ОБ АСИМПТОТИЧЕСКИХ СВОЙСТВАХ ОЦЕНКИ ПАРАМЕТРОВ ВЕКТОРНОЙ АВТОРЕГРЕССИИ ПРИ НАЛИЧИИ «ПРОПУСКОВ»

Ю. С. Харин, А. С. Гурин

Белорусский государственный университет
Минск, Беларусь
E-mail: kharin@bsu.by

В статье рассматривается проблема статистического оценивания параметров векторной авторегрессии при наличии «пропусков». Для построенных оценок параметров модели доказаны свойства состоятельности и асимптотической нормальности. Приводятся результаты численных экспериментов на реальных данных.

Ключевые слова: векторная авторегрессия, «пропуски», оценивание параметров.

Исследования были частично поддержаны грантом INTAS 03/51/3714.

1. ВВЕДЕНИЕ

«Пропуски» являются типичными искажениями модельных предположений в анализе данных [3, 12]. Существует множество причин, по которым данные могут иметь «пропуски» [6, 10]: 1) данные не существуют в некоторые моменты времени; 2) ошибки регистрации данных; 3) удаление выбросов.

Векторная авторегрессия (VAR) часто используется на практике для статистического анализа (статистического оценивания параметров, статистического тестирования гипотез, статистического прогнозирования) временных рядов в эконометрике [9], биометрике [5], технике [7] и во многих других приложениях.

Существует три подхода к оцениванию параметров векторной авторегрессии при наличии «пропусков»: 1) оценивание по методу максимального правдоподобия (МП-оценивание); 2) применение ЕМ-алгоритма [3]; 3) модификация известных статистических оценок параметров для случая «пропусков».

Первый и второй подход характеризуются значительной вычислительной сложностью из-за нелинейности целевой функции и отсутствием гарантированных асимптотических свойств. По этой причине мы развиваем в статье третий подход.

2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Пусть наблюдаемый d-векторный временной ряд Y_t описывается VAR(1) моделью:

$$Y_t = BY_t + U_t,\ t \in Z, \qquad (1)$$

где Z – множество целых чисел, $Y_t = (Y_{t1},...,Y_{td})' \in R^d$, $B = (B_{kl}) \in R^{d \times d}$ – матрица неизвестных коэффициентов авторегрессии (все собственные числа матрицы B лежат в единичном круге), $U_t = (U_{t1},...,U_{td})' \in R^d$, $\{U_t\}$ – н. о. р. случайные векторы, $\mathbf{E}\{U_t\} = 0_d$ – нулевой d-вектор, $\mathbf{E}\{U_t U_t'\} = \Sigma = \mathrm{Const}_{\bar{t}}$, $|\Sigma| \neq 0$, где $\mathrm{Const}_{\bar{i}_1,...,\bar{i}_n}$ обозначает математический объект (переменная, вектор, матрица и т. д.), который не зависит от переменных $i_1,...,i_n$, $n \in N$. В наблюдениях $\{Y_t\}$ имеются пропуски. Для каждого вектора Y_t укажем бинарный вектор («шаблон пропусков») $O_t = (O_{t1},...,O_{td})' \in \{0,1\}^d$, где $O_{ti} = \{1, \text{если } Y_{ti} \text{ наблюдается}; \ 0, \text{если } Y_{ti} - \text{пропуск}\}$. Определим дискретное множество $M = \{(t,i), t \in Z, i \in \{1,...,d\} : O_{ti} = 1\}$, его элементы лексикографически упорядочены в возрастающем порядке, $k = |M|$ – общее число наблюдаемых компонент, $t_- = \min\left\{t : \sum_{i=1}^{d} O_{ti} > 0\right\}$ – минимальный момент времени с наблюдаемыми компонентами, $t_+ = \max\left\{t : \sum_{i=1}^{d} O_{ti} > 0\right\}$ – максимальный момент времени с наблюдаемыми компонентами. Без потери общности предположим, что $t_- = 1$, $t_+ = T$. Определим биекцию $M \leftrightarrow \{1,...,K\} : k = \chi(t,i)$ и обратную функцию $(t,i) = \bar{\chi}(k)$. Составим K-вектор всех наблюдаемых компонент: $X = (X_1,...,X_K)' \in R^K$, $X_K = Y_{\bar{\chi}(k)}$, $k \in \{1,...,K\}$. Если $O_{ti} = 1$, $t \in \{1,...,T\}, i \in \{1,...,d\}$, тогда процесс Y_t наблюдается на $[1,T]$ без пропусков, $K = Td$, $X = (Y_1',...,Y_T')'$, $\chi(t,i) = i + (t-1)d$, $\bar{\chi}(k) = ([(k-1)/d] + 1, \ (k-1) \bmod d + 1)$, $k \in \{1,...,K\}$. Определим матрицы: $F = (F_{ij}) = \mathbf{cov}\{X, X\} \in R^{K \times K}$, $G = (G_{ij}) = \mathbf{cov}\{Y_1, Y_1\} \in R^{d \times d}$, $G_1 = \mathbf{cov}\{Y_2, Y_1\} \in R^{d \times d}$. Заметим, что согласно [1] $G = \sum_{i=0}^{\infty} B^i \Sigma (B')^i$, $G_1 = BG$.

Заметим, что AR(p)-модель и VAR(p)-модель могут быть преобразованы к VAR(p)-модели увеличением числа компонент [1].

Определим дополнительные условия на инновационный процесс U_t и «шаблон пропусков» $\{O_t\}$, которые будут использоваться далее:

У1. Моменты порядка три и четыре инновационного процесса ограничены: $\left|\mathbf{E}\{U_{ti_1} \cdots U_{ti_3}\}\right| \leq \mathrm{Const}_{\bar{t},\bar{i}_1,...,\bar{i}_3}$, $\left|\mathbf{E}\{U_{ti_1} \cdots U_{ti_4}\}\right| \leq \mathrm{Const}_{\bar{t},\bar{i}_1,...,\bar{i}_4}$, $t \in Z$, $i_1,...,i_4 \in \{1,...,d\}$.

У2. Момент четвертого порядка инновационного процесса не зависит от времени, и моменты порядков от 5 до 8 инновационного процесса ограничены: $\mathbf{E}\{U_{ti_1} \cdots U_{ti_4}\} = \Sigma^{(4)}_{i_1,...,i_4} = \mathrm{Const}_{\bar{t}}$, $\left|\mathbf{E}\{U_{ti_1} \cdots U_{ti_5}\}\right| \leq \mathrm{Const}_{\bar{t},\bar{i}_1,...,\bar{i}_5}$, $\left|\mathbf{E}\{U_{ti_1} \cdots U_{ti_6}\}\right| \leq \mathrm{Const}_{\bar{t},\bar{i}_1,...,\bar{i}_6}$, $\left|\mathbf{E}\{U_{ti_1} \cdots U_{ti_7}\}\right| \leq \mathrm{Const}_{\bar{t},\bar{i}_1,...,\bar{i}_7}$, $\left|\mathbf{E}\{U_{ti_1} \cdots U_{ti_8}\}\right| \leq \mathrm{Const}_{\bar{t},\bar{i}_1,...,\bar{i}_8}$, $t \in Z$, $i_1,...,i_8 \in \{1,...,d\}$.

У3. U_t – гауссовский случайный вектор: $\mathcal{L}\{U_t\} = \mathcal{N}_d(0_d, \Sigma)$.

У4. «Шаблон пропусков» $\{O_t\}$ удовлетворяет асимптотике ($T \to \infty$, $i, j \in \{1,\dots,d\}$):

$$T^{-1} \sum_{t=1}^{T} O_{ti} O_{tj} \to \nu_{ij}^{(0)} \in (0,1], \quad (T-1)^{-1} \sum_{t=1}^{T-1} O_{t+1,i} O_{tj} \to \nu_{ij}^{(1)} \in (0,1], \tag{2}$$

$\nu_{ij}^{(0)}$ – предельная частота пары компонент (i,j), наблюдаемых в один и тот же момент времени, $\nu_{ij}^{(1)}$ – предельная частота пары компонент (i,j), наблюдаемых в соседние моменты времени.

У5. «Шаблон пропусков» $\{O_t\}$ удовлетворяет асимптотике ($T \to \infty$):

$$\left(T - |\tau| - 1\right)^{-1} \sum_{t,t'=1}^{T-1} O_{ti} O_{tj} O_{t'i'} O_{t'j'} \delta_{t-t',\tau} \to \nu_{i,j,i',j'}^{(2)}(\tau) \in [0,1],$$

$\nu_{i,j,i',j'}^{(2)}(\tau)$ – предельная частота пары компонент (i,j), наблюдаемой вместе с парой (i',j') с лагом τ, $\nu_{i,j,i,j}^{(2)}(0) \equiv \nu_{ij}^{(0)}$, $\delta_{i,j}$ – символ Кронекера;

$$\left(T - |\tau| - 1\right)^{-1} \sum_{t,t'=1}^{T-1} O_{ti} O_{tj} O_{t'+1,i'} O_{t'j'} \delta_{t-t',\tau} \to \nu_{i,j,i',j'}^{(3)}(\tau) \in [0,1],$$

$\nu_{i,j,i',j'}^{(3)}(\tau)$ – предельная частота пары компонент (i,j), наблюдаемой вместе с компонентами (i',j') с лагами $\tau - 1$ и τ;

$$\left(T - |\tau| - 1\right)^{-1} \sum_{t,t'=1}^{T-1} O_{t+1,i} O_{tj} O_{t'+1,i'} O_{t'j'} \delta_{t-t',\tau} \to \nu_{i,j,i',j'}^{(4)}(\tau) \in [0,1],$$

$\nu_{i,j,i',j'}^{(4)}(\tau)$ – предельная частота пары компонент (i,j), наблюдаемой в соседние моменты времени, и пары (i',j'), наблюдаемой в соседние моменты времени с лагом τ, $\nu_{i,j,i,j}^{(4)}(0) \equiv \nu_{ij}^{(1)}$, где $\tau \in Z$, $i, j, i', j' \in \{1,\dots,d\}$.

Задача состоит в построении оценок $\hat{B} = \hat{B}(X)$: $R^K \to R^{d \times d}$, $\hat{\Sigma} = \hat{\Sigma}(X)$: $R^K \to R^{d \times d}$ для параметров B, Σ. Обозначим вариацию оценки матрицы параметров $\nu = \mathbf{E}\left\{ \sum_{i,j=1}^{d} \left(\hat{B}_{ij} - B_{ij}\right)^2 \right\}$.

3. МП-ОЦЕНИВАНИЕ

Следующая лемма дает явное выражение для элементов матрицы F.

Лемма 1. Пусть имеет место модель (1). Тогда

$$F = \left(F_{ij}\right), \; F_{ij} = F_{ji} = \left(B^{\bar{\chi}_1(i) - \bar{\chi}_1(j)} G\right)_{\bar{\chi}_2(i), \bar{\chi}_2(j)}, \; i, j \in \{1,\dots,K\}, i \ge j. \tag{3}$$

Доказательство. Используя выражение для ковариационной матрицы VAR(1)-модели [1]: $\mathbf{cov}\{Y_i, Y_j\} = B^{i-j}G$, $i \geq j$, находим ковариации: $F_{ij} =$ $\mathbf{cov}\{X_i, X_j\} =$ $\mathbf{cov}\{Y_{\overline{\chi}(i)}, Y_{\overline{\chi}(j)}\} =$ $\left(\mathbf{cov}\{Y_{\overline{\chi}_1(i)}, Y_{\overline{\chi}_1(j)}\}\right)_{\overline{\chi}_2(i),\overline{\chi}_2(j)} =$ $\left(B^{\overline{\chi}_1(i)-\overline{\chi}_1(j)}G\right)_{\overline{\chi}_2(i),\overline{\chi}_2(j)}$. ■

Теорема 1. Пусть имеют место модель (1) и условие У3. Тогда МП-оценки $\hat{B}_{OMП}, \hat{\Sigma}_{OMП}$ модельных параметров B, Σ есть решение оптимизационной задачи:

$$\ln|F| + X^T F^{-1} X \to \min_{B,\Sigma}. \tag{4}$$

Доказательство. Согласно предположениям теоремы 1, вектор X имеет гауссовское распределение: $l(B,\Sigma) = n_K(X \mid 0_K, F)$. Взяв логарифм, приходим к утверждению теоремы. ■

4. ОЦЕНИВАНИЕ, ОСНОВАННОЕ НА КОВАРИАЦИЯХ

Из-за вычислительной сложности оптимизационной задачи (4) мы предлагаем построить другую статистическую оценку параметра B. Определим минимальную допустимую длительность наблюдения $T_0 = \min\left\{T \in N : \min\limits_{i,j} \sum\limits_{t=1}^{T} O_{ti}O_{tj} > 0,\right.$ $\left.\min\limits_{i,j} \sum\limits_{t=1}^{T-1} O_{t+1,i}O_{tj} > 0\right\}$ и матрицы $\hat{G}, \hat{G}_1 \in R^{d\times d}$ для $T \geq T_0$:

$$\hat{G}_{ij} = \frac{\sum\limits_{t=1}^{T} Y_{ti}Y_{tj}O_{ti}O_{tj}}{\sum\limits_{t=1}^{T} O_{ti}O_{tj}}, \quad \left(\hat{G}_1\right)_{ij} = \frac{\sum\limits_{t=1}^{T-1} Y_{t+1,i}Y_{tj}O_{t+1,i}O_{tj}}{\sum\limits_{t=1}^{T-1} O_{t+1,i}O_{tj}}, \quad i,j \in \{1,\ldots,d\}. \tag{5}$$

Из модели (1) следует, что $|G| \neq 0$ и матрицы B, G, G_1 удовлетворяют матричному уравнению $G_1 = BG$, следовательно, $B = G_1 G^{-1}$. Следуя «plug-in» принципу и используя предыдущее уравнение, построим матричную статистику (если $|\hat{G}| \neq 0$):

$$\hat{B}_{OOK} = \hat{G}_1\left(\hat{G}\right)^{-1} \tag{6}$$

и будем называть ее оценкой, основанной на ковариациях (ООК). Проанализируем асимптотические свойства ($T \to \infty$) предложенных оценок (5), (6). Определим функции «шаблона пропусков» ($t, t', \tau \in Z$, $T \in N$, $i, j, i', j' \in \{1,\ldots,d\}$):

$$c^{(1)}_{\tau,i,j,i',j'}(T) = T\sum_{t,t'=1}^{T-1} O_{ti}O_{tj}O_{t'i'}O_{t'j'}\delta_{t-t',\tau}\left(\sum_{t=1}^{T} O_{ti}O_{tj}\sum_{t=1}^{T} O_{ti'}O_{tj'}\right)^{-1},$$

$$c^{(2)}_{\tau,i,j,i',j'}(T) = T\sum_{t,t'=1}^{T-1} O_{ti}O_{tj}O_{t'+1,i'}O_{t'j'}\delta_{t-t',\tau}\left(\sum_{t=1}^{T} O_{ti}O_{tj}\sum_{t=1}^{T-1} O_{t+1,i'}O_{tj'}\right)^{-1},$$

$$c^{(3)}_{\tau,i,j,i',j'}(T)=T\sum_{t,t'=1}^{T-1} O_{t+1,i}O_{tj}O_{t'i'}O_{t'j'}\delta_{t-t',\tau}\left(\sum_{t=1}^{T-1}O_{t+1,i}O_{tj}\sum_{t=1}^{T}O_{ti'}O_{tj'}\right)^{-1},$$

$$c^{(4)}_{\tau,i,j,i',j'}(T)=T\sum_{t,t'=1}^{T-1} O_{t+1,i}O_{tj}O_{t'+1,i'}O_{t'j'}\delta_{t-t',\tau}\left(\sum_{t=1}^{T-1}O_{t+1,i}O_{tj}\sum_{t=1}^{T-1}O_{t+1,i'}O_{tj'}\right)^{-1},$$

$$C^{(1)}_{\tau,i,j,i',j'}=\nu^{(2)}_{i,j,i',j'}(\tau)\left(\nu^{(0)}_{ij}\nu^{(0)}_{i'j'}\right)^{-1},\ C^{(2)}_{\tau,i,j,i',j'}=\nu^{(3)}_{i,j,i',j'}(\tau)\left(\nu^{(0)}_{ij}\nu^{(1)}_{i'j'}\right)^{-1},$$

$$C^{(3)}_{\tau,i,j,i',j'}=\nu^{(3)}_{i',j',i,j}(-\tau)\left(\nu^{(0)}_{i'j'}\nu^{(1)}_{ij}\right)^{-1},\ C^{(4)}_{\tau,i,j,i',j'}=\nu^{(4)}_{i,j,i',j'}(\tau)\left(\nu^{(1)}_{ij}\nu^{(1)}_{i'j'}\right)^{-1},$$

и ковариации:

$$g^{(1)}_{t-t',i,j,i',j'}=\mathbf{E}\left\{\left(Y_{ti}Y_{tj}-\mathbf{E}\left\{Y_{ti}Y_{tj}\right\}\right)\left(Y_{t'i'}Y_{t'j'}-\mathbf{E}\left\{Y_{t'i'}Y_{t'j'}\right\}\right)\right\},$$

$$g^{(2)}_{t-t',i,j,i',j'}=\mathbf{E}\left\{\left(Y_{ti}Y_{tj}-\mathbf{E}\left\{Y_{ti}Y_{tj}\right\}\right)\left(Y_{t'+1,i'}Y_{t'j'}-\mathbf{E}\left\{Y_{t'+1,i'}Y_{t'j'}\right\}\right)\right\},$$

$$g^{(3)}_{t-t',i,j,i',j'}=\mathbf{E}\left\{\left(Y_{t+1,i}Y_{tj}-\mathbf{E}\left\{Y_{t+1,i}Y_{tj}\right\}\right)\left(Y_{t'i'}Y_{t'j'}-\mathbf{E}\left\{Y_{t'i'}Y_{t'j'}\right\}\right)\right\},$$

$$g^{(4)}_{t-t',i,j,i',j'}=\mathbf{E}\left\{\left(Y_{t+1,i}Y_{tj}-\mathbf{E}\left\{Y_{t+1,i}Y_{tj}\right\}\right)\left(Y_{t'+1,i'}Y_{t'j'}-\mathbf{E}\left\{Y_{t'+1,i'}Y_{t'j'}\right\}\right)\right\}.$$

Следующая лемма непосредственно следует из введенных обозначений.

Лемма 2. Пусть имеют место модель (1) и условия У4, У5. Тогда при $T\to\infty$ имеет место следующее асимптотическое поведение «шаблона пропусков» $\{O_t\}$:

$$\left(T-|\tau|-1\right)^{-1}\sum_{t,t'=1}^{T-1}O_{t+1,i}O_{tj}O_{t'i'}O_{t'j'}\delta_{t-t',\tau}\to\nu^{(3)}_{i',j',i,j}(-\tau)\in[0,1];$$

$$c^{(k)}_{\tau,i,j,i',j'}(T)\to C^{(k)}_{\tau,i,j,i',j'},\ k\in\{1,2,3,4\};$$

$$C^{(k)}_{\tau,i',j',i,j}=C^{(k)}_{-\tau,i,j,i',j'},\ k\in\{1,4\},\ C^{(3)}_{\tau,i',j',i,j}=C^{(2)}_{-\tau,i,j,i',j'},\ \tau\in Z,\ i,j,i',j'\in\{1,\dots,d\}.$$

Лемма 3. Пусть имеют место модель (1) и условие У2. Тогда ковариация

$$\mathbf{E}\left\{\left(Y_{t+u,i}Y_{tj}-\mathbf{E}\left\{Y_{t+u,i}Y_{tj}\right\}\right)\left(Y_{t'+u',i'}Y_{t'j'}-\mathbf{E}\left\{Y_{t'+u',i'}Y_{t'j'}\right\}\right)\right\}$$

функционально зависит от моментов времени t,t' только через их разность $t-t'$,

$$\left|\mathbf{E}\left\{\left(Y_{t+u,i}Y_{tj}-\mathbf{E}\left\{Y_{t+u,i}Y_{tj}\right\}\right)\left(Y_{t'+u',i'}Y_{t'j'}-\mathbf{E}\left\{Y_{t'+u',i'}Y_{t'j'}\right\}\right)\right\}\right|\le\lambda^{|t-t'|}\mathrm{Const}_{\bar{\iota},\bar{\iota}'},$$

где $t,t',u,u'\in Z$, $\lambda\in[0,1)$, и справедливы следующие соотношения для ковариаций:

$$g^{(k)}_{\tau,i',j',i,j}=g^{(k)}_{-\tau,i,j,i',j'},\ k\in\{1,4\},\ g^{(3)}_{\tau,i',j',i,j}=g^{(2)}_{-\tau,i,j,i',j'},\ \tau\in Z,\ i,j,i',j'\in\{1,\dots,d\}.$$

Доказательство. Первое утверждение леммы следует из свойства независимости моментов $\Sigma,\Sigma^{(4)}$ от времени t. Второе – из неравенства для степеней матрицы B [1]: $\left|\left(B^{\tau}\right)_{ij}\right|\le\lambda^{\tau}\mathrm{Const}_{\bar{\tau},\bar{i},\bar{j}}$, где $\lambda\in[0,1)$, $\tau\in N$, $i,j\in\{1,..,d\}$. Заменой индексов легко получаем третье утверждение леммы. ■

Теорема 2. Пусть имеют место модель (1) и условия У1, У4. Тогда оценки (5), (6) состоятельны при $T \to \infty$:

$$\hat{B}_{OOK} \xrightarrow{\mathbf{P}} B,\ \hat{G} \xrightarrow{\mathbf{P}} G,\ \hat{G}_1 \xrightarrow{\mathbf{P}} G_1.$$

Доказательство. Утверждение следует из выражения (6) для матрицы B и свойств оценок ковариаций (5). ∎

Дальнейшие результаты об асимптотической нормальности $\hat{B}_{OOK}$ основаны на следующей центральной предельной теореме для m_T-зависимых случайных векторов [11, 8].

Теорема 3. Пусть $\left\{Z_t^{(T)}, t = 1,2,...,k_T\right\}$ – последовательность m_T-зависимых случайных величин с $\mathbf{E}\left\{Z_t^{(T)}\right\} = 0$ для всех t и T, $k_T \to \infty$ при $T \to \infty$, $S_T = \sum_{t=1}^{k_T} Z_t^{(T)}$, $D_T = \mathbf{E}\left\{S_T^2\right\}$, $\overline{D}_T^2 = \sum_{t=1}^{k_T} \mathbf{E}\left\{\left(Z_t^{(T)}\right)^2\right\}$, $F_T(x) = \mathbf{P}(S_T \le D_T x)$ – функция распределения S_T, $\Delta_T(x) = |F_T(x) - \Phi(x)|$, $\Phi(x)$ – функция стандартного нормального распределения, $\gamma_\delta = \frac{\delta(\delta+2)}{2\left(\delta^2 + 4\delta + 2\right)}$, Const – константа, $0 < \delta \le 1$, $\varepsilon_T = D_T^{-2} m_T^{\frac{3\delta+2}{\delta}}$. Предположим, что $\mathbf{E}\left\{\left|Z_t^{(T)}\right|^{2+\delta}\right\} < \infty$ и выполнены следующие условия: 1) $D_T^2 \to \infty$; 2) $\overline{D}_T^2 = O\left(D_T^2\right)$; 3) $\sum_{t=1}^{k_T} \mathbf{E}\left\{\left|Z_t^{(T)}\right|^{2+\delta}\right\} = O\left(D_T^2\right)$; 4) $k_T = O\left(D_T^2\right)$; 5) $D_T^8 m_T^{-6} \le k_T^7$ для больших T; 6) $\varepsilon_T \to \infty$. Тогда $\Delta_T(x) \le \frac{\text{Const}}{(1+|x|)^{2+\delta}} \varepsilon_T^{\gamma_\delta}$ для всех x.

Пусть последовательность случайных векторов $\xi_T = (\xi_{Ti}) \in R^d$ сходится по распределению к случайному вектору $\xi = (\xi_i) \in R^d$ при $T \to \infty$. Если ковариации $\mathbf{cov}\left\{\xi_i, \xi_j\right\}$ конечны $\forall i, j \in \{1,...,d\}$, тогда будем называть эти ковариации асимптотическими ковариациями последовательности ξ_T и обозначать их следующим образом:

$$\mathbf{acov}\left\{(\xi_T)_i, (\xi_T)_j\right\} \underset{T\to\infty}{:=} \mathbf{cov}\left\{\xi_i, \xi_j\right\},\ i, j \in \{1,...,d\}.$$

Аналогичным образом определим асимптотическую дисперсию и асимптотическое математическое ожидание последовательности ξ_T: $\mathbf{aE}\{\xi_T\} \underset{T\to\infty}{:=} \mathbf{E}\{\xi\}$, $\mathbf{aD}\{\xi_{Ti}\} \underset{T\to\infty}{:=} \mathbf{D}\{\xi_i\}$, $i \in \{1,...,d\}$.

Лемма 4. Пусть имеют место модель (1) и условия У2, У4, У5. Тогда при $T \to \infty$ вектор, составленный из элементов матриц $\sqrt{T}\left(\hat{G} - G\right)$, $\sqrt{T}\left(\hat{G}_1 - G_1\right)$ имеет асимптотически нормальное распределение с нулевым средним и асимптотическими ковариациями:

$$\mathbf{acov}\left\{\sqrt{T}\left(\hat{G}-G\right)_{ij},\sqrt{T}\left(\hat{G}-G\right)_{i'j'}\right\}=\sum_{\tau=-\infty}^{\infty} g^{(1)}_{\tau,i,j,i',j'}C^{(1)}_{\tau,i,j,i',j'}, \tag{7}$$

$$\mathbf{acov}\left\{\sqrt{T}\left(\hat{G}_1-G_1\right)_{ij},\sqrt{T}\left(\hat{G}_1-G_1\right)_{i'j'}\right\}=\sum_{\tau=-\infty}^{\infty} g^{(4)}_{\tau,i,j,i',j'}C^{(4)}_{\tau,i,j,i',j'},$$

$$\mathbf{acov}\left\{\sqrt{T}\left(\hat{G}-G\right)_{ij},\sqrt{T}\left(\hat{G}_1-G_1\right)_{i'j'}\right\}=\sum_{\tau=-\infty}^{\infty} g^{(2)}_{\tau,i,j,i',j'}C^{(2)}_{\tau,i,j,i',j'},$$

где $i,j,i',j' \in \{1,\ldots,d\}$.

Доказательство. Согласно теореме из [4], вектор имеет асимптотически нормальное распределение тогда и только тогда, когда произвольная линейная комбинация его элементов имеет асимптотически нормальное распределение. Определим линейную комбинацию с произвольными коэффициентами $\alpha_{ij},\beta_{ij}\in R$:

$$\sqrt{T}\sum_{i,j=1}^{d}\left(\alpha_{ij}\left(\hat{G}-G\right)_{ij}+\beta_{ij}\left(\hat{G}_1-G_1\right)_{ij}\right)=\eta_1(T)+\eta_2(T)+\eta_3(T),$$

где

$$\eta_1(T)=\sqrt{T}\sum_{t=1}^{T-1}\sum_{i,j=1}^{d}\left(\alpha_{ij}\left(Y^{(m_T)}_{ti}Y^{(m_T)}_{tj}-\mathbf{E}\left\{Y^{(m_T)}_{ti}Y^{(m_T)}_{tj}\right\}\right)O_{ti}O_{tj}\Big/\sum_{t_1=1}^{T}O_{t_1 i}O_{t_1 j}+\right.$$

$$\left.\beta_{ij}\left(Y^{(m_T)}_{t+1,i}Y^{(m_T)}_{tj}-\mathbf{E}\left\{Y^{(m_T)}_{t+1,i}Y^{(m_T)}_{tj}\right\}\right)O_{t+1,i}O_{tj}\Big/\sum_{t_1=1}^{T-1}O_{t_1+1,i}O_{t_1 j}\right),$$

$$\eta_2(T)=\sqrt{T}\sum_{t=1}^{T-1}\sum_{i,j=1}^{d}\left(\alpha_{ij}\left(Y^{(m_T)}_{ti}U^{(m_T)}_{tj}-\mathbf{E}\left\{Y^{(m_T)}_{ti}U^{(m_T)}_{tj}\right\}\right)O_{ti}O_{tj}\Big/\sum_{t_1=1}^{T}O_{t_1 i}O_{t_1 j}+\right.$$

$$\alpha_{ij}\left(U^{(m_T)}_{ti}Y^{(m_T)}_{tj}-\mathbf{E}\left\{U^{(m_T)}_{ti}Y^{(m_T)}_{tj}\right\}\right)O_{ti}O_{tj}\Big/\sum_{t_1=1}^{T}O_{t_1 i}O_{t_1 j}+$$

$$\alpha_{ij}\left(U^{(m_T)}_{ti}U^{(m_T)}_{tj}-\mathbf{E}\left\{U^{(m_T)}_{ti}U^{(m_T)}_{tj}\right\}\right)O_{ti}O_{tj}\Big/\sum_{t_1=1}^{T}O_{t_1 i}O_{t_1 j}+$$

$$\beta_{ij}\left(Y^{(m_T)}_{t+1,i}U^{(m_T)}_{tj}-\mathbf{E}\left\{Y^{(m_T)}_{t+1,i}U^{(m_T)}_{tj}\right\}\right)O_{t+1,i}O_{tj}\Big/\sum_{t_1=1}^{T-1}O_{t_1+1,i}O_{t_1 j}+$$

$$\beta_{ij}\left(U^{(m_T)}_{t+1,i}Y^{(m_T)}_{tj}-\mathbf{E}\left\{U^{(m_T)}_{t+1,i}Y^{(m_T)}_{tj}\right\}\right)O_{t+1,i}O_{tj}\Big/\sum_{t_1=1}^{T-1}O_{t_1+1,i}O_{t_1 j}+$$

$$\left.\beta_{ij}\left(U^{(m_T)}_{t+1,i}U^{(m_T)}_{tj}-\mathbf{E}\left\{U^{(m_T)}_{t+1,i}U^{(m_T)}_{tj}\right\}\right)O_{t+1,i}O_{tj}\Big/\sum_{t_1=1}^{T-1}O_{t_1+1,i}O_{t_1 j}\right),$$

$$\eta_3(T)=\sqrt{T}\sum_{i,j=1}^{d}\alpha_{ij}\left(Y_{Ti}Y_{Tj}-\mathbf{E}\left\{Y_{Ti}Y_{Tj}\right\}\right)O_{Ti}O_{Tj}\Big/\sum_{t=1}^{T}O_{ti}O_{tj},$$

$Y_t = Y_t^{(m_T)} + U_t^{(m_T)}$, $Y_t^{(m_T)} = \sum\limits_{s=0}^{m_T} B^s U_{t-s}$, $U_t^{(m_T)} = \sum\limits_{s=m_t+1}^{\infty} B^s U_{t-s}$, $m_T \in N$ – параметр разбиения.

Выбирая $\delta = 1$, $k_T = T - 1$, $m_T = \left[T^{\frac{7}{36}} \right]$, где $[\cdot]$ означает целую часть,

$$S_T = \sum_{t=1}^{T-1} T \sum_{i,j=1}^{d} \left(\alpha_{ij} \left(Y_{ti}^{(m_T)} Y_{tj}^{(m_T)} - \mathbf{E}\left\{ Y_{ti}^{(m_T)} Y_{tj}^{(m_T)} \right\} \right) O_{ti} O_{tj} \Big/ \sum_{t_1=1}^{T} O_{t_1 i} O_{t_1 j} + \right.$$

$$\left. \beta_{ij} \left(Y_{t+1,i}^{(m_T)} Y_{tj}^{(m_T)} - \mathbf{E}\left\{ Y_{t+1,i}^{(m_T)} Y_{tj}^{(m_T)} \right\} \right) O_{t+1,i} O_{tj} \Big/ \sum_{t_1=1}^{T-1} O_{t_1+1,i} O_{t_1 j} \right),$$

и используя леммы 2, 3, легко проверить условия теоремы 3 и, тем самым, доказать, что $\eta_1(T)$ имеет асимптотически нормальное распределение с нулевым средним и асимптотической дисперсией:

$$\sum_{\tau=-\infty}^{\infty} \sum_{i,j,i',j'=1}^{d} \left(\alpha_{ij} \alpha_{i'j'} g_{\tau,i,j,i',j'}^{(1)} C_{\tau,i,j,i',j'}^{(1)} + \alpha_{ij} \beta_{i'j'} g_{\tau,i,j,i',j'}^{(2)} C_{\tau,i,j,i',j'}^{(2)} + \right.$$

$$\left. \beta_{ij} \alpha_{i'j'} g_{\tau,i,j,i',j'}^{(3)} C_{\tau,i,j,i',j'}^{(3)} + \beta_{ij} \beta_{i'j'} g_{\tau,i,j,i',j'}^{(4)} C_{\tau,i,j,i',j'}^{(4)} \right).$$

Второе и третье слагаемые сходятся по вероятности к нулю: $\eta_1(T) \xrightarrow{\mathbf{P}} 0$, $\eta_3(T) \xrightarrow{\mathbf{P}} 0$ при $T \to \infty$. Тогда согласно [4] вектор, составленный из элементов матриц $\sqrt{T}\left(\hat{G} - G\right)$, $\sqrt{T}\left(\hat{G}_1 - G_1\right)$, имеет асимптотически нормальное распределение с нулевым средним и асимптотическими ковариациями (7). ■

Теорема 4. Пусть имеют место модель (1) и условия У2, У4, У5. Тогда при $T \to \infty$ вектор, составленный из элементов матрицы $\sqrt{T}\left(\hat{B}_{OOK} - B\right)$, имеет асимптотически нормальное распределение с нулевым средним и ковариациями:

$$\mathbf{acov}\left\{ \sqrt{T}\left(\hat{B}_{OOK} - B\right)_{ij}, \sqrt{T}\left(\hat{B}_{OOK} - B\right)_{i'j'} \right\} = \sum_{l,l'=1}^{d} \sum_{\tau=-\infty}^{\infty} \left(G^{-1}\right)_{lj} \left(G^{-1}\right)_{l'j'} \times \qquad (8)$$

$$\left(\sum_{k,k'=1}^{d} B_{ik} B_{i'k'} g_{\tau,k,l,k',l'}^{(1)} C_{\tau,k,l,k',l'}^{(1)} + g_{\tau,i,l,i',l'}^{(4)} C_{\tau,i,l,i',l'}^{(4)} - 2 \sum_{k=1}^{d} B_{ik} g_{\tau,k,l,i',l'}^{(2)} C_{\tau,k,l,i',l'}^{(2)} \right),$$

где $i, j, i', j' \in \{1, \ldots, d\}$.

Доказательство. Преобразуем нормированное отклонение, используя (6):

$$\sqrt{T}\left(\hat{B}_{OOK} - B\right) = \sqrt{T}\left(-\hat{G}_1 \hat{G}^{-1}\left(\hat{G} - G\right) G^{-1} + \left(\hat{G}_1 - G_1\right) G^{-1} \right).$$

Определим линейную комбинацию с произвольными коэффициентами $\alpha_{ij} \in R$:

$$\sum_{i,j=1}^{d} \alpha_{ij} \sqrt{T}\left(\hat{B}_{OOK} - B\right)_{ij} = \sqrt{T} \sum_{i,j=1}^{d} \alpha_{ij} \times$$

$$\left(\sum_{k,l=1}^{d}(-1)\left(\hat{G}_1\hat{G}^{-1}\right)_{ik}\left(\hat{G}-G\right)_{kl}\left(G^{-1}\right)_{lj}+\sum_{l=1}^{d}\left(\hat{G}_1-G_1\right)_{il}\left(G^{-1}\right)_{lj}\right).$$

Используя теорему 2, лемму 4 и известную теорему о непрерывных функциональных преобразованиях случайных векторов [2], приходим к асимптотической нормальности линейной комбинации с нулевым средним и асимптотической дисперсией:

$$\mathbf{aD}\left\{\sqrt{T}\sum_{i,j=1}^{d}\alpha_{ij}\left(\sum_{k,l=1}^{d}(-1)\left(G_1G^{-1}\right)_{ik}\left(\hat{G}-G\right)_{kl}\left(G^{-1}\right)_{lj}+\right.\right.$$

$$\left.\left.\sum_{l=1}^{d}\left(\hat{G}_1-G_1\right)_{il}\left(G^{-1}\right)_{lj}\right)\right\}=\sum_{i,j,i',j'=1}^{d}\alpha_{ij}\alpha_{i'j'}\sum_{l,l'=1}^{d}\sum_{\tau=-\infty}^{\infty}\left(G^{-1}\right)_{lj}\left(G^{-1}\right)_{l'j'}\times$$

$$\left(\sum_{k,k'=1}^{d}B_{ik}B_{i'k'}g^{(1)}_{\tau,k,l,k',l'}C^{(1)}_{\tau,k,l,k',l'}+g^{(4)}_{\tau,i,l,i',l'}C^{(4)}_{\tau,i,l,i',l'}-2\sum_{k=1}^{d}B_{ik}g^{(2)}_{\tau,k,l,i',l'}C^{(2)}_{\tau,k,l,i',l'}\right).$$

Таким образом, вектор, составленный из элементов матрицы $\sqrt{T}\left(\hat{B}_{OOK}-B\right)$, имеет асимптотически нормальное распределение с нулевым средним и ковариациями (8). ■

5. ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

Для того чтобы оценить точность оценок (6), был проведен эксперимент на реальных данных «Индекс Бевериджа цен на пшеницу 1500–1869» [1]. AR(3)-модель имела вид: $y_{t+1}=0.7489y_t-0.3397y_{t-1}+0.0388y_{t-2}+\xi_t$, где $\{\xi_t\}$ – независимые одинаково распределенные случайные величины с нулевым средним и дисперсией $\sigma^2=4$, $t_-=1$, $t_+=T=100$. Эта модель была преобразована к трехмерной VAR(1)-модели (1) с $Y_t=\left(y_t,y_{t-1},y_{t-2}\right)'\in R^3$. «Шаблон пропусков» имел вид: $O_t=\{0,$ $t=\left[\frac{T-d}{\left[(T-d)\gamma\right]+1}\right]i,\ i\in\{1,\ldots,[(T-d)]\gamma\};$ 1,иначе$\}$, где γ – доля «пропусков», $d=3$. Зависимость выборочной вариации $\hat{v}=\frac{1}{400}\sum_{k=1}^{400}\sum_{i,j=1}^{d}\left(\left(\hat{B}_{OOK}\right)^{(k)}_{ij}-B_{ij}\right)^2$ (для 400 «прогонов» метода Монте-Карло) оценки (6) от длительности наблюдения T представлена на рис. 1 для различных значений доли «пропусков» $\gamma\in\{0,0.07,0.1\}$.

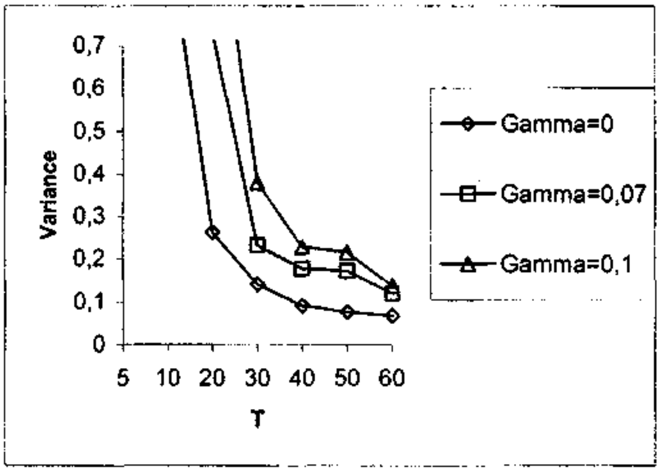

Рис.1. Выборочная вариация оценки $\hat{B}_{OOK}$

ЛИТЕРАТУРА

1. *Андерсон Т.* Статистический анализ временных рядов. М.: Мир, 1976.

2. *Барндорф-Нильсен О., Кокс Д.* Асимптотические методы в математической статистике. М.: Мир, 1999.

3. *Литтл Р. Дж. А., Рубин Д. Б.* Статистический анализ данных с пропусками. М: Финансы и статистика, 1991.

4. *Ширяев А. Н.* Вероятность. М.: Наука, 1980.

5. *Beran J., Bhansali R. J., Ocker D.* On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes // Biometrica. 1998. 85(4). P. 921–934.

6. *Greene W. H.* Econometric Analysis. New York: Macmillan, 2000.

7. *Jones R. H.* Maximum likelihood fitting of ARMA models to time series with missing observations // Technometrics. 1980. 22(3). P. 389–395.

8. *Maejima M.* A non-uniform estimate in the central limit theorem for m-dependent random variables // KEIO Engineering Reports. 1978. 31(2). P. 15–20.

9. *Pantula S. G., Shin D.* Testing for a unit root in autoregressive processes with systematic but incomplete sampling // Stat.&Prob. Letters. 1993. 18. P. 183–190.

10. *Shafer J. L.* Analysis of incomplete data. London: Chapman and Hall, 1997.

11. *Shergin V. V.* Estimate of the error term in the central limit theorem for m-dependent random variables // Lietuvos Matematikos Rinkinys. 1976. 16(4). P. 245–250.

12. *Stockinger N.* Dutter R. Robust time series analysis: A survey // Kybernetika. 1987. 23. P. 3–90.