ECON 5350 Solutions to the Midterm Exam – Fall 2013

1. (**50 pts**) Consider the following ("true") population regression model without a constant (i.e., $\beta_1 = 0$):

$$y_i = \beta_2 x_i + \epsilon_i,$$

where $i = 1, ..., n$ and $\epsilon \sim i.i.d.N(0, \sigma^2)$.

(a) (10 pts) Derive the OLS estimator for β_2 in summation form. Check the second-order condition.

Solution. The objective function is

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - b_2 x_i)^2.$$

The derivative with respect to b_2 is

$$\frac{\partial(\bullet)}{\partial b_2} = -2\sum_{i=1}^{n} (y_i - b_2 x_i)x_i = 0.$$

Solving for b_2 gives

$$b_2 = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}.$$

The second-order condition for a minimum

$$\frac{\partial^2(\bullet)}{\partial (b_2)^2} = 2\sum_{i=1}^{n} x_i^2 \geq 0$$

is satisfied.

(b) (10 pts) Write down the regression model in matrix form, carefully defining all variables. Show the equivalence of the matrix-based OLS estimator to the estimator in part (a).

Solution. The regression model in matrix form is $Y = X\beta + \epsilon$, where $Y = (y_1, ..., y_n)'$, $X = (x_1, ..., x_n)'$, $\beta = \beta_2$, and $\epsilon = (\epsilon_1, ..., \epsilon_n)'$. The matrix-based OLS estimator is $b = (X'X)^{-1}(X'Y)$. Substituting the X and Y vectors into b immediately produces the formula in summation form above.

(c) (10 pts) Consider a sample of data $X = (2, 3, 3, 0)'$ and $Y = (7, 6, 4, 3)'$. Calculate the R^2 value and comment on the results.

Solution. Substituting the X and Y values into the summation formula from part (a) produces $b_2 = 2$. Using the slope estimate, the residual vector is $e = (3, 0, -2, 3)'$. The R^2 value is

$$R^2 = 1 - \frac{e'e}{(y - \bar{y})'(y - \bar{y})} = 1 - \frac{22}{10} = -\frac{6}{5}.$$

This value is negative because the intercept is suppressed and the formula is no longer bounded between zero and one.

(d) (10 pts) Calculate the slope estimate and R^2 with an intercept.

Solution. The slope estimator with an intercept is

$$b_2 = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{2}{3}.$$

The intercept estimator is $b_1 = \bar{y} - b_2\bar{x} = 5 - (2/3)2 = 11/3$. The new residual vector is

$$e = (6/3, 1/3, -5/3, -2/3)'$$

and the new R^2 value is

$$R^2 = 1 - \frac{(66/9)}{10} = 1 - \frac{33}{45} = \frac{4}{15} \simeq 0.27.$$

(e) (10 pts) The slope estimate in part (d) is different than the estimate in part (a). Does that imply one is biased? Explain.

Solution. Given that the "true" regression model does not have an intercept and the errors are mean zero, both estimators are unbiased. Two things to keep in mind though. First, you cannot determine unbiasedness given the results from one sample. You need to calculate the expected value of the estimator from its entire sampling distribution. Second, although both estimators are unbiased, the estimator without an intercept is likely to be preferred from an efficiency standpoint. The intercept is unnecessary in this model and its estimation will only add to the variability of the slope estimates in small samples.

2. (**50 pts**) Consider the beta(2,2) probability density function (pdf):

$$f(x) = Ax(1-x),$$

where $0 \leq x \leq 1$.

(a) (10 pts) Find the value of the constant A that ensures $f(x)$ is a valid pdf.

Solution. The pdf must integrate to one over the range $0 \leq x \leq 1$:

$$A\int_{x=0}^{1} x(1-x)dx = A\left(\frac{x^2}{2}|_{x=0}^{1} - \frac{x^3}{3}|_{x=0}^{1}\right) = \frac{A}{6}.$$

This implies that $A = 6$.

(b) (10 pts) Find the cdf for X and verify the properties.

Solution. The cdf for X is given by

$$F(x) = \int_{t=0}^{x} 6t(1-t)dt = x^2(3-2x)$$

over the range $0 \le x \le 1$. The cdf satisfies the required properties of a cdf: $F(0) = 0$, $F(1) = 1$, and $F(x) \ge 0$.

(c) (10 pts) Find the mean and the variance of X.

Solution. The mean is given by

$$E(X) = \mu = \int_{x=0}^{1} 6x^2(1-x)dx = \frac{1}{2}.$$

The variance is given by $var(X) = \sigma^2 = E(X^2) - \mu^2$ where

$$E(X^2) = \int_{x=0}^{1} 6x^3(1-x)dx = \frac{3}{10}.$$

This implies that

$$var(X) = \frac{3}{10} - \frac{1}{4} = \frac{1}{20}.$$

(d) (10 pts) Find the distribution of $Y = \ln X$.

Solution. The pdf for Y can be found using the change-of-variable technique:

$$g(y) = 6e^y(1-e^y)|dx/dy|$$

where $X = e^Y$. The pdf for Y

$$g(y) = 6e^{2y}(1-e^y),$$

for $-\infty \le y \le 0$ and zero elsewhere.

(e) (10 pts) Graph the distribution for X and describe a real-world random variable that could be described by the beta(2,2) distribution.

Solution.