

РОБАСТНОЕ ОЦЕНИВАНИЕ КОЭФФИЦИЕНТОВ АВТОРЕГРЕССИИ В УСЛОВИЯХ «ВЫБРОСОВ» И «ПРОПУСКОВ»

Ю. С. Харин, В. А. Волошко

Белорусский государственный университет, Минск, Беларусь
E-mail: Kharin@bsu.by, ValeraVoloshko@yandex.ru

Рассматривается задача статистического оценивания коэффициентов гауссовского авторегрессионного временного ряда порядка p , наблюдаемого с пропущенными значениями и со случайными «выбросами», имеющими симметричное распределение вероятностей. Разработан новый метод параметрического робастного оценивания значений корреляционной функции и коэффициентов авторегрессионного временного ряда. Исследованы асимптотические свойства оценок. Представлены численные результаты сравнения предлагаемого метода с оценками Хьюбера и МНК – оценками.

Ключевые слова: временной ряд, стационарность, процессы авторегрессии, оценивание параметров, искажения, робастность.

ВВЕДЕНИЕ

В приложениях часто в качестве вероятностной модели наблюдаемого временного ряда используется модель $AR(p)$ гауссовского авторегрессионного временного ряда [1]. При этом возникает задача статистического оценивания неизвестных параметров этой модели. Ввиду того, что на практике наблюдаемый ряд обычно содержит некоторые искажения [2–5], классические оценки параметров, не учитывающие искажения, становятся несостоительными и не могут эффективно применяться. Поэтому необходимо строить оценки, устойчивые (робастные) к искажениям.

В данной статье рассматривается случай двух одновременно влияющих типов искажений: аддитивных «выбросов» и «пропусков». Дадим краткий обзор результатов по робастному оцениванию параметров $AR(p)$. Основной подход заключается в робастном оценивании ковариационной функции и нахождении оценок коэффициентов авторегрессии при решении уравнений Юла – Уокера. Известны следующие робастные оценки ковариаций: оценка Хьюбера [2], медианный коэффициент корреляции и его обобщения [6], оценки, основанные на непараметрических мерах [6], на отсеивании «выбросов», на робастных ковариациях [2] и др. Отличительной особенностью данной статьи является использование новых (параметрических) робастных оценок корреляционной функции с использованием специальных свойств распределения вероятностей Коши.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ И ПОСТАНОВКА ЗАДАЧИ

Модель $AR(p)$ гауссовского авторегрессионного временного ряда y_t задается на вероятностном пространстве (Ω, \mathcal{F}, P) стохастическим разностным уравнением порядка $p \in N$ [1]:

$$y_t + b_1 y_{t-1} + \dots + b_p y_{t-p} = \xi_t, t \in Z, \quad (1)$$

где $b = (b_1, \dots, b_p)' \in R$ – вектор-столбец неизвестных коэффициентов авторегрессии, удовлетворяющий условию стационарности [1], $\{\xi_t : t \in Z\}$ – последовательность независимых в совокупности одинаково распределенных гауссовых случайных величин с законом распределения $L\{\xi_t\} = N(0, \sigma^2), 0 < \sigma^2 < +\infty$ – неизвестная дисперсия инновационного процесса.

Временной ряд с “выбросами” z_t имеет вид:

$$z_t = (1 - \eta_t)y_t + \eta_t v_t, t \in Z, \quad (2)$$

где $\{\eta_t \in \{0,1\}\}$ – независимые в совокупности случайные величины Бернулли,

$$P\{\eta_t = 1\} = 1 - P\{\eta_t = 0\} = \varepsilon, \quad (3)$$

$0 \leq \varepsilon \ll 1$ – достаточно малая вероятность появления “выброса”, $\{v_t \in R\}$ – независимые в совокупности одинаково распределенные случайные величины с некоторой неизвестной симметричной функцией распределения $H(v) = 1 - H(-v)$, нулевым средним $E\{v_t\} = 0$ и дисперсией $D\{v_t\} \gg D\{y_t\}$.

Искажения, порождаемые “пропусками”, условимся характеризовать при помощи неслучайной двоичной индексной последовательности

$$\alpha_t = \{1, \text{ если } z_t \text{ наблюдается}; 0, \text{ если } z_t \text{ пропущено}\}, t \in Z. \quad (4)$$

Обозначим $\{t_k\}$ – упорядоченную в порядке возрастания последовательность моментов времени, для которых наблюдения доступны; $t_- = \min t_k$, $t_+ = \max t_k$ (без потери общности $t_- = 1$, $t_+ = T$). Тогда наблюдаемый временной ряд представим в виде

$$x_k = z_{t_k}, k = 1, \dots, K,$$

где K – число наблюдений. Задача заключается в статистическом оценивании коэффициентов b по наблюдениям $\{x_k\}$.

ПОСТРОЕНИЕ РОБАСТНЫХ ОЦЕНОК КОЭФФИЦИЕНТОВ АВТОРЕГРЕССИИ

Обозначим корреляционную функцию временного ряда y_t :

$$\theta_t = \text{Corr}\{y_t, y_{t+\tau}\}, \tau \in Z. \quad (5)$$

Уравнения Юла – Уокера для модели (1) имеют вид:

$$\begin{cases} \sigma_0 + b_1 \sigma_1 + \dots + b_p \sigma_p = \sigma^2, \\ \sigma_{-1} + b_1 \sigma_0 + \dots + b_p \sigma_{p-1} = 0, \\ \dots \\ \sigma_{-p} + b_1 \sigma_{-1-p} + \dots + b_p \sigma_0 = 0, \end{cases}$$

где $\sigma_t = \text{Cov}\{y_t, y_{t+\tau}\}$. Заметим, что решение $\{b_i\}$ этой системы уравнений не зависит от σ , поэтому вместо ковариации σ_t можно подставить коэффициент корреляции θ_t и без потери общности полагать $\sigma=1$. Если вместо коэффициентов корреляции $\{\theta_t\}$ подставить их оценки $\{\hat{\theta}_t\}$, то решением будет оценка $\hat{b} = (\hat{b}_i)$ вектора коэффициентов авторегрессии.

Пусть $\text{med}\{z_1, \dots, z_T\}$ – выборочная медиана. Согласно [2], робастная оценка коэффициента корреляции θ_t может быть вычислена при наличии “пропусков” с помощью следующего алгоритма.

1. Вычисляем вспомогательные коэффициенты:

$$a = 1/\text{med}\{|z_t| : o_t = 1, t = 1, \dots, T - \tau\}, \quad b = 1/\text{med}\{|z_t| : o_t = 1, t = 1 + \tau, \dots, T\}.$$

2. Находим

$$S_+ = \text{med}\{|az_t + bz_{t+\tau}| : o_t = o_{t+\tau} = 1, t = 1, \dots, T - \tau\}, \quad S_- = \text{med}\{|az_t - bz_{t+\tau}| : o_t = o_{t+\tau} = 1, t = 1, \dots, T - \tau\}.$$

3. Вычисляем оценку Хьюбера:

$$\hat{\theta}_\tau = (S_+^2 - S_-^2) / (S_+^2 + S_-^2). \quad (6)$$

Для построения (параметрической) робастной оценки $\hat{\theta}_\tau$ воспользуемся специальным свойством распределения Коши.

Лемма 1. Если $y_t, t \in Z$, – стационарный гауссовский временной ряд с нулевым средним и некоторой корреляционной функцией (5), то для любого $\tau \neq 0$ отношение случайных величин $\zeta = y_\tau / y_{\tau+\tau}$ имеет распределение Коши $C(\theta_\tau, \sqrt{1-\theta_\tau^2})$.

Доказательство. По условию имеем:

$$L\left(\begin{pmatrix} y_\tau \\ y_{\tau+\tau} \end{pmatrix}\right) = N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \theta_\tau \sigma^2 \\ \theta_\tau \sigma^2 & \sigma^2 \end{pmatrix}\right).$$

В силу известного регрессионного свойства многомерного нормального распределения [7] справедливо представление:

$$y_\tau = \theta_\tau y_{\tau+\tau} + \sqrt{1-\theta_\tau^2} \cdot \sigma \cdot \eta_\tau,$$

где η_τ не зависит от $y_{\tau+\tau}$ и имеет стандартное распределение $N_1(0, 1)$. Следовательно, $\zeta = \theta_\tau + \sqrt{1-\theta_\tau^2} \cdot \eta_\tau / (y_{\tau+\tau}/\sigma)$, где случайная величина $\zeta_0 = \eta_\tau / (y_{\tau+\tau}/\sigma)$ имеет стандартное распределение Коши $C(0, 1)$, поэтому согласно свойствам этого распределения [8] линейное преобразование $\zeta = \theta_\tau + \sqrt{1-\theta_\tau^2} \zeta_0$ приводит к распределению $C(\theta_\tau, \sqrt{1-\theta_\tau^2})$. ■

Лемма 2. Если $\phi: R^2 \rightarrow R$ – ограниченная, нечетная по обеим переменным функция, а z_t – временной ряд, определяемый моделью (1) – (3), то

$$E\{\phi(z_\tau, z_{\tau+\tau})\} = (1-\varepsilon)^2 E\{\phi(y_\tau, y_{\tau+\tau})\}. \quad (7)$$

Доказательство. В силу (2) по формуле полного математического ожидания

$$E\{\phi(z_\tau, z_{\tau+\tau})\} = (1-\varepsilon)^2 E\{\phi(y_\tau, y_{\tau+\tau})\} + \varepsilon(1-\varepsilon)(E\{\phi(y_\tau, v_{\tau+\tau})\} + E\{\phi(v_\tau, y_{\tau+\tau})\}) + \varepsilon^2 E\{\phi(v_\tau, v_{\tau+\tau})\}.$$

Так как y_t – стационарный гауссовский процесс с симметричным относительно $y=0$ распределением $L\{y_t\} = N_1(0, D\{y_t\})$, а v_t также имеет симметричное относительно $v=0$ распределение и $\{y_t\}, \{v_t\}$ – независимы, то в силу нечетности $\phi(\cdot)$ все математические ожидания, кроме первого, в этом выражении равны нулю. ■

Выберем функцию $\phi(\cdot)$ в (7) следующим специальным образом:

$$\phi(z_\tau, z_{\tau+\tau}) = \psi(z_\tau / z_{\tau+\tau}), \quad (8)$$

где $\psi(\cdot)$ – некоторая пока произвольная ограниченная, нечетная функция: $|\psi(u)| \leq c_0 < +\infty$, $\psi(-u) = -\psi(u)$, $u \in R$. С помощью этой функции и Леммы 1 построим ограниченную, нечетную функцию $[-1, 1] \rightarrow R$:

$$f_\psi(\theta_\tau) := E\{\psi(\zeta)\} = \frac{\sqrt{1-\theta_\tau^2}}{\pi} \int_{-\infty}^{+\infty} \frac{\psi(z)}{1-\theta_\tau^2 + (z-\theta_\tau)^2} dz = \frac{4\theta_\tau \sqrt{1-\theta_\tau^2}}{\pi} \int_0^{+\infty} \frac{z\psi(z)}{z^4 - 2z^2(2\theta_\tau^2 - 1) + 1} dz. \quad (9)$$

Теперь возьмем в качестве оценки для величины $E\{\psi(z_\tau / z_{\tau+\tau})\} = (1-\varepsilon)^2 f_\psi(\theta_\tau)$ статистику

$$\hat{E}_\varepsilon = \sum_{t=1}^{T-1} \psi(z_t / z_{t+\tau}) o_t o_{t+\tau} / \sum_{t=1}^{T-1} o_t o_{t+\tau}, \quad (10)$$

из которой получаем соответствующую оценку коэффициента корреляции:

$$\hat{\theta}_\varepsilon = f_\psi^{-1}(\hat{E}_\varepsilon / (1 - \varepsilon)^2). \quad (11)$$

Выбирая различные функции $\psi(\cdot)$ в (11), можно получить семейство оценок для θ_ε :

s – оценка: $\psi(x) = sign(x), f_\psi(y) := E{sign(\zeta)} = \int_R^{\sqrt{1-y^2}} sign(x) dx = \frac{2}{\pi} \arcsin y,$

$$\hat{\theta}_\varepsilon = \sin\left(\frac{\pi}{2(1-\varepsilon)^2} \sum_{t=1}^{T-1} sign(z_t / z_{t+\tau}) o_t o_{t+\tau} / \sum_{t=1}^{T-1} o_t o_{t+\tau}\right); \quad (12)$$

a – оценка: $\psi(x) = \arctan(x), f_\psi(y) := E{\arctan(\zeta)} = \int_R^{\sqrt{1-y^2}} \arctan x dx = \frac{1}{2} \arcsin y,$

$$\hat{\theta}_\varepsilon = \sin\left(\frac{2}{(1-\varepsilon)^2} \sum_{t=1}^{T-1} \arctan(z_t / z_{t+\tau}) o_t o_{t+\tau} / \sum_{t=1}^{T-1} o_t o_{t+\tau}\right).$$

Однако оценка (11) использует долю “выбросов” ε как априорную информацию о модели. Это ограничивает ее применение на практике, т.к. такая информация редко бывает заранее известной.

Укажем способ статистического оценивания ε . Будем дополнительно предполагать, что “выбросы” $\{v_t\}$ имеют гауссовское распределение: $L\{v_t\} = N_1(0, \beta_1 \sigma^2)$, где $\beta_1 > 0$ – коэффициент, определяющий уровень “выбросов”. Тогда распределение наблюдений $\{z_t\}$ является смесью нормальных распределений с нулевыми средними и дисперсиями

$$\sigma_1^2 = D\{v_t\} = \sigma^2 \beta_1 \quad \text{и} \quad \sigma_2^2 = D\{y_t\} = \sigma^2 \beta_2, \quad \text{где} \quad \beta_2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |B(e^{ix})|^{-2} dx. \quad \text{При этом}$$

соответствующая характеристическая функция имеет вид:
 $f_z(\lambda) = \varepsilon \exp\{-0.5\sigma_1^2\lambda^2\} + (1-\varepsilon)\exp\{-0.5\sigma_2^2\lambda^2\}$. Если $\beta_1 \gg \beta_2$, то при достаточно больших значениях λ можно использовать приближение

$$f_z(\lambda) \approx (1-\varepsilon)\exp\{-0.5\sigma_2^2\lambda^2\}. \quad (13)$$

Это приближение тем точнее, чем больше уровень “выбросов” β_1 . Далее, используя (13), имеем приближение: $f_z(\lambda_2)/f_z(\lambda_1) = \exp\{-0.5\sigma_2^2(\lambda_2^2 - \lambda_1^2)\}$. Возьмем некоторые конкретные значения λ_1 и λ_2 . Тогда в левой части соотношения можно статистически оценить отдельно числитель и знаменатель, и затем из правой части получить оценку $\hat{\sigma}_2$. Подставив ее вместо σ_2 в формулу (13), получим равенство, из которого и найдем оценку $\hat{\varepsilon}$. Таким образом, алгоритм оценивания ε имеет вид:

- 1) выбираем λ_1 и λ_2 ;
- 2) строим выборочную характеристическую функцию и вычисляем

$$\hat{f}_z(\lambda_i) = \sum_{t=1}^T o_t \cos\{\lambda_i z_t\} / \sum_{t=1}^T o_t;$$

- 3) строим оценку $\hat{\sigma}_2^2$: $\hat{\sigma}_2^2 = 2(\lambda_1^2 - \lambda_2^2)^{-1} \ln \hat{f}_z(\lambda_2) / \hat{f}_z(\lambda_1);$
- 4) в формуле (13) делаем подстановки: $\sigma_2 \rightarrow \hat{\sigma}_2, \lambda \rightarrow \hat{\sigma}_2^{-1}$ и вычисляем

$$\hat{\varepsilon} = 1 - \sqrt{e} \cdot \hat{f}_z(\hat{\sigma}_2^{-1}), \text{ где } \hat{f}_z(\hat{\sigma}_2^{-1}) = \sum_{t=1}^T o_t \cos(z_t/\hat{\sigma}_2) / \sum_{t=1}^T o_t.$$

В этом алгоритме остается одна неопределенность – выбор λ_1 и λ_2 . В численных экспериментах хорошие результаты показал алгоритм со следующим подбором этих параметров: $\lambda_1^2 = \sum_{t=1}^T o_t / \sum_{t=1}^T o_t z_t^2 \approx (\varepsilon \sigma_1^2 + (1-\varepsilon) \sigma_2^2)^{-1}$, $\lambda_2 = 2\lambda_1$.

Таким образом, нами построен следующий алгоритм оценивания вектора $b = (b_i)$:

Шаг 1) Если неизвестна доля “выбросов”, строим её оценку $\hat{\varepsilon}$.

Шаг 2) Согласно (12) строим S – оценки коэффициентов корреляции $\hat{\theta} = (\hat{\theta}_{ij})$, подставляя в них либо ε (если известно), либо $\hat{\varepsilon}$.

Шаг 3) Строим оценку матрицы системы Юла – Уокера: $\hat{\Xi}_y = \hat{\theta}_{i-j}, \hat{\theta}_{00} = 1, i, j = \overline{0, p}$.

Шаг 4) Вычисляем обратную матрицу и оценки параметров авторегрессии:

$$\hat{b}_i = (\hat{\Xi}_y^{-1})_{i0} / (\hat{\Xi}_y^{-1})_{00}, i = \overline{1, p}.$$

Шаг 5) Находим численным методом наибольший по модулю корень λ_{\max} характеристического многочлена $\hat{B}(\lambda) = \lambda^p + \hat{b}_1 \lambda^{p-1} + \dots + \hat{b}_p$ и, если нарушено условие стационарности и $|\lambda_{\max}| \geq 1$, нормируем оценки следующим образом: $\hat{b}_i \rightarrow \hat{b}_i / (|\lambda_{\max}| + \Delta)^i$, где $\Delta = 0.1 / |\lambda_{\max}|$. Этим мы масштабируем корни $\hat{B}(\lambda)$, поместив их в пределы единичного круга $|\lambda| < 1$.

СВОЙСТВА РОБАСТНЫХ ОЦЕНОК

Прежде чем привести основной результат, докажем две вспомогательные леммы.

Лемма 3. Если $\tau \in N, f(\cdot): R^{\tau+1} \rightarrow R$ – произвольная борелевская функция, а $\{z_t : t \in N\}$ – произвольная стационарная случайная последовательность, обладающая сильным перемешиванием с коэффициентом $\alpha_z(n)$, то случайная последовательность $s_t = f(z_t, \dots, z_{t+\tau})$ обладает сильным перемешиванием с коэффициентом

$$\alpha_s(n) \leq \begin{cases} 1/4, & 0 \leq n < \tau, \\ \alpha_z(n-\tau), & n \geq \tau. \end{cases}$$

Доказательство. По определению сильного перемешивания [9, 10]

$$\alpha_s(n) = \sup_{t \geq 1} \sup_{A \in G_t^{\sigma}, B \in G_{t+n}^{\sigma}} |P(AB) - P(A)P(B)|,$$

где $G_a^{\sigma} = \sigma\{s_t : t \in [a, b]\}$ – σ – алгебра, порожденная траекторией s_t на отрезке $[a, b]$. Пусть $F_a^b = \sigma\{z_t : t \in [a, b]\}$. Тогда по построению $G_t^{\sigma} \subset F_1^{t+\tau}, G_{t+n}^{\sigma} \subset F_{t+n}^{t+\tau}$, так что при $n \geq \tau$ справедливо неравенство: $\alpha_s(n) \leq \alpha_z(n-\tau)$. Так как всегда $|P(AB) - P(A)P(B)| \leq 1/4$, то $\alpha_s(n) \leq 1/4$ при $n > \tau$. ■

Лемма 4. Если случайная последовательность y_t обладает сильным перемешиванием с коэффициентом $\alpha_y(n)$, $\{\mu_t\}$ – независимые одинаково распределенные случайные величины, не зависящие от $\{y_t\}$, а $g(\cdot)$ – произвольная борелевская функция, то

$z_t = g(y_t, \mu_t)$ также обладает сильным перемешиванием с коэффициентом $\alpha_s(n) \leq \alpha_y(n), n \in N$.

Доказательство. Указанный результат следует из Теоремы 5.2 [9]. ■

Теорема 1. Если имеет место модель (1) – (4) и функция $\psi(\cdot)$ такова, что функция $f_\psi(\cdot)$, определяемая (9), имеет непрерывную обратную функцию $f_\psi^{-1}(\cdot)$, и $N_{\tau\tau} = \sum_{t=1}^{\tau-\tau} o_t o_{t+\tau} \rightarrow \infty$ при $T \rightarrow \infty$, то оценка (11) является состоятельной оценкой коэффициента корреляции θ_τ :

$$\hat{\theta}_\tau \xrightarrow{P} \theta_\tau. \quad (14)$$

Доказательство. В силу непрерывности $f_\psi^{-1}(\cdot)$ и свойства сходимости по вероятности функционально преобразованной сходящейся последовательности достаточно доказать справедливость следующего закона больших чисел:

$$S_T = \sum_{t=1}^{\tau-\tau} \psi(z_t / z_{t+\tau}) o_t o_{t+\tau} / N_{\tau\tau} \xrightarrow{P} E\{\psi(z_t / z_{t+\tau})\}. \quad (15)$$

Действительно, в силу Леммы 2, (8), (9), $E\{\psi(z_t / z_{t+\tau})\} = (1 - \varepsilon)^2 f_\psi(\theta_\tau)$, и для получения (14) достаточно обратить (15).

Для доказательства (15) проверим условие Маркова [11]. Исследуем дисперсию

$$D\{S_T\} = \frac{1}{N_{\tau\tau}^2} \left(\sum_{t=1}^{\tau-\tau} o_t o_{t+\tau} \sigma_\psi(0) + 2 \sum_{t=1}^{\tau-\tau-1} \sum_{t'=t+1}^{\tau-\tau} o_t o_{t+\tau} o_{t'} o_{t'+\tau} \sigma_\psi(t' - t) \right),$$

где $\sigma_\psi(t' - t) = \text{Cov}\{\psi(z_t / z_{t+\tau}), \psi(z_{t'}, z_{t'+\tau})\}$. Так как $|\psi(u)| \leq c_0 < +\infty$, то справедлива оценка сверху:

$$\begin{aligned} D\{S_T\} &\leq \frac{1}{N_{\tau\tau}^2} \left(N_{\tau\tau} \cdot c_0^2 + 2 \sum_{t=1}^{\tau-\tau-1} o_t o_{t+\tau} \sum_{t'=t+1}^{\tau-\tau} |\sigma_\psi(t' - t)| \right) = \\ &= \frac{1}{N_{\tau\tau}^2} \left(N_{\tau\tau} c_0^2 + 2 \sum_{t=1}^{\tau-\tau-1} o_t o_{t+\tau} \sum_{i=1}^{\tau-\tau} |\sigma_\psi(i)| \right). \end{aligned} \quad (16)$$

В силу Леммы 1.2.1 [12] $|\sigma_\psi(t' - t)| \leq 4c_0^2 \alpha_s(|t' - t|)$, где $\alpha_s(n), n \in N$, – коэффициент перемешивания [12] для стационарной случайной последовательности $s_i = \psi(z_i / z_{i+\tau})$. Согласно Лемме 3, если $\alpha_s(n)$ – коэффициент перемешивания для стационарной случайной последовательности z_t , то

$$\alpha_s(n) \leq \begin{cases} 1/4, & 0 \leq n < \tau, \\ \alpha_y(n - \tau), & n \geq \tau. \end{cases}$$

Наконец, в силу Леммы 4 $\alpha_s(n) \leq \alpha_y(n)$, где $\alpha_y(n)$ – коэффициент перемешивания для $AR(p)$ – временного ряда (1).

Согласно [10], для $AR(p)$ – временного ряда (1), удовлетворяющего условию стационарности, справедлива экспоненциальная асимптотическая оценка коэффициента перемешивания (при $n \rightarrow +\infty$): $\alpha_y(n) = O(e^{-bn})$, где $b > 0$ – некоторая константа.

Объединяя приведенные выше оценки и подставляя результат в (16), получаем искомую оценку дисперсии: $D\{S_T\} \leq (c_0^2 + 2 \cdot S) / N_{\tau\tau}$, где ряд $S = \sum_{i=1}^{+\infty} |\sigma_\psi(i)| \leq c_* < +\infty$.

сходится в силу установленной асимптотики. В результате находим $D\{S_{\tau}\} \leq (c_0^2 + 2 \cdot c_s) / N_{\tau}$, что и означает справедливость (15), влекущую (14). ■

Исследуем свойства s – оценки (12). Для произвольной последовательности $f: \mathbb{Z} \rightarrow \mathbb{R}$ обозначим $\Sigma^f = \sum_{s \in \mathbb{Z}} f(s)$.

Лемма 5. Пусть $a_t = sign(z_t z_{t+\tau})$, $u_t = sign(y_t y_{t+\tau})$, $\sigma(s) = cov\{a_t, a_{t+s}\}$, $\rho(s) = cov\{u_t, u_{t+s}\}$. Тогда справедливо неравенство:

$$\Sigma^\sigma \leq \varepsilon(2-\varepsilon)(1+3(1-\varepsilon)^2) + (1-\varepsilon)^4 \Sigma^\rho.$$

Доказательство. Пусть $s > 0, s \neq \tau$. Согласно (2) $\eta_t = 1$, если z_t – “выброс”. Очевидно $E\{a_t | \eta_t + \eta_{t+\tau} > 0\} = 0$, т.е. если произошел “выброс” в момент t или $t+\tau$. Легко видеть также, что при таком условии имеет место условная независимость a_t от a_{t+s} . Обозначим событие $\Omega_t = \{\eta_t + \eta_{t+\tau} > 0\}$. Тогда по формуле полного матожидания $E\{*\} = P(\Omega_t)E\{*\mid \Omega_t\} + P(\Omega_{t+s})E\{*\mid \Omega_{t+s}\} + P(\bar{\Omega}_t \bar{\Omega}_{t+s})E\{*\mid \bar{\Omega}_t \bar{\Omega}_{t+s}\} - P(\Omega_t \Omega_{t+s})E\{*\mid \Omega_t \Omega_{t+s}\}$, где \bar{A} значит дополнение A . Вместо $*$ подставим $a_t a_{t+s}$: $E\{a_t a_{t+s} | \Omega_t\} = E\{a_t | \Omega_t\}E\{a_{t+s} | \Omega_t\} = 0 = E\{a_t a_{t+s} | \Omega_{t+s}\} = E\{a_t a_{t+s} | \Omega_t \Omega_{t+s}\}$. Тогда $E\{a_t a_{t+s}\} = (1-\varepsilon)^4 E\{a_t a_{t+s} | \bar{\Omega}_t \bar{\Omega}_{t+s}\} = \sigma(s) + E^2\{a_t\}$. Поскольку a_t и Ω_{t+s} – независимы и $E\{a_t | \Omega_t\} = 0$, имеем $E\{a_t\} = P(\bar{\Omega}_t)E\{a_t | \bar{\Omega}_t\} = P(\bar{\Omega}_t)E\{a_t | \bar{\Omega}_t \bar{\Omega}_{t+s}\} = P(\bar{\Omega}_{t+s})E\{a_t | \bar{\Omega}_t \bar{\Omega}_{t+s}\}$. Получаем $\sigma(s) = (1-\varepsilon)^4 cov\{a_t, a_{t+s} | \bar{\Omega}_t \bar{\Omega}_{t+s}\} = (1-\varepsilon)^4 \rho(s)$.

При $s=0$ имеем $\sigma(0) = D\{a_t\} = 1 - E^2\{a_t\} = 1 - (1-\varepsilon)^4 E^2\{u_t\} = 1 + (1-\varepsilon)^4 (\rho(0) - 1)$. Наконец $\sigma(\tau) = E\{a_t a_{t+\tau}\} - E^2\{a_t\} = E\{sign(z_t z_{t+2\tau})\} - (1-\varepsilon)^4 E^2\{u_t\} = (1-\varepsilon)^2 E\{u_t u_{t+2\tau}\} - (1-\varepsilon)^4 E^2\{u_t\} = ((1-\varepsilon)^2 - (1-\varepsilon)^4) E\{u_t u_{t+2\tau}\} + (1-\varepsilon)^4 \rho(\tau)$. Тогда $\sigma(0) + 2\sigma(\tau) - (\rho(0) + 2\rho(\tau))(1-\varepsilon)^4 < 1 + 2(1-\varepsilon)^2 - 3(1-\varepsilon)^4 = \varepsilon(2-\varepsilon)(1+3(1-\varepsilon)^2)$. Учитывая, что $\sigma(s) = (1-\varepsilon)^4 \rho(s)$ для $s \neq 0, \pm\tau$, получаем требуемое неравенство для сумм рядов. ■

Теорема 2. Если имеет место модель (1) – (4) и $N_{\tau} \xrightarrow[T \rightarrow \infty]{} +\infty$, то s – оценка (12) асимптотически нормально распределена:

$$\sqrt{N_{\tau}} \cdot (\hat{\theta}_{\tau} - \theta_{\tau}) \xrightarrow[T \rightarrow \infty]{D} N_1(0, \Sigma^{\sigma} \cdot \pi^2 (1 - \theta_{\tau}^2) / 4(1 - \varepsilon)^4).$$

Доказательство. Для простоты будем считать без потери общности, что “выбросы” отсутствуют и $N_{\tau} = T - \tau$. Пусть снова $a_t = sign(z_t z_{t+\tau})$. Обозначим $S_{\tau} = (a_1 + \dots + a_{T-\tau}) / (T-\tau) = E\{a_t\} + \eta$. Согласно Теореме 5.19 [13], если существуют такие $C > 0$, $C_0 > 0$, $K > 0$, $0 < \delta < 1$, $\beta > 1$, что $\max_{k \leq n} E|a_k - Ea_k|^{2+\delta} \leq C$, $T \cdot D\{\eta\} \geq C_0$ и коэффициент перемешивания $\alpha_a(n) \leq Kn^{-\beta(2+\delta)/\delta}$, то η имеет асимптотическое нормальное распределение при $T \rightarrow \infty$. Первое условие, очевидно, выполняется. Второе следует из того, что $D\{\eta\} \xrightarrow[T \rightarrow \infty]{D} T^{-1} \cdot \Sigma^{\sigma}$.

Теперь из доказательства Теоремы 1 $\alpha_y(n) = O(e^{-hn})$, откуда по Лемме 4 $\alpha_z(n) = O(e^{-hn})$ и по Лемме 3 $\alpha_a(n) = O(e^{-hn}) \leq Kn^{-\beta(2+\delta)/\delta}$ для некоторого K . Значит $\sqrt{T} \cdot \eta \xrightarrow[T \rightarrow \infty]{D} N_1(0, \Sigma^{\sigma})$. Поскольку S – оценка θ_{τ} вычисляется как $\hat{\theta}_{\tau} = \text{Sin}(\pi \cdot S_{\tau} / 2(1-\varepsilon)^2)$, то из теоремы о функциональном преобразовании асимптотически нормальной случайной величины получаем

$$\sqrt{T} \cdot (\hat{\theta}_{\tau} - \theta_{\tau}) \xrightarrow[T \rightarrow \infty]{D} N_1(0, \text{Cos}^2(\arcsin(\theta_{\tau})) \cdot \Sigma^{\sigma} \cdot \pi^2 / 4(1 - \varepsilon)^4).$$

что эквивалентно утверждению теоремы. ■

ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

Для двух моделей авторегрессии M_1, M_2 с “выбросами” с параметрами $T=500, \varepsilon=0.1, \beta=10, \sigma=1$ и порождающими уравнениями:

$$M_1: y_t + 0.3y_{t-1} + 0.2y_{t-2} = u_t, p=2,$$

$$M_2: y_t - 0.2y_{t-1} + 0.1y_{t-2} - 0.4y_{t-4} = u_t, p=4,$$

были смоделированы серии по 10000 реализаций. Для каждой реализации коэффициенты авторегрессии b оценивались четырьмя разными процедурами: робастной с известным ε , робастной с неизвестным ε , процедурой, использующей оценки Хьюбера (6) и классической МНК – оценкой [1]. В таблицах 1 – 4 приведены численные результаты для оценок $\{\hat{\theta}_r\}, \{\hat{b}_r\}$ (Sm – смещение, Var – вариация). Гистограммы для абсолютных уклонений оценок $\delta = \max_i |b_i - \hat{b}_i|$ представлены на рис.1 и расположены сверху вниз в том же порядке, что и перечисленные выше процедуры оценивания.

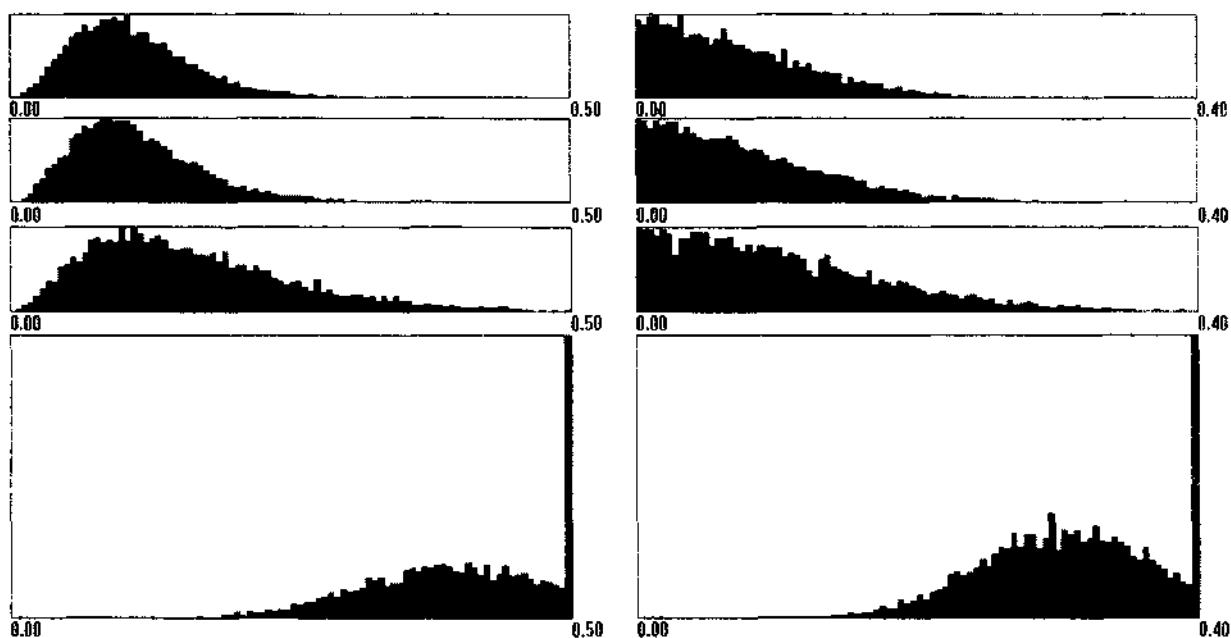


Рис.1. Гистограммы величин δ для моделей M_1 (справа), M_2 (слева)

Модель M_1

Процедура	$Var(\theta 1)$	$Sm(\theta 1)$	$Var(\theta 2)$	$Sm(\theta 2)$
роб. ε - изв.	0.017	0.007	0.023	0.009
роб. ε - неизв.	0.023	0.004	0.019	0.004
Хьюбера	0.029	0.012	0.031	0.017
МНК	0.197	-0.059	0.092	0.132

Таблица 1

Модель M_2

Процедура	$Var(b1)$	$E(b1)$	$Var(b2)$	$E(b2)$
роб. ε - изв.	0.011	0.301	0.013	0.207
роб. ε - неизв.	0.015	0.312	0.012	0.217
Хьюбера	0.014	0.284	0.015	0.188
МНК	0.097	-0.011	0.058	0.007

Таблица 2

Модель M_2 **Таблица 3**

Процедура	$Var(\theta 1)$	$Sm(\theta 1)$	$Var(\theta 2)$	$Sm(\theta 2)$	$Var(\theta 3)$	$Sm(\theta 3)$	$Var(\theta 4)$	$Sm(\theta 4)$
роб. ε - изв.	0.019	-0.009	0.026	0.004	0.017	0.007	0.029	-0.009
роб. ε - неизв.	0.017	-0.007	0.021	0.003	0.014	0.003	0.022	-0.008
Хьюбера	0.027	-0.014	0.031	0.008	0.022	0.012	0.034	-0.013
МНК	0.073	0.071	0.085	0.143	0.068	-0.052	0.273	0.039

Модель M_2 **Таблица 4**

Процедура	$Var(b1)$	$E(b1)$	$Var(b2)$	$E(b2)$	$Var(b3)$	$E(b3)$	$Var(b4)$	$E(b4)$
роб. ε - изв.	0.013	-0.209	0.011	0.115	0.010	0.008	0.017	-0.384
роб. ε - неизв.	0.012	-0.210	0.013	0.114	0.009	0.004	0.015	-0.390
Хьюбера	0.016	-0.189	0.017	0.121	0.013	0.003	0.021	-0.426
МНК	0.042	0.053	0.023	0.014	0.053	-0.02	0.191	0.008

ЛИТЕРАТУРА

1. *Андерсон Т. Статистический анализ временных рядов / Андерсон Т. М.: Мир, 1976.*
2. *Хьюбер П. Робастность в статистике / Хьюбер П. М.: Мир, 1984.*
3. *Хампель Ф. Робастность в статистике: подход на основе функций влияния / Хампель Ф. М.: Мир, 1989.*
4. *Тьюки Дж. Анализ данных и регрессия / Мостеллер Ф., Тьюки Дж. Москва: Статистика, 1982.*
5. *Kharin Yu. Robustness in statistical pattern recognition / Kharin Yu. Dordrecht-Boston-London: Kluwer Academic Publishers, 1996.*
6. *Shevlyakov G.L. Robustness Data Analysis: criteria and methods / Shevlyakov G.L., Vilchevski N.O. N.Y.: VSP, 2002.*
7. *Андерсон Т. Введение в многомерный статистический анализ данных / Андерсон Т. М.: ФМЛ, 1963.*
8. *Вадзинский Р.Н. Справочник по вероятностным распределениям / Вадзинский Р.Н. М.: Наука, 2001.*
9. *Bradley R.C. Introducion to strong mixing conditions / Bradley R.C. N.Y.: Kendrick Press, 2007.*
10. *Doukhan P. Mixing: properties and examples / Doukhan P. N.Y.: Springer-Verlag, 1994.*
11. *Ширяев А.Н. Вероятность / Ширяев А.Н. М.: Наука, 1980.*
12. *Lin Zh. Limit theory for mixing dependent random variables / Lin Zh., Lu Ch. Dordrecht-Boston-London: Kluwer Academic Publishers, 1996.*
13. *Харин Ю.С. Теория вероятностей / Харин Ю.С., Зуев Н.М. Минск: БГУ, 2004.*