

# Микроэконометрика

## Системы уравнений

Потанин Богдан Станиславович

старший преподаватель, кандидат экономических наук

2021-2022

# Системы бинарных уравнений

## Мотивация

- Системы бинарных уравнений предполагают одновременное оценивание параметров нескольких бинарных уравнений.
- Например, можно одновременно оценивать модели, предсказывающие, будет ли индивид курить и употреблять алкоголь. В таком случае эффективность оценок параметров каждого из уравнений может быть повышена за счет учета информации о связи между этими уравнениями: решениями о курении и употреблении алкоголя.
- Как правило предполагается, что уравнения связаны через случайные ошибки. Например, положительная корреляция случайных ошибок уравнений курения и употребления алкоголя может быть связана с наличием общих, не наблюдаемых характеристик, влияющих на принятие соответствующих решений (интеллект, психологические особенности и т.д.).
- С помощью иерархических систем можно в явном виде учитывать эндогенность одной бинарной переменной, влияющие на другую бинарную переменную. Например, можно оценивать влияние высшего образования на вероятность занятости с учетом эндогенности образования за счет оценивания дополнительного уравнения на наличие высшего образования.

# Системы бинарных уравнений

## Общая постановка

- Рассмотрим систему из  $J$  бинарных уравнений:

$$y_{ji}^* = x_{ji}\beta_j + \varepsilon_{ji}$$
$$y_i = \begin{cases} 1, & \text{если } y_{ji}^* > 0 \\ 0, & \text{в противном случае} \end{cases}$$

- Уравнения системы различаются регрессорами  $x_{ji}$ , коэффициентами  $\beta_j$  и случайными ошибками  $\varepsilon_{ji}$ .
- Случайные ошибки  $\varepsilon_i$  независимы между наблюдениями (по  $i$ ), но могут быть зависимы между уравнениями (по  $j$ ).
- Из соображений идентифицируемости предполагается, что  $E(\varepsilon_{ji}) = 0$  и  $Var(\varepsilon_{ji}) = 1$ .
- Для простоты без потери общности далее будем рассматривать двумерную систему бинарных уравнений  $J = 2$ .

# Многомерная пробит модель

## Общая постановка

- Предположим, что совместное распределение случайных ошибок является многомерным нормальным:

$$\begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right),$$

где  $\rho$  является коэффициентом корреляции между случайными ошибками уравнений.

- Вероятности считаются следующим образом:

$$P(y_{1i} = 1, y_{2i} = 1 | x_{1i}, x_{2i}) = P(-\varepsilon_{1i} < x_{1i}\beta_1, -\varepsilon_{2i} < x_{2i}\beta_2 | x_{1i}, x_{2i}) = \Phi(x_{1i}\beta, x_{2i}\beta; \rho)$$

$$P(y_{1i} = 1, y_{2i} = 0 | x_{1i}, x_{2i}) = P(-\varepsilon_{1i} < x_{1i}\beta_1, \varepsilon_{2i} < -x_{2i}\beta_2 | x_{1i}, x_{2i}) = \Phi(x_{1i}\beta, -x_{2i}\beta; -\rho)$$

$$P(y_{1i} = 0, y_{2i} = 1 | x_{1i}, x_{2i}) = P(\varepsilon_{1i} < -x_{1i}\beta_1, -\varepsilon_{2i} < x_{2i}\beta_2 | x_{1i}, x_{2i}) = \Phi(-x_{1i}\beta, x_{2i}\beta; -\rho)$$

$$P(y_{1i} = 0, y_{2i} = 0 | x_{1i}, x_{2i}) = P(\varepsilon_{1i} < -x_{1i}\beta_1, \varepsilon_{2i} < -x_{2i}\beta_2 | x_{1i}, x_{2i}) = \Phi(-x_{1i}\beta, -x_{2i}\beta; \rho)$$

Где  $\Phi(a, b; \rho)$  является функцией распределения двумерного стандартного нормального распределения с коэффициентом корреляции  $\rho$ .

# Многомерная пробит модель

## Оценивание

- Для краткости обозначим  $P_{j,k,i} = P(y_{1i} = j, y_{2i} = k)$ , где  $j, k \in \{0, 1\}$ .
- Как правило, для оценивания используется метод максимального правдоподобия, предполагающий максимизацию следующей функции правдоподобия по  $\beta_1$ ,  $\beta_2$  и  $\rho$ :

$$\begin{aligned} \ln L(\beta_1, \beta_2, \rho; x_1, x_2) &= \prod_i^n P_{y_{1i}, y_{2i}, i} = \\ &= \prod_{i=1}^n \Phi((2y_{1i} - 1)x_{1i}\beta_1, (2y_{2i} - 1)x_{2i}\beta_1; (2y_{1i} - 1)(2y_{2i} - 1)\rho) \end{aligned}$$

- При  $\rho \neq 0$  оценки данной модели более эффективны, чем оценки, полученные за счет оценивания каждого из бинарных уравнений по отдельности, поскольку мы учитываем дополнительную информацию о связи между уравнениями.
- Для того, чтобы понять, есть ли смысл в оценивании системы, достаточно протестировать гипотезу  $H_0 : \rho = 0$ , например, LR-тестом. Если нулевая гипотеза не отвергается, то без потери в эффективности уравнения можно оценивать по отдельности.

# Многомерная пробит модель

## Предельные эффекты

- Для расчета предельных эффектов будем пользоваться следующими формулами:

$$\frac{\partial \Phi(a, b; \rho)}{\partial a} = \phi(a) \Phi\left(\frac{b - \rho a}{\sqrt{1 - \rho^2}}\right), \quad \frac{\partial \Phi(a, b; \rho)}{\partial b} = \phi(b) \Phi\left(\frac{a - \rho b}{\sqrt{1 - \rho^2}}\right)$$

- Предельные эффекты на маргинальные вероятности  $P(y_{1i} = 1|x_{1i})$  и  $P(y_{2i} = 1|x_{2i})$  рассчитываются точно так же, как в обычной пробит модели.
- Рассмотрим предельный эффект регрессора  $v_i$ , входящего и в  $x_{1i}$  и в  $x_{2i}$  с коэффициентами  $\beta_{1v}$  и  $\beta_{2v}$  соответственно, на совместную вероятность:

$$\begin{aligned} \frac{\partial P(y_{1i} = 1, y_{2i} = 1|x_{1i}, x_{2i})}{\partial v} &= \frac{\partial \Phi(x_{1i}\beta_1, x_{2i}\beta_2; \rho)}{\partial v} = \\ &= \beta_{1v}\phi(x_1\beta_1)\Phi\left(\frac{x_2\beta_2 - \rho x_1\beta_1}{\sqrt{1 - \rho^2}}\right) + \beta_{2v}\phi(x_2\beta_2)\Phi\left(\frac{x_1\beta_1 - \rho x_2\beta_2}{\sqrt{1 - \rho^2}}\right) \end{aligned}$$

Если знаки  $\beta_{1v}$  и  $\beta_{2v}$  совпадают, то таким же будем и знак предельного эффекта.

# Многомерная пробит модель

## Иерархическая модель

- Предположим, что одна зависимая бинарная переменная в системе влияет на другую:

$$y_{1i}^* = x_{1i}\beta_1 + \alpha y_{2i} + \varepsilon_{1i}$$

$$y_{2i}^* = x_{2i}\beta_2 + \varepsilon_{2i},$$

где без потери общности так же можно было бы предположить, что  $y_{2i}$  входит в  $x_{1i}$ .

- Обычно  $y_{2i}$  именуют эндогенным бинарным регрессором. Иерархические модели позволяют в явном виде учесть эндогенность  $y_{2i}$  по отношению к уравнению  $y_{1i}$ .
- В подобных системах представляет интерес оценивание условных вероятностей:

$$P(y_{1i} = k | x_{1i}, x_{2i}, y_{2i} = j), \text{ где } k, j \in \{0, 1\}$$

- Данные вероятности нетрудно рассчитать по формуле Байеса, например:

$$P(y_{1i} = 1 | x_{1i}, x_{2i}, y_{2i} = 1) = \frac{P(y_{1i} = 1, y_{2i} = 1 | x_{1i}, x_{2i})}{P(y_{2i} = 1 | x_{1i}, x_{2i})} = \frac{\Phi(x_{1i}\beta_1 + \alpha, x_{2i}\beta_2; \rho)}{\Phi(x_{2i}\beta_2)}$$

# Многомерная пробит модель

## Иерархическая модель: предельный эффекты и эффект воздействия

- Предельный эффект условной вероятности по  $y_{2i}$  рассчитывается как разница условных вероятностей:

$$\begin{aligned} & P(y_{1i} = 1 | x_{1i}, x_{2i}, y_{2i} = 1) - P(y_{1i} = 1 | x_{1i}, x_{2i}, y_{2i} = 0) = \\ &= \frac{\Phi(x_{1i}\beta_1 + \alpha, x_{2i}\beta_2; \rho)}{\Phi(x_{2i}\beta_2)} - \frac{\Phi(x_{1i}\beta_1, -x_{2i}\beta_2; -\rho)}{1 - \Phi(x_{2i}\beta_2)} \end{aligned}$$

- Недостаток предельного эффекта с точки зрения содержательной интерпретации заключается в том, что он отражает влияние  $y_{2i}$  на  $y_{1i}$  не только непосредственно, то есть через  $\alpha$ , но и косвенно – через условное распределение случайной ошибки  $\varepsilon_{2i} | \varepsilon_{2i} > -x_{2i}\beta_2$ .
- Эффект воздействия  $y_{2i}$  на подвергнутых воздействию (ТЕТ) измеряет то, как  $y_{2i}$  подействовал непосредственно на тех, у кого  $y_{2i} = 1$ :

$$\begin{aligned} & P(x_{1i}\beta + \alpha + \varepsilon_{1i} > 0 | x_{1i}, x_{2i}, x_{2i}\beta_2 + \varepsilon_{2i} > 0) - P(x_{1i}\beta + \varepsilon_{1i} > 0 | x_{1i}, x_{2i}, x_{2i}\beta_2 + \varepsilon_{2i} > 0) = \\ &= \frac{\Phi(x_{1i}\beta_1 + \alpha, x_{2i}\beta_2; \rho)}{\Phi(x_{2i}\beta_2)} - \frac{\Phi(x_{1i}\beta_1, x_{2i}\beta_2; \rho)}{\Phi(x_{2i}\beta_2)} = \frac{\Phi(x_{1i}\beta_1 + \alpha, x_{2i}\beta_2; \rho) - \Phi(x_{1i}\beta_1, x_{2i}\beta_2; \rho)}{\Phi(x_{2i}\beta_2)} \end{aligned}$$



# Модели мультиномиального выбора

## Мотивация

- Иногда зависимая переменная может быть номинальной, то есть принимать одно из конечного числа взаимоисключающих значений.
- Часто рассматриваются номинальные переменные, отражающие выбор индивида, например, в отношении:
  - типа получаемого образования: техническое, гуманитарное или никакого
  - марки приобретаемого телефона
  - типа используемого браузера
  - бронируемого отеля из числа доступных в городе альтернатив
  - музыкального сервиса
- Модели мультиномиального (множественного) выбора позволяет оценить вероятности выбора индивидом той или иной альтернативы, а также то, как различные факторы влияют на эти вероятности.

# Модели мультиномиального выбора

## Общая постановка

- Индивид выбирает одну из  $J$  альтернатив:

$$y_{ji}^* = x_i \beta_j + \varepsilon_{ji},$$

где  $y_{ji}^*$  отражает склонность к выбору  $j$ -й альтернативы  $i$ -м индивидом.

- Индивид выбирает альтернативу, к выбору которой он проявил наибольшую склонность:

$$y_i = \begin{cases} 1 & \text{если } y_{1i}^* > y_{2i}^*, y_{1i}^* > y_{3i}^*, \dots, y_{1i}^* > y_{Ji}^* \\ 2 & \text{если } y_{2i}^* > y_{1i}^*, y_{2i}^* > y_{3i}^*, \dots, y_{2i}^* > y_{Ji}^* \\ \vdots & \\ J & \text{если } y_{Ji}^* > y_{1i}^*, y_{Ji}^* > y_{2i}^*, \dots, y_{Ji}^* > y_{(J-1)i}^* \end{cases}$$

- На практике мы наблюдаем только  $y_i$ , то есть для каждого индивида наблюдается лишь наиболее предпочтительный для него выбор.

# Мультиномиальная логит модель

## Формулировка

- В мультиномиальной логит модели предполагается, что случайные ошибки  $\varepsilon_{ji}$  независимы и одинаково распределены в соответствии с распределением Гумбеля, имеющим следующие функцию распределения и функцию плотности:

$$F_{\varepsilon_{ji}}(t) = e^{-e^{-t}}, \quad f_{\varepsilon_{ji}}(x) = e^{-t-e^{-t}}, \quad E(\varepsilon_{ji}) \approx 0.577$$

- С помощью формулы свертки можно вывести выражение для условной вероятности выбора  $k$ -й альтернативы:

$$\begin{aligned} P(y_i = k | x_i) &= P(y_{ik}^* > y_{i1}^*, \dots, y_{ik}^* > y_{iK}^*) = \\ &= P(x_i \beta_k + \varepsilon_{ki} > x_i \beta_1 + \varepsilon_{1i}, \dots, x_i \beta_k + \varepsilon_{ki} > x_i \beta_J + \varepsilon_{Ji} | x_i) = \\ &= P(\varepsilon_{1i} - \varepsilon_k < x_k \beta_k - x_i \beta_1, \dots, \varepsilon_{Ji} - \varepsilon_k < x_i \beta_k - x_i \beta_J | x_i) = e^{x_i \beta_k} / \sum_{j=1}^J e^{x_i \beta_j} \end{aligned}$$

- Сформировав с помощью этих вероятностей функцию правдоподобия можно получить ММП оценки параметров  $\beta_1, \dots, \beta_J$ .

# Мультиномиальная логит модель

## Идентификация

- Вычитание из каждого из векторов регрессионных коэффициентов  $\beta_j$  любого вектора  $v$  соответствующих размеров не приведет к изменению вероятностей:

$$P(y_i = k | x_i) = \frac{e^{x_i(\beta_k - v)}}{\sum_{j=1}^J e^{x_i(\beta_j - v)}} = \frac{e^{-x_i v} e^{x_i \beta_k}}{e^{-x_i v} \sum_{j=1}^J e^{x_i \beta_j}} = \frac{e^{x_i \beta_k}}{\sum_{j=1}^J e^{x_i \beta_j}}$$

- Следовательно, максимизируя функцию правдоподобия по всем  $\beta_j$  мы получим бесконечное число решений, что порождает проблему неидентифицируемости. Поэтому без потери общности полагают  $\beta_J = (0, \dots, 0)$ , откуда:

$$P(y_i = k | x_i) = \frac{e^{x_i \beta_k}}{1 + \sum_{j=1}^{J-1} e^{x_i \beta_j}}$$

# Мультиномиальная логит модель

## Предельные эффекты

- Рассмотрим предельный эффект вероятности выбора  $k$ -й альтернативы по  $t$ -му регрессору:

$$\frac{\partial P(y_i = k|x_i)}{\partial x_{ti}} = P(y_i = k|x_i) \times \left[ \beta_{kt} - \sum_{j=1}^J P(y_i = j|x_i) \beta_{jt} \right]$$

во сколько раз возрастает вероятность

- Знак предельного эффекта не определяется знаком  $\beta_{kt}$ , за исключением двух случаев:
  - Если  $\beta_{kt} = \max(\beta_{1t}, \dots, \beta_{Jt})$ , то предельный эффект положительный.
  - Если  $\beta_{kt} = \min(\beta_{1t}, \dots, \beta_{Jt})$ , то предельный эффект отрицательный.

- Для осуществления предсказания необходимо рассчитать вероятность выбора индивидом каждой из альтернатив.
- Предполагается, что индивид выбирает ту альтернативу, вероятность выбора которой является наибольшей.
- Это эквивалентно тому, чтобы присвоить индивиду ту альтернативу, склонность к выбору которой является максимальной.
- Например, Лаврентий выбирает между походом в кино, в театр и в оперу. Если в соответствии с нашими оценками он пойдет в кино с вероятностью 0.5, в театр с вероятностью 0.3 и в оперу с вероятностью 0.2, то мы предполагаем, что Лаврентий выберет кино.

# Мультиномиальная логит модель

Условная мультиномиальная логит модель (conditional multinomial logit)

- Предположим, что склонность формируется следующим образом:

$$y_{ji}^* = x_i \beta_j + x_i^j \beta^* + \varepsilon_{ji}$$

где:

- $x_i$  – регрессоры, общие для всех альтернатив
- $x_i^j$  – регрессоры, различающиеся между альтернативами
- $\beta_j$  – регрессионные коэффициенты, учитывающие, что не различающиеся регрессоры могут по разному влиять на вероятность выбора той или иной альтернативы
- $\beta^*$  – регрессионные коэффициенты, учитывающие, что разные регрессоры могут одинаковым образом влиять на вероятность выбора той или иной альтернативы
- Например, вы можете рассматривать модель, предсказывающую выбор индивидом средства поездки на работу: машина, такси или общественный транспорт. Тогда  $x_i$  будут отражать характеристики индивида (пол, возраст, доход), а  $x_i^j$  – характеристики альтернатив (стоимость и комфорт поездки на каждом из видов транспорта).
- Например, доход может по разному влиять на склонность к поездке на том или ином виде транспорта, а цена поездки – одинаково.

# Мультиномиальная логит модель

## Предельные эффекты в условной мультиномиальной логит модели

- Предельные эффекты по регрессорам  $x_i$  аналогичны рассмотренным ранее.
- Для краткости обозначим  $P_{ki} = P(y_i = k | x_i^1, \dots, x_i^J)$ .
- Предельный эффект  $t$ -го регрессора из  $x_i^k$  на условную вероятность выбора  $k$ -й альтернативы:

$$\frac{\partial P_{ki}}{\partial x_{ti}^k} = P_{ki} (1 - P_{ki}) \beta_t^*$$

Данный предельный эффект может отражать как, например, изменение стоимости поездки на такси влияет на вероятность поездки на такси. Знак данного предельного эффекта совпадает со знаком коэффициента  $\beta_t^*$  при стоимости поездки на такси.

- Предельный эффект  $t$ -го регрессора из  $x_i^j$  на условную вероятность выбора  $k$ -й альтернативы:

$$\frac{\partial P_{ki}}{\partial x_{ti}^j} = -P_{ki} P_{ji} \beta_t^*$$

Данный предельный эффект может отражать как, например, влияет изменение стоимости поездки на такси на вероятность поездки на машине. Знак данного предельного эффекта противоположен знаку коэффициента  $\beta_t^*$  при стоимости поездки на такси. Например, при  $\beta_t^* < 0$  мы получаем, что повышение стоимости поездки на такси приводит, при прочих равных, к увеличению вероятности поездки на машине.



# Мультиномиальная логит модель

## Независимость от посторонних альтернатив (IIA)

- Из-за того, что случайные ошибки независимы между альтернативами, отношения вероятностей выбора для двух альтернатив не зависят от вероятностей выбора других альтернатив:

$$\frac{P(y_i = k|x_i)}{P(y_i = j|x_i)} = \frac{e^{x_i\beta_k}}{e^{x_i\beta_j}}, \quad \forall k, j \in \{1, \dots, J\}$$

- Это свойство мультиномиальной логит модели именуют независимостью от посторонних альтернатив, поскольку выбор между двумя альтернативами не зависит от наличия других альтернатив (IIA - independence of irrelevant alternatives), что считается нереалистичным.
- Для проверки соблюдения данного допущения применяется тест Хаусмана-Макфаддена. Нулевая гипотеза теста предполагает выполнение (IIA). Если нулевая гипотеза отклоняется, то оценки мультиномиальной логит модели могут оказаться несостоятельными.

# Мультиномиальная пробит модель

## Основные особенности

- Допущение о независимости от посторонних альтернатив можно ослабить за счет снятия допущения о независимости случайных ошибок.
- Мультиномиальная пробит модель предполагает, что случайные ошибки имеют многомерное нормальное распределение (для простоты и без потери общности положим  $J = 3$ ):

$$\varepsilon_i = \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_1\sigma_3 \\ \rho_{13}\sigma_1\sigma_2 & \rho_{23}\sigma_1\sigma_3 & \sigma_3^2 \end{bmatrix} \right)$$

- Из соображений идентифицируемости полагают  $y_{3i}^* = 0$  и  $\sigma_1 = 1$ , что позволяет рассматривать совместное распределение случайных ошибок лишь первых двух альтернатив:

$$\begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{12}\sigma_2 \\ \rho_{12}\sigma_2 & \sigma_2^2 \end{bmatrix} \right)$$

# Мультиномиальная пробит модель

## Недостатки

- Максимизация функция правдоподобия является достаточно сложной технической задачей, особенно при большом числе альтернатив  $J \geq 5$ . Получение оценок параметров данной модели может занять несколько часов даже на достаточно продвинутых компьютерах.
- Формально параметры модели идентифицируются, однако (грубо говоря), параметры  $\beta_j$  коллинеарны параметрам ковариационной матрицы случайных ошибок, что осложняет получение точных оценок  $\beta_j$ . Однако, это не мешает получать точные предсказания вероятностей.