

АНАЛИЗ ЭФФЕКТИВНОСТИ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ С L-BFGS-B ПРИ ОЦЕНИВАНИИ ПАРАМЕТРОВ БИЛИНЕЙНОЙ АВТОРЕГРЕССИИ В УСЛОВИЯХ РАЗЛИЧНЫХ РАСПРЕДЕЛЕНИЙ ШУМА

В.Б. Горяинов

М.М. Масягин

В.М. Семерня

vb-goryainov@bmstu.ru

masyagin@bmstu.ru

semernyavm@student.bmstu.ru

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

Аннотация

Исследовано влияние различных распределений обновляющего процесса на точность оценок параметров билинейной авторегрессионной модели, полученных методом наименьших квадратов с применением алгоритма оптимизации L-BFGS-B. Исследование эффективности оценивания параметров проведено с использованием метода наименьших квадратов для нормального и загрязненного нормального (распределение Тьюки) распределений, распределений Стьюдента с различными степенями свободы и Лапласа, равномерного распределения. Особое внимание уделено анализу зависимости точности оценок от степеней свободы в распределении Стьюдента и параметров загрязнения в распределении Тьюки. Расчеты ошибок и смещений параметров модели проведены на основе компьютерного моделирования с генерацией стационарных временных рядов. Показано, что точность оценок существенно зависит от типа распределения обновляющего процесса, при этом наибольшая чувствительность наблюдается для параметра при нелинейном члене модели. Для распределения Стьюдента установлена нелинейная зависимость ошибок оценивания от числа степеней свободы, для загрязненного нормального распределения выявлено критическое влияние параметров загрязнения на точность оценок. Полученные результаты могут быть использованы при построении билинейных моделей временных рядов в условиях нестандартных распределений остатков

Ключевые слова

Билинейная авторегрессия, робастные функции потерь, оптимизация L-BFGS-B, среднеквадратичная ошибка, нормальное распределение, распределение Лапласа, распределение Стьюдента, равномерное распределение

Поступила 22.08.2025

Принята 17.04.2026

© Автор(ы), 2026

Введение. Все большую значимость в различных областях (экономика, финансы, метеорология, гидрология и технические системы) приобретает моделирование нелинейных временных рядов. В последние десятилетия стало очевидно, что многие реальные процессы не могут быть адекватно описаны линейными моделями, что привело к активному развитию нелинейных методов анализа временных рядов [1]. Среди различных классов нелинейных моделей особое место занимают билинейные авторегрессионные модели, которые представляют собой естественное обобщение классических линейных моделей ARMA, включая дополнительные члены, описывающие произведение прошлых значений ряда и обновляющего процесса.

Билинейные модели обладают способностью эффективно описывать явления, содержащие нелинейные взаимодействия и проявляющие асимметричную реакцию на различные возмущения, что делает их особенно ценными при моделировании экономических и финансовых процессов, где присутствуют структурные сдвиги, внезапные всплески и затухающие колебания [2]. Кроме того, билинейные модели успешно применяют для описания явлений, относящихся к природе (речной сток, температурные ряды, концентрации загрязняющих веществ), а также к техническим системам, включая обработку сигналов и управление. Недавние исследования подтверждают эффективность билинейных структур в современных методах прогнозирования [3].

При построении билинейных моделей ключевым этапом является оценивание их параметров. Для этого широко используют метод наименьших квадратов (МНК), оптимальность которого теоретически обоснована для случая, когда обновляющий процесс имеет нормальное распределение. Однако на практике нередко наблюдаются отклонения от нормальности, проявляющиеся в асимметрии распределения или наличии тяжелых хвостов. Это может быть вызвано присутствием выбросов, структурными изменениями во временном ряде или фундаментальными свойствами порождающего процесса. В таких случаях эффективность МНК может существенно снижаться, что приводит к смещенным оценкам параметров и ошибочным выводам.

В литературе достаточно подробно исследована робастность различных методов оценивания для линейных моделей авторегрессии и скользящего среднего [4]. Однако влияние различных типов распределения обновляющего процесса на точность оценок параметров билинейных моделей недостаточно изучено. В частности, открытым остается вопрос о зависимости ошибок оценивания от параметров распределения Стьюдента и загрязнен-

ного нормального распределения, которые часто используются для моделирования данных с тяжелыми хвостами.

Отдельная проблема — выбор алгоритма оптимизации для минимизации функции потерь МНК. Традиционно используемый алгоритм Гаусса — Ньютона может иметь проблемы сходимости при нелинейных моделях и нестандартных распределениях. Альтернативой является алгоритм L-BFGS (Limited-memory Broyden — Fletcher — Goldfarb — Shanno) [5], который имеет потенциальные преимущества при работе с негладкими функциями и сложной геометрией поверхности потерь.

Цель работы — представить результаты исследования влияния различных распределений обновляющего процесса на точность оценок параметров билинейной авторегрессионной модели (БАРМ), полученных МНК с применением алгоритма оптимизации L-BFGS. Объект исследования — частная модель билинейной авторегрессии первого порядка. В рамках поставленной цели решаются следующие задачи:

- сравнение точности оценок при нормальном и загрязненном нормальном распределениях, распределении Стьюдента с различными степенями свободы, распределении Лапласа и равномерном распределении обновляющего процесса;
- детальный анализ зависимости ошибок оценивания от степеней свободы в распределении Стьюдента;
- исследование влияния параметров загрязнения в распределении Тьюки на точность оценок;
- анализ смещения оценок параметров при различных распределениях.

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

Билинейная авторегрессионная модель. Билинейная авторегрессионная модель представляет собой расширение традиционной авторегрессии, включающее в себя нелинейные взаимодействия между переменными, что позволяет учесть более сложные зависимости во временных рядах [6]. В стандартной авторегрессионной модели (АРМ) значения временного ряда зависят только от собственных предыдущих значений, тогда как в БАРМ они зависят от предыдущих значений ряда и взаимодействий между предыдущими значениями и случайными ошибками.

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

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

где X_t — значение наблюдаемой величины на шаге t ; p, r, m, k — параметры, определяющие порядок модели; a_j, b_{ij}, c_j — параметры, подлежащие оценке, причем $c_0 = 1$.

Обновляющий процесс (шум) e_t является последовательностью независимых, одинаково распределенных 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 обновляющего процесса, что важно для моделирования нелинейных временных зависимостей.

В исследовании рассмотрена стационарная БАРМ, для которой предполагается выполнение условий стационарности. В общем случае эти условия довольно сложны, но для частной модели, например, $X_t = aX_{t-1} + e_t + bX_{t-1}e_{t-1}$, условие стационарности можно сформулировать как $a^2 + b^2\sigma_e^2 < 1$. Это условие гарантирует, что значения случайного процесса, описываемого моделью, будут ограниченными и не будут расходиться со временем, что важно для корректного оценивания параметров модели.

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

Для рассматриваемой билинейной модели $X_t = aX_{t-1} + ce_t + bX_{t-1}e_{t-1}$ функция потерь МНК определяется как $Q(\theta) = \sum_{t=\gamma+1}^N e_t^2(\theta)$, где γ — индекс начала суммирования, обычно $\gamma = 1$; $e_t(\theta)$ — ошибки модели, вычисляемые рекурсивно, $e_t(\theta) = X_t - aX_{t-1} - ce_{t-1}(\theta) - bX_{t-1}e_{t-1}(\theta)$; $\theta = (a, b, c)$ — вектор параметров модели, подлежащих оцениванию.

Ключевая особенность функции потерь для билинейной модели — ее нелинейная зависимость от параметров, что существенно усложняет задачу минимизации. В частности, невыпуклость критерия $Q(\theta)$ может приводить ко множеству локальных минимумов, поэтому для надежного

поиска глобального решения зачастую требуются специальные (глобальные или гибридные) методы оптимизации [7].

Еще одна особенность заключается в рекурсивном характере вычисления ошибок $e_t(\theta)$, что связано с присутствием в модели прошлых значений ряда и обновляющего процесса. Это требует задания начальных условий для итеративного вычисления ошибок. Как правило, $e_0 = 0$ или используется некоторая процедура инициализации на основе предварительных оценок параметров.

Гибридный метод оптимизации для оценивания параметров билинейных моделей. Для оценивания параметров БАРМ 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.

Метод дифференциальной эволюции. Дифференциальная эволюция — стохастический популяционный метод глобальной оптимизации [8]. Метод основан на принципах эволюционных алгоритмов и эффективен для поиска глобального оптимума в многомерных пространствах с невыпуклыми целевыми функциями. Основные этапы метода следующие.

1. *Инициализация популяции.* Создание начальной популяции из NP векторов параметров x_i , $i = 1, \dots, NP$, равномерно распределенных в допустимой области $x_i = l + \text{rand}(0, 1) \times (u - l)$, где l , u — векторы нижних и верхних границ параметров.

2. *Мутация.* Для каждого члена популяции (целевого вектора) создается мутантный вектор путем линейной комбинации нескольких случайно выбранных векторов из текущей популяции: $v_i = x_{r_1} + F(x_{r_2} - x_{r_3})$, где r_1 , r_2 , r_3 — случайные различные индексы, отличные от i ; F — коэффициент мутации.

3. *Кроссовер.* Формирование пробного вектора u_i смешиванием компонент мутантного вектора v_i и целевого вектора x_i :

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

где CR — вероятность кроссовера; j_{rand} — случайно выбранный индекс, обеспечивающий переход хотя бы одного элемента от мутантного вектора в пробный.

4. *Селекция.* Сравнение целевой функции для пробного вектора u_i и текущего вектора x_i , лучший из которых сохраняется в новом поколении:

$$x_{i, new} = \begin{cases} u_i, & \text{если } f(u_i) \leq f(x_i); \\ x_i, & \text{иначе.} \end{cases}$$

Этапы 1–4 повторяются для каждого поколения до достижения критерия останова.

Преимущества метода дифференциальной эволюции для билинейных моделей:

- глобальный характер поиска (с высокой вероятностью находит область глобального минимума даже при сложной топологии поверхности потерь [9]);
- 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).

Метод L-BFGS-B. L-BFGS-B [10] — квазиньютоновский метод локальной оптимизации, разработанный для задач с ограничениями в виде границ на переменные. Метод эффективно аппроксимирует обратный гессиан без необходимости его явного вычисления и хранения.

Основные принципы L-BFGS-B. Итерационная формула метода: $x_{k+1} = x_k - \alpha_k H_k \nabla f(x_k)$, 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 $j = k - m$, где $s_j = x_{j+1} - x_j$ — изменение в точках; $y_j = \nabla f(x_{j+1}) - \nabla f(x_j)$ — изменение в градиенте.

Преимущества L-BFGS-B для локального уточнения:

- высокая скорость сходимости вблизи оптимума ввиду использования информации второго порядка [11];
- 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.

Реализация гибридного метода дифференциальной эволюции и L-BFGS-B. Комбинированный подход реализуется в два этапа.

1. Глобальный поиск методом дифференциальной эволюции для нахождения области глобального минимума. Используется популяция векторов параметров и эволюционные операторы для обширного исследования пространства параметров.

2. Локальное уточнение методом L-BFGS-B, начиная с лучшего решения, найденного методом дифференциальной эволюции. Используется градиентная информация для быстрого уточнения решения с высокой точностью.

Математическая формулировка. Пусть $f(\theta)$ — функция потерь (например, сумма квадратов остатков билинейной модели), которую необходимо минимизировать, где $\theta = (a, b, c)$ — вектор параметров билинейной модели.

1. Этап дифференциальной эволюции: находим приближенное решение θ_{DE} , минимизирующее $f(\theta)$ с использованием метода дифференциальной эволюции $\theta_{DE} \approx \arg \min_{\theta} f(\theta)$.

2. Этап L-BFGS-B: используем θ_{DE} в качестве начального приближения для L-BFGS-B и находим уточненное решение θ^* : $\theta^* = \arg \min_{\theta} f(\theta)$, начиная с θ_{DE} .

Особенности реализации:

– обработка ограничений стационарности, для билинейной модели условие стационарности имеет вид $a^2 + b^2 \sigma_e^2 < 1$. Эти ограничения реализуются на этапе дифференциальной эволюции — через проекцию решений на допустимую область и на этапе L-BFGS-B через явные ограничения на параметры;

– многократные запуски, для повышения надежности оценок используется подход с несколькими запусками оптимизации из различных начальных точек, включая начальные значения и случайно генерируемые точки;

– адаптивный контроль параметров дифференциальной эволюции, коэффициент мутации F может адаптивно изменяться в процессе оптимизации, что позволяет балансировать между исследованием пространства параметров и уточнением найденных решений;

– условия сходимости, используются стандартные критерии остановки: достижение максимального числа итераций; достаточно малая норма градиента; недостаточное улучшение значения функции потерь на последовательных итерациях.

Преимущества гибридного метода для оценивания параметров билинейных моделей:

- уверенное нахождение глобального минимума (дифференциальная эволюция обеспечивает широкий обзор пространства параметров и устойчивость к локальным минимумам, что 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 (учитывает ограничения стационарности, обеспечивая корректное оценивание параметров билинейной модели);
- вычислительная эффективность (позволяет сбалансировать глобальный поиск и локальное уточнение, обеспечивая высокую точность при приемлемых вычислительных затратах).

Гибридный метод дифференциальной эволюции и L-BFGS-B представляет собой мощный инструмент для оценивания параметров БАРМ, особенно в условиях нестандартных распределений обновляющего процесса, когда поверхность функции потерь имеет сложную структуру со множеством локальных минимумов. Эффективность гибридных методов оптимизации для нелинейных временных рядов подтверждена результатами исследований [12].

Распределения шума. *Стандартное нормальное распределение.* Нормальное (гауссово) распределение традиционно является отправной точкой при статистическом моделировании ввиду его математических свойств и центральной роли в теории вероятностей. Плотность нормального распределения с нулевым средним и дисперсией σ^2 задается формулой

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2 / (2\sigma^2)}.$$

В контексте моделирования временных рядов предположение о нормальном распределении обновляющего процесса является стандартным и обладает несколькими теоретическими преимуществами:

– для линейных моделей с нормально распределенными инновациями оценка наименьших квадратов совпадает с оценкой максимального правдоподобия, обеспечивая оптимальность в классе несмещенных оценок [13];

– нормальное распределение полностью определяется первыми двумя моментами (средним и дисперсией), что упрощает теоретический анализ и интерпретацию результатов;

– при нормальном распределении обновляющего процесса для линейных моделей распределение самого процесса также является нормальным в стационарном режиме [14].

Однако для билинейных моделей ситуация существенно сложнее. Даже при нормальном распределении шума процесс не будет нормальным в результате наличия нелинейного члена. Это приводит к тому, что статистические свойства билинейного процесса, включая моменты высших порядков, существенно отличаются от случая линейных моделей.

Несмотря на перечисленные сложности, нормальное распределение остается важным базовым случаем для анализа билинейных моделей и служит точкой отсчета при сравнении с другими распределениями. В проведенном авторами исследовании в качестве референтного случая использовано стандартное нормальное распределение $\sigma^2 = 1$.

Распределение Стьюдента — одно из наиболее используемых распределений для моделирования данных с тяжелыми хвостами, часто наблюдаемыми в финансовых и экономических временных рядах. Применение этого распределения в контексте финансовых временных рядов активно исследуется в литературе [15]. Плотность распределения с числом ν степеней свободы задается формулой

$$f(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + x^2/\nu\right)^{-(\nu+1)/2},$$

где Γ — гамма-функция.

Ключевой параметр распределения Стьюдента — число степеней свободы, которое определяет его форму и статистические свойства:

– при $\nu = 1$ распределение Стьюдента совпадает с распределением Коши, которое имеет настолько тяжелые хвосты, что у него нет математического ожидания;

– при $1 < \nu \leq 2$ у распределения существует математическое ожидание (равное нулю в центрированном случае), но отсутствует конечная дисперсия;

- при $2 < \nu \leq 4$ распределение имеет конечные первый и второй моменты, но не имеет конечного четвертого момента;
- при $\nu > 4$ существуют все моменты до четвертого включительно, после четвертого момента дальнейшие моменты появляются по мере роста числа степеней свободы, каждый следующий требует ν строго больше соответствующего порядка;
- при $\nu \rightarrow \infty$ распределение Стьюдента стремится к стандартному нормальному.

Такая зависимость статистических свойств от числа степеней свободы делает распределение Стьюдента гибким инструментом для моделирования различных типов отклонений от нормальности, от экстремальных случаев с очень тяжелыми хвостами до практически нормального распределения.

Загрязненное нормальное распределение (Тьюки) представляет собой смесь двух нормальных распределений и часто используется для моделирования данных с локальными отклонениями от нормальности. Плотность загрязненного нормального распределения задается формулой

$$f(x, \gamma, \tau) = (1 - \gamma)\varphi(x) + \gamma\varphi\left(\frac{x}{\tau}\right)\frac{1}{\tau},$$

где γ — доля загрязнения; τ — масштаб загрязнения; $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ — плотность стандартного нормального распределения.

Загрязненное нормальное распределение представляет собой гибкий способ моделирования различных степеней отклонения от нормальности, от незначительных (при малых γ и τ) до существенных (при больших γ и τ). Влияние загрязненных распределений на оценки параметров остается активной областью исследований [16]. В контексте билинейных моделей это позволяет исследовать влияние на оценки параметров как редких, но сильных выбросов (малое γ , большое τ), так и частых, но умеренных отклонений от нормальности (большие γ , τ) [17].

Распределение Лапласа (двойное экспоненциальное распределение) характеризуется более острым пиком и более тяжелыми хвостами по сравнению с нормальным распределением. Плотность распределения Лапласа с нулевым средним и параметром масштаба b задается формулой

$$f(x; b) = \frac{1}{2b}e^{-|x|/b}.$$

Дисперсия распределения Лапласа $2b^2$. Здесь использован параметр $b = 1/\sqrt{2}$, что дает дисперсию, равную единице, для сопоставимости с другими распределениями.

Распределение Лапласа имеет коэффициент эксцесса, равный трем, что указывает на более тяжелые хвосты по сравнению с нормальным распределением. В отличие от распределения Стьюдента с малым числом степеней свободы, у распределения Лапласа существуют все моменты.

Особый интерес к распределению Лапласа при оценивании параметров связан с тем, что при таком распределении обновляющего процесса метод наименьших модулей (минимизация суммы абсолютных отклонений) является оптимальным в смысле максимального правдоподобия. Это создает естественный контраст с МНК, оптимальным при нормальном распределении, и позволяет оценить потерю эффективности МНК при отклонении от нормальности в сторону распределения Лапласа.

Равномерное распределение представляет собой простейший пример распределения с ограниченной областью определения и отсутствием хвостов. Плотность равномерного распределения на отрезке $[a, b]$ постоянна и равна

$$f(x) = \frac{1}{b-a} \quad \text{для } a \leq x \leq b.$$

Включение равномерного распределения в исследование связано с желанием охватить различные типы отклонений от нормальности как в сторону утяжеления хвостов, так и в сторону их облегчения.

Теоретические предпосылки влияния распределений на точность оценок. Влияние распределения обновляющего процесса на точность оценок параметров билинейной модели можно проанализировать по нескольким ключевым позициям.

Эффективность МНК. Для билинейной модели с гауссовым распределением ошибок МНК совпадает с оценкой максимального правдоподобия и достигает нижней границы Крамера — Рао [18]. При отклонении от нормальности эффективность снижается, особенно для распределений с тяжелыми хвостами.

Моменты распределения шума. Условия стационарности и свойства оценок зависят от моментов распределения шума, причем для некоторых распределений (например, Стьюдента с малыми степенями свободы) не существует отдельных высших моментов.

Чувствительность к выбросам. Метод наименьших квадратов известен слабой робастностью к выбросам, особенно в билинейных моделях,

ввиду нелинейного характера взаимодействия процесса с шумом [19]. Современные исследования робастных методов для временных рядов с тяжелыми хвостами показывают значительные преимущества альтернативных подходов [20, 21].

Геометрия поверхности потерь. При нормальном распределении функция потерь имеет регулярную структуру с единственным глобальным минимумом. При тяжелых хвостах возникают множественные локальные минимумы, что существенно усложняет оптимизацию.

С учетом этих позиций проверены следующие гипотезы: 1) точность оценок наивысшая при нормальном распределении; 2) параметр b наиболее чувствителен к типу распределения; 3) для распределения Стьюдента ошибка нелинейно зависит от числа степеней свободы; 4) для загрязненного нормального распределения критичным является сочетание параметров загрязнения.

Методология исследования. Для оценивания параметров БАРМ применен МНК с комбинированной оптимизацией. Исследование проведено на основе синтетических временных рядов, генерируемых по модели $X_t = aX_{t-1} + e_t + ce_{t-1} + bX_{t-1}e_{t-1}$, где X_t — значение временного ряда в момент времени t ; e_t — обновляющий процесс (случайная составляющая) с заданным распределением; a, b, c — параметры модели, которые необходимо оценить.

Для моделирования использованы следующие распределения обновляющего процесса: нормальное (стандартное гауссово) в качестве базового для сравнения; загрязненное нормальное (Тьюки) с различными параметрами загрязнения ($\gamma = 0,05-0,15$, $\tau = 3-10$); Стьюдента с различными степенями свободы ($\nu = 1-10$); Лапласа (двойное экспоненциальное); равномерное.

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 соблюдения стационарности выполнялось условие $a^2 + b^2\sigma^2 < 1$, где σ^2 — дисперсия обновляющего процесса (нормализована до единицы).

Параметры модели генерировались с учетом данного условия, чтобы обеспечить стационарность временного ряда. Для каждого типа распределения методом последовательных попыток выбирались такие комби-

нации параметров a , b и c , которые удовлетворяли условию стационарности.

Метрики оценки качества. Качество параметров модели оценивалось с помощью следующих метрик:

- среднеквадратическая ошибка (СКО) по всем параметрам

$$MSE = \left(\frac{1}{3} \right) \sum \{i \in \{a, b, c\}\} (\theta_i^{est} - \theta_i^{true})^2;$$

- абсолютные ошибки оценок каждого параметра модели

$$|\theta_i^{est} - \theta_i^{true}|, i \in \{a, b, c\};$$

- смещение оценок параметров

$$\text{Bias}(\theta_i) = E[\theta_i^{est}] - \theta_i^{true}, i \in \{a, b, c\};$$

- СКО прогноза как значение функции потерь в точке оптимума.

Для количественной оценки влияния различных типов распределений на точность оценивания параметров проведено сравнительное исследование с использованием описанных метрик. Результаты компьютерного моделирования для каждого типа распределения обновляющего процесса приведены в таблице (ДИ — доверительный интервал).

Сравнение распределений (ДИ 95 %)

СКО	Ошибка			Смещение		
	a	b	c	a	b	c
<i>Нормальное</i>						
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]
<i>Лапласа</i>						
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]
<i>Стюдента $d = 3$</i>						
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]

Окончание таблицы

СКО	Ошибка			Смещение		
	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>
<i>Стьюдента $d = 5$</i>						
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]
<i>Стьюдента $d = 10$</i>						
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]
<i>Равномерное</i>						
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]

Детальный анализ влияния числа степеней свободы распределения Стьюдента на точность оценивания параметров выявил существенную нелинейную зависимость. Наиболее значительное ухудшение качества оценок наблюдается при малых значениях степеней свободы (1–3), что объясняется наличием тяжелых хвостов распределения. Графическое представление этой зависимости показано на рис. 1.

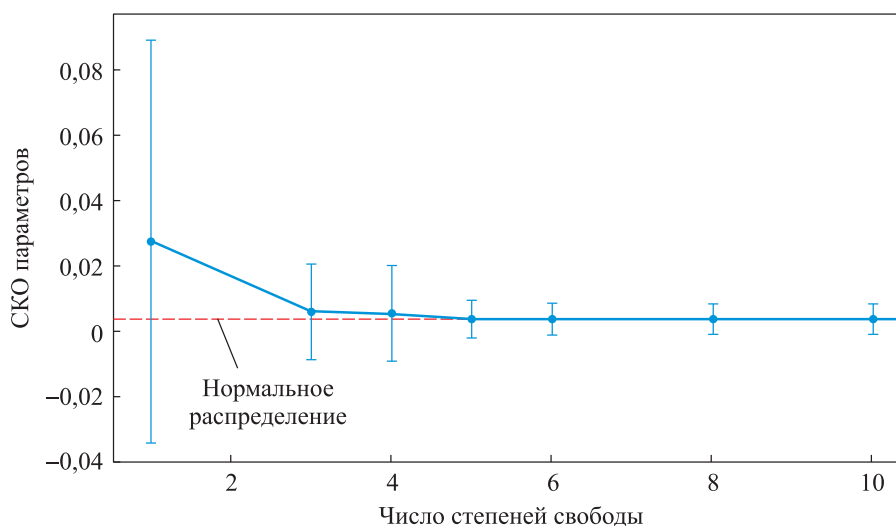


Рис. 1. Зависимость СКО параметров от числа степеней свободы распределения Стьюдента (—)

Исследование загрязненного нормального распределения (Тьюки) показало критическую важность сочетания параметров загрязнения γ (доля загрязнения) и τ (масштаб загрязнения). Результаты анализа различных комбинаций этих параметров продемонстрировали, что наибольшие ошибки оценивания возникают при определенных сочетаниях значений γ и τ . Зависимость СКО от параметров загрязнения показана на рис. 2.

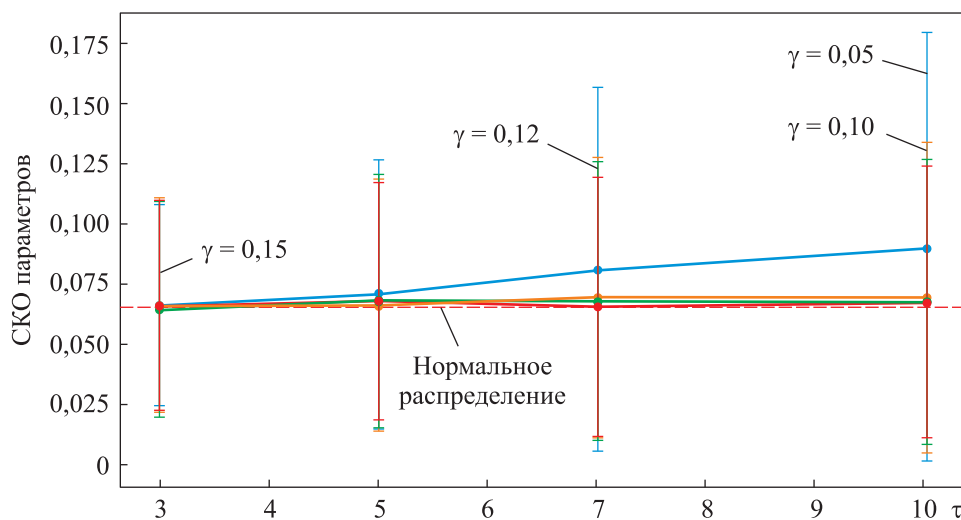


Рис. 2. Зависимость СКО параметров от γ и τ для распределения Тьюки

Заключение. Проведенное исследование эффективности МНК с комбинированной оптимизацией для оценивания параметров БАРМ выявило значительное влияние распределения обновляющего процесса на точность оценок.

Разработанная методология, сочетающая глобальный поиск методом дифференциальной эволюции с локальным уточнением L-BFGS-B, продемонстрировала высокую надежность при распределениях с тяжелыми хвостами.

Ключевые результаты показали, что нормальное распределение обеспечивает наименьшие ошибки оценивания. Для загрязненного нормального распределения критичным является сочетание малой доли загрязнения с большим масштабом хвоста. Распределение Стюдента демонстрирует нелинейную зависимость ошибок от числа степеней свободы. Параметр билинейного члена оказался наиболее чувствительным к типу распределения.

Рекомендуется осторожное отношение к оценкам билинейного параметра при подозрении на тяжелые хвосты в данных и применение методов компенсации смещений для финансовых временных рядов.

Перспективными направлениями являются разработка адаптивных процедур оценивания и применение методологии к реальным данным различных предметных областей.

ЛИТЕРАТУРА

- [1] Brockwell P.J., Davis R.A. Introduction to time series and forecasting. Springer, 2002.
- [2] Hamilton J.D. Time series analysis. Princeton Univ. Press, 1994.
- [3] Gao S., Zhou M., Wang Y., et al. Dendritic neuron model with effective learning algorithms for classification, approximation, and prediction. *IEEE Trans. Neural Netw. Learn. Syst.*, 2019, vol. 30, no. 2, pp. 601–614.
DOI: <https://doi.org/10.1109/TNNLS.2018.2846646>
- [4] Huber P.J., Ronchetti E.M. Robust statistics. Wiley, 2009.
- [5] Fletcher R. Practical methods of optimization. Wiley, 2013.
- [6] Granger C.W.J., Andersen A.P. An introduction to bilinear time series models. Vandenhoeck & Ruprecht, 1978.
- [7] Brunner A.D., Hess G.D. Potential problems in estimating bilinear time-series models. *J. Econ. Dyn. Control*, 1995, vol. 19, no. 4, pp. 663–681.
- [8] Storn R., Price K. Differential evolution — a simple and efficient heuristic for global optimization over continuous spaces. *J. Glob. Optim.*, 1997, vol. 11, no. 4, pp. 341–359.
DOI: <https://doi.org/10.1023/A:1008202821328>
- [9] Price K.V., Storn R.M., Lampinen J.A. Differential evolution. Springer, 2005.
- [10] Liu D.C., Nocedal J. On the limited-memory BFGS method for large scale optimization. *Math. Program.*, 1989, vol. 45, pp. 503–528.
DOI: <https://doi.org/10.1007/BF01589116>
- [11] Riani M., Atkinson A.C., Corbellini A., et al. Robust regression with density power divergence: theory, comparisons, and data analysis. *Entropy*, 2020, vol. 22, no. 4, art. 399. DOI: <https://doi.org/10.3390/e22040399>
- [12] Zhu C., Byrd R.H., Lu P., et al. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.*, 1997, vol. 23, no. 4, pp. 550–560. DOI: <https://doi.org/10.1145/279232.279236>
- [13] Agostinelli C., Valdora M., Yohai V.J. Initial robust estimation in generalized linear models. *Comput. Stat. Data Anal.*, 2019, vol. 134, pp. 144–156.
DOI: <https://doi.org/10.1016/j.csda.2018.12.010>
- [14] Yao W., Li L. A new regression model: modal linear regression. *Scand. J. Stat.*, 2014, vol. 41, no. 3, pp. 656–671. DOI: <https://doi.org/10.1111/sjos.12054>
- [15] Броквелл П.Дж., Дэвис Р.А. Введение в анализ временных рядов. М., Мир, 2004.

- [16] Chen Y., Xie Y., Ma X., et al. Probabilistic forecasting with temporal convolutional neural network. *Neurocomputing*, 2020, vol. 399, pp. 491–501.
DOI: <https://doi.org/10.1016/j.neucom.2020.03.011>
- [17] Billard L. Time series models. In: *Biometrics*. Vol. 1. UNESCO, 1997.
- [18] Sesay S.A.O., Subba Rao T. Frequency-domain estimation of bilinear time series models. *J. Time Ser. Anal.*, 1992, vol. 13, no. 6, pp. 521–545.
DOI: <https://doi.org/10.1111/j.1467-9892.1992.tb00124.x>
- [19] Горяинов В.Б., Горяинова Е.Р. Робастное оценивание в пороговой авторегрессии. *Вестник МГТУ им. Н.Э. Баумана. Сер. Естественные науки*, 2017, № 6 (75), с. 19–30. DOI: <http://dx.doi.org/10.18698/1812-3368-2017-6-19-30>
- [20] Zhu D., Galbraith J.W. Modeling and forecasting expected shortfall with the generalized asymmetric Student-*t* and asymmetric exponential power distributions. *J. Empir. Finance*, 2011, vol. 18, no. 4, pp. 765–778.
DOI: <https://doi.org/10.1016/j.jempfin.2011.05.006>
- [21] Insolia L., Kenney A., Chiaromonte F., et al. Simultaneous feature selection and outlier detection with optimality guarantees. *Biometrics*, 2022, vol. 78, no. 4, pp. 1592–1603. DOI: <https://doi.org/10.1111/biom.13553>

Горяинов Владимир Борисович — д-р физ.-мат. наук, профессор кафедры «Математическое моделирование» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1).

Масягин Михаил Михайлович — аспирант, ассистент кафедры «Математическое моделирование» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1).

Семерня Владислав Михайлович — аспирант кафедры «Математическое моделирование» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1).

Просьба ссылаться на эту статью следующим образом:

Горяинов В.Б., Масягин М.М., Семерня В.М. Анализ эффективности метода наименьших квадратов с L-BFGS-B при оценивании параметров билинейной авторегрессии в условиях различных распределений шума. *Вестник МГТУ им. Н.Э. Баумана. Сер. Естественные науки*, 2026, № 3 (126), с. 4–23. EDN: VVLPJN

**ANALYSIS OF LEAST SQUARES METHOD EFFICIENCY
WITH L-BFGS-B OPTIMIZATION FOR BILINEAR
AUTOREGRESSION PARAMETER ESTIMATION
UNDER VARIOUS NOISE DISTRIBUTIONS**

V.B. Goryainov

M.M. Masyagin

V.M. Semernya

vb-goryainov@bmstu.ru

masyagin@bmstu.ru

semernyavm@student.bmstu.ru

BMSTU, Moscow, Russian Federation

Abstract

The article investigates the influence of various distributions of the updating process on the accuracy of parameter estimates for a bilinear autoregressive model, obtained using the least squares method with the L-BFGS-B optimization algorithm. The research examines the effectiveness of parameter estimation using the least squares method for normal, contaminated normal (Tukey distribution), Student's t -distribution with various degrees of freedom, Laplace distribution, and uniform distribution. Special attention is given to analyzing the dependence of estimation accuracy on the degrees of freedom in the Student t -distribution and contamination parameters in the Tukey distribution. Calculations of errors and biases of the model parameters are conducted based on extensive computer modeling with the generation of stationary time series. The results show that the accuracy of estimates significantly depends on the type of distribution of the updating process, with the highest sensitivity observed for the parameter associated with the nonlinear term of the model. For the Student's t -distribution, a nonlinear relationship between estimation errors and the number of degrees of freedom is established, while for the contaminated normal distribution, a critical influence of contamination parameters on the accuracy of estimates is revealed. The obtained results can be used in constructing bilinear time series models under conditions of non-standard residual distributions

Keywords

Bilinear autoregression, robust loss functions, L-BFGS-B optimization, mean square error, normal distribution, Laplace distribution, Student's t -distribution, uniform distribution

Received 22.08.2025

Accepted 17.04.2026

© Author(s), 2026

REFERENCES

- [1] Brockwell P.J., Davis R.A. Introduction to time series and forecasting. Springer, 2002.
- [2] Hamilton J.D. Time series analysis. Princeton Univ. Press, 1994.
- [3] Gao S., Zhou M., Wang Y., et al. Dendritic neuron model with effective learning algorithms for classification, approximation, and prediction. *IEEE Trans. Neural Netw. Learn. Syst.*, 2019, vol. 30, no. 2, pp. 601–614.
DOI: <https://doi.org/10.1109/TNNLS.2018.2846646>
- [4] Huber P.J., Ronchetti E.M. Robust statistics. Wiley, 2009.
- [5] Fletcher R. Practical methods of optimization. Wiley, 2013.
- [6] Granger C.W.J., Andersen A.P. An introduction to bilinear time series models. Vandenhoeck & Ruprecht, 1978.
- [7] Brunner A.D., Hess G.D. Potential problems in estimating bilinear time-series models. *J. Econ. Dyn. Control*, 1995, vol. 19, no. 4, pp. 663–681.
- [8] Storn R., Price K. Differential evolution — a simple and efficient heuristic for global optimization over continuous spaces. *J. Glob. Optim.*, 1997, vol. 11, no. 4, pp. 341–359.
DOI: <https://doi.org/10.1023/A:1008202821328>
- [9] Price K.V., Storn R.M., Lampinen J.A. Differential evolution. Springer, 2005.
- [10] Liu D.C., Nocedal J. On the limited-memory BFGS method for large scale optimization. *Math. Program.*, 1989, vol. 45, pp. 503–528.
DOI: <https://doi.org/10.1007/BF01589116>
- [11] Riani M., Atkinson A.C., Corbellini A., et al. Robust regression with density power divergence: theory, comparisons, and data analysis. *Entropy*, 2020, vol. 22, no. 4, art. 399. DOI: <https://doi.org/10.3390/e22040399>
- [12] Zhu C., Byrd R.H., Lu P., et al. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.*, 1997, vol. 23, no. 4, pp. 550–560. DOI: <https://doi.org/10.1145/279232.279236>
- [13] Agostinelli C., Valdora M., Yohai V.J. Initial robust estimation in generalized linear models. *Comput. Stat. Data Anal.*, 2019, vol. 134, pp. 144–156.
DOI: <https://doi.org/10.1016/j.csda.2018.12.010>
- [14] Yao W., Li L. A new regression model: modal linear regression. *Scand. J. Stat.*, 2014, vol. 41, no. 3, pp. 656–671. DOI: <https://doi.org/10.1111/sjos.12054>
- [15] Brockwell P.J., Davis R.A. Time series: theory and methods. Springer, 1996.
- [16] Chen Y., Xie Y., Ma X., et al. Probabilistic forecasting with temporal convolutional neural network. *Neurocomputing*, 2020, vol. 399, pp. 491–501.
DOI: <https://doi.org/10.1016/j.neucom.2020.03.011>
- [17] Billard L. Time series models. In: *Biometrics*. Vol. 1. UNESCO, 1997.

- [18] Sesay S.A.O., Subba Rao T. Frequency-domain estimation of bilinear time series models. *J. Time Ser. Anal.*, 1992, vol. 13, no. 6, pp. 521–545.
DOI: <https://doi.org/10.1111/j.1467-9892.1992.tb00124.x>
- [19] Goryainov V.B., Goryainova E.R. Robust estimation in threshold autoregression. *Herald of the Bauman Moscow State Technical University, Series Natural Sciences*, 2017, no. 6 (75), pp. 19–30 (in Russ.).
DOI: <http://dx.doi.org/10.18698/1812-3368-2017-6-19-30>
- [20] Zhu D., Galbraith J.W. Modeling and forecasting expected shortfall with the generalized asymmetric Student-*t* and asymmetric exponential power distributions. *J. Empir. Finance*, 2011, vol. 18, no. 4, pp. 765–778.
DOI: <https://doi.org/10.1016/j.jempfin.2011.05.006>
- [21] Insolia L., Kenney A., Chiaromonte F., et al. Simultaneous feature selection and outlier detection with optimality guarantees. *Biometrics*, 2022, vol. 78, no. 4, pp. 1592–1603. DOI: <https://doi.org/10.1111/biom.13553>

Goryainov V.B. — Dr. Sc. (Phys.-Math.), Professor, Department of Mathematical Simulation, BMSTU (2-ya Baumanskaya ul. 5, str. 1, Moscow, 105005 Russian Federation).

Masyagin M.M. — Post-Graduate Student, Assistant, Department of Mathematical Simulation, BMSTU (2-ya Baumanskaya ul. 5, str. 1, Moscow, 105005 Russian Federation).

Semernya V.M. — Post-Graduate Student, Department of Mathematical Simulation, BMSTU (2-ya Baumanskaya ul. 5, str. 1, Moscow, 105005 Russian Federation).

Please cite this article in English as:

Goryainov V.B., Masyagin M.M., Semernya V.M. Analysis of least squares method efficiency with L-BFGS-B optimization for bilinear autoregression parameter estimation under various noise distributions. *Herald of the Bauman Moscow State Technical University, Series Natural Sciences*, 2026, no. 3 (126), pp. 4–23 (in Russ.).
EDN: VVLPJN