Prediction - Time Series

Jordi Tarroch Mejón

***Abstract*—Hotels have multiple ways to increase profits using data science but is not so clear how. Doing a good analysis of time series based on the data given by the hotel's Property Management Systems data can help to take the right decisions at the right time. Different time series can be extracted from the date such as the number of reservations for the arrival date or the number of reservations by the reservation date among others. Each one of them can be managed by different hotel teams. Results show different time series forecast based on different modeling techniques. Depending on the features of each time series analyzed one model performs better than the other when forecasting the future of the time series.**

CONTENTS

I. INTRODUCTION

The goal of this study is to predict:

• The Number of Reservations for the Arrival Date (independently and by splitting its data based on 'canceled bookings' and 'not canceling bookings/ Checkout'). Information that would be of interest for Hotel's Management in order to increase hotel room efficiency.

- The Number of Reservations for the Reservation Date (independently and by splitting its data based on 'canceled bookings' and 'not canceling bookings'). Information that would be of interest for Hotel's Marketing in order to detect anomalies and to increase marketing the dates in which the number of reservations seem to decrease.

A comparison of multiple time series prediction models is done to choose the best performing model and conclude with the best possible forecast of the time series. In order to compare the models the MAE and Bias % has been calculated:

- MAE (mean absolute error) is a measure of difference between two continuous variables.
- Bias %: occurs when there are consistent differences between actual outcomes and previously generated forecasts of those quantities; that is: forecasts may have a general tendency to be too high or too low. A normal property of a good forecast is that it is not biased.

The models applied for time series forecasting are:

- ETS (Exponential Smoothing)
- ARIMA(Autoregressive integrated moving average).

Jordi Tarroch Mejón is with MS Data Science for Finance, CUNEF, e-mail: jordi.tarroch@cunef.edu

II. MODELING

A. Time Series

1) Train - Test Data Set: For training the model, the last 20 weeks have been removed from the time series, weeks that will be used as test.

2) ETS: Exponential smoothing is a time series forecasting method for univariate data that can be extended to support data with a systematic trend or seasonal component. In this model a prediction is a weighted sum of past observations, but the model explicitly uses an exponentially decreasing weight for past observations. Specifically, past observations are weighted with a geometrically decreasing ratio, in other words, the more recent the observation the higher the associated weight.

a) Training the model: The training data set is used to train the model for each of the time series studied.

The resulting trained model is an exponential smoothing model defined with three letters:

- The first letter denotes the error type ("A", "M" or "Z")
- The second letter denotes the trend type ("N","A","M" or "Z")
- Third letter denotes the season type ("N","A","M" or "Z").
- In all cases, "N"=none, "A"=additive, "M"=multiplicative and "Z"=automatically selected.

b) Forecast: The last 20 weeks has been forecast for each of the models.

Number of Reservations for Arrival Date:

[Fig. 1 about here.]

Number of Reservations for Reservation Date:

[Fig. 2 about here.]

Number of Reservations for Arrival Date Canceled:

[Fig. 3 about here.]

Number of Reservations for Arrival Date Checkout:

[Fig. 4 about here.]

Number of Reservations at Reservation Date Canceled:

[Fig. 5 about here.]

Number of Reservations at Reservation Date Checkout:

[Fig. 6 about here.]

c) Error Measurements: The test data set has been compared with the forecast done by the model to measure the error and bias.

TABLE I
ERROR TABLE ETS

Time Series	MAE	BIAS %
Arrival Date	40.77	16.04
Reservation Date	49.53	35.07
Arrival Date Canceled	31.42	-20.43
Arrival Date Checkout	33.20	20.12
Arrival Date Checkout + Canceled	36.9	-0.30
Reservation Date Canceled	79.77	79.77
Reservation Date Checkout	35.35	27.46
Reservation Date Checkout + Canceled	107.23	107.23

3) ARIMA: Autoregressive integrated moving average is a model where the prediction is a weighted linear sum of recent past observations or lags.

a) Training the model: The auto.arima function is used to return the best ARIMA model according to either AIC, AICc or BIC value. The function conducts a search over possible model within the order constraints provided and the final model is defined by these three variables:

p: The number of lag observations included in the model, also called the lag order.

d: The number of times that the raw observations are differenced, also called the degree of differencing. In an ARIMA model the time series is transformed into a stationary one(series without trend or seasonality) using differencing. Stationary time series are when the **mean and variance are constant over time.** It is easier to predict when the series is stationary.

q: The size of the moving average window, also called the order of moving average.

b) Forecast:

Number of Reservations for Arrival Date:

[Fig. 7 about here.]

Number of Reservations for Reservation Date:

[Fig. 8 about here.]

Number of Reservations for Arrival Date Canceled:

[Fig. 9 about here.]

Number of Reservations for Arrival Date Checkout:

[Fig. 10 about here.]

c) Error Measurements:

d) Final Prediction:

TABLE II
ERROR TABLE ARIMA

Time Series	MAE	BIAS %
Arrival Date	60.74	-45.46
Reservation Date	44.90	22.71
Arrival Date Canceled	54.32	-51.83
Arrival Date Checkout	30.31	-9.68
Arrival Date Checkout + Canceled	74.47	-61.52
Reservation Date Canceled	31.65	30.36
Reservation Date Checkout	33.83	24.27
Reservation Date Checkout + Canceled	60.66	54.63

Forecast: The resulting best model for the Number of Reservations for an Arrival Date forecasting is the ETS model with the sum of the partitioned data of 'canceled bookings' and 'not canceled bookings' with a MAE of 36.9 and a -0.30% Bias that means that higher Number of Reservations for that Arrival Date could be expected.

These are the predicted Number of Reservations for the Arrival Date for the next 20 weeks using an ETS model:

[Fig. 11 about here.]

The resulting best model for the Number of Reservations done an exact Reservation Date forecasting is the ARIMA model without partitioned data of 'canceled bookings' and 'not canceled bookings' with a MAE of 44.90 and a 22.71% Bias that means that lower Number of Reservations for that Reservation Date could be expected.

These are the predicted Number of Reservations for the Reservation Date for the next 20 weeks using an ARIMA model:

[Fig. 12 about here.]

The ARIMA model ends up using a 1 lag order of difference, which is stationary but at the same time is white noise, thus the modeling and forecasting is quite bad.

In the following figures, ACF shows how similar is this first lag order of difference to the white noise.

[Fig. 13 about here.]

III. CONCLUSION

Before running any models, plotting the time series data and taking a good look at it must be done. If there is a trend and/or seasonality during the period to be predicted, an exponential smoothing method (ETS) explicitly model these components. In case there is autocorrelation in the data (seen in the ACF), i.e. the past explains the present, going for the ARIMA methodology is the best option.

However, running both ARIMA and ETS models and comparing the methods based on the error measurements is the best option and helps to choose the 'best' method.

IV. REFERENCES

- https://medium.com/brillio-data-science/forecasting-hierarchical-time-series-usingr-
- 598828dba435 https://cran.r-project.org/web/packages/hts/index.html

- https://robjhyndman.com/publications/hierarchical/
- https://machinelearningmastery.com/exponential-smoothing-for-time-series-forecasting-in-python/
- https://towardsdatascience.com/time-series-forecasting-arima-models-7f221e9eee06

V. Annexes

List of Figures

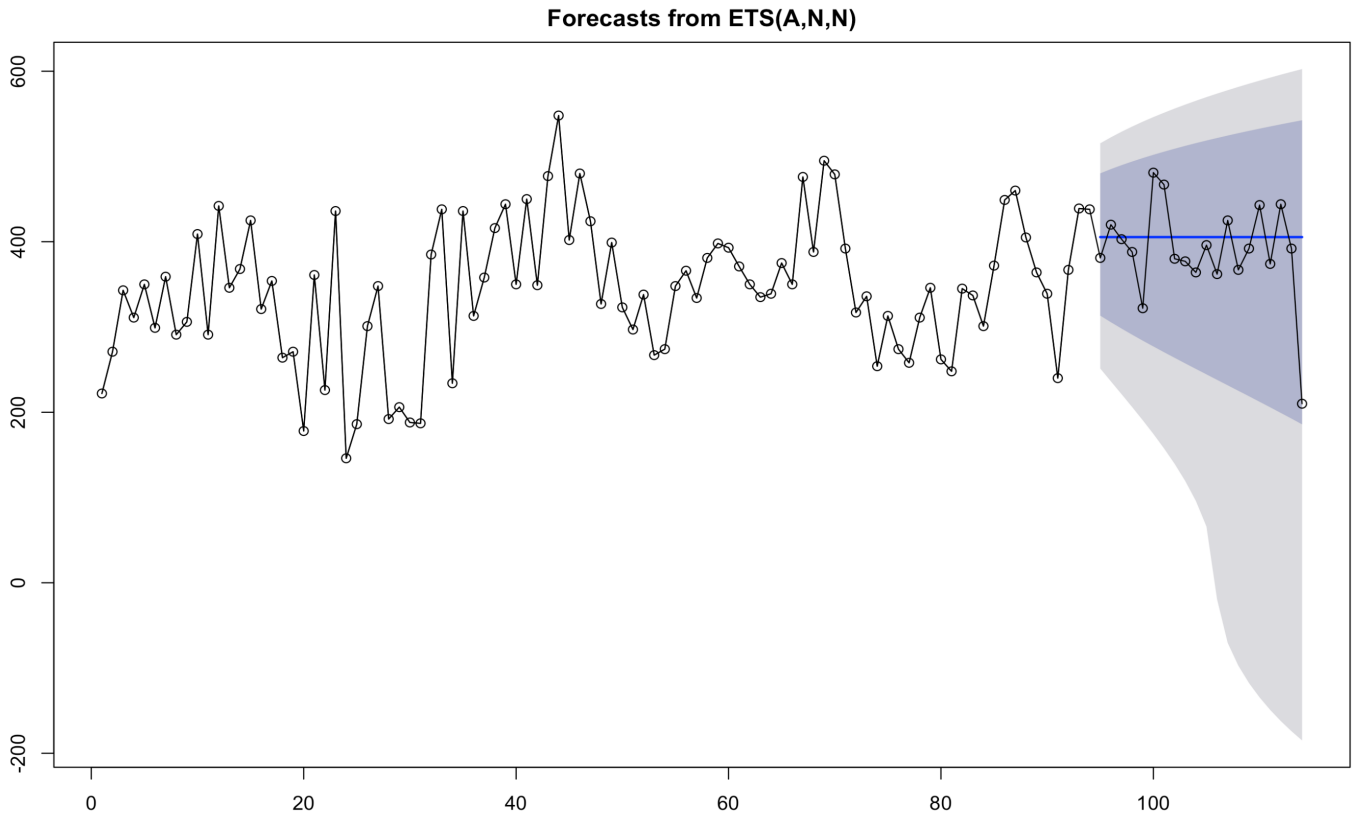

Fig. 1. Number of Reservations for Arrival Date Forecast - Test

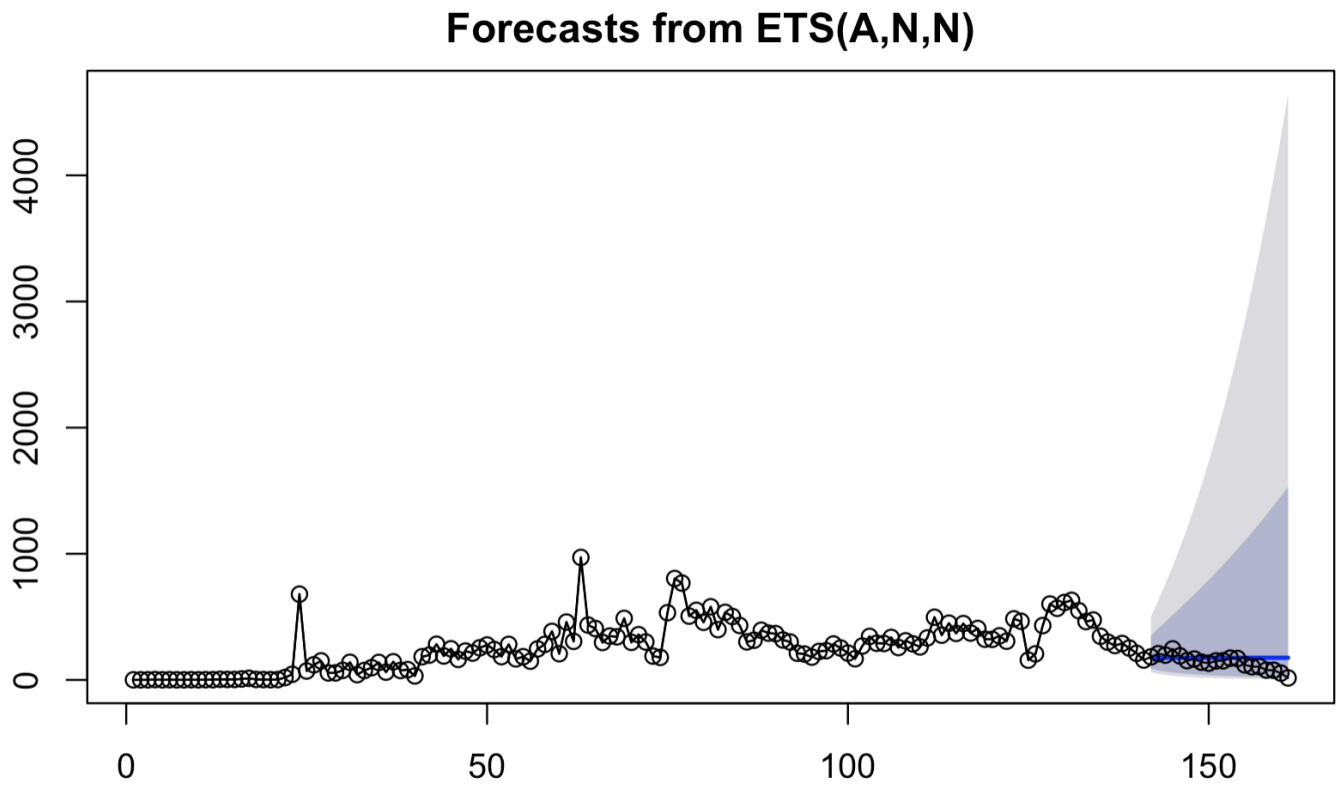

Fig. 2. Number of Reservations for Reservation Date Forecast - Test

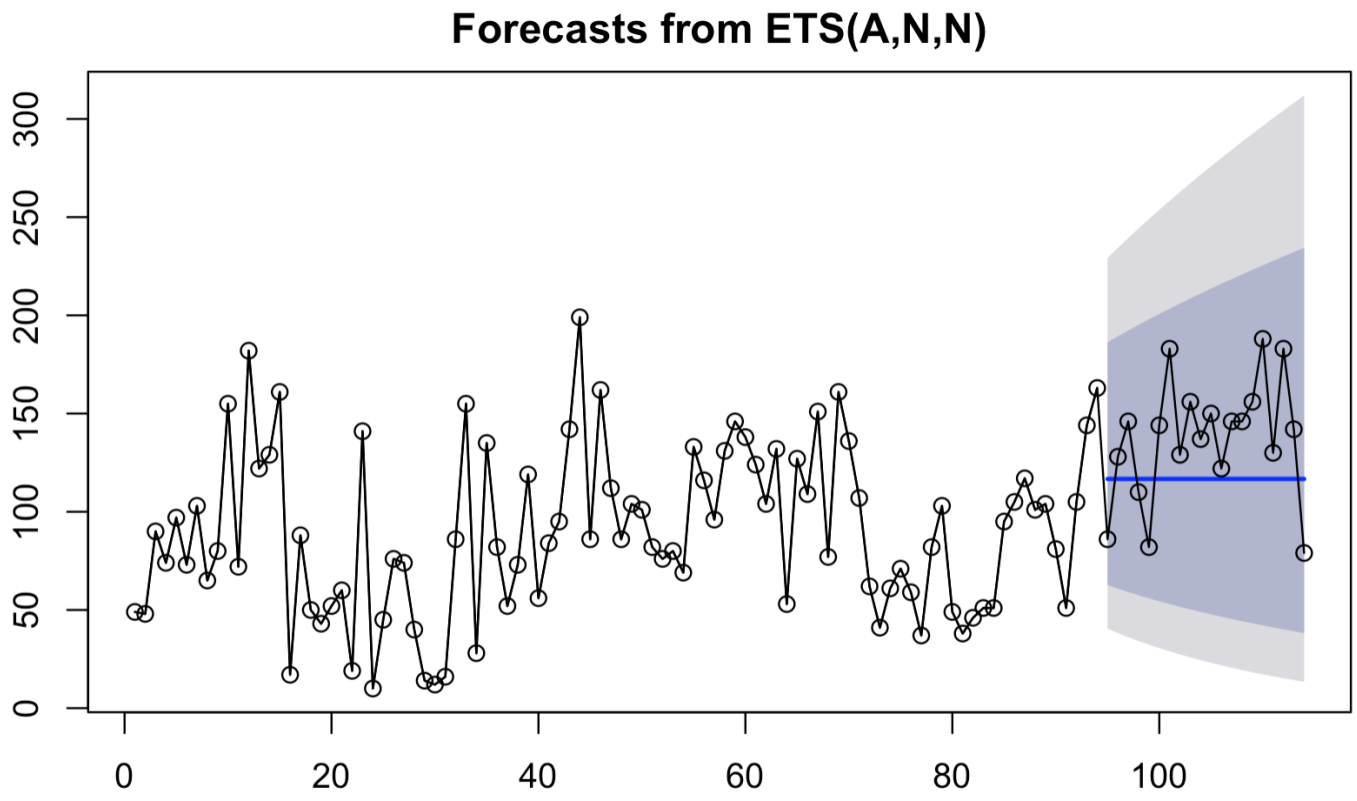

Fig. 3. Number of Reservations for Arrival Date Canceled Forecast - Test

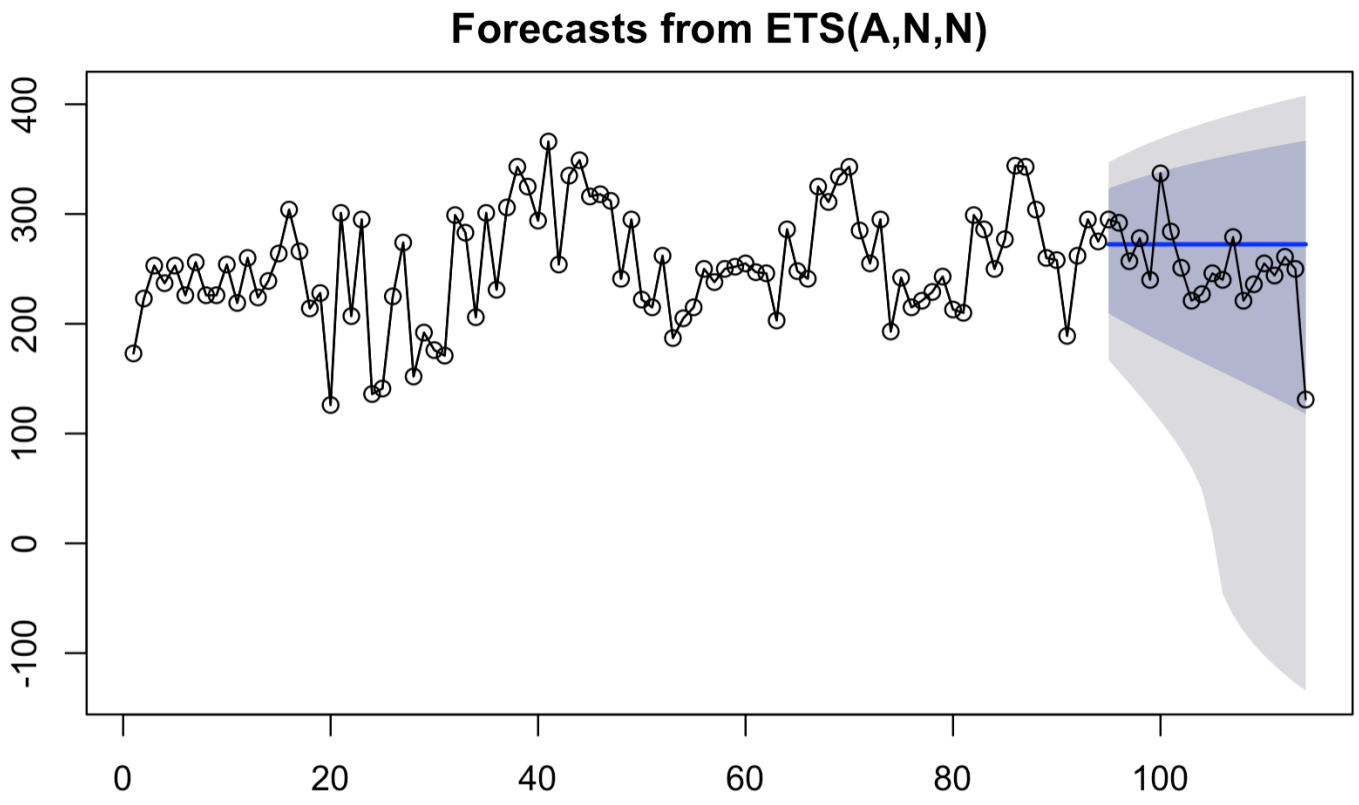

Fig. 4. Number of Reservations for Arrival Date Checkout Forecast - Test

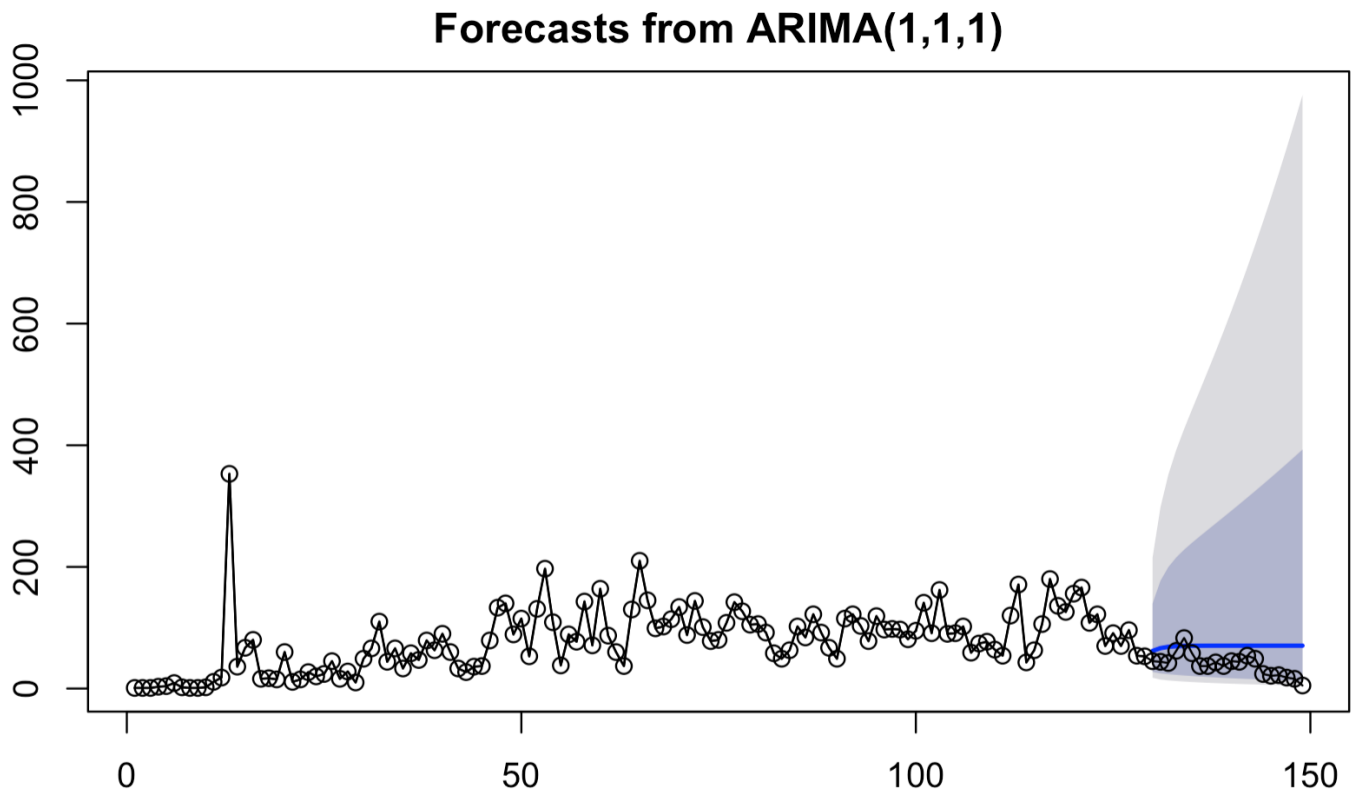

Fig. 5. Number of Reservations at Reservation Date Canceled Forecast - Test

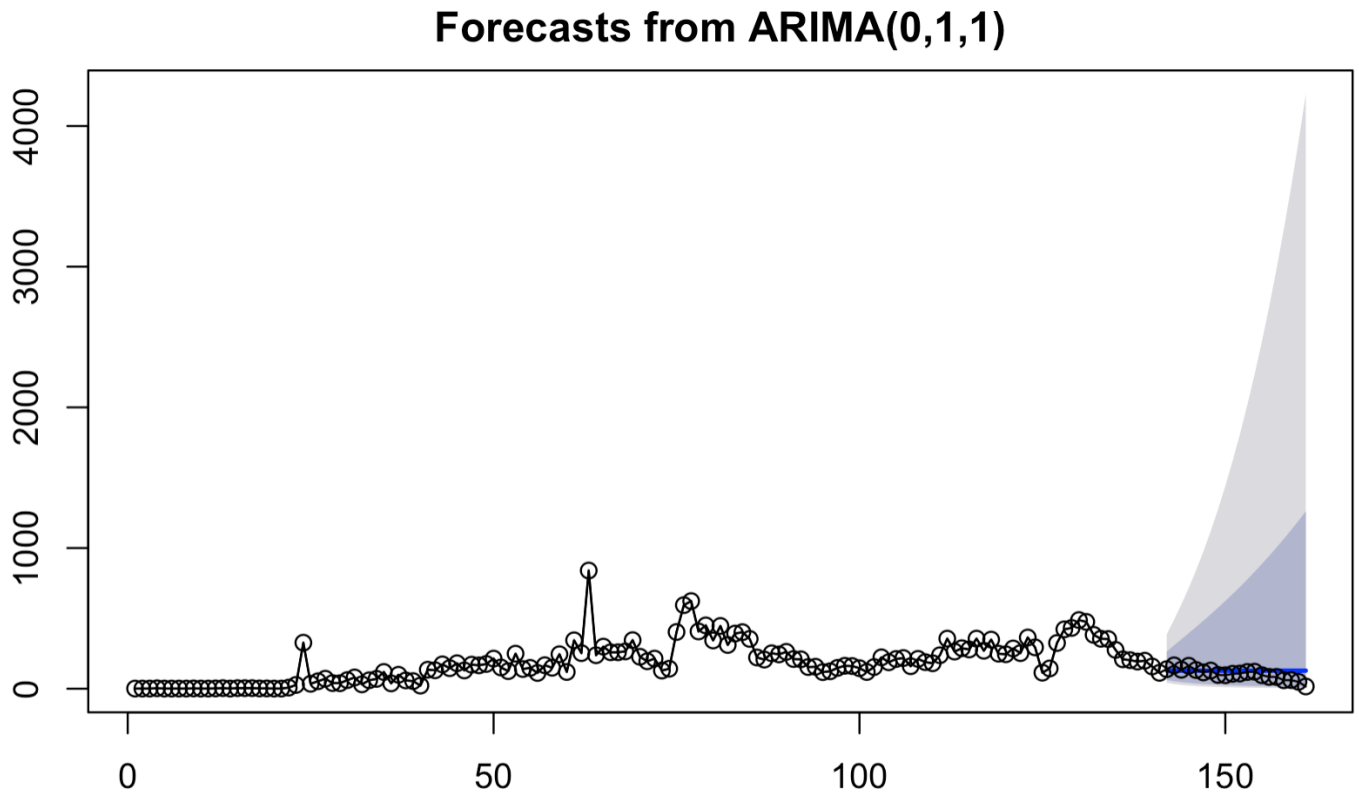

Fig. 6. Number of Reservations at Reservation Date Checkout Forecast - Test

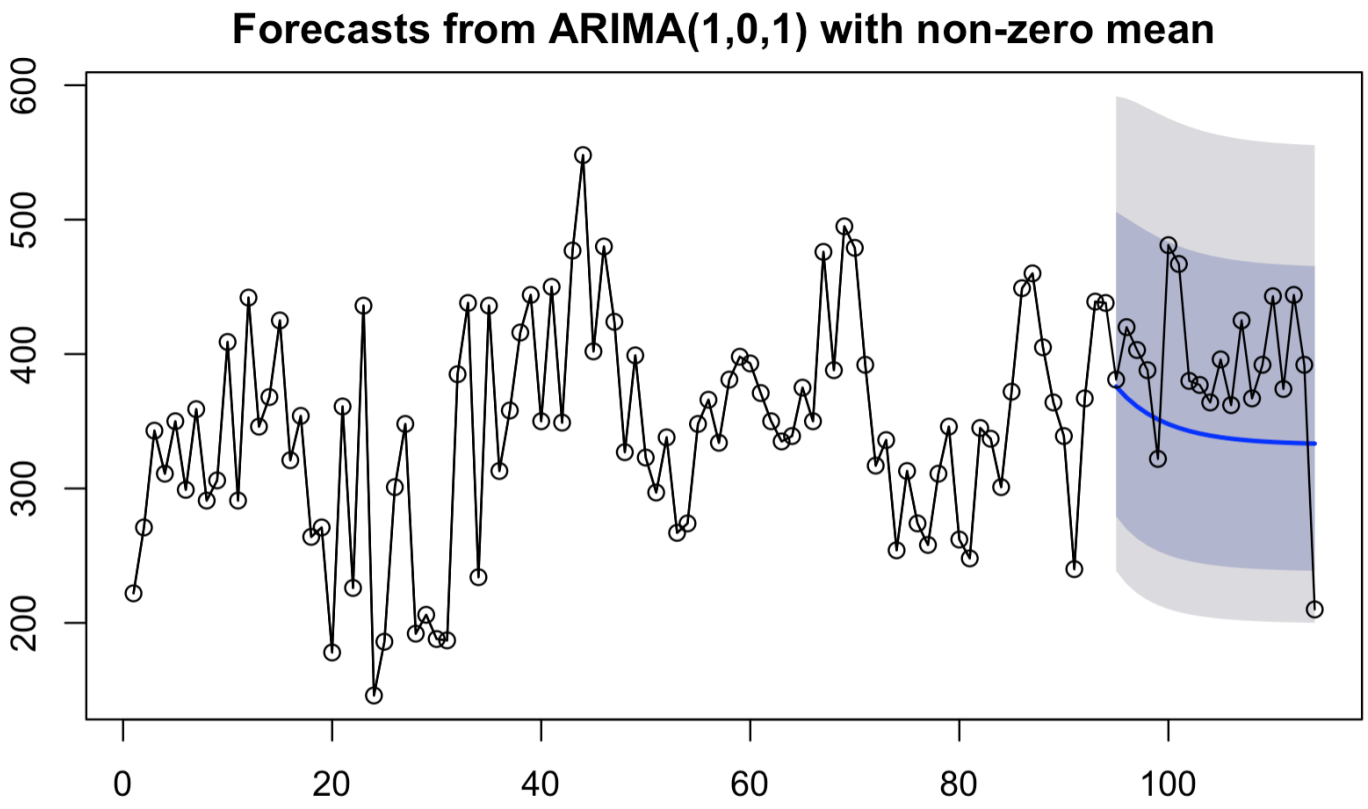

Fig. 7. Number of Reservations for Arrival Date Forecast - Test

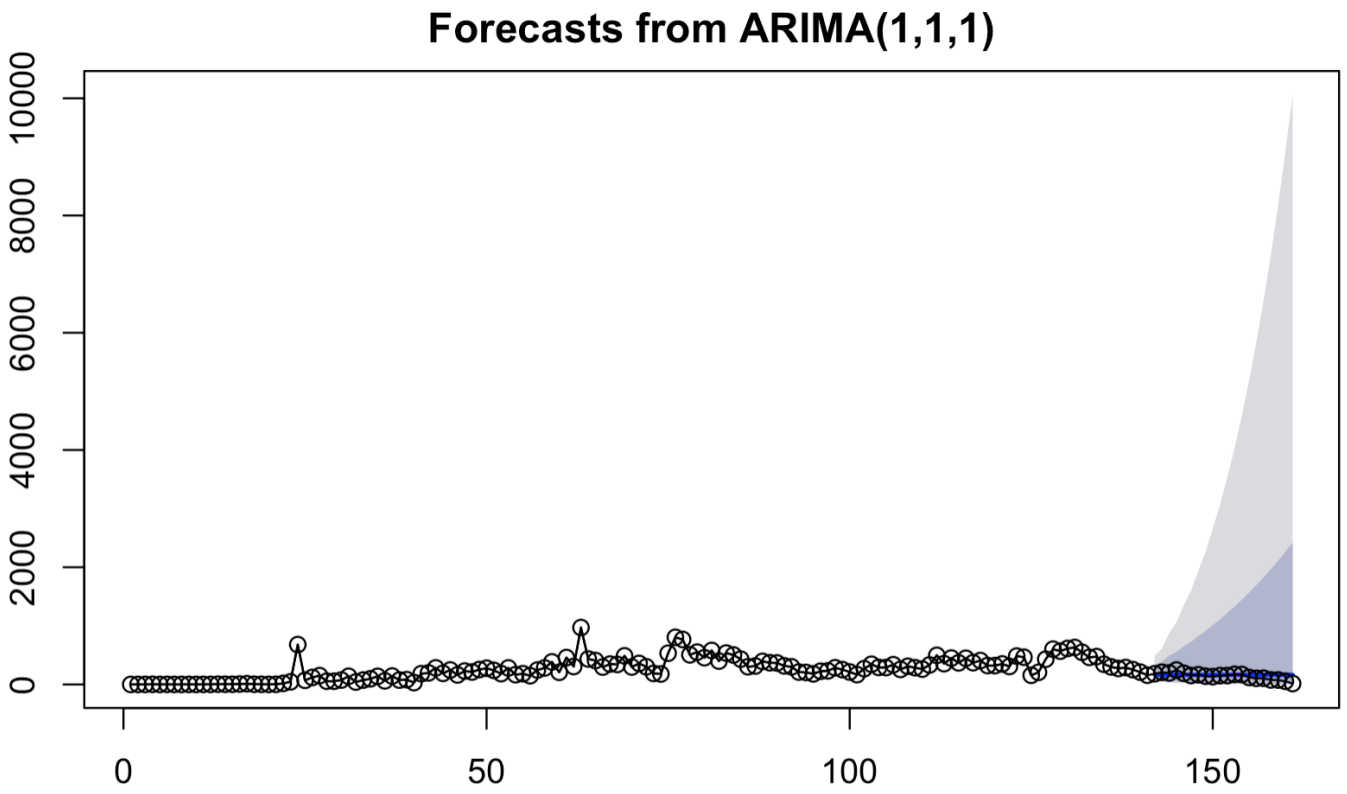

Fig. 8. Number of Reservations for Reservation Date Forecast - Test

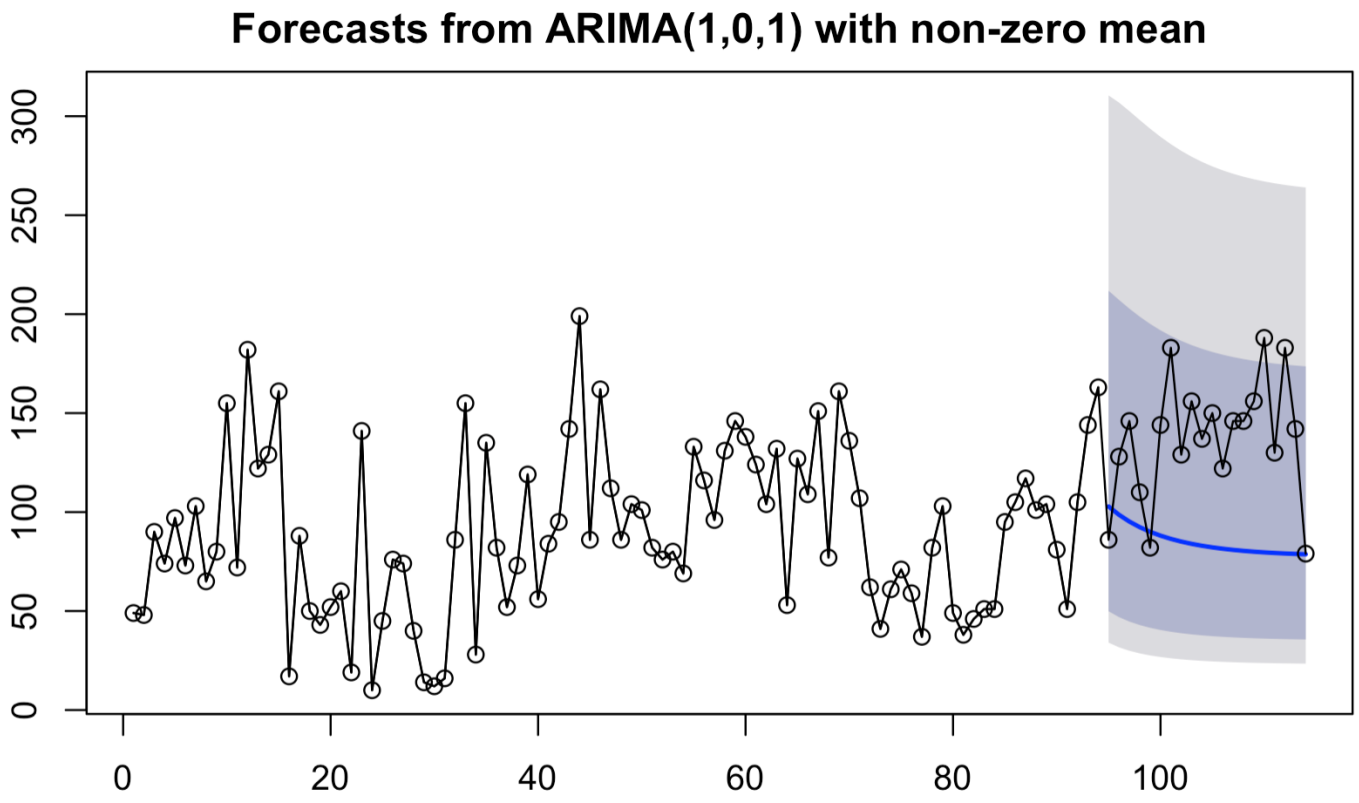

Fig. 9. Number of Reservations for Arrival Date Canceled Forecast - Test

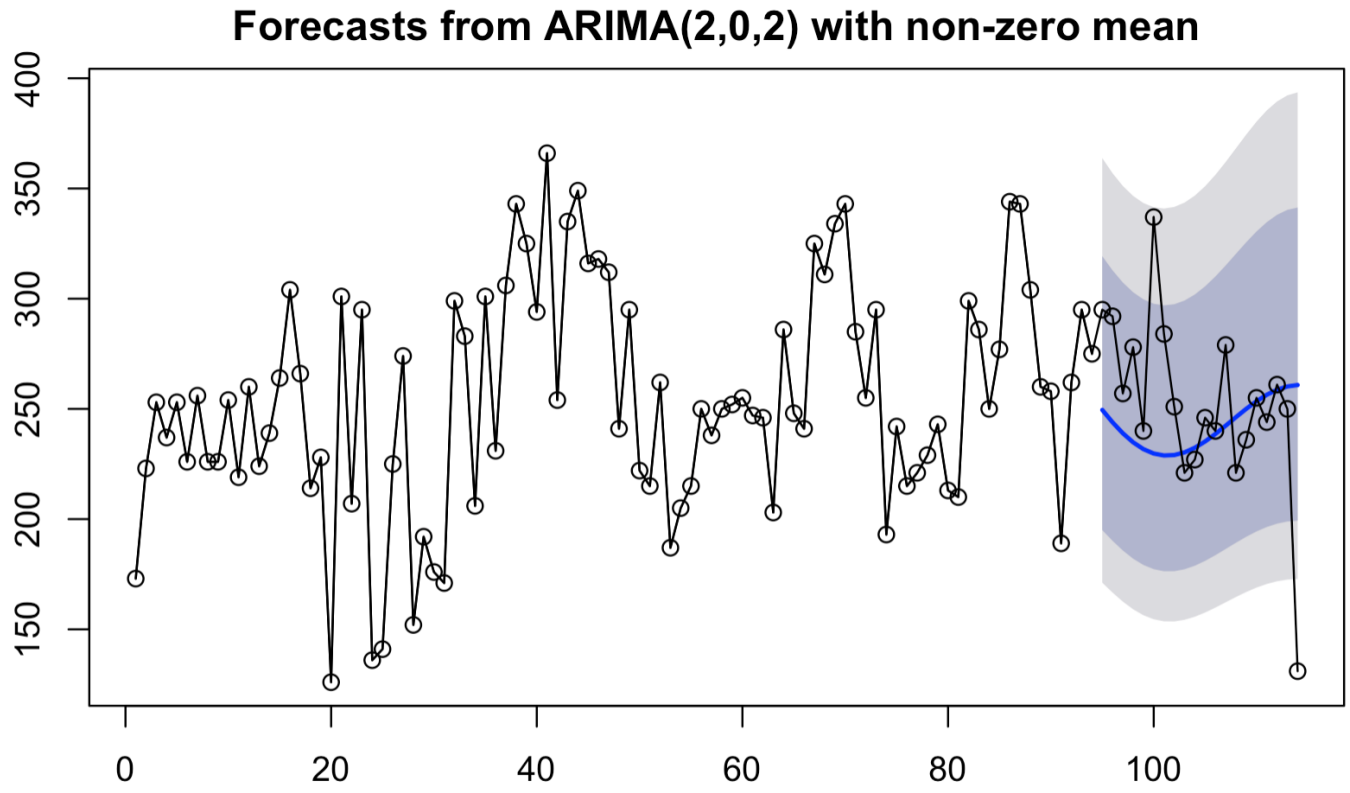

Fig. 10. Number of Reservations for Arrival Date Checkout Forecast - Test

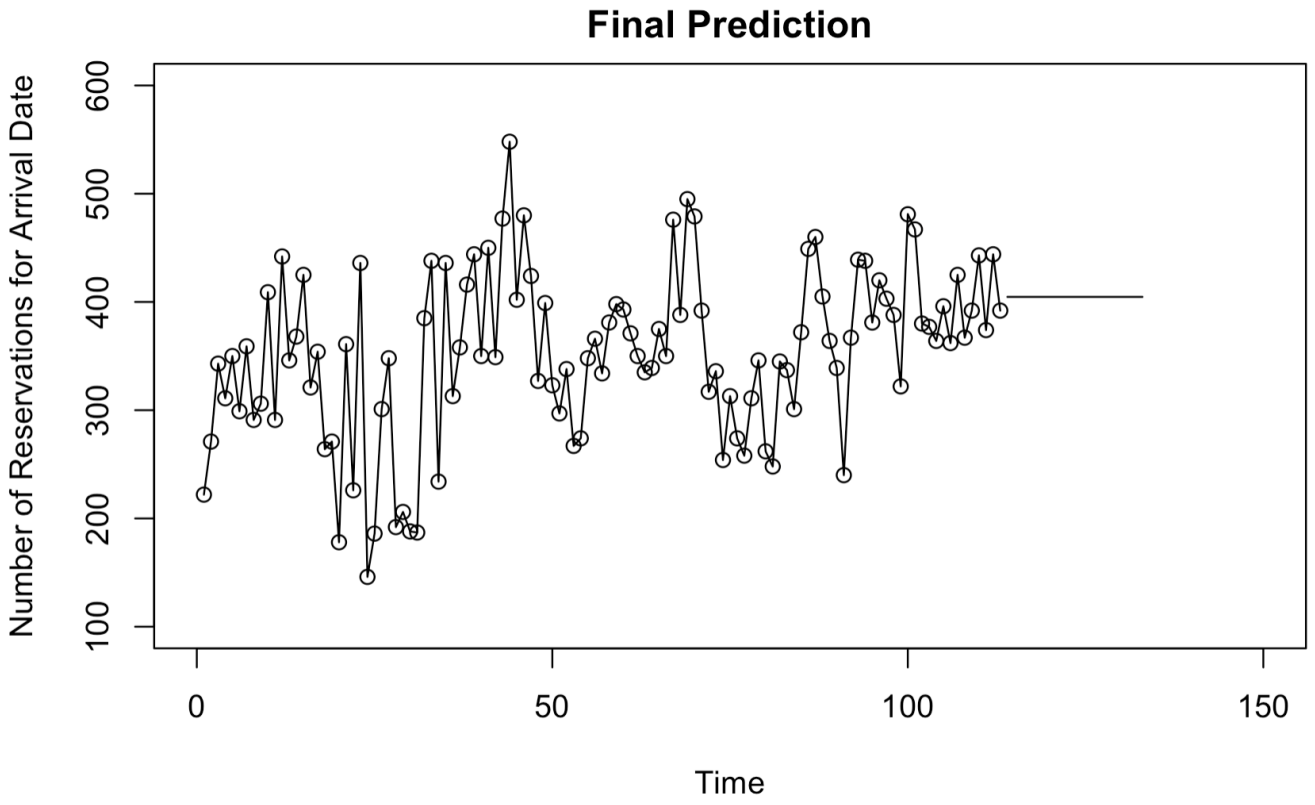

Fig. 11. Final Prediction Forecast of Number of Reservations for the Arrival Date as sum of Canceled and Not Canceled Bookings Forecast

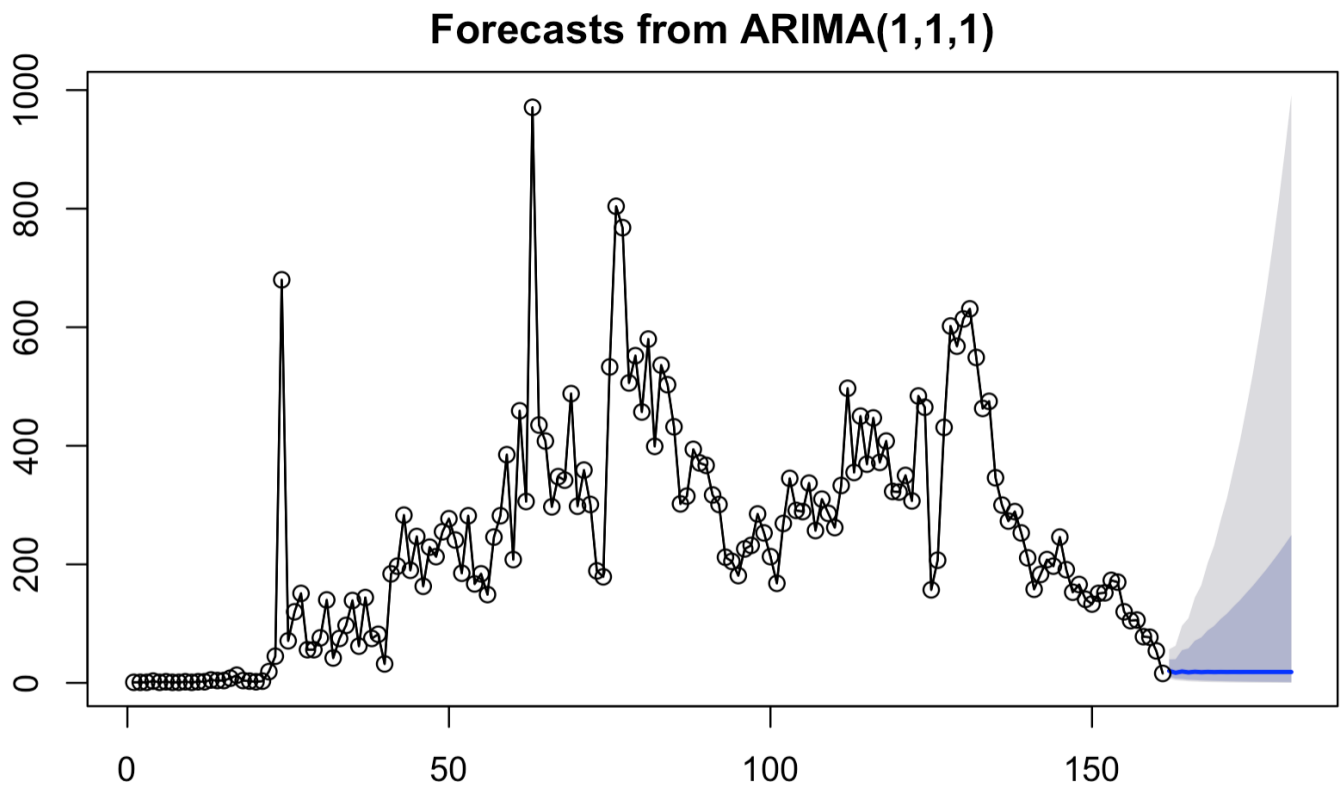

Fig. 12. Final Prediction Forecast of Number of Reservations for the Reservation Date Forecast

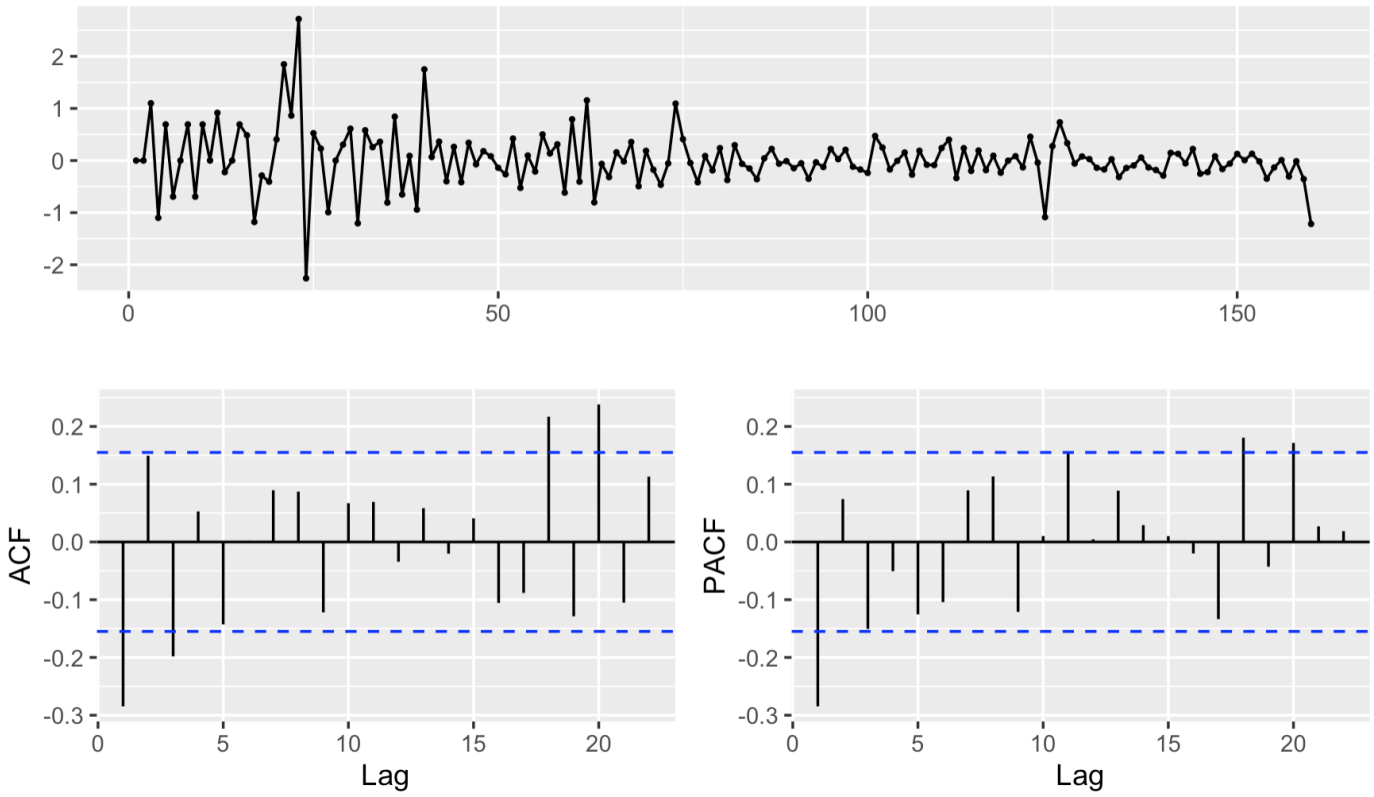

Fig. 13. White noise time series 1 order lag used in the ARIMA model