Конспект лекции 1 по микроэконометрике

Buchko Daniil, BEC175

14 октября 2020 г.

Введение

Что изучает микроэконометрика?

Мы пытаемся ответить на вопрос, а как вероятность какого-то события зависит от выбранных факторов. Рассмотрим такой пример: преподаватель выбирает, как добраться до вуза. Его решение – бинарный выбор:

$$Y = \begin{cases} 0, & \text{Если человек пешком} \\ 1, & \text{Если человек на машине} \end{cases}$$

Предположим, мы хотим моделировать такие процессы принятия решения. Мы хотим научиться отвечать на вопрос: а как поведет себя человек, если мы имеем определенные факты о внешнем мире.

Линейные вероятностные модели.
На эконометрике были линейные вероятностные модели:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_{1_i} + \beta_2 x_{2_i} + \dots + \beta_{k-1} x_{k-1_i} + \varepsilon_i \\ &= x_i^{'}\beta + \varepsilon_i \end{aligned}$$

Почему модель называется линейной вероятностной? Чтобы ответить на этот вопрос возьмем матожидание обоих частей верхнего равенства:

$$\mathbb{E}\, Y_i = \mathbb{E}\left(x_i^{'}\beta + \varepsilon_i\right)$$

Учитывая, что в левой части уравнения стоит вероятностное распределение Бернулли с произвольным параметром p, а также тот факт, что $\mathbb{E}(\varepsilon_i) = 0$, получим:

$$p_i = x_i^{'}\beta$$

Получается, что вероятность является линейной функцией объясняющих переменных. Поэтмоу мы называем модель линейной вероятностной.

Что будет, если мы попробуем оценить параметры модели через МНК? Оценки будут состоятельны, но возникнут проблемы. По условиям *Теоремы Гауса Маркова*, чтобы модель была *BLUE (Best, Linear, Unbiased estimator)*, необходима гомоскедастичность остатков. Посмотрим, что получается у нас:

$$\begin{aligned} D(Y_i) &= p_i(1 - p_i) \\ &= x_i^{'}\beta(1 - x_i^{'}\beta) \end{aligned}$$

То есть дисперсия различных наблюдений различна и, как следствие, оценки не эффективны. Кроме того, мы не можем тестировать гипотезы о коэффициентах моделей, потому что ошибки не распределены нормально:

$$\varepsilon_i = \begin{cases} 1 - x_i^{'}\beta, & \text{если } Y_i = 1 \\ -x_i^{'}\beta, & \text{если } Y_i = 0 \end{cases}$$

Ну и самый последний недостаток состоит в том, что модель может выдать значения предсказания вероятности за пределами $[0, 1]$.

При всех недостатках, значит ли это, что модель бесполезна? Нет. Более сложные модели используют в качестве стартовой точки оценки, полученные с этой регрессии, что позволяет нсизить число итераций для нахождения оптимальных значений параметров.

Придумали делать так, чтобы вероятности всегда лежали в промежутке $[0, 1]$. Для этого над линейной вероятностной моделью произвели функциональное преобразование:

$$p_i = F(x_i^{'}\beta + \varepsilon_i)$$

F - некоторая функция, ограничивающая выводы линейной комбинации парметров с факторами. Два наиболее популярных вида F:

$$F(z) = \begin{cases} \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{z} e^{\frac{-t^2}{2}} dt, & \text{стандартное нормальное} \\ \dfrac{e^z}{1+e^z}, & \text{сигмоида} \end{cases}$$

Почему именно такие виды F? Для вывода функции распределения воспользуемся подходом с латентной пременной. Пусть есть модель:

$$Y_i^* = x_i^{'}\beta + \varepsilon_i$$

Эта модель ничем не отличается от предыдущей, кроме отсутствия преобразования. Мы говорим, что если $Y_i^* > 0$, то значение целевой переменной будем считать равной единице, ну а если $Y_i^* \leqslant 0$, то нулем соответсвтенно:

$$Y_i = \begin{cases} 1, & Y_i^* > 0 \\ 0, & Y_i^* \leqslant 0 \end{cases}$$

Тогда можем записать следующую цепочку рассуждений:

$$\begin{aligned} p_i &= P(Y_i = 1) = P(Y_i^* > 0) = P(x_i^{'}\beta + \varepsilon_i > 0) \\ &= P(\varepsilon_i > -x_i^{'}\beta) = 1 - P(\varepsilon_i \leqslant -x_i^{'}\beta) = 1 - F_{\varepsilon_i}(-x_i^{'}\beta) \end{aligned}$$

Если распределение ошибки симметрично, то итоговое значение вероятности можно записать в следующем виде:

$$p_i = F_{\varepsilon_i}(x_i^{'}\beta)$$

Почему мы считаем, что ε_i распределено нормально? Потому что по ЦПТ вклад какждого из невключенных факторов имеет маленькое влияние на общую величину ошибки $\implies$ общая ошибка распределена нормально.
Почему используем логистическую функцию? Очень похожа на стандартное нормальное распределение (более тяжелые хвосты). В 1930-х использовали вместо нормального, потому что ММП было сложно оценить для нормальной функции распрделения.

Оценка параметров вероятностных моделей

Для оценки параметров пользуются ММП:

$$\mathcal{L} = \prod_{i=1}^{n} P(Y_i = y_i) = \prod_{i=1}^{n} \left(F(x_i^{'}\beta)\right)^{y_i} \left(1 - F(x_i^{'}\beta)\right)^{1-y_i}$$

Поскольку существует единственный максимум функции правдоподобия, то получаемые оценки $\hat{\beta}$ являются:

1. состоятельными

2. асимптотически несмещенные
3. асимптотически эффективные
4. асимптотически нормальные
5. инвариантными относительно гладких преобразований

Асимптотика в том смысле, что оценки такие только при большом количестве наблюдений. Получив оценки параметров, можно использовать их в гладких преобразованиях. К примеру, зная бэты, можно посчить вероятность объекта не из обучающей выборки принадлежать к определенному классу:

$$\hat{P}(Y_j = 1 \mid \hat{\beta}) = F(x_j^{'}\hat{\beta})$$

Как интерпетировать полученные результаты? Построив бинарную модель, мы хотим понять зависимость между данными и целевой переменной. Для этого, считаются предельные эффекты, а именно:

$$\frac{\partial \hat{P}(\bar{Y} = 1)}{\partial x_j} = \frac{\partial F(\bar{x}^{'}\hat{\beta})}{\partial x_j} = f(\bar{x}'\hat{\beta})\hat{\beta}_j, \quad for\ x_j \in R$$

$$\Delta P = P(\bar{Y} = 1 \mid D_j = 1) - P(\bar{Y} = 1 \mid D_j = 0), \quad for\ D_j \in \{0, 1\}$$

В качестве точки для подсчета производной берется либо типичный представитель по всем признакам, либо усредненное значение всех примеров в выборке.

Свойства логистической модели

1. $\dfrac{\sum_{i=1}^{n} \hat{P}(Y_i = 1)}{n} = \dfrac{\sum_{i=1}^{n} Y_i}{n}$

   Доказывается через максимизацию L, условие первого порядка.

2. $\dfrac{\partial F(z)}{\partial z} = F(z)(1 - F(z))$

   Доказывается через прямое дифференциирование