Эконометрический ликбез: ограниченные зависимые переменные

Оценивание моделей дискретного выбора и моделей с цензурированием*

Эрик Бьорн†

Университет Осло, Осло, Норвегия

В настоящих заметках содержится обзор вопросов спецификации модели, функции правдоподобия и структуры задач максимального правдоподобия для моделей дискретного выбора и моделей с цензурированием. Первая часть касается оценивания в случае одного уравнения с одномерными (кросс-секционными) наблюдениями. Другая часть расширяет постановку на случай двух уравнений. Последняя часть рассматривает расширение на ситуацию панельных данных.

1 Одномерные (кросс-секционные) данные

1.1 Отправная точка

Отправной точкой является следующее уравнение:

$$y_i^* = \boldsymbol{x}_i\boldsymbol{\beta} + \sigma\varepsilon_i, \quad \varepsilon_i \sim \mathsf{IIN}(0,1), \quad i = 1,\ldots,N, \tag{1}$$

где i обозначает номер наблюдения, IIN символизирует «одинаково, независимо и нормально распределены», y_i^* – значение эндогенной переменной для наблюдения i, $\boldsymbol{x}_{it}$ – вектор-строка наблюдаемых ковариат (экзогенных переменных), $\boldsymbol{\beta}$ – вектор-столбец коэффициентов, σ – положительная константа, и ε_i – ненаблюдаемый случайный шум. Мы не наблюдаем $(y_i^*, \boldsymbol{x}_i)$ для всех i. *Различия между тремя моделями, представленными ниже, определяются тем, как наблюдаются пары* (y_i^*, x_i). Эти три модели представляют из себя вариации на одну тему: *модели с ограниченной наблюдаемостью эндогенных переменных.* Мы будем обозначать наблюдения как (y_i, x_i). Нашей целью является несмещенная инференция о $\boldsymbol{\beta}$, т.е. об эффекте изменений в $\boldsymbol{x}_i$ на y_i^*, исходя из множества наблюдений $\{y_i, \boldsymbol{x}_i\}_{i=1}^n$.

1.2 Что мы наблюдаем?

Мы рассмотрим три случая, отличающихся по тому, как наблюдается латентная эндогенная переменная y_i^*.

Случай 1: Случай дискретного выбора. Мы наблюдаем $\boldsymbol{x}_i$ и

$$y_i = \begin{cases} 1 \quad \text{для} \quad y_i^* > 0 \quad \Leftrightarrow \quad -\varepsilon_i < \dfrac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}, \\ 0 \quad \text{для} \quad y_i^* \leq 0 \quad \Leftrightarrow \quad -\varepsilon_i \geq \dfrac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}, \end{cases} \qquad i = 1,\ldots,N. \tag{2}$$

*Перевод С. Анатольева. Цитировать как: Бьорн, Эрик (2009) «Оценивание моделей дискретного выбора и моделей с цензурированием», Квантиль, №6, стр. 49–57. Citation: Biørn, Erik (2009) "Estimation of discrete choice and censoring models," Quantile, No.6, pp. 49–57.

†Адрес: Department of Economics, University of Oslo, P.O. Box 1095 Blindern, 0317 Oslo, Norway. Электронная почта: *erik.biorn@econ.uio.no*

Формально, y_i – это ступенчатая функция от y_i^*, со *ступенькой в нуле*. Если мы в общем случае определим функцию $z = \mathbb{I}\{\mathcal{A}\}$, равную единице если событие $\mathcal{A}$ верно и нулю если событие $\mathcal{A}$ неверно, мы сможем записать (2) компактно как

$$y_i = \mathbb{I}\{y_i^* > 0\} = \mathbb{I}\{-\varepsilon_i < \boldsymbol{x}_i\boldsymbol{\beta}/\sigma\}, \qquad i = 1, \ldots, N.$$

Случай 2: Случай с цензурированием. В этом Случае мы предполагаем, что знаем больше об y_i^*, чем в Случае 1. Мы наблюдаем $\boldsymbol{x}_i$ и

$$y_i = \max\{y_i^*, 0\} = \begin{cases} y_i^* & \text{для} \quad y_i^* > 0 \quad \Leftrightarrow \quad -\varepsilon_i < \dfrac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}, \\ 0 & \text{для} \quad y_i^* \leq 0 \quad \Leftrightarrow \quad -\varepsilon_i \geq \dfrac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}, \end{cases} \qquad i = 1, \ldots, N. \tag{3}$$

Особенность этого Случая в том, что y_i *наблюдаема частично непрерывно (для* $y_i^* = y_i > 0$*), а частично дискретно (для* $y_i^* \leq 0$, $y_i = 0$*)*. Формально, y_i – непрерывная функция от y_i^*, с *изломом в нуле*. Наблюдения по y_i характеризуются *нагромождением нулей*.

Случай 3: Случай с отсечением. В этом Случае мы знаем меньше, чем в Случае 2, но больше, чем в Случае 1, только если наблюдается положительное y_i. У нас нет наблюдений для всех N значений i. Наблюдения определяются следующим образом:

$$\begin{array}{lll} (y_i, \boldsymbol{x}_i) = (y_i^*, \boldsymbol{x}_i) \text{ и наблюдаема,} & \text{если} \quad y_i^* > 0 \quad \Leftrightarrow \quad -\varepsilon_i < \dfrac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}, \\ (y_i, \boldsymbol{x}_i) \text{ ненаблюдаема,} & \text{если} \quad y_i^* \leq 0 \quad \Leftrightarrow \quad -\varepsilon_i \geq \dfrac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}, \end{array} \qquad i = 1, \ldots, N. \tag{4}$$

В этом Случае процесс, определяющий, будут ли у нас наблюдения для конкретного значения i или нет, случаен, и *этот выбор есть результат решений респондентов, определяемых рассматриваемой моделью*, (1).

1.3 Вероятностная структура откликов в Случаях 1 и 2

Случай 1: Модель дискретного выбора

Точечные вероятности двух возможных исходов для y_i, условно на $\boldsymbol{x}_i$, равны

$$\begin{array}{l} \mathbb{P}\{y_i = 1 \,|\, \boldsymbol{x}_i\} = \mathbb{P}\left\{-\varepsilon_i < \dfrac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma} \middle| \boldsymbol{x}_i\right\} = \Phi\left(\dfrac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right) \equiv \mathcal{L}_{1i} \\ \mathbb{P}\{y_i = 0 \,|\, \boldsymbol{x}_i\} = \mathbb{P}\left\{-\varepsilon_i \geq \dfrac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma} \middle| \boldsymbol{x}_i\right\} = 1 - \Phi\left(\dfrac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right) \equiv \mathcal{L}_{0i}, \end{array} \qquad i = 1, \ldots, N. \tag{5}$$

где $\Phi(\cdot)$ – КФР (кумулятивная функция распределения) распределения $\mathsf{N}(0,1)$, $\equiv$ следует интерпретировать как равенство по определению, и где первый индекс у функций $\mathcal{L}$ обозначает «Режим 1» когда $y_i = 1$, и «Режим 0» когда $y_i = 0$, соответственно. (Заметим, что мы здесь используем тот факт, что $-\varepsilon_i$ и ε_i имеют одну и ту же функцию плотности, поскольку нормальное распределение симметрично.)

Часть функции правдоподобия, «относящуюся» к наблюдению i, можно переписать как

$$\mathcal{L}_i \equiv \mathcal{L}_{1i}^{y_i}\mathcal{L}_{0i}^{1-y_i} \equiv \begin{cases} \mathcal{L}_{1i} & \text{для} \quad y_i = 1, \\ \mathcal{L}_{0i} & \text{для} \quad y_i = 0, \end{cases} \qquad i = 1, \ldots, N. \tag{6}$$

Случай 2: Регрессионная модель с цензурированием

Рассмотрим вначале Режим 1, в котором y_i непрерывна и имеет ту же КФР, что и у y_i^*:

$$\Phi\left(\frac{y_i - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right) = \Phi\left(\frac{y_i^* - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right) \tag{7}$$

и плотность, выводимую дифференцированием (7) по y_i для $y_i > 0$:

$$\frac{1}{\sigma}\phi\left(\frac{y_i - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right) = \frac{1}{\sigma}\phi\left(\frac{y_i^* - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right) \equiv \mathcal{M}_{1i}, \tag{8}$$

где $\phi(\cdot) \equiv \Phi'(\cdot)$.

Рассмотрим теперь Режим 0, в котором y_i наблюдается дискретно. Этот Режим совпадает с откликом $y_i = 0$ в Случае 1. Тогда у y_i нет плотности, а есть вероятностная масса, которую можно получить из КФР y_i^* следующим образом (см. вторую часть (5)):

$$\mathbb{P}\{y_i = 0 \,|\, \boldsymbol{x}_i\} = \mathbb{P}\{y_i^* \leq 0 \,|\, \boldsymbol{x}_i\} = \mathbb{P}\left\{-\varepsilon_i \geq \frac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\middle| \boldsymbol{x}_i\right\} = 1-\Phi\left(\frac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right) \equiv \mathcal{M}_{1i}. \tag{9}$$

Часть функции правдоподобия, «относящаяся» к наблюдению i, теперь выглядит как

$$\mathcal{M}_i \equiv \begin{cases} \mathcal{M}_{1i} & \text{для} \quad y_i > 0, \\ \mathcal{M}_{0i} & \text{для} \quad y_i = 0, \end{cases} \qquad i = 1, \ldots, N. \tag{10}$$

Эта функция, таким образом, является смесью функций плотности и КФР.

1.4 Задача максимизации правдоподобия в Случаях 1 и 2

Случай 1: Модель дискретного выбора

Поскольку наши N наблюдений независимы, полная функция правдоподобия является произведением функций правдоподобия в (6) для всех наблюдений, что дает

$$\mathcal{L} \equiv \prod_{i=1}^{N} \mathcal{L}_i \equiv \prod_{i:y_i=1} \mathcal{L}_{1i} \prod_{i:y_i=0} \mathcal{L}_{0i}. \tag{11}$$

Максимизируя $\mathcal{L}$, или, что несколько проще, максимизируя

$$\ln(\mathcal{L}) \equiv \sum_{i=1}^{N} \ln(\mathcal{L}_i) = \sum_{i:y_i=1} \ln(\mathcal{L}_{1i}) + \sum_{i:y_i=0} \ln(\mathcal{L}_{0i}) \equiv H\left(\frac{\boldsymbol{\beta}}{\sigma}\right)$$

по $\boldsymbol{\beta}/\sigma$, получаем оценки максимального правдоподобия этих параметров. Эту задачу приходится решать численно.

Решение данной максимизиционной задачи подразумевает интегрирование: КФР распределения $\mathsf{N}(0,1)$, $\Phi(\cdot)$, определяется как некий интеграл. Заметим, что *мы не можем оценить уровень вектора коэффициентов $\boldsymbol{\beta}$, а можем лишь отношение этого вектора к стандартному отклонению возмущения в (1).* Интерпретация: можно измерить (идентифицировать) лишь отклик на переменные в $\boldsymbol{x}_i$ в «единицах стандартного отклонения». Это следствие того, что наблюдаемы лишь качественные свойства y_i^*.

Случай 2: Регрессионная модель с цензурированием

Поскольку наши N наблюдений независимы, полная функция правдоподобия является произведением функций правдоподобия в (10) для всех наблюдений, что дает

$$\mathcal{M} \equiv \prod_{i=1}^{N} \mathcal{M}_i = \prod_{i:y_i>0} \mathcal{M}_{1i} \prod_{i:y_i=0} \mathcal{M}_{0i}. \tag{12}$$

Максимизируя $\mathcal{M}$, или, что несколько проще, максимизируя

$$\ln(\mathcal{M}) \equiv \sum_{i=1}^{N} \ln(\mathcal{M}_i) = \sum_{i:y_i>0} \ln(\mathcal{M}_{1i}) + \sum_{i:y_i=0} \ln(\mathcal{M}_{0i}) \equiv G(\boldsymbol{\beta}, \sigma)$$

по $(\boldsymbol{\beta}, \sigma)$, получаем оценки максимального правдоподобия этих параметров.

Еще раз отметим, что решение подразумевает интегрирование: КФР $\mathsf{N}(0,1)$-распределения, $\Phi(\cdot)$, реализующегося в Режиме 0, определяется как некий интеграл. *В данном случае мы уже можем оценить абсолютное значение вектора коэффициентов $\boldsymbol{\beta}$ вместе с параметром σ.* Это происходит благодаря тому, что, в отличие от полностью дискретного Случая 1, y_i^* на некотором отрезке наблюдаема как количественная (непрерывная) переменная. Этого достаточно для раздельной идентификации $\boldsymbol{\beta}$ и σ.

2 Первое расширение: модель из двух уравнений

Следующая модель, которую мы рассмотрим, состоит из двух уравнений в форме (1):

$$\begin{array}{l} y_{1i}^* = \boldsymbol{x}_{1i}\boldsymbol{\beta}_1 + \sigma_1\varepsilon_{1i}, \\ y_{2i}^* = \boldsymbol{x}_{2i}\boldsymbol{\beta}_2 + \sigma_2\varepsilon_{2i}, \end{array} \qquad \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \tag{13}$$

$$\begin{aligned} y_{1i} &= \begin{cases} y_{1i}^* & \text{если} \quad y_{1i}^* > 0, \\ 0 & \text{если} \quad y_{1i}^* \le 0, \end{cases} \\ y_{2i} &= \begin{cases} y_{2i}^* & \text{если} \quad y_{1i}^* > 0, \\ 0 & \text{если} \quad y_{1i}^* \le 0, \end{cases} \end{aligned} \tag{14}$$

где (y_{1i}, y_{2i}) цензурированы в нуле в зависимости от знака y_{1i}^*: строго положительные значения (y_{1i}, y_{2i}) наблюдаемы только если $y_{1i} > 0$.

Из (13)–(14) находим, что

$$\begin{aligned} \mathbb{E}(y_{1i}|y_{1i}>0) &= \boldsymbol{x}_{1i}\boldsymbol{\beta}_1 - \sigma_1\mathbb{E}[-\varepsilon_{1i}| - \varepsilon_{1i} < \boldsymbol{x}_{1i}(\boldsymbol{\beta}_1/\sigma_1)] \\ &\equiv \boldsymbol{x}_{1i}\boldsymbol{\beta}_1 - \sigma_1\mathbb{E}[\varepsilon_{1i}|\varepsilon_{1i} < \boldsymbol{x}_{1i}(\boldsymbol{\beta}_1/\sigma_1)], \end{aligned} \tag{15}$$

$$\begin{aligned} \mathbb{E}(y_{2i}|y_{1i}>0) &= \boldsymbol{x}_{2i}\boldsymbol{\beta}_2 - \sigma_2\mathbb{E}[-\varepsilon_{2i}| - \varepsilon_{1i} < \boldsymbol{x}_{1i}(\boldsymbol{\beta}_1/\sigma_1)] \\ &\equiv \boldsymbol{x}_{2i}\boldsymbol{\beta}_2 - \sigma_2\mathbb{E}[\varepsilon_{2i}|\varepsilon_{1i} < \boldsymbol{x}_{1i}(\boldsymbol{\beta}_1/\sigma_1)], \end{aligned} \tag{16}$$

$$\begin{aligned} \mathbb{E}(y_{2i}|y_{1i}=0) &= \boldsymbol{x}_{2i}\boldsymbol{\beta}_2 - \sigma_2\mathbb{E}[-\varepsilon_{2i}| - \varepsilon_{1i} > \boldsymbol{x}_{1i}(\boldsymbol{\beta}_1/\sigma_1)] \\ &\equiv \boldsymbol{x}_{2i}\boldsymbol{\beta}_2 - \sigma_2\mathbb{E}[\varepsilon_{2i}|\varepsilon_{1i} > \boldsymbol{x}_{1i}(\boldsymbol{\beta}_1/\sigma_1)]. \end{aligned} \tag{17}$$

Поскольку предположение о нормальности в (13) влечет за собой

$$\mathbb{E}[\varepsilon_{ji}|\varepsilon_{ji}<a] = -\frac{\phi(a)}{\Phi(a)}, \quad \mathbb{E}[\varepsilon_{ji}|\varepsilon_{ji}>a] = \frac{\phi(a)}{1-\Phi(a)}, \quad j=1,2;\ a\in(-\infty,+\infty),$$

$$\mathbb{E}[\varepsilon_{ji}|\varepsilon_{ki}] = \rho\,\varepsilon_{ki}, \quad j,k=1,2;\ j\ne k,$$

то из (15)–(17) следует в результате использования правила повторных ожиданий, что

$$\mathbb{E}(y_{1i}|y_{1i}>0) = \boldsymbol{x}_{1i}\boldsymbol{\beta}_1 + \sigma_1\lambda_{Ai}, \tag{18}$$

$$\mathbb{E}(y_{2i}|y_{1i}>0) = \boldsymbol{x}_{2i}\boldsymbol{\beta}_2 + \rho\sigma_2\lambda_{Ai}, \tag{19}$$

$$\mathbb{E}(y_{2i}|y_{1i}=0) = \boldsymbol{x}_{2i}\boldsymbol{\beta}_2 - \rho\sigma_2\lambda_{Bi}, \tag{20}$$

где

$$\lambda_{Ai} = \frac{\phi\left(\dfrac{\boldsymbol{x}_{1i}\boldsymbol{\beta}_1}{\sigma_1}\right)}{\Phi\left(\dfrac{\boldsymbol{x}_{1i}\boldsymbol{\beta}_1}{\sigma_1}\right)}, \qquad \lambda_{Bi} = \frac{\phi\left(\dfrac{\boldsymbol{x}_{1i}\boldsymbol{\beta}_1}{\sigma_1}\right)}{1-\Phi\left(\dfrac{\boldsymbol{x}_{1i}\boldsymbol{\beta}_1}{\sigma_1}\right)}. \tag{21}$$

Теперь можно легко объяснить двухшаговую процедуру Хекмана.

Шаг 1: Провести пробит-анализ на первом уравнении (13), задействуя только знак y_{1i} как качественную переменную и наблюдаемые значения $\boldsymbol{x}_{1i}$. Это дает $\widehat{\boldsymbol{\beta}_1/\sigma_1}$, откуда можно найти $\widehat{\lambda}_{Ai}$ и $\widehat{\lambda}_{Bi}$, используя (21).

Шаг 2:
(i) Для *нецензурированных* наблюдений, т.е. тех, для которых $y_{1i} > 0$, прорегрессировать y_{1i} на $\boldsymbol{x}_{1i}$ и $\widehat{\lambda}_{Ai}$, используя (18). Это дает $(\widehat{\boldsymbol{\beta}}_1, \widehat{\sigma}_1)$.
(ii) Для *нецензурированных* наблюдений, т.е. тех, для которых $y_{1i} > 0$, прорегрессировать y_{2i} на $\boldsymbol{x}_{2i}$ и $\widehat{\lambda}_{Ai}$, используя (19). Это дает $(\widehat{\boldsymbol{\beta}}_2, \widehat{\rho\sigma}_2)$. Или же для *цензурированных* наблюдений, т.е. тех, для которых $y_{1i} = 0$, прорегрессировать y_{2i} на $\boldsymbol{x}_{2i}$ и $\widehat{\lambda}_{Bi}$, используя (20). Это дает $(\widehat{\boldsymbol{\beta}}_2, \widehat{\rho\sigma}_2)$.

Обозначим через $f(u_{1i}, u_{2i}; \sigma_1, \sigma_2, \rho)$ плотность возмущений $(u_{1i}, u_{2i}) = (\sigma_1\varepsilon_{1i}, \sigma_2\varepsilon_{2i})$ в (13). Если бы цензурирования не было, то $f(y_{1i} - \boldsymbol{x}_{1i}\boldsymbol{\beta}_1, y_{2i} - \boldsymbol{x}_{2i}\boldsymbol{\beta}_2; \sigma_1, \sigma_2, \rho)$ была бы плотностью (y_{1i}, y_{2i}) (условно на $\boldsymbol{x}_{1i}, \boldsymbol{x}_{2i}$) на всей области (y_{1i}, y_{2i}). Но (y_{1i}, y_{2i}) распределены не непрерывно на той области, где цензурирование действует. В этой ситуации правдоподобия для наблюдений выводятся следующим образом.

Обозначим через $\mathcal{M}_i$ часть функции правдоподобия, относящуюся к наблюдению i. Имеем: $\mathcal{M}_i = f(y_{1i} - x_{1i}\beta_1, y_{2i} - x_{2i}\beta_2; \sigma_1, \sigma_2, \rho)$ для наблюдений с $y_{1i} > 0$ и $\mathcal{M}_i = \mathbb{P}\{y_{1i}^* \leq 0\} = \mathbb{P}\{-\varepsilon_{1i} > \boldsymbol{x}_{1i}\boldsymbol{\beta}_1/\sigma_1\} = 1 - \Phi(\boldsymbol{x}_{1i}\boldsymbol{\beta}_1/\sigma_1)$ для наблюдений с $y_{i1} = 0$.

В предположении о независимости наблюдений все это вместе означает, что полная функция правдоподобия имеет вид

$$\mathcal{M} = \prod_{i:y_{1i}>0} f(y_{1i} - \boldsymbol{x}_{1i}\boldsymbol{\beta}_1, y_{2i} - \boldsymbol{x}_{2i}\boldsymbol{\beta}_2; \sigma_1, \sigma_2, \rho) \prod_{i:y_{1i}=0} \left[1 - \Phi\left(\frac{\boldsymbol{x}_{1i}\boldsymbol{\beta}_1}{\sigma_1}\right)\right], \tag{22}$$

где $\prod_{i:y_{1i}>0}$ и $\prod_{i:y_{1i}=0}$ символизируют произведения по тем значениям i, для которых $y_{1i} > 0$, и по тем значениям i, для которых $y_{1i} = 0$, соответственно. Максимизация $\mathcal{M}$ по неизвестным параметрам дает оценки максимального правдоподобия.

3 Второе расширение: панельные данные

3.1 Отправная точка

Отправной точкой при расширении модели первого раздела на случай (сбалансированных) панельных даных является следующее уравнение:

$$y_{it}^* = \boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i + \sigma\varepsilon_{it}, \quad \varepsilon_{it} \sim \mathsf{IIN}(0,1), \quad i = 1, \ldots, N, \ t = 1, \ldots, T. \tag{23}$$

Мы, правда, не наблюдаем $(y_{it}^*, \boldsymbol{x}_{it})$ для всех (i,t). Предполагается, что α_i – латентный индивидуальный эффект, который можно рассматривать как фиксированный и полностью неизвестный и неструктурированный, или же как случайный и порожденный распределением вероятностей с определенными свойствами. Мы будем придерживаться его интерпретации как случайного эффекта, хотя в определенных местах и будем действовать условно на α_i, что можно рассматривать как подражание случаю с фиксированными эффектами. Нашей целью является несмещенная инференция о $\boldsymbol{\beta}$, т.е., об эффекте изменений в $\boldsymbol{x}_{it}$ на y_{it}^*, исходя из множества наблюдений $\{\{y_{it}, \boldsymbol{x}_{it}\}_{i=1}^N\}_{t=1}^T$.

3.2 Что мы наблюдаем?

Мы рассмотрим три случая, отличающихся по тому, как наблюдается латентная эндогенная переменная y_{it}^*.

Случай 1: Случай дискретного выбора. Мы наблюдаем $\boldsymbol{x}_{it}$ и

$$y_{it} = \begin{cases} 1 \quad \text{для} \quad y^*_{it} > 0 \quad \Leftrightarrow \quad -\varepsilon_{it} \, < \, \dfrac{\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i}{\sigma}, \\ 0 \quad \text{для} \quad y^*_{it} \leq 0 \quad \Leftrightarrow \quad -\varepsilon_{it} \, \geq \, \dfrac{\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i}{\sigma}, \end{cases} \qquad \begin{matrix} i = 1, \ldots, N, \\ t = 1, \ldots, T, \end{matrix} \tag{24}$$

где i индексирует индивида, а t – период времени. Формально, y_{it} – это ступенчатая функция от y^*_{it}, со *ступенькой в нуле*. Можно записать (24) компактно как

$$y_{it} = \mathbb{I}\{y^*_{it} > 0\} = \mathbb{I}\{-\varepsilon_{it} \, < \, (\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i)/\sigma\}, \qquad i = 1, \ldots, N, \; t = 1, \ldots, T.$$

Случай 2: Случай с цензурированием. В этом Случае мы знаем больше об y^*_{it}, чем в Случае 1. Мы наблюдаем $\boldsymbol{x}_{it}$ и

$$y_{it} = \max\{y^*_{it}, 0\} = \begin{cases} y^*_{it} \quad \text{для} \quad y^*_{it} > 0 \quad \Leftrightarrow \quad -\varepsilon_{it} < \dfrac{\boldsymbol{x}_{it}\boldsymbol{\beta}+\alpha_i}{\sigma}, \\ 0 \quad \text{для} \quad y^*_{it} \leq 0 \quad \Leftrightarrow \quad -\varepsilon_{it} \geq \dfrac{\boldsymbol{x}_{it}\boldsymbol{\beta}+\alpha_i}{\sigma}, \end{cases} \quad \begin{matrix} i=1, \ldots, N, \\ t=1, \ldots, T. \end{matrix} \tag{25}$$

Особенность этого Случая в том, что y_{it} *наблюдаема частично непрерывно (для* $y^*_{it} = y_{it} > 0$*), а частично дискретно (для* $y^*_{it} \leq 0$, $y_{it} = 0$*)*. Формально, y_{it} – непрерывная функция от y^*_{it}, с *изломом в нуле*. Наблюдения по y_{it} характеризуются *нагромождением нулей.*

Случай 3: Случай с отсечением. В этом Случае мы знаем меньше, чем в Случае 2, но больше, чем в Случае 1, только для тех индивидов в те периоды времени, когда они ответили положительно. У нас нет наблюдений для всех NT комбинаций (i, t). Наблюдения определяются следующим образом:

$$\begin{matrix} (y_{it}, \boldsymbol{x}_{it}) = (y^*_{it}, \boldsymbol{x}_{it}) \text{ и наблюдаема,} & \text{если} \quad y^*_{it} > 0 \quad \Leftrightarrow \quad -\varepsilon_{it} < \dfrac{\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i}{\sigma}, \\ (y_{it}, \boldsymbol{x}_{it}) \text{ ненаблюдаема,} & \text{если} \quad y^*_{it} \leq 0 \quad \Leftrightarrow \quad -\varepsilon_{it} \geq \dfrac{\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i}{\sigma}, \end{matrix} \quad \begin{matrix} i=1, \ldots, N, \\ t=1, \ldots, T. \end{matrix}$$

В этом Случае процесс, определяющий, будут ли у нас наблюдения для конкретных комбинаций (i, t) или нет, случаен, и *этот выбор есть результат индивидуальных решений респондентов, определяемых рассматриваемой моделью*, (1). Выборка представляет собой набор несбалансированных панельных данных, и *вид несбалансированности определяется эндогенно.*

3.3 Вероятностная структура откликов в Случаях 1 и 2

Случай 1: Модель дискретного выбора

Точечные вероятности двух возможных исходов для y_{it}, условно не только на $\boldsymbol{x}_{it}$, но также на индивидуальном эффекте α_i, равны

$$\begin{aligned} \mathbb{P}\{y_{it} = 1 \,|\, \boldsymbol{x}_{it}, \alpha_i\} &= \mathbb{P}\left\{-\varepsilon_{it} \, < \, \frac{\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i}{\sigma} \middle| \boldsymbol{x}_{it}, \alpha_i\right\} \\ &= \Phi\left(\frac{\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i}{\sigma}\right) \equiv \mathcal{L}_{1it}(\alpha_i), \\ \mathbb{P}\{y_{it} = 0 \,|\, \boldsymbol{x}_{it}, \alpha_i\} &= \mathbb{P}\left\{-\varepsilon_{it} \, \geq \, \frac{\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i}{\sigma} \middle| \boldsymbol{x}_{it}, \alpha_i\right\} \\ &= 1-\Phi\left(\frac{\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i}{\sigma}\right) \equiv \mathcal{L}_{0it}(\alpha_i), \end{aligned} \qquad \begin{matrix} i = 1, \ldots, N, \\ t = 1, \ldots, T. \end{matrix} \tag{26}$$

Вновь первый индекс у функций $\mathcal{L}$ обозначает «Режим 1» когда $y_{it} = 1$, и «Режим 0» когда $y_{it} = 0$, соответственно.

Согласно определению имеем

$$\mathcal{L}_{it}(\alpha_i) \equiv \mathcal{L}_{1it}(\alpha_i)^{y_{it}} \mathcal{L}_{0it}(\alpha_i)^{1-y_{it}} \equiv \begin{cases} \mathcal{L}_{1it}(\alpha_i) & \text{для} \quad y_{it} = 1, \quad i=1,\dots,N, \\ \mathcal{L}_{0it}(\alpha_i) & \text{для} \quad y_{it} = 0, \quad t=1,\dots,T. \end{cases} \tag{27}$$

Часть функции правдоподобия, «относящуюся» к индивиду i, условную на α_i, можно переписать из-за независимости $y_{i1}, \dots, y_{iT}$ ввиду (23) как

$$\mathcal{L}_i(\alpha_i) \equiv \prod_{t=1}^{T} \mathcal{L}_{1it}(\alpha_i)^{y_{it}} \mathcal{L}_{0it}(\alpha_i)^{1-y_{it}} \equiv \prod_{t:y_{it}=1} \mathcal{L}_{1it}(\alpha_i) \prod_{t:y_{it}=0} \mathcal{L}_{0it}(\alpha_i). \tag{28}$$

Здесь $t : y_{it} = 1$ и $t : y_{it} = 0$ внизу оператора взятия произведения $\prod$ означают, для индивида i, взятие произведений по всем t таким, что $y_{it} = 1$, и по всем t таким, что $y_{it} = 0$, соответственно.

Соответствующая часть *маржинальной* функции правдоподобия выводится следующим образом. Предположим, что α_i имеет функцию плотности $g(\alpha_i; \gamma)$, где γ – вектор неизвестных параметров, включающий среднее и стандартное отклонение. Тогда маржинальным аналогом (28) является

$$\mathcal{L}_i^* \equiv \int_{-\infty}^{+\infty} \mathcal{L}_i(\alpha_i)\, g(\alpha_i; \gamma) d\,\alpha_i \equiv \int_{-\infty}^{+\infty} \left[\prod_{t:y_{it}=1} \mathcal{L}_{1it}(\alpha_i) \prod_{t:y_{it}=0} \mathcal{L}_{0it}(\alpha_i) \right] g(\alpha_i; \gamma) d\,\alpha_i. \tag{29}$$

Случай 2: Регрессионная модель с цензурированием

Рассмотрим вначале Режим 1, в котором y_{it} непрерывна и имеет ту же КФР, что и у y_{it}^*:

$$\Phi\left(\frac{y_{it} - \boldsymbol{x}_{it}\boldsymbol{\beta} - \alpha_i}{\sigma}\right) = \Phi\left(\frac{y_{it}^* - \boldsymbol{x}_{it}\boldsymbol{\beta} - \alpha_i}{\sigma}\right) \tag{30}$$

и плотность, выводимую дифференцированием (30) по y_{it} для $y_{it} > 0$:

$$\frac{1}{\sigma}\phi\left(\frac{y_{it} - \boldsymbol{x}_{it}\boldsymbol{\beta} - \alpha_i}{\sigma}\right) = \frac{1}{\sigma}\phi\left(\frac{y_{it}^* - \boldsymbol{x}_{it}\boldsymbol{\beta} - \alpha_i}{\sigma}\right) \equiv \mathcal{M}_{1it}(\alpha_i). \tag{31}$$

Рассмотрим теперь Режим 0, в котором y_{it} наблюдается дискретно. Этот Режим совпадает с откликом $y_{it} = 0$ в Случае 1. Тогда у y_{it} нет плотности, а есть вероятностная масса, которую можно получить из КФР y_{it}^* следующим образом (см. вторую часть (26)):

$$\begin{aligned} \mathbb{P}\{y_{it} = 0 \,|\, \boldsymbol{x}_{it}, \alpha_i\} = \mathbb{P}\{y_{it}^* \leq 0 \,|\, \boldsymbol{x}_{it}, \alpha_i\} &= \mathbb{P}\left\{ -\varepsilon_{it} \;\geq\; \frac{\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i}{\sigma} \middle| \boldsymbol{x}_{it}, \alpha_i \right\} \\ &= 1 - \Phi\left(\frac{\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i}{\sigma}\right) \equiv \mathcal{M}_{1it}(\alpha_i) \end{aligned} \tag{32}$$

Часть функции правдоподобия, «относящуюся» к индивиду i, условную на α_i, можно переписать из-за независимости $y_{i1}, \dots, y_{iT}$ ввиду (23) как

$$\mathcal{M}_i(\alpha_i) \equiv \prod_{t:y_{it}>0} \mathcal{M}_{1it}(\alpha_i) \prod_{t:y_{it}=0} \mathcal{M}_{0it}(\alpha_i). \tag{33}$$

Эта функция правдоподобия для конкретного индивида, таким образом, является смесью функций плотности и КФР. Здесь $t : y_{it} > 0$ и $t : y_{it} = 0$ внизу оператора взятия произведения $\prod$ означают, для индивида i, взятие произведений по всем t таким, что $y_{it} > 0$, и по всем t таким, что $y_{it} = 0$, соответственно.

Соответствующая часть *маржинальной* функции правдоподобия выводится следующим образом. Предположим, что α_i имеет функцию плотности $g(\alpha_i;\gamma)$, где γ – вектор неизвестных параметров, включающий среднее и стандартное отклонение. Тогда маржинальный аналог (34) выводится интегрированием (27) по области определения α_i:

$$\mathcal{M}_i^* \equiv \int_{-\infty}^{+\infty} \mathcal{M}_i(\alpha_i) g(\alpha_i;\gamma) d\alpha_i \equiv \int_{-\infty}^{+\infty} \left[\prod_{t:y_{it}>0} \mathcal{M}_{1it}(\alpha_i) \prod_{t:y_{it}=0} \mathcal{M}_{0it}(\alpha_i) \right] g(\alpha_i;\gamma) d\alpha_i. \tag{34}$$

3.4 Задача максимизации правдоподобия в Случаях 1 и 2

Случай 1: Модель дискретного выбора

Поскольку N индивидов наблюдаются независимо, функция правдоподобия является произведением индивидуальных функций правдоподобия в (29), что дает

$$\mathcal{L}^* \equiv \prod_{i=1}^{N} \mathcal{L}_i^* = \prod_{i=1}^{N} \left[\int_{-\infty}^{+\infty} \left(\prod_{t:y_{it}=1} \mathcal{L}_{1it}(\alpha_i) \prod_{t:y_{it}=0} \mathcal{L}_{0it}(\alpha_i) \right) g(\alpha_i;\gamma) d\,\alpha_i \right]. \tag{35}$$

Максимизируя $\mathcal{L}^*$, или, что несколько проще, максимизируя

$$\ln(\mathcal{L}^*) = \sum_{i=1}^{N} \ln \left[\int_{-\infty}^{+\infty} \left(\prod_{t:y_{it}=1} \mathcal{L}_{1it}(\alpha_i) \prod_{t:y_{it}=0} \mathcal{L}_{0it}(\alpha_i) \right) g(\alpha_i;\gamma) d\alpha_i \right] \equiv H\left(\frac{\boldsymbol{\beta}}{\sigma}, \gamma \right)$$

по $(\boldsymbol{\beta}/\sigma, \gamma)$, получаем оценки максимального правдоподобия этих параметров. Эту задачу приходится решать численно. Заметим, что поскольку интегрирование – это что-то типа суммирования, нельзя переместить ln-оператор под интеграл!

Решение данной максимизационной задачи подразумевает интегрирование в двух местах. Во-первых, (логарифмическая) функция правдоподобия содержит интегрирование по отношению к случайному индивидуальному эффекту α_i. Во-вторых, КФР $\Phi(\cdot)$ определяется как некий интеграл. Заметим, что *мы не можем оценить уровень вектора коэффициентов $\boldsymbol{\beta}$, а можем лишь отношение этого вектора к стандартному отклонению возмущения в (23).* Интерпретация: можно измерить (идентифицировать) лишь отклик на переменные в $\boldsymbol{x}_{it}$ в «единицах стандартного отклонения». Это следствие того, что наблюдаемы лишь качественные свойства y_{it}^*.

Случай 2: Регрессионная модель с цензурированием

Поскольку N индивидов наблюдаются независимо, функция правдоподобия является произведением индивидуальных функций правдоподобия в (34), что дает

$$\mathcal{M}^* \equiv \prod_{i=1}^{N} \mathcal{M}_i^* = \prod_{i=1}^{N} \left[\int_{-\infty}^{+\infty} \left(\prod_{t:y_{it}>0} \mathcal{M}_{1it}(\alpha_i) \prod_{t:y_{it}=0} \mathcal{M}_{0it}(\alpha_i) \right) g(\alpha_i;\gamma) d\,\alpha_i \right]. \tag{36}$$

Максимизируя $\mathcal{M}^*$, или, что несколько проще, максимизируя

$$\ln(\mathcal{M}^*) = \sum_{i=1}^{N} \ln \left[\int_{-\infty}^{+\infty} \left(\prod_{t:y_{it}>0} \mathcal{M}_{1it}(\alpha_i) \prod_{t:y_{it}=0} \mathcal{M}_{0it}(\alpha_i) \right) g(\alpha_i;\gamma) d\alpha_i \right] \equiv G(\boldsymbol{\beta}, \sigma, \gamma)$$

по (β, σ, γ), получаем оценки максимального правдоподобия этих параметров.

И вновь решение подразумевает интегрирование в двух местах. Во-первых, (логарифмическая) функция правдоподобия содержит интегрирование по отношению к случайному индивидуальному эффекту α_i. Во-вторых, КФР $\Phi(\cdot)$, реализующегося в Режиме 0, определяется как некий интеграл. *В данном случае мы уже можем оценить абсолютное значение вектора коэффициентов $\boldsymbol{\beta}$ вместе с σ и γ.* Это происходит благодаря тому, что, в отличие от полностью дискретного Случая 1, y^*_{it} на некотором отрезке наблюдаема как количественная (непрерывная) переменная. Этого достаточно для раздельной идентификации $\boldsymbol{\beta}$ и σ.

Estimation of discrete choice and censoring models

Erik Biørn

University of Oslo, Oslo, Norway

This expository note gives an overview of model specifications, likelihood functions and a structure of maximum likelihood problems for discrete choice and censoring models. One part deals with estimation in a single equation case with unidimensional (cross-sectional) observations. Another part extends the framework to a two-equation case. The last part is concerned with an extension to a panel data situation.