, maintained and trusted by mutually suspicious groups*. Electronics Research Laboratory, University of California.

Cheah, E. T., Mishra, T., Parhi, M., & Zhang, Z. (2018). Long memory interdependency and inefficiency in Bitcoin markets. *Economics Letters*, *167*, 18-25.

Chiu, J., & Koeppl, T. V. (2017). The economics of cryptocurrencies–bitcoin and beyond. *Available at SSRN 3048124*.

Comtois, D. (2020). summarytools: Tools to Quickly and Neatly Summarize Data. R package version 0.9.6. https://CRAN.R-project.org/package=summarytools

European Central Bank (2019). *Crypto-assets – trends and implications* [Online]. Available at: https://www.ecb.europa.eu/paym/intro/mip-online/2019/html/1906_crypto_assets.en.html

Gardner Jr, E. S., & McKenzie, E. D. (1985). Forecasting trends in time series. *Management science*, *31*(10), 1237-1246.

Holt, Charles C. (1957). Forecasting Trends and Seasonal by Exponentially Weighted Averages. *Office of Naval Research Memorandum*.

Hotz-Behofsits, C., Huber, F., & Zörner, T. O. (2018). Predicting crypto-currencies using sparse non-Gaussian state space models. *Journal of Forecasting*, *37*(6), 627-640.

Hu, Y., Valera, H. G. A., & Oxley, L. (2019). Market efficiency of the top market-cap cryptocurrencies: Further evidence from a panel framework. *Finance Research Letters*, *31*, 138-145.

Hyndman, R. J. (2020). fpp3: Data for "Forecasting: Principles and Practice" (3rd Edition). R package version 0.3. https://CRAN.R-project.org/package=fpp3

Hyndman, R.J., & Athanasopoulos, G. (2019). *Forecasting: principles and practice*, 3rd Ed. Melbourne, Australia: OTexts [Online]. Available at: https://otexts.com/fpp3/

International Monetary Fund (2019). *Treatment of Crypto Assets in Macroeconomic Statistics* [Online]. Available at: https://www.imf.org/external/pubs/ft/bop/2019/pdf/Clarification0422.pdf

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, *6*(3), 255-259.

Jevons, W. S. (1875). *Money and the Mechanism of Exchange*. New York: D. Appleton and Co.

Kristoufek, L. (2018). On Bitcoin markets (in) efficiency and its evolution. *Physica A: Statistical Mechanics and its Applications*, *503*, 257-262.

Kumar, S. (2019). Forecasting Cryptocurrency Prices using ARIMA and Neural Network: A Comparative Study. *Journal of Prediction Markets*, *13*(2).

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?. *Journal of Econometrics*, *54*(1-3), 159-178.

Ljung, G. M., & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297-303.

López-de-Lacalle, J. (2019). tsoutliers: Detection of Outliers in Time Series. R package version 0.6-8. https://CRAN.R-project.org/package=tsoutliers

Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System* [Online]. Available at: https://bitcoin.org/bitcoin.pdf

Rebane, J., Karlsson, I., Denic, S., & Papapetrou, P. (2018). Seq2seq rnns and arima models for cryptocurrency prediction: A comparative study. *SIGKDD Fintech*, *18,* 2-6.

Vent, J. (2020). crypto: Cryptocurrency Market Data. https://github.com/JesseVent/crypto

Wickham, H., & Grolemund, G. (2017). R for Data Science. Canada.

Wickham, H., Hester, J. & Chang, W. (2020). devtools: Tools to Make Developing R Packages Easier. R package version 2.3.2. https://CRAN.R-project.org/package=devtools

Appendix A – Data and Results

i. Figures

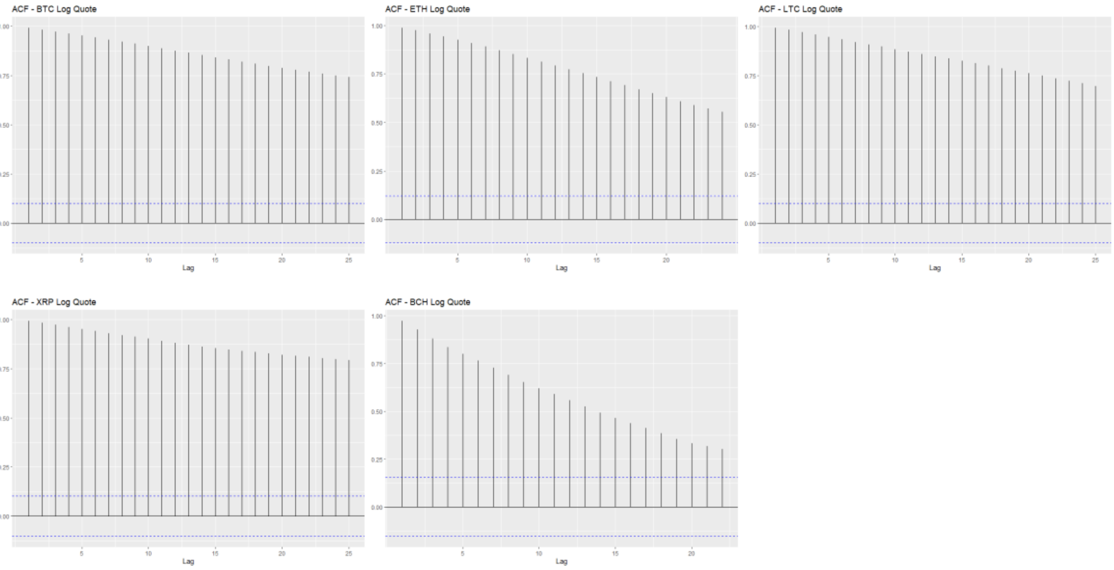

Figure 5 - ACF's of the logarithmic quotes of each cryptocurrency. The dashed lines represent the limits above/below which the coefficients are statistically significant.

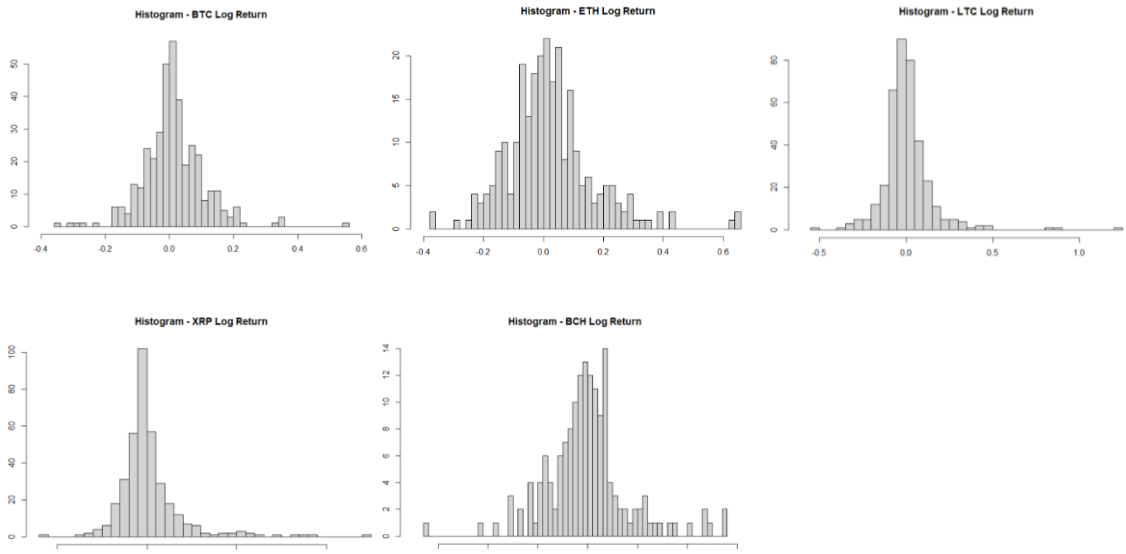

Figure 6 - Histograms of the logarithmic returns of each cryptocurrency. The analysis periods correspond to the periods announced through Section 3.

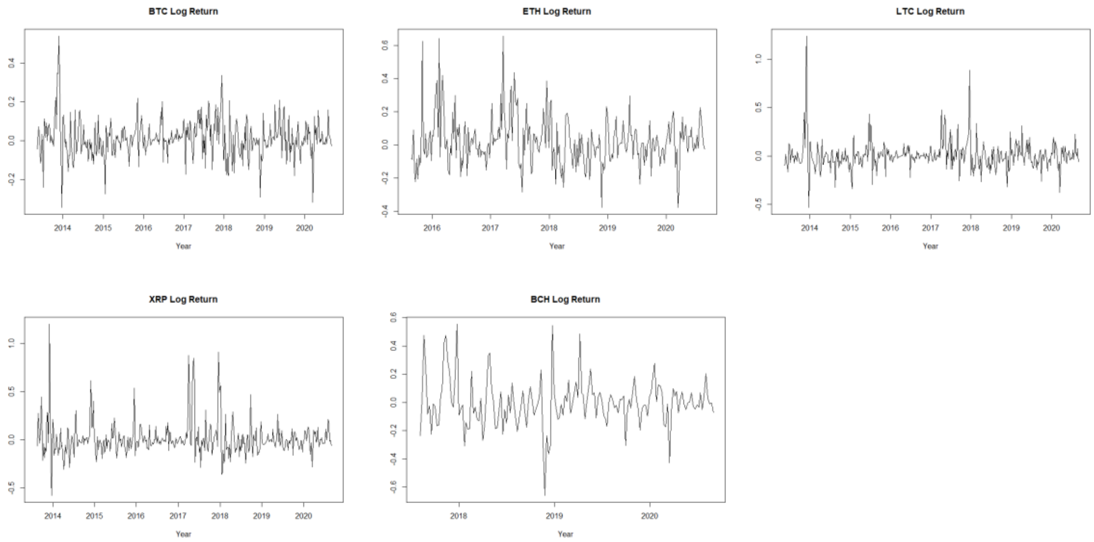

Figure 7 - Time-series of the logarithmic returns of each cryptocurrency.

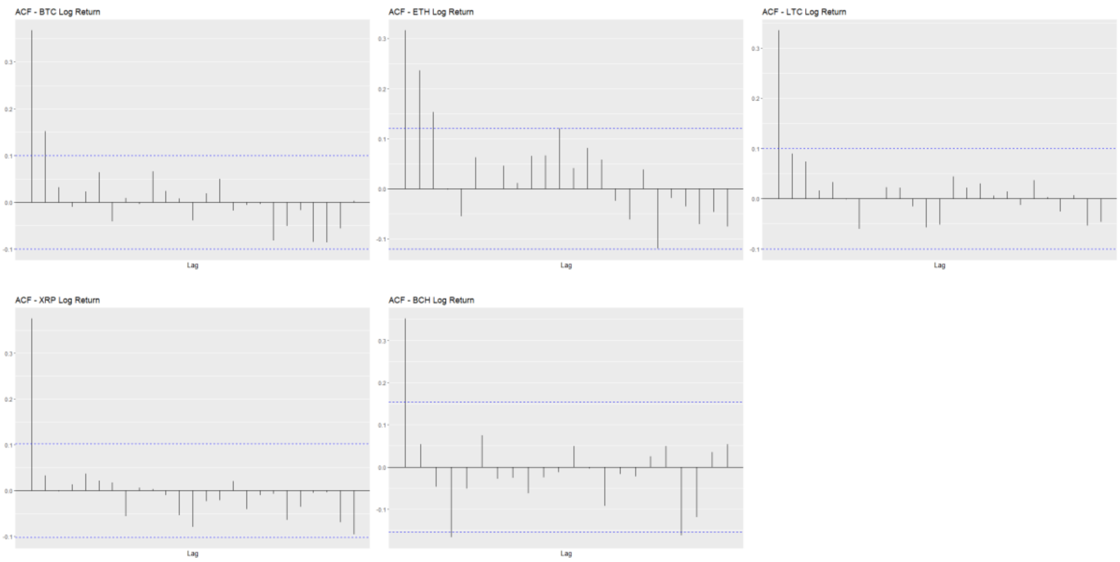

Figure 8 - ACF's of the logarithmic returns of each cryptocurrency.

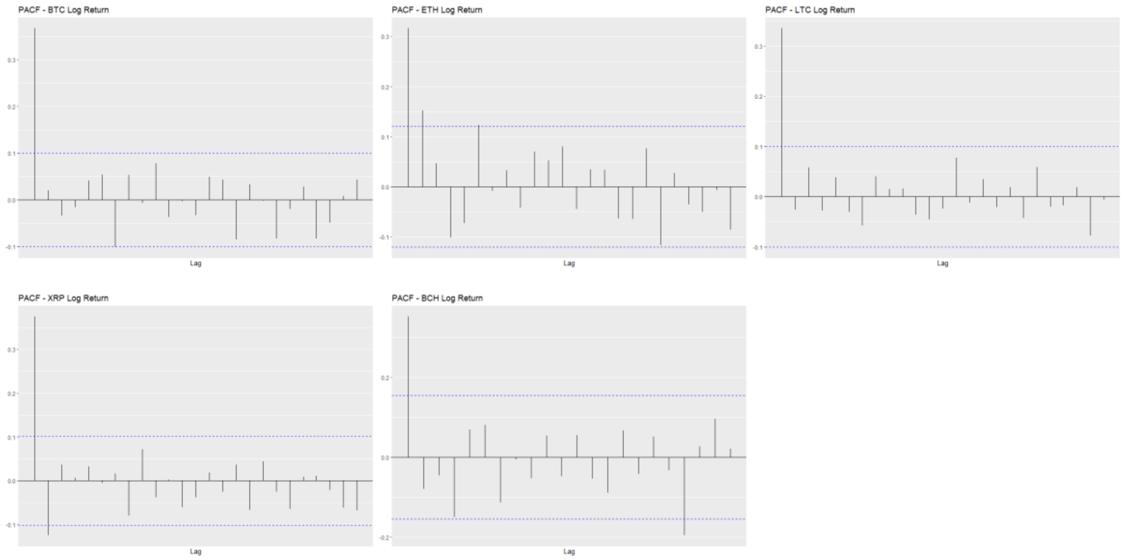

Figure 9 - PACF's of the logarithmic returns of each cryptocurrency.

```
> BTC.weekly %>% features(log_return, box_pierce, lag = 10)
# A tibble: 1 x 2
  bp_stat bp_pvalue
    <dbl>     <dbl>
1    64.7  4.61e-10
> BTC.weekly %>% features(log_return, ljung_box, lag = 10)
# A tibble: 1 x 2
  lb_stat lb_pvalue
    <dbl>     <dbl>
1    65.3  3.53e-10
```

```
> ETH.weekly %>% features(log_return, box_pierce, lag = 10)
# A tibble: 1 x 2
  bp_stat    bp_pvalue
    <dbl>        <dbl>
1    50.5 0.000000213
> ETH.weekly %>% features(log_return, ljung_box, lag = 10)
# A tibble: 1 x 2
  lb_stat    lb_pvalue
    <dbl>        <dbl>
1    51.3 0.000000153
```

```
> LTC.weekly %>% features(log_return, box_pierce, lag = 10)
# A tibble: 1 x 2
  bp_stat    bp_pvalue
    <dbl>        <dbl>
1    50.2 0.000000240
> LTC.weekly %>% features(log_return, ljung_box, lag = 10)
# A tibble: 1 x 2
  lb_stat    lb_pvalue
    <dbl>        <dbl>
1    50.7 0.000000199
```

```
> XRP.weekly %>% features(log_return, box_pierce, lag = 10)
# A tibble: 1 x 2
  bp_stat     bp_pvalue
    <dbl>         <dbl>
1    54.1 0.0000000464
> XRP.weekly %>% features(log_return, ljung_box, lag = 10)
# A tibble: 1 x 2
  lb_stat     lb_pvalue
    <dbl>         <dbl>
1    54.6 0.0000000378
```

```
> BCH.weekly %>% features(log_return, box_pierce, lag = 10)
# A tibble: 1 x 2
  bp_stat bp_pvalue
    <dbl>     <dbl>
1    27.4   0.00222
> BCH.weekly %>% features(log_return, ljung_box, lag = 10)
# A tibble: 1 x 2
  lb_stat lb_pvalue
    <dbl>     <dbl>
1    28.1   0.00171
```

Figure 10 - BP and LB test statistics, applied to each logarithmic return.

```
        Jarque Bera Test

data:  BTC.weekly$log_return
X-squared = 258.42, df = 2, p-value < 0.00000000000000022


        Skewness

data:  BTC.weekly$log_return
statistic = 0.54406, p-value = 0.00001418


        Kurtosis

data:  BTC.weekly$log_return
statistic = 6.8796, p-value < 0.00000000000000022
```

```
        Jarque Bera Test

data:  ETH.weekly$log_return
X-squared = 135.79, df = 2, p-value < 0.00000000000000022


        Skewness

data:  ETH.weekly$log_return
statistic = 0.99713, p-value = 0.00000000004065


        Kurtosis

data:  ETH.weekly$log_return
statistic = 5.9007, p-value < 0.00000000000000022
```

```
        Jarque Bera Test

data:  LTC.weekly$log_return
X-squared = 5659.3, df = 2, p-value < 0.00000000000000022


        Skewness

data:  LTC.weekly$log_return
statistic = 2.6209, p-value < 0.00000000000000022


        Kurtosis

data:  LTC.weekly$log_return
statistic = 21.113, p-value < 0.00000000000000022
```

```
        Jarque Bera Test

data:  XRP.weekly$log_return
X-squared = 2161.4, df = 2, p-value < 0.00000000000000022


        Skewness

data:  XRP.weekly$log_return
statistic = 2.4251, p-value < 0.00000000000000022


        Kurtosis

data:  XRP.weekly$log_return
statistic = 13.837, p-value < 0.00000000000000022
```

```
        Jarque Bera Test

data:  BCH.weekly$log_return
X-squared = 49.447, df = 2, p-value = 0.00000000001831


        Skewness

data:  BCH.weekly$log_return
statistic = 0.39887, p-value = 0.03881


        Kurtosis

data:  BCH.weekly$log_return
statistic = 5.5951, p-value = 0.00000000001799
```

```
> JarqueBera.test(BCH.daily$log_return)

        Jarque Bera Test

data:  BCH.daily$log_return
X-squared = 4585, df = 2, p-value < 0.00000000000000022


        Skewness

data:  BCH.daily$log_return
statistic = 0.25143, p-value = 0.0005503


        Kurtosis

data:  BCH.daily$log_return
statistic = 12.842, p-value < 0.00000000000000022
```

Figure 11 - JB test statistics of all cryptocurrencies' logarithmic returns. JB test statistic to the daily logarithmic returns of BCH (bottom right).

```
> # - Box-Pierce / Ljung-Box
> BTC.fit.ses.aug %>% features(.resid, box_pierce, lag = 10, dof = 1)
# A tibble: 1 x 3
  .model          bp_stat bp_pvalue
  <chr>             <dbl>     <dbl>
1 SES - ETS(A,N,N)   65.0  1.45e-10
> BTC.fit.ses.aug %>% features(.resid, ljung_box, lag = 10, dof = 1)
# A tibble: 1 x 3
  .model          lb_stat lb_pvalue
  <chr>             <dbl>     <dbl>
1 SES - ETS(A,N,N)   65.6  1.10e-10
```

```
> ETH.fit.ses.aug %>% features(.resid, box_pierce, lag = 10, dof = 1)
# A tibble: 1 x 3
  .model          bp_stat    bp_pvalue
  <chr>             <dbl>        <dbl>
1 SES - ETS(A,N,N)   50.8 0.0000000752
> ETH.fit.ses.aug %>% features(.resid, ljung_box, lag = 10, dof = 1)
# A tibble: 1 x 3
  .model          lb_stat    lb_pvalue
  <chr>             <dbl>        <dbl>
1 SES - ETS(A,N,N)   51.6 0.0000000537
```

```
> # Box-Pierce / Ljung-Box
> LTC.fit.ses.aug %>% features(.resid, box_pierce, lag = 10, dof = 1)
# A tibble: 1 x 3
  .model          bp_stat    bp_pvalue
  <chr>             <dbl>        <dbl>
1 SES - ETS(A,N,N)   50.4 0.0000000890
> LTC.fit.ses.aug %>% features(.resid, ljung_box, lag = 10, dof = 1)
# A tibble: 1 x 3
  .model          lb_stat    lb_pvalue
  <chr>             <dbl>        <dbl>
1 SES - ETS(A,N,N)   50.9 0.0000000731
```

```
> XRP.fit.ses.aug %>% features(.resid, box_pierce, lag = 10, dof = 1)
# A tibble: 1 x 3
  .model          bp_stat    bp_pvalue
  <chr>             <dbl>        <dbl>
1 SES - ETS(A,N,N)   54.3 0.0000000168
> XRP.fit.ses.aug %>% features(.resid, ljung_box, lag = 10, dof = 1)
# A tibble: 1 x 3
  .model          lb_stat    lb_pvalue
  <chr>             <dbl>        <dbl>
1 SES - ETS(A,N,N)   54.7 0.0000000136
```

```
> BCH.fit.ses.aug %>% features(.resid, box_pierce, lag = 10, dof = 1)
# A tibble: 1 x 3
  .model          bp_stat bp_pvalue
  <chr>             <dbl>     <dbl>
1 SES - ETS(A,N,N)   27.6   0.00111
> BCH.fit.ses.aug %>% features(.resid, ljung_box, lag = 10, dof = 1)
# A tibble: 1 x 3
  .model          lb_stat lb_pvalue
  <chr>             <dbl>     <dbl>
1 SES - ETS(A,N,N)   28.3  0.000850
```

Figure 12 - BP and LB test statistics, applied to the residuals of the ETS(A,N,N) ~ Log Quote.

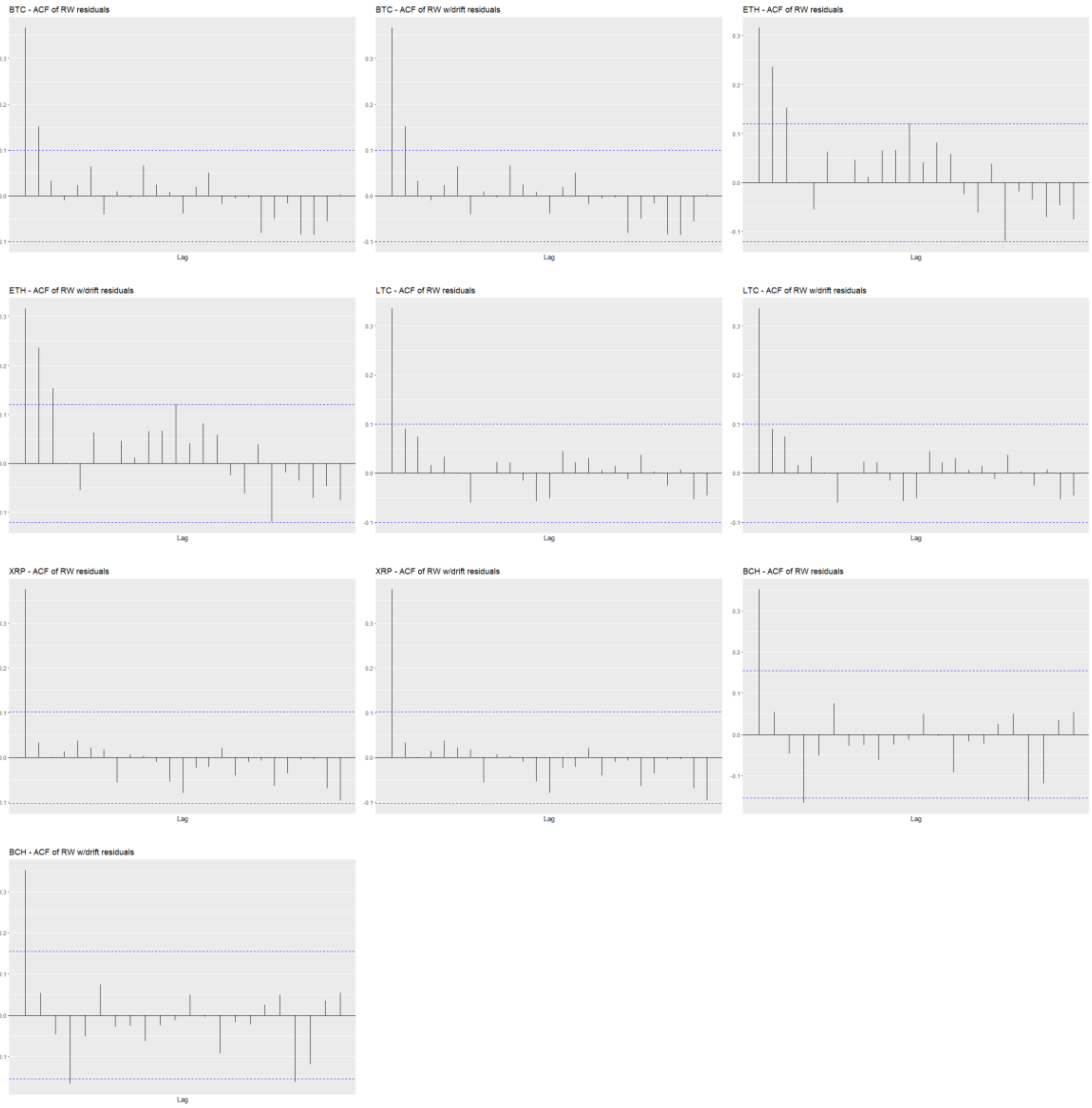

Figure 13 - ACF's of the residuals of RW and RW w/drift models, of each cryptocurrency.

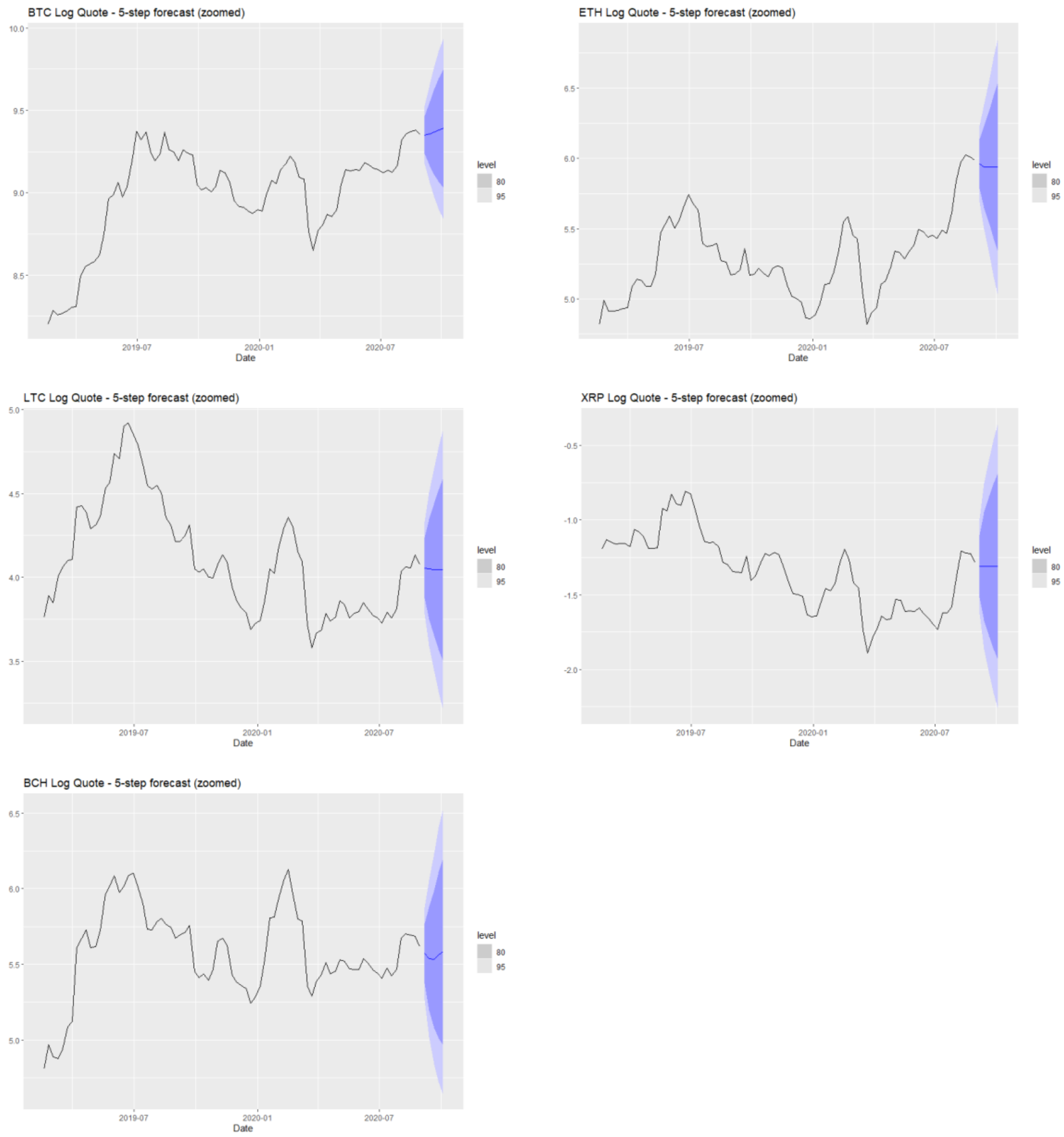

Figure 14 - 5-step ahead forecast of the logarithmic quote of each cryptocurrency.

ii. Tables

	BTC	ETH	LTC	XRP	BCH
Mean	0.01197630	0.007240315	0.007240315	0.01115366	-0.002181586
Std.Dev	0.09563647	0.147831585	0.147831585	0.17757359	0.169612678
Min	-0.34342348	-0.532739843	-0.532739843	-0.57839562	-0.657934417
Max	0.54035928	1.241510724	1.241510724	1.20234548	0.554308079
Skewness	0.54192323	2.610631501	2.610631501	2.41521600	0.395163105
Kurtosis	3.84366637	18.002655053	18.002655053	10.76186312	2.525815224

Table 1 - Summary table with descriptive statistics of all cryptocurrencies' logarithmic returns.

CRYPTO	kpss_stat	kpss_pvalue	ndiffs	nsdiffs
BTC	5.626468	0.01	1	0
ETH	2.853838	0.01	1	0
LTC	4.241672	0.01	1	0
XRP	4.669113	0.01	1	0
BCH	1.627812	0.01	1	0

kpss_stat	kpss_pvalue
0.07322650	0.1
0.26020047	0.1
0.07424031	0.1
0.07808082	0.1
0.07262732	0.1

Table 2 - Stationarity tests. KPSS test statistic and p-value, number of unit roots and number of seasonal differencing required of log quotes (left), and KPSS test statistic. and p-value of log returns.

Bitcoin	alpha	l	beta	b	phi	sigma2	log_lik	AIC	AICc	BIC	bp_pvalue	lb_pvalue
SES - ETS(A,N,N)	0.9999	4.794487	NA	NA	NA	0.009291574	-242.0851	490.1702	490.2336	502.0143	0.0000000001450882	0.0000000001103994
Holt - ETS(A,A,N)	0.9999	4.667851	0.1297648	0.008090524	NA	0.009154819	-238.2377	486.4755	486.6346	506.2157	0.0000001617936699	0.0000001248535768
Damped - ETS(A,Ad,N)	0.9999	4.648742	0.2056185	-0.113339584	0.8007774	0.008604349	-225.8564	463.7128	463.9362	487.4010	0.0000972896557964	0.0000814944434718

Ether	alpha	l	beta	b	phi	sigma2	log_lik	AIC	AICc	BIC	bp_pvalue	lb_pvalue
SES - ETS(A,N,N)	0.9998996	0.3442951	NA	NA	NA	0.02246377	-233.9691	473.9382	474.0305	484.6660	0.0000000751662	0.00000005371026
Holt - ETS(A,A,N)	0.9999000	0.2824913	0.1522886	-0.028887971	NA	0.02145713	-226.9058	463.8116	464.0441	481.6913	0.0001465645101	0.00011023292753
Damped - ETS(A,Ad,N)	0.9384050	0.3850127	0.4462825	0.002412589	0.8000001	0.02005306	-217.4640	446.9280	447.2548	468.3837	0.0462248996282	0.04046024426273

Litecoin	alpha	l	beta	b	phi	sigma2	log_lik	AIC	AICc	BIC	bp_pvalue	lb_pvalue
SES - ETS(A,N,N)	0.9998999	1.311389	NA	NA	NA	0.02190822	-406.3450	818.6899	818.7532	830.5340	0.00000008895229	0.00000007311999
Holt - ETS(A,A,N)	0.9999000	1.208794	0.07557772	-0.003097614	NA	0.02207579	-406.7962	823.5925	823.7516	843.3326	0.00000604982229	0.00000507368644
Damped - ETS(A,Ad,N)	0.9998999	1.192359	0.32085427	0.095060486	0.8	0.02045450	-391.6830	795.3659	795.5893	819.0542	0.01084681576834	0.00982655253177

XRP	alpha	l	beta	b	phi	sigma2	log_lik	AIC	AICc	BIC	bp_pvalue	lb_pvalue
SES - ETS(A,N,N)	0.9998999	-5.388350	NA	NA	NA	0.03165951	-452.5130	911.0261	911.0918	922.7585	0.00000001678806	0.00000001362322
Holt - ETS(A,A,N)	0.9999000	-5.257464	0.1523665	0.08906132	NA	0.03234812	-455.4748	920.9496	921.1149	940.5036	0.00000041021169	0.00000032980381
Damped - ETS(A,Ad,N)	0.9998999	-5.837190	0.3110547	0.34679944	0.8000001	0.02978553	-439.7412	891.4824	891.7144	914.9472	0.00001727251864	0.00001438569967

BitcoinCash	alpha	l	beta	b	phi	sigma2	log_lik	AIC	AICc	BIC	bp_pvalue	lb_pvalue
SES - ETS(A,N,N)	0.9999000	5.968125	NA	NA	NA	0.02877527	-123.6817	253.3634	253.5153	262.6262	0.0011121108	0.0008495224
Holt - ETS(A,A,N)	0.9999000	5.905339	0.0001000573	-0.002825265	NA	0.02916520	-123.7531	257.5062	257.8908	272.9442	0.0007102439	0.0005400620
Damped - ETS(A,Ad,N)	0.9997874	5.961012	0.1544319290	-0.095790483	0.8	0.02836192	-120.9766	253.9532	254.4951	272.4787	0.0011050414	0.0008142534

Table 3 - SES, TES and DES model comparison, applied to logarithmic quotes of each cryptocurrency. Includes BP and LB test p-values of the residuals.

BTC	.type	RMSE	MAE	MASE
Damped - ETS(A,Ad,N)	Test	0.09261854	0.06542939	0.9734945
Holt - ETS(A,A,N)	Test	0.09889644	0.06913391	1.0286124
SES - ETS(A,N,N)	Test	0.09609355	0.06689785	0.9953431

ETH	.type	RMSE	MAE	MASE
Damped - ETS(A,Ad,N)	Test	0.1486929	0.1056597	0.9982839
Holt - ETS(A,A,N)	Test	0.1570050	0.1108057	1.0469043
SES - ETS(A,N,N)	Test	0.1500891	0.1052212	0.9941415

LTC	.type	RMSE	MAE	MASE
Damped - ETS(A,Ad,N)	Test	0.1614849	0.09527368	1.0724328
Holt - ETS(A,A,N)	Test	0.1667055	0.09954685	1.1205331
SES - ETS(A,N,N)	Test	0.1487852	0.08876825	0.9992055

XRP	.type	RMSE	MAE	MASE
Damped - ETS(A,Ad,N)	Test	0.1750729	0.1057544	1.0033728
Holt - ETS(A,A,N)	Test	0.1848090	0.1138641	1.0803157
SES - ETS(A,N,N)	Test	0.1769708	0.1041038	0.9877122

BCH	.type	RMSE	MAE	MASE
Damped - ETS(A,Ad,N)	Test	0.1691803	0.1140646	0.9751139
Holt - ETS(A,A,N)	Test	0.1775670	0.1226441	1.0484581
SES - ETS(A,N,N)	Test	0.1663277	0.1150854	0.9838404

Table 4 - ETS model comparison, using error measurements, applied to logarithmic quotes of each cryptocurrency.

Bitcoin	constant	ar1	ar2	ar3	ma1	ma2	AICc	bp_pvalue	lb_pvalue
ARIMA(3,1,2) w/drift	0.01491 (0.10274)	0.28607 (0)	-0.87211 (0)	0.3208 (0)	0.06848 (0.02688)	0.96256 (0)	-758.8633	0.1871563	0.1763426
ARIMA(2,1,0) w/drift	0.00735 (0.10592)	0.35923 (0)	0.0193 (0.70578)	NA	NA	NA	-757.3862	0.2003934	0.1865253
ARIMA(1,1,0) w/drift	0.0075 (0.09935)	0.36639 (0)	NA	NA	NA	NA	-759.2861	0.2572056	0.2407545

Ether	ar1	ma1	ma2	ma3	ma4	ma5	AICc	bp_pvalue	lb_pvalue
ARIMA(1,1,5)	-0.92301 (0)	1.20591 (0)	0.51391 (0)	0.45042 (0.00001)	0.20289 (0.03198)	-0.04201 (0.5065)	-286.2130	0.9234256	0.9194952
ARIMA(1,1,4)	-0.92775 (0)	1.22361 (0)	0.53624 (0)	0.46887 (0.00001)	0.25179 (0.00004)	NA	-287.8828	0.9243808	0.9202364
ARIMA(0,1,3)	NA	0.2602 (0.00001)	0.26635 (0.00003)	0.20065 (0.00074)	NA	NA	-286.8740	0.5674538	0.5452076

Litecoin	ar1	ar2	ma1	AICc	bp_pvalue	lb_pvalue
ARIMA(2,1,0)	0.34443 (0)	-0.02437 (0.63363)	NA	-416.5848	0.7808980	0.7709098
ARIMA(1,1,0)	0.33621 (0)	NA	NA	-418.3892	0.8301471	0.8213706
ARIMA(0,1,1)	NA	NA	0.33644 (0)	-417.2389	0.7623413	0.7526075

XRP	ar1	ar2	ma1	ma2	AICc	bp_pvalue	lb_pvalue
ARIMA(2,1,0)	0.42379 (0)	-0.12435 (0.01693)	NA	NA	-283.3359	0.7133634	0.6980654
ARIMA(0,1,2)	NA	NA	0.42865 (0)	0.04099 (0.45084)	-283.8777	0.7150981	0.6992278
ARIMA(0,1,1)	NA	NA	0.41376 (0)	NA	-285.3451	0.7326693	0.7172440

BitcoinCash	ar1	ar2	ma1	ma2	ma3	ma4	ma5	AICc	bp_pvalue	lb_pvalue
ARIMA(1,1,5)	-0.60713 (0.00185)	NA	1.00891 (0)	0.34455 (0.01055)	0.07358 (0.51481)	-0.18772 (0.09998)	-0.26402 (0.00185)	-133.0246	0.9705098	0.9670995
ARIMA(2,1,3)	0.71717 (0)	-0.87859 (0)	-0.33917 (0.02409)	0.75267 (0)	0.2535 (0.00738)	NA	NA	-133.8126	0.3876510	0.3570424
ARIMA(0,1,1)	NA	NA	0.36435 (0)	NA	NA	NA	NA	-133.0761	0.5112372	0.4740724

Table 5 - ARIMA model comparison, applied to logarithmic quotes of each cryptocurrency. Includes BP and LB test p-values of the residuals.

Bitcoin	.type	RMSE	MAE	MASE
ARIMA(1,1,0) w/ drift	Test	0.09116333	0.06471128	0.9612550
ETS(A,Ad,N)	Test	0.09261854	0.06542939	0.9719221
RW	Test	0.09607943	0.06688119	0.9934879
RW w/drift	Test	0.09629871	0.06849643	1.0174815

Ether	.type	RMSE	MAE	MASE
ARIMA(1,1,4)	Test	0.1464123	0.1064643	1.0029179
ETS(A,Ad,N)	Test	0.1486929	0.1056597	0.9953383
RW	Test	0.1500842	0.1052180	0.9911778
RW w/drift	Test	0.1514440	0.1084653	1.0217679

Litecoin	.type	RMSE	MAE	MASE
ARIMA(1,1,0)	Test	0.1374919	0.08812872	0.9911475
ETS(A,Ad,N)	Test	0.1614849	0.09527368	1.0715039
RW	Test	0.1487802	0.08876496	0.9983031
RW w/drift	Test	0.1498962	0.09094671	1.0228404

XRP	.type	RMSE	MAE	MASE
ARIMA(0,1,1)	Test	0.1639814	0.09701921	0.9193576
ETS(A,Ad,N)	Test	0.1750729	0.10575438	1.0021324
RW	Test	0.1769640	0.10410001	0.9864555
RW w/drift	Test	0.1787065	0.10861439	1.0292340

BCH	.type	RMSE	MAE	MASE
ARIMA(2,1,3)	Test	0.1732803	0.1236549	1.0544739
ETS(A,Ad,N)	Test	0.1691803	0.1140646	0.9726921
ETS(A,N,N)	Test	0.1663277	0.1150854	0.9813969
RW	Test	0.1663217	0.1150797	0.9813486
RW w/drift	Test	0.1685351	0.1193431	1.0177047

Table 6 - Model comparison, using error measurements, applied to logarithmic quotes of each cryptocurrency.

Appendix B – *R* Code Repository

l45618/CryptoPriceForecasting.R at main · rmoreiraserra/l45618 (github.com)