АВТОРЕГРЕССИОННЫЕ ВРЕМЕННЫЕ РЯДЫ ПРИ НАЛИЧИИ КЛАССИФИКАЦИИ НАБЛЮДЕНИЙ: МНК-ОЦЕНКА И ЕЕ УЛУЧШЕНИЕ

А. В. Рудаковская

ВВЕДЕНИЕ

Модель авторегрессионных временных рядов (AR(p)) часто используется в математической статистике для исследования процессов и явлений в различных сферах. К сожалению, на практике наблюдения часто регистрируются с искажениями.

Одним из типов искажений авторегрессионных временных рядов является классификация наблюдений – регистрация вместо истинного значения временного ряда лишь номера класса, к которому относится исходное наблюдение. Искажения такого типа описаны в [1–3].

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ И ПОСТАНОВКА ЗАДАЧИ

Пусть на (Ω, F, P) определен временной ряд авторегрессии порядка p: $x_t = \theta_1 x_{t-1} + \theta_2 x_{t-2} + \cdots + \theta_p x_{t-p} + \xi_t, t \in \mathbb{Z}$, где $p \in \mathbb{N}$ – порядок авторегрессии, $\theta = (\theta_i)' \in \mathbb{R}^p$ – вектор коэффициентов авторегрессии, удовлетворяющий условиям устойчивости, $\theta_p \neq 0$, $\{\xi_t\}$ – дискретный «белый шум», $\mathcal{L}\{\xi_t\} = N_1(0, \sigma^2)$. Задано борелевское разбиение числовой прямой на $2 \leq L < +\infty$ числовых промежутков $\mathbb{R} = \bigcup_{i=0}^{L-1} A_{i+1}, A_{i+1} = (a_i; a_{i+1}], -\infty = a_0 << a_1 < \cdots < a_{L-1} < a_L = +\infty$.

Вместо исходного временного ряда x_t наблюдается классифицированный временной ряд y_t, $y_t = \sum_{i=1}^{L} i I_{A_i}(x_t)$, где y_t – номер класса, в который попало значение x_t ($I_A(x)$ – индикаторная функция множества A).

Полагаем известными порядок авторегрессии p, значения классифицированного временного ряда y_t, $t = 0, \ldots T$, а также интервалы $(a_i; a_{i+1}], i = 0, \ldots L-1$, по которым проводилась классификация.

Ставится задача оценивания неизвестных параметров $\theta_1, \theta_2, \cdots, \theta_p$.

Представим результаты исследования для модели AR(1) с классификацией наблюдений при $L = 2$, $A_1 = (-\infty, 0]$, $A_2 = (0, +\infty)$:

$$x_t = \theta x_{t-1} + \xi_t; \;\; \mathcal{L}\{\xi_t\} = N_1(0, \sigma^2); -1 < \theta < 1;$$
$$y_t = sign(x_t) = \begin{cases} 1, & x > 0; \\ -1, & x \leq 0. \end{cases} \quad (1)$$

МНК-ОЦЕНКА И ЕЕ СВОЙСТВА

МНК-оценка для модели (1) имеет вид:

$$\hat{\theta} = (y_0 y_{-1} + y_1 y_0 + \cdots + y_T y_{T-1})/(y_{-1}^2 + y_0^2 + \cdots + y_{T-1}^2). \quad (2)$$

Теорема 1 [3]. Оценка (2), построенная по методу наименьших квадратов для модели (1), является асимптотически смещенной, причем смещение является функцией параметра θ и верно предельное соотношение

$$\lim_{T\to\infty} b(\hat{\theta}) = 4p_{++}(\theta) - 1 - \theta = \beta(\theta), \quad (3)$$

где $p_{++}(\theta) = \sqrt{1-\theta^2}\int_0^{+\infty} \varphi\left(\sqrt{1-\theta^2}x\right)\Phi(\theta x)dx$, $\varphi(\cdot)$, $\Phi(\cdot)$ – плотность и функция распределения стандартного нормального закона $\mathcal{N}_1(0,1)$.

Теорема 2 [3]. МНК-оценка (2) имеет вариацию, асимптотически равную к квадрату смещения оценки:

$$\lim_{T\to\infty} v(\hat{\theta}) = \left(\beta(\theta)\right)^2 = \nu(\theta). \quad (4)$$

На рис. 1 представлены значения смещения и вариации оценки (2), полученные численно и аналитически, в зависимости от значения параметра θ:

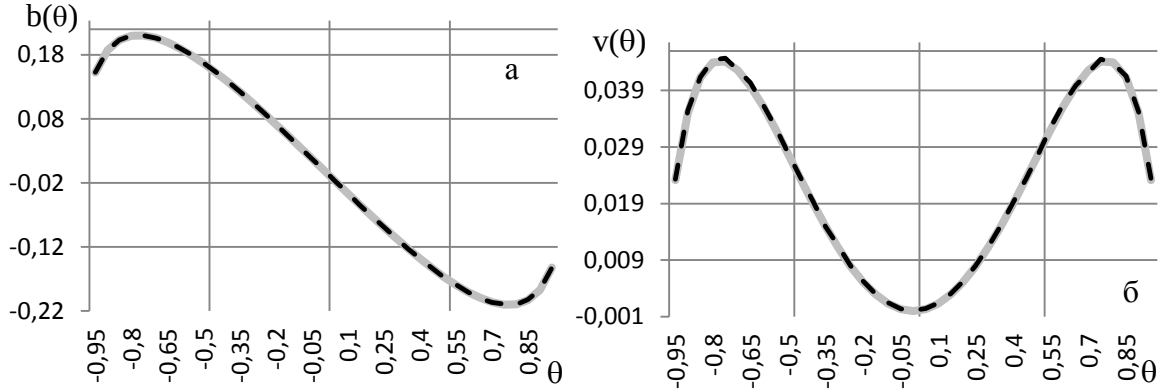

Рис. 1. а) смещение и *б*) вариация оценки в зависимости от значения θ.
График серого цвета – теоретически полученная кривая, черный пунктир – практически полученные значения

УЛУЧШЕНИЕ МНК-ОЦЕНКИ

Как видно из рисунка 1, смещение оценки достаточно велико. Поэтому ставится задача: улучшить существующую МНК-оценку путем уменьшения ее смещения.

Если бы была возможность построить оценку $\tilde{\theta}^{(0)} = \hat{\theta} - b(\hat{\theta}) = \hat{\theta} - \beta(\theta)$, то такая оценка оказалась бы несмещенной. Действительно,

$$E(\tilde{\theta}^{(0)}) = E\left(\hat{\theta} - \beta(\theta)\right) = E(\hat{\theta}) - E(\beta(\theta)) = \theta + \beta(\theta) - \beta(\theta) = \theta.$$

Однако для вычисления $\beta(\theta)$ необходимо знать истинное значение параметра θ, которое неизвестно. Поэтому воспользуемся разложением функции $\beta(\theta)$ в ряд Тейлора в окрестности построенной оценки $\hat{\theta}$:

$$\tilde{\theta}^{(0)} = \hat{\theta} - \beta(\theta) = \hat{\theta} - \beta\big(\hat{\theta} + (\theta - \hat{\theta})\big) = \hat{\theta} - \beta\big(\hat{\theta}\big) + O((\theta - \hat{\theta})^2).$$

Тогда улучшенная в первом приближении оценка будет иметь вид (сразу же учтем (2) и упростим вид оценки):

$$\tilde{\theta}^{(1)} = \hat{\theta} - \beta\big(\hat{\theta}\big) = 2\hat{\theta} - 4\mathrm{p}_{++}\big(\hat{\theta}\big) + 1. \qquad (5)$$

Результаты статистического моделирования для оценки (5) представлены на рис. 2 (график черного цвета – истинное значение параметра θ, серый пунктир – МНК-оценка, серая линия – улучшенная МНК-оценка, полученная при помощи формулы (5)). Количество реализаций временного ряда $K = 10^3$; длина каждого временного ряда $T \in \{9800, 9850, \cdots, 10000\}$.

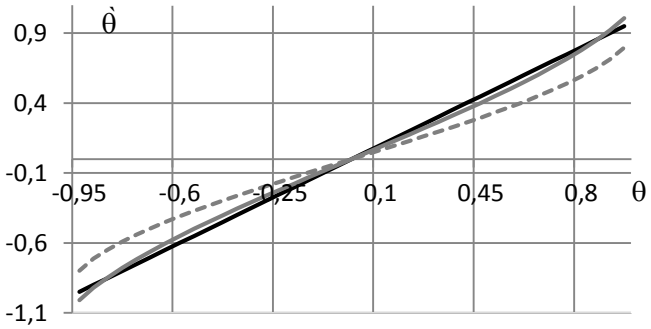

Рис. 2. Результаты статистического моделирования

СМЕЩЕНИЕ УЛУЧШЕННОЙ МНК-ОЦЕНКИ

Исследуем смещение полученной оценки, $b\big(\tilde{\theta}^{(1)}\big) = E\big(\tilde{\theta}^{(1)}\big) - \theta$. Для его нахождения необходимо найти $E\big(\tilde{\theta}^{(1)}\big)$:

$$E\big(\tilde{\theta}^{(1)}\big) = E\big(2\hat{\theta} - 4\mathrm{p}_{++}\big(\hat{\theta}\big) + 1\big) = 8\mathrm{p}_{++}(\theta) - 4E\left(\mathrm{p}_{++}\big(\hat{\theta}\big)\right) - 1.$$

Для нахождения $E\left(\mathrm{p}_{++}\big(\hat{\theta}\big)\right)$ были применены 4 различных способа:

1) разложение в ряд Тейлора в окрестности θ с точностью $O((\theta - \hat{\theta})^2)$;
2) разложение в ряд Тейлора в окрестности θ с точностью $O((\theta - \hat{\theta})^3)$;
3) разложение в ряд Тейлора в окрестности 0 с точностью $O(\hat{\theta}^3)$;
4) вычисление $E\left(\mathrm{p}_{++}\big(\hat{\theta}\big)\right)$ по определению (как интеграл).

Каждый из этих способов позволяет построить оценки, в какой-то мере улучшающие МНК-оценку. На рис. 3 представлены результаты сравнения смещения новых оценок с МНК-оценкой в зависимости от истинного значения θ. График черного цвета – смещение для МНК-оценки,

точки – для новой оценки, полученной способом 1, светло-серая линия – способом 2 (на рисунке *б* скрыты значения с большим смещением), темно-серая линия – способом 3, ромбы – способом 4 (на рисунке *б* устранена особенность в точке 0):

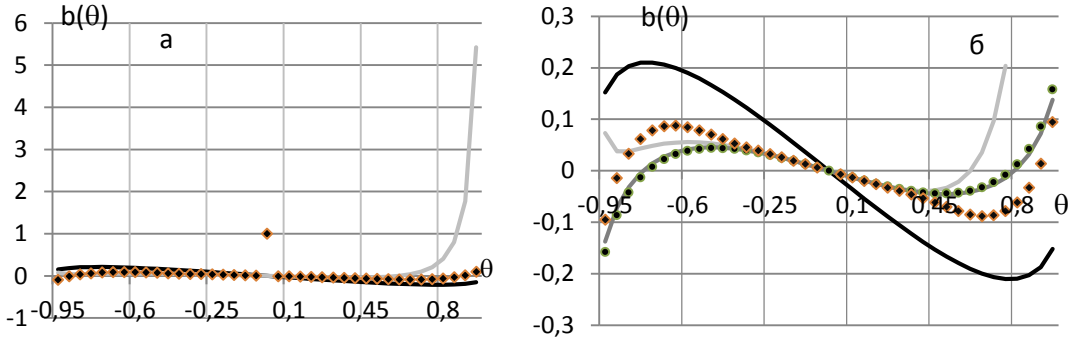

Рис. 3. Смещение новых оценок в сравнении с МНК-оценкой

Таким образом, МНК-оценка (2) для параметра θ модели (1) может быть улучшена при помощи формулы (5). Смещение оценки может быть найдено приближенно несколькими способами, каждый из которых имеет различную точность для разных θ.

ЗАКЛЮЧЕНИЕ

В работе рассмотрена модель авторегрессионных временных рядов при наличии искажений – классификации наблюдений. Рассмотрены свойства классической МНК-оценки параметра θ. Построено первое улучшение МНК-оценки и исследованы его свойства. С помощью найденных выражений для смещения улучшенной оценки, проводя аналогичные рассуждения, можно построить второе и последующие улучшения МНК-оценки.

Литература

1. *Агеева, Е. С., Харин, Ю. С.* Статистическое оценивание параметров множественной регрессии при наличии классификации наблюдений // Международный конгресс по информатике. – Минск:БГУ, 2011. – С . 22–26.
2. *Харин Ю. С.* Оптимальность и робастность в статистическом прогнозировании. – Мн:БГУ, 2008.
3. *Рудаковская А.В.* Авторегрессионные временные ряды при наличии классификации наблюдений // Сб. работ 70-й научн. конф. студентов и аспирантов БГУ. – Минск:БГУ, 2013. – С. 231–234.