Оценка структурных параметров и эффектов воздействия в квантильных регрессиях с использованием метода двойного машинного обучения

А. М. Ченцов[1]

[1] Московский физико-технический институт, г. Долгопрудный, Россия,
chentsov.am@mipt.ru

Модель. Рассматривается модель структурной частично-линейной регрессии с вектором контрольных переменных $X = (X_1, \ldots, X_p)$, в основном уравнении которой идентификационное условие накладывается на квантиль уровня $\tau \in (0, 1)$ условного распределения зависимой переменной Y:

$$Q_{Y|X,D}(\tau) := \alpha_0(\tau)D + g_0(X, \varepsilon), \quad f_\varepsilon(0) > 0, \tag{1}$$

$$D = m_0(X) + v, \quad \mathbb{E}(v \mid X) = 0. \tag{2}$$

Здесь $Q_{Y|X,D}(\tau)$ – квантильная функция условного распределения зависимой переменной Y; D – выделенная переменная (обычно в таком контексте называемая переменной воздействия), g_0 и m_0 – неизвестные функции, ε – *н.о.р.* ошибка, имеющая абсолютно-непрерывное распределение, v – ошибка во вспомогательной модели. Исследователю доступна случайная выборка $\{(Y_i, D_i, X_i)\}_{i=1}^n$, причём размер выборки может быть меньше числа контрольных переменных p. Интересующим исследователя параметром является коэффициент α_0, который обычно в такой модели интерпретируется как эффект воздействия на τ-квантиль распределения Y.

Ортогонализованная оценка α_0. Для построения асимптотически нормальной оценки α_0 используется функция потерь $\psi(\alpha_0, \eta_0; X, Y, D)$, имеющая нулевую производную Гато по отношению к мешающему параметру $\eta_0 := (m_0, g_0)$. Это позволяет получать оценку α_0 методом подстановки $\hat\eta_0$ – состоятельных оценок, основанных на методах машинного обучения, причем неточность используемых оценок в конечной выборке компенсируется низкой чувствительностью статистики к отклонениям $\hat\eta_0$ от η_0 в окрестности истинного значения η_0.

Предположение 1. Существуют $a, A \in \mathbb{R}$:

$$\liminf_{n\to\infty} \inf_{\|u\|=1} \mathbb{E}_n \mathbb{1}\{|vu| < a\} = 0, \quad \liminf_{n\to\infty} \sup_{\|u\|=1} \mathbb{E}_n (vu)^2 \le A,$$

где $\mathbb{E}_n$ – обозначение для выборочного среднего.

Предположение 2. Существуют $b_0, b_1(\tau) > 0$: $\lim_{n\to\infty} \mathbb{E}_n v^2 = b_0$, $\lim_{n\to\infty} \mathbb{E}_n \big[f_\varepsilon\big(F_\varepsilon^{-1}(\tau)\big)v^2\big] = b_1(\tau)$, $\max_{i\in\{1,\ldots,n\}} |v_i|/\sqrt{n} \to 0$.

Теорема 1. Пусть $\{P_n\}$ – семейство распределений, порождающих данные модели (1)-(2), для которых выполнены предположения 1 и 2. Тогда оценка $\hat\alpha_0$, заданная выборочным аналогом уравнения $\alpha_0 = \min_\alpha \mathbb{E}\rho_\tau\big[Y - g_0(X) - \alpha\big(D - m_0(X)\big)\big]$, где $\rho_\tau(z) := \tau - \mathbb{1}\{z \le 0\}$, и в котором вместо истинных значений g_0, m_0 используются их $n^{-\phi_g}$, $n^{-\phi_m}$ состоятельные оценки соответственно, такие, что $\phi_g + \phi_m \ge \frac{1}{2}$, является асимптотически нормальной равномерно по всем элементам $\{P_n\}$:

$$\forall P \in \{P_n\}, \quad \sqrt{n}(\hat\alpha - \alpha_0) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \quad \sigma^2 = \frac{\tau(1-\tau)}{f_\varepsilon^2(0)}.$$

Литература

1. Chernozhukov V., Chetverikov D., Demirer M., Duflo E., Hansen C., Newey W., Robins J. *Double/debiased machine learning for treatment and structural parameters.* The Econometrics Journal. 2018. V. 21. № 1. P. 1-68.

2. Neyman J. *$C(\alpha)$ tests and their use.* Sankhya: The Indian Journal of Statistics, Series A. 1979. V. 41. № 1/2. P. 1–21.

3. Koenker R. *Quantile regression.* Cambridge University Press. 2005.

4. Ченцов А. М., Торопов Н. И. *Применение подхода двойного машинного обучения для задачи анализа зависимости между отклонениями от непокрытого паритета процентных ставок и степенью открытости экономики* // Труды МФТИ 2024, Т. 16, № 3. С. 1–11.