Sensitivity Analysis Math

Jason Premo

January 2024

Consider a scalar outcome metric y and vectors of (possibly endogenous) covariates x and instruments z. Given matrices Y, X and Z consisting of N data points each, the variance of the 2SLS estimator is

$$Var(\beta_{2SLS}) = \sigma^2(X'Z(Z'Z)^{-1}Z'X)^{-1}$$

where σ^2 is the variance of the 2SLS second stage residual. By the law of large numbers,

$$Z'Z \xrightarrow{p} Var(z) \equiv \sigma_z^2$$
$$X'Z \xrightarrow{p} Cov(x,z) = Cov(z,x) \xleftarrow{p} Z'X$$

Therefore,

$$Var(\beta_{2SLS}) = \sigma^2(X'Z(Z'Z)^{-1}Z'X)^{-1} \xrightarrow{p} \sigma^2\left(\frac{\sigma_z^2}{Cov(x,z)^2}\right)$$

Using 2SLS, the endogenous covariates x are estimated in a first stage regression.

$$x = \alpha_0 + \alpha_1 z + u$$

Therefore,

$$Cov(x,z)^2 = Cov(\alpha_0 + \alpha_1 z, z)^2 = (\alpha_1 Cov(z,z))^2 = (\alpha_1 \sigma_z^2)^2$$

Which implies

$$Var(\beta_{2SLS}) \xrightarrow{p} \frac{\sigma^2}{\alpha_1^2 \sigma_z^2}$$

By the central limit theorem, the test statistic under the null hypothesis ($\beta_{2SLS} = 0$) converges to a standard normal.

$$\tau_N = \sqrt{N}\frac{\beta_{2SLS}}{\frac{\sigma}{\alpha_2\sigma_z}} = \frac{\sqrt{N}\beta_{2SLS}\alpha_2\sigma_z}{\sigma} \overset{a}{\sim} N(0,1)$$

Note that if z is a binary treatment with probability p of adoption, then $\sigma_z = p(1-p)$. The calculations in the sensitivity analysis workbook perform power calculations on the above test statistic with the simplifying assumption that z is binary. Since σ is not known, it must be estimated using existing data, but in the interest of time and simplicity, this additional source of variance is not accounted for, and the parameter is treated as though it is known *a priori*.