Chapter 5 Exercise 4

Allen Church

10/21/2019

Load data

```
require(knitr)
require(haven)
require(AER)

opts_chunk$set(echo = TRUE)
options(digits = 3)

#Add working directory, assign data to lab variable
load("/Users/allenchurch/Ch5_Exercise4_Speeding_tickets.RData")
View(dta)
```

4a. Estimate bivariate OLS model in which ticket amount is a function of age. Is age statistically significant, is endogeneity possible?

```
ols1 <- lm(dta$Amount ~ dta$Age)
summary(ols1)
```

```
##
## Call:
## lm(formula = dta$Amount ~ dta$Age)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -123.2  -46.6   -5.9   32.6  600.2
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 131.7067     0.8865   148.6   <2e-16 ***
## dta$Age      -0.2893     0.0248   -11.7   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 56.1 on 31672 degrees of freedom
##   (36683 observations deleted due to missingness)
## Multiple R-squared:  0.00429,    Adjusted R-squared:  0.00425
## F-statistic:  136 on 1 and 31672 DF,  p-value: <2e-16
```

According to the results above, the Beta Hat One coefficient on age is -0.2893, which suggests that a 1 unit increase in age is associated with decrease of 0.2893 units in ticket amount. As the t statistic is 11.7, and the Beta Hat One coefficient on age is -0.2893, we can posit that these results are statistically significant. However, endogeneity is possible because other independent variables may influence whether age influences the ticket amount - namely race and gender in our data, as well as other observations that were not collected.

4b. Estimate the model form part a, also controlling for miles per hour over the speed limit. Explain what happens to the coefficient on age and why?

```r
ols2 <- lm(dta$Amount ~ dta$Age + dta$MPHover)
summary(ols2)
```

```
##
## Call:
## lm(formula = dta$Amount ~ dta$Age + dta$MPHover)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -308.76  -19.78    7.68   25.76  226.29
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.4939     0.9543    3.66  0.00025 ***
## dta$Age       0.0250     0.0176    1.42  0.15606
## dta$MPHover   6.8917     0.0387  178.13  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 39.7 on 31671 degrees of freedom
##   (36683 observations deleted due to missingness)
## Multiple R-squared:  0.503,  Adjusted R-squared:  0.503
## F-statistic: 1.6e+04 on 2 and 31671 DF,  p-value: <2e-16
```

After controlling for miles per hour over the speed limit, the coefficient on age increases - from -0.2893 to 0.0250. When adding MPHover to the model, the Beta Hat One estimate for age suggests that a one unit increase in age is associated with an increase of 0.0250 units in amount. This increase in the coefficient estimate suggests that the previous model was biased by leaving out MPHover, which clearly affected both the dependent variable and is correlated with the independent variable.

3c. Suppose we had only the first 1000 observations. Estimate model from part b and report what happens to standard errors and t stats when we have fewer observations.

```r
#Subset dta by first 1000 rows and assign to dta2
dta2 <- dta[1:1000,]

ols3 <- lm(dta2$Amount ~ dta2$Age + dta2$MPHover)
summary(ols3)
```

```
##
## Call:
## lm(formula = dta2$Amount ~ dta2$Age + dta2$MPHover)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -173.60   -4.62    4.41   24.90  102.57
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.780      8.959   -0.20     0.84
## dta2$Age        0.181      0.192    0.95     0.35
## dta2$MPHover    6.857      0.342   20.03   <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 42.9 on 336 degrees of freedom
##   (661 observations deleted due to missingness)
## Multiple R-squared:  0.545,  Adjusted R-squared:  0.542
## F-statistic:  201 on 2 and 336 DF,  p-value: <2e-16
```

The standard error on age increase from 0.0176 to 0.192. The standard error on MPHover increases from 0.0387 to 0.342. For age, the t stats decrease from 1.42 to 0.95. For MPHover, the t stats decrease from 178.13 to 20.03. As seen in the results above, a lower number of observations will increase the standard error, while decreasing the t stats.