Advanced

Models

HW2

Part 1

a)

a. For every combination of AR and MA orders $p, q \in \{1, 2, 3\}$ excluding $ARMA\,(2, 2)$ model, fit the $ARMA\,(p, q)$ model to the data. Report the AICc values corresponding to each fitted model and identify the model that best fits the time series using the AICc criterion. In answering this question, you should fit the various models to the full data set (do not split it into a training/test split) and assume that $\alpha = 0$.

AIC_c results

```
[1] "ARMA(p= 1 ,q= 2 ), with AICc =, 3.45689021512635"
[1] "ARMA(p= 1 ,q= 3 ), with AICc =, 3.38685237768826"
[1] "ARMA(p= 2 ,q= 1 ), with AICc =, 3.11260193696493"
[1] "ARMA(p= 2 ,q= 3 ), with AICc =, 3.02069324822333"
[1] "ARMA(p= 3 ,q= 1 ), with AICc =, 2.9641543985109"
[1] "ARMA(p= 3 ,q= 2 ), with AICc =, 2.96949751178786"
```

According to the AIC_c criteria, we pick the model with the lowest AIC_c value. Hence, the best model is

$ARMA(3, 1)$

b)

b. For the particular $ARMA(p, q)$ model you selected in Part a, provide the equation for the fitted model (i.e. the model for x_t with the estimated values of θ's and ϕ's plugged in, including the estimated value of σ_w^2).

Using R sarima package we get

```
all:
rima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
    xreg = xmean, include.mean = FALSE, transform.pars = trans, fixed = fixed,
    optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:
         ar1     ar2      ar3     ma1
      0.7996  0.5480  -0.7759  0.5535
s.e.  0.1142  0.1638   0.0832  0.1940

sigma^2 estimated as 1.028:  log likelihood = -217.17,  aic = 444.35

$degrees_of_freedom
[1] 146

$ttable
    Estimate     SE t.value p.value
ar1   0.7996 0.1142  7.0025  0.0000
ar2   0.5480 0.1638  3.3454  0.0010
ar3  -0.7759 0.0832 -9.3226  0.0000
ma1   0.5535 0.1940  2.8537  0.0049

$AIC
[1] 2.962315

$AICc
[1] 2.964154

$BIC
[1] 3.06267
```

Hence:

$$x_t = 0.7996\,x_{t-1} + 0.5480\,x_{t-2} - 0.7759\,w_{t-1} + 0.5535\,w_{t-2}$$

and estimated value of σ_w^2,

$$\sigma_w^2 = 1.028$$

c)

c. Separate the new time series into training and testing sets, where the first 150 observations belong to the training set and the remaining are the test set. Consider the $ARMA(2, 2)$ model. Fit this model to the training set and report the corresponding AICc value.

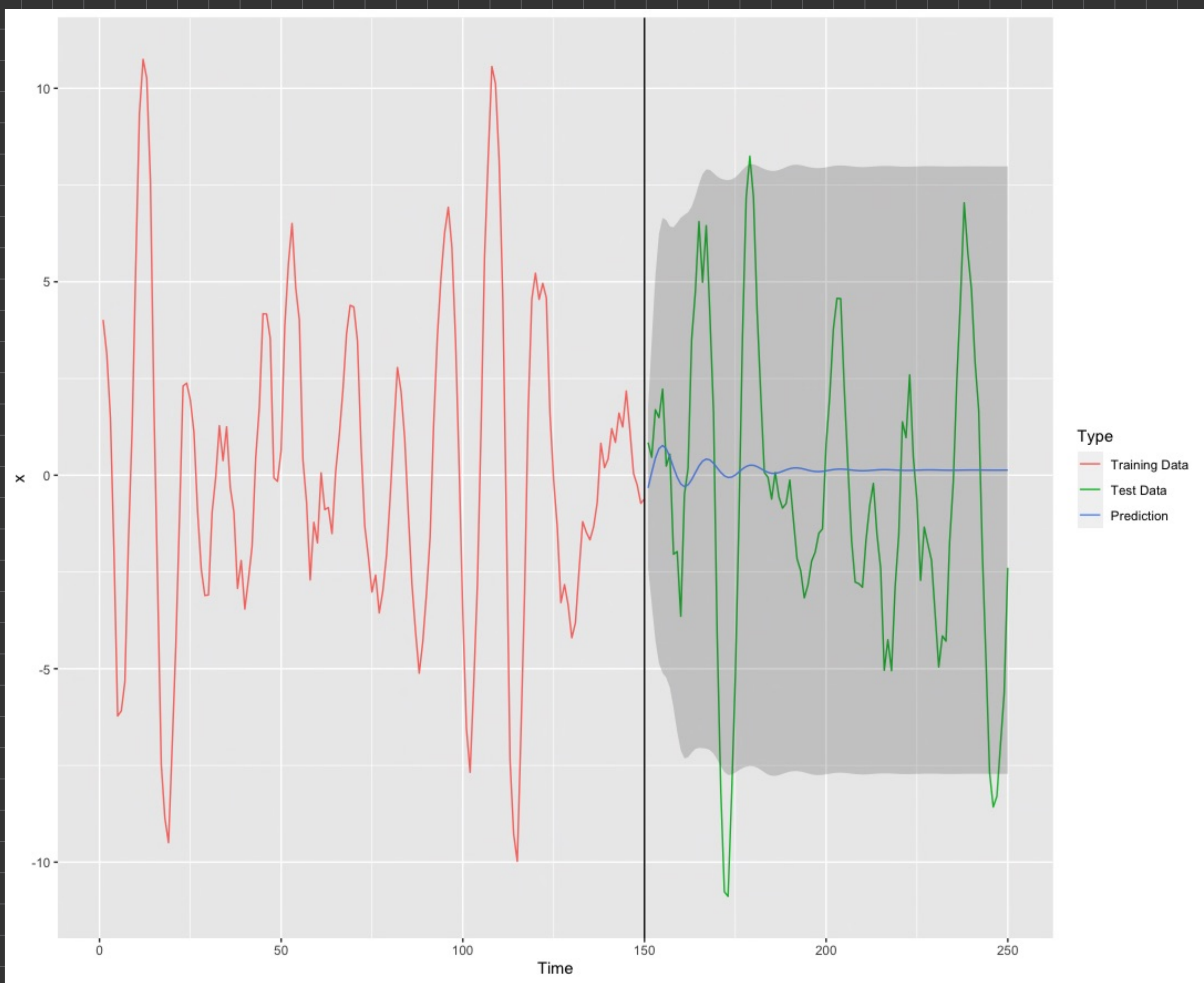
Using R SARIMA package, we obtain the $AIC_c = 2.964154$ on the training data.

```
$AICc
[1] 2.964154
```

d)

d. Using the model fitted to the training set in Part c, generate predictions for the values of the time series in the test set. Plot the full time series (training and test sets) and superimpose the predictions of the test set portion of the time series, along with 95% predictive interval bands. How well do the predictions match the test set? How does the quality of the predictions change over time? (Hint: Consider using the *geom_ribbon* function within *ggplot2* to create predictive interval bands)

Result Plot →



The prediction is not very accurate, as we can notice from the plot. There is a significant difference between the green line (test data) and the blue line (prediction). However, the prediction trend is correct for the first oscillations of the test data; and the quality of the trend prediction gets worse over time. As time passes, the prediction tends to zero, while the test data continue oscillating.

APPENDIX . CODE

```r
#-------------------------------question a--------------------------------------

data1 <- read.csv("/Users/rafa/Documents/Master Austin/MAESTRÍA_AUSTIN/Advanced Predictive Models/HW2/file1.xls", header=TRUE, stringsAsFactors=FALSE)

data1



d<- 0.0

for (p_ar in 1:3)
{
  for (q_ar in 1:3)
  {
    if (p_ar!=q_ar){

      # Fit model
      fit <- sarima(data1$x,
                    p = p_ar,
                    d = d,
                    q = q_ar,
                    no.constant = TRUE,
                    details = FALSE)

      # Examine estimated model parameters

      print(paste("ARMA(p=",p_ar,",q=", q_ar, "), with AICc =,", fit$AICc))
    }
  }

}

#-------------------------------question b--------------------------------------


# Selected ARMA model p=3, q=1

fit_best <- sarima(data1$x,
             p = 3,
             d = 0,
             q = 1,
             no.constant = TRUE,
             details = FALSE)
fit_best
```

```r
#-------------------------------question c--------------------------------------
x <- read.csv("/Users/rafa/Documents/Master Austin/MAESTRÍA_AUSTIN/Advanced Predictive Models/HW2/file1_2.xls", header=TRUE, stringsAsFactors=FALSE)
n<-250
n_train <-150
x_train<-x[1:150,]
data_test<-x[151:250,]

fit_train<- sarima(x_train$x,
                   p = 2,
                   d = 0,
                   q = 2,
                   no.constant = TRUE,
                   details = FALSE)
fit_best


fit_for <- sarima.for(x_train$x,
                      n.ahead = 100,
                      p = 2,
                      d = 0,
                      q = 2,
                      plot = F)
```

question d

```r
# Collect time series for plotting
fit_data <- bind_rows(
  data.frame(Time = 1:150,
             Type = factor(rep("Training Data", 150)),
             x = c(as.numeric(x_train$x))),
  data.frame(Time = 151:250,
             Type = factor(rep("Test Data", 100)),
             x = c( as.numeric(data_test$x))),
  data.frame(Time = 151:250,
             Type = factor(rep("Prediction", 100)),
             x = c(as.numeric(fit_for$pred ))))

fit_pred_data <- data.frame(Time = 151:n,
                            x = c(as.numeric(fit_for$pred)),
                            SE = c( as.numeric(fit_for$se)))

# Plot data and forecasts
gg_fit <- ggplot(fit_data,
                 aes(x = Time)) +
  geom_line(aes(y = x, col = Type)) +
  geom_ribbon(data = fit_pred_data,
              aes(x = Time,
                  ymin = x - 1.96*SE,
                  ymax = x + 1.96*SE),
              alpha = .2) +
  geom_vline(xintercept = 150)
```