Econometrics EC 424/524 Winter 2019

Problem Set 1

Due in Class, Tuesday, January 22.

1. (a) If X and Y are jointly normally distributed with means μ_x, μ_y and variances σ_x^2, σ_y^2, and covariance σ_{xy}, give the expected value and the variance of $Z = 5 + 3Y - 2X$. Is Z normally distributed? Obtain $cov(Z, X)$ and $cov(Z, Y)$.

(b) Suppose $X \sim N(\mu, \sigma^2)$ and $Y = (X - \mu)/\sigma$. Explain why $Y \sim N(0, 1)$.

(c) Suppose x_i and y_i, for $i = 1, ..., n$, are n observations of two variables, based on a random sample, with expected values $Ex_i = \mu_x$ and $Ey_i = \mu_y$. Let $\bar{x} = n^{-1} \sum_{i=1}^{n} x_i$ and $\bar{y} = n^{-1} \sum_{i=1}^{n} y_i$ denote the sample means. Suppose $z_i = ax_i + by_i + c$ for specified numbers a, b, c and $\bar{z} = n^{-1} \sum_{i=1}^{n} z_i$. (i) Show that $E(\bar{x}) = \mu_x$. (ii) Show that $\bar{z} = a\bar{x} + b\bar{y} + c$ and that $E\bar{z} = a\mu_x + b\mu_y + c$.

2. Suppose we have data $\{x_i, y_i\}_{i=1}^{n}$.

(a) Let s_x^2 and s_y^2 denote the sample variances of x and y, and let s_{xy} denote the sample covariance of x and y. If $z_i = a + bx_i$ show using the definitions of sample variances and covariances that $s_z^2 = b^2 s_x^2$ and $s_{zy} = bs_{xy}$.

(b) Show that $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i - \bar{x})y_i$.

(c) Let $\hat{y}_i = a + bx_i$ be the fitted values from a simple regression. Show that $\bar{\hat{y}} = \bar{y}$.

3. Let R^2 be the goodness of fit of the regression of y on an intercept and a single independent variable x using data $\{x_i, y_i\}_{i=1}^{n}$ and let r_{xy} be the sample correlation between x and y. (i) Show that $\hat{y}_i - \bar{y} = (x_i - \bar{x})b$ and thus that $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = b^2 \sum_{i=1}^{n}(x_i - \bar{x})^2$. (ii) Using (i) show that $R^2 = r_{xy}^2$.

4. Consider the simple regression model $y_i = \alpha + \beta x_i + \varepsilon_i$, $i = 1, \ldots, n$. Recall that the LS estimate b of β satisfies $b = \beta + \sum_{i=1}^{n} w_i \varepsilon_i$, where the w_i depends only on $x' = (x_1, \ldots, x_n)$ and where $\sum_{i=1}^{n} w_i = 0$. Suppose that as usual the $\varepsilon_i | x$ are *iid* but that instead of the usual assumption $E(\varepsilon_i | x) = 0$ we have $E(\varepsilon_i | x) = \gamma$, where $\gamma \neq 0$. Find $E(b|x)$. Is b a conditionally unbiased estimator of β? Is $a = \bar{y} - b\bar{x}$ a conditionally unbiased estimator of α?

5. (Matrix Review). (i) If A is an $n \times m$ matrix and $A'A = 0$, show that $A = 0$. (You may find it helpful to write $A = \begin{pmatrix} a_1 & \cdots & a_m \end{pmatrix}$ and note that $(A'A)_{ij} = a_i' a_j$) (ii) Let ι be an $n \times 1$ column vector of units, i.e. $\iota' = (1, \ldots, 1)$ and let I be the $n \times n$ identity matrix. Show that the matrix $I - n^{-1}\iota\iota'$ is symmetric and idempotent. (A matrix A is symmetric if $A = A'$ and idempotent if $AA = A$.) (iii) For any matrix X show that $X'X$ is positive semidefinite. (A $k \times k$ matrix A is said to be positive semidefinite if $y'Ay \geq 0$ for all $k \times 1$ vectors y).

6. The Stata file AUTO1.DTA contains quarterly data 1959q1 to 1992q1 on a number of variables, including gexp = log per capita real personal consumption expenditure on gasoline and oil and pg = log real price index for gasoline and oil. Define

$$\begin{aligned} \text{dgexp} &= 400 \cdot (\text{gexp - gexp(-1)}) \\ \text{dpg} &= 400 \cdot (\text{pg - pg(-1)}). \end{aligned}$$

The notation x(-1) denotes the lagged value of x, i.e. the value of x(-1) at time t is the value of x at time t-1.

Use LS to estimate the model

$$\text{dgexp} = \alpha + \beta \ \text{dpg} + \varepsilon$$

from 1959q2 to 1992q1. The intercept α and the slope β can be interpreted as the trend growth rate and the price elasticity, respectively, of per capita gas and oil consumption.

(i) Explain why dgexp represents the percentage growth rate (at annual rates) of per capita real personal consumption expenditure on gasoline and oil. Report the sample mean of dgexp.

(ii) Report LS estimates a, b, their standard errors SE(a), SE(b), and R^2.

(iii) Use your answer to (ii) and the t-distribution table posted on Canvas to test the null hypothesis $H_0 : \alpha = 0$ vs. $H_1 : \alpha \neq 0$ at the 5% significance level.

(iv) Use your answer to (ii) and the t-distribution table posted on Canvas to obtain a 90% confidence interval for β.

Note: For (iii) and (iv) use 100 in the t-distribution table as an approximation for the degrees of freedom. More precise values can be obtained using Stata, but your answers to (iii) and (iv) should be close to answers obtainable directly from Stata.