ARIMA Modelling and Forecasting of the Student Time Series

Universidad Politécnica de Madrid — Escuela Técnica Superior de Ingenieros Informáticos

Emanuele Alberti
Ottavia Biagi
Leandro Duarte
Master's in Computer Engineering and Data Science
Time Series Analysis — ARIMA Assignment
Universidad Politécnica de Madrid (UPM)

***Abstract*—This report follows the Box–Jenkins methodology to model the *Student series*, a short univariate time series of 197 observations. The analysis includes exploratory characterization, stationarity assessment, ACF/PACF-based model identification, ARIMA estimation, residual diagnostics, forecasting, and stochastic simulation. Among multiple candidates, ARIMA(1,0,1) provides the lowest AIC/BIC and exhibits residuals consistent with white noise according to the Ljung–Box test. Forecasts tend to revert to the series mean and simulations confirm the stationary nature of the model.**

I. Introduction

This assignment applies the Box–Jenkins methodology to identify the most appropriate ARIMA model for the Student series. The procedure includes:

1) characterizing and preprocessing the series,
2) identifying candidate models using ACF/PACF,
3) estimating and comparing ARIMA models,
4) diagnosing residuals and validating assumptions,
5) producing forecasts and stochastic simulations.

II. Exploratory Data Analysis

The dataset contains 197 numerical observations. Descriptive statistics are reported in Table I.

TABLE I: Descriptive statistics of the Student series.

Statistic	Value
Minimum	16.10
1st Quartile	16.80
Median	17.00
Mean	17.06
3rd Quartile	17.40
Maximum	18.20

The series plot (Fig. 1) shows no trend or seasonality and appears visually stationary.

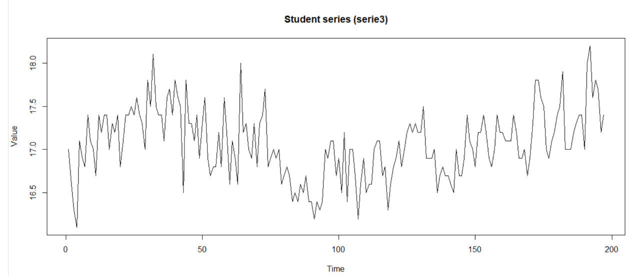

Fig. 1: Time series plot of the Student series.

III. Stationarity Assessment

The Augmented Dickey–Fuller test returned a statistic of -2.656 with $p = 0.3014$, failing to reject the unit-root null. However, the visual pattern, constant variance, and ACF decay strongly suggest stationarity. Given the low noise level and absence of trend, we proceed with $d = 0$ and consider non-differenced ARIMA models.

IV. Model Identification

Figure 2 shows the ACF and PACF of the series.

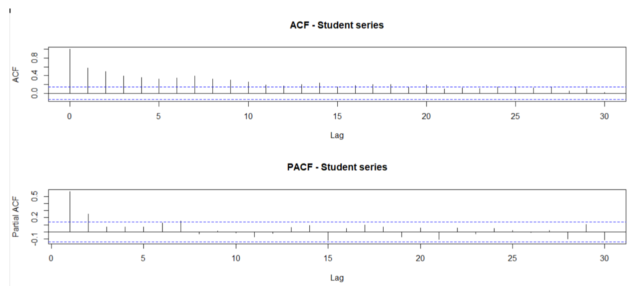

Fig. 2: ACF and PACF of the Student series.

The ACF exhibits a smooth decay, while the PACF shows a strong spike at lag 1. This pattern is compatible with AR(1) or ARMA(1,1) structures. Thus, the candidate models selected for estimation were:

- ARIMA(1,0,0),
- ARIMA(0,0,1),
- ARIMA(1,0,1).

V. Model Estimation and Comparison

Table II reports AIC and BIC values.

TABLE II: Model comparison using AIC and BIC.

Model	AIC	BIC
ARIMA(1,0,0)	124.88	134.73
ARIMA(0,0,1)	156.15	166.00
ARIMA(1,0,1)	**109.49**	**122.62**

ARIMA(1,0,1) provides the best fit. Estimated parameters are:

$$\phi_1 = 0.9087, \quad \theta_1 = -0.5758, \quad \mu = 17.0654.$$

The AR coefficient indicates strong persistence, while the MA component corrects short-term autocorrelation.

VI. Residual Diagnostics

A valid ARIMA model should produce residuals that behave as white noise: uncorrelated, zero mean, constant variance, and approximately normal.

Figure 3 shows the residuals over time. No trend or changing variance is visible, and residuals fluctuate around zero.

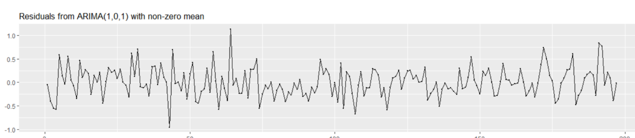

Fig. 3: Residual time series from ARIMA(1,0,1).

Further checks are provided by the residual ACF in Figure 4. All autocorrelation spikes lie within the 95% confidence bands, suggesting that no serial dependence remains in the residuals.

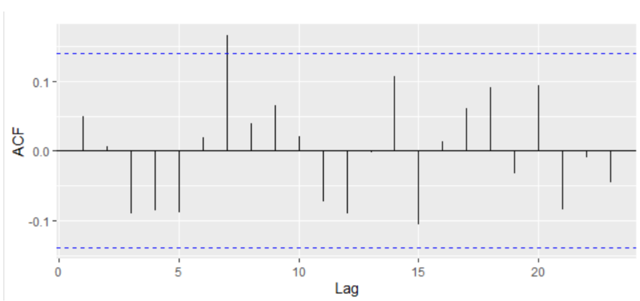

Fig. 4: Residual autocorrelation function (ACF).

Finally, the lag–residual plot in Figure 5 shows no systematic pattern, reinforcing the conclusion that residuals behave as white noise.

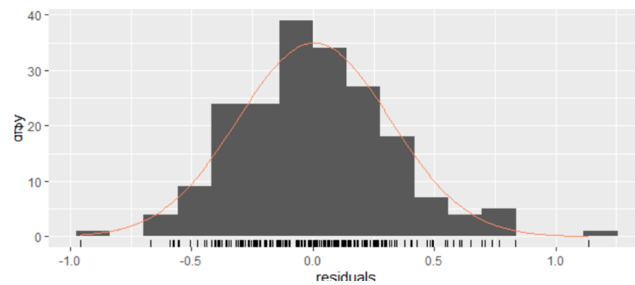

Fig. 5: Lag plot of residuals from ARIMA(1,0,1).

The Ljung–Box test yielded:

$$Q^* = 12.338, \quad \text{df} = 8, \quad p = 0.1368,$$

confirming that residuals exhibit no significant autocorrelation. Thus, ARIMA(1,0,1) meets the Box–Jenkins adequacy conditions.

VII. Forecasting

A 10-step forecast is shown in Figure 6.

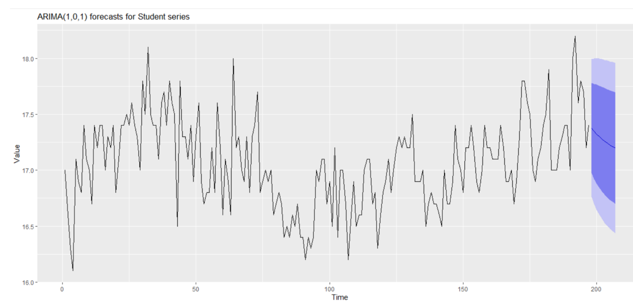

Fig. 6: ARIMA(1,0,1) 10-step forecast with confidence intervals.

Forecasts revert to the mean (around 17), and the widening bands reflect increasing uncertainty, typical of stationary ARMA processes.

VIII. Stochastic Simulation

Figure 7 shows 20 simulated future trajectories of length 10.

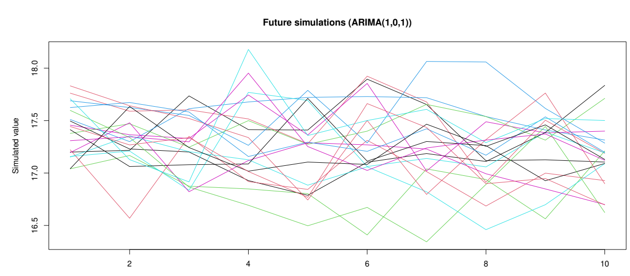

Fig. 7: Simulated trajectories generated from ARIMA(1,0,1).

All trajectories oscillate around the long-term mean, confirming stationarity and validating the selected model.

IX. Conclusion

Following the Box–Jenkins methodology, ARIMA(1,0,1) emerged as the best model for the Student series. It:

- achieved the lowest AIC and BIC among tested models,
- produced residuals consistent with white noise,
- generated stable forecasts reverting to the mean,
- and produced simulations compatible with a stationary ARMA process.

Despite the ADF result suggesting non-stationarity, the combination of visual evidence, ACF/PACF structure, and residual diagnostics supported the use of $d = 0$. This highlights the importance of combining statistical tests with expert judgement when applying the Box–Jenkins methodology.

Appendix A
R Code Used in the Analysis

```r
rm(list = ls())
library(forecast)
library(tseries)
library(ggplot2)

setwd("Working with ARIMA in R - 2023")

# Load Student series
raw <- read.table("Student series.txt", header = FALSE)
y <- as.numeric(raw$V1[-1])
ts_y <- ts(y, frequency = 1)

# Exploratory analysis
summary(y)
plot(ts_y,
     main = "Student series (serie3)",
     ylab = "Value",
     xlab = "Time")

# Stationarity test
adf.test(ts_y)

# ACF and PACF
par(mfrow = c(2, 1))
acf(ts_y, lag.max = 30,
    main = "ACF - Student series")
pacf(ts_y, lag.max = 30,
     main = "PACF - Student series")
par(mfrow = c(1, 1))

# Candidate ARIMA models
m_ar1    <- Arima(ts_y, order = c(1, 0, 0))
m_arma11 <- Arima(ts_y, order = c(1, 0, 1))
m_ma1    <- Arima(ts_y, order = c(0, 0, 1))

AIC(m_ar1);    BIC(m_ar1)
AIC(m_arma11); BIC(m_arma11)
AIC(m_ma1);    BIC(m_ma1)

# Residual diagnostics for ARIMA(1,0,1)
checkresiduals(m_arma11)

# Forecast and simulation
fc <- forecast(m_arma11, h = 10)
autoplot(fc)

set.seed(123)
sim_mat <- replicate(20,
                     simulate(m_arma11, nsim = 10))
matplot(sim_mat, type = "l")
```