

# Микроэконометрика

## Модели с усечением и цензурированием

Потанин Богдан Станиславович

доцент, научный сотрудник, кандидат экономических наук

2024-2025

# Усеченная регрессия (truncated regression)

## Мотивация

- Часто значение зависимой переменной наблюдается лишь при попадании в определенную область, например:

# Усеченная регрессия (truncated regression)

## Мотивация

- Часто значение зависимой переменной наблюдается лишь при попадании в определенную область, например:
  - Сумма заказа в интернет магазине известна лишь для индивидов, сделавших заказ на сумму не менее минимально установленного магазином порога.

# Усеченная регрессия (truncated regression)

## Мотивация

- Часто значение зависимой переменной наблюдается лишь при попадании в определенную область, например:
  - Сумма заказа в интернет магазине известна лишь для индивидов, сделавших заказ на сумму не менее минимально установленного магазином порога.
  - Информация о ежемесячных доходах может быть собрана по результатам опроса лишь людей со средним достатком, например, с доходом от ста тысяч до миллиона рублей.

# Усеченная регрессия (truncated regression)

## Мотивация

- Часто значение зависимой переменной наблюдается лишь при попадании в определенную область, например:
  - Сумма заказа в интернет магазине известна лишь для индивидов, сделавших заказ на сумму не менее минимально установленного магазином порога.
  - Информация о ежемесячных доходах может быть собрана по результатам опроса лишь людей со средним достатком, например, с доходом от ста тысяч до миллиона рублей.
  - В выборке может содержаться информация о расходах на покупку мороженого лишь среди тех домохозяйств, которые приобрели хотя бы одно мороженое.

# Усеченная регрессия (truncated regression)

## Мотивация

- Часто значение зависимой переменной наблюдается лишь при попадании в определенную область, например:
  - Сумма заказа в интернет магазине известна лишь для индивидов, сделавших заказ на сумму не менее минимально установленного магазином порога.
  - Информация о ежемесячных доходах может быть собрана по результатам опроса лишь людей со средним достатком, например, с доходом от ста тысяч до миллиона рублей.
  - В выборке может содержаться информация о расходах на покупку мороженого лишь среди тех домохозяйств, которые приобрели хотя бы одно мороженое.
- Такого рода зависимые переменные именуются усеченными и для их анализа используются специальные эконометрические методы.

# Усеченная регрессия (truncated regression)

Усеченное нормальное распределение (truncated normal distribution)

- Рассмотрим нормальную случайную величину  $X \sim \mathcal{N}(\mu, \sigma^2)$ .

# Усеченная регрессия (truncated regression)

Усеченное нормальное распределение (truncated normal distribution)

- Рассмотрим нормальную случайную величину  $X \sim \mathcal{N}(\mu, \sigma^2)$ .
- Случайная величина  $Y = X | (a \leq X \leq b)$  будет иметь усеченное нормальное распределение с нижней (левой) и верхней (правой) границами усечения  $a$  и  $b$  соответственно, где  $b > a$ .



# Усеченная регрессия (truncated regression)

Усеченное нормальное распределение (truncated normal distribution)

- Рассмотрим нормальную случайную величину  $X \sim \mathcal{N}(\mu, \sigma^2)$ .
- Случайная величина  $Y = X | (a \leq X \leq b)$  будет иметь усеченное нормальное распределение с нижней (левой) и верхней (правой) границами усечения  $a$  и  $b$  соответственно, где  $b > a$ .
- Математическое ожидание:

$$E(Y) = \mu + \frac{\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}\sigma$$

# Усеченная регрессия (truncated regression)

## Усеченное нормальное распределение (truncated normal distribution)

- Рассмотрим нормальную случайную величину  $X \sim \mathcal{N}(\mu, \sigma^2)$ .
- Случайная величина  $Y = X | (a \leq X \leq b)$  будет иметь усеченное нормальное распределение с нижней (левой) и верхней (правой) границами усечения  $a$  и  $b$  соответственно, где  $b > a$ .
- Математическое ожидание:

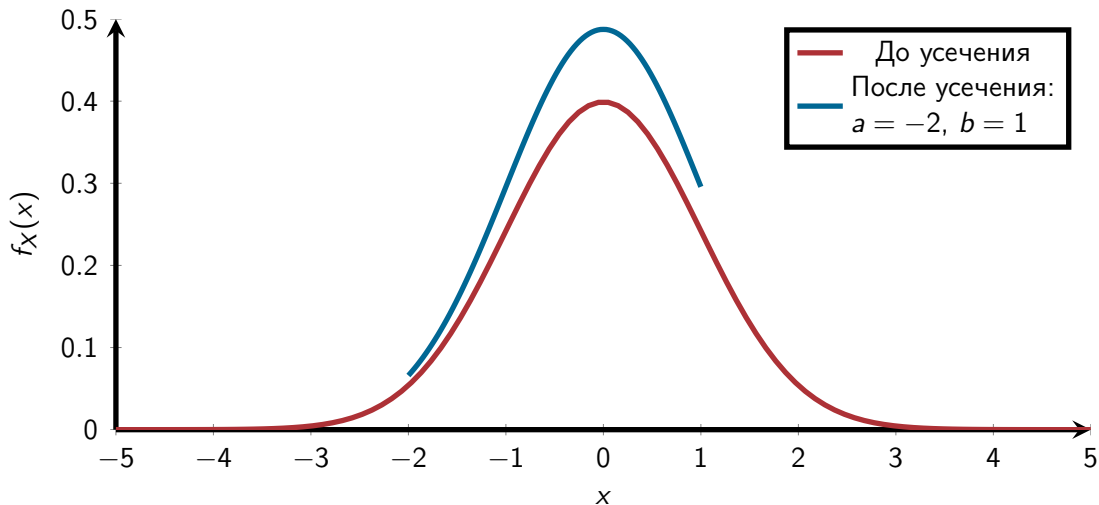
$$E(Y) = \mu + \frac{\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}\sigma$$

- Дисперсия:

$$\text{Var}(Y) = \sigma^2 \left( 1 + \frac{\frac{a-\mu}{\sigma}\phi\left(\frac{a-\mu}{\sigma}\right) - \frac{b-\mu}{\sigma}\phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} - \left( \frac{\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} \right)^2 \right) < \sigma^2$$

# Усеченная регрессия (truncated regression)

Визуализация усеченного нормального распределения



# Усеченная регрессия (truncated regression)

## Формулировка

- Имеется латентная переменная:

$$y_i^* = x_i\beta + \varepsilon_i,$$

где случайные ошибки  $\varepsilon_i$  одинаково распределены и независимы.

# Усеченная регрессия (truncated regression)

## Формулировка

- Имеется латентная переменная:

$$y_i^* = x_i\beta + \varepsilon_i,$$

где случайные ошибки  $\varepsilon_i$  одинаково распределены и независимы.

- Исследователь наблюдает лишь зависимую переменную, подвергнутую усечению:

$$y_i = \begin{cases} y_i^*, & \text{если } a \leq y_i^* \leq b \\ \text{не наблюдается, в противном случае} \end{cases},$$

где  $a$  и  $b$  именуются нижней (левой) и верхней (правой) границами усечения соответственно.

# Усеченная регрессия (truncated regression)

## Формулировка

- Имеется латентная переменная:

$$y_i^* = x_i\beta + \varepsilon_i,$$

где случайные ошибки  $\varepsilon_i$  одинаково распределены и независимы.

- Исследователь наблюдает лишь зависимую переменную, подвергнутую усечению:

$$y_i = \begin{cases} y_i^*, & \text{если } a \leq y_i^* \leq b \\ \text{не наблюдается, в противном случае} \end{cases},$$

где  $a$  и  $b$  именуются нижней (левой) и верхней (правой) границами усечения соответственно.

- Если исследователь изучает факторы, влияющие на ежемесячные доходы индивидов по выборке, собранной по результатам опроса лишь индивидов с доходом от ста тысяч до миллиона рублей, то  $a = 10^5$  и  $b = 10^6$ .

# Усеченная регрессия (truncated regression)

## Формулировка

- Имеется латентная переменная:

$$y_i^* = x_i\beta + \varepsilon_i,$$

где случайные ошибки  $\varepsilon_i$  одинаково распределены и независимы.

- Исследователь наблюдает лишь зависимую переменную, подвергнутую усечению:

$$y_i = \begin{cases} y_i^*, & \text{если } a \leq y_i^* \leq b \\ \text{не наблюдается, в противном случае} \end{cases},$$

где  $a$  и  $b$  именуются нижней (левой) и верхней (правой) границами усечения соответственно.

- Если исследователь изучает факторы, влияющие на ежемесячные доходы индивидов по выборке, собранной по результатам опроса лишь индивидов с доходом от ста тысяч до миллиона рублей, то  $a = 10^5$  и  $b = 10^6$ .
- При изучении факторов, влияющих на объем покупок в интернет магазине при минимальной сумме заказа в 500 рублей, границы усечения могут быть заданы как  $a = 500$  и  $b = \infty$ .

# Усеченная регрессия (truncated regression)

## Проблема оценивания

- Условное математическое ожидание имеет вид:

$$E(y_i^* | a \leq y_i^* \leq b) = x_i\beta + E(\varepsilon_i | a - x_i\beta \leq \varepsilon_i \leq b - x_i\beta)$$



# Усеченная регрессия (truncated regression)

## Проблема оценивания

- Условное математическое ожидание имеет вид:

$$E(y_i^* | a \leq y_i^* \leq b) = x_i\beta + E(\varepsilon_i | a - x_i\beta \leq \varepsilon_i \leq b - x_i\beta)$$

- Для простоты предположим, что случайные ошибки имеют нормальное распределение  $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ , откуда по формуле математического ожидания усеченного нормального распределения получаем:

$$E(\varepsilon_i | a - x_i\beta \leq \varepsilon_i \leq b - x_i\beta) = \frac{\phi\left(\frac{a - x_i\beta}{\sigma}\right) - \phi\left(\frac{b - x_i\beta}{\sigma}\right)}{\Phi\left(\frac{b - x_i\beta}{\sigma}\right) - \Phi\left(\frac{a - x_i\beta}{\sigma}\right)}\sigma = \sigma\lambda_i$$

# Усеченная регрессия (truncated regression)

## Проблема оценивания

- Условное математическое ожидание имеет вид:

$$E(y_i^* | a \leq y_i^* \leq b) = x_i\beta + E(\varepsilon_i | a - x_i\beta \leq \varepsilon_i \leq b - x_i\beta)$$

- Для простоты предположим, что случайные ошибки имеют нормальное распределение  $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ , откуда по формуле математического ожидания усеченного нормального распределения получаем:

$$E(\varepsilon_i | a - x_i\beta \leq \varepsilon_i \leq b - x_i\beta) = \frac{\phi\left(\frac{a - x_i\beta}{\sigma}\right) - \phi\left(\frac{b - x_i\beta}{\sigma}\right)}{\Phi\left(\frac{b - x_i\beta}{\sigma}\right) - \Phi\left(\frac{a - x_i\beta}{\sigma}\right)}\sigma = \sigma\lambda_i$$

- Очевидно, что в данном случае условное математическое ожидание случайной ошибки коррелирует с регрессорами  $x_i$ , что влечет несостоятельность МНК оценок  $\beta$ . При этом  $\lambda_i$  можно рассматривать как пропущенную значимую переменную с коэффициентом  $\sigma$  (omitted variable bias).

# Усеченная регрессия (truncated regression)

## Оценивание

- Параметры  $\beta$  и стандартное отклонение случайной ошибки  $\sigma$  можно оценить при помощи метода максимального правдоподобия.

# Усеченная регрессия (truncated regression)

## Оценивание

- Параметры  $\beta$  и стандартное отклонение случайной ошибки  $\sigma$  можно оценить при помощи метода максимального правдоподобия.
- Функция правдоподобия с учетом усечения будет иметь вид:

$$\begin{aligned} L(\beta, \sigma; y|X) &= \prod_{i=1}^n f_{y_i^* | (a \leq y_i^* \leq b, x_i)}(y_i^*) = \prod_{i=1}^n f_{\varepsilon_i | (a - x_i\beta \leq \varepsilon_i \leq b - x_i\beta, x_i)}(y_i^* - x_i\beta) = \\ &= \prod_{i=1}^n \frac{1}{\sigma} \phi\left(\frac{y_i^* - x_i\beta}{\sigma}\right) / \left[ \Phi\left(\frac{b - x_i\beta}{\sigma}\right) - \Phi\left(\frac{a - x_i\beta}{\sigma}\right) \right] \end{aligned}$$

# Усеченная регрессия (truncated regression)

## Оценивание

- Параметры  $\beta$  и стандартное отклонение случайной ошибки  $\sigma$  можно оценить при помощи метода максимального правдоподобия.
- Функция правдоподобия с учетом усечения будет иметь вид:

$$\begin{aligned} L(\beta, \sigma; y|X) &= \prod_{i=1}^n f_{y_i^* | (a \leq y_i^* \leq b, x_i)}(y_i^*) = \prod_{i=1}^n f_{\varepsilon_i | (a - x_i\beta \leq \varepsilon_i \leq b - x_i\beta, x_i)}(y_i^* - x_i\beta) = \\ &= \prod_{i=1}^n \frac{1}{\sigma} \phi\left(\frac{y_i^* - x_i\beta}{\sigma}\right) / \left[ \Phi\left(\frac{b - x_i\beta}{\sigma}\right) - \Phi\left(\frac{a - x_i\beta}{\sigma}\right) \right] \end{aligned}$$

- Оценки усеченной регрессии могут оказаться несостоятельными при нарушении допущения о нормальном распределении случайной ошибки или при наличии гетероскедастичности.

# Усеченная регрессия (truncated regression)

## Оценивание

- Параметры  $\beta$  и стандартное отклонение случайной ошибки  $\sigma$  можно оценить при помощи метода максимального правдоподобия.
- Функция правдоподобия с учетом усечения будет иметь вид:

$$\begin{aligned} L(\beta, \sigma; y|X) &= \prod_{i=1}^n f_{y_i^* | (a \leq y_i^* \leq b, x_i)}(y_i^*) = \prod_{i=1}^n f_{\varepsilon_i | (a - x_i\beta \leq \varepsilon_i \leq b - x_i\beta, x_i)}(y_i^* - x_i\beta) = \\ &= \prod_{i=1}^n \frac{1}{\sigma} \phi\left(\frac{y_i^* - x_i\beta}{\sigma}\right) / \left[ \Phi\left(\frac{b - x_i\beta}{\sigma}\right) - \Phi\left(\frac{a - x_i\beta}{\sigma}\right) \right] \end{aligned}$$

- Оценки усеченной регрессии могут оказаться несостоятельными при нарушении допущения о нормальном распределении случайной ошибки или при наличии гетероскедастичности.
- Модель можно по аналогии оценивать при альтернативных (в том числе гибких) предположениях о распределении случайной ошибки.

# Усеченная регрессия (truncated regression)

## Оценивание

- Параметры  $\beta$  и стандартное отклонение случайной ошибки  $\sigma$  можно оценить при помощи метода максимального правдоподобия.
- Функция правдоподобия с учетом усечения будет иметь вид:

$$\begin{aligned} L(\beta, \sigma; y|X) &= \prod_{i=1}^n f_{y_i^* | (a \leq y_i^* \leq b, x_i)}(y_i^*) = \prod_{i=1}^n f_{\varepsilon_i | (a - x_i\beta \leq \varepsilon_i \leq b - x_i\beta, x_i)}(y_i^* - x_i\beta) = \\ &= \prod_{i=1}^n \frac{1}{\sigma} \phi\left(\frac{y_i^* - x_i\beta}{\sigma}\right) / \left[ \Phi\left(\frac{b - x_i\beta}{\sigma}\right) - \Phi\left(\frac{a - x_i\beta}{\sigma}\right) \right] \end{aligned}$$

- Оценки усеченной регрессии могут оказаться несостоятельными при нарушении допущения о нормальном распределении случайной ошибки или при наличии гетероскедастичности.
- Модель можно по аналогии оценивать при альтернативных (в том числе гибких) предположениях о распределении случайной ошибки.
- Для учета гетероскедастичности  $\sigma$  может быть специфицирована по аналогии с гетероскедастичной пробит моделью.

# Усеченная регрессия (truncated regression)

Предельные эффекты на математическое ожидание

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

$$\frac{\partial E(y_i^* | x_i)}{\partial x_{ik}} = \beta_k$$



# Усеченная регрессия (truncated regression)

## Предельные эффекты на математическое ожидание

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

$$\frac{\partial E(y_i^* | x_i)}{\partial x_{ik}} = \beta_k$$

- Предельный эффект на усеченное математическое ожидание рассчитывается как:

$$\frac{\partial E(y_i^* | a < y_i^* < b, x_i)}{\partial x_{ik}} = \beta_k \underbrace{\text{Var}(\varepsilon_i | a - x_i\beta < \varepsilon_i < b - x_i\beta, x_i)}_{0 < \text{эффект усечения} < 1} / \sigma^2 =$$

# Усеченная регрессия (truncated regression)

## Предельные эффекты на математическое ожидание

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

$$\frac{\partial E(y_i^* | x_i)}{\partial x_{ik}} = \beta_k$$

- Предельный эффект на усеченное математическое ожидание рассчитывается как:

$$\begin{aligned} \frac{\partial E(y_i^* | a < y_i^* < b, x_i)}{\partial x_{ik}} &= \beta_k \underbrace{\text{Var}(\varepsilon_i | a - x_i\beta < \varepsilon_i < b - x_i\beta, x_i) / \sigma^2}_{0 < \text{эффект усечения} < 1} = \\ &= \beta_k \left( 1 + \frac{\frac{a - x_i\beta}{\sigma} \phi\left(\frac{a - x_i\beta}{\sigma}\right) - \frac{b - x_i\beta}{\sigma} \phi\left(\frac{b - x_i\beta}{\sigma}\right)}{\Phi\left(\frac{b - x_i\beta}{\sigma}\right) - \Phi\left(\frac{a - x_i\beta}{\sigma}\right)} - \left( \frac{\phi\left(\frac{a - x_i\beta}{\sigma}\right) - \phi\left(\frac{b - x_i\beta}{\sigma}\right)}{\Phi\left(\frac{b - x_i\beta}{\sigma}\right) - \Phi\left(\frac{a - x_i\beta}{\sigma}\right)} \right)^2 \right) \end{aligned}$$

# Усеченная регрессия (truncated regression)

## Предельные эффекты на математическое ожидание

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

$$\frac{\partial E(y_i^* | x_i)}{\partial x_{ik}} = \beta_k$$

- Предельный эффект на усеченное математическое ожидание рассчитывается как:

$$\begin{aligned} \frac{\partial E(y_i^* | a < y_i^* < b, x_i)}{\partial x_{ik}} &= \beta_k \underbrace{\text{Var}(\varepsilon_i | a - x_i\beta < \varepsilon_i < b - x_i\beta, x_i) / \sigma^2}_{0 < \text{эффект усечения} < 1} = \\ &= \beta_k \left( 1 + \frac{\frac{a-x_i\beta}{\sigma} \phi\left(\frac{a-x_i\beta}{\sigma}\right) - \frac{b-x_i\beta}{\sigma} \phi\left(\frac{b-x_i\beta}{\sigma}\right)}{\Phi\left(\frac{b-x_i\beta}{\sigma}\right) - \Phi\left(\frac{a-x_i\beta}{\sigma}\right)} - \left( \frac{\phi\left(\frac{a-x_i\beta}{\sigma}\right) - \phi\left(\frac{b-x_i\beta}{\sigma}\right)}{\Phi\left(\frac{b-x_i\beta}{\sigma}\right) - \Phi\left(\frac{a-x_i\beta}{\sigma}\right)} \right)^2 \right) \end{aligned}$$

Данный предельный эффект меньше  $\beta_k$  и совпадает с ним по знаку, но его величина зависит от  $x_i\beta$ .

# Усеченная регрессия (truncated regression)

Предельный эффект на вероятность усечения

- Предельный эффект переменной  $x_{ik}$  на вероятность усечения рассчитывается как:

$$\frac{\partial P(a < y_i^* < b | x_i)}{\partial x_{ik}} = \frac{\beta_k}{\sigma} \left( \phi \left( \frac{a - x_i \beta}{\sigma} \right) - \phi \left( \frac{b - x_i \beta}{\sigma} \right) \right)$$

# Усеченная регрессия (truncated regression)

## Предельный эффект на вероятность усечения

- Предельный эффект переменной  $x_{ik}$  на вероятность усечения рассчитывается как:

$$\frac{\partial P(a < y_i^* < b | x_i)}{\partial x_{ik}} = \frac{\beta_k}{\sigma} \left( \phi \left( \frac{a - x_i \beta}{\sigma} \right) - \phi \left( \frac{b - x_i \beta}{\sigma} \right) \right)$$

- Знак данного предельного эффекта может не совпадать со знаком  $\beta_k$ , поскольку разница функций плотности может оказаться как положительной, так и отрицательной.

- Помимо усечения в данных также часто встречается цензурирование – когда исследователю доступна некоторая информация об объектах, не попавших в выборку, но неизвестны точные значения зависимой переменной. Ситуации, в которых встречается цензурирование, часто схожи с теми, в которых имеет место усечение.

- Помимо усечения в данных также часто встречается цензурирование – когда исследователю доступна некоторая информация об объектах, не попавших в выборку, но неизвестны точные значения зависимой переменной. Ситуации, в которых встречается цензурирование, часто схожи с теми, в которых имеет место усечение.
- Рассмотрим разницу между усечением и цензурированием на нескольких примерах:
  - В усеченной выборке имеется информация лишь об индивидах с положительной заработной платой, а в цензурированной также могут иметься характеристики неработающих индивидов, у которых зарплата равна нулю.

- Помимо усечения в данных также часто встречается цензурирование – когда исследователю доступна некоторая информация об объектах, не попавших в выборку, но неизвестны точные значения зависимой переменной. Ситуации, в которых встречается цензурирование, часто схожи с теми, в которых имеет место усечение.
- Рассмотрим разницу между усечением и цензурированием на нескольких примерах:
  - В усеченной выборке имеется информация лишь об индивидах с положительной заработной платой, а в цензурированной также могут иметься характеристики неработающих индивидов, у которых зарплата равна нулю.
  - В усеченной выборке имеется информация о расходах на театр лишь для тех индивидов, которые ходят в театр, а в цензурированной выборке также известны характеристики индивидов, которые вообще не посещают театр.



- Помимо усечения в данных также часто встречается цензурирование – когда исследователю доступна некоторая информация об объектах, не попавших в выборку, но неизвестны точные значения зависимой переменной. Ситуации, в которых встречается цензурирование, часто схожи с теми, в которых имеет место усечение.
- Рассмотрим разницу между усечением и цензурированием на нескольких примерах:
  - В усеченной выборке имеется информация лишь об индивидах с положительной заработной платой, а в цензурированной также могут иметься характеристики неработающих индивидов, у которых зарплата равна нулю.
  - В усеченной выборке имеется информация о расходах на театр лишь для тех индивидов, которые ходят в театр, а в цензурированной выборке также известны характеристики индивидов, которые вообще не посещают театр.
  - Число проданных билетов на стадионе с ограниченным числом мест скорее подходит под цензурирование, чем под усечение, поскольку обычно имеется информация о случаях, когда стадион был полностью заполнен.

- Помимо усечения в данных также часто встречается цензурирование – когда исследователю доступна некоторая информация об объектах, не попавших в выборку, но неизвестны точные значения зависимой переменной. Ситуации, в которых встречается цензурирование, часто схожи с теми, в которых имеет место усечение.
- Рассмотрим разницу между усечением и цензурированием на нескольких примерах:
  - В усеченной выборке имеется информация лишь об индивидах с положительной заработной платой, а в цензурированной также могут иметься характеристики неработающих индивидов, у которых зарплата равна нулю.
  - В усеченной выборке имеется информация о расходах на театр лишь для тех индивидов, которые ходят в театр, а в цензурированной выборке также известны характеристики индивидов, которые вообще не посещают театр.
  - Число проданных билетов на стадионе с ограниченным числом мест скорее подходит под цензурирование, чем под усечение, поскольку обычно имеется информация о случаях, когда стадион был полностью заполнен.
- Модели, построенные на цензурированных данных, инкорпорируют больший объем информации, чем модели, построенные на схожих усеченных данных. Поэтому оценки моделей с цензурированием обычно эффективней оценок моделей с усечением.

- При цензурировании зависимая переменная имеет вид:

$$y_i = \begin{cases} a, & \text{если } y_i^* < a \\ y_i^*, & \text{если } a \leq y_i^* \leq b \\ b, & \text{если } y_i^* > b \end{cases}$$

- При цензурировании зависимая переменная имеет вид:

$$y_i = \begin{cases} a, & \text{если } y_i^* < a \\ y_i^*, & \text{если } a \leq y_i^* \leq b \\ b, & \text{если } y_i^* > b \end{cases}$$

- Условное математическое ожидание зависимой переменной имеет вид:

$$\begin{aligned} E(y_i|x_i) &= P(y_i^* \leq a|x_i) \times a + P(a < y_i^* < b|x_i) \times E(y_i^*|a < y_i^* < b, x_i) + P(y_i^* \geq b|x_i) \times b = \\ &= P(\varepsilon_i < a - x_i\beta|x_i) \times a + x_i\beta + E(\varepsilon_i|a - x_i\beta < \varepsilon_i < b - x_i\beta, x_i) + P(\varepsilon_i > b - x_i\beta|x_i) \times b \end{aligned}$$

- При цензурировании зависимая переменная имеет вид:

$$y_i = \begin{cases} a, & \text{если } y_i^* < a \\ y_i^*, & \text{если } a \leq y_i^* \leq b \\ b, & \text{если } y_i^* > b \end{cases}$$

- Условное математическое ожидание зависимой переменной имеет вид:

$$\begin{aligned} E(y_i|x_i) &= P(y_i^* \leq a|x_i) \times a + P(a < y_i^* < b|x_i) \times E(y_i^*|a < y_i^* < b, x_i) + P(y_i^* \geq b|x_i) \times b = \\ &= P(\varepsilon_i < a - x_i\beta|x_i) \times a + x_i\beta + E(\varepsilon_i|a - x_i\beta < \varepsilon_i < b - x_i\beta, x_i) + P(\varepsilon_i > b - x_i\beta|x_i) \times b \end{aligned}$$

- Условное математическое ожидание случайной ошибки аналогично тому, что было получено в случае с усеченной регрессией, из-за чего МНК оценки коэффициентов  $\beta$  окажутся несостоятельными.

- По аналогии с усеченной регрессией оценивание параметров  $\beta$  и  $\sigma$  осуществляется с помощью метода максимального правдоподобия со следующей функцией правдоподобия:

$$L(\beta, \sigma; y|X) = \prod_{i: a < y_i < b} f_{y_i|x_i}(y_i) \prod_{i: y_i = a} P(y_i = a|x_i) \prod_{i: y_i = b} P(y_i = b|x_i) =$$

- По аналогии с усеченной регрессией оценивание параметров  $\beta$  и  $\sigma$  осуществляется с помощью метода максимального правдоподобия со следующей функцией правдоподобия:

$$\begin{aligned} L(\beta, \sigma; y|X) &= \prod_{i:a < y_i < b} f_{y_i|x_i}(y_i) \prod_{i:y_i=a} P(y_i = a|x_i) \prod_{i:y_i=b} P(y_i = b|x_i) = \\ &= \prod_{i:a < y_i < b} f_{\varepsilon_i}(y_i - x_i\beta) \prod_{i:y_i=a} P(\varepsilon_i < a - x_i\beta|x_i) \prod_{i:y_i=b} P(\varepsilon_i > b - x_i\beta|x_i) = \end{aligned}$$

- По аналогии с усеченной регрессией оценивание параметров  $\beta$  и  $\sigma$  осуществляется с помощью метода максимального правдоподобия со следующей функцией правдоподобия:

$$\begin{aligned} L(\beta, \sigma; y|X) &= \prod_{i:a < y_i < b} f_{y_i|x_i}(y_i) \prod_{i:y_i=a} P(y_i = a|x_i) \prod_{i:y_i=b} P(y_i = b|x_i) = \\ &= \prod_{i:a < y_i < b} f_{\varepsilon_i}(y_i - x_i\beta) \prod_{i:y_i=a} P(\varepsilon_i < a - x_i\beta|x_i) \prod_{i:y_i=b} P(\varepsilon_i > b - x_i\beta|x_i) = \\ &= \prod_{i:a < y_i < b} \frac{1}{\sigma} \phi\left(\frac{y_i - x_i\beta}{\sigma}\right) \prod_{i:y_i=a} \Phi\left(\frac{a - x_i\beta}{\sigma}\right) \prod_{i:y_i=b} 1 - \Phi\left(\frac{b - x_i\beta}{\sigma}\right) \end{aligned}$$



- По аналогии с усеченной регрессией оценивание параметров  $\beta$  и  $\sigma$  осуществляется с помощью метода максимального правдоподобия со следующей функцией правдоподобия:

$$\begin{aligned} L(\beta, \sigma; y|X) &= \prod_{i:a < y_i < b} f_{y_i|x_i}(y_i) \prod_{i:y_i=a} P(y_i = a|x_i) \prod_{i:y_i=b} P(y_i = b|x_i) = \\ &= \prod_{i:a < y_i < b} f_{\varepsilon_i}(y_i - x_i\beta) \prod_{i:y_i=a} P(\varepsilon_i < a - x_i\beta|x_i) \prod_{i:y_i=b} P(\varepsilon_i > b - x_i\beta|x_i) = \\ &= \prod_{i:a < y_i < b} \frac{1}{\sigma} \phi\left(\frac{y_i - x_i\beta}{\sigma}\right) \prod_{i:y_i=a} \Phi\left(\frac{a - x_i\beta}{\sigma}\right) \prod_{i:y_i=b} 1 - \Phi\left(\frac{b - x_i\beta}{\sigma}\right) \end{aligned}$$

- Альтернативные формы распределения случайной ошибки и гетероскедастичность могут быть учтены по аналогии с усеченной регрессией.

- По аналогии с усеченной регрессией оценивание параметров  $\beta$  и  $\sigma$  осуществляется с помощью метода максимального правдоподобия со следующей функцией правдоподобия:

$$\begin{aligned} L(\beta, \sigma; y|X) &= \prod_{i:a < y_i < b} f_{y_i|x_i}(y_i) \prod_{i:y_i=a} P(y_i = a|x_i) \prod_{i:y_i=b} P(y_i = b|x_i) = \\ &= \prod_{i:a < y_i < b} f_{\varepsilon_i}(y_i - x_i\beta) \prod_{i:y_i=a} P(\varepsilon_i < a - x_i\beta|x_i) \prod_{i:y_i=b} P(\varepsilon_i > b - x_i\beta|x_i) = \\ &= \prod_{i:a < y_i < b} \frac{1}{\sigma} \phi\left(\frac{y_i - x_i\beta}{\sigma}\right) \prod_{i:y_i=a} \Phi\left(\frac{a - x_i\beta}{\sigma}\right) \prod_{i:y_i=b} 1 - \Phi\left(\frac{b - x_i\beta}{\sigma}\right) \end{aligned}$$

- Альтернативные формы распределения случайной ошибки и гетероскедастичность могут быть учтены по аналогии с усеченной регрессией.
- Для того, чтобы гарантировать нахождение глобального максимума данной функции правдоподобия, ее может сделать вогнутой за счет репараметризации Олсена, оценивая вместо  $\beta$  и  $\sigma$  такие  $\beta^*$  и  $\sigma^*$ , что  $\beta = \beta^*/\sigma^*$  и  $\sigma = 1/\sigma^*$ .

# Модель Тобина

## Сравнение модели Тобина и усеченной регрессии

- Состоятельные оценки параметров модели Тобина (с цензурированием) можно также получить и с помощью усеченной регрессии.

# Модель Тобина

## Сравнение модели Тобина и усеченной регрессии

- Состоятельные оценки параметров модели Тобина (с цензурированием) можно также получить и с помощью усеченной регрессии.
- Однако, оценки усеченной регрессии окажутся менее эффективными, поскольку игнорируют информацию о случаях, когда  $y_i = a$  и  $y_i = b$ .

# Модель Тобина

## Сравнение модели Тобина и усеченной регрессии

- Состоятельные оценки параметров модели Тобина (с цензурированием) можно также получить и с помощью усеченной регрессии.
- Однако, оценки усеченной регрессии окажутся менее эффективными, поскольку игнорируют информацию о случаях, когда  $y_i = a$  и  $y_i = b$ .
- Поскольку модель Тобина и усеченная регрессия оцениваются на разных выборках, то их непосредственное сравнение по информационным критериям является некорректным. В качестве альтернативы эти модели можно сравнить, например, по предсказательной силе, однако по возможности следует предпочитать модель Тобина в силу большей эффективности ее оценок.

- Предельный эффект переменной  $x_{ik}$  на обычное  $E(y_i^*|x_i)$  и усеченное  $E(y_i^*|a < y_i^* < b)$  математические ожидания латентной переменной аналогичны тем, что были рассмотрены в случае с усеченной регрессией.

- Предельный эффект переменной  $x_{ik}$  на обычное  $E(y_i^*|x_i)$  и усеченное  $E(y_i^*|a < y_i^* < b)$  математические ожидания латентной переменной аналогичны тем, что были рассмотрены в случае с усеченной регрессией.
- Предельный эффект на математическое ожидание наблюдаемой (с учетом цензурирования) зависимой переменной рассчитывается как:

$$\frac{\partial E(y_i|x_i)}{\partial x_{ik}} = \beta_k P(a < y_i^* < b|x_i) = \beta_k \left( \Phi\left(\frac{b - x_i\beta}{\sigma}\right) - \Phi\left(\frac{a - x_i\beta}{\sigma}\right) \right)$$

Данный предельный эффект меньше  $\beta_k$  и совпадает с ним по знаку, но его величина зависит от  $x_i\beta$ . То, что данный предельный эффект меньше, интуитивно связано с тем, что при попадании зависимой переменной под цензурирование малое изменение регрессора не изменит значения зависимой переменной.

- Предельный эффект переменной  $x_{ik}$  на обычное  $E(y_i^*|x_i)$  и усеченное  $E(y_i^*|a < y_i^* < b)$  математические ожидания латентной переменной аналогичны тем, что были рассмотрены в случае с усеченной регрессией.
- Предельный эффект на математическое ожидание наблюдаемой (с учетом цензурирования) зависимой переменной рассчитывается как:

$$\frac{\partial E(y_i|x_i)}{\partial x_{ik}} = \beta_k P(a < y_i^* < b|x_i) = \beta_k \left( \Phi\left(\frac{b - x_i\beta}{\sigma}\right) - \Phi\left(\frac{a - x_i\beta}{\sigma}\right) \right)$$

Данный предельный эффект меньше  $\beta_k$  и совпадает с ним по знаку, но его величина зависит от  $x_i\beta$ . То, что данный предельный эффект меньше, интуитивно связано с тем, что при попадании зависимой переменной под цензурирование малое изменение регрессора не изменит значения зависимой переменной.

- Предельные эффекты на вероятность цензурирования считаются как:

$$\frac{\partial P(y_i = b|x_i)}{\partial x_{ik}} = \frac{\beta_k}{\sigma} \phi\left(\frac{x_i\beta - b}{\sigma}\right), \quad \frac{\partial P(y_i = a|x_i)}{\partial x_{ik}} = -\frac{\beta_k}{\sigma} \phi\left(\frac{a - x_i\beta}{\sigma}\right)$$