# BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

## FACULTY OF MECHANICAL ENGINEERING

FAKULTA STROJNÍHO INŽENÝRSTVÍ

## INSTITUTE OF MATHEMATICS

ÚSTAV MATEMATIKY

# ANALYZE AND ECONOMIC TIME SERIES FORECASTING BY USING SELECTED STATISTICAL METHODS

ANALÝZA A PŘEDPOVĚĎ EKONOMICKÝCH ČASOVÝCH ŘAD POMOCÍ VYBRANÝCH STATISTICKÝCH METOD

**MASTER'S THESIS**

DIPLOMOVÁ PRÁCE


**AUTHOR** Bc. Martin Skopal

AUTOR PRÁCE

**SUPERVISOR** Ing. Tomáš Mauder, Ph.D.

VEDOUCÍ PRÁCE


**BRNO 2019**

# Specification Master's Thesis

| | |
|---|---|
| Department: | Institute of Mathematics |
| Student: | **Bc. Martin Skopal** |
| Study programme: | Applied Sciences in Engineering |
| Study field: | Mathematical Engineering |
| Supervisor: | **Ing. Tomáš Mauder, Ph.D.** |
| Academic year: | 2018/19 |

Pursuant to Act no. 111/1998 concerning universities and the BUT study and examination rules, you have been assigned the following topic by the institute director Master's Thesis:

## Analyze and economic time series forecasting by using selected statistical methods

### Concise characteristic of the task:

The forecast of stock market development, commodity prices, index values, energy prices, etc. using mathematical models is still a highly debated issue on the stock market. Nowadays, there are several statistical models, which try to predict the time series by evaluating previous trends. The thesis should deal with the analysis and description of time series focusing on the current situation in economic areas (stock values, exchange rates, etc.). Furthermore, the thesis should deal with the models of the time series forecast and their comparison on real data.

### Goals Master's Thesis:

The work will include a mathematical description and application of classical statistical methods for time series forecast (e.g. Box–Jenkins, ARIMA, SARIMA, etc.). The next step is to create a method combining individual approaches using a weight function with coefficients obtained by mathematical optimization methods. The verifying of mathematical forecast models will be on time series with a different random value in economic areas. Finally, a critical assessment of individual statistical approaches and their applicability on the real stock market should be made.

### Recommended bibliography:


LARSEN, R. J., MORRIS, L. M. An introduction to mathematical statistics and its applications, USA 2012, ISBN 978-0-321-69394-5.

MONTGOMERY, D. M., JENNINGS, CH. J., KULAHCI, M. Introduction to time series analysis and forecasting, New Jersey 2008, ISBN 978-0-4 71-65397-4.

Deadline for submission Master's Thesis is given by the Schedule of the Academic year 2018/19

In Brno,

L. S.

| prof. RNDr. Josef Šlapal, CSc.<br>Director of the Institute | doc. Ing. Jaroslav Katolický, Ph.D.<br>FME dean |
|---|---|

**Abstrakt**

V této diplomové práci se zaměřujeme na vytvoření plně automatizovaného algoritmu pro předpovědi finančních řad, který se snaží využít kombinační proceduru na dvou úrovních mezi dvěma rodinami předpovědních modelů, Box-Jenkins a Exponenciální stavové modely, které jsou schopny modelovat jak homoskedastické tak heteroskedastické časové řady. Pro tento účel jsme navrhli selekční proceduru v prostředí MATLAB pro modely ARIMA. Výsledný kombinovaný model je pak aplikován několik finančních časových řad a jeho výkonost je diskutována.



**Abstract**

In this thesis we aim to construct a fully automatic forecasting algorithm, which is trying to utilize a combining procedure on two levels between two families of forecasting models, Box-Jenkins and Exponential smoothing state space models, that is able to deal with homoscedastic and heteroscedastic time series. For this we devise a selection procedure in the MATLAB environment for ARIMA models. The resulting combined model is then applied several financial time series and its performance is discussed.



**Klíčová slova**

Analýza časových řad, předpověď, ARIMA, Box-Jenkins, ETS, Exponential smoothing, AFTER, kombinace předpovědí, GARCH.

**Keywords**

Time series analysis, forecasting, ARIMA, Box-Jenkins, ETS, Exponential smoothing, AFTER, Forecast combination, GARCH.



SKOPAL, M. Analyze and economic time series forecasting by using selected statistical methods. Brno: Brno University of Technology, Faculty of Mechanical Engineering, 2019. 63 p. Leader of diploma thesis Ing. Tomáš Mauder, Ph.D..

I hereby declare that I am the sole author of this master thesis *Analyze and economic time series forecasting by using selected statistical methods.* and that I have not used any sources other than those listed in the bibliography and identified as references.

Martin Skopal

I would like to hereby thank my supervisor Ing. Tomáš Mauder, Ph.D. for his patience and useful advice.

Martin Skopal

# Contents

# Introduction

The subject of time series analysis and forecasting has been for many years now a very vivid topic, that deserves all the attention it is given. Especially in the last twenty years, thanks to the exponential rise in computational power, which opens many doors that might have been shut, due to technical limitations. Forecasting combination is becoming more popular together with an optimization-like approach to forecasting itself.

Even though the aim to combine forecast has been a subject of research for many years, the topic itself might still be considered controversial. Many researchers, especially statisticians, object to the notion of forecast combination itself, as their view is that by combining all the information that is to be gathered by individual forecasters, one should be able to build a single model, that is considered to be the best and the optimal forecasts are to be found.

Nevertheless, the combination approach is still appealing, from a point of view that, possibly any forecasting procedure, might have something to contribute. There could be behaviour that perhaps is not easily seen or even hidden, so that a model, which would be ordinarily disregarded, might be able to catch it. Usually, the combination approach is somewhat more present in practice, such as industry or business.

We consider an algorithm proposed by Yang [8] to reduce the risk of model selection error, which we incorporate into a composite algorithm including the variance-covariance minimization approach. The aim is to end up with a fully automatic algorithm, that after loading any time series and a given forecast horizon, returns an optimized forecast.

In the first chapter we introduce some basic notions of statistics, closely linked to time series modeling.

In the second chapter we introduce two families of times series modeling methodologies, that are ARIMA and ETS, which will be applied in our combination procedure. We describe the ETS models in more detail, as the state space variation of the exponential smoothing method might be less known to the reader.

In chapter number five the procedure for automatic model selection and the subsequent combination of the corresponding forecasts is introduced.

The last chapter is dedicated real data application, describing the handling of datasets and the following employment of the defined procedure on selected time series from the financial market.

# 1 Basic Notions

We are going to implement some concepts and properties which we are going to use in the process of working with and assessing. We will assume some basic knowledge from statistics that are common knowledge for master studies students and rather limit ourselves to notions directly related to our topic.

These concepts are mostly derived from [1], [2] and [14].

## 1.1 Time Series

In this section we will introduce some basic concepts concerning time series in such a way as to be able to conduct analysis in the framework of our statistical forecasting models. We start with the definition of a time series.

Let $(\Omega, \mathcal{U}, P)$ be a probability space and $I$ an index set. A real valued stochastic process is a real valued function $y_t(\omega)$ such that for each fixed $t \in I$, $y_t(\omega)$ is a random variable on $(\Omega, \mathcal{U}, P)$.

Looking at this definition we see, fixing $\omega$ we get one realization of the stochastic process $y_t$ for $t = 1, \ldots, T$ which we will call a time series with period $T \in \mathbb{R}$ such that for each $t$, $y_t$ is called an observation.

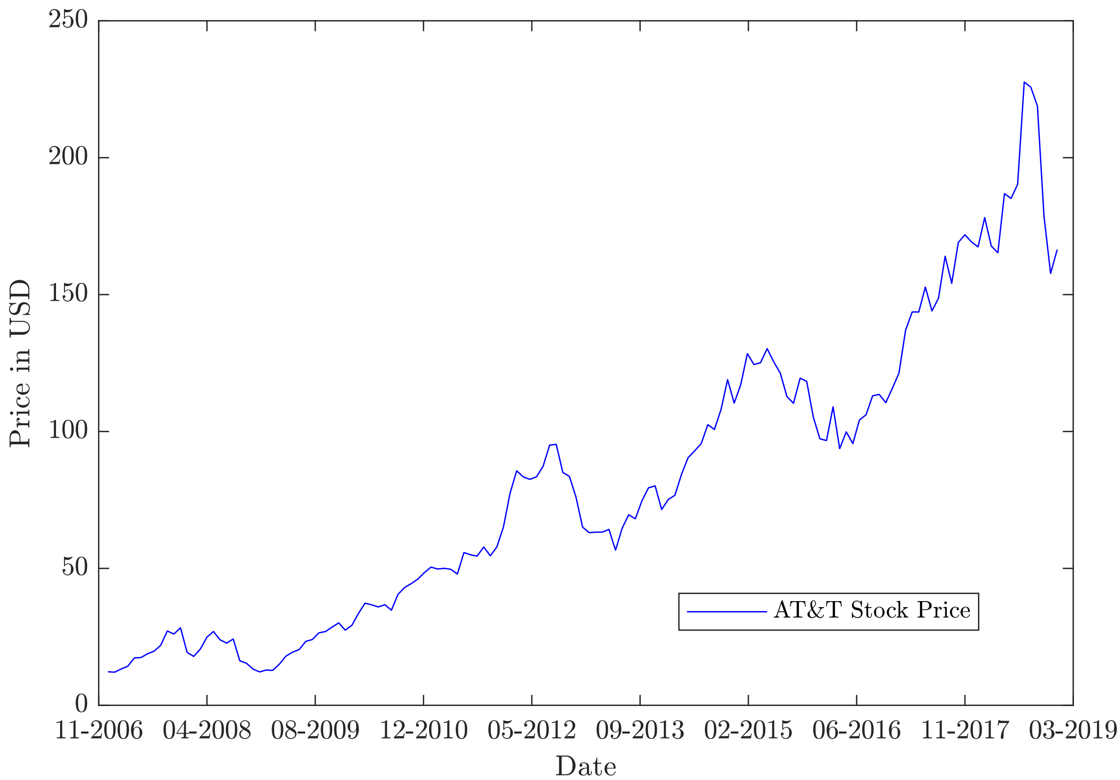


Figure 1: Time Series Plot

Time series come in many different shapes and forms. They vary in the data they are tracking, be it stock prices, sales numbers, imports for a country, etc. In the same way they can vary in the context of time, they are being collected over so that we can get yearly, quarterly, monthly or even second-by-second observations of the collected data. For the purpose of this thesis we will generally assume discrete time series that are observed at regular time intervals.

### 1.1.1 Stationarity

Stationarity is a very important property of a time series as it implies a kind of stability of the data. It is a vital assumption especially for the Box Jenkins methodology as we will see in later chapters. We firstly introduce the notion of strict stationarity.

We say that a strictly stationary stochastic process $\{y_t; t \in \mathbb{Z}\}$ is one for which the probabilistic behavior of every collection of values

$$\{y_{t_1}, y_{t_2}, \dots, y_{t_k}\},$$

is identical to that of the time shifted set

$$\{y_{t_{1+h}}, y_{t_{2+h}}, \dots, y_{t_{k+h}}\}.$$

That is, the joint probability distribution is of the form

$$P\{y_{t_1} \leq c_1, \dots, y_{t_k} \leq c_k\} = P\{y_{t_{1+h}} \leq c_1, \dots, y_{t_{k+h}} \leq c_k\},$$

for all $k = 1, 2, \dots,$ all time points $t_1, t_2, \dots, t_k$, all numbers $c_1, c_2, \dots, c_k$, and all time shifts $h = 0; \pm1, \pm2, \dots$.

In other words, we could say that a time series is said to be strictly stationary if a change in the time origin doesn't affect its properties, hence one could interpret it as a kind of stability. Nevertheless, such an assumption might be to strong or strict for the use in some forecasting methodologies. Therefore, we introduce the notion of weak stationarity.

So, we say that a stochastic process $\{y_t; t \in \mathbb{Z}\}$ is weakly stationary if

1. The second moment of $y_t$ is finite for all $t$, that is $E|y_t|^2 < \infty$, $\forall\, t \in \mathbb{Z}$.
2. The first moment of $y_t$ is independent of $t$, that is $E(y_t) = \mu$, $\forall\, t \in \mathbb{Z}$.
3. The cross moment $E(y_{t_1} y_{t_2})$ depends only on $t_1 - t_2$, that is $\text{Cov}(y_{t_1} y_{t_2}) = \text{Cov}(y_{t_{1+h}} y_{t_{2+h}})$, $\forall t_1, t_2, h \in \mathbb{Z}$.

From now on, when talking about stationarity we will mean the weak stationarity property.

#### White Noise Process

A special kind of a stationary time series process is the so-called white noise process. $\varepsilon_t$ is a stationary process with

$$E(\varepsilon_t) = \mu = 0,$$

$$Cov(\varepsilon_m, \varepsilon_n) = \begin{cases} \sigma^2, & m = n\,, \\ 0, & m \neq n\,, \end{cases} \tag{1}$$

It is said to be a white noise process with variance $\sigma^2$. Furthermore if we add the condition that $\varepsilon_t$ is a sequence of independent identically distributed (IID) normal random variables, then we get the so called Gaussian white noise, i.e.

$$\varepsilon_t \sim N(0, \sigma^2). \tag{2}$$

We will assume the Gaussian white noise if not stated otherwise from now on.

## 1.2 Heteroscedascity

More often than not, we are confronted by time series, which do not have constant variability, in finance we would say that such series is highly volatile. Such data we call heteroscedastic.

In time series analysis the term conditional heteroscedasticity is common. Usually we want to know if the conditional variance of $y_t$ conditioned on $y_{t-1}, \dots, y_1$ is constant or not. If it is not, then the data is said to be conditionally heteroscedastic.

It is important to take this behaviour into consideration, when building our model, as its neglection could cause misleading results.

## 1.3 Autocorrelation and Partial Autocorrelation

Autocorrelation is the linear dependence of a variable with itself at two points in time. For stationary processes, autocorrelation between any two observations only depends on the time lag $h$ between them.

Define $\gamma(h) = Cov(y_{t+h}, y_t)$ then lag-h autocorrelation is given by

$$\rho(h) = Corr(y_{t+h}, y_t) = \frac{Cov(y_{t+h}, y_t)}{\sqrt{Var(y_{t+h})Var(y_t)}} = \frac{\gamma(h)}{\gamma(0)} .$$

Then the values of $\rho(h)$ for $h = 0, 1, 2, \dots$ is called the autocorrelation function (ACF).

Assume the random variables $X, Y, Z$ and consider the simple linear regressions

$$\hat{X} = a_1 + b_1 Z, \ \ b_1 = \frac{Cov(Z, X)}{Var(Z)} ,$$
$$\hat{Y} = a_2 + b_2 Z, \ \ b_2 = \frac{Cov(Z, Y)}{Var(Z)} ,$$

So, the errors are calculated as follows

$$X^* = X - \hat{X} = X - (a_1 + b_1) ,$$
$$Y^* = Y - \hat{Y} = Y - (a_2 + b_2) ,$$

then the correlation between $X^*$ and $Y^*$, $corr(X^*, Y^*) = corr(X - \hat{X}, Y - \hat{Y})$ is the partial correlation between two random variables $X$ and $Y$ adjusted for a common factor $Z$.

Then by following the definition stated, we can define the partial autocorrelation function between $y_t$ and $y_{t+h}$ as the autocorrelation between $y_t$ and $y_{t+h}$ after adjusting for $y_{t+1}, \dots, y_{t+h-1}$.

The values of ACF and PACF are usually plotted against the lags into a so called correlogram. An example is given in this figure.

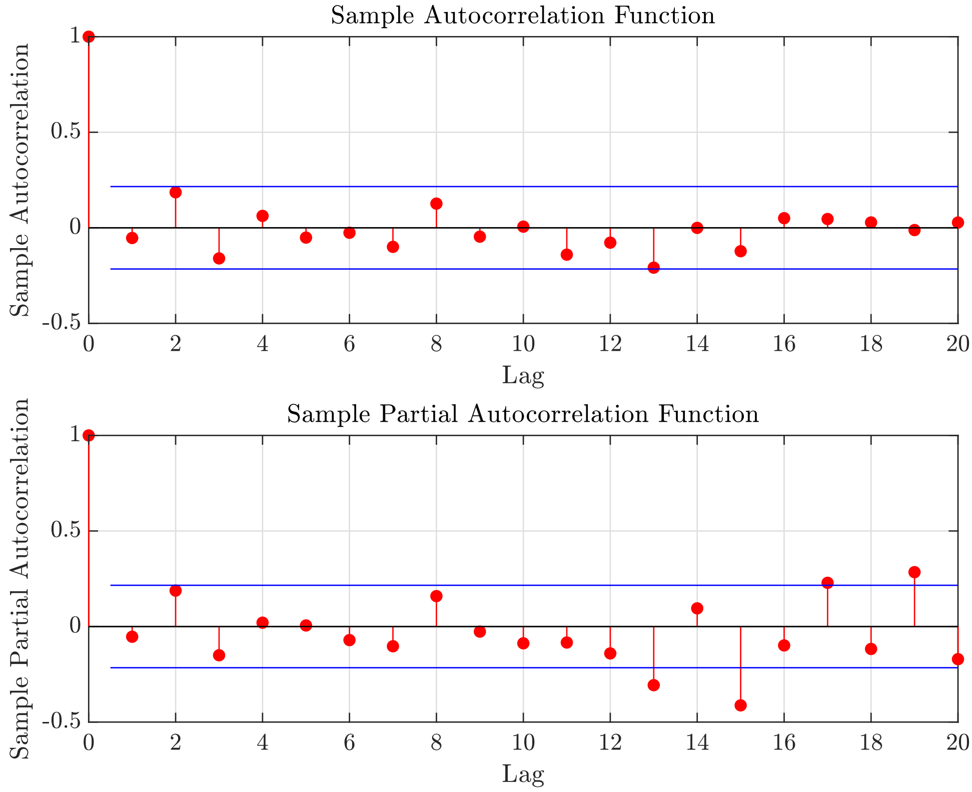


Figure 2: Correlograms of ACF and PACF

## 1.4 Time Series Patterns

When we are collecting any data throughout some time interval at any time step, we always end up with a set of numbers, which when plotted against their corresponding time stamps, can be visualized as points on a plane. This is an essential task when dealing with time series. What we end up with might vary in infinitely many ways depending not only on the data itself but also on the person who's observing it and making assumptions about. Such a fact should not be taken too lightly as human intervention and a kind of intuition, mostly derived by experience, plays a fairly large role in such endeavours.

We hinted the importance of plotting the data and the reason for it is, that through this simple step one might be able to obtain some preliminary idea about the descriptive properties of such a time series. A good way to do so might be to connect each point with a line. Such an undertaking might reveal features that are characteristic for any time series, namely trend, cycle, seasonality.

A trend can be defined as a long-term direction or change of the mean of a time series. Such a behaviour can be observed by fitting a line to the data points. The seasonality describes a pattern that repeats itself with a known periodicity, which is dependent on the frequency that the data is collected. The seasonal behaviour yields a wave-like pattern. Such a pattern is also characteristic for the cycle, but cyclic changes happen over a much larger period than for the seasonal.

It can also happen that a time series does not have any recognizable pattern or that it is purely random.

## 1.5 Time Series Decomposition

A way of looking at a time series is to see it as a combination of a variety of components such as trend, seasonal, cycle and error, each representing one of the underlying pattern categories. This can be useful as it helps us getting more insight on the behaviour of the time series and it may be even valuable when we try to predict it.

Assume that the time series $y_t$ consists of three components: a seasonal component, a trend component, and a remainder component. An intuitive approach would be to take these components and add them up all together, in which case we end up with an additive model.

$$y_t = S_t + T_t + E_t\,,$$

where $y_t$ is the data at time t, $S_t$ is the seasonal component, $T_t$ is the trend-cycle component and $E_t$ is the remainder (or irregular or error) component at time t. Note that even though here we assume a trend-cycle component, a separate cycle component could be introduced as well, nevertheless its incorporation into the trend is a likewise acceptable alternative.

Adding, of course, is not the only way of dealing with the them. By applying the multiplication operation the result yields a multiplicative model instead.

$$y_t = S_t \times T_t \times E_t\,.$$

Such a model is warranted if we want to express a higher growth rate magnitude. We can also combine these two approaches so that we can reflect different types of behaviour that might be unsuitable for the two pure decomposition models, such as for example.

$$y_t = (S_t + T_t) \times E_t\,. \tag{3}$$

This model in particular will be used further on in the case of exponential smoothing state space models.

## 1.6 Time Series Forecasting

The forecasting of a time series is one of the most important parts of working with them, as a certain desire to know the future, has been a part of the human experience throughout history. Not only that, but it can yield many practical benefits, such as saving money, resources and time itself.

A considerable amount of methods has been devised to tackle such a problem, based for example, on exploiting past observations to draw conclusions about future events, or also incorporating external influences that might help us in the forecasting endeavor.

When talking about forecasting of a time series, we often use the terms methods and models. It is important to firstly establish what is the difference between these two notions, before we delve any further. There are probably many ways to distinguish both, but we are going to use this interpretation. A forecasting method is an algorithm providing a single value prediction of the value at a future time period, in other words a point forecast. Contrarily a model allows us an entire probability distribution for future values which can be used to compute prediction or forecast confidence intervals, that are desirable as they convey much more information for example about the risks of an asset when we talk about financial time series.

Assuming a we have the observations of a time series $y_t$ for time $t = 1, 2, \dots T$. Consider that we want to forecast the $h$ future values $\hat{y}_t$ for $t = T+1, T+2, \dots, T+h$. We call $T$ the origin, as the value of time from which the forecasts are computed and $h$ as the horizon, or also leading time, that represents the number of future values that are to be predicted.

The calculation of the future values is then given by the corresponding method in use. A complete model can give us the confidence interval for a given percentage value.

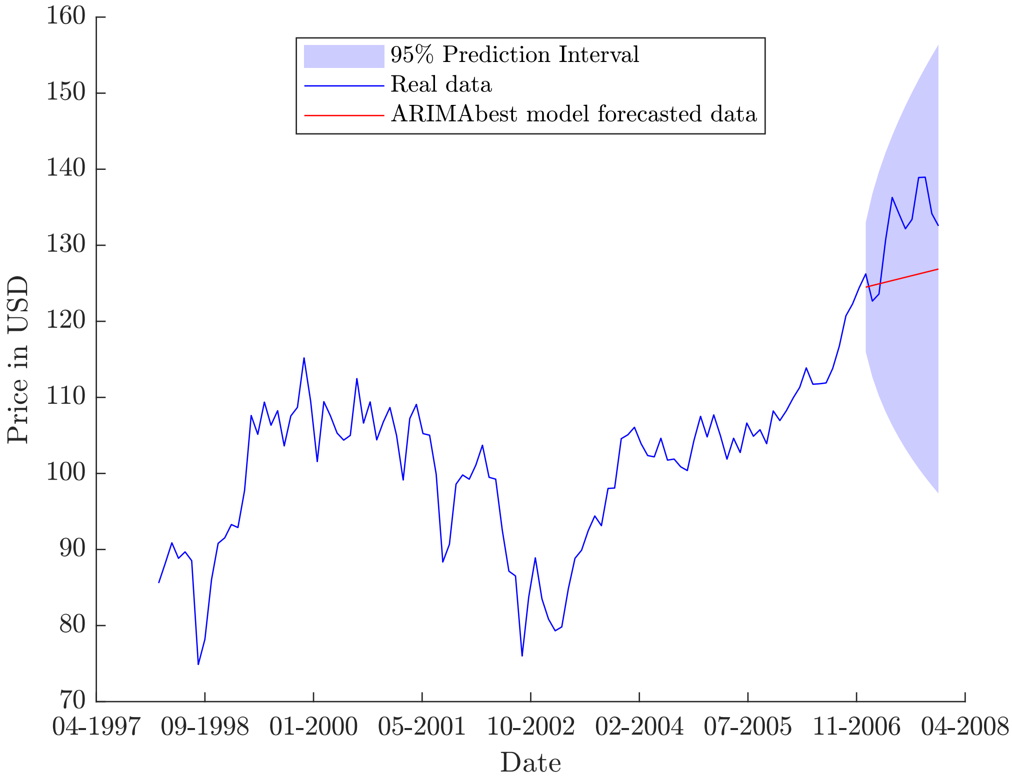


Figure 3: A forecasted time series plotted against real data

### 1.6.1 Forecast Accuracy

Forecasting accuracy is a measure, which expresses performance of forecasting model. It is a reverse value to the measure of forecasting error. There are several alternatives when it comes to calculating the measure of the forecasting error, with each of them measuring a different quality, so that by aggregating them, we get more detailed picture about the performance of our model. Firstly, we define the forecast error. It is expressed as a deviation of predicted value and actual value

$$\varepsilon_t = y_t - \hat{y}_t \,. \tag{4}$$

Then we can define some of the measures of forecasting error.

- Mean Squared Error

$$MSE = \frac{1}{N} \sum_{t=1}^{N} (y_t - \hat{y}_t)^2 \,. \tag{5}$$

- Mean Absole Percentage Error

$$MAPE = \frac{1}{N}\sum_{t=1}^{N} |\frac{(y_t - \hat{y}_t)}{y_t}| \cdot 100\,. \tag{6}$$

Where $y_t$ and $\hat{y}_t$ are the actual and predicted values respectively, $N$ is the size and $\mu$ is the mean of the test dataset. A lesser value of these measures indicates better forecast accuracy and an overall better performance of the model.

# 2 Forecasting Models

In this chapter we will introduce two forecasting techniques with different assumptions for arriving at the models, they yield. One of them will be the ARIMA or also known as the Box-Jenkins models and the exponential smoothing state space (or ETS) models. We firstly, introduce the class of ARIMA models in the following chapter.

## 2.1 Box-Jenkins Models

The class of ARIMA models or also known as Box-Jenkins models are one of the well known, well established and widely used forecasting models that can be. They were introduced by Box and Jenkins in 1971 and have been a staple in time series forecasting endeavors ever since.

The main idea is to express a dependency of the current value of a time series on the recent history. An important condition is on the stationarity of such a time series. As many conclusions in the theory behind the methodology rely on such a condition. The assumption of a weakly stationary time series is considered strong enough in this case.

The descriptions of the Box-Jenkins methodology models and their characteristics are acquired from [2] and [14].

### 2.1.1 Finite Order Moving Average model

In this subclass of models, the dependency of the current value of the time series is not sought on the past values of the time series itself, but rather on the past forecast errors.

Assume $y_t$ is a value of a time series at a given time $t$. We define the $q$ order MA($q$) as follows

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}\,, \tag{7}$$

where $\varepsilon_t$ is considered to be a white noise process. A more convenient way of representing this model can be shown by introducing the so-called backshift operator B, such that $By_t = y_{t-1}$, so that we get

$$\begin{aligned} y_t &= \mu + (1 - \theta_1 B - \cdots - \theta_q B^q)\varepsilon_t \\ &= \mu + (1 + \sum_{i=1}^{q} \theta_i B^i) \\ &= \mu + \Theta(B)\varepsilon_t\,, \end{aligned} \tag{8}$$

where $\Theta(B) = 1 - \sum_{i=1}^{q} \theta_i B^i$ Considering a MA($g$) process given by the equation above, it can be shown that such a process is stationary, for any values of $\theta_i$. So that by taking into consideration that $\varepsilon_t$ is a white noise process, we can express the mean value and variance of a MA($g$) as follows

$$E(y_t) = E(\mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}) = \mu\,, \tag{9}$$

$$\begin{aligned} Var(y_t) &= Var(\mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}) \\ &= \sigma^2 (1 + \theta_1^2 + \cdots + \theta_q^2)\,, \end{aligned} \tag{10}$$

Hence we are able to give the calculations of the autocorrelation function (ACF)

$$\rho(h) = \begin{cases} \frac{\gamma(h)}{\gamma(0)} = \frac{-\theta_h + \theta_1\theta_{h-1} + \cdots + \theta_{q-h}\theta_q}{1+\theta_1^2+\cdots+\theta_q^2}, & h = 1, 2, \ldots, q \\ 0, & h > q \end{cases} \tag{11}$$

given that the autocovariance at lag $h$ $\gamma(h)$ can be calculated as such

$$\begin{aligned} \gamma(h) &= Cov(y_t, y_{t+h}) \\ &= E[(\varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q})(\varepsilon_{t+k} - \theta_1\varepsilon_{t+k-1} - \cdots - \theta_q\varepsilon_{t+k-q})] \\ &= \begin{cases} \sigma^2(-\theta_h + \theta_1\theta_{h-1} + \cdots + \theta_{q-h}\theta_q), & h = 1, 2, \ldots, h\,, \\ 0, & h > q\,. \end{cases} \end{aligned} \tag{12}$$

So, we see that theoretically for a MA($q$) process after lag q, the ACF is equal to zero further on. Of course, in practice, this is not the case, but we take a certain confidence, usually 95%, that define bounds after which we consider it equal to zero. So, the last lag value above outside these bounds, defines the order of the MA($q$) model.

### 2.1.2 Finite Order Autoregressive model

Here we describe the actual dependency of the current value of a time series on a number of its preceding observations that are representing a distinct pattern.

Assume $y_t$ is a value of a time series at a given time $t$. We define the $p$ order AR($p$) as follows

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \tag{13}$$

where again $\varepsilon_t$ is a white noise process. Given the backshift operator B, we can rewrite the equation into

$$\phi(B)y_t = \delta + \varepsilon_t, \tag{14}$$

given that $\phi(B)y_t = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$.

Similarly, in the case of AR($p$), the order of $p$ can be found by looking at the PACF. So that the last lag value outside the confidence bounds, defines the order $p$.

### 2.1.3 Mixed Autoregressive Moving Average model

Simply by putting the two mentioned models together, we can produce a mixed version of both, called the ARMA model, that can be written as follows

$$\phi(B)y_t = \delta + \theta(B)\varepsilon_t\,, \tag{15}$$

where

$$\begin{aligned} \phi(B) &= 1 - \sum_{i=1}^{p} \phi_i B^i\,, \\ \theta(B) &= 1 - \sum_{i=1}^{q} \theta_i B^i\,. \end{aligned} \tag{16}$$

The procedures of finding the optimal value of $p, q$ are the same as for the separate procedures themselves by using the ACF and PACF as a guide.

### 2.1.4 Autoregressive Integrated Moving Average model

With respect to the stationarity condition, we are able to define a nonstationary process such that, we call $y_t$ homogeneous, nonstationary if $y_t$ itself is nonstationary, but the d-th differences $(1-B)^d y_t$ are. Such a process we call ARIMA$(p, d, q)$ and can be defined as

$$\phi(B)(1-B)^d y_t = \delta + \theta(B)\varepsilon_t\,, \tag{17}$$

where $\phi(B), \theta(B)$ are the same as in (16)

Usually no more than $d = 2$ differences are necessary to reduce a nonstationary series into a stationary one.

Something about unit roots etc.

### 2.1.5 Seasonal ARIMA Model

The models shown before are able to model non seasonal data, so with regards to this we call them non seasonal ARIMA models. But the Box-Jenkins methodology allows us also to study seasonal data, by applying seasonal ARIMA models, also known as SARIMA$(p, d, q)(P, D, Q)_m$ models, where $m$ is the period and by using the lower case and upper-case notation, we distinguish between the nonseasonal and seasonal parts respectively.

So, by applying the backshift operator we can write the model as follows

$$\phi(B)\Phi(B^m)(1-B)^d(1-B^m)^D y_t = \theta(B)\Theta(B^m)\varepsilon_t, \tag{18}$$

where

$$\Phi(B^m) = 1 - \sum_{i=1}^{P} \Phi_i B^{im}\,,$$

$$\Theta(B^m) = 1 - \sum_{i=1}^{Q} \Theta_i B^{im}\,,$$

and $\phi(B), \theta(B)$ are the same as in (16) and $m$ is the seasonal period.

The orders $P$ and $Q$ correspond to separate seasonal autoregressive and seasonal moving average models, SAR$(P)$ and SMA$(Q)$ respectively. They can be found by analyzing the PACF and ACF at the corresponding lags given by the period $m$, but usually, they are either equal to zero or one.

### 2.1.6 GARCH Models

So far we described processes where the residuals $\varepsilon_t$ were thought of as observations of a white noise process. Such a process is mainly defined by its two main properties, which are a constant zero mean and a constant variability. We have seen, that by differencing, we are able to address the problem of the nonconstant and nonzero mean, but not, what to do when the condition of constant variance is not met.

This property, as we mentioned before, is called heteroscedascity and there are several ways how to address it. One, more obvious approach, could be to somehow transform the data. Such an approach is viable if for example we see that the variability grows close to proportionally to the mean, then a log transform might be a good choice.

But generally, time series don't have to behave in such a fashion. In this case another approach has to be used. The idea behind it, is to model the nonconstant variability as, a possibly ARMA model itself. So, assuming that $\epsilon_t$ is an uncorrelated zero mean noise with changing variance, we can try to model $\epsilon_t^2$ as a pseudo-ARMA process dependent on the last $l$ values of $\epsilon_t^2$ and the last $k$ values of the conditional variance $v_t = Var(\epsilon_t|\epsilon_{t-1,\dots})$, we can define a process introduced by Bollersev [5] and is given as follows

$$v_t = \xi_0 + \zeta_1 v_{t-1} + \zeta_2 v_{t-2} + \dots + \zeta_k v_{t-k} + \xi_1 \epsilon_{t-1}^2 + \xi_2 \epsilon_{t-2}^2 + \dots + \xi_l \epsilon_{t-l}^2\,, \tag{19}$$

in which case we say that $\epsilon_t$ follows a generalized autoregressive conditional heteroscedastic process of order $k$ and $l$, i.e. GARCH$(k, l)$. We should note, that even though this model might resemble an ARMA process, it is not one, as there is no white noise error term with a constant variance for the MA part.

#### 2.1.7 Mixed Autoregressive Moving Average with Conditional Variance

To fully model the conditional mean and conditional variance, the mixed, or also sometimes called hybrid, ARIMA-GARCH model can be used. We define it by simply adding the two models together, obtaining an ARIMA(p,d,q)-GARCH(k,l) model.

$$(1 - \sum_{i=1}^{p} \phi_i B^i)(1 - B)^d y_t = \delta + (1 - \sum_{i=1}^{q} \theta_i B^i)\varepsilon_t\,, \tag{20a}$$

$$v_t = \xi_0 + \sum_{i=1}^{k} \zeta_i v_{t-i} + \sum_{i=1}^{l} \xi_i \epsilon_{t-i}^2 \tag{20b}$$

#### 2.1.8 Parameter Estimation

For the parameters given by the models mentioned before there are several methods by which they can be estimated. As we are using the MATLAB environment, this task is dealt with, within the confines of the arima class objects integrated in it as a maximum likelihood estimates problem.

### 2.2 Exponential Smoothing State Space Models

Exponential smoothing is one of the older and quite successful classes of forecasting methods. Possibly originating at about 1944, it gained traction during the 50s thanks to the work of Robert G. Brown, Charles Holt and later Peter Winters.

Nevertheless, not until fairly recently, there was no modeling framework that would include key qualities, for such stochastic models that is, prediction intervals or likelihood calculations. In the work of J.K. Ord and R.G. Hyndman a class of state space models, underlying the exponential smoothing methods has been developed.

ETS models, unlike ARIMA models, are based on the idea of decomposing a time series, so that they can be considered a complementary approach to them. Rather then trying to predict the future using some of the preceding observations, we instead try to model the individual parts of the time series itself, such as level, trend, seasonality and an error term.

The descriptions of the ETS models are acquired from Hyndman [1] and [14].

### 2.2.1 State Space Models

In general, a state space model describes the dependence of the observed measurement on the unobserved state variable. Such a description corresponds to the exponential smoothing notion, where the observed value of the time series depends on the unobserved terms for level, trend and seasonality, including some error term. In the case of ETS models, we consider them in the form of "single source of error" (or SSOE) state space models or the so-called *innovations* formulation.

Let us denote the observation at time $t$ as $y_t$ and $\boldsymbol{x}_t$ being the state vector, which consists of the unobserved parts of the time series, namely the level, trend and seasonality. Considering this we can write the general linear innovations state space model as follows

$$y_t = \boldsymbol{w}^{'}\boldsymbol{x}_{t-1} + \varepsilon_t, \tag{21a}$$

$$\boldsymbol{x}_t = \boldsymbol{F}\boldsymbol{x}_{t-1} + \boldsymbol{g}\varepsilon_t, \tag{21b}$$

where $\boldsymbol{w}$ and $\boldsymbol{g}$ are vectors of coefficients, $\boldsymbol{F}$ being a square matrix of coefficients and $\{\varepsilon_t\}$ being a white noise series. The equation (21a) is called the measurement or observation equation and which depicts the relation between the unobserved state $\boldsymbol{x}_t$ and the observation $y_t$. The equation (21b) is then called the state or transition equation, which describes the evolution of the states through time.

Alternatively, we can define the nonlinear form of the state space model as follows

$$y_t = w(\boldsymbol{x}_{t-1}) + r(\boldsymbol{x}_{t-1})\varepsilon_t, \tag{22a}$$

$$\boldsymbol{x}_t = f(\boldsymbol{x}_{t-1}) + g(\boldsymbol{x}_{t-1})\varepsilon_t, \tag{22b}$$

where the coefficient vectors and matrix are replaced by nonlinear functions dependent on the state vector $\boldsymbol{x}_{t-1}$. As one can clearly see, both sets of equations (21) and (22) have in their respected two equations the same source of error as we already hinted before. It is not uncommon to have two independent sources of error for each of those equations, in other words such a model has multiple sources of error (or MSOE). Maybe unintuitively, it can be shown, proving the utility of SSOE formulation of the models, that they are in fact less restrictive as their MSOE counterparts concerning this case.

### 2.2.2 Exponential Smoothing Methods

Before we show the actual ETS models, we firstly introduce the methods they underlie. We will limit ourselves to the models and the corresponding methods that we will later work with. Those methods are simple exponential smoothing, Holt's linear method, Holt-Winter's additive method and, for some, their corresponding forms with a damped trend. These are the well known names for these methods, but for a reason that will be obvious later, we will opt for a different classification given by the so called Pegel's taxonomy.

As already said, the idea behind exponential smoothing is to model the series by its terms, which are called level, trend and season. Starting with the trend component, that is described as a combination of the level and growth terms. Those can be combined in a variety of ways. We will consider only some of them, yielding us three types of the trend component: none, additive and damped additive.

Secondly, we consider the seasonal component. In this case we only distinguish between the types of none or additive, basically just distinguishing between the inclusion or exclusion of such a component in the method without a special form for it. Which is not to say that such forms don't exist as for both trend and seasonal components, a multiplicative type exists as well. But as we won't utilize them in our work later, we will omit them from further discussion eventhough they are worth mentioning at least at this point.

By combining our two components together, through simple combinatorics, we see that we end up with six methods described in the table below.

| Trend component | Seasonal component | |
|---|---|---|
| | None | Additive |
| None | (N,N) | (N,A) |
| Additive | (A,N) | (A,A) |
| Additive Damped | $(\mathrm{A_d},\mathrm{N})$ | $(\mathrm{A_d},\mathrm{A})$ |

Table 1: Types Exponential smoothing methods

The two letters in the brackets describe the method as follows. The first one describes the trend component and the second one the seasonal component.

Given these six methods, we can now describe them more precisely. For each of them we will show the equations governing the point forecasts they yield as they are the same as for their state space counterparts.

### Simple Exponential Smoothing - (N,N)

As this is the simplest method, there are several ways one can write the governing equations. We will inspect this method more closely as its simplicity offers us a look at the general methods that might be hidden or more difficult to see in the more complex ones.

This method is most suitable for time series with no clear trend or seasonality. The idea behind the method is to find some middle ground between a na?ve method where the forecasts are all equal to the last observation and the average method, where all future forecasts are equal to a simple average of all the values observed before. So, one might suggest thinking of more recent observations as more important and so attaching larger weights to them then the less recent ones. Hence, we get a method that where the weights decrease exponentially as we go further back in time.

Suppose that we have a time series $y_t$. The forecast is denoted by $\hat{y}_t$ for the corresponding time step $t$. The forecast error is then $y_t - \hat{y}_t$. Now if we want to forecast the next time step, we write

$$\hat{y}_{t+1} = \hat{y}_t + \alpha(y_t - \hat{y}_t)\,, \tag{23}$$

where $0 \leq \alpha \leq 1$ and by expanding the equation, we can rewrite it into

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t\,. \tag{24}$$

and by replacing the forecasted terms $\hat{y}_t$ we can get a clear picture of the exponential smoothing as seen here for the first step

$$\begin{aligned} \hat{y}_{t+1} &= \alpha y_t + (1-\alpha)\big[\alpha y_{t-1} + (1-\alpha)\hat{y}_{t-1}\big] \\ &= \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2\hat{y}_{t-1}, \end{aligned} \tag{25}$$

and by further substitution the whole expression

$$\hat{y}_{t+1} = \sum_{i=0}^{t} \alpha(1-\alpha)^i y_{t-i} + (1-\alpha)^t \hat{y}_1 \,. \tag{26}$$

We see that the problem now is not only, how to choose the constant $\alpha$, but also the initial forecast $\hat{y}_1$, which we will discuss later.

But most of the time we are interested in an actual forecast, so when we go beyond the scope of the observed time series. In this case the method produces a so called flat forecast function

$$\hat{y}_{t+h|t} = \hat{y}_{t+1}, \qquad h = 2, 3, \dots \,, \tag{27}$$

which ultimately leads us to the component form of representing the method which is generally, an assembly of two types of equations, namely a forecast equation and a number of smoothing equations coinciding with the number of components included in the method. In the case of SES, it is given by

$$\begin{aligned} &\text{Forecast:} & \hat{y}_{t+h|t} &= \ell_t \,, & \text{(28a)} \\ &\text{Level:} & \ell_t &= \alpha y_t + (1-\alpha)\ell_{t-1} \,, & \text{(28b)} \end{aligned}$$

where $\ell_t$ is called the estimate of the level and it is present in the only smoothing equation, that is the level equation. The usefulness of such a structure is clear as it resembles the form of a state equation, so that the transition to them is as clear as possible. For that reason the rest of the methods in further discussion will be given in this form only.

### Holt's Method - (A,N)

Extending the SES method to accommodate a trend behaviour, we introduce the Holt's linear method, which includes, besides the forecast equation, two smoothing equations and can be given by

$$\begin{aligned} &\text{Forecast:} & \hat{y}_{t+h|t} &= \ell_t + hb_t \,, & \text{(29a)} \\ &\text{Level:} & \ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}), & \text{(29b)} \\ &\text{Growth:} & b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}. & \text{(29c)} \end{aligned}$$

Here the components are as follows, the level estimate $\ell_t$ is the same as for SES and $b_t$ denotes the growth or also the slope estimate, which as we see is a weighted average between the the difference of the last two levels and the growth component of the previous step. We end up with two smoothing parameters $0 \leq \alpha, \beta^* \leq 1$ for each of the components respectively. As we clearly see, the forecast is no longer flat but can be increasing or decreasing, depending on the growth estimate, with each step.

### Damped Trend Method - ($\mathbf{A_d}$,N)

Given a slight modification of the previous method, we get a method that accommodates a specific behaviour of the trend, that is when the trend is damped over time and it is given by

$$\begin{aligned} \text{Forecast:} \quad & \hat{y}_{t+h|t} = \ell_t + (\phi + \phi^2 + \cdots + \phi^h) b_t, && (30a) \\ \text{Level:} \quad & \ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}), && (30b) \\ \text{Growth:} \quad & b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}. && (30c) \end{aligned}$$

The number of smoothing equations stays the same, but they are modified by the so called damping parameter affecting the growth estimate in each of them. The forecast itself is damped by a factor of $0 < \phi < 1$ for each future step and for $h \longrightarrow \infty$ the forecasts approach the asymptote $\ell_t + \phi b_t / (1 - \phi)$.

### Holt-Winters' Additive Method - (A,A)

To accommodates for not only a trend, but also a seasonal pattern with a period $m$, we introduce a method by extending the Holt's method by another smoothing equation estimating this seasonal pattern. It is given by these equations

$$\begin{aligned} \text{Forecast:} \quad & \hat{y}_{t+h|t} = \ell_t + h b_t + s_{t-m+h_m^+}, && (31a) \\ \text{Level:} \quad & \ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}), && (31b) \\ \text{Growth:} \quad & b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*) b_{t-1}, && (31c) \\ \text{Seasonal:} \quad & s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma) s_{t-m}. && (31d) \end{aligned}$$

The equation for the growth is the same as for (29c) together with the restrictions on the corresponding parameters. The difference lies in the level and the seasonal equations, where $s_t$ is the seasonal estimate with its parameter $0 < \gamma < 0$ and $h_m^+ = \big[(h-1) mod m\big] + 1$.

### Exponential Smoothing Methods Conclusion

The rest of the methods we introduced in Table 1 are basically combinations of the already described smoothing equations and as for not to repeat ourselves, we omit their closer discussion and introduce them in the form of a concluding table for all the given methods.

Table 2: Exponential smoothing methods

| Trend | Seasonal N | Seasonal A |
|---|---|---|
| N | $\hat{y}_{t+h\|t} = \ell_t$<br>$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$ | $\hat{y}_{t+h\|t} = \ell_t + s_{t-m+h_m^+}$<br>$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}$<br>$s_t = \gamma(y_t - \ell_{t-1}) + (1-\gamma)s_{t-m}$ |
| A | $\hat{y}_{t+h\|t} = \ell_t + hb_t$<br>$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$<br>$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ | $\hat{y}_{t+h\|t} = \ell_t + hb_t + s_{t-m+h_m^+}$<br>$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$<br>$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$<br>$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$ |
| $A_d$ | $\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t$<br>$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$<br>$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ | $\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t + s_{t-m+h_m^+}$<br>$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$<br>$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$<br>$s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m}$ |

Where $\phi_h = (\phi + \phi^2 + \cdots + \phi^h)$ and all other components and parameters are the same as mentioned before.

### 2.2.3 State Space Models for Exponential Smoothing Methods

For the methods mentioned earlier we now present the state space models which underlie these methods. This is done, simply said, by introducing an error term into the equations we defined. Such an error term can be either additive or multiplicative giving us two models for each method and a sum total of twelve models.

To distinguish between them, we extend our defined notation by an extra letter, which stands for the nature of the error term, which can be either A-Additive, or M-Multiplicative. The notation is then defined in the following way ETS($\cdot$,$\cdot$,$\cdot$), where the term in the brackets stand for the error,trend and seasonality correspondingly.

We will now closely examine the two classes, given by the nature of the error term.

#### Linear Innovations State Space models

Assuming $y_t$ are the values of an observed time series, $\boldsymbol{x}_t$ is a vector of unobserved state variables modeling the behaviour of the time series, then any linear innovations state space model can be written in the form

$$y_t = \boldsymbol{w}'\boldsymbol{x}_{t-1} + \varepsilon_t, \tag{32a}$$

$$\boldsymbol{x}_t = \boldsymbol{F}\boldsymbol{x}_{t-1} + \boldsymbol{g}\varepsilon_t, \tag{32b}$$

where $\boldsymbol{x}_t$ contains potentially all the mentioned components, such as level, trend and seasonality.

Equation (32a) is called the measurement equation. It describes, namely the term $\boldsymbol{w}'\boldsymbol{x}_{t-1}$ the dependency of the current observation $y_t$ on the past. The error term $\varepsilon_t$ is in this case a white noise, implying that it has a zero mean and a constant variance. $\varepsilon_t$ is also called the innovation, therefore the name linear innovations state space models.

By the equation (32a), which is called the transition equation, we understand a recurrent description of change of the state vector $\boldsymbol{x}_t$. Such a change is described by the transition matrix $\boldsymbol{F}$ that is a square matrix, so that the term $\boldsymbol{F}\boldsymbol{x}_{t-1}$ depicts this relationship of the current state vector to its past values. But this is not the only influence as the term $\boldsymbol{g}\varepsilon_t$ describes the unpredictable effect on the state vector $\boldsymbol{x}_t$, where $\boldsymbol{g}$, called the persistence vector regulates this effect. $\boldsymbol{g}$ is the vector of the smoothing parameters we discussed earlier.

The vectors $\boldsymbol{w}, \boldsymbol{g}$ and the matrix $\boldsymbol{F}$ are fixed and generally incorporate estimable parameters. The way of estimating these parameters will be shown later, together with the estimation of the seed state $x_0$.

For illustrative purposes we are going to show the assembly of the model for ETS(A,A,A) as it contains all the mentioned components, while the rest will be summarized in an all encompassing table, where all the models will be shown.

By referring to the beginning of the chapter on State Space Models for Exponential Smoothing Methods, we mentioned a simplistic approach how to get an underlying model by introducing an error term to the component form of a given method. Such a term is introduced quite easily, considering that $\hat{y}_{t|t-1}$ is the 1-step ahead forecast of the real value $y_t$, then by setting

$$\varepsilon_t = y_t - \hat{y}_{t|t-1}\,, \tag{33}$$

we obtain the residuals at time $t$. Moreover, expanding and rearranging the equations (31) we get

$$\begin{aligned}
\hat{y}_{t|t-1} &= \ell_{t-1} + b_{t-1} + s_{t-m}, && \text{(34a)}\\
\varepsilon_t &= y_t - \hat{y}_{t|t-1}\,, && \text{(34b)}\\
\ell_t &= \ell_{t-1} + b_{t-1} + \alpha(y_t - \ell_{t-1} + b_{t-1} + s_{t-m}), && \text{(34c)}\\
b_t &= b_{t-1} + \beta^*\alpha(y_t - \ell_{t-1} + b_{t-1} + s_{t-m}), && \text{(34d)}\\
s_t &= s_{t-m} + \gamma^*(1-\alpha)(y_t - \ell_{t-1} + b_{t-1} + s_{t-m}), && \text{(34e)}
\end{aligned}$$

and by setting $\beta = \beta^*\alpha$, $\gamma = \gamma^*(1-\alpha)$ and substituting, we get the governing equations for the ETS(A,A,A) model given by

$$\begin{aligned}
y_t &= \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t, && \text{(35a)}\\
\ell_t &= \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t, && \text{(35b)}\\
b_t &= b_{t-1} + \beta\varepsilon_t, && \text{(35c)}\\
s_t &= s_{t-m} + \gamma\varepsilon_t. && \text{(35d)}
\end{aligned}$$

Recalling the general form of the linear state space model, we define $\boldsymbol{w}, \boldsymbol{g}, \boldsymbol{F}$ and the state vector $\boldsymbol{x_t}$ as follows

$$\boldsymbol{w} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \boldsymbol{x_t} = \begin{bmatrix} \ell_t \\ b_t \\ s_t \\ s_{t-1} \\ \vdots \\ s_{t-m+2} \\ s_{t-m+1} \end{bmatrix}, \boldsymbol{g} = \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{F} = \begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$

In a similar fashion we can derive the rest of the linear innovations state space models so, we can give a complete overview of all of them in the next table.

Table 3: Linear Innovations State Space Models

| Trend | Seasonal N | Seasonal A |
|---|---|---|
| N | $y_t = \ell_{t-1} + \varepsilon_t$<br>$\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ | $y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$<br>$\ell_t = \ell_{t-1} + \alpha\varepsilon_t$<br>$s_t = s_{t-m} + \gamma\varepsilon_t$ |
| A | $y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$<br>$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$<br>$b_t = b_{t-1} + \beta\varepsilon_t$ | $y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$<br>$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$<br>$b_t = b_{t-1} + \beta\varepsilon_t$<br>$s_t = s_{t-m} + \gamma\varepsilon_t$ |
| $A_d$ | $y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$<br>$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$<br>$b_t = \phi b_{t-1} + \beta\varepsilon_t$ | $y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$<br>$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$<br>$b_t = \phi b_{t-1} + \beta\varepsilon_t$<br>$s_t = s_{t-m} + \gamma\varepsilon_t$ |

### Linear Heteroscedastic State Space Models

In this section we introduce a special case of generally speaking nonlinear and heteroscedastic innovations state space models that can be rewritten in a linear form. We are going to demonstrate this on one of such models and further conclude by listing all of the rest in a table.

We firstly present the general innovations form of a nonlinear state space model.

$$y_t = w(\boldsymbol{x}_{t-1}) + r(\boldsymbol{x}_{t-1})\varepsilon_t, \tag{36a}$$
$$\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_{t-1}) + \boldsymbol{g}(\boldsymbol{x}_{t-1})\varepsilon_t, \tag{36b}$$

where $\boldsymbol{f}(\cdot), \boldsymbol{g}(\cdot)$ are vector functions, $w(\cdot), r(\cdot)$ are scalar functions and $\varepsilon_t$ is a white noise process. It is reasonable enough to assume that the process is Gaussian, as in the case we are working with, it is good enough approximation.

In a similar way as we did in the case of linear innovations state space models, by setting

$$\varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}}, \tag{37}$$

where $\hat{y}_{t|t-1}$ is the 1-step ahead forecast of the real value $y_t$ and by similar substitutions we can derive the governing equations of ETS(M,A,N)

$$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t), \tag{38a}$$
$$\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t), \tag{38b}$$
$$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t, \tag{38c}$$

whereby recalling the general form (36) we can denote

$$\boldsymbol{x}_t = \left[\ell_t, b_t\right]', \quad w(\boldsymbol{x}_{t-1}) = r(\boldsymbol{x}_{t-1}) = \ell_{t-1} + b_{t-1},$$
$$\boldsymbol{f}(\boldsymbol{x}_{t-1}) = \left[\ell_{t-1} + b_{t-1}, b_t\right]', \quad \boldsymbol{g}(\boldsymbol{x}_{t-1}) = \left[\alpha(\ell_{t-1} + b_{t-1}), \beta(\ell_{t-1} + b_{t-1})\right]',$$

which defines the general state space structure for this model. But one can look at the equations in (38) and see that we can rewrite it into an alternative linear form

$$y_t = \boldsymbol{w}'\boldsymbol{x}_{t-1}(1 + \varepsilon_t), \tag{40a}$$
$$\boldsymbol{x}_t = (\boldsymbol{F} + \boldsymbol{g}\boldsymbol{w}'\varepsilon_t)\boldsymbol{x}_{t-1}, \tag{40b}$$

where

$$\boldsymbol{x}_t = \left[\ell_t, b_t\right]', \quad w = \left[1, 1\right]', \quad \boldsymbol{g}(\boldsymbol{x}_{t-1}) = \left[\alpha, \beta\right]', \quad \boldsymbol{F} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

In fact, we are able to rewrite in such a form all of the following models.

Table 4: Linear Heteroscedastic State Space Models

| Trend | Seasonal | |
|---|---|---|
| | N | A |
| N | $y_t = \ell_{t-1}(1 + \varepsilon_t)$<br>$\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ | $y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$<br>$\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$<br>$s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$ |
| A | $y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$,<br>$\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$,<br>$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ | $y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$<br>$\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$<br>$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$<br>$s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ |
| $\mathrm{A_d}$ | $y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$<br>$\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$<br>$b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ | $y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$<br>$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$<br>$b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$<br>$s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ |

#### 2.2.4 Model Estimation

So far, we introduced all the models we will be working with further on. To be able to produce forecasts we need to estimate not only the smoothing parameters $\alpha, \beta, \gamma$, but also the so-called seed state $x_0$. This we will divide into two procedures, namely initialization and estimation.

### Initialization

This section describes the way, how to find the initial seed state $x_0$. There are several ways of finding it. We follow a heuristic scheme which was designed by Hyndman.

The initial seed vector is of the form

$$x_0 = \left[l_0, b_0, s_0, \ldots, s_{-m+1}\right], \tag{42}$$

so, we start with the seasonal component, assuming the data is seasonal, by detrending the data $y_t$ with a moving average $f_t$ and computing each of the seasonal indices $s_0, \ldots, s_{-m+1}$ by averaging the detrended data $y_t - f_t$ for each season. The last step being the normalization of the found indices so they sum up to zero.

For the level and growth components, we fit a straight line $\ell + bt$ for the first ten observations of our, if necessary deseasonalized data, using the found seasonal indices, and then set $\ell_0 = \ell$ and $b_0 = b$. The choice of the number of observations being accounted for is quite arbitrary, ten leads to a good enough starting point with not much needless computational load.

This found seed vector is only a starting point for the next optimization procedure.

### Estimation

By computing the likelihood of the innovations state space models, one can arrive at the likelihood function dependent only on the parameters $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \phi)$, initial state vector $\boldsymbol{x}_0 = (\ell_0, b_0, s_0, \ldots, s_{-m+1})$.

$$\mathcal{L}^*(\boldsymbol{\theta}, \boldsymbol{x}_0) = n \log(\sum_{t=1}^{n} \varepsilon_t^2) + 2 \sum_{t=1}^{n} \log |r(\boldsymbol{x}_{t-1})|, \tag{43}$$

where the errors $\varepsilon_t$ are given by (33) for the homoscedastic case and by (37) for the heteroscedastic case.

So that by minimizing $\mathcal{L}^*$, we are able to acquire the optimal maximum likelihood estimates of $\boldsymbol{\theta}$ and $\boldsymbol{x}_0$.

Such an optimization procedure needs some starting values and in fact the optimality of these values is a problem in itself as a poor choice can lead to poor results. Therefore we use the heuristic approach for the seed vector $x_0$ shown before and past experience, starting values for $\boldsymbol{\theta}$ have been derived as such $\boldsymbol{\theta} = (0.1, 0.01, 0.01, 0.99)$.

We can also deduce some constraints for the values of each parameter as such

$$0 < \alpha < 1, \quad 0 < \beta < \alpha, \quad 0 < \gamma < 1 - \alpha, \quad 0 < \phi < 1, \tag{44}$$

but sometimes these might be too restrictive so we can co-opt a different set of constraints as we do in the case of ETS(A,N,N) and ETS(A,A,N)

$$0 < \alpha < 2, \quad 0 < \beta < 4 - 2\alpha. \tag{45}$$

### 2.2.5 Prediction Intervals

While point forecasts for each model were given by 2 it is valuable to compute in addition the corresponding prediction intervals, to give an idea about the uncertainty of the forecasted data.

For the classes of models we have mentioned, we are able to compute variance estimates analytically, which is a great advantage with respect to computing cost to the alternative of simulating these intervals.

Then for the linear innovations state space models the variance for each model is given by

$$v_{n+h|n} = Var(y_{n+h}|\boldsymbol{x}_n) = \begin{cases} \sigma^2, & \text{if } h = 1 \\ \sigma^2(1 + \sum_{j=1}^{h-1} c_j^2), & \text{if } h \geq 2, \end{cases} \tag{46}$$

where $c_j$ is given.

For the linear heteroscedastic state space models, the forecast variance is given by

$$v_{n+h|n} = (1 + \sigma^2)\theta_h - \hat{y}^2_{n+h|n} \,, \tag{47}$$

where

$$\theta_1 = \hat{y}^2_{n+h|n} \quad and \quad \theta_h = \hat{y}^2_{n+h|n} + \sigma^2 \sum_{j=1}^{h-1} c_j^2 \theta_{h-j} \,, \tag{48}$$

where $c_j$ is given.

For both cases we take the maximum likelihood estimator for $\sigma^2$, given by

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon_t}^2 \,, \tag{49}$$

where $\hat{\varepsilon_t}^2$ are the errors of the fitted values, that are obtained from the respective estimation procedures for each model as given by 2.2.4.

The $100(1-\alpha)\%$ prediction intervals are then obtained in the usual sense, that is $\hat{y}_{n+h|n} \pm z_{\alpha/2}\sqrt{v_{n+h|n}}$, where $z_{\alpha/2}$ is the $\alpha/2$ quantile of a Gaussian distribution.

# 3 Algorithm Implementation

In this chapter we will describe the implemented procedures and algorithms that were devised and are subsequently used for selecting suitable models for a given time series and for the combining of the forecasts they produce. Our aim was to come up with an algorithm that would do so completely automatically, without any human intervention, in the MATLAB environment, only specifying the time series to be forecasted and the forecast horizon.

## 3.1 Matlab

Matlab is a computational environment for matrix operations, technical calculations, and data visualization. Matlab has been developed for many years by MathWorks. Above the already implemented functions, MATLAB possesses a variety of different toolboxes, ranging across many fields of study.

When working with time series data, MATLAB offers many of these toolboxes, that include additional functionalities such as the Curve Fitting Toolbox, Optimization Toolbox Econometrics Toolbox, Signal Processing Toolbox, Financial Instruments Toolbox, Financial Toolbox.

We are using the MATLAB Version: 9.4.0.813654 (R2018a).

## 3.2 Model Selection

An important pre-step to the combination procedure itself, is the model selection. There is a variety of ways how to select such models. A possible approach would be to take every available model and proceed with all of them to the forecast combination. But a reasonable restriction to these values might help the procedure, as nonsensical models won't negatively affect the resulting combination.

Therefore, we implement an algorithm that selects reasonable models from both ARIMA and ETS types to be combined. As we want to be able to model any given time series, we need to be able to address these key qualities, that is heteroscedascity and seasonality. In other words given that a time series is either heteroscadastic or homoscedastic and either seasonal or nonseasonal, we want our algorithm to be able to sort any time series into these four categories:

1. Homoscedastic and nonseasonal.
2. Homoscedastic and seasonal.
3. Heteroscedastic and nonseasonal.
4. Heteroscadastic and seasonal.

The corresponding models in said categories can be summed up into the tables 5 and 6.

| Data | Nonseasonal | Seasonal |
|---|---|---|
| Homoscedastic | ETS{(A,N,N),(A,A,N), $(A,A_d,N)$} | ETS{$(A,N,A)_{12}$,$(A,A,A)_{12}$, $(A,A_d,A)_{12}$}) |
| Heteroscedastic | ETS{(M,N,N),(M,A,N), $(M,A_d,N)$}) | ETS{$(M,N,A)_{12}$,$(M,A,A)_{12}$, $(M,A_d,A)_{12}$}) |

Table 5: ETS model Categories

| Data | Nonseasonal | Seasonal |
|---|---|---|
| Homoscedastic | ARIMA | SARIMA |
| Conditonaly Heteroscedastic | ARIMA-GARCH | SARIMA-GARCH |

Table 6: ARIMA model Categories

#### 3.2.1 ETS

Even though MATLAB offers many helpful functionalities regarding time series analysis, a framework for the ETS models, as to our knowledge, is not yet present in any of the many toolboxes. The state space class unfortunately doesn't come in a SSOE variant, which is the case of ETS models, and therefore isn't suited for this task. We therefore implement our own models, together with the functions for initialization, parameter estimation, forecasting of the mean (point forecast) and the confidence intervals as they were described in the chapter about ETS models 2.2.

For the case of the ETS models, with the selection procedure we have a much easier task at hand. We have 12 models in total to choose from and an easy way of choosing between them at that. As stated in [1], [14], recalling the likelihood function (43) that we use for parameter estimation, we can utilize it in computing an information criterion

$$\mathrm{AIC} = -2\mathcal{L}^* + 2k\,, \tag{50}$$

where $\mathcal{L}^*$ is the log-likelihood of the model and $k$ is the total number of parameters and initial states that have been estimated. AIC is called the Akaike information criterion and is often used, not only in the ETS framework, for model selection, by choosing the model with the least value of AIC.

Therefore, by computing the AIC, we find the model with its least value and which then defines the category we are dealing with, as given by Table 5. This concludes the model selection process for the ETS models.

An important remark is to be made about this selection procedure, regarding seasonality. In fact, if we compare seasonal and nonseasonal models in this way, what we are actually asking is, what is the likelihood that a seasonal component, recall time series decomposition 1.5, is present in the data.

So, if the procedure points to a seasonal model, then we can say that it is likely that the data can be modeled by a seasonal model. We acknowledge that such an approach might not be rigorous, but we discovered that a possibly more proper technique, such as spectral analysis, might fail and even in some very obvious cases of seasonal behaviour.

Therefore, we use our findings concerning seasonality for further analysis, that is for model estimation of the ARIMA models.

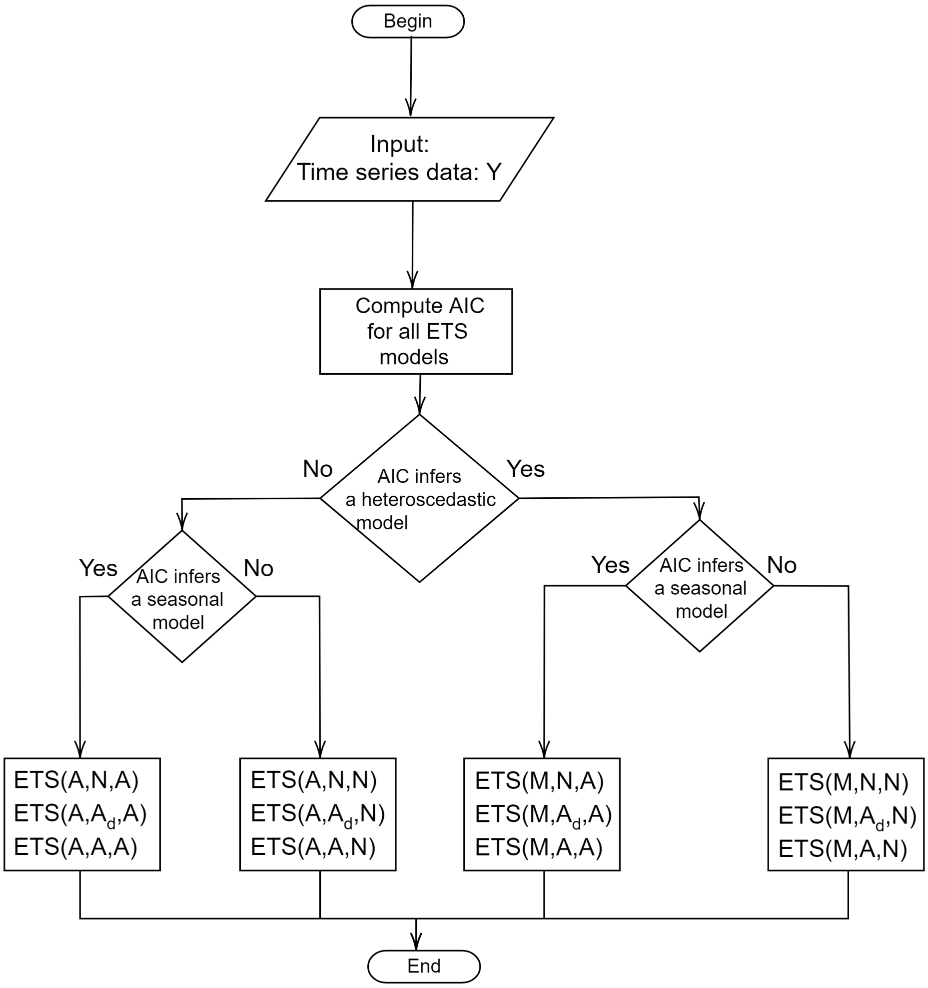


Figure 4: ETS model selection algorithm

### 3.2.2 Box-Jenkins Models

As we mentioned, the question of seasonality has been answered by the procedure detailed before. Of course, if we recall, it also concluded about the nature of the variance, that is if the data is heteroscedastic or not. Here we have to pause for a bit and elaborate on the way that we can deal with heteroscedascity in our Box-Jenkins framework and explain why we don't use the information about the behaviour of the variance the same way as about seasonality.

The nature of the heteroscedascity in the sense of GARCH modeling is understood as conditional heteroscedascity defined in chapter 1.2, also described as having an ARCH effect, while its nature in the scope of ETS modeling is broader and has a rather general meaning.

It is unclear whether these two notions can be equated, however we don't do so and instead we define a procedure that tests the data on conditional heteroscedascity, using the Engel's ARCH test, which will be described.

The now introduced was influenced by [11] that describes the auto.arima procedure in R and by [10] where a theoretical foundation was proposed for a procedure not unlike auto.arima.

### Stationarizing

We start by stationarizing the given data in the sense of trend. We recall the condition regarding Box-Jenkins models, that is the time series has to be stationary (trend-stationary). To do so we apply differencing, either seasonal or nonseasonal. The question is whether to difference or not and if so, how many times. To this we implement a procedure which checks for the existence of unit root, seasonal and/or nonseasonal, and defines the difference parameters $D$ and $d$ respectively, according to (18).

To do so we utilize the already in MATLAB implemented KPSS unit root test, which Assesses the null hypothesis that a univariate time series is trend stationary against the alternative that it is a nonstationary unit root process. As described in the MATLAB documentation [3], the test uses the structural model:

$$\begin{aligned} y_t &= c_t + \delta t + u_{1,t} \,, \\ c_t &= c_{t-1} + u_{2,t} \,, \end{aligned} \tag{51}$$

where $\delta$ is the trend coefficient, $u_{1,t}$ is a stationary process, $u_{2,t}$ is an independent and identically distributed process with mean 0 and variance $\sigma^2$.

Stating the null hypothesis that $\sigma^2 = 0$, implying that the random walk term $c_t$ is constant and functions as the models' intercept. On the other hand, the alternative hypothesis states that $\sigma^2 > 0$, which introduces the unit root in the random walk.

The idea is to run the test and difference the time series until the null hypotheses is accepted, setting $d$ as number of times the data was differenced, with the limit being set up as $d \leq 2$.

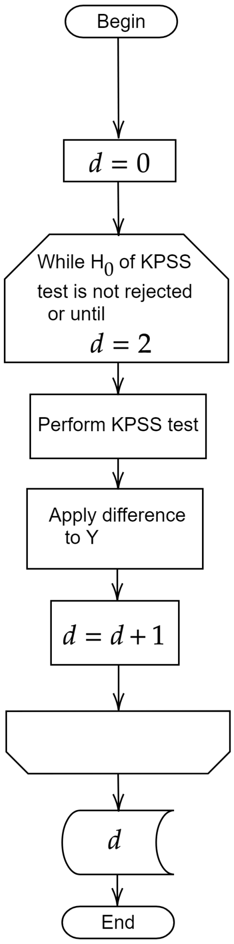


Figure 5: Scheme for KPSS testing and nonseasonal differencing order selection

If the data is seasonal, we also need to determine the seasonal difference. To this purpose we look at the ACF of the time series that was nonseasonaly differenced accordingly. If the values of the ACF decay very slowly, that that they do not cross the confidence bounds up to the given seasonal period, we consider a unit root present and regard it as to be seasonal. Thereby setting $D = 1$. Otherwise $D = 0$.

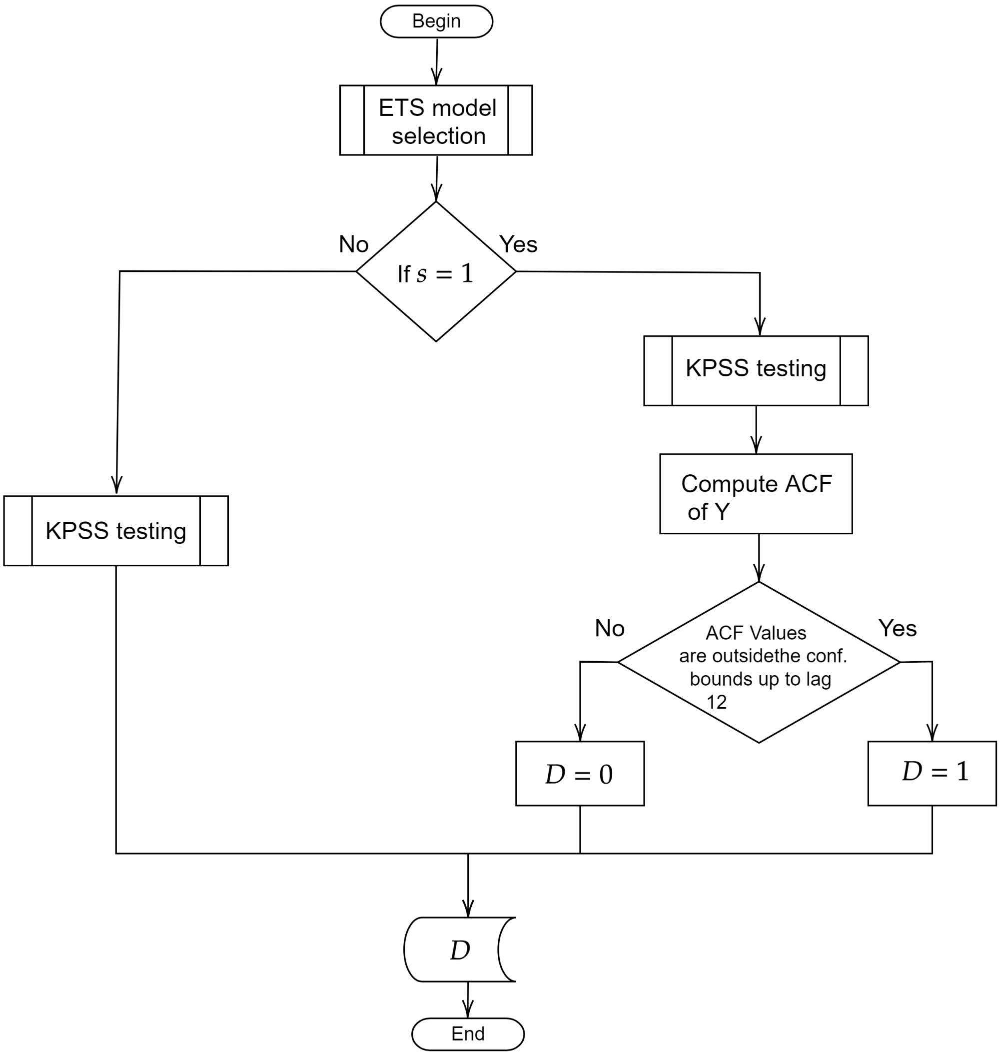


Figure 6: Scheme for stationarizing with respect to seasonal and nonseasonal differencing

But because of the danger of overdifferencing, if it is the case that $d = 2$, then $D$ is automatically set to zero, as it is convention for the total number of differences not to exceed two, i.e. $d + D \leq 2$.

### Checking for Conditional Heteroscedascity

As we already mentioned, it is necessary to check for heteroscedascity, with respect to the way it shall be treated. Consistent with

We apply the already in MATLAB implemented Engle's ARCH test, which is a Lagrange multiplier test to assess the significance of ARCH effects, that is time series exhibiting conditional heteroscedasticity or autocorrelation in the squared series, as described in the MATLAB documentation [3].

Consider a time series

$$y_t = \mu_t + \varepsilon_t \,, \tag{52}$$

where $\mu_t$ is the conditional mean of the process, and $\varepsilon_t$ is an innovation process with mean zero.

Consider that the innovations are given by

$$\varepsilon_t = \sigma_t z_t \,, \tag{53}$$

where $z_t$ is an independent and identically distributed process with mean zero and variance one. Therefore

$$E(\varepsilon_t \varepsilon_{t+h}) = 0 \,, \tag{54}$$

for all lags $h \neq 0$ and the innovations are uncorrelated.

Let $H_t$ be the history of the process at time $t$. The conditional variance of $y_t$ is

$$Var(y_t|H_{t-1}) = Var(\varepsilon_t|H_{t-1}) = E(\varepsilon_t^2|H_{t-1}) = \sigma_t^2 \,. \tag{55}$$

Thus, conditional heteroscedasticity in the variance process is equivalent to autocorrelation in the squared innovation process.

Define the residual series

$$\varepsilon_t = y_t - \hat{\mu}_t \,. \tag{56}$$

If all autocorrelation in the original series, $y_t$, is accounted for in the conditional mean model, then the residuals are uncorrelated with mean zero. However, the residuals can still be serially dependent.

The alternative hypothesis for Engle's ARCH test is autocorrelation in the squared residuals, given by the regression

$$H_a : \varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_1 \varepsilon_{t-m}^2 + u_t \,, \tag{57}$$

where $u_t$ is a white noise error process. Subsequently the null hypothesis is

$$H_0 = \alpha_0 = \alpha_1 = \cdots = \alpha_m = 0 \,, \tag{58}$$

where $m$ is the lag, given by the GARCH(P,Q) model, and we set $m = P + Q$.

Therefore, if the ARCH test fails to reject the null hypothesis, we consider the data heteroscedastic and try to model its conditional variance by a GARCH(1,1) and the conditional mean by an ARIMA (or SARIMA) process. The selection of GARCH(1,1) specificaly might seem superficial, but it was shown [**?**] that GARCH(1,1) outperforms all the other models most of the time in financial time series.

### Selecting parameters for the Autoregressive and Moving Average parts

The Box-Jenkins methodology offers a way of selecting the model parameters, by looking at the ACF and PACF of the time series, as the theory gives clear description of ACF and PACF in case of typical AR, MA or mixed ARMA processes. which defines a certain pre-model, which is then applied to the data. The residuals are again checked for any autocorrelation and the model is then updated accordingaly, if necessary. Such a procedure, rather tentative in nature, yielding usually more than one model, requires some experience in ARIMA model building.

Assuming that the data is differenced accordingly, we use already implemented functions, which give us the values of ACF and PACF for a given lags. Taking these maximal lags as $lag_{max} = 6$, we search for the last value of the lag for which the ACF value is outside the confidence bounds, before they cut off. This lag value sets the maximal order of the MA($q$) process $1 \leq q_{max} \leq 6$. The same is done for the PACF, which at the end gives us the maximal order of the AR($p$) process $1 \leq p_{max} \leq 6$.

For the orders of the seasonal parts of the SARIMA model, the orders $P, Q$ are chosen so that they both attain the value zero or one, $0 \leq P, Q \leq 1$. The whole process can be given by the flowchart.

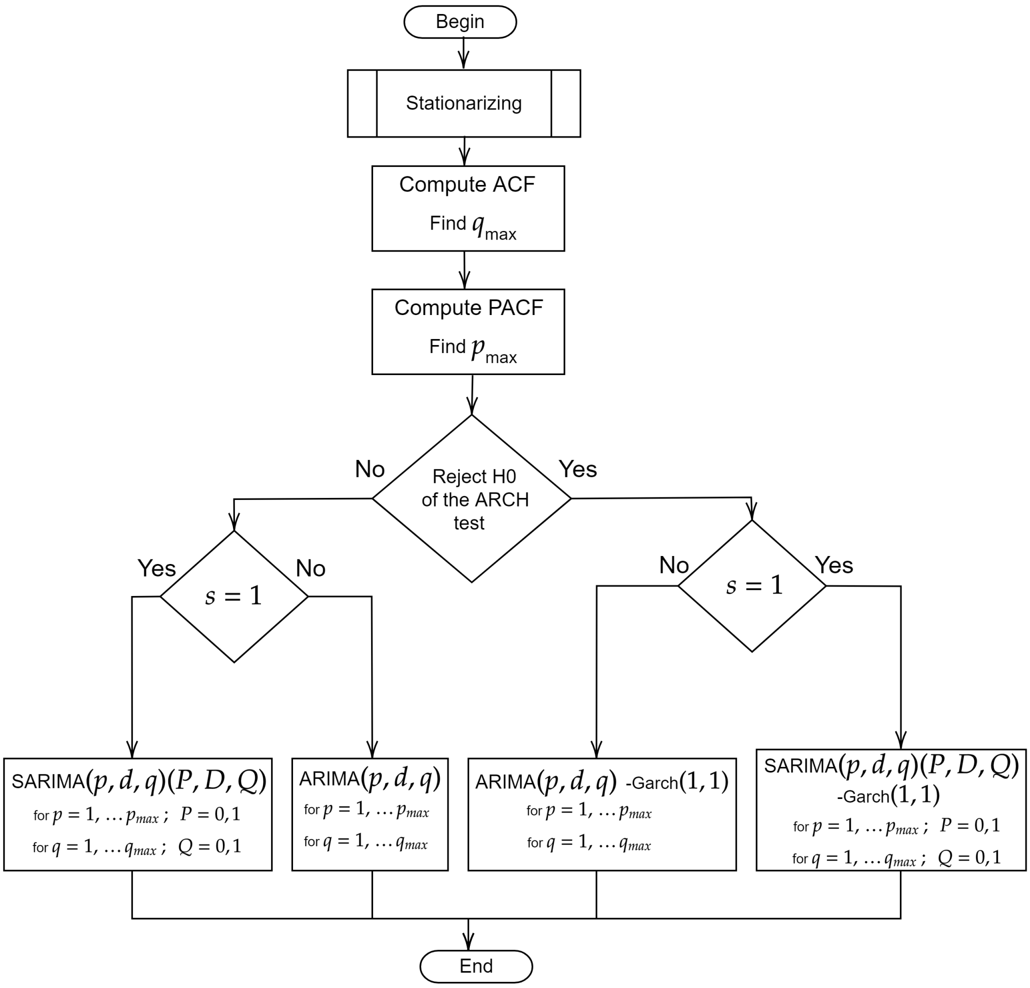


Figure 7: Box-Jenkins type model selection

## 3.3 Forecast Combination

Generally, as stated in [13], [17] the rule to forecast combination is to use either different data or methods. As we are concerned with the univariate case of time series analysis, the variablity in mehtods is what we have left. The aim is to use such methods, that vary in the assumptions, which they are based upon. It was shown that by such an approach, one is able to reduce error.

The majority of research in forecast combination have dealt with minimizing forecast error, i.e. combining for improvement of performance. Less thought has been dedicated to what Yang [8] called combining for adaptation, that is more concerned with the problem of model selection itself.

We consider an algorithm proposed by Yang [8] used by Yang and Zuo [9] to reduce the risk of model selection error and to produce an optimization procedure that deals with the potentialy ambiguous process of ARIMA model selection.

### 3.3.1 AFTER Algorithm

Given a time series with values $y_t$ at time $t$ that is to be forecasted. Let $\Delta = \{\delta_1, .., \delta_M\}$ be the set of all forecasting procedures. Consider that for each forecasting procedure $\delta_j$ along with the forecasts $\hat{y}_{j,t}$ the variance estimates $\hat{v}_{j,t}$ at time $t$, are obtained by the individual forecast approaches.

We define the Aggregated Forecast Through Exponential Re-weighting (AFTER) algorithm. Let $\underline{\pi} = \{\pi_j \geq 0 : j = 1, \ldots, M\}$ be the sequence of the primary weights, such that they sum up to zero. Set $W_{j,1} = \pi_j$ and for $n \geq 2$ let

$$W_{j,t} = \frac{\frac{\pi_j}{\prod_{i=1}^{t-1} \hat{v}_{j,i}^{1/2}} \exp(-\frac{1}{2} \sum_{i=1}^{t-1} \frac{(y_i - \hat{y}_{j,i})^2}{\hat{v}_{j,i}^{1/2}})}{\sum_k^M \frac{\pi_k}{\prod_{i=1}^{t-1} \hat{v}_{k,i}^{1/2}} \exp(-\frac{1}{2} \sum_{i=1}^{t-1} \frac{(y_i - \hat{y}_{k,i})^2}{\hat{v}_{k,i}^{1/2}})}, \tag{59}$$

be the weights for the $j$-th procedure at time $t$. The forecast combination is then found accordingly, to the weights

$$\hat{y}_{c,t} = \sum_{j=1}^{M} W_{j,t} \hat{y}_{j,t} \,. \tag{60}$$

We can clearly see that the distribution of weights (59) depends not only on the forecasts a model produces, but also on the variance estimates. So, the algorithm produces smaller weights for procedures with bad forecasts and/or variance estimation so that those prospectively don't affect the models, which perform better in that regard.

## 3.4 Variance Covariance Method

We introduce the variance covariance method as it is described in [16], given only two forecasts to be combined, as it will be the case for our approach.

The method itself is based on the minimization of the variance of the combined forecast errors, given by

$$\text{Var}(\epsilon_t^c) = w^2 Var(\epsilon_{1,t}) + (1-w)\text{Var}(\epsilon_{2,t}) + 2w(1-w)\text{Cov}(\epsilon_{1,t}, \epsilon_{2,t}) \,, \tag{61}$$

where $w$ is the weight to be found and the forecast error is given by

$$\epsilon_{t,c} = w\epsilon_{t,1} + (1-w)\epsilon_{t,2} \,, \tag{62}$$

such that $\epsilon_{t,1}$ and $\epsilon_{t,2}$ correspond to the forecast errors of the two forecast models. Minimizing $Var(\epsilon_t^c)$ over $w$ yields

$$w^* = \frac{\text{Var}(\epsilon_{2,t}) - \text{Cov}(\epsilon_{1,t}, \epsilon_{2,t})}{\text{Var}(\epsilon_{1,t}) + \text{Var}(\epsilon_{2,t}) - 2\text{Cov}(\epsilon_{1,t}, \epsilon_{2,t})}. \tag{63}$$

The procedure can be generalized for an arbitrary number of forecasts. The weights $W$ are given by a column vector, where each element of $W$ is a weight $w_i$. Then we can find $W$ by minimizing $W^T \Sigma W$, where $\Sigma$ is the covariance matrix. In practice then substitute instead of variances and covariances their respective samples.

## 3.5 Prediction Intervals

To give the final model more information about the overall behavior of the forecast, we add a prediction interval constructed by computing the variance of the combined model $\sigma^2_{c,t-h}$ as shown in [12].

The prediction intervals are given by the variance of the combined forecast, which is defined by

$$\hat{y}_{c,t-h} = \sum_{j=1}^{M} w_{j,t-h} \hat{y}_{j,t-h} \,, \tag{64a}$$

$$\sigma^2_{c,t-h} = \sum_{j=1}^{M} w_{j,t-h} (\hat{y}_{j,t-h} - \hat{y}_{c,t-h}) + \sum_{j=1}^{M} w_{j,t-h} \sigma^2_{j,t-h} \,, \tag{64b}$$

where $\hat{y}_{j,t-h}$ and $\sigma^2_{j,t-h}$ are the forecasts and variances by the individual models, which define, with respect to the weights $w_{j,t-h}$, the combined forecast $\hat{y}_{c,t-h}$ and its corresponding variance $\sigma^2_{c,t-h}$. The prediction intervals are then simply computed as usual, that is $\hat{y}_{n+h|n} \pm z_{\alpha/2} \sqrt{v_{n+h|n}}$, where $z_{\alpha/2}$ is the $\alpha/2$ quantile of a Gaussian distribution.

# 4 Financial Time Series Application and Results

In this chapter, we will apply our algorithm on a set of various financial time series data. Our aim is to compare the described approach with the individual models applied themselves and a combination of both. Therefore, we define the first competing model as the best ETS model, defined by minimizing the AIC, which has already been identified within the selection procedure.

The second model from the Box-Jenkins class is obtained by computing the BIC of all the models returned by our selection procedure and choosing the model minimizes this criterion. Such an approach is similar to an automatic ARIMA model selection, found in other data analysis software.

And the third competing model will be the combination of both the models mentioned above using the variance covariance method. The reason is to study, whether the AFTER algorithm is able improve the forecasts, so that this double combination outperforms, in the sense of our accuracy measures, this model.

We denote these models as follows.

1. Our combination model - CombMod.
2. The best Box-Jenkins model - ARIMAbest.
3. The best ETS model - ETSbest.
4. Var-Cov combination of ARIMAbest ETSbest - CombBest.

The individual real financial time series are chosen so that they represent a variety of different types of time series, to examine the performance under diverse behaviour, that one can be confronted with, while forecasting financial data.

## 4.1 Applying the Algorithm on Real Data

### 4.1.1 Data Preprocessing

The time series data before being used for estimation is at first divided into parts. The validation part is not considered in this description, as the validation process itself is done only in the case our application examples to assess the accuracy of the method. Generally, the aim of the function is to produce forecasts of yet unknown observations.

Keeping that in mind, we consider the time series $Y = \{y_t : t = 1, \ldots, n\}$ as the input dataset. Moreover, the forecast horizon $T$ is specified. This defines the partition of the data as such, $Y_{\text{Est}} = \{y_t : t = 1, \ldots, n - T\}$ being the estimation data and $Y_{\text{Train}} = \{y_t : t = n - T + 1, \ldots, n\}$ being the train data.

### 4.1.2 Estimation

The models are then selected as described in chapter 3.2 generating two sets of models ETSchosen and ARIMAchosen, those are estimated with respect to the estimation data $Y_{\text{Est}}$, which should contain enough values, so that a meaningful forecast can be produced. We recommend a rule of thumb for the is not to drop below a 75/25 ratio for the $Y_{\text{Est}}$ and $Y_{\text{Train}}$ respectively.

### 4.1.3 Forecasting

The estimated models from the sets ETSchosen and ARIMAchosen are then used to produce two sets of forecasts each ARIMAtrain, ARIMAfor, ETStrain, ETSfor, corresponding to the train and actual or target forecasts, which were obtained by using $Y$ as presample data. The train forecasts ARIMAtrain, ETStrain are then used in the AFTER algorithm to produce weights, with respect to $Y_{\text{Train}}$ data, which are subsequently used on $\hat{Y}$ to produce the weighted forecast, which is to be summed up to obtain the combined forecasts as shown in (60) and denoted ARIMAafter, ETSafter.

#### 4.1.4 Combination

We define two combination forecasts. The first one is used as a benchmark in the validation process to the actually studied forecast combination. By selecting two models from each of the sets ETSchosen, ARIMAchosen, so that one reduces the AIC and the other the BIC respectively, as it is custom. This selection generates two model, denoted ETSbest and ARIMAbest. The forecast CombBest is derived via the variance-covariance procedure applied on ETSbest and ARIMAbest. As we said, this model is devised as a competing model to our actual one denoted CombMod.

CombMod is defined as the combination of ARIMAafter and ETSafter, that is, two weighted averages, with weights returned by the AFTER algorithm applied to ARIMA-train and ETStrain, with respect to $Y_{\text{Train}}$.

#### 4.1.5 Validation

The validation procedure is done on data that hasn't been used in the estimation neither the training procedure, but is a set of real observations that continue beyond the time series $Y$ as we defined it before. We define $Y_{val} = \{y_t : t = n+1, \ldots, n+T\}$ as being the validation dataset, such that in the case of the following examples the real time series is $Y_{\text{real}} = Y + Y_{\text{val}}$.

The validation process itself done by studying the forecast error measures we defined 1.6.1 and infering information about the forecasting performance of CombMod.

## 4.2 Dow Jones Industrial Average Index

The first time series we will apply our algorithm to is the SPDR Dow Jones Industrial Average ETF, which is an exchange traded fund, following the Dow Jones Industrial Average Index - DIA, issued by the State Street corp., under the brand SPDR.

The DIA is the well known DOW Jones Industrial Average Index - DIA. It follows 30 of the biggest corporations on the market of the United States, such as Apple, Coca Cola, Microsoft and others. It's one of the more popular indexes as it is said to reflect the climate in the US industry, as the index encompasses corporation from all various industrial and business branches, except the cases of transportation and engineering networks.

It's one of the longest recorded indexes in the world, with more than hundred and twenty years of data has been a staple in the financial market.

A certain drawback might be the low number of contributors to the index, as from all of the more than five thousand stocks, only 30 are chosen. On the other hand, that ensures that the index is much less volatile, and it is in general considered to be very stable. Therefore, there exist several index fonds, such as the one we are using, following the DIA, thus being able to forecast it precisely surely be of great importance.

### 4.2.1 The Data

The data of SPDR Dow Jones Industrial Average ETF is collected as its price in USD over length of 119 months from February 1998 until December 2007. This decision is made because of the serious drop on value all over the financial market, due to the financial crisis of 2008. Our method is not generally devised to handle such shocks, but moreover we are interested to firstly model data with rather stable behaviour, leaving a more unstable behaviour to another example. The data obtained from the site investing.com and is available at [18].

We consider three trials of forecasting for three different forecast horizons, to study the performance of long-, medium- and short-term forecasts. These are $h = \{6, 12, 18\}$. Recalling data preprocessing 4.1.1 we see that the change of the horizon, changes the training and estimation sets, therefore different model selections are possible for the distinct horizons, while the estimated models are sure to be different.

### 4.2.2 Results

For the forecast horizon $h = 6$, the algorithm first selected the best ETS model, ETSbest, which was the ETS(A,N,N). That defined the group of ETS models to be used in the AFTER algorithm as given in the table 8. Knowing that there is no seasonal term, the corresponding selected Box-Jenkins models are given in table 8. Where ARIMA(0,1,0) was picked as the best model ARIMAbest.

| Chosen best models | |
|---|---|
| ARIMAbest | ARIMA(0,1,0) |
| ETSbest | ETS(A,N,N) |

Table 7: Best models returned by the selection procedure

| Chosen models | |
|---|---|
| ETS | ETS(A,N,N), ETS(A,A,N), ETS(A,$A_d$,N) |
| Box-Jenkins | ARIMA(0,1,0), ARIMA(0,1,1), ARIMA(1,1,0), ARIMA(1,1,1) |

Table 8: Models returned by the selection procedure

For all these models, their respective forecasts and forecast variances were computed. As we see the model selection didn't vary with the forecast horizons, we summarized it into only one table 8.

We compile the results of error measures into the table 9 containing the computations of the forecast error measures calculated on the validation data, for each of the competing models, at the respective forecast horizons.

| Horizon | Measure | CombMod | ARIMAbest | ETSbest | CombBest |
|---|---|---|---|---|---|
| $h = 6$ | MSE | 8.2428 | 8.3120 | 8.5353 | 8.5351 |
| | MAPE | 1.8780 | 2.0250 | 1.7321 | 1.7322 |
| $h = 12$ | MSE | 74.8491 | 59.4086 | 87.1666 | 87.1645 |
| | MAPE | 5.6090 | 5.0276 | 6.0412 | 6.0412 |
| $h = 18$ | MSE | 304.6319 | 222.7954 | 310.4736 | 310.4706 |
| | MAPE | 11.7517 | 10.0711 | 11.9007 | 11.9007 |

Table 9: The forecast error measures for given forecast horizons

We plot the the forecast of CombMod given by the different horizons.

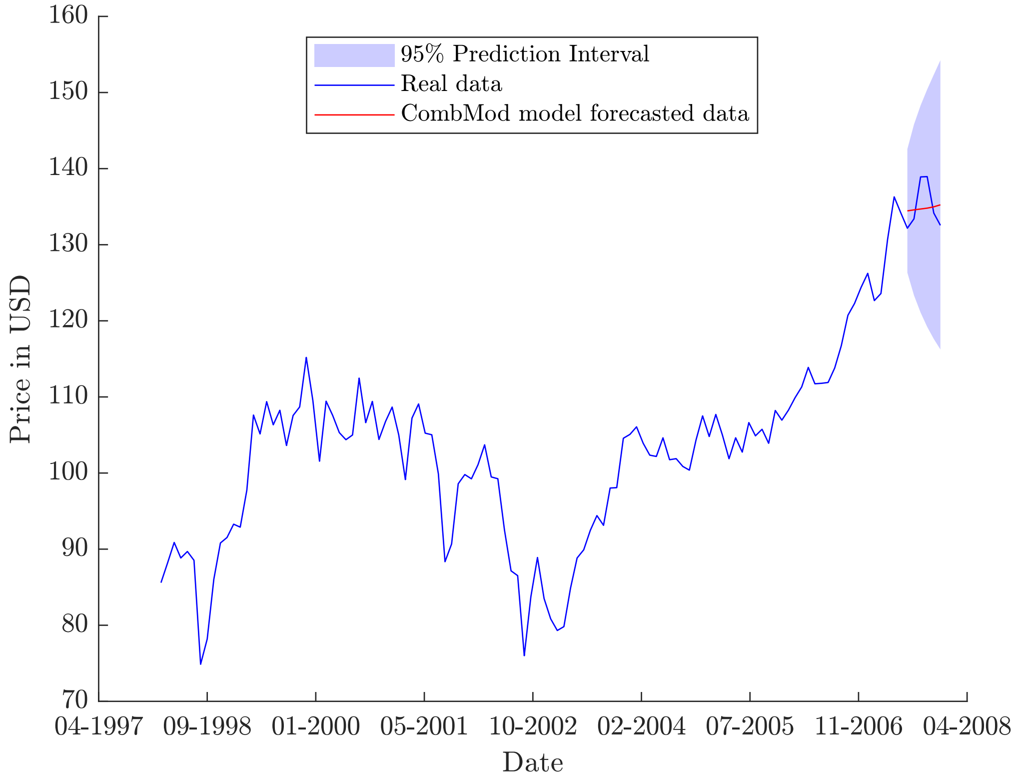


Figure 8: CombMod forecasts of DIA data for $h = 6$

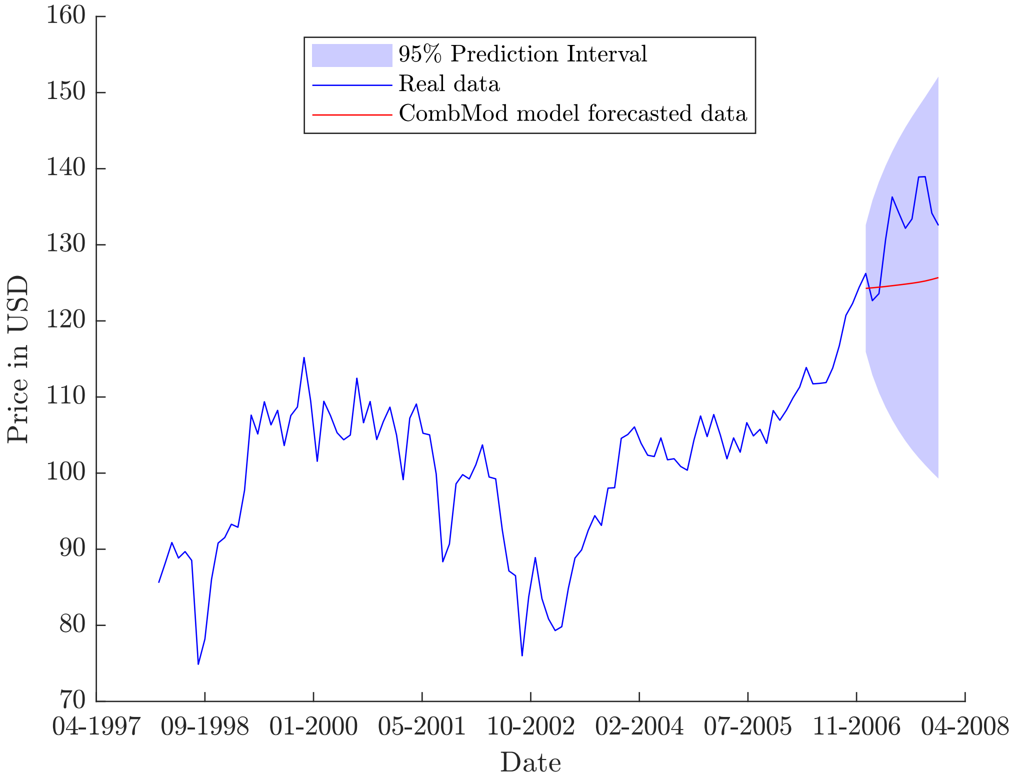


Figure 9: CombMod forecasts of DIA data for $h = 12$

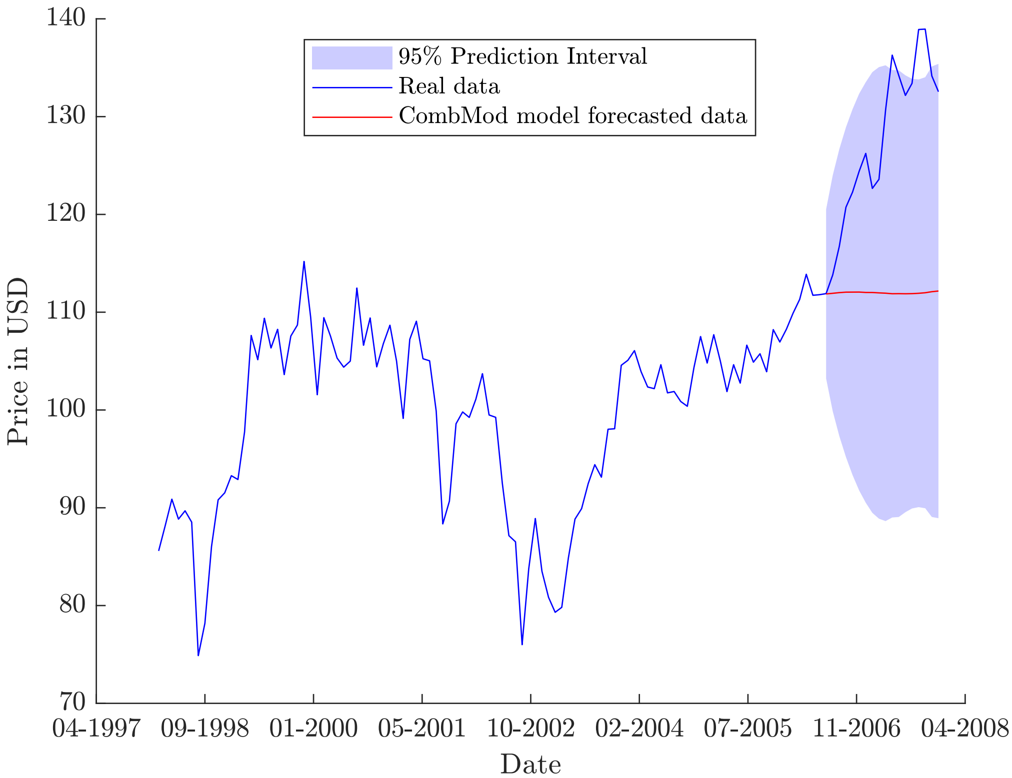


Figure 10: CombMod forecasts of DIA data for $h = 18$

Further we show the weights given to each model by the variance-covariance combination.

| h | CombMod | | CombBest | |
|---|---|---|---|---|
| | ARIMAafter | ETSafter | ARIMAbest | ETSbest |
| 6 | $1.86 \times 10^{-9}$ | 1 | $8.03 \times 10^{-5}$ | 0.99992 |
| 12 | $8.54 \times 10^{-10}$ | 1 | $6.50 \times 10^{-5}$ | 0.99994 |
| 18 | 0.081619 | 0.91838 | $3.07 \times 10^{-5}$ | 0.99997 |

Table 10: Weights given by the variance-covariance combination

#### 4.2.3 Discussion

We firstly examine the performance of our model with respect to the competing model, which combines the best models from the two methodologies. We clearly see that our model outperforms it at nearly every aspect and every forecast horizon. Overall, we can say that our model performed well at the short-term horizon, but didn't do so well, with respect to the ARIMAbest model on longer horizons.

The weights of the variance covariance combining method clearly favoured the combined ETSafter model giving it weight of nearly 1. The weights of the ETSafter were distributed fairly evenly, with a slight preference for the ETS(A,A,N) model, especially in the short- and medium-term. So as the ARIMA model clearly performed better at the actual forecast, the combined model couldn't due to the weights reflect that.

The rather bad performance of CombMod at the long-term horizon is even visible looking at forecasts and the prediction intervals, which even fails to contain the last values in itself.

### 4.3 Coca Cola

Coca Cola Company produces and distributes a wide range of beverages worldwide. It is one of America's largest corporations. Nowadays, there are over 500 different brands in more than 200 countries. Shares are traded on the New York Stock Exchange and are part of the Dow Jones Industrial average, the S&P 500 and the Russell 1000 Growth Index.

The Coca Cola Company stocks are well known to be a stable and in the long run very profitable investment. Even such a giant in the financial market as Warren Buffet is holding share in the company, already for many years. As for expert evaluation, this would certainly be a good argument.

#### 4.3.1 Data

The data of Coca Cola Company stocks is collected as its price in USD per share over length of 73 months from January 2013 until January 2019. A smaller time interval was chosen, to study as to what degree the short time interval affects the performance of the forecast, especially considering the long-term prediction. The dataset was collected from the site investing.com and is available at [18].

### 4.3.2 Results

For the short- and medium-term forecast given by the horizons $h = 6$ and $h = 12$, the selection procedure returns the best ETS model, ETSbest, as the ETS(A,N,N). That defined the group of ETS models to be used in the AFTER algorithm as given in the table 12, where the selection procedure for the ARIMA models returned a rather large number of models, given by the large value of $q_{\max}$. The reason as to why that might be, will be discussed. As the best ARIMA model ARIMAbest, ARIMA(1,0,1) was chosen for $h = 12$ ARIMA(1,0,0) was chosen instead.

For the pool of ARIMA models shrunk, but the same ARIMAbest model ARIMA(1,0,0) as before was chosen as well. On the other hand, the ETSbest was returned as ETS(M,N,N). Clearly the loss of information due to the longer horizon, caused problems in the model selection, which consequences shall be discussed in the conclusion.

| Horizon | Chosen best models | |
|---|---|---|
| h=6 | ARIMAbest | ARIMA(1,0,1) |
| | ETSbest | ETS(A,N,N) |
| h=12 | ARIMAbest | ARIMA(1,0,1) |
| | ETSbest | ETS(A,N,N) |
| h=18 | ARIMAbest | ARIMA(1,0,1) |
| | ETSbest | ETS(M,N,N) |

Table 11: Best models returned by the selection procedure

| Horizon | Chosen models | |
|---|---|---|
| h=6,12 | ETS | ETS(A,N,N), ETS(A,A,N), ETS(A,$A_d$,N) |
| | Box-Jenkins | ARIMA(1,0,1), ARIMA(1,0,0), ARIMA(1,0,2), ARIMA(1,0,4), ARIMA(1,0,3), ARIMA(1,0,5), ARIMA(1,0,6), ARIMA(0,0,3), ARIMA(0,0,6), ARIMA(0,0,4), ARIMA(0,0,5), ARIMA(0,0,2), ARIMA(0,0,1), ARIMA(0,0,0) |
| h=18 | ETS | ETS(M,N,N), ETS(M,A,N), ETS(M,$A_d$,N) |
| | Box-Jenkins | ARIMA(1,0,1), ARIMA(1,0,0), ARIMA(1,0,2), ARIMA(1,0,4), ARIMA(1,0,3), ARIMA(1,0,5), ARIMA(0,0,3), ARIMA(0,0,4), ARIMA(0,0,5), ARIMA(0,0,2), ARIMA(0,0,1), ARIMA(0,0,0) |

Table 12: Models returned by the selection procedure

| Horizon | Measure | CombMod | ARIMAbest | ETSbest | CombBest |
|---|---|---|---|---|---|
| $h = 6$ | MSE | 5.9897 | 8.6934 | 6.3107 | 6.3111 |
| | MAPE | 4.2612 | 5.2463 | 4.3920 | 4.3921 |
| $h = 12$ | MSE | 5.9020 | 20.8383 | 7.0924 | 7.0922 |
| | MAPE | 4.6465 | 7.8410 | 5.0882 | 5.0881 |
| $h = 18$ | MSE | 4.5565 | 26.1534 | 4.1153 | 4.1156 |
| | MAPE | 3.7523 | 9.9563 | 3.5865 | 3.5866 |

Table 13: The forecast error measures for given forecast horizons

We again plot the CombMod forecasts for the corresponding horizons.

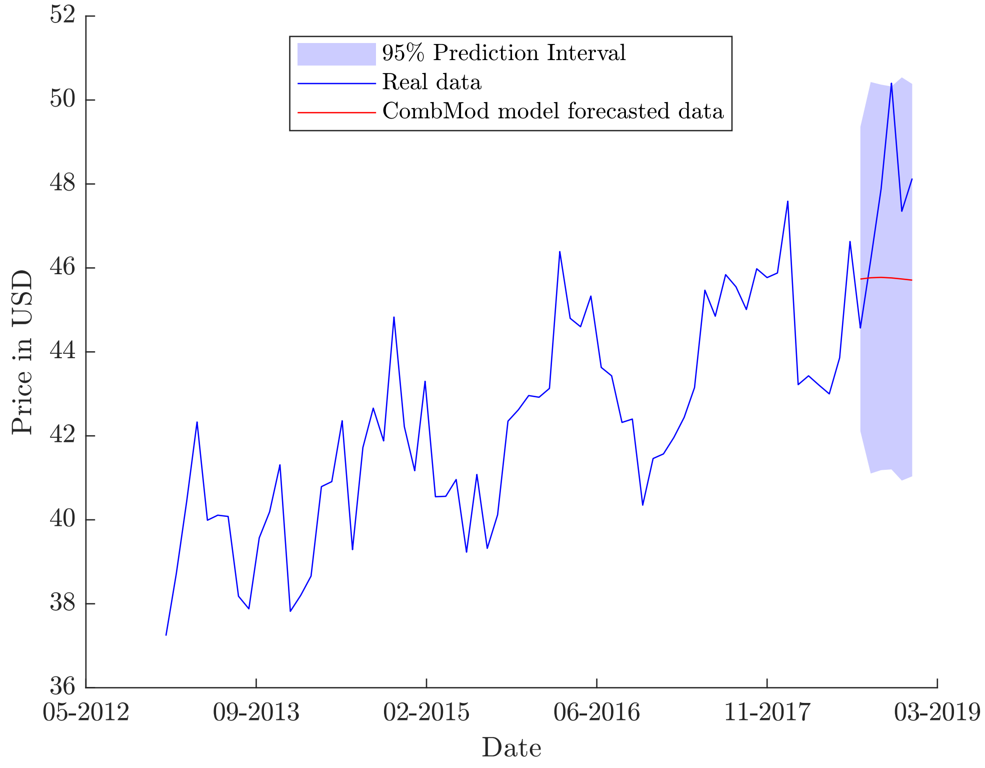


Figure 11: CombMod forecasts of Coca Cola data for $h = 6$

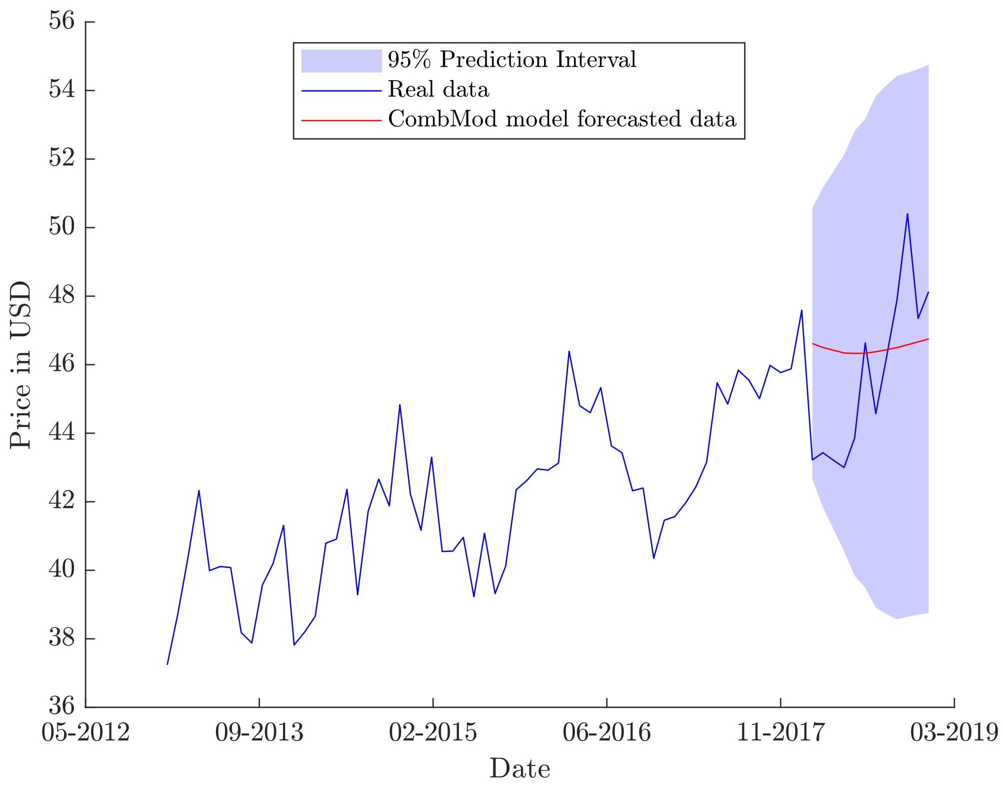


Figure 12: CombMod forecasts of Coca Cola data for $h = 12$

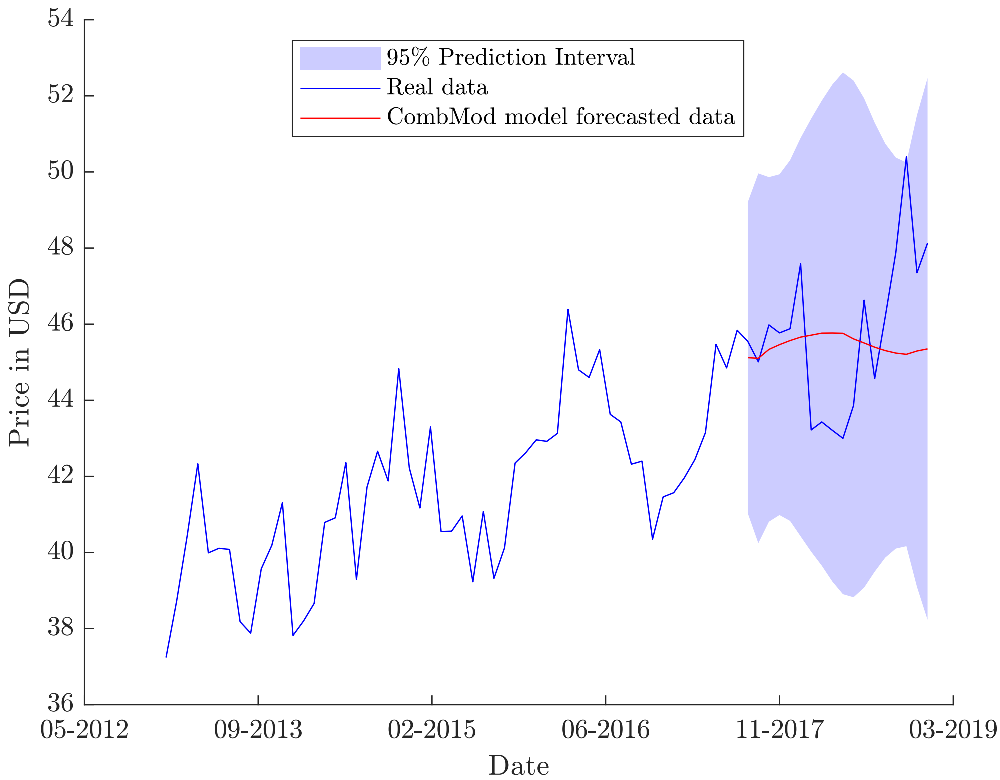


Figure 13: CombMod forecasts of Coca Cola data for $h = 18$

Further we show the weights given to each model by the variance-covariance combination.

| | CombMod | | CombBest | |
|---|---|---|---|---|
| h | ARIMAafter | ETSafter | ARIMAbest | ETSbest |
| 6 | 0.028839 | 0.97116 | $1.8306 \times 10^{-4}$ | 0.99982 |
| 12 | 0.14258 | 0.85742 | $3 \times 10^{-5}$ | 0.99997 |
| 18 | 0.088184 | 0.91182 | $6.51 \times 10^{-5}$ | 0.99993 |

Table 14: Weights given by the variance-covariance combination

### 4.3.3 Discussion

By looking at table 11, we have to address the discrepancy in the model selection for the two said models. At the forecasting horizon $h = 18$, the procedure selected a homoscedastic model ARIMA(1,0,1) and a heteroscedastic model ETS(M,N,N). Even though as we know, the multiplicative term doesn't affect the point forecast, we should proceed with caution, when trying to obtain any information under such circumstances. We can clearly, the rather strange behavior of the prediction interval.

Other than that, we see that the model is able to outperform any other model. The weight again was more on the ETSafter, being close 1. ETSbest itself was influential in this case only concerning the short-term forecast, when for the longer horizons, the two models with trend were more pronounced.

The ARIMAbest model on the other hand fails to play nearly any role in the ARIMAafter combination. Ultimately leading to its bad performance in the test.

Again our method outperforms its combinational counterpart, as it is closely linked ETSbest model, anew weighting it close to 1.

## 4.4 Brent Oil Futures

Derivatives and especially commodities, like oil in this case, are generally hard to predict, as they are very often open for speculations or hedging. Therefore, finding reliable forecasts could yield an advantage on the financial market or at least give more and better information for further action.

For this purpose, we examine the case of Brent Oil Futures, which follows the price of Brent crude oil in USD per barrel. Oil prices are besides the already mentioned influences, often affected artificial inflation or deflation by manipulating the market supply. This can cause highly volatile behavior.

Even if we are aware of the speculative approach to trading with derivatives, we nevertheless, examine the performance of our algorithm on long-, medium- and short-term forecasts, reasoning long term forecast of such strategical commodities might be rather helpful to countries dependent on the stability of such prices.

### 4.4.1 Data

The data of the Brent Oil Futures is collected as its price per barrel in USD over the length of 146 months from January 2007 untill January 2019. The time interval is rather arbitrary, that is, we didn't try to avoid some values corresponding to a well known outside influences, described in the paragraph before. The dataset was collected from the site investing.com and is available at [18].

### 4.4.2 Results

For all these models, their respective forecasts and forecast variances were computed. As the model selection of models didn't vary with the forecast horizons for $h = 6$ and $h = 12$, but in the in the case of $h = 18$, as the estimation dataset shortens it loses information and subsequently the pool of Box-Jenkins models is reduced.

| Chosen best models | |
|---|---|
| ARIMAbest | ARIMA(0,1,0)-GARCH(1,1) |
| ETSbest | ETS(M,N,N) |

Table 15: Best models returned by the selection procedure

<table>
<tr><th>Horizon</th><th colspan="2">Chosen models</th></tr>
<tr><td rowspan="2">h=6,12</td><td>ETS</td><td>ETS(M,N,N), ETS(M,A,N), ETS(M,$A_d$,N)</td></tr>
<tr><td>Box-Jenkins GARCH(1,1)</td><td>ARIMA(0,1,0), ARIMA(0,1,1), ARIMA(0,1,2), ARIMA(1,1,0), ARIMA(1,1,1), ARIMA(1,1,2)</td></tr>
<tr><td rowspan="2">h=18</td><td>ETS</td><td>ETS(M,N,N), ETS(M,A,N), ETS(M,$A_d$,N)</td></tr>
<tr><td>Box-Jenkins GARCH(1,1)</td><td>ARIMA(0,1,0), ARIMA(0,1,1), ARIMA(1,1,0), ARIMA(1,1,1)</td></tr>
</table>

Table 16: Models returned by the selection procedure

For all these models, their respective forecasts and forecast variances were computed and compiled into tables containing the computations of the forecast error measures calculated on the validation data, for each of the competing models. Further forecasts of CombMod were plotted for the corresponding horizon.

| Horizon | Measure | CombMod | ARIMAbest | ETSbest | CombBest |
|---|---|---|---|---|---|
| $h = 6$ | MSE | 159.2255 | 165.1412 | 137.9935 | 137.9953 |
| | MAPE | 17.4537 | 17.6657 | 16.4975 | 16.4976 |
| $h = 12$ | MSE | 79.1297 | 83.3840 | 77.7281 | 77.7282 |
| | MAPE | 11.5365 | 11.5610 | 11.5112 | 11.5112 |
| $h = 18$ | MSE | 285.1182 | 252.8048 | 283.6053 | 283.6024 |
| | MAPE | 19.9775 | 18.7442 | 19.8733 | 19.8731 |

Table 17: The forecast error measures for given forecast horizons

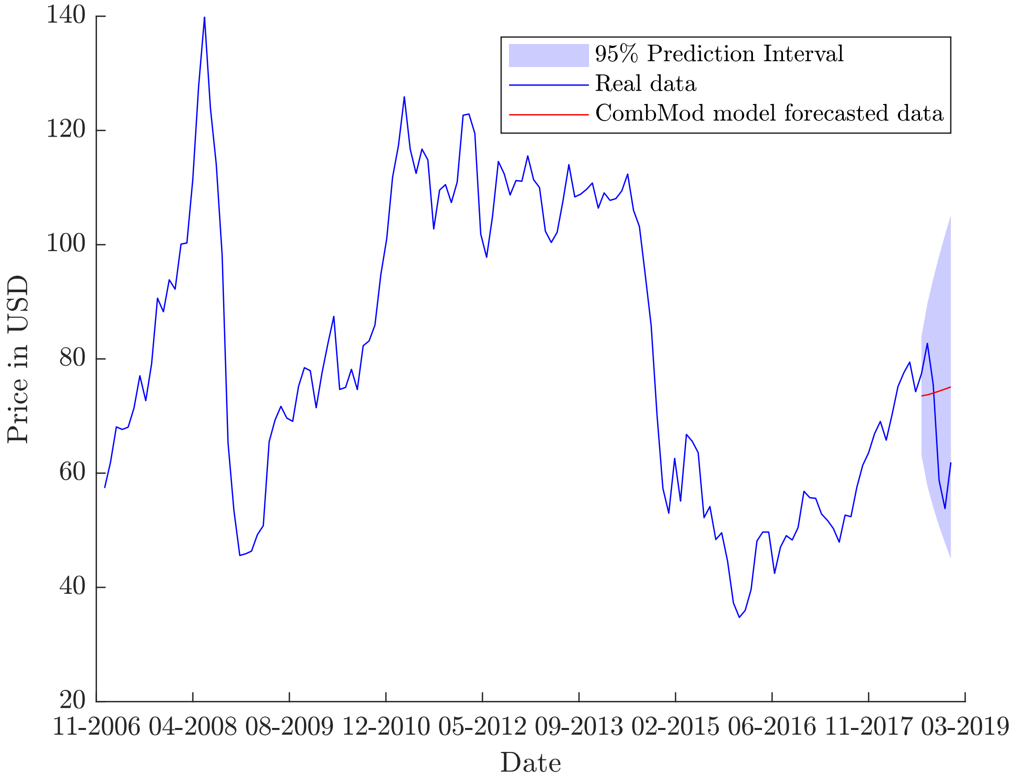


Figure 14: CombMod forecasts of Brent Oil data for $h = 6$

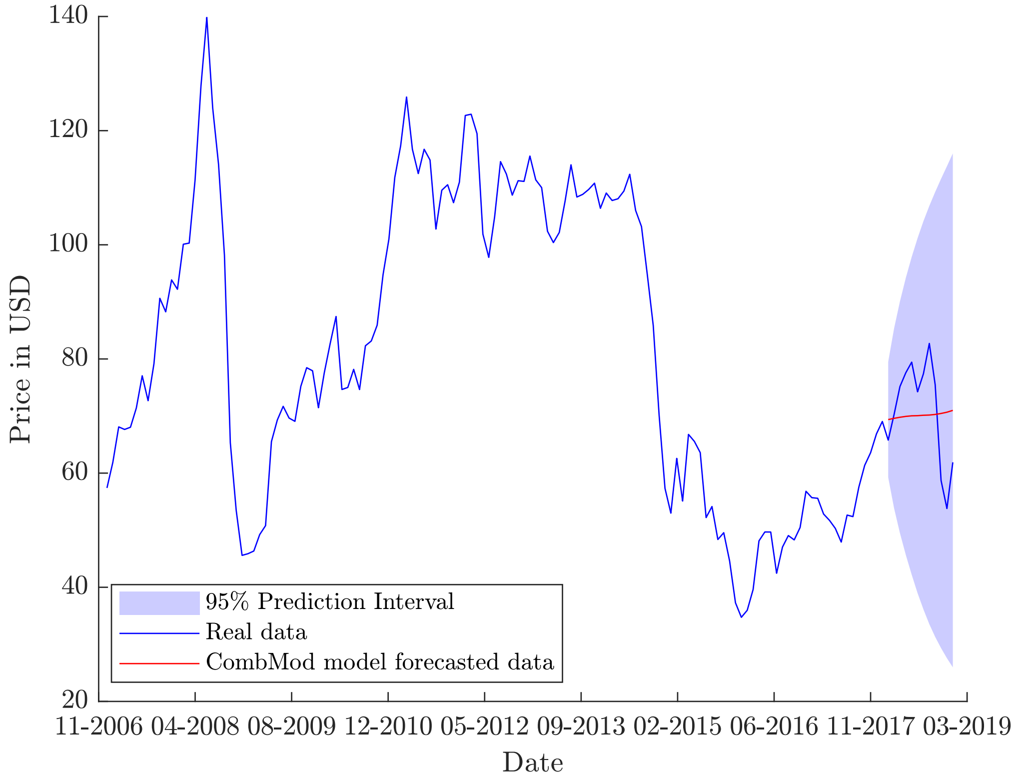


Figure 15: CombMod forecasts of Brent Oil data for $h = 12$

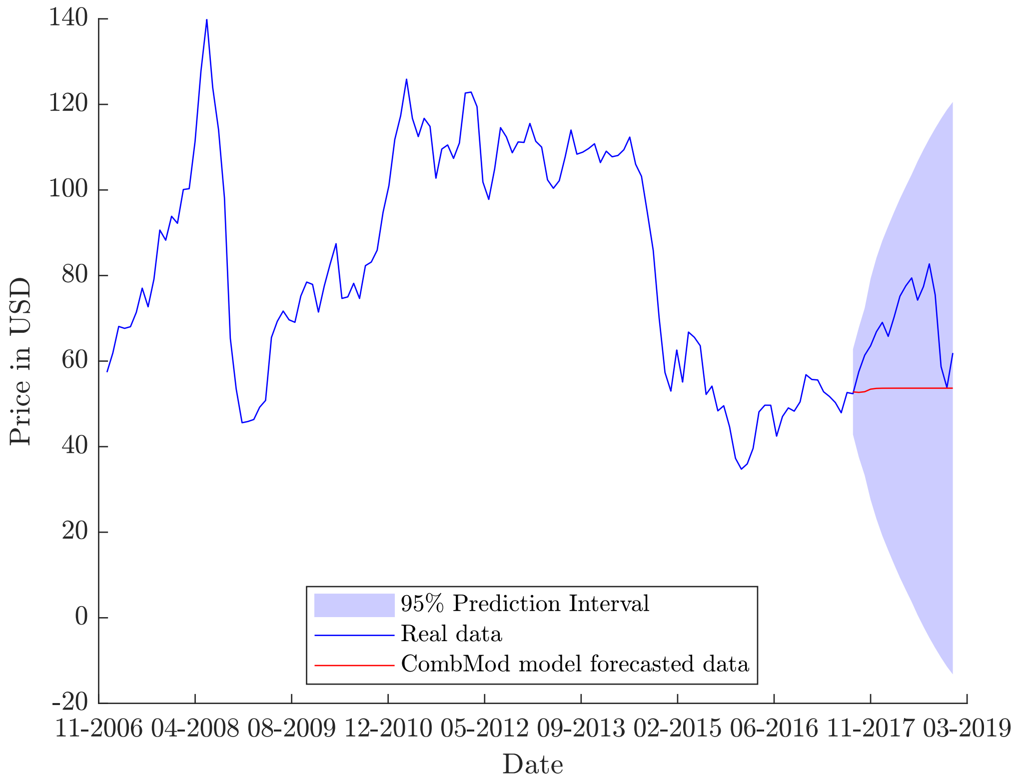


Figure 16: CombMod forecasts of Brent Oil data for $h = 18$

Further we show the weights given to each model by the variance-covariance combination.

| | CombMod | | CombBest | |
|---|---|---|---|---|
| h | ARIMAafter | ETSafter | ARIMAbest | ETSbest |
| 6 | 1 | $3.97 \times 10^{-11}$ | $7.19 \times 10^{-5}$ | 0.99993 |
| 12 | $2 \times 10^{-11}$ | 1 | $4.41 \times 10^{-5}$ | 0.99996 |
| 18 | $6.91 \times 10^{-7}$ | 1 | $9.14 \times 10^{-5}$ | 0.99991 |

Table 18: Weights given by the variance-covariance combination

### 4.4.3 Discussion

As we start to analyze heteroscedastic data, our combined model doesn't seem to perform that well anymore, with respect to its candidate models. In the case of the short-term forecast, the model combination ARIMAafter is weighted by 1, while for the rest of the horizons the other ETSafter is weighted with 1 instead. We see that in two of the cases, that is for $h = 6$ and $h = 18$, the pick wasn't so lucky, demonstrated by the poor performance. Only the medium-term forecast seems to bear any credibility.

By looking at the values of the MAPE, we might even conclude that none of the models, was that successful, but higher error terms are to be expected, considering the volatility of the data. Observing the prediction intervals, we can at least conclude that the model is able account for this volatility at least in this sense, as the real values of our data fall completely into the prediction interval, that could be considered even too pessimistic, but better be safe than sorry.

## 4.5 Bitcoin

Bitcoin, in general, is an open source peer-to-peer payment network and also the digital currency, or cryptocurrency, within it. The currency itself is completely decentralized, meaning that there is authority, central bank or as such, issuing it or in any way influencing it. Nevertheless, Bitcoin in the financial market is considered as a rather speculative asset, with high volatility.

Such a behaviour might suggests heteroscedascity. The selection of such a time series is made purposely, as to study, how our algorithm deals with such data. Looking at the plot we see a sudden drop, corresponding to the biggest crisis of cryptocurrencies yet. Considering this, we choose to forecast only at the horizon of $h = 6$, as for bigger values the information of the drop is lost, and the models fail. In fact, for $h = 12$ the returned forecast perform so badly, that the poor conditioning of the optimization procedures, caused by them, results in an error for the target model.

### 4.5.1 Data

The data of Bitcoin USD pricing is collected over length of 87 months from February 2012 until May 2019. This is basically the whole history of the asset. The dataset was collected from the site investing.com and is available at [18].

### 4.5.2 Results

For the forecast horizon $h = 6$, the algorithm first selected the best ETS model, ETSbest, which was the ETS(M,A,A)$_{12}$. That defined the group of ETS models to be used in the AFTER algorithm as given in the table 8 together with the corresponding selected Box-Jenkins models. Where SARIMA(1,1,1)(0,0,1)$_{12}$-GARCH(1,1) was picked as the best model ARIMAbest.

| Chosen best models | |
|---|---|
| ARIMAbest | SARIMA(1,1,1)(0,0,1)$_{12}$-GARCH(1,1) |
| ETSbest | ETS(M,A,A)$_{12}$ |

Table 19: Best models returned by the selection procedure

| Chosen models | |
|---|---|
| ETS | {ETS(M,N,A)$_{12}$, ETS(M,A,A)$_{12}$, ETS(M,A$_d$,A)$_{12}$} |
| Box-Jenkins GARCH(1,1) | SARIMA(1,1,1)(0,0,1)12, SARIMA(1,1,1)(1,0,0)12, SARIMA(1,1,0)(1,0,0)12, SARIMA(1,1,0)(0,0,1)12, SARIMA(1,1,1)(1,0,1)12, SARIMA(1,1,0)(1,0,1)12, SARIMA(0,1,1)(0,0,1)12, SARIMA(0,1,1)(1,0,0)12, SARIMA(0,1,0)(1,0,0)12, SARIMA(0,1,0)(0,0,1)12, SARIMA(0,1,1)(1,0,1)12, SARIMA(0,1,0)(1,0,1)12 |

Table 20: Models returned by the selection procedure

For all these models, their respective forecasts and forecast variances were computed and compiled into tables containing the computations of the forecast error measures calculated on the validation data, for each of the competing models.

| $h = 6$ | CombMod | ARIMAbest | ETSbest | CombBest |
|---|---|---|---|---|
| MSE $(1 \times 10^{-10})$ | 0.0896 | 4.6683 | 0.0873 | 0.1402 |
| MAPE | 13.6049 | 59.9544 | 13.6525 | 13.7103 |

Table 21: The forecast error measures for the forecast horizon $h = 6$

We conclude the result section with the plot of the values forecasted by our model.

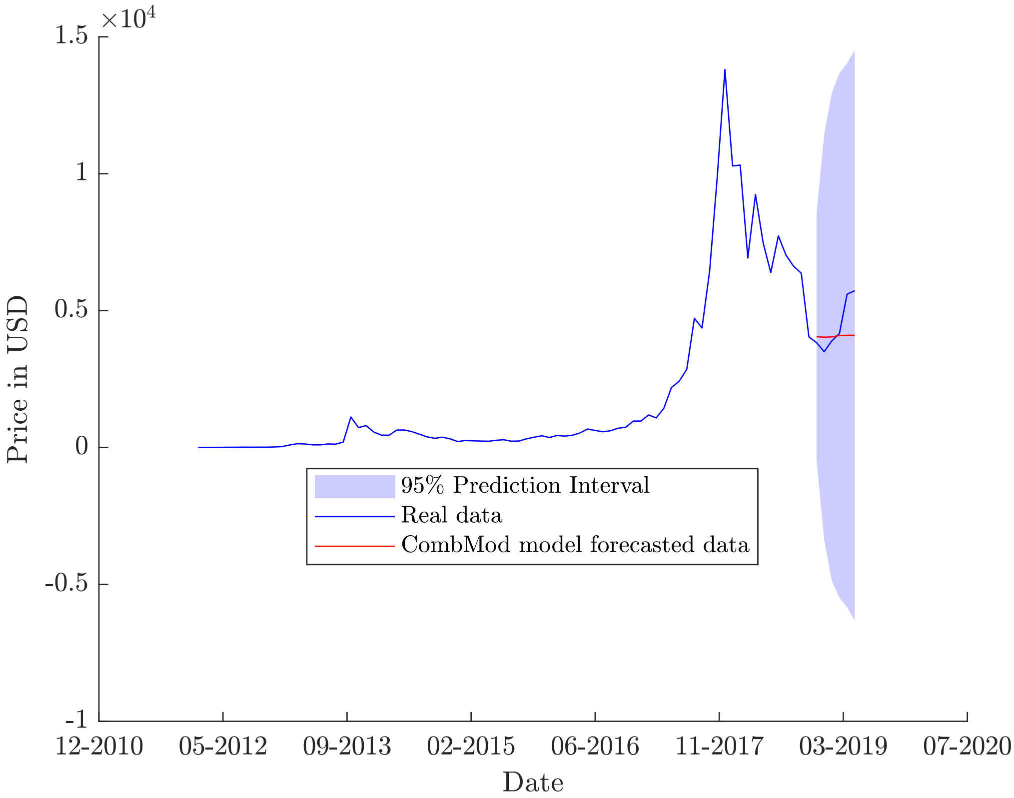


Figure 17: CombMod forecasts of BTC data for $h = 6$

| | CombMod | | CombBest | |
|---|---|---|---|---|
| h | ARIMAafter | ETSafter | ARIMAbest | ETSbest |
| 6 | $5.39 \times 10^{-17}$ | 1 | 0.042326 | 0.95767 |

Table 22: Weights given by the variance-covariance combination

#### 4.5.3 Discussion

Firstly, we should elaborate as to why seasonal models are being used. As already said, we deal within the algorithm with seasonality in a probabilistic sense, so that by computing the AIC for all the ETS methods and subsequently combining them, we pose the question: "Is it likely that the data can be modeled by a seasonal part?"

Nevertheless, it is definitely a striking result. The question arises whether a human intervention wouldn't make more sense and to this we say both yes and no. Yes, because the outcome truly is peculiar indeed and further analysis would definitely be warranted. But also, no, simply because beside the virtue of our described notion of seasonality we can look at the computed AIC's and see that the difference between the chosen model and the next nonseasonal model is quite big.

| ETS | (M,A,A) | (M,$A_d$,A) | (M,A,N) |
|---|---|---|---|
| AIC | 1171,99 | 1173,99 | 1222,38 |

Table 23: The forecast error measures for the forecast horizon $h = 6$

So following the decision to let this play itself out, we can look at the overall performance of the combined model. Observing the forecasts plotted in 17 we see a rather flat predictions, with a very faint wave-like behavior. As it is linked to the ETSbest rather closely, the forecast error measures are very similar, again the weight on the ETSafter being virtually equal to 1, and the weights within it are shared by the two seasonal models with trend (M,A,A) (M,$A_d$,A), which are, in a situation where the estimated trend term is negligible, very similar as well.

The ARIMAbest fails to correctly estimate the ascending trend and the error measures are correspondingly much higher than for the rest of the models. The same is true for the other Box-Jenkins models, so that they weighted with virtually zero correspondingly.

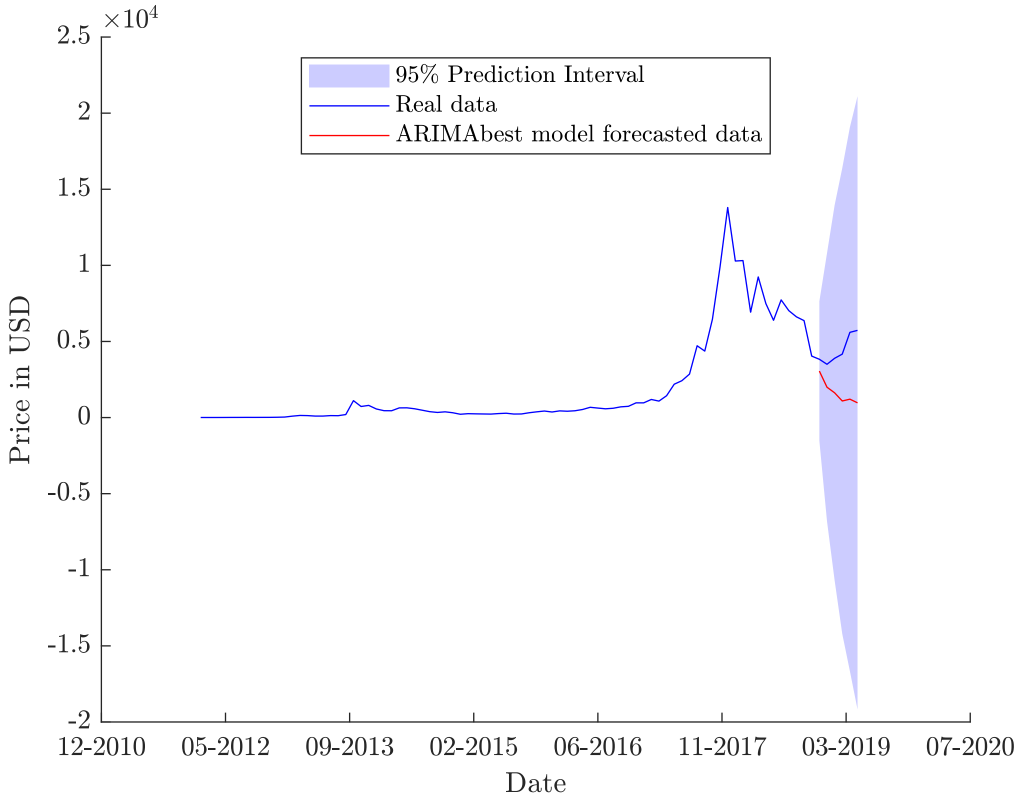


Figure 18: ARIMAbest forecasts of BTC data for $h = 6$

Including the prediction interval in our scope of study we can paint a complete picture of the performance of our model. We clearly see the immense span of the prediction interval, that accounts for the high volatility of the data and is as much as a practical tool, say for risk aversion, useless. Nevertheless, the point forecast, given the conditions, didn't perform that bad as it at least forecasted a slight ascend, which all the ARIMA models failed to do.

# 5 Conclusion

Recalling our aim was to construct an algorithm that would, if possible, fully automatically, by using combination procedures and optimization methods, produce reliable enough forecasts that could outperform its given counterparts. The combination was to be done on the so called classical statistical methods for time series forecasting. This begs the question. Were we successful?

In the sense of constructing such a method, certainly. Even more, we incorporated two different approaches of forecast combination, that are combining for adaptation and combining for performance and were able to show its mostly better performance then other analyzed models. By precombining a set of models of the same kind (combination for adaptation), say Box-Jenkins, and after that combining them into the final forecast with another precombined set of ETS models, we demonstrated its competency when forecasting especially homoscedastic time series, as they mostly performed better then the best models from the Box-Jenkins and ETS methodologies, which were found by classical selection procedures, found in a similar form in other statistical software, by minimizing some sort of information criterion, such as AIC or BIC. Here the term "best models" has to be taken lightly, as we do not wish to say that such selection procedures generate the best models possible, as it is usually not the case and we take it into account. But the idea as stated by Yang [8] is to eliminate not only the risk of possibly choosing the wrong model instead of the correct, considering there even exist one perfect model instead of perhaps multiple, but also reducing say computational load, as data for which the analysis, especially in the Box-Jenkins case, might take up a lot of time, as it can be rather tedious work. Moreover, if we end up at only one model, we may disregard a whole group of models, that might be useful, but were disregarded, for perhaps a minor difference in some decision statistic. On the other hand, we can just click a button.

Considering applicability in the real markert, we tried in the examples to test a variety of financial time series, exploring stocks, commodities, indexes and cryptocurrencies with characteristics such as homoscedascity, heteroscedascity. We saw better results with more stable data, such as the Doe Jones index, where the method performed very well, but at times it didn't, it usually could perform better than at least one of the models from a respective methodology. Therefore, one could think about the algorithm as a kind of risk aversion procedure, given that that we want to forecast using Box-Jenkins or ETS models. If one the models fails to forecast the data properly, we can make up for it, if the other doesn't.

# References


[1] HYNDMAN, Rob J., Anne B. KOEHLER, J. Keith ORD a Ralph D. SNYDER. *Forecasting with Exponential Smoothing: The State Space Approach* [online]. Springer-Verlag Berlin Heidelberg, 2008 [cit. 2019-05-21]. ISBN 978-3-540-71918-2. Available at: https://www.springer.com/gp/book/9783540719168.

[2] MONTGOMERY, Douglas C., Cheryl L. JENNINGS a Murat KULAHCI. *Introduction to time series analysis and forecasting.* Hoboken, N.J.: Wiley-Interscience, c2008. ISBN 0471653977.

[3] *MATLAB Documentation.* [online] © 1994-2019 The MathWorks, Inc. [cit. 2019-05-21]. Available at: https://www.mathworks.com/help/matlab/.

[4] MAKRIDAKIS, Spyros G., Steven C. WHEELWRIGHT and Rob J. HYNDMAN. Classical Decomposition. *Forecasting: Methods and Applications.* 3rd ed. John Wiley, 1997, p. 107-113, ISBN 978-0-471-53233-0.

[5] BOLLERSLEV, Tim. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* [online]. 1986, 31(3), 307-327 [cit. 2019-05-23]. DOI: 10.1016/0304-4076(86)90063-1. ISSN 03044076. Available at: https://linkinghub.elsevier.com/retrieve/pii/0304407686900631.

[6] HANSEN, Peter R. and Asger LUNDE. A forecast comparison of volatility models: does anything beat a GARCH(1,1)?. *Journal of Applied Econometrics* [online]. 2005, 20(7), 873-889 [cit. 2019-05-23]. DOI: 10.1002/jae.800. ISSN 0883-7252. Available at: http://doi.wiley.com/10.1002/jae.800

[7] BATES, J. M. and C. W. J. GRANGER. *The Combination of Forecasts.* OR [online]. 1969, 20(4) [cit. 2019-05-21]. DOI: 10.2307/3008764. ISSN 14732858. Available at: https://www.jstor.org/stable/3008764?origin=crossref.

[8] YANG, Y. Combining Forecasting Procedures: Some Theoretical Results. *Econometric Theory*, vol. 20, no. 1, 2004, pp. 176–222. JSTOR. [cit. 2019-05-22] https://pdfs.semanticscholar.org/be9d/ad6dface3607fa46a0e6a9be5a71066fd51e.pdf.

[9] ZOU, HUI and YANG, YUHONG. Combining time series models for forecasting. *International Journal of Forecasting.* 2004. Vol. 20, no. 1, p. 69-84, [cit. 2019-05-21] DOI 10.1016/s0169-2070(03)00004-9. Elsevier BV. Available at: http://users.stat.umn.edu/~zouxx019/Papers/after.pdf.

[10] HŘEBÍK, Radek and Jana SEKNIČKOVÁ. ARIMA model selection in Matlab. In: *Proceedings of 30th International Conference Mathematical Methods in Economics.* Karviná: Silesian University, School of Business Administration, 2012, pp. 332-337, [cit. 2019-05-21]. ISBN 978-80-7248-779-0. Available at: http://mme2012.opf.slu.cz/proceedings/pdf/057_Hrebik.pdf.

[11] HYNDMAN, Rob J. and Yeasmin KHANDAKAR. Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software* [online]. 2008, 27(3) [cit. 2019-05-21]. DOI: 10.18637/jss.v027.i03. ISSN 1548-7660. Available at: http://www.jstatsoft.org/v27/i03/.

[12] WALLIS, Kenneth. (2005), Combining Density and Interval Forecasts: A Modest Proposal, *Oxford Bulletin of Economics and Statistics*, 67, issue s1, p. 983-994.

[13] ELLIOTT, Graham, C. W. J. GRANGER and Allan TIMMERMANN. *Handbook of economic forecasting*. Boston: Elsevier North-Holland, 2013. ISBN 0-444-51395-7.

[14] HYNDMAN, Rob J., and George ATHANASOPOULOS. *Forecasting: principles and practice*, 2nd edition, OTexts: Melbourne, Australia, 2018. [cit. 2019-05-21]. Available at: https://otexts.com/fpp2/.

[15] MOHAMMADI,Shapour . *FORCASCOMB: MATLAB function to combine forecasts of various models*, 2009Statistical Software Components T741508, Boston College Department of Economics. [cit. 2019-05-21] Available at: https://ideas.repec.org/c/boc/bocode/t741508.html.

[16] The variance-covariance approach. CLEMENTS, Michael and David HENDRY. *Forecasting Economic Time Series* [online]. Cambridge: Cambridge University Press, 1998, s. 229-231 [cit. 2019-05-21]. ISBN 978-0521632423. Available at: https://books.google.cz/books?id=aJ966QLx0WAC&lpg=PR19&ots=uGIzDdftDt&dq=Forecasting%20Economic%20Time%20Series%20clements%20pdf&hl=cs&pg=PA229#v=onepage&q&f=false.

[17] ARMSTRONG, Jon Scott. *Principles of forecasting: a handbook for researchers and practitioners*. Boston, MA: Kluwer Academic, c2001. [cit. 2019-05-21] ISBN 0792379306.

[18] *Investing.com* [online] Fusion Media Limited © 2007-2019 [cit. 2019-05-21]. Available at: https://www.investing.com/.