TimeSeriesHW2

Fjolle Gjonbalaj

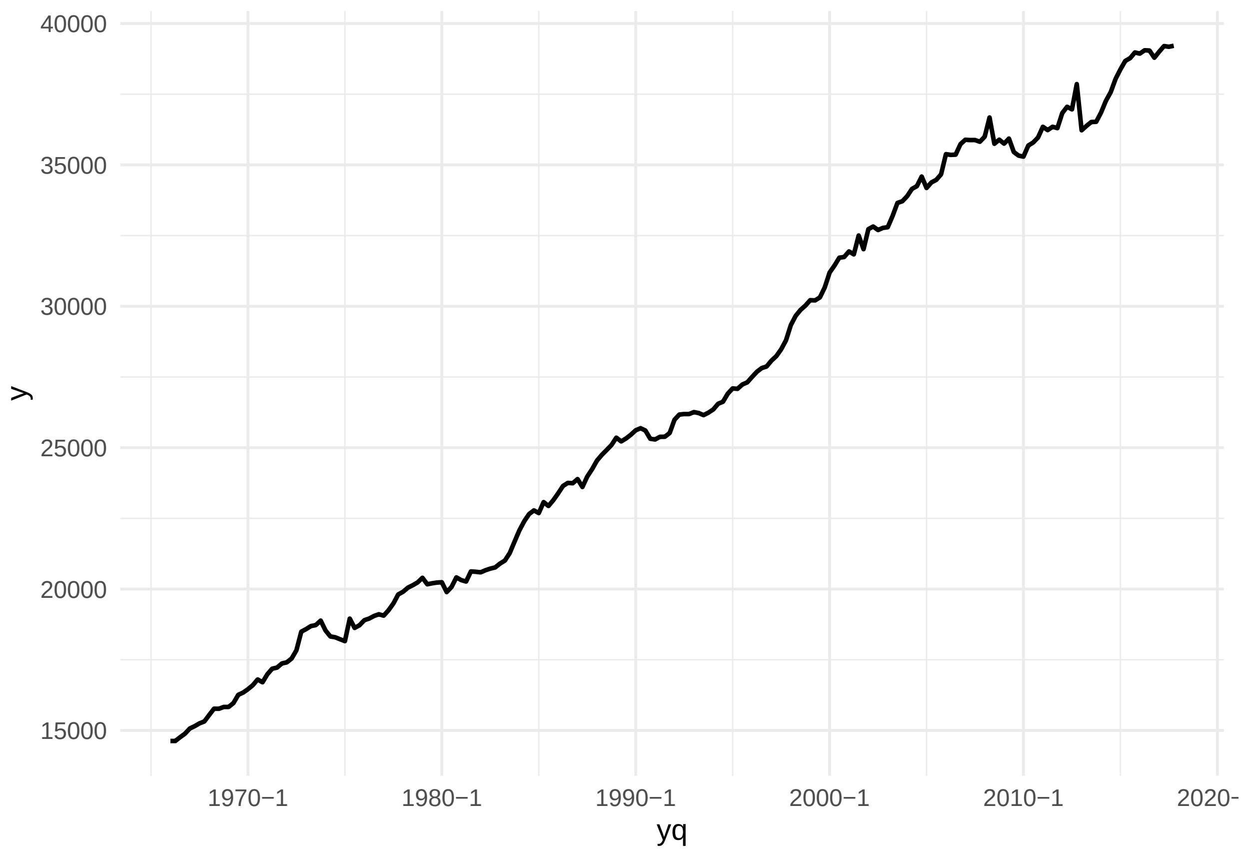
12 February 2021

Question 1

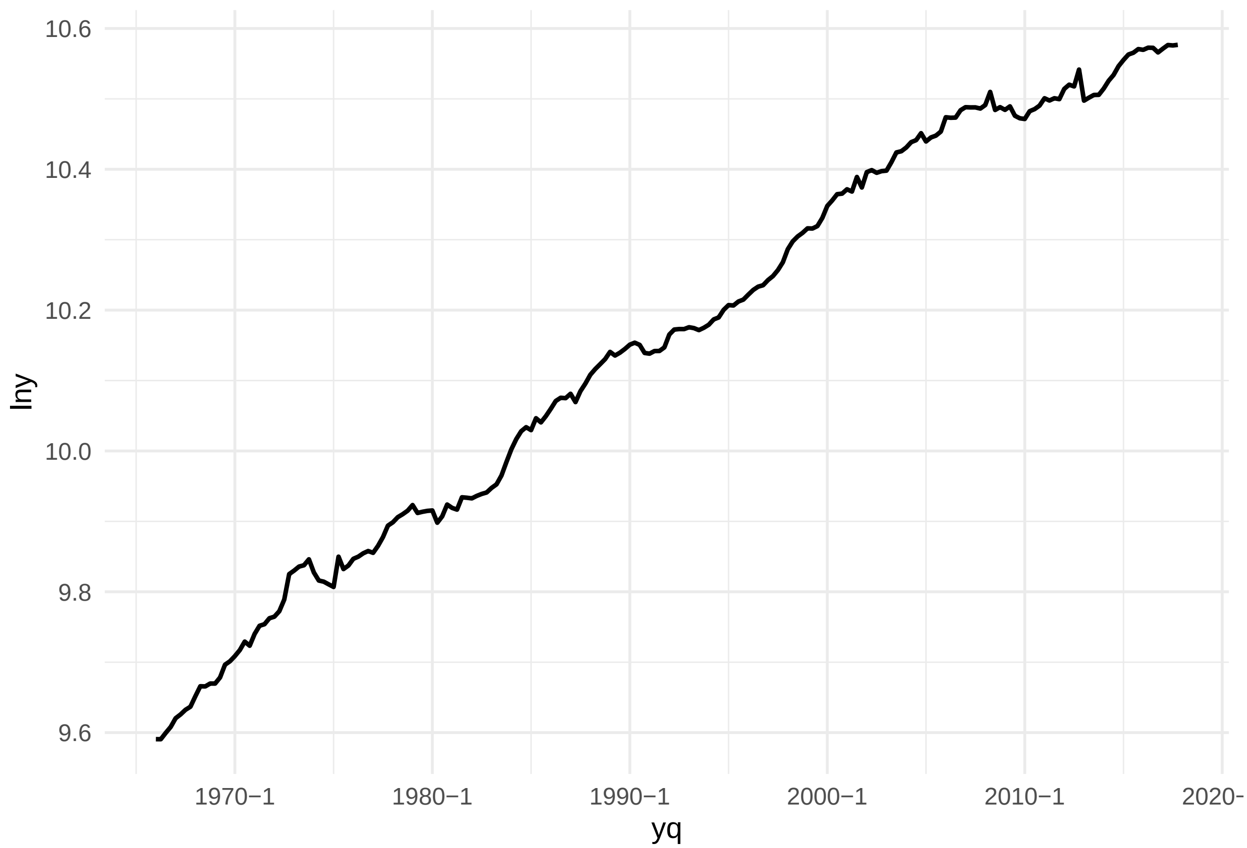
```
yq <- as.yearqtr(personal_income_series$observation_date, format = "%Y-%d-%m")

theme_set(theme_minimal())

ggplot(data = personal_income_series, aes(x = yq, y = y))+
  geom_line(color = "black", size = 0.8)
```

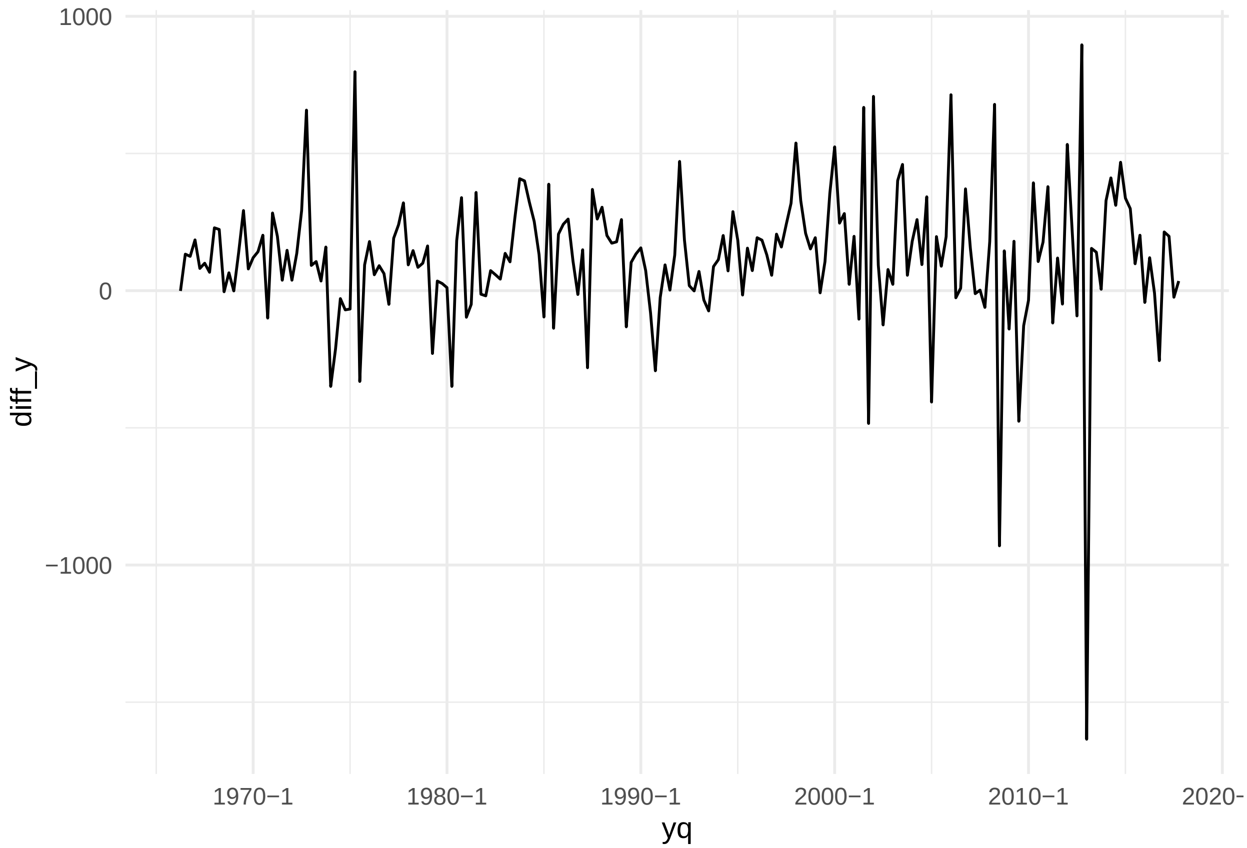



```
lny=log(personal_income_series$y)
ggplot(data = personal_income_series, aes(x = yq, y = lny))+
  geom_line(color = "black", size = 0.8)
```




```
personal_income_series %>%
  mutate(diff_y = c(NA, diff(y))) %>%
  ggplot(aes(x = yq, y = diff_y), color = "#00AFBB") +
  geom_line()
```

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```



Data show a linear trend in the series. This linear trend is not improved by taking the log of our series. On the other hand, differencting the data we remove the linear trend and the variablity in the data.

```
adf.test(personal_income_series$y)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  personal_income_series$y
## Dickey-Fuller = -2.111, Lag order = 5, p-value = 0.5295
## alternative hypothesis: stationary
```

```
dy=diff(personal_income_series$y)
adf.test(dy)
```

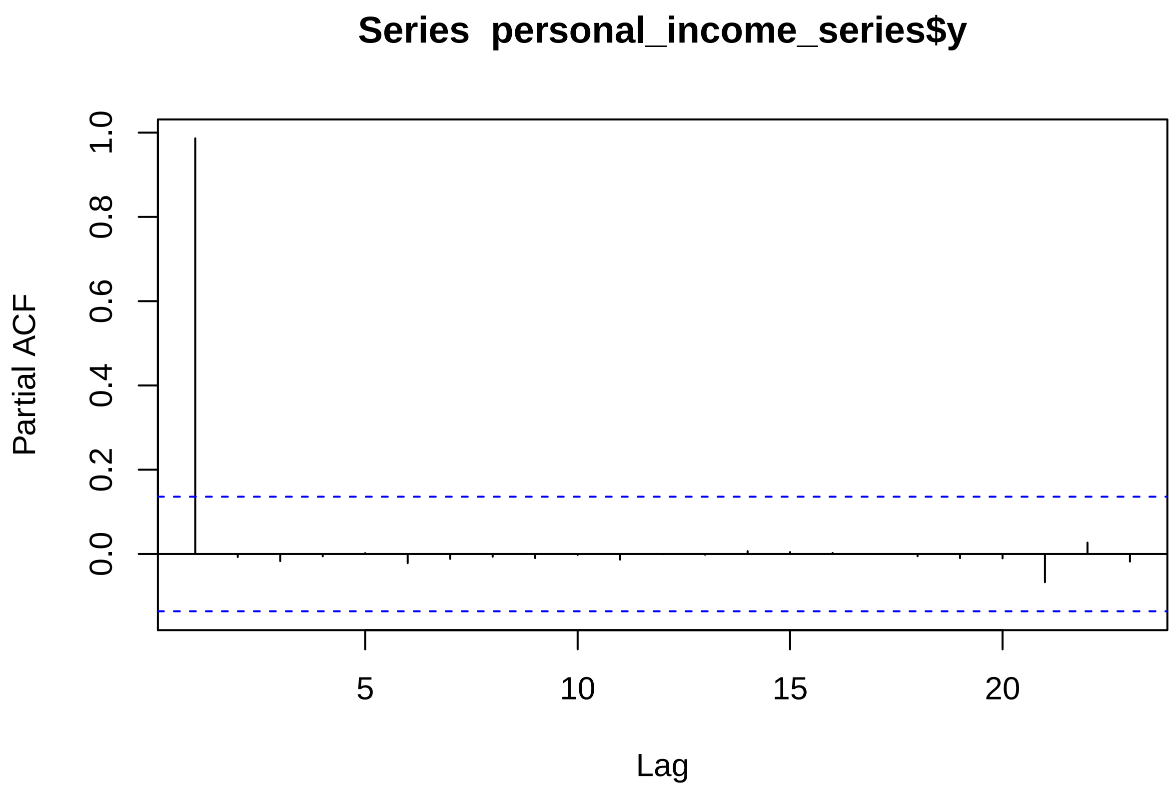
```
## Warning in adf.test(dy): p-value smaller than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  dy
## Dickey-Fuller = -6.2674, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

Performing DF and ADF tests on our initial time series data we get results such that there is no statistical evidence against our null hypothesis of unit root in our data series. We, hence, can conlude that we have non-stationarity in our data. On the other hand, performing the test on our transformed time series variable, we reject the null hypothesis of unit root. First difference of our time series results in stationarity.

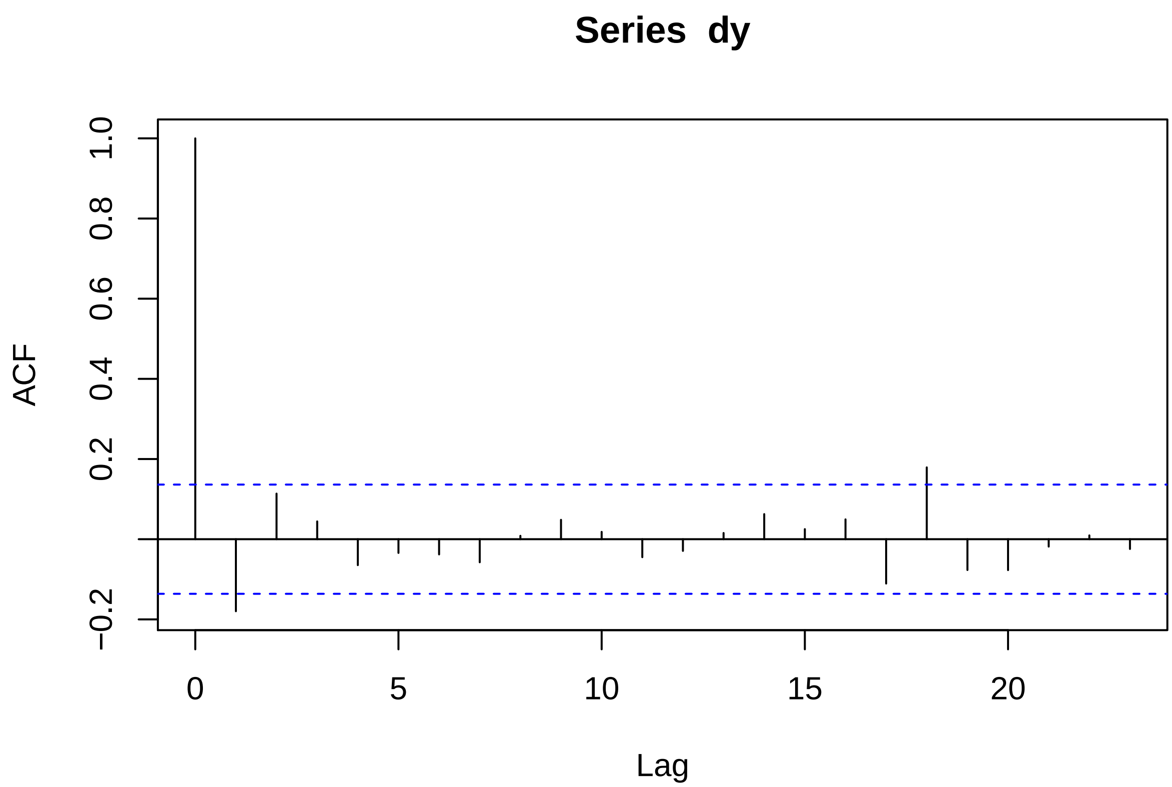
```
acf(personal_income_series$y)
```


```
pacf(personal_income_series$y)
```

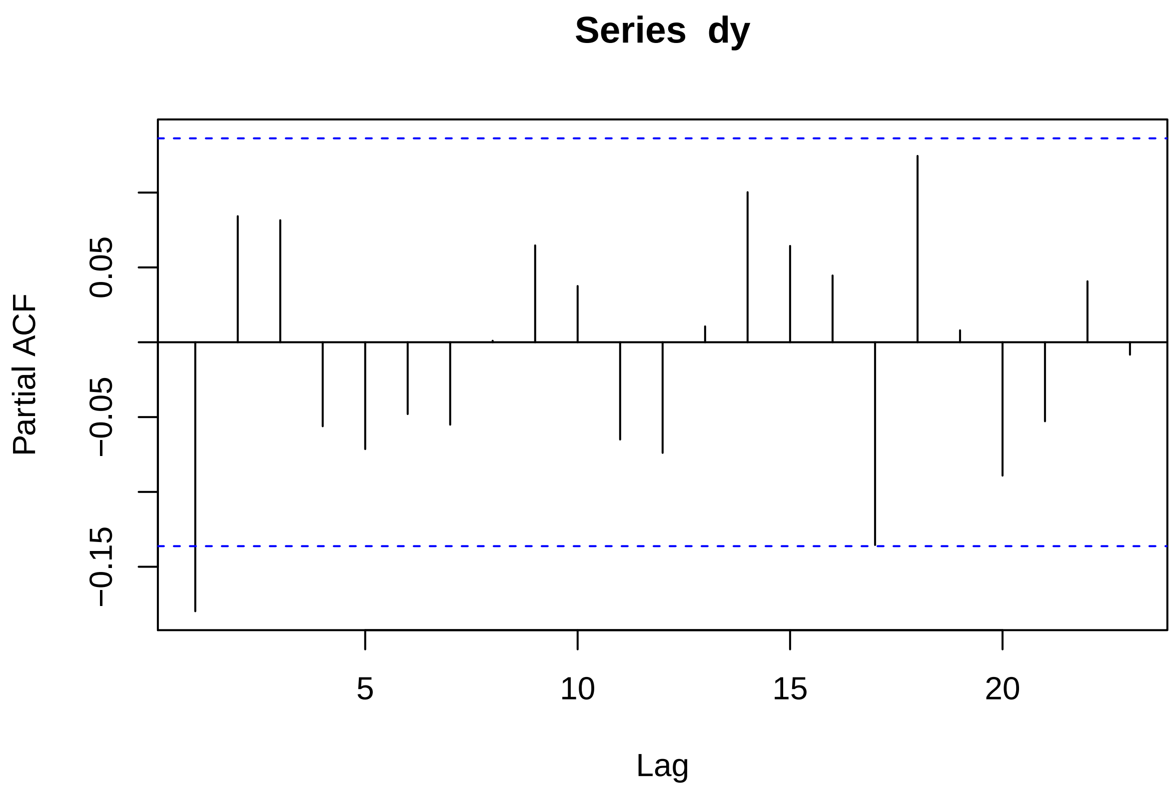



```
# transformed series

acf(dy)
```




```
pacf(dy)
```



Both the sample ACF and PACF decay relatively slowly for the original time series. This is consistent with an ARMA model. The ARMA lags cannot be selected solely by looking at the ACF and PACF, but it seems no more than four AR or MA terms are needed. However, we are interested about the ACF and PACF of the transformed time series. In this time series we no longer observe a gradual decay in the acf function. On the other hand, we do observe the PACF being significant only in its first lag. Hence, the first two models I would choose to start with are AR(1) and MA(1).

```
model1 <- arima(x=dy, order=c(1,0,1), method="ML")
model2 <- arima(x=dy, order=c(1,0,0), method="ML")
model3 <- arima(x=dy, order=c(0,0,1), method="ML")
model4 <- arima(x=dy, order=c(0,0,2), method="ML")

armaaic=forecast::auto.arima(dy, ic=c("aic"), trace=TRUE)
```

```
##
##  Fitting models using approximations to speed things up...
##
##  ARIMA(2,0,2) with non-zero mean : 2879.028
##  ARIMA(0,0,0) with non-zero mean : 2881.771
##  ARIMA(1,0,0) with non-zero mean : 2877.74
##  ARIMA(0,0,1) with non-zero mean : 2878.171
##  ARIMA(0,0,0) with zero mean     : 2921.086
##  ARIMA(2,0,0) with non-zero mean : 2879.275
##  ARIMA(1,0,1) with non-zero mean : 2879.04
##  ARIMA(2,0,1) with non-zero mean : 2880.807
##  ARIMA(1,0,0) with zero mean     : 2923.844
```

```
##
##  Now re-fitting the best model(s) without approximations...
##
##  ARIMA(1,0,0) with non-zero mean : 2876.997
##
##  Best model: ARIMA(1,0,0) with non-zero mean
```

```r
armabic=forecast::auto.arima(dy, ic=c("bic"), trace=TRUE)
```

```
##
##  Fitting models using approximations to speed things up...
##
##  ARIMA(2,0,2) with non-zero mean : 2899.024
##  ARIMA(0,0,0) with non-zero mean : 2888.437
##  ARIMA(1,0,0) with non-zero mean : 2887.738
##  ARIMA(0,0,1) with non-zero mean : 2888.17
##  ARIMA(0,0,0) with zero mean     : 2924.418
##  ARIMA(2,0,0) with non-zero mean : 2892.606
##  ARIMA(1,0,1) with non-zero mean : 2892.371
##  ARIMA(2,0,1) with non-zero mean : 2897.47
##  ARIMA(1,0,0) with zero mean     : 2930.509
##
##  Now re-fitting the best model(s) without approximations...
##
##  ARIMA(1,0,0) with non-zero mean : 2886.995
##
##  Best model: ARIMA(1,0,0) with non-zero mean
```

Model with lowest AIC is best. BIC punishes models for extra parameters. We see here that models with both lowest AIC and BIC are the AR(1) and MA(1) models. Although basing the decision on a model is not suggested to be done based on looking at the acf and pacf, in this case, I was successfully able to identity the best models based on graphs only. In general, however, more formal tests such as AIC and BIC must be performed.

```r
model2 <- arima(x=dy, order=c(1,0,0), method="ML")

Box.test(resid(model2), lag = 1, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  resid(model2)
## X-squared = 0.043609, df = 1, p-value = 0.8346
```

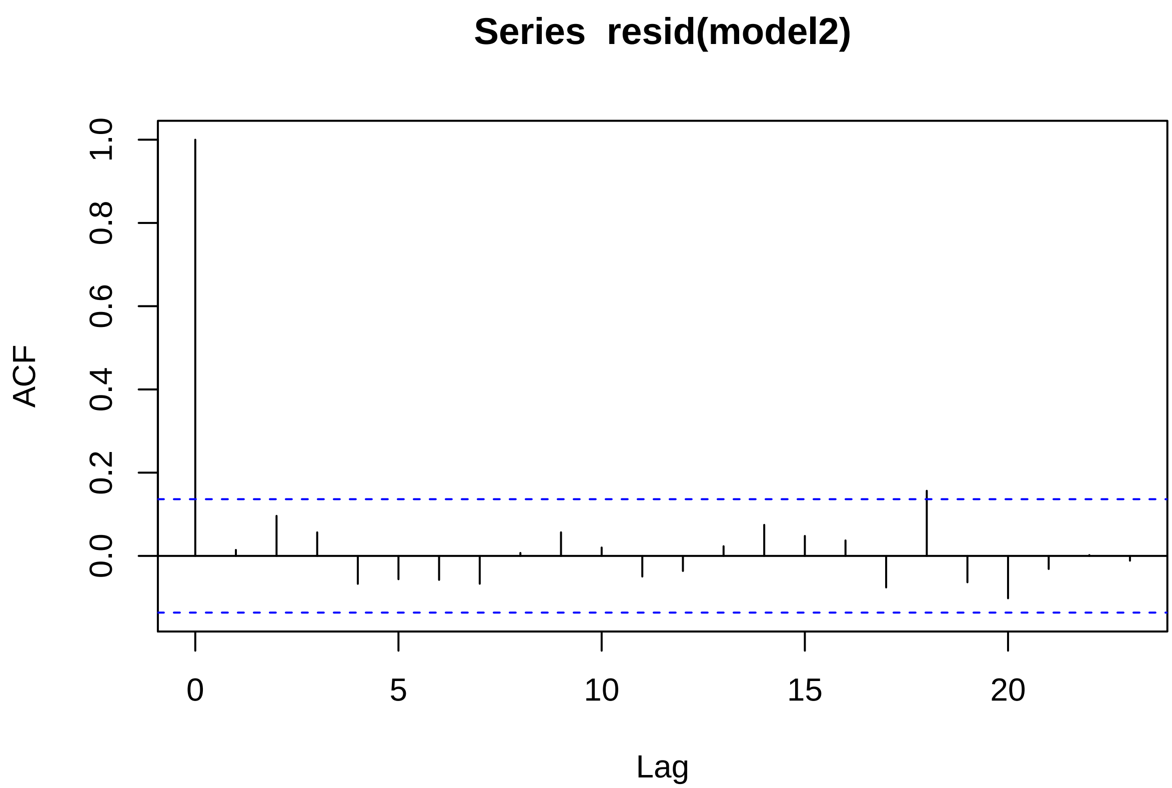
```r
model4 <- arima(x=dy, order=c(0,0,2), method="ML")

Box.test(resid(model4), lag = 1, type = "Ljung-Box")
```

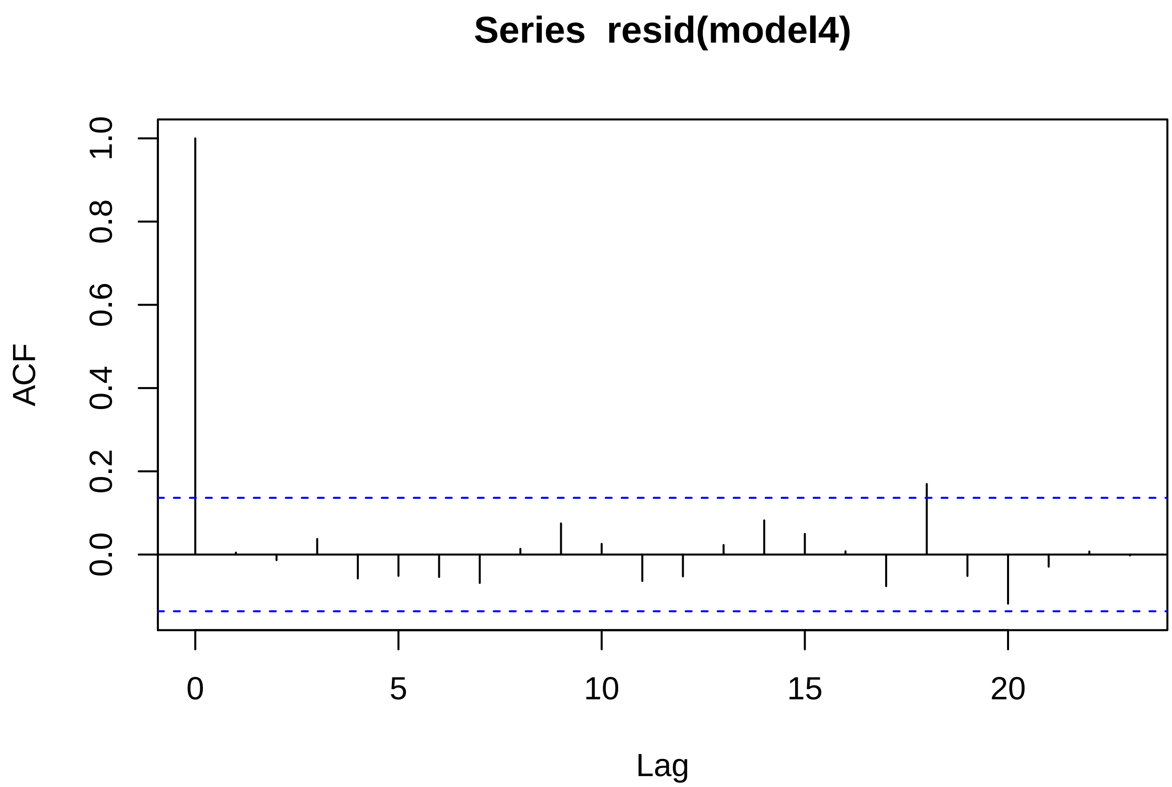
```
##
##  Box-Ljung test
##
## data:  resid(model4)
## X-squared = 0.0046121, df = 1, p-value = 0.9459
```

Looking at the Ljung-Box test we find out that the p-value is greater than 0.05, which implies that the residuals are independent at the 95% level and thus an AR(1) model provides a good model fit. Similarly, the Ljung-Box test for the MA(1) model has a p-value greater than 0.05, which means that the residuals are independent at the 95% level and thus the MA(1) model provides a good model fit.

```
acf(resid(model2))
```




```
acf(resid(model4))
```



Furthermore, looking at the autocorrelation functions of the two models we find out that both correlograms display realisations of discrete white noise.

```
ArchTest (dy, lags=1, demean = FALSE)
```

```
##
##  ARCH LM-test; Null hypothesis: no ARCH effects
##
## data:  dy
## Chi-squared = 10.802, df = 1, p-value = 0.001014
```

```
ArchTest (dy, lags=2, demean = FALSE)
```

```
##
##  ARCH LM-test; Null hypothesis: no ARCH effects
##
## data:  dy
## Chi-squared = 12.726, df = 2, p-value = 0.001724
```

```
ArchTest (dy, lags=3, demean = FALSE)
```

```
##
##  ARCH LM-test; Null hypothesis: no ARCH effects
##
## data:  dy
## Chi-squared = 12.652, df = 3, p-value = 0.005453
```

```r
ArchTest (dy, lags=4, demean = FALSE)
```

```
##
##  ARCH LM-test; Null hypothesis: no ARCH effects
##
## data:  dy
## Chi-squared = 13.281, df = 4, p-value = 0.009981
```

```r
ArchTest (dy, lags=8, demean = FALSE)
```

```
##
##  ARCH LM-test; Null hypothesis: no ARCH effects
##
## data:  dy
## Chi-squared = 13.682, df = 8, p-value = 0.09043
```

```r
ArchTest (dy, lags=10, demean = FALSE)
```

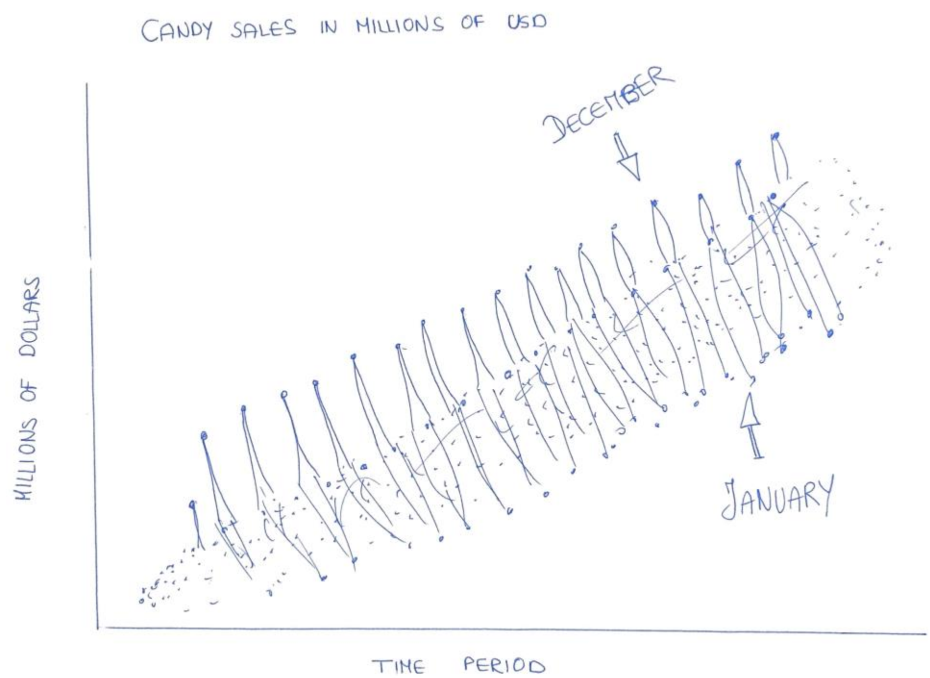
```
##
##  ARCH LM-test; Null hypothesis: no ARCH effects
##
## data:  dy
## Chi-squared = 14.237, df = 10, p-value = 0.1625
```

Running an ARCH test I find out that up to the third order, the ARCH test is significant at the significance level of 5%. This signifies that for the three first orders we reject the null hypothesis of 'no ARCH errors'. Rejecting the null thus means that such errors exist in the conditional variance. For the other orders, on the other hand, we fail to reject the null hypothesis.

The reason why model AR(1) and MA(1) have been chosen as two of my favorite models is that statistical indicators, such as R-squared, F-statistic, p-value as well as AIC and BIC indicate that in comparison to other models, these two models perform best. An economic interpretation to this choice would be the fact that forecasted values mostly depend on the first order p and q, and not higher orders. Thus, the choice of these two models would assist us in obtaining our best future predictions.

Answers to question 2.

(a). In time series data, heteroscedasticity is known as the occurrence where the standard deviations of a predicted variable, observed over different values of the independent variable or related to prior time periods, are non-constant. The reason why heteroskedasticity might be problematic for our data set is that the former often occurs in datasets that have a large range between the largest and smallest observed values. One of the main causes of heteroskedasticity is the proportional change in the error variance with a factor. In Time Series this factor is time itself.



Before testing for heteroskedasticity I would first check if there is any series correlation in the data set. Only after the serial correlation has been corrected for would I test for heteroskedasticity. If I were to suspect that the current value of my dependent variable is dependent on its past value, then I would first fit the residuals and then check if the are significantly different from zero or not. If I find out that the residuals are not free of autocorrelation, a possible solution would be to include a lag in my model. I will keep testing for autocorrelation on the residuals and add as many lags as needed, up until I solve for the serial correlation problem. Once there no longer is evidence of any serial correlation, I will move on with testing for heteroskedasticity. I will do so by running a Breusch–Pagan test by regressing the squared residuals on the independent variables, using an auxiliary regression equation, or a Lagrange

Multiplier test. If heteroskedasticity is present then, I will use heteroskedasticity/autocorrelation HAC corrected standard errors.

(b). Autocorrelation is the phenomenon of similarity among given time series and a lagged version of itself over successive time intervals. The OLS property that will be violated under the presence of serial correlation is the fifth Finite Sample assumption as well as the fifth asymptotic assumption for higher order AR (1). Both of these assume “No Serial Correlation”. Autocorrelation does not seem to pose a problem for an AR(1) model, as long as the sample size is large enough. When residuals are serially correlated it will be difficult to assume homoskedasticity. Besides Homoskedasticity and Strict exogeneity, all of the other OLS properties remain valid when there is autocorrelation.

The issues that arise when there is autocorrelation are the fact that OLS standard errors and tests will no longer be valid. Our inference will be incorrect. Furthermore, OLS will no longer be efficient. To solve for these problems that autocorrelation creates, it is first important to understand how to detect it. Usually with autocorrelation, positive values tend to be followed by positive ones, and negative values tend to be followed by negative ones. Or it might also happen that positive values are followed by negative ones, and negative values are followed by positive ones. Once autocorrelation has been detected we can run tests such as the Durbin-Watson test. We want dL to be smaller than our DW test so as to not be able to reject our null hypothesis of ‘no serial correlation’. This test, however, only operated under the assumption of Strict Exogeneity. To be able to run more generalized tests that allow for Strict Exogeneity to be violated, we can run a t-test and Lagrange multiplier test. If autocorrelation cannot be removed, we need to use corrected standard errors. We compute serial correlation-robust standard errors. An alternative method would be to take first differences of the dependent variable and test again for Autocorrelation.

(c). Because the Gauss-Markov Theorem requires both heteroskedasticity and no serial correlation in the errors, OLS will no longer be BLUE when serial correlation is present. OLS standard errors and test statistics are no loner valid. This not only applies to the Finite Sample Assumptions but also to the Asymptotic ones. The assumption being violated is “No Serial Correlation”.

The inconsistency in the OLS estimates with lagged dependent variables, caused by first order autocorrelated residuals, comes from writing the model in the error form, where

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + U_t \quad \text{and}$$

$\{U_t\}$ follows a stable AR(1) process, and where

$$E(e_t \mid U_{t-1}, U_{t-2}, \ldots) = E(e_t \mid Y_{t-1}, Y_{t-2}, \ldots) = 0$$

Because by assumption e_t is uncorrelated with Y_{t-1},

$$Cov(Y_{t-1}, U_t) = \rho \, Cov(Y_{t-1}, U_{t-1})$$

This will not be zero, unless $\rho = 0$.
This will, hence, cause OLS estimators β_0 and β_1 from the regression of Y_t on Y_{t-1} to be inconsistent.

Hence, when errors U_t follow an AR(1) model, and model contains a lagged dependent variable, OLS will be inconsistent

QUESTION 3 : PREDICTION WITH AR

OPTIMAL PREDICTOR FOR $\boxed{AR(1)}$ MODEL : $Y_t = \varepsilon_t + a Y_{t-1}$

where $\varepsilon_t \sim WN(0, \sigma^2)$

We have the following information set :

$$I_T = \{Y_1, Y_2, \dots Y_T ; \varepsilon_1, \varepsilon_2, \dots, \varepsilon_T\}.$$

Given the above, the one-period ahead optimal predictor can be found by computing

$$Y_{T+1} = a Y_t + \varepsilon_{T+1}$$

Applying the information from our information set to the above we obtain

$$Y_{T+1,T} = E(Y_{T+1} | I_T) = E(a Y_T + \varepsilon_{T+1} | I_T)$$

$$= E(a Y_T | I_T) + E(\varepsilon_{T+1} | I_T)$$

$$= a Y_t + 0$$

$$= \underline{\underline{a Y_t}}$$

THE PREDICTION ERROR IS :

$$e_{T+1,T} = Y_{T+1} - Y_{T+1,T} = \underline{\underline{\varepsilon_{T+1}}}$$

THE PREDICTION ERROR VARIANCE : $Var(e_{T+1,T}) = Var(\varepsilon_{T+1}) = \underline{\underline{\sigma^2}}$

The two-period ahead optimal predictor :

$$Y_{T+2} = a Y_{T+1} + \varepsilon_{T+2}$$

$$Y_{T+2} = a(a Y_t + \varepsilon_{T+1}) + \varepsilon_{T+2}$$

$$Y_{T+2} = a^2 Y_t + a \varepsilon_{T+1} + \varepsilon_{T+2}$$

Applying the information set, we obtain

$$Y_{T+2,T} = E(Y_{T+2} | I_T) = E(a^2 Y_t + a \varepsilon_{T+1} + \varepsilon_{T+2} | I_T)$$

$$= a^2 Y_t + a \cdot 0 + 0$$

The two-period ahead predictor is also a linear function of the final observed value, but with coefficient a^2.

The Prediction error is

$$e_{T+2,T} = Y_{T+2} - Y_{T+2,T} = \underline{a\,\varepsilon_{T+1} + \varepsilon_{T+2}}$$

The prediction error variance is:

$$Var(e_{T+2,T}) = Var(a\,\varepsilon_{T+1} + \varepsilon_{T+2}) = (1+a^2)\,\sigma^2$$

The three-period ahead optimal predictor:

$$Y_{T+3} = aY_{t+2} + \varepsilon_{T+3}$$

$$Y_{T+3} = a\left(aY_{T+1} + \varepsilon_{T+2}\right) + \varepsilon_{T+3}$$

$$Y_{T+3} = a\left(a\left(aY_T + \varepsilon_{T+1}\right) + \varepsilon_{T+2}\right) + \varepsilon_{T+3}$$

$$Y_{T+3} = a\left[a^2 Y_T + a\varepsilon_{T+1} + \varepsilon_{T+2}\right] + \varepsilon_{T+3}$$

$$Y_{T+3} = a^3 Y_T + a^2\varepsilon_{T+1} + a\varepsilon_{T+2} + \varepsilon_{T+3}$$

Applying the information set, we obtain

$$Y_{T+3,T} = E\left[Y_{T+3} / I_T\right] = E\left[a^3 Y_T + a^2\varepsilon_{T+1} + a\varepsilon_{T+2} + \varepsilon_{T+3}\right]$$

$$= a^3 Y_T + a^2 \cdot 0 + a \cdot 0 + 0$$

$$= \underline{\underline{a^3 Y_T}}.$$

The prediction error is:

$$e_{T+3,T} = Y_{T+3} - Y_{T+3,T} = \underline{a^2\varepsilon_{T+1} + a\varepsilon_{T+2} + \varepsilon_{T+3}}$$

The prediction error variance is:

$$Var\left(e_{T+3,T}\right) = Var\left(a^2\varepsilon_{T+1} + a\varepsilon_{T+2} + \varepsilon_{T+3}\right) = \sigma^2\left(a^4 + a^2 + 1\right)$$

OPTIMAL PREDICTOR FOR AR(2) MODEL: $Y_t = aY_{t-1} + a_2 Y_{t-2} + \varepsilon_t$

where $\varepsilon_t \sim WN(0, \sigma^2)$

$$Y_{t+1} = a_1 Y_t + a_2 Y_{t-1} + \varepsilon_{T+1}$$

The one-period ahead optimal predictor is found by computing

$$\begin{aligned} Y_{T+1,T} &= E[Y_{T+1} \mid I_T] \\ &= E(a_1 Y_T \mid I_T) + E(a_2 Y_{T-1} \mid I_T) + E(\varepsilon_{T+1} \mid I_T) \\ &= \underline{\underline{a_1 Y_T + a_2 Y_{T-1}}} \end{aligned}$$

The prediction error is :

$$e_{T+1,T} = Y_{T+1} - Y_{T+1 \mid T} = \varepsilon_{T+1}$$

The prediction error variance is :

$$Var(e_{T+1,T}) = \underline{\underline{\sigma^2}}$$

The two-period ahead optimal predictor is :

$$\begin{aligned} Y_{T+2,T} &= E(Y_{T+2} \mid I_T) \\ &= E(a_1 Y_{T+1} \mid I_T) + E(a_2 Y_T \mid I_T) + E(\varepsilon_{T+2} \mid I_T) \\ &= a_1 Y_{T+1,T} + a_2 Y_t \\ &= a_1 (a_1 Y_T + a_2 Y_{T-1}) + a_1 a_2 Y_{T-1} \\ &= a_1^2 Y_t + 2a_1 a_2 Y_{T-1} \end{aligned}$$

The prediction error is :

$$e_{T+2,T} = Y_{T+2} - Y_{T+2,T} = \underline{a_1 \varepsilon_{T+1} + \varepsilon_{T+2}}$$

The prediction error variance is :

$$Var(e_{T+2,T}) = Var(a_1 \varepsilon_{T+1} + \varepsilon_{T+2}) = \underline{(1 + a_1^2)\sigma^2}$$

Three - periods ahead optimal predictor:

$$
\begin{aligned}
Y_{T+3,T} &= E[Y_{T+3} \mid I_T] \\
&= E(a_1 Y_{T+2} \mid I_T) + E(a_2 Y_{T+1} \mid I_T) + E(\varepsilon_{T+3} \mid I_T) \\
&= a_1 Y_{T+2 \mid T} + a_2 Y_{T+1 \mid T} \\
&= a_1 (a_1^2 Y_T + 2 a_1 a_2 Y_{T-1}) + a_2 (a_1 Y_T + a_2 Y_{T-1}) \\
&= a_1^3 Y_T + 2 a_1^2 a_2 Y_{T-1} + a_2 a_1 Y_T + a_2^2 Y_{T-1}
\end{aligned}
$$

The prediction error is:

$$e_{T+3,T} = Y_{T+3} - Y_{T+3,T} = a^2 \varepsilon_{T+1} + a \varepsilon_{T+2} + \varepsilon_{T+3}$$

The prediction error variance is:

$$
\begin{aligned}
Var(e_{T+3,T}) &= Var(a^2 \varepsilon_{T+1} + a \varepsilon_{T+2} + \varepsilon_{T+3}) = a^4 \sigma^2 + a^2 \sigma^2 + \sigma^2 \\
&= \sigma^2 (a^4 + a^2 + 1)
\end{aligned}
$$

For a stationary series and model, the prediction error variance will gradually converge to the mean. once it reaches the mean, it will stay there for good.