Микроэконометрика

Модели с эндогенным переключением

Потанин Богдан Станиславович

доцент, научный сотрудник, кандидат экономических наук

2023-2024

- Иногда исследователь наблюдает лишь одно из двух возможных состояний зависимой переменной.

- Иногда исследователь наблюдает лишь одно из двух возможных состояний зависимой переменной.
- Например, исследователь наблюдает либо зарплату индивида в состоянии, когда у него есть высшее образование, либо когда у него нет высшего образования.

Эндогенное переключение
Мотивация

- Иногда исследователь наблюдает лишь одно из двух возможных состояний зависимой переменной.
- Например, исследователь наблюдает либо зарплату индивида в состоянии, когда у него есть высшее образование, либо когда у него нет высшего образования.
- Механизм формирования зарплаты (отдача от стажа и т.д.) может варьироваться в зависимости от наличия у индивида высшего образования. При этом может различаться отдача как от наблюдаемых, так и от ненаблюдаемых характеристик.

Эндогенное переключение

Мотивация

- Иногда исследователь наблюдает лишь одно из двух возможных состояний зависимой переменной.
- Например, исследователь наблюдает либо зарплату индивида в состоянии, когда у него есть высшее образование, либо когда у него нет высшего образования.
- Механизм формирования зарплаты (отдача от стажа и т.д.) может варьироваться в зависимости от наличия у индивида высшего образования. При этом может различаться отдача как от наблюдаемых, так и от ненаблюдаемых характеристик.
- Идейно, эндогенное переключение проистекает из неслучайного отбора, поскольку имеется неслучайный отбор в те состояния (режимы), в которых наблюдается зависимая переменная.

Эндогенное переключение

Мотивация

- Иногда исследователь наблюдает лишь одно из двух возможных состояний зависимой переменной.
- Например, исследователь наблюдает либо зарплату индивида в состоянии, когда у него есть высшее образование, либо когда у него нет высшего образования.
- Механизм формирования зарплаты (отдача от стажа и т.д.) может варьироваться в зависимости от наличия у индивида высшего образования. При этом может различаться отдача как от наблюдаемых, так и от ненаблюдаемых характеристик.
- Идейно, эндогенное переключение проистекает из неслучайного отбора, поскольку имеется неслучайный отбор в те состояния (режимы), в которых наблюдается зависимая переменная.
- Например, неслучайный отбор в число тех, кто получил высшее образование, определяет переключение между двумя типами зарплаты: при условии наличия высшего образования и без него.

Эндогенное переключение

Формулировка

- Имеются два целевых уравнения и уравнение, задающее переключение между ними:

$$\text{Целевое уравнение 1: } y_{1i}^{*} = x_i\beta_1 + \varepsilon_{1i}$$

$$\text{Целевое уравнение 0: } y_{0i}^{*} = x_i\beta_0 + \varepsilon_{0i}$$

$$\text{Уравнение переключения: } z_i^{*} = w_i\gamma + u_i$$

Эндогенное переключение

Формулировка

- Имеются два целевых уравнения и уравнение, задающее переключение между ними:

$$\text{Целевое уравнение 1: } y_{1i}^{*} = x_i\beta_1 + \varepsilon_{1i}$$

$$\text{Целевое уравнение 0: } y_{0i}^{*} = x_i\beta_0 + \varepsilon_{0i}$$

$$\text{Уравнение переключения: } z_i^{*} = w_i\gamma + u_i$$

- Наблюдаемое состояние целевого уравнения определяется эндогенным переключением, то есть зависит от z_i^{*}:

$$z_i = \begin{cases} 1, \text{ если } z_i^{*} \geq 0 \\ 0, \text{ в противном случае} \end{cases} \qquad y_i = \begin{cases} y_{1i}^{*}, \text{ если } z_i = 1 \\ y_{0i}^{*}, \text{ если } z_i = 0 \end{cases}$$

- Имеются два целевых уравнения и уравнение, задающее переключение между ними:

$$\text{Целевое уравнение 1: } y_{1i}^{*} = x_i\beta_1 + \varepsilon_{1i}$$
$$\text{Целевое уравнение 0: } y_{0i}^{*} = x_i\beta_0 + \varepsilon_{0i}$$
$$\text{Уравнение переключения: } z_i^{*} = w_i\gamma + u_i$$

- Наблюдаемое состояние целевого уравнения определяется эндогенным переключением, то есть зависит от z_i^{*}:

$$z_i = \begin{cases} 1, \text{ если } z_i^{*} \geq 0 \\ 0, \text{ в противном случае} \end{cases} \qquad y_i = \begin{cases} y_{1i}^{*}, \text{ если } z_i = 1 \\ y_{0i}^{*}, \text{ если } z_i = 0 \end{cases}$$

- Например, y_{1i}^{*} и y_{0i}^{*} могут отражать зарплату индивида при условии наличия или отсутствия у него высшего образования соответственно, а $z_i = 1$ в случаях, когда у индивида есть высшее образование. Различие в коэффициентах β_1 и β_0 отражает различную отдачу (влияние) от характеристик в зависимости от состояния. Например, отдача от стажа может быть выше для людей с высшим образованием.

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

Эндогенное переключение

Совместное распределение случайных ошибок

- Предположим, что совместное распределение случайных ошибок является многомерным нормальным.
- Совместное распределение случайных ошибок целевых уравнений ε_{1i} и ε_{0i} не идентифицируемо, поскольку мы наблюдаем значение целевой переменной лишь в одном из состояний.

Эндогенное переключение

Совместное распределение случайных ошибок

- Предположим, что совместное распределение случайных ошибок является многомерным нормальным.
- Совместное распределение случайных ошибок целевых уравнений ε_{1i} и ε_{0i} не идентифицируемо, поскольку мы наблюдаем значение целевой переменной лишь в одном из состояний.
- Однако, для оценивания параметров модели достаточно ввести допущение о совместном распределении этих случайных ошибок со случайной ошибкой уравнения переключения u_i:

$$(u_i, \varepsilon_{ji}) \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & \rho_j\sigma_j\\ \rho_j\sigma_j & \sigma_j^2\end{bmatrix}\right), \text{ где } j \in \{0, 1\}$$

Эндогенное переключение

Совместное распределение случайных ошибок

- Предположим, что совместное распределение случайных ошибок является многомерным нормальным.
- Совместное распределение случайных ошибок целевых уравнений ε_{1i} и ε_{0i} не идентифицируемо, поскольку мы наблюдаем значение целевой переменной лишь в одном из состояний.
- Однако, для оценивания параметров модели достаточно ввести допущение о совместном распределении этих случайных ошибок со случайной ошибкой уравнения переключения u_i:

$$(u_i, \varepsilon_{ji}) \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_j\sigma_j \\ \rho_j\sigma_j & \sigma_j^2 \end{bmatrix}\right), \text{ где } j \in \{0, 1\}$$

- Дисперсия случайной ошибки σ_j^2 и ее корреляция со случайной ошибкой уравнения переключения ρ_j различаются в зависимости от состояния (режима) целевого уравнения, что, в частности, может быть обусловлено различием в отдаче (влиянии) от ненаблюдаемых характеристик (на целевой показатель).

Эндогенное переключение

Проблема оценивания

- Для краткости обозначим $\tilde{z}_i = 2z_i - 1$.

- Для краткости обозначим $\tilde{z}_i = 2z_i - 1$.
- Поскольку y_i отражает лишь одно из наблюдаемых состояний, то:

$$E\left(y_i|w_i, x_i\right) = E\left(y^*_{z_i i}|z_i, w_i, x_i\right) = x_i\beta_{z_i} + E\left(\varepsilon_{z_i i}| - \tilde{z}_i u_i \leq \tilde{z}_i w_i\gamma, w_i, x_i\right),$$

- Для краткости обозначим $\tilde{z}_i = 2z_i - 1$.
- Поскольку y_i отражает лишь одно из наблюдаемых состояний, то:

$$E\left(y_i|w_i,x_i\right) = E\left(y^*_{z_i i}|z_i,w_i,x_i\right) = x_i\beta_{z_i} + E\left(\varepsilon_{z_i i}|-\tilde{z}_i u_i \leq \tilde{z}_i w_i\gamma, w_i, x_i\right),$$

где по свойствам усеченного двумерного нормального распределения:

$$E\left(\varepsilon_{z_i i}|-\tilde{z}_i u_i \leq \tilde{z}_i w_i\gamma, w_i, x_i\right) = \tilde{z}_i\rho_{z_i}\sigma_{z_i}\frac{\phi\left(w_i\gamma\right)}{\Phi\left(\tilde{z}_i w_i\gamma\right)} = \tilde{z}_i\rho_{z_i}\sigma_{z_i}\lambda_i,$$

Эндогенное переключение
Проблема оценивания

- Для краткости обозначим $\tilde{z}_i = 2z_i - 1$.
- Поскольку y_i отражает лишь одно из наблюдаемых состояний, то:

$$E\left(y_i | w_i, x_i\right) = E\left(y^*_{z_i i} | z_i, w_i, x_i\right) = x_i\beta_{z_i} + E\left(\varepsilon_{z_i i} | -\tilde{z}_i u_i \leq \tilde{z}_i w_i\gamma, w_i, x_i\right),$$

где по свойствам усеченного двумерного нормального распределения:

$$E\left(\varepsilon_{z_i i} | -\tilde{z}_i u_i \leq \tilde{z}_i w_i\gamma, w_i, x_i\right) = \tilde{z}_i\rho_{z_i}\sigma_{z_i}\frac{\phi\left(w_i\gamma\right)}{\Phi\left(\tilde{z}_i w_i\gamma\right)} = \tilde{z}_i\rho_{z_i}\sigma_{z_i}\lambda_i,$$

- В результате получаем регрессионные уравнения (где $j \in \{0, 1\}$ и $z_i = j$):

$$y^*_{ji} = x_i\beta_j + \tilde{z}_i\rho_j\sigma_j\lambda_i + v_{ji},$$

Эндогенное переключение

Проблема оценивания

- Для краткости обозначим $\tilde{z}_i = 2z_i - 1$.
- Поскольку y_i отражает лишь одно из наблюдаемых состояний, то:

$$E\left(y_i|w_i,x_i\right) = E\left(y^*_{z_i i}|z_i,w_i,x_i\right) = x_i\beta_{z_i} + E\left(\varepsilon_{z_i i}| - \tilde{z}_i u_i \leq \tilde{z}_i w_i\gamma, w_i, x_i\right),$$

где по свойствам усеченного двумерного нормального распределения:

$$E\left(\varepsilon_{z_i i}| - \tilde{z}_i u_i \leq \tilde{z}_i w_i\gamma, w_i, x_i\right) = \tilde{z}_i\rho_{z_i}\sigma_{z_i}\frac{\phi\left(w_i\gamma\right)}{\Phi\left(\tilde{z}_i w_i\gamma\right)} = \tilde{z}_i\rho_{z_i}\sigma_{z_i}\lambda_i,$$

- В результате получаем регрессионные уравнения (где $j \in \{0,1\}$ и $z_i = j$):

$$y^*_{ji} = x_i\beta_j + \tilde{z}_i\rho_j\sigma_j\lambda_i + v_{ji}, \qquad v_{ji} = \varepsilon_{ji} - (2j-1)\rho_j\sigma_j\lambda_i \implies E(v_{ji}|x_i,w_i) = 0$$

- Для краткости обозначим $\tilde{z}_i = 2z_i - 1$.
- Поскольку y_i отражает лишь одно из наблюдаемых состояний, то:

$$E\left(y_i|w_i,x_i\right) = E\left(y^*_{z_i i}|z_i,w_i,x_i\right) = x_i\beta_{z_i} + E\left(\varepsilon_{z_i i}| - \tilde{z}_i u_i \leq \tilde{z}_i w_i\gamma, w_i, x_i\right),$$

где по свойствам усеченного двумерного нормального распределения:

$$E\left(\varepsilon_{z_i i}| - \tilde{z}_i u_i \leq \tilde{z}_i w_i\gamma, w_i, x_i\right) = \tilde{z}_i\rho_{z_i}\sigma_{z_i}\frac{\phi\left(w_i\gamma\right)}{\Phi\left(\tilde{z}_i w_i\gamma\right)} = \tilde{z}_i\rho_{z_i}\sigma_{z_i}\lambda_i,$$

- В результате получаем регрессионные уравнения (где $j \in \{0,1\}$ и $z_i = j$):

$$y^*_{ji} = x_i\beta_j + \tilde{z}_i\rho_j\sigma_j\lambda_i + v_{ji}, \qquad v_{ji} = \varepsilon_{ji} - (2j-1)\rho_j\sigma_j\lambda_i \implies E(v_{ji}|x_i,w_i) = 0$$

- Без учета λ_i при $\rho \neq 0$ и наличии корреляции между λ_i и x_i МНК оценки коэффициентов β_j окажутся несостоятельными вследствие проблемы пропущенной переменной.

- Процедура оценивания аналогична двухшаговому методу Хекмана и производится по отдельности для каждого из целевых уравнений.

Оценивание

Двухшаговая процедура

- Процедура оценивания аналогична двухшаговому методу Хекмана и производится по отдельности для каждого из целевых уравнений.
- Двухшаговая процедура оценивания:
  - **Первый шаг**: при помощи пробит модели оцениваются параметры γ. В силу инвариантности ММП оценок состоятельная оценка λ_i рассчитывается как $\hat{\lambda}_i = \lambda_i\left(\tilde{z}_i w_i \hat{\gamma}\right)$. Этот шаг совпадает для обоих уравнений.

- Процедура оценивания аналогична двухшаговому методу Хекмана и производится по отдельности для каждого из целевых уравнений.
- Двухшаговая процедура оценивания:
  - **Первый шаг**: при помощи пробит модели оцениваются параметры γ. В силу инвариантности ММП оценок состоятельная оценка λ_i рассчитывается как $\hat{\lambda}_i = \lambda_i\left(\tilde{z}_i w_i \hat{\gamma}\right)$. Этот шаг совпадает для обоих уравнений.
  - **Второй шаг**: В регрессионное уравнение для y_{ji}^* подставляется $\hat{\lambda}_i$ в качестве дополнительного регрессора с коэффициентом $\rho_j\sigma_j$. Затем β_j и $\rho_j\sigma_j$ оцениваются при помощи МНК по выборке из наблюдений, для которых $z_i = j$.

Оценивание

Двухшаговая процедура

- Процедура оценивания аналогична двухшаговому методу Хекмана и производится по отдельности для каждого из целевых уравнений.
- Двухшаговая процедура оценивания:
  - **Первый шаг**: при помощи пробит модели оцениваются параметры γ. В силу инвариантности ММП оценок состоятельная оценка λ_i рассчитывается как $\hat{\lambda}_i = \lambda_i\left(\tilde{z}_i w_i \hat{\gamma}\right)$. Этот шаг совпадает для обоих уравнений.
  - **Второй шаг**: В регрессионное уравнение для y_{ji}^* подставляется $\hat{\lambda}_i$ в качестве дополнительного регрессора с коэффициентом $\rho_j\sigma_j$. Затем β_j и $\rho_j\sigma_j$ оцениваются при помощи МНК по выборке из наблюдений, для которых $z_i = j$.
- Оценивание параметров $\hat{\sigma}_j^2$ и $\hat{\rho}_j$, а также асимптотической ковариационной матрицы оценок коэффициентов (с учетом гетероскедастичности и оценок первого шага) осуществляется по аналогии с двухшаговой процедурой Хекмана.

- Процедура оценивания аналогична двухшаговому методу Хекмана и производится по отдельности для каждого из целевых уравнений.
- Двухшаговая процедура оценивания:
  - **Первый шаг**: при помощи пробит модели оцениваются параметры γ. В силу инвариантности ММП оценок состоятельная оценка λ_i рассчитывается как $\hat{\lambda}_i = \lambda_i\left(\tilde{z}_i w_i \hat{\gamma}\right)$. Этот шаг совпадает для обоих уравнений.
  - **Второй шаг**: В регрессионное уравнение для y_{ji}^* подставляется $\hat{\lambda}_i$ в качестве дополнительного регрессора с коэффициентом $\rho_j\sigma_j$. Затем β_j и $\rho_j\sigma_j$ оцениваются при помощи МНК по выборке из наблюдений, для которых $z_i = j$.
- Оценивание параметров $\hat{\sigma}_j^2$ и $\hat{\rho}_j$, а также асимптотической ковариационной матрицы оценок коэффициентов (с учетом гетероскедастичности и оценок первого шага) осуществляется по аналогии с двухшаговой процедурой Хекмана.
- Эффективность оценок данного метода в существенной степени зависит от наличия ограничений исключения.

- Функция правдоподобия:

$$L\left(\beta_1,\beta_0,\rho_1,\rho_0,\sigma_1,\sigma_0;y,z|X,W\right)=$$
$$=\prod_{i:z_i=1} f_{y_i|x_i,w_i}(y_i)P(z_i=1|y_i,x_i,w_i)\prod_{i:z_i=0} f_{y_i|x_i,w_i}(y_i)P(z_i=0|y_i,x_i,w_i)=$$

- Функция правдоподобия:

$$L\left(\beta_1,\beta_0,\rho_1,\rho_0,\sigma_1,\sigma_0;y,z|X,W\right)=$$

$$=\prod_{i:z_i=1} f_{y_i|x_i,w_i}(y_i)P(z_i=1|y_i,x_i,w_i)\prod_{i:z_i=0} f_{y_i|x_i,w_i}(y_i)P(z_i=0|y_i,x_i,w_i)=$$

$$=\prod_{i=1}^{n}\frac{1}{\sigma_{z_i}}\phi\left(\frac{y_i-x_i\beta_{z_i}}{\sigma_{z_i}}\right)\Phi\left(\frac{\rho_{z_i}\left(y_i-x_i\beta_{z_i}\right)/\sigma_{z_i}+w_i\gamma}{\tilde{z}_i\sqrt{1-\rho_{z_i}^2}}\right)$$

- Функция правдоподобия:

$$L\left(\beta_1,\beta_0,\rho_1,\rho_0,\sigma_1,\sigma_0;y,z|X,W\right)=$$

$$=\prod_{i:z_i=1}f_{y_i|x_i,w_i}(y_i)P(z_i=1|y_i,x_i,w_i)\prod_{i:z_i=0}f_{y_i|x_i,w_i}(y_i)P(z_i=0|y_i,x_i,w_i)=$$

$$=\prod_{i=1}^{n}\frac{1}{\sigma_{z_i}}\phi\left(\frac{y_i-x_i\beta_{z_i}}{\sigma_{z_i}}\right)\Phi\left(\frac{\rho_{z_i}\left(y_i-x_i\beta_{z_i}\right)/\sigma_{z_i}+w_i\gamma}{\tilde{z}_i\sqrt{1-\rho_{z_i}^2}}\right)$$

- Преимущества и недостатки двухшаговой и ММП процедур оценивания аналогичны тем, что имеют место для модели с неслучайным отбором.

Оценивание

Метод максимального правдоподобия

- Функция правдоподобия:

$$L\left(\beta_1,\beta_0,\rho_1,\rho_0,\sigma_1,\sigma_0;y,z|X,W\right)=$$

$$=\prod_{i:z_i=1} f_{y_i|x_i,w_i}(y_i)P(z_i=1|y_i,x_i,w_i)\prod_{i:z_i=0} f_{y_i|x_i,w_i}(y_i)P(z_i=0|y_i,x_i,w_i)=$$

$$=\prod_{i=1}^{n}\frac{1}{\sigma_{z_i}}\phi\left(\frac{y_i-x_i\beta_{z_i}}{\sigma_{z_i}}\right)\Phi\left(\frac{\rho_{z_i}\left(y_i-x_i\beta_{z_i}\right)/\sigma_{z_i}+w_i\gamma}{\tilde{z}_i\sqrt{1-\rho_{z_i}^2}}\right)$$

- Преимущества и недостатки двухшаговой и ММП процедур оценивания аналогичны тем, что имеют место для модели с неслучайным отбором.
- Для каждого из уравнений процедура тестирования наличия неслучайного отбора сводится к проверке гипотез $H_0:\rho_j\sigma_j=0$ и $H_0:\rho_j=0$ для двухшаговой процедуры и ММП соответственно. Уравнение, в котором данная нулевая гипотеза не отвергается, можно оценить при помощи обычного МНК.

- Предельный эффект переменной x_{ik} на обычное математическое ожидание имеет такой же вид, как в случае с обычной линейной регрессией:

$$\frac{\partial E\left(y_{ji}^{*}|x_i\right)}{\partial x_{ik}} = \beta_{jk},\text{ где } j \in \{0, 1\}.$$

- Предельный эффект переменной x_{ik} на обычное математическое ожидание имеет такой же вид, как в случае с обычной линейной регрессией:

$$\frac{\partial E\left(y_{ji}^{*}|x_i\right)}{\partial x_{ik}}=\beta_{jk},\text{ где } j\in\{0,1\}.$$

- Предельный эффект на условное математическое ожидание рассчитывается как:

$$\frac{\partial E\left(y_{ji}^{*}|z_i,x_i,w_i\right)}{\partial x_{ik}}=\beta_{jk}-\tilde{z}_i\gamma_{*}\rho_j\sigma_j\delta\left(\tilde{z}_iw_i\gamma\right),$$

$$\delta(a)=\lambda\left(a\right)\left(\lambda\left(a\right)+a\right),$$

где γ_{*} является коэффициентом при x_{ki} в уравнении отбора, если x_{ki} входит в w_i. В противном случае $\gamma_{*}=0$.

- Эффект воздействия z_i на целевую переменную i-го индивида определяется как:

$$\text{TE}_i = y_{1i} - y_{0i}$$

- Эффект воздействия z_i на целевую переменную i-го индивида определяется как:

$$\text{TE}_i = y_{1i} - y_{0i}$$

- Поскольку на практике мы наблюдаем либо y_{1i}, либо y_{0i}, то мы можем оценить лишь условный средний эффект воздействия:

$$\widehat{\text{CATE}}_i = E(y_{1i}|x_i) - E(y_{0i}|x_i) = x_i\beta_1 - x_i\beta_0$$

- Эффект воздействия z_i на целевую переменную i-го индивида определяется как:

$$\text{TE}_i = y_{1i} - y_{0i}$$

- Поскольку на практике мы наблюдаем либо y_{1i}, либо y_{0i}, то мы можем оценить лишь условный средний эффект воздействия:

$$\widehat{\text{CATE}}_i = E(y_{1i}|x_i) - E(y_{0i}|x_i) = x_i\beta_1 - x_i\beta_0$$

- Усредняя CATE можно получить оценку среднего эффекта воздействия:

$$\widehat{\text{ATE}} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\text{CATE}}_i = \frac{1}{n}\sum_{i=1}^{n}x_i\beta_1 - x_i\beta_0$$

- Эффект воздействия z_i на целевую переменную i-го индивида определяется как:

$$\text{TE}_i = y_{1i} - y_{0i}$$

- Поскольку на практике мы наблюдаем либо y_{1i}, либо y_{0i}, то мы можем оценить лишь условный средний эффект воздействия:

$$\widehat{\text{CATE}}_i = E(y_{1i}|x_i) - E(y_{0i}|x_i) = x_i\beta_1 - x_i\beta_0$$

- Усредняя CATE можно получить оценку среднего эффекта воздействия:

$$\widehat{\text{ATE}} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\text{CATE}}_i = \frac{1}{n}\sum_{i=1}^{n}x_i\beta_1 - x_i\beta_0$$

- Наконец, можно также оценить эффект воздействия на подвергшихся воздействию:

$$\text{ATET} = \text{E}(y_{1i}|z_i = 1) - \text{E}(y_{0i}|z_i = 1) = \frac{1}{n}\sum_{i=1}^{n}x_i\left(\beta_1 - \beta_0\right) + \left(\rho_1\sigma_1 - \rho_0\sigma_0\right)\frac{\phi\left(w_i\gamma\right)}{\Phi\left(w_i\gamma\right)}$$

Полупараметрическое оценивание

Краткие комментарии

- Полупараметрическое оценивание осуществляется по аналогии с моделью с неслучайным отбором.

- Полупараметрическое оценивание осуществляется по аналогии с моделью с неслучайным отбором.
- В частности, для оценивания можно применять метод Ньюи и метод Галланта и Нички.

Модель с эндогенным бинарным регрессором
Формулировка

- На практике в исследованиях часто предполагается, что уравнения различаются лишь константами.

- На практике в исследованиях часто предполагается, что уравнения различаются лишь константами.
- Это эквивалентно существованию одного уравнения:

$$y_i = x_i\beta + \alpha z_i + \varepsilon_i,$$

где эндогенность z_i учитывается за счет корреляции между u_i и ε_i.

Модель с эндогенным бинарным регрессором

Формулировка

- На практике в исследованиях часто предполагается, что уравнения различаются лишь константами.
- Это экивалентно существованию одного уравнения:

$$y_i = x_i\beta + \alpha z_i + \varepsilon_i,$$

где эндогенность z_i учитывается за счет корреляции между u_i и ε_i.
- Оценивание таких моделей по аналогии осуществляется при помощи метода максимального правдоподобия или с помощью двухшаговой процедуры, а CATE, ATE и ATET, как и сам эффект воздействия, будут совпадать с α.

Модель с эндогенным бинарным регрессором

Формулировка

- На практике в исследованиях часто предполагается, что уравнения различаются лишь константами.
- Это эквивалентно существованию одного уравнения:

$$y_i = x_i\beta + \alpha z_i + \varepsilon_i,$$

где эндогенность z_i учитывается за счет корреляции между u_i и ε_i.
- Оценивание таких моделей по аналогии осуществляется при помощи метода максимального правдоподобия или с помощью двухшаговой процедуры, а CATE, ATE и ATET, как и сам эффект воздействия, будут совпадать с α.
- Выбрать между данной моделью и моделью с эндогенным переключением можно при помощи LR теста, проверив гипотезу о том, что $\rho_0 = \rho_1$, $\sigma_0 = \sigma_1$, а также о том, что все коэффициенты β_1 и β_0, за исключением константы, совпадают.

- Для наглядности рассмотрим модель, в которой одновременно фигурируют и неслучайный отбор, и эндогенное переключение:

Целевое уравнение 1: $y_{1i}^{*} = x_i\beta_1 + \varepsilon_{1i}$ Целевое уравнение 0: $y_{0i}^{*} = x_i\beta_0 + \varepsilon_{0i}$

Уравнение отбора: $z_{0i}^{*} = w_{0i}\gamma + u_{0i}$ Уравнение переключения: $z_{1i}^{*} = w_{1i}\gamma + u_{1i}$

$$z_{0i} = \begin{cases}1, \text{ если } z_{0i}^{*} \geq 0\\ 0, \text{ в противном случае}\end{cases} \qquad z_{1i} = \begin{cases}1, \text{ если } z_{1i}^{*} \geq 0\\ 0, \text{ в противном случае}\end{cases}$$

$$y_i = \begin{cases}y_{0i}, \text{ если } z_{0i} = 1 \text{ и } z_{1i} = 0\\ y_{1i}, \text{ если } z_{0i} = 1 \text{ и } z_{1i} = 1\\ \text{не наблюдается, в противном случае}\end{cases}$$

- Для наглядности рассмотрим модель, в которой одновременно фигурируют и неслучайный отбор, и эндогенное переключение:

Целевое уравнение 1: $y_{1i}^* = x_i\beta_1 + \varepsilon_{1i}$ Целевое уравнение 0: $y_{0i}^* = x_i\beta_0 + \varepsilon_{0i}$

Уравнение отбора: $z_{0i}^* = w_{0i}\gamma + u_{0i}$ Уравнение переключения: $z_{1i}^* = w_{1i}\gamma + u_{1i}$

$$z_{0i} = \begin{cases} 1, \text{ если } z_{0i}^* \geq 0 \\ 0, \text{ в противном случае} \end{cases} \qquad z_{1i} = \begin{cases} 1, \text{ если } z_{1i}^* \geq 0 \\ 0, \text{ в противном случае} \end{cases}$$

$$y_i = \begin{cases} y_{0i}, \text{ если } z_{0i} = 1 \text{ и } z_{1i} = 0 \\ y_{1i}, \text{ если } z_{0i} = 1 \text{ и } z_{1i} = 1 \\ \text{не наблюдается, в противном случае} \end{cases}$$

- Оценивание этой системы возможно с помощью метода максимального правдоподобия или двухшаговой процедуры.

- Иногда целевых уравнений и латентных переменных может быть несколько:

$$y_{ji} = x_i\beta_j + \varepsilon_{ji} \qquad z_{ji}^* = w_{ji}\gamma_j + u_{ji} \qquad j \in \{0, ..., J-1\}$$

- Порядковый и множественный отборы обычно выглядит как (соответственно):

$$y_i = \begin{cases} y_{0i}, \text{ если } z_{ji}^* < c_1 \\ y_{1i}, \text{ если } c_2 > z_{ji}^* \geq c_1 \\ \vdots \\ y_{(J-1)i}, \text{ если } z_{ji}^* \geq c_{J-1} \end{cases} \qquad y_i = \begin{cases} y_{0i}, \text{ если } z_{0i}^* \geq y_{1i} = z_{1i}^*, ..., z_{0i}^* \geq z_{(J-1)i}^* \\ y_{1i}, \text{ если } z_{1i}^* \geq z_{0i}^*, ..., z_{1i}^* \geq z_{(J-1)i}^* \\ \vdots \\ y_{(J-1)i}, \text{ если } z_{(J-1)i}^* \geq z_{0i}^*, ..., z_{(J-1)i}^* \geq z_{(J-2)i}^* \end{cases}$$

Где $c_1 > c_2 > ... > c_{J-1}$.

- Иногда целевых уравнений и латентных переменных может быть несколько:

$$y_{ji} = x_i\beta_j + \varepsilon_{ji} \qquad z^*_{ji} = w_{ji}\gamma_j + u_{ji} \qquad j \in \{0, ..., J-1\}$$

- Порядковый и множественный отборы обычно выглядит как (соответственно):

$$y_i = \begin{cases} y_{0i}, \text{ если } z^*_{ji} < c_1 \\ y_{1i}, \text{ если } c_2 > z^*_{ji} \geq c_1 \\ \vdots \\ y_{(J-1)i}, \text{ если } z^*_{ji} \geq c_{J-1} \end{cases} \qquad y_i = \begin{cases} y_{0i}, \text{ если } z^*_{0i} \geq y_{1i} = z^*_{1i}, ..., z^*_{0i} \geq z^*_{(J-1)i} \\ y_{1i}, \text{ если } z^*_{1i} \geq z^*_{0i}, ..., z^*_{1i} \geq z^*_{(J-1)i} \\ \vdots \\ y_{(J-1)i}, \text{ если } z^*_{(J-1)i} \geq z^*_{0i}, ..., z^*_{(J-1)i} \geq z^*_{(J-2)i} \end{cases}$$

Где $c_1 > c_2 > ... > c_{J-1}$.

- Например, порядковый отбор возникает, когда уравнения зарплаты y_{ji} различаются в зависимости от уровня образования (базовое, бакалавриат и магистратура), а множественный – когда в зависимости от типа (гуманитарное, техническое, медицинское).