Econometrics 522A
Spring, 2019
Hidehiko Ichimura, TA: Zejin Shi

PS #2 (Due 23 Jan, 2019)

1. Suppose X_1 denotes years of education and X_2 denotes gender, where $X_2 = 1$ denotes males and $X_2 = 0$ denotes females. Y is wage.
   (a) Please specify a linear regression model which allows years of education to affect wage differently for males and females and also for whether one is in manufacturing sector or not.
   (b) If Y is log-wage, how does the interpretation of the coefficients change?
2. In addition to X_1 and X_2 above, assume that X_3 denotes a manufacturing industry dummy variable where $X_3 = 1$ if the person works in a manufacture industry and $X_3 = 0$ if not.
   (a) Specify the most general model in which X_1 enters linearly.
   (b) Specify the model in which all the variables enter linearly and explain the restrictions imposed in the less general model.
3. Let X_1 and X_2 be both binary random variables taking values x_{10} and x_{11}, and also x_{20} and x_{21}, respectively. Using the indicator variables $1\{x_1 = x_{10}\}$ and $1\{x_2 = x_{20}\}$, describe how one can use a linear regression model to formulate a satiated model.
4. Show that if we multiply the dependent variables by c, then all of the OLS estimate is multiplied by c as well.
5. Show that if we multiply the jth independent variable by $c \neq 0$, then the jth OLS coefficient will be multiplied by $1/c$ without changing any other OLS estimates.
6. Show that if we add a constant to any variable, it shifts the constant term alone. Show how the constant term shifts.
7. Suppose the dependent variable is $\log y$ and the regresssor vector is x where the first element of x is 1. If we change the unit of y to a new measure cy. How does the OLS coefficients change?
8. Suppose the dependent variable is $\log y$ and the regressor vector is x where the first element of x is 1 and the second element is $\log x_1$. If we change the unit of x_1 to a new measure cx. How does the OLS coefficients change?
9. Show that when A is full rank if and only if $A'A$ is invertible.
10. Assume that b_0 is a real number and b is a length K vector and X is a length K random vector.
    (a) When $E(XX')$ is an invertible matrix, find the solution β_0 and β to the following minimization problem
    $$\min_{b_0,b} E[(Y - b_0 - X'b)^2].$$
    and show that $\beta = Var(X)^{-1}Cov(X, Y)$ and $\beta_0 = E(Y) - E(X)'\beta$.
    (b) Define $U = Y - \beta_0 - X'\beta$. Show that $E(U) = 0$ and $Cov(U, X) = 0$.
11. Consider the geometry of the OLS estimator with 2 regressors and 3 observations.
    (a) Draw a graph showing that the projection of Y vector on the space spanned by the 2 regressor vectors are well defined even when 2 regressor vectors are proportional to one another.
    (b) Show that in this case $X'X$ has rank 1.

12. Consider a time series monthly observations of consumption and income, where the dependent variable is consumption. Compare the coeffient on income when the regression model is specified with 12 monthly dummy variables without a constant term with the coefficient on income when monthly adjusted consumption is regressed on monthly adjusted income, where monthly adjustment is obtained as a residual of regressing each variable on 12 monthly dummy variables.

13. If we include an additional regressor, R^2 becomes always strictly larger if the additional regressor is linearly independent from the rest of the regressors and the OLS estimate of the additional coefficient is not zero. To show this consider the regression model $y_i = x_i'\beta + \alpha z_i + \epsilon_i$ and compare the R^2 for this model with the R^2 for the model $y_i = x_i'\beta + u_i$. Consider the auxiliary regression of z_i on x_i and a constant term, if its already not included among regressors in x_i and define the predicted value as $\hat{z}_i = x_i'\hat{\pi} + \hat{\pi}_0$ and the residual as $\hat{v}_{zi}$. Let $\hat{\beta}$ be the OLS estimate of the smaller model and $\hat{\beta}_1$ and $\hat{\alpha}$ be the OLS estimates of the larger model. Also let $\hat{u}_i$ denote the OLS residual from the smaller model and denote the OLS residual from the larger model $\hat{\epsilon}_i$, so that

$$y_i = x_i'\hat{\beta} + \hat{u}_i, \qquad y_i = x_i'\hat{\beta}_1 + \hat{\alpha} z_i + \hat{\epsilon}_i$$

   (a) Show that $\hat{\beta} = \hat{\beta}_1 + \hat{\alpha}\hat{\pi}$ and that $y_i = x_i'\hat{\beta} + \hat{\alpha}\hat{v}_{zi} + \hat{\alpha}\hat{\pi}_0 + \hat{\epsilon}_i$, where $\hat{u}_i = \hat{\alpha}\hat{v}_{zi} + \hat{\alpha}\hat{\pi}_0 + \hat{\epsilon}_i$.

   (b) Let $\hat{y}_i = x_i'\hat{\beta}_1 + \hat{\alpha} z_i$ and $\hat{\hat{y}}_i = x_i'\hat{\beta}$. Show that $\hat{y}_i = x_i'\hat{\beta} + \hat{\alpha}\hat{v}_{zi} + \hat{\alpha}\hat{\pi}_0$ so that

$$\hat{y}_i = \hat{\hat{y}}_i + \hat{\alpha}\hat{v}_{zi} + \hat{\alpha}\hat{\pi}_0.$$

   (c) Show that $\bar{\hat{y}} = \bar{\hat{\hat{y}}} + \hat{\alpha}\hat{\pi}_0$ and that $\hat{y}_i - \bar{\hat{y}} = \hat{\hat{y}}_i - \bar{\hat{\hat{y}}} + \hat{\alpha}\hat{v}_{zi}$.

   (d) Prove the main statement.

   (e) Show, by inspecting the objective function, that R^2 becomes always weakly larger if a regressor is added.

14. Consider the following regression model: for $i = 1, \ldots, 1000$

$$y_i = 1 + x_{1i} + x_{2i} + \epsilon_i$$

where x_{1i}, x_{2i}, ϵ_i are mutually independent standard normal random variables. Use R or Stata, and Python to solve the following questions.

   (a) Generate data according to the model above, use the data on the regressors and the dependent variable and estimate the coefficients by OLS.

   (b) Obtain the OLS residual and verify that the OLS residual and regressors are orthogonal.

   (c) Verify that the vector of unobserved true residuals ϵ_i $(i = 1, \ldots, 1000)$ is not orthogonal with the regressors.

   (d) Run the auxiliary regression of x_{1i} on 1 and x_{2i}, obtain the residual and regress y_i on the obtained residual from the auxiliary regression to verify that it coincides with the OLS coefficient on x_{1i} obtained in (a).

   (e) What is $E(x_i x_i')$ in this model? What is the rank of this matrix?