Econometrics II, first problem set.

Note: "Matlab" can be read as "Stata", or any other software package you like, as long as it allows you to minimize functions of parameters and data, with and without constraints, and compute gradients, Hessians, and Jacobians.

1. **Practice loading data**

   (a) Obtain the Nerlove data in .xls format from http://fhayashi.fc2web.com/datasets.htm and import it into Matlab

   (b) Obtain the Nerlove data in plain text format from my github page at https://github.com/mcreel/Econometrics/blob/master/Examples/Data/nerlove.data, and import it into Matlab. The Matlab command to read the data is `data = importdata('nerlove.data');`

2. **Unconstrained minimization**. Using the data you have loaded, consider the the Nerlove model:

$$\ln C = \beta + \beta_Q \ln Q + \beta_L \ln P_L + \beta_F \ln P_F + \beta_K \ln P_K + \epsilon$$

   Note that the data are given in levels, for all variables. You need to compute the logarithms.

   (a) Compute the OLS estimator using the analytic formula.

   (b) Compute the OLS estimator using Matlab's fminunc, to minimize the sum of squared residuals. The important part of this problem is to learn how to minimize a function that depends on both parameters and data.

3. **Constrained minimization.** For the Nerlove model, consider the joint restrictions of constant returns to scale ($\beta_Q = 1$) and homogeneity of degree one ($\beta_L + \beta_F + \beta_K = 1$)

   (a) compute the restricted OLS estimator, imposing both restrictions, using Matlab's fmincon, <u>numerically</u> minimizing the sum of squared errors subject to the restrictions.

   (b) test the restrictions, using the F test, using any software you like. If in doubt about what package to use, I recommend using GRETL (Tests->Linear Restrictions in the model window), which is quick and easy, or consider Julia and the script https://github.com/mcreel/Econometrics/blob/main/Examples/Restrictions/NerloveRestrictions.jl. Report the p value of the test statistic, and whether or not you reject the restrictions.

4. **More practice with data.** Get the Card data set from https://github.com/mcreel/Econometrics/blob/master/Examples/Data/card.gdt or https://github.com/mcreel/Econometrics/blob/master/Examples/Data/card.csv, and use it to replicate Table 2, column 1 of Card's 1993 working paper "Using geographic variation in college proximity to estimate the return to schooling", National Bureau of Economic Research working paper No. w4483, 1993. (available at http://www.nber.org/papers/w4483.pdf). Note: the regressors in this specification are education, experience, experience squared (divided by 100), and dummy variables for black, smsa, and south.

5. **Computing gradients, Jacobians, and Hessians.** The exponential density is

$$f_Y(y; \lambda) = \begin{cases} \frac{e^{-\frac{y}{\lambda}}}{\lambda}, & y \geqslant 0 \\ 0, & y < 0 \end{cases}$$

Suppose we have a sample $Y = \{y_1, ..., y_7\} = \{1.2, 1.4, 1.5, 1.6, 2.3, 2.5, 3.5\}$

(a) Define, as a Matlab function, the average of the logarithm of $f_Y(y_i)$,

$$s(Y; \lambda) = \frac{1}{7} \sum_{i=1}^{7} \ln f_Y(y_i; \lambda)$$

Evaluate the function when $\lambda = 1.5$ and when $\lambda = 2.0$.

(b) Compute, using automatic differentiation or finite differences, the gradient of $s(Y; \lambda)$, and evaluate it at $\lambda = 1.5$ and $\lambda = 2.0$.

(c) Consider the vector valued function

$$t(Y, \lambda) = [\ln f_Y(y_1; \lambda), ..., \ln f_Y(y_7; \lambda)].$$

Compute, using automatic differentiation or finite differences, the Jacobian of this function, and evaluate it at $\lambda = 1.5$ and $\lambda = 2.0$.

(d) Compute, analytically, the maximizer of $s(Y; \lambda)$ and evaluate it at the sample data.

(e) Compute, using automatic differentiation or finite differences, the second derivative of $s(Y; \lambda)$.

(f) Discuss the results

6. **Review.** A Bernoulli trial is a random experiment that can have two outcomes, success or failure. The probability of success is p. Suppose we have data on n repetitions of the experiment. The data are y_t, $t = 1, 2, ..., n$, where $y_t = 1$ means trial t was a success, and $y_t = 0$ means that trial t was a failure.

(a) What is the density function of y_t?

(b) Write the logarithm of the density function of y_t.

(c) What is the expected value of y_t?

(d) What is the variance of y_t?

(e) What is the expected value of the logarithm of the density function of y_t?

(f) What is the expected value of $\frac{1}{n} \sum_{t=1}^{n} (y_t - a)$ for a some arbitrary constant?

(g) What is the expected value of $\frac{1}{n} \sum_{t=1}^{n} (y_t - p)$ where p is the true probability of a success?