Economics 231a — Fall, 2005

First Order Optimality of 2SLS: Alternative Derivation

Suppose that we have a linear model

$$y_i = x_i'\beta + u_i \qquad i = 1, 2, \ldots$$

where there exists some z_i such that $E[z_i u_i] = 0$ and $E[z_i x_i']$ is nonsingular. We will suppose that $\dim(z_i) \geq \dim(x_i)$, and analyze 2SLS. As before, we will assume that $(y_i, x_i', z_i')'$ is i.i.d.

Condition 1 $E[z_i u_i] = 0$, $E[u_i^2 z_i z_i'] = \sigma^2 E[z_i z_i']$

We will first consider a class of IV estimators using Lz_i as an instrument, where L is a $\dim(x_i) \times \dim(z_i)$ matrix. Note that the IV estimator can be written as

$$b_L = \left(\sum_{i=1}^{n} (Lz_i)\, x_i'\right)^{-1} \left(\sum_{i=1}^{n} (Lz_i)\, y_i\right)$$

We know that

$$\sqrt{n}\,(b_L - \beta) \xrightarrow{d} N(0, \Omega_L),$$

where

$$\begin{aligned}
\Omega_L &= A_L^{-1} B_L (A_L')^{-1} \\
A_L &= E[(Lz_i)\, x_i'] = LE[z_i x_i'] \\
B_L &= E\left[(Lz_i u_i)(Lz_i u_i)'\right] = LE\left[u_i^2 z_i z_i'\right] L' = \sigma^2 LE[z_i z_i']\, L'
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\Omega_L &= \sigma^2 (LE[z_i x_i'])^{-1} LE[z_i z_i']\, L' \left(E[x_i z_i']\, L'\right)^{-1} \\
&= \sigma^2 (LG)^{-1} LHL' (G'L')^{-1}
\end{aligned}$$

where we write

$$\begin{aligned}
G &= E[z_i x_i'] \\
H &= E[z_i z_i']
\end{aligned}$$

for simplicity of notation.

We would want to choose L to minimize Ω_L. It can be shown that $L^* = G'H^{-1}$ minimizes Ω_L. In order to appreciate it, let

$$\begin{aligned}
Q^* &= (L^*G)^{-1} L^* = \left(G'H^{-1}G\right)^{-1} G'H^{-1} \\
Q &= (LG)^{-1} L
\end{aligned}$$

Note that

$$\begin{aligned} H(Q^*)' &= G\left(G'H^{-1}G\right)^{-1} \\ Q^*G &= \left(G'H^{-1}G\right)^{-1} G'H^{-1}G = I \\ QG &= (LG)^{-1} LG = I \end{aligned}$$

It follows that

$$(Q - Q^*) H (Q^*)' = (Q - Q^*) G \left(G'H^{-1}G\right)^{-1} = QG - Q^*G = I - I = 0$$

from which we obtain

$$\begin{aligned} (LG)^{-1} LHL' (G'L')^{-1} &= QHQ' \\ &= [Q^* + (Q - Q^*)] H [Q^* + (Q - Q^*)]' \\ &= Q^*H(Q^*)' + (Q - Q^*) H (Q - Q^*)' \\ &\geq Q^*H(Q^*)' \\ &= \left(G'H^{-1}G\right)^{-1} G'H^{-1} \cdot H \cdot \left[\left(G'H^{-1}G\right)^{-1} G'H^{-1}\right]' \\ &= \left(G'H^{-1}G\right)^{-1} \end{aligned}$$

We can therefore see that the optimal choice of L is

$$L^* = G'H^{-1} = (E[x_i z_i'])(E[z_i z_i'])^{-1}$$

We now consider a class of estimators $b_{\widehat{L}}$, where $\widehat{L} = L^* + o_p(1)$. We can see that

$$\begin{aligned} b_{\widehat{L}} &= \left(\sum_{i=1}^{n} \left(\widehat{L} z_i\right) x_i'\right)^{-1} \left(\sum_{i=1}^{n} \left(\widehat{L} z_i\right) y_i\right) = \left(\widehat{L} \sum_{i=1}^{n} z_i x_i'\right)^{-1} \left(\widehat{L} \sum_{i=1}^{n} z_i y_i\right) \\ &= \beta + \left(\widehat{L} \sum_{i=1}^{n} z_i x_i'\right)^{-1} \left(\widehat{L} \sum_{i=1}^{n} z_i u_i\right) = \beta + \left[\widehat{L}\left(\frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right)\right]^{-1} \left[\widehat{L}\left(\frac{1}{n}\sum_{i=1}^{n} z_i u_i\right)\right] \end{aligned}$$

We therefore have

$$\sqrt{n}\left(b_{\widehat{L}} - \beta\right) = \left\{\left[\widehat{L}\left(\frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right)\right]^{-1} \widehat{L}\right\} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i u_i\right)$$

Note that

$$\begin{aligned} \left[\widehat{L}\left(\frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right)\right]^{-1} \widehat{L} &= [(L^* + o_p(1))(E[z_i x_i'] + o_p(1))]^{-1} (L^* + o_p(1)) \\ &= [(L^* + o_p(1))(G + o_p(1))]^{-1} (L^* + o_p(1)) \\ &= (L^*G)^{-1} L^* + o_p(1) \\ &= \left(G'H^{-1}G\right)^{-1} G'H^{-1} + o_p(1) \end{aligned}$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i u_i \xrightarrow{d} N\left(0, \sigma^2 E\left[z_i z_i'\right]\right) = N\left(0, \sigma^2 H\right)$$

It follows that

$$\sqrt{n}\left(b_{\widehat{L}} - \beta\right) \xrightarrow{d} N\left(0, \sigma^2 \left(G' H^{-1} G\right)^{-1}\right)$$

Note that

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i z_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-1} = \left(E\left[x_i z_i'\right]\right)\left(E\left[z_i z_i'\right]\right)^{-1} + o_p(1)$$

so we can take

$$\widehat{L} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i z_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-1}$$

We then have

$$\begin{aligned}
b_{\widehat{L}} &= \left(\widehat{L}\sum_{i=1}^{n} z_i x_i'\right)^{-1}\left(\widehat{L}\sum_{i=1}^{n} z_i y_i\right) \\
&= \left[\left(\frac{1}{n}\sum_{i=1}^{n} x_i z_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-1}\left(\sum_{i=1}^{n} z_i x_i'\right)\right]^{-1}\left[\left(\frac{1}{n}\sum_{i=1}^{n} x_i z_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-1}\left(\sum_{i=1}^{n} z_i y_i\right)\right] \\
&= \left[\left(\sum_{i=1}^{n} x_i z_i'\right)\left(\sum_{i=1}^{n} z_i z_i'\right)^{-1}\left(\sum_{i=1}^{n} z_i x_i'\right)\right]^{-1}\left[\left(\sum_{i=1}^{n} x_i z_i'\right)\left(\sum_{i=1}^{n} z_i z_i'\right)^{-1}\left(\sum_{i=1}^{n} z_i y_i\right)\right] \\
&= \left(X'Z\left(Z'Z\right)^{-1} Z'X\right)^{-1} X'Z\left(Z'Z\right)^{-1} Z'y \\
&= b_{2SLS}
\end{aligned}$$

and

$$\sqrt{n}\left(b_{2SLS} - \beta\right) \xrightarrow{d} N\left(0, \Omega\right)$$

where

$$\Omega = \sigma^2\left[\left(E\left[x_i z_i'\right]\right)\left(E\left[z_i z_i'\right]\right)^{-1}\left(E\left[z_i x_i'\right]\right)\right]^{-1}$$

Theorem 1 $\widehat{\Omega} = \Omega + o_p(1)$, *where*

$$\begin{aligned}
\widehat{\Omega} &= \widehat{A}^{-1}\widehat{B}\left(\widehat{A}'\right)^{-1} \\
\widehat{A} &= \left[\left(\frac{1}{n}\sum_{i=1}^{n} x_i z_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right)\right]^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i z_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-1} \\
\widehat{B} &= \frac{1}{n}\sum_{i=1}^{n} \widehat{u}_i^2 z_i z_i' \\
\widehat{u}_i &= y_i - x_i' b = y_i - x_i'\beta - x_i'\beta + x_i' b = u_i - x_i'\left(b - \beta\right)
\end{aligned}$$

Higher Order Asymptotics: Some Preliminaries

In this section, we assume that z_i are non-stochastic. We also assume that

$$\frac{1}{n} Z' \underset{n \times q}{Z} = \frac{1}{n} \sum_i z_i z_i'$$

is <u>fixed</u> at Υ. This simplifies some of the calculations without sacrificing too much generality. Finally, we assume that (ε_i, v_i) is bivariate normal.

Note that the 2SLS $\widehat{\beta}$ is such that

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) = \frac{\left(\frac{1}{n} X'Z\right)\left(\frac{1}{n} Z'Z\right)^{-1}\left(\frac{1}{\sqrt{n}} Z'\varepsilon\right)}{\left(\frac{1}{n} X'Z\right)\left(\frac{1}{n} Z'Z\right)^{-1}\left(\frac{1}{n} Z'X\right)} = \frac{\left(\frac{1}{n} X'Z\right)\Upsilon^{-1}\left(\frac{1}{\sqrt{n}} Z'\varepsilon\right)}{\left(\frac{1}{n} X'Z\right)\Upsilon^{-1}\left(\frac{1}{n} Z'X\right)}$$

Let

$$\begin{aligned}
\omega_1 &\equiv \frac{1}{\sqrt{n}} Z'\varepsilon \sim \mathcal{N}\left(0, \frac{\sigma_\varepsilon^2}{n} Z'Z\right) = \mathcal{N}\left(0, \sigma_\varepsilon^2 \Upsilon\right) = O_p(1) \\
\omega_2 &\equiv \frac{1}{\sqrt{n}} Z'v \sim \mathcal{N}\left(0, \frac{\sigma_v^2}{n} Z'Z\right) = \mathcal{N}\left(0, \sigma_v^2 \Upsilon\right) = O_p(1)
\end{aligned}$$

Writing

$$\frac{1}{n} Z'X = \frac{1}{n} Z'Z\pi + \frac{1}{n} Z'v = \Upsilon\pi + \frac{1}{\sqrt{n}} \omega_2$$

we can rewrite

$$\begin{aligned}
\sqrt{n}\left(\widehat{\beta} - \beta\right) &= \frac{\left(\Upsilon\pi + \frac{1}{\sqrt{n}}\omega_2\right)' \Upsilon^{-1}\omega_1}{\left(\Upsilon\pi + \frac{1}{\sqrt{n}}\omega_2\right)' \Upsilon^{-1}\left(\Upsilon\pi + \frac{1}{\sqrt{n}}\omega_2\right)} \\
&= \frac{(\Upsilon\pi)' \Upsilon^{-1}\omega_1 + \frac{1}{\sqrt{n}}\omega_2' \Upsilon^{-1}\omega_1}{(\Upsilon\pi)' \Upsilon^{-1}(\Upsilon\pi) + \frac{2}{\sqrt{n}}\omega_2' \Upsilon^{-1}(\Upsilon\pi) + \frac{1}{n}\omega_2' \Upsilon^{-1}\omega_2} \\
&= \frac{\pi'\omega_1 + \frac{1}{\sqrt{n}}\omega_2' \Upsilon^{-1}\omega_1}{\pi'\Upsilon\pi + \frac{2}{\sqrt{n}}\omega_2'\pi + \frac{1}{n}\omega_2' \Upsilon^{-1}\omega_2} \\
&\equiv \frac{a_1 + \frac{1}{\sqrt{n}} a_2}{b_1 + \frac{1}{\sqrt{n}}(2b_2) + \frac{1}{n} b_3}
\end{aligned}$$

Rewrite further $s = \frac{1}{\sqrt{n}}$

$$\frac{a_1 + \frac{1}{\sqrt{n}} a_2}{b_1 + \frac{1}{\sqrt{n}}(2b_2) + \frac{1}{n} b_3} = \frac{a_1 + a_2 s}{b_1 + (2b_2)s + b_3 s^2}$$

Because s is small, we can think of a Taylor series approximation at $s = 0$:

$$\frac{a_1 + a_2 s}{b_1 + (2b_2)s + b_3 s^2} = \frac{a_1}{b_1} + \left(\frac{a_2}{b_1} - \frac{2a_1 b_2}{b_1^2}\right) s + o(s)$$

which roughly means that we can rewrite

$$\frac{a_1 + \frac{1}{\sqrt{n}} a_2}{b_1 + \frac{1}{\sqrt{n}}(2b_2) + \frac{1}{n} b_3} \approx \frac{a_1}{b_1} + \frac{1}{\sqrt{n}} \left(\frac{a_2}{b_1} - \frac{2a_1 b_2}{b_1^2} \right)$$

or

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) \approx \frac{\pi'\omega_1}{\pi'\Upsilon\pi} + \frac{1}{\sqrt{n}} \left(\frac{\omega_2'\Upsilon^{-1}\omega_1}{\pi'\Upsilon\pi} - \frac{2(\pi'\omega_1)(\omega_2'\pi)}{(\pi'\Upsilon\pi)^2} \right)$$

Remark 1 *With some abuse of notation, we can write*

$$\pi'\omega_1 \sim \pi'\mathcal{N}\left(0, \sigma_\varepsilon^2 \Upsilon\right) = \mathcal{N}\left(0, \sigma_\varepsilon^2 \pi'\Upsilon\pi\right)$$

so that

$$\frac{\pi'\omega_1}{\pi'\Upsilon\pi} \sim \frac{\mathcal{N}(0, \sigma_\varepsilon^2 \pi'\Upsilon\pi)}{\pi'\Upsilon\pi} = \mathcal{N}\left(0, \frac{\sigma_\varepsilon^2}{\pi'\Upsilon\pi}\right)$$

We therefore have

$$\begin{aligned} E\left[\sqrt{n}\left(\widehat{\beta} - \beta\right)\right] &\approx \frac{E[\pi'\omega_1]}{\pi'\Upsilon\pi} + \frac{1}{\sqrt{n}} \left(\frac{E[\omega_2'\Upsilon^{-1}\omega_1]}{\pi'\Upsilon\pi} - \frac{2E[(\pi'\omega_1)(\omega_2'\pi)]}{(\pi'\Upsilon\pi)^2} \right) \\ &= \frac{1}{\sqrt{n}} \left(\frac{E[\omega_2'\Upsilon^{-1}\omega_1]}{\pi'\Upsilon\pi} - \frac{2E[(\pi'\omega_1)(\omega_2'\pi)]}{(\pi'\Upsilon\pi)^2} \right) \end{aligned}$$

Because

$$\begin{aligned} \omega_2'\Upsilon^{-1}\omega_1 &= \left(\frac{1}{\sqrt{n}} Z'v\right)' \left(\frac{1}{n} Z'Z\right)^{-1} \left(\frac{1}{\sqrt{n}} Z'\varepsilon\right) = v'Z(Z'Z)^{-1} Z'\varepsilon \\ &= \text{trace}\left(v'Z(Z'Z)^{-1} Z'\varepsilon\right) = \text{trace}\left(Z(Z'Z)^{-1} Z' \cdot \varepsilon v'\right) \\ (\pi'\omega_1)(\omega_2'\pi) &= \pi'\left(\frac{1}{\sqrt{n}} Z'\varepsilon\right)\left(\frac{1}{\sqrt{n}} v'Z\right)\pi = \frac{1}{n}\pi' Z'(\varepsilon v') Z\pi \end{aligned}$$

we have

$$\begin{aligned} E\left[\omega_2'\Upsilon^{-1}\omega_1\right] &= \text{trace}\left(Z(Z'Z)^{-1} Z' \cdot E[\varepsilon v']\right) = \text{trace}\left(Z(Z'Z)^{-1} Z' \cdot \sigma_{\varepsilon v} I_n\right) \\ &= \sigma_{\varepsilon v}\,\text{trace}\left(Z(Z'Z)^{-1} Z'\right) = \sigma_{\varepsilon v} q \\ E[(\pi'\omega_1)(\omega_2'\pi)] &= \frac{1}{n}\pi' Z'(E[\varepsilon v'])Z\pi = \frac{1}{n}\pi' Z'(\sigma_{\varepsilon v} I_n) Z\pi \\ &= \sigma_{\varepsilon v}\frac{1}{n}\pi' Z' Z\pi = \sigma_{\varepsilon v}\pi'\Upsilon\pi \end{aligned}$$

we obtain

$$\frac{E[\omega_2'\Upsilon^{-1}\omega_1]}{\pi'\Upsilon\pi} - \frac{2E[(\pi'\omega_1)(\omega_2'\pi)]}{(\pi'\Upsilon\pi)^2} = \frac{q-2}{\pi'\Upsilon\pi}\sigma_{\varepsilon v}$$

and

$$E\left[\sqrt{n}\left(\widehat{\beta}-\beta\right)\right] \approx \frac{1}{\sqrt{n}}\frac{q-2}{\pi'\Upsilon\pi}\sigma_{\varepsilon v}$$

To conclude, we have

$$\begin{aligned} E\left[\widehat{\beta}\right] &\approx \beta + \frac{q-2}{n\cdot\pi'\Upsilon\pi}\sigma_{\varepsilon v} = \beta + \frac{q-2}{\pi' Z' Z\pi}\sigma_{\varepsilon v} \\ &\approx \beta + \frac{1}{\sum_i x_i^2}\frac{q-2}{\mathbb{R}^2}\sigma_{\varepsilon v} \end{aligned}$$

Improvement over 2SLS?

Write

$$\widehat{\beta} = \frac{X'Py}{X'PX}$$

We will consider estimators of the form

$$\widetilde{\beta} \equiv \frac{X'Py - k\cdot X'My}{X'PX - k\cdot X'MX} = \beta + \frac{X'P\varepsilon - k\cdot X'M\varepsilon}{X'PX - k\cdot X'MX}$$

One possibility is LIML, where k is some eigenvalue. Another possibility is

$$k = \frac{\frac{q-2}{n}}{1-\frac{q-2}{n}}$$

This estimator is due to Nagar (1959).

Remark 2 (Intuition) *We noted above that the bias probem of 2SLS may be due to the fact that the expected value of the numerator is not equal to zero. Now, note that*

$$\begin{aligned} E\left[X'P\varepsilon - k\cdot X'M\varepsilon\right] &= E\left[(Z\pi+v)' P\varepsilon - k\cdot(Z\pi+v)' M\varepsilon\right] \\ &= E\left[v'P\varepsilon - k\cdot v'M\varepsilon\right] \\ &= \operatorname{trace}(P)\cdot\sigma_{\varepsilon v} - k\cdot\operatorname{trace}(M)\cdot\sigma_{\varepsilon v} \\ &= (q - k(n-q))\cdot\sigma_{\varepsilon v} \end{aligned}$$

Therefore, if we take

$$k \approx \frac{q}{n-q}$$

we expect the bias problem to be ameliorated.

Rothenberg (1983) and Donald and Newey (2001), based on even higher order expansion than we considered, established that

$$MSE_A(LIML) < MSE_A(Nagar) < MSE_A(2SLS)$$

Here, MSE_A denotes approximate mean squared error.

Digression: Choosing Instruments

Donald and Newey (2001) looked at a more complicated model

$$\begin{aligned} y_i &= x_i\beta + u_i \\ x_i &= f(z_i) + v_i \end{aligned}$$

where $f(\cdot)$ is an unknown function of z_i, and the researcher uses

$$\psi_i^K \equiv \psi^K(z_i) \equiv (\psi_{1K}(z_i), \ldots, \psi_{KK}(z_i))'$$

as an instrument.[1] They showed that the approximate MSEs of $\sqrt{n}$ (estimator $-\beta$) are such that

$$\begin{aligned} MSE_A\left(\sqrt{n}(\text{2SLS} - \beta)\right) &\approx \sigma_\varepsilon^2 \left(\frac{1}{H} + \frac{f'\left(I - P^K\right)f}{nH^2}\right) + \sigma_{\varepsilon v}^2 \frac{K^2}{nH^2} \\ MSE_A\left(\sqrt{n}(\text{LIML} - \beta)\right) &\approx \frac{\sigma_\varepsilon^2}{H}\left(\frac{1}{H} + \frac{f'\left(I - P^K\right)f}{nH^2}\right) + \left(\sigma_\varepsilon^2\sigma_v^2 - \sigma_{\varepsilon v}^2\right)\frac{K}{nH^2} \\ MSE_A\left(\sqrt{n}(\text{Nagar} - \beta)\right) &\approx \frac{\sigma_\varepsilon^2}{H}\left(\frac{1}{H} + \frac{f'\left(I - P^K\right)f}{nH^2}\right) + \left(\sigma_\varepsilon^2\sigma_v^2 + \sigma_{\varepsilon v}^2\right)\frac{K}{nH^2} \end{aligned}$$

Here, $H = E\left[f(z_i)^2\right]$, $f = (f(z_1), \ldots, f(z_n))'$, and P^K denotes the P matrix based on instruments ψ_i^K. In general, K need to be "large" for us to have a good approximation of $f(z_i)$ in the first stage. They suggest an empirically feasible method to choose K in such a way that MSE is minimized. Their proposal will not be discussed explicitly in class.

We now assume a special situation where we have some prior knowledge about $f(z_i)$. Suppose that there is a function $\phi(z_i)$, of which functional form is known, such that $f(z_i) = \pi \cdot \phi(z_i)$ for some unknown π. We can then use only one-dimensional IV $\phi(z_i)$. This will reduce MSE of various IV estimators in two ways. First, we can see that $f'\left(I - P^K\right)f = 0$. This is because we know that $f(z_i) = \pi \cdot \phi(z_i)$ so that the regression of $f(z_i)$ on $\phi(z_i)$ will leave zero residuals. Second, we can see that the additional terms $\sigma_{\varepsilon v}^2 \frac{K^2}{nH^2}$, $\left(\sigma_\varepsilon^2\sigma_v^2 - \sigma_{\varepsilon v}^2\right)\frac{K}{nH^2}$, or $\left(\sigma_\varepsilon^2\sigma_v^2 + \sigma_{\varepsilon v}^2\right)\frac{K}{nH^2}$ will be substantially reduced compared to the case where we have to use $K = 100$, e.g. This idea was exploited by Currie and Gruber (1996), who called $\phi(z_i)$ the "simulated instrument".

Understanding Nagar

2SLS solves

$$0 = \widehat{A}'(y - xb) = \sum_{i=1}^n \widehat{a}_i(y_i - x_i b)$$

where

$$\widehat{A} \equiv Px.$$

[1] For example, we can think of polynomials of z_i including $1, z_i, z_i^2, \ldots, z_i^K$. If K is large, then $f(z_i)$ will be approximated arbitrarily well by the first stage OLS.

Note that

$$\widehat{a}_i = z_i'\widehat{\pi}$$

where

$$\widehat{\pi} = (Z'Z)^{-1} Z'x$$

We would expect good finite sample performance if

$$E[x'P(y - x\beta)] = 0$$

But

$$E[x'P(y - x\beta)] = E[x'P\varepsilon] = E\left[(z\pi + v)' P\varepsilon\right] = E[v'P\varepsilon] = q\sigma_{\varepsilon v} \neq 0$$

The problem can be fixed by considering the fact that

$$\frac{q}{n-q} E[x'M\varepsilon] = \frac{q}{n-q} E[v'M\varepsilon] = q\sigma_{\varepsilon v}$$

and hence

$$E\left[x'\left(P - \frac{q}{n-q}M\right)\varepsilon\right] = 0$$

This suggests an estimator that solves

$$x'\left(P - \frac{q}{n-q}M\right)(y - xb) = 0$$

or

$$\frac{x'Py - \frac{q}{n-q}x'My}{x'Px - \frac{q}{n-q}x'Mx}$$

Note that we can fix the problem by considering a solution to

$$0 = \widetilde{A}'(y - xb) = \sum_{i=1}^{n} \widetilde{a}_i (y_i - x_i b)$$

where

$$\widetilde{a}_i = z_i'\widetilde{\pi}_{(i)}$$

Here, $\widetilde{\pi}_{(i)}$ denotes the OLS estimator for π based on every observation except the ith. Therefore, $\widetilde{\pi}_{(i)}$ and ε_i are independent of each other by construction, and therefore the solution is expected to have a good performance in finite sample. The estimator

$$\frac{\sum_{i=1}^{n}\left(z_i'\widetilde{\pi}_{(i)}\right) y_i}{\sum_{i=1}^{n}\left(z_i'\widetilde{\pi}_{(i)}\right) x_i}$$

is often called the JIVE (Jackknife Instrumental Variables Estimator) in the literature.

Exact Finite Sample Distribution under Normality

We will continue to assume that $\frac{1}{n}\sum_i z_i z_i'$ is fixed at Υ. We will also assume that

$$\begin{pmatrix} \varepsilon_i \\ v_i \end{pmatrix} \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon v} \\ \sigma_{\varepsilon v} & \sigma_v^2 \end{bmatrix}\right)$$

Write

$$\begin{aligned}
\widehat{\beta} &= \beta + \frac{\left(\frac{1}{\sqrt{n}}\sum_i z_i x_i\right)' \left(\frac{1}{n}\sum_i z_i z_i'\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_i z_i \varepsilon_i\right)}{\left(\frac{1}{\sqrt{n}}\sum_i z_i x_i\right)' \left(\frac{1}{n}\sum_i z_i z_i'\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_i z_i x_i\right)} \\
&= \beta + \frac{\left(\left(\frac{1}{n}\sum_i z_i z_i'\right)\left(\sqrt{n}\pi\right) + \frac{1}{\sqrt{n}}\sum_i z_i v_i\right)' \left(\frac{1}{n}\sum_i z_i z_i'\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_i z_i \varepsilon_i\right)}{\left(\left(\frac{1}{n}\sum_i z_i z_i'\right)\left(\sqrt{n}\pi\right) + \frac{1}{\sqrt{n}}\sum_i z_i v_i\right)' \left(\frac{1}{n}\sum_i z_i z_i'\right)^{-1} \left(\left(\frac{1}{n}\sum_i z_i z_i'\right)\left(\sqrt{n}\pi\right) + \frac{1}{\sqrt{n}}\sum_i z_i v_i\right)} \\
&= \beta + \frac{\left(\Upsilon c + \Phi_{zv}\right)' \Upsilon^{-1} \left(\Phi_{z\varepsilon}\right)}{\left(\Upsilon c + \Phi_{zv}\right)' \Upsilon^{-1} \left(\Upsilon c + \Phi_{zv}\right)}
\end{aligned}$$

where we write

$$\begin{aligned}
c &\equiv \sqrt{n}\pi \\
\begin{pmatrix} \Phi_{z\varepsilon} \\ \Phi_{zv} \end{pmatrix} &\equiv \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_i z_i \varepsilon_i \\ \frac{1}{\sqrt{n}}\sum_i z_i v_i \end{pmatrix}
\end{aligned}$$

Note that

$$\begin{pmatrix} \Phi_{z\varepsilon} \\ \Phi_{zv} \end{pmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon v} \\ \sigma_{\varepsilon v} & \sigma_v^2 \end{bmatrix} \otimes \Upsilon\right)$$

Higher Order Expansion for M-Estimator

Consider the M-estimator which solves

$$\frac{1}{n}\sum_{i=1}^{n} s\left(z_i, \widehat{\theta}\right) = 0$$

For simplicity, we will assume that $\dim(s) = \dim(\theta) = 1$.

We first review the usual (first order) asymptotics. From the first order Taylor expansion

$$0 \approx \frac{1}{n}\sum_{i=1}^{n} s\left(z_i, \theta_0\right) + \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\partial s\left(z_i, \theta_0\right)}{\partial \theta}\right)\left(\widehat{\theta} - \theta_0\right)$$

and the law of large numbers

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial s\left(z_i, \theta_0\right)}{\partial \theta} = E\left[\frac{\partial s\left(z_i, \theta_0\right)}{\partial \theta}\right] + o_p(1),$$

we obtain

$$0 \approx \frac{1}{n}\sum_{i=1}^{n} s\left(z_i,\theta_0\right) + E\left[\frac{\partial s\left(z_i,\theta_0\right)}{\partial \theta}\right]\left(\widehat{\theta}-\theta_0\right)$$

or

$$\begin{aligned}
\sqrt{n}\left(\widehat{\theta}-\theta_0\right) &\approx -\left(E\left[\frac{\partial s\left(z_i,\theta_0\right)}{\partial \theta}\right]\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s\left(z_i,\theta_0\right)\right) \\
&= -QJ
\end{aligned}$$

where, for simplicity of notation, we write

$$\begin{aligned}
Q &\equiv \left(E\left[\frac{\partial s\left(z_i,\theta_0\right)}{\partial \theta}\right]\right)^{-1} \\
J &\equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s\left(z_i,\theta_0\right) = O_p(1)
\end{aligned}$$

Higher order expansion is based on the higher order Taylor expansion:

$$\begin{aligned}
0 &\approx \frac{1}{n}\sum_{i=1}^{n} s\left(z_i,\theta_0\right) + \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial s\left(z_i,\theta_0\right)}{\partial \theta}\right)\left(\widehat{\theta}-\theta_0\right) + \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 s\left(z_i,\theta_0\right)}{\partial \theta^2}\right)\left(\widehat{\theta}-\theta_0\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} s\left(z_i,\theta_0\right) + \left(E\left[\frac{\partial s\left(z_i,\theta_0\right)}{\partial \theta}\right] + \frac{1}{\sqrt{n}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\partial s\left(z_i,\theta_0\right)}{\partial \theta} - E\left[\frac{\partial s\left(z_i,\theta_0\right)}{\partial \theta}\right]\right)\right)\left(\widehat{\theta}-\theta_0\right) \\
&\quad + \frac{1}{2}\left(E\left[\frac{\partial^2 s\left(z_i,\theta_0\right)}{\partial \theta^2}\right] + \frac{1}{\sqrt{n}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\partial^2 s\left(z_i,\theta_0\right)}{\partial \theta^2} - E\left[\frac{\partial^2 s\left(z_i,\theta_0\right)}{\partial \theta^2}\right]\right)\right)\left(\widehat{\theta}-\theta_0\right)^2 \\
&= \frac{1}{\sqrt{n}}J + \left(Q^{-1} + \frac{1}{\sqrt{n}}V\right)\left(\widehat{\theta}-\theta_0\right) + \left(H + \frac{1}{\sqrt{n}}W\right)\left(\widehat{\theta}-\theta_0\right)^2
\end{aligned}$$

where, for simplicity of notation, we write

$$\begin{aligned}
V &\equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\partial s\left(z_i,\theta_0\right)}{\partial \theta} - E\left[\frac{\partial s\left(z_i,\theta_0\right)}{\partial \theta}\right]\right) = O_p(1) \\
W &\equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\partial^2 s\left(z_i,\theta_0\right)}{\partial \theta^2} - E\left[\frac{\partial^2 s\left(z_i,\theta_0\right)}{\partial \theta^2}\right]\right) = O_p(1)
\end{aligned}$$

Now "guess"

$$\sqrt{n}\left(\widehat{\theta}-\theta_0\right) \approx -QJ + \frac{1}{\sqrt{n}}\Lambda$$

for some Λ, and obtain

$$\begin{aligned}
0 &\approx J + \left(Q^{-1} + \frac{1}{\sqrt{n}}V\right)\left(\sqrt{n}\left(\widehat{\theta}-\theta_0\right)\right) + \frac{1}{2}\frac{1}{\sqrt{n}}\left(H + \frac{1}{\sqrt{n}}W\right)\left(\sqrt{n}\left(\widehat{\theta}-\theta_0\right)\right)^2 \\
&= J + \left(Q^{-1} + \frac{1}{\sqrt{n}}V\right)\left(-QJ + \frac{1}{\sqrt{n}}\Lambda\right) + \frac{1}{2}\frac{1}{\sqrt{n}}\left(H + \frac{1}{\sqrt{n}}W\right)\left(-QJ + \frac{1}{\sqrt{n}}\Lambda\right)^2 \\
&= \frac{1}{\sqrt{n}}\left(-VQJ + \frac{1}{2}HQ^2J^2 + \frac{1}{Q}\Lambda\right) + O_p\left(\frac{1}{n}\right)
\end{aligned}$$

From

$$0 = -VQJ + \frac{1}{2}HQ^2J^2 + \frac{1}{Q}\Lambda$$

we obtain

$$\Lambda = Q^2JV - \frac{1}{2}Q^3J^2H$$

or

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \approx -QJ + \frac{1}{\sqrt{n}}\left(Q^2JV - \frac{1}{2}Q^3HJ^2\right)$$

We can now calculate the higher order bias of $\widehat{\theta}$ as

$$E\left[-\frac{1}{\sqrt{n}}QJ + \frac{1}{n}\left(Q^2JV - \frac{1}{2}Q^3HJ^2\right)\right] = \frac{1}{n}Q^2E\left[JV\right] - \frac{1}{2}\frac{1}{n}Q^3HE\left[J^2\right]$$