

Робастное, адаптивное и сетевое управление

© 2019 г. В.Б. ГОРЯИНОВ, д-р физ.-мат. наук (vb-goryainov@bmstu.ru)
(Московский государственный технический университет им. Н.Э. Баумана),
Е.Р. ГОРЯИНОВА, канд. физ.-мат. наук (el-goryainova@mail.ru)
(Национальный исследовательский университет
“Высшая школа экономики”, Москва)

СРАВНИТЕЛЬНЫЙ АНАЛИЗ РОБАСТНЫХ И КЛАССИЧЕСКИХ МЕТОДОВ ОЦЕНИВАНИЯ ПАРАМЕТРОВ УРАВНЕНИЯ ПОРОГОВОЙ АВТОРЕГРЕССИИ

При помощи компьютерного моделирования и исследования асимптотического распределения изучается относительная эффективность М-оценок коэффициентов порогового авторегрессионного уравнения по отношению к оценкам наименьших квадратов и наименьших модулей. Предполагается, что обновляющая последовательность авторегрессионного уравнения может иметь распределения Стьюдента, логистическое, двойное экспоненциальное, нормальное и загрязненное нормальное. Доказана асимптотическая нормальность М-оценок с выпуклой функцией потерь.

Ключевые слова: пороговая авторегрессионная модель, оценка наименьших квадратов, М-оценка, асимптотическая относительная эффективность, распределение Тьюки.

DOI: 10.1134/S0005231019040056

1. Введение

Среди нелинейных моделей временных рядов широко распространены пороговые модели, описываемые кусочно-линейными уравнениями в зависимости от величины пороговой переменной [1]. Пороговые модели применяются во многих областях науки и техники, например в биологии и медицине [2–5], психологии [6], экономике и финансах [2, 7], гидрологии [2, 8], океанологии [2, 9], астрономии [2], радиотехнике и обработке сигналов [2, 10], при изучении качки корабля и для управления сервоприводом и паровой машиной [2].

Когда каждый линейный режим описывается авторегрессионным уравнением, пороговая модель называется пороговой авторегрессионной моделью. Если пороговой переменной является функция прошлых наблюдений этого же временного ряда, то пороговая модель называется самовозбуждающейся.

Примеры, приведенные в [11], показывают, что пороговые модели обеспечивают лучшее описание реальных временных рядов, чем линейные модели. В частности, этими моделями может быть описан целый ряд нелинейных эффектов, например предельные циклы и резонансные скачки [11].

Одной из важнейших задач при исследовании пороговой авторегрессионной модели является оценивание ее параметров — коэффициентов соответствующего порогового уравнения. В линейных стохастических моделях наиболее известными оценками являются оценки наименьших квадратов и наименьших модулей. Однако известно, что в этих моделях оценки наименьших квадратов являются наилучшими лишь в гауссовском случае и резко теряют эффективность даже при небольшом нарушении предположения о гауссовости наблюдений, а эффективность оценок наименьших модулей для основных вероятностных распределений наблюдений невысока [12]. Хорошим компромиссом в этом случае являются робастные оценки [12–16], в частности M-оценки.

В настоящей статье M-оценки предлагается использовать для оценивания параметров порогового авторегрессионного уравнения. Проводится сравнительный анализ перечисленных оценок при наиболее распространенных вероятностных распределениях обновляющего процесса порогового уравнения. Исследование проводится как при помощи изучения асимптотического распределения оценок, так и с помощью компьютерного статистического эксперимента. При этом доказана асимптотическая нормальность M-оценок с выпуклой целевой функцией.

2. Постановка задачи

Рассмотрим случайный процесс X_t , описываемый кусочно-разностным уравнением [17, с. 77]

$$(1) \quad X_t = \begin{cases} a_{11}X_{t-1} + a_{21}X_{t-2} + \dots + a_{p1}X_{t-p} + \varepsilon_t, & U_t \geq r, \\ a_{12}X_{t-1} + a_{22}X_{t-2} + \dots + a_{p2}X_{t-p} + \varepsilon_t, & U_t < r, \end{cases}$$

с переключением в некоторой известной точке $r \in \mathbb{R}$, называемой порогом. Переключающая переменная U_t может иметь как внешнее (экзогенное) происхождение, так и может определяться самой моделью. В данной статье рассматривается самовозбуждающаяся модель, где переключение в момент t является измеримой функцией предыдущих наблюдений $U_t = U(X_1, \dots, X_{t-1})$. Авторегрессионные коэффициенты a_{ij} представляют собой действительные числа, являются параметрами авторегрессионной пороговой модели и считаются неизвестными. Далее всюду предполагается, что обновляющий процесс ε_t — последовательность независимых одинаково распределенных непрерывных случайных величин с плотностью распределения вероятности f , нулевым математическим ожиданием $E\varepsilon_t = 0$ и конечной дисперсией $D\varepsilon_t = \sigma^2$. Будем также считать, что процесс X_t является стационарным и эргодическим. В случае $p = 1$, $U_t = X_{t-1}$ и $r = 0$ для этого достаточно одновременное выполнение условий $a_{11} < 1$, $a_{12} < 1$ и $a_{11}a_{12} < 1$ [18]. В более общих случаях эти условия являются громоздкими (см. [19]).

Рассмотрим задачу оценивания по наблюдениям X_1, X_2, \dots, X_n процесса X_t параметров модели (1) — вектора коэффициентов $a = (a_{11}, \dots, a_{p1}, a_{12}, \dots, a_{p2})^\top$. Обозначим через \mathfrak{F}_k σ -алгебру, порожденную величинами X_1, X_2, \dots, X_k , а через $E[X_t | \mathfrak{F}_{t-1}]$ — условное математическое ожидание X_t относительно σ -алгебры \mathfrak{F}_{t-1} . M-оценку \hat{a} вектора a по наблюдениям

X_1, X_2, \dots, X_n определим как точку минимума функции

$$(2) \quad L(a) = \sum_{t=2}^n \rho(X_t - \mathbb{E}[X_t | \mathfrak{F}_{t-1}]),$$

где ρ — некоторая неотрицательная четная функция. При построении М-оценок ρ -функция обычно выбирается возрастающей при $x > 0$, удовлетворяет условию $\rho(0) = 0$ и растет на бесконечности медленнее, чем x^2 (см. [12, с. 31]). М-оценки образуют целое семейство, зависящее от вида функции ρ . Свое название М-оценка получила из-за того, что если $\rho(x) = -\ln f(x)$, то она совпадает с оценкой максимального правдоподобия. Оценки наименьших квадратов \tilde{a} и наименьших модулей a^* являются частными случаями М-оценок при $\rho(x) = x^2$ и $\rho(x) = |x|$ соответственно.

Необходимость в появлении М-оценок была обусловлена недостатками оценок наименьших квадратов и наименьших модулей. Основным недостатком метода наименьших квадратов заключается в его сильной чувствительности к большим значениям невязок $\varepsilon_t = X_t - \mathbb{E}[X_t | \mathfrak{F}_{n-1}]$, поскольку они влияют на минимизируемую функцию (2) квадратичным образом. Вследствие этого точность оценки наименьших квадратов достаточно сильно ухудшается при росте вероятности экстремальных значений ε_t , что происходит, если плотность распределения вероятности f случайных величин ε_t медленно убывает на бесконечности.

Этого недостатка лишен метод наименьших модулей, в котором влияние невязок на минимизируемую функцию (2) линейно. Однако в методе наименьших модулей недостаточно внимания уделяется небольшим значениям невязок, которые также искажают функцию (2). Например, в случае нормального распределения ε_t , плотность которого $f(x)$ с ростом $|x|$ стремится к нулю достаточно быстро, асимптотическая относительная эффективность оценки наименьших квадратов по отношению к оценке наименьших модулей равна $\pi/2$ (см., [12, 16, 20]), т.е. методу наименьших модулей требуется в полтора раза больше наблюдений для достижения точности метода наименьших квадратов.

В этом смысле М-оценки являются значительным обобщением, позволяющим использовать преимущества и оценки наименьших квадратов и оценки наименьших модулей. М-оценки, как и оценку наименьших модулей, целесообразно использовать при отклонении распределения ε_t от нормального, но в отличие от оценки наименьших модулей М-оценки почти не теряют эффективность и в случае нормального распределения ε_t . Механизм такого поведения М-оценок хорошо виден на примере М-оценки с наиболее распространенной ρ -функцией, называемой ρ -функцией Хьюбера [16], которая совпадает с x^2 в окрестности $(-k, k)$ начала координат и ведет себя линейно вне этой окрестности. Семейство ρ -функций Хьюбера имеет вид

$$(3) \quad \rho_H(x) = \begin{cases} x^2, & \text{если } |x| \leq k, \\ 2k|x| - k^2, & \text{если } |x| > k, \end{cases}$$

где k — параметр, принимающий положительные значения.

В М-методе с ρ -функцией Хьюбера $\rho_H(x)$ оптимальный выбор параметра k приводит к тому, что слагаемые в (2) с небольшими невязками становятся как и в методе наименьших квадратов еще меньше, а слагаемые с большими невязками становятся как и в методе наименьших модулей не слишком большими. Таким образом, вклад в сумму (2) наблюдений X_t , сформировавшихся под влиянием экстремальных значений ε_t , будет принижен по сравнению с вкладом остальных наблюдений.

Цель настоящей статьи — исследование свойств М-оценок вектора коэффициентов a по наблюдениям X_1, \dots, X_n случайного процесса (1) и сравнение М-оценки с ρ -функцией (3) с оценками наименьших квадратов и наименьших модулей при наиболее распространенных вероятностных распределениях обновляющего процесса ε_t .

3. Асимптотическое распределение оценок

Из двух оценок лучшей будем считать ту, которая меньше отклоняется от оцениваемого параметра. Тогда точность несмещенной оценки скалярного параметра будет определяться ее дисперсией — чем меньше дисперсия, тем точнее оценка. Обычно удается найти лишь асимптотические дисперсии оценок и установить их асимптотическую несмещенность. В этом случае лучшей из оценок следует считать оценку с наименьшей асимптотической дисперсией, а в качестве сравнительной характеристики их точности использовать асимптотическую относительную эффективность, определяемую как обратное отношение их асимптотических дисперсий. А именно предположим, что есть две последовательности \hat{b}_1 и \hat{b}_2 оценок параметра b , построенные по n наблюдениям. Если последовательность $\sqrt{n}(\hat{b}_i - b)$ слабо сходится при $n \rightarrow \infty$ к нормальной случайной величине с нулевым средним и дисперсией d_i , $i = 1, 2$, то асимптотическая относительная эффективность оценки \hat{b}_2 относительно \hat{b}_1 равна d_1/d_2 (см., например, [21, 22]).

Асимптотическая относительная эффективность одной оценки по отношению к другой показывает, во сколько раз меньше наблюдений нужно первой оценке для достижения точности второй оценки. Если оценки являются векторными, но их асимптотические ковариационные матрицы пропорциональны, то асимптотическая относительная эффективность определяется как обратное отношение этих коэффициентов пропорциональности.

Положим

$$Y_t = (Y_{11}(t), \dots, Y_{p1}(t), Y_{12}(t), \dots, Y_{p2}(t))^T,$$

где

$$Y_{i1}(t) = X_{t-i}I(U_t \geq r), \quad Y_{i2}(t) = X_{t-i}I(U_t < r), \quad i = 1, 2, \dots, p.$$

Обозначим через K ковариационную матрицу вектора Y_t , через \tilde{a} — оценку наименьших квадратов, а через a^* — оценку наименьших модулей параметра a порогового уравнения (1).

Известно [19, 23], что случайные последовательности $\sqrt{n}(\tilde{a} - a)$ и $\sqrt{n}(a^* - a)$ асимптотически нормальны с нулевым средним и ковариационными матрицами $\sigma^2 K^{-1}$ и $(4f^2(0)K)^{-1}$ соответственно. Таким образом, для

сравнения М-оценки \hat{a} с оценками наименьших квадратов и наименьших модулей нужно знать ее асимптотическое распределение.

Имеет место следующая теорема.

Теорема. Пусть \hat{a} – М-оценка параметра $a = (a_{11}, \dots, a_{p1}, a_{12}, \dots, a_{p2})^\top$ уравнения (1), определяемая как точка минимума функции (2). Пусть процесс X_t , описываемый уравнением (1), является стационарным и эргодическим, $\rho(x)$ в (2) – выпуклая функция, $\rho''(x)$ непрерывна почти всюду и ограничена на \mathbb{R} , плотность $f(x)$ независимых одинаково распределенных случайных величин ε_t в (1) удовлетворяет условиям $E\varepsilon_t = 0$, $E(\rho'(\varepsilon_t)^2) < \infty$, $E\rho'(\varepsilon_t) = 0$, $0 < E\rho''(\varepsilon_t) < \infty$ и $\sigma^2 = D\varepsilon_t < \infty$.

Тогда при $n \rightarrow \infty$ случайный вектор $\sqrt{n}(\hat{a} - a)$ является асимптотически нормальным с нулевым математическим ожиданием и ковариационной матрицей $B = K^{-1} \frac{E[\rho'(\varepsilon_1)^2]}{(E[\rho''(\varepsilon_1)])^2}$.

Доказательство теоремы приведено в Приложении. Отметим, что из выпуклости $\rho(x)$ вытекает выпуклость функции (2), поэтому М-оценка всегда существует.

Согласно теореме относительная асимптотическая эффективность М-оценки по отношению к оценкам наименьших квадратов и наименьших модулей равна $e(\hat{a}, \tilde{\alpha}) = \frac{\sigma^2 (E[\rho''(\varepsilon_1)])^2}{E[\rho'(\varepsilon_1)^2]}$ и $e(\hat{a}, \alpha^*) = \frac{(E[\rho''(\varepsilon_1)])^2}{4f^2(0)E[\rho'(\varepsilon_1)^2]}$ соответственно. Отметим, что асимптотическая относительная эффективность оценки наименьших квадратов по отношению к оценке наименьших модулей равна $e(\tilde{\alpha}, \alpha^*) = \frac{1}{4\sigma^2 f^2(0)}$.

4. Сравнение оценок

В приложениях обычно считается, что ε_t – нормальные случайные величины с нулевым математическим ожиданием, т.е. $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)$. Это допущение обосновывается центральной предельной теоремой.

С другой стороны, согласно центральной предельной теореме предположение о нормальности выполняется лишь приближенно. Общепринятой моделью приближенного нормального распределения является смесь распределений [24] с плотностью

$$(4) \quad f(x) = (1 - \gamma) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} + \gamma g f_\gamma(x), \quad \tau > 1, \quad 0 < \gamma < 1,$$

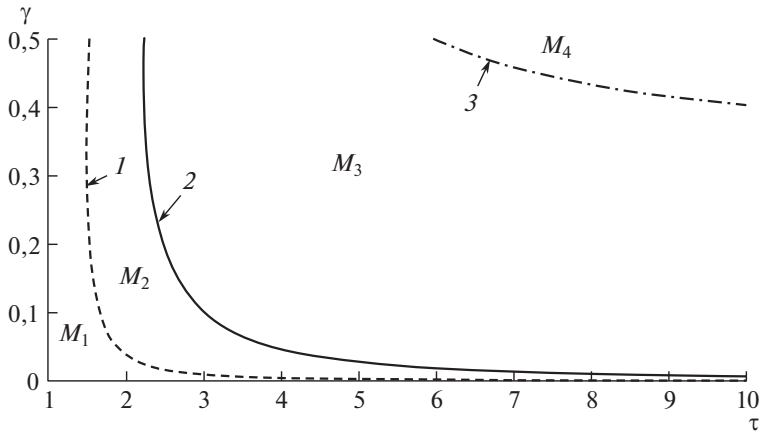
где $f_\gamma(x)$ – плотность распределения, моделирующего отклонение от нормального. Другими словами, среди ε_t с вероятностью $1 - \gamma$ встречаются стандартные (с нулевым математическим ожиданием и единичной дисперсией) нормальные величины, а с вероятностью γ – случайные величины с плотностью $f_\gamma(x)$. Будем полагать, что $f_\gamma(x) = \frac{1}{\sqrt{2\pi}\tau} e^{-\frac{x^2}{2\tau^2}}$, т.е. среди стандартных нормальных случайных величин ε_t с вероятностью γ встречаются нормальные величины с дисперсией $\tau^2 > 1$. Чем больше γ и τ , тем сильнее смесь гауссовских распределений (4), называемая в этом случае загрязненным или засоренным нормальным распределением, отклоняется от нормального.

Однако не только нормальное распределение возникает естественным образом при описании возмущений ε_t . Во многих случаях разумно полагать, что распределение случайных величин ε_t является нормальным, но со случайной дисперсией η (точнее, нормальным является условное распределение ε_t при условии, что случайная дисперсия η приняла какое-то конкретное значение y). Если о дисперсии η дополнительно не делается никаких предположений, кроме существования средней дисперсии $E\eta$, то, считая положение дисперсии η на числовой оси максимально неопределенным, методами теории информации можно получить [25], что безусловное распределение ε_t будет двусторонним экспоненциальным распределением, или распределением Лапласа с плотностью $f(x) = \frac{1}{\sqrt{2}\sigma} e^{-\frac{\sqrt{2}|x|}{\sigma}}$.

Вследствие изложенного было проведено сравнение оценок для перечисленных распределений. Кроме того, в качестве тестовых распределений рассматривались логистическое распределение с плотностью $f(x) = \frac{e^{-x}}{(1+e^{-x})^2}$ и распределение Стьюдента с числом степеней свободы $n = 3$ и $n = 30$ с плотностью $f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} (1 + x^2/n)^{-(n+1)/2}$, где Γ — гамма-функция.

Результаты вычислений асимптотической относительной эффективности М-оценки с ρ -функцией Хьюбера по отношению к оценкам наименьших квадратов и наименьших модулей при различных распределениях ε_t приведены в пятом и шестом столбцах таблицы. Постоянная k в (3) обычно выбирается таким образом (см., например, [12]), чтобы $e(\hat{\alpha}, \tilde{\alpha}) = 0,95$, если ε_t — нормальные случайные величины. Согласно этим рекомендациям $k = 1,345$.

Сравнение оценок посредством вычисления асимптотической относительной эффективности носит асимптотический характер. Чтобы понять, насколько надежно пользоваться асимптотическими результатами, величина $e(\hat{\alpha}, \tilde{\alpha})$ сопоставлялась с относительной эффективностью $e_n(\hat{\alpha}, \tilde{\alpha})$ оценок $\hat{\alpha}$ и $\tilde{\alpha}$, построенных по n наблюдениям X_1, \dots, X_n для n , равного 50, 100 и 500. Для простоты рассматривалась модель (1) с $p = 1$, $U_t = U_{t-1}$, $r = 0$, полагалось $a_{11} = -0,3$, $a_{12} = 0,5$, $D\varepsilon_t = 1$. Величины ε_t моделировались при помощи датчиков псевдослучайных чисел из MatLab. Минимум функции (2) находился итерационным методом наименьших квадратов с весами, веса определялись функцией (3), описание метода и доказательство его сходимости приведено, например, в [12, § 9.1]. Так как параметр a векторный, то для определенности величина $e_n(\hat{\alpha}, \tilde{\alpha})$ определялась как обратное отношение дисперсий оценок координаты a_{11} . Величина $e_n(\hat{\alpha}, \tilde{\alpha})$ оценивалась при помощи компьютерного моделирования. Для $n = 50$, $n = 100$ и $n = 500$ эксперимент проводился по $N = 10^3$ раз, реализации X_1, \dots, X_n длины n процесса X_t моделировались при помощи датчиков псевдослучайных чисел для вероятностных распределений из левого столбца таблицы. Для каждого $i = 1, \dots, N$ вычислялись оценки $\hat{\alpha}_i$ и $\tilde{\alpha}_i$ координаты a_{11} , затем величина $e_n(\hat{\alpha}, \tilde{\alpha})$ оценивалась обратным отношением выборочных дисперсий, т.е. числом $\frac{\sum_{i=1}^N (\hat{\alpha}_i - \bar{\alpha})^2}{\sum_{i=1}^N (\tilde{\alpha}_i - \bar{\alpha})^2}$, где $\bar{\alpha} = N^{-1} \sum_{i=1}^N \tilde{\alpha}_i$, $\bar{\alpha} = N^{-1} \sum_{i=1}^N \hat{\alpha}_i$. Оценки величин $e_{50}(\hat{\alpha}, \tilde{\alpha})$, $e_{100}(\hat{\alpha}, \tilde{\alpha})$ и $e_{500}(\hat{\alpha}, \tilde{\alpha})$, полученные по $N = 10^3$ компьютерным реализациям, приведены во втором, третьем и четвертом столбцах таблицы. Видно, что результаты



1 — кривая $e(\hat{\alpha}, \tilde{\alpha}) = 1$; 2 — кривая $e(\tilde{\alpha}, \alpha^*) = 1$; 3 — кривая $e(\hat{\alpha}, \alpha^*) = 1$; $e(\hat{\alpha}, \tilde{\alpha}) < 1$ на множестве M_1 , $e(\hat{\alpha}, \alpha^*) < 1$ на множестве M_4 , $e(\tilde{\alpha}, \alpha^*) > 1$ на множестве $M_1 \cup M_2$.

моделирования достаточно хорошо согласуются с асимптотическими показателями.

Анализ таблицы показывает, что метод наименьших квадратов предпочтителен только если распределение ε_t является нормальным или практически не отличается от него, как в случае распределения Стьюдента с большим числом степеней свободы, причем даже в этих случаях потеря эффективности М-оценок по отношению к оценке наименьших квадратов невелика. Для логистического распределения, которое статистический критерий хи-квадрат К. Пирсона иногда путает с нормальным, М-оценка лучше оценки наименьших квадратов. Также М-оценка становится лучше оценки наименьших квадратов, если в последовательности нормальных случайных величин ε_t в среднем у каждого члена в три раза увеличить среднеквадратическое отклонение. Таким образом, чем медленнее убывает f на бесконечности (чем “толще” у нее “хвосты”), тем эффективнее М-оценка по отношению к оценке наименьших квадратов. И наконец, М-оценка практически всегда лучше

Асимптотическая эффективность М-оценок по отношению к оценкам квадратов и наименьших модулей

Распределение ε_t	$e_{50}(\hat{\alpha}, \tilde{\alpha})$	$e_{100}(\hat{\alpha}, \tilde{\alpha})$	$e_{500}(\hat{\alpha}, \tilde{\alpha})$	$e(\hat{\alpha}, \tilde{\alpha})$	$e(\hat{\alpha}, \alpha^*)$
Нормальное	0,950	1,492	1,492	0,950	1,492
Логистическое	1,041	1,092	1,085	1,065	1,295
Тьюки ($\gamma = 0,01, \tau = 3$)	1,007	1,003	1,003	1,005	1,481
Тьюки ($\gamma = 0,01, \tau = 10$)	1,358	1,515	1,623	1,833	1,473
Тьюки ($\gamma = 0,10, \tau = 3$)	1,218	1,279	1,355	1,386	1,388
Тьюки ($\gamma = 0,10, \tau = 10$)	4,219	5,452	6,965	7,595	1,322
Стьюдента, 30 степеней свободы	1,202	1,124	1,077	0,976	1,455
Стьюдента, 3 степени свободы	1,267	1,589	1,703	1,927	1,189
Лапласа	1,190	1,286	1,368	1,406	0,703

оценки наименьших модулей, уступая ей лишь в случае экстремальных выбросов загрязненного нормального распределения (см. рисунок) и в случае распределения Лапласа.

Теперь предположим, что ε_t имеют загрязненное нормальное распределение и рассмотрим величины $e(\hat{\alpha}, \tilde{\alpha})$ и $e(\hat{\alpha}, \alpha^*)$ как функции от (τ, γ) . В этом случае наглядное сравнение оценок позволяет получить рисунок, на котором в координатной плоскости (τ, γ) изображены линии уровня единичной асимптотической относительной эффективности, т.е. кривые $e(\hat{\alpha}, \tilde{\alpha}) = 1$ и $e(\hat{\alpha}, \alpha^*) = 1$. Для сравнения на этом же рисунке приведена линия уровня $e(\alpha^*, \tilde{\alpha}) = 1$ асимптотической относительной эффективности оценки наименьших модулей по отношению к оценке наименьших квадратов.

Эти линии представляют собой множество точек (τ, γ) равной эффективности соответствующих оценок и разбивают плоскость (τ, γ) на четыре множества M_i , $i = 1, 2, 3, 4$, в каждом из которых одна из двух сравниваемых оценок всегда предпочтительнее конкурирующей. Оценка наименьших квадратов $\tilde{\alpha}$ на множестве M_1 лучше М-оценки $\hat{\alpha}$, а на множестве $M_1 \cup M_2$ лучше оценки наименьших модулей α^* . Другими словами, оценка $\tilde{\alpha}$ лучше оценки $\hat{\alpha}$, если точка (τ, γ) лежит левее кривой $e(\hat{\alpha}, \tilde{\alpha}) = 1$ и лучше оценки α^* , если точка (τ, γ) лежит левее кривой $e(\tilde{\alpha}, \alpha^*) = 1$. М-оценка $\hat{\alpha}$ лучше оценки наименьших квадратов $\tilde{\alpha}$ на множестве $M_2 \cup M_3 \cup M_4$ и лучше оценки наименьших модулей α^* на множестве $M_1 \cup M_2 \cup M_3$. Оценка наименьших модулей α^* лучше оценки наименьших квадратов $\tilde{\alpha}$ на множестве $M_3 \cup M_4$ и лучше М-оценки $\hat{\alpha}$ на множестве M_4 .

В [20, § 1.2] обсуждается вопрос о том, насколько часто встречаются большие ошибки в данных. Под большими ошибками подразумеваются “ошибки, которые редко, но с большой мощностью порождаются источником отклонений”, например ошибки в перфорации данных или в неправильном размещении десятичных точек, способных менять значения величин на порядок. Как пишут авторы в [20, § 1.2], “сделать общее заключение о том, насколько часто встречаются большие ошибки, по-видимому, нельзя, и в этом ничего удивительного нет, так как слишком многое зависит от конкретных обстоятельств”. Тем не менее из приведенного в [20] обзора литературы можно сделать вывод о том, что 1%–10% загрязнений (т.е. $\gamma \in (0,01; 0,1)$) уровня $\tau \in (3, 10)$ являются нередким на практике. Из рисунка видно, что при этих значениях γ и τ М-оценка практически всегда эффективнее и оценки наименьших квадратов, и оценки наименьших модулей. Таким образом, на практике при оценивании параметров модели пороговой авторегрессии М-оценку, по-видимому, следует предпочесть как оценке наименьших квадратов, так и оценке наименьших модулей.

ПРИЛОЖЕНИЕ

Доказательство теоремы. Основная идея доказательства состоит в том, чтобы аппроксимировать (2) квадратичной функцией, чей минимум имеет асимптотически нормальное распределение, и затем показать, что М-оценка $\hat{\alpha}$ близка к этому минимуму настолько, что имеет то же асимптотическое распределение.

Заметим, что вектор \hat{a} минимизирует (2) тогда и только тогда, когда вектор $\hat{\alpha} = \sqrt{n}(\hat{a} - a)$ является точкой минимума функции

$$\mathcal{L}(\alpha) = \sum_{t=2}^n \left(\rho \left(\varepsilon_t - \alpha^\top Y_t n^{-1/2} \right) - \rho(\varepsilon_t) \right).$$

Установим асимптотическую нормальность $\hat{\alpha}$, откуда будет вытекать утверждение теоремы. Раскладывая ρ в точке ε_t по формуле Тейлора до второй производной включительно, получим

$$\mathcal{L}(\alpha) = -A_n^\top \alpha + 2^{-1} \alpha^\top B_n \alpha + \beta_n(\alpha),$$

где вектор A_n , матрица B_n и остаточный член $\beta_n(\alpha)$ имеют соответственно вид:

$$A_n = n^{-1/2} \sum_{t=2}^n \rho'(\varepsilon_t) Y_t, \quad B_n = n^{-1} \sum_{t=2}^n \rho''(\varepsilon_t) Y_t Y_t^\top,$$

$$\beta_n(\alpha) = (2n)^{-1} \sum_{t=2}^n \left(\rho'' \left(\varepsilon_t - \tau \alpha^\top Y_t n^{-1/2} \right) - \rho''(\varepsilon_t) \right) \alpha^\top Y_t Y_t^\top \alpha, \quad 0 < \tau < 1.$$

Оценим $E|\beta_n(\alpha)|$. Из стационарности ε_t следует, что стационарным будет и Y_t . Поэтому

$$E|\beta_n(\alpha)| \leq (2)^{-1} E \left[\left| \rho'' \left(\varepsilon_1 - \tau \alpha^\top Y_1 n^{-1/2} \right) - \rho''(\varepsilon_1) \right| \alpha^\top Y_1 Y_1^\top \alpha \right].$$

Так как $\sigma^2 < \infty$ влечет $E|Y_1|^2 < \infty$ [26], ρ'' непрерывна почти всюду и ограничена, а случайная величина ε_1 непрерывна, то $E|\beta_n(\alpha)| \rightarrow 0$ при $n \rightarrow \infty$ по теореме Лебега о мажорируемой сходимости [27, с. 204]. Следовательно из неравенства Чебышёва вытекает, что $\beta_n(\alpha) \rightarrow 0$ при $n \rightarrow \infty$ по вероятности.

Последовательности $\rho'(\varepsilon_t)$, Y_t , $Y_t Y_t^\top$ и $\rho''(\varepsilon_t) Y_t Y_t^\top$ являются стационарными и эргодическими как преобразования белого шума ε_t (см. [28, с. 170 и 182]). Поэтому по закону больших чисел для эргодических последовательностей (см. [28, с. 181]) существуют пределы по вероятности:

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{t=2}^n Y_t Y_t^\top = K, \quad \lim_{n \rightarrow \infty} B_n = K E[\rho''(\varepsilon_1)].$$

Обозначим через $\tilde{\alpha}$ минимум функции $-A_n^\top \alpha + 2^{-1} \alpha^\top K E[\rho''(\varepsilon_1)] \alpha$, являющейся квадратичной аппроксимацией функции $\mathcal{L}(\alpha)$. Очевидно, что $\tilde{\alpha} = (K E[\rho''(\varepsilon_1)])^{-1} A_n$.

Найдем асимптотическое распределение $\tilde{\alpha}$. Заметим, что вектор Y_t измерим относительно σ -алгебры \mathfrak{F}_t , а ε_t независима от \mathfrak{F}_t . Отсюда, а также из $E\rho'(\varepsilon_t) = 0$ и свойств условных математических ожиданий (см. [27, § 7.4]) следует, что $E\rho'(\varepsilon_t) Y_t = 0$. Поэтому по центральной предельной теореме для мартингалов (см., например, [29, § 23] и [30, § 5.5]) случайный вектор A_n является асимптотически нормальным с нулевым математическим ожиданием и

ковариационной матрицей $K\mathbf{E}[\rho'(\varepsilon_1)^2]$. Следовательно (см. [31, § 6а.2]), случайная величина $\tilde{\alpha}$ асимптотически нормальна с нулевым математическим ожиданием и ковариационной матрицей $K^{-1} \frac{\mathbf{E}[\rho'(\varepsilon_1)^2]}{(\mathbf{E}[\rho''(\varepsilon_1)])^2}$.

Докажем теперь, что $\tilde{\alpha} - \hat{\alpha} \rightarrow 0$ по вероятности, откуда будет вытекать утверждение теоремы. Из изложенного следует, что

$$\mathcal{L}(\alpha) = -A_n^\top \alpha + 2^{-1} \mathbf{E}[\rho''(\varepsilon_1)] \alpha^\top K \alpha + \gamma_n(\alpha),$$

где $\gamma_n(\alpha) \rightarrow 0$ по вероятности при $n \rightarrow \infty$. Другими словами, последовательность выпуклых случайных функций $\mathcal{L}(\alpha) + A_n^\top \alpha$ стремится по вероятности в каждой точке α к выпуклой функции $2^{-1} \mathbf{E}[\rho''(\varepsilon_1)] \alpha^\top K \alpha$. Поэтому (см. [32]) $\sup_{\alpha \in \mathfrak{K}} |\gamma_n(\alpha)| \rightarrow 0$ по вероятности для любого компакта \mathfrak{K} .

Возьмем шар \mathfrak{B} произвольного радиуса δ с центром $\tilde{\alpha}$. Так как $\tilde{\alpha}$ сходится по распределению, то $\tilde{\alpha}$ стохастически ограничена и, следовательно, существует компакт \mathfrak{K} , содержащий \mathfrak{B} с вероятностью сколь угодно близкой к единице. Поэтому $\Delta = \sup_{\alpha \in \mathfrak{B}} |\gamma_n(\alpha)| \rightarrow 0$ по вероятности.

Рассмотрим поведение $\mathcal{L}(\alpha)$ вне \mathfrak{B} . Пусть $\alpha = \tilde{\alpha} + te$, где e — единичный вектор произвольного направления, $t > \delta$. Обозначим $\alpha^* = \tilde{\alpha} + \delta e$. Так как $\mathcal{L}(\alpha)$ выпукла, то

$$\mathcal{L}(\alpha^*) \leq \delta t^{-1} \mathcal{L}(\alpha) + (1 - \delta t^{-1}) \mathcal{L}(\tilde{\alpha}).$$

Следовательно,

$$\mathcal{L}(\alpha) \geq t\delta^{-1} \mathcal{L}(\alpha^*) - t\delta^{-1} (1 - \delta t^{-1}) \mathcal{L}(\tilde{\alpha}) = t\delta^{-1} (\mathcal{L}(\alpha^*) - \mathcal{L}(\tilde{\alpha})) + \mathcal{L}(\tilde{\alpha}).$$

Так как

$$\mathcal{L}(\alpha^*) = \mathcal{L}(\tilde{\alpha}) + \delta^2 \mathbf{E}[\rho''(\varepsilon_1)] e^T K e + \gamma_n(\alpha^*) - \gamma_n(\tilde{\alpha}),$$

то

$$\inf_{\alpha \notin \mathfrak{B}} \mathcal{L}(\alpha) = \mathcal{L}(\tilde{\alpha}) + \inf_{t > \delta} (t\delta^{-1} (\mathcal{L}(\alpha^*) - \mathcal{L}(\tilde{\alpha}))) \geq \mathcal{L}(\tilde{\alpha}) + \delta^2 \mathbf{E}[\rho''(\varepsilon_1)] e^T K e - 2\Delta.$$

Так как K положительно определена, то $\delta^2 \mathbf{E}[\rho''(\varepsilon_1)] e^T K e > 0$ в силу $0 < \mathbf{E}\rho''(\varepsilon_t) < \infty$. Поэтому с вероятностью, стремящейся к 1, точка минимума $\mathcal{L}(\alpha)$ не может быть вне \mathfrak{B} . Таким образом, $\mathbf{P}\{|\hat{\alpha} - \tilde{\alpha}| > \delta\} \rightarrow 0$ для любого $\delta > 0$ и, следовательно, асимптотическое распределение $\hat{\alpha}$ совпадает с асимптотическим распределением $\tilde{\alpha}$, т.е. $\hat{\alpha}$ асимптотически нормальна с нулевым средним и ковариационной матрицей $K^{-1} \frac{\mathbf{E}[\rho'(\varepsilon_{11})^2]}{(\mathbf{E}[\rho''(\varepsilon_{11})])^2}$. Теорема доказана.

СПИСОК ЛИТЕРАТУРЫ

1. *De Gooijer J.G.* Elements of Nonlinear Time Series Analysis and Forecasting. Cham: Springer, 2017.

2. *Tong H.* Threshold Models in Non-linear Time Series Analysis. N.Y.: Springer, 1983.
3. *Ozaki T.* Time series modeling of neuroscience data. N.Y.: CRC Press, 2012.
4. *Chavas J.-P.* Modeling Population Dynamics: a Quantile Approach // *Math. Biosci.* 2015. V. 262. P. 138–146.
5. *Yang K., Wang D., Li H.* Threshold Autoregression Analysis for Finite-Range Time Series of Counts with an Application on Measles Data // *J. Stat. Comput. Simulat.* 2018. V. 88. No. 3. P. 597–614.
6. *Hamaker E.L., Zhang Z, van der Maas H.L.J.* Using Threshold Autoregressive Models to Study Dyadic Interactions // *Psychometrika.* 2009. V. 74. No. 4. P. 727–745.
7. *Hansen B.E.* Threshold Autoregression in Economics // *Statistics and its interface.* 2011. V. 73. No. 8. P. 563–573.
8. *Bertone E., O' Halloran K., Stewart R.A., de Oliveira G.F.* Medium-term Storage Volume Prediction for Optimum Reservoir Management: A Hybrid Data-Driven Approach // *J. Clean. Prod.* 2017. V. 154. No. 15. P. 353–365.
9. *Pu Shuzhen, Yu Huiling.* Threshold Autoregression Models for Forecasting El Nino Events // *Acta Oceanologica Sinica.* 1990. V. 9. No. 1. P. 61–67.
10. *Kabiri S., Lotfollahzadeh T., Shayesteh M.G., Kalbkhani H.* Modelling and Forecasting of Signal-to-Interference Plus Noise Ratio in Femtocellular Networks Using Logistic Smooth Threshold Autoregressive Model // *IET Signal Process.* 2015. V. 9. No. 1. P. 48–59.
11. Howell Tong's contributions to statistics / ed. Kung-sik Chan, London: World Scientific, 2009.
12. *Maronna R.A., Martin D., Yohai V.* Robust Statistics: Theory and Methods. Chichester: Wiley, 2006.
13. *Поляк Б.Т., Хлебников М.В.* Метод главных компонент: робастные версии // *АиТ.* 2017. № 3. P. 130–148.
Polyak B.T., Khlebnikov M.V. The Method of Principal Components: Robust Versions // *Autom. Remote Control.* 2017. V. 78. No. 3. P. 490–596.
14. *Poljak B.T., Tsyupkin Ja.Z.* Robust Identification // *Automatica.* 1980. V. 16. No. 1. P. 53–63.
15. *Поляк Б.Т., Цыпкин Я.З.* Робастные псевдоградиентные алгоритмы адаптации // *АиТ.* 1980. № 10. P. 91–97.
Polyak B.T., Tsyupkin Ya.Z. Robust Pseudogradient Adaptation Algorithms // *Autom. Remote Control.* V. 41. No. 10. Part 1. P. 1404–1409.
16. *Huber P., Ronchetti E.M.* Robust statistics. Hoboken: Wiley, 2009.
17. *Douc R., Moulines E., Stoffer D.* Nonlinear Time Series: Theory, Methods and Applications with R Examples. Boca Raton: CRC Press, 2014.
18. *Petrucelli J.D., Woolford S.W.* A threshold AR(1) Model // *J. Appl. Probab.* 1984. V. 21. No. 2. P. 270–286.
19. *Li D., Ling S.* On the Least Squares Estimation of Multiple-Regime Threshold Autoregressive Models // *J. Econometrics.* 2012. V. 167. No. 1. P. 240–253.
20. Робастность в статистике. Подход на основе функций влияния / Пер. с англ. Хампель Ф., Рончетти Э., Рауссеу П., Штаэль В. М.: Мир, 1989.
21. *Леман Э.* Теория точечного оценивания. М.: Наука, 1991.
22. *Hettmansperger T.P., McKean J.W.* Robust nonparametric statistical methods. Boca Raton: CRC Press, 2011.

23. Wang L., Wang J. The Limiting Behavior of Least Absolute Deviation Estimators for Threshold Autoregressive Models // J. Multivariate Anal. 2004. V. 89. No. 2. P. 243–260.
24. Королев В.Ю. Вероятностно-статистический анализ хаотических процессов с помощью смешанных гауссовских моделей. М.: Изд-во МГУ, 2008.
25. Мудров В.И., Кушко В.Л. Метод наименьших модулей. М.: Знание, 1971.
26. Qian L. On Maximum Likelihood Estimators for a Threshold Autoregression // J. Statist. Plann. Inference. 1998. V. 75. No. 1. P. 21–46.
27. Ширяев А.Н. Вероятность. М.: Наука, 2011.
28. Stout W.F. Almost Sure Convergence. N.Y.: Acad. Press, 1974.
29. Биллингсли П. Сходимость вероятностных мер. М.: Наука, 1977.
30. Линцер Р.Ш., Ширяев А.Н. Теория мартингалов. М.: Наука, 1986.
31. Rao C.P. Линейные статистические методы и их применения. М.: Наука, 1968.
32. Andersen P.K., Gill R.D. Cox's Regression Model for Counting Processes: a Large Sample Study // Ann. Statist. 1982. V. 10. No. 4. P. 1100–1120.

Статья представлена к публикации членом редколлегии А.В. Назиным.

Поступила в редакцию 14.03.2018

После доработки 26.10.2018

Принята к публикации 08.11.2018