Problem Set 8

Economemtrics

2a.

```r
library(sandwich)

library(lmtest)
```

```
## Loading required package: zoo

##

## Attaching package: 'zoo'

## The following objects are masked from 'package:base':

##

##     as.Date, as.Date.numeric
```

```r
mc_simulate <- function(n, het = 1) {

    x <- rnorm(n)

    e <- runif(n, -1, 1)


    if (het == 0) {

        u <- e

    } else {

        u <- x * e

    }


    y <- 1 + x + u


    model <- lm(y ~ x)
```

```r
ci <- confint(model)[2, ]

h0 <- coeftest(model, vcov = vcovHC(model, type = "HC0"))
h2 <- coeftest(model, vcov = vcovHC(model, type = "HC2"))
h3 <- coeftest(model, vcov = vcovHC(model, type = "HC3"))

h0_se <- sqrt(diag(vcovHC(model, type = "HC0")))
h2_se <- sqrt(diag(vcovHC(model, type = "HC2")))
h3_se <- sqrt(diag(vcovHC(model, type = "HC3")))

h0_ci <- cbind(h0[, "Estimate"] - 1.96 * h0_se, h0[, "Estimate"] + 1.96 * h0_se)
h2_ci <- cbind(h2[, "Estimate"] - 1.96 * h2_se, h2[, "Estimate"] + 1.96 * h2_se)
h3_ci <- cbind(h3[, "Estimate"] - 1.96 * h3_se, h3[, "Estimate"] + 1.96 * h3_se)

colnames(h0_ci) <- colnames(h2_ci) <- colnames(h3_ci) <- c("2.5 %", "97.5 %")

true_slope <- 1
in_interval <- ci[1] <= true_slope & true_slope <= ci[2]
in_interval_h0 <- h0_ci[1, "2.5 %"] <= true_slope & true_slope <= h0_ci[1, "97.5 %"]
in_interval_h2 <- h2_ci[1, "2.5 %"] <= true_slope & true_slope <= h2_ci[1, "97.5 %"]
in_interval_h3 <- h3_ci[1, "2.5 %"] <= true_slope & true_slope <= h3_ci[1, "97.5 %"]

ci_length <- ci[2] - ci[1]
h0_length <- h0_ci[1, "97.5 %"] - h0_ci[1, "2.5 %"]
h2_length <- h2_ci[1, "97.5 %"] - h2_ci[1, "2.5 %"]
h3_length <- h3_ci[1, "97.5 %"] - h3_ci[1, "2.5 %"]

return(list(
    intervals = list(ci[1], ci[2], h0_ci[1, "2.5 %"], h0_ci[1, "97.5 %"], h2_ci[1, "
```

```
        includes_true = list(in_interval, in_interval_h0, in_interval_h2, in_interval_h3
        lengths = list(ci_length, h0_length, h2_length, h3_length)
    ))
}

results <- matrix(NA, nrow = 1000, ncol = 16)
colnames(results) <- c("ci_lower", "ci_upper", "h0_lower", "h0_upper", "h2_lower", "h2_u
```

```
n <- 30

for (i in 1:1000) {
    sim <- mc_simulate(n)
    results[i, ] <- unlist(c(sim$intervals, sim$includes_true, sim$lengths))
}

averages <- colMeans(results)

print(averages[1:8])
```

```
##  ci_lower  ci_upper  h0_lower  h0_upper  h2_lower  h2_upper  h3_lower  h3_upper
## 0.7992786 1.2192532 0.8163086 1.1847508 0.8044369 1.1966225 0.7912133 1.2098461
```

```
print(averages[9:12])
```

```
##    in_interval in_interval_h0 in_interval_h2 in_interval_h3
##          0.744          0.923          0.938          0.954
```

```
print(averages[13:16])
```

```
## ci_length h0_length h2_length h3_length
## 0.4199746 0.3684422 0.3921856 0.4186327
```

Only the interval for h3 is close to 0.95. The homoskedastic interval has the lowest incidence where the interval contains the true value with the incidences increasing as we go from h0 to h3. The homoskedastic interval has the lowest average length with the lengths increasing as we go from h0 to h3.

The h0 CI length is the shortest while the h3 length is the longest.

2b.

```
n <- 1000

for (i in 1:1000) {
    sim <- mc_simulate(n)
    results[i, ] <- unlist(c(sim$intervals, sim$includes_true, sim$lengths))
}

averages <- colMeans(results)

print(averages[1:8])
```

```
##  ci_lower  ci_upper  h0_lower  h0_upper  h2_lower  h2_upper  h3_lower  h3_upper
## 0.9627286 1.0343461 0.9649016 1.0361907 0.9648302 1.0362622 0.9647585 1.0363339
```

```
print(averages[9:12])
```

```
##   in_interval in_interval_h0 in_interval_h2 in_interval_h3
##         0.739          0.933          0.933          0.933
```

```
print(averages[13:16])
```

```
##  ci_length  h0_length  h2_length  h3_length
## 0.07161758 0.07128908 0.07143199 0.07157540
```

In this case, the homoskedastic interval still has the lowest incidence where the interval contains the true value while h0 to h3 have the true values in their intervals very close to 0.95.

h0 still has the shortest CI legth while h3 has the longest. However, all the CIs have very simmilar lengths in this experiment.

2c.

```r
n <- 30

for (i in 1:1000) {
    sim <- mc_simulate(n, 0)
    results[i, ] <- unlist(c(sim$intervals, sim$includes_true, sim$lengths))
}

averages <- colMeans(results)

print(averages[1:8])
```

```
##  ci_lower  ci_upper  h0_lower  h0_upper  h2_lower  h2_upper  h3_lower  h3_upper
## 0.7762940 1.2256310 0.7955764 1.2007266 0.7882566 1.2080464 0.7803611 1.2159419
```

```r
print(averages[9:12])
```

```
##    in_interval in_interval_h0 in_interval_h2 in_interval_h3
##          0.939          0.927          0.931          0.940
```

```r
print(averages[13:16])
```

```
## ci_length h0_length h2_length h3_length
## 0.4493370 0.4051503 0.4197897 0.4355807
```

In this case, h3 has the higehst incidence where the true value falls within its CI. h0 meanwhile has the lowest, however, compared to the heteroskedastic case, it is much closer to 0.95.

In this case, h0 has the shortest CI length while the homoskedastic CI has the longest. However, all the CIs have very simmilar lengths in this experiment.

2d.

```r
n <- 1000

for (i in 1:1000) {
    sim <- mc_simulate(n, 0)
    results[i, ] <- unlist(c(sim$intervals, sim$includes_true, sim$lengths))
}

averages <- colMeans(results)

print(averages[1:8])
```

```
##  ci_lower  ci_upper  h0_lower  h0_upper  h2_lower  h2_upper  h3_lower  h3_upper
## 0.9646453 1.0363856 0.9647832 1.0362555 0.9647474 1.0362913 0.9647116 1.0363271
```

```r
print(averages[9:12])
```

```
##    in_interval in_interval_h0 in_interval_h2 in_interval_h3
##          0.947          0.955          0.955          0.955
```

```r
print(averages[13:16])
```

```
##  ci_length  h0_length  h2_length  h3_length
## 0.07174025 0.07147228 0.07154384 0.07161554
```

In this case, all CIs are very close to 0.95. All CIs also have very simmilar lengths.

2e.

In both cases, h3 produces the interval that contains the true value most of the time, regardless of whether the variance is homoskedastic or heteroskedastic. Hence, it is the most reliable interval.