EconometricsPSet2

Anya Conti

10/23/2017

```
Q1 <- read.csv("/Users/Anya/Documents/Senior Year/Econometrics/PSet 2/dataQ1.csv")
Q3 <- read.csv("/Users/Anya/Documents/Senior Year/Econometrics/PSet 2/dataQ3.csv")


library(sandwich)
library(lmtest)
```

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

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

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

```
library(ggplot2)
library(plm)
```

```
## Loading required package: Formula
```

Question 1

a. The ordinary least squares estimates are shown below without heteroscedastic error.

```
Mod.NHE <- lm(y ~ x1 + x2, data = Q1)
summary(Mod.NHE)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2, data = Q1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.5606  -0.8098   0.7462   1.4269  24.0503
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.1904     0.9144   0.208    0.836
## x1            1.1311     0.9826   1.151    0.255
## x2            0.3768     0.4399   0.857    0.396
##
## Residual standard error: 6.378 on 47 degrees of freedom
## Multiple R-squared:  0.0379, Adjusted R-squared:  -0.003036
## F-statistic: 0.9258 on 2 and 47 DF,  p-value: 0.4033
```

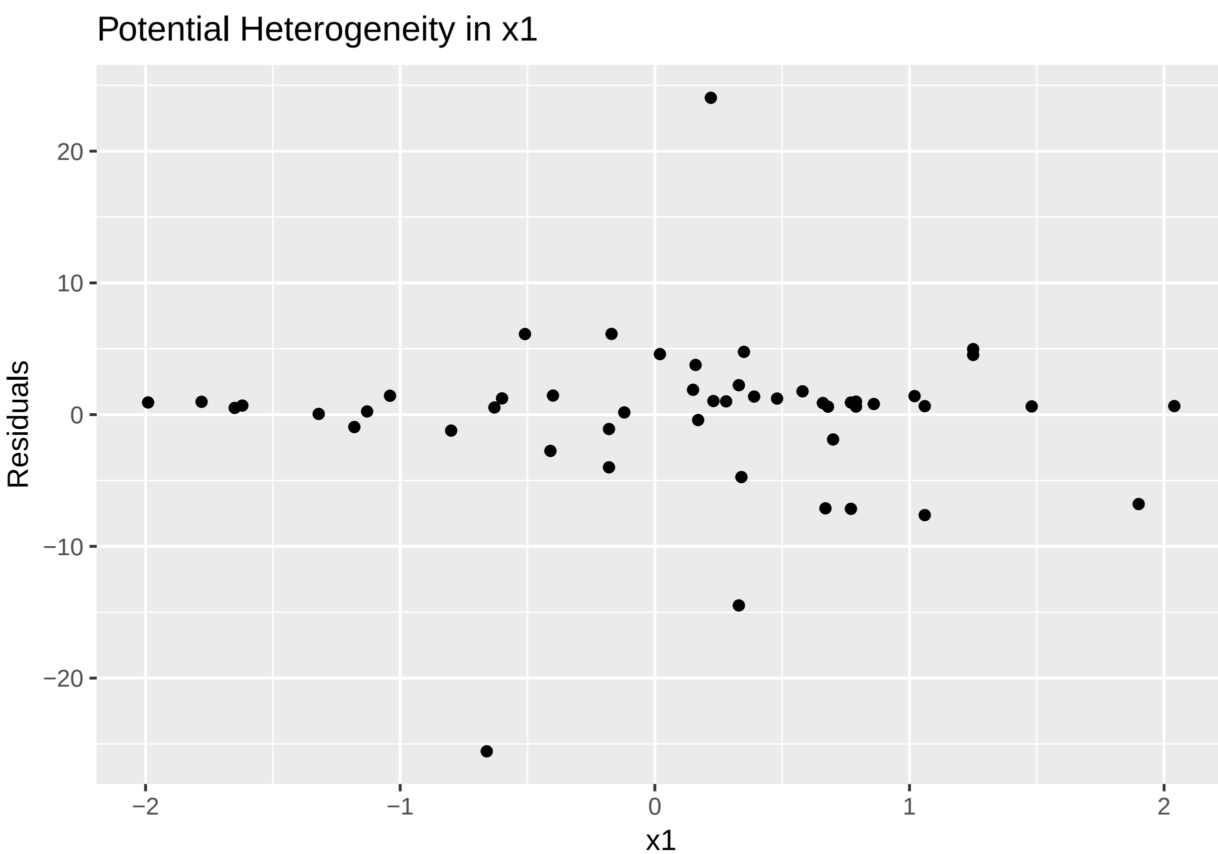
b. The ordinary least squares estimates are shown below with robust heteroscedastic error. The standard errors for the intercept and x_1 are lower, while the standard error for x_2 is higher.

```
RCovMat <- vcovHC(Mod.NHE, "HC1")
Mod.WHE <- coeftest(Mod.NHE, vcov. = RCovMat)
Mod.WHE
```

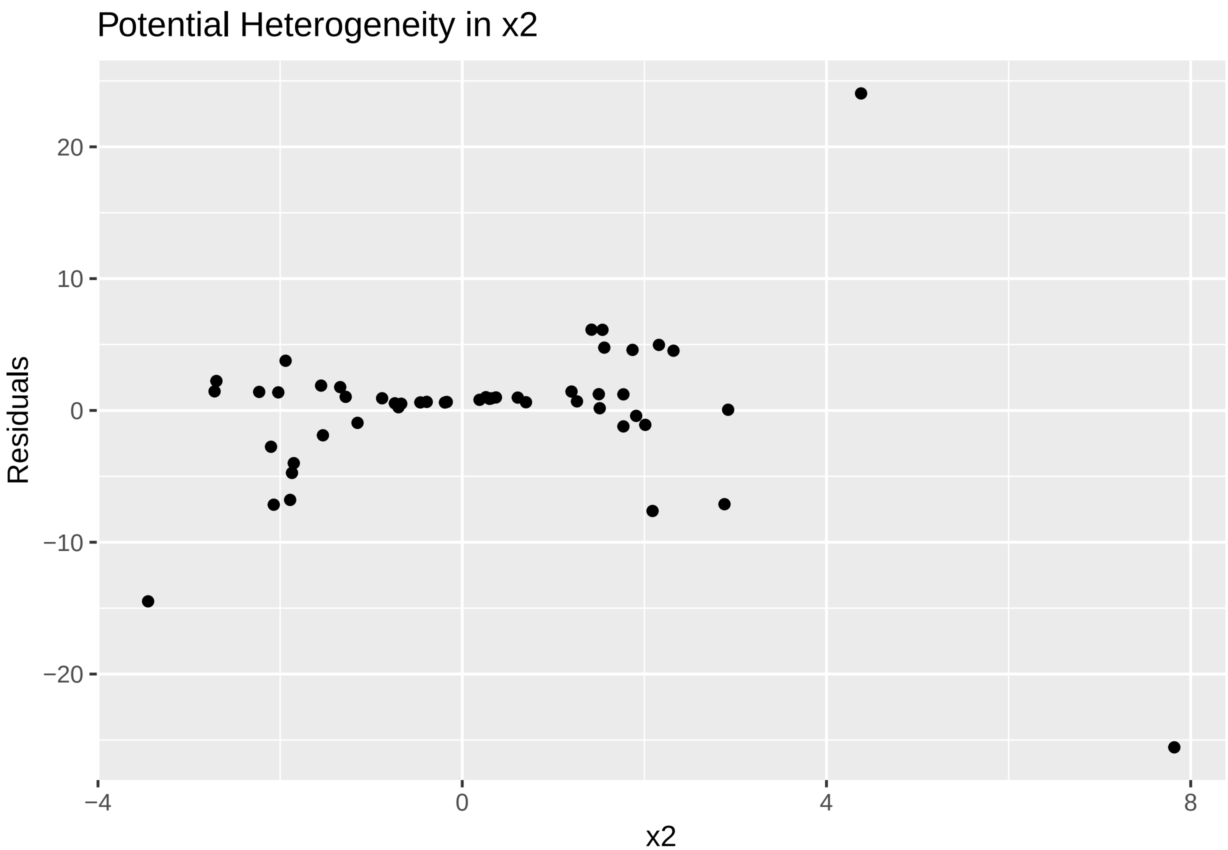
```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.19039    0.74704  0.2549  0.79994
## x1           1.13113    0.54808  2.0638  0.04458 *
## x2           0.37682    1.10342  0.3415  0.73424
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

c. The residuals plotted against both x_1 and x_2 are shown below. For the graph using x_1, there does not appear to any specific pattern to the residuals against x_1. However, for the graph using x_2, the residuals seem to be really small close to $x_2 = 0$, and to fan out on either end as x_2 gets smaller and as x_2 gets larger. This indicates the presence of heteroscedasticity in the data.

```r
Q1$resid <- Mod.NHE$residuals
ggplot(data = Q1) +
  geom_point(aes(x = x1, y = resid)) +
  ylab("Residuals") +
  ggtitle("Potential Heterogeneity in x1")
```




```r
ggplot(data = Q1) +
  geom_point(aes(x = x2, y = resid)) +
  ylab("Residuals") +
  ggtitle("Potential Heterogeneity in x2")
```



d. To test for heteroscedasticity, we need to use White's general test (which is equivalent to the Breusch-Pagan with all squares and products of each variable). The null hypothesis H_0 is that there is no heteroscedasticity (or specification error). The alternative hypothesis H_a is that there is heteroscedasticity (or specification error). The results are shown below. With a χ^2 statistic of 39.148, the p-value is 2.217×10^{-7}, and so we reject the null hypothesis H_0 in favor of the alternative hypothesis H_a that there is either heteroskedasticity or specifation bias.

```
bptest(Mod.NHE, ~x1*x2 + I(x1^2) + I(x2^2), data = Q1)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  Mod.NHE
## BP = 39.148, df = 5, p-value = 2.217e-07
```

e.

$$var[\epsilon_n|x_{n1}, x_{n2}] = \sigma^2 exp(\gamma_1 x_{n1} + \gamma_2 x_{n2})$$
$$ln(var[\epsilon_n|x_{n1}, x_{n2}]) = 2ln(\sigma) + \gamma_1 x_{n1} + \gamma_2 x_{n2}$$
$$var[\epsilon_n|x_{n1}, x_{n2}] \sim resid^2$$
$$ln(resid^2) \sim 2ln(\sigma) + \gamma_1 x_{n1} + \gamma_2 x_{n2}$$

```
Q1$residSQ <- (Q1$resid)^2
Q1$lnresidSQ <- log(Q1$residSQ)

LMweights <- lm(lnresidSQ ~ x1 + x2, data = Q1)

wt <- 1/(exp(LMweights$fitted.values))
```

```r
FGLS.mod <- glm(y ~ x1 + x2, data = Q1, weights = wt)

summary(FGLS.mod)
```

```
##
## Call:
## glm(formula = y ~ x1 + x2, data = Q1, weights = wt)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -10.9889   -0.8134    0.4856    1.4073    8.4540
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.1666     0.7198   0.231    0.818
## x1            0.7765     0.6388   1.215    0.230
## x2            0.8472     0.3633   2.332    0.024 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 9.213805)
##
##     Null deviance: 490.00  on 49  degrees of freedom
## Residual deviance: 433.05  on 47  degrees of freedom
## AIC: 308.15
##
## Number of Fisher Scoring iterations: 2
```

Question 2

$$y_n = \mu + \epsilon_n = \mu + (0)x_n + \epsilon_n$$
$$E[\epsilon_n|x_n] = 0$$
$$cov[\epsilon_m, \epsilon_n|x_m, x_n] = 0$$
$$var[\epsilon_n|x_n] = \sigma^2 x_n^2$$

a.

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

$$\hat{\beta}_{GLS} = (X_*'X_*)^{-1}X_*'y_*$$
$$\hat{\beta}_{GLS} = (X_*'X_*)^{-1}X_*'(X_*\beta + \epsilon_*)$$
$$\hat{\beta}_{GLS} = \beta + (X_*'X_*)^{-1}X_*'\epsilon_*$$
$$var(\hat{\beta}_{GLS}|X_*) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']$$
$$var(\hat{\beta}_{GLS}|X_*) = E[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon\epsilon'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}]$$
$$var(\hat{\beta}_{GLS}|X_*) = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}E[\epsilon\epsilon']\Omega^{-1}X(X'\Omega^{-1}X)^{-1}$$
$$var(\hat{\beta}_{GLS}|X_*) = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Omega\sigma^2\Omega^{-1}X(X'\Omega^{-1}X)^{-1}$$
$$var(\hat{\beta}_{GLS}|X_*) = \sigma^2(X'\Omega^{-1}X)^{-1}$$

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$
$$\Omega = x_n^2$$
$$\hat{\beta}_{GLS} = \frac{\sum \frac{y_n}{x_n^2}}{\sum \frac{1}{x_n^2}}$$

$$var(\hat{\beta}_{GLS}|X_*) = \sigma^2(X'\Omega^{-1}X)^{-1}$$
$$\Omega = x_n^2$$
$$var(\hat{\beta}_{GLS}|X_*) = \frac{\sigma^2}{\sum \frac{1}{x_n^2}}$$

b.

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y$$

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'(X\beta + \epsilon)$$
$$\hat{\beta}_{OLS} = \beta + (X'X)^{-1}X'\epsilon$$
$$var(\hat{\beta}_{OLS}|X) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']$$
$$var(\hat{\beta}_{OLS}|X) = E[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}]$$
$$var(\hat{\beta}_{OLS}|X) = (X'X)^{-1}X'E[\epsilon\epsilon']X(X'X)^{-1}$$
$$var(\hat{\beta}_{OLS}|X) = (X'X)^{-1}X'\sigma^2\Omega^{-1}X(X'X)^{-1}$$
$$var(\hat{\beta}_{OLS}|X) = \sigma^2(X'X)^{-1}(X'\Omega^{-1}X)(X'X)^{-1}$$

$$\frac{\partial SSE}{\partial \hat{\mu}} = \frac{\partial[\sum_{i=1}^{n}(y_i - \hat{\mu})^2]}{\partial \hat{\mu}}$$
$$\frac{\partial SSE}{\partial \hat{\mu}} = \sum_{i=1}^{n} -2(y_i - \hat{\mu})$$
$$\frac{\partial SSE}{\partial \hat{\mu}} = -2(\sum_{i=1}^{n} y_i - n\hat{\mu})$$
$$0 = -2(\sum_{i=1}^{n} y_i - n\hat{\mu})$$
$$0 = \sum_{i=1}^{n} y_i - n\hat{\mu}$$
$$n\hat{\mu} = \sum_{i=1}^{n} y_i$$
$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i$$
$$\hat{\mu} = \bar{y}$$

$$var(\hat{\mu}) = var(\frac{1}{n}\sum_{i=1}^{n} y_i)$$

$$var(\hat{\mu}) = \frac{1}{n^2}var(\sum_{i=1}^{n} \epsilon_i)$$

$$var(\hat{\mu}) = \frac{\sigma^2}{n^2}\sum_{i=1}^{n} x_i$$

c.

$$eff(\hat{\beta}_{OLS}, \hat{\beta}_{GLS}) = \frac{var(\hat{\beta}_{GLS})}{var(\hat{\beta}_{OLS})}$$

$$eff(\hat{\beta}_{OLS}, \hat{\beta}_{GLS}) = \frac{\sigma^2}{\sum \frac{1}{x_n^2}}(\frac{\sigma^2}{n^2}\sum_{i=1}^{n} x_i)^{-1}$$

$$eff(\hat{\beta}_{OLS}, \hat{\beta}_{GLS}) = \frac{\sigma^2}{\sum \frac{1}{x_n^2}}\frac{n^2}{\sigma^2 \sum x_n}$$

$$eff(\hat{\beta}_{OLS}, \hat{\beta}_{GLS}) = \frac{1}{\sum \frac{1}{x_n^2}}\frac{n^2}{\sum x_n}$$

$$eff(\hat{\beta}_{OLS}, \hat{\beta}_{GLS}) = \frac{n^2}{\sum \frac{1}{x_n^2} \times \sum x_n}$$

Since x_n represents the intercept, it is just a column of 1's in this case.

$$eff(\hat{\beta}_{OLS}, \hat{\beta}_{GLS}) = \frac{n^2}{\sum_n 1 \times \sum_n 1}$$

$$eff(\hat{\beta}_{OLS}, \hat{\beta}_{GLS}) = \frac{n^2}{n \times n}$$

$$eff(\hat{\beta}_{OLS}, \hat{\beta}_{GLS}) = 1$$

As such, the two estimators are equally as efficient.

Question 3

a. The estimates for the pooled effects model and the OLS model are shown below. They are both the same.

```
Mod.Pooled <- plm(y ~ x, data = Q3, model = "pooling")
Mod.LM <- lm(y~x, data = Q3)

summary(Mod.Pooled)
```

```
## Pooling Model
##
## Call:
## plm(formula = y ~ x, data = Q3, model = "pooling")
##
## Unbalanced Panel: n=30, T=1-1, N=30
##
## Residuals :
```

```
##    Min. 1st Qu.  Median 3rd Qu.    Max.
##  -3.600  -1.520  -0.195   1.060   3.390
##
## Coefficients :
##              Estimate Std. Error t-value Pr(>|t|)
## (Intercept) -0.747476   0.955953 -0.7819   0.4408
## x            1.058959   0.058656 18.0538   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1525.3
## Residual Sum of Squares: 120.67
## R-Squared:      0.92089
## Adj. R-Squared: 0.91807
## F-statistic: 325.94 on 1 and 28 DF, p-value: < 2.22e-16
```

```r
summary(Mod.LM)
```

```
##
## Call:
## lm(formula = y ~ x, data = Q3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5993 -1.5206 -0.1949  1.0637  3.3891
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.74748    0.95595  -0.782    0.441
## x            1.05896    0.05866  18.054   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.076 on 28 degrees of freedom
## Multiple R-squared:  0.9209, Adjusted R-squared:  0.9181
## F-statistic: 325.9 on 1 and 28 DF,  p-value: < 2.2e-16
```

b. The estimates for the fixed effects model are shown below.

```r
Mod.FE <- plm(y~x, data = Q3, model = "within", index = c("i"))
summary(Mod.FE)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = y ~ x, data = Q3, model = "within", index = c("i"))
##
## Balanced Panel: n=3, T=10, N=30
##
## Residuals :
##    Min. 1st Qu.  Median 3rd Qu.    Max.
##  -3.580  -1.090   0.228   1.230   2.350
##
## Coefficients :
##   Estimate Std. Error t-value  Pr(>|t|)
## x 1.102192   0.050719  21.732 < 2.2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1517.4
## Residual Sum of Squares: 79.183
## R-Squared:      0.94782
## Adj. R-Squared: 0.9418
## F-statistic: 472.258 on 1 and 26 DF, p-value: < 2.22e-16
```

Stata includes an intercept estimate, which is the mean of the intercept estimate for each individual group. As such, we can calculate that as is done below.

```r
mean(fixef(Mod.FE))
```

```
## [1] -1.394325
```

c. The estimates for the random effects model are shown below. This uses the Swamy-Arora method of estimating variance components. There seem to be many different methods for estimating this.

```r
Mod.RE <- plm(y~x, data = Q3, model = "random", effect = "time", index = c("i", "t"), random.method = "s
summary(Mod.RE)
```

```
## Oneway (time) effect Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = y ~ x, data = Q3, effect = "time", model = "random",
##     random.method = "swar", index = c("i", "t"))
##
## Balanced Panel: n=3, T=10, N=30
##
## Effects:
```

```
## Warning in sqrt(sigma2): NaNs produced
```

```
##                  var std.dev  share
## idiosyncratic  4.785   2.188  1.091
## time          -0.399      NA -0.091
## theta:  -0.1548
##
## Residuals :
##    Min. 1st Qu.  Median 3rd Qu.    Max.
## -3.7800 -1.5000 -0.0619  1.1300  3.6400
##
## Coefficients :
##              Estimate Std. Error t-value Pr(>|t|)
## (Intercept) -0.809666   0.918454 -0.8816   0.3855
## x            1.063115   0.056993 18.6536   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1751.1
## Residual Sum of Squares: 130.42
## R-Squared:      0.92552
## Adj. R-Squared: 0.92286
## F-statistic: 347.956 on 1 and 28 DF, p-value: < 2.22e-16
```

d.

Based on the information given with these particular results, I think Hausman would still be useful. There does appear to be some variation between the groups. As such, we can run the test as shown below, which results in rejecting the null hypothesis H_0 that both OLS and GLS are consistent in favor of the alternative hypothesis H_a that OLS is consistent and GLS is not.

```
phtest(Mod.FE, Mod.RE)
```

```
##
##  Hausman Test
##
## data:  y ~ x
## chisq = 2.2595, df = 1, p-value = 0.1328
## alternative hypothesis: one model is inconsistent
```