Problem Set 1: Basics of Estimation and Optimization
ECON 532
(Late submission is not allowed)

This problem set is designed to help you get your hands dirty with MATLAB/R/Python. **Read the problem set guideline in the syllabus before starting.**

1 OLS

In this problem you will be estimating OLS by minimizing the sum of squared errors rather than through matrix algebra.

Load airline.txt which contains on-time data for flights within the US. Our regression model is

$$\text{Delay}_i = \beta_0 + \beta_1 \text{Distance}_i + \beta_2 \text{Departure Delay}_i + \beta_{3-8} \text{Day of week fixed effects}_i + \varepsilon_i$$

Write a function that calculates the sum of squared errors

$$\sum_i (Y_i - X_i\beta)^2 \tag{1}$$

where $X_i = (\text{Distance}_i, \text{Departure Delay}_i, \text{Day of week fixed effects}_i)$ and $Y_i = \text{Delay}_i$ and search over β to minimize the objective function. In MATLAB you can use $fminsearch$. Compare your results to $\beta_{OLS} = (X'X)^{-1}X'Y$.

2 Maximum Likelihood

Now we will estimate a logit model using maximum likelihood. Let Y_i be the flight arrival time relative to the schedule. Generate a binary variable for a flight arriving more than 15 minutes late. We will use a logit model to study the probability of a flight arriving late. Let this probability be

$$P(Y_i > 15 | X_i; \beta) = \frac{\exp(\beta X_i)}{1 + \exp(\beta X_i)}$$

where X_i includes a constant, distance and departure delay.

Write a function to calculate the log likelihood of observing the flight arrival delays in the data:

$$L(\beta) = \sum \ln P(Y_i | X_i; \beta)$$

Search over the parameters β to maximize the log likelihood.

3 GMM

In this problem we will use GMM to estimate an instrumental variables model. Load the IV.mat matrix file. (If you don't plan to use MATLAB, you can google it: how to convert .mat file to .txt file without MATLAB?) X contains a vector of 3 observed variables and Z contains 4 instruments. Our model is

$$Y_i = \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \varepsilon_i$$

with 4 instruments $Z_1 ... Z_4$. Our moment conditions are

$$E[\varepsilon|Z] = 0$$

which implies the unconditional moment restriction

$$E[Z'(Y - X\beta)] = 0$$

Because we have more instruments than β's, we will need a weighting matrix W to minimize our objective.

Define $g_i(\beta) = Z_i'(Y_i - X_i\beta)$. Our objective function is

$$Q_n(\beta; W) = g_n(\beta)' W g_n(\beta)$$

where $g_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\beta)$.

To estimate an efficient weighting matrix W, use a two step estimator. Start with $W = I_4$ and minimize the objective function $Q_n(\beta; W)$ to obtain $\hat{\beta}$. Then set $\hat{W} = \Sigma^{-1}(\hat{\beta})$ where

$$\Sigma(\hat{\beta}) = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 z_i z_i'$$

The true data generating function was

$$Y_i = 2X_{i,1} + X_{i,2} + 4X_{i,3} + X_{i,4} + \varepsilon_i$$

Compare the point estimates and standard errors from the first-step and the second step estimator.

[Hint: The asymptotic distribution of a GMM estimator is given by

$$\sqrt{n}(\hat{\beta} - \beta_0) \to^d \mathcal{N}(0, (G'WG)^{-1} G'W\Omega W'G(G'WG)^{-1}),$$

where $\Omega = E[g_i(\beta_0) g_i(\beta_0')]$ and $G = E[\nabla_\beta g_i(\beta_0)]$.]