УДК 519.2

А. Ю. ХАРИН[1], *ТОН ТХАТ ТУ*[1, 2]

ПОСЛЕДОВАТЕЛЬНАЯ СТАТИСТИЧЕСКАЯ ПРОВЕРКА ГИПОТЕЗ О ПАРАМЕТРАХ ВРЕМЕННЫХ РЯДОВ С ТРЕНДОМ ПРИ ПРОПУСКАХ НАБЛЮДЕНИЙ

[1]*Белорусский государственный университет, Минск, Беларусь, e-mail: KharinAY@bsu.by, tthattu@gmail.com*
[2]*Университет г. Дананг, Вьетнам, e-mail: tthattu@gmail.com*

Рассматривается задача последовательной проверки простых гипотез для временных рядов с трендом в случае пропуска наблюдений. Построен последовательный тест, исследованы характеристики его эффективности. Приведены результаты вычислительных экспериментов.

Ключевые слова: последовательный тест, простые гипотезы, временные ряды, тренд, пропущенные значения, вероятности ошибок, математическое ожидание числа наблюдений.

A. Yu. KHARIN[1], *TON THAT TU*[1,2]

SEQUENTIAL STATISTICAL HYPOTHESES TESTING ON PARAMETERS OF TIME SERIES WITH A TREND UNDER MISSING VALUES

[1]*Belarusian State University, Minsk, Belarus, e-mail: KharinAY@bsu.by, tthattu@gmail.com*
[2] *Da Nang University, Vietnam, e-mail: tthattu@gmail.com*

The problem of sequential testing of simple hypotheses for time series with a trend is considered in case of missing observations. The sequential test is constructed and its performance characteristics are analysed. Numerical results of experiments are given.

Keywords: sequential test, simple hypotheses, time series, trend, missing values, error probabilities, expected number of observations.

Введение. Последовательный тест, предложенный А. Вальдом в 1947 г. (см. [1]), успешно применяется во многих задачах прикладной статистики благодаря его оптимальным свойствам. Характеристики теста хорошо исследованы при условии одинаковых распределений наблюдений (см. [2–4]). Однако на практике исследуемые данные часто описываются более сложными моделями, например моделью временного ряда с трендом (см. [2–4]). В рамках гипотетических предположений эти модели позволяют строить приемлемые статистические оценки для неизвестных параметров, например с использованием метода наименьших квадратов (см. [5]). Кроме того, некоторые наблюдения не могут быть зарегистрированы по каким-либо причинам, поэтому проблема пропущенных значений становится актуальной. В рамках гипотетической модели наблюдений при различных критериях найдены оптимальные решения задачи прогнозирования временных рядов с трендом (см. [2, 5]). Эти результаты используются в настоящей работе для решения задачи проверки простых гипотез о параметрах временных рядов с трендом в случае пропуска наблюдений.

1. Математическая модель. Рассмотрим гипотетическую вероятностную модель временного ряда с трендом [5]:

$$x_t = \theta^T \psi(t) + \xi_t, \ t = 1, 2, 3, ..., \tag{1}$$

где $\psi(t) = (\psi_1(t), \psi_2(t), ..., \psi_m(t))^T$, $t \geq 1$, – базисные функции тренда, $\theta = (\theta_1, \theta_2, ..., \theta_m)^T \in \mathbf{R}^m$ – неизвестный вектор параметров, $\{\xi_t, t \geq 1\}$ – последовательность независимых одинаково распределенных гауссовских случайных величин, $\xi_t \sim N(0, \sigma^2)$.

На практике иногда не могут наблюдаться некоторые выборочные значения. Пусть значение x_{k+1}, $k \geq m$, не наблюдается.

Рассматриваются две простые параметрические гипотезы:

$$H_0 : \theta = \theta^0, \; H_1 : \theta = \theta^1, \tag{2}$$

где θ^0, θ^1 – известные векторы. Обозначим статистику логарифмического отношения правдоподобия для гипотез H_0, H_1:

$$\Lambda_n = \Lambda_n(x_1, x_2, ..., x_n) = \sum_{t=1}^{n} \lambda_t, \tag{3}$$

где $\lambda_t = \ln\left(\dfrac{p_t(x_t, \theta^1)}{p_t(x_t, \theta^0)}\right)$ – логарифмическое отношение правдоподобия, вычисленное по наблюдению x_t; $p_t(x, \theta)$ – плотность распределения вероятностей случайной величины x_t.

В последовательном тесте Вальда [1] при проверке гипотез (2) после n наблюдений принимается решение

$$d = \mathbf{1}_{[C_+, +\infty)}(\Lambda_n) + 2 \cdot \mathbf{1}_{(C_-, C_+)}(\Lambda_n), \tag{4}$$

где $\mathbf{1}_D(\cdot)$ означает индикаторную функцию множества D. Решение d = 2 соответствует продолжению процесса наблюдения, поскольку заданная точность не может быть обеспечена. Решение d = 0 (d = 1) означает остановку процесса наблюдения и принятие гипотезы H_0(H_1). В (4) $C_-, C_+ \in \mathbf{R}, C_- < C_+$ – параметры теста, называемые порогами. В соответствии с [1] будем использовать следующие значения:

$$C_- = \ln\left(\frac{\beta_0}{1-\alpha_0}\right), \; C_+ = \ln\left(\frac{1-\beta_0}{\alpha_0}\right),$$

где α_0, β_0 – заданные предельно допустимые значения вероятностей ошибок первого (принять гипотезу H_1 при справедливой H_0) и второго (принять гипотезу H_0 при справедливой H_1) рода соответственно.

2. Вероятностные свойства статистики отношения правдоподобия при наличии пропущенного значения. Обозначим: $E^{(l)}(\cdot)$, $D^{(l)}(\cdot)$ – условное математическое ожидание и дисперсия при условии, что справедлива гипотеза $H_l, l \in \{0,1\}$; $H_n = \sum_{t=1}^{n} \psi(t)\psi^T(t)$;

$$X_k = (x_1, x_2, ..., x_k)^T, \; U = (\xi_1, \xi_2, ..., \xi_k)^T, \quad A = \begin{pmatrix} \psi_1(1) & \psi_2(1) & ... & \psi_m(1) \\ \psi_1(2) & \psi_2(2) & ... & \psi_m(2) \\ ... & ... & ... & ... \\ \psi_1(k) & \psi_2(k) & ... & \psi_m(k) \end{pmatrix}.$$

Пусть $\mathrm{rank}(A) = m$.

Лемма 1. *В условиях модели* (1) *для теста* (3), (4) *проверки гипотез* (2) *статистики* λ_t, Λ_n *имеют гауссовские распределения и справедливы соотношения:*

$$E(\lambda_t) = -\frac{1}{2\sigma^2}\left\{2(\theta^0 - \theta^1)^T \psi(t)\psi^T(t)\theta + (\theta^1)^T \psi(t)\psi^T(t)\theta^1 - (\theta^0)^T \psi(t)\psi^T(t)\theta^0\right\},$$

$$E(\Lambda_n) = -\frac{1}{2\sigma^2}\left\{2(\theta^0 - \theta^1)^T H_n \theta + (\theta^1)^T H_n \theta^1 - (\theta^0)^T H_n \theta^0\right\};$$

$$D(\lambda_t) = \frac{(\theta^0 - \theta^1)^T \psi(t)\psi^T(t)(\theta^0 - \theta^1)}{\sigma^2}, \; D(\Lambda_n) = \frac{(\theta^0 - \theta^1)^T H_n (\theta^0 - \theta^1)}{\sigma^2}.$$

Доказательство. Из (1) следует, что

$$x_t \sim N(\theta^T \psi(t), \sigma^2), \ t \geq 1, \ \ p_t(x, \theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2\sigma^2} (x - \theta^T \psi(t))^2 \right\},$$

$$\lambda_t = \lambda_t(x_t) = -\frac{1}{2\sigma^2} \left\{ 2x_t (\theta^0 - \theta^1)^T \psi(t) + (\theta^1)^T \psi(t) \psi^T(t) \theta^1 - (\theta^0)^T \psi(t) \psi^T(t) \theta^0 \right\}.$$

Далее доказательство состоит в использовании свойств нормального распределения и проведении тождественных преобразований. Лемма доказана.

Модель (1) в матричной форме записывается следующим образом:

$$X_k = A\theta + U. \tag{5}$$

Оценка $\hat{\theta}$ параметра θ по методу наименьших квадратов имеет вид

$$\hat{\theta} = (A^T A)^{-1} A^T X_k.$$

Из свойств метода наименьших квадратов следует, что $\hat{\theta}$ имеет нормальное распределение с параметрами

$$E(\hat{\theta}) = \theta, \ \ \mathrm{cov}(\hat{\theta}, \hat{\theta}) = \sigma^2 (A^T A)^{-1}.$$

Для построения статистического критерия (теста) проверки гипотез (2) в условиях пропуска наблюдения x_{k+1} будем использовать $\tilde{x}_{k+1} = \hat{\theta}^T \psi(k+1)$ вместо пропущенного наблюдения. В этом случае рассмотрим последовательный тест вида (4), где вместо Λ_n используется статистика

$$\tilde{\Lambda}_n = \begin{cases} \Lambda_n, & n \leq k, \\ \Lambda_n + \tilde{\lambda}_{k+1} - \lambda_{k+1}, & n > k, \end{cases} \quad \tilde{\lambda}_{k+1} = \lambda_{k+1}(\tilde{x}_{k+1}). \tag{3$'$}$$

Примем обозначения:

$$s_n^2 = \sum_{t=1}^{n} \sigma_t^2, \ \ m_n^{(l)} = \sum_{t=1}^{n} \mu_t^{(l)} = \frac{(-1)^{l+1} s_n^2}{2}, \ \ l = 0,1;$$

$$\sigma_n^2 = D^{(0)}(\lambda_n) = D^{(1)}(\lambda_n) = \frac{(\theta^0 - \theta^1)^T \psi(n) \psi^T(n) (\theta^0 - \theta^1)}{\sigma^2};$$

$$\mu_n^{(l)} = E^{(l)}(\lambda_n) = \frac{(-1)^{l+1} (\theta^0 - \theta^1)^T \psi(n) \psi^T(n) (\theta^0 - \theta^1)}{2\sigma^2} = \frac{(-1)^{l+1} \sigma_n^2}{2};$$

$$\Gamma = (\theta^0 - \theta^1)(\theta^0 - \theta^1), \ \ N = \inf\{n \in \mathbf{N} : \Lambda_n \notin (C_-, C_+)\}, \ \tilde{N} = \inf\{n \in \mathbf{N} : \tilde{\Lambda}_n \notin (C_-, C_+)\}.$$

Лемма 2. *В условиях модели* (1) *при истинности гипотезы* H_i (i = 0,1) *статистика* $\tilde{\Lambda}_n$ *имеет гауссовское распределение* $\tilde{\Lambda}_n \sim N(m_n^{(i)}, \tilde{s}_n^2)$, *где*

$$\tilde{s}_n^2 = \begin{cases} s_n^2, & n \leq k, \\ s_n^2 + \sigma_{k+1}^2 \left[1 + \psi^T(k+1)(A^T A)^{-1} \psi(k+1) \right], & n > k. \end{cases} \tag{6}$$

Доказательство. Имеем

$$\tilde{\Lambda}_{k+1} = \sum_{i=1}^{k} \lambda_i + \tilde{\lambda}_{k+1} = \sum_{i=1}^{k} \frac{(\theta^1 - \theta^0)^T \psi(i)}{\sigma^2} x_i + \frac{(\theta^1 - \theta^0)^T \psi(k+1)}{\sigma^2} \tilde{x}_{k+1} - \frac{(\theta^1)^T H_{k+1} \theta^1 - (\theta^0)^T H_{k+1} \theta^0}{2\sigma^2} =$$

$$= \left[V_k^T + \frac{(\theta^1 - \theta^0)^T \psi(k+1)}{\sigma^2} \psi^T(k+1)(A^T A)^{-1} A^T \right] X_k - \frac{(\theta^1)^T H_{k+1} \theta^1 - (\theta^0)^T H_{k+1} \theta^0}{2\sigma^2},$$

где

$$V_k=\left(\frac{(\theta^1-\theta^0)^T\psi(1)}{\sigma^2},\frac{(\theta^1-\theta^0)^T\psi(2)}{\sigma^2},\ldots,\frac{(\theta^1-\theta^0)^T\psi(k)}{\sigma^2}\right)^T=\frac{1}{\sigma^2}A(\theta^1-\theta^0).$$

Вычисляем дисперсию:

$$D(\tilde{\Lambda}_{k+1})=\sigma^2\left[V_k^T+\frac{(\theta^1-\theta^0)^T\psi(k+1)}{\sigma^2}\psi^T(k+1)(A^TA)^{-1}A^T\right]\times$$

$$\times\left[V_k^T+\frac{(\theta^1-\theta^0)^T\psi(k+1)}{\sigma^2}\psi^T(k+1)(A^TA)^{-1}A^T\right]^T=$$

$$=\sigma^2V_k^TV_k+2(\theta^1-\theta^0)^T\psi(k+1)\psi^T(k+1)(A^TA)^{-1}A^TV_k+$$

$$+\left[\frac{(\theta^1-\theta^0)^T\psi(k+1)}{\sigma}\right]^2\psi^T(k+1)(A^TA)^{-1}\psi(k+1).$$

Таким образом,

$$D(\tilde{\Lambda}_{k+1})=D(\Lambda_{k+1})+\sigma_{k+1}^2\left[1+\psi^T(k+1)(A^TA)^{-1}\psi(k+1)\right]. \qquad (7)$$

Для $n>k+1$,

$$D(\tilde{\Lambda}_n)=D(\tilde{\Lambda}_{k+1})+D(\lambda_{k+2})+\ldots+D(\lambda_n)=D(\Lambda_n)+\sigma_{k+1}^2\left[1+\psi^T(k+1)(A^TA)^{-1}\psi(k+1)\right].$$

Кроме того, $E^{(i)}(\tilde{\lambda}_{k+1})=E^{(i)}(\lambda_{k+1})=\mu_{k+1}^{(i)},\ i=0,1,$ и $E(\Lambda_n)=E(\tilde{\Lambda}_n),\forall n$. Лемма доказана.

В дальнейшем будем использовать следующий известный результат [6]: если X, Y – независимые случайные величины, $X\sim N(\mu_x,\sigma_x^2),\ Y\sim N(\mu_y,\sigma_y^2)$, то

$$f_{X+Y|Y}(x\,|\,y)=\frac{f_{X+Y,Y}(x,y)}{f_Y(y)}=n_1(x;y+\mu_x,\sigma_x^2), \qquad (8)$$

где $n_1(x;\mu,\sigma^2)$ – плотность распределения вероятностей, соответствующая $N(\mu,\sigma^2)$.

3. Анализ построенного последовательного теста.

Теорема 1. *Если в рамках модели* (1) $\mathrm{rank}(A)=m$, $k\geq m$, $\mathrm{tr}(\Gamma H_n)\to+\infty$ *при* $n\to+\infty$, *то с вероятностью* 1 *одна из гипотез* (2) *принимается тестом* (3′), (4) *на основе конечного числа наблюдений.*

Доказательство. В условиях теоремы $s_n^2\to+\infty$ при $n\to+\infty$. Для всех $n\in\mathbf{N}$ имеем

$$P_l(\tilde{N}>n)=P_l(\tilde{\Lambda}_i\in(C_-,C_+),\ i=\overline{1,n})\leq$$
$$\leq P_l(\tilde{\Lambda}_n\in(C_-,C_+))=\Phi\left(\frac{C_+-m_n^{(l)}}{\tilde{s}_n}\right)-\Phi\left(\frac{C_--m_n^{(l)}}{\tilde{s}_n}\right),\ l\in\{0,1\},$$

где $\Phi(x)$ – функция распределения $N(0,1)$. Поэтому $\lim\limits_{n\to+\infty}P_l(\tilde{N}>n)=0$, или эквивалентно, $P_l(\tilde{N}<+\infty)=1,\ \ l\in\{0,1\}$. Теорема доказана.

Замечание 1. Предположим, что значения k_0 последовательных наблюдений $x_{k+1},x_{k+2},\ldots,x_{k+k0}$, недоступны. Вместо них мы можем использовать $\tilde{x}_i=\hat{\theta}^T\psi(i),i=\overline{k+1,k+k_0}$, где $\hat{\theta}=(A^TA)^{-1}A^TX_k$. В этом случае теорема 1 также остается верной.

Замечание 2. Если имеется дополнительная информация $x_{k+1}\in[a,b]$, то оценка $\hat{\theta}$ параметра θ может быть получена решением следующей экстремальной задачи:

$$\begin{cases}(X_k-A\theta)^T(X_k-A\theta)\to\min,\\(\theta^T\psi(k+1)-a)(\theta^T\psi(k+1)-b)\leq 0.\end{cases}$$

Т е о р е м а 2. *В условиях теоремы* 1 *справедливы следующие неравенства:*

$$P_i(N=\tilde{N})\geq P_i(\tilde{N}\leq k+1)-\varphi_{k+1}^{(i)}P_i(\tilde{N}=k+1)+\sum_{l=k+2}^{+\infty}(1-\varphi_l^{(i)})P_i(\min\{N,\tilde{N}\}\geq l),\quad i=0,1, \tag{9}$$

где $\varphi_l^{(i)}=\Phi\left(\dfrac{C_+-C_-+\mu_l^{(i)}}{\sigma_l}\right)+\Phi\left(\dfrac{C_+-C_--\mu_l^{(i)}}{\sigma_l}\right)-1,\quad i=0,1.$

Д о к а з а т е л ь с т в о. Проведем преобразования:

$$P_i(N=\tilde{N})=\sum_{l=1}^{+\infty}P_i(N=\tilde{N}=l)=\sum_{l=1}^{k}P_i(N=\tilde{N}=l)+P_i(N=\tilde{N}=k+1)+\sum_{l=k+2}^{+\infty}P_i(N=\tilde{N}=l).$$

В силу $\Lambda_n=\tilde{\Lambda}_n,\ \forall n\leq k$, получаем:

$$\sum_{l=1}^{k}P_i(N=\tilde{N}=l)=\sum_{l=1}^{k}P_i(\tilde{N}=l)=P_i(\tilde{N}\leq k).$$

С другой стороны, $\{\lambda_n,\, n\geq 1\}$ – последовательность независимых случайных величин, поэтому

$$\begin{aligned}P_i(N=\tilde{N}=k+1)&=P_i(\Lambda_j\in(C_-,C_+),\ j=1,...k,\ \Lambda_{k+1},\tilde{\Lambda}_{k+1}\notin(C_-,C_+))=\\&=P_i(\Lambda_j\in(C_-,C_+),\ j=1,...,k,\ \tilde{\Lambda}_{k+1}\notin(C_-,C_+),\ \Lambda_k+\lambda_{k+1}\notin(C_-,C_+))\geq\\&\geq P_i(\Lambda_j\in(C_-,C_+),\ j=1,...,k,\ \tilde{\Lambda}_{k+1}\notin(C_-,C_+),\ \ |\lambda_{k+1}|\geq C_+-C_-)=\\&=\left(1-\varphi_{k+1}^{(i)}\right)P_i(\tilde{N}=k+1).\end{aligned}$$

Аналогично,

$$\begin{aligned}\sum_{l=k+2}^{+\infty}P_i(N=\tilde{N}=l)&=\sum_{l=k+2}^{+\infty}P_i(\Lambda_j,\ \tilde{\Lambda}_j\in(C_-,C_+),\ j=1,...,l-1,\ \Lambda_l,\tilde{\Lambda}_l\notin(C_-,C_+))=\\&=\sum_{l=k+2}^{+\infty}P_i(\Lambda_j,\ \tilde{\Lambda}_j\in(C_-,C_+),\ j=1,...,l-1,\ \Lambda_{l-1}+\lambda_l,\ \tilde{\Lambda}_{l-1}+\lambda_l\notin(C_-,C_+))\geq\\&\geq\sum_{l=k+2}^{+\infty}P_i(\Lambda_j,\ \tilde{\Lambda}_j\in(C_-,C_+),\ j=1,...,l-1,\ \ |\lambda_l|\geq C_+-C_-)=\\&=\sum_{l=k+2}^{+\infty}(1-\varphi_l^{(i)})P_i(\min\{N,\tilde{N}\}\geq l).\end{aligned}$$

Это завершает доказательство теоремы 2.

Пусть $\tilde{\alpha},\tilde{\beta}$ – фактические значения вероятностей ошибок первого и второго рода соответственно в модифицированном тесте (3′), (4); α,β – вероятности ошибок первого и второго рода соответственно в исходном тесте.

Л е м м а 3 (см. [7]). *Если случайная величина X принимает только целые положительные значения, то*

$$E(X)=\sum_{n=1}^{+\infty}P(X\geq n).$$

Т е о р е м а 3. *В условиях теоремы* 1 *имеют место следующие неравенства:*

$$|\tilde{\alpha}-\alpha|\leq P_0(N>k)\left[1+\Phi\left(\frac{\mu_{k+2}^{(0)}}{\sigma_{k+2}}\right)\right]+$$

$$+P_0(N>k)\sum_{i=k+3}^{+\infty}\int_0^{+\infty}\int_{C_--C_+}^{C_+-C_-}n_1(x,m_{i-1}^{(0)}-m_{k+1}^{(0)},s_{i-1}^2-s_{k+1}^2)n_1(y,x+\mu_i^{(0)},\sigma_i^2)dxdy; \tag{10}$$

$$|\tilde{\beta}-\beta| \le P_1(N>k)\left[1+\Phi\left(-\frac{\mu_{k+2}^{(1)}}{\sigma_{k+2}}\right)\right]+$$

$$+P_1(N>k)\sum_{i=k+3}^{+\infty}\int_{-\infty}^{0}\int_{C_--C_+}^{C_+-C_-} n_1(x,m_{i-1}^{(1)}-m_{k+1}^{(1)},s_{i-1}^2-s_{k+1}^2)n_1(y,x+\mu_i^{(1)},\sigma_i^2)dxdy; \quad (11)$$

$$|E^{(l)}(\tilde{N}-N)| \le P_l(N>k)\left[\Phi\left(\frac{C_+-C_--\mu_{k+2}^{(l)}}{\sigma_{k+2}}\right)+\Phi\left(\frac{C_+-C_-+\mu_{k+2}^{(l)}}{\sigma_{k+2}}\right)\right]+$$

$$+P_l(N>k)\sum_{i=k+3}^{+\infty}\int_{C_--C_+}^{C_+-C_-}\int_{C_--C_+}^{C_+-C_-} n_1(x,m_{i-1}^{(l)}-m_{k+1}^{(l)},s_{i-1}^2-s_{k+1}^2)n_1(y,x+\mu_i^{(l)},\sigma_i^2)dxdy, l\in\{0,1\}. \quad (12)$$

Доказательство. Заметим, что $\tilde{\alpha}=P_0(\tilde{\Lambda}_{\tilde{N}}\ge C_+)=\sum_{i=1}^{+\infty}P_0(\tilde{\Lambda}_i\ge C_+,\tilde{N}=i)$. Поэтому

$$0\le\tilde{\alpha}-\sum_{i=1}^{k}P_0(\tilde{\Lambda}_i\ge C_+,\tilde{N}=i)=$$

$$=P_0(\tilde{\Lambda}_{k+1}\ge C_+,\tilde{N}=k+1)+P_0(\tilde{\Lambda}_{k+2}\ge C_+,\tilde{N}=k+2)+\sum_{i=k+3}^{+\infty}P_0(\tilde{\Lambda}_i\ge C_+,\tilde{N}=i).$$

С другой стороны,

$$P_0(\tilde{\Lambda}_{k+1}\ge C_+,\tilde{N}=k+1)=P_0(\tilde{\Lambda}_j\in(C_-,C_+), j=\overline{1,k},\tilde{\Lambda}_{k+1}\ge C_+)\le$$

$$\le P_0(\tilde{\Lambda}_j\in(C_-,C_+), j=\overline{1,k})=P_0(\tilde{N}>k).$$

Аналогично,

$$P_0(\tilde{\Lambda}_{k+1}\ge C_+,\tilde{N}=k+2)=P_0(\tilde{\Lambda}_j\in(C_-,C_+), j=\overline{1,k+1},\tilde{\Lambda}_{k+2}\ge C_+)\le$$

$$\le P_0(\tilde{\Lambda}_j\in(C_-,C_+), j=\overline{1,k+1},\lambda_{k+2}>0)\le$$

$$\le P_0(\tilde{N}>k)P_0(\lambda_{k+2}>0)=P_0(\tilde{N}>k)\Phi\left(\frac{\mu_{k+2}^{(0)}}{\sigma_{k+2}}\right).$$

Для $i\ge k+3$ получаем

$$P_0(\tilde{\Lambda}_i\ge C_+,\tilde{N}=i)=P_0(\tilde{\Lambda}_j\in(C_-,C_+), j=\overline{1,i-1},\tilde{\Lambda}_i\ge C_+)\le$$

$$\le P_0(\tilde{\Lambda}_j,\in(C_-,C_+), j=\overline{1,k+1},\sum_{j=k+2}^{i-1}\lambda_j\in(C_--C_+,C_+-C_-),\sum_{j=k+2}^{i}\lambda_j>0)\le$$

$$\le P_0(\tilde{\Lambda}_j,\in(C_-,C_+), j=\overline{1,k})P(\sum_{j=k+2}^{i-1}\lambda_j\in(C_--C_+,C_+-C_-),\sum_{j=k+2}^{i}\lambda_j>0)\le$$

$$\le P_0(\tilde{N}>k)\int_{0}^{+\infty}\int_{C_--C_+}^{C_+-C_-} n_1(x,m_{i-1}^{(0)}-m_{k+1}^{(0)},s_{i-1}^2-s_{k+1}^2)n_1(y,x+\mu_i^{(0)},\sigma_i^2)dxdy.$$

Из этого следует

$$0\le\tilde{\alpha}-\sum_{i=1}^{k}P_0(\tilde{\Lambda}_i\ge C_+,\tilde{N}=i)\le P_0(\tilde{N}>k)\left[1+\Phi\left(\frac{\mu_{k+2}^{(0)}}{\sigma_{k+2}}\right)\right]+$$

$$+\sum_{i=k+3}^{+\infty}\int_{0}^{+\infty}\int_{C_--C_+}^{C_+-C_-} n_1(x,m_{i-1}^{(0)}-m_{k+1}^{(0)},s_{i-1}^2-s_{k+1}^2)n_1(y,x+\mu_i^{(0)},\sigma_i^2)dxdy. \quad (13)$$

Аналогично,

$$0 \le \alpha - \sum_{i=1}^{k} P_0(\Lambda_i \ge C_+, N = i) \le P_0(N > k)\left[1 + \Phi\left(\frac{\mu_{k+2}^{(0)}}{\sigma_{k+2}}\right)\right] +$$

$$+ \sum_{i=k+3}^{+\infty} \int_{0}^{+\infty} \int_{C_- - C_+}^{C_+ - C_-} n_1(x, m_{i-1}^{(0)} - m_{k+1}^{(0)}, s_{i-1}^2 - s_{k+1}^2) n_1(y, x + \mu_i^{(0)}, \sigma_i^2) dxdy. \quad (14)$$

Неравенство (10) получается из (13), (14) и того факта, что

$$\Lambda_j = \tilde{\Lambda}_j, \forall j = 1, ..., k. \quad (15)$$

Неравенство (11) доказывается аналогично.

Из леммы 3 имеем $E^{(l)}(\tilde{N}) = 1 + \sum_{i=1}^{+\infty} P_l(\tilde{N} > i)$. Очевидно, $P_l(\tilde{N} > k+1) \le P_l(\tilde{N} > k)$ и

$$P_l(\tilde{N} > k+2) = P_l(\tilde{\Lambda}_j \in (C_-, C_+), j = \overline{1, k+2}) \le P_l(\tilde{\Lambda}_j \in (C_-, C_+), j = \overline{1, k}, \lambda_{k+2} \in (C_- - C_+, C_+ - C_-)) =$$

$$= P_l(\tilde{N} > k)\left[\Phi\left(\frac{C_+ - C_- - \mu_{k+2}^{(l)}}{\sigma_{k+2}}\right) - \Phi\left(\frac{C_- - C_+ - \mu_{k+2}^{(l)}}{\sigma_{k+2}}\right)\right].$$

Для $i \ge k+3$ получаем

$$P_l(\tilde{N} > i) = P_l\left(\tilde{\Lambda}_j \in (C_-, C_+), j = \overline{1, i}\right) \le$$

$$\le P_l\left(\tilde{\Lambda}_j, \in (C_-, C_+), j = \overline{1, k+1}, \ \sum_{j=k+2}^{t} \lambda_j \in (C_- - C_+, C_+ - C_-), \ t = \overline{i-1, i}\right) \le$$

$$\le P_l(\tilde{\Lambda}_j, \in (C_-, C_+), j = \overline{1, k+1}) P\left(\sum_{j=k+2}^{t} \lambda_j \in (C_- - C_+, C_+ - C_-), \ t = \overline{i-1, i}\right) \le$$

$$\le P_l(\tilde{N} > k) \int_{C_- - C_+}^{C_+ - C_-} \int_{C_- - C_+}^{C_+ - C_-} n_1\left(x, m_{i-1}^{(l)} - m_{k+1}^{(l)}, s_{i-1}^2 - s_{k+1}^2\right) n_1\left(y, x + \mu_i^{(l)}, \sigma_i^2\right) dxdy.$$

Поэтому

$$0 \le E^{(l)}(\tilde{N}) - 1 - \sum_{i=1}^{k} P_l(\tilde{N} > i) \le P_l(\tilde{N} > k)\left[\Phi\left(\frac{C_+ - C_- - \mu_{k+2}^{(l)}}{\sigma_{k+2}}\right) + \Phi\left(\frac{C_+ - C_- + \mu_{k+2}^{(l)}}{\sigma_{k+2}}\right)\right] +$$

$$+ P_l(\tilde{N} > k) \sum_{i=k+3}^{+\infty} \int_{C_- - C_+}^{C_+ - C_-} \int_{C_- - C_+}^{C_+ - C_-} n_1\left(x, m_{i-1}^{(l)} - m_{k+1}^{(l)}, s_{i-1}^2 - s_{k+1}^2\right) n_1\left(y, x + \mu_i^{(l)}, \sigma_i^2\right) dxdy. \quad (16)$$

Аналогично,

$$0 \le E^{(l)}(N) - 1 - \sum_{i=1}^{k} P_l(N > i) \le P_l(N > k)\left[\Phi\left(\frac{C_+ - C_- - \mu_{k+2}^{(l)}}{\sigma_{k+2}}\right) + \Phi\left(\frac{C_+ - C_- + \mu_{k+2}^{(l)}}{\sigma_{k+2}}\right)\right] +$$

$$+ P_l(N > k) \sum_{i=k+3}^{+\infty} \int_{C_- - C_+}^{C_+ - C_-} \int_{C_- - C_+}^{C_+ - C_-} n_1\left(x, m_{i-1}^{(l)} - m_{k+1}^{(l)}, s_{i-1}^2 - s_{k+1}^2\right) n_1\left(y, x + \mu_i^{(l)}, \sigma_i^2\right) dxdy. \quad (17)$$

Из (15), (16) и (17) получаем (12). Теорема 3 доказана.

На практике иногда встречается следующая ситуация: значение x_{k+1} наблюдается с вероятностью $q \in (0,1)$. Пусть v – случайная величина, $v \in \{0,1\}$, $P(v=0) = 1-q$, $P(v=1) = q$, величина v независима от последовательности $\{\xi_t, t \ge 1\}$. В этом случае построим тест вида (4), использующий вместо (3) статистику

$$\overline{\Lambda}_n = \begin{cases} \Lambda_n, & v=1, \\ \tilde{\Lambda}_n, & v=0. \end{cases} \tag{3''}$$

Примем обозначения: $\overline{N} = \inf\{n \in \mathbf{N} : \overline{\Lambda}_n \notin (C_-, C_+)\}$, $\overline{\alpha} = P_0(\overline{\Lambda}_{\overline{N}} \ge C_+)$, $\overline{\beta} = P_1(\overline{\Lambda}_{\overline{N}} \le C_-)$.

Теорема 4. *В условиях теоремы* 1 *для теста* (3″), (4) *имеют место следующие неравенства:*

$$|\overline{\alpha} - \alpha| \le (1-q)P_0(N>k)\left[1 + \Phi\left(\frac{\mu_{k+2}^{(0)}}{\sigma_{k+2}}\right)\right] +$$

$$+(1-q)P_0(N>k)\sum_{i=k+3}^{+\infty}\int_0^{+\infty}\int_{C_- - C_+}^{C_+ - C_-} n_1\left(x, m_{i-1}^{(0)} - m_{k+1}^{(0)}, s_{i-1}^2 - s_{k+1}^2\right) n_1\left(y, x + \mu_i^{(0)}, \sigma_i^2\right) dxdy, \tag{18}$$

$$|\overline{\beta} - \beta| \le (1-q)P_1(N>k)\left[1 + \Phi\left(-\frac{\mu_{k+2}^{(1)}}{\sigma_{k+2}}\right)\right] +$$

$$+(1-q)P_1(N>k)\sum_{i=k+3}^{+\infty}\int_{-\infty}^{0}\int_{C_- - C_+}^{C_+ - C_-} n_1\left(x, m_{i-1}^{(1)} - m_{k+1}^{(1)}, s_{i-1}^2 - s_{k+1}^2\right) n_1\left(y, x + \mu_i^{(1)}, \sigma_i^2\right) dxdy, \tag{19}$$

$$|E^{(l)}(\overline{N} - N)| \le (1-q)P_l(N>k)\left[\Phi\left(\frac{C_+ - C_- - \mu_{k+2}^{(l)}}{\sigma_{k+2}}\right) + \Phi\left(\frac{C_+ - C_- + \mu_{k+2}^{(l)}}{\sigma_{k+2}}\right)\right] +$$

$$+(1-q)P_l(N>k)\sum_{i=k+3}^{+\infty}\int_{C_- - C_+}^{C_+ - C_-}\int_{C_- - C_+}^{C_+ - C_-} n_1(x, m_{i-1}^{(l)} - m_{k+1}^{(l)}, s_{i-1}^2 - s_{k+1}^2) n_1(y, x + \mu_i^{(l)}, \sigma_i^2) dxdy, l \in \{0,1\}. \tag{20}$$

Доказательство. Для всех $n \in \mathbf{N}, i \in \{0,1\}$, имеем

$$P_i(\overline{N} > n) = P_i(\overline{\Lambda}_j \in (C_-, C_+), j = 1,...,n) =$$
$$= P_i(\overline{\Lambda}_j \in (C_-, C_+), j = 1,...,n, v=0) + P_i(\overline{\Lambda}_j \in (C_-, C_+), j = 1,...,n, v=1) =$$
$$= qP_i(\Lambda_j \in (C_-, C_+), j = 1,...,n) + (1-q)P_i(\tilde{\Lambda}_j \in (C_-, C_+), j = 1,...,n) =$$
$$= qP_i(N > n) + (1-q)P_i(\tilde{N} > n).$$

Отсюда получаем $\lim\limits_{n\to+\infty} P_i(\overline{N} > n) = 0, i \in \{0,1\}$. Это означает, что тест, основанный на $\overline{\Lambda}_n$, конечен с вероятностью 1. С другой стороны,

$$\overline{\alpha} = P_0(\overline{\Lambda}_{\overline{N}} \ge C_+) = \sum_{i=1}^{+\infty} P_0(\overline{\Lambda}_i \ge C_+, \overline{N} = i) = \sum_{i=1}^{+\infty} P_0(\overline{\Lambda}_j \in (C_-, C_+), j = 1,...,i-1, \overline{\Lambda}_i \ge C_+) =$$
$$= q\sum_{i=1}^{+\infty} P_0(\Lambda_j \in (C_-, C_+), j = 1,...,i-1, \Lambda_i \ge C_+) + (1-q)\sum_{i=1}^{+\infty} P_0(\tilde{\Lambda}_j \in (C_-, C_+), j = 1,...,i-1, \tilde{\Lambda}_i \ge C_+) =$$
$$= qP_0(\Lambda_N \ge C_+) + (1-q)P_0(\tilde{\Lambda}_{\tilde{N}} \ge C_+) = q\alpha + (1-q)\tilde{\alpha}.$$

Из последнего соотношения и (10) получим (18). Неравенства (19) и (20) доказываются аналогично. Теорема доказана.

Замечание 3. Пусть $k \in [1, m-1]$, и хотя x_{k+1} не наблюдается, известно, что, $x_{k+1} \in [a,b]$. Пусть $\overline{\lambda}_{k+1} = \max\{\lambda_{k+1}(a), \lambda_{k+1}(b)\}$, $\underline{\lambda}_{k+1} = \min\{\lambda_{k+1}(a), \lambda_{k+1}(b)\}$. В этом случае тест может быть построен следующим образом. Если после k наблюдений окончательное решение $d = 0$ или $d = 1$ не может быть принято, то по результатам (k+1)-го наблюдения принимается H_0, если $\Lambda_k + \overline{\lambda}_{k+1} \le C_-$, и принимается H_1, если $\Lambda_k + \underline{\lambda}_{k+1} \ge C_+$, в противном случае отсутствующее значение x_{k+1} игнорируется с использованием статистики

$$L_n = \begin{cases} \Lambda_n, & n \le k, \\ \Lambda_n - \lambda_{k+1}, & n \ge k+2. \end{cases}$$

4. Результаты компьютерных экспериментов. Рассмотрим вероятностную модель (1) и гипотезы (2) при следующих значениях параметров:

$$q = 0,7,\ m = 4,\ \sigma = 10,\ \theta^0 = (1,2,2,2)^T,\ \theta^1 = (1,1,1,1)^T,\ \psi(t) = (1,\ t/10,\ t^2/100,\ t^3/1000).$$

Обозначим оценку характеристики $\tau \in \{\alpha, \overline{\alpha}, t_0, t_0'\}$ методом Монте-Карло через $\hat{\tau}$. Количество повторений в методе Монте-Карло составляло 50 000. Бесконечная сумма была заменена конечной суммой от 1 до 1000. Иллюстрация результатов теоремы 4 приведена в таблице, где $t_0 = E^{(0)}(N)$, $\overline{t_0} = E^{(0)}(\overline{N})$, t_0' – среднее число наблюдений, требуемых тестом (4), когда пропущенное наблюдение игнорируется, при справедливости гипотезы H_0.

Результаты вычислительных экспериментов

α_0	β_0	k	$\hat{\alpha}$	$\hat{\overline{\alpha}}$	$\lvert \alpha - \overline{\alpha} \rvert \le$	$\hat{t}_0$	$\hat{t}_0'$	$\hat{\overline{t}}_0$	$\lvert t_0 - \overline{t}_0 \rvert \le$
0,1	0,1	14	0,05576	0,06672	0,75200	17,9005	18,27068	17,78896	1,69637
		18		0,06272	0,22599		18,26831	17,89814	0,44214
		22		0,05592	0,01680		17,97683	17,88560	0,02834
0,05	0,1	14	0,02594	0,03440	0,76940	18,1521	18,52616	18,07926	1,87108
		18		0,03136	0,24901		18,51662	18,13772	0,52706
		22		0,02708	0,02382		18,24194	18,14634	0,04346

Из таблицы видно, что отсутствие одного наблюдения не приводит к значительным изменениям тестовых характеристик. При фиксированных значениях α_0, β_0 увеличение k приводит к уменьшению верхних границ значений $\lvert \alpha - \overline{\alpha} \rvert$ и $\lvert t_0 - \overline{t_0} \rvert$. Поскольку указанные верхние границы зависят от $P_l(N > k)$, $l \in \{0,1\}$, эти выражения теряют ценность, когда значение k становится меньше, чем среднее число наблюдений. Отметим, что по результатам экспериментов имеет место соотношение $\hat{t}_0' > \hat{\overline{t}}_0$ для любых α_0, β_0, k. Этот факт означает, что с использованием предсказанного значения пропущенного наблюдения число наблюдений может быть в среднем уменьшено по сравнению с игнорированием этого значения при обеспечении требуемой точности.

Заключение. В статье построен последовательный тест для временных рядов с трендом при наличии пропущенных наблюдений. Доказаны достаточные условия конечности теста. Построены верхние границы для абсолютных значений разностей между тестовыми характеристиками в модифицированных и оригинальных тестах. Результаты могут быть использованы для анализа робастности [8] построенного теста.

Работа выполнена при частичной поддержке по международной программе IMPULSE (проект № 698327-CDAMCSS).

Список использованной литературы

1. *Wald, A.* Sequential analysis / A. Wald. – New York: John Wiley and Sons, 1947.
2. *Харин, А. Ю.* Робастность байесовских и последовательных статистических решающих правил / А. Ю. Харин. – Минск: БГУ, 2013.
3. *Govindarajulu, Z.* Sequential statistics / Z. Govindarajulu. – Singapore: World Scientific Publishing, 2004.
4. *Kharin, A. Y.* An approach to performance analysis of the sequential probability ratio test for the simple hypotheses testing / A. Y. Kharin // Proceedings of the Belarusian State University. – 2002. – Vol. 1. – P. 92–96.
5. *Андерсон, Т.* Статистический анализ временных рядов / Т. Андерсон. – М.: Мир, 1976.
6. *Bilodeau, M.* Theory of multivariate statistics / M. Bilodeau, D. Brenner. – New York: Springer – Verlag, 1999.
7. *Chung, K. L.* A course in probability theory / K. L. Chung. – San Diego: Academic Press, 2001.
8. *Kharin, A. Y.* Performance and robustness evaluation in sequential hypotheses testing / A. Y. Kharin // Communications in Statistics – Theory and Methods. – 2016. – Vol. 45 (6). – P. 1663–1709.

Поступила в редакцию 30.06.2016