Моделі з дискретними й обмеженими змінними

Поняття «фіктивної змінної». Види моделей із фіктивними незалежними змінними.

Дослідження структурних змін за допомогою тесту Чоу.

Моделі з дискретними залежними змінними.

Лінійна модель бінарного вибору.

Логіт-модель і пробіт-модель.

Моделі множинного вибору.

Тобіт-модель.

Ключові слова: фіктивні незалежні змінні; фіктивна змінна зрушення; фіктивна змінна нахилу; моделі зі змінною структурою; тест Чоу; моделі з дискретними залежними змінними; лінійна модель ймовірності; логіт-модель; пробіт-модель; методи оцінювання параметрів; критерії якості; моделі множинного вибору; гніздові множинні логіт-моделі; порядкові логіт-, пробіт-моделі; моделі рахункових даних; тобіт-модель; модель Хекмана.

Поняття «фіктивної змінної». Види моделей із фіктивними незалежними змінними

У розглянутих регресійних моделях передбачалося, що факторні та результівні ознаки мають безперервні області зміни (валовий регіональний продукт, розмір середньомісячної заробітної платні тощо). Однак у ряді випадків факторні й/або результівні змінні є якісними ознаками, які можуть приймати дискретну множину значень. Зокрема, на динаміку фінансового результату від звичайної діяльності підприємства поряд з кількісними факторами можуть впливати такі якісні ознаки, як зміни податкової політики, регулятивного законодавства, форм власності та ін.

Фіктивні незалежні змінні вводяться в економетричну модель із метою урахування впливу якісних аспектів на закономірності розвитку розглянутих процесів. Як правило, якісні змінні є бінарними та приймають

значення «1» за наявності деякої властивості та значення «0» у випадку її відсутності. Такі змінні називають *dummy*-змінними. *Dummy*-змінні можуть використовуватися в економетричних моделях поряд з кількісними змінними, а можуть лежати в основі формування економетричних моделей, в яких усі змінні є якісними.

Для прикладу розглянемо модель оцінки вартості комерційної нерухомості. Поряд з такими кількісними змінними $x_i = (x_{i1}, \dots, x_{ik})$ – такими, як середньозважене зношування об'єкта нерухомості, висота та ін., на вартість об'єкта впливає його функційне призначення (магазин або офіс). Введемо фіктивну змінну d_i, яка приймає такі значення:

$$d_i = \begin{cases} 1, \text{якщо } i\text{-й об}'\text{єкт нерухомості є магазином;} \\ 0, \text{у протилежному випадку.} \end{cases}$$

Модель оцінювання вартості комерційної нерухомості має вигляд:

$$y_i = a_0 + \sum_{j=1}^{k} a_j\, x_{ij} + \beta d_i + \varepsilon_i$$

Коефіцієнт β при *dummy*-змінній називають *диференційованим коефіцієнтом перетинання.* Він показує, на скільки вартість торговельної нерухомості відрізняється від вартості офісної за інших рівних умов. Змінна d_i – *фіктивна змінна перетинання.* В якості базової категорії розглядаються об'єкти офісної нерухомості. Рівняння регресії для кожної з категорій має вигляд:

для офісної нерухомості: $y_i = a_0 + \sum_{j=1}^{k} a_j\, x_{ij} + \varepsilon_i$;

для торговельної нерухомості: $y_i = (a_0 + \beta) + \sum_{j=1}^{k} a_j\, x_{ij} + \varepsilon_i$.

Модель (6.1) містить як фіктивні, так і кількісні пояснювальні змінні. Такі моделі називають *ANCOVA-моделями (моделями коваріаційного аналізу).* Економетричні моделі, які містять у правій частині тільки фіктивні змінні, називають *ANOVA-моделями (моделями дисперсійного аналізу).*

Ще одним наочним прикладом застосування фіктивних змінних є модель регресії, що відображає проблему розриву в заробітній платні. На розмір заробітної платні впливають такі кількісні змінні, як вік (x_{i1}), кількість співробітників у компанії (x_{i2}), а також якісні змінні: наявність вищої освіти ($d_{i1} = 1$, якщо працівник має вищу освіту; 0 – якщо ні); стать

($d_{i2} = 1$, якщо стать чоловіча; 0 – якщо ні); родинний стан ($d_{i3} = 1$, якщо співробітник одружений; 0 – якщо ні); проходження курсів/тренінгів ($d_{i4} = 1$, якщо співробітник пройшов курси підвищення кваліфікації; 0 – якщо ні). З урахуванням припущень про тип зв'язку модель має вигляд:

$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + \sum_{j=1}^{4} \beta_j \, d_{ij} + \varepsilon_i.$$

Відповідні диференційовані коефіцієнти перетинання β_j дозволяють оцінити розрив у рівнях заробітної платні співробітників із вищою освітою та без неї, досліджувати проблему дискримінації в оплаті праці між чоловіками та жінками тощо.

Моделі типу і можуть бути сформовані на випадок більшої кількості груп фіктивних змінних. Якщо якісна ознака, що розглядається, має не два, а кілька значень, то можна було б ввести дискретну змінну, що приймає таку ж кількість значень. Але це робиться вкрай рідко, оскільки важко дати змістовну інтерпретацію відповідному коефіцієнту. У цьому випадку доцільно використовувати декілька бінарних змінних. Прикладом є дослідження сезонних коливань. Розглянемо регресійну модель аналізу обсягу споживання (y) від доходу (x). Модель можна подати у вигляді:

$$y_i = a_0 + a_1 x_i + \varepsilon_i$$

У даній моделі коефіцієнт a_1 називають «схильністю до споживання». Якщо є підстави вважати, що обсяг споживання (y) залежить від пори року, то для виявлення сезонності можна ввести три бінарні змінні d_{i1}, d_{i2}, d_{i3}:

$$d_{i1} = \begin{cases} 1, \text{якщо місяць є зимовим;} \\ 0, \text{у протилежному випадку;} \end{cases}$$

$$d_{i2} = \begin{cases} 1, \text{якщо місяць є весняним;} \\ 0, \text{у протилежному випадку;} \end{cases}$$

$$d_{i3} = \begin{cases} 1, \text{якщо місяць є літнім;} \\ 0, \text{у протилежному випадку.} \end{cases}$$

У цьому випадку рівняння регресії буде мати вигляд:

$$y_i = a_0 + a_1 x_i + \beta_1 d_{i1} + \beta_2 d_{i2} + \beta_3 d_{i3} + \varepsilon_i.$$

Слід зазначити, що вводити четверту змінну – осінню не можна, інакше виконувалася б тотожність $d_{i1} + d_{i2} + d_{i3} + d_{i4} = 1, i = \overline{1,n}$. Це означає лінійну залежність факторів (явище мультиколінеарності) і, як наслідок, неможливість отримання оцінок МНК. Ситуацію, коли сума фіктивних змінних тотожно дорівнює константі, називають *пасткою фіктивних змінних* («*dummy trap*»). Щоб уникнути такої пастки, дотримуються правила, згідно з яким, якщо якісна ознака має p альтернативних значень (градацій), то число введених бінарних фіктивних змінних повинне дорівнювати $p - 1$.

Модель), у яку включені всі фіктивні змінні, називають *моделлю регресії без обмежень (unrestricted regression). Базисною моделлю, або регресією з обмеженнями (restricted regression,)* є модель регресії, в якій усі значення фіктивних змінних дорівнюють нулю.

Для наведеного прикладу модель регресії виду $y_i = a_0 + a_1 x_i + +\beta_1 d_{i1} + \beta_2 d_{i2} + \beta_3 d_{i3} + \varepsilon_i$ буде моделлю регресії без обмежень, а модель регресії виду $y_i = a_0 + a_1 x_i + \varepsilon_i$ при $d_{i1} = 0;\ d_{i2} = 0;\ d_{i3} = 0$ буде моделлю регресії з обмеженнями. Базисна модель регресії відповідає регресійній залежності споживання в осінній період залежно від величини доходу.

Для моделі регресії без обмежень можна також побудувати часткові регресії. Наприклад, часткова модель регресії змінної споживання від змінної доходу в зимові місяці буде мати вигляд:

$$y_i = a_0 + \beta_1 + a_1 x_i + \varepsilon_i ,$$

де β_1 – це диференційований коефіцієнт перетинання, який характеризує, наскільки вище споживання в зимові місяці в порівнянні з осінніми за однакового рівня доходу.

Відповідно, часткові моделі регресії обсягу споживання у весняний і літній періоди мають вигляд:

$$y_i = a_0 + \beta_2 + a_1 x_i + \varepsilon_i ,$$
$$y_i = a_0 + \beta_3 + a_1 x_i + \varepsilon_i ,$$

де β_2, β_3 – це диференційовані коефіцієнти перетинання, які характеризують, наскільки змінюється обсяг споживання у весняні та літні місяці в порівнянні з осіннім періодом з однаковим рівнем доходу.

Фіктивні змінні, незважаючи на зовнішню простоту, є досить потужним інструментом дослідження впливу якісних ознак. Зокрема, за допомогою якісних змінних можна досліджувати зміни в рівнях ряду, викликані відмінностями в умовах розвитку економічних процесів у різні періоди часу зі збереженням їх загальних тенденцій [48].

Нехай для періоду (0, T_1) для розвитку процесу була характерна тенденція (1), а в період (T_1+1, T_2) – тенденція (2); динамічні характеристики цих тенденцій збігаються

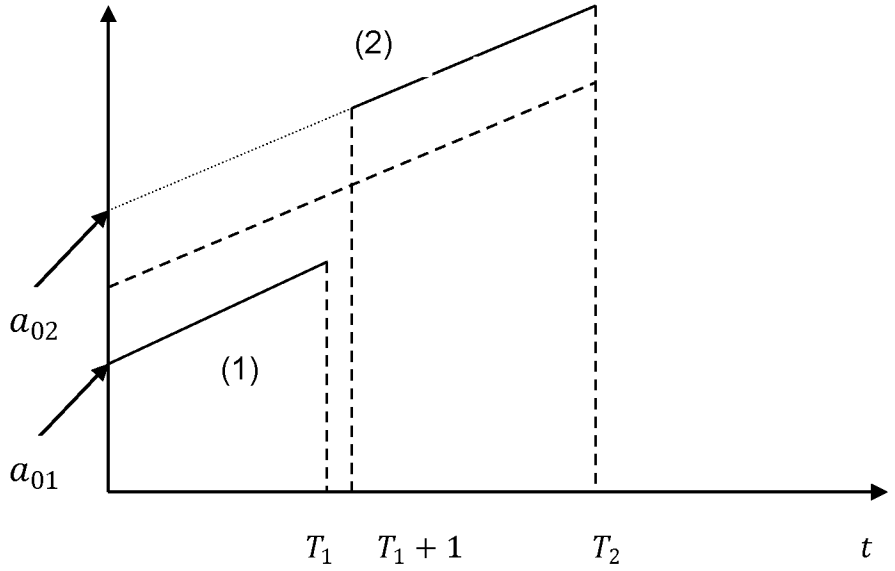

Рис. **Приклад відмінностей в умовах розвитку процесу**

Якщо не брати до уваги зазначені відмінності, спробувавши побудувати єдину, узагальнену модель для періоду (0, T_2), то її рівняння буде відповідати пунктирній лінії, що проходить між суцільними лініями, що характеризують реальні тенденції процесу в розглянутому періоді.

Якщо економетрична модель будується тільки на основі інформації першого періоду, то її рівняння буде мати такий вигляд:

$$y_i = a_{01} + a_1 x_i + \varepsilon_i,$$

а якщо за інформацією другого періоду, то:

$$y_i = a_{02} + a_1 x_i + \varepsilon_i .$$

Ці вираження відрізняються тільки величиною вільного коефіцієнта. Якщо ввести фіктивну змінну з такими властивостями:

$$d_i = \begin{cases} 0, \text{якщо розглядається перший період;} \\ 1, \text{у протилежному випадку,} \end{cases}$$

то вираження і можна об'єднати в рамках однієї моделі виду:

$$y_i = a_{01} + a_1 x_i + \beta d_i + \varepsilon_i .$$

Оцінки невідомих коефіцієнтів моделей регресії зі змінною структурою розраховують за допомогою методу найменших квадратів.

Фіктивні змінні можуть впливати не тільки на зрушення, але й на нахил, тобто регресії для різних категорій можуть відрізнятися за двома параметрами: a_0 – точка перетинання прямої з віссю ординат і a_1 – тангенс кута нахилу прямої до осі абсцис. У розглянутих прикладах вплив якісної ознаки позначався тільки на вільному члені рівняння регресії. За допомогою фіктивних змінних можна врахувати вплив якісної ознаки й на параметри при змінних регресійної моделі. Наприклад, сезон може впливати не тільки на обсяг споживання (6.4), але й на схильність до споживання. Для такого дослідження слід розглянути модель:

$$y_i = a_0 + \beta_1 d_{i1} + \beta_2 d_{i2} + \beta_3 d_{i3} + a_1 x_i + a_2 x_i d_{i1} + a_3 x_i d_{i2} + a_4 x_i d_{i3} + \varepsilon_i ,$$

де a_2, a_3, a_4 – диференційовані коефіцієнти перетинання, які відображають зміну схильності до споживання взимку, навесні, влітку, відповідно, в порівнянні з осіннім періодом.

Використовуючи наведену модель, можна перевіряти гіпотези про вплив сезонних факторів на схильність до споживання.

.2. Дослідження структурних змін за допомогою тесту Чоу

Доцільність застосування різних рівнянь регресії для різних категорій (підмножин, періодів) можна оцінити, не вдаючись до введення фіктивних змінних. Для цього використовують тест Чоу. Розглянемо покроковий алгоритм тесту.

І крок. На першому кроці поєднуються всі спостереження та розглядається загальне рівняння:

$$y_i = a_0 + a_1 x_i + \varepsilon_i.$$

Для отриманого регресійного рівняння визначається сума квадратів залишків: $SS_o = \sum_{i=1}^{T2} (y_i - \widehat{y_i})^2$.

ІІ крок. На другому кроці розглядаються регресійні рівняння для кожного періоду (категорії, підмножини):

$$y_i = a_{01} + a_{11} x_i + \varepsilon_{1i};$$
$$y_i = a_{02} + a_{12} x_i + \varepsilon_{2i}.$$

Для кожного з отриманих регресійних рівнянь (6.7) і (6.8) визначається сума квадратів залишків SS_1, SS_2.

ІІІ крок. На третьому кроці визначається значення *F-критерію:*

$$F = \frac{(SS_0 - (SS_1 + SS_2))}{(SS_1 + SS_2)} \cdot \frac{T_1 + T_2 - 2k}{k + 1},$$

де k – число оцінених параметрів (без вільного члена) у рівняннях, побудованих за підмножинами.

Якщо фактичне значення критерію виявиться більше табличного $F(\alpha; k + 1; T_1 + T_2 - 2k)$, то мають місце структурні зрушення; доцільно будувати рівняння регресії з відповідною фіктивною змінною.

Процедуру тесту Чоу можна спростити, якщо використовувати фіктивні змінні. Об'єднаємо всі спостереження й оцінимо регресію:

$$y_i = a_0 + a_1 x_i + \beta_1 d_i + \beta_2 x_i d_i + \varepsilon_i,$$

де β_1 – диференційований коефіцієнт перетинання, який показує, наскільки точка перетинання для другого періоду відрізняється від першого; β_2 – диференційований коефіцієнт нахилу, який показує, наскільки нахил функції другого періоду відрізняється від першого.

На наступному кроці оцінюється статистична значущість параметрів моделі за допомогою критерію Стьюдента. У результаті можливі такі ситуації:

- диференційовані коефіцієнти перетинання та нахилу статистично значущі – отже, регресії для двох періодів різні:

для першого періоду $y_i = a_0 + a_1 x_i + \varepsilon_i$,

для другого періоду: $y_i = a_0 + \beta_1 + (a_1 + \beta_2) x_i + \varepsilon_i$;

- диференційовані коефіцієнти перетинання та нахилу незначущі, отже, регресії ідентичні:

для першого та другого періодів: $y_i = a_0 + a_1 x_i + \varepsilon_i$;

- диференційований коефіцієнт перетинання значущий, а нахилу – ні. У цьому випадку регресії відрізняються тільки розміщенням:

для першого періоду: $y_i = a_0 + a_1 x_i + \varepsilon_i$,

для другого періоду: $y_i = a_0 + \beta_1 + a_1 x_i + \varepsilon_i$;

- диференційований коефіцієнт перетинання незначущий, а нахилу – так. У цьому випадку регресії мають однакові точки перетинання з віссю ординат, але різні кути нахилу:

для першого періоду: $y_i = a_0 + a_1 x_i + \varepsilon_i$,

для другого періоду: $y_i = a_0 + (a_1 + \beta_2) x_i + \varepsilon_i$.

Приклад У табл. 1 наведені дані про рівень доходу (y_t) і величину залучених коштів (x_t) великих і середніх банків.

Таблиця .1

Вихідні дані

Група I «Великі банки»	x_t	y_t	Група II «Середні банки»	x_t	y_t
1	16 988,84	426,24	11	131,91	27,83
2	13 632,10	342,32	12	354,32	72,36
3	12 345,48	310,16	13	632,04	127,85
4	9 232,78	232,33	14	431,60	87,73
5	6 263,36	158,10	15	67,97	15,00
6	1 209,92	31,76	16	122,71	25,96
7	6 456,91	162,94	17	225,29	46,52
8	4 310,09	109,27	18	308,78	63,23
9	6 962,56	175,58	19	189,06	39,23
10	5 351,18	135,30	20	339,71	69,39

Необхідно за допомогою тесту Чоу перевірити дані на наявність структурного зрушення.

Знайдемо оцінки параметрів моделі для трьох випадків:

1) регресія за всіма даними (1 – 20 спостереження):

$$y_i = 41{,}587 + 0{,}021x_i + \varepsilon_i,\ SS_o = 12990{,}84;$$

2) регресія за даними великих банків (1 – 10 спостереження):

$$y_i = 1{,}514 + 0{,}025x_i + \varepsilon_i,\ SS_1 = 0{,}0001;$$

3) регресія за даними середніх банків (11 – 20 спостереження):

$$y_i = 1{,}429 + 0{,}2x_i + \varepsilon_i,\ SS_1 = 0{,}0073.$$

Як видно з розрахунків, якість часткових регресій досить висока, тоді як регресія за всіма даними низької якості. Розрахуємо *F-тест:*

$$F = \frac{(12990{,}84 - (0{,}0001 + 0{,}0073))}{(0{,}0001 + 0{,}0073)} \cdot \frac{10 + 10 - 2 \cdot 2}{2 + 1} = 9362762.$$

Табличне значення *F-критерію* дорівнює 3,24. Гіпотеза про відсутність структурного зрушення у вибіркових даних відхиляється.

Оцінка моделі, яка включає як фіктивну змінну зрушення, так і фіктивну змінну нахилу, має вигляд:

$$y_i = 1{,}429 + 0{,}2x_i + 0{,}085d_i - 0{,}0175x_i d_i + \varepsilon_i;$$
$$R^2 = 0{,}99, t_{a0} = 104{,}543, t_{a1} = 4731{,}761, t_{\beta 1} = 4{,}325, t_{\beta 2} = -4137{,}885.$$

Порівняння фактичних значень критерію Стьюдента з табличним, яке дорівнює 2,12, дозволяє відкинути нульову гіпотезу про відсутність структурного зрушення, тобто регресії для двох груп банків різні:

- для великих банків ($d_i = 1$):

$$y_i = (1{,}429 + 0{,}085) + (0{,}2 - 0{,}0175)x_i + \varepsilon_i = 1{,}514 + 0{,}025x_i + \varepsilon_i;$$

- для середніх банків ($d_i = 0$):

$$y_i = 1{,}429 + 0{,}2x_i + \varepsilon_i.$$

Наведені регресії збігаються з регресійними рівняннями, отриманими раніше.

.3. Моделі з дискретними залежними змінними

Як було зазначено, на практиці часто виникає необхідність враховувати різного роду обмеження на значення не тільки факторних, але і залежних змінних. *Залежну змінну, яка приймає кілька значень, називають дискретною.*

Деякі показники приймають лише два значення («так» або «ні»), наприклад: перебування в зареєстрованому шлюбі, рішення працювати або не працювати, наявність власного житла, факт купівлі конкретного товару, факт вчасного погашення позики тощо.

Іноді значення є просто кодом однієї з можливих якісних альтернатив. Наприклад, результати голосування: 1 – за, 0 – проти; вибір покупцем каналу руху товару: 1 – купівля в магазині, 2 – на оптовому складі, 3 – замовлення поштою, 4 – замовлення через Інтернет.

Вибір альтернатив може бути ранжованим, наприклад: рівень доходу (1 – високий; 2 – середній, 3 – низький), рівень забрудненості та ін. В окрему категорію виділяють випадки, коли залежна змінна y є кількісною, але приймає тільки цілі додатні значення: 0, 1, 2, ... Наприклад, кількість дітей у сім'ї: $y = 0, 1, 2, \ldots$; кількість співробітників компанії: $y = 1, 2, \ldots$ Для подання значень залежної змінної в цілочисельному вигляді економетрична модель, що зв'язує ці значення з відповідним набором незалежних факторів, має специфічну змістовність. Зазвичай така модель визначає ймовірність здійснення події, яка полягає у тому, що з відомими рівнями незалежних факторів залежна змінна прийме конкретне значення із заданого набору значень $i = 0, 1, 2, \ldots$ Змістовне рівняння такої моделі виглядає таким чином:

$$P(\text{варіант } i) = P(y = i) = F(X_i^T \beta) = F(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}),$$

де X_j – вектор значень незалежних змінних для j-го спостереження;

β – вектор коефіцієнтів моделі:

$$\mathrm{X}_j = \begin{pmatrix} 1 \\ x_{j1} \\ \dots \\ x_{jn} \end{pmatrix}, \ \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_n \end{pmatrix}.$$

Моделі з дискретними залежними змінними можуть бути класифіковані залежно від типу змінних та обраного закону розподілу.

У свою чергу, в межах виділених груп може бути розгорнута більш детальна класифікація – залежно від більш детальних властивостей класифікаційних ознак.

За типом змінних моделі з дискретними залежними змінними розподіляють на моделі вибору серед кінцевої кількості альтернативних варіантів і моделі рахункових даних. Залежно від кількості варіантів, серед яких здійснюється вибір, розрізняють моделі бінарного вибору та моделі множинного вибору. На відміну від моделей множинного вибору, в моделях бінарного вибору результівний показник може приймати тільки два значення: 0 і 1.

До моделей множинного вибору відносять моделі з невпорядкованими та впорядкованими альтернативними варіантами.

4. Лінійна модель бінарного вибору

Моделі бінарного вибору широко використовують в економічних дослідженнях. Покажемо їх специфічні властивості на прикладі моделі трудової активності населення. Для цього необхідно розглянути вихідні передумови.

Індивідуум у певний період часу може працювати або шукати роботу $(y = 1)$ або не робити цього $(y = 0)$. Хоча замість нуля й одиниці можна використовувати інші числа, зазвичай «успіх» кодують одиницею, «неуспіх» – нулем. Припустимо, що стан «працювати» чи «не працювати» визначається набором факторів (вік, сімейний стан, освіта, досвід роботи тощо), а для оцінювання невідомих параметрів $\beta_i, i = \overline{1,n}$ рівняння регресії з генеральної сукупності взята випадкова вибірка обсягом m. Для моделювання ймовірності «успіху»:

$$P(y_i = 1|X_i) = F\left(X_i^T \beta\right)$$

необхідно підібрати функцію F таким чином, щоб її область зміни знаходилась між 0 і 1.

Однією з основних проблем під час побудови моделей бінарного вибору є обґрунтування функції F.

Функція $F(z)$ має відповідати таким вимогам:

а) $F(z)$ монотонно зростає по z;

б) $0 \leq F(z) \leq 1$;

в) $F(z) \to 1$ при $z \to \infty$;

г) $F(z) \to 0$ при $z \to -\infty$.

Імовірність того, що залежна змінна прийме значення, дорівнене нулю:

$$P(y_i = 0|X_i) = 1 - F\left(X_i^T \beta\right).$$

Якщо покласти $F\left(X_i^T \beta\right) = X_i^T \beta$, отримаємо модель лінійної регресії. Зауважимо, що для неї жодна з вимог не буде виконана.

Запишемо лінійну регресійну модель

$$y_i = X_i^T \beta + \varepsilon_i.$$

З огляду, що умовне математичне сподівання помилок дорівнює нулю:

$$M(\varepsilon_i|X_i) = 0,$$

отримаємо:

$$M(y_i|X_i) = 1 \cdot P(y_i = 1|X_i) + 0 \cdot P(y_i = 0|X_i) = P(y_i = 1|X_i) = X_i^T \beta.$$

З випливає, що ймовірність того, що із заданим значенням X_i залежна змінна прийме значення дорівнене одиниці, складає $X_i^T \beta$. Тому такі моделі називають лінійними ймовірнісними моделями.

Лінійна форма моделі представляє певну зручність для розкриття змісту доданків, що входять до неї. Перш за все, зауважимо, що між їх значеннями виконуються наступні співвідношення (табл. .2).

Таблиця .2

Властивості лінійних імовірнісних моделей

y_i	ε_i	P
1	$1 - X_i^T\beta$	$X_i^T\beta$
0	$-X_i^T\beta$	$1 - X_i^T\beta$

Очевидно, що розподіл помилки відрізняється від нормального, а її дисперсія залежить від пояснювальних змінних: $D(\varepsilon_i) = X_i^T\beta \cdot (1 - X_i^T\beta)$, тобто має місце гетероскедастичність. Це не дозволяє використовувати метод найменших квадратів і стандартні формули для перевірки гіпотез про коефіцієнти регресії. Більш істотний недолік полягає в такому: відповідно до формули прогнозні значення $\hat{y} = X_i^T\hat{\beta}$ показують ймовірність «успіху»; проте не гарантується для будь-якої оцінки $\beta \neq 0$, що прогноз не вийде за межі інтервалу [0; 1]. Це не дозволяє інтерпретувати прогнозне значення як ймовірність. Тому в якості функції $P(x)$ використовуються відомі функції розподілу ймовірностей.

.5. Логіт-модель і пробіт-модель

У деяких випадках модель бінарного вибору доцільно обґрунтовувати за допомогою введення прихованої (латентної) змінної. Розглянемо модель прийняття приватним підприємцем рішення: брати кредит на розвиток бізнесу або не брати. Будемо вважати, що його поведінка в бізнесі описується функцією корисності, що залежить від багатьох параметрів, насамперед – доходу, витрат, капіталу, конкурентоздатності, інноваційності, ціноутворення тощо. Підприємець може вибирати: взяти кредит, збільшивши капітал і отримавши можливості до розширення бізнесу, підвищення конкурентоспроможності, запровадження інновацій, тим самим підвищивши витрати на сплату відсотків, або не брати кредит. Кожній із цих ситуацій відповідає своя величина корисності.

Нехай U^1 – це величина корисності, якщо підприємець бере кредит, U^0 – якщо він відмовляється від такої можливості.

Якщо $U^1 > U^0$, то підприємцю вигідніше взяти кредит, оскільки додатковий дохід від упровадження інновацій та підвищення конкурентоспроможності продукції перевищує скорочення прибутку через додаткові витрати. Якщо $U^1 < U^0$, то підприємець відмовляється від кредиту.

Припустимо, що $y_i^* = U^1 - U^0$ – лінійна функція від спостережуваних факторів: доходу підприємця, ставки відсотка та ін. Тому її зручно подати як функцію спостережуваних і неспостережуваних характеристик:

$$y_i^* = X_i^T \beta + \varepsilon_i.$$

Значення змінної y_i^* не спостерігається, тому її називають *латентною*. Латентну змінну в загальному випадку можна інтерпретувати як «здатність», «схильність». Випадкова помилка ε_i включає вплив усіх неврахованих у моделі факторів. Якщо уявити ймовірність того, що підприємець візьме кредит за допомогою функції розподілу помилки, отримаємо:

$$P(y_i = 1) = P(y_i^* \geq 0) = P\left(X_i^T \beta + \varepsilon_i \geq 0\right) = P\left(-\varepsilon_i \leq X_i^T \beta\right) = F\left(X_i^T \beta\right).$$

Таким чином, отримана модель де F – функція розподілу випадкової величини (ε_i), яка у разі симетричного розподілу збігається з функцією розподілу випадкової величини ε_i. У побудованій моделі форма залежності бінарної змінної від факторів визначається законом розподілу випадкової помилки ε_i.

Досить часто в економічних задачах обґрунтування моделі дискретного вибору як різниці значень функції корисності доцільніше, але приховану змінну y_i^* часто вводять безпосередньо:

$$y_i^* = X_i^T \beta + \varepsilon_i,$$
$$y_i = \begin{cases} 1, \text{якщо } y_i^* > 0; \\ 0, \text{якщо } y_i^* \leq 0. \end{cases}$$

Передбачається, що випадкові помилки ε_i незалежні між собою та від x_i.

Використовуючи конкретні закони розподілу випадкової помилки, можна отримати різні варіації моделі бінарного вибору.

Пробіт (probit)-модель заснована на законі розподілу:

$$F(z) = \Phi(Z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt.$$

Пробіт-модель для бінарних даних має вигляд:

$$P(y_i = 1 \left| X_i^T = \int_{-\infty}^{x_i^T \beta} \varphi(t)dt = \Phi\left(X_i^T \beta\right); \right.$$

$$P(y_i = 0 \left| X_i^T = 1 - \Phi\left(X_i^T \beta\right), \right.$$

де y_i – дискретна залежна змінна, X_i^T – вектор незалежних змінних,

$\varphi(t)$ – диференціальна функція нормального закону розподілу;

$\Phi(t)$ – інтегральна функція нормального закону розподілу.

Логіт (logit)-модель ґрунтується на логістичному законі розподілу. Функція розподілу ймовірності логістичного закону:

$$\Lambda(z) = \frac{e^z}{1 + e^z}.$$

Логіт-модель для бінарних даних має вигляд:

$$P(y_i = 1 \left| X_i^T = \frac{e^{X_i^T \beta}}{1 + e^{X_i^T \beta}} = \Lambda\left(X_i^T \beta\right); \right.$$

$$P(y_i = 0 \left| X_i^T = 1 - \Lambda\left(X_i^T \beta\right). \right.$$

Питання про те, який із розподілів більш доречний для практичних досліджень, залишається відкритим. На ділянці $z \in$ [-1,2; 1,2] обидва вони поводяться практично однаково. Однак поза цією ділянкою на хвостах розподілу – значення функціоналів $\Phi\left(X_i^T \beta\right)$ і $\Lambda\left(X_i^T \beta\right)$ мають деякі

відмінності. Практика показує, що за умови відсутності суттєвого домінування однієї альтернативи над іншою, а також для вибірок з невеликим розкидом змінних висновки, отримані на основі пробіт- і логіт-моделей, як правило, збігаються.

Оцінки параметрів пробіт- і логіт-моделей найчастіше знаходять методом максимальної правдоподібності. З огляду на те, що залежна змінна приймає значення 0 або 1, логарифм функції правдоподібності для вибірки дорівнює:

$$\log L = \sum_{i=1}^{n} y_i \log F\left(X_i^T \beta\right) + \sum_{i=1}^{n} (1 - y_i) \log F\left(1 - F(X_i^T \beta)\right).$$

Для знаходження максимуму необхідно взяти перші похідні за всіма компонентами вектора β і прирівняти їх до нуля. У результаті для пробіт-моделі маємо:

$$\frac{\partial \log L(y_i)}{\partial \beta} = \sum_{i=1}^{n} \left| \frac{y_i - \Phi\left(X_i^T \beta\right)}{\Phi(X_i^T \beta)(1 - \Phi(X_i^T \beta)} \varphi(\left(X_i^T \beta\right) \right| X_i = 0.$$

З огляду на властивість логістичних функцій, для логіт-моделей відповідний вираз можна спростити:

$$\frac{\partial \log L(y_i)}{\partial \beta} = \sum_{i=1}^{n} \left| y_i - \frac{e^{X_i^T \beta}}{1 + e^{X_i^T \beta}} \right| X_i = \sum_{i=1}^{n} (y_i - \Lambda\left(X_i^T \beta\right) = 0.$$

Розв'язання систем рівнянь) і (знаходять чисельними методами, оцінка $\hat{\beta}$ є оцінкою максимальної правдоподібності.

Цікава також оцінка ймовірності певних станів, наприклад $y_i = 1$. Для пробіт-моделі вона виглядатиме таким чином:

$$\hat{p}_i = \Phi\left(X_i^T \hat{\beta}\right);$$

$$\hat{p}_i = \Lambda\left(X_i^T \hat{\beta}\right) = \frac{e^{X_i^T \hat{\beta}}}{1 + e^{X_i^T \hat{\beta}}}.$$

Приклад Є вибірка, що складається з 950-ти спостережень, в якій $y = 1$, якщо заробітна платня працівника нижче 5 дол. на годину, тобто працівник низькооплачуваний, ($y = 0$ – в іншому випадку). Передбачається, що рівень заробітної платні залежить від певних факторів: x_1 – освіта, років; x_2 – стать (1 – жіноча, 0 – чоловіча); x_3 – досвід роботи, років. У табл. 6.3 наведені коефіцієнти, отримані у ході оцінювання логіт-моделі за допомогою нелінійного МНК.

Таблиця .3

Коефіцієнти логіт-моделі

1	x_1	x_2	x_3
5,87	–0,56	1,26	–0,06

Потрібно визначити на основі логіт-моделі оцінку ймовірності для чоловіка та для жінки, які мають 12 років освіти і 15 років досвіду роботи, стати низькооплачуваними працівниками.

Спочатку відповідно до логіт-моделі розрахуємо значення функцій $\hat{z} = X_i^T \hat{\beta}$ для чоловіка і для жінки, відповідно:

$$\hat{z}_1 = 5{,}87 - 0{,}56 \cdot 12 + 1{,}26 \cdot 0 - 0{,}06 \cdot 15 = -1{,}75;$$
$$\hat{z}_2 = 5{,}87 - 0{,}56 \cdot 12 + 1{,}26 \cdot 1 - 0{,}06 \cdot 15 = -0{,}49.$$

Відповідно до (6.19) за логіт-моделлю ймовірності бути низькооплачуваним працівником складуть для чоловіка та жінки:

$$P(y_i = 1|x_1) = \frac{e^{-1{,}75}}{1 + e^{-1{,}75}} = 0{,}1480;$$
$$P(y_i = 1|x_2) = \frac{e^{-0{,}49}}{1 + e^{-0{,}49}} = 0{,}3799.$$

Таким чином, ймовірність для чоловіка та для жінки, з дванадцятьма роками освіти та п'ятнадцятьма – досвіду роботи потрапити в категорію низькооплачуваних працівників становить, відповідно, 0,148 і 0,3799.

У моделях коефіцієнти є маржинальним ефектом відповідної незалежної змінної. Цей ефект залежний від факторів:

$$\frac{\partial \Phi\left(X_i^T \beta\right)}{\partial X_{ik}} = \varphi((X_i^T \beta)\beta_k;$$

$$\frac{\partial \Lambda\left(X_i^T \beta\right)}{\partial X_{ik}} = \frac{e^{X_i^T \beta}}{(1+e^{X_i^T \beta})^2}\beta_k = \Lambda\left(X_i^T \beta\right)\left(1 - \Lambda\left(X_i^T \beta\right)\right)\beta_k.$$

У зв'язку з тим, що функції розподілу монотонно зростають, знаки коефіцієнтів інтерпретуються відповідним чином.

Приклад 3. Використовуючи дані приклада 2, на основі логіт-моделі визначити зміну оцінки ймовірності бути низькооплачуваним працівником для чоловіка, якщо він вчився на один рік більше.

Згідно з , за логіт-моделлю маржинальний ефект визначимо як:

$$P(y_i = 1|x_1) \cdot (1 - P(y_i = 1|x_1)\,\beta_1 = 0.148 \cdot (1 - 0.148) \cdot (-0.56) = -0.0706.$$

Тобто якщо чоловік з характеристиками з прикладу 6.2 навчатиметься на один рік більше, то оцінка ймовірності потрапити в категорію низькооплачуваних працівників зменшиться для нього на 7 %.

Приклад 4. Є вибірка, що складається з 150-ти спостережень за транснаціональною компанію Invest Inc., в якій $y = 1$, якщо працівник є менеджером середньої ланки, а $y = 0$ – в іншому випадку. Передбачається, що посада залежить від таких факторів: x_1 – освіта, років; x_2 – стать (1 – жіноча, 0 – чоловіча); x_3 – досвід роботи в попередніх компаніях, років; x_4 – досвід роботи в *Invest Inc.*, років.

На основі вибіркових даних була отримана така пробіт-модель:

$$\hat{z} = -0{,}9 + 0{,}015x_1 - 0{,}599x_2 - 0{,}003x_3 + 0{,}29x_4.$$

Визначити, наскільки підвищується ймовірність отримати роботу менеджера середньої ланки для співробітниці, яка має 5 років освіти, 6 років досвіду роботи в іншій компанії, 4 роки досвіду роботи в *Invest Inc.*, якщо вона пропрацює в даній компанії ще один рік.

Розрахуємо спочатку значення функції $\hat{z}$:

$$\hat{z} = -0{,}9 + 0{,}015 \cdot 5 - 0{,}599 \cdot 1 - 0{,}003 \cdot 6 + 0{,}29 \cdot 4 = -0{,}2658.$$

Граничний ефект фактора роботи в компанії відповідно до (6.20) розраховується таким чином:

$$\frac{\partial \Phi(y=1)}{\partial X_1} = \varphi(\hat{z})\hat{\beta}_1,$$

де $\varphi(\hat{z})$ – функція щільності нормального розподілу.

У даному випадку $P(-0{,}2658) = 0{,}604803$, тоді маржинальний ефект складе $0{,}604803 \cdot 0{,}29 = 0{,}077082$. Звідки робимо висновок, що, якщо співробітниця пропрацює в компанії *Invest Inc.* ще рік, то оцінка ймовірності отримати посаду менеджера середньої ланки збільшиться для неї на 8 %.

Для моделей бінарного вибору важко запропонувати природну міру якості, як, наприклад, R^2 для лінійної моделі. У цьому випадку доцільно будувати такі міри шляхом прямого або непрямого порівняння поточної моделі з тривіальною (що містить лише вільний член). Нехай $\ln L$, $\ln L_0$ – логарифм правдоподібності, відповідно, поточної та тривіальної моделей. Чим більша між ними різниця, тим краща повна модель в порівнянні з тривіальною. Таким чином, для оцінювання ступеня якості моделі використовують кілька так званих «псевдо-R^2»:

Псевдо-R^2 за Амемійя:

$$R^2 = 1 - \frac{1}{1 + \frac{2(\ln L - \ln L_0)}{n}}.$$

Псевдо-R^2 *за* Макфадденом (індекс відношення правдоподібності, *likelihood ratio index*):

$$R^2 = 1 - \frac{\ln L}{\ln L_0}.$$

Псевдо-R^2, засноване на коректних прогнозах. Визначимо прогнози таким чином: якщо передбачена за моделлю ймовірність більша 0,5, то прогнозне значення дорівнює 1, в іншому випадку – 0. Це, з огляду на властивості розподілів, відповідає такій системі:

$$\hat{y}_i = 1, \text{якщо } X_i^T \beta > 0;$$
$$\hat{y}_i = 0, \text{якщо } X_i^T \beta < 0.$$

Частка некоректних прогнозів становить:

$$wv_1 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2.$$

Для тривіальної моделі всі передбачені ймовірності будуть рівні між собою та дорівнюватимуть частці успіхів у вибірці $p = \frac{n_1}{n_2}$. Прогноз для всіх спостережень буде однаковий (всі нулі або одиниці, якщо p менше /більше 0,5). Число помилок прогнозу дорівнюватиме:

$$wv_0 = 1 - \hat{p}, \text{ якщо } \hat{p} > 0{,}5, \text{ і } wv_0 = \hat{p}, \text{ якщо } \hat{p} \leq 0{,}5.$$

Псевдо-R^2 в цьому випадку:

$$R^2 = 1 - \frac{wv_1}{wv_0}.$$

Приклад 5. Маємо дані, які складаються з шести спостережень (табл 4).

Таблиця .4

Вихідні дані

y	0	0	0	1	1	1
x	-1	-2	0	1	1	1

Необхідно:

оцінити лінійну модель ймовірності за допомогою МНК. Розрахувати R^2;

розрахувати псевдо-R^2, засноване на коректних прогнозах;

розрахувати кількість випадків правильного віднесення до відповідної групи, застосовуючи правило класифікації: група I $(y = 1)$, якщо $\hat{y} > 0{,}5$; група II $(y = 0)$, якщо $\hat{y} \leq 0{,}5$;

зіставити частку правильного потрапляння та коефіцієнт детермінації.

Рівняння лінійної моделі ймовірності, оцінене за допомогою МНК, виглядає таким чином: $\widehat{y_i} = 0{,}5 + 0{,}375x_i$. Розраховане значення $R^2 = 0{,}75$.

Підставляючи значення незалежної змінної в рівняння регресії, отримаємо, що оцінки ймовірності події $y = 1$ дорівнюють, відповідно: 0,125; -0,25; 0,5; 0,875; 0,875; 0,875.

Користуючись відповідним правилом, зазначимо, що в групу I увійдуть спостереження з четвертого по шосте; в групу II – перші три.

Усі спостереження правильно віднесені до відповідної групи, тому частка некоректних прогнозів становить:

$$wv_1 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 0.$$

Псевдо-R^2 у цьому випадку дорівнює 1, тобто є помітно вищим, ніж коефіцієнт детермінації.

6. Моделі множинного вибору

У багатьох економічних завданнях спостерігається множинність альтернатив. Наприклад: вибір менеджером конкретного постачальника, рішення працювати з повною зайнятістю, неповною або зовсім не працювати. Моделі множинного вибору дають можливість окреслити вірогідність кожної з альтернатив як функцію спостережуваних характеристик об'єкта. Для цього ймовірності повинні знаходитись в інтервалі від 0 до 1, а сума ймовірностей за всіма альтернативам повинна дорівнювати 1.

Припущення про існування латентної змінної допускає розширення допущення на моделі з декількома альтернативами. Припустимо, що існує m взаємовиключних і вичерпних альтернатив, $i = 1,2,\ldots,m$. Позначимо y – вектор розмірності m, $y = 1$, якщо спостерігається альтернатива i, і $y = 0$ – в іншому випадку.

Нехай y^* – багатовимірна латентна змінна:

$$y_i^* = X_i^T\beta + \varepsilon_i,$$

що приймає значення в області R^h; ε – вектор випадкових помилок; $F\left(\varepsilon\middle|X_i^T\right)$ – функція розподілу. Область значень R^h розбита на m підобла-

стей $A_1, A_2, \ldots, A_m$. Спостережуване значення i-ї компоненти вектора y дорівнює одиниці, якщо y_i^* потрапляє в область A_i, в іншому випадку – 0.

Відповідна ймовірність дорівнює:

$$p_i = P(y_i = 1|X, \beta) = P(X^T\beta + \varepsilon \in A_i = F(-\varepsilon|X^T\beta + \varepsilon \in A_i).$$

Конкретизуючи значення розмірностей h і m, правила формування областей A_i та вид функції розподілу F, можна отримати різні варіанти загальної моделі. Розглянемо основні з них.

1. **Множинний вибір (невпорядкований)**. Множинний вибір характеризує вибір індивідом з різних альтернатив, кожна з яких характеризується своїми властивостями. Наприклад, вибір товару серед різних виробників, марок, сортів; вибір постачальника; вибір сфери бізнесу для вкладення грошей тощо.

Припустимо, число розглянутих альтернатив m, у такому випадку $h = m$, а y^* – міра корисності, яку отримає індивідуум, зупинивши свій вибір на i-й альтернативі. Визначимо A_i як множину y^*, для яких i-та альтернатива краще інших:

$$A_i = y_i^* \left| y_i^* \geq y_j^*, \quad j = 1,2, \ldots, m. \right.$$

Відповідно, чинники, що визначають вибір, описуються змінними, що входять у вектор X. Ймовірність p_i можна інтерпретувати як частку індивідуумів, з характеристиками X, для яких максимум корисності досягається на множині A_i.

Залежно від конкретного виду функції $F\left(\varepsilon\middle|X_i^T\right)$ розрізняють варіанти формулювання моделей множинного вибору, серед яких найбільш популярними є множинні логіт- і пробіт-моделі, гніздові множинні логіт-моделі.

2. **Упорядкований вибір**. Варіанти в моделях множинного впорядкованого вибору є логічно та природно впорядкованими. Прикладами впорядкованих варіантів є: рейтинги цінних паперів, опитування громадської думки, рівні складності роботи тощо.

Зі збереженням введених позначень наведемо приклад для вибору товару, упорядкованого з точки зору фасування. У такому випадку $h = 1$ і $y_i^* = X_i^T\beta + \varepsilon_i$ – «ідеальний» попит за умови, що товар безмежно ділимий. Однак товар доступний лише в дискретному фасуванні λ_i, $i = 1,2, \ldots, m$.

Позначимо $U(y^* - \lambda_i)$ – корисність фасування λ_i за умови, що оптимальна кількість y^*. Нехай $U(y^* - \lambda_i) = U(y^* - \lambda_{i-1})$ і $A_i = [a_i; a_{i+1})$. Тоді ймовірність:

$$P_i = P(X^T\beta + \varepsilon \in A_i = F(a_i - X^T\beta|X) - F(a_{i-1} - X^T\beta|X).$$

Дану ймовірність можна інтерпретувати як частку покупців з характеристиками X, для яких оптимальним буде фасування λ_i.

Найбільш часто в якості закону розподілу ймовірностей обирають логістичний і нормальний закони. Відповідні моделі отримали назву *порядкові логіт-* і *пробіт-моделі*.

Приклад 6. Для поїздки на змагання спортсмен обирає тип страхового поліса з таких варіантів: 1 – відсутність такого, 2 – часткове покриття, 3 – повне покриття. Ці варіанти природно впорядковані за рівнем його безпеки. Латентна змінна $y_i^* = X_i^T\beta + \varepsilon_i$ відображає індивідуальні переваги спортсменів залежно від їх доходу, самопочуття, шансів на успіх, віку та ін. Нехай в якості закону розподілу використовується порядкова пробіт-модель. Будемо вважати, що існують такі числа a_1 і a_2, що:

$$y = 1, \text{ якщо } y^* \le a_1;$$
$$y = 2, \text{ якщо } a_1 < y^* \le a_2;$$
$$y = 3, \text{ якщо } y^* > a_2.$$

У даному випадку $a_1 = 0{,}35$, $a_2 = 0{,}7$.

Необхідно знайти ймовірності вибору спортсменом усіх видів страхового поліса, якщо для даного спортсмена за оцінками максимальної правдоподібності: $\hat{z} = X^T\beta = 0{,}1578$.

Згідно з (6.26), відповідні ймовірності такі:

$$P(y_i = 1) = \Phi(a_1 - X_i^T\beta) = \Phi(0{,}35 - 0{,}1578) = 0{,}576481;$$

$$P(y_i = 2) = \Phi(a_2 - X_i^T\beta) - \Phi(a_1 - X_i^T\beta) = \Phi(0{,}7 - 0{,}1578) - \Phi(0{,}35 - 0{,}1578) = \\ = 0{,}70616 - 0{,}576481 = 0{,}12968;$$

$$P(y_i = 3) = 1 - 0{,}576481 - 0{,}12968 = 0{,}293839.$$

Таким чином, найбільшу ймовірність має відмова спортсмена від укладання договору страхування.

Для деяких змінних (наприклад: кількість заявок на лінійних касах в супермаркеті, кількість відвідувань лікаря тощо), як закон розподілу помилки використовується закон Пуассона або від'ємний біноміальний закон. Ці моделі отримали назву *моделі рахункових даних*.

7. Тобіт-модель

Уперше тобіт-модель (*tobit model*) була розглянута Дж. Тобіном у 1958 р., але свою назву вона отримала завдяки А. Голбергу, який підкреслив її схожість із моделями *пробіт*.

У багатьох економічних завданнях залежна змінна може приймати значення з деякого обмеженого зверху чи знизу діапазону. Найчастіше зустрічаються випадки, коли для частини спостережень залежна змінна дорівнює нулю, а для інших приймає якесь додатне значення. Наприклад, витрати на товари тривалого користування, робочі години, іноземні інвестиції фірми та ін. Тобто змінні є сумішшю дискретних і безперервних величин [36].

Якщо вибірка зроблена не з усієї генеральної сукупності, а тільки з тих значень, які задовільнюють певним обмеженням, вона буде усіченою (урізаною). Надалі були запропоновані деякі узагальнення даної моделі. Однак слід розглянути стандартну тобіт-модель або *модель цензурованої регресії* (*censored regression*) для випадку $y_{min} = 0,\ y_{max} = \infty$:

$$y_i^* = X_i^T \beta + \varepsilon_i, \qquad i = 1,2,\dots,m,$$
$$y_i = \begin{cases} y_i^*, & y_i^* > 0 \\ 0, & y_i^* \le 0 \end{cases}.$$

У такому випадку $y_i^* = X_i^T \beta + \varepsilon_i$, інтерпретується як бажана кількість.

Розглянемо математичне очікування спостережуваної залежної змінної на прикладі тобіт-моделі з нормально розподіленою помилкою:

$$E(y) = P(y^* \le 0) \cdot E(y|y^* \le 0) + P(y^* > 0) \cdot E(y|y^* > 0) =$$
$$= P(y^* \le 0) \cdot 0 + P\left(\varepsilon > -\frac{X^T\beta}{\sigma}\right) \cdot \left(X^T\beta + \sigma E\left(\varepsilon \middle| \varepsilon > -\frac{X^T\beta}{\sigma}\right)\right).$$

Якщо f – щільність, а F – інтегральна функція розподілу випадкової помилки, то:

$$P\left(\varepsilon > -\frac{X^T\beta}{\sigma}\right) = \Phi\left(\frac{X^T\beta}{\sigma}\right),$$
$$E\left(\varepsilon \middle| \varepsilon > -\frac{X^T\beta}{\sigma}\right) = \frac{\phi\left(-\frac{X^T\beta}{\sigma}\right)}{\Phi\left(\frac{X^T\beta}{\sigma}\right)}.$$

Отже, остаточно маємо:

$$E(y) = \Phi\left(\frac{X^T\beta}{\sigma}\right) \cdot X^T\beta + \sigma \cdot \phi\left(\frac{X^T\beta}{\sigma}\right).$$

Очевидно, цей вираз не дорівнює $X^T\beta$, отже побудова звичайної регресії призведе до зміщених і необґрунтованих оцінок.

Оцінювання параметрів тобіт-моделей також здійснюють на основі методу максимальної правдоподібності.

Модель Тобіна має певний недолік. Справа в тому, що значення $y = 0$ може означати вибір «не брати участь» (наприклад, у витратах на відпочинок), а значення $y > 0$ можна інтерпретувати як «інтенсивність участі». У тобіт-моделі вибір «брати участь-не брати участь» та «інтенсивність участі» визначаються тими ж факторами, які діють в одному напрямку. Класичний приклад фактора і ситуації неоднозначного впливу – кількість дітей як фактор, що впливає на витрати сім'ї. Зрозуміло, що велика кількість дітей може негативно впливати на рішення «відпочивати чи ні» (через великі витрати). Однак, якщо прийняте таке рішення, то величина витрат («інтенсивність участі») на відпочинок прямо залежить від кількості дітей.

Дж. Хекман запропонував розділити модель на дві складові: модель бінарного вибору для участі, і лінійну модель для інтенсивності участі. Фактори цих двох моделей можуть бути різними. Таким чином, у моделі Хекмана є дві латентні змінні, що відповідають таким моделям:

$$y^* = X^T\beta + \varepsilon,$$
$$g^* = Z^T\gamma + u.$$

Випадкові помилки моделей передбачаються нормально розподіленими. Друга латентна змінна визначає вибір «брати участь/не брати участь» у рамках стандартної моделі бінарного вибору (наприклад, пробіт-моделі). Перша модель – це модель інтенсивності участі за умови вибору «брати участь». Якщо вибрати «не брати участь», то y не спостерігається (дорівнює нулю).

$$g = \begin{cases} 1, & g^* > 0 \\ 0, & g^* \le 0 \end{cases},$$
$$y = \begin{cases} y^*, & g = 1 \\ 0, & g = 0 \end{cases}.$$

Таку модель називають тобіт II (відповідно, вихідну тобіт-модель називають тобіт I), іноді за аналогією – хекіт (модель Хекмана). В англомовній літературі також зустрічається назва *sample selection model*.

Розглянемо математичне очікування спостережуваної залежної змінної (за умови $g = 1$):

$$E(y|g = 1) = X^T\beta + E(\varepsilon|g = 1) = X^T\beta + E(\varepsilon|u > -Z^T\beta).$$

Припускаючи, що випадкові помилки моделей латентних змінних корельовані та пов'язані співвідношенням:

$$\varepsilon = \sigma_{\varepsilon u}u + v,$$

то:

$$E(y|g = 1) = X^T\beta + \sigma_{\varepsilon u}E(u|u > -Z^T\beta) =$$
$$= X^T\beta + \sigma_{\varepsilon u}\frac{\phi(Z^T\beta)}{\Phi(Z^T\beta)} = X^T\beta + \sigma_{\varepsilon u}\lambda(Z^T\beta),$$

де $\lambda(Z^T\beta)$ – так звана лямбда Хекмана.

Оцінювання моделі Хекмана проводиться також методом максимальної правдоподібності, проте у зв'язку з нестандартністю даного завдання часто застосовують спрощену двокрокову процедуру оцінювання, запропоновану Джеймсом Хекманом. На першому кроці оцінюється модель бінарного вибору та визначаються параметри цієї моделі. На основі

цих параметрів можна визначити для кожного спостереження лямбду Хекмана. На другому кроці звичайним МНК оцінюється регресія:

$$y_i = X_i^T \beta + \sigma_{\varepsilon u} \lambda_i + \eta_i.$$

Отримані оцінки є неефективними, але можуть бути використані в якості початкових значень у методі максимальної правдоподібності.