Stat 443: Time Series and Forecasting
Lab 7

Caden Hewlett

March 01, 2024

```
df = read.csv("TempPG.csv")
summer = df$Summer

summerts = ts(summer, start = 1918,
              end = 2008, freq = 1)
```

Part 1

Fit the AR(2) model to the summer minimum temperatures using the `arima()` command, and write down your fitted model.

```
fitted = arima(summerts, order = c(2, 0, 0))
fitted
```

```
##
## Call:
## arima(x = summerts, order = c(2, 0, 0))
##
## Coefficients:
##          ar1     ar2  intercept
##       0.4026  0.3446     7.0669
## s.e.  0.1029  0.1041     0.3263
##
## sigma^2 estimated as 0.6672:  log likelihood = -111.07,  aic = 230.15
```

Write down your fitted model.

Our original model (in pure theory) would be:

$$(X_t - \mu) = \alpha_1(X_{t-1} - \mu) + \alpha_2(X_{t-2} - \mu) + Z_t$$

Where $Z_t \sim \text{WN}(0, \sigma^2)$.

We now would have:

$$(X_t - \hat{\mu}) = \hat{\alpha}_1(X_{t-1} - \mu) + \hat{\alpha}_2(X_{t-2} - \mu) + Z_t, \text{ where } Z_t \sim \text{WN}(0, \hat{\sigma}^2)$$

From the results, we see $\hat{\alpha}_1 = 0.4026$, $\hat{\alpha}_2 = 0.3446$, $\hat{\mu} = 7.0669$ and $\hat{\sigma}^2 = 0.6672$. In other words,
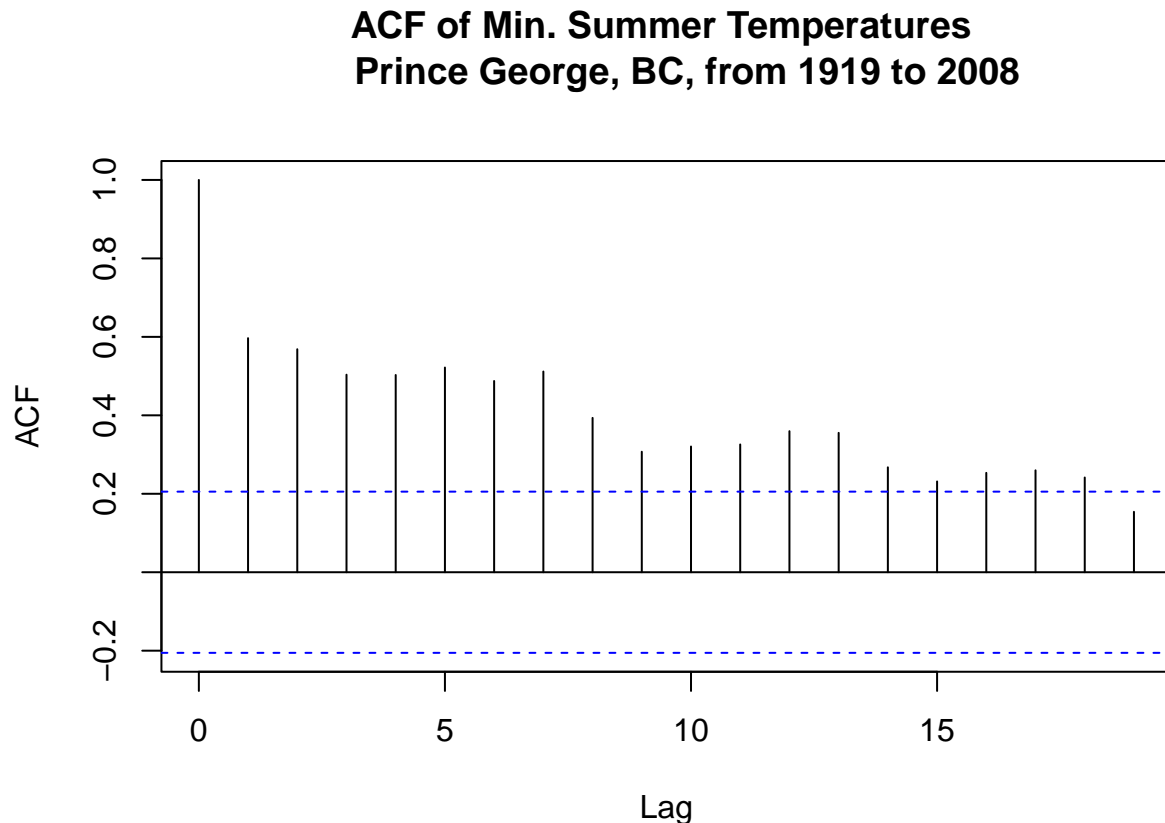
$$(X_t - 7.0669) = 0.4026(X_{t-1} - 7.0669) + 0.3446(X_{t-2} - 7.0669) + Z_t$$

Where $Z_t \sim \text{WN}(0, 0.6672)$.

Part 2

Look again at the sample acf of the summer minimum temperatures. In what way does the sample acf not behave as you would expect for the fitted AR(2) model?

```
acf(summerts, main = "ACF of Min. Summer Temperatures
    Prince George, BC, from 1919 to 2008")
```


Comments There doesn't appear to be the consistent exponential decay that we would expect for the fitted AR(2) model, given that both $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are greater than zero. We can see around $h \in [10, 15] \subset \mathbb{Z}$ that there is a periodic pattern (i.e. a rise and fall,) followed by another rise and fall for $h \in [15, 20] \subset \mathbb{Z}$. This slightly betrays the "gradual decay" we would expect for this process.

Part 3

Plot the series of first differences of minimum summer temperatures. Plot the sample acf and pacf of the differences. What model would you suggest for the differences?

We define $\{Y_t\}_{t \in \mathbb{Z}}$ to be the differenced minimum summer temperature series, where, $\forall t \in \mathbb{Z}, Y_t = X_t - X_{t-1}$

We create the plot for $\{Y_t\}_{t \in \mathbb{Z}}$ below:

```
diff_summ = diff(summerts)
plot(diff_summ, ylab = "Differenced Minimum Summer Temperatures (C)",
     xlab = "Year", main = "Differenced Minimum Temperatures (C)
     Mmeasured at Prince George, BC, from 1919 to 2008")
```

**Differenced Minimum Temperatures (C)
Mmeasured at Prince George, BC, from 1919 to 2008**
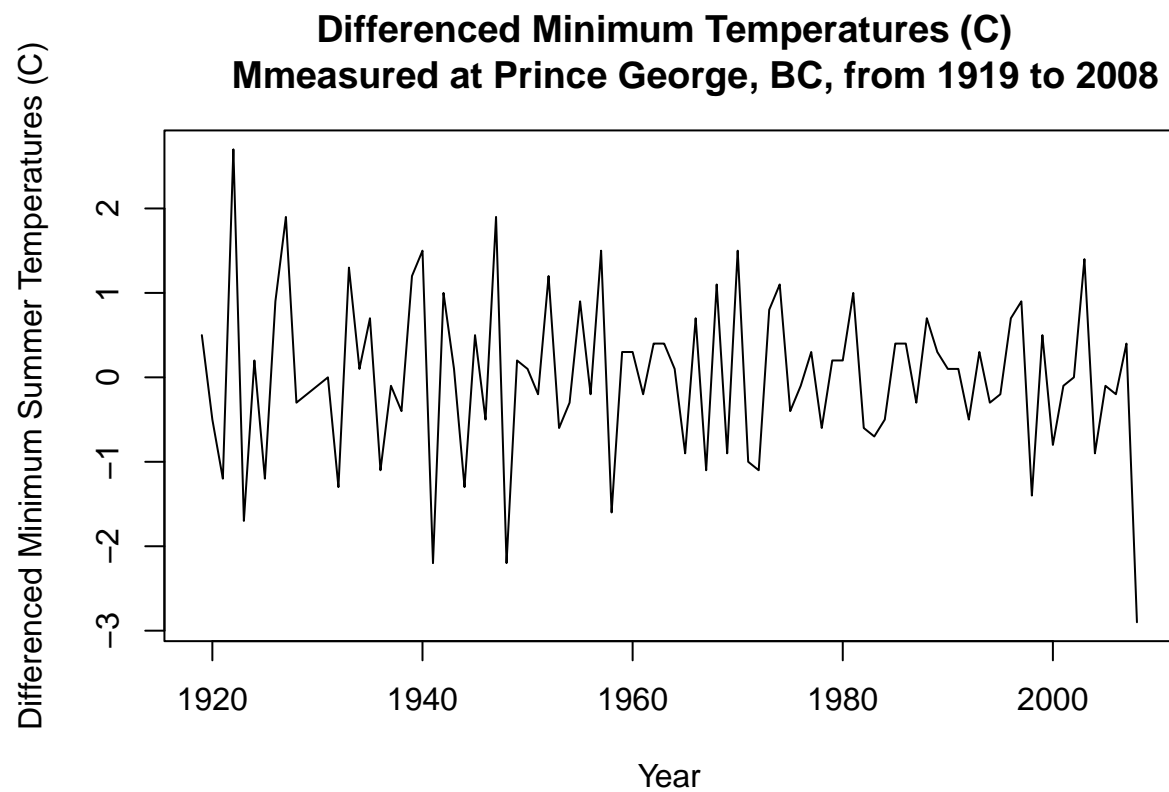
Then, we consider the ACF of the differenced data:

```r
acf(diff_summ, main = "ACF of Differenced Min. Summer Temperatures
    Prince George, BC, from 1919 to 2008")
```

ACF of Differenced Min. Summer Temperatures
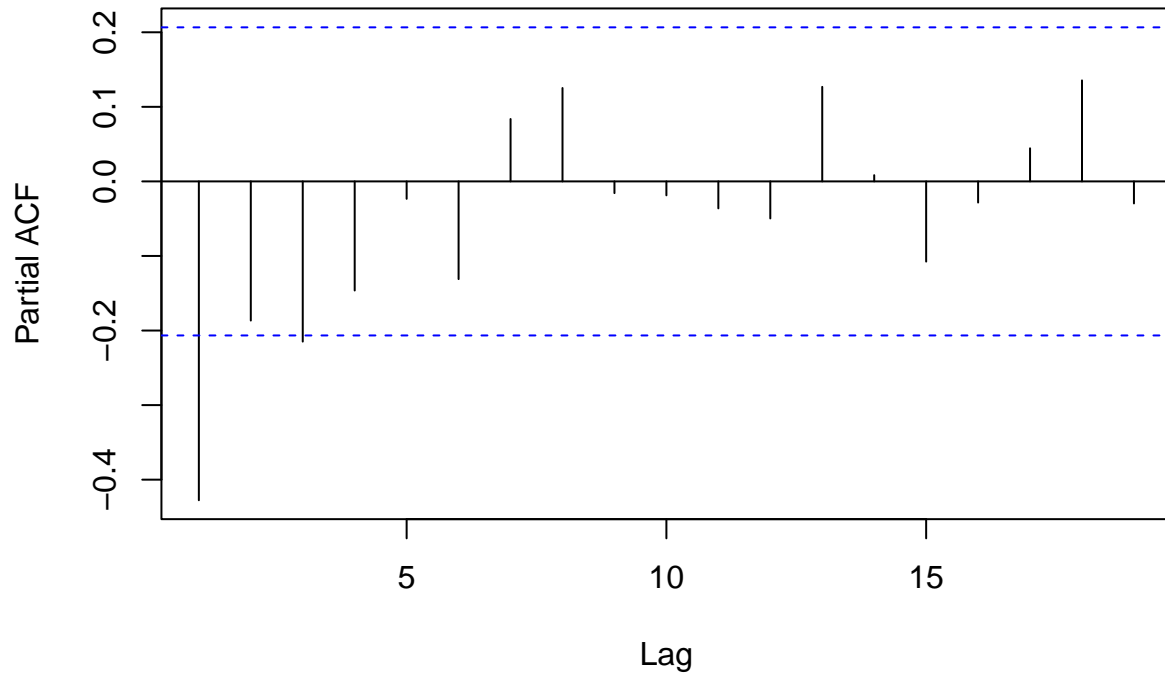## Prince George, BC, from 1919 to 2008
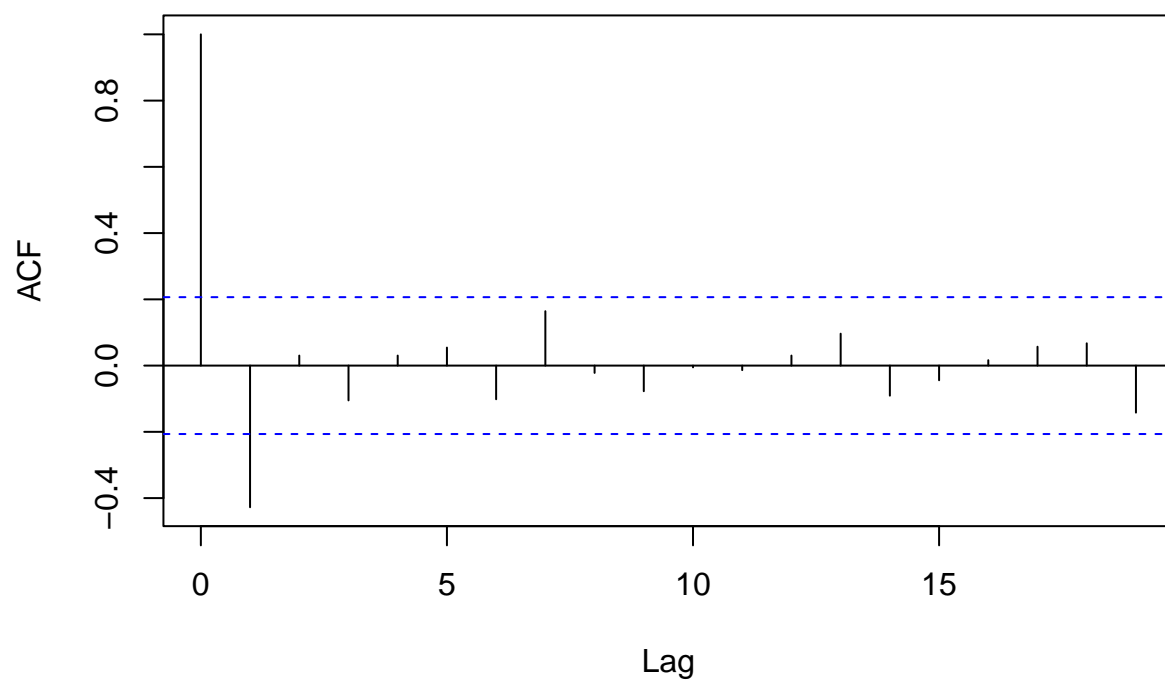

```r
pacf(diff_summ, main = "PACF of Differenced Min. Summer Temperatures
    Prince George, BC, from 1919 to 2008")
```

**PACF of Differenced Min. Summer Temperatures
Prince George, BC, from 1919 to 2008**

Simply from diagnostics, this ACF seems indicative of an MA(1) model, where $\beta_1 < 0$. We have some intuition towards this conclusion due to the fact that there is an extreme ACF value at lag $h = 1$, and it tails off sharply within the white noise bounds $\forall h \in \mathbb{Z}$ s.t. $h \geq 2$.

Part 4

In terms of ARIMA, I would fit a model of the following form (assuming $\mu = 0$):

$$\mathrm{ARIMA}(\underset{p}{0}, \underset{d}{1}, \underset{q}{1}) : \varphi(B)W_t = \theta(B)Z_t, \text{ where } W_t = \nabla^1 X_t, \theta(B) = (1 + \beta B), \text{ s.t. } \beta \in \mathbb{R}$$

So, the "expanded" model that we will fit is as follows, noting that there is no AR component. We express these values in terms of our *estimates.*

$$\mathrm{ARIMA}(\underset{p}{0}, \underset{d}{1}, \underset{q}{1}) : W_t = (1 + \hat{\beta}B)Z_t \text{ where } Z_t \sim \mathrm{WN}(0, \hat{\sigma}^2) \text{ and } \hat{\beta} \in \mathbb{R}$$

Now, we numerically fit the model below:

```
new = arima(summerts, order = c(0, 1, 1))
new
```

```
##
## Call:
## arima(x = summerts, order = c(0, 1, 1))
##
## Coefficients:
```

```
##            ma1
##        -0.7833
## s.e.    0.0678
##
## sigma^2 estimated as 0.6025:   log likelihood = -105.38,   aic = 214.76
```

With these estimates, we now have the fitted model in terms of our coefficients:

$$W_t = Z_t - 0.7833Z_{t-1} \text{ where } Z_t \sim \text{WN}(0, 0.6025)$$
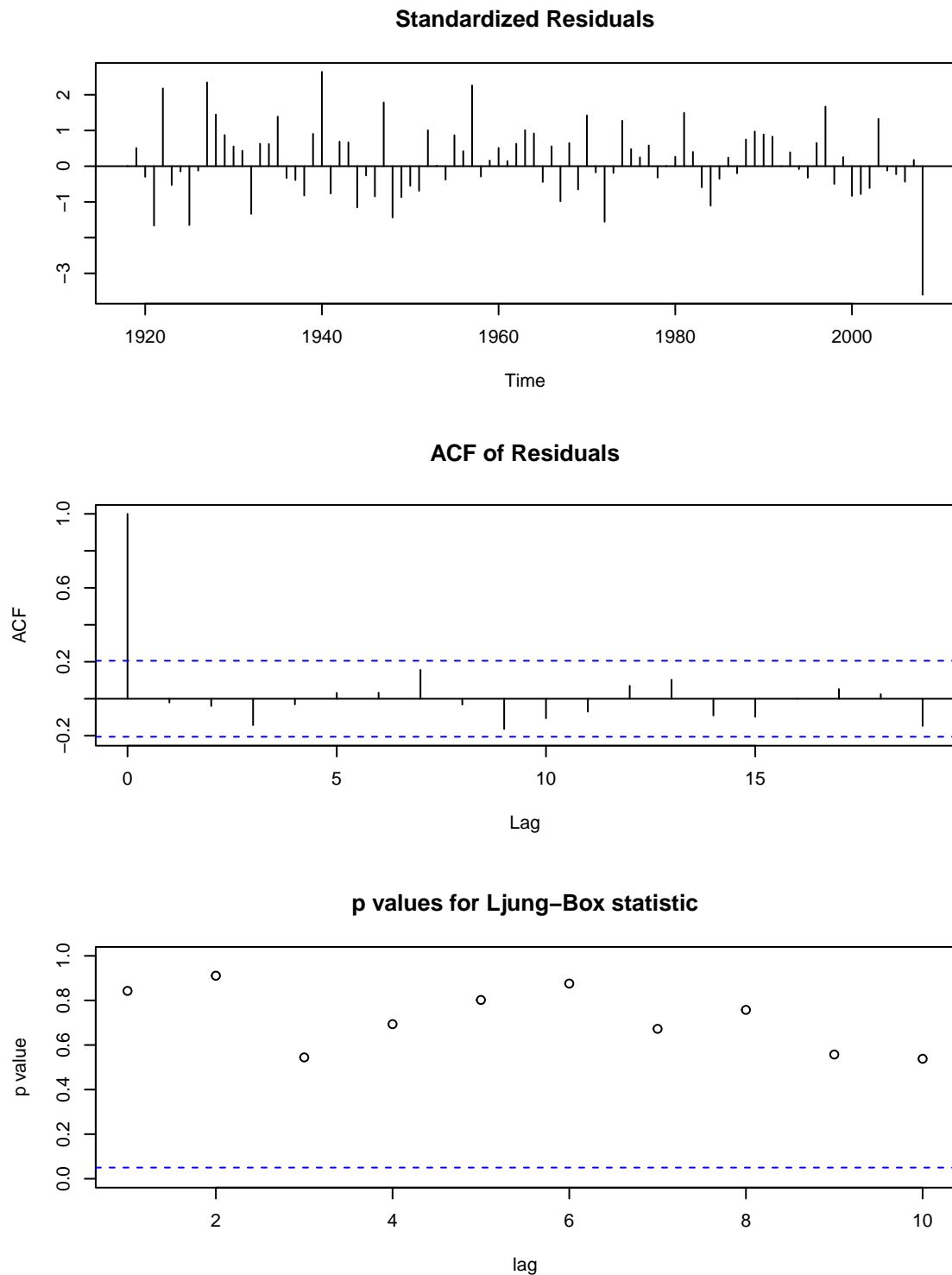
Understanding $W_t = X_t - X_{t-1}$, we could also write the model fitted as:

$$X_t = X_{t-1} + Z_t - 0.7833Z_{t-1} \text{ where } Z_t \sim \text{WN}(0, 0.6025)$$

Part 5

Use the `tsdiag()` function to see diagnostic plots for the model you have fitted. How well does the model appear to fit?

```
tsdiag(new)
```

Standardized Residuals

ACF of Residuals

p values for Ljung–Box statistic

Let's examine each of these diagnostic plots in turn:

Standardized Residuals: The first plot, containing the standardized residuals, shows roughly what we would like to see from a well-fitting model. The residual time series is centered about zero and exhibits no clear pattern. The final term in the series has a high standardized residual - this makes sense, since the differencing is more inconsistent for the tail end of the series.

Residual ACF: The second plot, containing the ACF of the residual time series. The behaviour of this ACF is even more encouraging. Beyond $\rho(0) = 1$, the ACF values are wholly within the white noise boundaries. This is indicative of a White Noise process for the residuals, which is what we would anticipate if the model is well-fitting.

Ljung-Box Statistic Plot: Finally, the observed p-values for the Ljung-Box statistic are greater than $\alpha = 0.05$ for all lags ($\in \mathbb{Z}$) up to 10 (in the plot, though theoretically we could plot further.) This shows us that there is not statistically significant evidence to suggest that the model is ill-fitting, which is also great news!

In all, from the diagnostic plots, the model seems to fit well.

Part 6

Recall the Akaike Information Criterion (AIC), defined to be proportional to the following.

Let $\mathrm{AIC}(m)$ be the Akaike Information Criterion for model $m \in M$, where M is a theoretical set of models being compared.

$$\mathrm{AIC}(m) \propto -\log\big(\mathrm{MLE}(m)\big) + 2r_m$$

where r_m is the number of independent parameters in the model.

This statistic can be used for model selection.

Models with smaller AIC values are often preferred. Compare the two competing models here via their AIC values.

Which model would you select?

We let m_1 be the original AR(2) model and m_2 be the new ARIMA(0, 1, 1) model. Hence, $M = \{m_1, m_2\}$.

The best performing model given the data, which I will call m_{best}, is hence given by:

$$m_{\mathrm{best}} = \operatorname*{argmin}_{m \in M}\big\{\mathrm{AIC}(m)\big\}, \text{ where } M = \Big\{\mathrm{AR}(\underset{p}{2}), \mathrm{ARIMA}(\underset{p}{0}, \underset{d}{1}, \underset{q}{1})\Big\}$$

In our case, the observed AIC values are:

```
c(fitted$aic, new$aic)
```

```
## [1] 230.1497 214.7597
```

So, by the equation above, we can find m_{best} given the data.

$$m_{\mathrm{best}} = \operatorname*{argmin}_{m \in M}\big\{\mathrm{AIC}(m)\big\}$$
$$m_{\mathrm{best}} = \operatorname*{argmin}_{m \in M}\Big\{\mathrm{AIC}(\mathrm{AR}(\underset{p}{2})), \mathrm{AIC}\big(\mathrm{ARIMA}(\underset{p}{0}, \underset{d}{1}, \underset{q}{1})\big)\Big\}$$
$$m_{\mathrm{best}} = \operatorname*{argmin}_{m \in M}\Big\{230.1497, 214.7597\Big\}$$
$$m_{\mathrm{best}} = \mathrm{ARIMA}(\underset{p}{0}, \underset{d}{1}, \underset{q}{1})$$

Therefore, by the AIC, the model we would select is the ARIMA model.