Advanced Econometrics II:

April 7, 2024

Assignment 1

(This assignment is due on April 12, 2024 at noon. Please submit answers to TA on time. The total marks are indicated in each question.)

1. Consider a linear regression model with $\mathrm{k} = 2$ and $\mathrm{x_{i1}} \equiv 1$, that is

   $\mathrm{y_i} = \beta_1 + x_i\beta_2 + u_i \qquad (\mathrm{i} = 1, \dots, \mathrm{n})$

   a) [5 points] In matrix notation, $\mathrm{Y} = \mathrm{X}\beta + \mathrm{u}$, give a detailed expression of the X matrix.

   b) [5 points] From $\hat{\beta} = (X'X)^{-1}X'Y$, show that

   $$\hat{\beta}_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

   where $\bar{x} = n^{-1}\sum_{i=1}^{n} x_i$ and $\bar{y} = n^{-1}\sum_{i=1}^{n} y_i$.

   c) [5 points] Show that $\hat{\beta}_1 = \bar{y} - \bar{x}\,\hat{\beta}_2$, where $\bar{x} = n^{-1}\sum_{i=1}^{n} x_i$ and $\bar{y} = n^{-1}\sum_{i=1}^{n} y_i$.

2. Consider the following linear regression model

   $$\mathrm{y_i} = \beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + x_{i4}\beta_4 + u_i \qquad (\mathrm{i} = 1, \dots, 10)$$

   where $u_i \sim N(0, \sigma^2)$. Suppose that $\mathrm{y_i}$ is the logarithm of output for firm i, x_{i2} is the logarithm of labor input, x_{i3} is the logarithm of capital input, x_{i4} is the firm's R&D input.

   a) [5 points] Discuss how to test the null hypothesis of the joint test: $H_0\colon \beta_2 + \beta_3 = 1$ and $\beta_4 = 0$ based on $R\beta - q = 0$. (You only need to show what R and q are equal to)

b) [5 points] Use $Y = X\beta + u$ to show the detailed steps about the construction of the test statistic for testing $R\beta - q = 0$.

c) [5 points] What is the economics interpretation of the null hypothesis specified in a)?

3. [10 points] For the classical linear regression model that we review in the class:

$$y = X\beta + u$$

where X is $n \times k$ and has no constant term. The $F - test$ for the restriction of $\beta = 0$ can be construct as $F[k, n-k] = \frac{R^2/k}{(1-R^2)/(n-k)}$, where the R^2 is the uncentered R^2.

Please calculate $plim\ F[k, n-k] = plim \frac{R^2/k}{(1-R^2)/(n-k)}$ =?

4. For the classical linear regression model with measurement error: $y^* = \beta x^* + \epsilon$, where y^* and x^* are two random variables standing for the true values and $\epsilon \sim N(0, \sigma_\epsilon^2)$. However, y^* and x^* are not observable. Instead we observe: $x \equiv x^* + u$, with $u \sim N(0, \sigma_u^2)$ and $y \equiv y^* + v$, with $v \sim N(0, \sigma_v^2)$. ϵ, u, and v are independent between each other and are independent of x^* and y^*.

a) [5 points] Please prove that $(corr(y^*, x))^2 < (corr(y^*, x^*))^2$

b) [5 points] Please prove that $(corr(y, x))^2 < (corr(y^*, x^*))^2$

5. Consider the model:

$$\mathrm{y} = \mathrm{Z}\delta + \epsilon \qquad \text{which can also be written as}$$

$$\mathrm{y} = \mathrm{X}_1\gamma_1 + X_2\gamma_2 + \epsilon,$$

where, $\epsilon \sim N(0, \sigma^2 I)$, $EX_1'\epsilon = 0$, and $EX_2'\epsilon \neq 0$. σ^2 can be estimated by $e'e/n$. Some instrumental variables are available, collected in a matrix X_{IV}, where $EX_{IV}'\epsilon = 0$ and $W = [X_1 \vdots X_{IV}]$. X_{IV} has more columns than does X_2. Z and W are nonrandom.

a) [5 points] Please derive $\hat{\delta}$ using efficient GMM and show that it is the same as the one derived using 2SLS.

b) [5 points] Basing on the GMM estimation, show that the test of over-identifying restrictions is equal to n times the uncentred R^2 from regressing the 2SLS residuals (2SLS residual is $e = y - Z\hat{\delta}_{2SLS}$) on W.

6. [10 points] In the classical linear regression with heteroscedasticity and endogenous variables, which is more efficient, 2 stage least squares of GMM? Obtain the two estimators and their respective asymptotic Var-Cov matrices, then prove your assertion.

(Note: The linear regression model is: $y = X\beta + u$, where X contains endogenous variables and we use W as instrument variables. W has the same number of columns as X.)

7. Two sets of observations each contain information about and an unknown parameter μ. The observations in sample A are assumed to have been drawn from a distribution with a mean equal to μ. The observations in sample B are assumed to have been drawn from a distribution with a mean equal to $\mu + 5$. The observations in sample A are not correlated with the observations in sample B. The sample means are $\bar{y}_A = 7$ and $\bar{y}_B = 9$, and their variances are estimated to be $\mathrm{Var}(\bar{y}_A) = 1$ and $\mathrm{Var}(\bar{y}_B) = 0.2$.

a) [5 points] Write a quadratic form that is a function of μ and is minimized at the asymptotically efficient GMM estimate of μ. Compute the numerical value of the GMM estimate.

b) [5 points] State the null hypothesis for testing the assumption about the relationship between the means of samples A and B. Compute the values of the GMM test statistic for it and state the asymptotic null distribution of this test statistic, including degrees of freedom.

8. We want to test that $H_0: \beta_1 = 0$ in the classical linear regression model $y = X_1\beta_1 + X_2\beta_2 + u$, where $u \sim N(0, \sigma^2 I)$, X_1 is nonrandom and $n \times k_1$, X_2 is nonrandom and $n \times k_2$, β_1 is $k_1 \times 1$, β_2 is $k_2 \times 1$.

   (a) [5 points] Please derive that: $Wald\ test = \frac{1}{\sigma^2} y' M_2 X_1 (X_1' M_2 X_1)^{-1} X_1' M_2 y$

   (b) [5 points] Please derive that: $LM\ test = \frac{1}{\sigma^2} y' M_2 X_1 (X_1' M_2 X_1)^{-1} X_1' M_2 y$

(Note: $M_2 = I - X_2(X_2'X_2)^{-1}X_2{}'$. For the derivation of (b), you are given the restricted estimator of β_1 and β_2 i.e.$\tilde{\beta}_1 = 0$ and $\tilde{\beta}_2 = (X_2'X_2)^{-1}X_2{}'y$)