Universidade de Brasília - UnB
Instituto de Ciências Exatas - IE
Departamento de Estatística - EST

ARIMA and ETS models for public transportation riders in Portland, Oregon

Marcos Augusto Daza Barbosa

Professor: **Dr. José Augusto Fiorucci**

Brasília - DF, 2023

Abstract

Predicting demand for a transportation public service is interesting for government institutions and associated companies. The aim of this work was to apply ARIMA and ETS models for a dataset of monthly count of riders for the Portland public transportation system, from January 1960 through June 1969 . The results evidence that ETS models did not satistfy the residuals quality criterions, but performed well in terms of prediction. Also the ARIMA model with Box-Cox transformation was considered adequate after the residuals diagnostics, but had a poor prediction performance when compared to the ETS models.

Keywords: time series; arima; ets; forecast; box-cox; public transportation.

1. Introduction

Predicting the number of riders using a public transportation system is crucial for government and citizens. Overload of the system means more delays and congestion. If the count of users of those transportation services can be predicted and monitored in advance, government and associated companies can plan some strategies to prevent those issues.

In this article, the time series of a monthly count of riders for the Portland public transportation system, from January 1960 through June 1969, is studied with Time Series Analysis. The main contributions of this paper are : (1) ARIMA and ETS models were applied for the time series. (2) The residuals for both classes of models were diagnosticated (3) The prediction performance was studied via a sliding window technique. (4) Point and interval predictions were computed on a test dataset and compared to other models and benchmarks.

There are 114 observations, from which 96 were used for training the models and the last 18 as the test dataset. The following figure shows the time series for the training dataset.

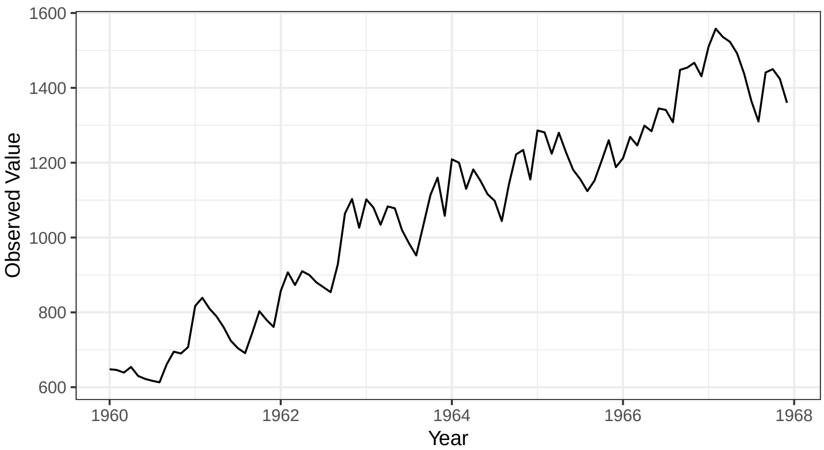

Figure 1: Portland, Oregon, public riders data

The dataset was retrieved from Kaggle [2023] website.

2. Methodology

2.1. ETS

ETS is a general class that describes the exponential smoothing models with types of components of error, trend and seasonality. For every component, there are the following possibilities:

- Error: addtive (A), multiplicative (M);
- Trend none (N), addtive (A), additive damped (Ad), multiplicative (M), multiplicative damped (Md);
- Seasonality: none (N), additive (A), multiplicative (M).

Figure 2 shows a table with the resulting methods of the combinations of trend (*Trend*) and seasonality (*Seasonal*), which is available in (Hyndman et al. [2008]).

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

In each case, ℓ_t denotes the series level at time t, b_t denotes the slope at time t, s_t denotes the seasonal component of the series at time t, and m denotes the number of seasons in a year; α, β^*, γ and ϕ are constants, $\phi_h = \phi + \phi^2 + \cdots + \phi^h$ and $h_m^+ = [(h-1) \bmod m] + 1$.

Figure 2: ETS: methods.
From Hyndman et al. [2008].

For each combination of trend and seasonality, the point forecasts are the same for added or multiplicative errors, but the interval forecasts are different. Figures 3 and 4 present tables (Hyndman et al. [2008]) with the resulting stochastic models for errors, respectively, additive and multiplicative.

Trend	Seasonal		
	N	A	M
N	$\mu_t = \ell_{t-1}$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$\mu_t = \ell_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$\mu_t = \ell_{t-1}s_{t-m}$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/\ell_{t-1}$
A	$\mu_t = \ell_{t-1} + b_{t-1}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$\mu_t = \ell_{t-1} + b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$\mu_t = (\ell_{t-1} + b_{t-1})s_{t-m}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$
A_d	$\mu_t = \ell_{t-1} + \phi b_{t-1}$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	$\mu_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$\mu_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$
M	$\mu_t = \ell_{t-1}b_{t-1}$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$	$\mu_t = \ell_{t-1}b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$\mu_t = \ell_{t-1}b_{t-1}s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1})$
M_d	$\mu_t = \ell_{t-1}b_{t-1}^{\phi}$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi} + \alpha\varepsilon_t$ $b_t = b_{t-1}^{\phi} + \beta\varepsilon_t/\ell_{t-1}$	$\mu_t = \ell_{t-1}b_{t-1}^{\phi} + s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi} + \alpha\varepsilon_t$ $b_t = b_{t-1}^{\phi} + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$\mu_t = \ell_{t-1}b_{t-1}^{\phi}s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1}^{\phi} + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1}^{\phi})$

Figure 3: ETS: additive errors.
From Hyndman et al. [2008].

Trend	Seasonal		
	N	A	M
N	$\mu_t = \ell_{t-1}$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	$\mu_t = \ell_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$\mu_t = \ell_{t-1}s_{t-m}$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$\mu_t = \ell_{t-1} + b_{t-1}$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$\mu_t = \ell_{t-1} + b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$\mu_t = (\ell_{t-1} + b_{t-1})s_{t-m}$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A_d	$\mu_t = \ell_{t-1} + \phi b_{t-1}$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$\mu_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$\mu_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M	$\mu_t = \ell_{t-1}b_{t-1}$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$	$\mu_t = \ell_{t-1}b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$	$\mu_t = \ell_{t-1}b_{t-1}s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M_d	$\mu_t = \ell_{t-1}b_{t-1}^{\phi}$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}^{\phi}(1 + \beta\varepsilon_t)$	$\mu_t = \ell_{t-1}b_{t-1}^{\phi} + s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi} + \alpha(\ell_{t-1}b_{t-1}^{\phi} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1}^{\phi} + \beta(\ell_{t-1}b_{t-1}^{\phi} + s_{t-m})\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1}^{\phi} + s_{t-m})\varepsilon_t$	$\mu_t = \ell_{t-1}b_{t-1}^{\phi}s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}^{\phi}(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

Figure 4: ETS: multiplicative errors.
From Hyndman et al. [2008].

The model description is given by ETS(·, ·, ·), where the first component is equivalent to the type of error, the second to the type of trend and the third to the type of seasonality. The SES model is equivalent to ETS(A, N, N), Holt to ETS(A, A, N), Holt with amortized trend to ETS(A, Ad, N), Holt-Winters additive to ETS(A, A, A) and multiplicative Holt-Winters to ETS(A, A, M).

2.2. ARMA

The autoregressive moving average ARMA(p, q) is defined as

$$y_t = \alpha + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

where $\alpha = (1 - \phi_1 - \cdots - \phi_p)\mu$ is the intercept and $\mu = E[y_t]$ is the unconditional mean of the time series. The parameters α, $\phi_1, \ldots,$ ϕ_p, $\theta_1, \ldots,$ θ_q are constants.

The punctual predictions for the model are

$$\hat{y}_{n+h|n} = E_n[y_{n+h}] = \alpha + \phi_1 y^*_{n+h-1} + \cdots + \phi_p y^*_{n+h-p} + \theta_1 \varepsilon^*_{n+h-1} + \cdots + \theta_q \varepsilon^*_{n+h-q},$$

where

$$y^*_j = \begin{cases} y_j,\ j \leq n \\ \hat{y}_{j|n},\ j > n \end{cases}$$

and

$$\varepsilon^*_j = \begin{cases} \varepsilon_j,\ j \leq n \\ 0,\ j > n \end{cases}$$

Recursively, the punctual predictions for the ARMA model can be computed with the above equations for any future horizon $h = 1, 2, 3, \ldots$.

With the assumption that the errors are i.i.d $\mathcal{N}(0, \sigma^2_\varepsilon)$, the interval predction for h-steps ahead with a confidence of $(1 - a)\%$ is given by

$$\hat{y}_{n+h|n} \pm z_{a/2} \sqrt{\sigma^2_\varepsilon \left[1 + \sum_{i=1}^{h-1} \psi^2_i\right]}, \qquad h = 1, 2, 3, \ldots,$$

where the coefficients ψ_i are given by

$$\psi_i = \theta_i + \sum_{j=1}^{i} \phi_j \psi_{i-j}, \qquad j = 1, 2, 3, \ldots,$$

com $\psi_0 = 1$, $\phi_j = 0$ para $j > p$ e $\theta_i = 0$ para $i > q$.

2.3. ARIMA

ARIMA models are applied only for stationary time series. If the time series is not stationary and not seasonal, it is required to take simple differences to make it stationary. Therefore, it is possible to use the ARMA model, because the ARIMA(p, d, q) is the model ARMA(p, q) applied to the $d - th$ difference of the time series. The model ARIMA(p, d, q) can be written as

$$\Phi_p(B)(1 - B)^d y_t = \Theta_q(B)\varepsilon_t.$$

Therefore, to compute punctual predictions for ARIMA models, just take the number of simple differences to make the original series stationary (d) and, then, apply the procedure described in Section 2.2. Finally, after calculating the punctual predictions for the stationary series, just take the inverse transformation that was used on the original series.

2.4. SARIMA

If the time series is not stationary and seasonal, it is required to takes simple and seasonal differences to make it stationary and non seasonal. Next, it is possible to use the ARMA model. The numbers of simple and seasonal differences are written as, respectively, d and D. The model SARIMA(p, d, q)x(P, D, Q)$_m$ can be written as

$$\Phi_P(B^m)\Phi_p(B)(1-B^m)^D(1-B)^d y_t = \Theta_Q(B^m)\Theta_q(B)\varepsilon_t.$$

Therefore, to compute punctual predictions SARIMA models, just take the d simple differences to make the series stationary , take D seasonal differences to make the series non seasonal and, then, apply the procedure described in Section 2.2. Finally, after calculating the punctual predictions for the non stationary and non seasonal series, just take the inverse transformations that were used on the original series.

2.5. Computational Implementation

All the results were computed with R, version 4.2.2 (R Core Team [2020]). The ARIMA and ETS models used are available on `forecast` package (Hyndman et al. [2017]), version 8.19. The machine had an Intel(R) Core(TM) i7-7500U CPU and 8GB of RAM, with a Windows 10 operating system.

3. Results

Before training the models, it was observed by Figure 1 that there is a growing trend throughout the years. Also, time series was decomposed by STL.

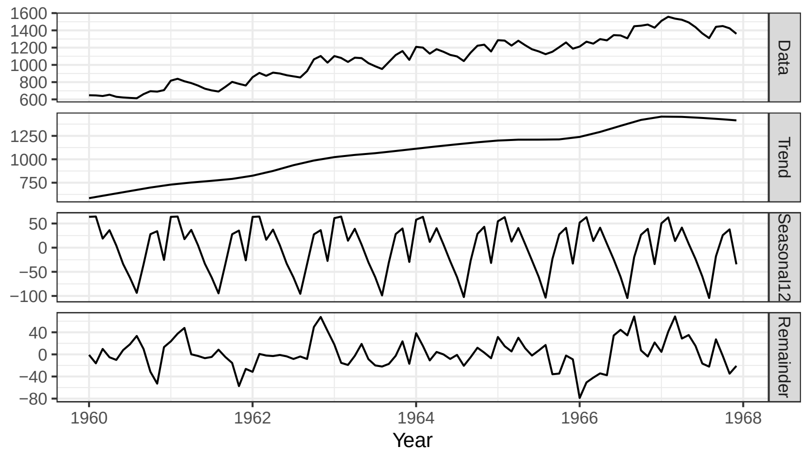

Figure 5: STL Decomposition

By Figure 5, the trend component grows almost every year. There is a slow decay starting from 1967. The seasonality trend has a seasonal cycle of size 12, which was expected because it is monthly data, and it seems that is constant. For the error component, the values generally fluctuates around zero, but close to 1966 they fall apart.

3.1. ARIMA models

In this section, ARIMA models will be chosen from a set of candidate models of that family.

After some statistical tests, it is verified that one simple difference and one seasonal difference are required to make the time series stationary.

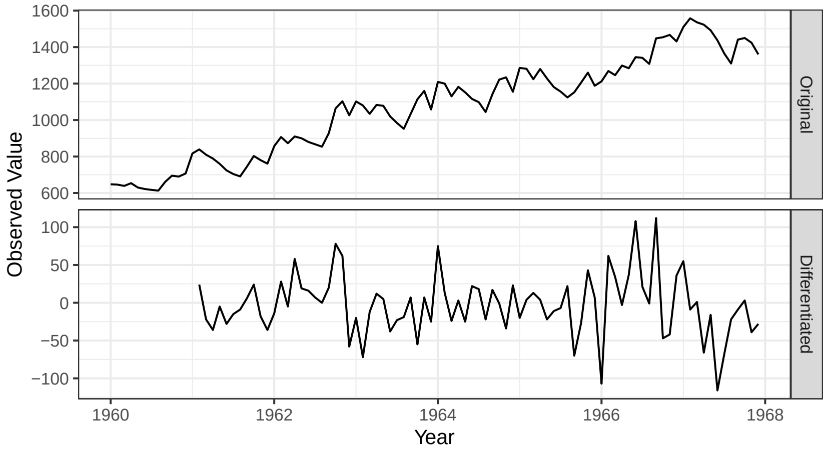

Figure 6: Differentiation

Figure 6 shows stationarity for the differenced time series. To check stationarity with a statistical test, the Augmented Dickey-Fuller procedure was executed that confirmed the assumption (p-value = 0.01492).

To select the parameters p, q, P and Q, the autocorrelation and partial autocorrelation plots were investigated.

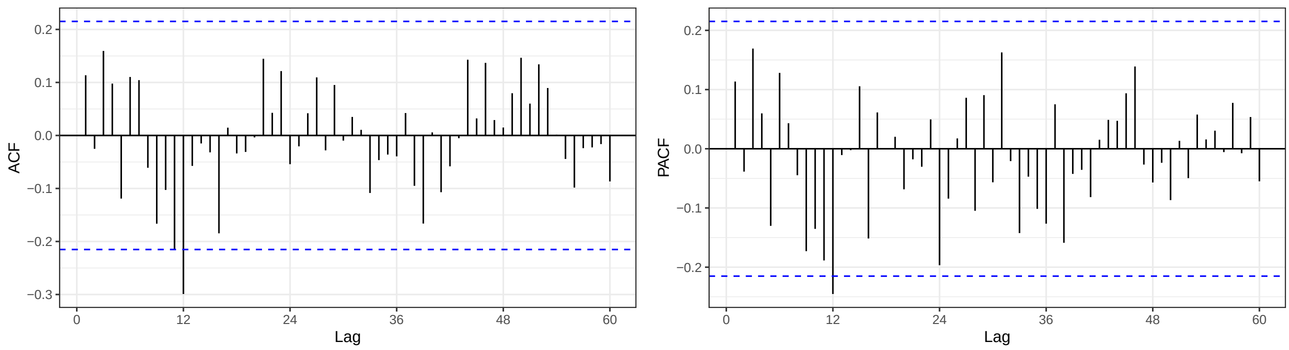

Figure 7: ACF and PACF

By analysing the ACF and PACF plots together, the non seasonal lag values generally are not significant, so the configuration $q = 0$ and $p = 0$ is an option. For the seasonal lags, particularly lag 12, the ACF value sticks out considerably from the confidence interval and for the PACF plot, looking the seasonal lags, clearly lag 12 has large partial correlation value. Therefore $P = 1$ and $Q = 1$ seems a good alternative, but $P = 0$, $P = 2$, $Q = 0$ and $Q = 2$ are going to be considered also.

After training all ARIMA models from the combination of possible values for p, q, P

and Q, the best model using the Box-Cox transformation and not using, with the lowest AICc, were the following.

Table 1: ARIMA models

Model	λ	**AICc**
$SARIMA(0,1,0) \times (0,1,2)_{12}$	0	836.98
$SARIMA(0,1,0) \times (1,1,1)_{12}$	0.174	-128.36

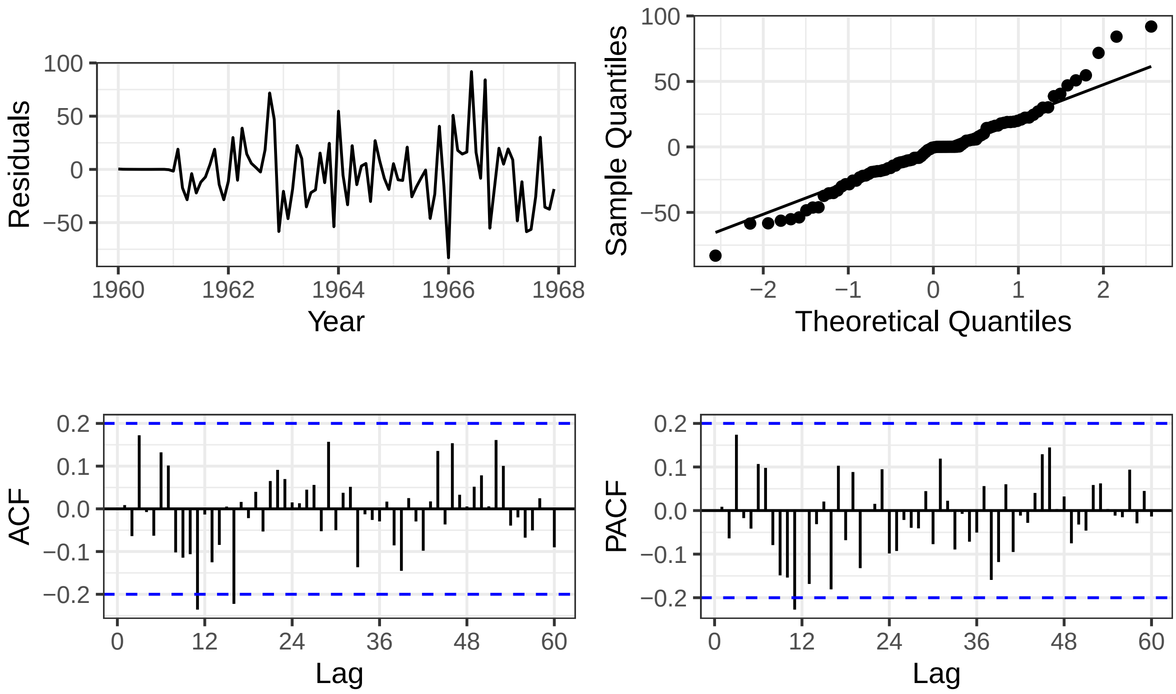

Figure 8: Residuals Analysis for ARIMA

By Figure 8, the ACF and PACF plots show some mild correlation at lags 11 and 16, but Ljung-Box does not rejects independence. Clearly, the QQ-plot shows some great deviance from the Normal distribution (Shapiro-Wilk rejects null hypothesis of normality). For stationarity the mean seems to be around zero, but the variance shows some unconstant behaviour, but Augmented Dickey-Fuller rejects null hypothesis of a unit root.

Therefore, the best ARIMA model without Box-Cox transformation does not attend every residual criterion.

Next, the ARIMA model with the Box-Cox transformation was chosen and the residuals behaviour are shown.

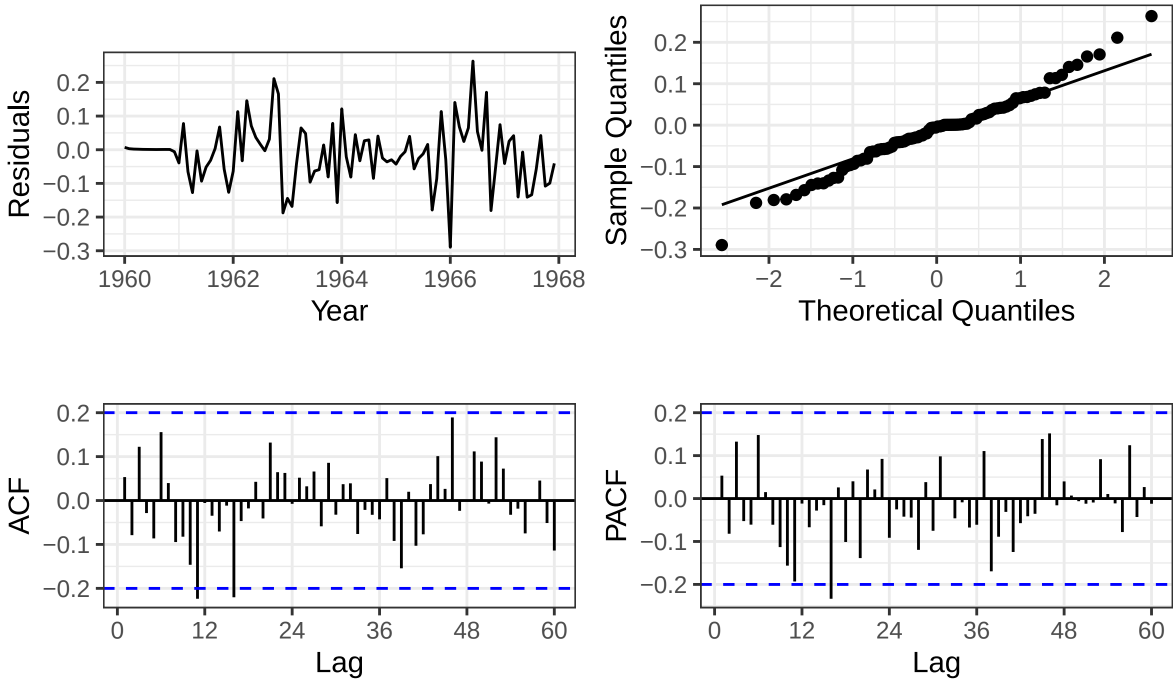

Figure 9: Residuals Analysis for ARIMA with Box-Cox

The ACF and PACF residuals correlations appear to not be significant in general, but two points are a little greater than the confidence interval limit (Ljung-Box does not rejects independence). For stationarity the mean seems to be around zero, but the variance shows some unconstant behaviour. The QQ-PLot shows a little deviance for the tails. Although the last assumptions are not clearly checked visually, Augmented Dickey-Fuller rejects null hypothesis of a unit root and Shapiro-Wilk does not reject null hypothesis of normality.

Therefore, the ARIMA model with Box-Cox transformation does satisfy all errors criteria, if only the statistical tests are considered.

3.2. ETS models

The ETS models were selected automatically by the associated function `ets` in `forecast` package, with and without Box-Cox transformation. The best models were the following.

Table 2: ETS models

Model	λ	**AICc**
$ETS(M, Ad, M)$	0	1221.841
$ETS(A, Ad, A)$	0.174	19.665

The best model, without Box-Cox, is an ETS(M, Ad, M), i.e, multiplicative errors, additive with damped trend and, finally, multiplicative seasonality, with and AICc of 1221.84.

For the best model, with Box-Cox, the selected model is an ETS model with additive errors, additive with damped trend and additive seasonality, with the $\lambda = 0.174$. The AICc is 19.665.

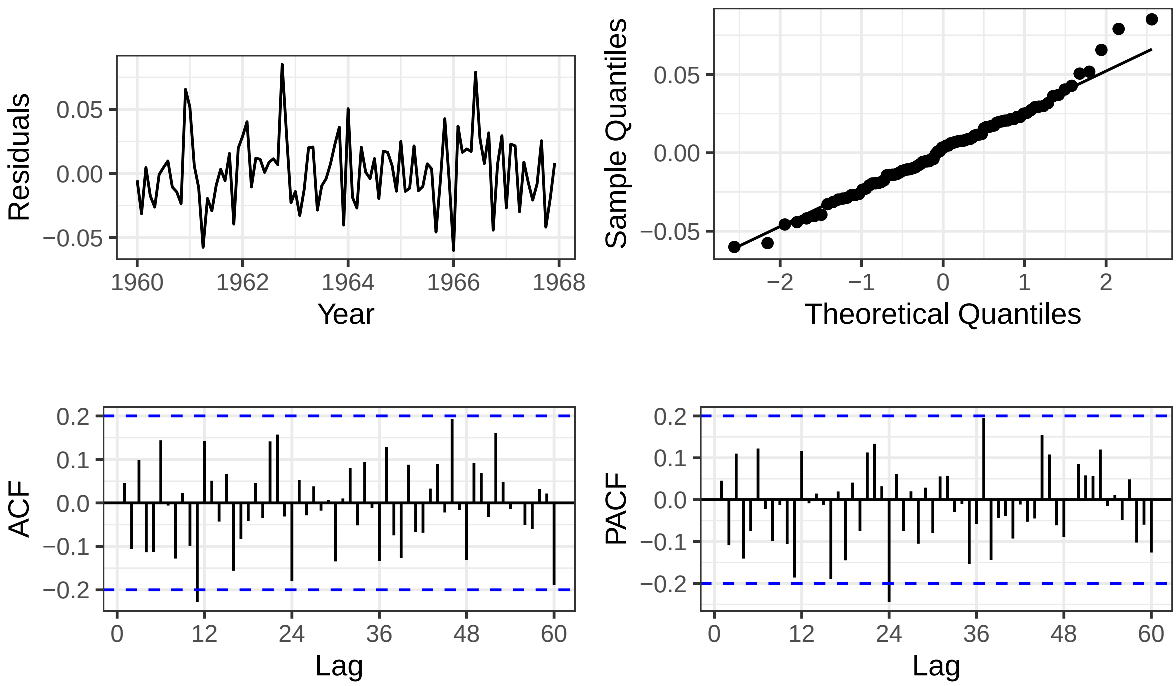

Figure 10: Residuals Analysis for ETS

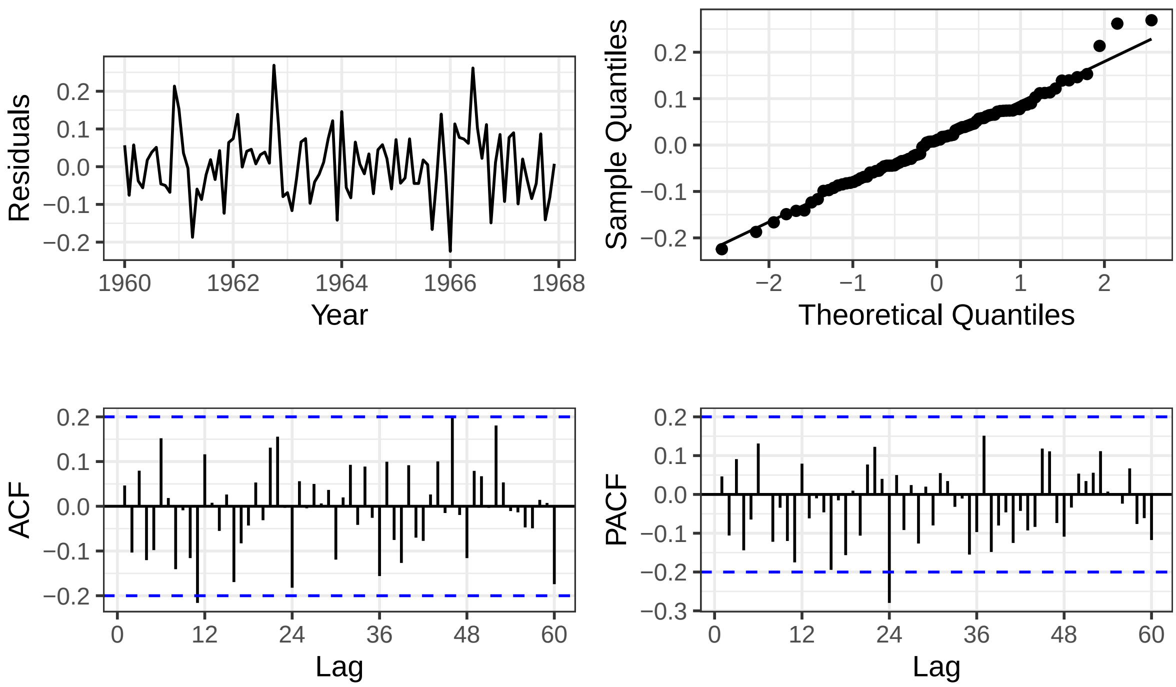

Figure 11: Residuals Analysis for ETS with Box-Cox

Figure 10 and Figure 11 have the same interpretation: the residuals appear to be stationary (Augmented Dickey-Fuller rejects null hypothesis of a unit root), following a normal distribution with some deviance at the tails (Shapiro-Wilk does not reject null hypothesis), but evidently there is still some correlation that can be seen in ACF and PACF plots (Ljung-Box rejects null hypothesis of errors independence).

Therefore, the ETS models are not adequate at every error criterion.

3.3. Prediction Performance

The four models prediction performance were evaluated using a sliding window, starting at the $n - 14$ point and forecasting for 5 steps ahead, only using the training data.

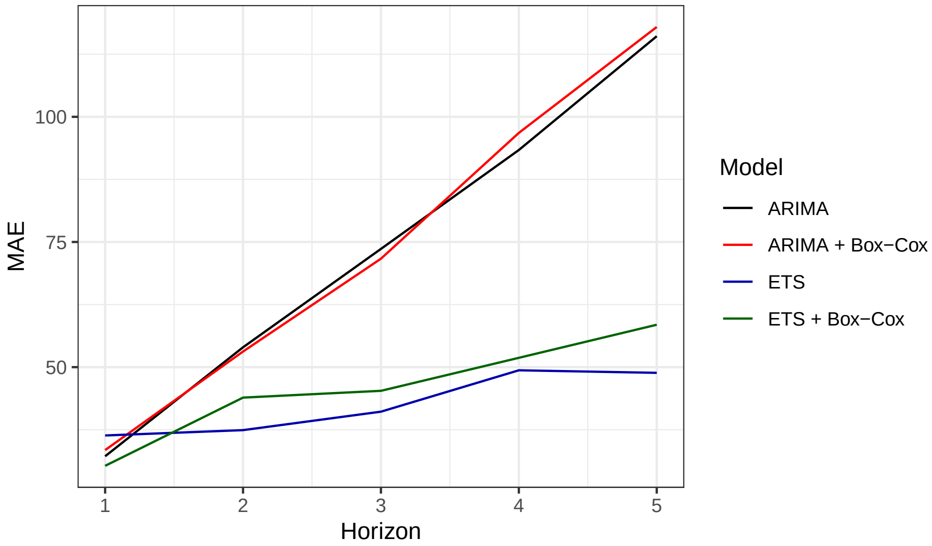

Figure 12: Prediction performance by horizon

By Figure 12, at the first horizon, the four models got Mean Absolute Error's close to each other. For the next 4 horizons, the ETS model had the best performance, followed by the ETS model with Box-Cox. The ARIMA models had a very similar prediction performance and were the worst ones.

Next, punctual and interval predictions were computed for each model, using the test data. All results were rounded to nearest whole number.

Table 3: Punctual predictions on test dataset

	Observed	ARIMA	ARIMA + Box-Cox	ETS	ETS + Box-Cox
1	1429	1457	1468	1474	1475
2	1440	1480	1499	1487	1486
3	1414	1447	1463	1434	1429
4	1424	1462	1480	1466	1459
5	1408	1434	1448	1423	1417
6	1337	1392	1408	1380	1378
7	1258	1352	1363	1338	1339
8	1214	1316	1322	1294	1299
9	1326	1406	1434	1388	1392
10	1417	1452	1486	1457	1459
11	1417	1454	1485	1459	1462
12	1329	1397	1424	1381	1391
13	1461	1493	1548	1497	1507
14	1425	1510	1574	1509	1518
15	1419	1476	1532	1456	1460
16	1432	1503	1561	1487	1489
17	1394	1478	1527	1443	1446
18	1327	1452	1489	1399	1405

Table 4: Interval predictions on test dataset

		ARIMA		ARIMA + Box-Cox		ETS		ETS + Box-Cox	
	Observed	**Inf**	**Sup**	**Inf**	**Sup**	**Inf**	**Sup**	**Inf**	**Sup**
1	1429	1390	1525	1386	1554	1389	1560	1396	1556
2	1440	1385	1575	1382	1623	1364	1609	1379	1600
3	1414	1331	1563	1324	1613	1290	1579	1305	1564
4	1424	1328	1595	1320	1657	1295	1636	1314	1616
5	1408	1284	1583	1273	1643	1238	1608	1260	1590
6	1337	1228	1556	1222	1617	1184	1576	1211	1563
7	1258	1175	1529	1169	1584	1133	1544	1164	1535
8	1214	1127	1505	1120	1553	1082	1507	1117	1503
9	1326	1205	1606	1206	1698	1147	1629	1189	1623
10	1417	1240	1663	1239	1773	1190	1724	1238	1712
11	1417	1233	1676	1227	1787	1179	1739	1230	1728
12	1329	1165	1629	1165	1730	1104	1659	1159	1659
13	1461	1243	1743	1248	1905	1185	1810	1250	1806
14	1425	1244	1777	1251	1964	1182	1836	1251	1830
15	1419	1194	1758	1198	1939	1129	1782	1193	1774
16	1432	1206	1799	1205	1999	1143	1831	1210	1819
17	1394	1167	1789	1162	1980	1099	1787	1166	1780
18	1327	1127	1776	1118	1956	1056	1742	1125	1742

The following figures demonstrate visually all the predictions for test dataset.

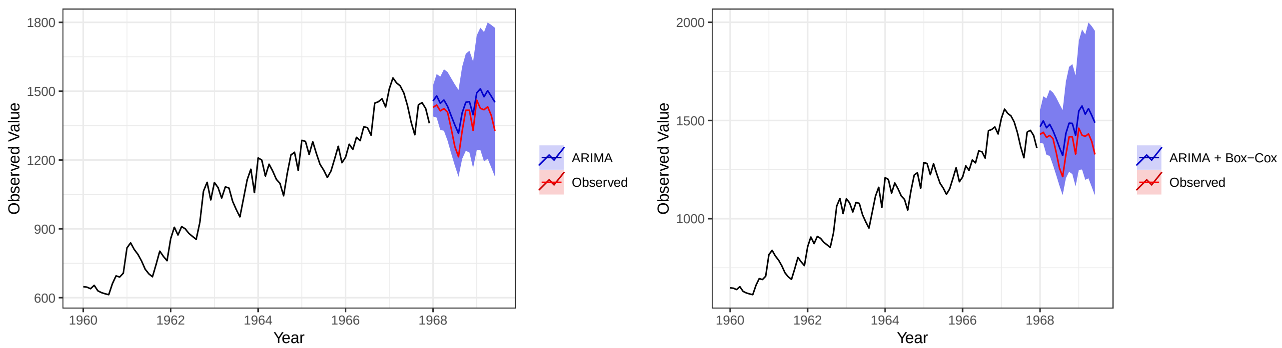

Figure 13: Forecast for ARIMA models on test dataset

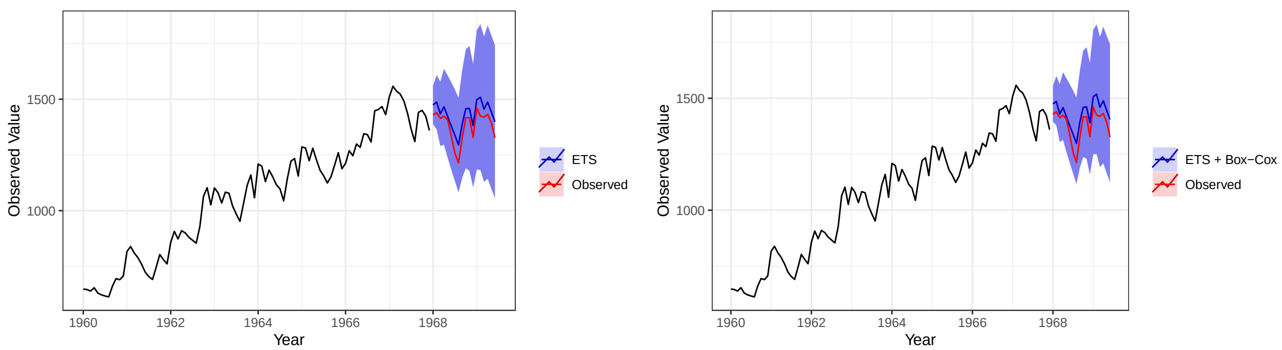

Figure 14: Forecast for ETS models on test dataset

The prediction plots show that both ETS models had closer predictions to the real observed values.

Finally, the performance of the four models were compared to BATS, TBATS models and the benchmarks available on `forecast` package (`auto.arima`, `ses`, `holt`, `stlf`).

Table 5: Predictive performance based on the test dataset

Modelo	**MAE**
BATS	37.51
TBATS	43.47
ETS	50.12
ETS + Box-Cox	52.15
ARIMA	60.50
`ses`	61.06
`holt`	77.84
ARIMA + Box-Cox	91.18
`sltf`	103.97
`auto.arima`	165.25

BATS and TBATS models had the lowest mean absolute error's, followed by the ETS and the ETS with Box-Cox transformation models. The ARIMA models performed considerably better than the associated benchmark `auto.arima`

4. Conclusion

In this article, SARIMA and ETS models were trained and it was observed that although ETS models did not satistfy the residuals quality criterions, they performed well in terms of prediction on the test dataset. Also the ARIMA model with Box-Cox transformation was considered adequate after the residuals diagnostics, but had a poor prediction performance when compared to the ETS models.

References

R Hyndman, G Athanasopoulos, C Bergmeir, G Caceres, L Chhay, M O'Hara-Wild, F Petropoulos, S Razbash, E Wang, and F Yasmeen. forecast: Forecasting functions for time series and linear models, 2019, r package version 8.7, 2017.

Rob Hyndman, Anne B Koehler, J Keith Ord, and Ralph D Snyder. *Forecasting with exponential smoothing: the state space approach.* Springer Science & Business Media, 2008.

Kaggle. Portland oregon riders monthly data, 2023. `https://www.kaggle.com/datasets/hsankesara/portland-oregon-avg-rider-monthly-data`, Last accessed on 2023-02-05.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2020. URL `https://www.R-project.org/`.