Семинар 16.

1. Докажите, что оценка эффекта воздействия $Y_i(1) - Y_i(0)$ может быть получена при помощи обычной парной регрессии вида:

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

где D_i — бинарная переменная, равная 1, если i-й объект вошел в группу, подвергшуюся воздействию (treatment group).

Решение:
Нужно показать, что МНК-оценка коэффициента при переменной в регрессии на бинарную переменную $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 \cdot D_i$ равна:

$$\hat{\beta}_2 = \overline{Y_1} - \overline{Y_0},$$

где $\overline{Y_1} = \frac{\sum_{D_i=1} Y_i}{n_1}$ — среднее выборочное значение зависимой переменной для тех наблюдений, для которых $D_i = 1$ (обозначим число таких наблюдений n_1);
$\overline{Y_0} = \frac{\sum_{D_i=0} Y_i}{n_0}$ — среднее выборочное значение зависимой переменной для тех наблюдений, для которых $D_i = 0$ (обозначим число таких наблюдений n_0).

Таким образом, общее число наблюдений составляет $n_0 + n_1 = n$. Воспользуемся формулой МНК-оценки коэффициента при переменной в модели парной регрессии:

$$\begin{aligned}
\overline{DY} - \bar{D} \cdot \bar{Y} &= \frac{0 \cdot \sum_{D_i=0} Y_i + 1 \cdot \sum_{D_i=1} Y_i}{n} - \frac{n_1}{n} \frac{\sum_{D_i=0} Y_i + \sum_{D_i=1} Y_i}{n} = \\
&= \frac{(n_0 + n_1) \sum_{D_i=1} Y_i - n_1 \sum_{D_i=0} Y_i - n_1 \sum_{D_i=1} Y_i}{n^2} = \\
&= \frac{n_0 \sum_{D_i=1} Y_i - n_1 \sum_{D_i=0} Y_i}{n^2};
\end{aligned}$$

$$\overline{D^2} - (\bar{D})^2 = \frac{n_1}{n} - \frac{n_1^2}{n^2} = \frac{n_1 (n_0 + n_1) - n_1^2}{n^2} = \frac{n_0 n_1}{n^2};$$

$$\begin{aligned}
\hat{\beta}_2 = \frac{\overline{DY} - \bar{D} \cdot \bar{Y}}{\overline{D^2} - (\bar{D})^2} &= \frac{n_0 \sum_{D_i=1} Y_i - n_1 \sum_{D_i=0} Y_i}{n_0 n_1} = \\
&= \frac{\sum_{D_i=1} Y_i}{n_1} - \frac{\sum_{D_i=0} Y_i}{n_0} = \overline{Y_1} - \bar{Y}_0.
\end{aligned}$$

2. Пусть в условиях предыдущей задачи α — доля наблюдений, относящихся к испытуемой группе, а $(1 - \alpha)$ — это соответственно доля наблюдений, относящихся к контрольной группе. Считая, что дисперсия случайной ошибки одинакова для всех наблюдений и равна σ^2, вычислите условную дисперсию МНК–оценки коэффициента при переменной $\operatorname{Var}\left(\hat{\beta}_2 \mid D_1, D_2, \ldots, D_n\right)$ (выразите ее

через σ^2, α, n). Какой должна быть доля наблюдений, относящихся к испытуемой группе, в общем числе наблюдений, чтобы МНК—оценка была наиболее точной?

Решение:
Дисперсия оценки коэффициента наклона в парной регрессии с константой для случая гомоскедастичных ошибок равна:

$$\operatorname{var}\left(\hat{\beta}_2 \mid D_1, D_2, \ldots, D_n\right) = \frac{\sigma^2}{\sum \left(D_i - \bar{D}\right)^2} = \frac{\sigma^2}{\sum \left(D_i - \alpha\right)^2} =$$
$$= \frac{\sigma^2}{n_1 \cdot (1-\alpha)^2 + n_0 \cdot \alpha^2} = \frac{\sigma^2}{n \cdot \alpha \cdot (1-\alpha)^2 + n \cdot (1-\alpha) \cdot \alpha^2} = \frac{\sigma^2}{n \cdot \alpha \cdot (1-\alpha)}.$$

Выражение $\alpha \cdot (1-\alpha)$ максимально при $\alpha = 1/2$. В этом случае дисперсия оценки будет минимальной. Следовательно, для получения максимально точной оценки при заданном объеме выборки необходимо включить в испытуемую группу половину объектов.

3. Исследователь анализирует воздействие закона, запрещающего продажу алкоголя после 23.00, на потребление алкоголя. Исследователь обладает информацией о подушевом потреблении алкоголя в восьми регионах в 2014 и 2015 гг. В 2014 г. во всех регионах алкоголь продавался без ограничений. В 2015 г. в регионах A, B, C, D был введен указанный закон, а в остальных регионах он не применялся. Данные о потреблении алкоголя (литров на человека в год) приведены в таблице.

Регион	***A***	***B***	***C***	***D***	***E***	***F***	***G***	***H***
2014	6	6	8	4	4	3	3	2
2015	6	8	9	5	6	5	5	4

a) Исследователь использует для оценки интересующего его влияния модель с фиксированными эффектами для регионов:

$$y_{it} = \beta x_{it} + \alpha_i + \varepsilon_{it},$$

где α_i — фиксированный эффект i-го региона (позволяет учесть особенности регионов);
x_{it} — фиктивная переменная, равная единице, если в i-м регионе в году t действовал закон об ограничении продажи алкоголя, и равная нулю в противном случае;
y_{it} — потребление алкоголя на душу населения в i-м регионе в году t. Используя внутригрупповое преобразование, найдите оценку параметра β и интерпретируйте полученный результат.

Id	t	Y	X	y^*	x^*	x^*y^*	$(x^*)^\wedge 2$
1	0	6	0	0	$-0,5$	0	$0,25$
1	1	6	1	0	$0,5$	0	$0,25$
2	0	6	0	-1	$-0,5$	$0,5$	$0,25$
2	1	8	1	1	$0,5$	$0,5$	$0,25$
3	0	8	0	$-0,5$	$-0,5$	$0,25$	$0,25$
3	1	9	1	$0,5$	$0,5$	$0,25$	$0,25$
4	0	4	0	$-0,5$	$-0,5$	$0,25$	$0,25$
4	1	5	1	$0,5$	$0,5$	$0,25$	$0,25$
5	0	4	0	-1	0	0	0
5	1	6	0	1	0	0	0
6	0	3	0	-1	0	0	0
6	1	5	0	1	0	0	0
7	0	3	0	-1	0	0	0
7	1	5	0	1	0	0	0
8	0	2	0	-1	0	0	0
8	1	4	0	1	0	0	0
					сумма	2	2

б) Теперь оцените эффект воздействия закона об ограничении продаж алкоголя, используя метод разности разностей. Интерпретируйте полученный результат. Дайте графическую иллюстрацию решения (не забудьте указать на рисунке координаты всех ключевых точек, а также величину эффекта воздействия).

в) Чем может быть вызвано подобное расхождение оценок?

Решение:

(a) Внутригрупповое преобразование позволяет устранить из модели индивидуальные эффекты (их оценка нас не интересует) следующим образом:

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i),$$

где $\bar{y}_i = \sum_t y_{it}$, $\bar{x}_i = \sum_t x_{it}$, $\bar{\varepsilon}_i = \sum_t \varepsilon_{it}$.

Теперь мы работаем с моделью парной регрессии без константы. МНК-оценка коэффициента наклона в такой модели будет иметь вид:

$$\hat{\beta} = \frac{\sum\sum x_{it}^* y_{it}^*}{\sum\sum (x_{it}^*)^2} = \frac{2}{2} = 1.$$

Мы получили, что введение закона приводит к увеличению потребления алкоголя на 1 литр на человека в год.

(б)

$$\begin{aligned}
\bar{Y}_{\text{treatment, before}} &= \frac{6+6+8+4}{4} = 6;\\
\bar{Y}_{\text{treatment, after}} &= \frac{6+8+9+5}{4} = 7;\\
\bar{Y}_{\text{control, before}} &= \frac{4+3+3+2}{4} = 3;\\
\bar{Y}_{\text{control, after}} &= \frac{6+5+5+4}{4} = 5;\\
\hat{\delta} = \left[\bar{Y}_{\text{treatment, after}} - \bar{Y}_{\text{treatment, before}}\right] &- \left[\bar{Y}_{\text{control, after}} - \bar{Y}_{\text{control, before}}\right] = -1.
\end{aligned}$$

Таким образом, введение закона приводит к снижению потребления алкоголя на 1 литр на человека в год.

(в) В спецификации модели из пункта (а) не учитывается временной эффект. Но если внимательно посмотреть на данные, то можно отметить рост потребления алкоголя с течением времени. Метод «разность разностей» учитывает этот эффект.

Список использованных источников

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