14.381, Fall 2006
Problem Set 5
Due: Nov. 16, 2006 (in class)

1. **Derive the OLS estimate for** β. *So much of what follows builds off of the basic ideas contained in this solution that you should be able to construct* $\hat{\beta}$ *in your sleep. Yes, you can just copy the results from class or one of many books, but you'll be thankful later if solving this on your own is second nature.*

   (a) What is OLS trying to minimize? Discuss (briefly) why the may or may not be a good idea.

   (b) What does it mean to be "the best linear predictor"?

   (c) Suppose you have the regression equation $y = X\beta + \varepsilon$. Derive the OLS estimate for β using

      1. Summation notation and
      2. Matrix notation.

2. **Interpreting regression coefficients**. *It's maddening to listen to someone present their results as "*β_1 *is 0.2 with a standard error of 0.08,* β_2 *is 1.4 ...." The point of econometrics is to be able to* say something *about your data. Here's a chance to practice interpreting regression coefficients on some numbers that admittedly don't say much.*

   Suppose you run the regression

$$G_i = \beta_1 + \beta_2 income_i + \beta_3 \ln gasprice_i + \beta_4 newcar_i + \varepsilon_i \tag{1}$$

   where G_i is individual i's gas consumption, $gasprice_i$ is the price of gas in her neighborhood, and $newcar_i$ is a dummy for whether or not she owns a new car. Suppose your estimate for β is [2.1, 0.01, -0.1, -150.2]. Use each of the coefficients in a sentence (for example, the coefficient on β_3 implies that when gas prices change by ... consumers spend ... more/less on gas). For interpreting log coefficients it may be useful to think about taking partial derivatives as in $\beta_3 = \partial G / \partial \ln gasprice$. What if the specification looked like

$$\ln G_i = \beta_1 + \beta_2 income_i + \beta_3 \ln gasprice_i + \beta_4 newcar_i + \varepsilon_i? \tag{2}$$

3. **Projection and residual matrices**. *You'll be using these all the time, so the goal of this problem is to get you comfortable working with them.*

   Define $P \equiv X(X'X)^{-1}X'$ and $M \equiv I - P$.

   (a) What does P do geometrically (hint: what does P stand for)? How about M?

   (b) Show that P and M are symmetric and idempotent. What is PM?

   (c) Suppose you have the model $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$. Define $P_j \equiv X_j(X_j'X_j)^{-1}X_j'$ and P as above with $X = [X_1 \vdots X_2]$. What is P_1P? How about M_1M? P_1X_1? M_1X_1? Explain intuitively why these answers make sense (think about your answers to a).

4. **Partitioned Regression**. *Once, and only once, you should work through by hand the math required to solve for β in the partitioned regression model. Not only might it become necessary on an exam, but the process captures many of the matrix manipulation you'll have to use elsewhere. Here you know what the answer is, and besides, it's a budding econometrician's Red Badge of Courage.*

   Consider the model

$$\underset{n\times 1}{y} = \underset{n\times(k_1+k_2)}{X}\ \underset{(k_1+k_2)\times 1}{\beta} + \underset{n\times 1}{\varepsilon} \tag{3}$$

   where X can be decomposed into $\begin{bmatrix} \underset{n\times k_1}{X_1} & \underset{n\times k_2}{X_2} \end{bmatrix}$ and β as $\begin{bmatrix} \underset{1\times k_1}{\beta_1'} & \underset{1\times k_2}{\beta_2'} \end{bmatrix}'$. Thus, the regression can be written as

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon. \tag{4}$$

   Let $P_j \equiv X_j(X_j'X_j)^{-1}X_j'$, i.e., our old friend the projection matrix onto X_j, and $M_j \equiv I - P_j$. Show the following:

   (a) The normal equations (the first order conditions for OLS) are

$$X_1'X_1\hat\beta_1 + X_1'X_2\hat\beta_2 = X_1'y, \tag{5}$$
$$X_2'X_1\hat\beta_1 + X_2'X_2\hat\beta_2 = X_2'y. \tag{6}$$

   Some helpful hints: (1) the transpose of a scalar equals the scalar (obvious but useful) and $\partial\mathbf{Ax}/\partial\mathbf{x} = \mathbf{A}'$.

   (b) $\hat\beta_2 = (\tilde X_2'\tilde X_2)^{-1}\tilde X_2'\tilde y$, where $\tilde X_2 \equiv M_1X_2$ and $\tilde y \equiv M_1y$.

   1. First, interpret what this form of the regression is doing, that is, if this were a simple univariate model, what would you be regressing on what.
   2. Next, try to show (b) by solving the normal equations (Hint: try solving (5) for $X_1\hat\beta_1$ and use the fact that M_1 is symmetric and idempotent).
   3. Finally, work through the gruesome math of matrix inversion (like I said, you should do this once and only one in your lifetime). The formula (for those of you who don't feel like deriving it) is in (A-74) on p. 824 of the fifth edition of Greene.[1] If working through this math seems too painful, just wait for the solution set, but you should work through the details once you see it.

5. **Cubic Spline Approximation.** The cubic spline takes the form:

$$x_t = f(w_t) = \left(1, w_t, w_t^2, w_t^3, (w_t - t_1)_+^3, ..., (w_t - t_r)_+^3\right)'.$$

   How many continuous derivatives does $f(w_t)'b$ have with respect to w_t, for any $(r+4)\times 1$ vector b?

[1] Appendix A of Greene is your friend: a great reference for all the matrix algebra few can remember.

6. **Monte Carlo**. *Through out the course, we'll be doing a number of Monte Carlo simulations, the main purpose of which is to build intuition. You can use Matlab, Stata (without the canned commands), Gauss, Mathmatica, or the like for the programming, but we'll be able to offer you more constructive feedback if you stick with Matlab or Stata. The purpose of this problem is to introduce the mechanics of OLS without relying on the canned* reg *command or its ilk and to introduce you to basic simulation programming. We strongly encourage you to write clean code (we'll distribute an example shortly) and to annotate your files so that when you return to them later you'll have some idea what the heck you were thinking.*

   Consider

$$y_i = x_i\beta + \varepsilon_i \tag{7}$$

   where x_1 is uniformly distributed on $[1, 5]$, ε_i is $N(0, 1)$, and $i = \{1, ..., I\}$. Suppose that the true value of β is 2. Repeat the following 1000 times:

   (a) Let $I = 10$, and generate 10 observations from (7). You'll use a random number generator to construct 10×1 vectors of i.i.d. random variables x and ε, then you'll construct y.

   (b) Calculate $\hat{\beta}$ and $s^2(X'X)^{-1}$. Store the results in a 1000×2 matrix (or 2 1000×1 vectors).

   (c) Plot your results for both $\hat{\beta}$ and $\widehat{\text{var}}(\hat{\beta})$ on a two histograms, each using equally spaced bins (say 25 or 51, your choice).

   Repeat this exercise for $I = 100, 1000$, and $100, 000$. Marvel at the power of larger samples and save your results for next week.

7. **Approximation Exercise**. The following table shows the conditional mean of wages $E[w|e]$ given education e:

$E[w\|e]$	5.84	5.94	5.97	6.08	6.24	6.41	6.45	6.76	6.57	6.88	7.31	6.98
e	8	9	10	11	12	13	14	16	17	18	19	20

   Try you best to find a parsimonious functional form $x = f(e) : R \to R^K$, such that $x'b$ approximates $E[w|e]$ for some constant $K \times 1$ vector b. Restrict the dimension of your proposed x to be no more than 8, i.e., $K \leq 8$. Calculate and show the RMSAE and the MAE.

   Again, you can use any software for the programming, but we'll be able to offer better feedback if you stick with Matlab. The example is the same as the one on the lecture handout. You might have noticed that there is an R-code attached to the notes. DO NOT try to decipher the code or to copy the code. It may take you a long time to figure it out! As a novice in R, you'll be better off to write your own code from scratch at this point. Be patient, and you'll get there eventually.