

Эконометрика–2

магистратура 1к ММАЭ 2010–2011

Семинар 21

Модели с урезанными и цензурированными выборками

15 марта 2011

“data_alcohol&tobacco.dta” – данные по расходам домохозяйств на табак и алкоголь.

Для анализа доли бюджетных расходов на табак и алкоголь у вас есть набор данных, содержащий результаты опроса 2724 семей в Бельгии, собранных Национальным институтом статистики Бельгии в 1995–96 гг. Рассмотрим один из продуктов: алкоголь.

Набор переменных:

Age – Возраст главы семейства: 1-4 категории (по увеличению возраста)

Flanders = 1 если семья живет в Фламандии, = 0 иначе

Walloon = 1 если семья живет в Валлонии, = 0 иначе

Bluecol = 1 глава семейства blue collar worker (работники, занятые на производстве), = 0 иначе

Whitecol = 1 глава семейства white collar (работники, не занятые непосредственным физическим трудом, служащие, управляющий персонал), = 0 иначе

Nadults – Количество взрослых членов семьи

Ninfants – Количество детей возраста до 2х лет в семье

Nkids – Количество детей возраста 2-18 лет в семье

Lnх – Логарифм общих расходов

Shalc – Доля расходов на алкоголь в общих расходах

Alc – Потребление алкоголя: 1 если shalc > 0, 0 - иначе

Shtob – Расходы на табак в общих расходах

Рассмотрим долю расходов на алкоголь.

NB! МНК дает смещенные и несостоятельные оценки в модели $shalc_i = x_i'\beta + \varepsilon_i$, т.к. нарушается свойство $E(\varepsilon_i) = 0$.

1 Цензурированные выборки. Tobit model.

$$shalc_i^* = x_i' \beta + \varepsilon_i \quad (1)$$

$$shalc_i = \begin{cases} shalc_i^* & \text{if } shalc_i^* > 0 \\ 0 & \text{if } shalc_i^* \leq 0 \end{cases} \quad (2)$$

$$E(y_i) = \Phi\left(\frac{x_i' \beta}{\sigma}\right) x_i' \beta + \sigma \phi\left(\frac{x_i' \beta}{\sigma}\right) \quad (3)$$

$$\varepsilon_i \sim N(0, \sigma^2) \quad (4)$$

tobit shalc age flanders bluecol whitecol nadults ninfa nkids lnx, ll

или

tobit shalc age flanders bluecol whitecol nadults ninfa nkids lnx, ll(0) ul(1)

Прогноз из Тобит-модели:

predict shalc_tobit

sum shalc_tobit

3 вида предельных эффектов:

1. Latent variable $\frac{\partial E(y^*|x)}{\partial x} = \beta$

2. Left-truncated (at 0): $\frac{\partial E(y|x, y>0)}{\partial x} = (1 - \omega \lambda(\omega) - \lambda(\omega)^2) \beta$

mfх, predict(e(.,.)) или **mfх, predict(e(0,.))** или **mfх, predict(e(0,1))**

3. Left-censored (at 0): $\frac{\partial E(y|x)}{\partial x} = \Phi(\omega) \beta$

mfх, predict(ystar(0,.)) или **mfх, predict(ystar(0,.))**

где $\omega = \frac{x' \beta}{\sigma}$, $\lambda(\omega) = \frac{\phi(\omega)}{\Phi(\omega)}$

2 Модель Хекмана. Selection model.

Ограничение Tobit-модели: наблюдение $shalc_i = 0$ можно интерпретировать как отказ от участия “в потреблении алкоголя”. $shalc_i > 0$ измеряет интенсивность участия. В Tobit-модели одни и те же факторы влияют как на вероятность, так и на интенсивность участия. Однако одинаковые факторы могут влиять на вероятность и интенсивность далеко не во всех случаях (кроме того, не факт, что факторы влияют однонаправленно).

Основное уравнение:

$$shalc_i^* = x_i' \beta + \varepsilon_i \quad (5)$$

$$shalc_i = \begin{cases} shalc_i^* & \text{if } alc_i = 1 \\ \emptyset & \text{if } alc_i = 0 \end{cases} \quad (6)$$

Уравнение участия:

$$alc_i^* = z_i' \gamma + u_i \quad (7)$$

$$alc_i = \begin{cases} alc_i^* & \text{if } alc_i^* \geq 0 \\ 0 & \text{if } alc_i^* < 0 \end{cases} \quad (8)$$

$$\begin{pmatrix} \varepsilon_i \\ u_i \end{pmatrix} \sim N \left(0, \begin{pmatrix} \sigma_\varepsilon^2 & \rho\sigma_\varepsilon \\ \rho\sigma_\varepsilon & 1 \end{pmatrix} \right) \quad (9)$$

$\sigma_{\varepsilon u} = \sigma_\varepsilon \rho$, если $\rho = 0$ (не отличается статистически значимо) – можно использовать линейную модель вместо модели Хекмана, т.к. нет самоотбора участников.

$$E(shalc_i | x_i, alc_i = 1) = x_i' \beta + \sigma_{\varepsilon u} \lambda(z_i' \gamma) \quad (10)$$

“Лямбда Хекмана” (Heckman lambda)

$$\lambda(z_i' \gamma) = \frac{\phi(z_i' \gamma)}{\Phi(z_i' \gamma)} \quad (11)$$

2.1 Модель Хекмана (5)–(9), оцененная с помощью метода максимального правдоподобия (ММП)

```
heckman shalc age nadults nkids lnx walloon, select(alc=age nadults nkids lnx walloon)
```

2.2 Двухшаговая модель Хекмана

Применение ММП для оценивания модели Хекмана требует, как правило, написания программы и существование решения оптимизационной задачи. В эмпирических исследованиях иногда ограничиваются двухшаговым методом оценивания, который позволяет получить несмещенные и состоятельные оценки параметров модели.

1. состоятельные оценки $\hat{\gamma}$ из бинарной модели (7)–(8)
2. уравнение (10) оценивается с учетом регрессора $\lambda(z_i' \hat{\gamma})$

Недостаток модели: неэффективные оценки, т.к. ошибки в (10) гетероскедастичны. Оценки двухшагового метода иногда используют при оценивании с помощью ММП в качестве начальных значений.

```
heckman shalc age nadults nkids lnx walloon, select(alc=age nadults nkids lnx walloon) twostep
```