УДК 62.501

А. В. КИТАЕВА

(Томск)

МЕДИАННЫЕ ОЦЕНКИ ПАРАМЕТРОВ КВАДРАТИЧНОГО ТРЕНДА ВРЕМЕННОГО РЯДА

С задачей выделения тренда временного ряда часто приходится сталкиваться при обработке экспериментальных данных, например, в научных исследованиях, при исследовании технических систем. Классический способ решения данной задачи — представление тренда в виде ряда по некоторой системе функций и последующее оценивание коэффициентов этого ряда методом наименьших квадратов (МНК) [1, 2]. Однако полученные таким образом оценки коэффициентов тренда неустойчивы по отношению к так называемым аномальным ошибкам наблюдений, когда отдельные измерения могут содержать очень большие погрешности. В такой ситуации следует отказаться от МНК и использовать робастные оценки параметров тренда [3], к классу которых и относятся рассматриваемые в работе оценки медианного типа.

Возьмем следующую модель наблюдений:

$$x_i = \sum_{k=0}^{2} a_k \psi_k \left(\frac{i}{N}\right) + n_i, \qquad i = \overline{1, N},$$

где шумы n_i — независимые одинаково распределенные случайные величины с дифференцируемой плотностью $p(\cdot)$; $\psi_0 = 1$, $\psi_1(x) = x - 1/2$, $\psi_2(x) = x^2 - x + 1/6$.

Классические медианные оценки параметров линейной регрессии — оценки метода наименьших модулей (ОМНМ) [4]. В данной работе предложены оценки $\widehat{a}_k$ параметров a_k, $k = \overline{0,2}$, аналогичные по своей структуре оценкам медианы, построенные на суммах и разностях наблюдений. Они обладают рядом преимуществ в точности оценивания по сравнению с ОМНМ, к тому же оценки параметров находятся последовательно, что существенно упрощает построение вычислительных процедур.

Будем искать $\widehat{a}_k$, $k = \overline{0,2}$, из уравнений

$$\xi = \frac{1}{N^2} \sum_{k=l}^{\left]\frac{N-1}{2}\right[} f\left(\frac{k}{N}\right) \sum_{i=k+1}^{N-k} \operatorname{sign}\left[x_{i+k} + x_{i-k} - x_{i+l} - x_{i-l} - 2\widehat{a}_2 \frac{k^2 - l^2}{N^2}\right] = 0.$$

$$0 < l < \left]\frac{N-1}{2}\right[; \tag{1}$$

$$\frac{1}{N^2} \sum_{i,j=1}^{N} \frac{i-j}{N} \operatorname{sign}\left[x_i - x_j - \widehat{a}_1 \frac{i-j}{N} - \widehat{a}_2 \left(\frac{i^2 - j^2}{N^2} + \frac{j-i}{N}\right)\right] = 0; \tag{2}$$

$$\frac{1}{N^2} \sum_{i,j=1}^{N} \operatorname{sign}\left[x_i + x_j - 2\widehat{a}_0 - \widehat{a}_1 \left(\frac{i+j}{N} - 1\right) - \widehat{a}_2 \left(\frac{i^2 + j^2}{N^2} - \frac{i+j}{N} + \frac{1}{3}\right)\right] = 0, \tag{3}$$

где $]c[$ обозначает целую часть числа c; $f(\cdot)$ — некоторая неотрицательная весовая функция на $[0, 1/2]$.

Заметим, что если достижение точного равенства в (1)—(3) невозможно, то в качестве оценок берутся те значения $\widehat{a}_k$, для которых отличие от нуля минимально.

Остановимся подробно на изучении асимптотических свойств оценки $\widehat{a}_2$. Отметим, что стандартные методы, используемые для исследования неявно задаваемых оценок, в данном случае неприменимы вследствие нерегулярности функции «sign».

Найдем математическое ожидание при $N \to \infty$:

$$L(\Delta a_2) = \lim_{N \to \infty} \mathscr{M}(\xi) = -2 \int_u^{1/2} f(x)(1 - 2x)\, \varphi\left[2\Delta a_2 (x^2 - u^2)\right] dx,$$

$$\varphi(x) = \int_0^x \int_R p_1(z)\, p_1(z + y)\, dz\, dy, \tag{4}$$

$$p_1(x) = \int_R p(v)\, p(x - v)\, dv, \quad 0 \leqslant u = \lim_{N \to \infty} l/N < 1/2, \quad \Delta a_2 = \widehat{a}_2 - a_2.$$

Теорема 1. Оценка $\widehat{a_2}$ сильно состоятельна.

Для доказательства достаточно показать:

а) сходимость почти наверное $\xi \xrightarrow[N\to\infty]{\text{ПН}} L(\Delta a_2)$, где $L(\Delta a_2)$ — непрерывная в нуле функция;

б) уравнение

$$L(\Delta a_2) = 0 \tag{5}$$

имеет единственное решение в точке $\Delta a_2 = 0$.

Доказательство утверждения п. а основано на сходимости ряда $\sum_{N=1}^{\infty} S_N = \sum_{N=1}^{\infty} \mathscr{M}[\xi - \mathscr{M}(\xi)]^4$ [5]. Нетрудно видеть, что $S_N = O(1/N^2)$ при $N \to \infty$. Подынтегральная функция в (4) сохраняет знак на $[u, 1/2]$, и так как $f(\cdot) \not\equiv 0$, то для удовлетворения (5) необходимо потребовать $\varphi(\cdot) \equiv 0$, что возможно только в случае $\Delta a_2 = 0$. Теорема 1 доказана.

Теорема 2. Величина $\sqrt{N}\Delta a_2$ имеет асимптотическое нормальное распределение с нулевым средним и дисперсией

$$\lambda^2 = \int_u^{1/2} f(x) \int_x^{1-x} \left\{ \frac{1}{3} \int_u^{\min(y,1-y)} f(z)\,dz + 2\left[\int_u^{(x+y)/2} f(z)\,dz + \int_u^{\tilde{u}} f(z)\,dz + \right.\right.$$

$$\left. + \int_u^{\max((y-x)/2,u)} f(z)\,dz - 2 \int_u^{(y+u)/2} f(z)\,dz - 2 \int_u^{\max(u,(y-u)/2)} f(z)\,dz \right] \times \tag{6}$$

$$\times \int_R \left[\int_x^{\infty} \tilde{p}(y)\,dy - \int_{-\infty}^{x} \tilde{p}(y)\,dy \right]^2 p(x)\,dx \Bigg\} dy\,dx \left[4 \int_u^{1/2} f(x)(1-2x)(x^2-u^2)\,dx \int_R p_1^2(v)\,dv \right]^{-2},$$

где

$$\tilde{p}(y) = \int p_1(u)\,p(u+y)\,du;$$

$$\tilde{u} = \max[\min(1/2,\ y+2u,\ 1-y-2u),\ u].$$

Доказательство. Обозначим $\eta = \xi(\Delta a_2) - \xi(0)$ и запишем уравнение (1) в виде

$$\xi(0) + [\eta - \mathscr{M}(\eta)] = -\mathscr{M}(\eta).$$

Разложим $\mathscr{M}(\eta)$ в правой части уравнения в ряд Маклорена по Δa_2, взяв остаточный член в форме Лагранжа. Получим $\sqrt{N}\Delta a_2 = \{\sqrt{N}\xi(0) + \sqrt{N}[\eta - \mathscr{M}(\eta)\}H^{-1}$, где $H \xrightarrow[N\to\infty]{\text{ПН}} -L'(0)$, так как $\Delta a_2 \xrightarrow[N\to\infty]{\text{ПН}} 0$ и $L''(\Delta a_2) < \infty$ в некоторой окрестности нуля. Нетрудно показать, что $\lim_{N\to\infty} N\mathscr{D}[\eta - \mathscr{M}(\eta)] = 0$, следовательно, $\sqrt{N}[\eta - \mathscr{M}(\eta)]$ сходится по вероятности к нулю при $N \to \infty$. По теореме о сходимости [6] получаем, что асимптотическое распределение случайной величины $\sqrt{N}\Delta a_2$ совпадает с асимптотическим распределением величины $-\sqrt{N}\xi(0)[L'(0)]^{-1}$, которое является гауссовым [7]. Очевидно, $\mathscr{M}[\xi(0)] = 0$. Рассмотрим $\lambda^2 = \lim_{N\to\infty} N\mathscr{D}[\xi(0)][L'(0)]^{-2}$:

$$\lim_{N\to\infty} N\mathscr{D}[\xi(0)] = \lim_{N\to\infty} \frac{1}{N^3} \sum_{k,m=l}^{\left]\frac{N-1}{2}\right[} f\left(\frac{k}{N}\right) f\left(\frac{m}{N}\right) \times \tag{7}$$

$$\times \sum_{i=k+1}^{N-k} \sum_{j=m+1}^{N-m} \mathscr{M}\{\text{sign}[(x_{i+k} + x_{i-k} - x_{i+l} - x_{i-l})(x_{j+m} + x_{j-m} - x_{j+l} - x_{j-l})]\}.$$

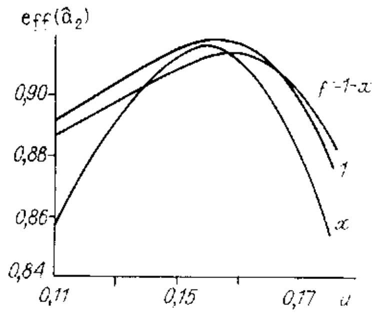
Главную часть в (7) при $N \to \infty$ дают совпадения двух наблюдений в произведениях под знаком «sign». Обозначим

$$S = \mathscr{M}\{\text{sign}[(x_l + x_{l_1} - x_{l_2} - x_{l_3})(x_l + x_{l_4} - x_{l_5} - x_{l_6})]\} =$$

$$= \int_R \left[\int_y^{\infty} \tilde{p}(y)\,dy - \int_{-\infty}^{y} \tilde{p}(y)\,dy \right]^2 p(y)\,dy, \quad l_1 \neq l_2 \neq l_3 \neq l_4 \neq l_5 \neq l_6 \neq l.$$

Так как l фиксировано, то возможен также случай

$$\mathscr{M}\left\{\operatorname{sign}\left[\left(x_{l_1}+x_{l_2}-x_l-x_p\right)\times\right.\right.$$
$$\left.\left.\times\left(x_{l_3}+x_{l_4}-x_l-x_p\right)\right]\right\}=1/3.$$



Рассматривая всевозможные варианты совпадения наблюдений в произведениях в (7) (всего их будет 15), получим соответствующие ограничения на область изменения индексов, что и отражено в пределах интегрирования в (6). Например, пусть $i+l=j-l$, тогда $m+1\leqslant j=i+2l\leqslant \leqslant N-m$, отсюда $m\leqslant\min(i+2l-1,\ N-i-2l)$. Учитывая, что, с другой стороны, $l\leqslant m\leqslant\left]\frac{N-1}{2}\right[$, найдем $l\leqslant m\leqslant \leqslant\max\left\{\min\left(i+2l-1,\ \left]\frac{N-1}{2}\right[,\ N-i-2l\right),\ l\right\}$. Таким образом, получаем в (7) слагаемое

$$S\int\limits_u^{1/2}\int\limits_x^{1-x}\int\limits_u^{\tilde{u}} f(x)\,f(z)\,dz\,dy\,dx.$$

Теорема 2 доказана.

Провести оптимизацию по функции $f(\cdot)$ не удается. Для гауссова случая на рисунке построены графики зависимости эффективности оценки $\widehat{a}_2$ $\mathrm{eff}(\widehat{a}_2)$ от параметра u в окрестности точки экстремума при $f(x)=1$, $f(x)=x$, $f(x)=1-x$.

Что касается оценок $\widehat{a}_0$ и $\widehat{a}_1$, то их асимптотические свойства можно исследовать аналогичными методами. Ортогональность полиномов $\{\psi_k(\cdot),\ k=\overline{0,2}\}$ на $[0,\ 1]$ обеспечивает независимость асимптотического распределения оценок $\widehat{a}_1$ и $\widehat{a}_0$ от распределения оценок $\widehat{a}_2$ и $\widehat{a}_1$, $\widehat{a}_2$ соответственно.

С другой стороны, можно показать, что оценки $\widehat{a}_k^{(1)}$, определяемые из системы

$$\frac{1}{N^2}\sum_{i,j=1}^{N}\left[\psi_m\left(\frac{i}{N}\right)-\psi_m\left(\frac{j}{N}\right)\right]\operatorname{sign}\left[x_i-x_j-\sum_k\widehat{a}_k^{(1)}\left[\psi_k\left(\frac{i}{N}\right)-\psi_k\left(\frac{j}{N}\right)\right]\right]=0,$$
$$m=1,\ 2,$$

эквивалентны ранговым оценкам параметров линейной регрессии с весовой функцией Вилкоксона, рассмотренным в [8]; оценка $\widehat{a}_0$ асимптотически эквивалентна оценке Ходжеса — Лемана параметра сдвига [9] (при этом необходимо потребовать, чтобы медиана распределения шумов находилась в нуле). Оценки $\widehat{a}_0$, $\widehat{a}_1$ сильно состоятельны, а случайные величины $\sqrt{N}(\widehat{a}_1-a_1)$, $\sqrt{N}(\widehat{a}_0-a_0)$ имеют асимптотическое нормальное распределение с параметрами $\left(0,\ \left(\int\limits_R p^2(x)\,dx\right)^{-2}\Big/12\right)$.

Таким образом, оценки $\widehat{a}_k$, $k=\overline{0,2}$, имеют высокую относительную асимптотическую эффективность (по сравнению с оценками МНК) в отличие от ОМНМ, асимптотическая дисперсия которых полностью определяется поведением плотности распределения шумов в нуле, и для гауссова случая асимптотическая эффективность оценок $\widehat{a}_0(\widehat{a}_1)$ и $\widehat{a}_2$ соответственно в 1,5 и 1,4 раза больше асимптотической эффективности ОМНМ.

СПИСОК ЛИТЕРАТУРЫ

1. **Кендалл М., Стьюарт А.** Многомерный статистический анализ.— М.: Наука, 1976.
2. **Андерсон Т.** Статистический анализ временных рядов.— М.: Мир, 1976.
3. **Хьюбер П.** Робастность в статистике.— М.: Мир, 1984.
4. **Huber P.-J.** The place of the L_1-norm in robust estimation // Comput. Statist. and Data Anal.— 1987.— 5.— P. 255.
5. **Боровков А. А.** Теория вероятностей.— М.: Наука, 1986.
6. **Крамер Г.** Математические методы статистики.— М.: Мир, 1975.
7. **Godwin H. J., Zaremba S. K.** A central limit theorem for partly dependent variables // Ann. Math. St.— 1961.— 32.— P. 677.
8. **Jaeckel L. A.** Estimating regression coefficients by minimizing dispersion of the residuals // Ann. Math. St.— 1972.— 43.— P. 1449.
9. **Hodges J. L., Lehmann E. L.** Estimates of location based on rank tests // Ann. Math. St.— 1963.— 34.— P. 598.

Поступило в редакцию 9 марта 1988 г.