Homework 11

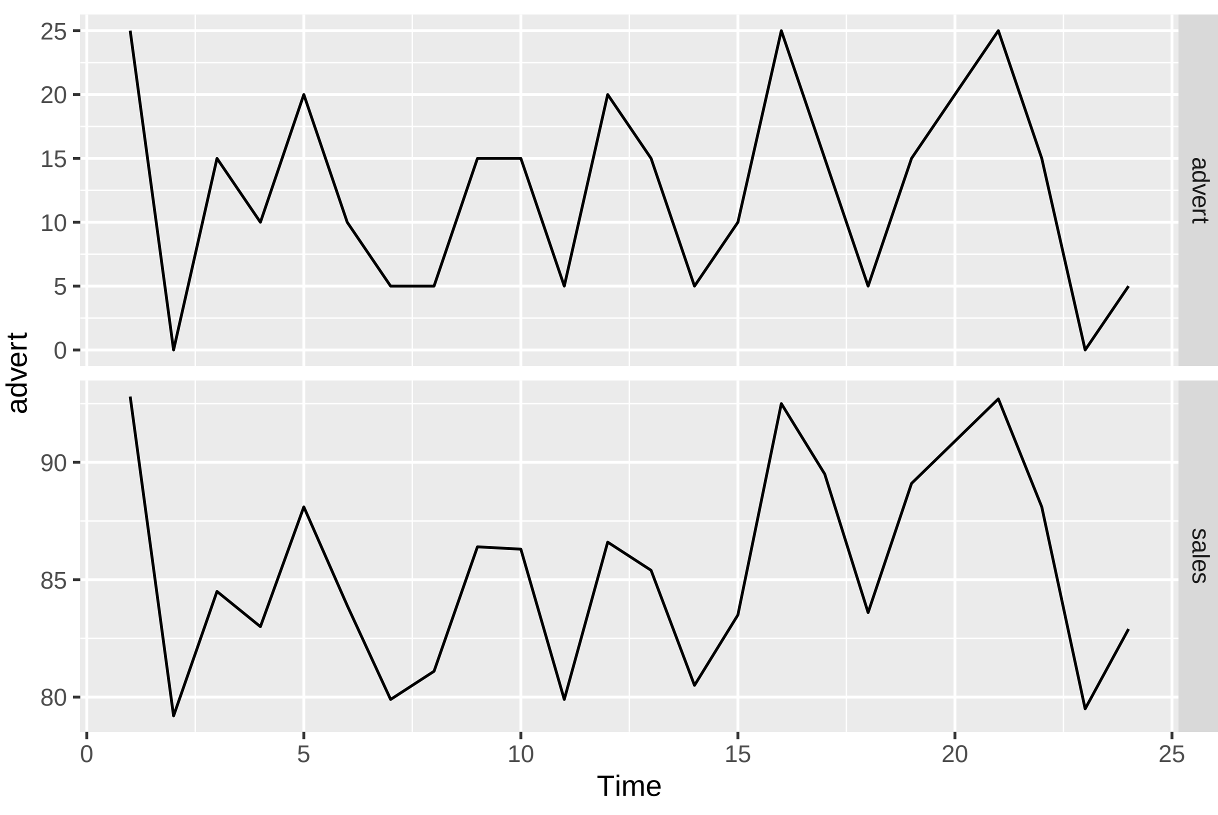
Joshua Ingram

11/30/2020

Problem 1

1.

```r
autoplot(advert, facets = TRUE)
```



2.

$$sales_t = \beta_0 + \beta_1 advert_t + \epsilon_t, \quad \epsilon_t \sim_{i.i.d.} N(0, \sigma^2)$$

```r
lm_1 <- lm(sales ~ advert, data = advert)
summary(lm_1)
```

```
##
## Call:
## lm(formula = sales ~ advert, data = advert)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8194 -1.1375 -0.2412  0.9123  2.7519
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 78.73426    0.59735  131.81  < 2e-16 ***
## advert       0.53426    0.04098   13.04 7.96e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.506 on 22 degrees of freedom
## Multiple R-squared:  0.8854, Adjusted R-squared:  0.8802
## F-statistic:   170 on 1 and 22 DF,  p-value: 7.955e-12
```

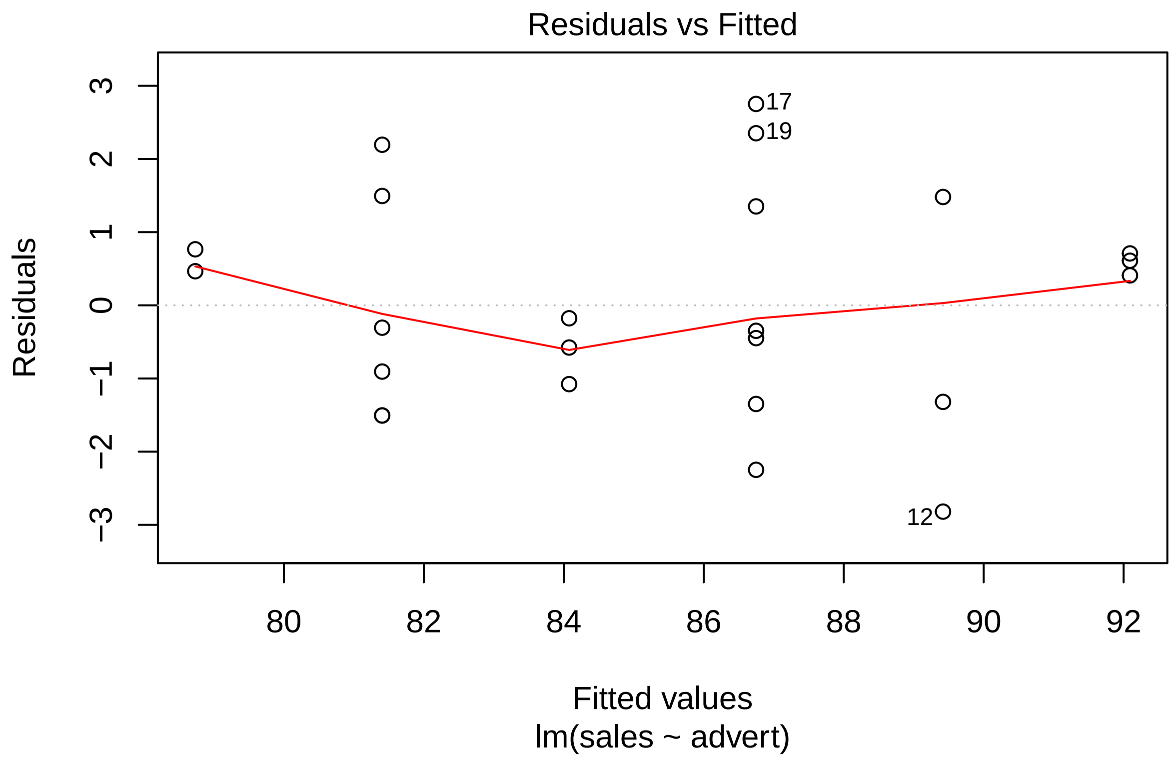
```r
confint(lm_1)
```

```
##                  2.5 %     97.5 %
## (Intercept) 77.4954320 79.9730865
## advert       0.4492764  0.6192421
```

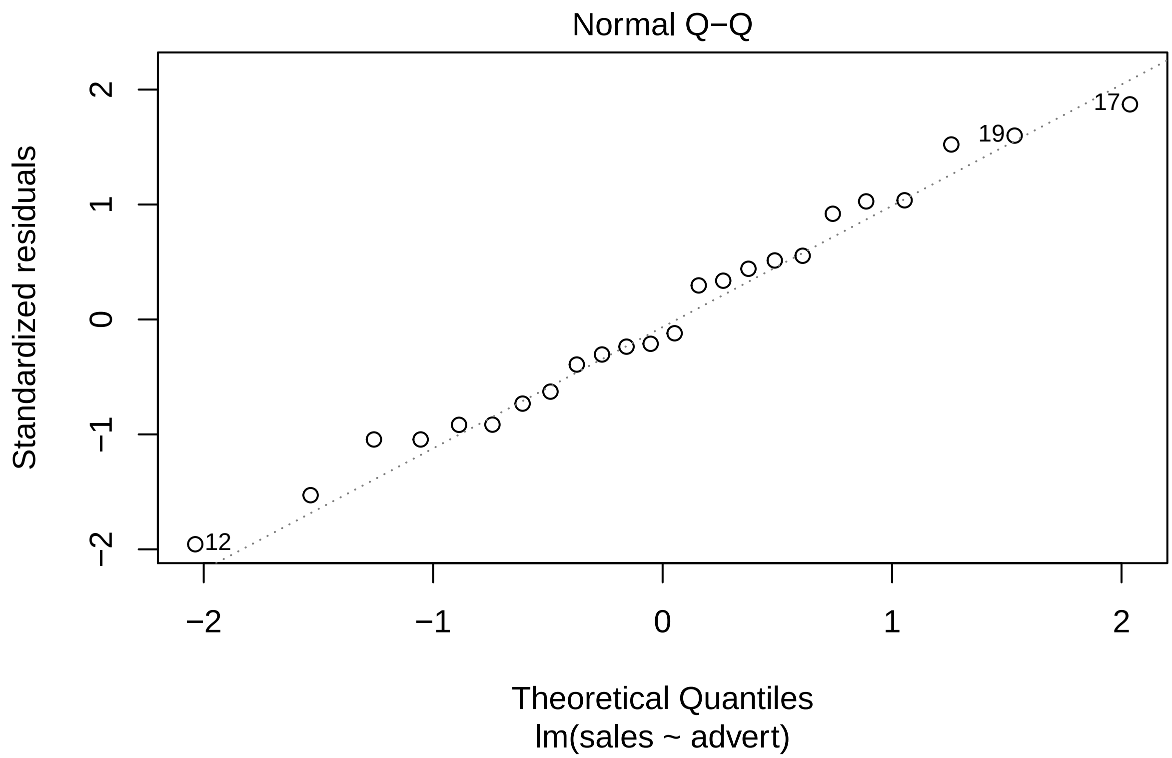
3.

We should not instantly trust the inference from part 2, as we need to check the assumptions on our residuals.

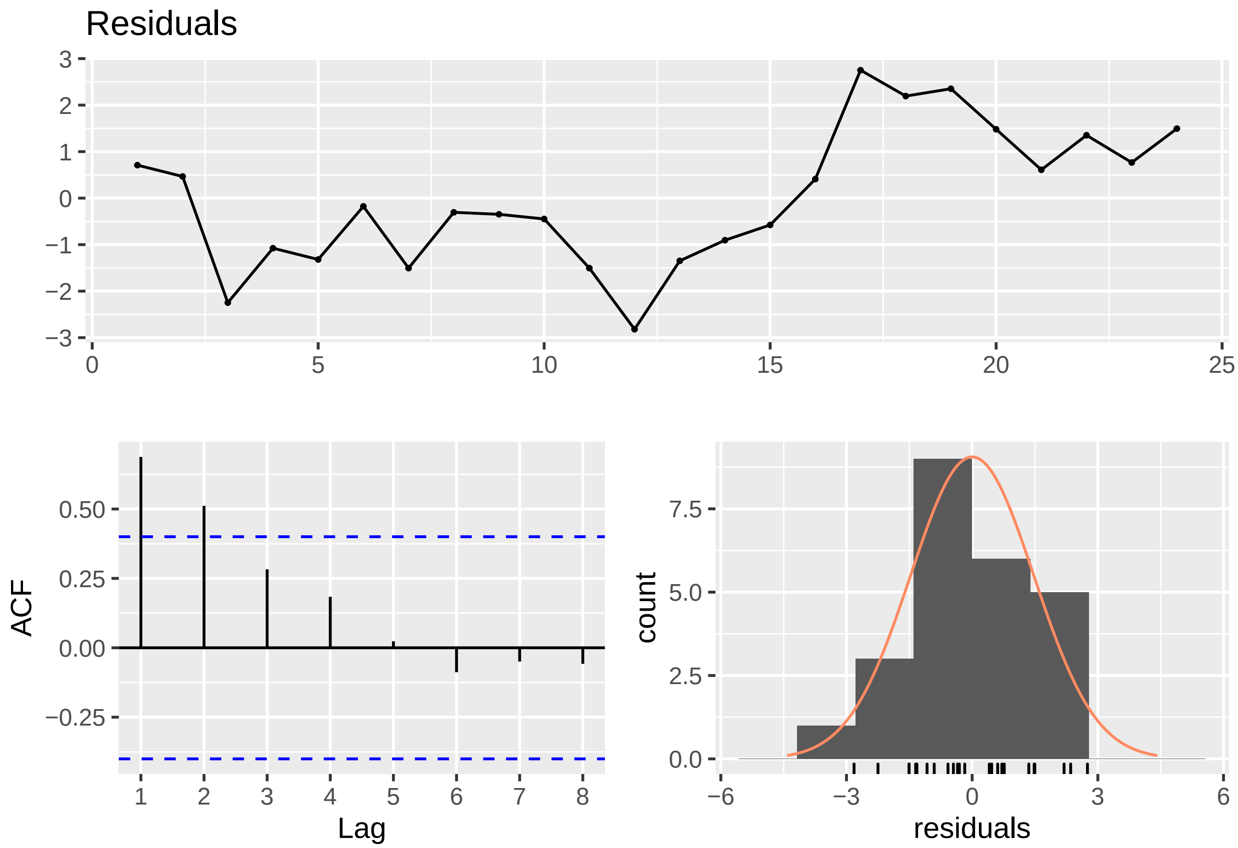
```r
plot(lm_1, 1)
```


Residuals vs Fitted
Residuals
3
2
1
0
−1
−2
−3
17
19
12
80
82
84
86
88
90
92
Fitted values
lm(sales ~ advert)

```
plot(lm_1, 2)
```




```r
checkresiduals(lm_1)
```




```
##
##  Breusch-Godfrey test for serial correlation of order up to 5
##
## data:  Residuals
## LM test = 12.498, df = 5, p-value = 0.02856
```

We find that there is autocorrelation in the residuals after conducting the Breusch-Godfrey test, as we received a small p-value (0.029). Our results are not reliable since the independence assumption of our residuals has been broken.

4.

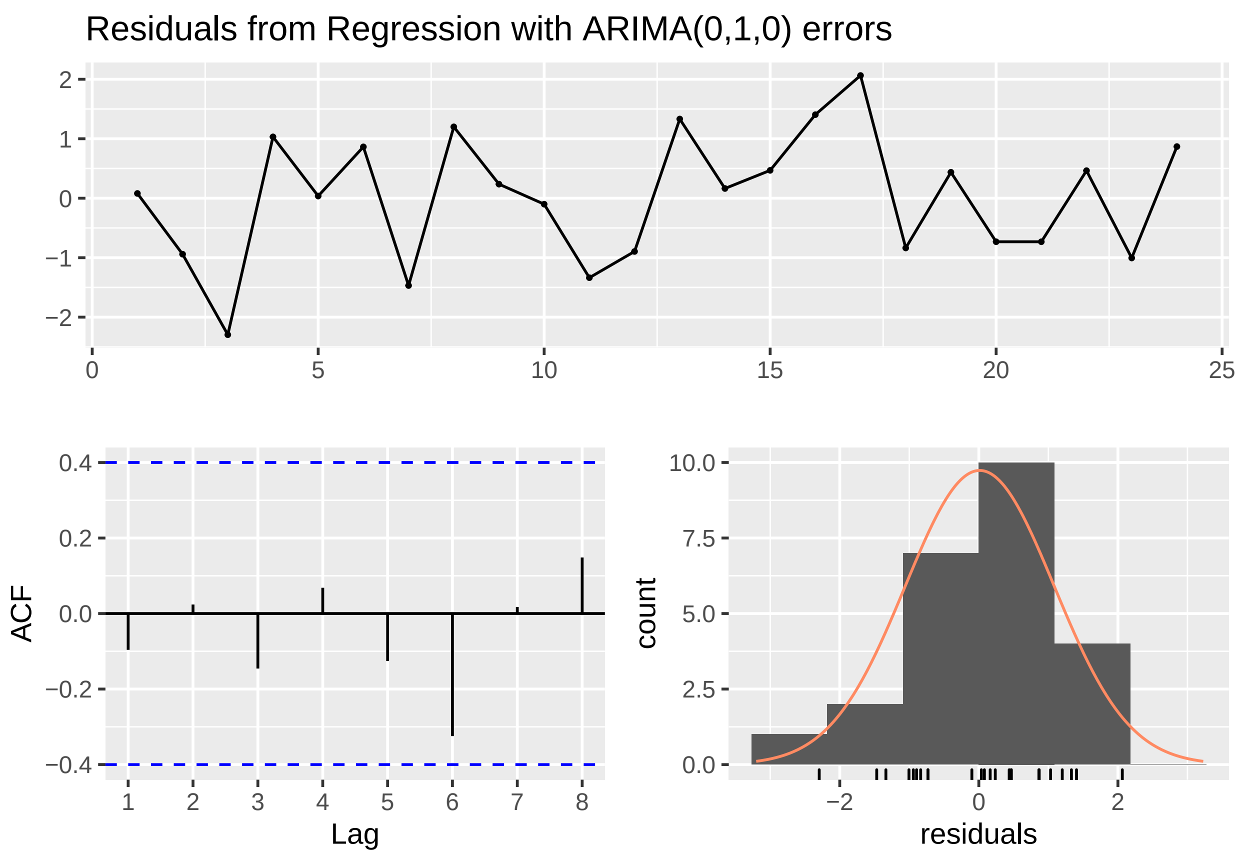
We need to utilize regression with ARIMA errors.

```
auto_1 <- auto.arima(advert[,"sales"],xreg = advert[,"advert"])
summary(auto_1)
```

```
## Series: advert[, "sales"]
## Regression with ARIMA(0,1,0) errors
##
## Coefficients:
##         xreg
##       0.5063
## s.e.  0.0210
```

```
##
## sigma^2 estimated as 1.201:  log likelihood=-34.22
## AIC=72.45   AICc=73.05   BIC=74.72
##
## Training set error measures:
##                      ME     RMSE       MAE          MPE     MAPE     MASE
## Training set 0.01279435 1.049041 0.8745732 -0.00247038 1.032833 0.189587
##                     ACF1
## Training set -0.09614401
```

```
checkresiduals(auto_1)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from Regression with ARIMA(0,1,0) errors
## Q* = 1.5622, df = 4, p-value = 0.8156
##
## Model df: 1.   Total lags used: 5
```

Taking a single-order difference makes the residuals uncorrelated. Written in ARMA form, we remove the intercept β_0 as we apply the differencing to all variables.

$$sales'_t = \beta_1 advert'_t + \epsilon'_t, \quad \epsilon'_t \sim_{i.i.d.} N(0, \sigma^2)$$

$$\epsilon_t' = \epsilon_t - \epsilon_{t-1}$$

```r
fit_2 <- Arima(advert[,"sales"],xreg = advert[,"advert"],order=c(0,1,0))
```

5.

Classical Regression

```r
summary(lm_1)
```

```
##
## Call:
## lm(formula = sales ~ advert, data = advert)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8194 -1.1375 -0.2412  0.9123  2.7519
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 78.73426    0.59735  131.81  < 2e-16 ***
## advert       0.53426    0.04098   13.04 7.96e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.506 on 22 degrees of freedom
## Multiple R-squared:  0.8854, Adjusted R-squared:  0.8802
## F-statistic:   170 on 1 and 22 DF,  p-value: 7.955e-12
```

```r
confint(lm_1)
```

```
##                  2.5 %     97.5 %
## (Intercept) 77.4954320 79.9730865
## advert       0.4492764  0.6192421
```

Dynamic Regression

```r
coeftest(fit_2)
```

```
##
## z test of coefficients:
##
##      Estimate Std. Error z value  Pr(>|z|)
## xreg 0.506346   0.021014  24.095 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
confint(fit_2)
```

```
##          2.5 %    97.5 %
## xreg 0.4651587 0.5475336
```

The classic regression and dynamic regression p-values and intervals differ. I'm hesitant to interpret the confidence intervals since this is for the DIFFERENCED time series, whereas the classical regression is not. The dynamic regression p-value is also much smaller than the classic regression p-value. The estimates differ slightly, as well.

6.

$$\hat{sales_t}' = 0.506346 \, advert'_t$$

Per one unit increase in the consecutive advertising monthly expenditures, the consecutive monthly change in sales will increase by 0.506 units, on average.

Problem 2

1.

```r
lm_2 <- lm(`log(GNP)` ~ ., data = USeconomic)
vif(lm_2)
```

```
## `log(M1)`        rs        rl
##  1.056049  6.373620  6.504435
```

```r
lm_3 <- lm(`log(GNP)` ~ . - rl, data = USeconomic)
vif(lm_3)
```

```
## `log(M1)`        rs
##  1.008113  1.008113
```

```r
lm_3
```

```
##
## Call:
## lm(formula = `log(GNP)` ~ . - rl, data = USeconomic)
##
## Coefficients:
## (Intercept)    `log(M1)`           rs
##      -4.601        1.929        7.092
```

We dropped rl as it had the largest VIF over 5.

2.

$$log(GNP)_t = \beta_0 + \beta_1 log(M1)_t + \beta_2 rs_t + \epsilon_t, \quad \epsilon_t \sim_{i.i.d.} N(0, \sigma^2)$$

```r
summary(lm_3)
```

```
##
## Call:
## lm(formula = `log(GNP)` ~ . - rl, data = USeconomic)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.269342 -0.086414 -0.000206  0.057088  0.251507
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.6007     0.7663  -6.004 1.73e-08 ***
## `log(M1)`     1.9285     0.1236  15.609  < 2e-16 ***
## rs            7.0916     0.3149  22.520  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1133 on 133 degrees of freedom
## Multiple R-squared:  0.8605, Adjusted R-squared:  0.8584
## F-statistic: 410.2 on 2 and 133 DF,  p-value: < 2.2e-16
```

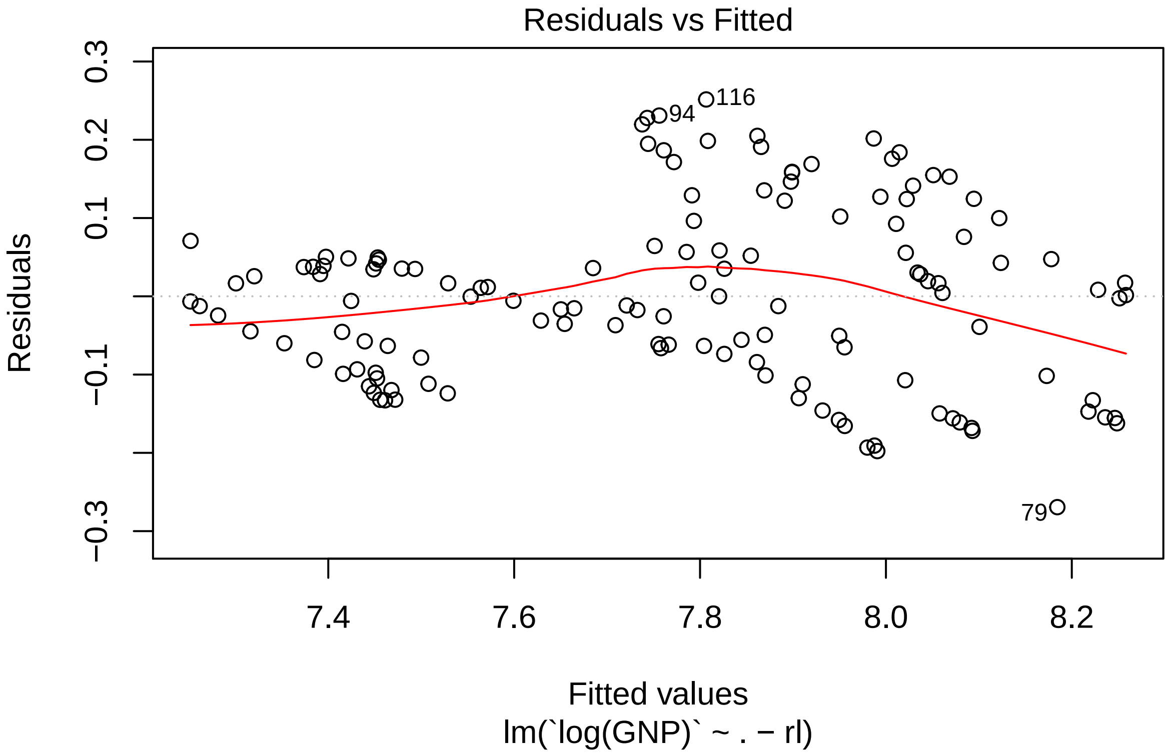
```r
confint(lm_3)
```

```
##                 2.5 %    97.5 %
## (Intercept) -6.116433 -3.085034
## `log(M1)`    1.684158  2.172915
## rs           6.468776  7.714517
```

3.

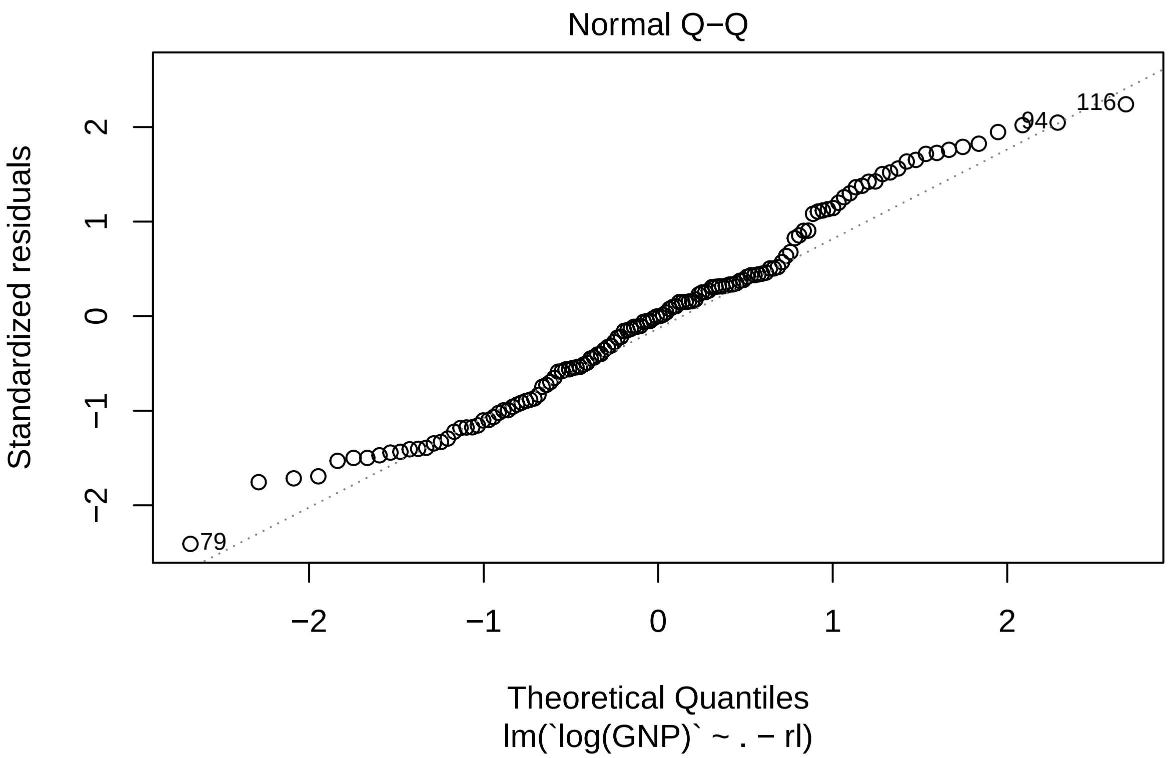
No, we should not trust the inference from part 2 at face value. Check residuals!

```r
plot(lm_3, 1)
```

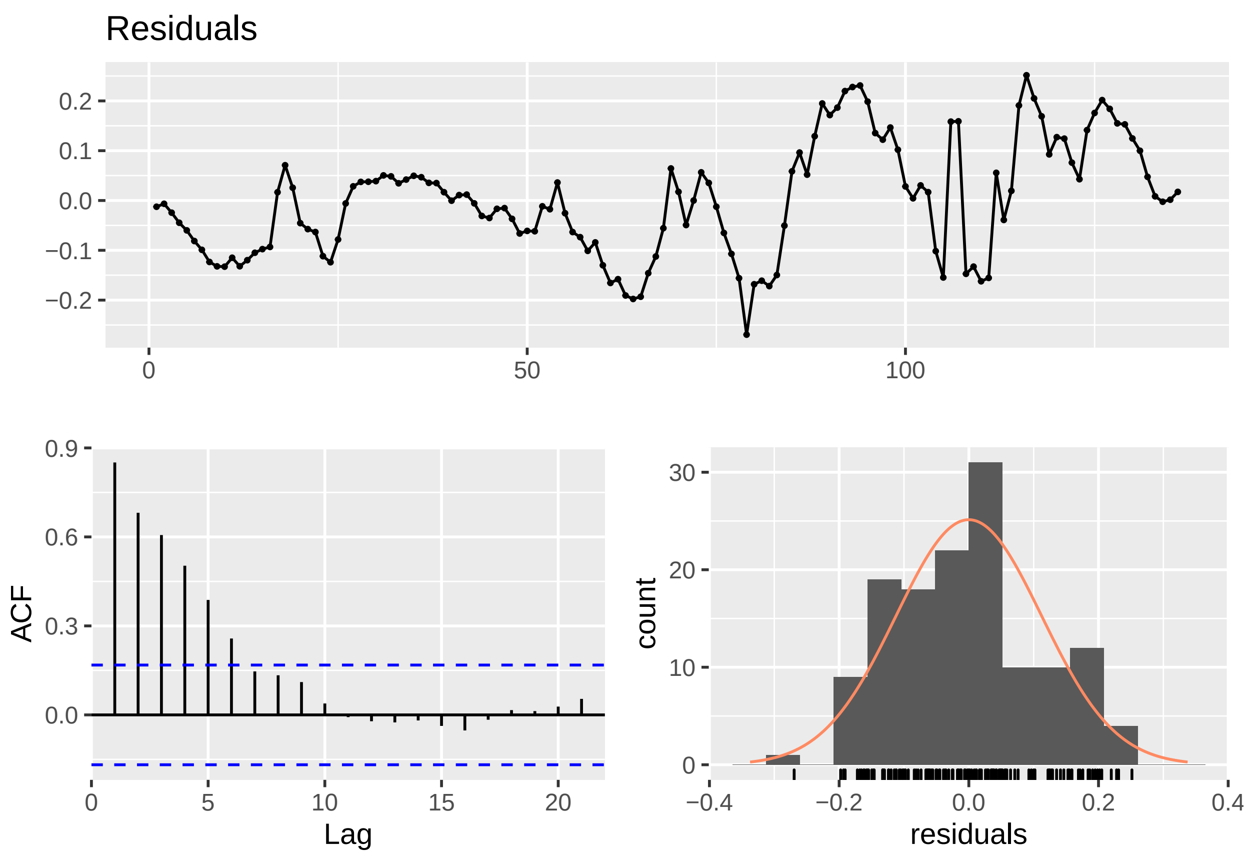


Residuals vs Fitted
Residuals
Fitted values
lm(`log(GNP)` ~ . − rl)
0.3
0.2
0.1
−0.1
−0.3
7.4
7.6
7.8
8.0
8.2
116
94
79


```r
plot(lm_3, 2)
```




```
checkresiduals(lm_3)
```




```
##
##  Breusch-Godfrey test for serial correlation of order up to 10
##
## data:  Residuals
## LM test = 107.69, df = 10, p-value < 2.2e-16
```

Residuals are not independently distributed by the Breusch-Godfrey test, as the p-value was small and we reject the null that they are not autocorrelated. Our inference from classical regression is not reliable.

4.

We should use dynamic regression to utilize ARIMA errors.

```
auto_2 <- auto.arima(USeconomic[,"log(GNP)"],xreg = USeconomic %>% as.data.frame %>% select(`log(M1)`,
summary(auto_2)
```

```
## Series: USeconomic[, "log(GNP)"]
## Regression with ARIMA(0,1,0) errors
##
## Coefficients:
##         drift  log(M1)      rs
##        0.0068   0.2791  0.4146
## s.e.   0.0008   0.0613  0.0883
##
```

```
## sigma^2 estimated as 8.055e-05:  log likelihood=446.45
## AIC=-884.9   AICc=-884.59   BIC=-873.28
##
## Training set error measures:
##                        ME        RMSE         MAE          MPE       MAPE
## Training set 4.067923e-05 0.008842105 0.006972419 0.0008151783 0.08953893
##                   MASE       ACF1
## Training set 0.199518 0.09159984
```

```r
checkresiduals(auto_2)
```


```
##
##  Ljung-Box test
##
## data:  Residuals from Regression with ARIMA(0,1,0) errors
## Q* = 4.2184, df = 5, p-value = 0.5184
##
## Model df: 3.   Total lags used: 8
```

$$log(GNP)'_t = \beta_1 log(M1)'_t + \beta_2 rs_t + \epsilon'_t$$

$$\epsilon'_t = \mu + \eta_t, \;\; \eta_t \sim_{i.i.d.} N(0, \sigma^2), \; \epsilon'_t = \epsilon_t - \epsilon_{t-1}$$

```r
fit_3 <- Arima(USeconomic[,"log(GNP)"],xreg = USeconomic %>% as.data.frame %>% select(`log(M1)`, rs) %>%
```

5.

Classical Regression

```r
summary(lm_3)
```

```
##
## Call:
## lm(formula = `log(GNP)` ~ . - rl, data = USeconomic)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.269342 -0.086414 -0.000206  0.057088  0.251507
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.6007     0.7663  -6.004 1.73e-08 ***
## `log(M1)`     1.9285     0.1236  15.609  < 2e-16 ***
## rs            7.0916     0.3149  22.520  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1133 on 133 degrees of freedom
## Multiple R-squared:  0.8605, Adjusted R-squared:  0.8584
## F-statistic: 410.2 on 2 and 133 DF,  p-value: < 2.2e-16
```

```r
confint(lm_3)
```

```
##                 2.5 %    97.5 %
## (Intercept) -6.116433 -3.085034
## `log(M1)`    1.684158  2.172915
## rs           6.468776  7.714517
```

Dynamic Regression

```r
coeftest(fit_3)
```

```
##
## z test of coefficients:
##
##         Estimate Std. Error z value  Pr(>|z|)
## log(M1) 0.383478   0.074955  5.1161 3.119e-07 ***
## rs      0.450252   0.110071  4.0905 4.304e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
confint(fit_3)
```

```
##              2.5 %    97.5 %
## log(M1) 0.2365687 0.5303881
## rs      0.2345164 0.6659884
```

The p-values are larger from the dynamic regression model than in the classical model… I'm hesitant to interpret the confidence intervals since this is for the DIFFERENCED time series, whereas the classical regression is not. The estimates are different for both models, as well.

We should trust the p-values from dynamic regression more than classical since the modeling assumptions are satisfied under the dynamic regression model.

6.

$$log(\hat{G}NP)'_t = 0.3835\, log(M1)'_t + 0.45 rs'_t, \;\; \epsilon'_t = 0.0068$$

log(M1)

Per one unit increase in the consecutive quarterly logged M1 money supply, the consecutive quarterly logged GNP will increase by 0.3835 units, on average, ceteris paribus. (holding all else constant)

rs

Per one percentage point increase in the consecutive quarterly discount rate on treasurey bills, the consecutive quarterly logged GNP will increase by 0.45 units, on average, ceteris paribus.