Section A: Least Squares and Generalized Least Squares (30 points)

Problem 1) Consider the following linear model:

$Y = X\beta + \varepsilon$

$\varepsilon \sim N(0, \sigma_\varepsilon^2)$

With the following matrices:

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}, X = \begin{bmatrix} x_{11} & \dots & x_{1K} \\ \vdots & \ddots & \vdots \\ x_{N1} & \dots & x_{NK} \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}$$

The OLS estimator is given by (as derived in class):

$\hat{\beta} = (X'X)^{-1}X'Y$

Derive the mean and variance of $\hat{\beta}$ and confirm the OLS estimator is unbiased.

Problem 2) Consider the model:

$$Y_t = X_t\beta + \varepsilon_t \, , t = 1,2, \dots, N$$

where $\varepsilon_t \sim N(0, \sigma_t^2)$ and $E(\varepsilon_t, \varepsilon_s) = 0 \; for \; all \; t \neq s$. Suppose N=100 and $Var(\varepsilon_t) = \sigma_1^2$ for all $t \leq 50$ and $Var(\varepsilon_t) = \sigma_2^2$ for all $t > 50$. Suppose $R = \sigma_1^2/\sigma_2^2$ is known.

Give the GLS estimator of β.

Problem 3) Suppose:

$$Y_t = X_t\beta + \varepsilon_t \, , t = 1,2$$

where

(i) $X_t = 1 \, , t = 1,2$
(ii) $E(\varepsilon_t) = 0, \; E(\varepsilon_1^2) = E(\varepsilon_2^2) = \sigma^2$ and $E(\varepsilon_1, \varepsilon_2) = \sigma_{12}$

Let $\rho = \sigma_{12}/\sigma^2$ and note that ρ is the correlation coefficient between ε_1 and ε_2. Finally, let $\hat{\beta}$ be the OLS estimator of β.

a) Determine $VC(\hat{\beta}) = (X'X)^{-1}X'(\sigma^2\Omega)X(X'X)^{-1}$ and $VC(think) = \sigma^2(X'X)^{-1}$
b) Are there values of ρ for which $VC(think) > VC(\hat{\beta})$?

Hint: Write out the var-cov matrix of ε_t.

Section B: Market Segmentation (70 points)

Instructions: Review the variable description for the data file hh2pub.csv, which is the same dataset we have been using in class. Use R/RStudio for the remainder of this section. Attach your final R script to your completed homework submission.

Market Segmentation Based on Income and Car Ownership (40 points)

Problem 1) Estimate a linear regression model of HHTRIPS as a function of the following variables: urban location (URBAN), household size (HHSIZE) and number of adults (NUMADLT) and children ages 0-4 years old (youngchild).

Note: HHTRIPS = cnttdhh

Note: For URBAN, use the binary variable created in class.

Use the same specification to estimate the same model separately for households in the entire data set and for each of the following three income groups based on the ranges for HHFAMINC_IMP:

HHFAMINC_IMP $\leq$ \$49,999

\$49,999 $<$ HHFAMINC_IMP $\leq$ \$124,999

\$124,999 $<$ HHFAMINC_IMP

Report the results of the pooled model and the results for each income subsample in a table. Apply the market segmentation test to determine if these three groups have different trip generation relationships. Independent of the results of the test, report ant major difference among the estimated models.

Problem 2) Next, develop a linear regression model that accommodates the market segmentation based on income, but which is estimated on the entire data set. Show the equivalence of this estimation with the results of the regression estimations on the three subsamples obtained earlier.

Problem 3) Estimate a linear regression model of HHTRIPS as a function of the following variables: urban location (URBAN), household size (HHSIZE) and number of adults and children ages 0-4 years old (ADLT0TO4), for each of the three car ownership levels"

HHVEHCNT = 0

HHVEHCNT = 1

HHVEHCNT $\geq$ 2

Apply the market segmentation test to determine if these three groups have different trip generation relationships.

Problem 4) Create three variable NCAR0, NCAR1 and NCAR 2 that take a value of one for households with 0, 1 and 2+ cars, respectively and zero otherwise. Now estimate a model with the entire sample with the following variables: urban location (URBAN), household size (HHSIZE), number of adults and children ages 0-4 years old (ADLT0TO4) and NCAR1 and NCAR2.

Compare this model with the segmented models of the three car ownership groups. What do you conclude?