КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ УСТОЙЧИВОСТИ МНК-ОЦЕНОК С ИСПОЛЬЗОВАНИЕМ L-BFGS-B ДЛЯ БИЛИНЕЙНОЙ АВТОРЕГРЕССИОННОЙ МОДЕЛИ ПРИ РАЗНЫХ РАСПРЕДЕЛЕНИЯХ ШУМА

Горяинов В.Б., Масягин М.М., Семерня В.М.
МГТУ им. Н.Э. Баумана (национальный исследовательский университет), Москва, Россия
vb-goryainov@bmstu.ru, masyagin@bmstu.ru, semernyavm@student.bmstu.ru

Аннотация. Исследуется влияние различных распределений остатков (нормального, Стьюдента, Лапласа, равномерного) на точность МНК-оценок параметров билинейной авторегрессионной модели. Компьютерное моделирование показало существенную зависимость точности от типа распределения, особенно для нелинейного параметра.

Ключевые слова: билинейная авторегрессия, робастные функции потерь, L-BFGS-B оптимизация, среднеквадратичная ошибка, нормальное распределение, распределение Лапласа, распределение Стьюдента, равномерное распределение.

Введение

Моделирование нелинейных временных рядов приобретает всё большую значимость в экономике, финансах, метеорологии и технических системах. Многие реальные процессы не могут быть адекватно описаны линейными моделями, что привело к активному развитию нелинейных методов анализа временных рядов [1]. Среди них особое место занимают билинейные авторегрессионные модели (БАРМ), представляющие естественное обобщение классических моделей ARMA за счет включения членов, описывающих произведение прошлых значений ряда и обновляющего процесса.

БАРМ эффективно описывают явления с нелинейными взаимодействиями и асимметричной реакцией на возмущения, что особенно ценно при моделировании экономических и финансовых процессов со структурными сдвигами, внезапными всплесками и затухающими колебаниями [2].

Ключевым этапом построения билинейных моделей является оценивание параметров. Широко используемый метод наименьших квадратов (МНК) теоретически оптимален при нормальном распределении обновляющего процесса, совпадая с оценкой максимального правдоподобия и достигая нижней границы Крамера-Рао [3]. Однако на практике часто наблюдаются отклонения от нормальности – асимметрия распределения, тяжелые хвосты, выбросы. При таких условиях эффективность МНК существенно снижается, особенно в билинейных моделях из-за нелинейного характера взаимодействия процесса с шумом [4, 5].

Влияние распределения обновляющего процесса на точность оценок определяется несколькими факторами. Во-первых, при распределениях с тяжелыми хвостами функция потерь МНК имеет множественные локальные минимумы вместо единственного глобального, что усложняет оптимизацию. Во-вторых, для некоторых распределений (например, Стьюдента с малыми степенями свободы) не существуют высшие моменты, что влияет на условия стационарности и свойства оценок. В-третьих, МНК обладает слабой робастностью к выбросам, что критично для билинейных моделей.

Для решения проблем оптимизации предлагается использовать гибридный подход, сочетающий глобальный поиск методом дифференциальной эволюции (DE) с локальным уточнением алгоритмом L-BFGS-B [6]. Такая комбинация обеспечивает надежное нахождение глобального минимума и высокую точность окончательного решения даже при сложной геометрии поверхности потерь.

Целью работы является исследование влияния различных распределений обновляющего процесса на точность МНК-оценок параметров билинейной авторегрессионной модели первого порядка.

В рамках исследования решаются следующие задачи:

- Сравнение точности оценок при нормальном, загрязненном нормальном, распределениях Стьюдента (с различными степенями свободы), Лапласа и равномерном распределении.
- Анализ зависимости ошибок оценивания от параметров распределений (степеней свободы Стьюдента, параметров загрязнения Тьюки).
- Исследование смещения оценок и чувствительности параметров модели к типу распределения.

Результаты позволят сформулировать практические рекомендации по применению МНК для оценивания параметров билинейных моделей в условиях нестандартных распределений обновляющего процесса.

1. Модели и методы параметрического оценивания

1.1. Билинейная авторегрессионная модель

Билинейная авторегрессионная модель (БАРМ) представляет собой расширение традиционной авторегрессии, включающее нелинейные взаимодействия между переменными, что позволяет учесть более сложные зависимости во временных рядах [7]. В стандартной авторегрессионной модели (АР) значения временного ряда зависят только от собственных предыдущих значений, тогда как в БАРМ они зависят как от предыдущих значений ряда, так и от взаимодействий между предыдущими значениями и случайными ошибками.

Общая формулировка модели БАРМ порядка p и q записывается следующим образом:

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \sum_{j=0}^{r} c_j e_{t-j} + \sum_{i=1}^{m} \sum_{j=1}^{k} b_{ij} X_{t-i} e_{t-j} \tag{1}$$

X_k – значение наблюдаемой величины на шаге t,

p, r, m, k – параметры, определяющие параметры модели,

$a_i, b_{i,j}, c_i$ – параметры, подлежащие оценке, причем $c_0 = 1$.

Обновляющий процесс (шум) e_t является последовательностью независимых одинаково распределённых случайных величин с нулевым математическим ожиданием и дисперсией σ_e^2. Такой подход позволяет учитывать как авторегрессионные, так и билинейные взаимодействия между значениями ряда и шумом, что важно для моделирования нелинейных временных зависимостей.

В данном исследовании рассматривается стационарная билинейная авторегрессионная модель, для которой предполагается выполнение условий стационарности. Эти условия, в общем случае, довольно сложны, но для частной модели

$$X_t = aX_{t-1} + e_t - bX_{t-1}e_{t-1} \tag{2}$$

условие стационарности можно сформулировать как: $a^2 + b^2\sigma_e^2 < 1$.

Это условие гарантирует, что значения модели будут ограниченными и не будут расходиться со временем, что важно для корректного оценивания параметров модели.

1.2. Метод наименьших квадратов

Метод наименьших квадратов (МНК) является одним из наиболее распространенных подходов к оцениванию параметров статистических моделей в силу его вычислительной простоты и оптимальности при определенных условиях. В основе метода лежит минимизация суммы квадратов отклонений между наблюдаемыми значениями временного ряда и значениями, предсказанными моделью.

Для рассматриваемой билинейной модели

$$X_t = aX_{t-1} + ce_t - bX_{t-1}e_{t-1}, \tag{3}$$

функция потерь МНК определяется как

$$Q(\theta) = \sum_{t=\gamma+1}^{N} e_t^2(\theta), \tag{4}$$

где $\theta = (a, b, c)$ – вектор параметров модели, подлежащих оцениванию,

$e_t(\theta)$ – ошибки модели, вычисляемые рекурсивно:

$$e_t(\theta) = X_t - aX_{t-1} - ce_{t-1}(\theta) - bX_{t-1}e_{t-1}(\theta), \tag{5}$$

а γ – индекс начала суммирования, обычно принимаемый равным 1.

Ключевой особенностью функции потерь для билинейной модели является её нелинейная зависимость от параметров, что существенно усложняет задачу минимизации.

В частности, невыпуклость критерия $Q(\theta)$ может приводить к множеству локальных минимумов, поэтому для надёжного поиска глобального решения зачастую требуются специальные (глобальные или гибридные) методы оптимизации [8].

Другая особенность заключается в рекурсивном характере вычисления ошибок $e_t(\theta)$, что связано с присутствием в модели как прошлых значений ряда, так и прошлых значений обновляющего процесса. Это требует задания начальных условий для итеративного вычисления ошибок. Обычно полагают $e_0 = 0$ или используют некоторую процедуру инициализации на основе предварительных оценок параметров.

1.3. Гибридный метод оптимизации для оценивания параметров билинейных моделей

Для оценивания параметров билинейных авторегрессионных моделей в работе используется комбинированный (гибридный) метод оптимизации, сочетающий глобальный поиск методом дифференциальной эволюции (DE) с последующим локальным уточнением решения методом L-BFGS-B. Такой подход позволяет преодолеть основные проблемы, возникающие при оптимизации нелинейных моделей: наличие множества локальных минимумов и сложную геометрию поверхности функции потерь.

1.4. Метод дифференциальной эволюции (DE)

Дифференциальная эволюция – стохастический популяционный метод глобальной оптимизации [9]. Метод основан на принципах эволюционных алгоритмов и особенно эффективен для поиска глобального оптимума в многомерных пространствах с невыпуклыми целевыми функциями.

Основные этапы метода DE:

- Инициализация популяции. Создание начальной популяции из NP векторов параметров $x_i, i = 1, \dots, NP$, равномерно распределенных в допустимой области:

$$x_i = l + rand(0{,}1) \cdot (u - l), \tag{6}$$

где l и u – векторы нижних и верхних границ параметров.

- Мутация. Для каждого члена популяции (целевого вектора) создается мутантный вектор путем линейной комбинации нескольких случайно выбранных векторов из текущей популяции:

$$v_i = x_{r1} + F \cdot (x_{r2} - x_{r3}), \tag{7}$$

где $r1, r2, r3$ – случайные различные индексы, отличные от i, F – коэффициент мутации.

- Кроссовер. Формирование пробного вектора u_i путем смешивания компонент мутантного вектора v_i и целевого вектора x_i:

$$u_{i,j} = \begin{cases} v_{i,j} & \text{если } rand(0{,}1) \le CR \text{ или } j = j_{rand} \\ x_{i,j} & \text{иначе} \end{cases}, \tag{8}$$

где CR – вероятность кроссовера, j_{rand} – случайно выбранный индекс, обеспечивающий, что хотя бы один элемент от мутантного вектора перейдет в пробный вектор.

- Селекция. Сравнение целевой функции для пробного вектора u_i и текущего вектора x_i лучший из них сохраняется в новом поколении:

$$x_{i,new} = \begin{cases} u_i, \text{ если } f(u_i) \le f(x_i) \\ x_i, \text{ иначе.} \end{cases} \tag{9}$$

Эти шаги повторяются для каждого поколения до достижения критерия останова.

Преимущества DE для билинейных моделей:

- Глобальный характер поиска. DE с высокой вероятностью находит область глобального минимума даже при сложной топологии поверхности потерь [10].
- Нетребовательность к свойствам функции. Не требует дифференцируемости или непрерывности целевой функции.
- Устойчивость к шуму. Высокая устойчивость к зашумленным данным и резким изменениям в функции потерь.
- Простота реализации. Относительно небольшое число контролируемых параметров и простая логика алгоритма.

1.5. Метод L-BFGS-B

L-BFGS-B [11] – это квазиньютоновский метод локальной оптимизации, специально разработанный для задач с ограничениями в виде границ на переменные. Метод эффективно аппроксимирует обратный гессиан без необходимости явного его вычисления и хранения.

Основные принципы L-BFGS-B.

Итерационная формула метода имеет вид:

$$x_{k+1} = x_k - \alpha_k H_k \nabla f(x_k), \tag{10}$$

где:

x_k – вектор параметров на итерации k,

α_k – параметр шага, определяемый линейным поиском с соблюдением условий Вольфе,

H_k – приближение обратного гессиана на итерации k,

$\nabla f(x_k)$ – градиент целевой функции.

Приближение обратного гессиана обновляется неявно с использованием ограниченного числа пар векторов $s_j, y_j{}_{j=k-m}^{k-1}$, где:

$s_j = x_{j+1} - x_j$ – изменение в точках,

$y_j = \nabla f\left(x_{j+1}\right) - \nabla f\left(x_j\right)$ – изменение в градиенте.

Преимущества L-BFGS-B для локального уточнения:

- Высокая скорость сходимости вблизи оптимума благодаря использованию информации второго порядка [12].
- Эффективное использование памяти за счет хранения только m последних пар векторов вместо полной матрицы гессиана.
- Поддержка границ на параметры, что критически важно для обеспечения стационарности билинейной модели.

1.6. Реализация гибридного метода DE + L-BFGS-B

Комбинированный подход реализуется в два последовательных этапа:

Глобальный поиск методом DE для нахождения области глобального минимума. На этом этапе используется популяция векторов параметров и эволюционные операторы для обширного исследования пространства параметров.

Локальное уточнение методом L-BFGS-B, начиная с лучшего решения, найденного методом DE. На этом этапе используется градиентная информация для быстрого уточнения решения с высокой точностью.

Преимущества гибридного метода для оценивания параметров билинейных моделей:

- Уверенное нахождение глобального минимума. DE обеспечивает широкий обзор пространства параметров и устойчивость к локальным минимумам, что особенно важно при сложной поверхности функции потерь, характерной для билинейных моделей.
- Высокая точность окончательного решения. L-BFGS-B обеспечивает быструю сходимость и высокую точность в окрестности найденного решения.
- Устойчивость к нестандартным распределениям. Комбинированный метод демонстрирует высокую эффективность при различных распределениях обновляющего процесса, включая распределения с тяжелыми хвостами (Стьюдента с малыми степенями свободы) и загрязненные распределения (Тьюки).
- Эффективная обработка ограничений. Метод естественным образом учитывает ограничения стационарности, обеспечивая корректное оценивание параметров билинейной модели.
- Вычислительная эффективность. Двухэтапный подход позволяет сбалансировать глобальный поиск и локальное уточнение, обеспечивая высокую точность при приемлемых вычислительных затратах.

Таким образом, гибридный метод DE + L-BFGS-B представляет собой мощный инструмент для оценивания параметров билинейных авторегрессионных моделей, особенно в условиях нестандартных распределений обновляющего процесса, когда поверхность функции потерь имеет сложную структуру с множеством локальных минимумов.

2. Методология исследования

Для оценивания параметров билинейной авторегрессионной модели применялся метод наименьших квадратов с комбинированной оптимизацией. Исследование проводилось на основе синтетических временных рядов, генерируемых согласно модели:

$$X_t = a \cdot X_{\{t-1\}} + e_t + c \cdot e_{\{t-1\}} + b \cdot X_{\{t-1\}} \cdot e_{\{t-1\}}, \tag{11}$$

где:

$\mathrm{X_t}$ – значение временного ряда в момент времени t,

$\mathrm{e_t}$ – обновляющий процесс (случайная составляющая) с заданным распределением,

a, b, c – параметры модели, которые нужно оценить.

Для моделирования использовались следующие типы распределений обновляющего процесса:

- Нормальное распределение (стандартное гауссовское) – как базовое для сравнений.
- Загрязнённое нормальное распределение (Тьюки) с различными параметрами загрязнения:

γ – доля загрязнения (от 0.05 до 0.15),

τ – масштаб загрязнения (от 3 до 10).

- Распределение Стьюдента с различными степенями свободы (df от 1 до 10).
- Распределение Лапласа (двойное экспоненциальное).
- Равномерное распределение.

Все распределения были нормализованы для обеспечения единичной дисперсии, что позволило корректно сравнивать эффективность оценивания параметров при различных типах распределений. Размер выборки каждого временного ряда составлял 500 наблюдений, что обеспечивало статистически значимые оценки параметров модели.

Для обеспечения стационарности соблюдалось условие:

$$a^2 + b^2 \cdot \sigma^2 < 1,$$

где σ^2 – дисперсия обновляющего процесса (в нашем случае нормализована до 1).

Параметры модели генерировались с учетом данного условия, чтобы обеспечить стационарность временного ряда. Для каждого типа распределения методом последовательных попыток выбирались такие комбинации параметров a, b и c, которые удовлетворяли условию стационарности.

Таблица 1. Сравнение распределений

	СКО (ДИ 95%)	**Ошибка a (ДИ 95%)**	**Ошибка b (ДИ 95%)**	**Ошибка c (ДИ 95%)**	**Смещение a (ДИ 95%)**	**Смещение b (ДИ 95%)**	**Смещение c (ДИ 95%)**
Нормальное	0.0055 [0.0049, 0.0061]	0.0660 [0.0615, 0.0705]	0.0291 [0.0270, 0.0312]	0.0713 [0.0665, 0.0761]	-0.0066 [-0.0139, 0.0007]	0.0002 [-0.0031, 0.0035]	0.0034 [-0.0045, 0.0114]
Лапласа	0.0056 [0.0049, 0.0064]	0.0698 [0.0650, 0.0746]	0.0209 [0.0189, 0.0230]	0.0700 [0.0651, 0.0749]	-0.0059 [-0.0137, 0.0019]	-0.0018 [-0.0046, 0.0009]	0.0016 [-0.0063, 0.0095]
Стьюдента d=10	0.0055 [0.0048, 0.0061]	0.0675 [0.0627, 0.0723]	0.0249 [0.0228, 0.0270]	0.0688 [0.0640, 0.0736]	0.0023 [-0.0054, 0.0099]	-0.0015 [-0.0045, 0.0015]	-0.0039 [-0.0116, 0.0038]
Стьюдента d=3	0.0115 [0.0085, 0.0145]	0.0791 [0.0734, 0.0847]	0.0262 [0.0223, 0.0301]	0.0893 [0.0791, 0.0995]	-0.0062 [-0.0151, 0.0027]	-0.0112 [-0.0157, -0.0068]	0.0098 [-0.0031, 0.0227]
Стьюдента d=5	0.0069 [0.0059, 0.0079]	0.0772 [0.0716, 0.0828]	0.0209 [0.0189, 0.0230]	0.0713 [0.0654, 0.0773]	-0.0127 [-0.0214, -0.0039]	-0.0035 [-0.0062, -0.0007]	0.0098 [0.0012, 0.0184]
Равномерное	0.0061 [0.0054, 0.0068]	0.0711 [0.0661, 0.0761]	0.0363 [0.0339, 0.0386]	0.0704 [0.0656, 0.0752]	-0.0042 [-0.0122, 0.0038]	-0.0028 [-0.0068, 0.0011]	-0.0024 [-0.0103, 0.0054]

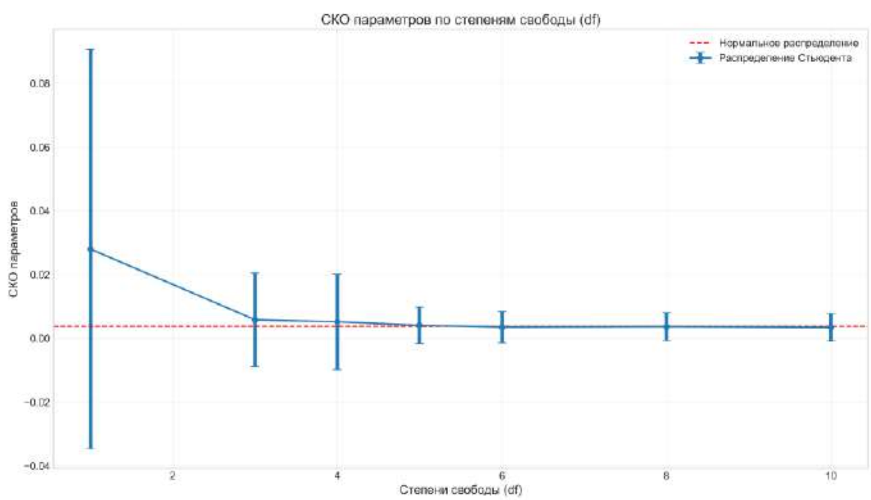

Рис. 1. Зависимость СКО параметров от степени свободы распределения Стьюдента

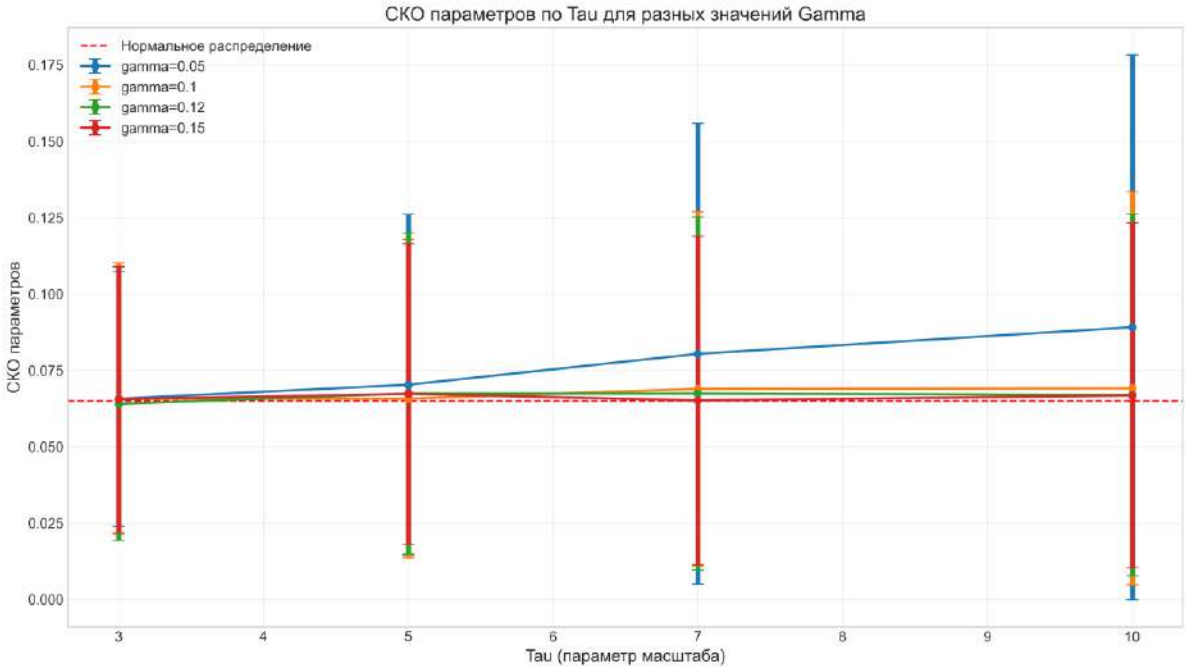

Рис. 2. Зависимость СКО параметров от γ и τ для распределения Тьюки

3. Заключение

Проведенное исследование выявило существенное влияние распределения обновляющего процесса на точность МНК-оценок параметров билинейной авторегрессионной модели. Гибридный метод DE + L-BFGS-B продемонстрировал высокую эффективность даже при распределениях с тяжелыми хвостами.

Основные результаты:

Подтверждена оптимальность МНК при нормальном распределении – получены минимальные ошибки оценивания (СКО = 0.0055) для всех параметров модели.

Установлена критическая чувствительность билинейного параметра b к типу распределения. При распределении Стьюдента с df=3 ошибка оценивания b увеличивается незначительно (с 0.0291 до 0.0262), тогда как ошибки параметров a и c возрастают на 20-25%.

Для загрязненного нормального распределения выявлено, что наибольшие сложности создает сочетание малой доли загрязнения (γ=0.05) с большим масштабом ($\tau \geq 7$), порождающее редкие крупные выбросы.

Обнаружена нелинейная зависимость точности оценок от числа степеней свободы распределения Стьюдента: резкое ухудшение при df<3 и асимптотическое приближение к нормальному случаю при df>5.

Апробация билинейной модели с использованием метода L-BFGS-B и МНК-оценивания была проведена на предприятии – участнике проекта «Сколково» – компании NeoDetect. Полученные результаты продемонстрировали более высокое качество предсказаний по сравнению с ранее применявшимся подходом, основанным на использовании трёх стандартных отклонений.

Практические рекомендации:

- При работе с финансовыми временными рядами (часто имеющими свойства распределения Стьюдента с малыми df) следует применять процедуры коррекции смещения оценок или использовать робастные методы оценивания.
- Для данных с потенциальными выбросами рекомендуется предварительная диагностика типа распределения и применение гибридных методов оптимизации для повышения надежности оценок.
- При подозрении на тяжелые хвосты в распределении следует с осторожностью интерпретировать оценки билинейного параметра и увеличивать объем выборки для компенсации потери эффективности.

Дальнейшие исследования целесообразно направить на разработку адаптивных процедур оценивания, автоматически учитывающих особенности распределения шума, и апробацию методологии на реальных данных различных предметных областей.

Литература

1. *Brockwell P.J., Davis R.A.* Introduction to Time Series and Forecasting. – Berlin: Springer, 2002. – С. 383–390.
2. *Hamilton J.D.* Time Series Analysis. – Princeton (NJ): Princeton University Press, 1994. – С. 72–108.
3. *Thavaneswaran A., Subba Rao T.* Least-squares estimation in bilinear models with Gaussian errors // Journal of Time Series Analysis. – 1990. – Vol. 11, № 1. — P. 95–102.
4. *Huber P.J., Ronchetti E.M.* Robust Statistics. - Hoboken (NJ): John Wiley & Sons, 2009. – С. 68–153.
5. *Горяинов В.Б., Горяинова Е.Р.* Робастное оценивание в пороговой авторегрессии // Вестник МГТУ им. Н.Э. Баумана. Серия Естественные науки. – 2017. – № 6 (75). – С. 19–30.
6. *Fletcher R.* Practical Methods of Optimization. - Hoboken (NJ): John Wiley & Sons, 2013. – С. 120–140.
7. *Granger C.W.J., Andersen A.P.* An Introduction to Bilinear Time Series Models. – Göttingen: Vandenhoeck & Ruprecht, 1978. – 88 p.
8. *Brunner A.D., Hess G.D.* Potential problems in estimati ng bilinear time-series models // Journal of Economic Dynamics and Control. – 1995. – Vol. 19, № 4. – P. 663–681.
9. *Storn R., Price K.* Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces // Journal of Global Optimization. – 1997. – Vol. 11, № 4. – P. 341–359.
10. *Price K.V., Storn R.M., Lampinen J.A.* Differential Evolution: A Practical Approach to Global Optimization. – Berlin: Springer, 2005. – 538 p.
11. *Liu D.C., Nocedal J.* On the limited-memory BFGS method for large scale optimization // Mathematical Programming. – 1989. – Vol. 45, № 1. — P. 503–528.
12. *Zhu C., Byrd R.H., Lu P., Nocedal J.* Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization // ACM Transactions on Mathematical Software. – 1997. – Vol. 23, № 4. – P. 550–560.