Эконометрика. Лекция 12 Модели бинарного выбора (2)

Д. С. Терещенко

НИУ ВШЭ, Санкт-Петербург

9 декабря 2022 г.

Основные источники

- SW, глава 11 (пункт 11.3 и приложение 11.2)
- HE, chapter 25

Содержание

1 Еще раз об интерпретациях и предельных эффектах

2 Оценивание параметров в моделях бинарного выбора

Содержание

1 Еще раз об интерпретациях и предельных эффектах

2 Оценивание параметров в моделях бинарного выбора

Модели бинарного выбора: повторение

• Мы хотим получить модель в привычном для нас виде:

$$Y = \mathbb{E}[Y|X_1, \dots, X_k] + u$$

$$\mathbb{E}[u|X_1, \dots, X_k] = 0, \dots$$

• Можно показать, что:

$$\mathbb{E}(Y) = 0 \times \Pr(Y = 0) + 1 \times \Pr(Y = 1) = \Pr(Y = 1)$$

$$\mathbb{E}(Y|X_1, \dots, X_k) = \Pr(Y = 1|X_1, \dots, X_k) \equiv P(\mathbf{x})$$

 Получается, что регрессия с бинарными зависимыми переменными моделирование вероятности того, что зависимая переменная равна 1

$$Y = P(\mathbf{x}) + u$$

Модели бинарного выбора: повторение

• Линейная вероятностная модель (ММЛР с бинарной зависимой переменной):

$$Y_{i} = \beta_{0} + \beta_{1}X_{1i} + \beta_{2}X_{2i} + \dots + \beta_{k}X_{ki} + u_{i}$$

$$Pr(Y = 1|X_{1}, \dots, X_{k}) = \beta_{0} + \beta_{1}X_{1} + \beta_{2}X_{2} + \dots + \beta_{k}X_{k}$$

 Пробит-модель — стандартная нормальная функция распределения вероятностей:

$$\Pr(Y = 1 | X_1, X_2, \dots, X_k) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)$$

• Логит-модель — логистическая функция распределения:

$$\Pr(Y = 1 | X_1, X_2, \dots, X_k) = \Lambda(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_{ki})$$

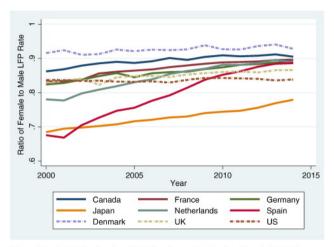


График из статьи: Cascio, E. U., Haider, S. J., & Nielsen, H. S. (2015). The effectiveness of policies that promote labor force participation of women with children: A collection of national studies. *Labour Economics*, 36, 64-71.

Notes: Authors' tabulations based on OECD labor force participation data (downloaded 9 July 2015).

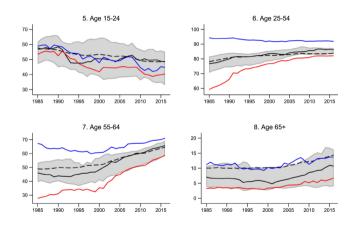


График из статьи: Grigoli, F., Koczan, Z., & Topalova, P. (2020). Automation and labor force participation in advanced economies: Macro and micro evidence. European Economic Review. 126, 103443.

Notes: Solid lines denote medians, dotted lines denote population-weighted averages, and shaded areas denote interquartile ranges. In panels 5 to 8, solid lines in blue and red denote medians for men and women, respectively.

- Пусть LFP_i равно 1, если индивид работает, и 0 в противном случае, а f_i равно 1 для женщин и 0 в противном случае.
- Рассмотрим сначала линейную вероятностную модель для парного случая:

$$LFP_i = \beta_0 + \beta_1 f_i + u_i$$

$$\mathbb{E}[LFP_i|f_i = f] = \Pr[LFP_i = 1|f_i = f] = \beta_0 + \beta_1 f$$

$$GAP_{LFP} = \mathbb{E}[LFP_i|f_i = 1] - \mathbb{E}[LFP_i|f_i = 0] =$$

$$= \Pr[LFP_i = 1|f_i = 1] - \Pr[LFP_i = 1|f_i = 0] =$$

$$= (\beta_0 + \beta_1) - \beta_0 =$$

$$= \beta_1$$

• Добавим контрольную переменную s_i — количество лет образования:

$$LFP_i = \beta_0 + \beta_1 f_i + \beta_2 s_i + u_i$$

 $\mathbb{E}[LFP_i|f_i = f, s_i = s] = \Pr[LFP_i = 1|f_i = f, s_i = s] = \beta_0 + \beta_1 f + \beta_2 s$

$$GAP_{LFP} = \mathbb{E}[LFP_{i}|f_{i} = 1, s_{i} = s] - \mathbb{E}[LFP_{i}|f_{i} = 0, s_{i} = s] =$$

$$= \Pr[LFP_{i} = 1|f_{i} = 1, s_{i} = s] - \Pr[LFP_{i} = 1|f_{i} = 0, s_{i} = s] =$$

$$= (\beta_{0} + \beta_{1} + \beta_{2}s) - (\beta_{0} + \beta_{2}s) =$$

$$= \beta_{1}$$

• Добавим кросс-эффект:

$$LFP_{i} = \beta_{0} + \beta_{1}f_{i} + \beta_{2}s_{i} + \beta_{3}f_{i}s_{i} + u_{i}$$

$$\mathbb{E}[LFP_{i}|f_{i} = f, s_{i} = s] = \Pr[LFP_{i} = 1|f_{i} = f, s_{i} = s] = \beta_{0} + \beta_{1}f + \beta_{2}s + \beta_{3}fs$$

$$GAP_{LFP} = \mathbb{E}[LFP_i|f_i = 1, s_i = s] - \mathbb{E}[LFP_i|f_i = 0, s_i = s] =$$

$$= \Pr[LFP_i = 1|f_i = 1, s_i = s] - \Pr[LFP_i = 1|f_i = 0, s_i = s] =$$

$$= (\beta_0 + \beta_1 + \beta_2 s + \beta_3 s) - (\beta_0 + \beta_2 s) =$$

$$= \beta_1 + \beta_3 s$$

• Рассмотрим пробит-модель для парного случая:

$$LFP_i = \Phi(\beta_0 + \beta_1 f_i) + u_i$$

$$\mathbb{E}[LFP_i|f_i = f] = \Pr[LFP_i = 1|f_i = f] = \Phi(\beta_0 + \beta_1 f)$$

$$GAP_{LFP} = \mathbb{E}[LFP_i|f_i = 1] - \mathbb{E}[LFP_i|f_i = 0] =$$
 $= \Pr[LFP_i = 1|f_i = 1] - \Pr[LFP_i = 1|f_i = 0] =$
 $= \Phi(\beta_0 + \beta_1) - \Phi(\beta_0) =$
 $\neq \beta_1$

• Рассмотрим пробит-модель с контрольной переменной без кросс-эффектов:

$$LFP_{i} = \Phi(\beta_{0} + \beta_{1}f_{i} + \beta_{2}s_{i}) + u_{i}$$

$$\mathbb{E}[LFP_{i}|f_{i} = f, s_{i} = s] = \Pr[LFP_{i} = 1|f_{i} = f, s_{i} = s] = \Phi(\beta_{0} + \beta_{1}f + \beta_{2}s)$$

$$GAP_{LFP} = \mathbb{E}[LFP_{i}|f_{i} = 1, s_{i} = s] - \mathbb{E}[LFP_{i}|f_{i} = 0, s_{i} = s] =$$

$$= \Pr[LFP_{i} = 1|f_{i} = 1, s_{i} = s] - \Pr[LFP_{i} = 1|f_{i} = 0, s_{i} = s] =$$

$$= \Phi(\beta_{0} + \beta_{1} + \beta_{2}s) - \Phi(\beta_{0} + \beta_{2}s) =$$

$$\neq \beta_{1}$$

• Рассмотрим пробит-модель с контрольной переменной и кросс-эффектом:

$$LFP_{i} = \Phi(\beta_{0} + \beta_{1}f_{i} + \beta_{2}s_{i} + \beta_{3}f_{i}s_{i}) + u_{i}$$

$$\mathbb{E}[LFP_{i}|f_{i} = f, s_{i} = s] = \Pr[LFP_{i} = 1|f_{i} = f, s_{i} = s] = \Phi(\beta_{0} + \beta_{1}f + \beta_{2}s + \beta_{3}fs)$$

$$GAP_{LFP} = \mathbb{E}[LFP_{i}|f_{i} = 1, s_{i} = s] - \mathbb{E}[LFP_{i}|f_{i} = 0, s_{i} = s] =$$

$$= \Pr[LFP_{i} = 1|f_{i} = 1, s_{i} = s] - \Pr[LFP_{i} = 1|f_{i} = 0, s_{i} = s] =$$

$$= \Phi(\beta_{0} + \beta_{1} + \beta_{2}s + \beta_{3}s) - \Phi(\beta_{0} + \beta_{2}s)$$

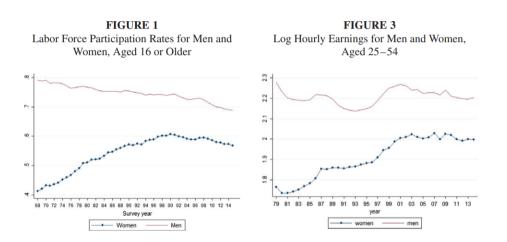


График из статьи: Huang, C. (2018). Why are Us Women decreasing their labor force participation if their wages are rising?. Economic Inquiry, 56(4), 2010-2026.

- Пусть w_i доход индивида.
- Рассмотрим пробит-модель для парного случая:

$$LFP_i = \Phi(\beta_0 + \beta_1 w_i) + u_i$$

$$\mathbb{E}[LFP_i|w_i = w] = \Pr[LFP_i = 1|w_i = w] = \Phi(\beta_0 + \beta_1 w)$$

• Тогда предельный эффект дохода на участие в рабочей силе можно выразить:

$$ME_{w} = \frac{\partial}{\partial w} \mathbb{E}[LFP_{i}|w_{i} = w] =$$

$$= \frac{\partial}{\partial w} \Pr[LFP_{i} = 1|w_{i} = w] =$$

$$= \beta_{1}\phi(\beta_{0} + \beta_{1}w)$$

• Рассмотрим пробит-модель с контрольной переменной:

$$LFP_{i} = \Phi(\beta_{0} + \beta_{1}w_{i} + \beta_{2}s_{i}) + u_{i}$$

$$\mathbb{E}[LFP_{i}|w_{i} = w, s_{i} = s] = \Pr[LFP_{i} = 1|w_{i} = w, s_{i} = s] = \Phi(\beta_{0} + \beta_{1}w + \beta_{2}s)$$

• Тогда предельный эффект дохода на участие в рабочей силе можно выразить:

$$ME_{w} = \frac{\partial}{\partial w} \mathbb{E}[LFP_{i}|w_{i} = w, s_{i} = s] =$$

$$= \frac{\partial}{\partial w} \Pr[LFP_{i} = 1|w_{i} = w, s_{i} = s] =$$

$$= \beta_{1}\phi(\beta_{0} + \beta_{1}w + \beta_{2}s)$$

• Рассмотрим пробит-модель с логарифмом дохода и контрольной переменной:

$$LFP_{i} = \Phi(\beta_{0} + \beta_{1} \ln(w_{i}) + \beta_{2}s_{i}) + u_{i}$$

$$\mathbb{E}[LFP_{i}|w_{i} = w, s_{i} = s] = \Pr[LFP_{i} = 1|w_{i} = w, s_{i} = s] = \Phi(\beta_{0} + \beta_{1} \ln(w) + \beta_{2}s)$$

• Тогда предельный эффект дохода на участие в рабочей силе можно выразить:

$$ME_w = \frac{\partial}{\partial w} \mathbb{E}[LFP_i|w_i = w, s_i = s] =$$

$$= \frac{\partial}{\partial w} \Pr[LFP_i = 1|w_i = w, s_i = s] =$$

$$= \frac{\beta_1}{w} \phi(\beta_0 + \beta_1 \ln(w) + \beta_2 s)$$

Средние предельные эффекты

• Предельный эффект в общем виде для простого случая:

$$ME_{X_1} = \beta_1 g(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)$$

• Средний предельный эффект:

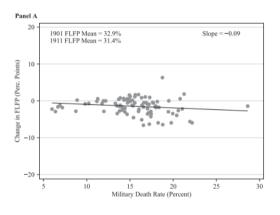
$$AME_{X_1} = \mathbb{E}[ME_{X_1}] = \beta_1 \, \mathbb{E}[g(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)]$$

• Оценка среднего предельного эффекта:

$$\widehat{AME}_{X_1} = \frac{1}{n} \sum_{i=1}^{n} \widehat{ME}_{X_{1i}} = \hat{\beta}_1 \frac{1}{n} \sum_{i=1}^{n} g(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_k X_{ki})$$

Дополнительный пример (не бинарный выбор!)

Boehnke, J., & Gay, V. (2022). The missing men World War I and female labor force participation. *Journal of Human Resources*, 57(4), 1209-1241.



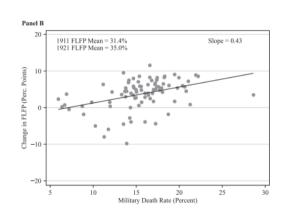


Рис.: 1901–1911 Рис.: 1911–1921

Дополнительный пример (не бинарный выбор!)

Boehnke, J., & Gay, V. (2022). The missing men World War I and female labor force participation. *Journal of Human Resources*, 57(4), 1209-1241.

Table 3
Impact of WWI Military Fatalities on FLFP

	Dependent Variable: FLFP								MLFP		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Death rate × post	0.35***	0.37***	0.40**	0.21***	0.23**	0.38***	0.45***	0.28***	0.55***	0.48***	-0.01
	[0.07]	[0.08]	[0.16]	[0.08]	[0.09]	[0.07]	[0.10]	[0.07]	[0.13]	[0.07]	[0.06]
Controls Specification	Yes Baseline	No No controls	Yes Dépspecific trends	Yes Region × year FE	Yes Residual measure	Yes Empl.	Yes Female farm owner	Yes No war dép.	Yes Corr. measure	Yes Pop. weights	Yes Male placebo
Difference	0.00	0.02	0.05	-0.15	-0.13	0.03	0.10	-0.07	0.20	0.13	-0.36***
from baseline		[0.10]	[0.17]	[0.10]	[0.11]	[0.10]	[0.12]	[0.10]	[0.15]	[0.10]	[0.09]
Observations	609	609	609	609	609	609	522	532	609	609	609
Départements	87	87	87	87	87	87	87	76	87	87	87
Within R ²	0.581	0.578	0.824	0.798	0.550	0.584	0.606	0.633	0.569	0.636	0.675
1911 mean	31.4	31.4	31.4	31.4	31.4	31.1	51.5	30.5	31.4	35.7	93.2

Notes: This table reports OLS coefficients from estimating Specification 1. The dependent variable is female labor force participation rate (FLFP) in percent except in Column 6, where it is female employment rate in percent, and in Column 11, where it is male labor force participation rate (MLFP) in percent. All regressions include departement and year fixed effects. Controls include the share of nural population in percent and the share of population born in the departement in percent. Census years are 1901, 1906, 1911, 1921, 1926, 1931, and 1936. Column 3 includes departement-specific linear time trends, and Column 4 includes region-by-year fixed effects. In Column 5, military death rates are purged from pre-war trends between 1901 and 1911 in FLFP and rurally. In Column 7, female labor force participation includes female farm owners and excludes census year 1901. In Column 8, all 11 departments that experienced war combats on their territory are excluded. In Column 9, military death rates are corrected for pre-war migration patterns (see Online Appendix E). In Column 10, departments' relative population sizes in 1911 are used as weights. Standard errors are in brackets and are clustered the department evel. Significance: **p> C.0.01. **p> **p> C.0.1.

Дополнительный пример (не бинарный выбор!)

Boehnke, J., & Gay, V. (2022). The missing men World War I and female labor force participation. *Journal of Human Resources*, 57(4), 1209-1241.

	Dependent Variable: Military Death Rate × Post					
	(1)	(2)	(3)	(4)		
Panel A: First Stage						
Ratio class 1911–1912×post	-0.28*** [0.05]			-0.17*** [0.03]		
Ratio class 1912–1913×post		-0.39*** [0.05]		-0.31*** [0.06]		
Ratio class 1913–1914×post			-0.31*** [0.10]	-0.20*** [0.05]		
Controls	Yes	Yes	Yes	Yes		
Observations Départements	609 87	609 87	609 87	609 87		

	Dependent Variable: FLFP							
	OLS	IV						
	(1)	(2)	(3)	(4)	(5)			
Panel B: Second Stage								
Death rate×post	0.35*** [0.07]	0.80*** [0.22]	0.48*** [0.14]	0.37** [0.18]	0.54*** [0.13]			
Instruments Controls	No Yes	1 Yes	2 Yes	3 Yes	1–3 Yes			
Difference with OLS	0.00	0.44* [0.23]	0.13 [0.16]	0.02 [0.19]	0.19 [0.15]			
Observations Départements Within R ²	609 87 0.581	609 87 0.507	609 87 0.575	609 87 0.581	609 87 0.567			
KPW F-statistic		29.159	55.847	10.638	27.350			

Содержание

① Еще раз об интерпретациях и предельных эффектах

2 Оценивание параметров в моделях бинарного выбора

Модели бинарного выбора: свойства оценок параметров

- Чтобы проверят гипотезы о коэффициентах нам нужно, чтобы оценки параметров были ...
 - состоятельными: $\hat{\beta} \xrightarrow[n \to \infty]{p} \beta$
 - асимптотически нормальными: $\sqrt{n}(\hat{\beta}-\beta) \xrightarrow[n \to \infty]{d} \mathcal{N}(0,\mathbf{V}_{\hat{\beta}})$
- Тогда мы могли бы проверять гипотезы и о предельных эффектах, опираясь на их свойства:

•
$$\widehat{ME} \xrightarrow{p} ME$$

•
$$\sqrt{n}(\widehat{ME} - ME) \xrightarrow{d} N(0,...)$$

- Также мы могли бы проверять и множественные гипотезы о совместной значимости и т. п., используя критерий Вальда.
- Можно использовать робастные оценки ковариационной матрицы оценок коэффициентов (но сложно).

Модели бинарного выбора: проблемы оценивания

- Пробит-модель и логит-модель наиболее распространенные модели бинарного выбора ...
 - уважительно относятся к границам вероятности;
 - относятся к моделям линейного индекса и являются нелинейными по параметрам:

$$\Pr(Y = 1 | X_1, \dots, X_k) = G(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)$$

- Параметры в этих моделях могут быть оценены при помощи:
 - нелинейного метода наименьших квадратов;
 - метода максимального правдоподобия.
- Оба метода позволяют получить состоятельные и асимптотические нормальные оценки.
- На практике наиболее распространен метод максимального правдоподобия (ММП).

Нелинейный метод наименьших квадратов: основные идеи

- Нелинейный МНК расширяет стандартные МНК-оценки на случай нелинейных по параметрам регрессионных моделей.
- Как и обычный МНК, он минимизирует сумму квадратов ошибок в используемой модели:

$$\sum_{i=1}^{n} [Y_i - G(b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k)]^2 \longrightarrow \min_{b_0, b_1, \dots}$$

• Аналитического решения, т. е. простых формул для оценок параметров $\hat{eta}_0,\hat{eta}_1,\dots$ нет, поэтому для решения используются численные методы.

Метод максимального правдоподобия: основные определения

- Функция правдоподобия совместная функция распределения вероятностей данных, рассматриваемая как функция неизвестных коэффициентов.
- Оценки максимального правдоподобия (ОМП, MLE) дают значения коэффициентов, которые максимизируют функцию правдоподобия.
- Поскольку ОМП коэффициентов максимизируют функцию правдоподобия, которая, в свою очередь, является совместной функцией распределения вероятностей, то ОМП позволяют получить значения параметров, которые максимизируют вероятность реализации именно тех данных, которые действительно наблюдаются (содержатся в используемой выборке).
- В этом смысле ОМП являются значениями параметров, которые с «наибольшей вероятностью» производят имеющиеся данные.

Метод максимального правдоподобия: пример

- Пусть Y случайная величина, имеющая распределение Бернулли.
- Наблюдения Y_i незавимые и одинаково распределенные.
- Для простоты рассмотрим два наблюдения Y_1 и Y_2 .
- Нам нужно оценить неизвестный параметр $p = \Pr(Y = 1)$.
- Совместная функция распределения:

$$Pr(Y_1 = y_1, Y_2 = y_2) = Pr(Y_1 = y_1) Pr(Y_2 = y_2) =$$

$$= [p^{y_1}(1-p)^{1-y_1}] [p^{y_2}(1-p)^{1-y_2}] =$$

$$= p^{y_1+y_2}(1-p)^{2-(y_1+y_2)}$$

• Функция правдоподобия:

$$f(p; Y_1, Y_2) = p^{Y_1 + Y_2} (1 - p)^{2 - (Y_1 + Y_2)}$$

• Задача — максимизировать функцию правдоподобия по неизвестному параметру p.

Метод максимального правдоподобия: пример

- MM Π : $p^{Y_1+Y_2}(1-p)^{2-(Y_1+Y_2)} \longrightarrow \max_p$
- Пусть $Y_1 = 1$, $Y_2 = 0$
- ullet Тогда функция правдоподобия принимает вид p(1-p)
- Максимум функции достигается при $\hat{p}^{MLE}=0.5$
- ullet В общем случае $\hat{p}^{MLE} = ar{Y}$

Оценки максимального правдоподобия для моделей линейного индекса

• В моделях линейного индекса

$$\Pr(Y_i = 1 | X_{1i}, \dots, X_{ki}) = G(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}) \equiv p_i$$

• Тогда совместная функция распределения Y_1, \ldots, Y_n имеет вид

$$Pr(Y_1 = y_1, ..., Y_n = y_n | X_{1i}, ..., X_{ki}, i = 1, ..., n) =$$

$$= Pr(Y_1 = y_1 | X_{11}, ..., X_{k1}) \times ... \times Pr(Y_n = y_n | X_{1n}, ..., X_{kn}) =$$

$$= (p_1^{y_1} (1 - p_1)^{1 - y_1}) \times ... \times (p_n^{y_n} (1 - p_n)^{1 - y_n})$$

• Функция правдоподобия:

$$f(\beta_0, ..., \beta_k; Y_1, ..., Y_n | X_{1i}, ..., X_{ki}, i = 1, ..., n) =$$

$$= \prod_{i=1}^n \left[G(\beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki}) \right]^{Y_i} \prod_{i=1}^n \left[1 - G(\beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki}) \right]^{1-Y_i}$$

Оценки максимального правдоподобия для моделей линейного индекса

• Логарифм функции правдоподобия:

$$\begin{aligned} & \ln \left[f(\beta_0, \dots, \beta_k; Y_1, \dots, Y_n | X_{1i}, \dots, X_{ki}, i = 1, \dots, n) \right] = \\ & = \sum_{i=1}^n Y_i \ln \left[G(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}) \right] + \\ & + \sum_{i=1}^n (1 - Y_i) \ln \left[1 - G(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}) \right] \end{aligned}$$

• Максимизация функции правдоподобия:

$$\ln \left[f(b_0,\ldots,b_k;Y_1,\ldots,Y_n|X_{1i},\ldots,X_{ki},i=1,\ldots,n) \right] \longrightarrow \max_{b_0,b_1,\ldots,b_k}$$

• Аналитического решения, т. е. простых формул для оценок параметров $\hat{\beta}_0^{MLE}, \hat{\beta}_1^{MLE}, \dots, \hat{\beta}_k^{MLE}$ нет, поэтому для решения используются численные методы.