Presentation, Q2014, 3-5 June, Vienna

Instrument Variable Selection in the Calibration Estimator under Nonresponse

Thomas Laitila

Örebro University and Statistics Sweden

thomas.laitila@oru.se

Calib. Est. for nonresponse adjustment

SL

Lundström and Särndal (1999)
Särndal and Lundström (2005)

(Linear calibration, bias reduction)

CK

Kott (2006)
Chang and Kott (2008)
Kott and Chang (2010)

Response probability function known up to unknown parameters.
Estimation of parameters via calibration.

- Consistent if est. of parameters are consistent

Can adjust for informative nonresponse.
SL estimator a special case.

A new consistent Calibration Estimator

$$\hat{Y}_C = \sum_r w_k y_k$$

$$\begin{cases} w_k = d_k v_k \\ v_k = \lambda_r^t z_k \\ \tilde{X} = \sum_r w_k x_k \end{cases}$$

$$w_k = \tilde{X}^t \left(\sum_r d_k z_k x_k^t \right)^{-1} d_k z_k$$

y_k	Study variable
x_k	Auxiliary variable vector, $\mu^t x_k = 1$ assumed
$\tilde{X}$	Population totals of x_k (estimated)
z_k	Instrument variable vector
r	Response set

- SL (2005) estimator defined with $v_k = 1 + \lambda^t z_k$ and is obtained if $\mu^t z_k = 1$.
- $\mu^t z_k = 1$ not assumed for $\hat{Y}_C$.

Instrument vectors

Define z_k so that $\hat{Y}_C$ is consistent.

Example 1:

$$\theta_k = \Pr(k \in r \mid k \in s, s)$$

$$\phi_k = \theta_k^{-1}$$

$$z_k = \phi_k x_k$$

$\hat{Y}_C$ equals the response propensity GREG estimator.

Example 2:

Assisting model for the response set

$$y_k = x_k^t B_U + \eta_k + \omega_k$$

where

$$\sum_U \theta_k x_k \eta_k = \sum_U \theta_k x_k e_k$$

$$e_k = y_k - x_k^t B_U$$

$$\sum_U \theta_k x_k \omega_k = 0$$

Define

$$z_k = x_k - \eta_k \delta$$

$$\delta = \left(\sum_U \theta_k \eta_k^2\right)^{-1} \sum_U \theta_k \eta_k x_k$$

Then $\hat{Y}_C$ is consistent.

In practice

θ_k and η_k unknown

Replace for estimates:

$$\hat{\theta}_k = \Phi(u_k^t \hat{\alpha})$$

$\Phi(\cdot)$=cdf of N(0,1) (Probit)

$$\hat{\eta}_k = f(u_k^t \hat{\alpha}) / \Phi(u_k^t \hat{\alpha})$$

$f(\cdot)$=pdf of N(0,1) (Inverse of Mills Ratio, Heckman, 1979)

Simulation

$$y = \beta_0 + \beta_1 x + \varepsilon$$
$$R = 1(\alpha_0 + \alpha_1 y + \alpha_2 x_2 + \varphi > 0)$$

i) *Population model Norm(ε)/Norm(φ):* (Heckman, 1979) Expected response rates are 60%.

ii) *Population model BinU(ε)/U(φ):*

Binary study variable $y_{obs}=1(y>6,5)$. Response rates are around 58%.

Table 1: Bias and st.dev (in parenthesis) of the $\hat{\bar{Y}}_C = \hat{Y}_C / N$ estimator and the SL calibration estimator $\hat{\bar{Y}}_W = \hat{Y}_W / N$.
Norm(ε)/Norm(φ) population

		Sample size (n)		
Estimator[a)]	ρ [b)]	200	800	1500
$\hat{Y}_W(x1)$	0	-.013	.009	.001
		(.092)	(.045)	(.032)
	0.3	.124	.153	.144
		(.092)	(.045)	(.033)
$\hat{Y}_C(z1)$	0	.002	.003	.000
		(.129)	(.062)	(.045)
	0.3	.223	.230	.229
		(.125)	(.061)	(.045)
$\hat{Y}_C(z2)$	0	-.041	.024	-.006
		(.140)	(.066)	(.046)
	0.3	-.043	.026	-.008
		(.147)	(.069)	(.048)

[a)] Auxiliary/Instrument vectors: $x1=(1\ x)^t$, $z1=\hat{\phi}x1$, $z2= x1-\hat{\delta}\hat{\eta}$, $u=(1\, x\, x_2)^t$
[b)] $\rho = Corr(\alpha_1\varepsilon+\varphi,\varepsilon)$.

Table 2: Bias and st.dev (in parenthesis) of the $\hat{\bar{Y}}_C = \hat{Y}_C / N$ estimator and the SL calibration estimator $\hat{\bar{Y}}_W = \hat{Y}_W / N$.

<u>BinU(ε)/U(φ) population</u>

		Sample size (n)		
Estimator[a)]	ρ [b)]	200	800	1500
$\hat{Y}_W(x1)$	0	-.005	-.000	-.001
		(.036)	(.018)	(.013)
	0.3	.038	.044	.043
		(.038)	(.019)	(.014)
$\hat{Y}_C(z1)$	0	-.003	-.000	-.000
		(.046)	(.023)	(.016)
	0.3	.066	.071	.071
		(.053)	(.026)	(.019)
$\hat{Y}_C(z2)$	0	-.002	-.003	-.002
		(.055)	(.028)	(.020)
	0.3	.002	-.001	-.000
		(.061)	(.031)	(.022)

[a-b)] See Table 1.

Final comments

- The suggested calibration estimator works well in the simulation.
- Can adapt and adjust for informative nonresponse (NMAR).
- Constructed confidence intervals of appropriate coverage (only for $\hat{Y}_C(z2)$).
- Instrument vectors based on Heckman's sample selection model more robust then implied in Heckman (1979).
- Still, alternative instrument vectors of interest.
- Comparison of $\hat{\bar{Y}}_C$ with $\hat{\bar{Y}}_W$ in a Hausman test yields a test of noninformative nonresponse.
- The two-step Heckman estimator is a calibration estimator
- Can be implemented using designweighted IV estimation.

Thanks for Your attention!

If You want to know more:

thomas.laitila@scb.se