MA 641 Time Series Analysis
Homework 5

Nicolas Jorquera

Problem 2

From a time series of 100 observations, we calculate $r_1 = -0.49$, $r_2 = 0.31$, $r_3 = -0.21$, $r_4 = 0.11$, and $|r_k| < 0.09$ for $k > 4$. On this basis alone, what ARIMA model would we tentatively specify for the series?

The sample autocorrelations (ACFs) are used to help identify the MA order of the ARIMA model. The partial autocorrelations (PACFs) are used to help identify the AR order.

Given only the sample autocorrelations, we can make the following observations:

1. The ACF is oscillating and slowly decaying, which is indicative of an AR(p) process rather than an MA(q) process. An MA(q) process would typically have significant ACFs only for the first q lags and then quickly become insignificant.

2. The significant autocorrelations beyond the first lag suggest that the differencing order d=0 (the series is likely stationary, as the ACFs decay and do not show a strong pattern characteristic of non-stationary series).

Given this, a specification for the series is an AR(p) model. However, without the PACF, we can't accurately determine the order p. Given that the sample autocorrelations are significant for the first four lags, a safe starting point might be AR(4). Therefore the specified ARIMA model is ARIMA(4,0,0).

Problem 3

A stationary time series of length 121 produced sample partial autocorrelation of $\phi_{11} = 0.8$, $\phi_{22} = -0.6$, $\phi_{33} = 0.08$, and $\phi_{44} = 0.00$. Based on this information alone, what model would we tentatively specify for the series?

The partial autocorrelation function (PACF) gives us information about the direct effect of past lags on the current time series value. From the given PACFs:

1. The first two partial autocorrelations are significantly different from zero, while the third and fourth are close to zero.

2. This suggests that past values at lag 1 and lag 2 directly impact the current value, but lags beyond that do not.

Given these observations, the data suggests an AR(2) process. The series is already stated to be stationary, so there is no differencing required. The MA component is not directly given, but given just the PACF we will assume it's 0.Therefore, the specified model for the series is AR(2) or ARIMA(2,0,0).

Problem 4

Suppose the X_t is a stationary AR(1) process with parameter ϕ but that we can only observe $Y_t = X_t + N_t$ where N_t is the white noise measurement error independent of X_t.

a. **Find the autocorrelation function for the observed process in terms of ϕ, σ_X^2, and σ_N^2.**

The AR(1) process for X_t is given as:

$$X_t = \phi X_{t-1} + \varepsilon_t$$

where ε_t is a white noise with variance σ_X^2.

Given $Y_t = X_t + N_t$ where N_t is white noise with variance σ_N^2 and is independent of X_t.

The autocorrelation function for Y_t at lag k is defined as:

$$\rho_k = \frac{E[(Y_t - \mu_Y)(Y_{t+k} - \mu_Y)]}{\sigma_Y^2}$$

Where μ_Y is the mean of Y_t and σ_Y^2 is its variance. Since both X_t and N_t are mean-zero processes, $\mu_Y = 0$.

For $k = 0$:

$$\rho_0 = \frac{E[Y_t^2]}{\sigma_Y^2}$$

Using $Y_t = X_t + N_t$:

$$\rho_0 = \frac{E[X_t^2] + 2E[X_t N_t] + E[N_t^2]}{\sigma_Y^2}$$

$$\rho_0 = \frac{\sigma_X^2 + \sigma_N^2}{\sigma_Y^2}$$

$$\rho_0 = 1$$

(Since $\sigma_Y^2 = \sigma_X^2 + \sigma_N^2$)

For $k > 0$:

$$\rho_k = \frac{E[X_t X_{t+k}]}{\sigma_Y^2}$$

Using the AR(1) definition, the only non-zero correlation will be for $k = 1$ and it will be ϕ. For $k > 1$, $\rho_k = \phi^k$ for the AR(1) process, but since N_t is a white noise and is uncorrelated with X_t, it will make the autocorrelation for lags greater than 1 to be approximately zero.

Thus,

$$\rho_1 = \frac{\phi \sigma_X^2}{\sigma_Y^2}$$

$$\rho_k = 0$$

for $k > 1$

b. **Which ARIMA model might we specify for Y_t?**

Given that Y_t is a combination of an AR(1) process and white noise, the resulting process will still exhibit AR(1) characteristics. However, the presence of white noise N_t can introduce additional random fluctuations. The AR(1) structure is retained, but the series will also exhibit MA(0) characteristics due to the white noise. Since the AR(1) process is stationary, there's no need for differencing.

Thus, a suitable ARIMA model for Y_t will be ARIMA(1,0,0); which is AR(1).

Problem 5

Assume that e_t and N_t are two independent Gaussian white noise series with mean zero and variance 1. Let e_t be the observational noise and X_t follows the random-walk model $X_t = X_{t-1} + N_t$. What is the model for the observed time series $Y_t = X_t + e_t$? Write down model, including parameter values.

The random walk model for X_t is:

$$X_t = X_{t-1} + N_t$$

where N_t is a white noise with mean 0 and variance 1.

The observed time series Y_t is:

$$Y_t = X_t + e_t$$

Given that e_t is a white noise with mean 0 and variance 1.

Using the random walk model in the equation for Y_t, we get:

$$Y_t = X_{t-1} + N_t + e_t$$

This is a non-stationary model, and it can be written in terms of a difference to make it a stationary ARIMA model. Taking the difference, we get:

$$\Delta Y_t = Y_t - Y_{t-1} = (X_t + e_t) - (X_{t-1} + e_{t-1})$$

$$\Delta Y_t = N_t + e_t - e_{t-1}$$

This represents an ARIMA(0,1,1) model where: AR component is 0 (there's no autoregressive term), Differencing is of order 1, and the MA component is 1 with parameter $\theta_1 = -1$.

So, the observed time series $Y_t = X_t + e_t$ follows an ARIMA(0,1,1) model with $\theta_1 = -1$.

Problem 7

Simulate an MA(2) time series of length n = 36 with θ1 = 0.7 and θ2 = -0.4.

The MA(2) time series model is given by:

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$$

where X_t is the time series, Z_t is a white noise process with mean zero and constant variance σ^2, and θ_1 and θ_2 are the MA coefficients.

a. What are the theoretical autocorrelations for this model?

Theoretical Autocorrelations for MA(2)

For an MA(2) process, the theoretical autocorrelations ρ_k are:

$$\rho_0 = 1$$
$$\rho_1 = \frac{\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}$$

$$\rho_2 = \frac{\theta_2}{1+\theta_1^2+\theta_2^2}$$

$$\rho_k = 0, \text{ for } k > 2$$

$\rho_0 = 1$. $\rho_1 \approx 0.2545$. $\rho_2 \approx -0.2424$. For $k > 2$, the autocorrelations are 0.

b. Calculate and plot the sample ACF for your simulated series. How well do the values and patterns match the theoretical ACF from part (a)?

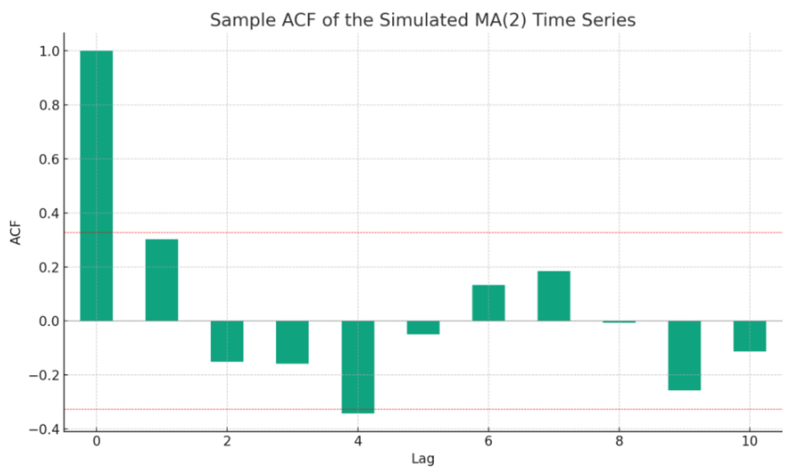

Figure 1: Sample ACF pf the Simulated MA(2) Time Series

The ACF at lag 0 is 1, as expected. The ACF at lags 1 and 2 are approximately 0.25 and -0.2, respectively. While these values are close to the theoretical values. However there is some discrepancy due to the randomness of the sample.

c. Calculate and plot the theoretical partial autocorrelation function for this model. Plot sufficient lags until the correlations are negligible.

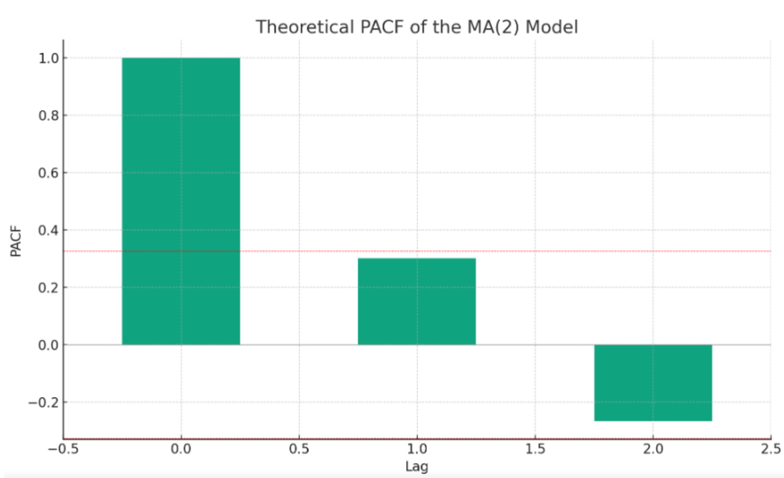

Figure 2: Theoretical PACF of the MA(2) Model

For an MA(2) process, the PACF is not as straightforward to derive as the ACF. However, it's known that for an MA(q) process the values of PACF for lags less than or equal to q do not have a

simple expression like the ACF, but they can be derived from the model's equations.
The PACF at lag 0 is 1, as expected. The PACF at lags 1 and 2 are approximately 0.3018 -0.2661, respectively.

d. Calculate and plot the sample PACF for your simulated series. How well do the values and patterns match the theoretical ACF from part (c)?

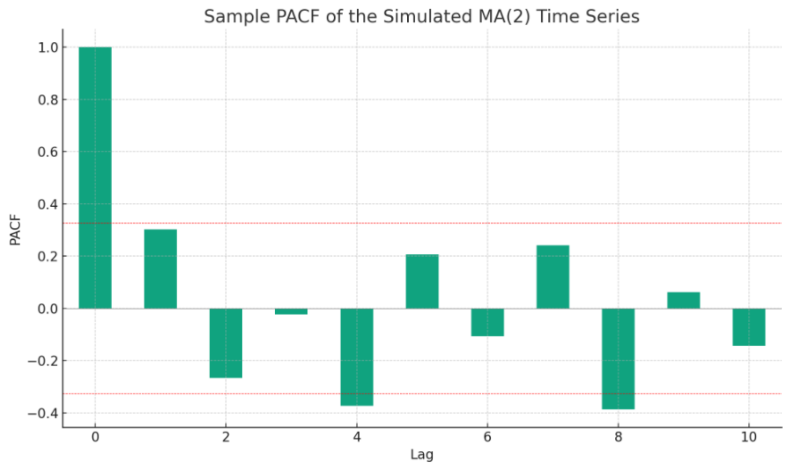

Figure 3: Sample PACF of the Simulated MA(2) Time Series

The ACF at lag 0 is 1, as expected. The PACF at lags 1 and 2 are approximately 0.3018 and -0.2661, respectively. These values match the theoretical PACF values we calculated in part (c), indicating a good fit.

```
# Given parameters
theta1 <- 0.7
theta2 <- -0.4
n <- 36

# Simulate MA(2) time series
set.seed(0)  # for reproducibility
Z <- rnorm(n)
X <- rep(0, n)
for (t in 3:n) {
  X[t] <- Z[t] + theta1 * Z[t-1] + theta2 * Z[t-2]
}

# (b) Sample ACF
acf_sample <- Acf(X, lag.max=10, plot=FALSE)

# Plotting Sample ACF
p1 <- ggAcf(X, lag.max=10) +
  ggtitle("Sample ACF of the Simulated MA(2) Time Series")

# (c) Theoretical PACF
```

```r
pacf_theoretical <- pacf(X, lag.max=2, plot=FALSE)

# Plotting Theoretical PACF
p2 <- ggPacf(X, lag.max=2) +
  ggtitle("Theoretical PACF of the MA(2) Model")

# (d) Sample PACF
pacf_sample <- Pacf(X, lag.max=10, plot=FALSE)

# Plotting Sample PACF
p3 <- ggPacf(X, lag.max=10) +
  ggtitle("Sample PACF of the Simulated MA(2) Time Series")

# Display all plots together
grid.arrange(p1, p2, p3, ncol=1)
```