# Ec122 - HW7

## Ben Juarez

## 11/15/2021

# (a)

## (1)

The main question/problem in this article is to determine the impact of Vietnam military service on the long-term labor market.

## (2)

The main result was that Vietnam veterans who are white earned significantly less than non-veterans who are white (up to 10 yrs. after service). This would be up to approx. 15% in the early-mid 80s. Although, this effect was not significantly significant for nonwhite veterans.

## (3)

The methods used involved estimating draft eligibility's impact on earnings using draft eligibility as a treatment effect and differences in earnings for each cohort as times series. Earnings, and other dummy/instrumental variables were estimated using a SLR in order to show the effect of military service. They also used GLS due to correlation between residuals and veteran status. The validity of certain variables was also tested appropriately.

## (4)

Yes, I do believe the result. This study was very thoroughly done using the proper methods and the appropriately correlations were considered which makes the results more valid. It is unfortunate to see this effect, but it is important to acknowledge. The conclusions drawn from this study were supported thoroughly with proper significance justifications, etc.

# (c)

# HGL 9.6

### Part a

Considering the given least squares regression estimates, let us interpret the estimate -53.51. The number of new houses sold between $t$ and $t-1$ decreases by 53,510 if there is a 1 unit (1 percentage point) increase in the mortgage rate from $t$ to $t-1$.

Let us construct a 95% confidence interval for the coefficient $DIRATE_{t-1}$ with 218observations ($t_c = 1.971$):

$$(b_2 - t_c \cdot se(b_2), b_2 + t_c \cdot se(b_2)) = (-53.51 - 1.971 \cdot 16.98, -53.51 + 1.971 \cdot 16.98) = (-86.98, -20.04)$$

So, we are 95% confident that the estimate for the decrease in number of new houses sold in this case falls within this range.

## Part b

Let $\hat{e}$ denote the residuals from the given equation. Using the other given estimated equation, let us conduct two separate tests for first-order autoregressive errors.

We can first use a $t$-test for the significance of $\hat{e}_{t-1}$ by testing $H_0 : \hat{e}_{t-1} = 0$ vs $H_1 : \hat{e}_{t-1} \neq 0$:

$$t = \frac{\hat{e}_{t-1} - 0}{se(\hat{e}_{t-1})} = \frac{-0.3306}{0.0649} = -5.09 < -t_c = -1.971 \implies \text{reject } H_0$$

This implies that $\hat{e}_{t-1} \neq 0$, suggesting autocorrelation.

Let us now use a LM test (testing $H_0 : \rho = 0$):

$$LM = T \cdot R^2 = 218 \cdot 0.1077 = 23.48$$

Since 23.48 is greater than 3.84 (5% critical value from a $\chi^2_{(1)}$ distribution), we reject the null hypothesis of no autocorrelation (implying evidence of autocorrelation).

## Part c

Given the estimated model with AR(1) errors, let us construct a 95% confidence interval for the coefficient of $DIRATE_{(t-1)}$:

$$(-58.61 - 1.971 \cdot 14.10, -58.61 + 1.971 \cdot 14.10) = (-86.40, -30.82)$$

We can see that ignoring autocorrelation (parts a vs. c) underestimates the standard error and gives us a slightly different coefficient value which seemingly explains the differences in the intervals. Comparing the intervals, we see that the lower bounds were more similar than the upper bounds because of this effect. Thus, we see the importance of not ignoring autocorrelation.

# HGL 9.7

Consider $e_t = \rho e_{t-1} + v_t$ (AR(1) model).

## Part a

Suppose $\rho = 0.9$ and $\sigma_v^2 = 1$.

### (i)

$\text{corr}(e_t, e_{t-1}) = \rho^1 = 0.9$

### (ii)

$\text{corr}(e_t, e_{t-4}) = \rho^4 = 0.9^4 = 0.66$

### (iii)

$\sigma_e^2 = \frac{\sigma_v^2}{1-\rho^2} = \frac{1}{1-0.9^2} = 5.26$

## Part b

Repeat part (a) with $\rho = 0.4$ and $\sigma_v^2 = 1$

### (i)

$\text{corr}(e_t, e_{t-1}) = \rho^1 = 0.4$

**(ii)**

$\text{corr}(e_t, e_{t-4}) = \rho^4 = 0.4^4 = 0.0256$

**(iii)**

$\sigma_e^2 = \frac{\sigma_v^2}{1-\rho^2} = \frac{1}{1-0..4^2} = 1.19$

Comparing the differences in values, we can see that with a smaller $\rho$, the correlations become quite increasingly smaller between $e_t$ and $e_{t-k}$ as $k$ increases (0.66 vs 0.0256). We also see a greater variance with a larger $\rho$ (5.26 vs 1.19).

## HGL 9.19

```
setwd("~/Desktop/R")
homes = read.table("dat/homes.dat")
colnames(homes) = c("homes", "irate")
```
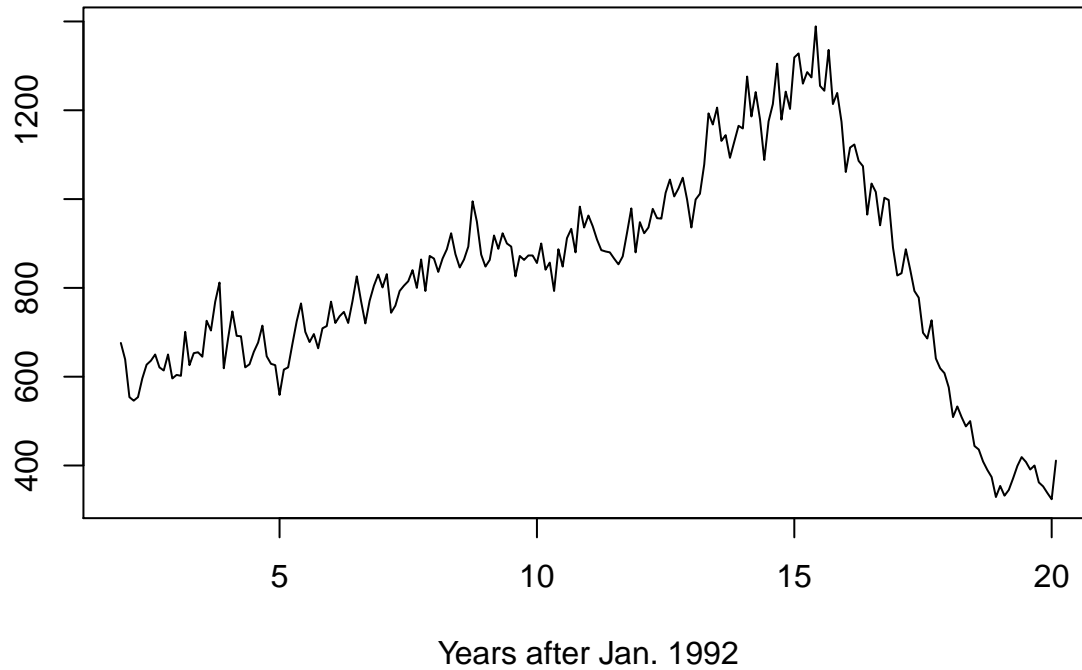
## Part a

```
HOMES = homes$homes
IRATE = homes$irate
DHOMES = data.frame(matrix(ncol = 1, nrow = 219))
DHOMES[1,] = 0
DIRATE = data.frame(matrix(ncol = 1, nrow = 219))
DIRATE[1,] = 0

for (i in 3:220) {
    DHOMES[i-1,] = homes$homes[i-1] - homes$homes[i-2]
    DIRATE[i-1,] = homes$irate[i-1] - homes$irate[i-2]
}

plot(ts(HOMES, start(1992,1), frequency = 12), ylab="",xlab="Years after Jan. 1992",main="HOMES")
```
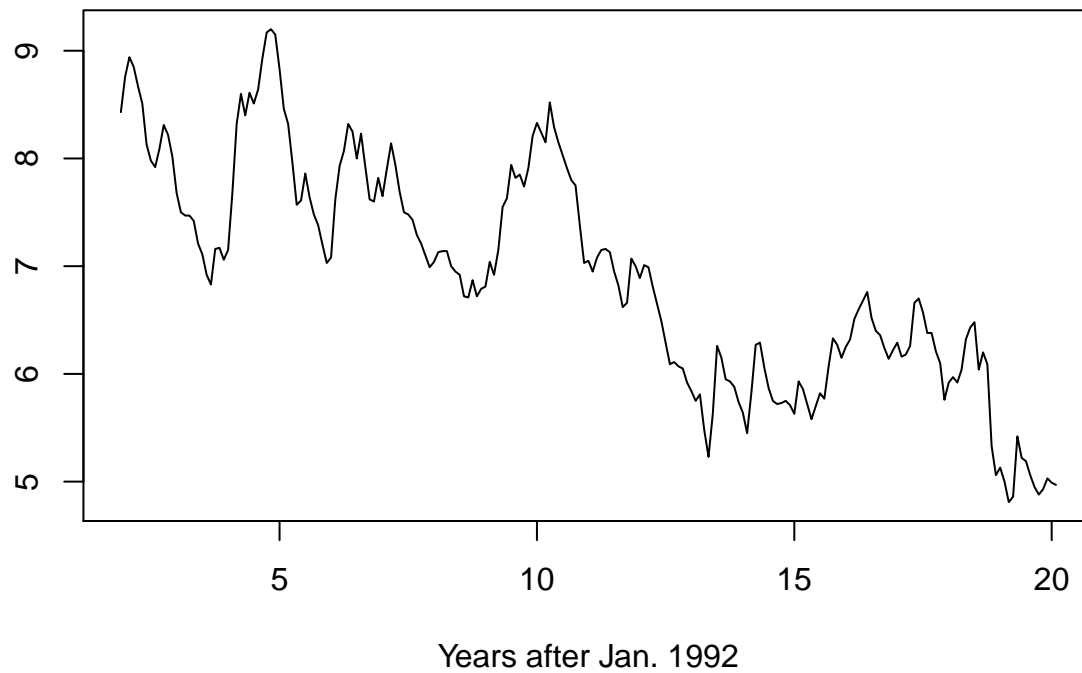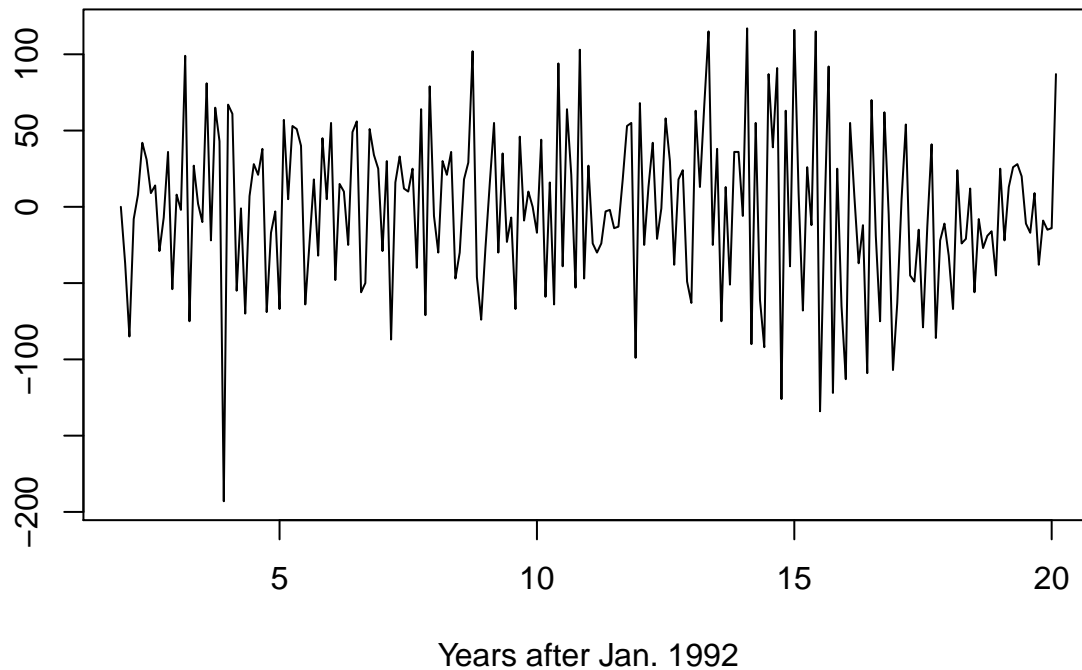
## HOMES



Years after Jan. 1992

```r
plot(ts(IRATE, start(1992,1), frequency = 12), ylab="",xlab="Years after Jan. 1992",main="IRATE")
```
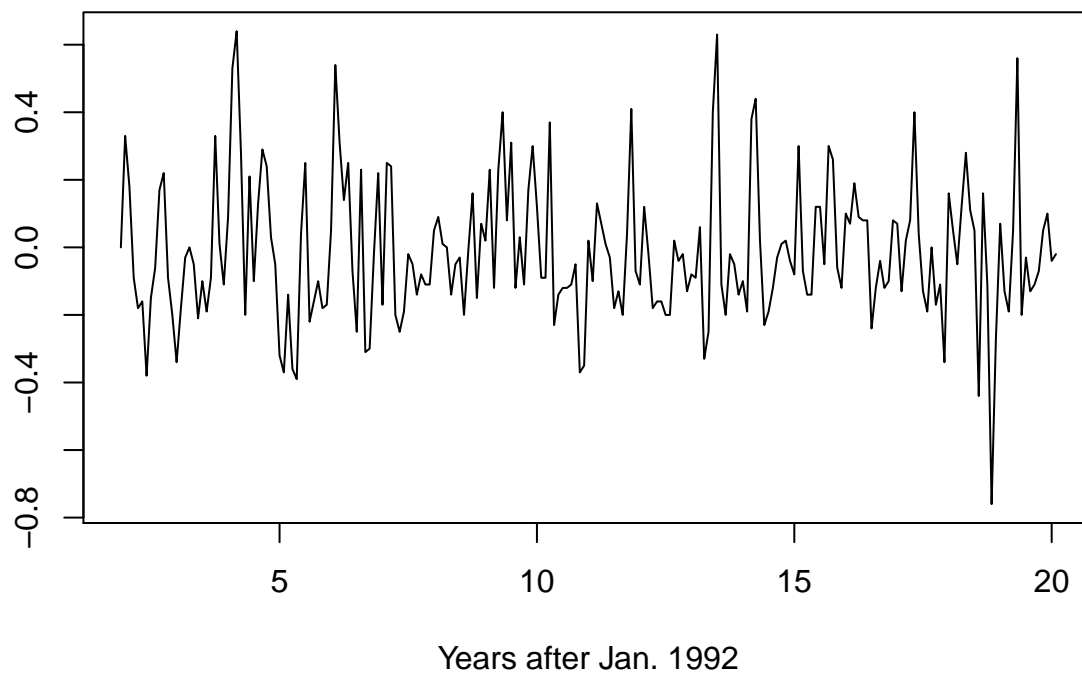
## IRATE



Years after Jan. 1992

```r
plot(ts(DHOMES, start(1992,1), frequency = 12), ylab="",xlab="Years after Jan. 1992",main="DHOMES")
```

## DHOMES



Years after Jan. 1992

```
plot(ts(DIRATE, start(1992,1), frequency = 12), ylab="",xlab="Years after Jan. 1992",main="DIRATE")
```

## DIRATE



Years after Jan. 1992

HOMES positively trends until about 15 years after Jan. 1992 (2005-2006 range) while IRATE's trend fluctuates positively and negatively consistently. It does not appear that DHOMES and DIRATE are trending since they both stay around a constant line.

## Part b

```r
library("dynlm")
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric
```

```r
DHOMES = ts(diff(HOMES))
DIRATE = ts(diff(IRATE))


homes_model = dynlm(DHOMES ~ L(DHOMES) + L(DIRATE) + L(DIRATE,2))
summary(homes_model)
```

```
##
## Time series regression with "ts" data:
## Start = 3, End = 218
##
## Call:
## dynlm(formula = DHOMES ~ L(DHOMES) + L(DIRATE) + L(DIRATE, 2))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -166.073  -29.618   -0.177   34.270  124.573
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.49122    3.33273  -0.748  0.45559
## L(DHOMES)     -0.33501    0.06489  -5.163 5.59e-07 ***
## L(DIRATE)    -50.78783   16.92832  -3.000  0.00302 **
## L(DIRATE, 2) -28.85502   17.12782  -1.685  0.09352 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 48.68 on 212 degrees of freedom
## Multiple R-squared:  0.1493, Adjusted R-squared:  0.1372
## F-statistic:  12.4 on 3 and 212 DF,  p-value: 1.666e-07
```

Examining the summary, we can see that $\delta$ (intercept) and $\delta_2$ ($DIRATE_{t-2}$) are the only estimates that are not significantly different from zero at the 5% significance level (by examining t-values with $t_c = 1.971$ or examining p-values).

## Part c

Let us test $H_0 : \theta_1\delta_1 = -\delta_2 \implies \theta_1\delta_1 + \delta_2 = 0$ vs $H_1 : \theta_1\delta_1 \neq -\delta_2 \implies \theta_1\delta_1 + \delta_2 \neq 0$ at a 5% significance level:

```r
sqrt(c(coef(homes_model)[3],coef(homes_model)[2],1) %*% vcov(homes_model)[2:4,2:4] %*% c(coef(homes_mod
```
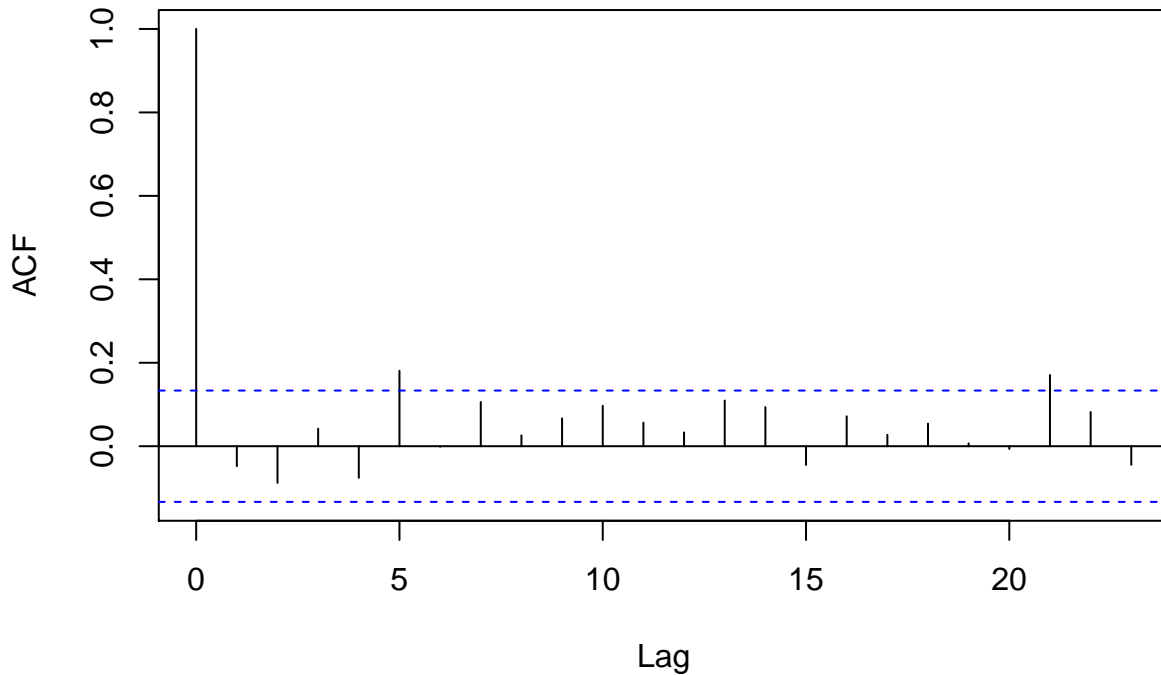
```
##          [,1]
## [1,] 19.26213
```

$$t = \frac{\theta_1\delta_1 + \delta_2}{se(\theta_1\delta_1 + \delta_2)} = \frac{(-0.335)(-50.788) - 28.86}{19.26} = -0.615 \implies |t| < t_c = 1.971$$

Thus, we fail to reject $H_0$ which suggests that we should use the AR(1) model due to the construction of our hypotheses.

## Part d

```
acf(ts(homes_model$residuals), main="")
```



Examining this correlogram, we see potential evidence of serial correlation since $r_5$ and $r_{21}$ (aside from $r_0$) extend past the significance line.

## Part e

```
library("lmtest")
bgtest(homes_model, order=2, type="Chisq", fill=NA)
```

```
##
##  Breusch-Godfrey test for serial correlation of order up to 2
##
## data:  homes_model
## LM test = 5.266, df = 2, p-value = 0.07186
```

Following the results of the LM test, we compare the value 5.266 to 5.991 (5% critical value from $\chi^2_{(1)}$ distribution). Since $5.266 < 5.991$, we fail to reject the null hypothesis, suggesting that the errors are serially uncorrelated.

## Part f

```
f_model = dynlm(DHOMES ~ L(DHOMES) + L(DHOMES,5) + L(DIRATE) + L(DIRATE,3))
summary(f_model)
```

```
##
## Time series regression with "ts" data:
## Start = 6, End = 218
##
## Call:
## dynlm(formula = DHOMES ~ L(DHOMES) + L(DHOMES, 5) + L(DIRATE) +
##     L(DIRATE, 3))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -178.349  -33.092    2.109   30.767  122.917
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.92154    3.28406  -0.890  0.37470
## L(DHOMES)     -0.30731    0.06345  -4.843 2.49e-06 ***
## L(DHOMES, 5)   0.20687    0.06331   3.268  0.00127 **
## L(DIRATE)    -64.32447   15.97394  -4.027 7.92e-05 ***
## L(DIRATE, 3) -46.63059   16.09396  -2.897  0.00417 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 47.5 on 208 degrees of freedom
## Multiple R-squared:  0.2026, Adjusted R-squared:  0.1873
## F-statistic: 13.22 on 4 and 208 DF,  p-value: 1.306e-09
```

```r
acf(ts(f_model$residuals), main="")
```
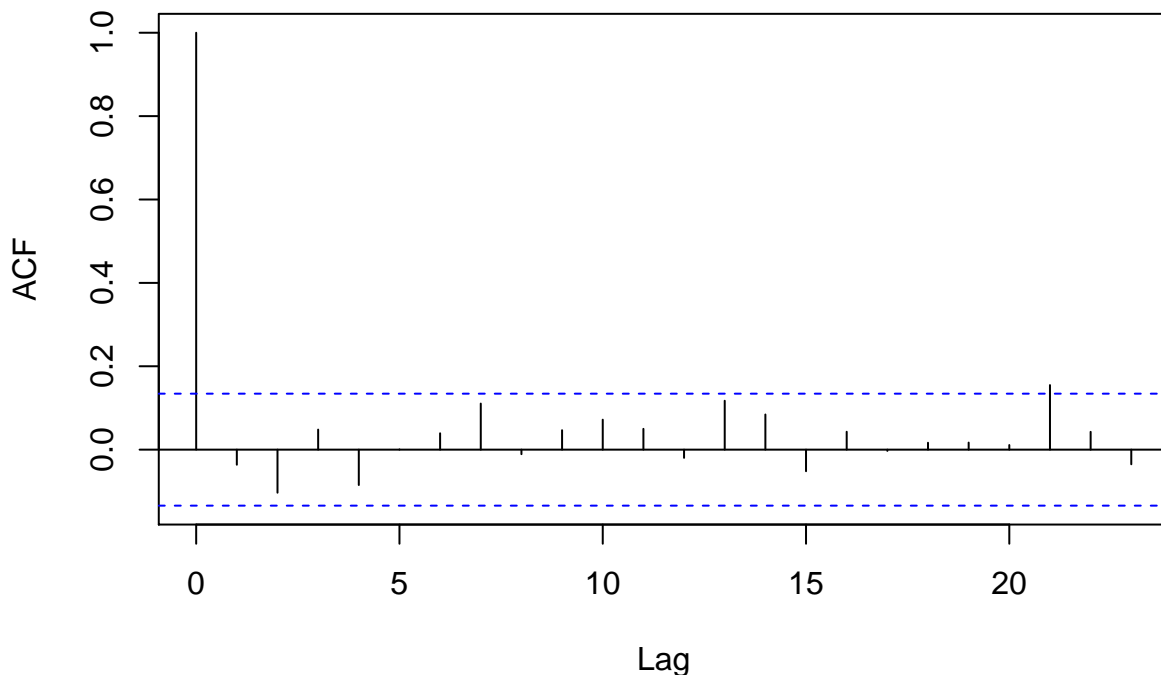


There seems to be improvement with this model over (9.92). Besides $\delta$ (intercept), each coefficient is significant at the 5% level (examining summary output). Also, by examining the correlogram, there is one less significant correlation since only $r_2 1$ passes the significance line (compared to both $r_5$ and $r_{21}$ in the other model).