ECON 4101 Econometrics
CM19 and CM 20 Homework

Pranav Singh

April 3, 2017

```r
df <- read.xlsx("../../Data/wheat.xlsx", 1, colIndex = 1:3)
str(df)
```

```
## 'data.frame':    26 obs. of  3 variables:
##  $ qty  : num  198 140 162 166 160 ...
##  $ price: num  1.47 1.3 1.59 1.44 1.89 1.49 1.94 1.52 2.15 2.09 ...
##  $ trend: num  1 2 3 4 5 6 7 8 9 10 ...
```

Part 1: CM19

Problem 1

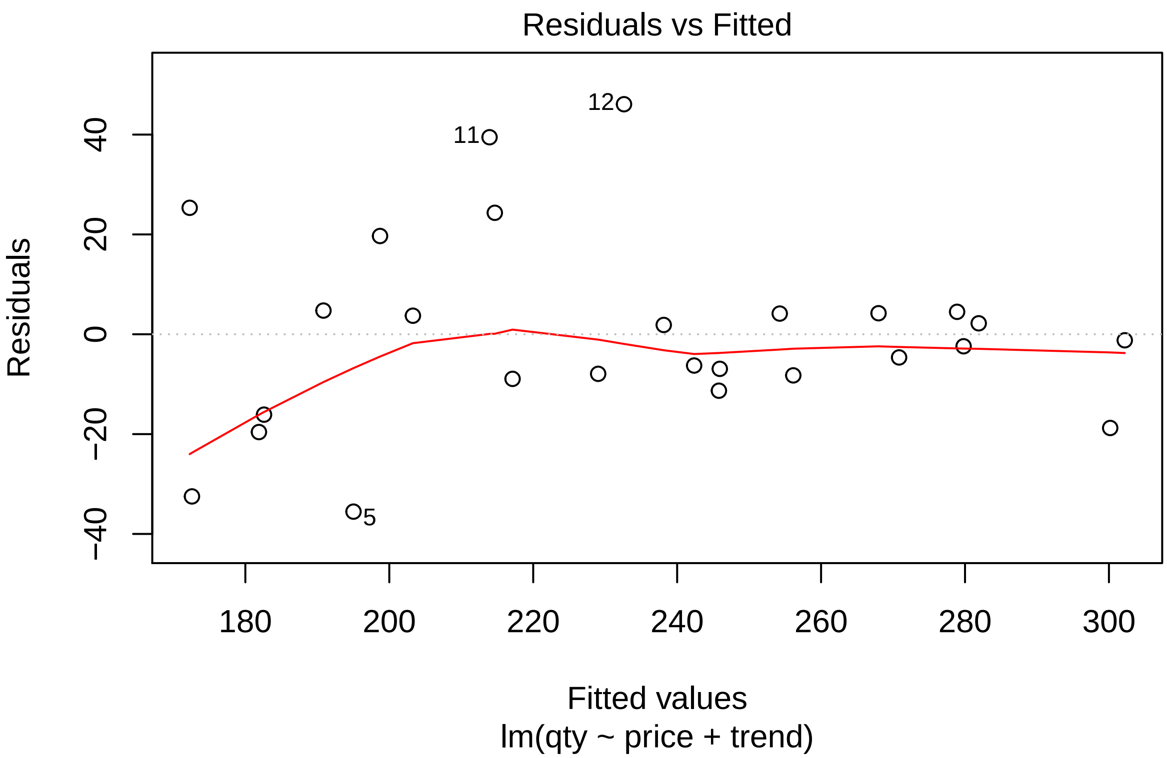
```r
mod1 <- lm(qty ~ price + trend, df)
summary(mod1)
```

```
##
## Call:
## lm(formula = qty ~ price + trend, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -35.528  -8.758  -1.802   4.430  46.083
##
## Coefficients:
##             Estimate Std. Error t value   Pr(>|t|)
## (Intercept)  139.901     23.218   6.026 0.00000382 ***
## price         19.541     17.415   1.122     0.2734
## trend          3.639      1.418   2.567     0.0172 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.97 on 23 degrees of freedom
## Multiple R-squared:  0.8089, Adjusted R-squared:  0.7923
## F-statistic: 48.67 on 2 and 23 DF,  p-value: 0.000000005431
```

The error term represents the cumulative effects of all omitted variables on the dependent variable, which is wheat supply in our case. So yes, it accounts for weather and all other factors besides the included price and technological trend variables.

Problem 2

```r
plot(mod1, 1)
```



Problem 3

```r
gqtest(qty ~ price + trend, order.by = -df$trend, data = df)
```

```
##
##  Goldfeld-Quandt test
##
## data:  qty ~ price + trend
## GQ = 11.109, df1 = 10, df2 = 10, p-value = 0.0003642
## alternative hypothesis: variance increases from segment 1 to 2
```

Problem 4

```r
se.uncorrected <- confint(mod1)
se.uncorrected
```

```
##                   2.5 %     97.5 %
## (Intercept)  91.8716477 187.930220
## price       -16.4852003  55.566203
## trend         0.7064496   6.571717
```

```r
# se.corrected <- mod1$coefficients + qt(.975,
# df=mod1$df.residual)*sqrt(diag(vcovHC(mod1, 'HC0'))) %*% t(c(-1,1));
# se.corrected
se.robust <- coefci(mod1, vcov = vcovHC(mod1, "HC0"))
se.robust
```

```
##                   2.5 %     97.5 %
```

```
## (Intercept)  89.0517303 190.750138
## price       -20.8044941  59.885497
## trend         0.3159286   6.962238
```

```
coeftest(mod1, vcov = vcovHC(mod1, "HC0"))
```

```
##
## t test of coefficients:
##
##             Estimate Std. Error t value    Pr(>|t|)
## (Intercept) 139.9009    24.5808  5.6915 0.000008539 ***
## price        19.5405    19.5030  1.0019     0.32681
## trend         3.6391     1.6064  2.2653     0.03322 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see that the confidence intervals using robust standard errors are centered around the same means but are wider than the one using uncorrected standard errors.

Part 2: CM20

Problem 1

Approach 1: Weighted Least Squares

```
df1 <- df[1:13, ]
df2 <- df[14:26, ]

m1 <- lm(qty ~ price + trend, df1)
e1.sd <- sd(m1$residuals)
m2 <- lm(qty ~ price + trend, df2)
e2.sd <- sd(m2$residuals)

w <- rep(c(1/e1.sd, 1/e2.sd), each = 13)
mod.wle <- lm(qty ~ price + trend, df, weights = w)
summary(mod.wle)
```

```
##
## Call:
## lm(formula = qty ~ price + trend, data = df, weights = w)
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -7.4751 -2.1070  0.3799  2.3962  9.6890
##
## Coefficients:
##             Estimate Std. Error t value     Pr(>|t|)
## (Intercept)  138.905     17.143   8.103 0.0000000344 ***
## price         20.974     12.538   1.673      0.10790
## trend          3.380      1.063   3.180      0.00418 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.471 on 23 degrees of freedom
```

```
## Multiple R-squared:  0.8564, Adjusted R-squared:  0.8439
## F-statistic: 68.57 on 2 and 23 DF,  p-value: 0.0000000002033
```

```r
confint(mod.wle)
```

```
##                  2.5 %     97.5 %
## (Intercept) 103.442323 174.368234
## price        -4.961831  46.909894
## trend         1.180860   5.578848
```

Approach 2: Feasible General Least Squares

```r
mod1 <- lm(qty ~ price + trend, df)
e <- mod1$residuals
le2 <- log(e^2)
mod2 <- lm(le2 ~ price + trend, df)
ghat <- mod2$fitted.values
hhat <- exp(ghat)
w <- 1/sqrt(hhat)
mod.fgls <- lm(qty ~ price + trend, df, weights = w)
summary(mod.fgls)
```

```
##
## Call:
## lm(formula = qty ~ price + trend, data = df, weights = w)
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -9.9768 -3.2783  0.0922  2.5084 14.4108
##
## Coefficients:
##             Estimate Std. Error t value     Pr(>|t|)
## (Intercept)  143.740     17.136   8.388 0.0000000188 ***
## price         21.805     12.441   1.753       0.0930 .
## trend          3.042      1.099   2.768       0.0109 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.829 on 23 degrees of freedom
## Multiple R-squared:  0.8319, Adjusted R-squared:  0.8173
## F-statistic:  56.9 on 2 and 23 DF,  p-value: 0.000000001244
```

```r
confint(mod.fgls)
```

```
##                   2.5 %     97.5 %
## (Intercept) 108.2912880 179.188518
## price        -3.9317757  47.542208
## trend         0.7685339   5.314862
```

The confidence intervals for WLS and GLS are not only tighter in bounds than those from Part 1, but they are also centered about different means. That is, whereas we only played around with the magnitude of the parameter estimates in Part 1 but not the paramter estimates themselves, the WLS and GLS models yield entirely different parameter estimates and standard errors.