**Problem Set 2**
**Due Monday, April 13 at the beginning of lecture**
**ECON 381-2, Northwestern University, Spring 2015**

You may work in groups of 2 or 3 as long as each group member turns in their own handwritten copy and all group members are clearly indicated at the top of every copy. No credit is given if no work is shown. For Stata/R problems include both your code and the output.

1. Suppose that

$$Y_i = \alpha + \beta X_i + \gamma W_i + U_i.$$

for an i.i.d. sample $(Y_i, X_i, W_i)$ with finite fourth moments and no perfect linear dependencies among the regressors. We are not comfortable in assuming $\mathbb{E}(U_i \mid X_i, W_i) = 0$, because we do not think that $\mathbb{E}(U_i \mid W_i) = 0$. However, we believe that $\mathbb{E}(U_i \mid X_i, W_i) = \mathbb{E}(U_i \mid W_i)$ is a good assumption.

(a) Suppose that we assume $\mathbb{E}(U_i \mid W_i) = \delta_0 + \delta_1 W_i + \delta_2 W_i^2 + \delta_3 W_i^3$. Find a consistent estimator of $\beta$ and justify its consistency. (*Hint: Remember that you can appeal to theorems in the notes as long as the relevant assumptions are satisfied.*)

(b) Suppose that $W_i$ is a binary variable, i.e. it only takes on two values, 0 or 1. Without making any additional assumptions, find a consistent estimator of $\beta$ and justify its consistency.

(c) Suppose that $W_i$ is a random variable that can be 0, 1 or 2. Without making any additional assumptions, find a consistent estimator of $\beta$ and justify its consistency.

2. Consider the multiple linear regression model in matrix notation,

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_K X_{Ki} + U_i \equiv \boldsymbol{X}_i' \boldsymbol{\beta} + U_i$$

and assume that all of Assumptions MLR are satisfied.

(a) Show that if $K = 1$ then

$$\boldsymbol{\beta}_{mlr} \equiv \mathbb{E}[\boldsymbol{X}_i \boldsymbol{X}_i']^{-1} \mathbb{E}[\boldsymbol{X}_i Y_i] = \begin{bmatrix} \mathbb{E}[Y_i] - (\operatorname{Cov}(X_{1i}, Y_i)/\operatorname{Var}(X_{1i}))\,\mathbb{E}[X_{1i}] \\ \operatorname{Cov}(X_{1i}, Y_i)/\operatorname{Var}(X_{1i}) \end{bmatrix}.$$

What does this calculation show? *Hint: If you've forgotten how to invert a 2x2 matrix the answer is just a quick Google search away.*

(b) Show that if $K = 1$ and Assumption MLR7 holds then

$$\mathbb{E}[\boldsymbol{X}_i \boldsymbol{X}_i']^{-1} \mathbb{E}[U_i^2 \boldsymbol{X}_i \boldsymbol{X}_i'] \mathbb{E}[\boldsymbol{X}_i \boldsymbol{X}_i']^{-1} = \begin{bmatrix} \frac{\operatorname{Var}(U_i)\,\mathbb{E}[X_{1i}^2]}{\operatorname{Var}(X_{1i})} & \frac{-\operatorname{Var}(U_i)\,\mathbb{E}[X_{1i}]}{\operatorname{Var}(X_{1i})} \\ \frac{-\operatorname{Var}(U_i)\,\mathbb{E}[X_{1i}]}{\operatorname{Var}(X_{1i})} & \frac{\operatorname{Var}(U_i)}{\operatorname{Var}(X_{1i})} \end{bmatrix}.$$

What does this calculation show?

(c) Let $\boldsymbol{X}$ be the $N \times (K+1)$ matrix with typical $j^{\text{th}}$ row given by $[1, X_{1j}, \ldots, X_{Kj}]$ and let $\boldsymbol{Y}$ be the $N$–dimensional column vector with $j^{\text{th}}$ entry given by $Y_j$. Show that

$$\widehat{\boldsymbol{\beta}}_{mlr} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}.$$

What does this calculation show?

(d) Show that $\widehat{\boldsymbol{\beta}}_{mlr}$ is an unbiased estimator of $\boldsymbol{\beta}$.

3. Consider a sample analog to the best linear approximation problem:

$$\widehat{\boldsymbol{\beta}}_{bla} = \underset{\boldsymbol{b} \in \mathbb{R}^{K+1}}{\arg\min} \sum_{i=1}^{N} (Y_i - \boldsymbol{X}_i'\boldsymbol{b})^2,$$

constructed from an i.i.d. sample $(Y_i, \boldsymbol{X}_i), i = 1, \ldots, N$. Assume that $\sum_{i=1}^{N} \boldsymbol{X}_i\boldsymbol{X}_i'$ is invertible.

(a) Show that $\widehat{\boldsymbol{\beta}}_{bla} \equiv \widehat{\boldsymbol{\beta}}_{mlr} = (\sum_{i=1}^{N} \boldsymbol{X}_i\boldsymbol{X}_i')^{-1}(\sum_{i=1}^{N} \boldsymbol{X}_i Y_i)$.

(b) Let $\widehat{U}_i \equiv Y_i - \boldsymbol{X}_i'\widehat{\boldsymbol{\beta}}_{mlr}$. Show that $\sum_{i=1}^{N} \boldsymbol{X}_i \widehat{U}_i = \mathbf{0}$.

4. Suppose that we observe an i.i.d. sample $(Y_i, X_i, \boldsymbol{W}_i), i = 1, \ldots, N$, where all variables have finite fourth moments, and where $\boldsymbol{W}_i$ is a $K$–dimensional vector that includes a constant term. Let $\widehat{\beta}_{mlr}$ denote the coefficient on $X_i$ in a regression of $Y_i$ on $X_i$ and $\boldsymbol{W}_i$. This problem will show you another numerically equivalent way to compute $\widehat{\beta}_{mlr}$.

(a) Regress $X_i$ on $\boldsymbol{W}_i$. Let the resulting coefficient vector be denoted by $\widehat{\boldsymbol{\pi}}$. What is the expression for $\widehat{\boldsymbol{\pi}}$?

(b) Let $\widehat{V}_i$ be the residual from the regression in part a). That is, define $\widehat{V}_i \equiv X_i - \boldsymbol{W}_i'\widehat{\boldsymbol{\pi}}$. Is $\widehat{V}_i$ generally correlated with $\boldsymbol{W}_i$?

(c) Regress $Y_i$ on $\widehat{V}_i$ without a constant and let the resulting coefficient be denoted by $\widehat{\beta}_{new}$. Define a matrix $\boldsymbol{M_W} \equiv \boldsymbol{I}_N - \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'$, where $\boldsymbol{W}$ is the data matrix associated with $\{\boldsymbol{W}_i : i = 1, \ldots, N\}$. Show that

$$\widehat{\beta}_{new} = (\boldsymbol{X}'\boldsymbol{M_W}'\boldsymbol{M_W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_W}'\boldsymbol{Y}.$$

(d) Use the expression in part c) to show that $\widehat{\beta}_{new} = \widehat{\beta}_{mlr}$.
*Hint: You may find it helpful to make use of another result on this problem set.*

(e) Now suppose that instead of $X_i$ being a scalar, we have $\boldsymbol{X}_i$ being an $L$–dimensional vector. Let $\widehat{\boldsymbol{\beta}}_{mlr}$ denote the coefficient on $\boldsymbol{X}_i$ in a regression of $Y_i$ on $\boldsymbol{X}_i$ and $\boldsymbol{W}_i$. Generalize your results in parts a)–d) to find a similar numerically equivalent way to compute $\widehat{\boldsymbol{\beta}}_{mlr}$.

**Stata/R**

5. This problem uses the dataset contained in `wage2.dta`, which is available on Canvas. The dataset is a cross-section of 935 men in 1980. For R users, the data is contained in `wage2.tab`. A description of the variables is contained in the Stata file, and also in `wage2desc.pdf`.
   (a) Familiarize yourself with the data.
   (b) Regress wages on education and a constant. What is the coefficient on education, and how do you interpret it? Instead, regress log wages on education and a constant. Now how do you interpret the coefficient on education?
   (c) Is education correlated with total work experience? Does it make sense that it would or wouldn't be? How does controlling for experience affect the coefficient on education in the log wage regression? Does this make sense?
   (d) Regress log wages on the log of IQ score. How do you interpret the coefficient on log IQ score?
   (e) Is IQ correlated with education? How does controlling for IQ affect the coefficient on education in a regression of log wages on education and experience? Does controlling for the log of IQ change this? What about adding in a squared term for IQ? Does the direction of this effect make sense?
   (f) Repeat the previous part replacing IQ with the score on the Knowledge of the World of Work (KWW) assessment test. How do your answers change? Does including the KWW score along with IQ and experience affect your estimates of the return to education?