What is the best forecast?

Francisco I. dos R. Ferreira

November 12, 2023

Abstract

A brief comparison of three time series forecasting models.

1 Introduction

Time Series Analysis is one of the compulsory subjects in the statistics course at the University of Brasilia (UnB). It introduces students to the main techniques, especially decomposition techniques, transformations, filters, model selection, adjustment and diagnosis, and the calculation of point and interval forecasts. By the end of the course, we have developed the ability to identify models, estimate the parameters involved and calculate forecasts for the main types of time series. In one of the assessment activities, students were divided into groups and asked to evaluate one of the series that made up the M3 competition database. My group and I achieved the maximum score. After a while I got to know the **Prophet**, and I was enthusiastic about using it. I then decided to use it as a basis for comparing three time series models. The first, chosen manually, the second automatically and the last using Prophet. In the end, we'll be able to answer the question behind this short article.

2 Residual Analysis

2.1 Manual model

In our study, we chose $SARIMA(0,1,3)x(0,1,1)_{12}$. The model's residuals can be seen in the figure below.

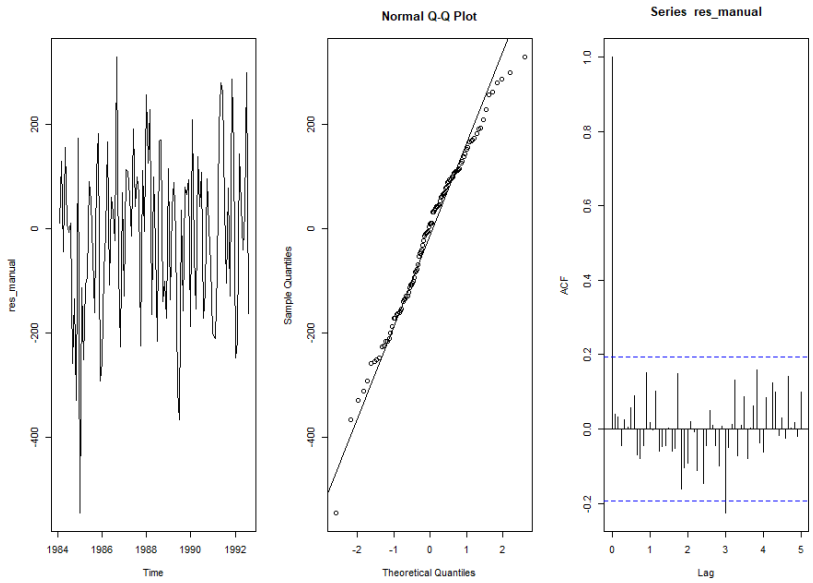

Figure 1: Residuals analysis of the manual model

The the graph shows various oscillations which, on average, are centered around zero. around zero. The extremities, both upper and lower, are equidistant from the average. from the average, and no behavior can be seen that indicates that the variance is correlated with time. Instead, it appears to be constant regardless of the time period. period, a behavior similar to that observed in white noise.

Comparing the quantiles of the normal distribution with the distributed points of the errors shows that the two distributions are very close. The straight line coincides almost with the random residuals, with the points having a higher concentration concentration around 0 and lower concentration at the extremities. In the autocorrelation graph, on the other hand, for lags other than zero (where the autocorrelation is always 1), the general behavior is autocorrelation contained within the interval. There is no graphical evidence of autocorrelation between the residuals.

2.2 Automatic model

The automatic model chosen by the software was $SARIMA(2,1,2)x(1,1,0)_{12}$

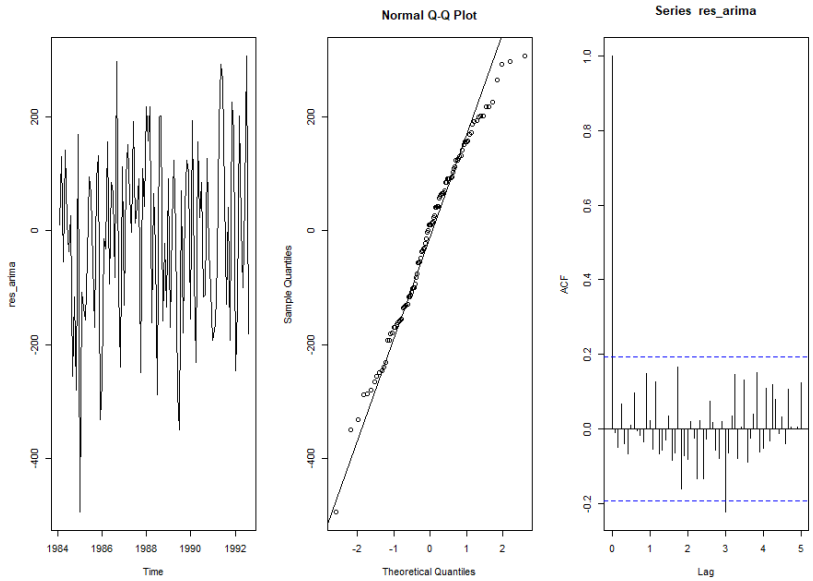

Figure 2: Residuals analysis of the Automatic model

we can see a distribution similar to the manual model. An average oscillation around zero over time, which indicates independence. A similarity in the distribution line of the residuals, with a lower concentration at the ends. As well as an autocorrelation contained within the interval, thus indicating that there is no autocorrelation in the residuals.

2.3 Prophet

The model adjusted via Prophet uses a different approach, relying on additive components to model the trend, seasonality and holidays. Therefore, as with the previous models, we will not be able to identify whether the model is, for example, a SARIMA(p,d,q)x(P,D,Q).

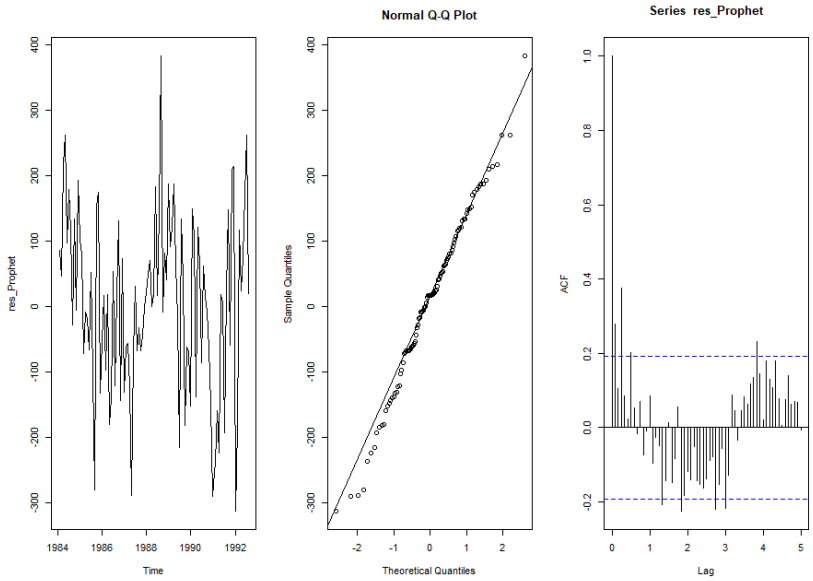

Figure 3: Residuals analysis of the Prophet model

We noticed a different behavior from the other models when we evaluated the residual part. Despite this, we can still see its variability around zero. The quantile graph shows the points well distributed along the straight line, with some points at the far end, which is to be expected. The autocorrelation graph, on the other hand, shows that there may have been autocorrelation in the residuals, since the observations tend to escape the limits.

2.4 Hypothesis testing

- KPSS test

MODEL	P-Value
Manual	0.1
Automatic	0.1
Prophet	0.1

- Shapiro-Wilk test

MODEL	W	P-Value
Manual	0.98538	0.318
Automatic	0.98366	0.2358
Prophet	0.99182	0.7932

- Box-Ljung test

MODEL	χ^2	DF	P-Value
Manual	9.1918	20	0.9806
Automatic	10.181	20	0.9648
Prophet	45.696	20	0.0008872

As seen in the graphs, all the residuals from the three models follow a normal distribution and are stationary. The manual and automatic models also show that there is no dependence in their residuals; on the other hand, it is clear that there is autocorrelation in the residuals of the Prophet model.

3 Evaluating MAE

Next, we will focus on evaluating the Mean Absolute Error of the three models, as a central metric to measure how close the models are to replicating the real values of the time series. You can see the forecasts for each one here[1].

MODEL	MAE	MAPE
Manual	778.3053	7.75%
Automatic	800.5752	7.97%
Prophet	1072.0791	10.71%

4 Conclusion

As you can see, the manual and automatic models performed better than the Prophet model. The manual model was the winner when we evaluated its mean absolute error. It is true that simply giving the software the power to decide which time series model to use is not always enough. In cases like this, it is extremely important that the researcher has the theoretical knowledge to interpret the results in a critical and informed manner. The choice between time series models, whether manual or automatic, must take into account not only performance metrics, but also an understanding of the specific context of the application. Theoretical knowledge makes it possible to adjust parameters in a more informed way, identify subtle patterns in the data and adapt modeling strategies to the particularities of the problem in question.In addition, the ability to interpret results in a contextualized way is crucial to avoiding hasty conclusions or misinterpretations. Therefore, the combination of theoretical expertise and critical analysis is essential to ensure the robustness and reliability of the conclusions drawn from the time series models used.

[1] https://github.com/FerreiraFrancisco/Forecast-2334/tree/main