Um mæliskekkjur og instrumental breytur

Helgi Tómasson

4. október 2015

Hvenær má framkvæma regression?

- Þ.e. hvenær mun gagnasöfnun á breytunni y og $\boldsymbol{x} = (x_1, \ldots, x_K)$ og mati á tengslum með mati á jöfnunni:

$$y_i = \boldsymbol{x}_i\boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, T.$$

gefa „réttvísandi"mat á stuðlunum $\boldsymbol{\beta}$,. þ.e. á hagrannsóknarmáli hvenær er:

$$\mathrm{plim}(\hat{\boldsymbol{\beta}})) = \boldsymbol{\beta}.$$

- Upprifjun úr hagrannsóknum, geri ráð fyrir að $\boldsymbol{x}_i$ séu „observational"(þ.e. ekki experimental) og stjórnist af einhverju slembiferli (þ.e. x-breytur random).
- Meðal forsendna er að $\boldsymbol{x}_i$ má ekki tengjast ε_i, og að fylkið

$$\mathrm{plim}(\boldsymbol{X}'\boldsymbol{X}/T) = Q_{XX},$$

sé full-rank. Hér er $\boldsymbol{X}$ TxK fylki þar sem fyrst dálkur eru T mælingar á x_1.

- Ef mæliskekkjur eru í $\boldsymbol{X}$ þá brestur forsendan um tengslaleysi skýribreytunnar við ómælda þáttinn ε_i.
- Lausnin sem kennd er í hagrannsóknum (econometríu) er að notast við instrumental breytur, $\boldsymbol{Z}$. Instrumental breyturnar þurfa að vera fleiri en X breyturnar, nægilega óháðar (order condition) og nægilega tengdar X breytunum og ótengdar ε. Á stærðfræðimáli:

$$\begin{aligned} \operatorname{plim}(\boldsymbol{Z}'\boldsymbol{\varepsilon}/T) &= 0, \\ \operatorname{plim}(\boldsymbol{Z}'\boldsymbol{X}/T) = Q_{ZX}, \quad \operatorname{rank} &= K, \\ \operatorname{plim}(\boldsymbol{Z}'\boldsymbol{Z}/T) = Q_{ZZ}, \quad \operatorname{rank} &= M > K. \end{aligned}$$

- $$\hat{\boldsymbol{\beta}}_{IV} \xrightarrow{d} N(\boldsymbol{\beta}, \frac{\sigma^2}{T} Q_{ZX}^{-1} Q_{ZZ} Q_{XZ}^{-1}) \tag{1}$$

- Ef mæliskekkjan er „stór" og instrumentið lélegt er söfnun gagna tilgangslítil.
- Hef gert sýnidæmi þar sem meta á tengsl hæðar og þyngdar. Fyrir liggja gögn úr spurningakönnun þar sem fólk (n manns) hefur gefið upp hæð og þyngd. Það ljúga allir til um hæð þannig að annað hvort er Δ bætt við hæðina eða Δ er dregið frá. Meðalllygin er 0.
- Notendur forritsins geta séð að það að meta:

  $$\text{þyngd}_i = \beta_0 + \beta_1 \text{hæð}_i + u_i,$$

  gefur ekki rétt mat á β_1. Gögnin eru búin til með $\beta_0 = -100$ og $\beta_1 = 1$.
- Þetta lagast ekki þó að n sé stækkað.
- Hægt er að nota skónúmer sem instrument. G.r.f. að lygi um skónúmer sé óháð lygi um hæð. Gæði instruments er stjórnað með `rho`. Stór gildi þýðir mikla fylgni skónúmers og hæðar, þ.e. gott instrument.
- Forritið hermir gögn (`nsim` sinnum) og ber saman eiginleika OLS og IV-metils. Hægt er að stilla fjölda mælinga (`n`) og gæði instruments (`rho`). Hægt er að sjá hvernig asymptotic/(large-sample) formúlan (1) stendur sig. Fræðilega eru bara til large-sample nálgunarformúlur.