Methods of Economic Investigation II (EC403)
Midyear Test

Selected Solutions

1. If we write the procedures in mean deviation form we get

$$b = \frac{\sum(x_i - \overline{x})y_i}{\sum(x_i - \overline{x})^2},$$

while in the second place we compute

$$\begin{pmatrix} \widehat{\beta} \\ \widehat{\gamma} \end{pmatrix} = \begin{pmatrix} \sum(x_i - \overline{x})^2 & \sum(x_i - \overline{x})(z_i - \overline{z}) \\ \sum(x_i - \overline{x})(z_i - \overline{z}) & \sum(z_i - \overline{z})^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum(x_i - \overline{x})y_i \\ \sum(z_i - \overline{z})y_i \end{pmatrix}.$$

   (a) If $\sum(x_i - \overline{x})(z_i - \overline{z}) = 0$, then $b = \widehat{\beta}$.

   (b) It is always the case that $\sum \widehat{\varepsilon}_i^2 \leq \sum e_i^2$ because there are more choice variables in the case corresponding to $\sum \widehat{\varepsilon}_i^2$, so that the least squares criterion can always be made as small.

   (c) This is quite likely to be the case when the true parameter is close to zero and not precisely estimable. The variance of $\widehat{\beta}$ is always greater than the variance of b. If the coefficient on z is zero, then b will be unbiased and the standard errors obtained in the first procedure will be correct. In this case we might expect to find significant b but insignificant $\widehat{\beta}$.

   (d) Suppose that the true coefficient on x is zero but the true coefficient on z is nonzero. Then $\widehat{\beta}$ will be unbiased and should be insignificantly different from zero. However, if the is sufficient correlation between x and z we might find that b is badly biased and can appear to be significantly different from zero.

2. This question is very easy, and the solutions are obvious.

3. Clearly, $(\widehat{\alpha}_1, \widehat{\beta}_1)'$ is unbiased. We can write

$$\widehat{\beta}^* = \frac{\sum_{i=1}^n (x_i^* - \overline{x}^*)y_i}{\sum_{i=1}^n (x_i^* - \overline{x}^*)^2} \quad ; \quad \widehat{\alpha}^* = \overline{y} - \widehat{\beta}^* \overline{x}^*.$$

Then note that $\overline{x}^* = \sum_{i=1}^n x_i^*/n = (\sum_{i=1}^{n_1} x_i^* + (n-n_1)\overline{x}_1)/n = \overline{x}_1$, so that $\sum_{i=1}^n (x_i^* - \overline{x}^*)^2 = \sum_{i=1}^{n_1}(x_i - \overline{x}_1)^2$ and $\sum_{i=1}^n (x_i^* - \overline{x}^*)y_i = \sum_{i=1}^{n_1}(x_i - \overline{x}_1)y_i$. In other words, $\widehat{\beta}^* = \widehat{\beta}_1$ so that $\widehat{\beta}^*$ is unbiased and of course has the same variance as $\widehat{\beta}_1$. We now write

$$\widehat{\alpha}^* = \overline{y} - \widehat{\beta}_1\overline{x}_1 = \alpha + \beta\overline{x} + \overline{\varepsilon} - \widehat{\beta}_1\overline{x}_1 = \alpha + \beta(\overline{x} - \overline{x}_1) + \overline{\varepsilon} - (\widehat{\beta}_1 - \beta)\overline{x}_1,$$

which has expectation $\beta(\overline{x} - \overline{x}_1)$. This is unconditionally mean zero, if in random design. Therefore in the fixed design there is a bias [which vanishes asymptotically] but in the random design there is no bias. The variance of $\widehat{\alpha}^*$ is smaller because the variance of $\overline{y}$ is smaller than the variance of $\overline{y}_1$ [it uses additional observations].

(a) If we take $z_i = 1$, then we can write $\overline{\beta} = \sum_{i=1}^n z_i y_i / \sum_{i=1}^n z_i x_i$, i.e., it is an instrumental variable estimator.

(b) Note that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n y_i &= \beta\frac{1}{n}\sum_{i=1}^n x_i + \frac{1}{n}\sum_{i=1}^n \varepsilon_i - \beta\frac{1}{n}\sum_{i=1}^n u_i \\ &= \beta\frac{1}{n}\sum_{i=1}^n x_i + o_p(1) \\ &\to {}^p \beta\mu^*\end{aligned}$$

by the law of large numbers using the facts that $E\varepsilon_i = Eu_i = 0$. Furthermore,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n x_i &= \frac{1}{n}\sum_{i=1}^n x_i^* + \frac{1}{n}\sum_{i=1}^n u_i \\ &= \frac{1}{n}\sum_{i=1}^n x_i^* + o_p(1) \\ &\to {}^p \mu^*\end{aligned}$$

also by the law of large numbers. Therefore, $\overline{\beta} \to^p \beta\mu^*/\beta = \beta$ by the Slutsky theorem. As for the asymptotic distribution, Note that

$$\sqrt{n}(\overline{\beta} - \beta) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i - \beta\frac{1}{\sqrt{n}}\sum_{i=1}^n u_i}{\frac{1}{n}\sum_{i=1}^n x_i} = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i - \beta\frac{1}{\sqrt{n}}\sum_{i=1}^n u_i}{\mu^* + o_p(1)} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \eta_i + o_p(1)$$

is a sum of independent mean zero random variables. Therefore a central limit theorem applies and $\sqrt{n}(\overline{\beta} - \beta)$ has asymptotic variance $\sigma_\eta^2 = (\sigma_\varepsilon^2 + \beta^2\sigma_u^2)/(\mu^*)^2$.

(c) A two-sided confidence interval would be $\overline{\beta} \pm z_{\alpha/2}\overline{\sigma}_\eta^2/\sqrt{n}$, where $\overline{\sigma}_\eta^2$ is a consistent estimate of σ_η^2. Specfically, $\overline{\sigma}_\eta^2 = \sum \overline{\eta}_i^2/n$, where $\overline{\eta}_i = (y_i - \overline{\beta}x_i)/\frac{1}{n}\sum_{i=1}^n x_i$.