Econ 220B, Winter 2020

Problem Set 1
Due Thursday, January 16

1.) Let $\mathbf{X}$ be a $(T \times k)$ matrix whose columns are linearly independent, and let $\mathbf{M} = \mathbf{I}_T - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Show that $\mathbf{M}$ is symmetric and idempotent. Calculate the eigenvalues and rank of $\mathbf{M}$, and show that it is positive semidefinite.

2.) Consider a regression of y_t on $\mathbf{x}_t$ where the first element of $\mathbf{x}_t$ is a constant term. The R^2 or coefficient of determination is defined as

$$R^2 = 1 - \frac{\sum_{t=1}^{T}(y_t - \mathbf{x}_t'\mathbf{b})^2}{\sum_{t=1}^{T}(y_t - \overline{y})^2}$$

where $\mathbf{b}$ is the OLS regression coefficient and $\overline{y}$ is the sample mean. Show that $0 \leq R^2 \leq 1$.

3.) Let $\mathbf{P}$ be a nonsingular symmetric $(k_1 \times k_1)$ matrix, $\mathbf{Q}$ a nonsingular symmetric $(k_2 \times k_2)$ matrix, and $\mathbf{R}$ an arbitrary $(k_1 \times k_2)$ matrix,. Verify the following formula for the inverse of a partitioned matrix:

$$\begin{bmatrix} \mathbf{P} & \mathbf{R} \\ \mathbf{R}' & \mathbf{Q} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{W} & -\mathbf{W}\mathbf{R}\mathbf{Q}^{-1} \\ -\mathbf{Q}^{-1}\mathbf{R}'\mathbf{W} & (\mathbf{Q}^{-1} + \mathbf{Q}^{-1}\mathbf{R}'\mathbf{W}\mathbf{R}\mathbf{Q}^{-1}) \end{bmatrix}$$

for $\mathbf{W} = (\mathbf{P} - \mathbf{R}\mathbf{Q}^{-1}\mathbf{R}')^{-1}$.

4.) Consider a regression of y_t on $\mathbf{x}_t$, where we partition the regressors into two groups: $\mathbf{x}_t = (\mathbf{x}_{1t}', \mathbf{x}_{2t}')'$ where k_1 of the variables are included in the subvector $\mathbf{x}_{1t}$ and the remaining k_2 variables are in $\mathbf{x}_{2t}$:

$$y_t = \mathbf{x}_{1t}'\boldsymbol{\beta}_1 + \mathbf{x}_{2t}'\boldsymbol{\beta}_2 + \varepsilon_t.$$

The usual OLS regression coefficients are given by

$$\begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

for $\mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_2]$ a $[T \times (k_1 + k_2)]$ matrix and $\mathbf{X}_i$ the $(T \times k_i)$ matrix whose tth row is $\mathbf{x}_{it}'$. Use the results from question (3) to show that the OLS estimate $\mathbf{b}_1$ could equivalently be calculated as follows: (a) regress y_t on $\mathbf{x}_{2t}$ alone and calculate the residuals e_{2t}, for e_{2t} the tth element of $\mathbf{e}_2 = \mathbf{M}_2\mathbf{y}$ with $\mathbf{M}_2 = \mathbf{I}_T - \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'$; regress each element of $\mathbf{x}_{1t}$ on $\mathbf{x}_{2t}$ and calculate the residuals $\tilde{\mathbf{x}}_{1t}$, where $\tilde{\mathbf{x}}_{1t}'$ is the tth row of $\mathbf{M}_2\mathbf{X}_1$; (c) regress e_{2t} on $\tilde{\mathbf{x}}_{1t}$, to obtain a $(k_1 \times 1)$ vector $\hat{\boldsymbol{\beta}}_1$ that is numerically identical to $\mathbf{b}_1$ given above.

5.) Consider a special case of the previous result when the second explanatory variable is the constant term in the regression, so that $x_{2t} = 1$ for $t = 1, ..., T$ and $k_2 = 1$. Describe in words the interpretation of e_{2t} and $\tilde{\mathbf{x}}_{1t}$ for this case.

6.) Suppose you have performed an initial OLS regression of a scalar y_t on a $(k \times 1)$ vector of explanatory variables $\mathbf{x}_t$,

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + \varepsilon_t$$

and obtained the OLS estimates $\mathbf{b}$, s^2, OLS residuals $\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$, and R^2. You are asked to predict the consequences for OLS estimation where you regress y_t not on the original $\mathbf{x}_t$ but instead on a linear transformation of the original regressors,

$$\mathbf{x}_t^* = \mathbf{Q}\mathbf{x}_t,$$

where $\mathbf{Q}$ is a nonsingular $(k \times k)$ matrix, and you now perform the OLS regression

$$y_t = \mathbf{x}_t^{*\prime} \boldsymbol{\beta}^* + \varepsilon_t^*.$$

a.) Write the simplest possible formulas for the OLS estimates on the transformed data $\mathbf{b}^*$, s^{*2}, $\mathbf{e}^*$, and R^{*2} as functions of the original $\mathbf{b}$, s^2, $\mathbf{e}$, and R^2.

b.) As a special case, consider the regression on a constant and a scalar x_t:

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t.$$

What happens to the estimated values for b_1 and b_2 if you multiply x_t by 10 and regress

$$y_t = \beta_1^* + \beta_2^* x_t^* + \varepsilon_t^*$$

for $x_t^* = 10x_t$? What is the relation between the t-statistic for testing the null hypothesis $\beta_2 = 0$ using the original regression and the t-statistic for testing the null hypothesis $\beta_2^* = 0$ on the transformed regression?