

Семинар 17.

1. Докажите теорему о LATE.

Решение:

См. учебник [J.D. Angrist \(2008\). Mostly Harmless Econometrics: An Empiricist's Companion.](#), стр. 112–115.

2. Докажите, что оценка $LATE$ может быть получена в результате оценивания регрессии на бинарную эндогенную переменную воздействия D с помощью двухшагового МНК. В качестве инструментальной переменной рассматривается экзогенная бинарная переменная Z (в примере с повестками D — служба в армии, а Z — это факт получения повестки).

Решение:

Рассмотрим регрессию

$$Y_i = \beta_1 + \beta_2 D_i + \varepsilon_i,$$

в которой коэффициент при бинарной переменной D_i оценивается при помощи 2МНК с бинарной переменной Z_i в качестве инструмента.

Наблюдения можно представить в виде таблицы:

	$D_i = 0$	$D_i = 1$
$Z_i = 0$	Число наблюдений = a Сумма соответствующих значений зависимой переменной $\sum_a Y_i$	Число наблюдений = b Сумма соответствующих значений зависимой переменной $\sum_b Y_i$
$Z_i = 1$	Число наблюдений = c Сумма соответствующих значений зависимой переменной $\sum_c Y_i$	Число наблюдений = d Сумма соответствующих значений зависимой переменной $\sum_d Y_i$

Посчитаем следующие выражения:

$$\begin{aligned} \overline{ZY} - \bar{Z} \cdot \bar{Y} &= \frac{\sum_c Y_i + \sum_d Y_i}{a + b + c + d} - \frac{(c + d)(\sum_a Y_i + \sum_b Y_i + \sum_c Y_i + \sum_d Y_i)}{(a + b + c + d)^2} = \\ &= \frac{(a + b)(\sum_c Y_i + \sum_d Y_i) - (c + d)(\sum_a Y_i + \sum_b Y_i)}{(a + b + c + d)^2}; \\ \overline{DZ} - \bar{D} \cdot \bar{Z} &= \frac{d}{a + b + c + d} - \frac{(c + d)(b + d)}{(a + b + c + d)^2} = \\ &= \frac{ad + bd + cd + d^2}{(a + b + c + d)^2} - \frac{bc + cd + bd + d^2}{(a + b + c + d)^2} = \frac{ad - bc}{(a + b + c + d)^2}. \end{aligned}$$

В этом случае оценка коэффициента при переменной равна:

$$\hat{\beta}_2 = \frac{\overline{ZY} - \bar{Z} \cdot \bar{Y}}{\overline{DZ} - \bar{D} \cdot \bar{Z}} = \frac{(a+b)(\sum_c Y_i + \sum_d Y_i) - (c+d)(\sum_a Y_i + \sum_b Y_i)}{ad - bc}$$

С другой стороны, оценка $LATE$ составляет:

$$\widehat{LATE} = \frac{\bar{Y}_1 - \bar{Y}_0}{\bar{D}_1 - \bar{D}_0}$$

где \bar{Y}_1 — среднее значение зависимой переменной для индивидов, которые получили предписание; \bar{Y}_0 — среднее значение зависимой переменной для индивидов, которые не получили предписание; \bar{D}_1 — доля тех, кто подвергся воздействию, среди тех, кто получил предписание. В нашем примере это доля победителей лотереи, которые пошли служить; \bar{D}_0 — доля тех, кто подвергся воздействию, среди тех, кто не получил предписание;

$$\begin{aligned} \bar{D}_1 - \bar{D}_0 &= \frac{d}{c+d} - \frac{b}{a+b} = \frac{ad - bc}{(a+b)(c+d)} \\ \bar{Y}_1 - \bar{Y}_0 &= \frac{\sum_c Y_i + \sum_d Y_i}{c+d} - \frac{\sum_a Y_i + \sum_b Y_i}{a+b} = \\ &= \frac{(a+b)(\sum_c Y_i + \sum_d Y_i) - (c+d)\sum_a Y_i - \sum_b Y_i}{(a+b)(c+d)} \\ \widehat{LATE} &= \frac{\bar{Y}_1 - \bar{Y}_0}{\bar{D}_1 - \bar{D}_0} = \frac{(a+b)(\sum_c Y_i + \sum_d Y_i) - (c+d)\sum_a Y_i - \sum_b Y_i}{ad - bc} = \hat{\beta}_2. \end{aligned}$$

Что и требовалось доказать.

Список использованных источников

- (а) Картаев Ф.С. Введение в эконометрику : Учебник / Ф.С. Картаев — Москва : МГУ, 2019. — 472 с. — ISBN 978-5-906932-22-8.