Econometrics

We have a random sample of size n from some population. For each individual i, there are variables (Y_i, Z_i, A_i), where we interpret Z_i as a noisy measure of A_i. We observe (Y_i, Z_i) for $i = 1, \ldots, n$, but we do not observe A_i. The structural model is

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 A_i + U_i \\ Z_i &= A_i + V_i \qquad (i = 1, \ldots, n), \end{aligned}$$

where the latent (unobserved) random variables A_i, U_i, and V_i have mean 0 and are all uncorrelated with each other. In addition, we assume that $\beta_1 > 0$.

a. **(10 points)** Work out the covariance matrix for (Y_i, Z_i) (i.e., express the two variances and one covariance in terms of the model parameters β_1, $\mathrm{Var}(A_i)$, $\mathrm{Var}(U_i)$, $\mathrm{Var}(V_i)$).

Answer:

$$\begin{pmatrix} \mathrm{Var}(Z_i) & \mathrm{Cov}(Z_i, Y_i) \\ \mathrm{Cov}(Z_i, Y_i) & \mathrm{Var}(Y_i) \end{pmatrix} = \begin{pmatrix} \mathrm{Var}(A_i) + \mathrm{Var}(V_i) & \beta_1 \mathrm{Var}(A_i) \\ \beta_1 \mathrm{Var}(A_i) & \beta_1^2 \mathrm{Var}(A_i) + \mathrm{Var}(U_i) \end{pmatrix}$$

b. **(25 points)** Consider the linear predictors

$$E^*(Y_i \,|\, 1, Z_i) = \pi_0 + \pi_1 Z_i,$$

$$E^*(Z_i \,|\, 1, Y_i) = \alpha_0 + \alpha_1 Y_i.$$

Show that

$$\pi_1 \;\leq\; \beta_1 \;\leq\; 1/\alpha_1\,.$$

Provide a procedure that uses the data on (Y_i, Z_i) for $i = 1, \ldots, n$ to provide estimates of these lower and upper bounds for β_1. Explain why the estimates are consistent.

Answer:

$$\pi_1 = \frac{\mathrm{Cov}(Z_i, Y_i)}{\mathrm{Var}(Z_i)} = \beta_1 \frac{\mathrm{Var}(A_i)}{\mathrm{Var}(A_i) + \mathrm{Var}(V_i)} \leq \beta_1$$

$$1/\alpha_1 = \frac{\mathrm{Var}(Y_i)}{\mathrm{Cov}(Z_i, Y_i)} = \frac{\beta_1^2 \mathrm{Var}(A_i) + \mathrm{Var}(U_i)}{\beta_1 \mathrm{Var}(A_i)} = \beta_1 + \frac{\mathrm{Var}(U_i)}{\beta_1 \mathrm{Var}(A_i)} \geq \beta_1$$

Use the least-squares estimates:

$$\hat{\pi}_1 = \frac{\text{sample Cov}(Z,Y)}{\text{sample Var}(Z)},$$

$$1/\hat{\alpha} = \frac{\text{sample Var}(Y)}{\text{sample Cov}(Z,Y)},$$

where, for example,

$$\text{sample Cov}(Z,Y) = \frac{1}{n}\sum_{i=1}^{n} Z_i Y_i - \frac{1}{n}\sum_{i=1}^{n} Z_i \cdot \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

These estimates are consistent because, under random sampling, sample means converge to population means, and so sample covariances converge to population covariances.

c. **(25 points)** Now suppose that there are two noisy measures of A_i, with

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 A_i + U_i \\ Z_{i1} &= A_i + V_{i1} \\ Z_{i2} &= A_i + V_{i2} \qquad (i = 1, \ldots, n), \end{aligned}$$

where the latent (unobserved) random variables A_i, U_i, V_{i1}, and V_{i2} have mean 0 and are all uncorrelated with each other. Show that now β_1 is identified, and use your analysis to provide a consistent estimator of β_1 using the data on (Y_i, Z_{i1}, Z_{i2}) for $i = 1, \ldots, n$.

Answer:

The covariance matrix for Z_{i1}, Z_{i2}, Y_i is

$$\text{Cov}\begin{pmatrix} Z_{i1} \\ Z_{i2} \\ Y_i \end{pmatrix} = \begin{pmatrix} \text{Var}(A_i) + \text{Var}(V_{i1}) & \text{Var}(A_i) & \beta_1 \text{Var}(A_i) \\ - & \text{Var}(A_i) + \text{Var}(V_{i2}) & \beta_1 \text{Var}(A_i) \\ - & - & \beta_1^2 \text{Var}(A_i) + \text{Var}(U_i) \end{pmatrix}.$$

So β_1 is identified from

$$\beta_1 = \frac{\text{Cov}(Z_{i1}, Y_i)}{\text{Cov}(Z_{i1}, Z_{i2})} \quad \text{or} \quad \beta_1 = \frac{\text{Cov}(Z_{i2}, Y_i)}{\text{Cov}(Z_{i1}, Z_{i2})}.$$

We can obtain a consistent estimator for β_1 by using sample covariances to estimate population covariances (as in (b)):

$$\hat{\beta}_1 = \frac{\text{sample Cov}(Z_1, Y)}{\text{sample Cov}(Z_1, Z_2)} \quad \text{or} \quad \hat{\beta}_1' = \frac{\text{sample Cov}(Z_2, Y)}{\text{sample Cov}(Z_1, Z_2)}.$$