ARE 212 Problem Set 3

Aline Abayo, Eleanor Adachi, Anna Cheyette, Karla Neri, and Stephen Stack

2024-02-16

```r
# Comment out after installing
# install.packages("pacman")

options(scipen = 999)

# Load packages
library(pacman)
# p_load(dplyr, haven, readr, knitr, psych, ggplot2,stats4, stargazer, lmSupport,
#        magrittr, qwraps2, Jmisc , fastDummies)

# Remove lmSupport, add tidyverse
p_load(dplyr, haven, readr, knitr, psych, ggplot2,stats4, stargazer, magrittr,
       qwraps2, Jmisc , fastDummies, tidyverse)

# get directory of current file
current_directory <-
  dirname(dirname(rstudioapi::getSourceEditorContext()$path))
```

Question 1

Read `pset3_2024.dta` *into R Please check for missing values (as in section 3).*

```r
# Load data
my_data_raw <- read_dta(file.path(current_directory, "data", "pset3_2024.dta"))
head(my_data_raw)
```

```
## # A tibble: 6 x 28
##    year country      co type    segment domestic firm     brand   loc          qu
##   <dbl> <dbl+lbl> <dbl> <chr>   <dbl+l>    <dbl> <dbl+lb> <dbl+l> <dbl+l>   <dbl>
## 1  1971 4 [Italy]    15 audi 1~ 4 [sta~        0 26 [VW]  2 [Aud~ 4 [Ger~   1016
## 2  1971 4 [Italy]    35 citroe~ 3 [int~        0  4 [Fia~ 4 [Cit~ 3 [Fra~   9659
## 3  1971 4 [Italy]    36 citroe~ 1 [sub~        0  4 [Fia~ 4 [Cit~ 3 [Fra~  12343
## 4  1971 4 [Italy]    61 fiat 1~ 1 [sub~        1  4 [Fia~ 7 [Fia~ 5 [Ita~ 100523
## 5  1971 4 [Italy]    64 fiat 1~ 1 [sub~        1  4 [Fia~ 7 [Fia~ 5 [Ita~ 197574
## 6  1971 4 [Italy]    71 ford e~ 2 [com~        0  5 [For~ 8 [For~ 4 [Ger~  45688
## # i 18 more variables: pr <dbl>, princ <dbl>, price <dbl>, horsepower <dbl>,
## #   fuel <dbl>, width <dbl>, height <dbl>, weight <dbl>, pop <dbl>, ngdp <dbl>,
## #   ngdpe <dbl>, country1 <dbl>, country2 <dbl>, country3 <dbl>,
## #   country4 <dbl>, country5 <dbl>, yearsquared <dbl>, luxury <dbl>
```

```r
# Check for missing values
my_data <- my_data_raw %>% drop_na()
```

```r
# Check if any missing values removed
all.equal(my_data, my_data_raw)
```

```
## [1] TRUE
```

Question 2

Get the summary statistics for price: sample mean, standard deviation, minimum and maximum. Construct a 99% confidence interval for the sample average of price (car price in thousands of Euros).

```r
# Get summary statistics for price
describe(my_data$price)
```

```
##    vars  n  mean   sd median trimmed  mad  min   max range skew kurtosis   se
## X1    1 57 20.77 8.54  18.37   19.67 6.37 8.13 48.08 39.95 1.23     1.28 1.13
```

```r
# Create summary table
summary_maker <-
  list("Price" =
         list("min" = ~ min(my_data$price),
              "max" = ~ max(my_data$price),
              "mean (sd)" = ~ qwraps2::mean_sd(my_data$price)))

whole <- summary_table(my_data, summary_maker)
whole
```

	my_data (N = 57)
Price	
min	8.13041400909424
max	48.0754890441895
mean (sd)	20.77 ± 8.54

```r
# Construct a 99% confidence interval for the sample average of price
xbar <- mean(my_data$price)
n <- nrow(my_data)
se_xbar <- sd(my_data$price)/sqrt(n)
cnn <- qt(0.005,n-1,lower.tail=FALSE)
bottom99ci <- xbar-cnn*se_xbar
top99ci <- xbar+cnn*se_xbar
bottom99ci
```

[1] 17.75517

```r
top99ci
```

[1] 23.79069

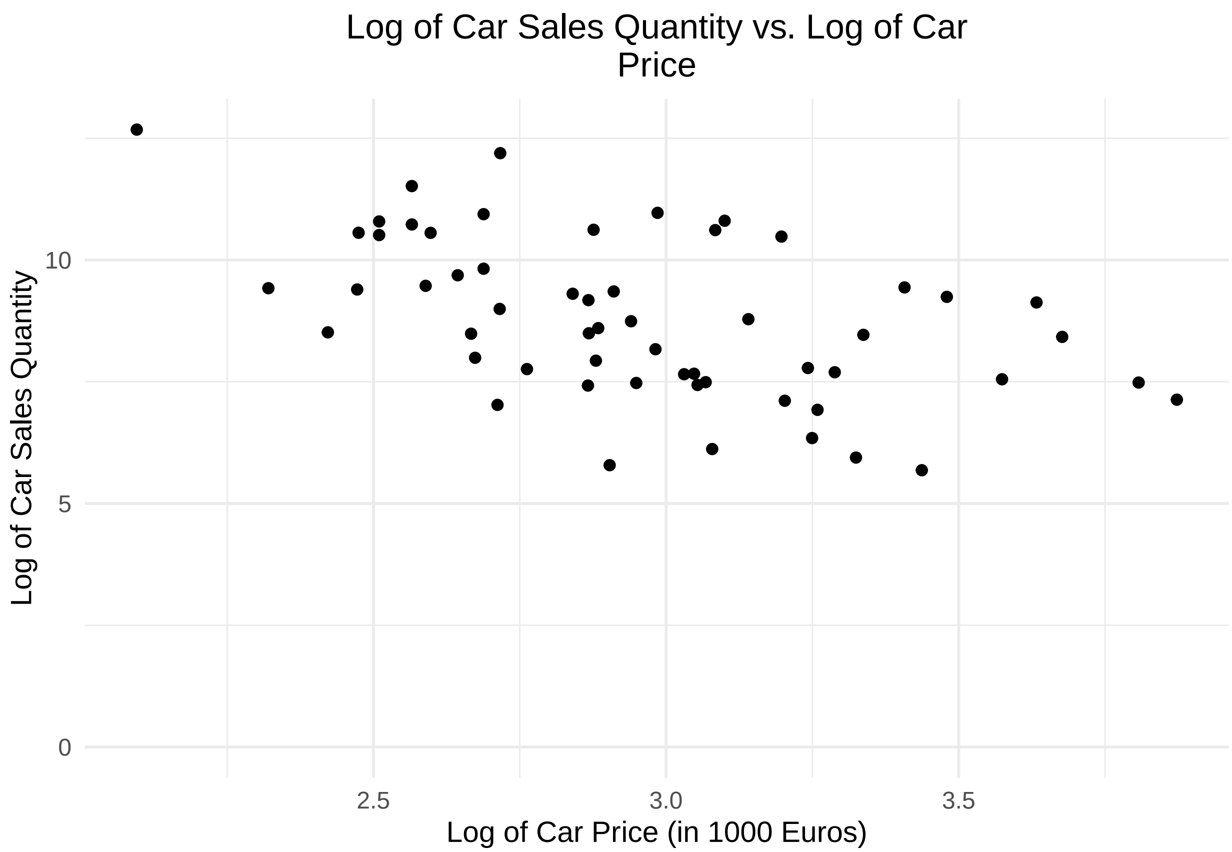
The 99% confidence interval for the sample mean of price is **17.7551718** to **23.7906866**.

Question 3

Create two new variables log of price and log of quantity, `lprice` *and* `lqu` *Create the scatter plot of the two variables* `lqu` *and* `lprice`*. What is the estimated OLS linear model slope associated with this scatter plot? Estimate a regression to answer this.*

```r
# Create lprice and lqu
my_data$lprice <- log(my_data$price)
my_data$lqu <- log(my_data$qu)
```

```r
# create lprice-lqu scatterplot
X3 <- my_data$lprice
y3 <- my_data$lqu
lscatter <- ggplot() +
  geom_point(aes(x = X3, y = y3)) +
  labs(x = "Log of Car Price (in 1000 Euros)",
       y = "Log of Car Sales Quantity",
       title = str_wrap(
         "Log of Car Sales Quantity vs. Log of Car Price",
         40)) +
  ylim(0,max(y3)) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
lscatter
```




```r
# estimate slope using OLS
ols_estimator <-function(y,X){
  # Sample size
  n <- length(y)
  # degrees of freedom
  if (is.null(dim(X))) {
    df <- n - 1
    } else {
```

```r
      df <- nrow(X) - ncol(X)
      }
  # Find coefficient vector
  b <- solve(t(X) %*% X) %*% t(X) %*% y
  # projection matrix of reg y on X
  P <- X%*%solve(t(X)%*%X)%*%t(X)
  # residual maker of reg y1 on x: M= I - P
  M <- diag(n)-P
  # sum of squared residuals, SSR=e'e
  e <- M%*%y
  SSR <- t(e)%*%e
  # construct demeaner
  i <- c(rep(1,n))
  M0 <- diag(n)-i%*%t(i)*(1/n)
  # demeaned y
  M0y <- M0%*%y
  # total sum of squares
  SST <- t(M0y)%*%M0y
  # calculate R squared
  Rsquared <- 1-(SSR/SST)
  # create predictions
  y_hat <- P%*%y
  #varcov matrix b
  s2 <- as.numeric(t(e)%*%e)/df
  vb <- s2*solve(t(X)%*%X)
  #std error b
  seb <- sqrt(diag(vb))
  return(list(b, Rsquared, y_hat, e, seb))
}

# Assume that we should use a constant
X3 <- cbind(1, my_data$lprice)
results3 <- ols_estimator(y3, X3)
b3 <- results3[[1]]
b3
```

```
##           [,1]
## [1,] 15.303260
## [2,] -2.203468
```

The estimated slope of this regression using OLS is **-2.2034678**.

Question 4

Regress lqu on fuel, luxury, domestic, and a constant, create the residuals elqu.

```r
# Regress lqu on fuel, luxury, domestic, and a constant
X4 <- cbind(1, my_data$fuel, my_data$luxury, my_data$domestic)
y4_lqu <- my_data$lqu
results4_lqu <- ols_estimator(y4_lqu, X4)

# create the residuals elqu
elqu <- results4_lqu[[4]]
```

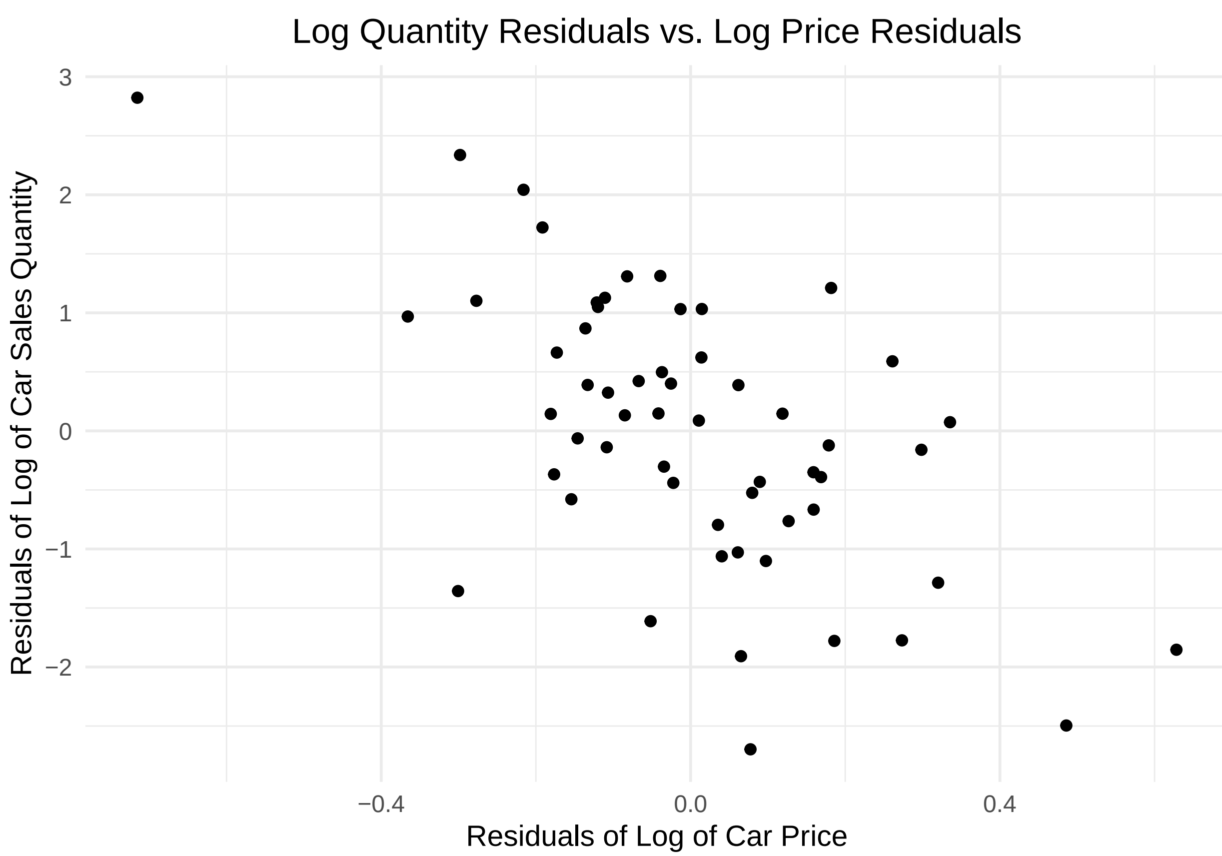
Regress lprice on fuel, luxury, domestic, and a constant, create the residuals elprice

```r
# Regress lprice on fuel, luxury, domestic, and a constant
y4_lprice <- my_data$lprice
results4_lprice <- ols_estimator(y4_lprice, X4)

# create the residuals elprice
elprice <- results4_lprice[[4]]
```

Scatter plot the residuals elqu on vertical axis and elprice on horizontal axis

```r
# Scatter plot the residuals elqu on vertical axis and elprice on horizontal axis
ggplot() +
  geom_point(aes(x = elprice, y = elqu)) +
  labs(x = "Residuals of Log of Car Price",
       y = "Residuals of Log of Car Sales Quantity",
       title = "Log Quantity Residuals vs. Log Price Residuals"
       ) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```



What is the estimated OLS slope associated with this scatter plot? Estimate a regression (no constant) to answer this and explain what theorem underlies the fact that this slope is the marginal effect of lprice on lqu in a regression that also features fuel, luxury, domestic, and a constant.

```r
# Sample size
X4c <- elprice
y4c <- elqu
n <- length(elqu)
df4c <- length(elqu) - ncol(X4) - 1
```

```r
# Find coefficient vector
b4c <- solve(t(X4c) %*% X4c) %*% t(X4c) %*% y4c
b4c
```

```
##           [,1]
## [1,] -3.335674
```

```r
# projection matrix of reg y on X
P <- X4c%*%solve(t(X4c)%*%X4c)%*%t(X4c)
# residual maker of reg y on x: M= I - P
M <- diag(n)-P
# sum of squared residuals, SSR=e'e
e <- M%*%y4c
s2 <- as.numeric(t(e)%*%e)/df4c
#varcov matrix b
vb <- s2*solve(t(X4c)%*%X4c)
#std error b
seb4c <- sqrt(diag(vb))
```

The OLS slope for the residuals `elqu` on the residuals `elprice`, **-3.3356739**, is equal to the coefficient of `lprice` in the regression of `lqu` on `lprice`, fuel, luxury, domestic, and a constant. This is a demonstration of the Frish-Waugh-Lovell Theorem.

Question 5

Why is the slope estimate in 3 not equal to the one in 4? Theoretically speaking, when would they be equal?

The slope estimates in 3 and 4 are not equal because there is some relationship between `lprice` and the other variables in 4. The estimates would be equal if the vectors were orthogonal, i.e., there was no relationship between `lprice` and the other variables.

In other words, there is omitted variable bias in the coefficient of `lprice` in question 3.

Question 6

Please interpret the OLS slope point estimate size, sign of the slope lprice estimate in 4.

The OLS slope for the residuals `elqu` on the residuals `elprice`, **-3.3356739**, means that for each 100% increase in `price` there is a ~333% decrease in `qu`. We know that it is a decrease because the slope is negative.

What is the pvalue for the estimated lprice coefficient? Use the stat tables for this. And then check with $pvalue_6 = 2 \cdot pt(abs(t_6), df)$ *, where t6 is the t stat value t6=-, and df are degrees of freedom*

```r
# pvalue for lprice coefficient
t6 <- b4c/seb4c
t6
```

```
##           [,1]
## [1,] -5.628982
```

The t statistic is **-5.6289824** and there are **52** degrees of freedom. Looking at the t-tables, this corresponds to a p-value of less than 0.000.

```r
# pvalue for lprice coefficient
pvalue6 <- 2*(1-pt(abs(t6),df4c))
pvalue6
```

```
##                  [,1]
```

```
## [1,] 0.0000007371973
```

The p-value is **0.0000007**.

Question 7

Can you reject that the marginal effect of lprice on lqu is -4 conditional on all else equal (fuel, luxury, domestic, and a constant)? Do five steps in Hypothesis Testing at the 5% significance level against a two-sided alternative. Get critical values from the relevant stats table.

$H_0 : \beta = -4$ $H_a : \beta \neq -4$

Remember, estimate of β is b, **-3.3356739** and standard error is **0.5925891**.

```r
# get t statistic
t7 <- (b4c - (-4))/seb4c
t7
```

```
##          [,1]
## [1,] 1.121057
```

Critical value of t for 5% significance and **52** degrees of freedom from table is between 2.000 and 2.021.

The t statistic is **1.1210569**, which is less than the critical value. Therefore, we cannot reject the null hypothesis.

Question 8

Estimate the sample data correlation of all these variables with each other: lqu, lprice, fuel, weight, luxury, domestic. Suppose the population model is given by

$lqu = \beta_0 + lprice\beta_1 + domestic\beta_2 + fuel\beta_3 + luxury\beta_4 + \epsilon (8.a)$

and you estimate the model

$lqu = \alpha_0 + lprice\alpha_1 + fuel\alpha_3 + luxury\alpha_4 + \epsilon (8.b)$

Based on the variables' correlation and without estimating any regression models, would the estimated coefficient for fuel in (8.b) have a negative or a positive bias? Explain briefly.

```r
# correlation between domestic and lqu
cor(my_data$domestic,my_data$lqu)
```

```
## [1] 0.5772312
```

```r
# correlation between fuel and domestic
cor(my_data$fuel,my_data$domestic)
```

```
## [1] 0.1531191
```

Omitting `domestic` from the model would cause omitted variable bias. Since `lqu` and `domestic` are positively correlated and `domestic` and `fuel` are positively correlated, then the coefficient for `fuel` in (8.b) would tend to be overestimated if `domestic` is omitted, i.e. the coefficient for `fuel` would have positive bias. Intuitively, increase in `lqu` from `domestic` would be falsely attributed to `fuel`.

Question 9

If I told you that research shows that advertising expenditures by car model are positively correlated with lprice and that when including advertising in addition to all factors in (8.b), the estimated weight coefficient does *not change at all. What does this imply about the sample correlation between advertising and weight of cars in the sample?*

The sample correlation between advertising expenditures and car weight is very weak.

Question 10

Suppose that research showed that the log of advertising is, on average, 5 times the log of price. Construct that advertising variable based on this fact and include it in a regression in addition to lprice and the other covariates in 8.b.

```
# Regress lqu on lprice, fuel, luxury, ladvertising and a constant
my_data$ladvertising <- 5*my_data$lprice
X10 <- cbind(1, my_data$lprice, my_data$fuel, my_data$luxury, my_data$ladvertising)
y10 <- my_data$lqu

# Commented out to knit document because returns error
#results10 <- ols_estimator(y10, X10)
```

Explain what happened.

This returns an error:

Error in solve.default(t(X) %*% X) : system is computationally singular: reciprocal condition number = 5.09482e-34

Since one of the explanatory variables is a linear combination of another explanatory variable, this violates one of the conditions of OLS.

Question 11

Please estimate a specification that allows you to test the following. Research shows that luxury goods have a different price elasticity than non-luxury goods. The null hypothesis is that the marginal effect in lprice on lqu does not differ by luxury classification of the car. Write out the regression model that allows you to estimate and perform a hypothesis test for this null. Do the five steps in hypothesis testing at the 5% significance level. What do you conclude?

```
# Regress lqu on lprice, luxury, interaction term, and a constant
reg11 <- lm(lqu~lprice+domestic+fuel+luxury+lprice*luxury, my_data)
summary(reg11)
```

```
##
## Call:
## lm(formula = lqu ~ lprice + domestic + fuel + luxury + lprice *
##     luxury, data = my_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.34137 -0.57374 -0.05842  0.51199  1.52226
##
## Coefficients:
##               Estimate Std. Error t value             Pr(>|t|)
## (Intercept)   17.18336    1.12682  15.249 < 0.0000000000000002 ***
## lprice        -3.68249    0.55393  -6.648      0.0000000195856 ***
## domestic       2.43494    0.27526   8.846      0.0000000000071 ***
## fuel           0.20469    0.09467   2.162              0.03532 *
```

```
## luxury         67.22358   20.07201   3.349              0.00153 **
## lprice:luxury -17.44827    5.31867  -3.281              0.00187 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.874 on 51 degrees of freedom
## Multiple R-squared:  0.739,  Adjusted R-squared:  0.7134
## F-statistic: 28.88 on 5 and 51 DF,  p-value: 0.00000000000009008
```

```r
# Find p-value
pval11 <- summary(reg11)$coefficients[4, 4]
pval11
```

```
## [1] 0.03531572
```

Null hypothesis: H_0: $\beta_3 = 0$

Alternative hypothesis: H_a: $\beta_3 \neq 0$

Since the p-value of the null hypothesis is **0.0353157** which is less than 0.05, we can reject the null hypothesis.

We conclude that the marginal effect of `lprice` on `lqu` differs between luxury and non-luxury cars.

Question 12

Regress lqu on a constant, fuel, lprice, luxury, domestic, weight. (eq 12) Test the joint hypothesis that $\beta_{domestic} = 1.5$ *and* $\beta_{fuel} = 60\beta_{weight}$ *at the 1 percent significance level.*

```r
# Regress lqu on a constant, fuel, lprice, luxury, domestic, weight
X12 <- cbind(1, my_data$fuel, my_data$lprice, my_data$luxury, my_data$domestic,
             my_data$weight)
y12 <- my_data$lqu
results12 <- ols_estimator(y12, X12)
b12 <- results12[[1]]
e12 <- results12[[4]]
df12 <- nrow(X12) - ncol(X12)
s2_12 <- as.numeric(t(e12)%*%e12)/df12
```

a) Perform a Fit based test and then also a Wald test.

```r
# Test joint hypothesis that beta_domestic=1.5 and beta_fuel=60*beta_weight
# F = (ssrr-ssr)/(2*(s2))
# ssr restricted
X12r <- cbind(1, (60*my_data$fuel + my_data$weight), my_data$lprice,
              my_data$luxury)
y12r <- my_data$lqu - 1.5*my_data$domestic
results12r <- ols_estimator(y12r, X12r)
e12r <- results12r[[4]]
ssrr <- as.numeric(t(e12r)%*%e12r)

# ssr unrestricted
ssru <- as.numeric(t(e12)%*%e12)
# F test
F_fitBased<-(ssrr-ssru)/(2*(s2_12))
F_fitBased
```

```
## [1] 4.131466
```

```r
# Test joint hypothesis that beta_domestic=1.5 and beta_fuel=60*beta_weight
Rr1 <- c(0, 0, 0, 0, 1, 0)
Rr2 <- c(0, 1, 0, 0, 0, -60)
R <- t(cbind(Rr1,Rr2))
q <- c(1.5, 0)
VRbq <- s2_12* R %*% solve(t(X12) %*% X12) %*% t(R)
Fw <- (t(R %*% b12-q) %*% solve(VRbq) %*% (R %*% b12-q))/2 # J = 2
Fw
```

```
##          [,1]
## [1,] 4.131466
```

b) Are the values of the fit and Wald test statistics equal?

Yes, the F-statistic values of the fit-based F test and the Wald test are equal.

Question 13

Without running any additional regressions and starting from the baseline regression in reg question 8.a,

a) Will omitting fuel create an OVB problem for the OLS estimator of lprice?

Population model given by (8.a):

$lqu = \beta_0 + lprice\beta_1 + domestic\beta_2 + fuel\beta_3 + luxury\beta_4 + \epsilon$

Sample model if fuel omitted:

$l\hat{q}u = \tilde{\beta_0} + lprice\tilde{\beta_1} + domestic\tilde{\beta_2} + luxury\tilde{\beta_4} + \tilde{e}$

```r
# correlation between fuel and lqu
cor(my_data$fuel,my_data$lqu)
```

```
## [1] -0.3030351
```

Increasing fuel efficiency tends to decrease quantity sold.

```r
# correlation between lprice and fuel
cor(my_data$lprice,my_data$fuel)
```

```
## [1] 0.7739287
```

```r
# correlation between domestic and fuel
cor(my_data$domestic,my_data$fuel)
```

```
## [1] 0.1531191
```

```r
# correlation between luxury and fuel
cor(my_data$luxury,my_data$fuel)
```

```
## [1] 0.3097369
```

Increasing price, increasing domestic (true = 1), and increasing luxury (true = 1) tends to increase fuel efficiency.

Since the coefficients have opposite signs, omitting fuel will lead to omitted variable bias and the bias will be negative. The coefficients for the other explanatory variables will tend to be underestimated.

b) Compute the variance inflated factor (VIF) for the variable height to be potentially also included into the (reg of question 8.a) model to explain the variation in lqu. Feel free to use the lm canned function to get what you need for the VIFj, for all j.

```r
# regress height on other explanatory variables
reg13 <- lm(height~lprice+domestic+fuel+luxury, my_data)
summary(reg13)
```

```
##
## Call:
## lm(formula = height ~ lprice + domestic + fuel + luxury, data = my_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.6176  -2.2278   0.1582   3.1342  14.6499
##
## Coefficients:
##             Estimate Std. Error t value            Pr(>|t|)
## (Intercept) 136.0579     8.0524  16.896 <0.0000000000000002 ***
## lprice        0.7625     3.8897   0.196               0.845
## domestic     -3.0639     1.9498  -1.571               0.122
## fuel          0.4377     0.6462   0.677               0.501
## luxury        1.3298     4.3586   0.305               0.762
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.252 on 52 degrees of freedom
## Multiple R-squared:  0.07552,    Adjusted R-squared:  0.004411
## F-statistic: 1.062 on 4 and 52 DF,  p-value: 0.3847
```

```r
rsq13 <- summary(reg13)$r.squared
vif13 <- 1/(1-rsq13)
vif13
```

```
## [1] 1.081694
```

The variance inflated factor (VIF) for height is **1.0816942**.

c) Will including this variable height with the others in (model in question 8.a) result in multicollinearity problems?

There is not strong evidence that including height in the model from 8.a will results in multicollinearity problems. The R-squared value is only 0.07552 and the VIF is **1.0816942**, which is only slightly more than 1.

Question 14

Suppose a car salesman told you that the conditional variance in the unobserved determinants of the log quantity (lqu) for luxury cars is three times the variance for nonluxury cars.

a. Which assumption no longer holds when we derive the statistical properties of the OLS estimators for the linear model in (reg 8.a) ?

Assumption 3–homoskedasticity and spherical disturbances–no longer holds.

b. Let the variance of the disturbance of log quantity for luxury=1 be 3 times the variance for luxury=0. In R create a matrix Omega, its inverse, and the positive definite matrix C such that the inverse of Omega = C C' as derived in lecture.

```r
# sort data by luxury
my_dataGLS <- my_data[order(my_data$luxury),]
head(my_dataGLS)
```

```
## # A tibble: 6 x 31
##    year country      co type    segment domestic firm     brand   loc          qu
##   <dbl> <dbl+lbl> <dbl> <chr>   <dbl+l>    <dbl> <dbl+lb> <dbl+l> <dbl+l>  <dbl>
## 1  1971 4 [Italy]    15 audi 1~ 4 [sta~        0 26 [VW]  2 [Aud~ 4 [Ger~   1016
## 2  1971 4 [Italy]    35 citroe~ 3 [int~        0  4 [Fia~ 4 [Cit~ 3 [Fra~   9659
## 3  1971 4 [Italy]    36 citroe~ 1 [sub~        0  4 [Fia~ 4 [Cit~ 3 [Fra~  12343
## 4  1971 4 [Italy]    61 fiat 1~ 1 [sub~        1  4 [Fia~ 7 [Fia~ 5 [Ita~ 100523
## 5  1971 4 [Italy]    64 fiat 1~ 1 [sub~        1  4 [Fia~ 7 [Fia~ 5 [Ita~ 197574
## 6  1971 4 [Italy]    71 ford e~ 2 [com~        0  5 [For~ 8 [For~ 4 [Ger~  45688
## # i 21 more variables: pr <dbl>, princ <dbl>, price <dbl>, horsepower <dbl>,
## #   fuel <dbl>, width <dbl>, height <dbl>, weight <dbl>, pop <dbl>, ngdp <dbl>,
## #   ngdpe <dbl>, country1 <dbl>, country2 <dbl>, country3 <dbl>,
## #   country4 <dbl>, country5 <dbl>, yearsquared <dbl>, luxury <dbl>,
## #   lprice <dbl>, lqu <dbl>, ladvertising <dbl>
```

```
# Count number of non-luxury and luxury
countlux <- sum(my_dataGLS$luxury == 1)
countnonlux <- nrow(my_dataGLS) - countlux

# Construct Omega: 54 non-luxury (luxury = 0), 3 luxury (luxury = 1)
topnonlux <- diag(countnonlux)
toplux <- matrix(data=0, nrow=countnonlux, ncol=countlux)
topOmega<-cbind(topnonlux, toplux)
bottomnonlux <- matrix(data=0, nrow=countlux, ncol=countnonlux)
bottomlux <- diag(countlux)*3 #identity times 3
bottomOmega <- cbind(bottomnonlux, bottomlux)
Omega <- rbind(topOmega,bottomOmega)

# Calculate inverse of Omega
invOmega <- solve(Omega)

# Calculate positive definite matrix C, sqrt of invOmega
invOm.eig <- eigen(invOmega)

D <- diag((invOm.eig$values))
V <- invOm.eig$vectors
C <- invOm.eig$vectors %*% diag(sqrt(invOm.eig$values)) %*% solve(invOm.eig$vectors)
```

c. Estimate the BLUE estimator in this setting of model 8.4 and test whether the marginal effect of lprice on lqu is equal to -1, at the 10 percent significance level, ceteris paribus.

Population model given by (8.a):

$lqu = \beta_0 + lprice\beta_1 + domestic\beta_2 + fuel\beta_3 + luxury\beta_4 + \epsilon$

```
# pre multiply all the variables by C, get ytilde and Xtilde
ytilde <- C %*% my_dataGLS$lqu
Xtilde <- C %*% cbind(1, my_dataGLS$lprice, my_dataGLS$domestic, my_dataGLS$fuel,
                      my_dataGLS$luxury)

bGLS <- solve(t(Xtilde)%*%Xtilde) %*% t(Xtilde)%*%ytilde
eGLS <- ytilde - Xtilde%*%bGLS

S2_GLS <- (t(eGLS) %*% eGLS)/(length(eGLS)-ncol(Xtilde))
S2_GLS <- as.numeric(S2_GLS)
```

```r
# variance of bGLS
VbGLS <- solve(t(Xtilde)%*%Xtilde) *  S2_GLS

# standard error of bGLS
seGLS <- sqrt(diag(VbGLS))
```

```r
t14 <- (bGLS[[2]] - (-1))/seGLS[[2]]
t14
```

```
## [1] -4.763834
```

```r
df14 <- nrow(Xtilde) - ncol(Xtilde)
tc14 <- qt(.05, df14, lower.tail=TRUE)
tc14
```

```
## [1] -1.674689
```

Null hypothesis: $H_0 : \beta_1 = -1$

Alternative hypothesis: $H_a : \beta_1 \neq -1$

The t score for the `lprice` coefficient, β_1, being equal to -1 is **-4.763834**. The critical value for the 10% significance level is **-1.6746892**. Since $|t| > |t_{critical}|$, we reject the null hypothesis and conclude that β_1 is not equal to -1.